



SKRIPSI

**KLASIFIKASI AUDIO *DEEPPFAKE* DENGAN
EKSTRAKSI FITUR YAMNET DAN *DEEP
NEURAL NETWORK***

HAKAM DZAKWAN DIASH
NPM 22083010069

DOSEN PEMBIMBING

Dr. Ir. Mohammad Idhom, S.P., S.Kom., M.T.
Alfan Rizaldy Pratama, S.Tr.T., M.Tr.Kom.

**KEMENTERIAN PENDIDIKAN TINGGI, SAINS, DAN TEKNOLOGI
UNIVERSITAS PEMBANGUNAN NASIONAL VETERAN JAWA TIMUR
FAKULTAS ILMU KOMPUTER
PROGRAM STUDI SAINS DATA
SURABAYA
2026**

LEMBAR PENGESAHAN

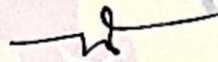
KLASIFIKASI AUDIO DEEPPAKE DENGAN EKSTRAKSI FITUR YAMNET DAN DEEP NEURAL NETWORK

Oleh:
HAKAM DZAKWAN DIASH
NPM. 22083010069

Telah dipertahankan di hadapan dan diterima oleh Tim Penguji Sidang Skripsi Program Studi Sains Data Fakultas Ilmu Komputer Universitas Pembangunan Nasional Veteran Jawa Timur pada Tanggal 23 Juni 2026:

Menyetujui,

Dr. Ir. Mohammad Idhom, S.P., S.Kom., M.T.
NIP. 19830310 202121 1 006


.....

(Pembimbing I)

Alfan Rizaldy Pratama, S.Tr.T., M.Tr.Kom.
NIP. 19990606 202406 1 001


.....

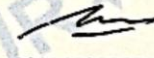
(Pembimbing II)

Kartika Maulida Hindravani, S.Kom., M.Kom.
NIP. 19920909 202203 2 009


.....

(Ketua Penguji)

Andri Fauzan Adziima, M. Si.
NIP. 19950512 202406 1 001


.....

(Penguji I)

Mengetahui,

Dekan Fakultas Ilmu Komputer




Prof. Dr. Ir. Novirina Hendrasarie, MT.
NIP. 19681126 199403 2 001

LEMBAR PERSETUJUAN

**KLASIFIKASI AUDIO *DEEFAKE* DENGAN EKSTRAKSI FITUR
YAMNET DAN *DEEP NEURAL NETWORK***

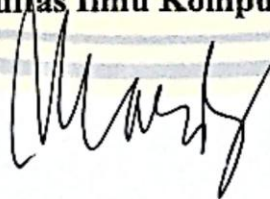
Oleh:

HAKAM DZAKWAN DIASH
NPM. 22083010069

Telah disetujui untuk mengikuti Ujian Skripsi

Menyetujui,

Plt Koordinator Program Studi Sains Data
Fakultas Ilmu Komputer



Dr. I Gede Susrama Mas Divasa, ST., MT.

NIP. 19700619 2021211 009

SURAT PERNYATAAN BEBAS PLAGIASI

Saya yang bertanda tangan di bawah ini:

Nama Mahasiswa : Hakam Dzakwan Diash
NPM : 22083010069
Program : Sarjana (S1)
Program Studi : Sains Data
Fakultas : Fakultas Ilmu Komputer

Menyatakan bahwa dalam dokumen ilmiah Skripsi ini tidak terdapat bagian dari karya ilmiah lain yang telah diajukan untuk memperoleh gelar akademik di suatu lembaga Pendidikan Tinggi, dan juga tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan oleh orang/lembaga lain, kecuali yang secara tertulis disitasi dalam dokumen ini dan disebutkan secara lengkap dalam daftar pustaka.

Dan saya menyatakan bahwa dokumen ilmiah ini bebas dari unsur-unsur plagiasi. Apabila di kemudian hari ditemukan indikasi plagiat pada Skripsi ini, saya bersedia menerima sanksi sesuai dengan peraturan perundang-undangan yang berlaku.

Demikian surat pernyataan ini saya buat dengan sesungguhnya tanpa ada paksaan dari siapapun juga dan untuk dipergunakan sebagaimana mestinya.



