SITEKNIK
Sistem Informasi, Teknik dan Teknologi Terapan
E-ISSN: 3032-3991 P-ISSN: 3090-1626
Vol. 2. No. 3 July 2025. Pages. 260-271

Implementasi Teknologi Mediapipe Menggunakan Metode CNN Berbasis Website Untuk Pengamanan VVIP Dalam Mobil

M. Ilham AlFatrah[1], Hery Sudaryanto[2], H. A. Danang Rimbawa[3]

milhamalfatrah@gmail.com[1], sudaryanto_h.casanova@yahoo.com[2], hadr71@gmail.com[3]
[123] Program Studi Teknik Elektro Fakultas Teknik dan Teknologi Pertahanan
Universitas Pertahanan Republik Indonesia

Kata kunci: Keamanan VVIP, MediaPipe, CNN, Deteksi Gestur, Pengawasan Real-Time

Dikirimkan: 25/07/2025
Direvisi: 28/07/2025
Diterima: 31/07/2025

Abstrak

Penelitian ini bertujuan mengembangkan sistem deteksi gestur tangan berbasis MediaPipe dan Convolutional Neural Network (CNN) guna meningkatkan efektivitas pengamanan VVIP. Mengingat ancaman modern yang semakin kompleks, sistem ini dirancang untuk mendeteksi gestur darurat secara real-time dan memungkinkan respons cepat. Metodologi yang digunakan meliputi pengumpulan dataset gestur tangan, anotasi data menggunakan MediaPipe, dan pelatihan model CNN di Google Colab. Kinerja model dievaluasi dengan metrik akurasi, presisi, recall, dan F1-score. Pengujian juga dilakukan dalam berbagai kondisi, seperti pencahayaan rendah dan gerakan cepat, untuk menilai ketangguhan sistem di dunia nyata. Hasilnya, sistem ini berhasil mendeteksi gestur tangan darurat dengan akurasi tinggi dan kecepatan kurang dari satu detik. Kinerja optimal, dengan akurasi mendekati 100%, tercapai pada kondisi pencahayaan yang baik. Meskipun akurasi sedikit menurun pada kondisi ekstrem, integrasi sistem pada platform website memungkinkan pengawasan dan pengambilan keputusan cepat di pusat komando. Penelitian ini membuktikan bahwa kombinasi MediaPipe dan CNN adalah solusi inovatif, namun optimasi lebih lanjut tetap dibutuhkan.

Penulis Korespondensi:
M. Ilham AlFatrah
Program Studi Teknik Elektro Fakultas Teknik dan Teknologi Pertahanan Universitas Pertahanan Republik Indonesia
Kawasan IPSC Sentul, Sukahati, Kec. Citeureup, Kabupaten Bogor, Jawa Barat 16810.
Email: milhamalfatrah@gmail.com

PENDAHULUAN

Penerapan kecerdasan buatan (AI) pada sistem pengawasan telah menjadi inovasi krusial dalam meningkatkan keamanan VVIP, seperti yang diuraikan oleh Arinuryadi (2018) dalam analisisnya mengenai keterbatasan pengamanan tradisional. Sistem

This is an open access article under the CC BY 4.0 license

pengamanan VVIP konvensional—yang mengandalkan pengawalan fisik, pemeriksaan makanan, dan pengamanan rute—sering kali tidak mampu merespons ancaman modern yang semakin kompleks, termasuk sabotase dan serangan siber. Kelemahan ini menyoroti pentingnya adopsi teknologi mutakhir. Sesuai temuan Kim & Lee (2021) serta Guo et al. (2022), sistem pengawasan cerdas berbasis AI mampu menganalisis data visual secara real-time, mengidentifikasi perilaku mencurigakan, dan mengeluarkan peringatan dini secara lebih cepat dan akurat dibandingkan dengan metode manual. Dengan demikian, integrasi AI menawarkan solusi efektif untuk memperkuat sistem keamanan VVIP agar lebih adaptif dan responsif terhadap berbagai ancaman.

Salah satu inovasi teknologi AI terkini yang sangat relevan untuk pengamanan VVIP adalah MediaPipe. Lugaresi et al. (2019) menjelaskan bahwa MediaPipe adalah kerangka kerja *machine learning* yang unggul dalam mendeteksi pose dan gerakan tubuh secara akurat dan *real-time*. Kemampuannya untuk mendeteksi titik-titik penting pada tubuh, termasuk tangan, memungkinkan pengenalan gestur yang cepat dan tepat, sebagaimana ditekankan oleh Zhang et al. (2020). Teknologi ini telah membuktikan keefektifannya dalam berbagai aplikasi, seperti sistem pengenalan bahasa isyarat yang mencapai akurasi lebih dari 94% (Wang et al., 2023). Studi lain oleh Wagh et al. (2023) juga menunjukkan kemampuan MediaPipe dalam melacak gerakan tangan 2D dengan tingkat kesalahan yang sangat rendah. Keunggulan MediaPipe lainnya adalah kemampuannya beroperasi pada perangkat *edge* (*on-device*), sehingga tidak memerlukan perangkat keras berkinerja tinggi. Hal ini menjadikannya solusi ideal untuk kebutuhan pengamanan VVIP yang menuntut sistem deteksi cepat, adaptif, dan andal dalam situasi darurat.

Penelitian ini berfokus pada implementasi teknologi MediaPipe dalam sistem keamanan VVIP berbasis *website* untuk mendeteksi gestur tangan kode darurat. Pertanyaan utamanya adalah bagaimana sistem ini dapat diterapkan secara efektif, serta sejauh mana akurasi dan kecepatan deteksi gestur dapat tercapai dalam berbagai skenario. Selain itu, penelitian ini juga akan mengevaluasi efektivitas integrasi sistem dengan *platform web* dan perangkat keras seperti Raspberry Pi 4 untuk mendukung pemantauan *real-time* pengawalan VVIP. Tujuan utama dari penelitian ini adalah mengembangkan prototipe sistem pengamanan VVIP inovatif yang (1) mengimplementasikan MediaPipe dan CNN untuk mendeteksi gestur tangan darurat secara otomatis pada *platform web*, (2) menganalisis tingkat akurasi dan kecepatan deteksi gestur dalam berbagai skenario, dan (3) mengevaluasi efektivitas integrasi sistem deteksi ini dengan *dashboard web* dan perangkat lapangan berbasis Raspberry Pi. Dengan demikian, diharapkan sistem yang dikembangkan dapat menjadi landasan bagi solusi pengamanan VVIP modern yang lebih adaptif, cepat, dan andal.

