Jurnal Penelitian Inovatif (JUPIN)
Vol. 6, No. 3, Agustus 2026, Hal. 2805-2828
https://jurnal-id.com/index.php/jupin

DOI: https://doi.org/10.54082/jupin.2712
p-ISSN: 2808-148X
e-ISSN: 2808-1366

Perancangan Sistem Deteksi Video Deepfake Menggunakan Arsitektur Multicue End-to-End dengan Pendekatan Design Science Research

Teguh Pribadi[*1], Zaenal Ma'aruf [2]

[1,2] Program Studi Teknik Informatika, Sekolah Tinggi Manajemen Informatika dan Komputer DCI, Indonesia
Email: [1]teguh.2022210063@gmail.com, [2]zaenal_maruf@stmik-dci.ac.id

Abstrak

Perkembangan model generatif berbasis deep learning membuat manipulasi wajah, suara, dan gerak bibir semakin realistis serta meningkatkan kebutuhan akan sistem deteksi yang dapat ditelusuri. Penelitian perancangan ini bertujuan menyusun spesifikasi sistem deteksi video deepfake multicue end-to-end yang siap diimplementasikan. Metode mengikuti enam tahapan Design Science Research: identifikasi masalah dan motivasi, penetapan tujuan solusi, perancangan dan pengembangan, pengujian melalui skenario penggunaan, evaluasi terhadap kriteria rancangan, serta diseminasi hasil. Capaian utama penelitian meliputi: (1) arsitektur multicue yang mengintegrasikan petunjuk spasial, frekuensi, temporal, dan audio-visual; (2) kontrak API dan alur layanan asinkron; (3) skema data, penyimpanan bukti, dan jejak audit; (4) rancangan deployment, registri model, dan pemantauan; serta (5) protokol evaluasi model, sistem, keamanan, dan manusia beserta kriteria penerimaannya. Spesifikasi juga mengatur pencegahan kebocoran data, kalibrasi dan ketidakpastian, pengujian lintas-dataset dan lintas-generator, serta penundaan keputusan ketika bukti tidak memadai. Implementasi direncanakan secara bertahap, dimulai dari MVP spasial-temporal, kemudian diperluas dengan cabang frekuensi dan audio-visual. Artikel ini tidak menyajikan pengukuran performa empiris karena fokus penelitian berada pada penyusunan spesifikasi sistem yang konsisten, dapat ditelusuri, dan siap divalidasi; pengukuran kinerja dilakukan setelah implementasi selesai dibangun.

Kata kunci: *AI generatif, deepfake, media digital, rancangan sistem.*

Abstract

The increasing realism of face, voice, and lip-motion manipulation creates a need for traceable detection systems. This design-oriented study develops an implementation-ready specification for an end-to-end multicue video deepfake detection system. The method follows six Design Science Research stages: problem identification and motivation, definition of solution objectives, design and development, demonstration through use scenarios, evaluation against design criteria, and communication of results. The main outcomes comprise: (1) a multicue architecture integrating spatial, frequency, temporal, and audio-visual evidence; (2) API contracts and an asynchronous service workflow; (3) data schemas, evidence storage, and audit trails; (4) deployment, model registry, and monitoring designs; and (5) model, system, security, and human evaluation protocols with acceptance criteria. The specification also addresses data-leakage prevention, calibration and uncertainty, cross-dataset and cross-generator testing, and decision deferral when evidence is insufficient. Implementation is planned incrementally, beginning with a spatial-temporal MVP and extending it with frequency and audio-visual branches. This paper does not report empirical performance measurements because its scope is the construction of a coherent, traceable, and validation-ready system specification; performance measurement is reserved for the post-implementation stage.

***Keywords**: Deepfake, digital media, generative AI, system design.*

1. PENDAHULUAN

Kemajuan kecerdasan buatan generatif telah memungkinkan pembuatan media sintetis yang semakin menyerupai rekaman autentik sehingga perbedaannya sulit dikenali secara visual. Pada media video, teknologi deepfake dapat digunakan untuk mengganti identitas seseorang, memanipulasi ekspresi wajah, menyesuaikan gerakan bibir dengan audio, maupun menghasilkan wajah sintetis secara keseluruhan. Perkembangan tersebut menimbulkan berbagai risiko, seperti penipuan identitas,

DOI: https://doi.org/10.54082/jupin.2712
p-ISSN: 2808-148X
e-ISSN: 2808-1366

penyebaran disinformasi, pencemaran nama baik, manipulasi bukti digital, dan menurunnya kepercayaan publik terhadap keaslian konten audiovisual (Mirsky & Lee, 2021; Nguyen et al., 2022; Juefei-Xu et al., 2022; Abbas & Taeihagh, 2024; Kaur et al., 2024). Kondisi ini menunjukkan bahwa deteksi deepfake tidak cukup dipandang hanya sebagai persoalan klasifikasi antara konten asli dan palsu. Dalam konteks forensik digital, sistem perlu menyajikan tingkat keyakinan prediksi, bukti pendukung, rekam jejak pemeriksaan, potensi kegagalan, serta mekanisme hasil peninjauan oleh manusia. Tinjauan forensik terkini juga menekankan perlunya validasi lintas-teknik dan data dunia nyata sebelum keluaran detektor digunakan sebagai dasar keputusan (van Lierop et al., 2026).

Berbagai pendekatan telah dikembangkan untuk meningkatkan kemampuan deteksi deepfake dengan memanfaatkan karakteristik manipulasi yang berbeda. Pendekatan berbasis pola spasial dan karakteristik wajah, seperti multi-attentional detection dan analisis hubungan facial action units, mampu mengekstraksi ketidakkonsistenan lokal pada area wajah yang sulit diamati secara langsung (Zhao et al., 2021; Bai et al., 2023). Pendekatan berbasis domain spasial dan frekuensi juga memperluas informasi yang digunakan model sehingga tidak hanya bergantung pada karakteristik RGB, tetapi juga pada pola frekuensi yang muncul akibat proses manipulasi (Dong et al., 2023). Meskipun efektif pada kondisi pengujian tertentu, pendekatan yang terlalu bergantung pada pola visual tertentu berpotensi mengalami penurunan performa ketika teknik generasi, tingkat kompresi, atau karakteristik dataset berubah. Sebaliknya, FakeCatcher memanfaatkan sinyal biologis yang terekam pada wajah, sedangkan ID-Reveal memanfaatkan pola temporal yang berkaitan dengan karakteristik identitas seseorang (Ciftci et al., 2020; Cozzolino et al., 2021). Pendekatan tersebut menawarkan informasi yang berbeda dari pola visual konvensional, tetapi efektivitasnya tetap dipengaruhi oleh kualitas video dan keterlihatan petunjuk temporal yang digunakan. AltFreezing kemudian menggabungkan pembelajaran pola spasial dan temporal untuk meningkatkan generalisasi terhadap manipulasi yang tidak terlihat pada proses pelatihan (Wang et al., 2023). Sementara itu, pendekatan multimodal seperti AVFakeNet serta analisis sinkronisasi audio-visual memanfaatkan ketidaksesuaian antara informasi suara dan gerakan visual sehingga memberikan sumber bukti tambahan yang tidak tersedia pada detektor visual tunggal (Ilyas et al., 2023; Javed et al., 2025). Namun, pendekatan multimodal memerlukan ketersediaan dan kualitas lebih dari satu modalitas sehingga tidak selalu dapat diterapkan pada seluruh jenis video. Secara lebih luas, kajian sistematis menunjukkan bahwa generalisasi lintas-dataset dan variasi kondisi akuisisi tetap menjadi tantangan utama deteksi deepfake (Rana et al., 2022).

Perbedaan karakteristik tersebut menunjukkan bahwa tidak terdapat satu jenis petunjuk yang selalu dominan untuk seluruh bentuk manipulasi deepfake. Temuan pada DeepfakeBench menunjukkan bahwa perbedaan pengelolaan data, konfigurasi eksperimen, protokol evaluasi, dan dataset dapat menghasilkan perbandingan kinerja yang tidak konsisten antarmetode (Yan et al., 2023). Selanjutnya, DF40 memperluas evaluasi terhadap 40 teknik pembuatan deepfake dan menunjukkan pentingnya keragaman jenis manipulasi, tingkat realisme, serta skenario evaluasi dalam menilai kemampuan generalisasi suatu detektor (Yan et al., 2024). Hasil tersebut memperkuat kebutuhan akan pendekatan yang tidak bergantung pada satu petunjuk saja, tetapi mengintegrasikan beberapa petunjuk atau multicue secara terkoordinasi. Integrasi tersebut diharapkan memungkinkan kelemahan suatu petunjuk dikompensasi oleh petunjuk lain ketika sistem menghadapi perubahan generator, identitas, kualitas video, kompresi, maupun karakteristik data yang berbeda.

Berdasarkan kajian tersebut, terdapat kesenjangan antara perkembangan model deteksi deepfake dan kebutuhan implementasinya sebagai suatu sistem yang dapat digunakan secara operasional. Pertama, performa yang tinggi pada satu dataset belum selalu menunjukkan kemampuan generalisasi terhadap dataset, identitas, generator, atau kondisi video yang berbeda (Tolosana et al., 2022; Stroebel et al., 2023; Kaur et al., 2024; Yan et al., 2023; Yan et al., 2024). Kedua, pengelolaan dan pembagian data perlu dirancang secara hati-hati agar informasi yang berasal dari video atau identitas yang sama tidak tersebar pada data pelatihan dan pengujian, karena kondisi tersebut dapat menghasilkan estimasi performa yang terlalu optimistis. Ketiga, sebagian besar penelitian berfokus pada pengembangan dan evaluasi model, sedangkan aspek penerimaan media, orkestrasi proses inferensi, pemrosesan asinkron, pengelolaan versi model, penyimpanan bukti, keamanan, kalibrasi hasil, serta audit keputusan belum selalu menjadi bagian integral dari rancangan.

DOI: https://doi.org/10.54082/jupin.2712
p-ISSN: 2808-148X
e-ISSN: 2808-1366

Tabel 1. Sintesis Pendekatan Deteksi Deepfake dan Implikasinya terhadap Rancangan Sistem

Pendekatan dan studi	Kelebihan utama	Keterbatasan	Sintesis untuk rancangan sistem
Karakteristik spasial dan relasi wajah: Zhao et al. (2021); Bai et al. (2023)	Mendeteksi ketidakkonsistenan lokal melalui mekanisme perhatian dan hubungan facial action units. Bukti spasial relatif mudah divisualisasikan.	Rentan mempelajari ciri dataset, kompresi, atau pola generator tertentu. Kinerja dapat turun ketika domain berubah.	Digunakan sebagai cabang spasial dan sumber bukti, tetapi tidak dijadikan satu-satunya dasar keputusan.
Spasial dan frekuensi: Dong et al. (2023)	Menggabungkan informasi RGB dan pola frekuensi sehingga manipulasi yang tidak jelas pada domain piksel dapat terdeteksi.	Jejak frekuensi dapat berubah akibat pengodean ulang, resizing, kompresi berat, dan generator baru.	Ditempatkan sebagai cabang komplementer dengan pencatatan kualitas input dan uji ketahanan terhadap transformasi.
Biologis dan temporal berbasis identitas: Ciftci et al. (2020); Cozzolino et al. (2021)	Memanfaatkan sinyal biologis atau pola gerak identitas yang berbeda dari pola visual konvensional.	Memerlukan wajah yang terlihat, durasi memadai, dan urutan video yang stabil. Sinyal melemah pada kualitas rendah atau oklusi.	Diterapkan melalui cabang temporal dengan quality gating dan status bukti tidak memadai ketika petunjuk tidak tersedia.
Pembelajaran spasial-temporal: Wang et al. (2023)	AltFreezing memperkuat pembelajaran kedua domain dan meningkatkan generalisasi pada manipulasi yang tidak terlihat saat pelatihan.	Hasil tetap dipengaruhi komposisi data, protokol eksperimen, dan biaya komputasi video yang lebih tinggi.	Memerlukan ablation study, pengujian lintas-generator, serta pemantauan latensi dan penggunaan sumber daya.
Audio-visual multimodal: Ilyas et al. (2023); Javed et al. (2025)	Menangkap ketidaksesuaian suara, fonem, gerak bibir, dan ekspresi sebagai bukti tambahan terhadap manipulasi lintas-modalitas.	Tidak dapat digunakan optimal ketika audio hilang, bising, tidak sinkron, atau tidak relevan terhadap jenis manipulasi.	Dirancang sebagai cabang opsional. Sistem harus tetap berfungsi dan mencatat alasan ketika modalitas audio tidak tersedia.
Benchmark terpadu dan beragam: Yan et al. (2023, 2024)	DeepfakeBench meningkatkan keterbandingan, sedangkan DF40 memperluas variasi teknik manipulasi untuk menguji generalisasi.	Benchmark tidak otomatis menyelesaikan kalibrasi, keamanan, penyimpanan bukti, dan kebutuhan penggunaan operasional.	Rancangan menggabungkan protokol evaluasi seragam dengan pengujian lintas-data, versioning, audit, dan tata kelola keputusan.

