

BAB I

PENDAHULUAN

1.1 Latar Belakang Masalah

Aspek penting dari interaksi sosial antarmanusia adalah komunikasi. Melalui komunikasi, setiap orang dapat menyampaikan pikiran, perasaan, gagasan, dan informasi kepada orang lain, yang memungkinkan mereka untuk membangun hubungan yang sehat dalam kehidupan sosial. Namun demikian, proses komunikasi tidak selalu dapat dilakukan secara verbal. Bagi penyandang tunarungu dan tunawicara, komunikasi verbal seringkali menjadi kendala utama karena adanya keterbatasan dalam mendengar maupun berbicara. Kondisi ini menyebabkan mereka mengandalkan bentuk komunikasi non-verbal untuk berinteraksi dengan lingkungan sekitar (Listiyani et al., 2024)

Salah satu bentuk komunikasi non-verbal yang menjadi media utama bagi komunitas tunarungu adalah bahasa isyarat (*sign language*). Bahasa isyarat adalah sistem komunikasi visual yang menyampaikan makna melalui gerakan tangan, ekspresi wajah, kontak mata, dan postur tubuh. (Huroniyah et al., 2023). Bahasa Isyarat Indonesia (BISINDO) adalah bahasa isyarat yang telah berkembang dan digunakan secara luas di Indonesia. BISINDO membantu orang tuna rungu berkomunikasi satu sama lain dan dengan masyarakat umum, yang tidak menguasai BISINDO. Kehadiran BISINDO menjadi instrumen penting dalam meningkatkan aksesibilitas dan inklusivitas sosial bagi penyandang disabilitas pendengaran.

Walaupun BISINDO memiliki fungsi yang penting dalam komunikasi penyandang Tuli, pemahaman masyarakat umum terhadap bahasa isyarat masih belum merata.. Sebagian besar orang tidak memahami simbol-simbol isyarat dalam BISINDO, sehingga interaksi sehari-hari antara penyandang tunarungu dan masyarakat non-penutur seringkali mengalami hambatan. Kondisi ini menyebabkan penyandang tunarungu menghadapi tantangan dalam mengakses layanan publik, pendidikan, maupun pekerjaan, karena komunikasi sering tidak berjalan efektif (Talisa, 2025). Oleh karena itu, diperlukan teknologi yang dapat menjembatani perbedaan komunikasi ini dan secara

otomatis mengubah isyarat tangan BISINDO menjadi kata-kata atau teks sehingga masyarakat umum dapat memahaminya dengan mudah.

Pengembangan sistem otomatis untuk pengenalan bahasa isyarat semakin meningkat sebagai akibat dari kemajuan dalam teknologi Kecerdasan Buatan (AI), khususnya di bidang *computer vision* dan *deep learning*. Ada sejumlah metode yang dapat digunakan untuk mengidentifikasi gerakan tangan dengan menggunakan gambar atau video. Pendekatan berbasis ekstraksi fitur telah menjadi salah satu solusi yang paling banyak digunakan seiring dengan peningkatan kebutuhan akan sistem yang mampu beroperasi secara real-time dan efisien. Salah satu teknologi yang banyak digunakan dalam proses tersebut adalah MediaPipe karena mampu mendeteksi dan mengekstraksi posisi tangan secara efisien, yang mampu mendeteksi dan mengekstraksi titik-titik penting (*landmark*) pada tangan, wajah, dan pose secara *real-time* dengan efisiensi komputasi yang tinggi. Representasi *landmark* ini memungkinkan data visual diubah menjadi bentuk koordinat numerik yang lebih ringan untuk diproses lebih lanjut oleh model pembelajaran mesin.

Dalam pengenalan bahasa isyarat, tidak hanya bentuk atau posisi tangan yang penting, tetapi juga pergerakannya dari waktu ke waktu, karena gesture bersifat dinamis dan terdiri dari urutan gerakan dalam setiap frame. Oleh karena itu, diperlukan model yang dapat mempelajari pola data berurutan. Untuk menangani data sekuensial, Long Short-Term Memory (LSTM), sebuah komponen Recurrent Neural Network (RNN), dapat mengingat informasi jangka panjang dan menangkap hubungan antar waktu. Oleh karena itu, LSTM sangat cocok untuk pengenalan gesture bahasa isyarat yang memiliki pola gerakan berurutan.

