



**PERBANDINGAN KINERJA MODEL 3D-CNN, MOBILENETV2, DAN
SWIN TRANSFORMER UNTUK KLASIFIKASI LIP READING**

SKRIPSI

ALBERT CHANDRA JUNIOR

41522110044

**UNIVERSITAS
MERCU BUANA
PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS MERCU BUANA
JAKARTA
2026**



**PERBANDINGAN KINERJA MODEL 3D-CNN, MOBILENETV2, DAN
SWIN TRANSFORMER UNTUK KLASIFIKASI LIP READING**

SKRIPSI

ALBERT CHANDRA JUNIOR
41522110044

Diajukan sebagai salah satu syarat untuk memperoleh gelar sarjana

**UNIVERSITAS
MERCU BUANA**
**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS MERCU BUANA
JAKARTA
2026**

HALAMAN PERNYATAAN KARYA SENDIRI

Saya yang bertanda tangan di bawah ini:

Nama : Albert Chandra Junior
NIM : 41522110044
Program Studi : Teknik Informatika
PERBANDINGAN KINERJA MODEL 3D-
CNN, MOBILENETV2, DAN SWIN
Judul Skripsi : TRANSFORMER UNTUK KLASIFIKASI
LIP READING

Menyatakan bahwa Laporan Skripsi ini adalah hasil karya saya sendiri dan bukan plagiat, serta semua sumber baik yang dikutip maupun dirujuk telah saya nyatakan dengan benar. Apabila ternyata ditemukan di dalam Skripsi saya terdapat unsur plagiat, maka saya siap mendapatkan sanksi akademis yang berlaku di Universitas Mercu Buana.



Jakarta, 13 Agustus 2026

METERAN
TEMPEL
64A22AOX149160322
Albert Chandra Junior.

UNIVERSITAS
MERCU BUANA

HALAMAN SURAT KETERANGAN HASIL UJI TURNITIN



072.423.4.03.01

PERNYATAAN SIMILARITY CHECK /SIMILARITY CHECK STATEMENT

Saya yang bertanda tangan dibawah ini menyatakan, bahwa karya ilmiah yang ditulis oleh
/The undersigned, hereby declare that the scientific work written by

Nama /Name : Albert Chandra Junior
NIM /Student ID Number : 41522110044
Program Studi /Study Program : Teknik Informatika

Dengan Judul Tugas Akhir

/The title:

“PERBANDINGAN KINERJA MODEL 3D-CNN, MOBILENETV2, DAN SWIN TRANSFORMER UNTUK KLASIFIKASI LIP READING”

telah dilakukan pengecekan *similarity* dengan sistem Turnitin pada tanggal:

/Similarity checks have been carried out with the Turnitin system on the date:

29 Juli 2026

dengan nilai persentase sebesar :

/with a percentage value of:

8%

dinyatakan memenuhi standar sesuai dengan ketentuan berlaku di **Fakultas Ilmu Komputer** Universitas Mercu Buana. */declared to meet standards in accordance with applicable regulations at the Faculty of Computer Science, Universitas Mercu Buana.*

File hasil cek similarity turnitin:

/Turnitin similarity report file

https://drive.google.com/file/d/1Cqu8QkDe3-lqUkqsvkj6li921hYMBUA/view?usp=drive_link



Jakarta, 29 Juli 2026
Admin Turnitin Fasilkom UMB

Agung Prawoto

Agung Prawoto, S.Kom., B.Sc

NIK : 322970503

Fakultas Ilmu Komputer

KAMPUS MENARA BHAKTI

Jl. Raya Meruya Selatan No. 1 Kembangan, Jakarta Barat 11650

Telp. 021-5840816 (Hunting), Psw : 5700 Fax. 021-5840813

<http://www.mercubuana.ac.id>, e-mail : fasilkom@mercubuana.ac.id

HALAMAN PENGESAHAN

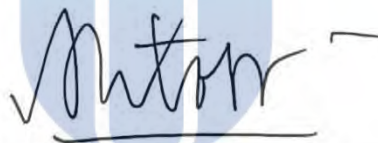
Tugas Akhir ini diajukan oleh:

Nama : Albert Chandra Junior
NIM : 41522110044
Fakultas/Program Studi : Fakultas Ilmu Komputer / Teknik Informatika
Judul Tugas Akhir : PERBANDINGAN KINERJA MODEL 3D-CNN, MOBILENETV2,
DAN SWIN TRANSFORMER UNTUK KLASIFIKASI LIP
READING

Telah berhasil dipertahankan pada sidang tanggal 13 Agustus 2026 dan diterima sebagai bagian persyaratan yang diperlukan untuk memperoleh gelar Sarjana pada Program Studi Teknik Informatika, Fakultas Ilmu Komputer Universitas Mercu Buana.

