

BAB I

PENDAHULUAN

1.1. Latar Belakang

Penyandang disabilitas merupakan kelompok masyarakat yang menghadapi keterbatasan dalam aspek fisik, mental, intelektual, maupun sensorik. Kondisi ini mengakibatkan adanya kesulitan dalam berinteraksi dengan lingkungan sekitar. Kondisi tersebut dapat menimbulkan hambatan dalam proses komunikasi dan bersosialisasi, sehingga membuat penyandang disabilitas sering kali mengalami keterbatasan dalam menyampaikan maupun menerima informasi. Salah satu bentuk disabilitas yang sering menimbulkan hambatan komunikasi adalah tuna rungu dan tuna wicara. Tuna rungu merupakan kondisi ketika seseorang mengalami gangguan pendengaran dengan tingkat keparahan yang beragam, sedangkan tuna wicara adalah kondisi ketika seseorang mengalami kesulitan atau ketidakmampuan dalam berbicara. Berdasarkan data yang tercatat hingga tahun 2009, terdapat sebanyak 2.126.000 jiwa penyandang disabilitas di Indonesia. Dari jumlah tersebut, 223.665 jiwa merupakan penyandang tuna rungu, 151.371 jiwa penyandang tuna wicara, dan 73.560 jiwa penyandang tuna rungu wicara. Secara keseluruhan, sekitar 21% dari total penyandang disabilitas di 24 provinsi di Indonesia mengalami disabilitas rungu wicara [1]. Data tersebut menunjukkan bahwa penyandang disabilitas rungu wicara merupakan kelompok yang cukup signifikan jumlahnya dan memerlukan perhatian khusus, terutama dalam aspek komunikasi sehari-hari.

Bahasa isyarat merupakan bentuk bahasa yang memanfaatkan gerakan bibir, tubuh, dan tangan untuk menyampaikan maksud dalam komunikasi. Bagi penyandang tuna rungu dan tuna wicara, bahasa isyarat menjadi sarana komunikasi utama yang memungkinkan mereka menyampaikan pesan dan berinteraksi dalam kehidupan sehari-hari. Dengan adanya bahasa isyarat, individu penyandang disabilitas rungu maupun wicara tetap dapat menjalin hubungan sosial dan berpartisipasi dalam aktivitas sehari-hari layaknya masyarakat pada umumnya. Namun, sebagian besar masyarakat Indonesia belum memahami bahasa isyarat dan cenderung enggan untuk mempelajarinya. Hal ini menimbulkan hambatan yang

signifikan ketika penyandang disabilitas harus berinteraksi dengan masyarakat umum. Kondisi tersebut tidak hanya menciptakan keterbatasan komunikasi, tetapi juga memperlebar kesenjangan dalam berbagai aspek kehidupan, seperti pendidikan, pekerjaan, hingga akses terhadap layanan publik [2]. Hal ini sejalan dengan survei yang menunjukkan bahwa penyandang disabilitas wicara memiliki tingkat partisipasi kerja lebih rendah dibandingkan dengan penyandang disabilitas visual maupun pendengaran, dengan selisih peluang masing-masing sebesar 0,149 poin dan 0,24 poin [3]. Temuan tersebut mengindikasikan bahwa prevalensi disabilitas di Indonesia memberikan dampak signifikan terhadap kesenjangan dalam partisipasi kerja.

Indonesia memiliki dua bentuk bahasa isyarat yang berbeda, yaitu SIBI (Sistem Bahasa Isyarat Indonesia) dan BISINDO (Bahasa Isyarat Indonesia). SIBI merupakan adaptasi dari *American Sign Language* yang banyak diterapkan di lingkungan sekolah luar biasa (SLB) serta ditetapkan sebagai sistem bahasa isyarat resmi di Indonesia [4]. Namun, dalam praktik sehari-hari, komunitas tuli lebih sering menggunakan BISINDO sebagai bahasa alami mereka untuk berkomunikasi, baik antar sesama maupun dengan individu dengar. Sebuah penelitian membahas mengenai penggunaan bahasa isyarat di Indonesia menunjukkan preferensi yang kuat terhadap BISINDO dibandingkan SIBI. Berdasarkan survei preferensi bahasa isyarat, ditemukan bahwa 91% dari 100 responden tuli berusia 16-50 tahun di wilayah Jawa dan Bali lebih memilih BISINDO, dengan alasan bahwa kompleksitas SIBI sering menjadi hambatan dalam berkomunikasi [5]. Meskipun demikian, BISINDO hingga kini belum memperoleh pengakuan formal sebagaimana SIBI. Kompleksitas penggunaan SIBI dianggap menyulitkan oleh komunitas tuli, sehingga berbagai organisasi, salah satunya Rumah Setara Palembang mendorong agar BISINDO diakui sebagai dialek asli penyandang tuli dan dipandang sebagai representasi identitas sejati mereka [6]. Oleh karena itu, BISINDO lebih banyak digunakan dari pada SIBI, sehingga penelitian ini fokus pada sistem penerjemahan BISINDO.

