

IMPLEMENTASI ARSITEKTUR CAPSNET DAN LSTM UNTUK KLASIFIKASI AUDIO DEEPPAKE BAHASA INDONESIA

SKRIPSI

Oleh:

VIVALDI GABRIEL MANURUNG

NIM. 221112090

JESSEN ADRYAN SEMBIRING

NIM. 221112897

CRISTIAN JOSEPHINE MANALU

NIM. 221112984



**PROGRAM STUDI S-1 TEKNIK INFORMATIKA
FAKULTAS INFORMATIKA
UNIVERSITAS MIKROSKIL
MEDAN
2026**

IMPLEMENTATION OF CAPSNET AND LSTM ARCHITECTURES FOR INDONESIAN DEEPFAKE AUDIO CLASSIFICATION

UNDERGRADUATE THESIS

By:

VIVALDI GABRIEL MANURUNG

NIM. 221112090

JESSEN ADRYAN SEMBIRING

NIM. 221112897

CRISTIAN JOSEPHINE MANALU

NIM. 221112984



**UNDERGRADUATE PROGRAM OF COMPUTER SCIENCE
FACULTY OF INFORMATICS
UNIVERSITAS MIKROSKIL
MEDAN
2026**

LEMBARAN PENGESAHAN

IMPLEMENTASI ARSITEKTUR CAPSNET DAN LSTM UNTUK KLASIFIKASI AUDIO DEEPPAKE BAHASA INDONESIA

SKRIPSI

Diajukan untuk Melengkapi Persyaratan Guna
Mendapatkan Gelar Sarjana
Program Studi S-I Teknik Informatika

Oleh:

VIVALDI GABRIEL MANURUNG
NIM. 221112090
JESSEN ADRYAN SEMBIRING
NIM. 221112897
CRISTIAN JOSEPHINE MANALU
NIM. 221112984

Disetujui Oleh:

Dosen Pembimbing I,



Sunaryo Winardi, S.Kom., M.T.

Dosen Pembimbing II,



Arwin Halim, S.Kom., M.Kom.

Medan, 14 Juli 2026

Diketahui dan Disahkan Oleh:

Ketua Program Studi
S-I Teknik Informatika



Charles Juliandy, S.Kom., M.Kom.

HALAMAN PERNYATAAN

Saya yang membuat pernyataan ini adalah mahasiswa Program Studi S-1 Teknik Informatika Universitas Mikroskil Medan dengan identitas mahasiswa sebagai berikut:

Nama : Vivaldi Gabriel Manurung

NIM : 221112090

Saya telah melaksanakan penelitian dan penulisan Skripsi dengan judul dan tempat penelitian sebagai berikut:

Judul Skripsi : Implementasi Arsitektur CapsNet dan LSTM untuk Klasifikasi Audio Deepfake Bahasa Indonesia

Tempat Penelitian : -

Alamat Tempat Penelitian : -

No. Telp. Tempat Penelitian : -

Sehubungan dengan Skripsi tersebut, dengan ini saya menyatakan dengan sebenar-benarnya bahwa penelitian dan penulisan Skripsi tersebut merupakan hasil karya saya sendiri (tidak menyuruh orang lain yang mengerjakannya) dan semua sumber, baik yang dikutip maupun dirujuk, telah saya nyatakan dengan benar. Bila di kemudian hari ternyata terbukti bahwa bukan saya yang mengerjakannya (membuatnya), maka saya bersedia dikenakan sanksi yang telah ditetapkan oleh Universitas Mikroskil Medan, yakni pencabutan ijazah yang telah saya terima dan ijazah tersebut dinyatakan tidak sah.

Selain itu, demi pengembangan ilmu pengetahuan, saya menyetujui untuk memberikan kepada Universitas Mikroskil Medan Hak Bebas Royalti Non-eksklusif (Non-exclusive Royalty Free Right) atas Skripsi saya beserta perangkat yang ada (jika diperlukan). Dengan hak ini, Universitas Mikroskil Medan berhak menyimpan, mengalihmedia/formatkan, mengelola dalam bentuk pangkalan data (database), merawat, dan mempublikasikan Skripsi saya, secara keseluruhan atau hanya sebagian atau hanya ringkasannya saja dalam bentuk format tercetak dan/atau elektronik, selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik hak cipta. Menyatakan juga bahwa saya akan mempertahankan hak eksklusif saya untuk menggunakan seluruh atau sebagian isi Skripsi saya guna pengembangan karya di masa depan, misalnya dalam bentuk artikel, buku, ataupun perangkat lunak/sistem informasi.