Disahkan oleh:

Pembimbing



Dr. Hadi Santoso, S.Kom., M.Kom

NIDN/NUPTK: 0225067701

Jakarta, 13 Agustus 2026

Mengetahui,

Dekan Fakultas Ilmu Komputer



Dr. Bambang Jokonowo, S.Si., MTI
NIDN/NUPTK: 0320037002

Ketua Program Studi
Teknik Informatika



Dr. Hadi Santoso, S.Kom., M.Kom
NIDN/NUPTK: 0225067701

KATA PENGANTAR

Puji syukur kehadiran Tuhan yang Maha Esa, atas segala rahmat dan ridha-Nya sehingga penulis dapat menyelesaikan skripsi yang merupakan salah satu persyaratan kelulusan Program Studi Strata Satu (S1) pada jurusan Teknik Informatika, Universitas Mercu Buana.

Penulis menyadari bahwa skripsi ini masih jauh dari sempurna, karena kesempurnaan sejatinya hanya milik Tuhan yang Maha Esa. Oleh karena itu, saran dan masukan yang membangun senantiasa penulis terima dengan senang hati. Serta berkat dukungan, motivasi, bantuan, bimbingan, dan doa dari banyak pihak, penulis mengucapkan terima kasih kepada:

1. Bapak Prof. Dr. Andi Adriansyah, M.Eng. selaku Rektor Universitas Mercu Buana.
2. Bapak Dr. Bambang Jokonowo, S.Si., M.TI selaku Dekan Fakultas Ilmu Komputer.
3. Bapak Dr. Hadi Santoso, S.Kom., M.Kom. selaku Ketua Program Studi Teknik Informatika Universitas Mercubuana.
4. Bapak Dr. Hadi Santoso, S.Kom., M.Kom. selaku dosen pembimbing MPTI yang telah memberikan pengarahan, motivasi, menyediakan waktu, tenaga, dan pikiran sehingga selama pembuatan skripsi ini terjadwal dengan baik.
5. Kedua Orang Tua saya yang selalu mensupport dan mendukung saya selama menjalani masa studi sebagai mahasiswa Universitas Mercubuana..
6. Semua teman kuliah yang selalu berbagi informasi dan memberikan dukungan dalam bentuk yang berbeda-beda.

Akhir kata, penulis berharap semoga Tuhan yang Maha Esa membalas kebaikan dan selalu mencurahkan rahmat, hidayah, serta panjang umur kepada kita semua, aamiin. Terima Kasih.

Jakarta, 13 Agustus 2026

Albert Chandra Junior

**HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI
TUGAS AKHIR DI REPOSITORI UMB**

Sebagai sivitas akademik Universitas Mercu Buana, saya yang bertanda tangan di bawah ini:

Nama : Albert Chandra Junior
NIM : 41522110044
Fakultas/Program Studi : Fakultas Ilmu Komputer / Teknik Informatika
Judul Laporan Skripsi : PERBANDINGAN KINERJA MODEL 3D-CNN,
MOBILENETV2, DAN SWIN TRANSFORMER UNTUK
KLASIFIKASI LIP READING

Demi pengembangan ilmu pengetahuan, dengan ini memberikan izin dan menyetujui untuk memberikan kepada Universitas Mercu Buana **Hak Bebas Royalti Non-Eksklusif** (*Non-exclusive Royalty-Free Right*) atas karya ilmiah saya yang berjudul di atas beserta perangkat yang ada (jika diperlukan).

Dengan Hak Bebas Royalti Non-Eksklusif ini Universitas Mercu Buana berhak menyimpan, mengalihmedia/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.

