E-ISSN: 3025-0994
Halaman 9389-9399
Volume 4 Nomor 3 Tahun 2026

# Sistem Otentikasi Biometrik Multimodal Berbasis Face Recognition Dan Voiceprint Pada Aplikasi Website

**M. C. Aprizal[1], M. Azwar[2], L. Widyawati[3]**

[1,2,3] Program Studi Teknologi Informasi, Fakultas Teknik, Universitas Bumigora
e-mail: carlieaprizal725@gmail.com

**Abstrak**

Keamanan sistem otentikasi berbasis kata sandi semakin rentan terhadap ancaman pencurian identitas, *phishing*, dan serangan modern lainnya. Penelitian ini merancang dan mengimplementasikan sistem otentikasi biometrik *multimodal* berbasis *website* yang mengintegrasikan pengenalan wajah, verifikasi *voiceprint*, dan verifikasi *passphrase* sebagai pendekatan autentikasi *passwordless*. Sistem dikembangkan dengan arsitektur *client-server* menggunakan *Flask* pada sisi *backend* serta *WebRTC* dan *MediaRecorder API* pada sisi *frontend*. Pengenalan wajah dilakukan menggunakan *DeepFace* dengan model *FaceNet*, sedangkan *voiceprint* diperoleh dari fusi berbobot tiga model *deep learning*, yaitu *ECAPA-TDNN*, *WavLM*, dan *Wav2Vec* 2.0. *Passphrase* yang diucapkan pengguna ditranskripsikan melalui *Whisper ASR* dan dicocokkan dengan teks referensi untuk memperkuat deteksi keaktifan (*liveness detection*). Keputusan akhir otentikasi menggunakan mekanisme *cascading* dengan ambang suara 0,70, wajah 0,75, *passphrase* 0,80, dan skor gabungan 0,85. Sistem diuji melalui skenario identifikasi 1:N, meliputi *black box testing*, pengujian *genuine*, pengujian *impostor*, pengujian performa, dan simulasi *replay attack*. Hasil pengujian menunjukkan *Rank-1 Accuracy* sebesar 72% dengan tingkat penolakan *impostor* 100% dari 250 percobaan penyusupan, serta rata-rata waktu pemrosesan 3,42 detik per autentikasi. Kegagalan pengguna sah pada pengujian *genuine* disebabkan oleh kesalahan pengucapan *passphrase*, bukan oleh kegagalan model biometrik. Penelitian ini membuktikan bahwa autentikasi biometrik *multimodal passwordless* dapat diimplementasikan secara praktis pada infrastruktur web standar tanpa perangkat keras tambahan, serta memberikan lapisan keamanan berlapis yang lebih tahan terhadap serangan dibandingkan autentikasi konvensional berbasis kata sandi.

**Kata Kunci:** *Otentikasi Biometrik, Multimodal, Voiceprint; Face Recognition,* Automatic Speech Recognition

**Abstract**

Password-based authentication has become increasingly vulnerable to identity theft, phishing, and modern cyberattacks. This study designed and implemented a website-based multimodal biometric authentication system that integrated face recognition, voiceprint verification, and passphrase verification as a passwordless authentication approach. The system was developed with a client-server architecture using Flask on the backend and WebRTC together with the MediaRecorder API on the frontend. Face recognition was performed using DeepFace with the FaceNet model, while the voiceprint was obtained from a weighted fusion of three deep learning models, namely ECAPA-TDNN, WavLM, and Wav2Vec 2.0. The passphrase spoken by the user was transcribed through Whisper ASR and matched against a reference text to strengthen liveness detection. The final authentication decision applied a cascading mechanism with thresholds of 0.70 for voice, 0.75 for face, 0.80 for passphrase, and 0.85 for the combined score. The system was tested through a 1:N identification scenario, covering black box testing, genuine trials, impostor trials, performance testing, and replay attack simulation. The results showed a Rank-1 Accuracy of 72% with a 100% impostor rejection rate across 250 intrusion attempts, and an average processing time of 3.42 seconds per

authentication. Failures among legitimate users in the genuine trials were caused by passphrase mispronunciation rather than biometric model failure. This research demonstrates that passwordless multimodal biometric authentication can be practically implemented on standard web infrastructure without additional hardware, providing a layered security mechanism more resistant to attacks than conventional password-based authentication.

**Keywords:** *Biometric Authentication, Multimodal, Voiceprint, Face Recognition,* Automatic Speech Recognition

## LATAR BELAKANG

Keamanan sistem informasi berbasis *website* menjadi isu yang semakin krusial seiring meningkatnya ancaman pencurian identitas, kebocoran data, dan serangan terhadap autentikasi berbasis kata sandi (Rachapalli et al., 2024). Metode autentikasi konvensional yang hanya mengandalkan *username* dan *password* memiliki keterbatasan mendasar, karena mayoritas insiden keamanan justru disebabkan oleh perilaku pengguna itu sendiri, seperti pemilihan kata sandi yang lemah, penggunaan ulang kata sandi pada berbagai layanan, serta kerentanan terhadap *phishing* dan *shoulder surfing* (Alharbi & Alshanbari, 2023; Rahman et al., 2024).Dengan kata lain, bukan sistem yang gagal mengenali pengguna sah, melainkan pengguna yang tidak mampu menjaga kerahasiaan kredensialnya, sehingga diperlukan pendekatan otentikasi yang mengalihkan beban keamanan dari pengguna ke sistem (Radytia, 2026; Utami, 2023).

