



Perancangan Sistem Penerjemah Bahasa Isyarat bagi Tunarungu dan Tunawicara Berbasis Pengolahan Citra Digital dan Text-to-Speech

Nafil Rizq Trianto¹, Alfarizi Wijaya², Arion Pardede³, Daniel Pandiangan⁴,
Hermawan Syahputra⁵

¹⁻⁵Program Studi Ilmu Komputer, Fakultas Matematika dan Ilmu Pengetahuan Alam,
Universitas Negeri Medan

Korespondensi penulis: mrarion2006@gmail.com

Abstract. Communication is an essential human right, yet a significant communication gap persists between individuals with sensory disabilities, specifically the deaf and speech-impaired, and the general public. This barrier primarily stems from a lack of widespread sign language literacy. While numerous technological solutions have been proposed to translate sign language, existing models primarily rely on heavy deep learning architectures such as Convolutional Neural Networks (CNN) or Long Short-Term Memory (LSTM) networks. These models, although highly accurate, demand substantial computational power and often require specific hardware configurations, leading to severe latency and frame-freezing issues during real-time execution on standard devices. To address these critical limitations, this study proposes a lightweight, fast, and highly responsive sign language translation system utilizing a dual-mode approach. The system is specifically designed to recognize static Indonesian Sign Language (BISINDO) alphabets (A-Z) and single-character air writing using a standard webcam. The proposed methodology utilizes MediaPipe for precise hand tracking, where feature extraction is intelligently processed by calculating 63 relative spatial coordinates of fingertips to the wrist (wrist-relative), significantly reducing the dependency on raw camera pixels. Classification is efficiently performed using a Support Vector Machine (SVM) equipped with a Radial Basis Function (RBF) kernel, prioritizing computational efficiency without sacrificing classification accuracy. Based on the evaluation of 3,214 images, the proposed system achieved a classification accuracy of 63.14% with an inference latency of only 0.25 milliseconds per frame, outperforming the Random Forest baseline. To enhance two-way communication, the recognized text is synthesized into audio using Google Text-to-Speech (gTTS), which is executed asynchronously via Daemon Threads to prevent video frame freezing. The findings indicate that the seamless integration of SVM, MediaPipe, OCR PyTesseract, and asynchronous audio playback creates a highly robust, accessible, and inclusive translation tool for standard hardware.

Keywords: Air Writing, Computer Vision, MediaPipe, Multithreading, Sign Language, Support Vector Machine.

Abstrak. Kesenjangan komunikasi antara penyandang tunarungu dan tunawicara dengan masyarakat umum masih menjadi tantangan yang sangat signifikan akibat minimnya pemahaman masyarakat luas terhadap bahasa isyarat. Meskipun berbagai solusi teknologi telah dikembangkan untuk menerjemahkan bahasa isyarat, sebagian besar model yang ada saat ini sangat bergantung pada arsitektur *deep learning* yang berat, seperti *Convolutional Neural Network* (CNN) atau *Long Short-Term Memory* (LSTM). Model-model tersebut menuntut daya komputasi yang masif dan spesifikasi perangkat keras tingkat tinggi, yang seringkali memicu masalah *latency* (jeda waktu) serta penurunan laju bingkai (*frame freeze*) saat diaplikasikan secara *real-time* pada perangkat standar. Untuk mengatasi batasan krusial tersebut, penelitian ini mengusulkan sebuah sistem penerjemah Bahasa Isyarat Indonesia (BISINDO) berbasis *computer vision* yang beroperasi secara *real-time*, ringan, dan responsif menggunakan *webcam* biasa tanpa dukungan unit pemrosesan grafis (GPU). Sistem ini mengusung kebaruan melalui pendekatan *dual-mode*, yakni klasifikasi gestur abjad statis (*Mode Sign*) dan pelacakan lukisan huruf di udara (*Mode Draw*). Pada tahap ekstraksi, modul *MediaPipe* digunakan untuk menghasilkan 63 fitur koordinat spasial yang dihitung relatif terhadap pergelangan tangan (*wrist-relative*). Fitur ini kemudian diklasifikasikan menggunakan algoritma *Support Vector Machine* (SVM) ber-kernel *Radial Basis Function* (RBF) yang dilatih menggunakan *dataset* sekunder tervalidasi sebanyak 3.214 sampel citra. Hasil evaluasi menunjukkan bahwa sistem usulan mencapai tingkat akurasi sebesar 63,14% dengan *latency* inferensi hanya 0,25 milidetik per sampel,

