TIN: Terapan Informatika Nusantara
Vol 7, No 2, July 2026, page 1168-1181
ISSN 2722-7987 (Media Online)
Website https://ejurnal.seminar-id.com/index.php/tin
DOI 10.47065/tin.v7i2.10711

Evaluasi Arsitektur Long Short-Term Memory untuk Klasifikasi Gestur Tangan Dinamis Berbasis MediaPipe Sebagai Kendali Presentasi

Putu Gede Dimas Witjaksana*, I Nyoman Saputra Wahyu Wijaya, Putu Hendra Suputra

Fakultas Teknik dan Kejuruan, Program Studi Ilmu Komputer, Universitas Pendidikan Ganesha, Singaraja, Indonesia
Email: [1,*]dimaswitjaksana068@gmail.com, [2]wahyu.wijaya@undiksha.ac.id, [3]hendra.suputra@undiksha.ac.id
Email Penulis Korespondensi: dimaswitjaksana068@gmail.com

Abstrak–Perkembangan *Human-Computer Interaction* (HCI) mendorong penggunaan metode interaksi yang lebih alami melalui teknologi pengenalan gestur tangan. Gestur dinamis memiliki keunggulan dibandingkan gestur statis karena mengandung informasi temporal berupa arah dan pola gerakan yang lebih representatif dalam menyampaikan suatu perintah. Namun, pengembangan model pengenalan gestur dinamis masih menghadapi tantangan dalam mempertahankan kemampuan generalisasi terhadap pengguna baru, sehingga diperlukan skema evaluasi yang mampu mengukur performa model secara lebih representatif. Penelitian ini bertujuan mengembangkan model klasifikasi gestur tangan dinamis untuk kendali presentasi PowerPoint berbasis MediaPipe Hands, MediaPipe Pose, dan *Long Short-Term Memory* (LSTM). Tahapan penelitian meliputi pengumpulan data video, ekstraksi landmark tangan dan bahu, normalisasi data menggunakan titik tengah kedua bahu sebagai titik acuan, pelatihan delapan variasi arsitektur LSTM, serta evaluasi menggunakan skema *Leave-One-Subject-Out Cross-Validation* (LOSO-CV), sehingga setiap partisipan secara bergantian menjadi data uji untuk mengevaluasi kemampuan generalisasi model terhadap pengguna baru. Hasil pengujian menunjukkan bahwa arsitektur Baseline 1 (B1) memberikan performa terbaik dengan rata-rata akurasi 95,77%, presisi 96,03%, *recall* 95,90%, dan *F1-score* 95,84%. Analisis *confusion matrix* menunjukkan bahwa sebagian besar gestur berhasil dikenali dengan benar, sedangkan kesalahan klasifikasi terutama terjadi pada kelas *idle* serta beberapa gestur yang memiliki kemiripan pose tangan. Hasil penelitian menunjukkan bahwa kombinasi MediaPipe dan LSTM mampu membangun model klasifikasi gestur tangan dinamis yang mempertahankan kemampuan klasifikasi secara konsisten pada pengujian lintas subjek.

Kata Kunci: MediaPipe; Long Short Term Memory; Gestur Tangan Dinamis; Kontrol Presentasi; Leave One Subject Out Cross Validation

Abstract–Advances in Human-Computer Interaction (HCI) are driving the use of more natural interaction methods through hand gesture recognition technology. Dynamic gestures have an advantage over static gestures because they contain temporal information such as direction and movement patterns hat is more representative in conveying a command. However, the development of dynamic gesture recognition models still faces challenges in maintaining generalization capabilities for new users, necessitating an evaluation scheme capable of measuring model performance more representatively. This study aims to develop a dynamic hand gesture classification model for controlling PowerPoint presentations based on MediaPipe Hands, MediaPipe Pose, and Long Short-Term Memory (LSTM). The research stages included video data collection, hand and shoulder landmark extraction, data normalization using the midpoint of both shoulders as a reference point, training of eight variations of the LSTM architecture, and evaluation using the Leave-One-Subject-Out Cross-Validation (LOSO-CV) scheme, such that each participant took turns serving as test data to evaluate the model's generalization ability toward new users. Test results show that the Baseline 1 (B1) architecture delivers the best performance with an average accuracy of 95.77%, precision of 96.03%, recall of 95.90%, and an F1-score of 95.84%, Analysis of the confusion matrix shows that most gestures were correctly recognized, while misclassifications occurred primarily in the "idle" class and for some gestures with similar hand poses. The results of the study indicate that the combination of MediaPipe and LSTM is capable of building a dynamic hand gesture classification model that maintains consistent classification performance in cross-subject testing.

Keywords: MediaPipe; Long Short Term Memory; Dynamic Hand Gesture; Presentation Control; Leave One Subject Out Cross Validation

1. PENDAHULUAN

Perkembangan teknologi dalam Human-Computer Interaction (HCI) dan Artificial Intelligence (AI) mendorong berbagai metode interaksi antarmanusia dan komputer yang lebih alami. Interaksi yang sebelumnya didominasi oleh penggunaan perangkat fisik seperti keyboard dan mouse kini berkembang menuju Natural User Interface (NUI) menggunakan suara, gerakan tubuh, maupun tangan (Zholshiyeva dkk., 2021). Dalam HCI, penggunaan perangkat fisik memiliki keterbatasan karena harus menggunakan perangkat tambahan dan kurangnya fleksibilitas karena harus berada dekat dengan perangkat. Salah satu pendekatan yang banyak digunakan untuk mengatasi keterbatasan ini adalah menggunakan teknologi computer vision. Penggunaan pendekatan ini memungkinkan sistem untuk mengenali perintah dengan adanya gerakan atau pose yang ditangkap melalui kamera tanpa adanya perangkat tambahan seperti sensor (Yasen & Jusoh, 2019). Pendekatan ini juga memungkinkan peningkatan kenyamanan dan efisiensi interaksi dalam melakukan presentasi. Penggunaan computer vision dinilai lebih praktis dan dapat memberikan interaksi yang lebih natural dibandingkan dengan pendekatan yang memerlukan perangkat tambahan (F. Zhang dkk., 2020).

Salah satu aktivitas yang masih didominasi oleh penggunaan perangkat fisik adalah kendali presentasi menggunakan aplikasi seperti Microsoft PowerPoint atau Canva. Pada umumnya presenter melakukan kendali prsentasi harus berada di dekat komputer atau menggunakan perangkat tambahan seperti remote untuk melakukan perpindahan slide. Kondisi ini membatasi mobilitas dan mengurangi naturalitas interaksi selama penyampaian materi. Sehingga, perlu adanya metode interaksi yang memungkinkan pengendalian presentasi secara hands-free. Salah satu solusi yang

TIN: Terapan Informatika Nusantara
Vol 7, No 2, July 2026, page 1168-1181
ISSN 2722-7987 (Media Online)
Website https://ejurnal.seminar-id.com/index.php/tin
DOI 10.47065/tin.v7i2.10711

dapat mendukung interaksi ini adalah Hand Gesture Recognition (HGR). Gestur tangan dapat dideteksi oleh komputer sehingga komputer dapat membaca gestur tertentu sebagai sebuah perintah (SHI dkk., 2021).

Berdasarkan karakteristiknya, terdapat dua jenis gestur tangan, yaitu gestur statis dan gestur dinamis (Herbert dkk., 2024). Gestur statis merupakan gestur yang merepresentasikan bentuk tangan pada satu waktu tertentu tanpa ada pergerakan. Sedangkan gestur dinamis melibatkan perubahan posisi dan pergerakan tangan dalam kurun waktu tertentu (Rahman dkk., 2025). Dalam kasus kendali presentasi, gestur tangan dinamis memiliki keunggulan dibandingkan dengan gestur statis. Hal ini karena gestur dinamis dapat merepresentasikan sebuah perintah dengan lebih baik karena adanya informasi temporal berupa arah, kecepatan, dan pola gerakan yang memberikan makna lebih jelas pada setiap perintahnya (Rehman dkk., 2022). Sebagai contoh, perintah perpindahan slide dapat direpresentasikan melalui gerakan menyapukan tangan ke kiri atau ke kanan. Atau perintah Zoom dapat direpresentasikan dengan perubahan bentuk atau gerakan tangan. Informasi gerakan yang lebih menggambarkan perintah ini sulit direpresentasikan jika hanya menggunakan gestur statis.

Penerapan HGR dengan computer vision semakin berkembang seiring dengan adanya framework MediaPipe yang mampu mendeteksi dan melacak landmark tangan secara real-time (Prayoga dkk., 2025; Uddin dkk., 2025). Beberapa penelitian telah menggunakan MediaPipe untuk membangun sistem kendali berbasis gestur tangan. Penelitian yang dilakukan oleh Sugiantari Sugiantari dkk. (2025), serta penelitian Y. Zhang & Notni (2025) menunjukkan bahwa MediaPipe mampu mendeteksi pose tangan secara akurat untuk melakukan klasifikasi pose dan gerakan. Dalam pengendalian presentasi penerapan HGR dengan gestur statis seperti pada penelitian Agustiani dkk. (2024), serta penelitian Setyowati & Granito (2025) menunjukkan bahwa MediaPipe mampu melakukan deteksi pose tangan untuk kendali presentasi. Namun, kedua penelitian tersebut masih berfokus pada penggunaan gestur statis dalam interaksinya. Keterbatasan penggunaan gestur statis muncul ketika sistem perlu mengenali banyak perintah yang memiliki kemiripan bentuk tangan. Pada sistem multiperintah, perbedaan antarkelas gestur dapat memiliki kemiripan yang tinggi sehingga berpotensi meningkatkan kesalahan klasifikasi. Selain itu, beberapa perintah secara alami lebih mudah direpresentasikan dalam bentuk gerakan daripada pose tunggal (Hax dkk., 2024).