Jakarta, 13 Agustus 2026

Yang menyatakan,



Albert Chandra Junior

PERBANDINGAN KINERJA MODEL 3D-CNN, MOBILENETV2, DAN SWIN TRANSFORMER UNTUK KLASIFIKASI LIP READING

ALBERT CHANDRA JUNIOR

ABSTRAK

Visual Speech Recognition (VSR) atau *lip reading* merupakan teknologi penting untuk mendukung komunikasi pada lingkungan dengan gangguan akustik tinggi maupun bagi penyandang tunarungu, namun menghadapi tantangan berupa ambiguitas visual (*homophenes*) serta variasi karakteristik gerak bibir antar-individu. Penelitian ini membandingkan kinerja tiga arsitektur *Deep Learning*, yaitu *3D Convolutional Neural Network (3D-CNN)* yang dilatih dari awal (*from scratch*), *MobileNetV2* dengan pendekatan *transfer learning*, dan *Swin Transformer* yang dipadukan dengan *Temporal Transformer Encoder*, dalam menyelesaikan tugas klasifikasi *lip reading* ke dalam 10 kelas kata pada Dataset MIRACL-VC1. Setiap video diproses melalui deteksi wajah dan pemotongan *region of interest (ROI)* pada area bibir menggunakan 68 *facial landmarks*, kemudian dilatih menggunakan skema *single split* selama 15 *epoch*. Kinerja ketiga model dievaluasi menggunakan metrik akurasi, presisi, *recall*, F1-score, *confusion matrix*, serta kecepatan inferensi (*Frames Per Second/FPS*) untuk menilai kelayakan *real-time*. Hasil pengujian menunjukkan bahwa Swin Transformer memberikan akurasi uji tertinggi sebesar 68,50% dan F1-score makro 0,6791, diikuti oleh MobileNetV2 (46,30%; 0,4536) dan 3D-CNN (35,31%; 0,3253); namun Swin Transformer memerlukan waktu pelatihan paling lama dan hanya mencapai 3,60 FPS saat inferensi sehingga belum memenuhi kriteria *real-time*, berbeda dengan 3D-CNN (53,33 FPS) dan MobileNetV2 (45,91 FPS) yang tetap memenuhi kriteria tersebut. Temuan ini menunjukkan adanya *trade-off* antara akurasi klasifikasi dan efisiensi komputasi pada ketiga arsitektur yang diuji.

Kata kunci: Visual Speech Recognition (VSR), Lip Reading, 3D Convolutional Neural Network (3D-CNN), MobileNetV2, Swin Transformer.

Performance Comparison of 3D-CNN, MobileNetV2, and Swin Transformer Models for Lip-Reading Classification

ALBERT CHANDRA JUNIOR

ABSTRACT

Visual Speech Recognition (VSR), or lip reading, is an important technology for supporting communication in acoustically noisy environments and for people with hearing impairments, although it still faces challenges such as visual ambiguity (homophenes) and variation in lip movement across individuals. This study compares the performance of three Deep Learning architectures, namely a 3D Convolutional Neural Network (3D-CNN) trained from scratch, MobileNetV2 with a transfer learning approach, and a Swin Transformer combined with a Temporal Transformer Encoder, on a 10-class lip reading classification task using the MIRACL-VC1 dataset. Each video is processed through face detection and lip region-of-interest (ROI) cropping using 68 facial landmarks, before being trained with a single-split scheme for 15 epochs. The three models are evaluated using accuracy, precision, recall, F1-score, confusion matrix, and inference speed (Frames Per Second/FPS) to assess real-time feasibility. The results show that the Swin Transformer achieves the highest test accuracy of 68.50% and a macro F1-score of 0.6791, followed by MobileNetV2 (46.30%; 0.4536) and 3D-CNN (35.31%; 0.3253). However, the Swin Transformer requires the longest training time and reaches only 3.60 FPS during inference, failing to meet the real-time criterion, unlike 3D-CNN (53.33 FPS) and MobileNetV2 (45.91 FPS), which remain real-time capable. These findings indicate a trade-off between classification accuracy and computational efficiency among the three tested architectures

Keywords: Visual Speech Recognition (VSR), Lip Reading, 3D Convolutional Neural Network (3D-CNN), MobileNetV2, Swin Transformer.