Perkembangan visi komputer awalnya berfokus pada klasifikasi citra yang hanya mengenali apa objeknya tanpa mengetahui posisinya. Namun, kebutuhan

aplikasi dunia nyata menuntut pemahaman posisi objek secara langsung. Hal ini mendorong lahirnya deteksi objek, yang kini lebih unggul karena mampu mengenali kelas sekaligus lokasi [7]. Deteksi objek telah menjadi salah satu fokus utama dalam bidang visi komputer sejak dua dekade terakhir. Pada tahap awal, deteksi objek masih mengandalkan metode *machine learning* konvensional yang bekerja dengan mengekstraksi fitur secara manual dari citra [8]. Meskipun metode tersebut relatif ringan, kinerjanya sangat terbatas karena sangat bergantung pada kondisi pencahayaan, latar belakang, dan sudut objek. Perkembangan signifikan terjadi dengan hadirnya *deep learning* berbasis *Convolutional Neural Network* (CNN) yang memungkinkan sistem untuk secara otomatis mempelajari fitur visual penting dari ribuan bahkan jutaan data pelatihan, tanpa perlu perancangan fitur manual. *Deep learning* terbukti jauh lebih unggul dibandingkan metode konvensional karena mampu mencapai akurasi tinggi, lebih adaptif terhadap variasi lingkungan, dan efektif dalam mendeteksi banyak objek sekaligus [9]. Oleh karena itu, penelitian ini mengembangkan model deteksi objek berbasis *deep learning* karena kemampuannya secara otomatis mempelajari pola visual kompleks langsung dari data.

Berbagai penelitian terdahulu telah berfokus pada pengembangan model deteksi objek untuk menerjemahkan BISINDO dengan memanfaatkan teknik *deep learning*. Sebuah penelitian telah mengembangkan sistem klasifikasi alfabet BISINDO menggunakan arsitektur *MobileNet V2 FPNLite SSD* yang diimplementasikan melalui *TensorFlow Object Detection API*. Sistem ini mampu mengenali 26 kelas huruf dengan performa rata-rata presisi sebesar 0,712 dan *recall* 0,747 serta mendukung pemrosesan *real-time* pada 30 *frames per second* [10]. Sementara itu, algoritma YOLOv8 digunakan untuk mendeteksi BISINDO, dan berhasil menunjukkan kinerja yang unggul dengan akurasi 95,8%, *recall* 97,4%, serta mAP50 mencapai 99,5%. Hasil ini mengindikasikan bahwa model berbasis YOLO memiliki potensi yang sangat baik dalam mendeteksi bahasa isyarat dengan tingkat keandalan yang tinggi [11]. Selanjutnya, optimasi model dilakukan pada model YOLOv5 untuk klasifikasi huruf BISINDO di kondisi lingkungan yang beragam. Model tersebut menunjukkan presisi yang sangat tinggi sebesar 0,9901

dan *recall* mencapai 0,9999, dengan hasil yang tetap stabil dalam pengaturan *indoor* serta hanya mengalami sedikit penurunan kinerja dalam kondisi *outdoor* [12]. Pendekatan berbasis YOLO dan arsitektur lainnya sangat menjanjikan untuk dikembangkan lebih lanjut dalam membangun sistem penerjemah bahasa isyarat secara *real-time*.