Pengamanan VVIP mencakup upaya perlindungan terhadap individu penting negara dari berbagai ancaman, baik fisik maupun non-fisik. Meskipun pendekatan tradisional berfokus pada keamanan fisik seperti pengawalan ketat dan kendaraan lapis baja, meningkatnya kompleksitas ancaman kini menuntut sistem pengawasan cerdas yang lebih proaktif. Konsep pengawasan cerdas untuk VVIP memanfaatkan teknologi sensor dan AI untuk mendeteksi potensi bahaya secara otomatis dan berkelanjutan, seperti yang dijelaskan oleh Khodadin & Pudaruth (2020). Dalam sistem ini, kamera dan sensor ditempatkan secara tersembunyi untuk memantau lingkungan sekitar VVIP. Data yang terkumpul kemudian dianalisis oleh algoritma AI untuk membedakan aktivitas normal dari aktivitas mencurigakan. Ketika gestur atau kejadian yang mengindikasikan ancaman terdeteksi, sistem akan segera memberikan peringatan tanpa menunggu respons manual. Konsep ini sejalan dengan tren keamanan modern, di mana kecepatan respons dan deteksi dini merupakan kunci utama dalam melindungi VVIP (Li et al., 2021).

Kecerdasan Buatan (AI) dan Computer Vision dalam Keamanan.

Kecerdasan buatan (AI) adalah cabang ilmu komputer yang fokus pada pengembangan mesin yang dapat meniru kecerdasan manusia, termasuk kemampuan

belajar dan mengambil keputusan (Chen et al., 2022). Dalam konteks pengawasan keamanan, AI berperan penting dalam mengotomatisasi proses deteksi dan pemantauan, sehingga sistem dapat bereaksi lebih cepat dibandingkan pengamatan manusia. Salah satu cabang AI yang paling relevan adalah computer vision, yaitu teknologi yang memungkinkan komputer untuk memahami dan menafsirkan data visual, seperti gambar atau video (Guo et al., 2022). Teknik computer vision mencakup deteksi objek, pengenalan pola, dan pelacakan gerak secara otomatis dalam citra. Untuk pengamanan VVIP, teknologi ini dapat digunakan untuk mendeteksi gerakan atau gestur mencurigakan secara *real-time* melalui kamera pengawas. Sebagai contoh, sistem pengawasan berbasis computer vision mampu mengenali gestur tangan tertentu—seperti tanda bahaya—di tengah kerumunan atau di dalam kendaraan, lalu segera mengirimkan sinyal alarm (Kim & Lee, 2021). Kombinasi AI dan computer vision ini telah terbukti meningkatkan keandalan dan ketepatan sistem keamanan, karena mampu menyaring kejadian penting dari aliran data video yang masif, sehingga mengurangi beban pemantauan manual (Guo et al., 2022).

MediaPipe untuk Deteksi Gestur Tangan

MediaPipe, sebuah *framework* sumber terbuka dari Google, dirancang untuk membangun alur pemrosesan sensoris dan visual secara efisien (Lugaresi et al., 2019). Salah satu modul utamanya, MediaPipe Hands, mampu mendeteksi 21 titik kunci (*landmark*) pada tangan manusia dari kamera secara *real-time* (Zhang et al., 2020). Keunggulan MediaPipe terletak pada sifatnya yang lintas *platform* dan kemampuannya untuk berjalan pada perangkat dengan sumber daya terbatas, menjadikannya ideal untuk aplikasi keamanan yang membutuhkan respons cepat tanpa ketergantungan pada koneksi internet. Banyak penelitian telah menggunakan MediaPipe sebagai dasar sistem pengenalan gestur karena akurasi tinggi dan latensinya yang rendah. Sebagai contoh, Indriani et al. (2021) berhasil menerapkan MediaPipe untuk pengenalan gerakan tangan pada aplikasi panduan dengan akurasi yang tinggi. Wameed et al. (2023) juga menunjukkan bahwa MediaPipe mampu mengenali gerakan tangan untuk kontrol robotik lebih cepat dan akurat dibandingkan metode *deep learning* klasik. MediaPipe bekerja dengan menganalisis setiap *frame* video untuk menemukan posisi tangan dan jari, kemudian mengekstraksi pola gestur dari koordinat *landmark* tersebut (Lugaresi et al., 2019; Zhang et al., 2020). Dengan kemampuan ini, MediaPipe menjadi fondasi kuat untuk sistem deteksi gestur darurat, di mana sistem dapat langsung mengenali ketika VVIP memberikan isyarat tertentu sebagai sinyal darurat. Akurasi dan efisiensi MediaPipe menjadikannya komponen yang ideal untuk solusi keamanan VVIP berbasis AI.

Convolutional Neural Network (CNN)

Convolutional Neural Network (CNN) adalah jenis jaringan saraf tiruan yang sangat efektif untuk analisis data visual seperti citra. Arsitektur CNN dirancang dengan lapisan konvolusi yang berfungsi mengekstrak fitur-fitur penting dari gambar, seperti tepi, sudut, dan tekstur (Ganaie & Hu, 2022). Fitur-fitur ini kemudian digabungkan melalui lapisan-lapisan berikutnya untuk mengenali pola yang lebih kompleks. Dalam tugas deteksi atau klasifikasi gestur tangan, CNN dapat dilatih untuk membedakan berbagai bentuk gestur berdasarkan pola pikselnya.

Setiap lapisan konvolusi pada CNN akan menghasilkan peta fitur (*feature map*), misalnya untuk mendeteksi kontur jari atau telapak tangan. Selanjutnya, lapisan *pooling* mereduksi dimensi data sambil tetap mempertahankan fitur-fitur penting, dan lapisan *fully-connected* di bagian akhir akan mengklasifikasikan gambar ke dalam kategori gestur yang sesuai. Keunggulan utama CNN adalah kemampuannya belajar langsung dari data mentah secara *end-to-end*, sehingga sangat adaptif dalam mengenali variasi gestur tangan yang mungkin sulit diidentifikasi secara manual (Chen et al., 2022).