MERCU BUANA

DAFTAR ISI

HALAMAN SAMPUL.....	0
HALAMAN JUDUL	i
HALAMAN PERNYATAAN KARYA SENDIRI	ii
HALAMAN SURAT KETERANGAN HASIL UJI TURNITIN	iii
HALAMAN PENGESAHAN.....	iv
KATA PENGANTAR.....	v
HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR DI REPOSITORI UMB.....	vi
ABSTRAK	vii
ABSTRACT	viii
DAFTAR ISI.....	ix
DAFTAR TABEL	xi
DAFTAR GAMBAR.....	xii
DAFTAR LAMPIRAN	xiii
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Perumusan Masalah	3
1.3 Tujuan Penelitian	3
1.4 Manfaat/Kontribusi Penelitian	3
BAB II TINJAUAN PUSTAKA	5
2.1 Teori Utama	5
2.2 Teori Pendukung.....	5
2.3 Penelitian Terdahulu	8
2.4 Gap Penelitian	10
BAB III METODE PENELITIAN	12
3.1 Pendekatan Penelitian	12
3.2 Desain Penelitian	12
3.3 Dataset dan Praproses Data.....	13
3.4 Arsitektur Model	15
3.5 Skema Pelatihan.....	18

3.6	Instrumen Penelitian	18
3.7	Metrik Evaluasi	19
BAB IV HASIL DAN PEMBAHASAN		20
4.1	Kurva Pelatihan dan Validasi.....	20
4.2	Kinerja pada Data Uji (Test Set).....	22
4.3	Analisis Laporan Klasifikasi per Kelas.....	24
4.4	Analisis Confusion Matrix	26
4.5	Analisis Efisiensi Komputasi dan Kelayakan Real-Time	28
4.6	Evaluasi Kualitas ROI Bibir (Intersection over Union).....	30
4.7	Perbandingan dengan Penelitian Terdahulu.....	32
4.8	Diskusi Umum	33
BAB V KESIMPULAN DAN SARAN		36
5.1	Kesimpulan	36
5.2	Saran	37
DAFTAR PUSTAKA.....		38
LAMPIRAN.....		43



DAFTAR TABEL

Tabel 2. 1 Penelitian Terkait	8
Tabel 3. 1 Parameter pelatihan model.....	18
Tabel 4. 1 Perbandingan kinerja ketiga model pada data uji	22
Tabel 4. 2 Precision (P), Recall (R), dan F1-score per kelas pada data uji.....	24
Tabel 4. 3 Hasil pengujian kecepatan inferensi (FPS) ketiga model	28
Tabel 4. 4 Rata-rata IoU deteksi ROI bibir pada data uji	30
Tabel 4. 5 Distribusi nilai IoU per kategori pada data uji.....	30
Tabel 4. 6 Rata-rata IoU per kelas kata pada data uji	31



DAFTAR GAMBAR

Gambar 3. 1 Diagram alur metodologi penelitian.....	12
Gambar 3. 2 Ilustrasi ringkas arsitektur model 3D-CNN, MobileNetV2, dan Swin Transformer.....	15
Gambar 3. 3 Rincian layer arsitektur 3D-CNN beserta perubahan dimensi tensor	16
Gambar 3. 4 Rincian layer arsitektur MobileNetV2 (transfer learning) beserta perubahan dimensi tensor.....	16
Gambar 3. 5 Rincian layer arsitektur Swin Transformer + Temporal Transformer beserta perubahan dimensi tensor	17
Gambar 4. 1 Perbandingan training loss, validation loss, training accuracy, dan validation accuracy ketiga model.....	20
Gambar 4. 2 Confusion matrix model 3D-CNN, MobileNetV2, dan Swin Transformer pada data uji	26
Gambar 4. 3 (a) Rata-rata IoU per kelas kata dengan standar deviasi; (b) distribusi kategori kualitas IoU pada data uji	31



DAFTAR LAMPIRAN

Lampiran 1 Kartu Asistensi	44
Lampiran 2 Curriculum Vitae	45
Lampiran 3 Surat Pernyataan & Pengalihan HAKI	46
Lampiran 3 Sertifikat BNSP	48