Dengan demikian, masih diperlukan rancangan yang menghubungkan model deteksi multicue dengan komponen perangkat lunak, pengelolaan data, mekanisme evaluasi, dan tata kelola keputusan sehingga hasil deteksi dapat digunakan secara lebih transparan dan dapat dipertanggungjawabkan.

DOI: https://doi.org/10.54082/jupin.2712
p-ISSN: 2808-148X
e-ISSN: 2808-1366

Tabel 2. Ringkasan Kesenjangan Penelitian dan Respons Rancangan Sistem

Kesenjangan penelitian	Dampak ilmiah dan operasional	Respons dalam rancangan sistem
Generalisasi lintas dataset, identitas, generator, dan kondisi video belum konsisten.	Akurasi pada satu benchmark dapat memberikan gambaran terlalu optimistis dan tidak mewakili penggunaan dunia nyata.	Mengintegrasikan petunjuk spasial, frekuensi, temporal, serta audio-visual dan menguji sistem pada skenario lintas-data serta lintas-generator.
Pembagian data pada tingkat frame atau tanpa kontrol identitas berisiko menimbulkan kebocoran informasi.	Model dapat mengenali kemiripan video atau identitas, bukan karakteristik manipulasi, sehingga estimasi performa menjadi bias.	Menerapkan pemisahan berdasarkan identitas, sumber video, dan keluarga generator serta memeriksa duplikasi melalui hash dan manifest data.
Sebagian penelitian berhenti pada model dan belum merancang alur operasional end-to-end.	Hasil deteksi sulit direproduksi, ditelusuri, diamankan, dipelihara, dan digunakan sebagai bukti pendukung.	Merancang layanan penerimaan media, antrean asinkron, prapemrosesan, registri model, penyimpanan bukti, API, keamanan, dan audit keputusan.
Kalibrasi, ketidakpastian, serta pengawasan manusia belum selalu menjadi bagian integral dari keluaran sistem.	Skor biner berpotensi ditafsirkan sebagai keputusan final meskipun bukti lemah atau distribusi input telah berubah.	Menambahkan kalibrasi, rentang keputusan, status insufficient-evidence, penjelasan per petunjuk, dan mekanisme human-in-the-loop.

Untuk menjawab kesenjangan tersebut, penelitian ini bertujuan untuk: (1) merumuskan landasan teoretis dan kebutuhan fungsional maupun nonfungsional bagi sistem deteksi video deepfake; (2) menerjemahkan pendekatan multicue ke dalam arsitektur end-to-end yang mengintegrasikan proses penerimaan video, ekstraksi dan analisis petunjuk, inferensi, agregasi hasil, serta penyimpanan bukti; (3) menyusun strategi pengelolaan dan pemisahan data untuk mengurangi risiko kebocoran informasi serta menguji kemampuan generalisasi pada dataset dan teknik manipulasi yang tidak digunakan pada proses pelatihan; (4) menetapkan protokol evaluasi yang mencakup performa model, ketahanan terhadap perubahan kondisi input, kalibrasi, keamanan, dan aspek operasional sistem; serta (5) menetapkan batas penggunaan dan mekanisme pengawasan manusia agar keluaran model diperlakukan sebagai bukti pendukung dan bukan sebagai dasar tunggal dalam pengambilan keputusan.

Kelima tujuan penelitian saling berhubungan dan membentuk satu kontribusi sistemik. Pertama menghasilkan dasar kebutuhan dan kriteria desain. Kedua menerjemahkannya menjadi arsitektur end-to-end, alur layanan, dan skema data multicue. Ketiga menyediakan tata kelola dataset, pencegahan kebocoran, serta skenario pengujian generalisasi. Keempat membentuk protokol verifikasi untuk menilai performa, ketahanan, kalibrasi, keamanan, dan kesiapan operasional. Kelima menetapkan batas penggunaan, jejak audit, status bukti tidak memadai, dan pengawasan manusia. Integrasi kelima tujuan tersebut menghasilkan rancangan sistem deteksi video deepfake yang dapat diwujudkan sebagai Minimum Viable Product (MVP) dan dievaluasi secara transparan sebelum digunakan pada keputusan berisiko tinggi.

Karena penelitian menghasilkan rancangan sistem dan prosedur evaluasinya, proses penelitian ditempatkan dalam kerangka Design Science Research yang menekankan pembangunan dan evaluasi solusi untuk menyelesaikan permasalahan yang teridentifikasi (Hevner et al., 2004; Peffers et al., 2007; Gregor & Hevner, 2013; Akoka et al., 2023). Dengan demikian, kontribusi penelitian tidak terbatas pada model klasifikasi, tetapi mencakup rancangan sistem deteksi video deepfake multicue end-to-end yang terukur, transparan, dan dapat dipertanggungjawabkan.

DOI: https://doi.org/10.54082/jupin.2712
p-ISSN: 2808-148X
e-ISSN: 2808-1366

2. METODE PENELITIAN

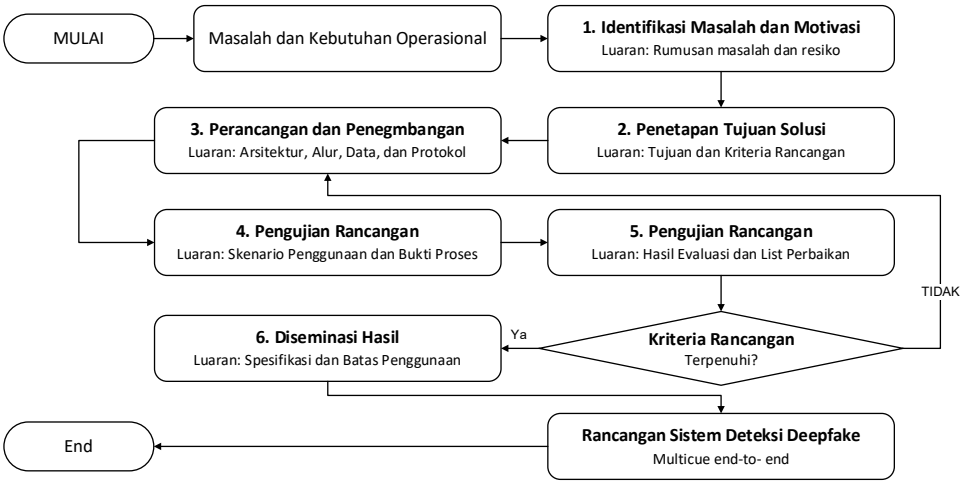
Status penelitian artikel ini merupakan penelitian perancangan sistem. Kegiatan penelitian dibatasi pada identifikasi masalah, perumusan kebutuhan, penyusunan arsitektur, pengujian melalui skenario penggunaan, serta penetapan protokol evaluasi. Penelitian ini belum mengunduh atau memproses dataset video, belum melatih maupun menjalankan model, dan belum menghasilkan metrik performa empiris. Seluruh nama dataset dan nilai target yang disampaikan pada bagian berikutnya merupakan rencana validasi dan kriteria penerimaan untuk tahap implementasi, bukan hasil pengujian sistem yang telah berjalan.

2.1 Tahapan Design Science Research

Penelitian ini menggunakan pendekatan Design Science Research (DSR) karena tujuan utama penelitian adalah menghasilkan dan mengevaluasi berupa rancangan sistem deteksi video deepfake berbasis arsitektur multicue end-to-end yang harus menjawab masalah operasional secara terstruktur, bukan hasil pengukuran performa model yang telah diimplementasikan. Prosedur penelitian mengadaptasi enam aktivitas Design Science Research Methodology (DSRM), yaitu identifikasi masalah dan motivasi, penetapan tujuan solusi, perancangan dan pengembangan, pengujian, evaluasi, serta komunikasi hasil (Hevner et al., 2004; Peffers et al., 2007). Hubungan antaraktivitas dan luaran setiap tahap ditunjukkan pada Gambar 1.

Tahap identifikasi menghasilkan rumusan masalah, risiko, dan kebutuhan pemangku kepentingan. Tujuan solusi kemudian diterjemahkan menjadi kriteria rancangan yang terukur. Tahap perancangan menghasilkan arsitektur layanan, model multicue, skema data, kontrak API, serta protokol evaluasi. Rancangan tersebut disimulasikan melalui skenario unggah video, inferensi, penyimpanan bukti, dan peninjauan hasil. Evaluasi memeriksa kelengkapan fungsi, keterlacakan, keamanan, kemampuan generalisasi yang direncanakan, dan kesesuaian dengan kriteria rancangan. Jika kriteria belum terpenuhi, hasil evaluasi menjadi masukan untuk iterasi perancangan; jika terpenuhi, spesifikasi akhir, batas penggunaan, dan agenda implementasi dikomunikasikan sebagai hasil penelitian.

Kerangka DSR pada penelitian ini juga diposisikan sebagai penghubung antara masalah praktis, pembangunan solusi, dan pembentukan pengetahuan desain. Hevner dan Chatterjee (2010) menekankan bahwa keluaran DSR tidak berhenti pada sistem yang dapat dijalankan, tetapi perlu disertai argumentasi desain dan evaluasi yang dapat dipertanggungjawabkan. Dresch et al. (2015) memperkuat kebutuhan akan keselarasan antara identifikasi masalah, pembangunan solusi, dan prosedur evaluasi, sedangkan vom Brocke et al. (2020) menempatkan solusi yang dirancang dan pengetahuan desain sebagai dua hasil yang saling melengkapi. Atas dasar itu, artikel ini memperlakukan arsitektur, kontrak API, skema data, serta protokol evaluasi sebagai satu paket kontribusi desain yang harus dapat ditelusuri ke kebutuhan awal.



Gambar 1. Alur Tahapan Penelitian Design Science Research

DOI: https://doi.org/10.54082/jupin.2712
p-ISSN: 2808-148X
e-ISSN: 2808-1366

Hubungan antarbagian pada Gambar 1 bersifat dependensial dan iteratif. Rumusan masalah menentukan tujuan solusi, sedangkan tujuan menjadi dasar kriteria penerimaan yang mengendalikan perancangan. Simulasi tidak diperlakukan sebagai hasil akhir, tetapi sebagai cara menemukan ketidaksesuaian antara rancangan dan skenario penggunaan. Evaluasi kemudian berfungsi sebagai gerbang keputusan: ketika kriteria belum terpenuhi, temuan dikembalikan ke tahap perancangan; ketika terpenuhi, luaran dikomunikasikan bersama batas penggunaan. Dengan demikian, rancangan akhir merupakan hasil rantai keputusan yang dapat ditelusuri, bukan sekadar urutan enam aktivitas.

2.2 Status Dataset dan Rencana Validasi

Penelitian ini berada pada tahap perancangan dan belum melaksanakan pelatihan, inferensi, validasi, atau pengujian empiris terhadap video. Dengan demikian, tidak ada dataset video yang digunakan sebagai data eksperimen pada penelitian ini. Sumber yang benar-benar digunakan adalah literatur ilmiah, dokumentasi benchmark, dan deskripsi dataset publik untuk menurunkan kebutuhan sistem, aturan pemisahan data, skenario evaluasi, serta kriteria penerimaan. Nama dataset video yang dicantumkan selanjutnya merupakan rencana validasi pada tahap implementasi dan tidak menjadi dasar klaim performa dalam artikel ini. Pemisahan status tersebut dirangkum pada Tabel 3.