Untuk dapat mengenali pola temporal, diperlukan model yang mampu mempelajari data dalam rentang tertentu, seperti data frame yang berurutan atau data time series yang merekam data dalam kurun waktu tertentu (Dewi dkk., 2022). Salah satu metode yang banyak digunakan adalah Long Short-Term Memory (LSTM), yaitu arsitektur deep learning yang dapat memproses data sekuensial dan mempertahankan informasi dalam jangka waktu tertentu (Saputri dkk., 2024). Dalam kombinasi MediaPipe dan LSTM terdapat beberapa penelitian yang menunjukkan hasil yang baik, seperti penelitian yang dilakukan oleh Artha dkk. (2025) yang mencapai akurasi 92,35% pada klasifikasi gerakan push-up. Selain itu, terdapat penelitian dari Ho & Santoso (2025) dengan akurasi 93,94% pada pengenalan bahasa isyarat, serta penelitian dari Putra dkk. (2025) yang berhasil memantau perubahan landmark sendi untuk mendeteksi kondisi jatuh pada lansia dengan akurasi 91%. Hasil tersebut menunjukkan bahwa MediaPipe efektif dalam melakukan ekstraksi fitur pada tubuh manusia, serta LSTM mampu dengan baik mempelajari data hasil ekstraksi pada data temporal. Meskipun arsitektur berbasis Transformer mulai banyak diterapkan pada berbagai tugas pengenalan gerakan, pemilihannya umumnya ditujukan pada dataset berskala besar dengan representasi sekuens yang lebih kompleks. Pada penelitian ini, data masukan berupa landmark MediaPipe memiliki dimensi fitur yang relatif rendah serta panjang sekuens yang terbatas, sehingga LSTM dipilih karena telah banyak digunakan pada pengenalan gestur berbasis landmark dan mampu memodelkan ketergantungan temporal dengan kompleksitas model yang lebih rendah. Meskipun berbagai penelitian menunjukkan bahwa kombinasi MediaPipe dan LSTM mampu memberikan performa yang baik pada berbagai tugas pengenalan gestur, sebagian besar penelitian tersebut masih berfokus pada aplikasi seperti bahasa isyarat, aktivitas olahraga, atau pemantauan kondisi medis. Penerapan kombinasi MediaPipe dan LSTM untuk kendali presentasi berbasis gestur dinamis masih relatif terbatas, khususnya yang mempertimbangkan normalisasi landmark terhadap posisi tubuh pengguna untuk mengurangi pengaruh variasi posisi pengguna di depan kamera (Kim dkk., 2020; Purbojo & Wijaya, 2025).

Berdasarkan permasalahan tersebut, penelitian ini mengembangkan sistem kendali presentasi PowerPoint berbasis gestur tangan dinamis menggunakan MediaPipe Hands dan MediaPipe Pose untuk ekstraksi fitur serta LSTM untuk klasifikasi gestur. Selain menerapkan normalisasi landmark menggunakan titik tengah kedua bahu untuk meningkatkan konsistensi representasi data, penelitian ini juga mengevaluasi delapan variasi arsitektur LSTM guna mengidentifikasi konfigurasi yang memberikan performa terbaik. Selain pengembangan model klasifikasi, evaluasi kemampuan generalisasi model juga menjadi perhatian dalam pengenalan gestur. Oleh karena itu, penelitian ini menggunakan skema Leave-One-Subject-Out Cross-Validation (LOSO-CV), sehingga setiap partisipan secara bergantian menjadi data uji untuk mengevaluasi kemampuan model pada pengguna yang tidak terlibat dalam proses pelatihan.

2. METODOLOGI PENELITIAN

2.1 Objek Penelitian

Objek penelitian ini adalah gestur tangan dinamis yang digunakan sebagai kendali presentasi PowerPoint. Sistem mengklasifikasikan enam gestur kendali, yaitu next, previous, zoom in, zoom out, fullscreen, dan cancel, serta satu kelas idle sebagai kondisi tanpa perintah. Setiap gestur dirancang untuk merepresentasikan aksi tertentu pada aplikasi

TIN: Terapan Informatika Nusantara
Vol 7, No 2, July 2026, page 1168-1181
ISSN 2722-7987 (Media Online)
Website https://ejurnal.seminar-id.com/index.php/tin
DOI 10.47065/tin.v7i2.10711

presentasi berdasarkan perubahan pose dan arah gerakan tangan. Dataset diperoleh melalui proses perekaman video dari empat partisipan menggunakan kamera dengan kondisi pengambilan data yang terkontrol. Selanjutnya setiap video diproses menjadi data landmark menggunakan MediaPipe sehingga menghasilkan representasi koordinat tangan dan bahu yang digunakan sebagai masukan model klasifikasi

2.2 MediaPipe

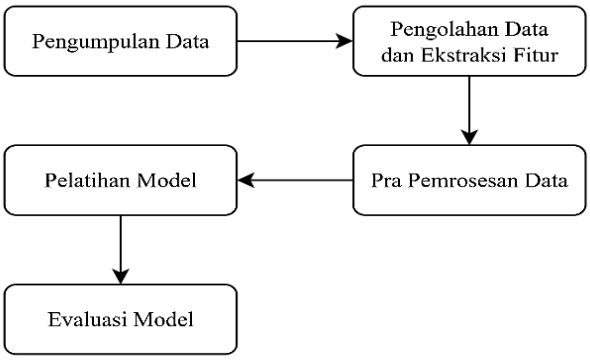
MediaPipe merupakan sebuah *library* yang digunakan untuk mengekstraksi landmark tangan dan bahu dari setiap frame video. Landmark tangan terdiri atas 21 titik dengan koordinat tiga dimensi (x, y, z), sedangkan MediaPipe Pose digunakan untuk memperoleh landmark bahu kiri dan kanan sebagai referensi normalisasi posisi. Pada penelitian ini digunakan 23 landmark sehingga diperoleh 69 fitur koordinat sebagai masukan model**.**

2.3 Long Short Term Memory

Long Short-Term Memory (LSTM) merupakan pengembangan dari Recurrent Neural Network (RNN) yang dirancang untuk mempelajari ketergantungan jangka panjang pada data sekuensial. LSTM menggunakan mekanisme *input gate*, *forget gate*, dan *output gate* untuk mengatur informasi yang disimpan maupun dihapus selama proses pembelajaran. Karakteristik tersebut membuat LSTM sesuai untuk mengklasifikasikan gestur dinamis karena mampu memanfaatkan urutan perubahan landmark tangan pada setiap frame**.**

2.4 Alur Penelitian

Pengembangan model LSTM untuk klasifikasi gestur tangan dinamis pada penelitian ini dilakukan melalui beberapa tahapan yang saling berkesinambungan, yaitu pengumpulan data, pengolahan data video dan ekstraksi fitur menggunakan MediaPipe, pemrosesan data, serta pelatihan dan evaluasi model. Setiap tahapan dirancang untuk menghasilkan data sekuensial yang siap digunakan pada proses pembelajaran model serta mengevaluasi kemampuan generalisasinya. Gambar 1 memperlihatkan alur penelitian secara keseluruhan, mulai dari proses pengumpulan dataset hingga evaluasi model.



Gambar 1. Alur Penelitian

2.4.1 Data Collection

Pengumpulan data citra dilakukan dengan perekaman video menggunakan webcam yang memiliki resolusi minimal 1080p dan 30 fps. Perekaman data dilakukan di dalam ruangan dengan pencahayaan yang terkontrol dan latar belakang dengan warna solid untuk mempermudah MediaPipe dalam melakukan deteksi landmark. Penelitian melibatkan empat partisipan sebagai peraga, dengan dua partisipan laki-laki dan dua partisipan perempuan, serta selama perekaman, partisipan menggunakan tangan kanan untuk memperagakan gestur.

Hasil pengumpulan data telah diperoleh. Secara keseluruhan, data dapat dikelompokkan menjadi 7 kelas gestur, yaitu gestur *next, previous, zoom in, zoom out, cancel*, dan fullscreen. Gestur *next* dan *previous* merupakan gestur berkebalikan, yaitu *next*: tangan bergerak secara horizontal dengan kelima jari terbuka dan meningkatkan nilai x. Sedangkan untuk *previous*, tangan bergerak dan koordinat x berkurang, kebalikan dari gerakan *next*. Kemudian, gestur *zoom in* diawali dengan gestur "OK" dengan jari telunjuk dan ibu jari bertemu, lalu kedua jari menjauh hingga membentuk gestur lima jari terbuka lebar. Gestur *zoom out* sendiri adalah kebalikan dari *zoom in*, diawali dengan gestur tangan terbuka, kemudian menjadi gestur "OK". Selanjutnya, gestur *fullscreen* diawali dengan tangan mengepal, kemudian dibuka lebar. Kelas gestur cancel diawali dengan tangan terbuka dengan menujukan telapak tangan ke kamera, kemudian bergerak ke bawah dengan perubahan orientasi ke punggung tangan sehingga jari-jari tangan mengarah ke bawah. Serta yang terakhir, kelas idle atau non-gestur memiliki empat variasi pose statis dari semua kelas seperti tangan terbuka lebar, tangan terbuka namun kelima jari menempel, tangan membentuk pose OK, tangan mengepal, dan tangan dengan kelima jari rapat yang menghadap ke bawah.