Autentikasi berbasis biometrik dikembangkan sebagai solusi karena karakteristik biometrik lebih sulit dipalsukan dibandingkan kata sandi (Atenco-vázquez et al., 2024). Pada sisi pengenalan wajah, model *FaceNet* telah terbukti menghasilkan *face embedding* yang diskriminatif untuk mengukur kemiripan antarwajah (Mugalu et al., 2021).Pada sisi suara, *ECAPA-TDNN* mampu menghasilkan *speaker embedding* dengan akurasi tinggi dan tahan terhadap derau (Dawalatabad & Ravanelli, 2021; Weng et al., 2025), sedangkan pendekatan *self-supervised learning* melahirkan model *Wav2Vec* 2.0 dan *WavLM* yang mampu mempelajari representasi suara yang kaya tanpa memerlukan anotasi berskala besar (Chen et al., 2022; Fan et al., 2021).

Namun, sistem biometrik *unimodal* tetap rentan terhadap serangan presentasi (*presentation attack*) berupa foto atau rekaman suara curian. Pendekatan *multimodal* yang menggabungkan wajah dan suara terbukti meningkatkan ketahanan sistem dibandingkan pendekatan *unimodal* (Farhadipour et al., 2024; Rahman et al., 2024).Meski demikian, solusi *multimodal* yang ada saat ini umumnya belum terintegrasi dalam satu ekosistem *website* yang ringan dan *passwordless*, masih beroperasi pada mode verifikasi 1:1 tanpa pengujian pada skenario identifikasi 1:N, belum dilengkapi mekanisme verifikasi *passphrase*, dan sering memerlukan perangkat keras tambahan (Choi et al., 2024; Unnithan et al., 2026).

Berdasarkan celah tersebut, penelitian ini merancang dan mengimplementasikan sistem otentikasi biometrik *multimodal* berbasis *website* yang mengintegrasikan *FaceNet* untuk pengenalan wajah, ansambel *ECAPA-TDNN*, *WavLM*, dan *Wav2Vec* 2.0 untuk verifikasi *voiceprint*, serta *Whisper ASR* untuk verifikasi *passphrase* sebagai mekanisme *liveness detection* (Radford et al., 2021).Sistem bekerja pada mode identifikasi 1:N sehingga pengguna tidak perlu memasukkan klaim identitas apa pun; pengguna cukup berbicara dan menatap kamera, kemudian sistem mencari dan mengenali identitas secara otomatis. Ketiga modalitas dikombinasikan melalui mekanisme keputusan *cascading* untuk membentuk pertahanan berlapis yang sulit ditembus, sementara seluruh komponen berjalan pada infrastruktur web standar (*Flask*, *WebRTC*, *SQLite*) tanpa memerlukan perangkat keras khusus (Dobrowolski et al., 2023). Penelitian ini bertujuan mengimplementasikan ketiga model biometrik tersebut, mengintegrasikannya ke dalam satu arsitektur sistem, serta mengevaluasi kinerja fusi keputusan secara *real-time*, sehingga dapat menjadi alternatif yang layak untuk menggantikan autentikasi berbasis kata sandi pada aplikasi web modern.

## METODE

Penelitian ini menggunakan dua pendekatan metodologis, yaitu studi literatur dan pengembangan sistem dengan model *Secure Software Development Life Cycle* (SSDLC), yang mengintegrasikan praktik keamanan pada setiap tahap siklus pengembangan perangkat lunak. Tahapan yang diterapkan meliputi perencanaan, analisis kebutuhan, desain, pengembangan (implementasi), dan pengujian, sebagaimana ditunjukkan pada Gambar 1.




**Gambar 1. Alur Metode SSDLC yang Digunakan**

Pada tahap analisis kebutuhan, ditetapkan enam kebutuhan fungsional utama: (1) pendaftaran biometrik (*enrollment*) berupa perekaman tiga sampel suara dan satu gambar wajah yang diekstraksi menjadi *voiceprint* dan *faceprint* referensi; (2) autentikasi wajah menggunakan *face embedding* 128 dimensi dari *DeepFace-FaceNet*; (3) autentikasi *voiceprint* melalui fusi berbobot *ECAPA-TDNN*, *WavLM*, dan *Wav2Vec* 2.0 dengan bobot 1,4 : 0,8 : 0,8 yang menghasilkan vektor 1.728 dimensi dan dinormalisasi L2; (4) verifikasi *passphrase* sebagai *liveness detection* menggunakan *Whisper ASR* dan pencocokan teks dengan *SequenceMatcher*; (5) kalkulasi skor dan keputusan *multimodal*; serta (6) manajemen log akses untuk kebutuhan audit keamanan. Kebutuhan nonfungsional mencakup keamanan penyimpanan data (hanya vektor *embedding* yang disimpan, bukan berkas audio/gambar mentah), target latensi pemrosesan di bawah 4 detik, serta kompatibilitas dengan peramban modern yang mendukung *WebRTC* dan *MediaRecorder API*.