Penelitian yang dilakukan oleh (Sarjianto et al., 2026) mengembangkan sistem penerjemah bahasa isyarat berbasis *computer vision* menggunakan kombinasi *OpenCV*, *Mediapipe*, dan *Long Short-Term Memory* (LSTM). Sistem yang dikembangkan mampu menerjemahkan gesture bahasa isyarat menjadi teks secara *real-time* melalui antarmuka berbasis web. Berdasarkan hasil pengujian, model yang digunakan mampu mencapai akurasi sebesar 85% untuk pengenalan kata dasar, namun hanya mencapai 45% pada pengenalan

alfabet. Penelitian tersebut menggunakan Bahasa Isyarat Sistem Indonesia (SIBI), yang memiliki perbedaan dengan Bahasa Isyarat Indonesia (BISINDO) yang digunakan dalam penelitian ini, sehingga masih diperlukan pengembangan lebih lanjut pada konteks bahasa isyarat yang lebih sesuai dengan penggunaan sehari-hari di masyarakat.

Selain itu, penelitian lain yang dilakukan oleh (Uddin & Boletsis, 2025) yang mengembangkan sistem pengenalan bahasa isyarat Norwegia (*Norwegian Sign Language/NSL*) secara *real-time* menggunakan kombinasi *Mediapipe* dan *Long Short-Term Memory* (LSTM). Dalam penelitian tersebut, *Mediapipe* digunakan untuk mengekstraksi fitur spasial berupa *landmark* tangan dan mulut, sedangkan LSTM digunakan untuk memodelkan hubungan temporal antar frame. Sistem yang dikembangkan mampu mencapai akurasi pengujian sebesar 95% dalam mengenali gesture angka 0 hingga 10. Hasil penelitian ini menunjukkan bahwa kombinasi *Mediapipe* dan LSTM efektif dalam menangani variasi gerakan, orientasi, serta kecepatan gesture, meskipun masih terdapat kendala dalam membedakan gesture yang memiliki kemiripan pola serta keterbatasan jumlah dataset yang digunakan.

Pendekatan berbasis *Mediapipe* dan LSTM juga dikembangkan oleh (Ahmad & Nurtanio, 2023). (2023) dalam penelitian berjudul *Dynamic Sign Language Recognition Using Mediapipe Library and Modified LSTM Method*. Penelitian tersebut memanfaatkan *Mediapipe Holistic* untuk mengekstraksi *landmark* pose, tangan, dan wajah yang kemudian diproses menggunakan metode Modified LSTM untuk meningkatkan akurasi pengenalan gesture dinamis. Dataset yang digunakan terdiri dari 20 kata dinamis yang umum digunakan oleh komunitas tunarungu, dengan total 2400 sampel gestur. Hasil evaluasi menunjukkan bahwa model Modified LSTM mampu mencapai akurasi rata-rata sebesar 85% pada pengenalan kata tunggal dan 80% pada pengenalan kalimat secara *real-time*.

Berdasarkan kajian terhadap penelitian terdahulu, dapat diketahui bahwa pendekatan berbasis *Mediapipe* dan LSTM telah menunjukkan performa yang baik dalam pengenalan gestur bahasa isyarat. Namun, sebagian besar penelitian masih berfokus pada tahap pengenalan gesture hingga menghasilkan *output*

berupa teks, serta belum mengintegrasikan sistem secara menyeluruh dalam bentuk sistem berbasis web yang mampu menyusun hasil pengenalan menjadi kalimat yang utuh dan menyampaikannya dalam bentuk suara secara *real-time*. Selain itu, penelitian terkait pengenalan gestur dinamis, khususnya dalam konteks Bahasa Isyarat Indonesia (BISINDO), masih terbatas dan memerlukan pengembangan lebih lanjut agar sistem dapat bekerja secara optimal dalam kondisi nyata.

