

BAB I

PENDAHULUAN

1.1 Latar Belakang

Komunikasi merupakan proses penyampaian dan penerimaan pesan untuk mencapai pemahaman bersama [1]. Dalam kehidupan sehari-hari, masyarakat umumnya menggunakan bahasa lisan berbasis suara, sedangkan komunitas Tuli menggunakan bahasa isyarat sebagai sarana komunikasi utama [2]. Di Indonesia, Bahasa Isyarat Indonesia (BISINDO) digunakan oleh komunitas Tuli dalam berbagai konteks interaksi sehari-hari dan berkembang secara alami sebagai bagian dari budaya Tuli di Indonesia [3], [4]. Komunitas Tuli di Indonesia yang diperkirakan mencakup sekitar 3–4% dari populasi Tuli global pada praktiknya menggunakan BISINDO sebagai bahasa komunikasi utama [5].

Perbedaan cara berkomunikasi antara pengguna bahasa lisan dan bahasa isyarat menimbulkan kendala dalam interaksi sosial [6]. Keterbatasan pemahaman masyarakat umum terhadap BISINDO menyebabkan komunikasi dengan komunitas Tuli belum berjalan secara optimal [7]. Secara visual, BISINDO memiliki karakteristik gestur BISINDO yang dinamis dengan variasi bentuk dan arah gerakan [8]. Salah satu tantangan utama adalah adanya kemiripan gestur BISINDO pada beberapa kosakata yang memiliki makna berbeda [9]. Kemiripan ini umumnya terjadi pada gestur yang memiliki bentuk tangan dan posisi yang serupa, namun hanya dibedakan oleh arah atau dinamika gerakan. Perbedaan ini menyebabkan kesulitan dalam mengidentifikasi dan membedakan gestur secara akurat, sehingga meningkatkan risiko kesalahan dalam proses pengenalan atau klasifikasi gestur, terutama pada sistem berbasis computer vision [9].

Beberapa penelitian terdahulu telah mengembangkan sistem pengenalan BISINDO dengan berbagai pendekatan, namun masing-masing memiliki keterbatasan yang belum sepenuhnya diatasi. Pertama, Saputra dan Hadiwandura (2024) mengusulkan pendekatan *deep learning* berupa *Convolutional Neural Network* (CNN) untuk mengenali 26 huruf dan 10 angka BISINDO, dengan akurasi pengujian sebesar 98,3%, *precision* 98,3%, *recall* 98,4%, dan *F1-score* 99,3% [9]. Meskipun akurasi yang dicapai tinggi, CNN konvensional memiliki keterbatasan mendasar dalam menangani gestur BISINDO yang bersifat dinamis karena arsitekturnya tidak dirancang untuk menangkap informasi temporal antar-*frame*, sehingga kurang mampu membedakan gestur yang hanya berbeda pada pola gerakannya.

Kedua, Maheswara dan Al-Sulthon (2023) mengembangkan sistem pengenalan bahasa isyarat menggunakan *framework MediaPipe* dengan kombinasi *Holistic Pipeline* dan *Recurrent Neural Network* (RNN). Model GRU tanpa *facial key points* MediaPipe pada penelitian tersebut digunakan untuk mengekstraksi titik-titik kunci (landmark) tangan, wajah, dan tubuh secara real-time, yang kemudian dijadikan representasi fitur bagi model RNN mencapai akurasi 99% pada tahap pelatihan, namun mengalami penurunan signifikan menjadi 83% pada pengujian langsung [10], [11]. Kesenjangan akurasi yang besar ini mengindikasikan adanya masalah *overfitting* dan ketidakmampuan model untuk melakukan generalisasi pada kondisi penggunaan nyata. Ketiga, Farhana (2025) mengembangkan sistem deteksi gestur BISINDO menggunakan CNN dengan arsitektur You Only Look Once versi 8 (YOLOv8) yang dilatih pada dataset alfabet BISINDO (A–Z), memperoleh akurasi 89,74% [11]. Namun, penelitian ini masih terbatas pada pengenalan huruf tunggal dan belum menangani gestur dinamis yang membentuk kata atau frasa. Keempat, Ogunlana *et al.* (2025) mengembangkan sistem pengenalan bahasa isyarat berbasis sensor *wearable*. Pendekatan ini dinilai memiliki keterbatasan dari segi biaya, kenyamanan penggunaan, serta ketidakmampuan dalam menangkap gestur wajah dan tubuh secara menyeluruh [12], [13]. Secara keseluruhan, kajian terhadap berbagai penelitian tersebut menunjukkan bahwa sistem pengenalan bahasa isyarat yang ada masih mengalami kesulitan dalam membedakan gestur yang memiliki kemiripan visual, terutama yang hanya dibedakan oleh arah atau pola gerakan [14].

