



## Pengenalan Gerakan Tangan untuk Kontrol Slide Presentasi Menggunakan Framework MediaPipe, OpenCV, dan Model LSTM

Muhammad Rizaldi<sup>1\*</sup>, Rudi Heriansyah<sup>2</sup>, Nazori Suhandi<sup>3</sup>

<sup>1,2,3</sup> Teknik Informatika, Ilmu Komputer dan Sains, Universitas Indo Global Mandiri, Palembang, Indonesia

Email: [2021110037@students.uigm.ac.id](mailto:2021110037@students.uigm.ac.id)<sup>1\*</sup>, [rudi@uigm.ac.id](mailto:rudi@uigm.ac.id)<sup>2</sup>, [nazori@uigm.ac.id](mailto:nazori@uigm.ac.id)<sup>3</sup>

Alamat: Jl. Jend. Sudirman Km.4 No. 62, 20 Ilir D. IV, Kec. Ilir Tim. I, Kota Palembang, Sumatera Selatan 30129

\*Korespondensi penulis

**Abstract.** *In presentation activities, the use of physical devices such as a mouse or remote often limits the presenter's mobility and reduces the effectiveness of interaction with the audience. This study aims to implement a hand gesture recognition system as an alternative solution to control presentation slides in real-time without additional devices. The system was developed using the MediaPipe framework for hand landmark detection, OpenCV for video image processing, and a Long Short-Term Memory (LSTM) model for sequential gesture classification. Three main gestures were defined as commands, namely "Next," "Previous," and "Idle," with input taken from live video streaming at a distance of 1–3 meters. The development process included manual labeling of gesture data from multiple users, training the LSTM model with sequential data, and testing the system in real-time integrated with Microsoft PowerPoint. Experimental results indicate that the system successfully recognized hand gestures with high accuracy across most scenarios, with optimal performance observed at distances of 1–2 meters. However, accuracy decreased under low-light conditions or when gestures were performed too quickly. These findings demonstrate that the combination of MediaPipe, OpenCV, and LSTM is effective in building a gesture-based presentation control system. Beyond enhancing flexibility and interactivity in presentations, this research also contributes to the development of more natural and practical human-computer interaction systems, while offering opportunities for broader applications in other domains.*

**Keywords:** LSTM; MediaPipe; OpenCV; Recognition Gesture; Slide Control

**Abstrak.** Dalam aktivitas presentasi, penggunaan perangkat fisik seperti mouse atau remote sering kali membatasi ruang gerak presenter dan mengurangi efektivitas interaksi dengan audiens. Penelitian ini bertujuan mengimplementasikan sistem pengenalan gerakan tangan sebagai solusi alternatif untuk mengontrol slide presentasi secara real-time tanpa memerlukan perangkat tambahan. Sistem dibangun dengan memanfaatkan framework MediaPipe untuk deteksi landmark tangan, OpenCV untuk pengolahan citra video, serta model Long Short-Term Memory (LSTM) untuk klasifikasi pola gerakan sekuensial. Tiga gestur utama yang digunakan sebagai perintah adalah "Next", "Previous", dan "Diam", dengan input berupa streaming video dari kamera pada jarak 1–3 meter. Proses pengembangan meliputi pelabelan manual data gestur dari berbagai pengguna, pelatihan model menggunakan data sekuensial, serta pengujian sistem secara real-time yang diintegrasikan dengan Microsoft PowerPoint. Hasil uji coba menunjukkan bahwa sistem mampu mengenali gestur tangan dengan tingkat akurasi yang tinggi pada sebagian besar skenario, dengan performa optimal pada jarak penggunaan 1–2 meter. Namun, sistem mengalami penurunan akurasi pada kondisi pencahayaan rendah atau gerakan tangan yang terlalu cepat. Temuan ini menunjukkan bahwa kombinasi MediaPipe, OpenCV, dan model LSTM efektif dalam membangun sistem kontrol presentasi berbasis gestur. Selain meningkatkan fleksibilitas dan interaktivitas, penelitian ini juga memberikan kontribusi pada pengembangan interaksi manusia-komputer yang lebih alami dan praktis, sekaligus membuka peluang penerapan lebih luas di bidang lain.

**Kata kunci:** Kontrol Geser; LSTM; MediaPipe; OpenCV; Pengenalan Gerakan

## 1. LATAR BELAKANG

Presentasi merupakan salah satu metode komunikasi yang digunakan untuk menyampaikan informasi secara visual dan verbal kepada audiens dalam berbagai konteks, seperti akademik, bisnis, dan seminar (Asmara et al., 2021). Presentasi memiliki peran penting dalam menyampaikan ide, menjelaskan konsep, serta meyakinkan audiens mengenai suatu topik tertentu.

Namun, di balik pentingnya presentasi, terdapat tantangan teknis yang sering kali menghambat efektivitas penyampaian materi. Pembicara sering mengalami kendala saat harus berpindah posisi atau bergerak aktif di depan audiens, karena perangkat kontrol seperti *mouse*, *remote*, atau *keyboard* tidak selalu berada dalam jangkauan yang optimal. Keterbatasan ini terutama terasa dalam lingkungan pendidikan dan bisnis yang menuntut interaksi aktif antara pembicara dan audiens, sehingga dapat mengurangi kualitas dan dinamika presentasi (Khairianto & Firdaus, 2024a).