Tabel 3. Pemisahan Sumber yang Digunakan dan Dataset Rencana Validasi

Status	Sumber atau dataset	Peran	Batas klaim
Digunakan pada penelitian ini	Literatur ilmiah, dokumentasi DeepfakeBench, serta deskripsi dan metadata dataset publik	Menurunkan kebutuhan fungsional dan nonfungsional, arsitektur, aturan pembagian data, metrik, skenario uji, dan batas penggunaan	Sumber perancangan; tidak digunakan untuk melatih model, menjalankan inferensi, atau menghasilkan nilai performa
Rencana validasi internal	FaceForensics++, Celeb-DF, DFDC, dan DeeperForensics-1.0	Pelatihan, validasi, pengujian dalam dataset, serta uji gangguan terkontrol pada tahap implementasi	Belum diproses dalam penelitian ini; komposisi akhir mengikuti lisensi, ketersediaan, dan audit duplikasi
Rencana validasi generalisasi	WildDeepfake dan DF40	Pengujian lintas-dataset, lintas-generator, dan lintas-teknik manipulasi	Belum menjadi bukti empiris; generator, identitas, dan video sumber harus ditahan dari data pelatihan
Rencana validasi eksternal bersyarat	Dataset evaluasi dunia nyata terkini yang memiliki izin penggunaan dan dokumentasi memadai	Menguji validitas eksternal, pergeseran distribusi, kompresi platform, serta kondisi operasional	Dilaksanakan hanya jika lisensi, privasi, provenance, dan hak penggunaan memungkinkan

Pada tahap implementasi, pembagian data direncanakan berdasarkan identitas, video sumber, dan keluarga generator. Frame dari video yang sama tidak boleh tersebar ke beberapa split; duplikasi diperiksa menggunakan SHA-256 dan perceptual hash; sedangkan manifest, lisensi, checksum, seed, transformasi, serta versi kode dicatat dalam dokumentasi dataset dan registri model (Gebru et al., 2021; Yan et al., 2023; Yan et al., 2024). Perumusan dalam bentuk rencana ini menjaga agar artikel tidak menyamakan spesifikasi validasi dengan pelaksanaan eksperimen.

2.3 Rasional Pemilihan Komponen Arsitektur

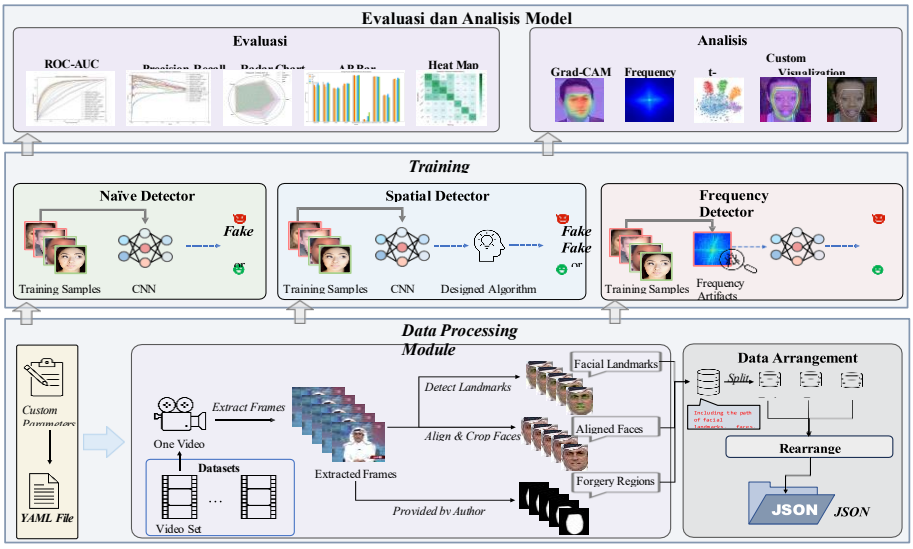
Komponen arsitektur dipilih berdasarkan empat kriteria: cakupan petunjuk manipulasi yang saling melengkapi, efisiensi komputasi untuk MVP, kemampuan diuji melalui ablation study, dan modularitas agar cabang yang tidak memiliki input dapat dinonaktifkan tanpa menghentikan analisis. EfficientNet-B0 dan Xception diposisikan sebagai kandidat backbone spasial yang dibandingkan, bukan sebagai dua

DOI: https://doi.org/10.54082/jupin.2712
p-ISSN: 2808-148X
e-ISSN: 2808-1366

model yang wajib dijalankan bersamaan. GRU, DCT, dan modul audio-visual menambahkan bukti temporal, frekuensi, dan sinkronisasi lintasmodalitas. Rasional serta strategi verifikasi setiap komponen dirangkum pada Tabel 4.

Tabel 4. Rasional Pemilihan Komponen Arsitektur Multicue

Komponen	Alasan pemilihan	Peran dalam rancangan	Keterbatasan dan strategi verifikasi
Efficient Net-B0	Memberikan kompromi antara kapasitas representasi dan biaya komputasi sehingga sesuai sebagai baseline MVP.	Mengekstraksi ciri spasial per frame dan menyediakan dasar visualisasi Grad-CAM.	Berisiko mempelajari ciri khas dataset; diverifikasi melalui perbandingan lintas-dataset dan dengan Xception (Juefei-Xu et al., 2022; Kaur et al., 2024).
Xception	Konvolusi depthwise separable menyediakan pembanding spasial yang mapan pada studi deteksi manipulasi wajah.	Menjadi backbone alternatif untuk menguji apakah hasil bergantung pada satu keluarga arsitektur.	Biaya komputasi relatif lebih besar; dibandingkan berdasarkan generalisasi, latensi, memori, dan kalibrasi, bukan AUC saja (Kaur et al., 2024).
GRU	Meringkas dependensi urutan dengan parameter yang lebih ringkas daripada recurrent stack yang lebih kompleks.	Mengagregasi vektor fitur antarframe untuk mendeteksi ketidakkonsistenan temporal.	Sensitif terhadap kualitas pelacakan dan panjang urutan; diuji melalui ablation spasial saja versus spasial-temporal (Ismail et al., 2022; Wang et al., 2023).
DCT	Membuka pola frekuensi tinggi dan jejak pemrosesan yang tidak selalu terlihat pada citra RGB.	Menyediakan cabang frekuensi yang melengkapi petunjuk spasial.	Dapat melemah setelah pengodean ulang atau kompresi; diuji pada beberapa codec, bitrate, dan tingkat kompresi (Dong et al., 2023).
Modul audio-visual	Ketidaksesuaian fonem-visem, gerak bibir, dan audio memberi bukti tambahan saat petunjuk visual ambigu.	Mengukur sinkronisasi lintasmodalitas sebagai cabang opsional.	Tidak tersedia atau tidak andal pada semua video; sistem menonaktifkan cabang dan mencatat alasan tanpa mengganti sinyal yang hilang (Ilyas et al., 2023; Javed et al., 2025).

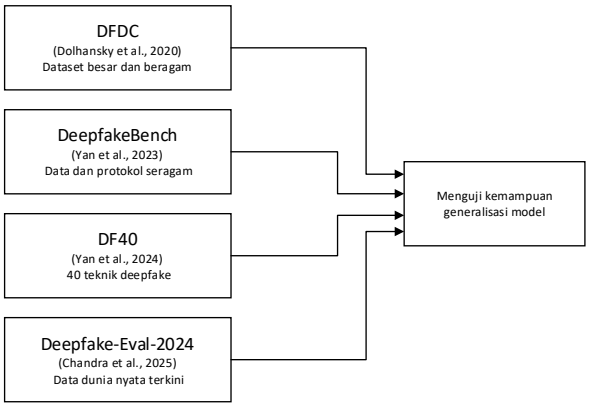


Gambar 2. Struktur Modular DeepfakeBench.

DOI: https://doi.org/10.54082/jupin.2712
p-ISSN: 2808-148X
e-ISSN: 2808-1366

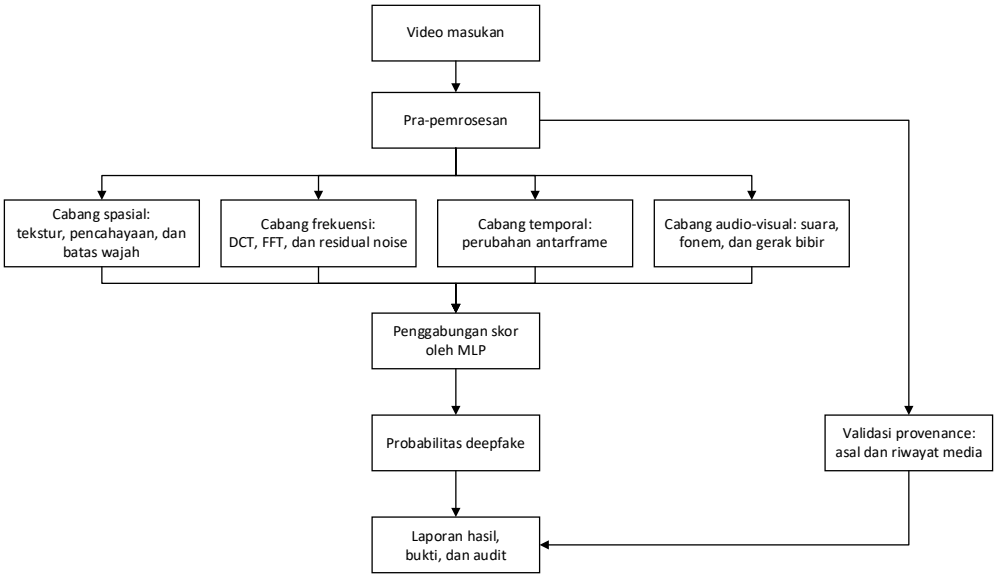
Urutan implementasi mengikuti tingkat risiko dan ketergantungan komponen. MVP membandingkan EfficientNet-B0 dan Xception sebagai backbone spasial, kemudian menambahkan GRU apabila pelacakan wajah dan urutan frame telah stabil. DCT serta modul audio-visual ditempatkan pada tahap pengayaan setelah baseline spasial-temporal memenuhi kriteria fungsi. Setiap penambahan diuji secara terpisah dan gabungan agar peningkatan performa dapat dibedakan dari pertambahan biaya komputasi atau ketergantungan terhadap modalitas tertentu.

Pada Gambar 2, modul pengolahan data, pelatihan, evaluasi, dan analisis membentuk rantai reprodusibilitas. Pengaturan data menentukan distribusi input yang diterima model; keluaran model baru dapat dibandingkan secara sah jika split, transformasi, dan unit evaluasinya seragam. Hasil analisis seperti Grad-CAM, frekuensi, dan proyeksi fitur kemudian memberikan umpan balik terhadap pemilihan transformasi maupun arsitektur. Hubungan ini penting karena kebocoran identitas atau ketidakkonsistenan prapemrosesan pada lapisan data dapat menghasilkan performa semu yang tidak dapat diperbaiki hanya dengan mengganti classifier.



Gambar 3. Gambaran Benchmark Penelitian

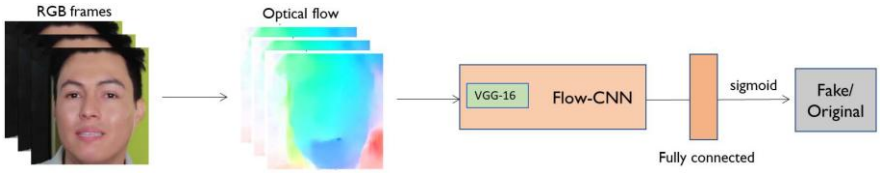
Gambar 3 menempatkan benchmark dalam fungsi yang berlapis, bukan sebagai satu kumpulan data yang langsung digabungkan. DFDC direncanakan untuk variasi internal berskala besar, DeepfakeBench untuk menyeragamkan protokol, DF40 untuk menahan keluarga generator yang lebih beragam, dan data dunia nyata untuk menguji pergeseran distribusi eksternal. Hubungan berjenjang ini memisahkan data yang digunakan untuk pengembangan dari data uji akhir sehingga keputusan arsitektur dan ambang tidak disesuaikan berdasarkan hasil pengujian eksternal.



