

PENERAPAN KECERDASAN BUATAN DALAM SISTEM PENGENALAN GERAK TANGAN UNTUK Mendukung Komunikasi pada Penyandang Tunawicara

Rija Ardiansyah¹, Rio Zamroni², Muhammad Azhar³, Kevin Indra Mahesa⁴, Syaiful Fan Reza⁵,
 Yudhistira Arkananta⁶, Eron Richardo Nainggolan⁷, Herlambang Brawijaya⁸

^{1,2,3,4,5,6,7,8} Program Studi Teknologi Informasi, Universitas Bina Sarana Informatika
 Jl. Kamal Raya No 18, Cengkareng – DKI Jakarta, Indonesia

Email: ¹rija493@gmail.com, ²zamronirio12@gmail.com, ³muhammadazhar2935@gmail.com,
⁴kevinindramahesw@gmail.com, ⁵fanrezaa05@gmail.com, ⁶yudhistira.arkananta12@gmail.com,
⁷eron.ekg@bsi.ac.id, ⁸herlambang.hba@bsi.ac.id

ABSTRAK

Perkembangan teknologi kecerdasan buatan (AI) telah mendorong peningkatan signifikan dalam interaksi antara manusia dan mesin, khususnya dalam menyediakan akses komunikasi yang lebih baik bagi individu dengan gangguan bicara. Penelitian ini mengusulkan sebuah sistem penerjemah Bahasa Isyarat Indonesia (BISINDO) berbasis computer vision yang mengintegrasikan deteksi *landmark* tangan secara *real-time* menggunakan MediaPipe dengan proses klasifikasi gestur berbasis *Convolutional Neural Network* (CNN). Sistem ini dirancang untuk mengenali pola gerakan tangan secara dinamis dan mengubahnya menjadi teks atau suara sintesis melalui modul Google Text-to-Speech (gTTS). Pendekatan yang digunakan menggabungkan analisis spasial dan temporal untuk menghasilkan interpretasi gestur yang akurat dan responsif terhadap konteks. Penelitian ini juga mengidentifikasi tantangan implementasi terkait kemampuan generalisasi model terhadap perbedaan pengguna, kondisi pencahayaan, dan lingkungan, serta menawarkan solusi melalui teknik augmentasi data dan optimalisasi arsitektur jaringan saraf. Dengan desain yang fleksibel dan adaptif, sistem ini memiliki potensi besar untuk menjadi dasar pengembangan teknologi bantuan komunikasi inklusif berbasis AI di Indonesia serta mendorong kolaborasi antara bidang visi komputer, bahasa isyarat, dan teknologi Internet of Things (IoT).

Kata kunci: kecerdasan buatan, pengenalan gerakan tangan, tunawicara, MediaPipe, CNN

ABSTRACT

Advances in artificial intelligence (AI) technology have driven significant improvements in human-machine interaction, particularly in enhancing communication accessibility for individuals with speech impairments. This study proposes an Indonesian Sign Language (BISINDO) interpreter system based on computer vision, integrating real-time hand landmark detection using the MediaPipe framework with gesture classification powered by a Convolutional Neural Network (CNN). The system is designed to dynamically recognize hand gesture patterns and convert them into text or synthetic speech through the Google Text-to-Speech (gTTS) module. This approach combines spatial and temporal analysis to produce accurate and contextually responsive gesture interpretations. The study also identifies implementation challenges related to the model's generalizability across different users, lighting conditions, and environments, while offering solutions through data augmentation techniques and neural network architecture optimization. With its flexible and adaptive design, the proposed system has strong potential to serve as a foundation for the development of AI-based inclusive communication technologies in Indonesia and to foster collaboration across the fields of computer vision, sign language, and Internet of Things (IoT) applications.

Keywords: Artificial intelligence, Hand gesture recognition, Speech impairment, MediaPipe, CNN.

1. PENDAHULUAN

Bagi individu yang menghadapi tantangan dalam mendengar dan berbicara, hambatan komunikasi seringkali menjadi kendala yang cukup besar. Salah satu cara penyampaian informasi yang sangat penting dan krusial bagi kelompok ini adalah melalui bahasa isyarat [1]. Namun, bagi penyandang tunawicara, keterbatasan kemampuan berbicara menyebabkan kesulitan dalam mengekspresikan ide dan keinginan. Untuk itu, bahasa isyarat menjadi



media komunikasi utama bagi mereka. Bahasa isyarat sendiri merupakan sistem komunikasi visual yang mengandalkan gerakan tangan, ekspresi wajah, dan gerak tubuh untuk menyampaikan arti tertentu [2].

Di Indonesia, dikenal dua sistem bahasa isyarat utama, yaitu Sistem Isyarat Bahasa Indonesia (SIBI) dan Bahasa Isyarat Indonesia (BISINDO). Penelitian ini mengajukan sistem untuk mengubah abjad BISINDO menjadi teks, dengan tujuan membantu komunikasi bagi individu tunarungu, meskipun ada kendala terkait jumlah *dataset* [3]. Perbedaan sistem ini sering menimbulkan kebingungan bagi masyarakat umum, karena tidak semua orang dapat memahami kedua bentuk bahasa isyarat tersebut [4].

