

BAB 1

PENDAHULUAN

1.1 Latar Belakang Masalah

Komunikasi merupakan kebutuhan dasar dalam interaksi sosial. Bagi penyandang disabilitas pendengaran, keterbatasan fungsi pendengaran berdampak pada akses terhadap komunikasi verbal. Data *Long Form Sensus Penduduk 2020* menunjukkan bahwa prevalensi penduduk dengan kesulitan mendengar berat pada usia ≥ 5 tahun di Indonesia mencapai 0,53% [3]. Dalam praktik sosial, komunitas tunarungu menggunakan bahasa isyarat sebagai sarana komunikasi. Penelitian terhadap Gerakan Kesejahteraan Tunarungu Indonesia (Gerkatan) menunjukkan bahwa organisasi tersebut melakukan sosialisasi Bahasa Isyarat Indonesia (BISINDO) kepada masyarakat umum, namun efektivitasnya masih terbatas karena rendahnya minat dan dukungan media [4].

Dalam konteks pendidikan, komunikasi non-verbal seperti gerakan tangan dan ekspresi visual digunakan dalam proses penyampaian pesan pada penyandang rungu-wicara ketika komunikasi verbal tidak optimal [5]. Temuan ini menunjukkan pentingnya komunikasi visual dalam interaksi penyandang disabilitas pendengaran. Secara struktural, BISINDO melibatkan konfigurasi tangan, orientasi, posisi, serta pergerakan yang dapat bersifat statis maupun dinamis [6].

Seiring perkembangan teknologi kecerdasan buatan, pendekatan berbasis *Deep Learning* dan *Computer Vision* menjadi relevan untuk pengembangan sistem pengenalan bahasa isyarat secara otomatis [7]. Berbagai penelitian telah menggunakan arsitektur deteksi objek berbasis *You Only Look Once* (YOLO), termasuk YOLOv8, untuk deteksi gesture secara *real-time* dengan akurasi tinggi [8]. Penelitian terdahulu menunjukkan bahwa YOLO efektif dalam mengenali pola huruf bahasa isyarat pada citra statis [9], dan pendekatan berbasis YOLO dapat diperluas ke skenario pengenalan frasa secara *real-time* [10].

Selain deteksi berbasis citra, beberapa penelitian mengintegrasikan representasi *landmark* tangan untuk memperkaya informasi spasial gesture. Ekstraksi *landmark* menggunakan MediaPipe terbukti efektif merepresentasikan pola tangan BISINDO [11], sementara MediaPipe Holistic juga telah digunakan untuk mengklasifikasikan gesture BISINDO menggunakan Random Forest dengan hasil yang menjanjikan [12]. Namun, pada pendekatan-pendekatan tersebut,

landmark digunakan sebagai fitur numerik yang dipisah dari citra aslinya.

Pendekatan yang berbeda diusulkan dalam penelitian ini, yaitu *keypoint annotation preprocessing*: memvisualisasikan 21 titik *landmark* tangan secara langsung ke atas citra pelatihan sebelum model YOLOv8s dilatih. Dengan cara ini, model mempelajari representasi gabungan antara tekstur tangan dari piksel asli dan geometri rangka tangan dari visualisasi *landmark* secara simultan. Pendekatan ini berbeda dari metode yang menggunakan *landmark* sebagai fitur terpisah [11, 12] maupun yang menggunakan deteksi objek tanpa informasi *landmark* sama sekali [8, 9].

Pendekatan tersebut memiliki tiga keterbatasan. Pertama, koordinat kedalaman (z) yang dihasilkan MediaPipe Hands merupakan estimasi kedalaman relatif terhadap pergelangan tangan dari citra monokuler, bukan hasil pengukuran sensor kedalaman langsung [13], sehingga menjadikannya sebagai fitur klasifikasi utama berisiko menurunkan keandalan model terhadap variasi sudut kamera dan oklusi antar-jari yang umum terjadi pada gesture BISINDO. Kedua, representasi berbasis koordinat semata menghilangkan informasi tekstur dan kontekstual citra, padahal sejumlah pasangan gesture BISINDO memiliki topologi *landmark* yang serupa namun berbeda pada orientasi tangan atau bentuk permukaan jari yang hanya dapat direpresentasikan melalui piksel asli. Ketiga, *classifier* berbasis fitur numerik (mis. Random Forest) tidak menghasilkan lokalisasi objek (*bounding box*), sehingga tetap memerlukan komponen deteksi terpisah apabila sistem hendak menampilkan visualisasi area tangan secara *real-time*.

Pada tahap inferensi *real-time*, MediaPipe berfungsi sebagai *gating mechanism*: YOLOv8s hanya dijalankan ketika tangan terdeteksi dalam *frame*, sehingga beban komputasi berkurang dan *false positive* pada kondisi tanpa tangan dapat diminimalkan. Selain itu, sistem dilengkapi dengan keluaran *text-to-speech* (TTS) berbahasa Indonesia untuk menyuarakan hasil prediksi secara otomatis, menjadikannya sebagai alat bantu komunikasi yang dapat digunakan secara langsung.