2.4.2 Pengolahan Data dan Ekstraksi Fitur

Setelah data terkumpul, langkah selanjutnya adalah melakukan pengolahan data video menggunakan software Adobe Premiere Pro. Pemrosesan yang dilakukan adalah memotong durasi video sehingga video yang digunakan berfokus

TIN: Terapan Informatika Nusantara
Vol 7, No 2, July 2026, page 1168-1181
ISSN 2722-7987 (Media Online)
Website https://ejurnal.seminar-id.com/index.php/tin
DOI 10.47065/tin.v7i2.10711

pada gerakan inti dari gestur yang digunakan. Hal ini penting dilakukan untuk menghilangkan data yang tidak bermakna akibat durasi berlebih pada saat perekaman data (Sindu dkk., 2024). Setelah semua data video selesai dibersihkan, dilanjutkan dengan tahap ekstraksi landmark dengan menggunakan MediaPipe Hands dan MediaPipe Pose, menghasilkan 21 keypoint dari telapak tangan dan 2 keypoint dari bahu kiri dan kanan (Google AI Edge, 2025). Data keypoint terdiri dari nilai koordinat x, y, dan z. Dalam proses ekstraksi ini juga dilakukan deteksi kegagalan ekstraksi landmarks dengan menambahkan satu kolom yang menyimpan data apakah landmarks terekstraksi dengan baik atau tidak. Dari hasil ekstraksi tersebut, setiap landmark tangan menghasilkan tiga nilai koordinat (x, y, z), sehingga 21 landmark tangan menghasilkan 63 nilai per frame. Ditambah dengan 2 landmark bahu yang masing-masing juga memiliki tiga koordinat, total fitur yang digunakan dalam penelitian ini adalah 69 fitur.

2.4.3 Pre-Processing Data

Langkah awal yang dilakukan dalam tahapan ini adalah melakukan pembersihan data. Metode pembersihan yang digunakan adalah penghapusan data jika dalam sebuah video terdapat lebih dari 3 frame yang mengalami kegagalan ekstraksi landmark. Namun, jika terdapat 3 atau kurang, maka akan dilakukan metode interpolasi. Selanjutnya, dilakukan normalisasi data dalam dua tahap. Tahap pertama adalah normalisasi translasi, di mana seluruh nilai koordinat landmark digeser sehingga titik tengah antara bahu kanan dan kiri menjadi titik asal baru bernilai (0, 0, 0). Titik tengah bahu dipilih sebagai titik acuan karena posisinya relatif stabil secara anatomis selama gestur tangan dilakukan, sehingga efektif digunakan untuk menghilangkan pengaruh posisi pengguna terhadap kamera. Pemilihan pendekatan normalisasi berbasis titik acuan tubuh yang stabil ini dilandasi oleh temuan Kim dkk. (2020) membuktikan bahwa normalisasi keypoint berdasarkan panjang bagian tubuh yang relatif stabil ukurannya di sepanjang video, seperti leher dan bahu, menghasilkan performa yang signifikan lebih baik dibandingkan dengan normalisasi berbasis statistik rata-rata dan standar deviasi seluruh titik tubuh. Titik tengah bahu yang digunakan pada penelitian ini secara fungsional setara dengan titik leher pada kerangka OpenPose yang digunakan Kim dkk., mengingat MediaPipe Pose tidak menyediakan landmark leher secara eksplisit.

Tahap kedua adalah normalisasi skala, di mana seluruh nilai koordinat hasil translasi dibagi dengan jarak antara landmark bahu kiri dan kanan (lebar bahu) sebagai faktor skala. Pendekatan ini berbeda dengan Kim dkk. (2020) yang menggunakan jarak leher ke bahu kanan sebagai referensi skala dengan performa terbaik pada pengujian mereka. Selain itu, berbeda dengan Kim dkk. yang melakukan normalisasi pada ruang koordinat dua dimensi (x, y), penelitian ini memanfaatkan koordinat tiga dimensi (x, y, z) yang disediakan MediaPipe Pose, sehingga perhitungan faktor skala turut mempertimbangkan dimensi kedalaman, alih-alih terbatas pada bidang dua dimensi seperti pada pendekatan Kim dkk. (2020). Normalisasi ini membuat representasi data menjadi invariant terhadap jarak fisik pengguna dari kamera, sehingga gestur yang dilakukan dari jarak dekat maupun jauh menghasilkan representasi fitur yang setara. Setelah itu, skema pembagian data menggunakan metode LOSO (Leave One Subject Out) yang membagi data untuk latih dan uji berdasarkan subjek, di mana masing-masing dari keempat partisipan akan berperan sebagai data uji untuk melihat kemampuan generalisasi model. Khusus pada data training, dilakukan augmentasi dengan melakukan horizontal flip sehingga diperoleh data dengan penggunaan tangan kiri. Mengingat kelas next dan prev didefinisikan berdasarkan arah pergerakan tangan pada sumbu horizontal (x), proses augmentasi horizontal flip pada kedua kelas ini turut disertai penukaran label: data next yang di-flip diberi label prev, dan sebaliknya, untuk menjaga kesesuaian antara arah lintasan hasil flip dengan makna gestur yang sesungguhnya

Pada setiap iterasi skema LOSO, baik data latih setelah augmentasi, data validasi internal, maupun data uji diberi padding untuk menyamakan panjang seluruh sekuens, dengan panjang sekuens maksimum ditetapkan berdasarkan panjang sekuens terpanjang yang ditemukan dalam keseluruhan dataset. Sampel dengan panjang kurang dari panjang maksimum tersebut diberi padding menggunakan teknik *centered padding*, sehingga data asli berada di tengah sekuens dan padding mengisi bagian awal serta akhir. Nilai padding yang digunakan dipilih berada di luar rentang nilai koordinat hasil normalisasi, sehingga dapat dibedakan dengan jelas dari data asli. Selanjutnya, diterapkan lapisan Masking agar model hanya fokus mempelajari frame yang berisi data asli dan mengabaikan frame padding selama proses pelatihan pada setiap iterasi(Zhu & Chollet, 2021).

2.4.4 Pelatihan Model

Tahapan pelatihan model menggunakan metode LOSO. Dengan memanfaatkan framework TensorFlow, dilakukan 8 percobaan pelatihan model dengan susunan layer yang berbeda untuk mendapatkan model dengan susunan layer yang menghasilkan performa terbaik. 8 percobaan susunan layer ini didasarkan pada dua penelitian terdahulu yang dilakukan oleh Huu dkk. (2023), serta Biswas dkk. (2024). Adapun susunan layer dari 8 model yang diuji dapat dilihat pada Tabel 1. Untuk memudahkan pembacaan, model Baseline 1 dan Baseline 2 selanjutnya disingkat menjadi B1 dan B2, sedangkan Model 1 hingga Model 6 disingkat menjadi M1 hingga M6 sesuai urutannya pada Tabel 1.

Tabel 1. Rancangan Arsitektur Pelatihan Model

Nama Model	Susuna Layer
Baseline 1	LSTM (64) LSTM (128) Dense (64)

TIN: Terapan Informatika Nusantara
Vol 7, No 2, July 2026, page 1168-1181
ISSN 2722-7987 (Media Online)
Website https://ejurnal.seminar-id.com/index.php/tin
DOI 10.47065/tin.v7i2.10711

Nama Model	Susuna Layer
	Dense (output)
Baseline 2	LSTM (64) LSTM (128) LSTM (64) Dense (64) Dense (32) Dense (output)
Model 1	LSTM (64) Dense (32) Dense (output)
Model 2	LSTM (128) Dense (64) Dense (output)
Model 3	LSTM (64) LSTM (64) Dense (32) Dense (output)
Model 4	LSTM (32) LSTM (64) Dense (32) Dense (output)
Model 5	LSTM (64) LSTM (128) LSTM (64) Dense (32) Dense (output)
Model 6	LSTM (32) LSTM (64) LSTM (32) Dense (32) Dense (output)

Model dilatih dengan pengaturan *hyperparameter* yang sama. Menggunakan epoch sebanyak 200, optimizer Adam dengan learning rate sebesar 0.001, serta menerapkan mekanisme Callback Early Stopping dengan patience 15 yang memantau nilai loss validasi saat pelatihan model. Kemudian, dalam proses pengujian model menggunakan data testing. Untuk memastikan model tidak hanya menghafal karakteristik unik dari setiap individu, evaluasi model dilakukan menggunakan metode pengujian *Leave-One-Subject-Out Cross-Validation* (LOSO-CV). Mengingat dataset terdiri dari empat partisipan, pengujian dilakukan dalam empat iterasi. Pada setiap iterasi, data dari tiga partisipan digunakan sebagai data latih (*training*), sementara data dari satu partisipan yang belum pernah dilihat oleh model digunakan sepenuhnya sebagai data uji (*testing*). Pendekatan LOSO ini secara ketat mencegah terjadinya *data leakage* dan merepresentasikan skenario dunia nyata di mana sistem diuji oleh pengguna baru.

2.4.5 Evaluasi Model

Setelah proses pelatihan selesai, dilakukan pengujian terhadap masing-masing model menggunakan data testing dari masing-masing partisipan yang telah dibagi menggunakan metode LOSO untuk mengukur performa akhir. Pengujian menggunakan beberapa metrik evaluasi yang umum digunakan dalam klasifikasi, yaitu akurasi, presisi, recall, dan F1-score. Akurasi adalah proporsi keseluruhan prediksi yang benar terhadap total data yang diuji. Presisi menunjukkan sejauh mana prediksi positif yang dikeluarkan model dapat dipercaya, yakni dengan melihat proporsi prediksi positif yang memang benar adanya; semakin tinggi nilainya, semakin sedikit kasus salah deteksi (*false positive*) yang dihasilkan model. Sementara itu, recall yang juga dikenal sebagai sensitivitas menilai kelengkapan model dalam menangkap seluruh data positif yang sesungguhnya ada pada ground truth; nilai recall yang tinggi berarti model jarang melewatkan objek yang seharusnya terdeteksi (*false negative*). Adapun F1-score dihitung sebagai rata-rata harmonik antara presisi dan recall, sehingga memberi satu ukuran tunggal yang mencerminkan keseimbangan antara ketepatan dan kelengkapan deteksi. Model metrik ini terutama berguna ketika distribusi kelas pada dataset tidak seimbang (Saputra dkk., 2026). Presisi, recall, dan F1-score dihitung menggunakan skema rata-rata makro (macro average) di antara ketujuh kelas gestur, sehingga setiap kelas memberikan kontribusi yang setara terhadap nilai akhir metrik, terlepas dari jumlah sampel pada masing-masing kelas.