melampaui performa *baseline Random Forest*. Guna menstabilkan akurasi antar-*frame*, sistem dilengkapi dengan metode *hold system* berbasis pewaktu 1,0 detik sebagai protokol anti-duplikasi. Sementara itu, fitur *air writing* diterjemahkan menjadi teks menggunakan *Optical Character Recognition* (PyTesseract). Untuk menyempurnakan inklusivitas komunikasi, teks disintesis menjadi suara menggunakan *Text-to-Speech* (gTTS) yang dieksekusi secara asinkron menggunakan *daemon thread*. Pendekatan ini terbukti mampu mengeliminasi *frame freeze*, menghasilkan solusi teknologi asistif yang sangat efisien dan adaptif bagi perangkat keras konsumen standar.

Kata kunci: *Air Writing, Bahasa Isyarat, Computer Vision, MediaPipe, Multithreading, Support Vector Machine.*

1. LATAR BELAKANG

Komunikasi merupakan hak dasar manusia yang esensial untuk berinteraksi dan berpartisipasi dalam kehidupan sosial. Namun, kesenjangan komunikasi yang signifikan masih terjadi antara penyandang disabilitas sensorik, khususnya tunarungu dan tunawicara, dengan masyarakat umum. Kendala utama yang dihadapi adalah minimnya literasi dan pemahaman masyarakat luas terhadap bahasa isyarat. Di Indonesia, Bahasa Isyarat Indonesia (BISINDO) telah diakui sebagai bahasa alami yang tumbuh dan berkembang di dalam komunitas Tuli. Berbeda dengan bahasa lisan, BISINDO menggunakan modalitas komunikasi visual-gestural. Perbedaan modalitas ini kerap kali menimbulkan hambatan dalam interaksi sehari-hari, sehingga membatasi aksesibilitas dan inklusivitas ruang publik bagi komunitas Tuli.

Seiring dengan pesatnya perkembangan artificial intelligence (AI) dan pengolahan citra digital (computer vision), teknologi komputasi kini memiliki kapabilitas untuk menjembatani kesenjangan komunikasi tersebut (Ardiansyah et al., 2026; Betri, 2026). Mayoritas pendekatan pengenalan bahasa isyarat saat ini memanfaatkan arsitektur Convolutional Neural Network (CNN) karena kemampuannya memproses fitur spasial citra secara hierarkis (Darmatasia, 2021; Sholawati et al., 2022; Therry et al., 2024). Meskipun mampu menghasilkan akurasi klasifikasi yang tinggi, CNN memerlukan beban komputasi yang masif dan sangat sensitif terhadap variasi pencahayaan, sehingga rentan mengalami latensi tinggi pada perangkat konsumen standar tanpa unit pemrosesan grafis (GPU) (Damatraseta et al., 2021; Fadillah et al., 2022). Beberapa peneliti berupaya mengatasi keterbatasan fitur statis dengan beralih ke jaringan Long Short-Term Memory (LSTM) untuk menangkap pola gestur sekuensial (Dewanto et al., 2024; Prayoga 2025). Namun, pendekatan ini membutuhkan volume data temporal yang jauh lebih besar dan waktu inferensi yang lebih panjang, yang justru memperburuk masalah latency pada skenario penggunaan nyata.

Upaya simplifikasi komputasi mulai dilakukan melalui arsitektur transfer learning (Fadillah et al., 2022) dan Feedforward Neural Network (Firdaus et al., 2025). Walaupun menawarkan alternatif yang sedikit lebih ringan, metode tersebut tetap beroperasi pada paradigma deep learning yang membutuhkan dukungan perangkat keras spesifik untuk berjalan secara real-time. Di sisi lain, pendekatan ekstraksi fitur berbasis landmark menggunakan MediaPipe telah terbukti sangat efisien dalam memetakan topologi tangan tanpa harus memproses seluruh piksel citra secara mentah (Budiman et al., 2022; Nur'azizan et al., 2024). Riset komparatif menunjukkan bahwa ketika fitur landmark ini dikombinasikan dengan algoritma machine learning konvensional seperti Support Vector Machine (SVM) atau Random Forest, sistem mampu memberikan klasifikasi yang stabil dan cepat pada CPU standar (Mahendra et al., 2025; Syahputra & Wibowo, 2023; Wardana et al., 2025). Namun, literatur yang ada menunjukkan bahwa sistem-sistem tersebut umumnya belum mengintegrasikan mekanisme kontrol duplikasi input maupun sinkronisasi audio nirkekakuan (non-blocking) dalam satu pipeline terpadu yang siap digunakan oleh pengguna akhir.