Surabaya, 3 Juli 2026
Yang Membuat Pernyataan,



HAKAM DZAKWAN DIASH
NPM. 22083010069

ABSTRAK

Nama Mahasiswa / NPM : Hakam Dzakwan Diash / 22083010069
Judul Skripsi : Klasifikasi Audio *Deepfake* dengan Ekstraksi Fitur
YAMNet dan *Deep Neural Network*
Dosen Pembimbing : 1. Dr. Ir. Mohammad Idhom, S.P., S.Kom., M.T.
2. Alfian Rizaldy Pratama, S.Tr.T., M.Tr.Kom.

Perkembangan teknologi kecerdasan buatan dalam sintesis suara telah meningkatkan kemampuan pembuatan audio *deepfake* yang sulit dibedakan dari suara asli sehingga berpotensi menimbulkan penyebaran informasi palsu, penipuan identitas, dan penyalahgunaan media digital. Penelitian ini bertujuan untuk menguji efektivitas kombinasi ekstraksi fitur YAMNet dan *Deep Neural Network* (DNN) dalam klasifikasi audio *deepfake*, menganalisis penerapan metode agregasi statistik terhadap *embedding* YAMNet, mengukur kemampuan generalisasi lintas *dataset*, serta mengimplementasikan model ke dalam prototipe berbasis web menggunakan Flask. Metode penelitian dilakukan dengan mengekstraksi *embedding* audio menggunakan YAMNet, kemudian menerapkan agregasi statistik seperti mean dan standar deviasi untuk menghasilkan representasi fitur yang lebih ringkas sebelum digunakan sebagai input model DNN. Hasil penelitian menunjukkan bahwa metode agregasi statistik berhasil mengurangi kompleksitas data sekaligus mempertahankan performa klasifikasi yang tinggi dengan rentang akurasi 97,58% hingga 99,24%, di mana model terbaik diperoleh pada metode Mean + Std Model 1 dengan akurasi sebesar 99,24%. Pada pengujian lintas *dataset*, model awal menunjukkan performa yang rendah dengan akurasi primer sekitar 61,00% hingga 62,37%, namun setelah dilakukan *fine-tuning* menggunakan pendekatan *freeze layer*, model mampu mencapai keseimbangan performa yang lebih stabil dengan akurasi primer sebesar 92,81% dan akurasi sekunder sebesar 91,41%. Selain itu, model berhasil diimplementasikan ke dalam sistem berbasis Flask bernama VocalSentry yang menyediakan fitur deteksi audio *deepfake* secara interaktif melalui antarmuka yang mudah digunakan. Berdasarkan hasil penelitian, kombinasi ekstraksi fitur YAMNet dan DNN terbukti efektif dalam klasifikasi audio *deepfake* serta mampu mempertahankan performa yang baik setelah proses *fine-tuning* lintas *dataset* dilakukan.

Kata Kunci: *Audio Deepfake, Klasifikasi Audio, Ekstraksi Fitur, YAMNet, Deep Neural Network*

ABSTRACT

Student Name / NPM : Hakam Dzakwan Diash / 22083010069
Undergraduate thesis title : *Audio Deepfake Classification Based on YAMNet Feature Extraction and Deep Neural Network*
Advisors : 1. Dr. Ir. Mohammad Idhom, S.P., S.Kom., M.T.
2. Alfian Rizaldy Pratama, S.Tr.T., M.Tr.Kom.