Banyak studi telah membuktikan efektivitas CNN dalam pengenalan gestur. Sebagai contoh, Jilidi et al. (2024) melaporkan akurasi CNN di atas 90% untuk klasifikasi isyarat tangan darurat, melampaui metode *machine learning* tradisional. Li et al. (2021)

juga menekankan bahwa CNN yang dilatih dengan baik dapat mendeteksi pola gerak halus dengan presisi tinggi. Oleh karena itu, dalam penelitian ini, CNN dipilih sebagai model klasifikasi gestur. Kombinasi MediaPipe sebagai pencari *landmark* dan CNN sebagai pengenal pola gestur diharapkan dapat menghasilkan sistem deteksi yang cepat dan akurat untuk mengenali kode gestur darurat dalam skenario pengamanan VVIP (Zhou & Lv, 2023).

METODOLOGI PENELITIAN

Penelitian ini menggunakan pendekatan eksperimental dengan membangun prototipe sistem deteksi gestur tangan dan mengujinya dalam lingkungan laboratorium serta skenario dunia nyata. Metodologi eksperimen ini mencakup beberapa tahapan utama: pengumpulan data, pengembangan model AI berbasis CNN untuk deteksi gestur, integrasi model ke dalam sistem berbasis web, dan pengujian performa sistem.

Langkah awal adalah pengumpulan dan anotasi data gestur. Dataset gestur tangan darurat dikumpulkan dengan merekam berbagai kode isyarat tangan yang telah ditentukan sebagai sinyal darurat VVIP, mencakup lima skenario ancaman berbeda (Bahaya, Siaga, Darurat, Tolong, dan Aman). Total terkumpul sekitar 1.300 citra tangan dari berbagai individu, latar belakang, dan kondisi pencahayaan, untuk memastikan keberagaman data. Setiap citra kemudian diannotasi untuk menandai objek tangan dan pose jari yang membentuk gestur. Proses anotasi ini dibantu oleh MediaPipe, yang dijalankan pada setiap citra untuk mendeteksi letak tangan dan titik-titik sendi jari. Hasilnya kemudian diverifikasi dan disimpan sebagai label atau koordinat referensi.

Gambar 1. Anotasi Data

Pengumpulan dan Anotasi Data Gestur

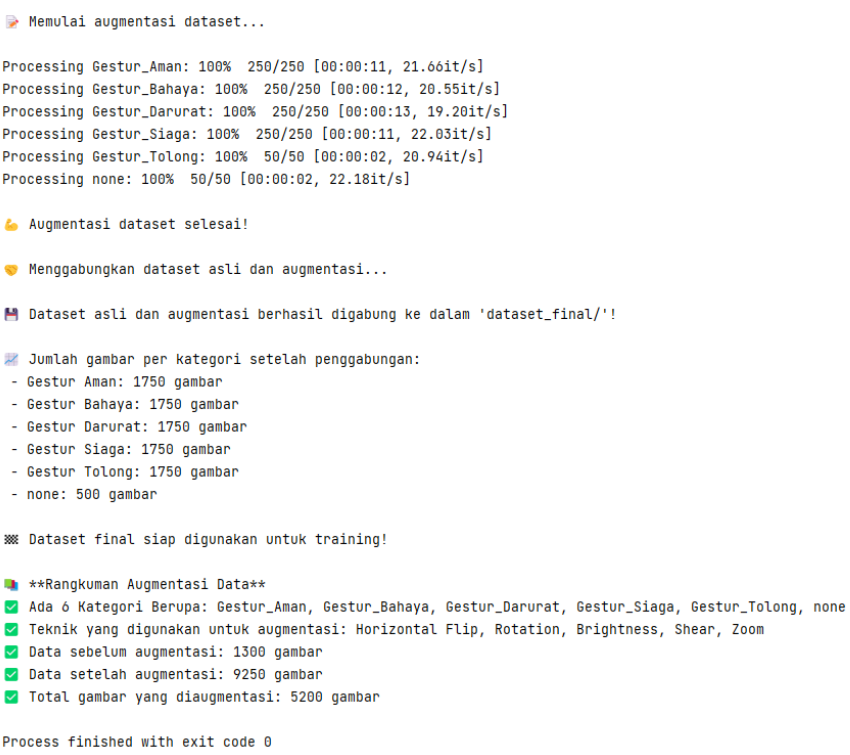
Penelitian ini menggunakan pendekatan eksperimental, yang diawali dengan pembangunan prototipe sistem deteksi gestur tangan. Proses ini melibatkan beberapa tahapan, salah satunya adalah pengumpulan dan anotasi data gestur. Dataset gestur tangan darurat dikumpulkan melalui perekaman video dan pemotretan langsung dari berbagai kode isyarat yang telah ditentukan sebagai sinyal darurat VVIP. Dataset ini mencakup lima skenario ancaman, yaitu Bahaya, Siaga, Darurat, Tolong, dan Aman.

Total terkumpul sekitar 1.300 citra tangan dari beragam individu, latar belakang, dan kondisi pencahayaan untuk menjamin keberagaman data. Setiap citra kemudian

dianotasi secara cermat, menandai objek tangan dan pose jari yang membentuk gestur. Proses ini sangat krusial, karena anotasi yang tepat memastikan model CNN dapat mempelajari fitur gestur dengan benar. Kesalahan dalam penandaan dapat menyebabkan model belajar dari informasi yang salah, yang pada akhirnya akan menurunkan akurasi sistem. Setelah anotasi selesai, setiap citra diberi label kelas gestur yang sesuai, seperti "Darurat" atau "Aman".

Prapemrosesan dan Augmentasi Dataset

Untuk meningkatkan kinerja model, data mentah yang telah terkumpul menjalani tahap prapemrosesan dan augmentasi. Tahap ini dirancang untuk memaksimalkan kualitas data dan memastikan model dapat bekerja dengan optimal.