Sebagai mekanisme *liveness detection*, sistem menerapkan verifikasi *passphrase* berbasis *Whisper ASR* dengan alur sebagaimana ditunjukkan pada Gambar 2. Sistem membangkitkan *passphrase* berupa deretan angka acak yang hanya diperdengarkan melalui *speaker* perangkat tanpa ditampilkan pada antarmuka, sehingga pengguna harus mengucapkannya secara langsung. Ucapan pengguna direkam melalui mikrofon, kemudian ditranskripsikan oleh *Whisper ASR* (model *small*) menjadi teks mentah. Teks tersebut dinormalisasi dengan mengubah kata angka (Inggris/Indonesia) menjadi digit dan menyeragamkan spasi, sebelum dibandingkan dengan *passphrase* referensi menggunakan *SequenceMatcher*. Apabila skor kemiripan mencapai ambang minimal 0,80, verifikasi *passphrase* dinyatakan lolos dan dilanjutkan ke pemeriksaan skor suara dan wajah; sebaliknya, sistem langsung menolak akses dengan status *PASSPHRASE_FAIL*.

Pada tahap pengembangan, seluruh model *deep learning* (*ECAPA-TDNN*, *WavLM*, *Wav2Vec* 2.0, *Silero VAD*, *DeepFace-FaceNet*, dan *Whisper*) diinisialisasi satu kali saat *server Flask* dijalankan agar dapat digunakan berulang tanpa *overhead* pemuatan ulang. Audio mentah berformat *WebM* dikonversi menjadi *WAV* 16 kHz *mono*, direduksi deraunya dengan *NoiseReduce*, kemudian dipotong segmen sunyinya menggunakan *Silero VAD* sebelum diekstraksi fiturnya. Keputusan akhir otentikasi diambil melalui mekanisme *cascading AND* yang mengevaluasi empat syarat secara berurutan: skor suara ≥ 0,70, skor wajah ≥ 0,75, skor *passphrase* ≥ 0,80, dan skor gabungan (Persamaan 1) ≥ 0,85. Apabila salah satu syarat tidak terpenuhi, akses ditolak tanpa melanjutkan pemeriksaan berikutnya.

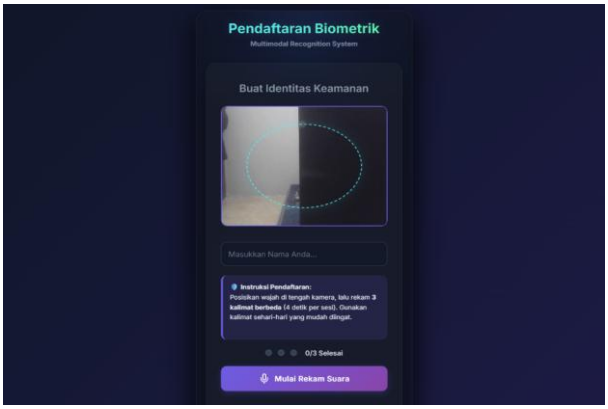
$$SkorGabungan = (0{,}60 \times SkorSuara) + (0{,}40 \times SkorWajah) \quad (1)$$

Pengujian sistem dilakukan pada laptop dengan prosesor Intel Core i5-10300H, RAM 16 GB, dan GPU NVIDIA GeForce GTX 1650 4 GB, menggunakan lima skema pengujian: (1)

*black box testing* untuk memverifikasi kesesuaian fungsional sistem; (2) pengujian *genuine* untuk mengukur *Rank-1 Accuracy* pada mode identifikasi 1:N; (3) pengujian *impostor* untuk mengukur tingkat penolakan penyusup; (4) pengujian performa untuk mengukur waktu pemrosesan *end-to-end*; dan (5) simulasi *replay attack* pada enam skenario serangan untuk menguji ketahanan keamanan sistem. Sebagai pembanding, dilakukan pula pengujian terhadap otentikasi konvensional berbasis *username* dan *password*.

