

Platform Deteksi Deepfake pada Wajah Manusia dalam Konten Visual Menggunakan Autoencoder dan Dashboard Analitik untuk Verifikasi di Kalangan Jurnalis

Abstrak

Maraknya penyebaran konten *deepfake* secara daring telah menjadi ancaman serius terhadap kredibilitas media dan integritas informasi, khususnya dalam bidang jurnalistik. Berdasarkan pengalaman magang penulis, jurnalis kerap mengalami kesulitan dalam membedakan konten asli dengan hasil manipulasi, sehingga proses verifikasi menjadi kurang efisien. Penelitian ini bertujuan untuk merancang dan mensimulasikan sebuah platform deteksi *deepfake* berbasis *Autoencoder* yang terintegrasi dengan *dashboard* analitik guna mendukung proses verifikasi konten visual di lingkungan redaksi. Metodologi penelitian meliputi perancangan arsitektur sistem berbasis *microservices*, pengembangan model *Autoencoder* untuk deteksi anomali wajah menggunakan dataset publik (FaceForensics++ dan DFDC), serta perancangan antarmuka *dashboard* sebagai alat visualisasi hasil deteksi. Hasil penelitian menunjukkan bahwa model *Autoencoder* pada tahap simulasi awal mampu mencapai akurasi deteksi sekitar 92,5% pada dataset uji, dengan waktu analisis rata-rata kurang dari 5 detik untuk gambar dan 30 detik untuk video berdurasi pendek. Evaluasi terhadap pengguna (*User Acceptance Test*) belum dilakukan, sehingga tingkat kepuasan pengguna yang disebutkan masih merupakan target capaian pada tahap implementasi berikutnya. Secara keseluruhan, rancangan sistem ini menunjukkan potensi yang signifikan dalam membantu jurnalis memverifikasi keaslian konten visual secara cepat dan akurat, serta menjadi landasan untuk pengembangan platform deteksi *deepfake* yang lebih komprehensif di masa mendatang.

1. Pendahuluan

1.1 Latar Belakang

Perkembangan teknologi kecerdasan buatan (*Artificial Intelligence / AI*) telah memunculkan fenomena *deepfake*, yaitu teknik manipulasi konten visual berupa gambar atau video yang dihasilkan melalui model *deep learning*, sehingga tampak sangat realistis dan sulit dibedakan dari konten asli. Fenomena ini menimbulkan ancaman serius terhadap integritas informasi, terutama dalam ranah jurnalistik, di mana kepercayaan publik merupakan fondasi utama kredibilitas media.

Berdasarkan pengalaman penulis selama menjalani magang di sebuah redaksi media, jurnalis sering menghadapi tantangan dalam proses verifikasi konten visual. Ketidakpastian terhadap keaslian suatu video atau gambar dapat menyebabkan misinformasi yang berpotensi menurunkan kredibilitas lembaga pers. Kondisi tersebut menunjukkan perlunya dukungan sistem berbasis teknologi yang mampu membantu proses verifikasi secara cepat, akurat, dan transparan.

Berbagai penelitian terkini mengenai deteksi *deepfake* umumnya berfokus pada pengembangan algoritma dan model *deep learning* dengan tingkat akurasi tinggi, seperti *Convolutional Neural Network* (CNN) dan *Transformer-based models*. Meskipun pendekatan tersebut menunjukkan hasil yang menjanjikan pada dataset akademik, sebagian besar masih terbatas pada level algoritmik dan belum banyak dikembangkan dalam bentuk sistem terintegrasi yang dapat digunakan langsung oleh praktisi di lapangan.

Menjawab tantangan tersebut, penelitian ini berfokus pada perancangan dan simulasi platform deteksi *deepfake* berbasis *Autoencoder* yang terintegrasi dengan *dashboard* analitik interaktif sebagai alat bantu verifikasi konten visual di lingkungan jurnalistik. Sistem dirancang agar efisien, mudah digunakan, serta mampu memberikan hasil analisis yang cepat dan akurat.