Banyak penelitian telah memanfaatkan kecerdasan buatan, menggunakan pembelajaran mesin, pembelajaran mendalam, dan teknik visi komputer untuk mengidentifikasi sinyal tangan yang dihasilkan oleh jari [5]. Pendekatan awal banyak menggunakan sensor fisik seperti sensor flex untuk membaca lekukan jari, sensor accelerometer dan gyroscope untuk mendeteksi kemiringan tangan, serta modul MP3 dan LCD untuk menampilkan hasil keluaran dalam bentuk suara atau teks [6]. Seiring perkembangan teknologi kecerdasan buatan, metode berbasis pengolahan citra digital dan jaringan saraf tiruan (*Neural Network*) mulai diterapkan untuk meningkatkan akurasi pengenalan bahasa isyarat [7]. Dengan memanfaatkan kamera dan algoritma pengenalan pola, sistem mampu mendeteksi bentuk tangan secara visual tanpa menggunakan perangkat sensor tambahan. Proses ini memungkinkan interaksi yang lebih alami dan efisien dibandingkan sistem berbasis perangkat keras. Salah satu metode yang digunakan dalam penelitian terkini adalah *Convolutional Neural Network* (CNN) karena kemampuannya dalam mengenali pola visual dan citra tangan secara otomatis [8]. CNN mampu melakukan ekstraksi fitur secara mendalam sehingga sistem dapat mengenali berbagai bentuk gestur tangan yang merepresentasikan huruf atau kata. Beberapa penelitian mengembangkan arsitektur CNN dengan berbagai variasi lapisan konvolusi untuk meningkatkan ketepatan deteksi bahasa isyarat. Selain itu, penelitian juga mengadopsi arsitektur GoogLeNet (Inception v3) yang memiliki efisiensi tinggi dalam klasifikasi citra kompleks [9].

Model ini digunakan untuk mengenali pola bahasa isyarat berdasarkan data citra tangan yang diambil dari kamera. Hasilnya menunjukkan tingkat akurasi yang lebih tinggi dibandingkan model konvensional. Pengujian juga dilakukan menggunakan beberapa optimizer seperti Adam, RMSProp, dan SGD untuk menemukan performa pelatihan terbaik dalam mendeteksi gestur tangan. Komunikasi berbasis bahasa isyarat secara umum bertujuan untuk menghubungkan penyandang tunarungu dengan masyarakat yang menggunakan bahasa verbal [10]. Dalam konteks teknologi, penerjemahan otomatis bahasa isyarat menjadi sarana penting yang dapat meminimalisir kesalahpahaman dan membantu integrasi sosial penyandang disabilitas wicara. Sistem seperti ini juga memiliki potensi besar untuk digunakan dalam pendidikan inklusif maupun layanan publik. Keberadaan teknologi penerjemah bahasa isyarat akan sangat membantu komunikasi penyandang tuna wicara untuk berinteraksi dengan masyarakat pengguna bahasa verbal yang umumnya tidak mengerti makna bahasa isyarat [11].

Oleh karena itu, sasaran dari penelitian ini adalah untuk menyelesaikan batasan tersebut dengan menggabungkan model deep learning yang telah diciptakan ke dalam sebuah platform pengenalan bahasa isyarat yang berbasis web, yang dirancang dengan fokus pada elemen praktis dan kenyamanan penggunaan [12]. Dengan dukungan sistem ini, komunikasi dua arah antara penyandang tunawicara dan masyarakat umum dapat dilakukan dengan lebih mudah dan cepat. Bahasa sebagai alat komunikasi memiliki fungsi penting dalam mengekspresikan pesan melalui bentuk verbal maupun nonverbal, termasuk gerak tubuh dan ekspresi wajah [13].

Beberapa penelitian sebelumnya memang membahas tentang pengenalan gestur tangan menggunakan machine learning, menggunakan metode CNN, dan dikombinasikan dengan mediapipe, seperti yang dilakukan oleh [1][2]. Namun, penelitian tersebut menurut kami masih kurang lengkap dan memiliki keterbatasan pada aspek dataset, skenario pengujian, maupun perhitungan TP/FP, dan dibuatnya sebuah table Confusion Matriks. Oleh karena itu, penelitian ini difokuskan pada aspek pengujian dataset yang ditambahkan secara primer, mengikuti beberapa *dataset* yang tersedia di kaggle berupa *dataset* BISINDO, lalu kami uji kembali *dataset* tersebut dan di *train* secara mandiri sehingga memunculkan output perhitungan True Positive Rate dan True Negative Rate pada data *confusion* matriks yang tersedia.