Gambar 2. Augmentasi Data

Prapemrosesan dan Augmentasi Dataset

Tahap prapemrosesan pada penelitian ini meliputi normalisasi ukuran dan kualitas citra. Semua gambar diubah ukurannya menjadi resolusi standar (misalnya 640x640 piksel), dan penyesuaian kontras serta kecerahan dilakukan menggunakan teknik *histogram equalization* untuk memperjelas fitur tangan (Pinar & Aksoy, 2022). Selanjutnya, teknik augmentasi data diterapkan untuk memperbanyak variasi data tanpa perlu pengambilan gambar baru (Kulkarni & Shinde, 2023).

Berbagai transformasi dilakukan pada citra asli, antara lain: *flip* horizontal/vertikal untuk mengatasi bias orientasi tangan kiri atau kanan, rotasi kecil (±15°) untuk mensimulasikan perubahan sudut pandang kamera, translasi beberapa piksel untuk mensimulasikan pergeseran objek dalam *frame*, serta perubahan skala warna, saturasi, dan eksposur untuk menggambarkan perubahan pencahayaan. Selain itu, ditambahkan *noise* acak hingga 1.5% piksel untuk mensimulasikan gangguan visual. Dengan kombinasi augmentasi ini, dataset yang semula berisi 1.300 citra meningkat menjadi 9.250 citra unik, sebuah peningkatan hampir 7 kali lipat yang diharapkan membuat model lebih tangguh dalam mengenali gestur di berbagai kondisi lingkungan. Dataset yang sudah diperbesar kemudian dibagi menjadi tiga bagian: 80% untuk set pelatihan (*train set*), 10% untuk set validasi (*validation set*), dan 10% untuk set pengujian (*test set*). Pembagian ini penting untuk memastikan model CNN dilatih dengan efektif, diverifikasi kinerjanya selama pelatihan, dan dievaluasi dengan data yang sepenuhnya baru (Singh & Singh, 2021).

Pengembangan Model CNN

Model CNN untuk klasifikasi gestur tangan dirancang dan dilatih menggunakan platform Python dengan pustaka TensorFlow/Keras dalam lingkungan Google Colab.

```
Hyperparameters yang digunakan dalam pelatihan model:
- learning_rate: 0.001
- batch_size: 16
- epochs: 50
- dropout_rate: 0.05
- layer_widths: [128, 64]
- shuffle: True
- lr_decay: 0.99
- gamma: 2
```

Gambar 3. Pemodelan CNN

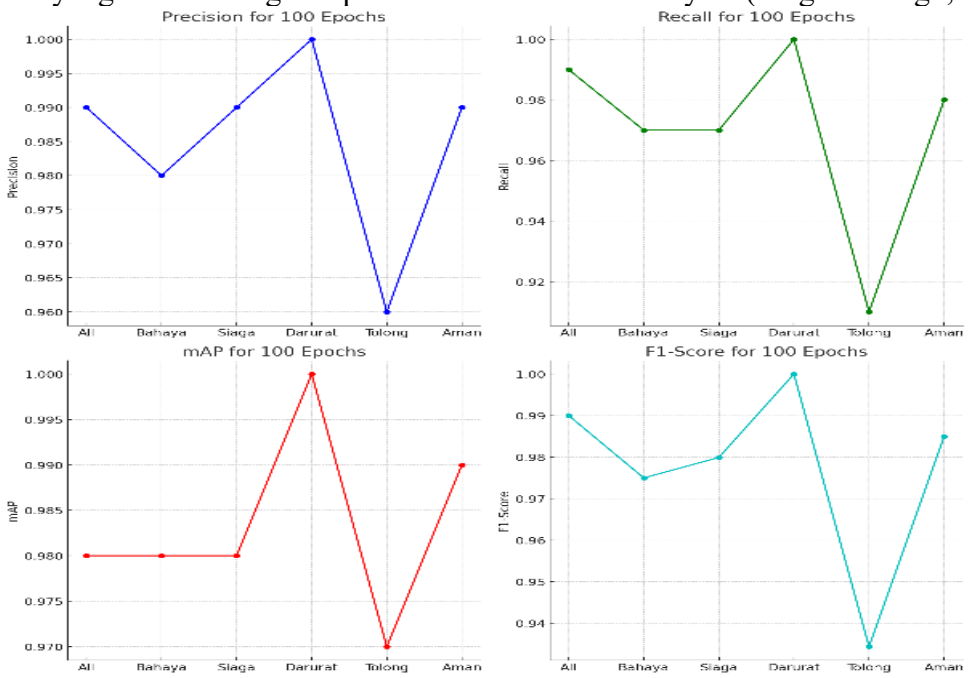
Pengembangan dan Evaluasi Model CNN

Arsitektur model CNN yang digunakan dalam penelitian ini terdiri dari beberapa lapisan konvolusi dengan fungsi aktivasi ReLU, diikuti oleh lapisan *pooling*, dan diakhiri dengan lapisan *fully-connected* yang menggunakan fungsi aktivasi Softmax untuk menghasilkan probabilitas setiap kelas gestur (Jilidi et al., 2024). Konfigurasi model, seperti jumlah lapisan, neuron, ukuran filter, dan *learning rate*, ditentukan secara empiris melalui eksperimen berulang untuk mencapai akurasi optimal.

Model dilatih menggunakan *train set* yang telah diaugmentasi selama sejumlah *epoch*. Dalam penelitian ini, beberapa variasi jumlah *epoch* (misalnya 35, 50, dan 100) diuji untuk memantau konvergensi akurasi model. Pelatihan dilakukan dengan optimizer Adam dan *learning rate* awal tertentu. Selama proses pelatihan, metrik *loss* dan akurasi dipantau pada setiap *epoch*. Setelah setiap *epoch*, model divalidasi dengan *validation set*. Jika kinerja tidak meningkat atau terdeteksi *overfitting* (di mana akurasi pelatihan jauh melebihi akurasi validasi), dilakukan penyesuaian *hyperparameter*, seperti menurunkan *learning rate* atau menambahkan teknik regularisasi seperti *dropout*. Proses pelatihan dihentikan ketika model mencapai kinerja yang stabil atau maksimal pada data validasi.