Untuk mengatasi permasalahan ini, penelitian ini mengeksplorasi pendekatan inovatif dengan memanfaatkan teknologi pengolahan citra dan pembelajaran mesin. Secara khusus, penelitian ini akan mengintegrasikan *MediaPipe* untuk mendeteksi *landmark* tangan secara *real-time*, *OpenCV* untuk pemrosesan citra, serta *Long Short-Term Memory* (LSTM) untuk mengenali pola gerakan berurutan. Kombinasi ketiga teknologi ini bertujuan mengembangkan sistem kontrol slide presentasi berbasis gestur tangan tanpa memerlukan perangkat fisik tambahan.

Pendekatan berbasis teknologi pengolahan citra dan gestur telah dieksplorasi dalam berbagai penelitian sebelumnya. Misalnya, penelitian oleh (Sruthi & Swetha, 2023) telah membuktikan efektivitas *MediaPipe* dalam mendeteksi *landmark* tangan secara *real-time*, sementara (Khairianto & Firdaus, 2024a) menunjukkan keunggulan *OpenCV* dalam pemrosesan citra digital untuk pengenalan pola gerakan. Namun, dalam konteks pengenalan gerakan, berbagai metode telah dikembangkan dengan tingkat keberhasilan yang berbeda.

Metode konvensional seperti *Convolutional Neural Network* (CNN) dan *Support Vector Machine* (SVM) memiliki keterbatasan dalam mengolah data bergerak. CNN cenderung efektif untuk citra statis, sementara SVM kurang optimal dalam mengklasifikasikan gerakan berurutan. Sebaliknya, penelitian mutakhir menunjukkan keunggulan model *Long Short-Term Memory* (LSTM). Studi yang dilakukan oleh (Nor et al., 2022) dan (Fakhrudin et al., 2019) secara konsisten membuktikan bahwa LSTM jauh lebih unggul dalam mengenali dan mengklasifikasikan pola gerakan kompleks, terutama dalam konteks data berurutan.

Meskipun demikian, celah penelitian masih terbuka lebar. Mayoritas studi sebelumnya hanya berfokus pada pengenalan gestur secara umum atau aplikasinya dalam domain seperti gaming dan kontrol perangkat mobile. Integrasi spesifik *MediaPipe*, *OpenCV*, dan LSTM untuk kontrol slide presentasi masih sangat terbatas. Padahal, potensi dampak teknologi ini sangat signifikan, terutama dalam meningkatkan kualitas presentasi di lingkungan pendidikan dan bisnis.

Dalam konteks pendidikan, sistem pengenalan gerakan tangan dapat memfasilitasi gaya mengajar yang lebih dinamis dan interaktif. Pendidik dapat fokus sepenuhnya pada penyampaian materi tanpa terbatas oleh perangkat kontrol fisik. Di lingkungan bisnis, teknologi ini dapat meningkatkan profesionalisme presentasi dengan memberikan kebebasan gerak yang lebih luas kepada presenter. Lebih dari itu, sistem ini berpotensi memberikan aksesibilitas yang lebih baik bagi individu dengan keterbatasan mobilitas yang mungkin mengalami kesulitan menggunakan perangkat kontrol konvensional.

Berdasarkan permasalahan dan potensi tersebut, penelitian ini bertujuan untuk mengimplementasikan sistem pengenalan gerakan tangan menggunakan *MediaPipe*, *OpenCV*, dan model LSTM untuk mengontrol slide presentasi secara *real-time*. Sistem yang diharapkan tidak sekadar menjadi solusi teknis, melainkan terobosan yang dapat meningkatkan kualitas komunikasi dan interaksi dalam presentasi, baik di lingkungan akademik maupun profesional.

## 2. KAJIAN TEORITIS

### Deteksi dan Pengenalan Gerakan Tangan (*Hand Gesture Recognition*)

Gestur Tangan (*Hand Gesture*) merupakan gerakan atau posisi tangan yang memiliki makna tertentu dan berfungsi sebagai bentuk komunikasi non-verbal manusia. Gestur ini dapat digunakan untuk menyampaikan pesan atau instruksi tanpa menggunakan kata-kata, melainkan melalui gerakan tangan yang dapat dikenali dan diinterpretasikan oleh sistem komputer. Gestur tangan mencakup berbagai bentuk ekspresi seperti gerakan tangan, posisi jari, maupun bentuk tangan yang spesifik, dan memiliki arti yang dapat berbeda tergantung pada budaya, konteks sosial, maupun situasi komunikasi (Khairianto & Firdaus, 2024b).

Selain sebagai alat komunikasi antar manusia, gestur tangan juga memainkan peran penting dalam pengenalan emosi, di mana gerakan dan posisi jari-jari dapat merefleksikan ekspresi emosional individu seperti kegembiraan, kesedihan, atau kemarahan. Hal ini memungkinkan adanya aplikasi teknologi yang mendukung pengenalan emosi secara lebih mendalam dan interaktif (Prananta et al., 2023).

Gestur tangan juga telah banyak diimplementasikan dalam berbagai aplikasi teknologi, seperti sistem pengenalan gestur untuk mengontrol perangkat elektronik, antarmuka *Virtual Reality* (VR), permainan interaktif, dan interaksi manusia–robot (*Human-Robot Interaction*). Dalam konteks ini, gestur tangan digunakan untuk memberikan instruksi atau perintah kepada robot melalui gerakan yang direkam dan dikenali oleh sistem, yang selanjutnya dapat digunakan dalam berbagai aplikasi seperti asisten pribadi, perangkat rumah pintar, maupun otomasi industri (Arif et al., 2024).

