Jurnal Rekayasa Sistem Informasi dan Teknologi
Volume 3, No 4 – Mei 2026
e-ISSN : 3025-888X

DETEKSI DEEPFAKE: BERDASARKAN INKONSISTENSI GERAKAN KEPALA DAN LEHER MENGGUNAKAN VARIATIONAL AUTOENCODER

Muhammad Fikri Maulana[1*], Bagus Satrio Waluyo Poetro[2],
[1,2] Teknik Informatika, Universitas Islam Sultan Agung
Fikrimaulana18122003@gmail.com[1*] Bagusswp@unissula.ac.id[2]

Received: 25-04- 2026	Revised: 20-05-2026	Approved: 26-05-2026

ABSTRAK

Teknologi deepfake saat ini mampu memproduksi manipulasi video yang sangat realistis, sehingga sulit dideteksi oleh metode berbasis spasial statis. Penelitian ini mengusulkan metode deteksi deepfake melalui analisis inkonsistensi biomekanik gerakan kepala dan wajah menggunakan Variational Autoencoder (VAE). Dengan pendekatan unsupervised anomaly detection, model dilatih hanya menggunakan video asli untuk mempelajari pola pergerakan alami manusia. Lima fitur utama diekstraksi menggunakan MediaPipe, yakni Pitch, Yaw, Roll, Eye Aspect Ratio (EAR), dan Mouth Aspect Ratio (MAR), yang diproses dengan teknik sliding window 60 frame. Hasil pengujian menggunakan dataset FaceForensics++ dan Celeb-DF menunjukkan performa model dengan akurasi 66% dan nilai AUC sebesar 0,69. Penelitian ini membuktikan bahwa inkonsistensi gerakan biomekanik merupakan indikator kuat dalam mendeteksi manipulasi video sintesis tingkat lanjut.
Kata kunci: Biomekanik, Deepfake, Deteksi Anomali, Variational Autoencoder.

PENDAHULUAN

Kemajuan pesat teknologi *deep learning* seperti GANs dan VAEs telah mendorong realisme *deepfake* ke tingkat yang mengancam keamanan informasi global [1], [2]. Sayangnya, metode deteksi konvensional berbasis klasifikasi biner (*supervised learning*) yang mengandalkan pengamatan visual tingkat piksel sering kali gagal dan rentan mengalami *overfitting* saat menghadapi jenis manipulasi baru atau *unseen attacks* [3], [4]. Detektor yang hanya berfokus pada fitur spasial statis (bingkai per bingkai) kini menjadi kurang efektif karena algoritma *deepfake* modern mampu menghilangkan cacat visual dengan sangat halus [5]. Oleh karena itu, literatur terkini menuntut pergeseran paradigma deteksi dari analisis gambar statis menuju analisis koherensi spatiotemporal [6], [7].

Merespons tantangan tersebut, kelemahan fundamental *deepfake* generasi terbaru kini sering kali terungkap pada aspek biomekanik dan konsistensi fisik. Meskipun tekstur wajah sintetis tampak sangat realistis, generator AI sering kali gagal mempertahankan konsistensi orientasi wajah (*face orientation*) secara wajar saat subjek bergerak [8]. Ketidaksinkronan sinyal fisiologis dan gerakan antara wajah buatan dengan wilayah leher (*neck region*) asli target telah terbukti menjadi indikator anomali yang sangat kuat [9]. Selain itu, celah biomekanik juga sering terekspos melalui inkonsistensi pada fusi *landmark* wajah seperti mata, hidung, dan mulut yang gagal ditiru secara presisi oleh generator[10].

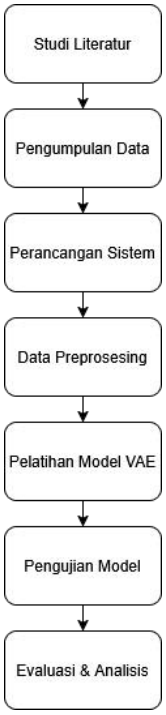
Untuk mengeksploitasi celah biomekanik tersebut, penelitian ini mengusulkan pendekatan *Unsupervised Anomaly Detection* menggunakan *Variational Autoencoder* (VAE). Pendekatan ini secara eksklusif memodelkan distribusi data video asli, sehingga sistem tetap tangguh terhadap metode manipulasi yang belum pernah dilihat sebelumnya [11], [12], [13]. Berbeda dengan penelitian terdahulu yang umumnya menggunakan arsitektur *Convolutional Neural Network* (CNN) untuk citra, model VAE yang diusulkan memanfaatkan *Dense Layer* untuk memproses data spatiotemporal berupa matriks fitur biomekanik wajah.

Pemilihan *Dense Layer* didasarkan pada argumen teknis bahwa fitur yang diinputkan sudah berupa representasi numerik abstrak tingkat tinggi hasil ekstraksi, sehingga operasi konvolusi spasial CNN tidak lagi diperlukan dan justru tidak efisien untuk data non-citra. *Dense Layer* memungkinkan model untuk langsung menangkap korelasi global serta interdependensi non-linier antar-variabel biomekanik yang heterogen secara langsung, dengan beban komputasi yang jauh lebih ringan. Fitur ini diekstraksi menggunakan MediaPipe, yang mencakup parameter orientasi kepala (*Pitch, Yaw, Roll*) serta rasio bukaan mata (EAR) dan mulut (MAR) secara sekuensial menggunakan teknik *Sliding Window*.