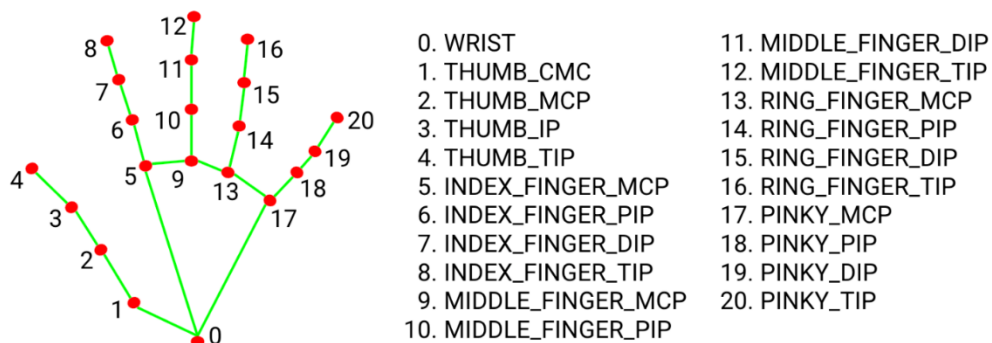
2. METODE PENELITIAN

Pendekatan Penelitian

Penelitian ini dikategorikan sebagai penelitian terapan dan menggunakan pendekatan kuantitatif dengan metode eksperimen. Proses penelitian berlangsung dalam beberapa langkah yang menitikberatkan pada pengenalan bahasa isyarat melalui *Convolutional Neural Network* dan MediaPipe. Keduanya berfungsi untuk mengevaluasi kinerja sistem penerjemah Bahasa Isyarat Indonesia saat menerjemahkan gerakan tangan ke dalam teks dan suara secara langsung menggunakan *Google Text to Speech* sehingga suara yang dikeluarkan menjadi jelas. Metode eksperimen dipilih karena penelitian ini melibatkan pengujian model kecerdasan buatan dengan mengukur variabel kuantitatif seperti akurasi klasifikasi, sehingga tingkat akurasi dapat terlihat ketika program yang telah diuji dijalankan secara langsung.

Mekanisme MediaPipe Hands

Selanjutnya, MediaPipe adalah sebuah *framework* sumber terbuka yang dibuat oleh Google untuk mengembangkan proses pengolahan data sensor, terutama data video dan gambar. Dalam hal pengenalan gerakan tangan, MediaPipe Hands berperan penting dalam mendeteksi dan melacak titik-titik *landmark* tangan secara langsung. Proses tersebut dimulai dengan model deteksi telapak tangan yang menentukan area tangan, lalu dilanjutkan dengan model landmark tangan yang mengambil koordinat tiga dimensi dari 21 titik penting pada tangan. Output dari proses ini sangat berguna untuk mengidentifikasi posisi dan gerakan tangan dengan tingkat akurasi yang tinggi [14], 21 titik *landmark* dapat dilihat pada Gambar 1.



Gambar 1. Hand Landmark Mediapipe Model

Sumber:

<https://mediapipe.readthedocs.io/en/latest/solutions/hands.html>

Studi Literatur

Penelitian ini menerapkan kajian literatur yang berasal dari berbagai artikel ilmiah dan publikasi, baik itu BISINDO, CNN dan MediaPipe, salah satunya adalah Mediapipe. MediaPipe merupakan sebuah kerangka kerja yang memanfaatkan model pembelajaran mesin dengan tingkat akurasi tinggi untuk mendeteksi dan melacak objek, termasuk titik-titik pada bagian tubuh manusia yang dikenal sebagai penanda. Penanda ini digambarkan sebagai koordinat tiga dimensi yang telah dinormalisasi dalam rentang nilai [0, 1]. Dengan memanfaatkan model-model bawaan yang tersedia di MediaPipe, peneliti dan pengembang dapat menyesuaikan jalur pemrosesan dan mengembangkan sistem sesuai dengan kebutuhan mereka. Salah satu penerapan MediaPipe yang banyak digunakan dalam penelitian saat ini adalah pembuatan dataset bahasa isyarat yang didasarkan pada data penanda tiga dimensi, karena informasi koordinat dari titik tubuh dapat berfungsi sebagai representasi dari gerakan tangan dan tubuh. MediaPipe juga menawarkan API khusus yang mempermudah integrasi dan pengembangan berbagai aplikasi untuk pelacakan pose, gerakan tangan, dan ekspresi wajah secara waktu nyata [15].

Pengumpulan Data

Dataset yang digunakan pada penelitian ini merupakan dataset primer yang dikumpulkan secara mandiri menggunakan kamera/webcam berbentuk open cv di dalam program. Dataset terdiri dari 6 kelas dengan total 1250 data, yang kemudian dibagi menjadi data latih/data uji di masing-masing kelas. Pada kelas aku cinta kamu memiliki 231 *dataset*, pada kelas bertemu memiliki 148 *dataset*, pada kelas hai memiliki 120 *dataset*, pada kelas haloo memiliki 117 *dataset*, pada kelas makan memiliki 211 *dataset*, pada kelas pagi memiliki 423 *dataset*. Sehingga pada output TP/FP menghasilkan nilai 100%, karena data yang dilatih atau diuji lebih dari 100 data perkelas, karena hal itulah tingkat akurasi data dapat tercapai.