Demikian pernyataan ini saya perbuat dengan sungguh-sungguh, dalam keadaan sadar dan tanpa ada tekanan dari pihak manapun.

Medan, 01 Juli 2026

Saya yang membuat pernyataan,



Vivaldi Gabriel Manurung

HALAMAN PERNYATAAN

Saya yang membuat pernyataan ini adalah mahasiswa Program Studi S-1 Teknik Informatika Universitas Mikroskil Medan dengan identitas mahasiswa sebagai berikut:

Nama : Jessen Adryan Sembiring

NIM : 221112897

Saya telah melaksanakan penelitian dan penulisan Skripsi dengan judul dan tempat penelitian sebagai berikut:

Judul Skripsi : Implementasi Arsitektur CapsNet dan LSTM untuk Klasifikasi Audio Deepfake Bahasa Indonesia

Tempat Penelitian : -

Alamat Tempat Penelitian : -

No. Telp. Tempat Penelitian : -

Sehubungan dengan Skripsi tersebut, dengan ini saya menyatakan dengan sebenar-benarnya bahwa penelitian dan penulisan Skripsi tersebut merupakan hasil karya saya sendiri (tidak menyuruh orang lain yang mengerjakannya) dan semua sumber, baik yang dikutip maupun dirujuk, telah saya nyatakan dengan benar. Bila di kemudian hari ternyata terbukti bahwa bukan saya yang mengerjakannya (membuatnya), maka saya bersedia dikenakan sanksi yang telah ditetapkan oleh Universitas Mikroskil Medan, yakni pencabutan ijazah yang telah saya terima dan ijazah tersebut dinyatakan tidak sah.

Selain itu, demi pengembangan ilmu pengetahuan, saya menyetujui untuk memberikan kepada Universitas Mikroskil Medan Hak Bebas Royalti Non-eksklusif (Non-exclusive Royalty Free Right) atas Skripsi saya beserta perangkat yang ada (jika diperlukan). Dengan hak ini, Universitas Mikroskil Medan berhak menyimpan, mengalihmedia/formatkan, mengelola dalam bentuk pangkalan data (database), merawat, dan mempublikasikan Skripsi saya, secara keseluruhan atau hanya sebagian atau hanya ringkasannya saja dalam bentuk format tercetak dan/atau elektronik, selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik hak cipta. Menyatakan juga bahwa saya akan mempertahankan hak eksklusif saya untuk menggunakan seluruh atau sebagian isi Skripsi saya guna pengembangan karya di masa depan, misalnya dalam bentuk artikel, buku, ataupun perangkat lunak/sistem informasi.

Demikian pernyataan ini saya perbuat dengan sungguh-sungguh, dalam keadaan sadar dan tanpa ada tekanan dari pihak manapun.

Medan, 01 Juli 2026

Saya yang membuat pernyataan,



Jessen Adryan Sembiring

HALAMAN PERNYATAAN

Saya yang membuat pernyataan ini adalah mahasiswa Program Studi S-1 Teknik Informatika Universitas Mikroskil Medan dengan identitas mahasiswa sebagai berikut:

Nama : Cristian Josephine Manalu
NIM : 221112984

Saya telah melaksanakan penelitian dan penulisan Skripsi dengan judul dan tempat penelitian sebagai berikut:

Judul Skripsi : Implementasi Arsitektur CapsNet dan LSTM untuk Klasifikasi Audio Deepfake Bahasa Indonesia

Tempat Penelitian : -

Alamat Tempat Penelitian : -

No. Telp. Tempat Penelitian : -

Sehubungan dengan Skripsi tersebut, dengan ini saya menyatakan dengan sebenar-benarnya bahwa penelitian dan penulisan Skripsi tersebut merupakan hasil karya saya sendiri (tidak menyuruh orang lain yang mengerjakannya) dan semua sumber, baik yang dikutip maupun dirujuk, telah saya nyatakan dengan benar. Bila di kemudian hari ternyata terbukti bahwa bukan saya yang mengerjakannya (membuatnya), maka saya bersedia dikenakan sanksi yang telah ditetapkan oleh Universitas Mikroskil Medan, yakni pencabutan ijazah yang telah saya terima dan ijazah tersebut dinyatakan tidak sah.