Kebaruan (*novelty*) utama dalam penelitian ini terletak pada perumusan penentuan skor anomali. Karena karakteristik video *deepfake* umumnya memiliki pergerakan wajah yang kaku, statis, dan minim dinamika fisiologis, model VAE justru sangat mudah merekonstruksinya, yang ironisnya menghasilkan nilai *Mean Squared Error* (MSE) yang rendah. Menyadari paradoks tersebut, penelitian ini mentransformasikan nilai MSE menjadi *Inverse Score* (1/MSE). Melalui formulasi matematis ini, video *deepfake* dengan pergerakan yang kaku akan menghasilkan nilai anomali yang melonjak secara eksponensial, sehingga dapat secara presisi melampaui batas ambang (*threshold*) pergerakan normal manusia. Kontribusi ilmiah utama dari penelitian ini adalah untuk Mengembangkan kerangka kerja deteksi anomali tanpa pengawasan (*unsupervised*) berbasis pergerakan spatiotemporal kepala dan leher yang adaptif terhadap arsitektur generator *deepfake* masa depan (*zero-day attacks*), mengimplementasikan arsitektur VAE berbasis *Dense Layer* yang ringan dan efisien untuk memodelkan interdependensi fitur biomekanik abstrak tanpa membutuhkan beban komputasi besar seperti model berbasis konvolusi citra, dan juga mengusulkan formulasi fungsi skor anomali baru menggunakan metode inversi (*Inverse Score*) untuk membalikkan paradoks rendahnya *reconstruction error* pada video manipulasi yang kaku menjadi sinyal deteksi yang kuat.

METODE PENELITIAN

Untuk mencapai tujuan penelitian secara sistematis dan terstruktur, penelitian ini dilaksanakan mengikuti alur kerja (*framework*) yang terdiri dari beberapa tahapan utama yaitu studi literatur, Pengumpulan data, perancangan sistem, data preprocessing, pelatihan model variational autoencoder pengujian model evaluasi dan analisis. Tahapan-tahapan tersebut digambarkan dalam diagram alir ditunjukkan pada gambar 1



Gambar 1 Alur Penelitian

Pengumpulan Data

Gambar 2 Data Bersih dan Data Deepfakes FaceForensics++_C23

Pada Gambar 2 bisa terlihat bahwa yang kiri adalah data bersih dan yang kanan adalah data *Deepfakes*. Dataset FaceForensics++ versi c23 digunakan sebagai sumber data primer yang terbagi menjadi data bersih (*original sequences*) dan data manipulasi (*Deepfakes*)

Gambar 3 FaceShifter dan FaceSwap FaceForensics++_C23

Pada Gambar 3 yang kiri adalah *FaceShifter* dan yang kanan adalah *FaceSwap*. Dalam dataset FaceForensics++ versi c23, kategori *FaceSwap* dan *FaceShifter* merepresentasikan metode manipulasi pertukaran identitas (*identity swap*) yang digunakan dalam skenario pengujian internal untuk mengevaluasi keandalan deteksi model.

Sumber data primer dalam penelitian ini adalah *FaceForensics++* (FF++) dengan tingkat kompresi *c23 (High Quality/Light Compression)*. Versi c23 dipilih

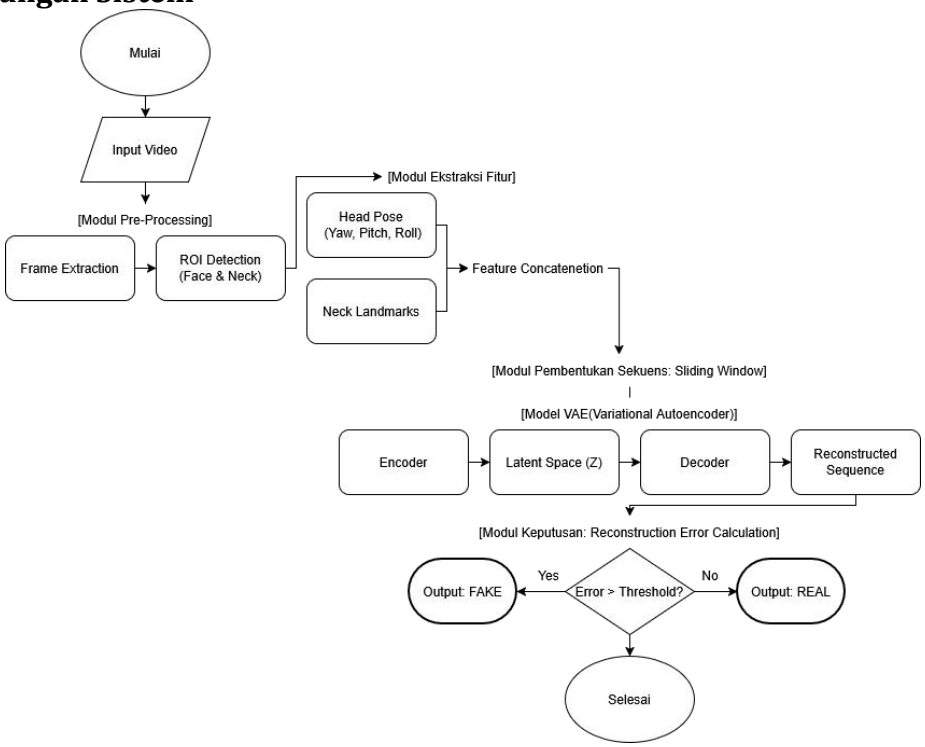
karena menawarkan keseimbangan optimal antara kualitas visual yang mendekati video asli dan efisiensi penyimpanan dibandingkan versi *c0 (Raw)* yang tidak terkompresi.

