

**ANALISIS PERBANDINGAN KINERJA ARSITEKTUR 3D  
CNN, VIDEO VISION TRANSFORMER, DAN 3D  
CONVNEXT DALAM KLASIFIKASI  
VIDEO DEEPPFAKE**

**SKRIPSI**

Diajukan untuk memenuhi salah satu syarat mencapai derajat Sarjana  
Program Studi Sarjana Informatika



disusun oleh

**Alvin Nauval Harits**

**22.11.4546**

Kepada

**FAKULTAS ILMU KOMPUTER  
UNIVERSITAS AMIKOM YOGYAKARTA  
YOGYAKARTA**

**2026**

**ANALISIS PERBANDINGAN KINERJA ARSITEKTUR 3D  
CNN, VIDEO VISION TRANSFORMER, DAN 3D  
CONVNEXT DALAM KLASIFIKASI  
VIDEO DEEPPAKE**

**SKRIPSI**

untuk memenuhi salah satu syarat mencapai derajat Sarjana  
Program Studi Sarjana Informatika



disusun oleh  
**Alvin Nauval Harits**  
**22.11.4546**

Kepada  
**FAKULTAS ILMU KOMPUTER**  
**UNIVERSITAS AMIKOM YOGYAKARTA**  
**YOGYAKARTA**  
**2026**

**HALAMAN PERSETUJUAN**

**SKRIPSI**

**ANALISIS PERBANDINGAN KINERJA ARSITEKTUR 3D  
CNN, VIDEO VISION TRANSFORMER, DAN 3D  
CONVNEXT DALAM KLASIFIKASI  
VIDEO DEEPPFAKE**

yang disusun dan diajukan oleh

**Alvin Nauval Harits**

**22.11.4546**

telah disetujui oleh Dosen Pembimbing Skripsi  
pada tanggal 21 Mei 2026

Dosen Pembimbing,



**Ike Verawati, S.Kom., M.Kom.**  
**NIK. 190302237**

**HALAMAN PENGESAHAN**

**SKRIPSI**

**ANALISIS PERBANDINGAN KINERJA ARSITEKTUR 3D  
CNN, VIDEO VISION TRANSFORMER, DAN 3D  
CONVNEXT DALAM KLASIFIKASI  
VIDEO DEEPFAKE**

yang disusun dan diajukan oleh

**Alvin Nauval Harits**

**22.11.4546**

Telah dipertahankan di depan Dewan Penguji  
pada tanggal 21 Mei 2026

Susunan Dewan Penguji

**Nama Penguji**

**Tanda Tangan**

**Nuri Cahyona, M.Kom.**  
NIK. 190302278

**Bambang Pili Hartato, S.Kom., M.Eng.**  
NIK. 190302707

**Ike Verawati, S.Kom., M.Kom.**  
NIK. 190302237

Skrripsi ini telah diterima sebagai salah satu persyaratan  
untuk memperoleh gelar Sarjana Komputer  
Tanggal 21 Mei 2026

**DEKAN FAKULTAS ILMU KOMPUTER**



**Prof. Dr. Kusriani, M.Kom.**  
NIK. 190302106

## HALAMAN PERNYATAAN KEASLIAN SKRIPSI

Yang bertandatangan di bawah ini,

Nama mahasiswa : Alvin Nauval Harits  
NIM : 22.11.4546

Menyatakan bahwa Skripsi dengan judul berikut:

**Analisis Perbandingan Kinerja Arsitektur 3D CNN, Video Vision Transformer, dan 3D Convnext Dalam Klasifikasi Video Deepfake.**

Dosen Pembimbing : Ike Verawati, S.Kom., M.Kom.