Selain itu, demi pengembangan ilmu pengetahuan, saya menyetujui untuk memberikan kepada Universitas Mikroskil Medan Hak Bebas Royalti Non-eksklusif (Non-exclusive Royalty Free Right) atas Skripsi saya beserta perangkat yang ada (jika diperlukan). Dengan hak ini, Universitas Mikroskil Medan berhak menyimpan, mengalihmedia/formatkan, mengelola dalam bentuk pangkalan data (database), merawat, dan mempublikasikan Skripsi saya, secara keseluruhan atau hanya sebagian atau hanya ringkasannya saja dalam bentuk format tercetak dan/atau elektronik, selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik hak cipta. Menyatakan juga bahwa saya akan mempertahankan hak eksklusif saya untuk menggunakan seluruh atau sebagian isi Skripsi saya guna pengembangan karya di masa depan, misalnya dalam bentuk artikel, buku, ataupun perangkat lunak/sistem informasi.

Demikian pernyataan ini saya perbuat dengan sungguh-sungguh, dalam keadaan sadar dan tanpa ada tekanan dari pihak manapun.

Medan, 01 Juli 2026

Saya yang membuat pernyataan,

Cristian Josephine Manalu

IMPLEMENTASI ARSITEKTUR CAPSNET DAN LSTM UNTUK KLASIFIKASI AUDIO DEEPPAKE BAHASA INDONESIA

Abstrak

Audio deepfake berbahasa Indonesia menjadi ancaman keamanan siber yang dapat dimanfaatkan untuk penipuan finansial dan penyebaran informasi palsu. Penelitian ini mengimplementasikan arsitektur hybrid CapsNet-LSTM untuk mengklasifikasikan audio berbahasa Indonesia ke dalam kelas asli dan palsu. Dataset yang digunakan bersumber dari Speech Dataset of Human and AI-Generated Voices berisi 42 file audio dari satu penutur, terdiri atas 21 rekaman suara asli dan 21 suara hasil voice cloning. Seluruh audio disegmentasi menjadi potongan 4 detik tanpa overlap dan disiapkan dalam dua varian, yaitu Dataset A tanpa pembuangan keheningan awal dan Dataset B dengan pembuangan keheningan awal. Fitur diekstraksi menggunakan Mel-Frequency Cepstral Coefficients (MFCC) sebagai input model. Pada dataset yang digunakan, model terbaik Dataset A mencapai F1-Score 95,76% dan model terbaik Dataset B mencapai F1-Score 97,38%, keduanya dengan batch size 16. Pembuangan keheningan awal pada Dataset B meningkatkan performa klasifikasi, dan validasi K-Fold Cross-Validation menghasilkan F1-Score rata-rata di atas 98%. Seluruh data berasal dari satu penutur, sehingga hasil ini merepresentasikan konsistensi model pada karakteristik suara penutur tersebut dan belum menggambarkan generalisasi terhadap penutur lain.

Kata kunci: *audio deepfake, Capsule Network, Long Short-Term Memory, MFCC*

Abstract

Indonesian-language audio deepfakes have become a cybersecurity threat that can be exploited for financial fraud and the spread of false information. This study implements a hybrid CapsNet-LSTM architecture to classify Indonesian-language audio into genuine and fake classes. The dataset used is drawn from the Speech Dataset of Human and AI-Generated Voices, comprising 42 audio files from a single speaker, consisting of 21 genuine voice recordings and 21 voice-cloned samples. All audio was segmented into 4-second clips without overlap and prepared in two variants: Dataset A without initial silence removal and Dataset B with the initial silence removed. Features were extracted using Mel-Frequency Cepstral Coefficients (MFCC) as the model input. On the dataset used, the best Dataset A model reached a 95.76% F1-Score and the best Dataset B model reached a 97.38% F1-Score, both with a batch size of 16. Removing the initial silence in Dataset B improved classification performance, and K-Fold Cross-Validation produced an average F1-Score above 98%. All data originates from a single speaker, so these results represent the model's consistency on that speaker's voice characteristics and do not yet reflect generalization to other speakers.