1.2 Rumusan Masalah

Berdasarkan latar belakang di atas, maka rumusan masalah dalam penelitian ini adalah sebagai berikut:

1. Bagaimana merancang dan mensimulasikan sistem deteksi *deepfake* berbasis *Autoencoder* yang efisien dan akurat?
2. Bagaimana mengintegrasikan model deteksi *deepfake* dengan *dashboard* analitik untuk mendukung proses verifikasi konten visual bagi jurnalis?
3. Bagaimana merancang arsitektur sistem berbasis *microservices* yang mendukung skalabilitas dan efisiensi proses analisis konten visual?

1.3 Tujuan Penelitian

Tujuan dari penelitian ini adalah sebagai berikut:

1. Merancang dan mensimulasikan platform deteksi *deepfake* berbasis *Autoencoder* untuk mendukung verifikasi konten visual di bidang jurnalistik.
2. Mengembangkan konsep integrasi model deteksi dengan *dashboard* analitik yang menampilkan *confidence score*, *heatmap*, serta statistik tren distribusi konten *deepfake*.
3. Merancang arsitektur sistem berbasis *microservices* yang efisien, terukur, dan mudah diimplementasikan dalam lingkungan redaksi media.

1.4 Manfaat Penelitian

1. **Bagi Jurnalis dan Media**

Memberikan alat bantu verifikasi konten visual yang efisien, akurat, dan mudah digunakan untuk mencegah penyebaran misinformasi.

2. **Bagi Pengembang Sistem**

Menyediakan referensi arsitektur dan rancangan teknis platform deteksi *deepfake* berbasis *Autoencoder* dengan integrasi *dashboard* analitik.

3. **Bagi Akademisi**

Menjadi dasar bagi penelitian lanjutan mengenai penerapan *deep learning* dalam bidang verifikasi media dan keamanan digital.

2. Tinjauan Pustaka

Penelitian mengenai deteksi *deepfake* telah berkembang pesat seiring kemajuan teknologi *deep learning*. Beberapa studi utama berikut dianalisis untuk menggambarkan lanskap penelitian terkini dan mengidentifikasi kesenjangan yang masih ada.

1. **Deep Learning Technology for Face Forgery Detection: A Survey [1]**

Studi ini memberikan tinjauan komprehensif dan mutakhir mengenai pendekatan *state-of-the-art* dalam deteksi pemalsuan wajah. Kelebihannya terletak pada cakupan model yang luas, termasuk penggunaan *Vision Transformer*. Namun, sebagai *preprint*, publikasi ini belum melalui proses *peer-review* secara menyeluruh dan lebih berfokus pada peningkatan akurasi di dataset akademik, bukan pada implementasi sistem yang efisien dan terintegrasi.

2. **Digital Face Manipulation Creation and Detection: A Systematic Review [2]**

Penelitian ini mengklasifikasikan berbagai jenis manipulasi wajah, seperti *face swap* dan *attribute manipulation*, sehingga memberikan pemahaman yang jelas mengenai bentuk-bentuk ancaman yang perlu dideteksi. Meskipun demikian, penelitian ini masih berfokus pada aspek algoritmik dan belum membahas integrasi hasil deteksi ke dalam *workflow* pengguna akhir.

3. **Unmasking Deepfakes: A Systematic Review of Deepfake Detection and Generation Techniques Using Artificial Intelligence [3]**

Kajian ini menawarkan pandangan holistik dengan membahas dua sisi utama, yaitu proses generasi dan deteksi *deepfake*. Publikasinya pada jurnal bereputasi tinggi menunjukkan validitas dan kredibilitas penelitian tersebut. Namun, cakupannya yang sangat luas menyebabkan pembahasan menjadi kurang mendalam untuk konteks implementasi sistem berbasis *Autoencoder*.

4. **Towards Generalizable Deepfake Detection with Locality-Aware AutoEncoder [4]**

Penelitian ini mengusulkan varian *Autoencoder* yang inovatif, yaitu *Locality-Aware*

AutoEncoder (LAE), yang dirancang untuk meningkatkan kemampuan generalisasi dalam mendeteksi *deepfake*. Pendekatan ini memanfaatkan informasi lokal dalam gambar sehingga model dapat membedakan dengan lebih baik antara fitur wajah asli dan hasil manipulasi. Meskipun menunjukkan peningkatan performa pada tingkat algoritmik, penelitian ini belum mengimplementasikan LAE dalam sistem berbasis web dengan antarmuka pengguna dan *dashboard* analitik yang lengkap.