Berdasarkan analisis kritis terhadap literatur di atas, terdapat dua celah penelitian (gap analysis) yang signifikan. Pertama, sistem penerjemah bahasa isyarat berbasis deep learning yang ada saat ini belum mampu berjalan secara real-time pada perangkat keras standar tanpa dukungan GPU karena beban komputasi yang tinggi. Kedua, belum ada penelitian yang secara komprehensif mengintegrasikan mekanisme anti-duplikasi berbasis waktu (hold system) dan eksekusi audio asinkron (daemon thread) ke dalam satu sistem penerjemah yang ringan dan responsif. Oleh karena itu, penelitian ini hadir untuk merancang sistem penerjemah bahasa isyarat berbasis SVM dan MediaPipe yang secara khusus mengatasi kedua celah tersebut guna menyediakan solusi komputasi yang terjangkau bagi perangkat keras umum.

Kontribusi ilmiah dari penelitian ini dirumuskan ke dalam empat aspek kebaruan (novelty) utama. Pertama, ekstraksi fitur wrist-relative sistem tidak lagi bergantung pada koordinat piksel absolut kamera, melainkan menghitung posisi setiap landmark secara relatif terhadap pergelangan tangan (wrist), sehingga menghasilkan fitur yang invarian terhadap skala dan posisi tangan di dalam frame. Kedua, arsitektur dual-mode sistem mengintegrasikan dua modalitas input dalam satu pipeline, yaitu mode isyarat statis (A-Z) berbasis SVM dan mode air writing berbasis heuristik geometri serta pindaian Optical

Character Recognition (OCR) (Carlson et al., 2023; Darpito et al., 2025). Ketiga, penerapan hold system sebagai protokol anti-duplikasi berbasis ambang waktu 1 detik untuk mencegah sistem mendaftarkan huruf yang sama secara berulang. Keempat, penggunaan daemon thread nirkekakuan pemutaran audio Google Text-to-Speech (gTTS) dieksekusi secara asinkron pada thread terpisah untuk memastikan aliran video webcam utama terbebas dari masalah frame freeze selama proses sintesis suara berlangsung.

2. KAJIAN TEORITIS

Pengembangan sistem visi komputer untuk pengenalan bahasa isyarat telah bertransformasi dari pemrosesan citra berbasis piksel mentah menuju ekstraksi fitur berbasis topologi tangan. Penggunaan kerangka kerja MediaPipe Hands memungkinkan pelacakan 21 titik koordinat kunci atau landmark secara tiga dimensi dengan efisiensi tinggi (Budiman et al., 2022). Metode ini terbukti lebih unggul dibandingkan arsitektur Deep Learning konvensional karena mampu mereduksi beban komputasi dengan hanya memproses data koordinat numerik daripada seluruh matriks citra (Nur'azizan et al., 2024). Melalui pemanfaatan koordinat spasial tersebut, sistem dapat melakukan kalkulasi jarak relatif guna menghasilkan deskripsi fitur yang presisi sebelum tahap klasifikasi dilakukan (Syahputra et al., 2014).

Algoritma Support Vector Machine (SVM) merupakan mesin klasifikasi yang sangat andal dalam menangani data berdimensi tinggi dengan ruang sampel yang terbatas. Penggunaan kernel Radial Basis Function (RBF) pada SVM memungkinkan pemetaan fitur ke dalam ruang dimensi yang lebih tinggi untuk menemukan bidang pemisah optimal atau hyperplane (Syahputra & Wibowo, 2023). Dalam konteks pengenalan gestur berbasis landmark, SVM terbukti memiliki tingkat akurasi yang kompetitif namun dengan waktu inferensi yang jauh lebih cepat dibandingkan model berbasis saraf tiruan (Wardana et al., 2025). Proses standarisasi data menggunakan StandardScaler sebelum klasifikasi menjadi krusial untuk memastikan distribusi fitur yang seimbang dalam proses pembelajaran model (Dewanto et al., 2024).

Pada aspek interaksi dinamis melalui lukisan jari di udara (air writing), mesin Optical Character Recognition (OCR) Tesseract menyediakan kapabilitas ekstraksi teks yang efisien dari pola visual (Carlson et al., 2023). Implementasi konfigurasi Page Segmentation Mode 10 memungkinkan mesin untuk fokus pada pengenalan karakter tunggal, yang sangat krusial dalam memproses hasil lukisan jari pada kanvas virtual

(Darpito et al., 2025). Integrasi metode ini memberikan fleksibilitas tambahan pada sistem untuk menerjemahkan masukan yang bersifat temporal menjadi unit informasi yang bermakna.