The advancement of artificial intelligence technology in voice synthesis has increased the capability of generating deepfake audio that is difficult to distinguish from genuine human speech, potentially leading to misinformation, identity fraud, and misuse of digital media. This study aims to evaluate the effectiveness of combining YAMNet feature extraction and Deep Neural Network (DNN) for deepfake audio classification, analyze the application of statistical aggregation methods on YAMNet embeddings, measure cross-dataset generalization capability, and implement the trained model into a Flask-based web prototype. The research method was conducted by extracting audio embeddings using YAMNet, followed by applying statistical aggregation methods such as mean and standard deviation to generate more compact feature representations before being used as input for the DNN classification model. The results show that statistical aggregation methods successfully reduced data complexity while maintaining high classification performance with accuracy ranging from 97.58% to 99.24%, where the best model was achieved by the Mean + Std Model 1 method with an accuracy of 99.24%. In cross-dataset testing, the initial model demonstrated low performance with primary dataset accuracy ranging from 61.00% to 62.37%, however, after fine-tuning using the freeze layer approach, the model achieved a more balanced performance with primary dataset accuracy of 92.81% and secondary dataset accuracy of 91.41%. Furthermore, the trained model was successfully implemented into a Flask-based system called VocalSentry, which provides interactive deepfake audio detection features through an informative and user-friendly interface. Based on the results, the combination of YAMNet feature extraction and DNN proved effective for deepfake audio classification and was able to maintain good performance after cross-dataset fine-tuning.

Keywords: *Deepfake Audio, Audio Classification, Feature Extraction, YAMNet, Deep Neural Network*

KATA PENGANTAR

Puji syukur kehadirat Allah SWT atas segala rahmat, hidayah dan karunia-Nya kepada penulis sehingga skripsi dengan judul **“Klasifikasi Audio *Deepfake* dengan Ekstraksi Fitur YAMNet dan *Deep Neural Network*”** dapat terselesaikan dengan baik.

Penulis mengucapkan terima kasih kepada Bapak Dr. Ir. Mohammad Idhom, S.P., S.Kom., M.T. selaku Dosen Pembimbing I, Bapak Alfian Rizaldy Pratama, S.Tr.T., M.Tr.Kom. selaku Dosen Pembimbing II, serta Bapak Dr. Eng. Ir. Dwi Arman Prasetya, ST., MT., IPU., ASEAN Eng. selaku Dosen Pembimbing sebelumnya yang telah meluangkan waktu, tenaga, dan pikirannya untuk memberikan bimbingan, arahan, saran, motivasi, serta masukan yang sangat berharga kepada penulis selama proses penyusunan skripsi ini. Penulis juga banyak menerima bantuan dari berbagai pihak, baik berupa dukungan moril, spiritual, maupun materiil. Oleh karena itu, penulis mengucapkan terima kasih kepada:

1. Ibu Prof. Dr. Ir. Novirina Hendrasarie, MT selaku Dekan Fakultas Ilmu Komputer Universitas Pembangunan Nasional Veteran Jawa Timur.
2. Seluruh dosen dan staff prodi Sains Data yang selalu mengusahakan dan menyediakan lingkungan belajar yang mendukung bagi seluruh mahasiswa.
3. Kedua orang tua dan seluruh keluarga yang senantiasa memberikan doa, dukungan, motivasi, serta semangat selama proses perkuliahan.
4. Ika yang selalu memberikan dukungan, motivasi, serta menjadi tempat berbagi cerita dan semangat dalam menghadapi berbagai tantangan selama masa studi.
5. Teman-teman dekat penulis, yaitu Akmal, Amirah, Raihan, Zufar, Yoga dan Vannesa, yang senantiasa memberikan bantuan, dukungan, serta berbagai masukan yang berharga selama menjalani proses perkuliahan.
6. Seluruh mahasiswa Angkatan 2022 Program Studi Sains Data Universitas Pembangunan Nasional “Veteran” Jawa Timur yang telah menjadi rekan belajar, berdiskusi, dan bertumbuh bersama selama menempuh pendidikan di bangku perkuliahan.

7. Penulis sendiri, yang telah memilih untuk terus berjuang dan berkembang di setiap proses yang dijalani. Terima kasih karena tetap bertahan, tetap belajar, dan tidak menyerah meskipun banyak hal yang tidak berjalan sesuai dengan keinginan. Semoga setiap usaha dan pengorbanan yang telah dilakukan menjadi langkah menuju masa depan yang lebih baik.

Penulis menyadari bahwa skripsi ini masih jauh dari kata sempurna dan memiliki banyak kekurangan. Oleh karena itu, penulis sangat terbuka terhadap kritik dan saran konstruktif dari berbagai pihak demi penyempurnaan karya ini. Dengan segala keterbatasan yang ada, penulis memiliki harapan agar skripsi ini dapat memberi manfaat bagi pembaca dan penulis.