Berbagai penelitian terdahulu telah mengembangkan sistem penerjemah bahasa isyarat berbasis aplikasi maupun perangkat *Internet of Things* (IoT) dengan dukungan teknologi kecerdasan buatan. Penelitian terdahulu mengusulkan aplikasi penerjemah BISINDO berbasis *hand landmark recognition* dengan model LSTM yang mencapai akurasi 85% pada 1500 data [13]. Sementara itu, perancangan aplikasi *android* berbasis *TensorFlow Lite* dengan metode CNN menggunakan 1820 citra ASL yang menghasilkan akurasi pelatihan hingga 97%, tetapi menurun menjadi 73% setelah di-*deploy* [2]. Selain itu, pengembangan sarung tangan penerjemah dengan sepuluh sensor *flex* yang terhubung ke mikrokontroler. Pengembangan tersebut menghasilkan luaran suara dengan keberhasilan 86,67% pada abjad A-G, meskipun akurasinya menurun saat terjadi peralihan posisi [14]. Lebih lanjut, sebuah perancangan alat pendeteksi bahasa isyarat berbasis Raspberry Pi dengan metode *Blazepose* mampu mengenali huruf, angka, dan kata dengan akurasi serta kecepatan optimal untuk pengguna terdaftar [15]. Selain itu, pengembangan sistem penerjemah BISINDO berbasis kamera *real-time* dengan CNN-LSTM yang terhubung ke ESP32 untuk menghasilkan keluaran suara, mencapai akurasi 91,4% dengan latensi rata-rata 3,1 detik dan tingkat keberhasilan suara 86,7% [16].

Berdasarkan beberapa penelitian terkini, perangkat IoT secara umum dikategorikan portabel apabila memiliki dimensi yang ringkas, bobot yang ringan, dan dirancang untuk mudah dibawa atau digunakan tanpa instalasi kompleks. Selain itu, perangkat harus mampu beroperasi secara mandiri menggunakan catu daya baterai *rechargeable*, tanpa ketergantungan terhadap sumber listrik eksternal permanen, sehingga dapat digunakan dalam berbagai skenario mobilitas tinggi, baik *indoor* maupun *outdoor* [17]. Suatu perangkat juga disebut rendah biaya atau *low-cost* apabila total biaya komponen utamanya berada di bawah harga pasar,

termasuk modul pemrosesan, sensor, dan sistem *output*, sehingga memungkinkan untuk direplikasi dan diproduksi secara massal oleh institusi pendidikan, komunitas, maupun pengguna umum [18]. Sementara itu, sistem dikategorikan *real-time* apabila mampu memproses dan mengambil data secara berkelanjutan tanpa jeda atau penghentian proses, sehingga aliran informasi berlangsung secara konsisten dan terus-menerus (*streaming*), bukan berdasarkan proses perekaman terlebih dahulu [19]. Ketiga indikator ini menjadi pertimbangan utama dalam perancangan sistem penerjemah BISINDO berbasis IoT agar tidak hanya unggul dari sisi akurasi algoritma, tetapi juga praktis, ekonomis, mudah diimplementasikan, dan responsif dalam konteks penggunaan dunia nyata.

Meskipun berbagai penelitian terdahulu telah berhasil mengimplementasikan sistem penerjemah BISINDO berbasis *deep learning* maupun IoT, masih terdapat sejumlah keterbatasan signifikan yang menjadi celah pengembangan pada penelitian ini. Pertama, minimnya studi komparatif yang secara langsung membandingkan performa antar keluarga YOLO maupun model CNN lainnya menyebabkan pemilihan arsitektur yang benar-benar optimal untuk BISINDO belum dapat ditentukan secara objektif berdasarkan metrik akurasi dan latensi. Kedua, sebagian besar prototipe hanya berfokus pada satu jenis keluaran, yaitu teks saja atau suara saja, sehingga belum menyediakan pengalaman multimedia yang lebih informatif melalui kombinasi keduanya secara simultan. Ketiga, penggunaan perangkat seperti Raspberry Pi atau laptop membuat biaya produksi relatif tinggi dan kurang sesuai untuk implementasi skala besar atau distribusi massal. Keempat, banyak sistem yang dikembangkan masih bergantung pada format aplikasi Android, yang berarti pengguna tetap membutuhkan *smartphone* sebagai perangkat tambahan. Kelima, sebagian besar penelitian hanya mendukung deteksi alfabet dasar atau gestur terbatas, belum mencakup kombinasi alfabet dan kosakata dasar BISINDO yang lebih relevan untuk komunikasi nyata. Terakhir, masih jarang ditemukan rancangan prototipe yang bersifat sepenuhnya portabel, murah, dan mampu bekerja secara *real-time* melalui *streaming* langsung tanpa proses perekaman terlebih dahulu.