Pengembangan Sistem

Sistem ini dibangun dengan menggunakan beberapa komponen utama yang terintegrasi, mendukung proses pengenalan gerakan tangan serta penerjemahan bahasa isyarat secara langsung. Komponen tersebut adalah:

- OpenCV berfungsi dalam pemrosesan citra dan deteksi gerakan. Perpustakaan ini menangani pengambilan gambar dari kamera RGB, konversi warna, penghilangan gangguan, serta penyesuaian ukuran gambar sebelum tahap deteksi fitur. Selain itu, OpenCV juga meningkatkan pengolahan citra agar posisi tangan dan jari dapat terdeteksi pada setiap *frame* video.
- Python dipilih sebagai bahasa pemrograman utama karena banyak pustakanya, mudah untuk dipahami, dan dapat bekerja dengan baik bersama berbagai pustaka kecerdasan buatan seperti TensorFlow, NumPy, dan MediaPipe.
- Kamera RGB berperan sebagai alat input utama yang merekam gambar atau video dari gerakan tangan pengguna secara langsung. Data yang diperoleh menjadi sumber utama untuk deteksi, ekstraksi fitur, dan klasifikasi gerakan.

- MediaPipe adalah kerangka kerja sumber terbuka yang dikembangkan oleh Google, digunakan untuk mendeteksi dan melacak titik-titik (*landmark*) tangan secara waktu nyata. Modul MediaPipe Hands dalam sistem ini mengambil koordinat tiga dimensi (3D) dari 21 titik penting pada tangan, seperti ujung jari, sendi, dan telapak. Data landmark ini selanjutnya digunakan sebagai fitur input untuk model CNN dalam klasifikasi gerakan. MediaPipe memiliki keunggulan dalam kecepatan, efisiensi pemrosesan, dan kemampuannya beroperasi dengan baik pada perangkat yang memiliki sumber daya terbatas.
- *Convolutional Neural Network* (CNN) berfungsi sebagai model klasifikasi utama yang mempelajari pola serta karakteristik gerakan tangan menggunakan data landmark yang dihasilkan oleh MediaPipe. CNN mampu mengenali berbagai bentuk tangan dan gerakannya melalui pelatihan pada dataset gerakan, sehingga dapat mengklasifikasikan jenis bahasa isyarat dengan tepat saat dijalankan di waktu nyata.
- NumPy digunakan untuk pengolahan data numerik dan perhitungan matriks, seperti normalisasi koordinat landmark, pengukuran jarak antar titik tangan, serta pengaturan format data sebelum dimasukkan ke model CNN. NumPy sangat penting untuk memastikan bahwa proses perhitungan berlangsung cepat dan efisien.
- Pandas merupakan sebuah pustaka di Python yang berfungsi untuk mengolah, menganalisis, serta mengubah data dalam format tabel yang terdiri dari baris dan kolom. Dengan Pandas, pengguna dapat dengan mudah melakukan tugas seperti mengimpor data dari berkas (CSV, Excel), membersihkan data yang tidak rapi, menghitung berbagai statistik, serta mempersiapkan data guna analisis atau penerapan machine learning.
- gTTS (*Google Text-to-Speech*) digunakan untuk mengubah hasil klasifikasi gerakan menjadi suara. Setelah sistem mengenali dan menerjemahkan gerakan tangan menjadi teks, gTTS mengonversi teks tersebut menjadi suara dalam bahasa Indonesia, memungkinkan komunikasi dua arah antara pengguna tunawicara dan lawan bicaranya secara langsung dan mudah.

Diagram Alir

Diagram alir penelitian ini menunjukkan urutan terencana yang dilakukan dalam penelitian, mulai dari langkah awal hingga hasil akhir yang berupa klasifikasi data. Penelitian dimulai pada langkah awal, yang menandakan dimulainya seluruh proses penelitian. Kemudian, ada tahap Input Data, yaitu memasukkan data awal berupa gestur tangan yang diperoleh menggunakan perangkat input seperti kamera atau *webcam*. Langkah selanjutnya adalah Pengumpulan Data, di mana gestur-gestur tersebut direkam dan dikelompokkan sesuai dengan kategori yang telah ditentukan. Data yang sudah terkumpul kemudian diproses pada tahap Preprocessing Data, yang bertujuan untuk meningkatkan mutu data sebelum digunakan dalam pelatihan. Proses preprocessing bisa mencakup normalisasi, penyesuaian ukuran data, serta ekstraksi fitur. Setelah melakukan preprocessing, data bergerak ke tahap Pelatihan Data, yaitu melatih model pembelajaran mesin dengan data latih agar sistem bisa mengenali pola dari setiap kategori gestur. Setelah itu, dilanjutkan dengan tahap Akurasi Data, yang bertujuan untuk menentukan apakah hasil pelatihan telah mencapai tingkat akurasi yang diinginkan. Jika akurasi belum memenuhi standar, proses bisa diulang kembali ke tahap pelatihan atau membuat data uji baru pada tahap input data jika ingin mendapatkan akurasi yang tinggi. Namun, apabila akurasi sudah memenuhi kriteria, proses dilanjutkan ke tahap Pengujian Data menggunakan data uji yang sudah ditambahkan di tahap input data, agar pengguna dapat mengetahui akurasi dan validitas data yang telah di uji.