Surabaya, 3 Juli 2026

Penulis

DAFTAR ISI

HALAMAN SAMPUL.....	i
LEMBAR PENGESAHAN	ii
LEMBAR PERSETUJUAN	iii
SURAT PERNYATAAN BEBAS PLAGIASI.....	iv
ABSTRAK	v
<i>ABSTRACT</i>	vi
KATA PENGANTAR.....	vii
DAFTAR ISI.....	ix
DAFTAR GAMBAR.....	xii
DAFTAR TABEL	xiv
DAFTAR LAMPIRAN	xv
DAFTAR NOTASI.....	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
BAB II TINJAUAN PUSTAKA.....	6
2.1. Penelitian Terdahulu	6
2.2. Kerangka Teori.....	12
2.2.1. <i>Audio Deepfake</i>	12
2.2.2. <i>Waveform</i>	13

2.2.3.	<i>Log-Mel Spectrogram</i>	14
2.2.4.	<i>Deep Learning</i>	15
2.2.5.	<i>Depthwise Separable Convolution</i>	20
2.2.6.	<i>Yet another Audio Mobilenet Network (YAMNet)</i>	21
2.2.7.	<i>Deep Neural Network</i>	23
2.2.8.	Metrik Evaluasi	24
2.2.9.	Flask	26
BAB III METODOLOGI PENELITIAN.....		28
3.1.	Variabel Penelitian dan Sumber Data.....	28
3.1.1.	Data Primer.....	28
3.1.2.	Data Sekunder	30
3.2.	Langkah Analisis	32
3.2.1.	Pengumpulan dan Pelabelan Data	33
3.2.2.	<i>Embedding</i> dengan YAMNet	34
3.2.3.	<i>Splitting Data</i>	36
3.2.4.	Perancangan Arsitektur Jaringan Klasifikasi.....	37
3.2.5.	<i>Training Model</i>	37
3.2.6.	<i>Model Evaluation</i>	38
3.2.7.	<i>Hyperparameter Tuning</i>	38
3.3.	Desain Sistem	39
BAB IV HASIL DAN PEMBAHASAN		42
4.1.	Hasil Penelitian.....	42
4.1.1.	Pengumpulan Data.....	42
4.1.2.	<i>Data Preprocessing</i>	42
4.1.3.	Mendapatkan Vektor <i>Embedding</i>	45
4.1.4.	Agregasi Statistik.....	47
4.1.5.	<i>Splitting Data</i>	57
4.1.6.	Perancangan Arsitektur Jaringan Klasifikasi.....	58
4.1.7.	<i>Training Model</i>	65

4.1.8.	Evaluasi Awal pada Data Uji Sekunder	66
4.1.9.	Pengujian Generalisasi pada Data Primer	70
4.1.10.	<i>Fine-Tuning</i> Model Terpilih pada <i>Dataset</i> Primer.....	72
4.1.11.	Evaluasi Akhir dan Pemilihan Model	83
4.2.	Tampilan Antar Muka (<i>Graphical User Interface</i>).....	85
4.2.1.	Halaman Awal.....	86
4.2.2.	Hasil Prediksi	87
4.2.3.	Interpretasi Prediksi	88
4.2.4.	Fitur Tanya Jawab	88
BAB V PENUTUP		90
5.1.	Kesimpulan	90
5.2.	Saran Pengembangan	91
DAFTAR PUSTAKA		93
LAMPIRAN.....		97