Gambar 4. Arsitektur Deteksi Multicue

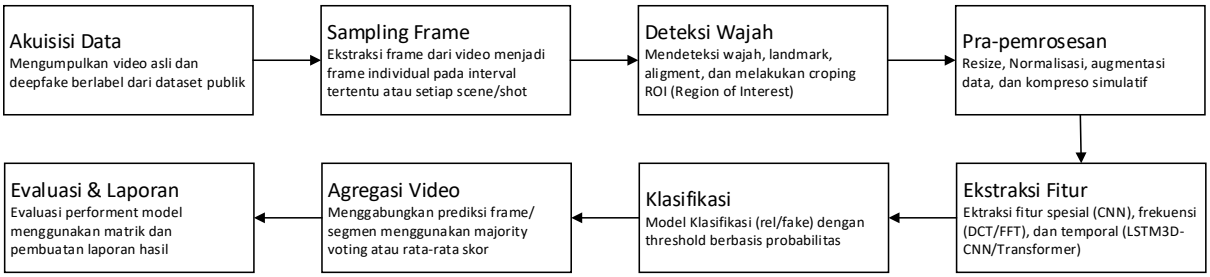
DOI: https://doi.org/10.54082/jupin.2712
p-ISSN: 2808-148X
e-ISSN: 2808-1366

Hubungan utama pada Gambar 4 terletak pada pemrosesan paralel dan penggabungan bukti. Cabang spasial, frekuensi, temporal, dan audio-visual menerima representasi yang diselaraskan, tetapi mempertahankan skor terpisah agar kontribusi serta konflik antarpetunjuk dapat diperiksa. Fusi hanya dilakukan setelah kualitas input dan ketersediaan modalitas dinilai. Provenance tidak digabungkan sebagai skor klasifikasi, melainkan menjadi bukti integritas yang berdiri sejajar. Ketika cabang menghasilkan bukti yang bertentangan, hubungan tersebut harus menurunkan keyakinan dan mengalihkan kasus ke review manusia.



Gambar 5. Metode Pendeteksian Video Deepfake Berbasis Optical Flow dan CNN.

Gambar 5 menunjukkan ketergantungan langsung antara frame RGB dan optical flow: representasi gerak hanya valid jika urutan frame, pelacakan, dan estimasi perpindahan cukup stabil. Flow-CNN kemudian mengubah pola gerak menjadi skor, tetapi tidak menggantikan petunjuk spasial. Dalam rancangan multicue, jalur ini berfungsi sebagai pembanding temporal terhadap hasil citra tunggal. Jika kompresi, frame drop, atau gerakan kamera merusak optical flow, ketidakandalan tersebut harus dicatat sebagai kualitas input dan tidak boleh diterjemahkan otomatis menjadi bukti manipulasi.



Gambar 6. Alur Data dan Inferensi sebagai Alur Implementasi.

Hubungan antartahap pada Gambar 6 membentuk transformasi data yang dapat diaudit. Sampling menghasilkan indeks frame; deteksi wajah menghasilkan jalur dan crop; prapemrosesan menghasilkan tensor serta catatan transformasi; ekstraksi fitur menghasilkan bukti per petunjuk; dan agregasi mengubah skor segmen menjadi keputusan tingkat video. Karena keluaran satu tahap menjadi masukan tahap berikutnya, kegagalan deteksi wajah atau kualitas crop harus diteruskan sebagai status bukti tidak memadai. Evaluasi menggunakan semua rekaman antara tersebut agar penyebab keputusan dapat direkonstruksi.

3. HASIL DAN PEMBAHASAN

3.1 Rancangan Sistem Implementas

3.1.1. Aktor dan Kebutuhan Fungsional

Sistem memiliki tiga aktor. Pengguna mengunggah media dan menerima status pekerjaan. Service worker memproses media, menjalankan model, dan menyimpan file. Reviewer atau analis memeriksa skor, kualitas input, segmen mencurigakan, peta perhatian, serta informasi asal-usul media sebelum membuat keputusan akhir. Administrator model mengelola versi model, nilai batas untuk pengambilan keputusan, dan proses pengembalian ke versi sebelumnya. Pemisahan peran ini mencegah keputusan forensik ditutup secara otomatis oleh satu skor model.

DOI: https://doi.org/10.54082/jupin.2712
p-ISSN: 2808-148X
e-ISSN: 2808-1366

Tabel 5. Kebutuhan Fungsional Utama

ID	Kebutuhan	Masukan	Keluaran / kriteria penerimaan
F1	Unggah dan validasi media	MP4/MOV/WebM, ukuran, checksum	Job ID; file ilegal ditolak; hash tersimpan
F2	Prapemrosesan	Video valid	Frame, face crop, landmark, metadata
F3	Inferensi multicue	Segmen dan audio opsional	Skor tiap cue dan skor gabungan
F4	Agregasi dan kalibrasi	Logit per segmen	Probabilitas terkalibrasi; status uncertain
F5	Pelaporan bukti	Skor, heatmap, provenance	JSON/PDF ringkas; frame bukti
F6	Review manusia	Laporan job	Catatan, keputusan, dan alasan reviewer
F7	Audit dan penghapusan	Job ID / kebijakan retensi	Riwayat akses; media dapat dihapus
F8	Penilaian ketidakpastian	Skor, kualitas input, perbedaan antar cabang	uncertain/insufficient-evidence beserta kode alasan
F9	Kelola versi model	Model, data, konfigurasi, ambang	Versi dapat ditelusuri, diuji regresi, dan dikembalikan

3.1.2 Kebutuhan Nonfungsional dan Batasan

Kebutuhan nonfungsional mencakup keamanan, keterlacakan, kinerja, ketersediaan, dan kemudahan pemeliharaan. Media disimpan terenkripsi, akses API memakai token, serta file dibatasi berdasarkan ukuran, durasi, format, dan MIME type. Setiap pekerjaan memiliki idempotency key berbasis SHA-256 agar unggahan ulang tidak membuat pekerjaan berulang. Worker berjalan asinkron agar API tetap responsif. Target prototipe adalah P95 respons status API di bawah satu detik, penyelesaian analisis video 10 detik di bawah 10 detik pada satu GPU, percobaan ulang maksimum tiga kali, dan kebijakan retensi yang dapat dikonfigurasi. Target tersebut harus diukur kembali pada perangkat keras dan beban yang sebenarnya.

Batasan sistem harus terlihat pada interface dan laporan. Model hanya memberi indikasi kemungkinan manipulasi berdasarkan sinyal yang dipelajari; skor tidak menyatakan pembuat, waktu manipulasi, niat, atau kebenaran isi secara semantik. File tanpa wajah yang cukup jelas, video terlalu pendek, pelacakan wajah tidak stabil, atau audio tidak sesuai harus diberi status bukti-tidak-cukup (insufficient-evidence), bukan otomatis diberi label real. Sistem juga harus menampilkan alasan status tersebut agar pengguna dapat memperbaiki input.

3.1.3. Arsitektur Deployment dan Alur Layanan

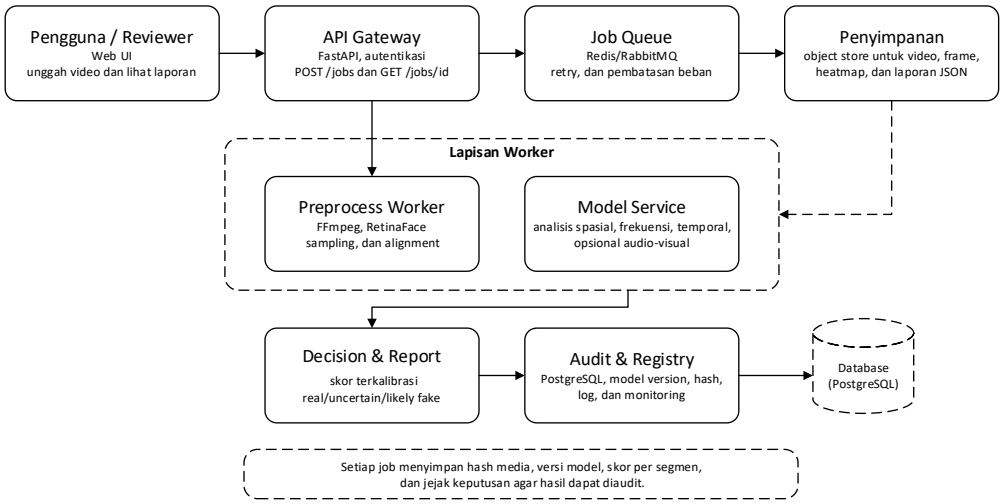
Arsitektur layanan memisahkan jalur kendali dan komputasi. API Gateway menerima permintaan, memeriksa token serta metadata, menghitung hash, lalu membuat pekerjaan pada antrean. Worker mengambil pekerjaan secara idempoten, membaca video dari penyimpanan objek, menjalankan FFmpeg dan RetinaFace untuk sampling, pelacakan, serta alignment, kemudian memanggil layanan model. Layanan model mengembalikan skor per frame, per segmen, dan per jalur wajah. Aggregator menggabungkan skor, menjalankan kalibrasi, menilai kualitas input, menentukan status, dan menghasilkan laporan. PostgreSQL menyimpan metadata serta kejadian, sedangkan penyimpanan objek menampung video, crop wajah, peta perhatian, dan laporan.

Gambar 7 menghubungkan jalur kendali dan jalur komputasi. API Gateway mengelola identitas permintaan dan membuat job; queue memisahkan waktu respons API dari beban GPU; worker mengambil media dari penyimpanan dan memanggil layanan model; sedangkan komponen keputusan, audit, dan registri menyimpan konteks hasil. Media hash dan job ID menjadi pengikat lintas-komponen sehingga retry tidak membuat analisis ganda. Pemisahan ini juga membatasi dampak kegagalan: layanan model dapat diulang tanpa kehilangan permintaan, sementara laporan hanya diterbitkan setelah bukti dan versi model tersimpan.

DOI: https://doi.org/10.54082/jupin.2712
p-ISSN: 2808-148X
e-ISSN: 2808-1366

Tabel 6. Komponen Perangkat Lunak dan Teknologi Referensi

Komponen	Pilihan referensi	Tanggung jawab
API	FastAPI + JWT	Validasi, job lifecycle, akses laporan
Queue	Redis/RabbitMQ	Asinkron, retry, rate limiting
Preprocess	FFmpeg, OpenCV, RetinaFace	Decode, sampling, face alignment
Model	PyTorch + TorchScript/ONNX	Inferensi spasial, frekuensi, temporal
Storage	S3-compatible object store	Media dan file bukti berukuran besar
Metadata	PostgreSQL	Job, skor, model version, audit
Registry	MLflow atau setara	Eksperimen, model, lineage, rollback
Observability	Prometheus/Grafana + log terstruktur	Latency, error, drift, queue depth
Security	Sandbox, KMS/secret manager	Validasi file, isolasi proses, enkripsi, rotasi secret



Gambar 7. Arsitektur Deployment Referensi untuk Sistem yang dibangun.

3.1.4. Alur inferensi, Endpoint, dan Status Keputusan

Pada sistem deteksi deepfake, simbol p(fake) menyatakan probabilitas atau tingkat kemungkinan bahwa media yang dianalisis merupakan deepfake. Nilai tersebut berada pada rentang 0 sampai 1. Nilai yang mendekati 0 menunjukkan kemungkinan deepfake yang rendah, sedangkan nilai yang mendekati 1 menunjukkan kemungkinan deepfake yang tinggi. Sebagai contoh, nilai p(fake)=0,80 berarti model memperkirakan kemungkinan media tersebut sebagai deepfake sebesar 80%.

Simbol τ atau tau digunakan untuk menyatakan ambang keputusan (threshold). Dalam sistem ini digunakan dua ambang, yaitu τlow sebagai ambang bawah dan τhigh sebagai ambang atas. Ambang bawah digunakan untuk menentukan kondisi real-likely, sedangkan ambang atas digunakan untuk menentukan kondisi fake-likely. Nilai di antara kedua ambang tersebut dikategorikan sebagai uncertain karena model belum memiliki bukti yang cukup untuk memberikan keputusan yang tegas.

Aturan keputusan dapat ditulis sebagai berikut:

$$\begin{aligned} & real - likely\ jika\ p(fake) < \tau_{\text{low}} \\ & uncertain\ jika\ \tau_{low} \leq p\,(fake) \leq \tau high \\ & fake\text{-}likely\ jika\ p\,(fake) > \tau_{high} \end{aligned} \qquad (1)$$

Dalam prototipe awal, digunakan τlow = 0,35 dan τhigh = 0,65. Dengan demikian, media dengan nilai p(fake)<0,35 dikategorikan sebagai real-likely, media dengan nilai antara 0,35 sampai 0,65

DOI: https://doi.org/10.54082/jupin.2712
p-ISSN: 2808-148X
e-ISSN: 2808-1366

dikategorikan sebagai uncertain, sedangkan media dengan nilai p(fake)>0,65 dikategorikan sebagai fake-likely.