Hasil dari pengujian kemudian dianalisis pada tahap Evaluasi Hasil, dengan cara mengukur performa model menggunakan matrik evaluasi seperti akurasi, precision, recall, dan confusion matrix. Langkah berikutnya adalah Hasil Klasifikasi, yaitu keluaran sistem yang berupa klasifikasi atau pengenalan gestur tangan berdasarkan model yang telah dilatih. Proses penelitian ditutup pada tahap Selesai, yang menandakan bahwa semua rangkaian penelitian telah diselesaikan.



Gambar 2. Diagram Alir Penelitian

3. HASIL DAN PEMBAHASAN

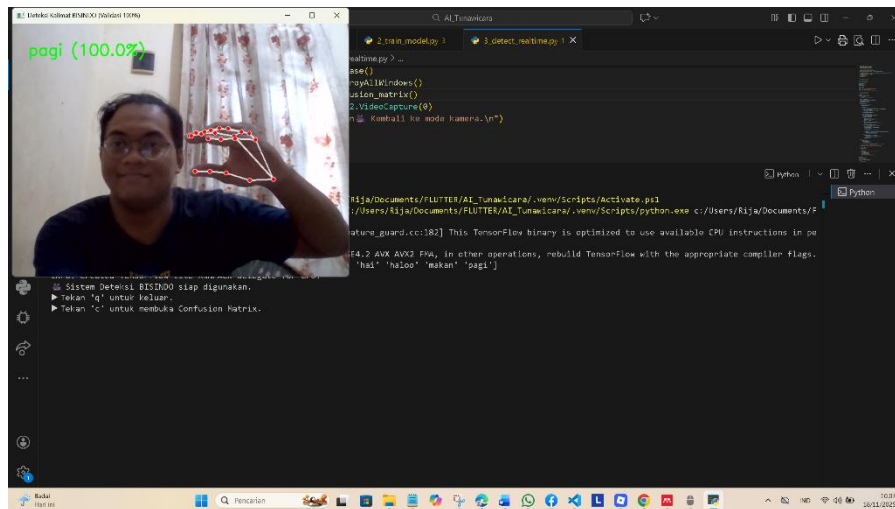
Tahapan implementasi

Pada file `1_collect-dataset_auto.py`, proses dimulai dengan akuisisi dataset yang akan dijadikan landasan untuk melatih model kecerdasan buatan. Pengguna harus terlebih dahulu memasukkan nama gestur BISINDO yang ingin direkam, seperti "halo," "pagi," atau "makan." Setelah menerima input, sistem mengaktifkan kamera dan memanfaatkan MediaPipe Hands untuk mendeteksi posisi tangan secara langsung. MediaPipe menghasilkan 21 titik koordinat tiga dimensi (x, y, z) untuk masing-masing tangan, yang merepresentasikan struktur anatomi jari dan telapak tangan. Setiap kali pengguna melakukan gerakan sesuai dengan gestur yang dipilih, koordinat tersebut dicatat dalam file CSV. Proses ini dilakukan berulang kali untuk mendapatkan variasi data yang cukup, karena perbedaan pose, kondisi pencahayaan, dan variasi gerakan akan meningkatkan kemampuan model untuk melakukan generalisasi di situasi nyata. Tahapan pengumpulan dataset ini sangat penting, karena kualitas data merupakan faktor utama yang mempengaruhi akurasi model di tahap berikutnya.

Dalam file `2_train_model.py`, *dataset* yang sudah dikumpulkan kemudian diproses untuk melatih model klasifikasi. File CSV yang berisi koordinat tangan dibaca dengan library Pandas, lalu dipisahkan menjadi fitur berupa nilai koordinat dan label berupa nama gestur. Teks label tersebut diubah menjadi representasi numerik menggunakan LabelEncoder agar dapat dimengerti oleh algoritma pembelajaran mesin. Data kemudian dibagi menjadi data latih dan data uji untuk memastikan model tidak hanya menghafal data namun dapat memahami pola gestur secara lebih umum. Model jaringan saraf tiruan dibangun dengan TensorFlow menggunakan beberapa lapisan Dense yang dilengkapi dengan fungsi aktivasi ReLU dan Softmax. Selama pelatihan yang berlangsung dalam beberapa epoch, model belajar untuk mengidentifikasi pola antar titik koordinat tangan. Hasil dari pelatihan menunjukkan tingkat akurasi validasi yang tinggi, yang menunjukkan bahwa data yang dikumpulkan cukup representatif dan model mampu membedakan gestur dengan baik. Tahap ini menghasilkan file model berformat `.h5` yang menyimpan bobot hasil pembelajaran dan berfungsi sebagai dasar untuk prediksi di tahap implementasi.