Berdasarkan tinjauan penelitian terdahulu, terdapat tiga celah penelitian (*research gap*) yang belum ditangani secara memadai. Pertama, sebagian besar penelitian masih berfokus pada pengenalan gestur statis, alfabet, atau gestur umum dengan dataset terbatas, tanpa mempertimbangkan kompleksitas gestur dinamis dalam kosakata BISINDO [9], [10], [11]. Kedua, belum ada penelitian yang secara khusus merancang dan mengevaluasi sistem berdasarkan kemampuannya membedakan gestur BISINDO yang memiliki kemiripan visual berdasarkan arah dan dinamika gerakan [14]. Ketiga, implementasi sistem pengenalan BISINDO pada perangkat mobile yang efisien belum banyak dieksplorasi, padahal aksesibilitas perangkat mobile sangat relevan bagi pengguna sehari-hari [10], [15]. Kemiripan pola gerakan merupakan salah satu tantangan utama dalam proses klasifikasi gestur BISINDO berbasis computer vision [14], sehingga diperlukan pendekatan yang mampu memodelkan informasi spasial dan temporal secara bersamaan.

Untuk mengatasi keterbatasan tersebut, penelitian ini mengusulkan integrasi YOLOv8, MediaPipe, dan Long Short-Term Memory (LSTM) dalam sebuah aplikasi mobile berbasis Android. YOLOv8 dipilih karena kemampuannya mendeteksi dan mengklasifikasikan objek dengan akurasi tinggi pada berbagai kondisi pencahayaan dan latar belakang [14], serta menunjukkan performa lebih baik dibandingkan YOLOv5 dengan akurasi 95,6% dalam 50 epoch pelatihan dibandingkan 92,8% dalam 100 epoch [16]. YOLOv8 berfungsi untuk mendeteksi gestur tangan secara spasial pada setiap frame [11], [16]. Hasil deteksi tersebut kemudian diproses menggunakan MediaPipe Hands untuk mengekstraksi titik-titik kunci (*landmark*) pada jari dan telapak tangan, sehingga diperoleh representasi fitur numerik yang lebih ringkas, invarian terhadap variasi latar belakang dan pencahayaan, serta lebih efisien untuk diproses pada perangkat mobile dibandingkan citra piksel penuh [10], [11]. Rangkaian koordinat *landmark* hasil ekstraksi MediaPipe pada setiap frame selanjutnya disusun menjadi sekuens temporal yang menjadi input bagi LSTM. Sedangkan LSTM digunakan untuk memodelkan urutan gerakan secara temporal sehingga sistem dapat memahami pola gerakan dari waktu ke waktu [10], [15]. LSTM mampu menangkap dependensi jangka panjang pada rangkaian gerakan dinamis, sehingga tepat digunakan untuk mengenali pola gestur yang berurutan [10]. Menurut Aljabar dan Suharjito (2020), kombinasi CNN dan LSTM merupakan pendekatan yang efektif untuk pengenalan bahasa isyarat seperti BISINDO [15]. Dengan demikian, kombinasi YOLOv8 sebagai pendeteksi spasial, MediaPipe sebagai pengekstraksi fitur *landmark*, dan LSTM sebagai pemodel temporal diharapkan dapat meningkatkan akurasi pengenalan gestur pengenalan gestur, khususnya pada kosakata yang memiliki kemiripan visual [15], [21]. Pengembangan aplikasi berbasis mobile dipilih karena lebih praktis, efisien, dan fleksibel, serta memungkinkan penerjemahan gestur dilakukan melalui kamera smartphone tanpa memerlukan perangkat tambahan [10], [11], [21].

Berdasarkan uraian permasalahan yang telah dijelaskan, diusulkan penelitian berjudul “Pengembangan Aplikasi Mobile Penerjemah Bahasa Isyarat Indonesia (BISINDO) Menggunakan YOLOv8.” sebagai upaya untuk mengatasi kesulitan dalam membedakan gestur BISINDO yang memiliki kemiripan visual. Sistem yang dikembangkan menggunakan menggunakan YOLOv8 untuk mendeteksi gestur tangan secara spasial, MediaPipe untuk mengekstraksi titik-titik kunci (*landmark*) tangan sebagai representasi fitur, serta LSTM untuk memahami urutan gerakan secara temporal. Dengan pendekatan ini, sistem diharapkan mampu meningkatkan akurasi pengenalan gestur, khususnya pada kosakata yang memiliki

kemiripan gestur BISINDO, serta membantu mempermudah komunikasi antara komunitas Tuli dan masyarakat umum secara lebih efektif dan efisien.

1.2 Rumusan Masalah

Berdasarkan uraian latar belakang yang telah disampaikan, maka rumusan masalah dalam penelitian ini adalah kesulitan dalam membedakan gestur Bahasa Isyarat Indonesia (BISINDO) yang memiliki kemiripan visual, khususnya pada kosakata yang hanya dibedakan oleh arah atau dinamika gerakan, sehingga meningkatkan risiko kesalahan dalam proses klasifikasi.