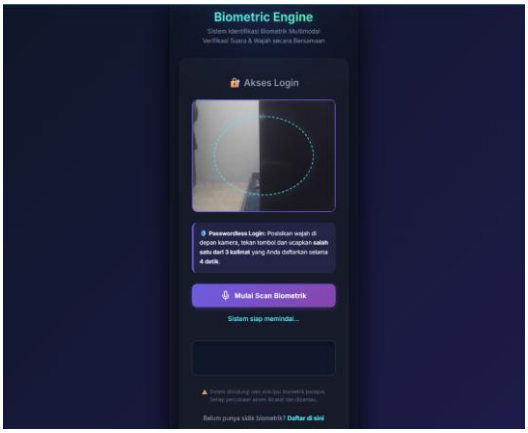
## HASIL DAN PEMBAHASAN

Implementasi Antarmuka Pengguna. Sistem otentikasi biometrik *multimodal* berhasil diimplementasikan dalam bentuk aplikasi *website* yang menggabungkan autentikasi wajah, suara, dan *passphrase*. Halaman Registrasi (Gambar 2) menggunakan tema *dark mode* dan memandu pengguna melalui pengisian *username*, perekaman tiga sampel suara, serta pengambilan satu gambar wajah. Data yang ditangkap diproses menggunakan fusi model *ECAPA-TDNN*, *WavLM*, *Wav2Vec* 2.0, dan *FaceNet* untuk menghasilkan *voiceprint* dan *faceprint* permanen yang disimpan pada basis data.




**Gambar 2. Tampilan Halaman Registrasi**

Halaman *Login* (Gambar 3) mendukung mekanisme identifikasi 1:N secara *passwordless*: pengguna cukup menekan tombol mulai rekam suara tanpa perlu memasukkan identitas teks apa pun. Pada sisi *backend*, sampel suara dan wajah yang ditangkap langsung diekstraksi dan dibandingkan dengan seluruh basis data *voiceprint* dan *faceprint* menggunakan *cosine similarity* untuk menentukan otoritas akses.



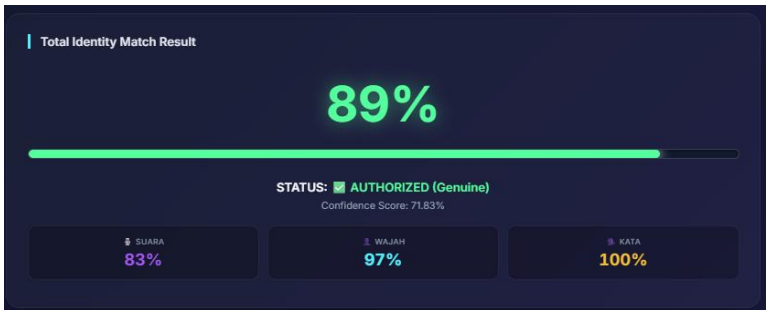

**Gambar 3. Tampilan Halaman Login**

Halaman *Dashboard* (Gambar 4) berfungsi sebagai pusat pemantauan keamanan (*security monitoring*) yang menampilkan hasil verifikasi identifikasi 1:N secara *real-time*, mencakup persentase tingkat kemiripan (*match rate*) dan status keputusan akses (*granted/denied*), dilengkapi tabel *Recent Access Logs* untuk kebutuhan audit.



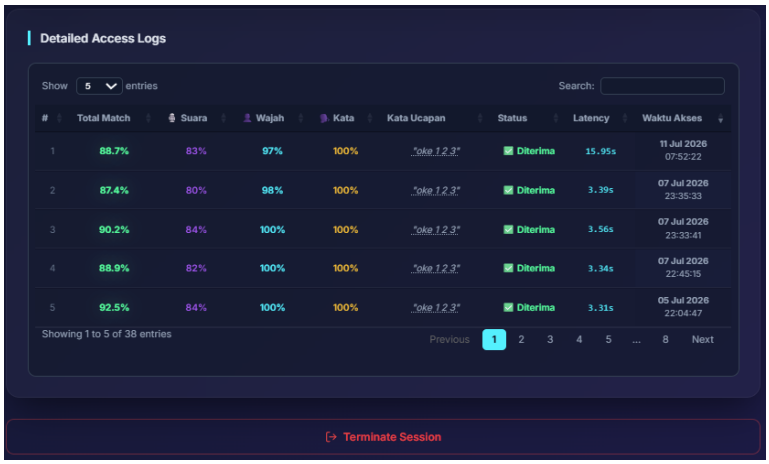

**Gambar 4. Tampilan Halaman Dashboard**

Komponen *Total Identity Match Result* (Gambar 5) menyajikan visualisasi hasil akhir fusi biometrik *multimodal*. Pada kondisi akses ditolak, antarmuka menampilkan peringatan berwarna merah dengan status "ACCESS DENIED (Impostor)" dan *confidence score* 0%, sedangkan panel suara, wajah, dan *passphrase* otomatis dikosongkan karena sistem langsung memblokir pemrosesan fitur penyusup. Sebaliknya, pada akses yang diterima, komponen ini berubah warna hijau dan menampilkan rincian nilai persentase kemiripan pada masing-masing modalitas sebagai bukti transparansi hasil pencocokan.