Keywords: *audio deepfake, Capsule Network, Long Short-Term Memory, MFCC,*

KATA PENGANTAR

Puji dan syukur penulis panjatkan ke hadirat Tuhan Yang Maha Esa atas berkat, rahmat dan karunia-nya yang senantiasa menyertai penulis sehingga dapat menyelesaikan skripsi yang berjudul “IMPLEMENTASI ARSITEKTUR CAPSNET DAN LSTM UNTUK KLASIFIKASI AUDIO DEEPPAKE BAHASA INDONESIA” dengan baik dan tepat waktu.

Skripsi ini disusun sebagai salah satu syarat untuk menyelesaikan Pendidikan Strata 1 pada Program Studi S-1 Teknik Informatika di Universitas Mikroskil Medan. Melalui penelitian ini, penulis dapat memperoleh manfaat dan wawasan baik bagi penulis maupun pembaca.

Selama proses penyusunan skripsi ini, penulis menerima banyak dukungan, bimbingan dan bantuan dari berbagai pihak. Penulis ingin menyampaikan terima kasih sebesar-besarnya kepada:

1. Bapak Sunaryo Winardi, S.Kom., M.T., selaku Dosen Pembimbing I dan juga Dekan Fakultas Informatika Universitas Mikroskil Medan.
2. Bapak Arwin Halim, S.Kom., M.Kom., selaku Dosen Pembimbing II.
3. Bapak Hardy, S.Kom., M.Sc., Ph.D., selaku Rektor Universitas Mikroskil Medan.
4. Bapak Carles Juliandy, S.Kom., M.Kom., selaku Ketua Program Studi S-1 Teknik Informatika Fakultas Informatika Universitas Mikroskil Medan.
5. Ungkapan rasa syukur dan terima kasih mendalam atas dukungan moral, materi, dan doa selama proses penyusunan skripsi kepada Bapak Eben Halasan Manurung dan Ibu Rumondang Siregar, selaku orang tua dari Vivaldi Gabriel Manurung.
6. Kepada kedua orang tua dari Jessen Adryan Sembiring, yaitu Bapak Rukun Sembiring dan Ibu Diana Ginting, terima kasih atas segala kasih sayang, doa, nasehat, serta atas kesabaran yang luar biasa dalam setiap langkah hidup penulis. Penulis berharap dapat menjadi anak yang membanggakan.
7. Terima kasih kepada Bapak Jonear Manalu dan Ibu Mawarni Lumban Batu selaku orang tua dari Cristian Josephine Manalu atas segala doa dan dukungan yang diberikan.

Penulis menyadari bahwa skripsi ini masih memiliki kekurangan dan jauh dari kata sempurna, sehingga penulis mengharapkan kritik dan saran yang bersifat membangun dari pembaca.