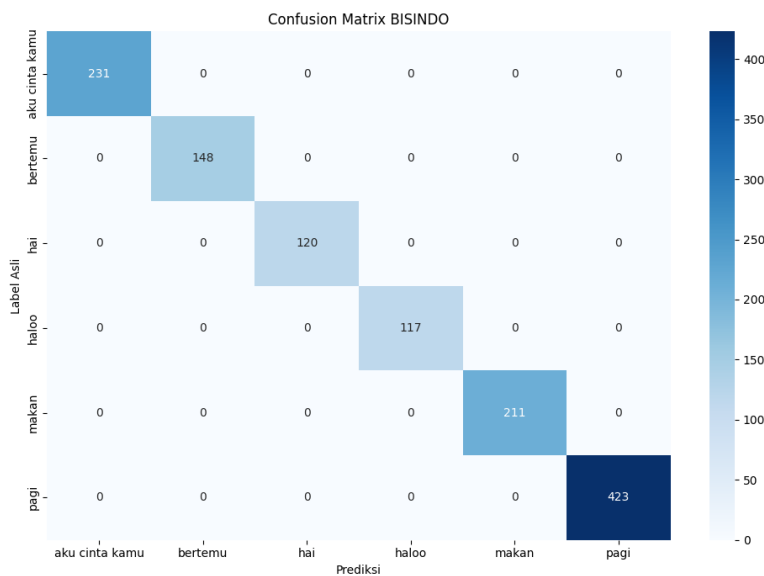
Tahap implementasi dilakukan melalui file `3_detect_realtime.py`, yang mengintegrasikan model yang sudah dilatih dengan sistem deteksi video secara real-time. Saat program berjalan, kamera diaktifkan untuk menampilkan video secara langsung, kemudian MediaPipe digunakan lagi untuk mendeteksi posisi tangan di setiap *frame*. Koordinat tangan yang terdeteksi dikirimkan ke model `.h5` untuk diprediksi sebagai salah satu gestur BISINDO. Hasil dari prediksi ditampilkan dalam format teks di layar, dan jika hasilnya valid, sistem menghasilkan keluaran audio menggunakan Google *Text-to-Speech* sebagai bentuk terjemahan lisan. Jika model tidak dapat menemukan kecocokan pola gerakan dengan data pelatihan, sistem memberikan informasi bahwa gestur tidak dikenali. Tahap ini menunjukkan kemampuan sistem untuk mengubah bahasa isyarat menjadi informasi verbal dengan cepat dan akurat, sehingga dapat membantu komunikasi bagi penyandang tunawicara tanpa perlu penerjemah manual. Pada tahap terakhir, model diuji menggunakan data uji yang telah dipisahkan sebelumnya. Evaluasi dilakukan untuk menilai kemampuan model dalam mengenali pola gestur, menjaga konsistensi prediksi, dan melakukan generalisasi terhadap data baru. Pengujian menggunakan metrik seperti akurasi, presisi, recall, dan tingkat kesalahan untuk memastikan performa model stabil dan dapat diandalkan sebelum diterapkan dalam aktivitas

sehari-hari. Hasil evaluasi menunjukkan bahwa model dapat melakukan klasifikasi dengan efektif, sehingga layak digunakan dalam sistem deteksi real-time untuk mendukung proses penerjemahan BISINDO dengan tepat.



Gambar 3. Hasil *dataset* yang berhasil didaftarkan

Confussion Matrix



Gambar 4. Confussion Matrix BISINDO

Gambar Confussion matriks ini menggambarkan enam kategori sinyal BISINDO yang diuji, yaitu: aku cinta kamu, bertemu, hai, haloo, makan, dan pagi. Pada matriks ini, hanya elemen di diagonal utama yang memiliki nilai, sementara sel-sel lainnya di luar diagonal memiliki nilai 0. Hal ini menunjukkan bahwa model berhasil mengklasifikasikan semua sampel uji dengan akurasi sempurna tanpa kesalahan prediksi.

Penjelasan Setiap Baris (Berdasarkan Kelas)
 Baris = label asli
 Kolom = prediksi dari model

Contoh penjelasan:

Untuk kelas "aku cinta kamu", ada 231 sampel dan semuanya diprediksi sebagai "aku cinta kamu" (nilai di diagonal 231). Tidak ada sampel dari kelas ini yang salah diprediksi ke kelas yang lain.

Demikian juga dengan kelas "bertemu" (148 sampel), "hai" (120 sampel), "haloo" (117 sampel), "makan" (211 sampel), dan "pagi" (423 sampel), semuanya terletak di diagonal tanpa ada yang terprediksi ke kelas lain.

Implikasi Ilmiah Dari Matriks Diagonal Sempurna

Tidak ada kesalahan klasifikasi (error = 0%)



Tidak ada false positif (FP = 0 untuk seluruh kelas)

Tidak ada false negatif (FN = 0 untuk seluruh kelas)

Model menunjukkan kemampuan diskriminasi yang sangat baik, dengan kemampuan untuk membedakan setiap kelas secara sempurna pada data uji yang disediakan.

Perhitungan TP, FP, FN, TN (Per Kelas)

Tabel 1. Kelas TP, FP, FN, TN

Kelas	TP	FP	FN	TN
aku cinta kamu	231	0	0	1.019
bertemu	148	0	0	1.102
hai	120	0	0	1.130
haloo	117	0	0	1.133
makan	211	0	0	1.039
pagi	423	0	0	827

Keterangan:

- TP (True Positive): sampel kelas tertentu yang diprediksi benar
- FP (False Positive): sampel kelas lain yang salah diprediksi sebagai kelas tersebut
- FN (False Negative): sampel kelas tersebut yang salah diprediksi menjadi kelas lain
- TN (True Negative): seluruh sampel selain TP, FP, FN

Pada *dataset* ini, semua kelas memiliki FP=0 dan FN=0, sehingga model tidak pernah melakukan kesalahan baik secara komisi (FP) maupun omisi (FN).