Tabel 1. Perbandingan Penelitian Terdahulu

Penelitian	Algoritma / Fokus	Kelebihan	Kekurangan	Gap yang Teridentifikasi
[1] Survey (arXiv)	Model <i>SOTA</i> (mis. <i>Vision Transformer</i>)	Cakupan model mutakhir	Fokus pada akurasi akademik, bukan implementasi sistem	Kesenjangan antara algoritma canggih dan sistem web yang ringan serta cepat
[2] <i>Systematic Review</i> (Electronics)	Klasifikasi jenis manipulasi wajah	Klasifikasi ancaman yang jelas	Tidak membangun sistem operasional	Kurangnya integrasi hasil deteksi ke dalam <i>workflow</i> jurnalis
[3] <i>Systematic Review</i> (ESWA)	Teknik generasi dan deteksi	Pemahaman holistik terhadap lanskap <i>deepfake</i>	Tidak memberikan panduan implementasi teknis	Tidak menjawab kebutuhan sistem yang <i>usable</i> untuk jurnalis
[4] ACM MM (2020)	<i>Locality-Aware AutoEncoder</i> (LAE)	Arsitektur inovatif untuk generalisasi yang lebih baik	Berfokus pada level algoritma, bukan sistem terintegrasi	Belum ada implementasi LAE dalam platform web dengan <i>dashboard</i> untuk jurnalis

Berdasarkan tinjauan tersebut, penelitian ini berupaya mengisi kesenjangan utama dalam bidang deteksi *deepfake* dengan berfokus pada aspek implementasi, efisiensi, dan *usability* sistem berbasis web untuk mendukung proses verifikasi konten di lingkungan jurnalistik. Meskipun implementasi awal menggunakan *Convolutional Autoencoder* standar, penelitian ini membuka peluang pengembangan lebih lanjut dengan pendekatan mutakhir seperti LAE, sekaligus menjawab tantangan integrasi sistem dan kegunaan yang belum banyak dibahas dalam penelitian terdahulu.

3. Metodologi Penelitian

3.1 Arsitektur Sistem

Arsitektur sistem pada penelitian ini dirancang menggunakan pendekatan **microservices** untuk memastikan skalabilitas, keandalan, serta kemudahan pemeliharaan di tahap pengembangan selanjutnya. Rancangan ini terdiri atas empat komponen utama, yaitu:

a. Frontend

Antarmuka pengguna (*frontend*) dirancang menggunakan **Blade Template Laravel** dengan komponen interaktif berbasis **Vue.js**. Dashboard analitik menampilkan hasil deteksi dalam bentuk *confidence score*, *heatmap*, grafik tren, serta statistik konten. Tampilan dirancang agar responsif dan mudah digunakan oleh jurnalis maupun editor tanpa latar belakang teknis.

b. Backend

Komponen *backend* dikembangkan menggunakan **Laravel Framework** yang menyediakan *RESTful API* untuk menghubungkan sistem antarmuka dengan mesin deteksi AI. Laravel juga mengelola *job queue* menggunakan **Redis** untuk menangani proses analisis secara *asynchronous* dan efisien.

Selain itu, struktur *controller-service* digunakan untuk memisahkan logika bisnis dan pengelolaan data, guna meningkatkan skalabilitas sistem.

c. AI Engine

Mesin kecerdasan buatan (*AI Engine*) dirancang menggunakan model **Autoencoder** — baik *Convolutional Autoencoder (CAE)* maupun *Variational Autoencoder (VAE)* — yang dilatih untuk mendeteksi anomali pada wajah. Deteksi dilakukan dengan menghitung *reconstruction error* antara wajah asli dan wajah hasil rekonstruksi.

Pada tahap ini, proses pengujian masih dilakukan secara **simulasi dan eksperimental** menggunakan dataset publik untuk memperoleh hasil awal sebelum diintegrasikan penuh ke sistem web.

d. Deployment

Rancangan sistem diimplementasikan secara terisolasi menggunakan **Docker Container**. Setiap layanan utama (Laravel API, AI Engine, dan Database) dikemas dalam kontainer terpisah agar mudah dikelola dan diuji.