**Gambar 5. Tampilan Total Identity Match Result**

Antarmuka *Detailed Access Logs* (Gambar 6) dirancang sebagai fungsi audit dan pengawasan *real-time* terhadap setiap aktivitas otentikasi. Tabel ini merekam skor fusi akhir, skor kemiripan individual (suara dan wajah), verifikasi kesesuaian *passphrase*, status akses, latensi waktu komputasi, dan *timestamp*, dilengkapi fitur pencarian serta tombol *terminate session* untuk memutus sesi yang mencurigakan.




**Gambar 6. Tampilan Detailed Access Logs**

Pengujian Fungsional (Black Box Testing). Pengujian *black box* dilakukan untuk memverifikasi kesesuaian fungsional sistem dari sudut pandang pengguna akhir tanpa memperhatikan struktur kode internal. Sepuluh skenario pengujian mencakup proses registrasi hingga keamanan *endpoint* API. Seluruh skenario dinyatakan berhasil sesuai dengan hasil yang diharapkan, sebagaimana dirangkum pada Tabel 1.

**Tabel 1. Hasil Pengujian Black Box**

| No | Skenario Pengujian | Hasil Pengujian | Status |
|---|---|---|---|
| 1 | Registrasi (username + 3 sampel suara + wajah) | Data tersimpan, diarahkan ke halaman Login | Berhasil |
| 2 | Login Genuine | Identitas dikenali, akses ke Dashboard | Berhasil |
| 3 | Login Impostor | Ditolak, pesan “Identitas tidak dikenali” | Berhasil |
| 4 | Login dengan Passphrase Salah | Ditolak, alasan PASSPHRASE_FAIL tercatat | Berhasil |
| 5 | Akses Dashboard | Skor dan riwayat akses ditampilkan | Berhasil |
| 6 | Logout | Session terhapus, kembali ke Login | Berhasil |
| 7 | Evaluasi Genuine | Log tersimpan, is_correct = true | Berhasil |
| 8 | Evaluasi Impostor | Log tersimpan, is_correct = false | Berhasil |
| 9 | Riwayat Akses | Tabel tanggal, skor, dan status tampil | Berhasil |
| 10 | Keamanan API tanpa sesi | Mengembalikan array kosong | Berhasil |

Pengujian Genuine (Rank-1 Accuracy). Pengujian ini mengukur kemampuan sistem mengenali pengguna sah pada peringkat pertama (*Rank-1*) dalam mode identifikasi 1:N. Sebanyak 100 percobaan *login genuine* dilakukan menggunakan suara dan wajah pengguna terdaftar; Tabel 2 menampilkan sepuluh sampel representatif dari hasil pengujian tersebut.

**Tabel 2. Sampel Hasil Pengujian Genuine**

| No | Voice | Face | Passphrase | Combined | Keterangan |
|---|---|---|---|---|---|
| 1 | 0.86 | 0.99 | 1.00 | 0.93 | Dikenali |
| 2 | 0.88 | 0.99 | 1.00 | 0.94 | Dikenali |
| 3 | 0.83 | 0.93 | 0.94 | 0.88 | Dikenali |
| 4 | 0.82 | 0.96 | 0.94 | 0.90 | Dikenali |
| 5 | 0.80 | 0.95 | 0.87 | 0.88 | Dikenali |
| 6 | 0.81 | 0.91 | 0.86 | 0.86 | Dikenali |
| 7 | 0.78 | 0.95 | 1.00 | 0.87 | Dikenali |
| 8 | 0.75 | 0.94 | 1.00 | 0.85 | Dikenali |
| 9 | 0.82 | 0.95 | 1.00 | 0.87 | Dikenali |
| 10 | 0.85 | 0.96 | 1.00 | 0.89 | Dikenali |

Dari keseluruhan 100 percobaan, sistem berhasil menempatkan identitas pengguna sah pada peringkat pertama sebanyak 72 kali sehingga diperoleh *Rank-1 Accuracy* sebesar 72%. Kegagalan pada percobaan sisanya seluruhnya disebabkan oleh ketidakcocokan *passphrase* (PASSPHRASE_FAIL) meskipun skor suara dan wajah telah melampaui ambang batas, bukan oleh kegagalan model biometrik wajah maupun suara. Skor kemiripan suara pengguna sah berkisar 0,75–0,88, sedangkan skor wajah konsisten tinggi pada rentang 0,91–0,99, menunjukkan bahwa *FaceNet* dan fusi model suara mampu menghasilkan representasi yang diskriminatif.

Pengujian Impostor (Impostor Rejection). Pengujian ini bertujuan memastikan sistem tidak pernah mengidentifikasi individu tidak terdaftar sebagai pengguna sah. Tabel 3 menyajikan sepuluh sampel percobaan *impostor* menggunakan suara dan wajah penyerang sendiri.