Oleh karena itu, Studi ini menggunakan *Mediapipe* sebagai metode ekstraksi fitur untuk menghasilkan sistem pengenalan Bahasa Isyarat Indonesia (BISINDO) secara *real-time*. Ini juga menggunakan *Long Short-Term Memory* (LSTM) sebagai model untuk mempelajari pola gerakan secara temporal. Sistem baru tidak hanya dapat mengenali gerakan secara individual, tetapi juga dapat mengubah hasil pengenalan menjadi kata atau kalimat yang bermakna. Selain itu, sistem ini memiliki fitur *Text-to-Speech* (TTS) yang dapat mengubah hasil deteksi menjadi suara dalam waktu nyata..

Dengan adanya sistem ini, diharapkan dapat dibuat solusi yang lebih komprehensif untuk membantu masyarakat umum dan penyandang tunarungu berkomunikasi satu sama lain. Diharapkan bahwa integrasi antara pemodelan temporal, konversi teks ke suara, ekstraksi fitur berbasis landmark, dan pemodelan temporal dapat meningkatkan akurasi sistem dan membantu pengembangan teknologi asistif yang lebih inklusif di Indonesia.

1.2 Rumusan Masalah

Berdasarkan uraian latar belakang tersebut, penelitian ini bertujuan untuk menyelidiki berbagai masalah yang terkait dengan pengembangan sistem pengenalan Bahasa Isyarat Indonesia (BISINDO) secara real-time dengan menggunakan MediaPipe sebagai metode ekstraksi fitur dan Long Short-Term Memory (LSTM) sebagai model klasifikasi. Dengan demikian, masalah penelitian dirumuskan sebagai berikut:

1. Bagaimana proses penggunaan MediaPipe untuk mengekstraksi fitur landmark tangan dan Long Short-Term Memory (LSTM) sebagai model klasifikasi data sekuensial untuk perancangan dan implementasi sistem pengenalan Bahasa Isyarat Indonesia (BISINDO) berbasis real-time?

2. Bagaimana merancang mekanisme penyusunan hasil pengenalan gesture menjadi rangkaian kata yang membentuk kalimat sederhana secara otomatis?
3. Bagaimana mengintegrasikan fitur Text-to-Speech (TTS) untuk mengubah hasil pengenalan menjadi keluaran suara dalam sistem berbasis web secara real-time?
4. Bagaimana performa sistem yang dikembangkan berdasarkan parameter akurasi, *precision*, *recall*, dan kecepatan pemrosesan dalam skenario penggunaan *real-time*?
5. Apa saja faktor-faktor yang mempengaruhi kinerja sistem dalam mengenali gesture BISINDO, seperti variasi pencahayaan, latar belakang, jarak kamera, serta pergerakan tangan?

1.3 Batasan Masalah

Penelitian dibatasi pada beberapa elemen berikut untuk menjaga fokusnya pada tujuan yang telah ditetapkan :

1. Penelitian ini hanya menggunakan BISINDO (Bahasa Isyarat Indonesia) dan tidak mencakup sistem bahasa isyarat lain seperti SIBI atau ASL.
2. Gesture yang dikenali dibatasi pada sejumlah kosakata tertentu sesuai dengan dataset yang disiapkan peneliti, dan tidak mencakup seluruh kosakata BISINDO secara menyeluruh.
3. Penelitian ini tidak membandingkan metode atau algoritma lain, hanya menggunakan *Mediapipe* untuk ekstraksi fitur *landmark* tangan dan *Long Short-Term Memory* (LSTM) sebagai model klasifikasi berbasis data sekuensial.
4. Sistem difokuskan pada pengenalan gesture tangan dan tidak secara eksplisit memodelkan ekspresi wajah, gerakan tubuh penuh, atau komponen *non-manual* lainnya dalam struktur bahasa isyarat.
5. Penyusunan kalimat dilakukan berdasarkan urutan hasil pengenalan gesture secara sekuensial, tanpa melibatkan analisis linguistik mendalam, tata bahasa kompleks, atau pemodelan *Natural Language Processing* (NLP).