1. Karya tulis ini adalah benar-benar ASLI dan BELUM PERNAH diajukan untuk mendapatkan gelar akademik, baik di Universitas AMIKOM Yogyakarta maupun di Perguruan Tinggi lainnya.
2. Karya tulis ini merupakan gagasan, rumusan dan penelitian SAYA sendiri, tanpa bantuan pihak lain kecuali arahan dari Dosen Pembimbing.
3. Dalam karya tulis ini tidak terdapat karya atau pendapat orang lain, kecuali secara tertulis dengan jelas dicantumkan sebagai acuan dalam naskah dengan disebutkan nama pengarang dan disebutkan dalam Daftar Pustaka pada karya tulis ini.
4. Perangkat lunak yang digunakan dalam penelitian ini sepenuhnya menjadi tanggung jawab SAYA, bukan tanggung jawab Universitas AMIKOM Yogyakarta.
5. Pernyataan ini SAYA buat dengan sesungguhnya, apabila di kemudian hari terdapat penyimpangan dan ketidakbenaran dalam pernyataan ini, maka SAYA bersedia menerima SANKSI AKADEMIK dengan pencabutan gelar yang sudah diperoleh, serta sanksi lainnya sesuai dengan norma yang berlaku di Perguruan Tinggi.

Yogyakarta, 21 Mei 2026

Yang Menyatakan,



Alvin Nauval Harits

## HALAMAN PERSEMBAHAN

Dengan rasa syukur yang mendalam kepada Tuhan Yang Maha Esa, karya sederhana ini sepenuhnya saya persembahkan kepada orang-orang terkasih yang menjadi alasan, kekuatan, dan penyokong hidup saya untuk terus melangkah:

Almarhum Ayahanda tercinta **Kholrul Anam** & Almarhumah Ibunda tercinta **Siti Rokayah** Persembahan ini adalah bukti bakti atas setiap tetes keringat, doa, kasih sayang, serta jerih payah luar biasa Ayah dan Ibu semasa hidup. Terima kasih yang tak terhingga karena atas bekal, warisan perjuangan, dan hasil kerja keras yang kalian siapkanlah yang senantiasa membiayai dan menopang seluruh proses perkuliahan penulis hingga selesai sampai saat ini. Meski raga kita telah terpisah jarak, cinta, restu, dan perjuangan nyata kalian tetap hidup mengalir di setiap langkah penulis. Semoga Tuhan memberikan tempat paling mulia dan damai di sisi-Nya.

Kakak tersayang, **Aroby Ihsan** Pilar kekuatan utama, pelindung, sekaligus pengganti orang tua bagi penulis. Terima kasih yang tak terhingga atas pengorbanan luar biasa yang tidak pernah mengeluh, bantuan tanpa pamrih baik moril maupun materil, serta nasihat dan doa yang selalu menguatkan penulis dalam melewati setiap badai dan tantangan selama masa perkuliahan hingga skripsi ini selesai. Karya ini juga milikmu.

## KATA PENGANTAR

Puji dan syukur penulis panjatkan ke hadirat Tuhan Yang Maha Esa atas rahmat dan karunia-Nya, sehingga penulis dapat menyelesaikan skripsi yang berjudul **"Analisis Perbandingan Kinerja Arsitektur 3D CNN, Video Vision Transformer, dan 3D ConvNeXt dalam Klasifikasi Video Deepfake"**.

Skripsi ini disusun sebagai salah satu syarat kelulusan untuk memperoleh gelar Sarjana Komputer (S.Kom.) pada Program Studi S1 Informatika, Fakultas Ilmu Komputer, Universitas AMIKOM Yogyakarta.

Penulis menyampaikan ucapan terima kasih yang sebesar-besarnya atas bimbingan dan bantuan dalam penyelesaian skripsi ini kepada:

1. Eli Pujastuti, M.Kom., selaku Ketua Program Studi S1 Informatika Universitas AMIKOM Yogyakarta yang telah memberikan dukungan dan fasilitas administratif selama masa perkuliahan dan penyusunan skripsi.
2. Ike Verawati, S.Kom., M.Kom., selaku Dosen Pembimbing yang telah memberikan bimbingan, arahan, dan motivasi berharga selama penelitian.
3. Seluruh Dosen dan Staf Pengajar Program Studi S1 Informatika Universitas AMIKOM Yogyakarta yang telah mendidik penulis selama masa perkuliahan.
4. Sahabat-sahabat yang telah memberikan dukungan, dan bantuan hingga skripsi ini dapat diselesaikan.