**Tabel 3. Sampel Hasil Pengujian Impostor**

| No | Voice | Face | Passphrase | Combined | Dikenali Sebagai |
|---|---|---|---|---|---|
| 1 | 0.49 | 0.56 | 0.62 | 0.52 | Unknown |
| 2 | 0.49 | 0.56 | 0.66 | 0.52 | Unknown |
| 3 | 0.49 | 0.58 | 0.11 | 0.51 | Unknown |
| 4 | 0.50 | 0.63 | 0.51 | 0.55 | Unknown |
| 5 | 0.56 | 0.59 | 0.72 | 0.51 | Unknown |
| 6 | 0.59 | 0.58 | 0.85 | 0.59 | Unknown |
| 7 | 0.48 | 0.58 | 0.15 | 0.52 | Unknown |
| 8 | 0.54 | 0.54 | 0.51 | 0.54 | Unknown |
| 9 | 0.44 | 0.56 | 0.57 | 0.48 | Unknown |
| 10 | 0.53 | 0.54 | 0.52 | 0.53 | Unknown |

Berdasarkan Tabel 3, seluruh percobaan *impostor* berhasil ditolak sistem (tingkat penolakan 100%) dari total 150 percobaan yang dilakukan pada mode identifikasi 1:N. Skor kemiripan suara dan wajah penyusup secara konsisten berada di bawah ambang batas (suara < 0,70; wajah < 0,75), meskipun beberapa percobaan sempat mengucapkan *passphrase* dengan benar, sehingga skor gabungan tidak pernah mencapai ambang keputusan 0,85.

Pengujian Performa (Waktu Pemrosesan). Pengujian performa mengukur efisiensi waktu sistem dalam menjalankan *pipeline* otentikasi biometrik *multimodal* secara *end-to-end*, mencakup konversi format audio, reduksi derau, deteksi aktivitas suara, ekstraksi *embedding* tiga model suara, transkripsi *passphrase*, ekstraksi *embedding* wajah, hingga pencarian 1:N terhadap seluruh *template* pada basis data. Sebanyak 30 sampel data waktu pemrosesan dikumpulkan dari log otentikasi, dengan hasil dirangkum pada Tabel 4 dan divisualisasikan pada Gambar 7.

**Tabel 4. Analisis Rata-rata Waktu Proses**

| Metrik | Nilai |
|---|---|
| Jumlah data | 3,00 detik |
| Rata-rata | 3,42 detik |
| Minimum | 3,23 detik |
| Maksimum | 3,62 detik |
| Standar deviasi | 0,12 detik |




**Gambar 7. Grafik Garis Waktu Proses**

Rata-rata waktu pemrosesan sistem adalah 3,42 detik, dengan rentang antara 3,23 detik (tercepat) hingga 3,62 detik (terlama). Standar deviasi sebesar 0,12 detik menunjukkan variasi yang kecil, sehingga *pipeline multimodal* berjalan konsisten dan seluruh proses selesai dalam waktu kurang dari 4 detik sesuai batas durasi rekaman yang diterapkan.

Pengujian Keamanan (Replay Attack). Pengujian ini membuktikan ketahanan sistem terhadap serangan pemutaran ulang (*replay attack*), yaitu ketika penyerang menggunakan rekaman suara, foto, atau video pengguna sah untuk mencoba masuk. Pengujian dilaksanakan dalam enam skenario serangan yang masing-masing diulang beberapa kali, dengan hasil dirangkum pada Tabel 5.

**Tabel 5. Hasil Pengujian Replay Attack**

| No | Metode Serangan | Jumlah Percobaan | Gagal | Berhasil |
|---|---|---|---|---|
| 1 | Rekaman suara + foto asli | 15 | 15 | 0 |
| 2 | Rekaman suara + video asli | 15 | 15 | 0 |
| 3 | Suara impostor + foto asli | 15 | 15 | 0 |
| 4 | Suara impostor + video asli | 15 | 15 | 0 |
| 5 | Rekaman suara asli + wajah impostor | 20 | 20 | 0 |

| No | Metode Serangan | Jumlah Percobaan | Gagal | Berhasil |
|---|---|---|---|---|
| 6 | Impostor meniru passphrase + wajah impostor | 20 | 20 | 0 |

Dari total 100 percobaan serangan pada enam skenario, seluruhnya berhasil ditolak oleh sistem (tingkat penolakan 100%). Pada skenario rekaman suara dan foto/video asli, suara dan wajah pengguna sah dikenali dengan skor tinggi, tetapi *passphrase* dari rekaman lama tidak cocok dengan tantangan baru yang diperdengarkan sistem sehingga *Whisper ASR* gagal mentranskripsikan ucapan secara sesuai. Pada skenario penyerang mengucapkan *passphrase* baru menggunakan foto/video pengguna sah, wajah menghasilkan skor tinggi namun suara penyerang tidak dikenali sehingga skor gabungan tidak mencapai ambang 0,85. Skenario paling berbahaya, yaitu penyerang menirukan *passphrase* dengan benar sekaligus mencoba wajah palsu, tetap gagal karena suara dan wajah penyerang tidak cocok dengan *template* mana pun pada basis data. Hasil ini membuktikan bahwa arsitektur *cascading* dengan tiga lapis biometrik saling melengkapi: apabila satu lapisan berhasil dikompromikan, lapisan lain tetap menutup celah keamanan tanpa memerlukan perangkat keras tambahan untuk *liveness detection*.