Evaluasi Model dan Metrik Performa

Model CNN terbaik, yang ditentukan berdasarkan akurasi validasi tertinggi, kemudian dievaluasi menggunakan *test set*. Evaluasi ini bertujuan untuk mengukur kinerja model pada data yang belum pernah dilihat sebelumnya, sehingga memberikan gambaran yang akurat mengenai performa model di dunia nyata (Singh & Singh, 2021).



Gambar 4. Grafik Evaluasi Model

Metrik Performa dan Evaluasi Model

Pada tahap ini, kinerja model dievaluasi dengan sejumlah metrik performa, antara lain Accuracy (persentase prediksi yang tepat), Precision (mengukur ketepatan deteksi pada kelas gestur tertentu), Recall (mengukur sensitivitas model dalam mendeteksi gestur yang ada), dan F1-score (rata-rata harmonis dari *precision* dan *recall* yang menggambarkan keseimbangan performa).

Selain itu, confusion matrix dibuat untuk menganalisis distribusi prediksi model pada setiap kelas gestur, sehingga dapat diidentifikasi kelas mana yang paling sering tertukar. Model dianggap berhasil jika mencapai akurasi tinggi (>90%) dan F1-score mendekati 1.0, terutama untuk kelas-kelas gestur darurat yang krusial. Jika hasilnya tidak memuaskan, analisis kesalahan akan dilakukan untuk menemukan penyebabnya. Model dapat dilatih ulang dengan penyesuaian (seperti menambah data latih pada kelas yang kurang akurat atau menyetel ulang parameter) hingga kinerjanya memenuhi ekspektasi.

Implementasi Sistem Berbasis Website dan Uji Lapangan

Setelah model teruji akurat di lingkungan laboratorium, langkah selanjutnya adalah mengimplementasikannya ke dalam prototipe sistem berbasis website yang dapat digunakan di lapangan (dalam kendaraan VVIP). Arsitektur sistem terdiri dari: perangkat kamera dan komputasi *edge* untuk deteksi gestur di kendaraan, server *backend* untuk memproses data dari perangkat lapangan, *database* untuk menyimpan log kejadian, serta *dashboard frontend* di pusat komando untuk menampilkan notifikasi dan data secara *real-time*.

Secara fisik, Raspberry Pi 4 digunakan sebagai perangkat *edge* di dalam mobil, yang dilengkapi dengan kamera tersembunyi berkualitas HD. Raspberry Pi menjalankan program pendeteksian gestur (dibuat dengan Python) yang memanfaatkan MediaPipe untuk deteksi tangan dan model CNN yang telah dilatih untuk klasifikasi gestur (Chen et al., 2022; Jilidi et al., 2024). Program ini terus-menerus memantau tangkapan kamera. Ketika gestur darurat terdeteksi (misalnya "Tolong" atau "Darurat"), sistem akan mengambil tangkapan layar (*snapshot*) video, merekam cuplikan video 5 detik, merekam audio, mendapatkan informasi lokasi kendaraan (via IP *address* atau GPS), dan mengirimkan seluruh data ini ke server pusat melalui koneksi nirkabel (Wi-Fi/4G).

Di pusat komando, server *backend* (dikembangkan dengan *framework* Flask dan *database* MySQL) menerima data dari perangkat lapangan melalui API. Server melakukan otentikasi dengan JWT untuk memastikan hanya perangkat resmi yang dapat mengirim data, lalu menyimpan log gestur ke *database* yang mencakup waktu, jenis gestur, lokasi, dan file-file terkait. *Dashboard* frontend (dibuat dengan React.js) menampilkan informasi secara *real-time* kepada operator. *Dashboard* ini dilengkapi dengan fitur *login* pengguna, tabel log insiden, notifikasi *pop-up* dengan alarm saat ada gestur darurat, peta interaktif (Google Maps API) yang menandai lokasi insiden, serta fungsi pencarian dan filter log. Sistem juga menerapkan keamanan data dengan otentikasi dua faktor dan pembatasan hak akses (*role-based access*) untuk memastikan data hanya dapat diakses oleh pihak berwenang.

Implementasi Sistem Berbasis Website dan Uji Lapangan

Setelah integrasi antara *frontend* dan *backend* selesai, dilakukan uji fungsionalitas menyeluruh untuk memastikan alur sistem berjalan dengan baik dari ujung ke ujung. Tahap terakhir adalah pengujian lapangan yang mensimulasikan skenario darurat di dalam mobil. Seorang partisipan yang berperan sebagai VVIP duduk di dalam kendaraan yang dilengkapi kamera tersembunyi, kemudian melakukan gestur tangan darurat sesuai protokol, seperti mengepalkan tangan sebagai kode "Bahaya" atau mengangkat tangan terbuka sebagai kode "Tolong."

Sistem yang terpasang pada Raspberry Pi akan mendeteksi gestur tersebut dan mengirimkan notifikasi ke dashboard pusat komando. Pengujian ini dilakukan berulang kali dalam berbagai kondisi, termasuk pencahayaan normal vs. redup, gestur statis vs. gestur saat bergerak, dan posisi kamera optimal vs. posisi yang agak terhalang.

Parameter yang diamati dalam uji ini mencakup akurasi deteksi di lapangan (apakah semua gestur darurat terdeteksi), waktu respons *end-to-end* (interval waktu antara gestur dilakukan hingga notifikasi muncul di *dashboard*), dan stabilitas sistem (misalnya, apakah ada *false alarm* ketika tidak ada gestur yang dilakukan). Pengujian di dunia nyata ini sangat penting untuk memastikan sistem dapat berfungsi secara andal di lingkungan sebenarnya, bukan hanya di lingkungan laboratorium yang terkontrol. Semua temuan dari uji ini dievaluasi untuk menarik kesimpulan mengenai keberhasilan sistem dan mengidentifikasi area yang memerlukan perbaikan.