Tabel 7. Kontrak API

Endpoint	Metode	Respons
/v1/jobs	POST	job_id, media_hash, status=queued
/v1/jobs/{id}	GET	status, progress, error_code
/v1/jobs/{id}/report	GET	label, p_fake, cue_scores, evidence
/v1/jobs/{id}/review	POST	reviewer_decision, note, timestamp
/v1/jobs/{id}	DELETE	retention status dan penghapusan
/v1/health	GET	API, queue, storage, model readiness
/v1/models/active	GET	model_id, checksum, threshold, limitations

3.1.5. Skema Data dan Penyimpanan Bukti

Penyimpanan dipisah menjadi metadata relasional dan bukti file. Tabel media_assets menyimpan hash, ukuran, durasi, MIME type, status provenance, dan kebijakan retensi. Tabel jobs menyimpan pemohon, waktu, status, kode alasan, serta konfigurasi. Tabel face_tracks dan segment_scores menyimpan jalur wajah, rentang waktu, kualitas, skor tiap petunjuk, skor gabungan, dan URI peta perhatian. Tabel model_versions menyimpan checksum model, pembagian dataset, parameter, ambang, dan metrik validasi. Tabel review_decisions serta audit_events menyimpan keputusan, alasan, pelaku, waktu, dan hasil. Pemisahan ini memungkinkan laporan direkonstruksi tanpa menaruh video besar di basis data.

Tabel 8. Entitas Data Minimal

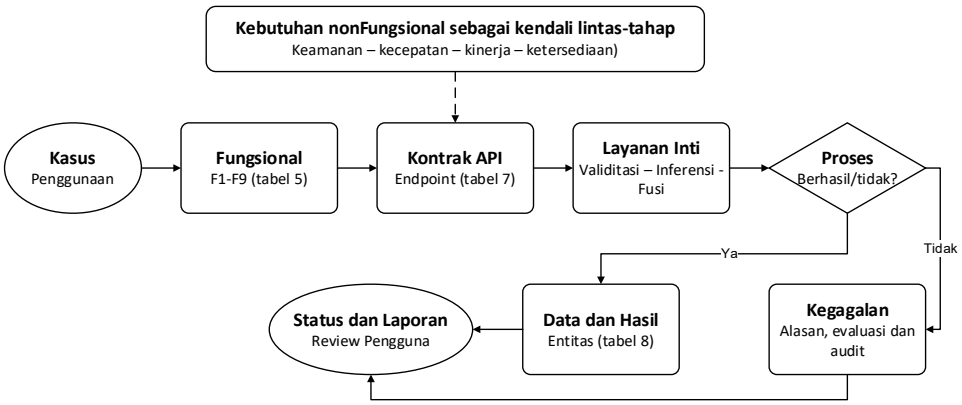
Entitas	Kunci	Atribut	Retensi
media_assets	asset_id	sha256, mime, duration, provenance	sesuai kebijakan
jobs	job_id	asset_id, status, config, created_at	metadata lebih lama
segment_scores	score_id	t_start, t_end, p_spatial, p_freq, p_temp	ikut job
model_versions	model_id	git_sha, checksum, dataset, metrics	permanen
audit_events	event_id	actor, action, timestamp, result	sesuai audit
face_tracks	track_id	asset_id, bbox, quality, detection_rate	ikut job
review_decisions	review_id	job_id, decision, reason, reviewer	sesuai audit

3.1.6. Alur Keterhubungan Spesifikasi Implementasi

Tabel 5, uraian kebutuhan nonfungsional, Tabel 7, dan Tabel 8 tidak berdiri sebagai spesifikasi yang terpisah. Kebutuhan fungsional menetapkan kemampuan yang harus tersedia; kontrak API menerjemahkan kemampuan tersebut menjadi permintaan dan respons yang dapat dipanggil; layanan pemrosesan menjalankan validasi, inferensi, agregasi, dan pelaporan; sedangkan skema data menyimpan status, bukti, serta riwayat keputusan. Pada seluruh tahapan, kebutuhan nonfungsional bertindak sebagai kendali yang membatasi keamanan, latensi, keterlacakan, ketersediaan, dan retensi. Hubungan tersebut dirangkum pada Gambar 8.

Hubungan pada Gambar 8 menunjukkan bahwa setiap kebutuhan harus dapat ditelusuri sampai ke endpoint, proses, dan entitas penyimpanannya. Sebagai contoh, F1 diwujudkan melalui POST /v1/jobs, validasi media, serta pencatatan media_assets dan jobs; F5 diwujudkan melalui endpoint report, agregasi skor, serta segment_scores dan file bukti; sedangkan F6 menghubungkan endpoint review dengan review_decisions dan audit_events. Cabang kegagalan tetap menghasilkan kode alasan dan jejak audit sehingga kegagalan tidak menjadi keadaan tanpa informasi. Dengan alur ini, perubahan pada satu tabel dapat diperiksa dampaknya terhadap kontrak layanan dan struktur data sebelum implementasi dilakukan.

DOI: https://doi.org/10.54082/jupin.2712
p-ISSN: 2808-148X
e-ISSN: 2808-1366



Gambar 8. Alur Keterhubungan Kebutuhan, API, Layanan, dan Skema Data

3.1.7. Keamanan, Pemantauan, dan Pengelolaan Perubahan

Permulaan serangan mencakup file media berbahaya, paket video, URL internal, antrean, model, dan tautan laporan. Mitigasi meliputi allowlist format, pemeriksaan magic bytes, batas waktu decode, sandbox tanpa akses jaringan, pemindaian malware, URL bertanda, pembatasan laju, serta pemisahan hak akses antar layanan. Kegagalan validasi dicatat tanpa menyimpan isi sensitif yang tidak diperlukan.

Setiap perubahan data, kode, dependensi, bobot, ambang, atau konfigurasi menghasilkan versi baru. Promosi model mensyaratkan lulus pengujian regresi, keamanan, kalibrasi, dan performa lintas-dataset. Pemantauan produksi membandingkan distribusi kualitas input, tingkat wajah terdeteksi, skor, status uncertain, dan kesalahan layanan terhadap baseline. Peringatan tidak langsung memicu pelatihan ulang; perubahan harus ditinjau dan dapat dikembalikan ke versi sebelumnya.

3.2 Rancangan Model Multicue dan Training

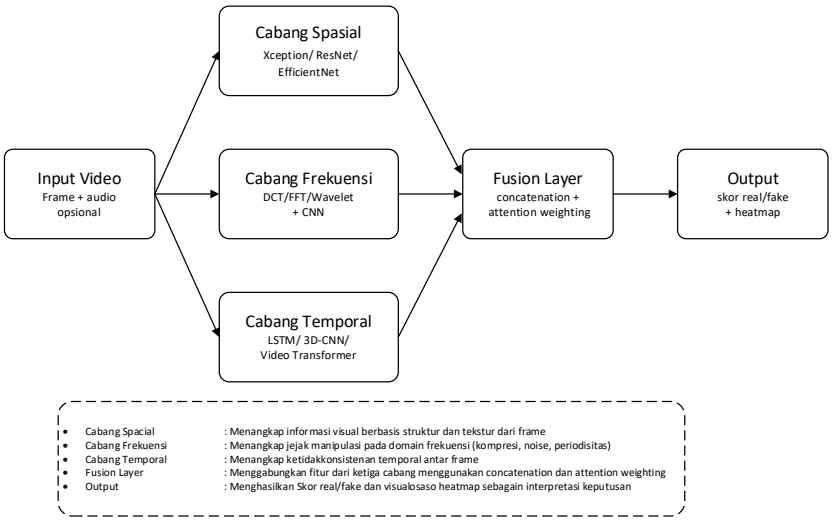
Input video dibagi menjadi segmen dua detik dengan maksimum 16 frame terpilih per segmen. RetinaFace mendeteksi wajah, landmark dipakai untuk alignment, dan crop diubah menjadi 224×224 piksel. Wajah dilacak antarframe sehingga beberapa orang dalam satu video tidak tercampur. Jika ada beberapa jalur wajah, sistem menyimpan skor per jalur dan menandai jalur yang paling berpengaruh terhadap hasil. Augmentasi mencakup pembalikan horizontal, perubahan pencahayaan, blur ringan, kompresi JPEG, perubahan ukuran, dan penghilangan frame. Seluruh transformasi disimpan dalam konfigurasi eksperimen agar pengujian ketahanan dapat diulang.

Pada MVP, backbone spasial menghasilkan vektor fitur sepanjang 1280 nilai dari EfficientNet-B0 atau Xception. Cabang temporal menerima urutan vektor tersebut dan menghasilkan ringkasan sepanjang 256 nilai melalui GRU. Cabang frekuensi mengubah crop menjadi kanal DCT frekuensi tinggi dan memprosesnya dengan CNN ringkas. Cabang audio opsional mengubah suara menjadi log-mel spectrogram dan mengukur keselarasan dengan gerak mulut. Vektor dari cabang aktif digabungkan pada fusion MLP, sedangkan indikator kualitas input dipakai untuk mencegah skor tinggi dari frame yang buram atau wajah yang terlalu kecil.

Pada Gambar 9, setiap cabang menerima crop wajah yang sama tetapi menghasilkan representasi berbeda. Cabang spasial menekankan tekstur, cabang frekuensi menekankan jejak transformasi, dan cabang temporal menekankan perubahan antarframe. Fusion layer tidak sekadar menggabungkan vektor; bobot fusi harus mempertimbangkan kualitas input dan ketersediaan cabang. Skor per cabang, segmen dominan, dan heatmap disimpan bersama agar skor akhir dapat ditelusuri. Konflik yang besar antar cabang menjadi sinyal ketidakpastian, bukan alasan untuk memilih skor tertinggi.

Penanganan ketidakseimbangan kelas dilakukan dengan menerapkan weighted binary cross-entropy sebagai fungsi loss, sementara focal loss dievaluasi sebagai pembanding. Jika label manipulasi tidak lengkap, Self-Blended Images dapat digunakan sebagai sumber pseudo-fake, tetapi data tersebut dipisahkan dari data uji. Early stopping dipantau dengan AUC validasi, PR-AUC, dan log-loss, bukan accuracy saja. Model terbaik didaftarkan bersama hash dataset, konfigurasi augmentasi, seed, commit kode, checksum bobot, hasil kalibrasi, serta ambang keputusan.

DOI: https://doi.org/10.54082/jupin.2712
p-ISSN: 2808-148X
e-ISSN: 2808-1366



Gambar 9. Rancangan Arsitektur Model Multicue dan Keluaran Bukti.

Penggunaan pseudo-fake tersebut mengikuti strategi Self-Blended Images dan harus dilaporkan terpisah dari evaluasi pada manipulasi nyata (Shiohara & Yamasaki, 2022).

Audit kebocoran dilakukan sebelum pelatihan dan sebelum evaluasi. Frame dari video yang sama tidak boleh tersebar ke beberapa split; identitas serta video sumber dipisahkan; kemiripan tinggi diperiksa dengan perceptual hash; dan generator untuk pengujian cross-generator ditahan seluruhnya dari pelatihan. Data uji tidak digunakan untuk memilih epoch, ambang, transformasi, atau bobot fusi. Setiap hasil harus dapat dihubungkan kembali ke data teramati dan konfigurasi yang tepat.