Perbandingan Akurasi Mode Verifikasi (1:1) dan Identifikasi (1:N).Pengujian ini membandingkan akurasi sistem pada dua mode operasi menggunakan data pengguna dan ambang keputusan yang seragam. Hasil perbandingan disajikan pada Tabel 6.

**Tabel 6. Perbandingan Akurasi Mode 1:1 dan 1:N**

| Metrik | Mode 1:1 (Verifikasi) | Mode 1:N (Identifikasi) |
|---|---|---|
| Genuine diterima | 100% | 72% |
| Impostor ditolak | 100% | 100% |

Mode verifikasi (1:1) menunjukkan tingkat penerimaan pengguna sah yang lebih tinggi karena sistem hanya membandingkan satu *template* acuan, sedangkan mode identifikasi (1:N) mengharuskan sistem mencari kecocokan di antara seluruh *template* pada basis data sehingga skor relatif dapat menurun. Meskipun demikian, dari segi keamanan kedua mode menunjukkan hasil sempurna, yaitu seluruh percobaan *impostor* berhasil ditolak tanpa kecuali pada kedua mode.

Perbandingan dengan Otentikasi Konvensional. Sebagai pembanding, dilakukan pengujian otentikasi tradisional berbasis *username* dan *password* pada lima percobaan *genuine* dan lima percobaan *impostor*. Tabel 7 merangkum hasil pengujian tersebut.

**Tabel 7. Perbandingan dengan Otentikasi Password Konvensional**

| Metrik | Otentikasi Password | Otentikasi Biometrik Multimodal |
|---|---|---|
| Akurasi | 100% | 72% (Rank-1, mode 1:N) |
| Rata-rata waktu proses | 0,08 detik | 3,42 detik |
| Ketahanan terhadap pencurian kredensial/phishing | Rendah | Tinggi |

Otentikasi konvensional mencapai akurasi 100% dengan waktu komputasi sangat cepat (0,08 detik), namun memiliki kerentanan inheren terhadap serangan pencurian kredensial seperti *brute-force* dan *phishing*. Waktu komputasi biometrik yang lebih lama merupakan *trade-off* yang sepadan karena menawarkan proteksi tiga lapis (wajah, suara, dan *passphrase*), menjamin bukti kehadiran fisik (*non-repudiation*), serta menghilangkan beban kognitif pengguna dalam menghafal kata sandi.

Secara keseluruhan, hasil implementasi dan pengujian menunjukkan bahwa sistem otentikasi biometrik *multimodal* berbasis *website* dapat diimplementasikan secara praktis pada

infrastruktur web standar (*Flask*, *WebRTC*, *SQLite*) tanpa memerlukan perangkat keras tambahan. Integrasi wajah, suara, dan verifikasi *passphrase* melalui mekanisme *cascading decision* terbukti memberikan pertahanan berlapis yang mampu menolak 100% percobaan penyusupan dan serangan *replay*, sekalipun tingkat penerimaan pengguna sah pada mode identifikasi 1:N (72%) masih dapat ditingkatkan melalui optimasi akurasi transkripsi *passphrase*. Temuan ini mendukung penelitian terdahulu bahwa pendekatan *multimodal* lebih tahan terhadap serangan presentasi dibandingkan sistem *unimodal* (Farhadipour et al., 2024; Rahman et al., 2024), sekaligus memperkuat argumen bahwa pendekatan *passwordless* dapat mengalihkan beban keamanan dari pengguna ke sistem (Dobrowolski et al., 2023).

## SIMPULAN

Penelitian ini berhasil merancang dan mengimplementasikan sistem otentikasi biometrik multimodal berbasis website yang mengintegrasikan FaceNet, ansambel Wav2Vec 2.0-WavLM-ECAPA-TDNN, dan Whisper ASR dalam satu arsitektur cascading decision menggunakan Flask, WebRTC, dan SQLite. Sistem mencapai Rank-1 Accuracy 72% pada mode identifikasi 1:N dengan tingkat penolakan impostor 100% pada seluruh pengujian genuine, impostor, dan replay attack. Kegagalan penerimaan pengguna sah semata-mata disebabkan kesalahan pengucapan passphrase, bukan kegagalan model biometrik. Disarankan pengembangan lanjutan pada verifikasi passphrase yang lebih toleran, penambahan anti-spoofing visual, serta pengujian pada populasi dan lingkungan akustik yang lebih beragam.