Sebagai data uji sekunder untuk mengukur ketahanan (*robustness*), digunakan Celeb-DF v2. Dataset ini memiliki kualitas visual *State-of-the-Art* dengan minimnya artefak visual kasar (*ghosting* atau *jitter*) yang sering ditemukan pada dataset generasi lama.

Gambar 4 *Data real* dan *synthesis* Celeb DF(v2)

Pada Gambar 4 yang kiri adalah data *real* dan yang kanan adalah data *synthesis*. Dataset Celeb-DF (v2) difungsikan sebagai data sekunder untuk pengujian generalisasi, yang mencakup video asli (*Celeb-real*) dan video manipulasi (*Celeb-synthesis*) dengan kualitas visual tingkat lanjut (*State-of-the-Art*) dan minim artefak visual kasar

Perancangan Sistem



Gambar 5 Perancangan Sistem

Pada Gambar 5 *flowchart* ini menjelaskan tentang perancangan sistem. Sistem deteksi deepfake ini diawali dengan memproses input video melalui tahap *pre-processing* untuk mengekstraksi frame dan mendeteksi area wajah serta leher (*Region of Interest*), yang kemudian dilanjutkan dengan ekstraksi fitur *head pose* dan *neck landmarks* untuk menangkap pola gerakan fisik. Fitur-fitur tersebut disusun menjadi data sekuensial menggunakan metode *sliding window* agar memuat konteks waktu, lalu dimasukkan ke dalam model *Variational Autoencoder* (VAE) yang bertugas mempelajari dan merekonstruksi pola gerakan normal dari

video asli. Penentuan keaslian video dilakukan di tahap akhir dengan menghitung *reconstruction error* antara input dan hasil output VAE apabila nilai error melebihi ambang batas (*threshold*) yang ditentukan, video diklasifikasikan sebagai palsu (*Fake*) karena adanya inkonsistensi gerakan, sedangkan error yang rendah menandakan video tersebut asli (*Real*).

Strategi Pembagian dan Pra-pemrosesan Data

1. Data Pelatihan (*Training Set*)
   - Komposisi Fase pelatihan menggunakan total 1.440 video asli (*real videos*), yang terdiri dari 925 video asli dari *dataset* FaceForensics++ dan 515 video asli dari Celeb-DF v2. Sama sekali tidak ada sampel video manipulasi (*fake*) yang dimasukkan ke dalam fase ini.
   - Pra-pemrosesan (*Sliding Window*) untuk menjaga efisiensi komputasi dan konsistensi dimensi temporal, video asli tidak diinputkan secara utuh. Matriks fitur biomekanik yang telah diekstraksi (EAR, MAR, *Head Pose*) dipotong menggunakan teknik *sliding window* dengan durasi 60 *frame* secara berkelanjutan.
   - Tujuan penggunaan 1.440 video asli dari dua sumber yang berbeda bertujuan untuk memastikan model VAE mampu mempelajari, merepresentasikan, dan menggeneralisasi pola pergerakan kepala serta dinamika wajah manusia yang natural dalam berbagai variasi resolusi dan pencahayaan.
   - Komposisi data validasi diambil 10% dari total 1.440 video asli secara acak yang telah disiapkan pada tahap pelatihan yaitu 144 video asli dari dataset FaceForensics++ maupun dari dataset Celeb-DF v2.
   - Fungsi kontrol data ini digunakan sebagai instrumen pemantau konvergensi *loss function* selama *epoch* pelatihan berjalan, serta menjadi basis distribusi *reconstruction error* untuk menentukan titik ambang batas (*threshold*) anomali yang optimal
2. Data Pengujian (*Testing Set*)
   - Tahap pengujian dilakukan untuk mengukur kemampuan deteksi model menggunakan 300 video yang belum pernah dilihat oleh model sebelumnya. Pengujian ini dirancang secara seimbang (150 video asli dan 150 video manipulasi) dengan rincian sebagai berikut
   - Pengujian Video Asli (150 sampel): Terdiri dari 75 video asli dari FaceForensics++ dan 75 video asli dari Celeb-DF v2.
   - Pengujian Video Palsu (150 sampel): Terdiri dari 113 video manipulasi dari FaceForensics++ (dengan rincian spesifik: 38 video *Deepfakes*, 38 video *FaceShifter*, dan 37 video *FaceSwap*) serta 37 video manipulasi dari Celeb-DF v2.
   - Hipotesis Pengujian: Model diharapkan mampu mengklasifikasikan 150 video asli dengan *Reconstruction Error* (MSE) yang selaras dengan *latent space* normal, sementara 150 video *deepfake* akan menghasilkan *error* rekonstruksi yang rendah akibat kekakuan pergerakan. Nilai MSE ini kemudian dikonversi menjadi *Inverse Score* (1/MSE) yang akan melonjak tinggi melampaui *threshold*, sehingga terklasifikasi sebagai anomali/palsu.