Penyampaian informasi hasil terjemahan melalui luaran suara didukung oleh teknologi Text-to-Speech yang disintesis secara real-time. Penggunaan modul Google Text-to-Speech (gTTS) yang diintegrasikan dengan teknik pemrograman asinkron terbukti mampu meningkatkan aksesibilitas bagi pengguna (Hikmatia & Zul, 2021). Penerapan daemon thread dalam eksekusi audio menjadi solusi teknis untuk mengatasi permasalahan sinkronisasi yang sering menyebabkan gangguan pada aliran bingkai video (frame freeze) saat aplikasi sedang memproses keluaran suara secara simultan (Ardiansyah et al., 2026; Betri, 2026).

3. METODE PENELITIAN

Penelitian ini dirancang menggunakan pendekatan eksperimental melalui tahapan machine learning pipeline yang terstruktur. Alur penelitian dimulai dari pengumpulan data sekunder, tahap prapemrosesan citra, ekstraksi fitur koordinat tangan, pelatihan model klasifikasi, hingga tahap pengujian sistem secara waktu nyata. Seluruh arsitektur sistem dibangun dengan fokus pada efisiensi komputasi agar dapat dioperasikan pada perangkat keras kelas konsumen tanpa memerlukan dukungan unit pemrosesan grafis (GPU).

Data penelitian bersumber dari dataset publik bertajuk BISINDO Indonesian Sign Language Alphabet Image Data yang dipublikasikan melalui repositori Mendeley Data (Sanjaya, 2024). Dataset ini mencakup total 3.214 sampel citra yang mewakili 26 kelas abjad BISINDO (A hingga Z). Sampel tersebut memiliki variasi latar belakang, pencahayaan, dan subjek yang beragam guna melatih tingkat generalisasi model. Untuk proses evaluasi, data dibagi menjadi dua bagian dengan proporsi 80% sebagai data latih (2.571 sampel) dan 20% sebagai data uji (643 sampel). Pembagian ini dilakukan secara acak menggunakan parameter `random_state=42` untuk memastikan hasil pengujian yang konsisten dan dapat direproduksi (reproducibility).

Tahap ekstraksi fitur memanfaatkan kerangka kerja MediaPipe Hands untuk memetakan anatomi tangan menjadi 21 titik koordinat kunci atau landmark dalam ruang tiga dimensi (Budiman et al., 2022). Sistem tidak menggunakan koordinat piksel absolut kamera guna menghindari ketergantungan pada posisi objek di dalam layar. Sebagai

gantinya, sistem menghitung 63 fitur koordinat spasial yang bersifat relatif terhadap pergelangan tangan (wrist-relative). Pada fase prapemrosesan, sistem juga menerapkan normalisasi skala (scale-normalization) berdasarkan jarak Euclidean antara pergelangan tangan dan pangkal jari tengah (Middle Finger MCP). Pendekatan matematis ini memastikan susunan vektor fitur tetap stabil meskipun ukuran tangan pengguna bervariasi atau jarak tangan terhadap sensor kamera berubah (Mulyana et al., 2022).

Klasifikasi abjad isyarat dijalankan menggunakan algoritma Support Vector Machine (SVM) dengan kernel Radial Basis Function (RBF). Parameter regularisasi ditetapkan pada nilai $C = 1.0$ dengan parameter $\gamma = 'scale'$. Algoritma SVM dipilih karena kemampuannya yang sangat baik dalam menangani klasifikasi multikelas pada ruang fitur berdimensi tinggi dengan waktu inferensi yang cepat (Syahputra & Wibowo, 2023; Wardana et al., 2025). Model klasifikasi ini diinkapsulasi dalam sebuah pipeline bersama modul StandardScaler untuk memastikan setiap fitur memiliki rentang nilai yang seragam sebelum proses pembelajaran. Model final diekspor ke dalam format berkas pickle agar dapat dimuat secara instan saat aplikasi dijalankan pada lingkungan waktu nyata.