Tabel 9. Komponen Training dan Output

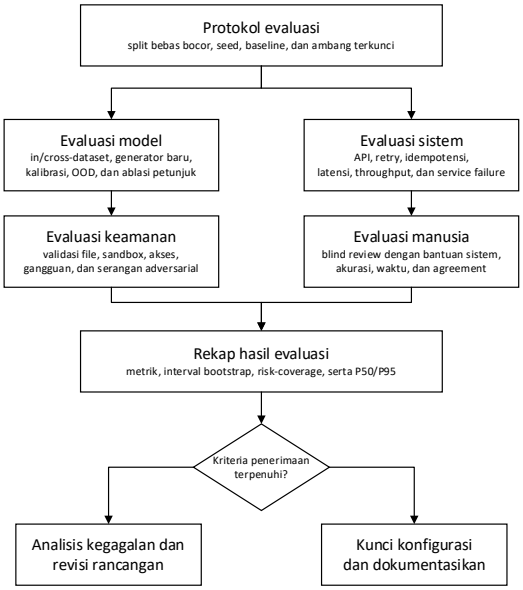
Tahap	Konfigurasi awal	Hasil yang disimpan
Split	by-identity; train/val/test	manifest dan checksum
Sampling	2 detik; maksimum 16 frame	index frame per segmen
Baseline	EfficientNet/Xception + GRU	checkpoint dan log
Fusion	MLP; cue dapat diablasikan	fusion weights
Calibration	temperature scaling	temperature T dan ECE
Export	TorchScript/ONNX	model version dan signature
Leakage audit	hash, identitas, sumber, generator	laporan duplikat dan manifest split
Ragu/OOD	temperature scaling + risk–coverage	temperature, ECE, ambang, coverage

3.3 Rancangan Evaluasi

Evaluasi dipisahkan menjadi evaluasi model, sistem, keamanan, dan manusia. Evaluasi model mencakup in-dataset, cross-dataset, cross-generator, cross-manipulation, gangguan alami, kalibrasi, data di luar distribusi latih, serta pengujian kontribusi tiap cabang. Evaluasi sistem menguji fungsi API, retry, idempotensi, validasi file, latensi, throughput, dan ketahanan saat layanan model atau penyimpanan tidak siap. Evaluasi keamanan mencakup gangguan kecil yang sengaja dirancang untuk menipu detektor. Evaluasi manusia menilai apakah reviewer dapat menemukan bukti, memahami ketidakpastian, dan menghindari ketergantungan berlebihan pada label otomatis. Alur rancangan evaluasi dirangkum pada Gambar 10.

Hubungan pada Gambar 10 menegaskan bahwa evaluasi model, sistem, keamanan, dan manusia merupakan gerbang yang saling melengkapi. Performa model yang baik tidak dapat menutup kegagalan keamanan atau latensi, sedangkan sistem yang cepat belum layak digunakan jika reviewer tidak memahami ketidakpastian. Semua jalur bertemu pada rekap hasil dan kriteria penerimaan. Kegagalan salah satu jalur mengembalikan temuan ke analisis kegagalan dan revisi rancangan; hanya konfigurasi yang memenuhi seluruh persyaratan yang dikunci dan didokumentasikan.

DOI: https://doi.org/10.54082/jupin.2712
p-ISSN: 2808-148X
e-ISSN: 2808-1366



Gambar 10. Flow Rancangan Evaluasi Model, Sistem, Keamanan, dan Manusia.

Metrik klasifikasi meliputi accuracy, precision, recall, F1-score, ROC-AUC, PR-AUC, EER, Brier score, dan log-loss. ECE serta reliability plot digunakan untuk kalibrasi; risk–coverage curve digunakan untuk keputusan yang dapat ditunda; median dan P95 digunakan untuk latensi. Setiap hasil dilaporkan dengan interval kepercayaan bootstrap dan dipisahkan menurut dataset, generator, teknik manipulasi, kompresi, kualitas wajah, dan kelompok demografis yang tersedia secara sah. Hasil rata-rata tidak boleh menutupi kegagalan pada satu kelompok atau kondisi.

Tabel 10. Protokol Uji dan Kriteria Penerimaan Desain

Uji	Prosedur	Output	Target awal
Fungsional	Upload, invalid file, retry, delete	Pass/fail per F1–F7	100% kasus kritis lulus
Dalam-data	Train/test pada dataset sama	AUC, F1, confusion matrix	baseline terdokumentasi
Lintas-data	Train FF++; test Celeb-DF/WildDeepfake	gap generalisasi	gap dilaporkan, bukan disembunyikan
Ketahanan	JPEG, blur, noise, resize, frame drop	ΔAUC dan ΔF1	degradasi terukur
Kalibrasi	Temperature scaling validation	ECE, reliability plot	ECE ≤ 0,05 sebagai target
Operasional	10 detik video; satu GPU	P50/P95 latency	P95 ≤ 10 detik sebagai target
Penjelasan	Grad-CAM dan top segments	heatmap serta agreement reviewer	bukti dapat ditelusuri
Generator baru	Latih pada sebagian teknik; uji generator baru	AUC/F1 dan gap	Gap dan interval dilaporkan
Shift/OOD	Uji data terkini dan kualitas berbeda	ECE, AURC, coverage	Kasus ragu dialihkan ke review
Serangan	Gangguan kecil dan serangan transfer	ΔAUC, attack success	Kerentanan dan mitigasi dicatat
Keadilan	Pisahkan hasil menurut kelompok yang sah	FPR/FNR per kelompok	Gap, ukuran sampel, coverage
Manusia	Review dibutakan dengan/tanpa sistem	Akurasi, waktu, agreement	Tidak menambah ketergantungan salah

DOI: https://doi.org/10.54082/jupin.2712
p-ISSN: 2808-148X
e-ISSN: 2808-1366

3.4 Dasar Penetapan Target Penerimaan

Nilai ECE <= 0,05 dan P95 <= 10 detik pada Tabel 10 tidak diambil sebagai ambang universal dari penelitian terdahulu dan bukan hasil pengukuran penelitian ini. Keduanya ditetapkan a priori sebagai target rancangan agar prototipe memiliki kriteria lulus yang dapat diuji. ECE <= 0,05 menyatakan sasaran agar perbedaan rata-rata antara tingkat keyakinan dan proporsi prediksi benar tidak melebihi lima poin persentase pada skema binning yang dilaporkan. P95 <= 10 detik untuk video berdurasi 10 detik pada satu GPU merupakan sasaran layanan mendekati waktu nyata pada skenario batch, sedangkan P95 respons status API di bawah satu detik merupakan sasaran respons jalur kendali.

Pada tahap implementasi, ECE harus dihitung bersama jumlah bin, ukuran sampel, interval kepercayaan, dan reliability plot; P95 harus diukur secara end-to-end sejak job diterima sampai laporan tersedia, dengan spesifikasi GPU, durasi video, jumlah wajah, dan beban antrean yang dinyatakan. Apabila target tidak tercapai, hasil harus dilaporkan sebagai kegagalan terhadap kriteria rancangan dan digunakan untuk memperbaiki model, pipeline, atau kapasitas layanan. Target dapat direvisi setelah baseline empiris tersedia, tetapi perubahan harus dilakukan sebelum uji akhir dan disertai justifikasi agar tidak menjadi penyesuaian pascahasil.

Pengujian kontribusi dilakukan pada konfigurasi spasial saja, spasial–temporal, spasial–frekuensi, spasial–audio, dan gabungan penuh. Tujuannya bukan mencari model terbesar, melainkan kombinasi yang paling seimbang antara generalisasi, latensi, kebutuhan memori, dan kemudahan penafsiran. Ambang dipilih pada data validasi, dikunci sebelum pengujian, dan tidak diubah berdasarkan hasil data uji. Setiap konfigurasi dijalankan dengan sedikitnya tiga seed serta dilaporkan bersama rata-rata dan variasinya.

Perbandingan model menggunakan bootstrap berpasangan pada unit video, bukan frame, agar frame yang berkorelasi tidak dianggap sebagai sampel independen. Selain nilai titik, laporan memuat interval kepercayaan, jumlah video, jumlah identitas, jumlah generator, proporsi kelas, dan seed. Jika banyak konfigurasi dibandingkan, pemilihan model dilakukan pada data validasi dan hanya model terpilih yang dievaluasi sekali pada data uji akhir.

Evaluasi manusia menggunakan tugas yang dibutakan dengan dua kondisi: reviewer tanpa bantuan sistem dan reviewer dengan laporan sistem. Ukuran yang dicatat meliputi ketepatan keputusan, waktu pemeriksaan, tingkat kesepakatan antar-reviewer, perubahan keputusan setelah melihat bukti, serta jumlah kasus ketika reviewer mengikuti model yang salah. Hasil ini menentukan apakah heatmap, skor segmen, dan status uncertain benar-benar membantu, bukan sekadar terlihat meyakinkan.

3.5. Pembahasan Rancangan

3.5.1 Hasil Perancangan

Hasil utama penelitian ini adalah spesifikasi sistem yang dapat dibangun tanpa mengubah tujuan deteksi. Pertama, alur kerja telah dipetakan dari unggah media sampai laporan dan review. Kedua, batas layanan antara API, queue, worker, model, penyimpanan, dan audit telah ditentukan agar pengembangan dapat dibagi antar modul. Ketiga, kebutuhan fungsional, kontrak endpoint, skema data, kriteria penerimaan, dan alasan kegagalan ditulis secara jelas. Keempat, model tidak hanya menghasilkan label, tetapi juga skor per petunjuk, skor per segmen dan jalur wajah, peta perhatian, hash, versi model, status provenance, serta ukuran ketidakpastian.

Rancangan tersebut memungkinkan implementasi bertahap. Sprint pertama menghasilkan API, penyimpanan, prapemrosesan, pelacakan wajah, dan baseline spasial. Sprint kedua menambahkan GRU temporal, kalibrasi, status ragu, serta laporan bukti. Sprint ketiga menambahkan DCT dan audio–visual. Sprint keempat mengaktifkan registri model, pemantauan perubahan distribusi, pengujian keamanan, dan dashboard reviewer. Setiap sprint dapat disimulasikan dengan dataset kecil sebelum skala diperbesar.

3.5.2 Analisis Pemilihan Multicue dibandingkan Single-Cue

Pemilihan multicue didasarkan pada perbedaan jenis bukti dan pola kegagalan yang dilaporkan penelitian terdahulu. Pendekatan spasial mampu menyoroti ketidakkonsistenan lokal dan hubungan gerak wajah (Zhao et al., 2021; Bai et al., 2023), tetapi cirinya dapat berubah akibat generator dan

DOI: https://doi.org/10.54082/jupin.2712
p-ISSN: 2808-148X
e-ISSN: 2808-1366

kompresi. Dong et al. (2023) menunjukkan bahwa petunjuk frekuensi menambah informasi di luar RGB, sedangkan AltFreezing menggabungkan pembelajaran spasial-temporal untuk memperbaiki generalisasi pada manipulasi yang tidak terlihat saat pelatihan (Wang et al., 2023). Ilyas et al. (2023) dan Javed et al. (2025) memperlihatkan bahwa sinkronisasi audio-visual menyediakan bukti tambahan yang tidak tersedia pada detektor visual tunggal. Sementara itu, DeepfakeBench dan DF40 menunjukkan bahwa peringkat metode dapat berubah menurut protokol, dataset, dan keluarga generator (Yan et al., 2023; Yan et al., 2024).

Tabel 11. Analisis Pendekatan Single-Cue dan Multicue

Aspek	Pendekatan single-cue	Implikasi pendekatan multicue	Rujukan langsung
Cakupan bukti	Mengandalkan satu sumber, misalnya tekstur spasial, frekuensi, gerak, atau sinkronisasi.	Menggabungkan bukti yang berbeda dan menyimpan skor per cabang agar kontribusinya dapat diperiksa.	Zhao et al. (2021); Dong et al. (2023); Ilyas et al. (2023).
Pergeseran distribusi	Kinerja dapat turun tajam ketika ciri utama berubah akibat generator, dataset, atau kompresi baru.	Cabang lain dapat mempertahankan bukti ketika satu petunjuk melemah, tetapi manfaatnya harus diuji lintas-dataset.	Wang et al. (2023); Yan et al. (2023, 2024).
Input hilang atau rusak	Keputusan kehilangan dasar ketika petunjuk tunggal tidak tersedia atau kualitasnya rendah.	Cabang dinonaktifkan berdasarkan quality gating; sistem tetap bekerja dan mencatat bukti yang tidak tersedia.	Ilyas et al. (2023); Javed et al. (2025).
Keterlacakan	Lebih sederhana dijelaskan, tetapi tidak menunjukkan apakah petunjuk lain mendukung atau menolak prediksi.	Konflik antarskor dapat digunakan untuk menurunkan keyakinan, memunculkan status uncertain, dan memicu review.	Bai et al. (2023); Momin et al. (2025).
Biaya dan validasi	Lebih ringan dan sesuai sebagai baseline, tetapi cakupan kegagalannya lebih sempit.	Lebih mahal; setiap cabang hanya dipertahankan jika ablation study menunjukkan manfaat generalisasi, kalibrasi, atau bukti.	Kaur et al. (2024); Yan et al. (2023).