1.3 Tujuan

Tujuan penelitian ini adalah mengembangkan sistem penerjemah Bahasa Isyarat Indonesia (BISINDO) berbasis *computer vision* menggunakan metode metode YOLOv8, MediaPipe, dan LSTM untuk membedakan gestur yang memiliki kemiripan visual secara akurat.

1.4 Manfaat

Adapun hasil dari pelaksanaan penelitian ini memberikan manfaat dengan rincian sebagai berikut:

1. Memberikan solusi teknologi berupa aplikasi penerjemah Bahasa Isyarat Indonesia (BISINDO) ke teks yang dapat membantu menjembatani komunikasi antara komunitas Tuli dan masyarakat umum.
2. Menjadi referensi dan dasar pengembangan penelitian selanjutnya terkait sistem penerjemah bahasa isyarat berbasis *computer vision*

1.5 Ruang Lingkup

Agar proses penelitian dan pengembangan sistem ini lebih terarah, fokus, dan tidak menyimpang dari tujuan awal yang telah ditetapkan, maka ruang lingkup dan batasan masalah dalam penelitian ini ditetapkan sebagai berikut:

1. Penelitian ini berfokus pada perancangan dan pengembangan sistem penerjemah Bahasa Isyarat Indonesia (BISINDO) berbasis *computer vision* yang mampu membedakan gestur BISINDO dengan kemiripan visual, khususnya kosakata yang hanya dibedakan oleh arah atau dinamika gerakan, menggunakan metode *You Only Look Once* versi 8 (YOLOv8) untuk deteksi gestur tangan secara spasial MediaPipe untuk ekstraksi titik-titik kunci (landmark) tangan sebagai representasi fitur dan *Long Short-Term Memory* (LSTM) untuk pemodelan urutan gerakan secara temporal.

2. Hasil penelitian diwujudkan dalam bentuk aplikasi *mobile* berbasis Android yang diberi nama Jemari Bicara, dirancang untuk menerjemahkan gestur BISINDO menjadi teks melalui kamera *smartphone* tanpa memerlukan perangkat tambahan.
3. Algoritma menerjemahkan gerakan menggunakan algoritma LSTM yang menerima input berupa sekuens landmark tangan hasil ekstraksi MediaPipe pada area yang telah dideteksi oleh YOLOv8, dengan memanfaatkan konversi model ke format TFLite agar dapat berjalan secara efisien pada perangkat *mobile*.
4. Sistem dirancang untuk mendeteksi gestur tangan yang dilakukan oleh satu pengguna tunggal (*single signer*) dengan batasan kondisi pencahayaan yang memadai dan latar belakang visual yang tidak terlalu kompleks.
5. Dataset yang digunakan dalam penelitian ini adalah WL-BISINDO, yaitu dataset video Bahasa Isyarat Indonesia yang terdiri dari 1.600 video *RGB* dengan 32 gestur terisolasi yang dilakukan oleh 5 penutur dari Banten [4], guna merepresentasikan variasi gerakan dalam penggunaan bahasa isyarat.
6. Perancangan antarmuka pengguna (*user interface*) aplikasi dilakukan menggunakan Figma untuk menghasilkan desain *mockup* yang menjadi acuan dalam implementasi tampilan aplikasi.
7. Pengujian sistem dilakukan secara menyeluruh yang mencakup: pengujian fungsionalitas aplikasi menggunakan *Black Box Testing*, evaluasi performa model melalui *Confusion Matrix* (dengan metrik *precision*, *recall*, *mAP*, dan *F1-score*), serta pengukuran kecepatan inferensi dalam satuan *Frames Per Second* (FPS).
8. Validasi gestur mirip dilakukan oleh tim peneliti untuk memastikan kesesuaian antara gestur yang dikenali sistem dengan makna yang dimaksud, sedangkan pengujian penerimaan pengguna (User Acceptance Testing/UAT) melibatkan responden terbatas, yaitu dua orang narasumber dari SLB-B Karya Murni di Medan.
9. Bahasa pemrograman yang digunakan dalam pengembangan sistem ini meliputi *Python* untuk keseluruhan proses pelatihan model (praproses data, pelatihan YOLOv8s, ekstraksi *landmark* menggunakan MediaPipe, dan pelatihan Stacked LSTM pada notebook Google Colab) dan Kotlin serta layanan inferensi *backend* menggunakan *FastAPI*, serta *Dart* dengan *framework Flutter* untuk pengembangan antarmuka aplikasi *mobile* berbasis Android, termasuk integrasi model hasil konversi *TFLite* agar dapat dijalankan secara *on-device* tanpa bergantung pada koneksi internet.

10. Penelitian ini hanya mencakup penerjemahan gestur BISINDO ke dalam bentuk teks dan tidak membahas penerjemahan ke dalam bentuk suara.