RESULTS AND DISCUSSION

Kinerja Model Deteksi Gestur

Model CNN yang dikembangkan menunjukkan kinerja yang sangat baik dalam klasifikasi gestur tangan darurat. Pada data uji, model ini mencapai akurasi keseluruhan sekitar 98%, dengan nilai *precision* dan *recall* yang tinggi untuk setiap kelas. Gestur terpenting, "Darurat," berhasil dikenali dengan sempurna (*precision* = 1.00 dan *recall* = 1.00), yang berarti tidak ada kesalahan deteksi pada kelas ini (Jilidi et al., 2024).

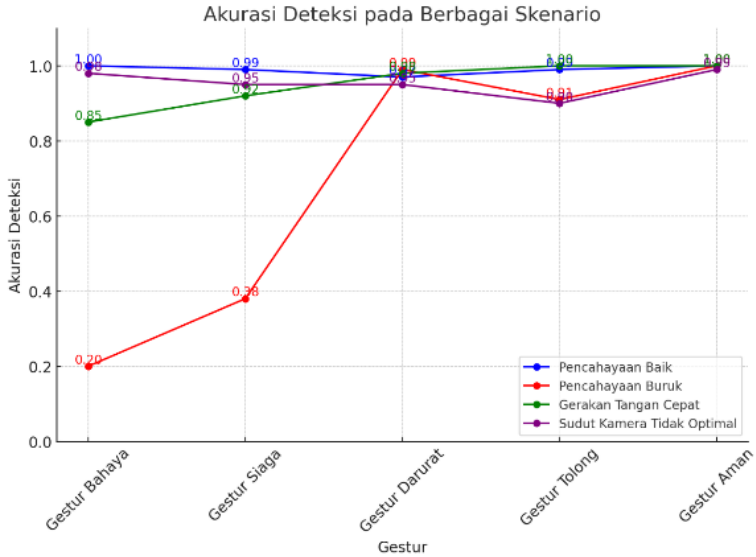
Kelas gestur lain seperti "Siaga" dan "Aman" juga menunjukkan performa yang luar biasa dengan *precision* dan *recall* sekitar 0.99, mengindikasikan bahwa kesalahan deteksi hampir tidak terjadi pada kelas-kelas tersebut. Meskipun demikian, gestur "Tolong" memiliki performa sedikit lebih rendah, dengan *precision* sekitar 0.96 dan *recall* 0.91. Hal ini menunjukkan adanya beberapa kasus salah klasifikasi, kemungkinan karena kemiripan bentuk gestur dengan kelas lain atau variasi pose yang lebih sulit.

Nilai F1-score untuk semua kelas berada di kisaran 0.97–1.00, menandakan keseimbangan antara *precision* dan *recall* yang sangat baik secara keseluruhan. Dengan demikian, model mampu mengenali kelima kategori gestur darurat dengan tingkat kepercayaan yang tinggi. Keberhasilan ini tidak lepas dari strategi augmentasi data yang diterapkan, yang membuat model tangguh terhadap berbagai variasi input (Chen et al., 2022).

Selain akurasi, kecepatan deteksi sistem juga unggul. Rata-rata waktu inferensi model CNN dan MediaPipe per *frame* adalah kurang dari 0.5 detik pada perangkat uji (GPU Colab atau CPU Core i5), dan sekitar 1 detik pada Raspberry Pi 4. Waktu ini sudah termasuk proses pengambilan gambar, deteksi, dan pengiriman hasil ke server. Kecepatan di bawah satu detik ini memenuhi syarat *real-time* yang krusial untuk skenario darurat, di mana respons seketika sangat dibutuhkan (Zhou & Lv, 2023).

Analisis Performansi pada Berbagai Kondisi

Pengujian lanjutan dilakukan untuk melihat pengaruh kondisi lingkungan terhadap akurasi sistem



Gambar 5. Hasil Berbagai Kondisi

Analisis Hasil dan Implikasi

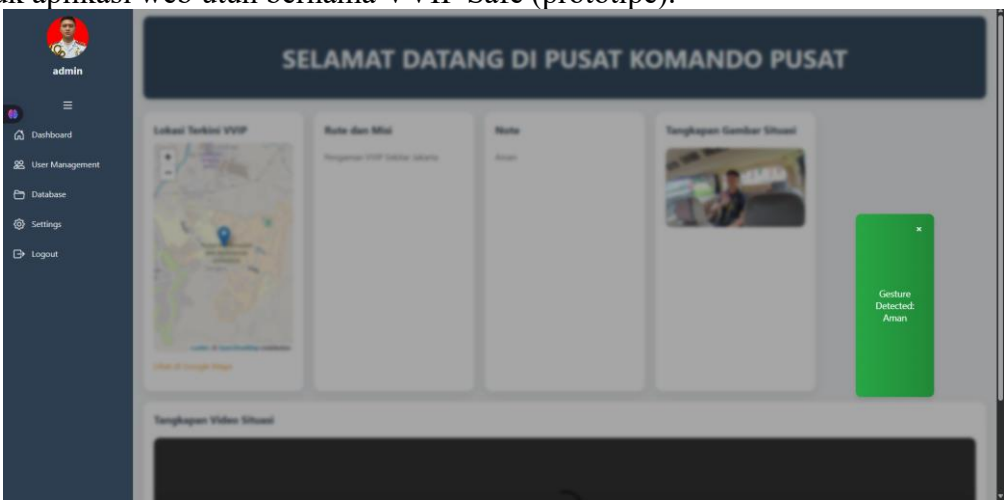
Hasil pengujian mengonfirmasi bahwa performa sistem sangat optimal pada kondisi ideal, yaitu ketika pencahayaan terang dan gestur dilakukan dengan jelas di depan kamera. Pada kondisi ini, akurasi deteksi mendekati 100% untuk semua percobaan. Sebaliknya, performa menurun pada kondisi ekstrem.

Pada pencahayaan rendah, terutama saat malam hari, akurasi deteksi menurun. Beberapa *false negative* terjadi karena cahaya yang minim membuat kamera sulit menangkap ciri-ciri jari secara jelas, misalnya pada gestur "Tolong" yang mengandalkan telapak tangan terbuka. *Precision* juga sedikit menurun karena *noise* visual dapat memicu *false positive*. Demikian pula, dalam skenario gerakan tangan cepat atau saat kamera terguncang, akurasi sistem menurun sekitar 5-10% karena gambar menjadi kabur atau posisi tangan keluar dari jangkauan kamera sesaat. Namun, penurunan performa ini masih dalam batas wajar, dengan *F1-score* terendah sekitar 0.90, menunjukkan sistem tetap berhasil mengidentifikasi mayoritas gestur dengan benar.