Sistem yang dibangun mendukung dua modalitas interaksi, yaitu Mode Sign untuk abjad statis dan Mode Draw untuk lukisan jari di udara (air writing). Pada Mode Sign, diimplementasikan protokol anti-duplikasi berupa hold system dengan ambang waktu 1,0 detik untuk mencegah masukan karakter ganda. Pada Mode Draw, sistem melacak koordinat ujung jari telunjuk berdasarkan aturan heuristik geometri dan menerjemahkan hasil lukisan pada kanvas virtual menggunakan mesin Optical Character Recognition (PyTesseract) dengan konfigurasi `--psm 10` untuk pengenalan karakter tunggal (Carlson et al., 2023; Darpito et al., 2025). Luaran hasil terjemahan dikonversi menjadi suara melalui Google Text-to-Speech (gTTS) yang dieksekusi secara asinkron menggunakan daemon thread guna mencegah terjadinya fenomena jendela video yang membeku (frame freeze) (Ardiansyah et al., 2026; Betri, 2026).

Evaluasi kinerja sistem diukur berdasarkan empat metrik klasifikasi utama, yaitu Accuracy, Precision, Recall, dan F1-Score. Selain evaluasi model, dilakukan juga pengujian kinerja waktu nyata untuk mengukur kelancaran sistem melalui metrik Frame per Second (FPS) dan latency inferensi (waktu prediksi per sampel dalam milidetik). Pengujian ini krusial untuk membuktikan bahwa integrasi algoritma ringan dan

manajemen utas konkurensi dapat menghasilkan sistem penerjemah yang responsif dan inklusif pada perangkat keras standar.

4. HASIL DAN PEMBAHASAN

Seluruh rangkaian pengujian dan evaluasi model dijalankan pada lingkungan komputasi standar dengan spesifikasi perangkat keras Dell Latitude 3420 yang mencakup prosesor Intel Core i5-1135G7 (4 core, 8 thread, base clock 2,4 GHz), kapasitas RAM sebesar 8 GB DDR4, kartu grafis terintegrasi Intel Iris Xe Graphics, dan sistem operasi Windows 11 Home. Spesifikasi kelas konsumen ini sengaja dipilih untuk memvalidasi kelayakan operasional sistem di perangkat tanpa dukungan *Graphical Processing Unit* (GPU) diskret. Pada pengujian *baseline*, algoritma *Random Forest* diinisialisasi menggunakan *hyperparameter* standar meliputi $n_estimators=100$, $max_depth = None$, $criterion='gini'$, dan penetapan $random_state=42$.

Evaluasi Kuantitatif Klasifikasi SVM

Pengujian dilakukan terhadap model *Support Vector Machine* (SVM) ber-*kernel Radial Basis Function* (RBF) yang dioptimasi menggunakan *StandardScaler*. Evaluasi pada data uji yang berjumlah 643 sampel (mewakili 20% dari total *dataset*) menghasilkan tingkat akurasi sebesar 63,14%. Rincian performa model mencatatkan nilai *precision* berbobot sebesar 65,48%, *recall* berbobot 63,14%, dan *F1-score* berbobot 61,51%. Rangkuman hasil pengujian metrik klasifikasi disajikan pada Tabel 1.

Tabel 1. Hasil Evaluasi Metrik Model SVM RBF

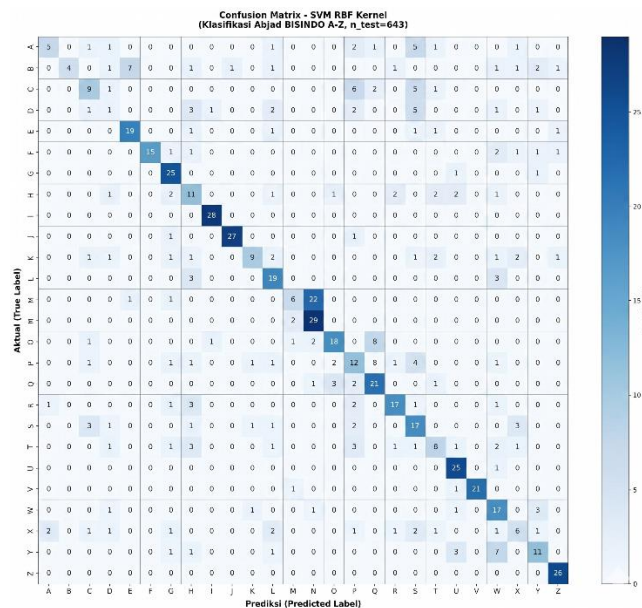
Metrik Pengujian	Nilai Capaian
<i>Accuracy</i>	63,14%
<i>Precision (Weighted)</i>	65,48%
<i>Recall (Weighted)</i>	63,14%
<i>F1-Score (Weighted)</i>	61,51%
<i>Latency Inferensi</i>	0,25 ms/sampel

Guna memvalidasi efisiensi, performa model SVM dibandingkan dengan algoritma *baseline machine learning* konvensional lainnya, yakni *Random Forest*. Tabel 2 menguraikan perbandingan performa antara sistem usulan dan *baseline*.