DAFTAR GAMBAR

Gambar 2. 1 <i>Convolutional Layer</i> [23].....	16
Gambar 2. 2 Ilustrasi jaringan saraf tiruan standar (A) dan setelah menerapkan <i>dropout</i> (B) [26]	18
Gambar 2. 3 <i>Depthwise Separable Convolution</i> [33]	21
Gambar 2. 4 Arsitektur YAMNet [8].....	22
Gambar 2. 5 Ilustrasi <i>Confusion Matrix</i> [40]	25
Gambar 3. 1 Perbandingan <i>Waveform</i> Kedua Label.....	32
Gambar 3. 2 Alur Kerja Penelitian.....	33
Gambar 3. 3 Sub alur <i>Embedding</i> dengan YAMNet	34
Gambar 3. 4 Rancangan Desain Sistem	40
Gambar 3. 5 Rancangan Antarmuka Pengguna Aplikasi VocalSentry.....	41
Gambar 4. 1 Arsitektur Model DNN dengan Dimensi <i>Input</i> 1024.....	64
Gambar 4. 2 Arsitektur Model DNN dengan Dimensi <i>Input</i> 2048.....	64
Gambar 4. 3 Arsitektur Model DNN dengan Dimensi <i>Input</i> 4096.....	64
Gambar 4. 4 <i>Confusion Matrix</i> Hasil Pengujian Model Mean + Std 1 Pada <i>Dataset</i> Primer.....	70
Gambar 4. 5 <i>Confusion Matrix</i> Hasil Pengujian Model Mean 1 Pada <i>Dataset</i> Primer	71
Gambar 4. 6 <i>Confusion Matrix</i> Hasil Pengujian Model Mean 2 Pada <i>Dataset</i> Primer	71
Gambar 4. 7 Hasil <i>Confusion Matrix</i> Pengujian Model Mean + Std 1 Setelah <i>Fine-Tuning</i> Tanpa <i>Freeze Layer</i> pada Data Primer dan Sekunder (Skenario 20% Data Latih dan 80% Data Uji).....	74
Gambar 4. 8 Hasil <i>Confusion Matrix</i> Pengujian Model Mean 1 Setelah <i>Fine-Tuning</i> Tanpa <i>Freeze Layer</i> pada Data Primer dan Sekunder (Skenario 20% Data Latih dan 80% Data Uji).....	74
Gambar 4. 9 Hasil <i>Confusion Matrix</i> Pengujian Model Mean 2 Setelah <i>Fine-Tuning</i> Tanpa <i>Freeze Layer</i> pada Data Primer dan Sekunder (Skenario 20% Data Latih dan 80% Data Uji).....	75

Gambar 4. 10 Hasil <i>Confusion Matrix</i> Pengujian Model Mean + Std 1 Setelah <i>Fine-Tuning</i> Tanpa <i>Freeze Layer</i> pada Data Primer dan Sekunder (Skenario 40% Data Latih dan 60% Data Uji)	76
Gambar 4. 11 Hasil <i>Confusion Matrix</i> Pengujian Model Mean 1 Setelah <i>Fine-Tuning</i> Tanpa <i>Freeze Layer</i> pada Data Primer dan Sekunder (Skenario 40% Data Latih dan 60% Data Uji)	77
Gambar 4. 12 Hasil <i>Confusion Matrix</i> Pengujian Model Mean + Std 1 Setelah <i>Fine-Tuning</i> Tanpa <i>Freeze Layer</i> pada Data Primer dan Sekunder (Skenario 40% Data Latih dan 60% Data Uji)	77
Gambar 4. 13 Hasil <i>Confusion Matrix</i> Pengujian Model Mean + Std 1 Setelah <i>Fine-Tuning</i> dengan <i>Freeze Layer</i> pada Data Primer dan Sekunder (Skenario 20% Data Latih dan 80% Data Uji)	79
Gambar 4. 14 Hasil <i>Confusion Matrix</i> Pengujian Model Mean 1 Setelah <i>Fine-Tuning</i> dengan <i>Freeze Layer</i> pada Data Primer dan Sekunder (Skenario 20% Data Latih dan 80% Data Uji)	80
Gambar 4. 15 Hasil <i>Confusion Matrix</i> Pengujian Model Mean 2 Setelah <i>Fine-Tuning</i> dengan <i>Freeze Layer</i> pada Data Primer dan Sekunder (Skenario 20% Data Latih dan 80% Data Uji)	80
Gambar 4. 16 Hasil <i>Confusion Matrix</i> Pengujian Model Mean + Std 1 Setelah <i>Fine-Tuning</i> dengan <i>Freeze Layer</i> pada Data Primer dan Sekunder (Skenario 40% Data Latih dan 60% Data Uji)	81
Gambar 4. 17 Hasil <i>Confusion Matrix</i> Pengujian Model Mean 1 Setelah <i>Fine-Tuning</i> dengan <i>Freeze Layer</i> pada Data Primer dan Sekunder (Skenario 40% Data Latih dan 60% Data Uji)	82
Gambar 4. 18 Hasil <i>Confusion Matrix</i> Pengujian Model Mean 2 Setelah <i>Fine-Tuning</i> dengan <i>Freeze Layer</i> pada Data Primer dan Sekunder (Skenario 40% Data Latih dan 60% Data Uji)	82
Gambar 4. 19 Tampilan Awal GUI	86
Gambar 4. 20 Tampilan Hasil Prediksi Klasifikasi Audio <i>Deepfake</i>	87
Gambar 4. 21 Tampilan Interpretasi Hasil Prediksi Klasifikasi Audio <i>Deepfake</i>	88
Gambar 4. 22 Tampilan Fitur Tanya Jawab AI	89