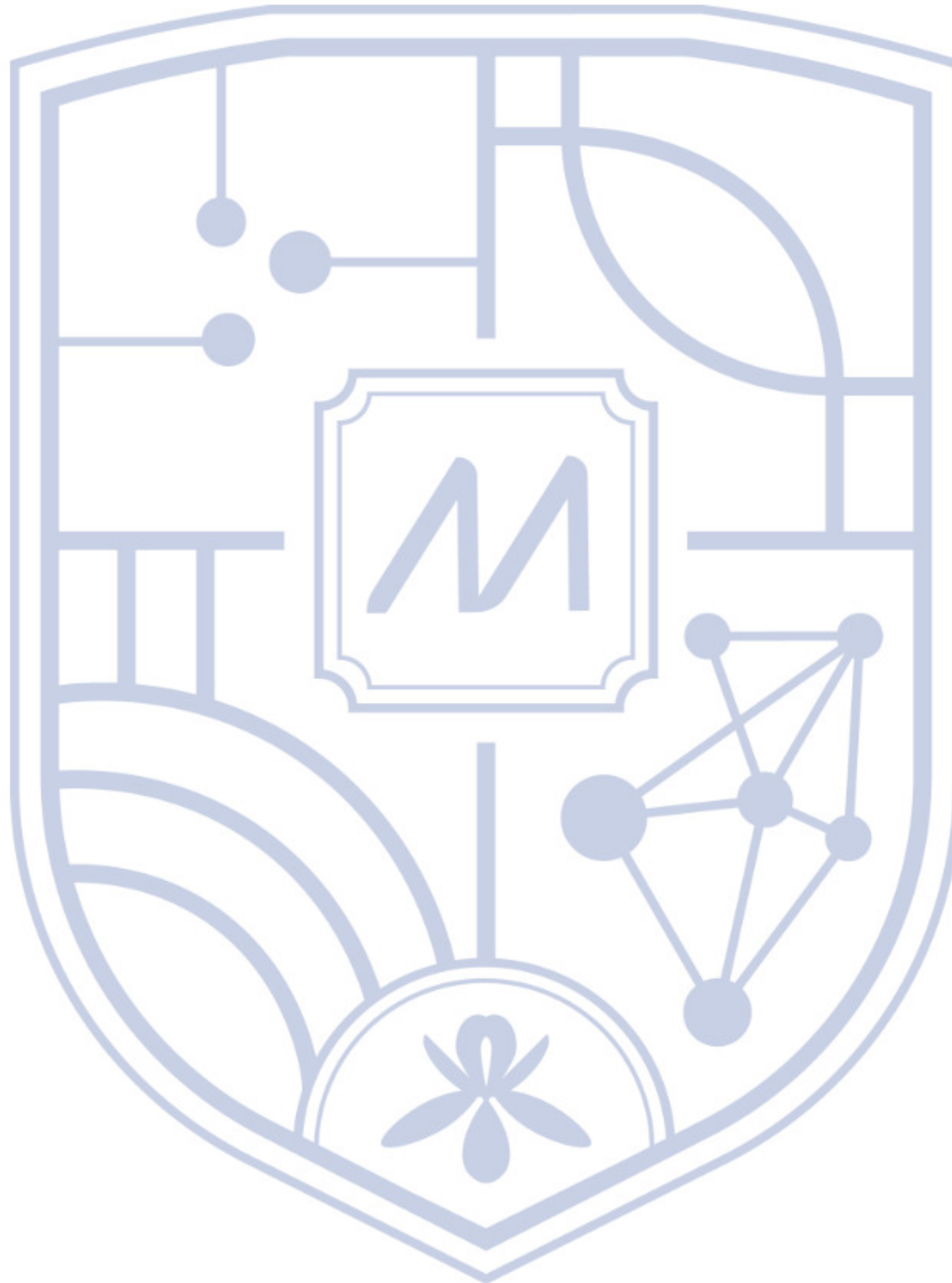
Medan, 01 Juli 2026

Penulis,

Vivaldi Gabriel Manurung

Jessen Adryan Sembiring

Cristian Josephine Manalu



DAFTAR ISI

Abstrak	i
KATA PENGANTAR	ii
DAFTAR ISI	iv
DAFTAR GAMBAR	vii
DAFTAR TABEL	viii
DAFTAR LAMPIRAN	x
BAB I PENDAHULUAN	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah.....	2
1.3 Tujuan.....	2
1.4 Manfaat.....	3
1.5 Ruang Lingkup	3
BAB II KAJIAN LITERATUR.....	5
2.1 Suara	5
2.2 Sinyal Audio Digital.....	6
2.2.1 <i>Sampling</i> dan Teorema Nyquist	6
2.2.2 <i>Bit Depth</i>	7
2.3 <i>Deepfake</i>	7
2.3.1 Definisi dan Kategori <i>Deepfake</i>	7
2.3.2 <i>Voice Cloning</i>	8
2.4 Segmentasi dan Pemotongan	9
2.5 <i>Mel-Frequency Cepstral Coefficients</i> (MFCC)	9
2.5.1 Prinsip Kerja dan Skala Mel.....	10
2.5.2 Tahapan Pra-pemrosesan Sinyal	10
2.5.3 Tahapan Transformasi Inti	12
2.5.4 Normalisasi Fitur	13
2.6 <i>Deep Learning</i>	14
2.6.1 Artificial Neural Network (ANN)	14
2.6.2 Fungsi Aktivasi.....	15
2.6.3 <i>Optimization Algorithm</i>	16
2.6.4 <i>Binary Cross-Entropy</i>	16
2.6.5 <i>Dropout</i>	17