Ekstraksi Fitur Biomekanik dan Temporal

1. *Eye Aspect Ratio* (EAR)
   EAR diekstraksi untuk mengukur rasio bukaan kelopak mata, yang merepresentasikan dinamika kedipan mata subjek. Generator *deepfake* sering kali menghasilkan pola kedipan yang tidak wajar atau tidak sinkron. Berdasarkan penelitian [10] , metrik EAR dihitung menggunakan 6 titik koordinat *landmark* mata dengan persamaan matematis berikut:

$$EAR = \frac{||P_2 - P_6|| + ||P_3 - P_5||}{2||P_1 - P_4||}$$

   Di mana P_1 ,..., P_6 merupakan titik koordinat 2D pada area mata, pembilang menghitung jarak vertikal antara kelopak mata atas dan bawah, sedangkan penyebut menghitung jarak horizontal antara sudut mata kiri dan kanan. Nilai EAR ini dihitung untuk mata kiri dan kanan, lalu diambil nilai rata-ratanya.
2. *Mouth Aspect Ratio* (MAR)
   MAR digunakan untuk mengukur rasio bukaan mulut yang menangkap dinamika pergerakan bibir, terutama saat subjek berbicara. Inkonsistensi antara suara dan pergerakan bibir merupakan salah satu artefak utama pada *deepfake*. MAR dihitung dengan prinsip jarak Euclidean yang serupa dengan EAR, diformulasikan sebagai:

$$MAR = \frac{||P_2 - P_8|| + ||P_3 - P_7|| + ||P_4 - P_6||}{2||P_1 - P_5||}$$

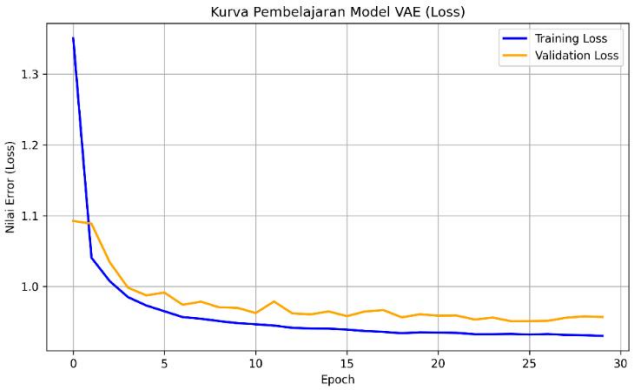
   Di mana pembilang mewakili jarak vertikal antara bibir atas dan bawah pada beberapa titik tengah, dan penyebut mewakili jarak horizontal antara ujung bibir kiri dan kanan.
3. *Head Pose Estimation (Pitch, Yaw, Roll)*
   Untuk mendeteksi inkonsistensi pergerakan antara kepala dan leher—yang merupakan fokus utama penelitian ini—dilakukan estimasi orientasi kepala 3D (*Head Pose*). Mengacu pada temuan [8] , algoritma *Perspective-n-Point* (PnP) diterapkan pada titik *landmark* utama (seperti ujung hidung, dagu, sudut mata, dan sudut mulut) untuk menghitung tiga sudut rotasi kepala:
   - Pitch: Gerakan mengangguk (rotasi pada sumbu X).
   - Yaw: Gerakan menoleh ke kiri/kanan (rotasi pada sumbu Y).
   - Roll: Gerakan memiringkan kepala (rotasi pada sumbu Z).
4. Pembentukan Matriks Spatiotemporal (*Sliding Window*)
   Ekstraksi per bingkai (*frame*) di atas menghasilkan sebuah vektor berisikan 5 nilai metrik *(EAR, MAR, Pitch, Yaw, Roll).* Namun, anomali pergerakan tidak dapat dideteksi hanya dari satu bingkai statis. Oleh karena itu, penelitian ini menerapkan teknik segmentasi *sliding window* dengan ukuran jendela *(window size)* sebanyak 60 *frame* berkelanjutan.
   Melalui teknik ini, aliran gerakan wajah selama kurang lebih 2 detik direpresentasikan menjadi sebuah matriks spatiotemporal berdimensi 60 X 5 Matriks sekuensial inilah yang kemudian diumpankan (*fed*) ke dalam model *Dense-Variational Autoencoder* (Dense-VAE) untuk dipelajari pola keluwesannya dan dihitung nilai *Reconstruction Error*-nya.

Evaluasi Model

Evaluasi model dilakukan secara komprehensif menggunakan 300 video pengujian untuk mengukur seberapa baik model dapat memisahkan antara video asli dan video *deepfake* (anomali) menggunakan ambang batas tertentu.