Penulis menyadari bahwa skripsi ini masih memiliki banyak kekurangan. Oleh karena itu, kritik dan saran yang membangun sangat diharapkan demi perbaikan di masa depan. Semoga karya ini bermanfaat bagi perkembangan ilmu teknologi informasi.

Yogyakarta, 21 Mei 2026



Penulis

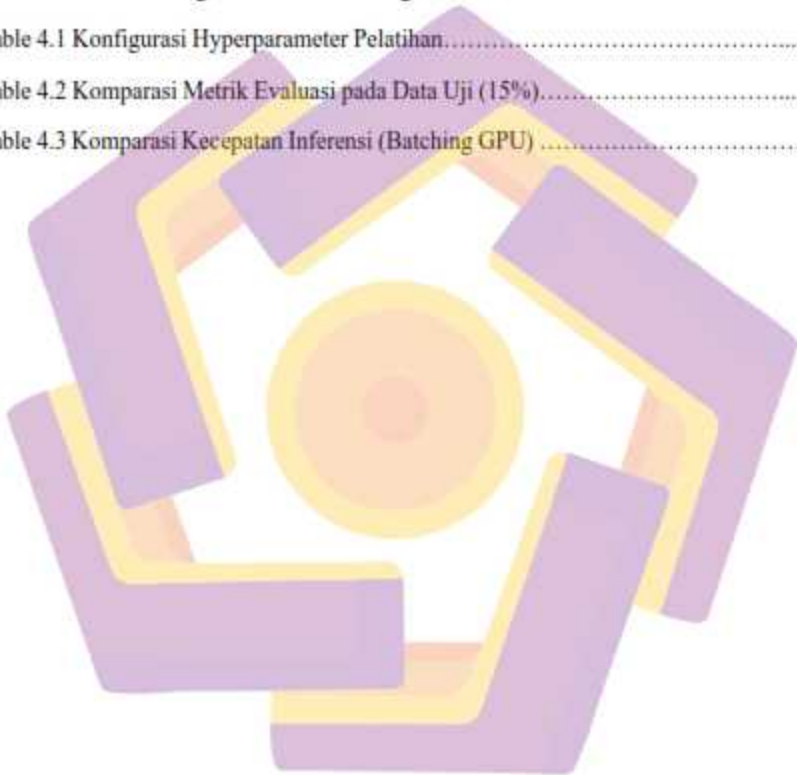
## DAFTAR ISI

|                                                                 |      |
|-----------------------------------------------------------------|------|
| HALAMAN JUDUL .....                                             | I    |
| HALAMAN PERSETUJUAN .....                                       | II   |
| HALAMAN PENGESAHAN .....                                        | III  |
| HALAMAN PERNYATAAN KEASLIAN SKRIPSI .....                       | IV   |
| HALAMAN PERSEMBAHAN .....                                       | V    |
| KATA PENGANTAR .....                                            | VI   |
| DAFTAR ISI .....                                                | VII  |
| DAFTAR TABEL .....                                              | IX   |
| DAFTAR GAMBAR .....                                             | X    |
| INTISARI .....                                                  | XII  |
| ABSTRACT .....                                                  | XIII |
| BAB I PENDAHULUAN .....                                         | 1    |
| 1.1 Latar Belakang .....                                        | 1    |
| 1.2 Rumusan Masalah .....                                       | 3    |
| 1.3 Batasan Masalah .....                                       | 4    |
| 1.4 Tujuan Penelitian .....                                     | 5    |
| 1.5 Manfaat Penelitian .....                                    | 6    |
| 1.6 Sistematika Penulisan .....                                 | 6    |
| BAB II TINJAUAN PUSTAKA .....                                   | 8    |
| 2.1 Studi Literatur .....                                       | 8    |
| 2.2 Dasar Teori .....                                           | 17   |
| 2.2.1 Deep Learning .....                                       | 17   |
| 2.2.2 Deepfake dan Generative Adversarial Networks (GANs) ..... | 19   |
| 2.2.3 3D Convolutional Neural Network (3D CNN) .....            | 21   |
| 2.2.4 Video Vision Transformer (ViViT) .....                    | 23   |
| 2.2.5 3D ConvNeXt .....                                         | 25   |
| 2.2.6 Lapisan Klasifikasi ( Classification Layer ) .....        | 28   |
| 2.2.7 Metrik Evaluasi Kinerja .....                             | 29   |
| 2.2.8 Pra-pemrosesan Data (Data Preprocessing) .....            | 32   |
| 2.2.9 Pembagian Data (Data Splitting) .....                     | 33   |