TIN: Terapan Informatika Nusantara
Vol 7, No 2, July 2026, page 1168-1181
ISSN 2722-7987 (Media Online)
Website https://ejurnal.seminar-id.com/index.php/tin
DOI 10.47065/tin.v7i2.10711

3. HASIL DAN PEMBAHASAN

Bagian ini memaparkan hasil dari setiap tahapan penelitian yang telah dilakukan, mulai dari pengumpulan dan pengolahan data, pelatihan model, hingga pengujian performa seluruh variasi arsitektur LSTM yang dikembangkan. Hasil disajikan secara bertahap mengikuti alur penelitian, kemudian dilanjutkan dengan pembahasan yang menganalisis temuan tersebut secara lebih mendalam. Evaluasi model dilakukan berdasarkan empat metrik utama, yaitu akurasi, presisi, recall, dan F1-score, serta membandingkan performa antarmodel untuk menentukan arsitektur yang paling optimal dalam mengenali pola temporal gestur tangan dinamis.

3.1 Pengumpulan Data

Dalam proses pengumpulan data, pengambilan video dilakukan secara langsung dengan partisipan melakukan repetisi gerakan setiap gestur sebanyak 50-60 kali. Hasil pengumpulan data menunjukkan sebanyak 1696 video dengan variasi posisi tangan dan kecepatan gerak. Dengan contoh gestur seperti pada Gambar 2

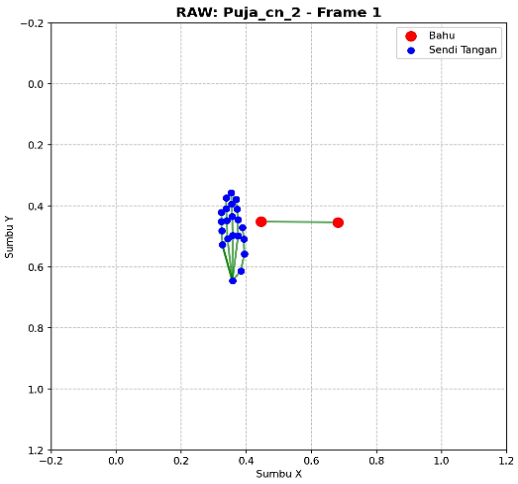
(a) (b) (c)

(d) (e) (f)

Gambar 2. Contoh Dataset: (a) Next; (b) Previous; (c) Zoom in; (d) Zoom Out; (e) Fullscreen; (f) Cancel

3.2 Pengolahan Data dan Ekstraksi Fitur

Setelah tahap pengumpulan data, seluruh rekaman video gestur diseleksi melalui proses segmentasi. Proses ini bertujuan untuk mengisolasi bagian inti dari setiap gerakan dan mengeliminasi durasi frame awal atau akhir yang tidak relevan. Video yang telah tersegmentasi kemudian diproses ke tahap ekstraksi landmark mentah menggunakan pustaka MediaPipe. Tahapan ini berfungsi untuk mentransformasikan informasi piksel dari setiap frame video menjadi koordinat numerik dua dimensi yang merepresentasikan titik-titik sendi tubuh. Melalui transformasi ini, data visual diubah menjadi urutan representasi spasial terstruktur yang dapat diproses lebih lanjut oleh model Long Short-Term Memory (LSTM). Representasi visual dari hasil ekstraksi koordinat landmark pada satu frame tunggal dapat dilihat pada Gambar 3.



Gambar 3. Hasil Ekstraksi Landmark

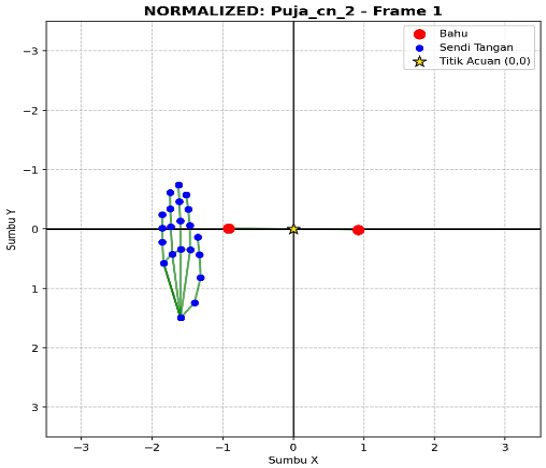
Berdasarkan visualisasi pada Gambar 3 ekstraksi landmark dipetakan ke dalam ruang koordinat dua dimensi yang merepresentasikan posisi spasial titik tubuh di dalam frame video. Klaster titik berwarna biru menunjukkan 21

TIN: Terapan Informatika Nusantara
Vol 7, No 2, July 2026, page 1168-1181
ISSN 2722-7987 (Media Online)
Website https://ejurnal.seminar-id.com/index.php/tin
DOI 10.47065/tin.v7i2.10711

titik anatomi sendi tangan (hand landmarks) yang saling terhubung membentuk kerangka pose tangan. Sebagai contoh observasi pada frame pertama dari eksekusi gestur Cancel tersebut, titik-titik tangan terekam membentuk formasi telapak tangan yang terbuka lebar, yang merupakan tahapan awal sebelum tangan digerakkan ke bawah. Selain landmark tangan, proses ekstraksi juga menangkap posisi bahu kiri dan kanan yang direpresentasikan oleh dua titik berwarna merah. Kehadiran landmark bahu ini berperan sebagai titik acuan tak bergerak (reference point) untuk mengetahui posisi relatif pergerakan tangan terhadap postur tubuh pengguna secara keseluruhan.

3.3 Pre-Processing Data

Setelah melalui tahapan ekstraksi landmark, selanjutnya masuk ke tahapan pre-processing yang diawali dengan penanganan missing values akibat kegagalan sistem dalam mendeteksi landmark tangan pada frame tertentu. Sekuens video yang mengalami kehilangan data spasial secara fatal dihapus dari dataset untuk menghindari masuknya pola anomali ke dalam model. Untuk missing values minor pada sebagian kecil frame, dilakukan penambalan koordinat menggunakan teknik interpolasi guna merekonstruksi jalur gerakan secara mulus dan menjaga kontinuitas data time-series. Untuk merekonstruksi jalur gerakan secara mulus dan menjaga kontinuitas data time-series. Selanjutnya, khusus untuk keenam kelas gestur kendali, ditemukan bahwa salah satu partisipan hanya memiliki 50 repetisi tersisa pada salah satu kelas gestur setelah proses pembersihan data. Karena keterbatasan ini tidak memungkinkan penambahan data pada kelas tersebut, dilakukan random undersampling menggunakan seed = 42 untuk menyamakan jumlah repetisi pada partisipan lain menjadi 50 repetisi per kelas gestur kendali, untuk menjaga keseimbangan jumlah data antarkelas sekaligus memastikan hasil seleksi dapat direplikasi. Penyetaraan ini tidak diterapkan pada kelas idle, mengingat kelas ini berfungsi sebagai kelas negatif/penolakan sehingga tidak terikat pada protokol repetisi yang sama dengan kelas gestur kendali. Data yang telah melalui tahap penanganan missing value selanjutnya dinormalisasi mengikuti prosedur yang telah dijelaskan pada bagian 2.3. Hasil dari normalisasi divisualisasikan pada Gambar 4.



Gambar 4. Hasil Normalisasi Landmark

Setelah melalui keseluruhan tahapan pembersihan data dan penyetaraan jumlah repetisi antarpartisipan sebagaimana dijelaskan sebelumnya, diperoleh data akhir sebanyak 1.419 video yang tersebar ke dalam tujuh kelas gestur, terdiri dari enam kelas gestur kendali dengan jumlah yang seragam pada tiap partisipan, serta satu kelas idle yang jumlahnya bervariasi mengikuti hasil perekaman asli. Rincian persebaran data akhir tersebut disajikan pada Tabel 2.