HASIL DAN PEMBAHASAN

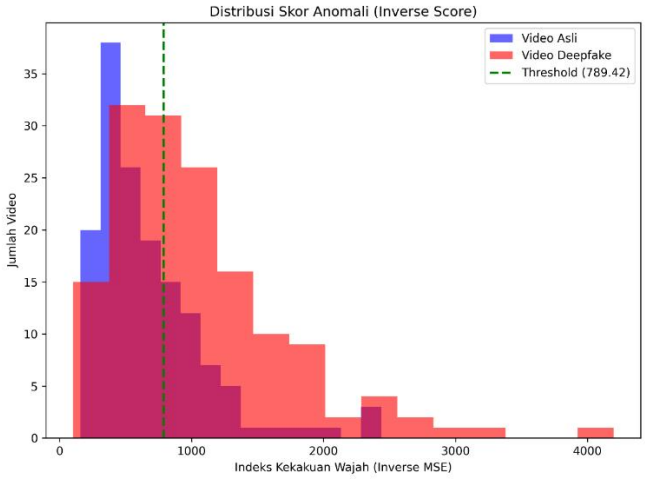
Pelatihan Model



Gambar 6 *Kurva Loss Model* VAE

Pada Gambar 6 Model VAE dilatih hanya menggunakan 1.440 video asli agar sistem dapat mempelajari secara mendalam distribusi gerakan manusia yang wajar dan natural. Evaluasi proses pembelajaran ini divisualisasikan melalui Kurva *Loss.* Pada grafik tersebut, terlihat bahwa *Training Loss* (garis biru) dan *Validation Loss* (garis oranye) mengalami penurunan tajam pada *epoch* awal. Model mencapai stabilitas konvergensi saat memasuki *epoch* ke-10 hingga akhir, di mana garis latih dan validasi bergerak beriringan tanpa adanya jarak (*gap*) yang melebar. Hal ini membuktikan bahwa pelatihan berjalan optimal, model mampu merekonstruksi gerakan asli dengan galat yang sangat kecil, dan terhindar dari indikasi *overfitting*.

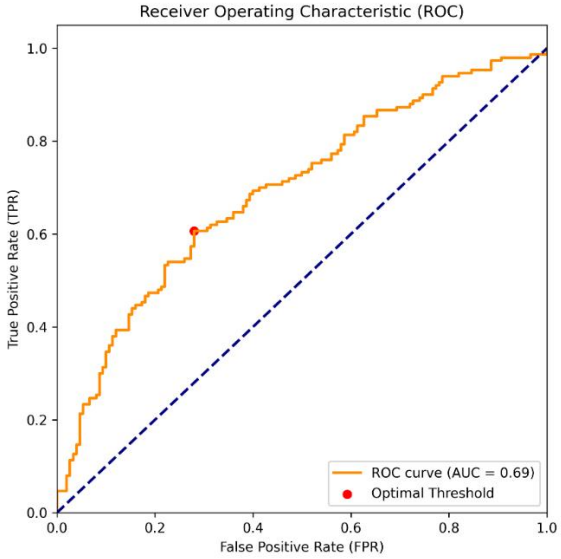
Distribusi Skor dan Penentuan *Threshold*



Gambar 7 Distribusi Skor Anomali *(Inverse Score)*

Pada Gambar 7 adalah Penentuan apakah sebuah video diklasifikasikan sebagai *deepfake* didasarkan pada Indeks Kekakuan Wajah (yang diturunkan dari nilai *Reconstruction Error*). Distribusi skor menunjukkan pemisahan antara kelas asli dan palsu. Ambang batas (*threshold*) optimal ditetapkan pada nilai 789.42. Sampel dengan skor di bawah ambang batas ini diidentifikasi sebagai pergerakan wajar (Asli), sedangkan sampel dengan skor 789.42 diklasifikasikan sebagai anomali (*Deepfake*).

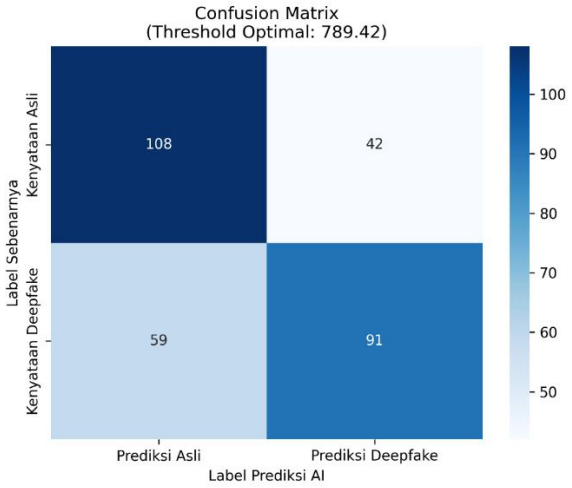
Kurva Receiver Operating Characteristic (ROC)



Gambar 8 Kurva ROC-AUC

Pada Gambar 8 Kemampuan model dalam membedakan kedua kelas dievaluasi melalui kurva ROC. Kurva ini menghasilkan nilai *Area Under Curve* (AUC) sebesar 0.69. Angka ini merepresentasikan kapasitas diskriminatif model dalam mendeteksi anomali pergerakan lintas metode manipulasi (*FaceSwap, FaceShifter, Deepfakes*, dll) di atas probabilitas acak, memvalidasi bahwa pendekatan biomekanik memberikan sinyal (*cue*) forensik yang berharga.

Confusion Matrix dan Laporan Klasifikasi



Gambar 9 *Confusion Matrix*

Pada Gambar 9 Hasil akhir pengujian pada 300 sampel video divisualisasikan pada *Confusion Matrix*. Sistem berhasil memprediksi *True Negative* (video asli yang diprediksi benar) sebanyak 108 sampel dan *True Positive* (video *deepfake* yang diprediksi benar) sebanyak 91 sampel.

