

BAB I

PENDAHULUAN

1.1 Latar Belakang

Bahasa merupakan sarana utama yang digunakan oleh manusia untuk berinteraksi dan berkomunikasi dalam kehidupan sehari-hari [1]. Komunikasi dapat dikelompokkan secara umum menjadi dua bentuk, yaitu komunikasi verbal dan nonverbal. Komunikasi verbal dapat didefinisikan sebagai aktivitas penyampaian pesan yang melibatkan penggunaan kata-kata lisan. Sedangkan komunikasi nonverbal dapat didefinisikan sebagai rangkaian ekspresi yang tidak melibatkan penggunaan kata-kata lisan, seperti gestur yang dilakukan melalui gerakan tangan, wajah, atau bagian tubuh lainnya untuk menyampaikan pesan khusus. Contoh komunikasi nonverbal yang paling sering digunakan adalah penggunaan bahasa isyarat [2].

Bahasa isyarat adalah bentuk komunikasi nonverbal yang umumnya dipergunakan oleh individu yang mengalami disabilitas pendengaran (tuna rungu) dan disabilitas bicara (tuna wicara). Di Indonesia, terdapat dua varian utama bahasa isyarat, yakni Sistem Isyarat Bahasa Indonesia (SIBI) yang secara resmi ditetapkan oleh pemerintah Indonesia dan Bahasa Isyarat Indonesia (BISINDO) yang banyak diterapkan oleh kaum disabilitas secara umum [1] [3]. Walau demikian, tidak semua individu dalam masyarakat umum memiliki pemahaman terhadap bahasa isyarat tersebut. Permasalahan ini sering dikenal sebagai *language barrier*. *Language barrier* atau hambatan bahasa merupakan kondisi di mana komunikasi antara dua atau lebih individu terhambat karena adanya perbedaan dalam bahasa, budaya, kebiasaan, dan faktor-faktor lainnya. Dampaknya, kondisi ini dapat menciptakan batasan antara anggota masyarakat umum dan mereka yang berkebutuhan khusus [4] [5].

Berbagai pendekatan penelitian digunakan untuk mencoba memecahkan masalah hambatan bahasa. Mayoritas penelitian tersebut menghasilkan sistem terjemahan bahasa isyarat yang memecahkan masalah hambatan bahasa.

Mayoritas sistem dirancang menggunakan *computer vision* dan *algoritma machine learning* untuk mendeteksi gerakan bahasa isyarat ke dalam bentuk komunikasi verbal berupa tulisan atau suara [1] [4] [6]. Dengan demikian, masyarakat umum dapat dengan lebih mudah memahami makna dari komunikasi menggunakan bahasa isyarat yang berhasil terdeteksi. Meskipun demikian, mencapai tingkat ketepatan atau akurasi optimal dalam menerjemahkan gestur bahasa isyarat masih merupakan tantangan yang kompleks. Hal ini disebabkan oleh kesulitan dalam mengkonfigurasi sistem hingga mencapai tingkat akurasi yang optimal selama pengujian, dimana tidak semua sistem yang telah dirancang dapat mencapai tingkat akurasi yang diharapkan.

Salah satu kendala yang muncul dalam proses menerjemahkan gestur bahasa isyarat adalah tingkat keakuratan atau ketepatan. Pada penelitian [7] menghasilkan aplikasi deteksi beberapa kata pada BISINDO berbasis Android melalui pemanfaatan *library* TensorFlow Lite yang mencapai tingkat akurasi sebesar 77%. Pada penelitian [8] menerapkan model YOLOv5 untuk mengklasifikasikan simbol gestur tangan yang memberikan hasil akurasi sebesar 80%. Pada penelitian [9] mengembangkan sistem klasifikasi alfabet BISINDO dengan menerapkan metode *Few-Shot Learning* dan menghasilkan tingkat akurasi sebesar 85,75%. Penelitian lain [1] melakukan pengembangan sistem deteksi alfabet SIBI dengan memanfaatkan pendekatan *Single Shot Multibox Detector* dengan mengonfigurasi 300 *epochs* dan mencapai nilai akurasi sebesar 91,79%. Penelitian lainnya [2] telah menciptakan suatu sistem pengenalan gestur tangan dengan mengimplementasikan beberapa varian YOLO, mulai dari YOLOv3 hingga YOLOv5, dan mencapai tingkat akurasi tertinggi pada YOLOv5, yaitu 98,47%. Berdasarkan analisis perbandingan penelitian yang telah disajikan di atas, model YOLOv5 dapat dianggap sebagai alternatif yang layak, mempertimbangkan tingkat akurasi yang mencapai 98,47% dalam konteks pengenalan atau deteksi bahasa isyarat pada tingkat kata. Namun demikian, hasil penelitian [10] mengindikasikan bahwa YOLOv8 menunjukkan nilai akurasi dan nilai *loss* yang lebih optimal bila