Tabel 2. Jumlah Sebaran Data

Gestur	P1	P2	P3	P4	Total Video
Next	50	50	50	50	200
Previous	50	50	50	50	200
Zoom in	50	50	50	50	200
Zoom out	50	50	50	50	200
Fullscreen	50	50	50	50	200
Cancel	50	50	50	50	200
Idle	56	54	55	54	219
Total Video					1419

Sebanyak 1.419 sampel data yang telah diekstraksi kemudian dibagi menggunakan metode LOSO, dengan 4 iterasi. Iterasi 1 menggunakan data dari P1 sebagai test set, lalu P2, P3, dan P4. Sebagaimana dijelaskan pada bagian 2.4, pembagian data pada penelitian ini menggunakan skema *Leave-One-Subject-Out* (LOSO), yang menghasilkan empat iterasi pembagian data berdasarkan partisipan yang dijadikan data uji. Pada setiap iterasi, data gabungan dari ketiga partisipan yang tidak menjadi data uji terlebih dahulu dipisahkan menjadi 90% data latih dan 10% data validasi

TIN: Terapan Informatika Nusantara
Vol 7, No 2, July 2026, page 1168-1181
ISSN 2722-7987 (Media Online)
Website https://ejurnal.seminar-id.com/index.php/tin
DOI 10.47065/tin.v7i2.10711

secara stratifikasi per kelas gestur, sebelum data latih tersebut diperkaya melalui augmentasi *horizontal flip* sehingga volumenya berlipat ganda. Rincian pembagian data pada setiap iterasi disajikan pada Tabel 3

Tabel 3. Pembagian Data Skema LOSO

Iterasi	Data Pelatihan	Data Uji	Jumlah Data Latih	Augmentasi	Jumlah Data Validasi	Jumlah Data Uji
1	P2, P3, P4	P1	956	1912	107	356
2	P1, P3, P4	P2	958	1916	107	354
3	P1, P2, P4	P3	957	1914	107	355
4	P1, P2, P3	P4	958	1916	107	354

Berdasarkan Tabel 3, jumlah data uji pada setiap iterasi berkisar antara 354 hingga 356 video, mengikuti jumlah data asli milik masing-masing partisipan yang dijadikan data uji pada iterasi tersebut. Jumlah data latih sebelum augmentasi berkisar antara 956 hingga 958 video, meningkat dua kali lipat menjadi 1.912 hingga 1.916 sampel setelah augmentasi *horizontal flip* diterapkan, sementara jumlah data validasi internal konsisten sebesar 107 video pada seluruh iterasi. Konsistensi ini menunjukkan bahwa keempat iterasi memiliki proporsi data yang relatif seimbang, meskipun komposisi partisipan pada data latih berbeda-beda di setiap iterasi.

Mengingat setiap eksekusi gestur memiliki durasi waktu yang bervariasi, dilakukan analisis terhadap panjang sekuens dan diidentifikasi bahwa sekuens terpanjang dalam dataset terdiri dari 60 frame. Oleh karena itu, teknik padding diterapkan guna menyeragamkan dimensi masukan menjadi 60 frame. Berdasarkan pemeriksaan terhadap rentang nilai koordinat hasil normalisasi, diperoleh nilai berkisar antara -2,6553 hingga 3,1701, sehingga nilai padding ditetapkan sebesar 99,0 karena berada jauh di luar rentang tersebut dan tidak berpotensi tumpang tindih dengan data asli. Agar penambahan elemen padding ini tidak mendistorsi proses pembelajaran, mekanisme masking turut diimplementasikan. Lapisan masking ini berfungsi untuk menginstruksikan model LSTM agar mengabaikan nilai padding buatan tersebut dan memusatkan perhatian sepenuhnya pada urutan data gerakan yang asli.

3.4 Pelatihan Model

Sebagaimana telah dijelaskan pada bagian 2.4, proses pelatihan dilakukan terhadap delapan arsitektur model (B1, B2, serta M1 hingga M6) menggunakan skema LOSO Cross-Validation, sehingga setiap arsitektur menghasilkan empat nilai akurasi dan loss validasi yang berbeda pada tiap iterasi-nya. Nilai rata-rata akurasi dan loss validasi dari keempat iterasi tersebut kemudian dihitung untuk masing-masing arsitektur, guna memperoleh gambaran performa yang lebih representatif dan tidak hanya bergantung pada pembagian data tunggal. Hasil lengkap dari proses pelatihan kedelapan arsitektur tersebut disajikan pada Tabel 4.

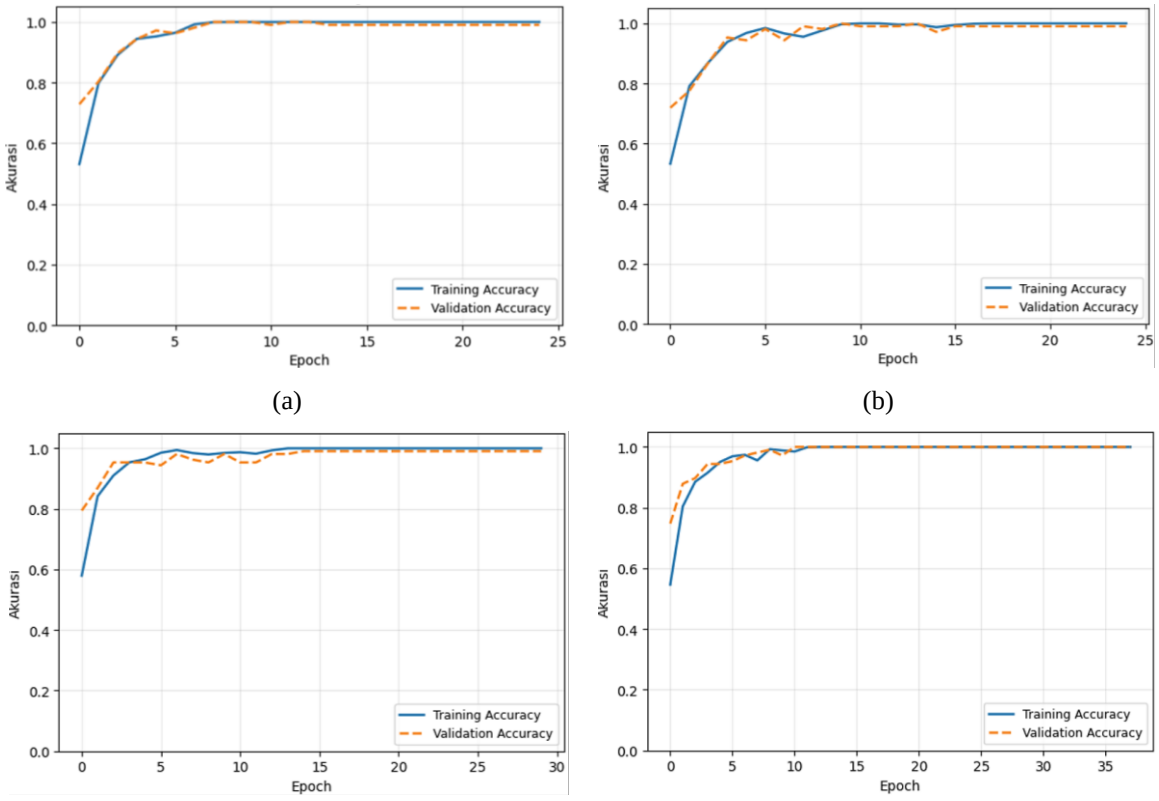
Tabel 4. Akurasi dan Loss Validasi Pelatihan Model

Model	Iterasi	Loss Validasi	Akurasi Validasi(%)	Rata-Rata Loss	Rata-Rata Akurasi(%)
B1	1	0.0095	100	0.0206	99.77
	2	0.0114	100		
	3	0.0602	99.07		
	4	0.0014	100		
B2	1	0.0025	100	0.0040	100
	2	0.0013	100		
	3	0.0106	100		
	4	0.0016	100		
M1	1	0.0050	100	0.0187	99.53
	2	0.0241	99.07		
	3	0.0012	100		
	4	0.0445	99.07		
M2	1	0.0021	100	0.0021	100
	2	0.0024	100		
	3	0.0009	100		
	4	0.0032	100		
M3	1	0.0011	100	0.0032	100
	2	0.0064	100		
	3	0.0026	100		
	4	0.0026	100		
M4	1	0.0128	99.07	0.0106	99.30
	2	0.0086	99.07		
	3	0.0194	99.07		
	4	0.0017	100		
M5	1	0.0019	100	0.0032	100
	2	0.0074	100		

TIN: Terapan Informatika Nusantara
Vol 7, No 2, July 2026, page 1168-1181
ISSN 2722-7987 (Media Online)
Website https://ejurnal.seminar-id.com/index.php/tin
DOI 10.47065/tin.v7i2.10711

Model	Iterasi	Loss Validasi	Akurasi Validasi(%)	Rata-Rata Loss	Rata-Rata Akurasi(%)
	3	0.0018	100		
	4	0.0015	100		
M6	1	0.0210	99.07	0.0102	99.53
	2	0.0128	99.07		
	3	0.0044	100		
	4	0.0023	100		