Dalam sistem berbasis visual, dua proses penting terkait pengolahan gestur tangan adalah deteksi tangan dan pelacakan tangan. Deteksi tangan merujuk pada proses identifikasi keberadaan tangan dalam citra atau video, sedangkan pelacakan tangan mencakup pemantauan pergerakan dan posisi tangan dari satu *frame* ke *frame* berikutnya secara berkelanjutan (Amanda Muchsin Chalik et al., 2021).

Pengenalan isyarat tangan (*Hand Gesture Recognition*) merujuk pada proses interaksi antara manusia dan sistem komputer melalui gerakan tangan. Konsep dasarnya adalah menciptakan bentuk interaksi yang alami dan sesuai dengan kebiasaan manusia, sehingga pengguna dapat mengontrol perangkat atau menyampaikan informasi dengan lebih intuitif dan efisien melalui gerakan yang dikenali oleh sistem (Arif et al., 2024).

### **Framework *MediaPipe***

*MediaPipe* adalah sebuah *framework* open-source yang dikembangkan oleh *Google Inc.* untuk mendukung pengembangan aplikasi multimodal, seperti pengolahan data video, gambar, audio, dan jenis data lainnya secara *real-time*. *Framework* ini dirancang khusus untuk memudahkan pengembang dalam membangun aplikasi di bidang *computer vision*, machine learning, dan image processing (Arif et al., 2024). (Heriansyah & Utomo, 2021) menunjukkan efisiensi pemrosesan citra penting di lingkungan *real-time*: *Scilab* unggul dalam konversi *RGB* ke *grayscale* dibanding *Octave* dan *FreeMat*, yang memperkuat pentingnya penggunaan *framework* cepat seperti *MediaPipe* + *OpenCV*.

*MediaPipe* mendukung berbagai platform, seperti Android, iOS, JavaScript, dan Python, sehingga sangat fleksibel dalam implementasinya. Instalasi *MediaPipe* pada Python dapat dilakukan dengan mudah melalui perintah `pip install MediaPipe` di terminal atau melalui fitur pengelolaan package otomatis pada IDE seperti *PyCharm* (Amanda Muchsin Chalik et al., 2021)

Salah satu keunggulan utama *MediaPipe* adalah tersedianya berbagai modul siap pakai (*pre-built solutions*) seperti *Face Mesh*, *Hand Tracking*, *Object Detection*, *Pose Estimation*,

dan lainnya. Modul-modul ini memungkinkan deteksi dan pelacakan fitur visual seperti wajah, tangan, dan pose tubuh secara cepat dan akurat (Amanda Muchsin Chalik et al., 2021).

Dalam proses pengolahan data, *MediaPipe* menggunakan algoritma *Convolutional Neural Networks* (CNN) untuk mengekstraksi dan mengenali pola dari data time-series seperti video dan audio. Dengan memanfaatkan kemampuan ini, *MediaPipe* dapat digunakan untuk menghasilkan kerangka (*skeleton*) dari gestur tangan, yang sangat berguna dalam aplikasi seperti pengenalan isyarat tangan, interaksi manusia-komputer, hingga kontrol berbasis gerakan (Arif et al., 2024).

### **OpenCV (*Open Source Computer vision*)**

*OpenCV (Open Source Computer vision Library)* adalah perpustakaan sumber terbuka yang dirancang untuk visi komputer dan pembelajaran mesin. Perpustakaan ini pertama kali dikembangkan oleh *Intel* pada tahun 1999 dan kini tersedia secara *open-source* di bawah lisensi BSD, sehingga memungkinkan bisnis dan pengembang untuk memanfaatkan serta memodifikasi kode secara bebas (Maulana et al., 2024).

Tujuan utama *OpenCV* adalah menyediakan infrastruktur umum bagi aplikasi visi komputer dan mempercepat adopsi teknologi persepsi mesin dalam produk komersial. *OpenCV* menawarkan lebih dari 2500 algoritma yang dioptimalkan, mencakup beragam metode klasik maupun mutakhir dalam bidang visi komputer dan pembelajaran mesin (Zebua & Rosyani, 2024).

OpenCV tersedia dalam berbagai bahasa pemrograman seperti C++, Python, Java, dan MATLAB. Hal ini membuat *OpenCV* mudah diintegrasikan ke berbagai *platform* dan digunakan dalam berbagai aplikasi, termasuk visi mesin, pengolahan citra medis, dan kontrol industri (Zebua & Rosyani, 2024).

### **Long Short-Term Memory (LSTM)**

Long Short-Term Memory (LSTM) adalah jenis jaringan saraf tiruan yang dirancang khusus untuk menangani dan memprediksi data yang bentuknya berurutan atau sekuensial. Keunggulan utama LSTM adalah kemampuannya dalam mengingat dan menyimpan informasi masa lampau serta mampu mempelajari suatu data yang bersifat sekuensial (Husna Moetia Putri et al., 2022). LSTM memiliki struktur dasar yang terdiri dari *input layer*, *hidden layer*, dan *output layer*.

Pada arsitektur LSTM terdapat dua fungsi aktivasi utama yang digunakan, yaitu sigmoid dan tanh. LSTM juga dilengkapi dengan *memory cell* dan gerbang (*gates*). Gerbang-gerbang

tersebut tersusun dari tiga komponen utama yaitu *forget gate*, *input gate*, dan *output gate* (Husna Moetia Putri et al., 2022).