Berdasarkan kekurangan tersebut, penelitian ini berfokus pada pembangunan prototipe penerjemah Bahasa Isyarat Indonesia (BISINDO) berbasis *Internet of Things* (IoT) yang mampu bekerja secara langsung (*real-time*) sebagai solusi untuk menjembatani komunikasi antara penyandang tuna rungu/tuna wicara dan masyarakat umum. Prototipe ini dirancang agar mudah diakses dan dapat diproduksi secara massal dengan biaya yang efisien. Secara fungsional, sistem ini memungkinkan masyarakat umum untuk memahami gestur BISINDO tanpa perlu memiliki kemampuan bahasa isyarat, sehingga berpotensi diterapkan pada lingkungan publik seperti pelayanan administrasi, kesehatan, transportasi, hingga pendidikan. Alur kerja sistem dimulai ketika kamera ESP32-CAM menangkap gestur BISINDO secara langsung dalam bentuk *streaming* video, kemudian data tersebut dikirimkan ke server lokal melalui jaringan Wi-Fi tanpa perlu disimpan atau direkam terlebih dahulu. Server menjalankan model YOLO untuk melakukan deteksi objek terhadap gestur tangan dan menerjemahkannya menjadi teks. Hasil deteksi kemudian dikirim kembali ke perangkat dan ditampilkan melalui OLED *Graphic Display* serta dibunyikan secara langsung menggunakan *mini speaker*. Selain itu, informasi yang sama juga ditampilkan pada antarmuka web agar dapat diakses oleh perangkat lain dalam jaringan yang sama.

Inovasi utama dari penelitian ini terletak pada rancangan prototipe yang bersifat portabel, hemat daya, serta mendukung pemrosesan *real-time*. Selain itu, sistem ini dirancang untuk mampu mengenali kombinasi gestur BISINDO dalam bentuk abjad sekaligus kata, sehingga tidak hanya terbatas pada alfabet per huruf, tetapi juga mampu mendukung penerjemahan kosakata dasar agar lebih natural dan komunikatif. Efisiensi biaya dicapai dengan penggunaan ESP32-CAM sebagai alternatif yang jauh lebih murah dibanding Raspberry Pi atau *smartphone*. Portabilitas didukung oleh desain fisik yang ringkas serta penggunaan baterai *rechargeable* sehingga dapat digunakan secara mandiri tanpa ketergantungan pada sumber listrik tetap. Kemampuan *real-time* dalam konteks penelitian ini merujuk pada proses deteksi berbasis *streaming* langsung (*live streaming*) tanpa harus merekam atau menyimpan video terlebih dahulu sebelum dilakukan inferensi. Selain itu, penelitian ini melakukan studi komparatif terhadap model YOLO v8, v9,

v10, 11, dan 26 untuk menentukan model terbaik berdasarkan nilai mAP tertinggi dan waktu inferensi tercepat, kemudian membandingkannya dengan arsitektur *Faster R-CNN* dan SSD sebagai pembanding objektif. Model YOLO12 tidak diikutsertakan dalam penelitian ini karena membutuhkan sumber daya komputasi yang lebih besar dan waktu pelatihan yang lebih lama, sehingga kurang sesuai untuk kebutuhan sistem deteksi *real-time* [20].

1.2. Rumusan Masalah

Rumusan masalah yang menjadi fokus dalam penelitian ini adalah sebagai berikut:

1. Bagaimana menyusun skema pengambilan data primer yang mampu menghasilkan *dataset* gestur BISINDO yang beragam dan representatif untuk pelatihan model deteksi objek?
2. Bagaimana implementasi model deteksi gestur BISINDO berbasis arsitektur YOLO?
3. Bagaimana mengukur performa model YOLO dalam mendeteksi gestur BISINDO berdasarkan akurasi mAP dan latensi inferensi?
4. Bagaimana mengimplementasikan model YOLO terbaik ke dalam sistem IoT berbasis ESP32-CAM dengan *output* berupa tampilan teks dan suara secara *real-time*?