## DAFTAR PUSTAKA

Alharbi, B., & Alshanbari, H. S. (2023). *Face-Voice Based Multimodal Biometric Authentication System Via Facenet And Gmm*. Https://Doi.Org/10.7717/Peerj-Cs.1468

Atenco-Vázquez, J. C., Moreno-Rodríguez, J. C., & Ramírez-Cortés, J. M. (2024). *Bimodal Biometric Recognition System Using Convolutional Neural Networks And Fusion Of Deep Audiovisual Feature Vectors Biometric Verification And Identification Multimodal Biometric Systems*. *15*(3), 4–14.

Chen, S., Wang, C., Chen, Z., Wu, Y., Liu, S., & Chen, Z. (2022). *Wavlm : Large-Scale Self-Supervised Pre-Training For Full Stack Speech Processing*. 1–14.

Choi, H., Kim, J., Woo, S. S., & Kim, H. (2024). Blind-Match : Efficient Homomorphic Encryption-Based 1 : N Matching For Privacy-Preserving Biometric Identification. In *Proceedings Of The 33rd Acm International Conference On Information And Knowledge Management (Cikm '24), October 21â•fi25, 2024, Boise, Id, Usa* (Vol. 1, Issue 1). Association For Computing Machinery. Https://Doi.Org/10.1145/3627673.3680017

Dawalatabad, N., & Ravanelli, M. (2021). *Ecapa-Tdnn Embeddings For Speaker Diarization*.

Dobrowolski, A. P., Piotrowski, Z., & Adam, K. (2023). *Enhancing Web Application Security : Advanced Biometric Voice Verification For Two-Factor Authentication*. 1–19.

Fan, Z., Li, M., Zhou, S., & Xu, B. (2021). *Exploring Wav2vec 2.0 On Speaker Verification And Language Identification*. *Lid*, 1–5.

Farhadipour, A., Chapariniya, M., Vukovic, T., & Dellwo, V. (2024). *Comparative Analysis Of Modality Fusion Approaches For Audio-Visual Person Identification And Verification*.

Mugalu, B. W., Wamala, R. C., Serugunda, J., & Katumba, A. (2021). *Face Recognition As A Method Of Authentication In A Web-Based System*.

Rachapalli, D. R., Dondeti, V., & Kalluri, H. K. (2024). Multimodal Cancellable Biometric Template Protection And Person Verification In Transformed Domain. *Ieee Access*, *12*(November), 173557–173582. Https://Doi.Org/10.1109/Access.2024.3501368

Radford, A., Wook, J., Tao, K., Greg, X., Christine, B., & Ilya, M. (2021). *Robust Speech Recognition Via Large-Scale Weak Supervision*.

Radytia, A. A. (2026). *Keamanan Informasi Di Era Digital: Tanggung Jawab Kita Bersama*. Https://Www.Djkn.Kemenkeu.Go.Id/Kpknl-Biak/Baca-Artikel/18443/Keamanan-Informasi-Di-Era-Digital-Tanggung-Jawab-Kita-Bersama.Html

Rahman, S. A. El, Saleh, A., & Id, A. (2024). *Enhanced Multimodal Biometric Recognition*

*Systems Based On Deep Learning And Traditional Methods In Smart Environments*. 1–24. Https://Doi.Org/10.1371/Journal.Pone.0291084

Unnithan, M. R., Sabir, M., Sherief, S., Afiya, S., & Aziz, A. A. (2026). *Fusion Strategies And Open-Set Recognition In Multimodal Biometric Authentication Systems : A Comprehensive Review*. *04*(April), 2350–2361.

Utami, S. N. (2023). *8 Macam Ancaman Pada Keamanan Informasi Digital*. Https://Www.Kompas.Com/Skola/Read/2023/01/28/150000069/8-Macam-Ancaman-Pada-Keamanan-Informasi-Digital

Weng, S., Yang, L., & Mao, J. (2025). *Effective Modeling Of Critical Contextual Information For Tdnn-Based Speaker Verification*. 1–5.

Baevski, A., Zhou, Y., Mohamed, A., & Auli, M. (2020). Wav2vec 2.0: A Framework For Self-Supervised Learning Of Speech Representations. Advances In Neural Information Processing Systems, 33, 12449–12460.

Desplanques, B., Thienpondt, J., & Demuynck, K. (2020). Ecapa-Tdnn: Emphasized Channel Attention, Propagation And Aggregation In Tdnn Based Speaker Verification. Interspeech 2020, 3830–3834. Https://Doi.Org/10.21437/Interspeech.2020-2650

Schroff, F., Kalenichenko, D., & Philbin, J. (2015). Facenet: A Unified Embedding For Face Recognition And Clustering. Proceedings Of The Ieee Conference On Computer Vision And Pattern Recognition (Cvpr), 815–823. Https://Doi.Org/10.1109/Cvpr.2015.7298682

Snyder, D., Garcia-Romero, D., Povey, D., & Khudanpur, S. (2017). Deep Neural Network Embeddings For Text-Independent Speaker Verification. Interspeech 2017, 999–1003. Https://Doi.Org/10.21437/Interspeech.2017-620