Pendekatan awal sistem monitoring otomatis oleh (Wibowo & Heriansyah, 2014) menunjukkan bahwa deteksi gerak berbasis citra digital dapat diintegrasikan dengan sistem peringatan nyata, menegaskan bahwa *gesture-based control* bisa diterapkan secara efisien. Sistem deteksi bahasa isyarat berbasis LSTM yang dikombinasikan dengan penggunaan *MediaPipe* dapat mengenali gerakan tangan secara *real-time* dengan akurasi yang tinggi. Irviantina et al. (2024) menunjukkan bahwa model LSTM memiliki performa yang baik dalam mendeteksi bahasa isyarat berdasarkan abjad dengan akurasi mencapai 98%. Sementara itu, (Husna Moetia Putri et al., 2022) dalam penelitiannya menggunakan *bidirectional LSTM* untuk model 10 kelas dengan *epoch* 1000, *hidden layer* 64, dan *batch size* 32 mendapatkan akurasi sebesar 92%, serta 65% untuk model 30 kelas dengan 2 *layer LSTM epoch* 500, *hidden layer* 64, dan *batch size* 64.

Dalam implementasinya, model LSTM mampu menangkap pola temporal pada gerakan tangan yang mewakili huruf dan kata dalam bahasa isyarat, menjadikannya pilihan yang tepat untuk aplikasi pengenalan bahasa isyarat (Irviantina et al. 2024). Meskipun demikian, tantangan dalam penerapan LSTM untuk pengenalan bahasa isyarat masih ada, seperti variasi pencahayaan dan posisi tangan yang dapat memengaruhi akurasi pengenalan.

### 3. METODE PENELITIAN

Fokus utama penelitian adalah penerapan *MediaPipe* untuk deteksi landmark tangan secara *real-time*, *OpenCV* untuk pemrosesan *frame*, dan model LSTM untuk mengklasifikasikan urutan gerakan "*Next*", "*Previous*", dan "*Diam*" sebagai perintah slide. Penelitian ini bertujuan mengeksplorasi kemampuan model LSTM dalam mengenali pola gestur tangan serta mengintegrasikannya ke dalam aplikasi presentasi tanpa kontak perangkat fisik.

Penelitian ini mencakup serangkaian tahapan yang terstruktur mulai dari identifikasi permasalahan hingga evaluasi hasil. Pada tahap awal, dilakukan identifikasi masalah mengenai kendala presenter dalam mengoperasikan perangkat fisik, serta kajian literatur terkait gestur tangan, deteksi *landmark*, dan arsitektur LSTM. Selanjutnya, ditetapkan protokol pengumpulan dan anotasi data gerakan tangan, meliputi ekstraksi 21 titik *landmark* per *frame* dan pelabelan gerakan menggunakan *callback mouse*.

Setelah data terkumpul, dilaksanakan *preprocessing* untuk normalisasi, interpolasi *frame* hilang, dan *smoothing landmark*. Data kemudian diaugmentasi melalui teknik *time-warp*,

*jittering*, *flipping*, dan rotasi ringan untuk meningkatkan variasi. Pada tahap berikutnya, dataset dibagi menjadi data latih, validasi, dan uji agar model LSTM dapat dilatih secara optimal dan diuji akurasinya.

(Wibowo & Heriansyah, 2021) telah membuktikan bahwa pipeline image processing dengan background subtraction dan chain-code detection mampu mendeteksi obyek bergerak secara real-time, lalu mengeksekusi tindakan sistem melalui notifikasi SMS; ini menjadi relevan saat membangun sistem kontrol presentasi yang responsif dan adaptif. Proses pelatihan model meliputi desain arsitektur LSTM dua lapis, penentuan *hyperparameter*, dan validasi model. Setelah pelatihan, model diintegrasikan secara *real-time* ke *pipeline OpenCV*, di mana setiap *frame* diolah, diprediksi kelas gerakan, dan diterjemahkan menjadi perintah *keyboard* untuk mengendalikan slide.

Tahap akhir penelitian berfokus pada evaluasi kinerja model menggunakan metrik akurasi, *precision*, *recall*, *F1-score*, serta pengukuran latensi *end-to-end*. Hasil evaluasi menjadi dasar penarikan kesimpulan dan rekomendasi pengembangan lebih lanjut.

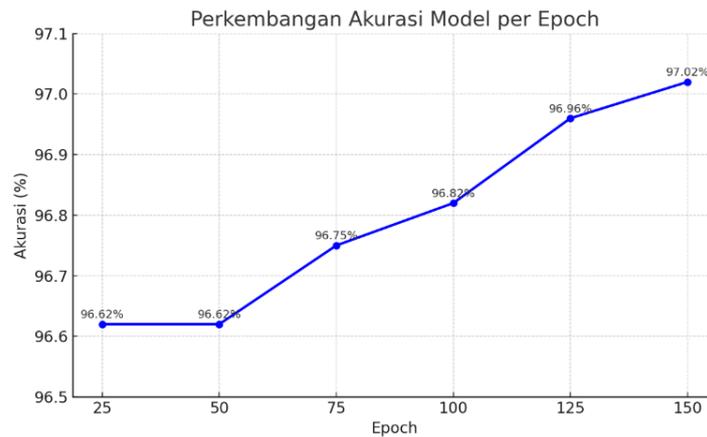
#### 4. HASIL DAN PEMBAHASAN

##### Hasil Penelitian

Penelitian ini menghasilkan sistem pengenalan gestur tangan berbasis *MediaPipe*, *OpenCV*, dan *Long Short-Term Memory* (LSTM) yang digunakan untuk mengontrol slide presentasi. Dataset diperoleh dari 20 partisipan dengan tiga kategori gerakan utama (*Next*, *Previous*, dan *Diam*), yang direkam menggunakan webcam beresolusi 640×480 pada 10 fps. Setiap *frame* diproses oleh *MediaPipe* untuk mengekstraksi 21 titik landmark tangan dalam koordinat tiga dimensi. Data yang gagal terdeteksi digantikan dengan nilai nol agar dimensi fitur tetap konsisten.