|                                   |                                     |    |
|-----------------------------------|-------------------------------------|----|
| 2.2.10                            | Perancangan Arsitektur Model .....  | 34 |
| 2.2.11                            | Pelatihan Model (Training) .....    | 36 |
| 2.2.12                            | Analisis Komparatif .....           | 37 |
| BAB III METODE PENELITIAN .....   |                                     | 39 |
| 3.1                               | Alur Penelitian .....               | 39 |
| 3.1.1                             | Pengumpulan Data (Kaggle API) ..... | 40 |
| 3.1.2                             | Pra-pemrosesan .....                | 41 |
| 3.1.3                             | Pembagian Data .....                | 41 |
| 3.1.4                             | Membangun Arsitektur .....          | 42 |
| 3.1.5                             | Training Model .....                | 48 |
| 3.1.6                             | Pengujian dan Evaluasi .....        | 51 |
| 3.1.7                             | Analisis Komparatif .....           | 59 |
| 3.2                               | Alat dan Bahan .....                | 60 |
| 3.2.1                             | Alat Penelitian .....               | 60 |
| 3.2.2                             | Bahan Penelitian .....              | 62 |
| BAB IV HASIL DAN PEMBAHASAN ..... |                                     | 65 |
| 4.1                               | Pengumpulan Data (Kaggle API) ..... | 65 |
| 4.3                               | Hasil Pra-Pemrosesan data .....     | 66 |
| 4.4                               | Pembagian Data .....                | 67 |
| 4.5                               | Membangun Arsitektur .....          | 68 |
| 4.6                               | Training Model .....                | 73 |
| 4.7                               | Hasil Testing Model .....           | 76 |
| 4.8                               | Deployment .....                    | 81 |
| BAB V PENUTUP .....               |                                     | 84 |
| 5.1                               | Kesimpulan .....                    | 84 |
| 5.2                               | Saran .....                         | 85 |
| REFERENSI .....                   |                                     | 87 |
| LAMPIRAN .....                    |                                     | 90 |