Tabel 2. Perbandingan Performa Algoritma Klasifikasi

Metode Klasifikasi	Accuracy	F1-Score (Weighted)	Latency Inferensi
SVM RBF (Sistem Usulan)	63,14%	61,51%	0,25 ms
<i>Random Forest (Baseline)</i>	58,63%	57,49%	0,03 ms

Berdasarkan perbandingan pada Tabel 2, model SVM mengungguli *Random Forest* dengan selisih akurasi sebesar 4,51 poin persentase. Hal ini membuktikan bahwa proyeksi *hyperparameter* pada *kernel* RBF lebih efektif dalam memisahkan batas keputusan antar 26 kelas isyarat yang kompleks dibandingkan algoritma pohon keputusan, meskipun *Random Forest* memiliki waktu inferensi yang sedikit lebih cepat. Distribusi prediksi tepat dan kesalahan klasifikasi untuk setiap abjad divisualisasikan melalui *confusion matrix* pada Gambar 1.



Gambar 1. Confusion Matrix Hasil Klasifikasi SVM RBF

Analisis Kinerja Waktu Nyata (Real-Time)

Pengujian kinerja operasional secara waktu nyata menunjukkan bahwa proses inferensi model SVM membutuhkan waktu komputasi rata-rata hanya 0,25 milidetik per bingkai. Metrik ini secara teoritis memungkinkan sistem melakukan hingga 4.000 prediksi per detik. Dalam implementasi menggunakan *webcam* standar yang beroperasi pada 30 *Frame per Second* (FPS), keterbatasan sistem bukan terletak pada inferensi SVM, melainkan pada siklus akuisisi citra dan pelacakan oleh *MediaPipe*. Implementasi arsitektur *daemon thread* terbukti efektif mengisolasi pemutaran audio di latar belakang, sehingga tampilan video tetap berjalan mulus tanpa gangguan jendela yang membeku saat suara sedang diputar. Observasi empiris mengonfirmasi bahwa feed webcam tetap berjalan mulus tanpa indikasi frame freeze selama sintesis suara berlangsung. Tampilan antarmuka sistem saat menjalankan kedua mode utama disajikan pada Gambar 2.



Gambar 2. Tampilan antarmuka pendeteksi bahasa isyarat BISINDO menggunakan webcam secara real-time

Evaluasi Hold System dan Analisis Kegagalan

Mekanisme *hold system* dengan ambang waktu penguncian selama 1 detik diuji secara fungsional untuk mengukur ketahanannya terhadap duplikasi masukan. Hasil pengujian menunjukkan bahwa tanpa *hold system*, satu gestur statis yang ditahan selama 3 detik menghasilkan puluhan registrasi huruf ganda. Dengan aktifnya protokol ini, setiap gestur berdurasi 3 detik hanya meregistrasi tepat satu karakter secara konsisten, yang setara dengan fungsi *debouncing* pada perangkat keras.

Meskipun sistem beroperasi secara responsif, analisis kegagalan terhadap matriks pengujian mengungkap sejumlah batasan. Identifikasi huruf D (7,4%), M (30,0%), dan B (32,0%) mencatatkan performa terendah karena kemiripan geometri titik kunci pada isyarat BISINDO yang nyaris identik dari sudut pandang kamera dua dimensi. Selain itu, akurasi pelacakan tangan menurun pada kondisi pencahayaan ruangan yang minim atau ketika latar belakang memiliki warna yang menyerupai rona kulit pengguna. Pada mode *air writing*, keandalan mesin OCR juga mengalami penurunan apabila pengguna melukis garis dengan kecepatan tinggi yang memicu distorsi kontur karakter.

5. KESIMPULAN DAN SARAN

Kesimpulan

Berdasarkan rangkaian proses perancangan, implementasi, dan evaluasi yang telah dilakukan, dapat disimpulkan bahwa sistem penerjemah bahasa isyarat BISINDO berbasis komputasi ringan berhasil dikembangkan dengan performa yang stabil. Penggunaan algoritma Support Vector Machine (SVM) dengan kernel Radial Basis Function (RBF) yang bersinergi dengan ekstraksi fitur landmark tangan dari MediaPipe terbukti efektif untuk klasifikasi abjad statis secara waktu nyata. Sistem mencapai tingkat

akurasi sebesar 63,14% dengan waktu inferensi yang sangat cepat, yakni 0,25 milidetik per bingkai. Capaian ini membuktikan bahwa metode machine learning konvensional mampu memberikan hasil yang responsif pada perangkat keras standar tanpa memerlukan dukungan unit pemrosesan grafis (GPU).