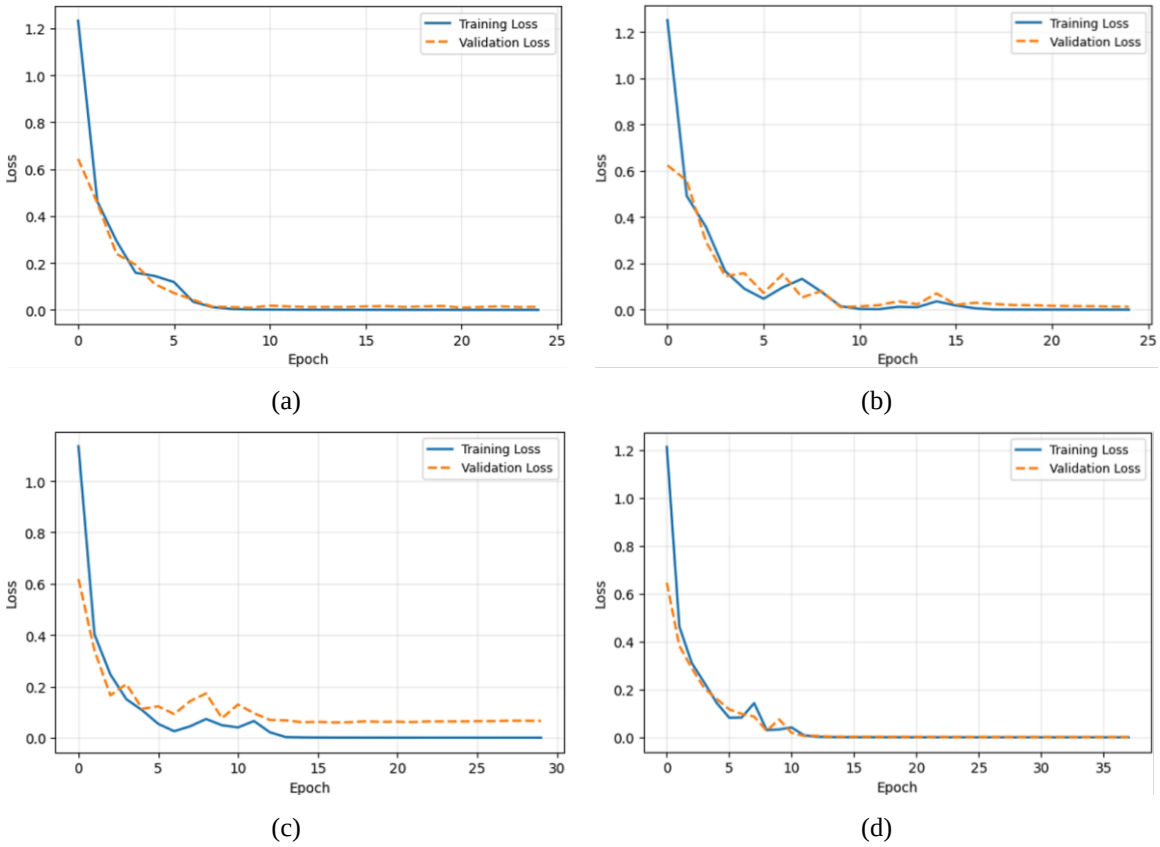
Dari data hasil pelatihan model pada Tabel 4, nilai akurasi validasi internal pada seluruh arsitektur berada pada rentang yang hampir seragam (99,30% - 100%), mengingat data validasi internal berasal dari sampel yang belum dipelajari model, tetapi tetap berasal dari partisipan yang sama dengan data latih, sehingga metrik ini kurang mampu membedakan performa antararsitektur secara signifikan. Nilai loss validasi menunjukkan variasi yang lebih informatif, dengan M2 mencatatkan nilai terendah, yaitu 0.0021, sementara model B1 justru mencatatkan nilai loss validasi tertinggi pada 0.0206 di antara kedelapan arsitektur. hal ini disebabkan oleh lonjakan *validation loss* pada iterasi ketiga LOSO, sementara tiga iterasi lainnya menunjukkan nilai *validation loss* yang rendah dan relatif konsisten. Hal ini menunjukkan bahwa nilai *validation loss* internal tidak selalu berkorelasi langsung dengan kemampuan generalisasi lintas subjek. Hal tersebut disebabkan oleh data validasi yang masih berasal dari partisipan yang sama dengan data pelatihan, sehingga evaluasi pada tahap ini lebih mencerminkan kemampuan model dalam mempelajari distribusi data pelatihan daripada kemampuannya mengenali pengguna yang benar-benar baru. Kurva akurasi dan loss dari model B1 selama pelatihan ditunjukkan pada Gambar 5 dan 6.



Gambar 5. Kurva Akurasi Pelatihan Model B1: (a) Iterasi 1; (b) Iterasi 2; (c) Iterasi 3; (d) Iterasi 4

Gambar 5 memperlihatkan kurva akurasi pelatihan dan validasi pada empat skenario. Seluruh skenario menunjukkan pola pembelajaran yang serupa, di mana akurasi meningkat secara cepat pada beberapa epoch pertama sebelum kemudian melandai hingga mencapai kondisi yang stabil. Pada awal pelatihan, akurasi pelatihan berada pada kisaran 53-58%, sedangkan akurasi validasi telah berada pada kisaran 72-80%. Setelah sekitar epoch ke-5, kedua kurva mengalami peningkatan yang signifikan dan pada umumnya telah mencapai akurasi di atas 95%. Memasuki sekitar epoch ke-10 hingga ke-15, kurva akurasi pelatihan dan validasi mulai mendatar dengan nilai yang mendekati 100%. Jarak antara kedua kurva juga relatif kecil sepanjang proses pelatihan. Pada beberapa skenario, seperti iterasi ke-2 dan ke-3, terlihat fluktuasi kecil pada kurva validasi di beberapa epoch, namun perubahan tersebut bersifat sementara dan kembali mengikuti tren peningkatan hingga mencapai kondisi stabil. Kesamaan pola antara kurva pelatihan dan validasi menunjukkan bahwa model mampu mempertahankan performa yang konsisten pada data pelatihan maupun data validasi, tanpa adanya penurunan akurasi validasi pada akhir proses pelatihan yang umumnya menjadi indikasi overfitting.

TIN: Terapan Informatika Nusantara
Vol 7, No 2, July 2026, page 1168-1181
ISSN 2722-7987 (Media Online)
Website https://ejurnal.seminar-id.com/index.php/tin
DOI 10.47065/tin.v7i2.10711



Gambar 6. Kurva Loss Pelatihan Model: (a) Iterasi 1; (b) Iterasi 2; (c) Iterasi 3; (d) Iterasi 4

Kurva loss pada Gambar 6 menunjukkan bahwa nilai kesalahan pelatihan maupun validasi menurun secara konsisten selama proses pelatihan. Pada seluruh iterasi, penurunan loss berlangsung paling cepat pada beberapa epoch pertama, kemudian berangsur melambat hingga mendekati nol seiring meningkatnya akurasi model. Pola ini menunjukkan bahwa model secara bertahap mampu meminimalkan kesalahan prediksi selama proses optimasi menggunakan fungsi loss *Categorical Crossentropy*. Kurva *training loss* dan *validation loss* juga memperlihatkan pola yang berdekatan sepanjang pelatihan. Meskipun terdapat sedikit fluktuasi pada *validation loss* di beberapa epoch, khususnya pada iterasi ke-2 dan ke-3, fluktuasi tersebut tidak berkembang menjadi peningkatan loss yang berkelanjutan. Setelah sekitar epoch ke-10 hingga ke-15, kedua kurva cenderung stabil dengan nilai loss yang sangat rendah. Selain itu, proses pelatihan berhenti sebelum mencapai batas maksimum 200 epoch karena mekanisme *early stopping* yang menghentikan pelatihan ketika tidak lagi diperoleh perbaikan nilai validasi dalam sejumlah epoch berturut-turut. Hal ini menunjukkan bahwa model telah mencapai kondisi konvergen sehingga pelatihan tambahan tidak lagi memberikan peningkatan performa yang berarti.

3.5 Evaluasi Model

Setelah proses pelatihan selesai, tahap selanjutnya adalah pengujian terhadap kedelapan model menggunakan data testing untuk mengetahui performa akurasi, presisi, *recall*, dan F1-score dalam mengenali data baru. Hasil dari pengujian model ditampilkan pada Tabel 5. Evaluasi yang sesungguhnya terhadap kemampuan generalisasi model (cross-subject generalization) dilakukan pada tahap pengujian menggunakan 100% himpunan data dari subjek hold-out. Subjek uji ini sama sekali tidak dilibatkan dalam pembaruan bobot (training) maupun pemantauan penghentian awal (validation), sehingga merepresentasikan pengguna baru di skenario dunia nyata.

Tabel 5. Nilai Akurasi dan Presisi Pengujian Model

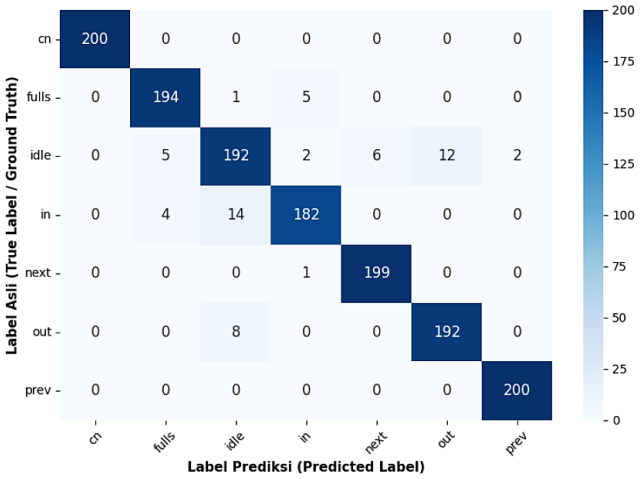
Model	Partisipan Uji	Akurasi Uji	Presisi	Recall	F1-Score	Rata-Rata Akurasi Uji	Rata-Rata Presisi	Rata-Rata Recall	Rata-Rata F1-Score
B1	P1	0.9691	0.9705	0.9713	0.9695	0.9577	0.9603	0.9590	0.9584
	P2	0.9548	0.9556	0.9551	0.9552				
	P3	0.9127	0.9203	0.9151	0.9142				
	P4	0.9944	0.9949	0.9943	0.9945				
B2	P1	0.9494	0.9540	0.9510	0.9509	0.9450	0.9497	0.9455	0.9456
	P2	0.8983	0.9065	0.8976	0.8973				
	P3	0.9352	0.9410	0.9364	0.9371				
	P4	0.9972	0.9974	0.9971	0.9972				

TIN: Terapan Informatika Nusantara
Vol 7, No 2, July 2026, page 1168-1181
ISSN 2722-7987 (Media Online)
Website https://ejurnal.seminar-id.com/index.php/tin
DOI 10.47065/tin.v7i2.10711