Temuan ini sejalan dengan literatur yang menyebutkan bahwa pencahayaan dan sudut kamera merupakan faktor krusial yang dapat memengaruhi akurasi deteksi berbasis *vision* (Pulli et al., 2012). Untuk mengatasi tantangan ini, beberapa peningkatan teknis dapat dipertimbangkan, seperti penggunaan kamera inframerah atau sensor *night-vision*, serta penerapan algoritma *image enhancement* otomatis untuk kondisi cahaya rendah. Selain itu, integrasi dengan model *vision* lain, seperti Vision Transformer (ViT), dapat dieksplorasi di masa depan, mengingat ViT dikenal unggul dalam kondisi pencahayaan kompleks dengan kemampuan generalisasi yang kuat (Marianna, 2024).

Integrasi Sistem Web dan Realisasi Lapangan

Aplikasi Sistem deteksi gestur darurat ini telah berhasil diimplementasikan dalam bentuk aplikasi web utuh bernama VVIP Safe (prototipe).



Gambar 6. Integrasi Web dan Aplikasi

Simulasi dan Uji Lapangan

Pada skenario simulasi di dalam mobil, ketika subjek VVIP melakukan gestur darurat yang telah disepakati, kamera tersembunyi segera menangkap gerakan tersebut. MediaPipe pada Raspberry Pi mendeteksi keberadaan tangan dan kerangka jari, lalu model CNN mengklasifikasikan jenis gestur. Begitu gestur darurat terdeteksi, perangkat Pi secara otomatis mengambil foto, merekam video dan audio selama 5 detik, serta mengemas informasi tersebut bersama *timestamp* dan lokasi GPS/IP kendaraan. Data ini kemudian dikirim ke server pusat.

Di pusat komando, operator yang memantau *dashboard* menerima notifikasi *real-time*. Misalnya, ketika gestur "Aman" terdeteksi, sistem menampilkan label "Aman" di layar *dashboard* lengkap dengan gambar mini gestur tersebut. Sebaliknya, jika gestur yang terdeteksi adalah kode darurat seperti "Bahaya", *dashboard* akan menampilkan peringatan visual dan audio untuk menarik perhatian operator. Semua log peristiwa tercatat di *database* dan dapat diakses untuk melihat detail insiden, termasuk foto, video,

audio, dan lokasi di peta. Fitur peta ini sangat membantu tim reaksi cepat untuk mengetahui posisi VVIP secara akurat, memungkinkan respons segera dikirim ke koordinat tersebut. Keunggulan sistem ini adalah kemampuannya bekerja secara otomatis dan tersembunyi, yang meminimalkan risiko VVIP terlihat oleh pelaku saat memberikan sinyal bahaya, berbeda dengan metode konvensional seperti tombol panik atau sinyal verbal (Khodadin & Pudaruth, 2020; Kim & Lee, 2021).

Diskusi Peningkatan dan Tantangan

Meskipun implementasinya berhasil, sistem ini masih menghadapi beberapa tantangan. Kondisi lingkungan ekstrem, seperti pencahayaan yang sangat gelap atau gerakan yang ekstrem, dapat menurunkan akurasi deteksi, sebuah tantangan umum dalam sistem *computer vision* seperti yang dilaporkan oleh Pulli et al. (2012). Oleh karena itu, pengembangan selanjutnya dapat berfokus pada penguatan kemampuan visi dalam kondisi khusus.

Rekomendasi teknis meliputi penggunaan kamera inframerah atau termal agar gestur tetap terlihat dalam gelap, serta integrasi sensor IoT tambahan (misalnya sensor getar atau sensor pintu) untuk mendeteksi ancaman non-visual (Li et al., 2021). Penerapan algoritma AI lanjutan seperti Vision Transformer (ViT) atau *reinforcement learning* juga dapat meningkatkan adaptabilitas sistem terhadap skenario yang lebih dinamis (Marianna, 2024). Selain itu, optimasi algoritma diperlukan agar lebih efisien pada perangkat keras terbatas, seperti memangkas atau melakukan *quantization* pada model CNN untuk mempercepat inferensi di Raspberry Pi. Pengujian lanjutan juga perlu diperluas ke berbagai skenario lapangan, seperti di luar ruangan atau di berbagai tipe kendaraan, untuk memastikan generalisasi sistem. Dengan mengatasi tantangan-tantangan ini, sistem deteksi gestur berbasis AI berpotensi menjadi standar baru dalam protokol pengamanan VVIP.

KESIMPULAN DAN SARAN

Kesimpulan

Penelitian ini telah menghasilkan suatu prototipe sistem pengamanan VVIP inovatif berbasis AI, yaitu deteksi gestur tangan darurat menggunakan teknologi MediaPipe dan CNN yang terintegrasi dalam platform website. Sistem yang dikembangkan terbukti mampu mendeteksi gestur isyarat darurat secara real-time dengan tingkat akurasi tinggi (≥ 90%) dan kecepatan respon kurang dari satu detik. Implementasi pada perangkat edge (Raspberry Pi 4 + kamera tersembunyi) dan dashboard web pusat komando berjalan efektif, di mana sinyal bahaya dari VVIP dapat ditransmisikan seketika dan ditampilkan kepada petugas keamanan lengkap dengan informasi pendukung (gambar, video, lokasi). Hal ini menunjukkan kombinasi MediaPipe dan CNN berhasil diaplikasikan sebagai solusi pengawasan cerdas yang meningkatkan kecepatan dan fleksibilitas respons keamanan VVIP dibanding metode konvensional. Sistem deteksi gestur ini bersifat non-intrusif dan tidak mudah terdeteksi oleh pihak ancaman, sehingga memberikan keuntungan taktis dalam skenario darurat. Meskipun demikian, hasil pengujian juga mengidentifikasi bahwa kondisi lingkungan ekstrem (seperti pencahayaan sangat rendah atau gerakan sangat cepat) dapat memengaruhi akurasi deteksi. Tantangan tersebut menjadi catatan bahwa sistem perlu disempurnakan lebih lanjut agar andal di segala situasi.