2.6.6 Hyperparameter	17
2.6.7 Early Stopping.....	18
2.7 Capsule Network.....	18
2.7.1 Convolutional Layer.....	19
2.7.2 Primary Capsule.....	20
2.7.3 Dynamic Routing.....	21
2.7.4 Squashing Function.....	22
2.8 Long Short-Term Memory	23
2.9 Arsitektur Hibrida.....	25
2.10 Evaluasi Model	26
2.11 Shortcut Learning	29
2.12 K-Fold Cross-Validation	29
BAB III TAHAPAN PELAKSANAAN	31
3.1 Alur Penelitian	31
3.2 Dataset	33
3.2.1 Sumber Dataset.....	33
3.2.2 Komposisi Dataset.....	33
3.2.3 Pembagian Data.....	34
3.2.4 Pemotongan Segmen Audio.....	35
3.2.5 Analisis Dataset	36
3.3 Ekstraksi Fitur MFCC.....	38
3.3.1 Tahapan Pra-pemrosesan Sinyal.....	39
3.3.2 Tahapan Transformasi Inti.....	41
3.3.3 Ringkasan Parameter MFCC	44
3.3.4 Normalisasi Matriks Fitur.....	45
3.4 Perancangan Arsitektur CapsNet.....	46
3.4.1 Convolutional Layer	47
3.4.2 Primary Capsule.....	48
3.4.3 DigitCaps	50
3.4.4 Ringkasan Parameter CapsNet.....	53
3.5 Perancangan Arsitektur LSTM	54
3.5.1 Reshaping Layer	54
3.5.2 LSTM Layer	55
3.5.3 Ringkasan Parameter LSTM.....	57

3.6 Integrasi Arsitektur Hybrid CapsNet-LSTM	58
3.6.1 Arsitektur CapsNet	59
3.6.2 Arsitektur LSTM.....	60
3.6.3 Mekanisme Penggabungan dan Klasifikasi	61
3.7 Perancangan Proses Pelatihan Model	63
3.7.1 Loss Function.....	63
3.7.2 Optimization Algorithm.....	64
3.7.3 Hyper Parameter	65
3.7.4 Ringkasan Parameter Pelatihan	67
3.8 Evaluasi.....	68
BAB IV HASIL DAN PEMBAHASAN.....	72
4.1 Hasil.....	72
4.1.1 Hasil Pembagian dan Segmentasi Dataset A dan B	72
4.1.2 Hasil Ekstraksi Fitur MFCC.....	74
4.1.3 Hasil Pelatihan Model	76
4.1.4 Hasil Model Terbaik.....	77
4.1.5 Hasil Pengujian Antar Varian Dataset.....	79
4.1.6 Hasil Validasi K-Fold pada Model Terbaik	80
4.2 Pembahasan	81
4.2.1 Karakteristik Dataset dan Hasil Persiapan Data.....	81
4.2.2 Pengaruh Hyperparameter terhadap Performa Model.....	82
4.2.3 Perbandingan Performa Dataset A dan Dataset B.....	83
4.2.4 Pola Kesalahan Klasifikasi Model	84
4.2.5 Pengujian Antar Varian Dataset.....	85
4.2.6 Pembahasan Validasi K-Fold.....	87
4.2.7 Ringkasan Hasil Pembahasan.....	87
BAB V PENUTUP	90
5.1 Kesimpulan.....	90
5.2 Saran	91
DAFTAR PUSTAKA	92
DAFTAR RIWAYAT HIDUP	102
LAMPIRAN	105
Lampiran A. Confusion Matrix	105
Lampiran B. Grafik Pelatihan Model	106