Untuk mengukur kinerja aktual dari model klasifikasi berbasis *Variational Autoencoder* (VAE) secara mendetail, hasil prediksi dari 300 sekuens data uji (terdiri dari 150 sekuens Asli dan 150 sekuens *Deepfake*) dievaluasi menggunakan *Confusion Matrix*. Matriks ini menggunakan ambang batas (*threshold*) optimal sebesar 789.42 yang telah didapatkan pada tahap analisis sebelumnya.

Analisis Laporan Klasifikasi (*Classification Report*)

Tabel 1 Laporan Klasifikasi Akhir

Kelas	Precision	Recall	F1-Score
Real (Asli)	0.65	0.72	0.68
Deepfake	0.68	0.61	0.64
Accuracy			0.66
Macro Avg	0.66	0.66	0.66
Weighted Avg	0.67	0.66	0.66

Berdasarkan Tabel 1 laporan klasifikasi, model mencapai tingkat Akurasi keseluruhan sebesar 66%. Model menunjukkan performa yang sedikit lebih superior dalam mengenali kelas video Asli dengan *Recall* sebesar 72%. Meskipun mendeteksi 150 sampel *deepfake* dari berbagai jenis algoritma manipulasi merupakan tantangan yang sangat sulit (*highly challenging*), pencapaian skor presisi 68% dan F1-Score 66% pada kelas *deepfake* membuktikan bahwa inkonsistensi temporal yang dianalisis oleh VAE terbukti efektif sebagai parameter deteksi.

Untuk melengkapi hasil evaluasi dari *Confusion Matrix*, kinerja model VAE dalam mendeteksi anomali pergerakan juga diukur menggunakan metrik evaluasi standar klasifikasi mesin pembelajaran, yaitu *Precision, Recall, F1-Score*, dan *Accuracy*. Pengujian dilakukan pada total 300 sampel (150 sampel video Asli dan 150 sampel video *Deepfake*).

Tabel 2 Sebagai perbandingan kinerja deteksi *deepfake* dengan Metode Terkini

Metode / Referensi	Pendekatan	Fitur Utama	Dataset	AUC (%)
Lord Sen dkk. (2026)	Supervised	Swin Transformer + BERT (Cross-Attention)	FF++	99.80
Lord Sen dkk. (2026)	Supervised	Swin Transformer + BERT (Cross-Attention)	Celeb-DF	99.88
Yunzhuo Chen dkk. (2025)	Supervised	Spatio-Temporal Consistency & Attention	FF++	98,10
Metode Usulan (VAE)	Unsupervised	Biomechanical (Head Pose, EAR, MAR) + Inverse Score	FF++ / Celeb-DF	69.0

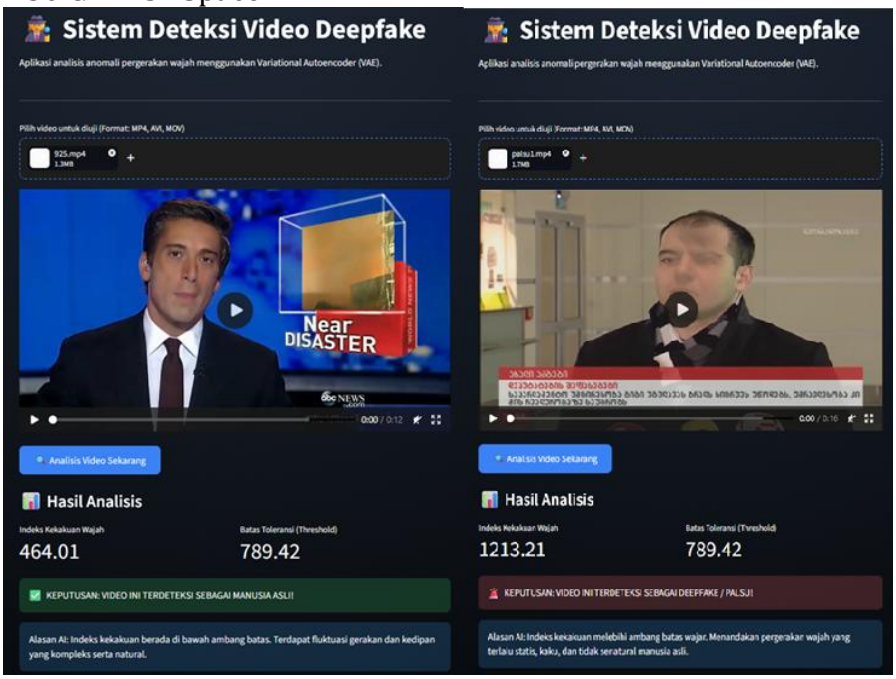
Berdasarkan Tabel 2 tingginya nilai AUC pada metode SOTA sangat dipengaruhi oleh penggunaan pendekatan klasifikasi biner (*supervised*), di mana model dilatih menggunakan data *deepfake* secara langsung. Pendekatan ini memicu risiko *overfitting* yang tinggi model cenderung menghafal artefak visual spesifik dari algoritma pembuat *deepfake* pada dataset latih, sehingga kinerjanya terancam anjlok secara drastis ketika dihadapkan pada jenis manipulasi baru di dunia nyata (*zero-day attacks*). Sebaliknya, metode usulan bersifat *unsupervised*—hanya dilatih menggunakan data manusia normal. Pencapaian AUC 69,0% tanpa model pernah melihat satupun data *deepfake* saat fase pelatihan membuktikan bahwa sistem secara mandiri mampu mengidentifikasi anomali pergerakan baru, memberikan tingkat generalisasi yang lebih logis dan aman terhadap evolusi generator AI di masa depan.