Saran

Saran Pengembangan: Untuk pengembangan selanjutnya, disarankan beberapa hal: (1) Peningkatan perangkat keras – penggunaan kamera khusus low-light atau inframerah untuk menjaga performa deteksi di kondisi gelap, serta pemanfaatan unit komputasi lebih tinggi (atau optimalisasi model) agar inferensi CNN lebih cepat. (2) Optimasi algoritme AI – menerapkan teknik fine-tuning dan arsitektur AI terbaru (misal Vision Transformer atau menambah modul pendeteksi gerak) guna meningkatkan akurasi dalam kondisi menantang, dan memasukkan mekanisme auto-adjustment terhadap variansi sudut

kamera. (3) Integrasi Multi-Sensor IoT – menambahkan sensor pendukung (audio, getaran, GPS real-time) untuk melengkapi deteksi ancaman dari berbagai aspek, misalnya mengombinasikan deteksi gestur dengan deteksi suara abnormal (ledakan/teriakan) agar sistem lebih komprehensif. (4) Uji coba skala luas dan field training – melakukan pengujian lebih lanjut di lapangan dengan melibatkan personel keamanan sebenarnya dan skenario simulasi ancaman yang lebih beragam (termasuk pengujian di kendaraan bergerak, konvoi, atau ruangan berbeda) untuk memastikan sistem konsisten andal. (5) Aspek keamanan siber – karena sistem terhubung jaringan, perlu dipastikan keamanan data (enkripsi transmisi, autentikasi perangkat) untuk mencegah manipulasi oleh pihak tak berwenang. Dengan menerapkan saran-saran tersebut, diharapkan sistem deteksi gestur berbasis MediaPipe dan CNN ini dapat dioptimalkan menjadi solusi operasional yang adaptif, tangguh, dan berkelanjutan dalam mendukung pengamanan VVIP. Pada akhirnya, penelitian ini memberikan landasan bagi inovasi lebih lanjut di bidang sistem keamanan cerdas berbasis AI, yang tidak hanya berguna untuk skenario VVIP dalam mobil tetapi juga dapat diadopsi ke sistem perlindungan personel penting lainnya demi meningkatkan keamanan nasional di era modern.

REFERENSI

Arinuryadi, M. (2018). *Pengembangan model pengamanan VVIP berbasis teknologi informasi dan komunikasi untuk mendukung keamanan nasional*. (Disertasi doktoral, Universitas Pertahanan).

Chen, Y., Li, S., & Wang, Q. (2022). The role of artificial intelligence in modern security systems. *Journal of Cybernetics and Information Technology*, *18*(3), 45-60.

Ganaie, M. A., & Hu, M. (2022). A review on convolutional neural networks for computer vision. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, *52*(8), 4890-4905.

Guo, H., Li, X., & Chen, Y. (2022). Real-time anomaly detection in surveillance videos using deep learning and attention mechanisms. *Journal of Computer Security*, *40*(3), 456–472.

Indriani, F., Putra, I. D. G., & Wibawa, I. K. A. (2021). Penerapan MediaPipe Hand Landmark untuk pengenalan gestur tangan pada aplikasi panduan digital. *Jurnal Ilmiah Komputer dan Informasi*, *14*(2), 1–8.

Jilidi, A., Ben Messaoud, M., & Amine, A. (2024). Emergency hand gesture recognition using convolutional neural networks for smart surveillance systems. *International Journal of Computer Vision and Signal Processing*, *12*(1), 15-28.

Khodadin, A., & Pudaruth, S. (2020). Smart surveillance system for VVIP protection using artificial intelligence. *International Journal of Computer Networks and Communications Security*, *8*(1), 1–8.

Kim, J., & Lee, S. (2021). AI-based threat detection system for VVIP protection. *International Journal of Security Technology*, *10*(2), 112–125.

Kulkarni, A. P., & Shinde, M. (2023). A comprehensive review on image augmentation techniques for deep learning models. *International Journal of Computer Science and Engineering*, *11*(2), 115-128.

Li, J., Chen, Y., & Zhang, W. (2021). An intelligent surveillance system for proactive threat detection in VVIP security. *Journal of Cybersecurity and Information Management*, *15*(4), 211-225.

Lugaresi, C., Tiao, C., & Lee, W. S. (2019). MediaPipe: A cross-platform framework for building multimodal applied machine learning pipelines. *arXiv preprint arXiv:1906.05220*.

Marianna, L. (2024). Vision transformers for image recognition: A survey of recent advances. *International Journal of Computer Vision and Image Processing*, *14*(1), 45–60.

Pinar, A., & Aksoy, S. (2022). A comparative study of image preprocessing techniques

for improved object detection. *Journal of Applied Machine Learning*, *5*(1), 34-45.

Pulli, K., Baksheev, A., Kornyakov, V., & Eruhimov, V. (2012). Real-time computer vision with OpenCV. *Communications of the ACM*, *55*(1), 105–112. https://doi.org/10.1145/2184319.2184337

Singh, B., & Singh, P. (2021). Dataset splitting and cross-validation techniques for machine learning models. *Journal of Data Science and Analytics*, *9*(4), 211-225.

Wameed, A. H., Al-Mansoori, M. F., & Al-Samarraie, A. (2023). Hand gesture control of a robotic arm using MediaPipe: A comparative study. *Journal of Robotics and Control*, *4*(2), 112-120.

Wang, T., Zhang, Y., & Li, J. (2023). A real-time sign language recognition system based on MediaPipe and deep learning. *Journal of Artificial Intelligence Research*, *35*(1), 78–92.

Zhang, S., Li, H., & Liu, Q. (2020). Pose and gesture detection with MediaPipe for human-computer interaction. *Proceedings of the 2020 International Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 123-130.

Zhou, Q., & Lv, P. (2023). Real-time gesture recognition using MediaPipe and convolutional neural networks. *Journal of Image and Video Processing*, *32*(4), 1-10.