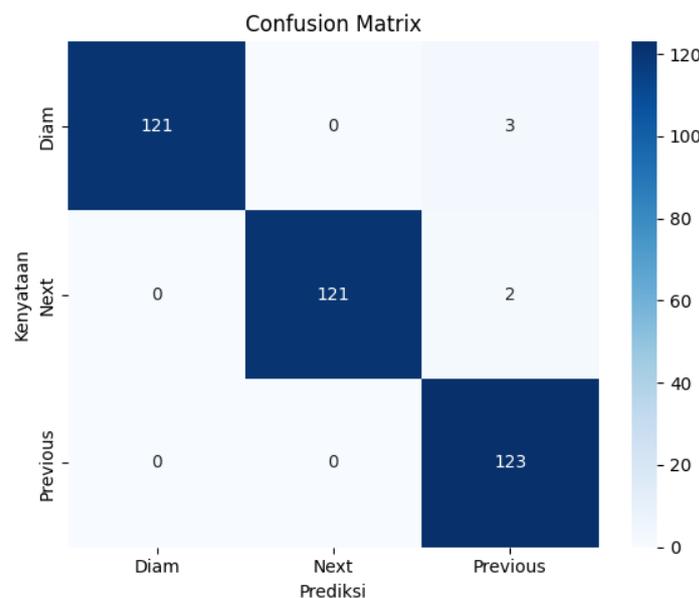
Tahap pemrosesan data meliputi pembentukan sekuens sepanjang 10 frame, penyesuaian dimensi ke dalam format tiga dimensi, serta konversi label dengan *one-hot encoding*. Dataset kemudian dibagi dengan rasio 80:20 untuk pelatihan dan pengujian. Model LSTM dibangun dengan dua lapisan, masing-masing berisi 64 unit, dan dilengkapi *dropout* sebesar 30% untuk mengurangi risiko *overfitting*.

Proses pelatihan menunjukkan peningkatan akurasi yang stabil hingga mencapai titik optimal pada *epoch* ke-50. Seperti ditunjukkan pada Gambar 1, akurasi pelatihan dan validasi meningkat konsisten, dengan perbedaan relatif kecil antara keduanya, yang menandakan model tidak mengalami *overfitting* secara signifikan.



**Gambar 1.** Grafik Akurasi Model LSTM per Epoch.

Evaluasi model dengan *confusion matrix* memperlihatkan performa klasifikasi yang sangat baik. Seperti ditunjukkan pada Gambar 2, gestur *Diam* memiliki tingkat pengenalan tertinggi dengan prediksi yang hampir sempurna. Sementara itu, sebagian kecil kesalahan klasifikasi ditemukan pada gestur *Next* dan *Previous*, yang sesekali tertukar akibat kemiripan pola pergerakan tangan.



**Gambar 2.** Confusion Matrix Model LSTM.

Hasil uji coba sistem secara real-time dengan *black-box testing* memperlihatkan bahwa jarak penggunaan optimal berada pada 1–2 meter dari kamera, dengan akurasi menurun pada jarak lebih dari 2 meter. Kondisi pencahayaan juga terbukti memengaruhi performa; pencahayaan yang buruk mengurangi kejelasan deteksi landmark dan meningkatkan tingkat kesalahan klasifikasi.

Setelah proses pelatihan dan integrasi model LSTM ke dalam sistem *real-time* selesai dilakukan, langkah selanjutnya adalah mengevaluasi performa sistem secara keseluruhan

menggunakan pendekatan *black-box testing*. Setiap pengujian didokumentasikan dalam bentuk tabel pengujian *black-box*, yang mencakup skenario uji, kondisi input, hasil yang diharapkan, hasil aktual, serta status keberhasilan pengujian. Pendekatan ini diharapkan mampu memberikan gambaran menyeluruh mengenai keandalan sistem dalam penggunaan nyata. Berikut Tabel 1. menyajikan hasil pengujian *black-box* pada integrasi sistem kontrol slide presentasi.