Perbandingan ini menyoroti *trade-off* (tarik-ulur) yang signifikan antara efisiensi komputasi dan akurasi. Metode seperti Swin Transformer [14]

memproses anomali piksel dalam ruang spasial berdimensi raksasa yang membutuhkan sumber daya komputasi (*GPU power*) yang sangat masif. Sebagai alternatif yang jauh lebih ringan (*lightweight*), metode usulan memanfaatkan *Dense Layer* untuk memproses ekstraksi fitur biomekanik abstrak tingkat tinggi (orientasi kepala, EAR, dan MAR) yang volumenya sangat kecil. Oleh karena itu, capaian 69,0% ini menegaskan kelayakan formulasi *Inverse Score* dalam membalikkan paradoks rekonstruksi menjadi sinyal anomali. Hal ini memberikan landasan empiris yang kuat bahwa inkonsistensi spatiotemporal biomekanik dapat dieksploitasi sebagai indikator *deepfake* yang valid tanpa harus mengorbankan efisiensi komputasi sistem.

Deployment Model

Platform *Deployment* dan *Hosting* menggunakan Hugging Face (khususnya fitur Hugging Face Spaces), digunakan sebagai infrastruktur *cloud* untuk mengimplementasikan model yang telah dilatih ke dalam bentuk aplikasi berbasis web (*web-based application*). Pemilihan Hugging Face didasarkan pada kemampuannya dalam menyediakan lingkungan *server* yang stabil untuk menjalankan model *Machine Learning*, serta memfasilitasi antarmuka pengguna (UI) yang interaktif agar sistem deteksi *deepfake* dapat diakses dan diuji coba secara publik secara *real-time*, dan juga Hugging Face Spaces (Free Tier) memberikan alokasi yang sangat besar untuk ukuran gratis, yaitu 16 GB RAM, 2 vCPU, dan 50 GB Disk Space.



Gambar 10 Prediksi video asli dan palsu *(Website)*

Pada Gambar 10 yang kiri adalah prediksi video asli dan yang kanan adalah prediksi video palsu yang sudah dideploy di website ini memungkinkan interaksi langsung untuk menganalisis anomali pergerakan wajah menggunakan arsitektur *Variational Autoencoder* (VAE).

KESIMPULAN

Penelitian ini berhasil membuktikan bahwa optimalisasi desain deteksi *deepfake* dapat dicapai secara efisien tanpa bergantung pada piksel gambar mentah, melainkan melalui ekstraksi koordinat 3D *landmark* wajah dan leher menggunakan MediaPipe dengan metode *sliding window* 60 frame. Signifikansi hasil riil menunjukkan bahwa model VAE yang dilatih secara *unsupervised* murni menggunakan 1.440 video asli mampu beroperasi secara stabil dengan *validation loss* konvergen di angka ~0.95, serta berhasil diimplementasikan menjadi sistem *end-to-end* interaktif di Hugging Face Spaces. Meskipun akurasi keseluruhan saat ini berada di angka 66%, capaian ini memberikan validasi empiris penting bahwa formulasi *Inverse Score* layak digunakan sebagai indikator awal untuk mendeteksi inkonsistensi biomekanik pada *zero-day attacks* secara ringan (*lightweight*). Namun, batasan utama sistem ini terletak pada karakteristik model VAE konvensional yang rentan mengalami *over-smoothing* saat merekonstruksi data spatiotemporal, sehingga variasi gerakan fisiologis yang terlalu halus terkadang gagal memicu lonjakan skor anomali melampaui ambang batas (*threshold*). Sebagai saran optimasi untuk penelitian selanjutnya, disarankan untuk mengintegrasikan arsitektur temporal yang lebih kompleks seperti LSTM-VAE atau *Transformer-based Autoencoder*, serta menerapkan mekanisme *adaptive thresholding* guna meningkatkan akurasi dan sensitivitas sistem deteksi.

DAFTAR PUSTAKA

[1] Dr. J. Upadhyay, V. Chaudhari, R. Darpe, R. Desai, and S. Gujar, "Deepfake Detection: A Comprehensive Review of Techniques and Challenges," vol. Volume 7, Issue 2, Apr. 2025, doi: 10.36948/ijfmr.2025.v07i02.41306.

[2] Y. Chen, N. Akhtar, N. A. H. Haldar, and A. Mian, "Deepfake Detection with Spatio-Temporal Consistency and Attention," pp. 1–8, Nov. 2022, doi: 10.1109/DICTA56598.2022.10034609.

[3] S. Belguesmia, M. S. Allili, and A. Hamadene, "Unmasking Facial DeepFakes: A Robust Multiview Detection Framework for Natural Images," May 2025, doi: https://doi.org/10.21428/d82e957c.879d9210.

[4] G. Pei *et al.*, "Deepfake Generation and Detection: A Benchmark and Survey," *ACM Comput. Surv.*, vol. 58, no. 11, pp. 1–41, Aug. 2026, doi: 10.1145/3801962.