## DAFTAR TABEL

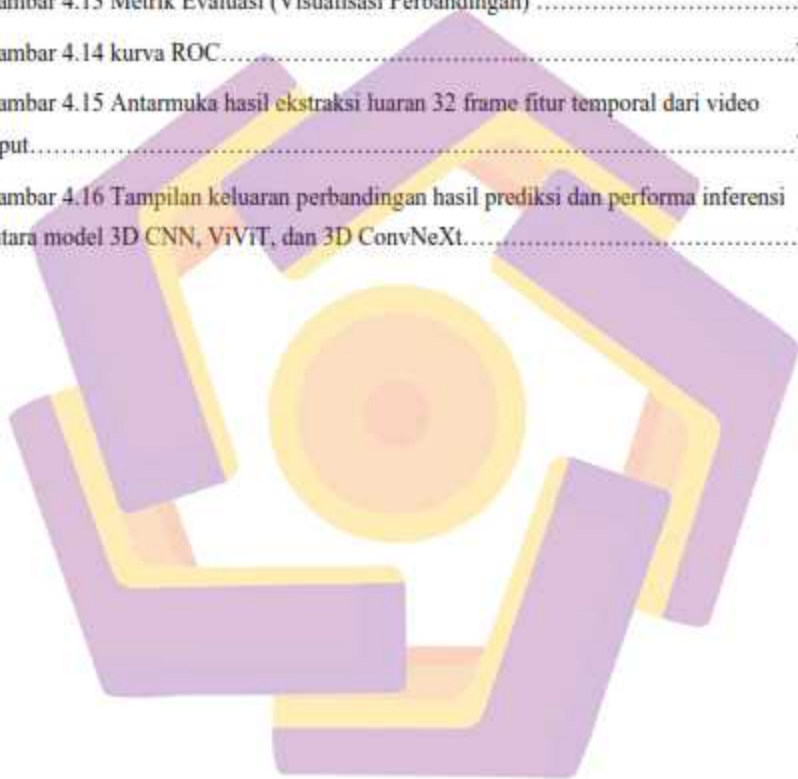
|                                                                    |    |
|--------------------------------------------------------------------|----|
| Table 2.1 Keaslian Penelitian.....                                 | 13 |
| Table 2.2 Matriks Konfusi (Confusion Matrix) Deteksi Deepfake..... | 29 |
| Table 3.1 Daftar Perangkat Keras dan Perangkat Lunak .....         | 57 |
| Table 4.1 Konfigurasi Hyperparameter Pelatihan.....                | 70 |
| Table 4.2 Komparasi Metrik Evaluasi pada Data Uji (15%).....       | 73 |
| Table 4.3 Komparasi Kecepatan Inferensi (Batching GPU) .....       | 77 |



## DAFTAR GAMBAR

|                                                                                                                                                 |    |
|-------------------------------------------------------------------------------------------------------------------------------------------------|----|
| Gambar 2.1 Perbandingan Arsitektur: (a) Jaringan Saraf Dangkal (Shallow Network ) dan (b) Deep Neural Network dengan Banyak Lapisan Berada..... | 18 |
| Gambar 2.2 Arsitektur Dasar Generative Adversarial Networks (GANs).....                                                                         | 20 |
| Gambar 2.3 Ilustrasi Arsitektur TD-3DCNN dengan Mekanisme Temporal Dropout.....                                                                 | 21 |
| Gambar 2.4 Arsitektur Video Vision Transformer (ViViT) dengan Ekstraksi Tubelet.....                                                            | 23 |
| Gambar 2.5 Perbandingan Struktur Blok ResNet (kiri) dan ConvNeXt (kanan) yang Menunjukkan Perbedaan Rasio Dimensi.....                          | 25 |
| Gambar 3.1 Alur Penelitian.....                                                                                                                 | 38 |
| Gambar 3.2 Distribusi Pembagian Dataset (70:15:15). ....                                                                                        | 40 |
| Gambar 3.3 Sampel data 32 frame kelas real pada dataset FaceForencis++ extracted frames.....                                                    | 60 |
| Gambar 3.4 Sampel data 32 frame kelas fake pada dataset FaceForencis++ extracted frames.....                                                    | 61 |
| Gambar 4.1 Tampilan Output Pengunduhan dan Ekstraksi Dataset.....                                                                               | 62 |
| Gambar 4.2 Visualisasi Contoh Urutan 32 Frame Wajah dalam Satu Folder Sekuens.....                                                              | 63 |
| Gambar 4.3 Output log dari script Python yang menampilkan jumlah akhir pembagian data.....                                                      | 64 |
| Gambar 4.4 Output Model Summary dari arsitektur 3D CNN.....                                                                                     | 66 |
| Gambar 4.5 Output Model Summary dari arsitektur ViViT.....                                                                                      | 67 |
| Gambar 4.6 Output Model Summary dari arsitektur 3D ConvNeXt.....                                                                                | 70 |
| Gambar 4.7 Grafik Training & Validation Accuracy dan Loss 3D CNN.....                                                                           | 71 |
| Gambar 4.8 Grafik Training & Validation Accuracy dan Loss ViViT.....                                                                            | 72 |