**Tabel 1.** Hasil Pengujian Black-box.

No	Skenario Uji	Deskripsi	Hasil Diharapkan	Validasi 1m	Validasi 2m	Validasi 3m
1	Deteksi gestur "Next"	Gerakan tangan sesuai gestur "Next"	Slide berpindah ke kanan	Sesuai	Sesuai	Sesuai
2	Deteksi gestur "Prev"	Gerakan tangan sesuai gestur "Prev"	Slide berpindah ke kiri	Sesuai	Sesuai	Sesuai
3	Deteksi gestur "Diam"	Tidak ada gerakan atau posisi tangan netral	Slide tetap	Sesuai	Sesuai	Sesuai
4	Tidak ada tangan dalam frame	Kamera menyala tapi tidak ada tangan	Sistem idle	Sesuai	Sesuai	Sesuai
5	Tangan tidak jelas	Gerakan tangan tidak lengkap / noise	Tidak ada prediksi	Sesuai	Sesuai	Sesuai
6	Respon cepat	Gestur dilakukan dalam waktu $\pm 1$ detik	Sistem merespon setelah gestur	Sesuai	Sesuai	Sesuai
7	Gestur cepat berulang	"Next" dilakukan dua kali cepat	Hanya 1 respon yang dikirim	Sesuai	Sesuai	Sesuai
8	Gestur berubah di tengah	Gestur berubah dari "Next" ke "Diam" dengan cepat	Sistem hanya merespon yang valid	Sesuai	Sesuai	Sesuai
9	Lama penggunaan	Sistem dijalankan selama 10–30 menit	Tidak crash, tetap responsif	Sesuai	Sesuai	Sesuai
10	Resolusi kamera diubah	Uji resolusi di 480p dan 720p	Sistem tetap stabil	Sesuai	Sesuai	Sesuai
11	Gestur mirip	Gerakan menyerupai "Next" tapi tidak valid	Tidak ada aksi	Sesuai	Sesuai	Sesuai

No	Skenario Uji	Deskripsi	Hasil Diharapkan	Validasi 1m	Validasi 2m	Validasi 3m
12	Dua tangan dalam frame	Tangan kanan dan kiri muncul bersamaan	Sistem hanya mengenali satu tangan	Sesuai	Sesuai	Sesuai
13	Kamera tertutup sebagian	Kamera terhalang sebagian objek	Sistem idle	Sesuai	Sesuai	Sesuai
14	Presentasi sambil bergerak	Posisi badan bergeser saat presentasi	Sistem tetap mengenali gestur	Sesuai	Sesuai	Sesuai
15	Menjelaskan tanpa gestur	Tangan diam/turun saat berbicara	Slide tidak berpindah	Sesuai	Sesuai	Sesuai
16	Pencahayaan terang	Gerakan dilakukan di ruangan terang	Slide berpindah	Sesuai	Sesuai	Sesuai
17	Pencahayaan redup	Gerakan dilakukan di ruangan agak redup	Slide berpindah	Sesuai	Sesuai	Tidak Sesuai
18	Pencahayaan gelap	Gerakan dilakukan di ruangan gelap	Slide berpindah	Tidak Sesuai	Tidak Sesuai	Tidak Sesuai

Berdasarkan Tabel 1, pengujian *black-box* terhadap sistem pengenalan gerakan tangan yang terintegrasi dengan kontrol slide presentasi menunjukkan bahwa sistem secara umum telah berfungsi dengan baik dalam berbagai kondisi, meskipun terdapat keterbatasan pada kondisi pencahayaan gelap. Pengujian dilakukan melalui 18 skenario, mencakup aspek fungsionalitas gestur utama, respons sistem, ketahanan, serta interaksi pengguna.

## Pembahasan

Pembahasan difokuskan pada efektivitas pendekatan yang digunakan dalam mengenali tiga jenis gestur, yaitu “*Next*”, “*Previous*”, dan “*Diam*”. Tahap awal dimulai dari proses akuisisi data menggunakan *webcam* yang merekam pergerakan tangan dalam format video dengan resolusi 640×480 piksel dan kecepatan 10 *frame* per detik. Data yang diambil kemudian diproses menggunakan *MediaPipe Hands* untuk mengekstraksi 21 titik koordinat tiga dimensi dari tangan (x, y, z). Untuk membentuk data sekuensial, digunakan struktur *buffer deque* sepanjang 10 *frame* yang memastikan model memperoleh konteks temporal dari gestur.

Sebanyak 1.848 data berhasil dikumpulkan dan diberi label secara manual, lalu diproses menjadi input model melalui tahap pembuatan *sekuens*, *reshaping*, konversi label ke *one-hot encoding*, dan pembagian ke dalam data latih dan uji (80:20). Model LSTM kemudian dilatih

menggunakan teknik *grid search* dengan 18 kombinasi parameter berbeda (*units*, *dropout*, dan *batch size*) pada enam nilai *epoch* (25, 50, 75, 100, 125, dan 150). Proses ini menghasilkan data yang sangat komprehensif untuk mengevaluasi performa dan stabilitas pelatihan model.

Evaluasi menunjukkan bahwa dua kandidat model terbaik berasal dari *epoch* 25 dan *epoch* 50, keduanya memiliki akurasi uji yang sama tinggi yaitu 98.65% serta nilai *precision*, *recall*, dan *F1-score* yang identik (98.70%, 98.65%, dan 98.65%). Namun, setelah dilakukan analisis lebih dalam terhadap stabilitas pelatihan dari masing-masing *epoch*, diperoleh fakta bahwa model dengan *epoch* 50 lebih unggul secara konsistensi karena hanya terdapat 3 dari 18 percobaan yang memiliki akurasi latih di bawah 96%, dibandingkan dengan *epoch* 25 yang memiliki 8 percobaan di bawah ambang tersebut. Hal ini menunjukkan bahwa *epoch* 50 tidak hanya unggul dalam hasil akhir, tetapi juga lebih stabil pada berbagai konfigurasi parameter. Karena itu, *epoch* 50 ditetapkan sebagai model terbaik secara defensif dan berbasis data konkret.