Implementasi fitur pendukung berupa hold system dengan ambang waktu 1,0 detik terbukti secara fungsional mampu memecahkan permasalahan duplikasi masukan karakter atau spamming yang sering terjadi pada sistem visi komputer. Selain itu, penggunaan arsitektur asinkron melalui daemon thread pada modul audio gTTS berhasil mengeliminasi kendala jendela video yang membeku saat proses sintesis suara berlangsung. Secara keseluruhan, integrasi teknik ekstraksi fitur wrist-relative, klasifikasi SVM, dan manajemen konkurensi ini menghasilkan solusi teknologi asistif yang inklusif dan handal bagi penyandang tunarungu serta tunawicara.

Saran

Meskipun sistem telah berfungsi sesuai dengan tujuan penelitian, terdapat beberapa aspek yang dapat ditingkatkan pada penelitian selanjutnya. Pertama, akurasi klasifikasi sistem masih perlu dioptimalkan melalui penambahan volume dataset atau penerapan teknik augmentasi citra, khususnya untuk kondisi pencahayaan rendah guna meningkatkan stabilitas pelacakan titik kunci oleh MediaPipe. Kedua, pengembangan ke depan dapat mengeksplorasi penggunaan data temporal atau model hibrida untuk meningkatkan kemampuan pemisahan hyperplane pada abjad yang memiliki kemiripan geometri tinggi seperti huruf D dan E. Ketiga, disarankan untuk mengintegrasikan sensor tambahan atau algoritma yang lebih adaptif terhadap variasi latar belakang yang kompleks agar sistem dapat digunakan secara lebih fleksibel di berbagai lingkungan ruang publik. Terakhir, perluasan kosa kata dari sekadar abjad menjadi kata atau kalimat pendek akan sangat meningkatkan nilai praktis sistem ini dalam memfasilitasi komunikasi sehari-hari.

UCAPAN TERIMA KASIH

Penulis mengucapkan puji syukur kepada Tuhan Yang Maha Esa atas kelancaran penelitian ini. Penulis juga menyampaikan terima kasih yang sebesar-besarnya kepada Program Studi Ilmu Komputer, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Negeri Medan atas dukungan fasilitas, lingkungan akademis yang kondusif, serta bimbingan yang diberikan selama proses perancangan sistem hingga penyusunan

artikel ini. Apresiasi juga ditujukan kepada para kontributor *dataset* publik (Mendeley Data) serta pihak-pihak lain yang telah memberikan dukungan moral maupun teknis sehingga penelitian ini dapat diselesaikan dengan baik.