6. Pengujian sistem dilakukan dalam kondisi lingkungan yang terkendali dengan variasi tertentu pada pencahayaan, latar belakang, dan jarak kamera, serta tidak mencakup skenario dunia nyata yang sangat ekstrem atau tidak terkontrol.
7. Implementasi sistem dibatasi pada lingkungan komputasi lokal (PC/laptop) dan tidak mencakup optimasi khusus untuk perangkat mobile, embedded system, atau integrasi dengan layanan berbasis cloud.
8. Output sistem difokuskan pada tampilan teks hasil pengenalan gesture, sedangkan fitur *text-to-speech* bersifat tambahan dan tidak menjadi fokus utama dalam evaluasi penelitian.
9. Sistem penelitian ini masih dalam tahap pengembangan dan dimaksudkan untuk bekerja secara *real-time*. Namun, kecepatan dan respons sistem dapat dipengaruhi oleh spesifikasi perangkat yang digunakan, dan belum dilakukan optimasi komprehensif untuk memastikan bahwa sistem akan bekerja dengan optimal dalam berbagai kondisi perangkat.

1.4 Tujuan

Tujuan penelitian ini adalah sebagai berikut berdasarkan rumusan masalah:

1. Mengembangkan dan menerapkan sistem pengenalan Bahasa Isyarat Indonesia (BISINDO), yang bergantung pada *Mediapipe* sebagai metode ekstraksi fitur, dan *Long Short-Term Memory* (LSTM) sebagai model klasifikasi real-time berbasis data sekuensial.
2. Mengembangkan mekanisme penyusunan hasil pengenalan gesture menjadi rangkaian kata yang membentuk kalimat sederhana.
3. Menggunakan fitur Text-to-Speech (TTS) dalam sistem berbasis web untuk mengubah hasil pengenalan menjadi keluaran suara secara real-time.
4. Menganalisis performa sistem berdasarkan parameter akurasi, *precision*, *recall*, serta kecepatan pemrosesan dalam skenario penggunaan *real-time*.
5. Mengidentifikasi pengaruh kondisi lingkungan, seperti variasi pencahayaan, jarak kamera, serta pergerakan tangan / gesture secara *real-time*.

1.5 Manfaat Penelitian

a. Manfaat Teoritis

Penelitian ini diharapkan dapat membantu kemajuan penelitian ilmiah di bidang penglihatan komputer dan pembelajaran mendalam, terutama berkaitan dengan pengembangan sistem pengenalan Bahasa Isyarat Indonesia (BISINDO), yang menggunakan *Mediapipe* sebagai metode ekstraksi fitur berbasis *landmark* dan *Long Short-Term Memory* (LSTM) sebagai model untuk mempelajari pola gerakan secara temporal. Diharapkan juga bahwa temuan penelitian ini akan berguna untuk penelitian yang akan datang yang berfokus pada pengembangan sistem pengenalan gesture secara real-time yang menggunakan teknik ekstraksi fitur dan pemodelan data sekuensial.

b. Manfaat Praktis

1. Bagi pengguna Tuli dan masyarakat umum, sistem ini dapat membantu menjembatani komunikasi secara lebih cepat dan praktis melalui pengenalan gesture BISINDO secara *real-time*, tanpa harus bergantung sepenuhnya pada penerjemah bahasa isyarat dalam situasi tertentu.
2. Sistem ini dapat digunakan oleh lembaga pendidikan dan komunitas difabel untuk mengajar bahasa isyarat dan membantu siswa berkomunikasi dalam kegiatan belajar mengajar dan interaksi sehari-hari.
3. Penelitian ini diharapkan menjadi referensi bagi peneliti dan pengembang teknologi dalam pengembangan teknologi pengenalan gesture berbasis *Mediapipe* dan LSTM, terutama dalam bentuk sistem web dan *real-time*.
4. Diharapkan bahwa penelitian ini akan meningkatkan pengetahuan dan keterampilan penulis tentang proses pengumpulan dataset, ekstraksi fitur menggunakan *Mediapipe*, pelatihan model Long Short-Term Memory (LSTM), evaluasi performa sistem, dan penerapan aplikasi berbasis web dalam bidang komunikasi untuk penyandang tuli.

1.6 Metode Penelitian

Karena tahapan pengembangannya dilakukan secara sistematis dan berurutan sesuai kebutuhan penelitian, penelitian ini menggunakan pendekatan Waterfall sebagai metode pengembangan sistem.