Akurasi

$$Accuracy = \frac{\sum TP}{Total\ Sampel} = \frac{1250}{1250} = 1.0 = 100\% \mid \text{Akurasi} = 100\%$$

True Positive Rate (TPR / Recall)

Rumus:

$$TPR = \frac{TP}{TP + FN}$$

Pada *confusion matrix*, FN = 0 untuk semua kelas, sehingga:

$$TPR = \frac{TP}{TP + 0} = 1.0$$

Jadi untuk setiap kelas:

Kelas	TPR
aku cinta kamu	1.0
bertemu	1.0
hai	1.0
haloo	1.0
makan	1.0
pagi	1.0

Semua kelas memiliki TPR = 1.0 (100%)

True Negative Rate (TNR / Specificity)

Rumus:

$$TNR = \frac{TN}{TN + FP}$$

Karena FP = 0 untuk semua kelas, maka:

$$TNR = \frac{TN}{TN + 0} = 1.0$$

Dengan TN yang sudah dihitung sebelumnya:

Tabel 2. True Negative Rate
(Sumber: Penelitian Mandiri, 2025)

Kelas	TN	FP	TNR
aku cinta kamu	1019	0	1.0
bertemu	1102	0	1.0
hai	1130	0	1.0
haloo	1133	0	1.0
makan	1139	0	1.0
pagi	827	0	1.0

Semua kelas memiliki TNR = 1.0 (100%)

4. SIMPULAN

Penelitian ini berhasil menciptakan alat penerjemah Bahasa Isyarat Indonesia (BISINDO) berbasis kecerdasan buatan, yang menggabungkan teknik visi komputer MediaPipe untuk menentukan titik acuan tangan dan model Jaringan Saraf Tiruan Konvolusional (CNN) untuk mengkategorikan gerakan tangan secara instan. Efektivitas sistem ini dibuktikan melalui penilaian menggunakan matriks konfusi, yang menunjukkan hasil luar biasa dengan akurasi 100%, karena semua kategori yang diuji teridentifikasi secara akurat tanpa menghasilkan hasil positif atau negatif yang salah. Hasil ini menunjukkan kemahiran model dalam mengekstrapolasi informasi uji yang diberikan dan mengidentifikasi gestur BISINDO secara tepat.

Lebih lanjut, dengan integrasi Google Text-to-Speech (gTTS), sistem ini mampu menghasilkan hasil terjemahan yang dapat didengar, sehingga memfasilitasi interaksi timbal balik antara penyandang disabilitas verbal dan masyarakat luas. Pemanfaatan pustaka OpenCV, MediaPipe, TensorFlow, Pandas, dan NumPy memfasilitasi penyederhanaan fase implementasi, memungkinkan identifikasi, kategorisasi, dan penilaian model yang efektif melalui input kamera langsung. Oleh karena itu, penelitian ini menyimpulkan bahwa penggunaan CNN dan MediaPipe menawarkan strategi yang berharga untuk sistem penerjemahan BISINDO yang berbasis visi komputer. Sistem ini memiliki potensi substansial untuk diimplementasikan dalam lingkungan pendidikan inklusif, layanan dukungan komunal, dan alat komunikasi yang mendukung bagi individu yang mengalami hambatan bicara. Meskipun demikian, penelitian ini mengakui adanya kendala terkait variabilitas *dataset*, kerentanan terhadap kondisi cahaya, dan perbedaan antar pengguna, yang semuanya perlu diatasi dalam studi selanjutnya dengan memperluas jangkauan klasifikasi gestur, memperluas cakupan situasi lingkungan, dan mengevaluasi dengan kelompok partisipan yang lebih besar agar sistem dapat menyesuaikan diri dengan keadaan dunia nyata yang lebih beragam dan autentik dalam masyarakat.