Model terbaik dari *epoch* 50 kemudian dievaluasi lebih lanjut menggunakan *confusion matrix* dan metrik klasifikasi lainnya. Hasil menunjukkan bahwa semua kelas dikenali dengan akurat, dengan akurasi keseluruhan mencapai 99%. Kesalahan klasifikasi minor hanya terjadi pada gestur “*Diam*” dan “*Next*” yang kadang tertukar dengan “*Previous*”, tetapi secara umum, *precision* dan *recall* tetap sangat tinggi.

Juliansyah et al., (2025) melaporkan skor usability 84,25% dan kualitas informasi 84,45% pada evaluasi portal mahasiswa baru UIGM, menegaskan bahwa sistem digital yang efektif bukan hanya responsif secara teknis, tetapi juga evaluatif terhadap pengalaman pengguna. Sistem ini kemudian diintegrasikan secara *real-time* dengan menggunakan *OpenCV* untuk menangkap video secara langsung dan *MediaPipe* sebagai pelacak *landmark* tangan. LSTM digunakan untuk inferensi klasifikasi gestur yang dikaitkan dengan perintah *keyboard* menggunakan pustaka *pyautogui*, seperti “*Next*” untuk slide maju, “*Previous*” untuk mundur, dan “*Diam*” untuk tetap. Sistem juga dilengkapi mekanisme *cooldown* selama 1 detik dan *frame skip* untuk menghindari perintah ganda serta menjaga performa *respons*. (Puspasari et al., 2021) berhasil membangun aplikasi AR edukatif dengan waktu respons deteksi marker sebesar 0,599 detik dan skor usability SUS 78,3 menegaskan bahwa interaksi visual real-time dapat dioptimalkan tanpa mengorbankan pengalaman pengguna.

Uji coba sistem secara *black-box* menunjukkan bahwa sistem dapat berjalan dengan baik dalam berbagai kondisi normal. Namun, terdapat beberapa kelemahan dalam mendeteksi gestur saat dua tangan muncul bersamaan, atau saat gerakan tangan tidak stabil, serta adanya potensi *error* saat tangan bergerak cepat atau pengguna berbicara sambil menggunakan tangan. Sejalan

Studi oleh Gustriansyah et al., (2025), termasuk Suhandi, memperlihatkan efektivitas model *tree-based* dan optimasi *hyperparameter* dalam memprediksi kinerja manusia ini menambah bobot teknikal dalam validasi algoritma klasifikasi gestur Anda. Ini menjadi catatan penting untuk pengembangan di masa depan agar sistem lebih adaptif terhadap kondisi dunia nyata.

Secara keseluruhan, sistem yang dikembangkan berhasil mengimplementasikan pengenalan gestur tangan untuk kontrol slide presentasi secara *real-time* dengan akurasi dan stabilitas tinggi. Model LSTM menunjukkan performa baik dalam mengenali pola temporal, dan pendekatan berbasis *MediaPipe* sangat efisien dalam ekstraksi fitur tangan. Hasil ini menjadi fondasi kuat untuk pengembangan sistem pengenalan gestur yang lebih canggih dan kontekstual di masa mendatang.

## 5. KESIMPULAN DAN SARAN

Berdasarkan hasil penelitian, implementasi, dan pengujian yang telah dilakukan, dapat disimpulkan bahwa penelitian ini berhasil mengembangkan sistem kontrol slide presentasi berbasis pengenalan gestur tangan menggunakan *MediaPipe*, *OpenCV*, dan *Long Short-Term Memory* (LSTM), yang mampu mengenali tiga jenis gestur utama, yaitu “*Next*”, “*Previous*”, dan “*Diam*”. Model terbaik diperoleh pada *epoch* ke-50 dengan akurasi uji sebesar 98.65% serta *precision*, *recall*, dan *F1-score* rata-rata sebesar 0.99, yang menunjukkan performa klasifikasi yang seimbang antar kelas. Hasil pengujian *black-box* pada jarak 1 m, 2 m, dan 3 m menunjukkan bahwa jarak optimal untuk kinerja terbaik sistem berada pada rentang 1–2 meter. Pada jarak ini, sistem mampu bekerja dengan stabil, cepat, dan akurat pada hampir semua skenario, termasuk gerakan cepat, perubahan gestur di tengah, penggunaan berulang, serta perubahan resolusi kamera. Sementara itu, pada jarak 3 meter, sistem masih dapat berfungsi, tetapi performanya cenderung menurun terutama ketika dipadukan dengan kondisi pencahayaan rendah. Selain itu, sistem dapat mengenali gestur dengan baik pada pencahayaan terang dan redup, namun mengalami penurunan akurasi yang signifikan pada pencahayaan gelap di semua jarak. Secara keseluruhan, sistem telah berhasil diimplementasikan secara *real-time* dan menunjukkan kinerja yang konsisten dalam mendukung aktivitas presentasi berbasis kontrol gestur tangan.