Dalam penelitian ini, tahapan metode Waterfall meliputi:

1. Analisis Kebutuhan

Pada tahap ini, dilakukan identifikasi kebutuhan sistem, yang mencakup data, perangkat keras, dan perangkat lunak. Selain itu, ditentukan apa yang diperlukan sistem untuk berfungsi. Ini termasuk proses pengambilan video secara real-time, penggunaan Mediapipe untuk mengekstraksi fitur berupa landmark tangan, penggunaan Long Short-Term Memory (LSTM) untuk klasifikasi gesture, dan konversi hasil menjadi teks dan suara (Text-to-Speech). Selain itu, penelitian literatur tentang penelitian sebelumnya yang relevan mendukung langkah ini.

2. Desain Sistem

Tahap ini mencakup perancangan arsitektur sistem yang terdiri dari proses input video, ekstraksi fitur menggunakan *Mediapipe*, pemodelan data sekuensial menggunakan LSTM, hingga menghasilkan output berupa teks dan suara. Selain itu, dirancang alur kerja sistem berbasis web yang memungkinkan pengguna mengakses sistem secara *real-time*.

3. Implementasi

Pada titik ini, desain yang telah dibuat digunakan untuk membangun sistem. Ini termasuk proses pengumpulan dan pengolahan dataset gesture, pelatihan model LSTM, integrasi Mediapipe dengan model klasifikasi, dan pengembangan antarmuka sistem berbasis web untuk mendukung proses pengenalan gesture secara real-time.

4. Pengujian

Proses pengujian dilakukan untuk mengevaluasi kinerja sistem secara keseluruhan. Ini dilakukan dengan mengukur akurasi, ketepatan, recall, skor F1, dan kecepatan pemrosesan (FPS) dalam kondisi penggunaan khusus.

5. *Deployment* dan Pemeliharaan

Tahap ini meliputi penyempurnaan sistem berdasarkan hasil pengujian serta evaluasi terhadap kemungkinan pengembangan lebih lanjut. Sistem yang dikembangkan masih berada pada tahap prototype dan dapat dikembangkan lebih lanjut sesuai kebutuhan.

1.7 Sistematika Penulisan

Untuk membuat penyusunan Laporan Kerja Praktik ini lebih mudah, sistem penulisan yang digunakan adalah sebagai berikut:

BAB 1 : PENDAHULUAN

Latar belakang masalah, rumusan dan batasan masalah, tujuan dan keuntungan penelitian, serta metode dan sistematika penulisan yang digunakan sebagai dasar dan arah penelitian dibahas dalam bab ini.

BAB II : TINJAUAN PUSTAKA

Dalam bab ini, kami membahas teori dan konsep seperti Bahasa Isyarat Indonesia (BISINDO), *computer vision*, *machine learning*, *deep learning*, *Mediapipe* sebagai metode ekstraksi fitur, *Long Short-Term Memory* (LSTM) sebagai model klasifikasi sekuensial, dan penelitian sebelumnya yang berkaitan dengan subjek ini.

BAB III : METODOLOGI PENELITIAN

Dalam bab ini, metode penelitian yang digunakan dibahas. Ini termasuk tahapan pengembangan sistem, perancangan sistem, pengumpulan dan pengolahan data, dan perancangan dataset. Selain itu, proses pelatihan dan pengujian model berbasis *Mediapipe* dan LSTM untuk pengenalan gesture BISINDO secara *real-time* juga dijelaskan..

BAB IV : IMPLEMENTASI DAN PENGUJIAN

Bab ini berisi implementasi sistem yang telah dirancang, meliputi proses pelatihan model, integrasi sistem deteksi *real-time*, perangkaian kata menjadi kalimat, integrasi *text-to-speech*, serta hasil pengujian dan evaluasi kinerja sistem berdasarkan metrik yang telah ditentukan.

BAB V : PENUTUP

Bab penutup menyajikan kesimpulan dan rekomendasi untuk seluruh isi laporan, disesuaikan dengan hasil dan diskusi di bab-bab sebelumnya.