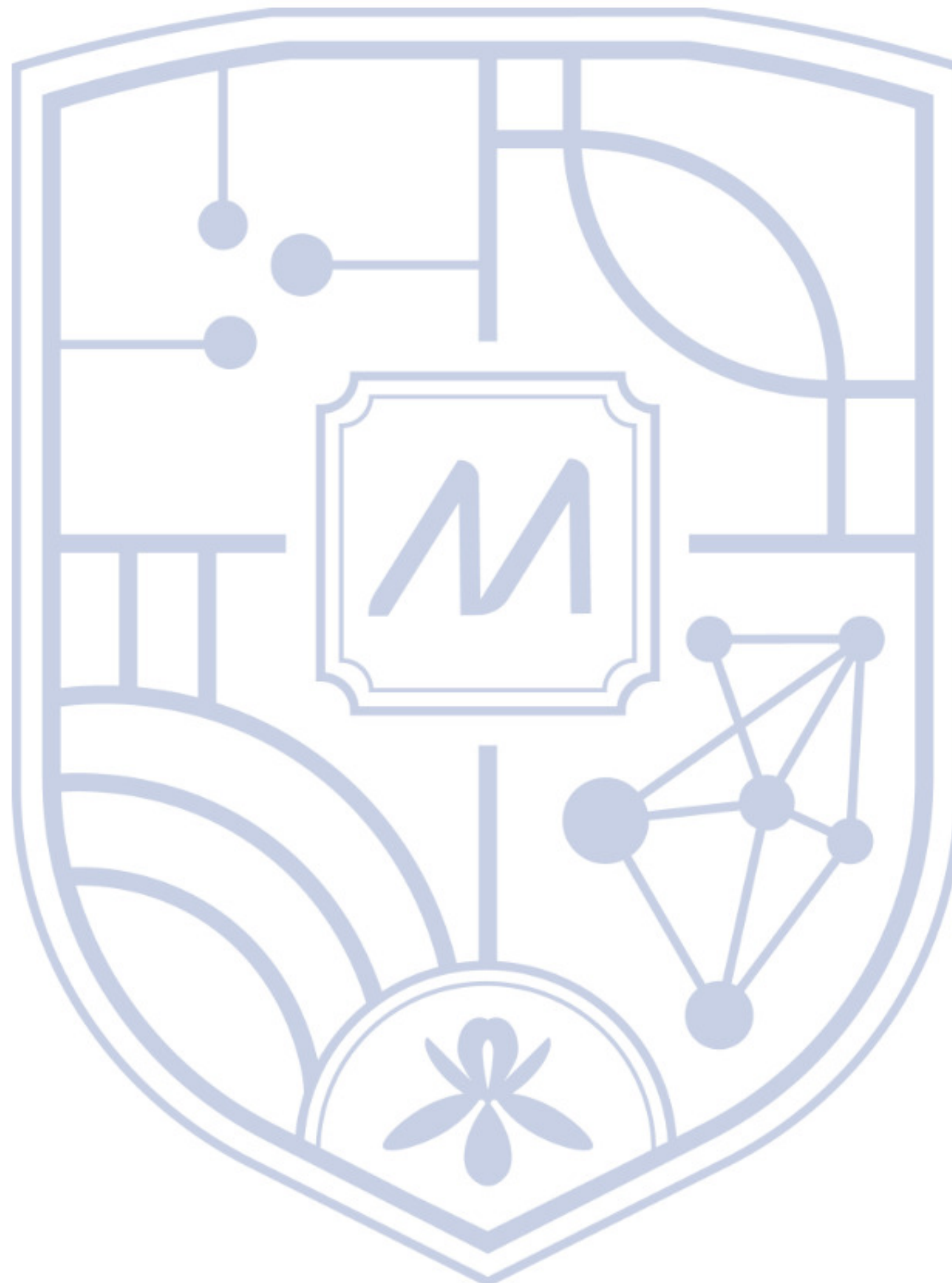
DAFTAR GAMBAR

Gambar 2.1 Perbandingan Spektrogram Audio Asli dan Audio <i>Deepfake</i> [37]	8
Gambar 2.2 Diagram Alur Tahapan Ekstraksi Fitur MFCC [42].....	10
Gambar 2.3 Arsitektur <i>Capsule Network</i> [59].....	19
Gambar 2.4 Arsitektur LSTM [63].....	23
Gambar 3.1 Diagram Alur Penelitian.....	31
Gambar 3.2 Alur Arsitektur CapsNet.....	47
Gambar 3.3 Alur Arsitektur LSTM.....	54
Gambar 3.4 Flowchart Arsitektur Hybrid CapsNet-LSTM.....	59
Gambar 4.1 MFCC	75
Gambar 4.2 Kurva Pelatihan Model Terbaik Dataset A.....	77
Gambar 4.3 Kurva Pelatihan Model Terbaik Dataset B.....	78
Gambar 4.4 Confusion Matrix Model Terbaik Dataset A.....	78
Gambar 4.5 Confusion Matrix Model Terbaik Dataset B	79
Gambar 4.6 Heatmap F1-Score	82
Gambar 4.7 Perbandingan Dataset A dan Dataset B.....	84
Gambar 4.8 Perbandingan Pengujian Akurasi Antar Dataset	86

DAFTAR TABEL

Tabel 3.1 Komposisi Dataset.....	34
Tabel 3.2 Distribusi file per subset.....	34
Tabel 3.3 Distribusi Segmen Dataset A.....	35
Tabel 3.4 Distribusi Segmen Dataset B.....	35
Tabel 3.5 Sensitivitas Ambang Pembuangan Keheningan.....	36
Tabel 3.6 Sepuluh Segmen dengan Energi Terendah.....	37
Tabel 3.7 ilustrasi 5 Sampel Pertama Sinyal Audio	39
Tabel 3.8 Contoh Perhitungan <i>Windowing</i> pada 5 Sampel Pertama	41
Tabel 3.9 Perhitungan nilai frekuensi <i>frame</i>	42
Tabel 3.10 Perhitungan logaritma pada 5 filter pertama	43