DAFTAR REFERENSI

- Ardiansyah, R., Zamroni, R., Azhar, M., Mahesa, K. I., Reza, S. F., Arkananta, Y., Brawijaya, H. (2026). Penerapan kecerdasan buatan dalam sistem pengenalan gerak tangan untuk mendukung komunikasi pada penyandang tunawicara. *HOAQ: Jurnal Teknologi Informasi*, 17(1), 53-61. <https://doi.org/10.52972/hoaq.vol17no1>
- Betri, T. J. (2026). Gesture to speech transformation: A real-time sign language alphabet recognition system for digital inclusion of the speech impaired community. *CITIZEN: Jurnal Ilmiah Multidisiplin Indonesia*, 6(1), 1-10. <https://doi.org/10.53866/jimi.v6i1.1247>
- Budiman, S. N., Lestanti, S., Evvandri, S. M., & Putri, R. K. (2022). Pengenalan gestur gerakan jari untuk mengontrol volume di komputer menggunakan library OpenCV dan MediaPipe. *ANTIVIRUS: Jurnal Ilmiah Teknik Informatika*, 16(2), 223-232.
- Carlson, J., Bryan, T., & Dell, M. (2023). Efficient OCR for building a diverse digital history. *arXiv preprint arXiv:2304.02737*. <https://doi.org/10.48550/arXiv.2304.02737>
- Darmatasia. (2021). Pengenalan Sistem Isyarat Bahasa Indonesia (SIBI) menggunakan Gradient-Convolutional Neural Network. *JURNAL INSTEK (Informatika Sains dan Teknologi)*, 6(1), 56-65.
- Darpito, M. N., Firdausy, K., & Fadlil, A. (2025). Perbandingan unjuk kerja library Optical Character Recognition (OCR) dalam pengenalan teks pada dokumen digital. *JIP (Jurnal Informatika Polinema)*, 11(3), 273-281.
- Dewanto, F. M., Harjanta, A. T. J., Nada, N. Q., & Herlambang, B. A. (2024). Pengenalan gestur bahasa isyarat Indonesia dengan Mediapipe keypoints. *Jurnal Informatika dan Rekayasa Perangkat Lunak*, 6(2), 427-432.
- Fadillah, R. Z., Irawan, A., & Susanty, M. (2022). Model penerjemah bahasa isyarat Indonesia (BISINDO) menggunakan pendekatan transfer learning. *PETIR: Jurnal Pengkajian dan Penerapan Teknik Informatika*, 15(1), 1-9. <https://doi.org/10.33322/petir.v15i1.1289>
- Damatraseta, F., Novariany, R., & Ridhani, M. A. (2021). Real-time BISINDO Hand Gesture Detection and Recognition with Deep Learning CNN. *Jurnal Informatika Kesatuan*, 1(1), 71-80. <https://doi.org/10.37641/jikes.v1i1.774>
- Firdaus, A. B., Atmojo, F. S., Ananda, N. C., Pramudito, V. G., & Irawan, R. D. (2025). Deteksi bahasa isyarat Indonesia (SIBI) dengan MediaPipe dan Feedforward Neural Network. *Seminar Nasional Teknologi Informasi dan Bisnis (SENATIB) 2025*, 86-93.
- Hikmatia, N. A. E., & Zul, M. I. (2021). Aplikasi penerjemah bahasa isyarat Indonesia menjadi suara berbasis Android menggunakan Tensorflow. *Jurnal Komputer Terapan*, 7(1), 74-83.
- Mahendra, A., Wardana, A. B., Armyanto, J., & Rahman, M. R. D. S. (2025). Klasifikasi gestur BISINDO berbasis landmark tangan-lengan menggunakan Mediapipe Holistic dan Random Forest. *JSITIK: Jurnal Sistem Informasi dan Teknologi Informasi Komputer*, 4(1), 31-39.

- Mulyana, D. I., Lazuardi, M. F., & Yel, M. B. (2022). Deteksi bahasa isyarat dalam pengenalan huruf Hijaiyah dengan metode YOLOV5. *ELKOM: Jurnal Teknik Elektro dan Komputasi*, 4(2), 145-151.
- Nur'azizan, A. H., Ardiansyah, A. R., & Fernandis, R. (2024). Implementasi deteksi bahasa isyarat tangan menggunakan OpenCV dan MediaPipe. *Prosiding Seminar Nasional Teknologi dan Sains*, 3(1), 331-337.
- Prayoga, I. M. P. (2025). Pengenalan kata Kolok secara real time menggunakan Mediapipe dan algoritma Long Short-Term Memory (LSTM). *SINTECH JOURNAL*, 1-102.
- Sholawati, M., Auliasari, K., & Ariwibisono, F. X. (2022). Pengembangan aplikasi pengenalan bahasa isyarat abjad SIBI menggunakan metode Convolutional Neural Network (CNN). *JATI (Jurnal Mahasiswa Teknik Informatika)*, 6(1), 134-144.
- Sanjaya, S. A. (2024). *BISINDO Indonesian Sign Language: Alphabet Image Data. Mendeley Data, Version 1.* <https://doi.org/10.17632/ywnjpbcz8m.1>
- Syahputra, H., Harjoko, A., Wardoyo, R., & Pulungan, R. (2014). Plant recognition using stereo leaf image using gray-level co-occurrence matrix. *Journal of Computer Science*, 10(4), 697-704.
- Syahputra, H., & Wibowo, A. (2023). Comparison of Support Vector Machine (SVM) and Random Forest algorithm for detection of negative content on websites. *Jurnal Ilmiah Teknik Elektro Komputer dan Informatika (JITEKI)*, 9(1), 165-173.
- Therry, R. W. L., Junaidi, A., & Sihananto, A. N. (2024). Program penerjemah bahasa isyarat Indonesia (BISINDO) secara real time menggunakan Convolutional Neural Network dan Mediapipe. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 8(6), 11642-11649.
- Wardana, A. B., Armyanto, J., & Daniati, E. (2025). Perancangan algoritma SVM untuk pengembangan model pendeteksi bahasa isyarat berbasis landmark. *INOTEK*, 9(1), 665-673.