Oleh karena itu, penelitian ini mengusulkan sistem pengenalan gesture kata BISINDO secara *real-time* menggunakan pendekatan *keypoint annotation preprocessing* yang mengintegrasikan MediaPipe Hands dan YOLOv8s. Penelitian ini bertujuan untuk mengeksplorasi efektivitas integrasi informasi geometris *landmark* ke dalam citra pelatihan sebagai upaya meningkatkan akurasi pengenalan gesture, sekaligus mengevaluasi performa sistem pada skenario deteksi *real-time* berbasis webcam.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah diuraikan, komunikasi menggunakan Bahasa Isyarat Indonesia (BISINDO) masih menghadapi kendala, khususnya bagi masyarakat yang belum memahami bahasa isyarat. Oleh karena itu, diperlukan suatu metode untuk mengenali gesture kata BISINDO secara real-time. Rumusan masalah dalam penelitian ini dirumuskan sebagai berikut:

- a. Bagaimana penerapan metode YOLOv8 dan MediaPipe dalam pengenalan gesture kata Bahasa Isyarat Indonesia (BISINDO) secara real-time?
- b. Seberapa akurat metode YOLOv8 dan MediaPipe dalam mengenali gesture kata Bahasa Isyarat Indonesia (BISINDO) secara real-time?

1.3 Batasan Permasalahan

Agar penelitian lebih terarah dan tidak meluas dari ruang lingkup yang telah ditentukan, maka batasan permasalahan dalam penelitian ini adalah sebagai berikut:

- a. Gesture yang dikenali terbatas pada 33 kelas BISINDO dalam dataset, mencakup 25 huruf alfabet (A–Z, tanpa Q) dan delapan kata dasar: Halo, Nama, Bagus, Salam Kenal, Saya, Siapa, Terima Kasih, dan Ya.
- b. Sistem hanya mengenali gesture pada tingkat kata/huruf dan tidak mencakup penerjemahan kalimat atau struktur linguistik BISINDO secara kompleks.
- c. Dataset yang digunakan berupa gesture huruf dan kata dasar BISINDO yang diperoleh dari platform Roboflow dengan judul dataset BISINDO Computer Vision Model oleh BICARA PINTAR.
- d. *Keypoint annotation preprocessing* dilakukan menggunakan MediaPipe Hands dalam mode `static_image_mode=True` yang memproses setiap citra secara independen tanpa informasi temporal antara *frame*.
- e. Sistem memproses gesture tangan dan tidak mempertimbangkan ekspresi wajah maupun gerakan tubuh secara menyeluruh.
- f. Input sistem *real-time* berupa video dari webcam dengan kondisi pencahayaan standar ruangan.

1.4 Tujuan Penelitian

Tujuan yang ingin dicapai dalam penelitian ini adalah sebagai berikut:

- a. Menerapkan metode YOLOv8 dan MediaPipe dalam sistem pengenalan gesture kata Bahasa Isyarat Indonesia (BISINDO) secara real-time.
- b. Mengevaluasi tingkat akurasi metode YOLOv8 dan MediaPipe dalam mengenali gesture kata Bahasa Isyarat Indonesia (BISINDO) secara real-time.

1.5 Manfaat Penelitian

Manfaat yang diharapkan dari penelitian ini adalah sebagai berikut:

- a. Memberikan kontribusi dalam pengembangan metode pengenalan bahasa isyarat melalui integrasi deteksi objek dan representasi landmark tangan pada sistem real-time.
- b. Mendukung pengembangan teknologi asistif yang berpotensi membantu komunikasi antara penyandang tunarungu dan masyarakat umum.
- c. Menjadi referensi bagi penelitian selanjutnya terkait sistem pengenalan bahasa isyarat Indonesia.

1.6 Sistematika Penulisan

Berisikan uraian singkat mengenai struktur isi penulisan laporan penelitian, dimulai dari Pendahuluan hingga Simpulan dan Saran.

Sistematika penulisan laporan adalah sebagai berikut:

- a. Bab 1 PENDAHULUAN
Pendahuluan memaparkan latar belakang penelitian yang mendasari pengembangan sistem pengenalan gesture Bahasa Isyarat Indonesia (BISINDO) menggunakan *keypoint annotation preprocessing*. Selain itu dijelaskan rumusan masalah, tujuan penelitian, batasan masalah, manfaat penelitian, serta sistematika penulisan.

b. Bab 2 LANDASAN TEORI

Landasan teori menguraikan konsep-konsep yang mendukung penelitian, meliputi Bahasa Isyarat Indonesia (BISINDO), *gesture recognition*, arsitektur YOLOv8, MediaPipe Hands, *transfer learning*, serta penelitian terdahulu yang relevan sebagai dasar pengembangan penelitian.

c. Bab 3 METODOLOGI PENELITIAN

Metodologi penelitian menjelaskan tahapan penelitian yang disusun dalam bentuk *pipeline* sistem, mencakup pengumpulan dan kurasi dataset, *keypoint annotation preprocessing*, pelatihan model YOLOv8s, evaluasi model, serta perancangan sistem deteksi *real-time*.

d. Bab 4 HASIL DAN DISKUSI

Hasil dan diskusi menyajikan implementasi sistem serta hasil yang mencakup kurasi dataset, proses pelatihan, evaluasi kuantitatif pada *validation set* dan *test set*, analisis per kelas, pengujian sistem *real-time*, dan perbandingan dengan penelitian terdahulu.

e. Bab 5 KESIMPULAN DAN SARAN

Kesimpulan dan saran merangkum temuan utama dari penelitian berdasarkan hasil implementasi dan pengujian, serta menyampaikan saran untuk pengembangan penelitian selanjutnya.