Tabel 3.11 Perhitungan koefisien pada MFCC	44
Tabel 3.12 Ringkasan Parameter Ekstraksi MFCC.....	44
Tabel 3.13 Contoh Perhitungan <i>StandardScaler</i> pada 5 <i>Frame</i> Pertama.....	46
Tabel 3.14 Perhitungan output <i>DigitCaps</i>	53
Tabel 3.15 Ringkasan Parameter Arsitektur CapsNet.....	53
Tabel 3.16 input LSTM Setelah <i>reshaping</i>	55
Tabel 3.17 ilustrasi aliran data pada 3 <i>timestep</i> pertama.....	55
Tabel 3.18 Ringkasan Parameter Arsitektur LSTM.....	57
Tabel 3.19 Transformasi Dimensi Cabang CapsNet	60
Tabel 3.20 Transformasi Dimensi Cabang LSTM	60
Tabel 3.21 Keputusan ambang batas	62
Tabel 3.22 Dimensi Output setiap Tahapan	63
Tabel 3.23 Contoh Perhitungan <i>Loss</i>	63
Tabel 3.24 Ilustrasi Pengaruh <i>Learning Rate</i> terhadap Pembaruan Parameter	66
Tabel 3.25 Ringkasan Parameter Pelatihan Model.....	68
Tabel 4.1 Hasil Distribusi Segmen Dataset A dan Dataset B.....	72
Tabel 4.2 Hasil Distribusi Segmen Dataset A dan Dataset B.....	73
Tabel 4.3 Pembuangan keheningan	73
Tabel 4.4 Hasil Pemeriksaan Kualitas Segmen.....	74
Tabel 4.5 Dimensi Hasil Ekstraksi MFCC	75
Tabel 4.6 Hasil Pelatihan dan Pengujian 12 Skenario.....	76
Tabel 4.7 Tabel Model Terbaik	77

Tabel 4.8 Hasil Pengujian Ketahanan Model terhadap Data Hening	80
Tabel 4.9 Hasil K-Fold	80
Tabel 4.10 Nilai Probabilitas Segmen	85



DAFTAR LAMPIRAN

Gambar A. 12 Skenario Confusion Matrix.....	106
Gambar B. Grafik Pelatihan Model.....	109