Dengan demikian, multicue dipilih bukan karena penambahan cabang selalu meningkatkan akurasi, melainkan karena rancangan perlu menghadapi kegagalan yang tidak berkorelasi dan menyediakan bukti yang dapat dibandingkan. Single-cue tetap dipertahankan sebagai baseline dan sebagai konfigurasi fallback berbiaya rendah. Suatu cabang hanya layak masuk ke konfigurasi akhir apabila ablation study menunjukkan perbaikan pada generalisasi, kalibrasi, atau keterlacakan yang sebanding dengan tambahan latensi dan penggunaan memori. Jika manfaat tersebut tidak muncul, cabang harus dihapus atau tetap bersifat opsional.

3.5.3 Pembahasan Generalisasi dan Ketidakpastian

Penggunaan beberapa petunjuk merupakan respons terhadap perbedaan kegagalan tiap cabang. Cabang spasial dapat mempelajari seperti kompresi atau latar dataset; cabang frekuensi dapat rusak oleh pengodean ulang; cabang temporal memerlukan frame berurutan dan lebih mahal; cabang audio–visual sensitif terhadap sinkronisasi serta bahasa. Fusi tidak menghapus kelemahan tersebut, tetapi menyediakan cara untuk mengukur kontribusi dan menandai kasus ketika hasil cabang saling bertentangan.

Rentang keyakinan dan status uncertain menjadi bagian penting dari rancangan. Jika skor spasial tinggi tetapi temporal rendah, sistem tidak boleh langsung menaikkan keyakinan. Aggregator menyimpan distribusi skor dan menampilkan segmen yang paling berpengaruh. Kalibrasi membantu menafsirkan probabilitas, tetapi pergeseran distribusi dapat kembali merusak kalibrasi. Oleh sebab itu,

DOI: https://doi.org/10.54082/jupin.2712
p-ISSN: 2808-148X
e-ISSN: 2808-1366

reviewer tetap memeriksa konteks, sumber, metadata, kualitas input, dan provenance. Bukti visual yang terlacak membantu manusia memeriksa keputusan model secara kritis (Mathews et al., 2023).

3.5.4 Pembahasan Kesiapan Implementasi

Secara rekayasa, rancangan dapat diimplementasikan dengan perangkat lunak open-source dan komponen umum: FFmpeg/OpenCV untuk video, RetinaFace untuk deteksi serta pelacakan wajah, PyTorch untuk model, FastAPI untuk layanan, Redis atau RabbitMQ untuk antrean, PostgreSQL untuk metadata, dan penyimpanan objek yang kompatibel dengan S3 untuk file bukti. Pilihan tersebut bukan klaim bahwa satu stack selalu terbaik; komponen dapat diganti selama kontrak API, skema data, keamanan, dan kriteria evaluasi dipertahankan.

Kesiapan produksi ditentukan oleh pengujian dan pemantauan, bukan oleh satu nilai AUC. Sistem memantau perubahan resolusi, durasi, rasio wajah terdeteksi, kualitas crop, skor model, persentase status, panjang antrean, error rate, dan latensi. Model baru menjalani shadow deployment, pengujian regresi, persetujuan reviewer, serta uji pengembalian versi sebelum menjadi aktif. Praktik ini mengikuti gagasan ML Test Score bahwa kesiapan produksi memerlukan serangkaian pengujian dan kebutuhan pemantauan yang jelas. Rangkaian pengujian, deployment, registri versi, dan pemantauan tersebut sejalan dengan MLOps sebagai siklus operasional terpadu (Kreuzberger et al., 2023).

3.5.5 Batasan Hasil

Artikel ini belum menjalankan eksperimen empiris sehingga angka AUC, F1, latensi, ECE, dan risk–coverage belum dapat dilaporkan sebagai hasil penelitian. Target pada Tabel 10 adalah kriteria penerimaan awal yang harus diuji pada tahap implementasi. Tidak adanya angka empiris merupakan batas klaim yang sengaja dijaga agar rancangan tidak disalahartikan sebagai sistem yang telah tervalidasi. Validasi berikutnya harus mencakup DF40 dan data dunia nyata terkini, bukan hanya benchmark lama.

3.5.6 Analisis Kegagalan dan Mitigasi

Mode kegagalan yang perlu dicatat mencakup wajah terlalu kecil, oklusi, gerakan cepat, banyak wajah, frame hilang, audio tidak tersedia, sinkronisasi rusak, kompresi berat, dan teknik pembuatan yang tidak dikenali. Respons sistem dibedakan menurut kondisi: melakukan sampling ulang, memilih jalur wajah tertentu, menonaktifkan cabang yang tidak memiliki input, menurunkan cakupan keputusan otomatis, atau memberi status insufficient-evidence. Sistem tidak boleh mengganti sinyal yang hilang dengan skor netral tanpa mencatat alasannya.

Kegagalan juga dapat berasal dari ketergantungan pada ciri dataset, kalibrasi yang rusak setelah pergeseran distribusi, dan serangan yang dirancang untuk menghindari deteksi. Mitigasi mencakup data yang lebih beragam, pengujian generator baru, pengodean ulang acak saat pengujian ketahanan, model pembanding, pemantauan perubahan distribusi, serta review manusia. Setiap mitigasi harus dievaluasi karena peningkatan pada satu gangguan dapat menurunkan performa pada kondisi lain.

3.6 Implikasi Etika, Privasi, dan Keamanan

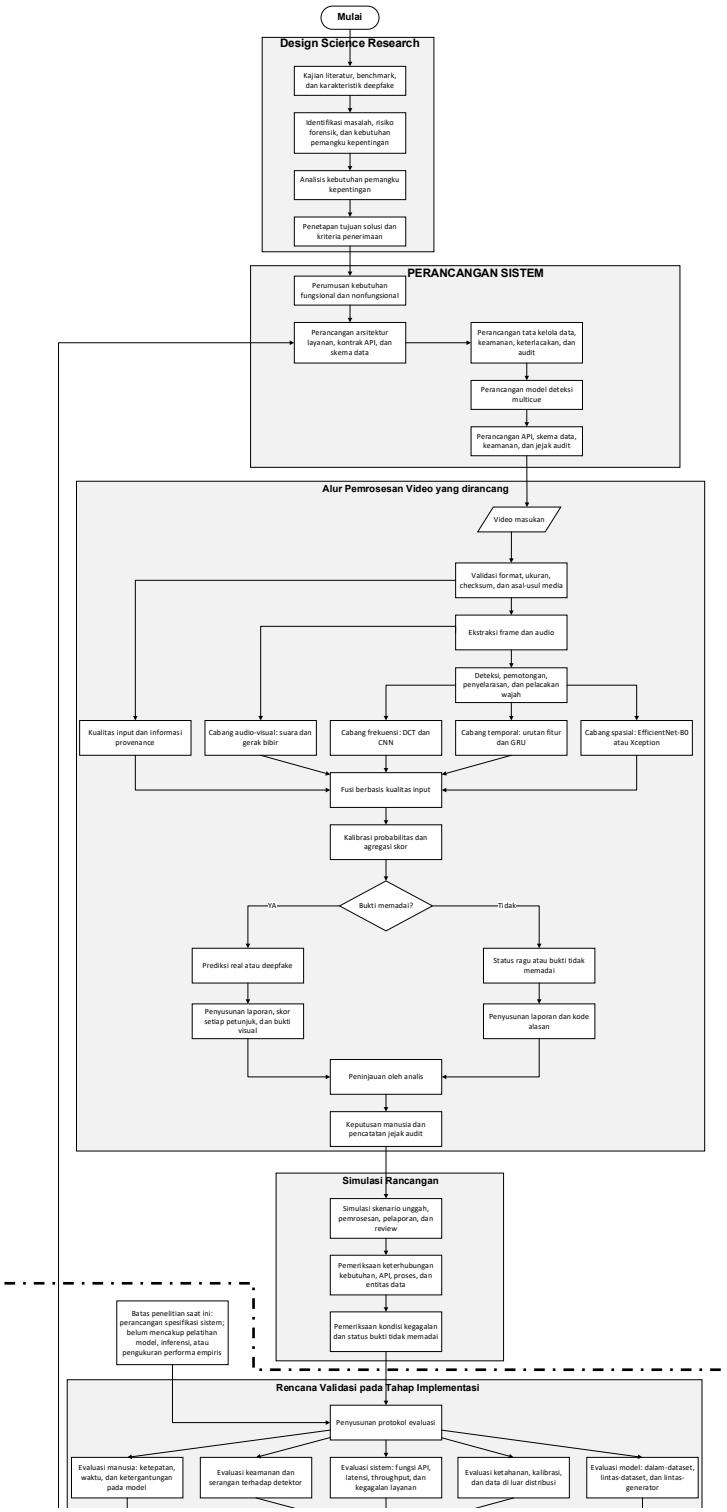
Video wajah dan suara merupakan data sensitif. Dataset harus memiliki lisensi serta dasar penggunaan yang jelas, partisipan perlu mendapat informasi yang memadai, dan contoh konten tidak boleh dipublikasikan ulang tanpa alasan yang sah. Sistem menerapkan minimisasi data: hash dan metadata dapat disimpan lebih lama daripada video asli, tautan file bukti memiliki masa berlaku, dan penghapusan mengikuti kebijakan. Setiap dataset disertai lembar dokumentasi yang menjelaskan tujuan, komposisi, proses pengumpulan, batas penggunaan, lisensi, serta riwayat pemeliharaan.

Keadilan performa perlu diuji ketika atribut demografis tersedia dan penggunaannya sah. Perbedaan false positive atau false negative antar kelompok dapat menimbulkan tuduhan palsu atau perlindungan yang tidak merata. Laporan harus menyebut cakupan data, kelompok yang tidak terwakili, ukuran sampel, dan kondisi ketika model tidak layak digunakan. Model card mencatat tujuan, data, batasan, metrik, ambang, risiko, dan prosedur pengawasan manusia.

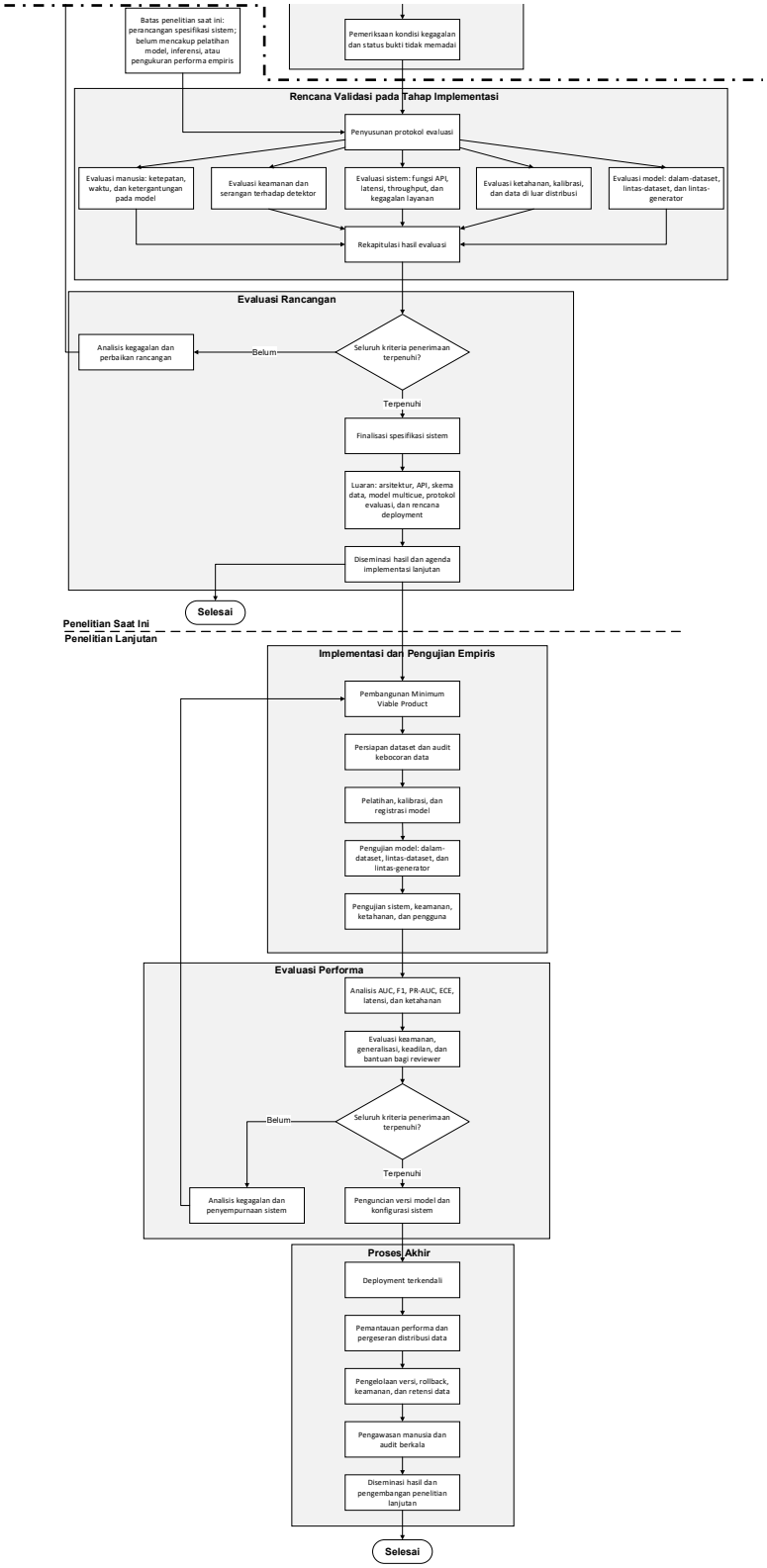
DOI: https://doi.org/10.54082/jupin.2712
p-ISSN: 2808-148X
e-ISSN: 2808-1366