DAFTAR TABEL

Tabel 2.1. Penelitian Terdahulu	6
Tabel 3. 1 Daftar Kata dan Kalimat yang Digunakan dalam Perekaman Data Primer	29
Tabel 3. 2 Skenario Eksperimen dengan Data Primer	36
Tabel 3. 3 Skenario <i>Hyperparameter Tuning</i>	38
Tabel 4. 1 Data Hasil Pengujian Waktu Pemrosesan Metode Ekstraksi Fitur	56
Tabel 4. 2 Kinerja Klasifikasi Model DNN pada Fitur <i>Embedding</i> Berdimensi 1024	66
Tabel 4. 3 Kinerja Klasifikasi Model DNN pada Fitur <i>Embedding</i> Berdimensi 2048	67
Tabel 4. 4 Kinerja Klasifikasi Model DNN pada Fitur <i>Embedding</i> Berdimensi 4096	68
Tabel 4. 5 Tiga Model Terbaik Berdasarkan Tingkat Akurasi pada Data Uji Sekunder.....	69
Tabel 4. 6 Hasil Evaluasi Akurasi <i>Fine-Tuning</i> Tanpa <i>Freeze Layer</i> pada Data Primer dan Sekunder (Skenario 20% Data Latih dan 80% Data Uji)	73
Tabel 4. 7 Hasil Evaluasi Akurasi <i>Fine-Tuning</i> Tanpa <i>Freeze Layer</i> pada Data Primer dan Sekunder (Skenario 40% Data Latih dan 60% Data Uji)	76
Tabel 4. 8 Hasil Evaluasi Akurasi <i>Fine-Tuning</i> dengan <i>Freeze Layer</i> pada Data Primer dan Sekunder (Skenario 20% Data Latih dan 80% Data Uji)	79
Tabel 4. 9 Hasil Evaluasi Akurasi <i>Fine-Tuning</i> dengan <i>Freeze Layer</i> pada Data Primer dan Sekunder (Skenario 40% Data Latih dan 60% Data Uji)	81
Tabel 4. 10 Hasil Evaluasi Performa Model Terbaik pada Data Uji Primer.....	83
Tabel 4. 11 Hasil Evaluasi Performa Model Terbaik pada Data Uji Sekunder	84

DAFTAR LAMPIRAN

Lampiran 1. <i>Letter of Acceptance</i> (LoA) Jurnal	97
Lampiran 2. Kode Script Penelitian	98
Lampiran 3. Kode Script GUI.....	99
Lampiran 4. Dataset	100

DAFTAR NOTASI

$x(n)$	:	Sinyal audio dalam domain waktu
$X(m,k)$	:	Koefisien STFT
$LMS(k)$	:	Log-Mel Spectrogram
W	:	Matriks bobot (<i>Weights</i>)
b	:	Vektor bias
$f(x)$	:	Fungsi aktivasi
y_i	:	Label asli
$P(x_i)$	:	Probabilitas hasil prediksi
α	:	<i>Learning rate</i>
g_t	:	Gradien dari <i>loss function</i>
TP	:	<i>True positive</i>
TN	:	<i>True negative</i>
FP	:	<i>False positive</i>
FN	:	<i>False negative</i>
$F1 - Score$	:	Metrik evaluasi F1-Score