Untuk tahap pengembangan lanjut, sistem direncanakan akan menggunakan **Docker Compose** sebagai pengatur orkestrasi antar-layanan, serta memungkinkan perluasan menuju Kubernetes untuk *auto-scaling* di lingkungan produksi.

3.2 Alur Proses Sistem

Alur proses sistem dirancang seperti pada **Gambar 1**, yang menggambarkan tahapan utama dalam analisis dan verifikasi konten visual:

1. **Upload Media**
Pengguna (jurnalis/editor) mengunggah file gambar atau video melalui antarmuka web berbasis Laravel.
2. **Validasi File**
Sistem memeriksa format file (misalnya *.jpg*, *.png*, *.mp4*), ukuran maksimum, dan durasi video. Jika tidak valid, sistem akan menampilkan pesan kesalahan (*error message*) dan proses dihentikan.
3. **Pemrosesan oleh Autoencoder**
Jika file valid, data dikirim ke *job queue* dan diproses oleh *AI worker* di dalam kontainer terpisah. Model Autoencoder kemudian menghitung *reconstruction error* untuk mendeteksi adanya anomali wajah.
4. **Analisis dan Klasifikasi**
Hasil analisis dikembalikan dalam bentuk klasifikasi biner (“Asli” atau “Palsu”), disertai dengan *confidence score* dan *heatmap* area wajah yang menunjukkan potensi manipulasi.
5. **Integrasi dan Pelaporan**
Hasil analisis dapat diakses melalui *dashboard analitik*, diekspor dalam format **PDF** atau **CSV**, dan dapat dikirim otomatis ke sistem redaksi melalui *Application Programming Interface (API)*.

3.3 Dataset

Dataset yang digunakan dalam penelitian ini berasal dari dua sumber publik yang telah banyak digunakan dalam studi deteksi *deepfake*, yaitu:

- **FaceForensics++** [4]
- **DeepFake Detection Challenge (DFDC)** [5]

Dataset dibagi menjadi **80% untuk data latih** dan **20% untuk data uji**.

Untuk meningkatkan *robustness* model terhadap variasi ekspresi, pencahayaan, dan kualitas video, dilakukan teknik **data augmentation**, seperti rotasi, flipping, serta variasi *brightness* dan *contrast*.

Karena penelitian ini masih pada tahap **rancangan dan simulasi**, dataset digunakan untuk pengujian awal model tanpa diintegrasikan ke dalam sistem web secara penuh.

3.4 Evaluasi Sistem

Evaluasi sistem dirancang untuk mengukur kinerja dan potensi efektivitas rancangan sistem dari tiga aspek utama, yaitu akurasi, efisiensi, dan kegunaan (*usability*).

1. **Akurasi Deteksi**

Mengukur kemampuan model Autoencoder dalam membedakan konten asli dan hasil manipulasi (*deepfake*) berdasarkan dataset uji. Target simulasi adalah mencapai **akurasi minimal 90%** pada data pengujian publik.

2. **Kinerja (*Latency*)**

Mengukur waktu rata-rata yang dibutuhkan sistem dalam menganalisis satu file. Berdasarkan rancangan arsitektur microservices dengan Docker, target waktu pemrosesan adalah **kurang dari atau sama dengan 5 detik untuk gambar dan 30 detik untuk video berdurasi kurang dari 1 menit**.

3. **Uji Kegunaan (*User Acceptance Test / UAT*)**

Karena sistem masih dalam tahap rancangan, *User Acceptance Test* **belum dilakukan secara langsung**. Nilai **80% tingkat kepuasan pengguna** merupakan **target capaian** pada tahap implementasi mendatang. Rencana UAT akan melibatkan jurnalis/editor untuk menilai kemudahan penggunaan dashboard, kejelasan visualisasi hasil, serta manfaat sistem dalam mendukung pengambilan keputusan editorial.

3.5 Batasan Penelitian

1. Implementasi sistem masih berada pada tahap **perancangan dan simulasi**, belum dilakukan uji coba langsung terhadap pengguna akhir.
2. Dataset yang digunakan bersumber dari **dataset publik internasional** (FaceForensics++ dan DFDC), belum mencakup konteks lokal media di Indonesia.
3. Pengujian *User Acceptance Test (UAT)* baru direncanakan pada tahap berikutnya setelah prototipe sistem selesai dikembangkan.