|                                                                                                                                   |    |
|-----------------------------------------------------------------------------------------------------------------------------------|----|
| Gambar 4.9 Grafik Training & Validation Accuracy dan Loss 3D ConvNeXt.....                                                        | 73 |
| Gambar 4.10 Confusion Matrix 3D CNN.....                                                                                          | 74 |
| Gambar 4.11 Confusion Matrix ViViT.....                                                                                           | 75 |
| Gambar 4.12 Confusion Matrix 3D ConvNeXt.....                                                                                     | 75 |
| Gambar 4.13 Metrik Evaluasi (Visualisasi Perbandingan) .....                                                                      | 76 |
| Gambar 4.14 kurva ROC.....                                                                                                        | 76 |
| Gambar 4.15 Antarmuka hasil ekstraksi luaran 32 frame fitur temporal dari video input.....                                        | 78 |
| Gambar 4.16 Tampilan keluaran perbandingan hasil prediksi dan performa inferensi antara model 3D CNN, ViViT, dan 3D ConvNeXt..... | 79 |



## INTISARI

Pesatnya perkembangan teknologi *Generative AI* telah memunculkan ancaman serius dalam bentuk video *deepfake*, yang berpotensi besar untuk penyebaran misinformasi dan aktivitas jahat lainnya. Kemampuan untuk mendeteksi video *deepfake* secara akurat dan efisien menjadi sangat krusial dalam menjaga keamanan digital dan kepercayaan publik. Penelitian ini bertujuan untuk melakukan analisis perbandingan kinerja dari tiga arsitektur *deep learning* modern *3D Convolutional Neural Network* (3D CNN), *Video Vision Transformer* (ViViT), dan 3D ConvNeXt untuk tugas klasifikasi video *deepfake*. Ketiga model dilatih dan dievaluasi menggunakan dataset standar industri, FaceForensics++, dengan metrik evaluasi meliputi akurasi, presisi, *recall*, dan *F1-Score*. Hasil eksperimen menunjukkan bahwa arsitektur 3D CNN murni menghasilkan kinerja paling unggul dengan tingkat akurasi mencapai 80,00% dan waktu inferensi 0,0531 detik per video. Sebaliknya, arsitektur *State-of-the-Art* (SOTA) seperti *Video Vision Transformer* (ViViT) dan 3D ConvNeXt hanya mampu mencatatkan akurasi masing-masing sebesar 51,00% dan 49,67% karena keterbatasan skala dataset (*data starvation*). Meskipun demikian, ViViT menawarkan efisiensi komputasi tertinggi pada GPU dengan kecepatan inferensi mencapai 911,10 FPS, sementara 3D ConvNeXt rentan terhadap masalah *over-prediction* dalam dimensi temporal. Penelitian ini memberikan wawasan mendalam mengenai kekuatan dan kelemahan masing-masing arsitektur, serta menjadi acuan dalam pemilihan model untuk aplikasi keamanan multimedia.

**Kata kunci:** Deteksi *Deepfake*, Keamanan Multimedia, 3D CNN, *Video Vision Transformer*, 3D ConvNeXt

## ABSTRACT

The rapid advancement of Generative AI technology has introduced a serious threat in the form of deepfake videos, which hold significant potential for spreading misinformation and other malicious activities. The ability to accurately and efficiently detect deepfake videos has become crucial for maintaining digital security and public trust. This research aims to conduct a comparative performance analysis of three modern deep learning architectures 3D Convolutional Neural Network (3D CNN), Video Vision Transformer (ViT), and 3D ConvNeXt for the task of deepfake video classification. The three models were trained and evaluated using the industry-standard dataset, FaceForensics++, with evaluation metrics including accuracy, precision, recall, and F1-Score. Experimental results indicate that the Video Vision Transformer architecture achieved the highest accuracy, albeit at the most significant computational cost. Meanwhile, 3D ConvNeXt offered the best balance between accuracy performance and inference efficiency, making it a promising candidate for implementation in real-world detection systems. This study provides in-depth insights into the strengths and weaknesses of each architecture, serving as a reference for model selection in multimedia security applications.

**Keyword:** Deepfake Detection, Multimedia Security, 3D CNN, Video Vision Transformer, 3D ConvNeXt