Model	Partisipan Uji	Akurasi Uji	Presisi	Recall	F1-Score	Rata-Rata Akurasi Uji	Rata-Rata Presisi	Rata-Rata Recall	Rata-Rata F1-Score
M1	P1	0.9382	0.9478	0.9417	0.9390	0.9048	0.9241	0.9064	0.9057
	P2	0.7910	0.8274	0.7943	0.7899				
	P3	0.9239	0.9530	0.9229	0.9272				
	P4	0.9661	0.9683	0.9666	0.9667				
M2	P1	0.9579	0.9619	0.9593	0.9581	0.9507	0.9583	0.9516	0.9507
	P2	0.9209	0.9270	0.9236	0.9199				
	P3	0.9352	0.9550	0.9343	0.9359				
	P4	0.9887	0.9891	0.9894	0.9889				
M3	P1	0.9691	0.9743	0.9689	0.9700	0.9387	0.9492	0.9396	0.9385
	P2	0.8559	0.8737	0.8600	0.8531				
	P3	0.9296	0.9490	0.9294	0.9309				
	P4	1.0000	1.0000	1.0000	1.0000				
M4	P1	0.9185	0.9217	0.9227	0.9193	0.9394	0.9451	0.9413	0.9403
	P2	0.9407	0.9482	0.9428	0.9413				
	P3	0.9296	0.9367	0.9312	0.9315				
	P4	0.9689	0.9738	0.9688	0.9693				
M5	P1	0.9382	0.945	0.9433	0.9374	0.9465	0.9529	0.9478	0.9466
	P2	0.8983	0.9083	0.8984	0.8981				
	P3	0.9493	0.9581	0.9494	0.9508				
	P4	1.0000	1.0000	1.0000	1.0000				
M6	P1	0.9185	0.9215	0.9236	0.9164	0.9444	0.9518	0.9464	0.9446
	P2	0.952	0.9571	0.9546	0.9516				
	P3	0.9099	0.9311	0.9104	0.9129				
	P4	0.9972	0.9974	0.9971	0.9972				

Temuan paling signifikan dari evaluasi ini terlihat pada performa Model Baseline 1 (B1). Meskipun B1 memiliki rata-rata *validation loss* paling tinggi selama proses pelatihan, arsitektur ini justru menunjukkan kemampuan generalisasi terbaik saat dihadapkan pada subjek asing. Model B1 unggul secara konsisten pada seluruh metrik pengujian dengan memperoleh akurasi uji 95,77%, presisi 96,03%, recall 95,90%, dan F1-score 95,84%, tertinggi di antara kedelapan arsitektur. Oleh karena itu, pemilihan B1 sebagai arsitektur terbaik pada penelitian ini didasarkan pada performa generalisasinya terhadap data uji lintas subjek, yang merupakan indikator paling relevan untuk mengukur kemampuan model dalam mengenali pengguna baru, dibandingkan dengan nilai loss validasi internal yang lebih rentan dipengaruhi oleh variasi kesulitan pada iterasi tertentu. Temuan ini menunjukkan bahwa peningkatan kompleksitas arsitektur tidak selalu menghasilkan kemampuan generalisasi yang lebih baik terhadap subjek baru. Meskipun beberapa arsitektur memperoleh validation loss yang lebih rendah selama pelatihan, performa pengujian lintas subjek menunjukkan bahwa arsitektur yang lebih sederhana justru mampu mempertahankan kemampuan klasifikasi yang lebih konsisten pada data yang benar-benar belum pernah dilihat sebelumnya.

Meskipun metrik evaluasi menunjukkan performa klasifikasi yang tinggi, metrik tersebut belum menggambarkan distribusi kesalahan prediksi pada setiap kelas gestur. Oleh karena itu, analisis dilanjutkan menggunakan confusion matrix untuk mengidentifikasi pola klasifikasi dan kesalahan antarkelas. Gambar 7 menyajikan confusion matrix teragregasi yang diperoleh dari seluruh skenario pengujian LOSO.



Gambar 7. Hasil Confusion Matrix Model B1


TIN: Terapan Informatika Nusantara
Vol 7, No 2, July 2026, page 1168-1181
ISSN 2722-7987 (Media Online)
Website https://ejurnal.seminar-id.com/index.php/tin
DOI 10.47065/tin.v7i2.10711

Hasil pengujian LOSO pada model B1 menunjukkan sebagian besar sampel berada pada diagonal utama, yang menunjukkan bahwa model mampu mengklasifikasikan mayoritas gestur ke kelas yang sesuai. Sebagian besar sampel berada pada diagonal utama, yang menunjukkan bahwa model mampu mengklasifikasikan mayoritas gestur dengan benar. Kelas cancel dan previous berhasil dikenali seluruhnya, sedangkan kelas next hanya mengalami satu kesalahan klasifikasi. Sebaliknya, kesalahan prediksi lebih banyak ditemukan pada kelas idle, zoom in, zoom out, dan fullscreen, meskipun sebagian besar sampel pada kelas-kelas tersebut tetap berhasil dikenali dengan benar. Pola kesalahan menunjukkan bahwa kesalahan klasifikasi tidak terjadi secara acak, tetapi terkonsentrasi pada kelas-kelas yang memiliki kemiripan konfigurasi pose tangan. Kelas idle merupakan kelas yang paling sering mengalami kesalahan karena terdiri atas beberapa variasi pose statis yang juga digunakan sebagai pose awal maupun pose akhir beberapa gestur dinamis. Akibatnya, sebagian kecil sampel idle diprediksi sebagai gestur tertentu, sedangkan gestur yang memiliki karakteristik gerakan yang lebih khas, seperti next, previous, dan cancel, hampir tidak mengalami kesalahan klasifikasi.

3.6 Pembahasan

Hasil pengujian menggunakan skema Leave-One-Subject-Out (LOSO) menunjukkan bahwa model mampu mempertahankan performa yang konsisten ketika diuji pada partisipan yang tidak dilibatkan selama pelatihan. Dengan rata-rata akurasi uji 95,77%, presisi 96,03%, recall 95,90%, dan F1-score 95,84%, model B1 dipilih sebagai arsitektur terbaik karena memiliki kemampuan generalisasi paling baik terhadap pengguna baru. Menariknya, model B1 tidak memperoleh *validation loss* internal terendah, tetapi menghasilkan performa terbaik pada pengujian LOSO. Hal ini menunjukkan bahwa *validation loss* internal tidak selalu berkorelasi langsung dengan kemampuan generalisasi lintas subjek karena data validasi masih berasal dari partisipan yang sama dengan partisipan dalam data pelatihan. Temuan ini juga menunjukkan bahwa peningkatan kompleksitas arsitektur tidak selalu menghasilkan kemampuan generalisasi yang lebih baik terhadap pengguna baru.

Jika dibandingkan dengan penelitian terdahulu yang sama-sama memanfaatkan kombinasi MediaPipe dan LSTM, hasil penelitian ini menunjukkan performa yang kompetitif. Artha dkk. (2025), melaporkan akurasi sebesar 92,35% pada klasifikasi gerakan *push-up*, sedangkan Ho & Santoso (2025) memperoleh akurasi sebesar 93,94% pada pengenalan Bahasa Isyarat Indonesia (SIBI). Meskipun demikian, perbandingan tersebut perlu diinterpretasikan secara hati-hati karena masing-masing penelitian menggunakan karakteristik dataset, jumlah kelas, skenario pengujian, serta tingkat kompleksitas gestur yang berbeda. Agustiani dkk. (2024) sendiri mengidentifikasi keterbatasan variasi kontrol pada sistem gestur statis mereka dan menyarankan pengembangan ke arah gestur dinamis untuk penelitian mendatang. Penelitian ini dapat dipandang sebagai upaya untuk menjawab arah tersebut, dengan capaian akurasi uji 95,77% pada tujuh kelas gestur dan lebih banyak variasi kelas dibandingkan dengan sistem Agustiani dkk. (2024), sekaligus tetap mempertahankan akurasi tinggi. Analisis *confusion matrix* menunjukkan bahwa sebagian besar kesalahan klasifikasi terkonsentrasi pada kelas *idle* yang memiliki kemiripan pose dengan beberapa gestur dinamis. Temuan ini menunjukkan bahwa tantangan utama penelitian bukan terletak pada kemampuan model mempelajari pola temporal, melainkan pada kemiripan konfigurasi pose antarkelas yang memang terdapat pada rancangan dataset. Meskipun demikian, sebagian besar gestur tetap berhasil dikenali dengan benar sehingga model mampu mempertahankan kemampuan generalisasi yang baik terhadap pengguna baru.

Penelitian ini masih memiliki beberapa keterbatasan yang dapat memengaruhi ruang lingkup generalisasi hasil. Dataset hanya melibatkan empat partisipan dengan penggunaan tangan kanan serta dikumpulkan pada kondisi pencahayaan, latar belakang, dan jarak kamera yang relatif terkontrol. Selain itu, kelas *idle* dirancang menggunakan variasi pose yang memiliki kemiripan dengan beberapa gestur dinamis, sehingga menjadi kelas yang paling menantang untuk diklasifikasikan. Oleh karena itu, penelitian selanjutnya dapat memperluas jumlah dan keberagaman partisipan, melibatkan variasi kondisi lingkungan yang lebih kompleks, serta memperkaya variasi sampel pada kelas *idle* untuk mengurangi kemiripan antarkelas dan meningkatkan kemampuan generalisasi model.