[5] Y. Luo, Y. Zhang, J. Yan, and W. Liu, "Generalizing Face Forgery Detection with High-frequency Features," *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2021, doi: 10.1109/CVPR46437.2021.01605.

[6] M. A. Amin, Y. Hu, and J. Hu, "Analyzing temporal coherence for deepfake video detection," *Electronic Research Archive*, vol. 32, no. 4, pp. 2621–2641, 2024, doi: 10.3934/ERA.2024119.

[7] Z. Gu *et al.*, "Spatiotemporal Inconsistency Learning for DeepFake Video Detection," in *MM 2021 - Proceedings of the 29th ACM International Conference on Multimedia*, Association for Computing Machinery, Inc, Oct. 2021, pp. 3473–3481. doi: 10.1145/3474085.3475508.

[8] C.-M. Rosca and A. Stancu, "AI Anomaly-Based Deepfake Detection Using Customized Mahalanobis Distance and Head Pose with Facial Landmarks," vol. Vol 15, Issue 17, 2025, doi: 10.3390/app15179574.

[9] B. S. An, H. Lim, H. A. Seong, and E. C. Lee, "Facial and Neck Region Analysis for Deepfake Detection Using Remote Photoplethysmography Signal Similarity," *IET Biom.*, vol. 2024, no. 1, 2024, doi: 10.1049/bme2/7095412.

[10] S. Sohail, S. M. Sajjad, A. Zafar, Z. Iqbal, Z. Muhammad, and M. Kazim, "Deepfake Image Forensics for Privacy Protection and Authenticity Using Deep Learning," *Information (Switzerland)*, vol. 16, no. 4, Apr. 2025, doi: 10.3390/info16040270.

[11] R. Yao, Z. Bai, J. Tong, and K. Rezaee, "Deepfake Detection in Image Sequences: A Temporal Approach for Anomaly Detection," *International Journal of Intelligent Systems*, vol. 2025, no. 1, 2025, doi: 10.1155/int/8566328.

[12] Y. Tian *et al.*, "Unsupervised Anomaly Detection in Medical Images with a Memory-augmented Multi-level Cross-attentional Masked Autoencoder," vol. 14349 LNCS, 2023, doi: 10.1007/978-3-031-45676-3_2.

[13] H. H. Nguyen, C. A. Nguyen, X. T. Dao, Q. T. Duong, D. P. T. Kim, and M.-T. Pham, "Variational Autoencoder for Anomaly Detection: A Comparative Study," *ArXiv preprint*, vol. abs/2408.13561, 2024, doi: 10.48550/arXiv.2408.13561.

[14] Lord Sen and S. Mukherjee, "A Novel Unified Approach to Deepfake Detection of Images," in *2024 12th International Conference on Intelligent Systems and Embedded Design (ISED)*, 2024, pp. 1–6. doi: 10.1109/ISED63599.2024.10957205.

[15] K. N. Ramadhani, R. Munir, and N. P. Utama, "Improving Video Vision Transformer for Deepfake Video Detection Using Facial Landmark, Depthwise Separable Convolution and Self Attention," *IEEE Access*, vol. 12, pp. 8932–8939, 2024, doi: 10.1109/ACCESS.2024.3352890.

[16] Y. Wang, T. Liu, J. Zhou, and J. Guan, "Video anomaly detection based on spatio-temporal relationships among objects," *Neurocomputing*, vol. 532, pp. 141–151, May 2023, doi: 10.1016/j.neucom.2023.02.027.

[17] E. Tchaptchet, E. F. Tagne, J. Acosta, D. B. Rawat, and C. Kamhoua, "Deepfakes Detection by Iris Analysis," *IEEE Access*, vol. 13, pp. 8977–8987, Jan. 2025, doi: 10.1109/ACCESS.2025.3527868.

[18] S. M. Yasir and H. Kim, "Lightweight Deepfake Detection Based on Multi-Feature Fusion," Feb. 2025, doi: 10.3390/app15041954.

[19] K. P. Rahatwal, S. Pundir, M. Wazid, and V. Bhat K., "A Novel Approach to Deepfake Detection: Leveraging Fused Facial and Body Dynamics With a CNN–Transformer Hybrid Network," *IEEE Access*, vol. 13, pp. 197085–197108, 2025, doi: 10.1109/ACCESS.2025.3632155.

[20] D. W. Deressa, H. Mareen, P. Lambert, S. Atnafu, Z. Akhtar, and G. Van Wallendael, "GenConViT: Deepfake Video Detection Using Generative Convolutional Vision Transformer," *Applied Sciences (Switzerland)*, vol. 15, no. 12, Jun. 2025, doi: 10.3390/app15126622.

[21] Y. Xu, J. Liang, L. Sheng, and X.-Y. Zhang, "Learning Spatiotemporal Inconsistency via Thumbnail Layout for Face Deepfake Detection," *Int. J. Comput. Vis.*, vol. 132, no. 12, pp. 5663–5680, 2024, doi: 10.1007/s11263-024-02054-2.

[22] C. Feng, Z. Chen, and A. Owens, "Self-Supervised Video Forensics by Audio-Visual Anomaly Detection," in *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2023, pp. 10491–10503. doi: 10.1109/CVPR52729.2023.01011.