1.3. Batasan Masalah

Batasan masalah dimaksud untuk membuat studi lebih jelas. Batasan studi ini adalah sebagai berikut:

1. Sistem penerjemah bahasa isyarat yang dikembangkan hanya berfokus pada BISINDO, tidak mencakup bahasa isyarat lainnya.
2. Pengambilan data hanya terbatas pada data citra (gambar/visual), tidak mencakup data suara maupun sensor lainnya.
3. Sistem yang dirancang menggunakan server lokal sebagai pusat pemrosesan data dan deteksi objek.

4. Metode deteksi objek yang digunakan terbatas pada keluarga algoritma YOLO(v8, v9, v10, 11, dan 26), *Faster R-CNN*, dan SSD untuk perbandingan performa.
5. *Dataset* yang digunakan merupakan kombinasi dari sumber *open source* dan data primer yang dikumpulkan oleh penulis.
6. Pengujian sistem difokuskan pada aspek akurasi, presisi, kecepatan deteksi, dan *real-time streaming*.

1.4. Tujuan Penelitian

Tujuan penelitian ini menjawab pertanyaan dari rumusan masalah sebelumnya, yakni:

1. Menyusun dan menerapkan skema pengambilan data primer yang mencakup variasi peraga, lokasi, dan pakaian untuk menghasilkan *dataset* BISINDO yang beragam dan representatif bagi pelatihan model deteksi gestur.
2. Mengimplementasikan model deteksi gestur BISINDO berbasis arsitektur YOLO.
3. Mengevaluasi performa model YOLO menggunakan metrik akurasi (mAP) dan latensi inferensi untuk memastikan kemampuan model dalam mendeteksi gestur secara tepat dan cepat.
4. Menerapkan model YOLO terbaik ke dalam prototipe sistem IoT berbasis ESP32-CAM yang menghasilkan *output* deteksi dalam bentuk teks dan suara secara *real-time*.

1.5. Manfaat Penelitian

Diharapkan studi ini mampu memberi manfaat untuk beragam pihak yang membutuhkan, di antaranya:

1. Manfaat Teoritis
Berkontribusi dalam pengembangan ilmu pengetahuan dan teknologi, khususnya di bidang sistem penerjemah bahasa isyarat berbasis IoT dan visi komputer. Memperkaya literatur mengenai penerapan metode deteksi objek YOLO dalam konteks penerjemahan Bahasa Isyarat Indonesia (BISINDO),

sekaligus memperluas kajian tentang pemanfaatan perangkat IoT berbiaya rendah seperti ESP32-CAM dalam mendukung aplikasi kecerdasan buatan. Selain itu, penelitian ini dapat menjadi referensi dalam pengembangan model penerjemah bahasa isyarat yang portabel, hemat energi, dan mampu bekerja secara *real-time*.

2. Manfaat Praktis

Manfaat penelitian yang dapat diperoleh dari hasil penelitian ini adalah sebagai berikut:

a. Bagi penyusun

Penelitian ini memberikan pengalaman langsung dalam merancang dan mengimplementasikan sistem penerjemah bahasa isyarat berbasis IoT dengan metode deteksi objek YOLO. Penulis dapat memperdalam pemahaman mengenai pengolahan citra, *deep learning*, serta pemanfaatan perangkat IoT seperti ESP32-CAM. Selain itu, penelitian ini juga melatih keterampilan dalam mengintegrasikan *hardware*, *software*, dan model AI untuk menghasilkan sistem yang portabel, biaya rendah, dan aplikatif.

b. Bagi pembaca

Hasil penelitian ini dapat menjadi referensi dan sumber pengetahuan bagi pembaca yang ingin mempelajari atau mengembangkan sistem penerjemah bahasa isyarat maupun aplikasi IoT berbasis visi komputer. Pembaca juga dapat menjadikannya sebagai acuan untuk penelitian selanjutnya, khususnya dalam mengeksplorasi penerapan metode YOLO, optimalisasi model deteksi, serta inovasi sistem komunikasi bagi penyandang tunarungu dan tunawicara.

Halaman ini sengaja dikosongkan