**DAFTAR REFERENSI**

- Amanda Muchsin Chalik, Bilal Abdul Qowy, Faiz Hanafi, & Ahlijati Nuraminah. (2021). Mouse tracking tangan dengan klasifikasi gestur menggunakan OpenCV dan MediaPipe. *Jurnal Ilmiah Teknik Informatika Dan Komunikasi*, 1(2), 10-18. <https://doi.org/10.55606/juitik.v1i2.323>
- Arif, M., Haryono, G. S., Arsyad, N. F., Ramadhani, R., Sahid, A., Rosyani, P., Kunci, K., Tangan, P., Gerakan, P., & Manusia-Komputer, I. (2024). Teknik dan multimedia sistem pendeteksi tangan berbasis MediaPipe dan OpenCV untuk pengenalan gerakan. *Biner: Jurnal Ilmu Komputer*, 2(2), 173-177.
- Asmara, G. I., Erdina, N., & Ariza, R. A. (2021). Urgensi pertemuan dan presentasi dalam organisasi bisnis. *Da'watuna: Journal of Communication and Islamic Broadcasting*, 1(2), 109-119. <https://doi.org/10.47467/dawatuna.v1i2.487>
- Fakhrudin, M. A., Pratikno, H., & Kusumawati, W. I. (2019). Kontrol level kecepatan kipas melalui deteksi gestur jari tangan menggunakan MediaPipe dan Faster-RCNN. *Jurnal Teknologi Informasi Dan Ilmu Komputer*. <https://doi.org/10.25126/jtiik.2023107345>
- Gustriansyah, R., Puspasari, S., Sanmorino, A., Suhandi, N., & Sartika, D. (2025). Tree-based models and hyperparameter optimization for assessing employee performance. *Indonesian Journal of Electrical Engineering and Computer Science*, 38(1), 569-577. <https://doi.org/10.11591/ijeecs.v38.i1.pp569-577>
- Heriansyah, R., & Utomo, W. M. (2021). Performance evaluation of digital image processing by using Scilab. *JUITA: Jurnal Informatika*, 9(2), 239-247. <https://doi.org/10.30595/juita.v9i2.8434>
- Husna Moetia Putri, Fadlisyah, & Wahyu Fuadi. (2022). Pendeteksian bahasa isyarat Indonesia secara real-time menggunakan Long Short-Term Memory (LSTM).
- Irviantina, S., Wijaya, D. A., Situmorang, D. R., & Nasution, N. M. J. (2024). Deteksi bahasa isyarat berdasarkan abjad menggunakan metode LSTM (Long Short Term Memory). *Majalah Ilmiah METHODODA*, 14(3), 371-376. <https://doi.org/10.46880/methoda.Vol14No3.pp371-376>
- Juliansyah, M. A., Terttiaavini, Heryati, A., & Sanmorino, A. (2025). Evaluasi kinerja portal new student UIGM berdasarkan perspektif mahasiswa menggunakan metode use questionnaire dan importance performance analysis (IPA). *Jurnal Ilmiah Informatika Global*, 16(2), 98-107. <https://doi.org/10.36982/jiig.v16i2.5416>
- Khairianto, D., & Firdaus, R. (2024a). Penerapan hand gesture recognition sebagai media kontrol presentasi aplikasi PowerPoint. *JATI: Jurnal Mahasiswa Teknik Informatika*, 8(2), 1852-1860. <https://doi.org/10.36040/jati.v8i2.9167>
- Khairianto, D., & Firdaus, R. (2024b). Penerapan hand gesture recognition sebagai media kontrol presentasi aplikasi Powerpoint. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 8(2), 1852-1860. <https://doi.org/10.36040/jati.v8i2.9167>
- Maulana, A., Auliatunnajah, F., Rosidin, N., Ramadien Rizki Darmawan, M., & Rosyani, P. (2024). Implementasi OpenCV dengan metode image thresholding pada gambar. *Jurnal Artificial Inteligent Dan Sistem Penunjang Keputusan*, 2(1), 27-32.
- Nor, S., Muslim, M. A., & Aswin, M. (2022). Pengenalan pola dasar angka berdasarkan gerakan tangan menggunakan machine learning. *Elkomika*, 10(3), 596-608. <https://doi.org/10.26760/elkomika.v10i3.596>

- Prananta, G. B., Azzikri, H. A., & Rozikin, C. (2023). Deteksi dan pengenalan gesture tangan secara real-time menggunakan jaringan saraf tiruan konvolusional. *Methodika*, 9(2), 30-34. <https://doi.org/10.46880/mtk.v9i2.1911>
- Puspasari, S., Suhandi, N., & Iman, J. N. (2021). Evaluation of augmented reality application development for cultural artefact education. *International Journal of Computing*, 20(2). <https://doi.org/10.47839/ijc.20.2.2171>
- Sruthi, S., & Swetha, S. (2023). Hand gesture controlled presentation using OpenCV and MediaPipe. *International Journal of Engineering Technology & Management Sciences*, 7(4), 338-341. <https://doi.org/10.46647/ijetms.2023.v07i04.046>
- Wibowo, A. D. A., & Heriansyah, R. (2014). Automated vehicle monitoring system. *1st International Conference on Computer Science and Engineering (ICON-CSE 2014)*.
- Wibowo, A. D. A., & Heriansyah, R. (2021). Real-time vehicle surveillance system based on image processing and short message service. *JUITA: Jurnal Informatika*, 9(2), 249-257. <https://doi.org/10.30595/juita.v9i2.8728>
- Zebua, E. T. P., & Rosyani, P. (2024). Perancangan deteksi objek kendaraan bermotor berbasis OpenCV Python menggunakan metode HOG-SVM untuk analisis lalu lintas cerdas. *Jurnal AI Dan SPK: Jurnal Artificial Intelligent Dan Sistem Penunjang Keputusan*, 2(1), 16-26.