Keamanan sistem mengikuti pertahanan berlapis: validasi file dan sandbox FFmpeg, pembatasan ukuran serta durasi, isolasi worker GPU, token akses, enkripsi, rotasi secret, log append-only, dan pemindaian malware. Hash media dan manifest provenance tidak menggantikan deteksi; keduanya menjaga integritas dan membantu verifikasi. Serangan terhadap model, termasuk gangguan kecil yang menurunkan performa detektor, harus masuk dalam pengujian keamanan. Sistem tidak boleh menjadi satu-satunya dasar keputusan hukum atau sanksi.

3.7 Lampiran



DOI: https://doi.org/10.54082/jupin.2712
p-ISSN: 2808-148X
e-ISSN: 2808-1366



Gambar 11. Flowchart Penelitan saat ini dan Penelitian Lanjutan

4. KESIMPULAN

Penelitian ini menghasilkan tiga kontribusi utama: arsitektur sistem deteksi deepfake multicue yang mengintegrasikan petunjuk spasial, frekuensi, temporal, dan audio-visual dengan kalibrasi serta penilaian ketidakpastian; rancangan implementasi yang mencakup alur layanan asinkron, kontrak API, skema data, penyimpanan bukti, audit, dan review manusia; serta protokol evaluasi yang memisahkan pengujian model, sistem, keamanan, dan manusia. Ketiga keluaran tersebut membentuk spesifikasi sistem yang siap diimplementasikan dan diuji tanpa mengklaim performa empiris pada tahap perancangan ini.

Validasi performa merupakan tahap penelitian berikutnya setelah implementasi sistem selesai dibangun. Tahap tersebut perlu menguji generalisasi lintas-dataset dan lintas-generator, kalibrasi, latensi, ketahanan, keamanan, serta kegunaan laporan bagi reviewer berdasarkan protokol dan target penerimaan yang telah ditetapkan.

Rancangan ini dapat digunakan sebagai acuan teknis dan metodologis bagi penelitian selanjutnya dalam mengembangkan, membandingkan, dan memvalidasi sistem deteksi deepfake yang modular, terlacak, serta berorientasi pada bukti.

DAFTAR PUSTAKA

Abbas, F., & Taeihagh, A. (2024). Unmasking deepfakes: A systematic review of deepfake detection and generation techniques using artificial intelligence. *Expert Systems with Applications, 252*, Article 124260. https://doi.org/10.1016/j.eswa.2024.124260

Akoka, J., Comyn-Wattiau, I., Prat, N., & Storey, V. C. (2023). Knowledge contributions in design science research: Paths of knowledge types. *Decision Support Systems, 166*, Article 113898. https://doi.org/10.1016/j.dss.2022.113898

Bai, W., Liu, Y., Zhang, Z., Li, B., & Hu, W. (2023). AUNet: Learning relations between action units for face forgery detection. In *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 24709–24719). IEEE. https://doi.org/10.1109/CVPR52729.2023.02367

Ciftci, U. A., Demir, I., & Yin, L. (2020). FakeCatcher: Detection of synthetic portrait videos using biological signals. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. Advance online publication. https://doi.org/10.1109/TPAMI.2020.3009287

Cozzolino, D., Rössler, A., Thies, J., Nießner, M., & Verdoliva, L. (2021). ID-Reveal: Identity-aware deepfake video detection. In *2021 IEEE/CVF International Conference on Computer Vision (ICCV)* (pp. 15088–15097). IEEE. https://doi.org/10.1109/ICCV48922.2021.01483

Dong, F., Zou, X., Wang, J., & Liu, X. (2023). Contrastive learning-based general deepfake detection with multi-scale RGB frequency clues. *Journal of King Saud University—Computer and Information Sciences, 35*(4), 90–99. https://doi.org/10.1016/j.jksuci.2023.03.005

Dresch, A., Lacerda, D. P., & Antunes, J. A. V., Jr. (2015). *Design science research: A method for science and technology advancement*. Springer. https://doi.org/10.1007/978-3-319-07374-3

Gebru, T., Morgenstern, J., Vecchione, B., Vaughan, J. W., Wallach, H., Daumé III, H., & Crawford, K. (2021). Datasheets for datasets. *Communications of the ACM, 64*(12), 86–92. https://doi.org/10.1145/3458723

Gregor, S., & Hevner, A. R. (2013). Positioning and presenting design science research for maximum impact. *MIS Quarterly, 37*(2), 337–355. https://doi.org/10.25300/MISQ/2013/37.2.01

Hevner, A. R., & Chatterjee, S. (2010). *Design research in information systems: Theory and practice*. Springer. https://doi.org/10.1007/978-1-4419-5653-8

Hevner, A. R., March, S. T., Park, J., & Ram, S. (2004). Design science in information systems research. *MIS Quarterly, 28*(1), 75–105. https://doi.org/10.2307/25148625

Ilyas, H., Javed, A., & Malik, K. M. (2023). AVFakeNet: A unified end-to-end Dense Swin Transformer deep learning model for audio–visual deepfakes detection. *Applied Soft Computing, 136*, Article 110124. https://doi.org/10.1016/j.asoc.2023.110124

Ismail, A., Elpeltagy, M., Zaki, M. S., & Eldahshan, K. (2022). An integrated spatiotemporal-based methodology for deepfake detection. *Neural Computing and Applications, 34*, 21777–21791. https://doi.org/10.1007/s00521-022-07633-3

Javed, M., Zhang, Z., Dahri, F. H., Laghari, A. A., Krajčík, M., & Almadhor, A. (2025). Audio–visual synchronization and lip movement analysis for real-time deepfake detection. *International Journal of Computational Intelligence Systems, 18*, Article 170. https://doi.org/10.1007/s44196-025-00911-7

Juefei-Xu, F., Wang, R., Huang, Y., Guo, Q., Ma, L., & Liu, Y. (2022). Countering malicious deepfakes: Survey, battleground, and horizon. *International Journal of Computer Vision, 130*(7), 1678–1734. https://doi.org/10.1007/s11263-022-01606-8

Kaur, A., Noori Hoshyar, A., Saikrishna, V., Firmin, S., & Xia, F. (2024). Deepfake video detection: Challenges and opportunities. *Artificial Intelligence Review, 57*, Article 159. https://doi.org/10.1007/s10462-024-10810-6

Kreuzberger, D., Kühl, N., & Hirschl, S. (2023). Machine learning operations (MLOps): Overview, definition, and architecture. *IEEE Access, 11*, 31866–31879. https://doi.org/10.1109/ACCESS.2023.3262138

Mathews, S., Trivedi, S., House, A., Povolny, S., & Fralick, C. (2023). An explainable deepfake detection framework on a novel unconstrained dataset. *Complex & Intelligent Systems, 9*(4), 4425–4437. https://doi.org/10.1007/s40747-022-00956-7

Mirsky, Y., & Lee, W. (2021). The creation and detection of deepfakes: A survey. *ACM Computing Surveys, 54*(1), Article 7. https://doi.org/10.1145/3425780

Momin, M. S., Sufian, A., Barman, D., Leo, M., Distante, C., & Damer, N. (2025). Explainable deepfake detection across different modalities: An overview of methods and challenges. *Image and Vision Computing, 163*, Article 105738. https://doi.org/10.1016/j.imavis.2025.105738

Nguyen, T. T., Nguyen, Q. V. H., Nguyen, D. T., Nguyen, D. T., Huynh-The, T., Nahavandi, S., Nguyen, T. T., Pham, Q.-V., & Nguyen, C. M. (2022). Deep learning for deepfakes creation and detection: A survey. *Computer Vision and Image Understanding, 223*, Article 103525. https://doi.org/10.1016/j.cviu.2022.103525

Peffers, K., Tuunanen, T., Rothenberger, M. A., & Chatterjee, S. (2007). A design science research methodology for information systems research. *Journal of Management Information Systems, 24*(3), 45–77. https://doi.org/10.2753/MIS0742-1222240302

Rana, M. S., Nobi, M. N., Murali, B., & Sung, A. H. (2022). Deepfake detection: A systematic literature review. *IEEE Access, 10*, 25494–25513. https://doi.org/10.1109/ACCESS.2022.3154404

Shiohara, K., & Yamasaki, T. (2022). Detecting deepfakes with self-blended images. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 18699–18708). IEEE. https://doi.org/10.1109/CVPR52688.2022.01816

Stroebel, L., Llewellyn, M., Hartley, T., Ip, T. S., & Ahmed, M. (2023). A systematic literature review on the effectiveness of deepfake detection techniques. *Journal of Cyber Security Technology, 7*(2), 83–113. https://doi.org/10.1080/23742917.2023.2192888

Tolosana, R., Romero-Tapiador, S., Vera-Rodriguez, R., Gonzalez-Sosa, E., & Fierrez, J. (2022). Deepfakes detection across generations: Analysis of facial regions, fusion, and performance evaluation. *Engineering Applications of Artificial Intelligence, 110*, Article 104673. https://doi.org/10.1016/j.engappai.2022.104673

van Lierop, S., Najdenkoska, I., Worring, M., & Geradts, Z. (2026). Interpol review of detection of AI-generated image and video deepfakes, 2022–2025. *Forensic Science International: Synergy, 13*, Article 100704. https://doi.org/10.1016/j.fsisyn.2026.100704

vom Brocke, J., Hevner, A., & Maedche, A. (2020). Introduction to design science research. In J. vom Brocke, A. Hevner, & A. Maedche (Eds.), *Design science research: Cases* (pp. 1–13). Springer. https://doi.org/10.1007/978-3-030-46781-4_1

DOI: https://doi.org/10.54082/jupin.2712
p-ISSN: 2808-148X
e-ISSN: 2808-1366

Wang, Z., Bao, J., Zhou, W., Wang, W., & Li, H. (2023). AltFreezing for more general video face forgery detection. In *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 4129–4138). IEEE. https://doi.org/10.1109/CVPR52729.2023.00402

Yan, Z., Yao, T., Chen, S., Zhao, Y., Fu, X., Zhu, J., Luo, D., Wang, C., Ding, S., Wu, Y., & Yuan, L. (2024). DF40: Toward next-generation deepfake detection. In *Advances in Neural Information Processing Systems* (Vol. 37, pp. 29387–29434). Curran Associates, Inc. https://doi.org/10.52202/079017-0925

Yan, Z., Zhang, Y., Yuan, X., Lyu, S., & Wu, B. (2023). DeepfakeBench: A comprehensive benchmark of deepfake detection. In *Advances in Neural Information Processing Systems* (Vol. 36, pp. 4534–4565). Curran Associates, Inc. https://doi.org/10.52202/075280-0201

Zhao, H., Zhou, W., Chen, D., Wei, T., Zhang, W., & Yu, N. (2021). Multi-attentional deepfake detection. In *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 2185–2194). IEEE. https://doi.org/10.1109/CVPR46437.2021.00222

DOI: https://doi.org/10.54082/jupin.2712
p-ISSN: 2808-148X
e-ISSN: 2808-1366

Halaman ini dikosongkan