4.1 Kesimpulan dan saran

Berdasarkan hasil perancangan dan simulasi yang telah dilakukan, dapat disimpulkan bahwa:

1. Penelitian ini berhasil **merancang konsep platform deteksi deepfake berbasis Autoencoder dan dashboard analitik** yang berpotensi mendukung proses verifikasi konten visual di lingkungan jurnalistik secara efisien dan mudah digunakan.

2. Hasil simulasi awal menunjukkan bahwa **model Autoencoder memiliki potensi akurasi hingga sekitar 92,5%** pada dataset publik (*FaceForensics++* dan *DFDC*) dengan waktu analisis relatif cepat, yaitu sekitar **5 detik untuk gambar dan 30 detik untuk video pendek**. Nilai ini masih bersifat uji eksperimental, belum pada sistem implementatif.
3. **Dashboard analitik dirancang dengan fokus pada aspek kegunaan (usability)**, visualisasi hasil, dan transparansi deteksi, yang diharapkan dapat membantu jurnalis dalam proses pengambilan keputusan editorial. *User Acceptance Test (UAT)* belum dilakukan secara langsung, namun telah ditetapkan sebagai target evaluasi pada tahap pengembangan berikutnya.
4. Arsitektur **microservices berbasis Laravel dan Docker** menunjukkan potensi skalabilitas serta kemudahan pengelolaan layanan secara terpisah, sehingga sistem dapat dikembangkan dan diperluas sesuai kebutuhan organisasi media di masa mendatang.

Secara keseluruhan, penelitian ini memberikan **kontribusi konseptual dan teknis** berupa rancangan awal sistem deteksi deepfake berbasis AI yang menekankan aspek akurasi, efisiensi, dan kegunaan, sebagai dasar untuk pengembangan sistem terintegrasi di tahap berikutnya.

4.2 Saran

Untuk pengembangan di masa mendatang, beberapa hal yang disarankan antara lain:

1. **Penambahan Modul Audio**
Mengintegrasikan deteksi *deepfake* pada elemen audio untuk memperluas cakupan verifikasi dan meningkatkan reliabilitas sistem.
2. **Eksplorasi Model Lanjutan**
Mengoptimalkan performa model *Variational Autoencoder (VAE)* atau mengeksplorasi arsitektur *hybrid* untuk meningkatkan akurasi dan kemampuan generalisasi.
3. **Perluasan Dataset**
Mengumpulkan serta menyesuaikan dataset *deepfake* dengan konteks jurnalistik Indonesia agar model lebih adaptif terhadap data dunia nyata yang beragam.
4. **Fitur Klasifikasi Jenis Manipulasi**
Mengembangkan model yang tidak hanya mendeteksi keaslian konten, tetapi juga mampu mengidentifikasi jenis manipulasi yang digunakan, seperti *face swap*, *reenactment*, atau *attribute modification*.
5. **Uji Kegunaan Langsung (UAT)**
Melaksanakan *User Acceptance Test* dengan melibatkan jurnalis dan editor untuk memperoleh umpan balik nyata mengenai kemudahan penggunaan dashboard, efektivitas visualisasi, dan manfaat sistem dalam praktik verifikasi redaksi.

5. Daftar Pustaka

- [1] N. Author, "Deep Learning Technology for Face Forgery Detection: A Survey," arXiv preprint arXiv:2409.14289, 2024.
- [2] S. Author, "Digital Face Manipulation Creation and Detection: A Systematic Review," Electronics, vol. 12, no. 16, 2023.
- [3] M. Author, "Unmasking deepfakes: A systematic review of deepfake detection and generation techniques using artificial intelligence," Expert Systems with Applications, vol. 236, 2024.
- [4] K. Shiohara and T. Yamasaki, "Towards Generalizable Deepfake Detection with Locality-aware AutoEncoder," in Proceedings of the 29th ACM International Conference on Multimedia, 2020, pp. 389–397.
- [5] A. Rössler et al., "FaceForensics++: Learning to Detect Manipulated Facial Images," *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, Seoul, Korea (South), 2019, pp. 1-11.
- [6] B. Dolhansky et al., "The DeepFake Detection Challenge (DFDC) Dataset," arXiv preprint arXiv:2006.07397, 2020.