4. KESIMPULAN

Penelitian ini mengembangkan sistem kendali presentasi PowerPoint berbasis gestur tangan dinamis menggunakan MediaPipe Hands, MediaPipe Pose, dan Long Short-Term Memory (LSTM). Landmark dinormalisasi menggunakan titik tengah kedua bahu sebagai titik acuan untuk mengurangi pengaruh variasi posisi pengguna terhadap kamera, kemudian delapan variasi arsitektur LSTM dievaluasi menggunakan skema Leave-One-Subject-Out (LOSO) Cross-Validation. Hasil pengujian menunjukkan bahwa arsitektur Baseline 1 (B1) memberikan performa terbaik dengan rata-rata akurasi 95,77%, presisi 96,03%, recall 95,90%, dan F1-score 95,84%. Analisis confusion matrix menunjukkan bahwa sebagian besar gestur berhasil dikenali dengan baik, sedangkan kesalahan klasifikasi terutama terjadi pada kelas idle serta beberapa gestur yang memiliki kemiripan konfigurasi pose tangan. Temuan ini menunjukkan bahwa kombinasi MediaPipe dan LSTM mampu membangun model klasifikasi gestur tangan dinamis dengan kemampuan generalisasi yang baik terhadap pengguna yang tidak terlibat pada proses pelatihan. Penelitian selanjutnya dapat mengevaluasi model pada jumlah partisipan yang lebih banyak, kondisi lingkungan yang lebih beragam, serta membandingkan pendekatan ini dengan arsitektur lain untuk memperoleh pemahaman yang lebih komprehensif mengenai kemampuan generalisasi model.


TIN: Terapan Informatika Nusantara
Vol 7, No 2, July 2026, page 1168-1181
ISSN 2722-7987 (Media Online)
Website https://ejurnal.seminar-id.com/index.php/tin
DOI 10.47065/tin.v7i2.10711

REFERENCES

Agustiani, A. D., Sholahuddin, M. R., Putri, S. M., & Hidayatullah, P. (2024). Penggunaan MediaPipe untuk Pengenalan Gesture Tangan Real-Time dalam Pengendalian Presentasi. *Media Jurnal Informatika*, *16*(2), 147. https://doi.org/10.35194/mji.v16i2.4788

Artha, I. K. B. D., Arthana, I. K. R., & Suputra, P. H. (2025). Pengembangan Model Long Short-Term Memory Berbasis MediaPipe Pose untuk Klasifikasi dan Penilaian Gerakan Push-Up. *Jutisi : Jurnal Ilmiah Teknik Informatika dan Sistem Informasi*, *14*(3), 2106. https://doi.org/10.35889/jutisi.v14i3.3363

Biswas, S., Nandy, A., Naskar, A. K., & Saw, R. (2024). *MediaPipe with LSTM Architecture for Real-Time Hand Gesture Recognization*. https://doi.org/10.1007/978-3-031-58174-8_36

Dewi, N. P. N. P., Kertiasih, N. K., & Sintiari, N. L. D. (2022). Modifikasi Fruit Fly Optimiziation Algorithm untuk Optimasi General Regression Neural Network pada Kasus Prediksi Time-Series. *Jurnal Nasional Pendidikan Teknik Informatika (JANAPATI)*, *11*(3), 192–204. https://doi.org/10.23887/janapati.v11i3.54521

Google AI Edge. (2025, Februari 26). *MediaPipe Solutions guide*. https://ai.google.dev/edge/mediapipe/solutions/guide

Hax, D. R. T., Penava, P., Krodel, S., Razova, L., & Buettner, R. (2024). A Novel Hybrid Deep Learning Architecture for Dynamic Hand Gesture Recognition. *IEEE Access*, *12*, 28761–28774. https://doi.org/10.1109/ACCESS.2024.3365274

Herbert, O. M., Pérez-Granados, D., Ruiz, M. A. O., Cadena Martínez, R., Gutiérrez, C. A. G., & Antuñano, M. A. Z. (2024). Static and Dynamic Hand Gestures: A Review of Techniques of Virtual Reality Manipulation. Dalam *Sensors* (Vol. 24, Nomor 12). Multidisciplinary Digital Publishing Institute (MDPI). https://doi.org/10.3390/s24123760

Ho, P., & Santoso, H. (2025). LSTM-Based Hand Gesture Recognition for Indonesian Sign Language System (SIBI) on Affix, Alphabet, Number, and Word. Dalam *Journal of Applied Informatics and Computing (JAIC)* (Vol. 9, Nomor 3). http://jurnal.polibatam.ac.id/index.php/JAIC

Huu, P. N., Hong, P. D. Le, Dang, D. D., Quoc, B. V., Bao, C. N. Le, & Minh, Q. T. (2023). Proposing Hand Gesture Recognition System Using MediaPipe Holistic and LSTM. *International Conference on Advanced Technologies for Communications*, 433–438. https://doi.org/10.1109/ATC58710.2023.10318885

Kim, S., Kim, C. J., Park, H.-M., Jeong, Y., Jang, J. Y., & Jung, H. (2020). *Robust Keypoint Normalization Method for Korean Sign Language Translation using Transformer*. IEEE.

Prayoga, I. M. P., Indradewi, I. G. A. A. D., & Suputra, P. H. (2025). Pengenalan Kata Kolok Secara Real Time Menggunakan Mediapipe dan Algoritma Long Short-Term Memory (LSTM). *SINTECH JOURNAL*. https://doi.org/10.31598

Purbojo, T., & Wijaya, A. P. (2025). Enhancing Pose-Based Sign Language Recognition: A Comparative Study of Preprocessing Strategies with GRU and LSTM. *Advance Sustainable Science, Engineering and Technology*, *7*(2). https://doi.org/10.26877/asset.v7i2.1658

Putra, G. B. P., Suputra, P. H., Marti, N. W., & Sugiantari, K. F. (2025). Deteksi Jatuh Lansia Berbasis Landmark Sendi Pada Model LSTM. *JIP (Jurnal Informatika Polinema*, (ISSN: 2614-6371 E-ISSN: 2407-070X). https://www.kaggle.com.

Rahman, M. M., Uzzaman, A., Khatun, F., Aktaruzzaman, M., & Siddique, N. (2025). A comparative study of advanced technologies and methods in hand gesture analysis and recognition systems. Dalam *Expert Systems with Applications* (Vol. 266). Elsevier Ltd. https://doi.org/10.1016/j.eswa.2024.125929

Rehman, M. U., Ahmed, F., Khan, M. A., Tariq, U., Alfouzan, F. A., Alzahrani, N. M., & Ahmad, J. (2022). Dynamic hand gesture recognition using 3D-CNN and LSTM networks. *Computers, Materials and Continua*, *70*(3), 4675–4690. https://doi.org/10.32604/cmc.2022.019586

Saputra, I. P. G. D., Kesiman, M. W. A., & Sunarya, I. M. G. (2026). Deteksi Pemalsuan QRIS MPM Statis Menggunakan YOLO, PaddleOCR dan Metode Berbasis Aturan. *Bulletin of Computer Science Research*, *6*(2), 763–775. https://doi.org/10.47065/bulletincsr.v6i2.1020

Saputri, N. K. T. A., Gunadi, I. G. A., & Sunarya, I. M. G. (2024). Analisis Sentimen Pelayanan Daring di Fakultas Teknik dan Kejuruan Universitas Pendidikan Ganesha Menggunakan Algoritma Naïve Bayes dan LSTM. *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, *4*(3), 1120–1129. https://doi.org/10.57152/malcom.v4i3.1336

Setyowati, L., & Granito, H. N. (2025). *Smart Presentation Control Using Hand Gesture Recognition*. (2), 168–176. https://doi.org/10.56127/ijst.v4i

SHI, Y., LI, Y., FU, X., Kaibin, M. I. A. O., & Qiguang, M. I. A. O. (2021). Review of dynamic gesture recognition. Dalam *Virtual Reality and Intelligent Hardware* (Vol. 3, Nomor 3, hlm. 183–206). KeAi Communications Co. https://doi.org/10.1016/j.vrih.2021.05.001

Sindu, I. G. P., Sudarma, M., Hartati, R. S., & Gunantara, N. (2024). Classification of Tri Pramana learning activities in virtual reality environment using convolutional neural network. *IAES International Journal of Artificial Intelligence (IJ-AI*, *13*(3), 2840–2853. https://doi.org/10.11591/ijai.v

Sugiantari, K. F., Suputra, P. H., Dewi, L. J. E., & Putra, G. B. P. (2025). Pendekatan MLP Dalam Klasifikasi Bahasa Isyarat: Analisis Jarak Euclidean Landmark Tangan. Dalam *JIKA* (Vol. 9, Nomor 2). https://doi.org/10.31000/jika.v9i2.13368.

TIN: Terapan Informatika Nusantara
Vol 7, No 2, July 2026, page 1168-1181
ISSN 2722-7987 (Media Online)
Website https://ejurnal.seminar-id.com/index.php/tin
DOI 10.47065/tin.v7i2.10711

Uddin, Md. Z., Boletsis, C., & Rudshavn, P. (2025). Real-Time Norwegian Sign Language Recognition Using MediaPipe and LSTM. *Multimodal Technologies and Interaction*, *9*(3), 23. https://doi.org/10.3390/mti9030023

Yasen, M., & Jusoh, S. (2019). A systematic review on hand gesture recognition techniques, challenges and applications. *PeerJ Computer Science*, *2019*(9). https://doi.org/10.7717/peerj-cs.218

Zhang, F., Bazarevsky, V., Vakunov, A., Tkachenka, A., Sung, G., Chang, C.-L., & Grundmann, M. (2020). *MediaPipe Hands: On-device Real-time Hand Tracking*. http://arxiv.org/abs/2006.10214

Zhang, Y., & Notni, G. (2025). 3D geometric features based real-time American sign language recognition using PointNet and MLP with MediaPipe hand skeleton detection. *Measurement: Sensors*, *38*. https://doi.org/10.1016/j.measen.2024.101697

Zholshiyeva, L. Z., Zhukabayeva, T. K., Turaev, S., Berdiyeva, M. A., & Jambulova, D. T. (2021, Oktober 11). Hand Gesture Recognition Methods and Applications: A Literature Survey. *ACM International Conference Proceeding Series*. https://doi.org/10.1145/3492547.3492578

Zhu, S., & Chollet, F. (2021). *Understanding masking & padding _ TensorFlow Core*.