DAFTAR PUSTAKA

- [1] I. V. Lemmuela, M. Ayub, and O. Karnalim, "Dynamic Sign Language Recognition in Bahasa using MediaPipe, Long Short-Term Memory, and Convolutional Neural Network," *Journal of Information Systems Engineering and Business Intelligence*, vol. 11, no. 1, pp. 17–29, Feb. 2025, doi: <https://doi.org/10.20473/jisebi.11.1.17-29>.
- [2] O. Dwi Nurhayati, D. Eridani, and M. Hafiz Tsalavin, "Sistem Isyarat Bahasa Indonesia (SIBI) Metode Convolutional Neural Network Sequential Secara Real Time," *Jurnal Teknologi Informasi dan Ilmu Komputer (JTIK)*, vol. 9, No 4, pp. 819–828, Aug. 2022, doi: <https://doi.org/10.25126/jtiik.2022944787>.
- [3] F. Damatraseta, R. Novariyany, and M. A. Ridhani, "Real-time BISINDO Hand Gesture Detection and Recognition with Deep Learning CNN," *Jurnal Informatika Kesatuan*, vol. 1, no. 1, pp. 71–80, Jul. 2021, doi: <https://doi.org/10.37641/jikes.v1i1.774>.
- [4] S. Suki, C. Lubis, and J. Pragantha, "Implementasi Virtual Mouse Berbasis Hand Gesture Recognition dengan Mediapipe dan Convolutional Neural Network under a Creative Commons Attribution-NonCommercial ShareAlike 4.0 International (CC BY-NC-SA 4.0)," *Jurnal INSTEK (Informatika Sains dan Teknologi Informasi)*, vol. 10, no 1, no. 1, pp. 40–51, Apr. 2025, doi: <https://doi.org/10.24252/instek.v10i1.53138>.
- [5] Heri Pratikno, Muhammad Rifki Pratama, Yosefine Triwidayastuti, and Musayyanah, "Pengenalan Gestur Jari Tangan Sebagai Media Pembelajaran Berhitung Bagi PAUD Berbasis Visi Komputer Dan Deep Learning," *Journal of Computer Electronic and Telecommunication*, vol. 4, no. 1, Aug. 2023, doi: <https://doi.org/10.52435/complete.v4i1.355>.
- [6] Muhammad Rizaldi, Rudi Heriansyah, and Nazori Suhandi, "Pengenalan Gerakan Tangan untuk Kontrol Slide Presentasi Menggunakan Framework Mediapipe, OpenCV, dan Model LSTM," *Jurnal Publikasi Teknik Informatika (Jupti)*, vol. 4, no. 3, pp. 98–111, Sep. 2025, doi: <https://doi.org/10.55606/juption.v4i3.5332>.



- [7] S. Nur Budiman, S. Lestanti, S. Marselius Evvandri, and R. Kartika Putri, "Pengenalan Gesture Gerakan Jari untuk Mengontrol Volume di Komputer Menggunakan Library OpenCV dan Mediapipe," *Antivirus : Jurnal Ilmiah Teknik Informatika*, vol. 16, no. 2, pp. 223–232, Nov. 2022, doi: <https://doi.org/10.35457/antivirus.v16i2.2508>.
- [8] A. A. H. A. A. S. C. Sofyan, "Pengenalan Bahasa Isyarat Bahasa Indonesia Real-time Menggunakan Metode SP-Tree," *Jurnal Indonesia Manajemen Informatika dan Komunikasi*, vol. 6, no. 2, pp. 1007–1014, May 2025, doi: <https://doi.org/10.63447/jimik.v6i2.1371>.
- [9] E. U. E. A. D. Tikasni, "Analisis Akurasi Object Detection Menggunakan Tensorflow Untuk Pengenalan Bahasa Isyarat Tangan Menggunakan Metode SSD," *Jurnal Fasilkom (Teknologi Informasi dan Ilmu Komputer)*, vol. 14, no. 2, pp. 385–393, Aug. 2024, doi: <https://doi.org/10.37859/jf.v14i2.7512>.
- [10] W. N. W. G. K. Z. A. Saputra R, "Sistem Klasifikasi Alfabet Bahasa Isyarat Indonesia Menggunakan CNN dengan MobileNetV2 berbasis Android," *JUPITER (Jurnal Penelitian Ilmu dan Teknologi Komputer)*, 2025, doi: <https://doi.org/10.5281/zenodo.14686025>.
- [11] D. Yahya, B. C. Putra, K. Anam, and M. Farkhan, "Alat Penerjemah Bahasa Isyarat Huruf Hijaiyah Menggunakan Sensor Fleksibel dan Mpu 6050 Berbasis Arduino," *BIOS : Jurnal Teknologi Informasi dan Rekayasa Komputer*, vol. 5, no. 1, pp. 46–52, 2023, doi: <https://doi.org/10.37148/bios.v5i1.92>.
- [12] R. M. P. A. Z. M. Libra Kelana E, "Integrating the CNN Model with the Web for Indonesian Sign Language (BISINDO) Recognition," *JAIC (Journal of Applied Informatics and Computing)*, 2025, doi: <https://doi.org/10.30871/jaic.v9i3.9345>.
- [13] I. I. F. F. N. P. G. Q. O. Arifah, "Deteksi Tangan Otomatis Pada Video Percakapan Bahasa Isyarat Indonesia Menggunakan Metode YOLO Dan CNN," *Journal of Applied Informatics and Computing*, vol. 6, no. 2, pp. 171–176, Dec. 2022, doi: <https://doi.org/10.30871/jaic.v6i2.4694>.
- [14] D. R. P. Endang Ratnawati Djuwitaningrum*, "Implementasi MouseVirtualBerbasis Pengenalan Gerakan Tangan MenggunakanMediaPipe dan OpenCV," *Jurnal Ilmu Pengetahuan dan Teknologi (IPTEK)*, vol. 9, no. 2, Aug. 2025, doi: <https://doi.org/10.31543/jii.v9i2.402>.
- [15] I. Suyudi, S. Sudadio, and S. Suherman, "Pengenalan Bahasa Isyarat Indonesia menggunakan Mediapipe dengan Model Random Forest dan Multinomial Logistic Regression," *Jurnal Ilmu Siber dan Teknologi Digital*, vol. 1, no. 1, pp. 65–80, Feb. 2023, doi: <https://doi.org/10.35912/jisted.v1i1.1899>.