dibandingkan dengan YOLOv5 dalam konteks deteksi bahasa isyarat. Penemuan ini terkonfirmasi dengan pencapaian akurasi sebesar 96% pada YOLOv8, sedangkan YOLOv5 hanya mencapai nilai akurasi sebesar 93,6%.

YOLOv8 (*You Only Look Once* versi 8) adalah salah satu model *computer vision* yang sangat cocok untuk pekerjaan pendeteksian objek karena memberikan kompromi antara kecepatan dan akurasi secara *real-time* [11]. YOLOv8 dibangun di atas *Convolutional Neural Network* yang kuat, yang biasanya digunakan dalam tugas pemrosesan gambar dan video. Model YOLO berbeda dengan model CNN karena proses pada arsitektur YOLO lebih bersifat *one-stage* atau paralel, sedangkan CNN memerlukan proses pelatihan berulang (*multi-stage*) [11] [12]. Hal ini menjadi keunggulan model YOLO karena pelatihan dan inferensi atau pengujian tidak memakan waktu lama [11]. Peneliti memilih model YOLOv8 karena cukup efektif dalam penerjemahan otomatis bahasa isyarat, dan hasil pelatihannya juga lebih akurat dan cepat dibandingkan metode algoritma *deep learning* yang lainnya, termasuk Model YOLO versi sebelumnya. Berdasarkan penelitian [10] [13] dapat disimpulkan bahwa YOLOv8 menunjukkan performa yang lebih unggul jika dibandingkan dengan YOLOv1 hingga YOLOv5. Serta pada penelitian [14] [15] menyimpulkan bahwa YOLOv5 berhasil memperoleh nilai akurasi yang lebih tinggi bila dibandingkan dengan metode *deep learning* lainnya, seperti MobileNet-SSD, LSTM, dan CNN.

Secara mendalam, hasil penelitian [16] [17] [18] menggambarkan dampak *hyperparameter* terhadap performa model yang telah dilatih. Penelitian [16] menyatakan bahwa pilihan untuk menggunakan fungsi *optimizer Stochastic Gradient Descent* (SGD) terbukti paling optimal. Hal ini terlihat dari peningkatan akurasi mencapai 91% pada nilai *Intersection over Union* (IoU) sebesar 0,5, dibandingkan dengan penggunaan fungsi *optimizer* lain seperti *Adaptive Moment Estimation* (Adam) dan *Adam with Weight Decay Regularization* (AdamW) pada model YOLOv5 [16]. Pada penelitian [17] mengindikasikan bahwa penyesuaian *hyperparameter*, termasuk ukuran *batch* 16, jumlah *anchor boxes* 4, dan pemilihan fungsi *optimizer* Adam,

dapat meningkatkan tingkat akurasi hingga 90,6% dengan nilai *Intersection over Union* (IoU) sebesar 0,5 pada model YOLOv3 [17]. Berikutnya pada penelitian [18] mengemukakan bahwa penyesuaian *hyperparameter*, termasuk jumlah *epochs* maksimal 20, ukuran *batch* 8, *learning rate* $1e^{-5}$, dan *weight decay* 10^{-4} , dapat mengurangi nilai *error rates* hingga 4,18% pada *Convolutional Neural Network* [18]. Oleh karena itu, untuk mendapatkan *weight* (hasil dari proses pelatihan) yang optimal, peneliti menjalankan *hyperparameter tuning* pada model YOLOv8.

Pemanfaatan data secara *real-time* dalam penelitian ini menjadi sangat penting untuk memastikan bahwa sistem penerjemah dapat memberikan respons terhadap input bahasa isyarat yang diterima. Pengolahan data *real-time* memungkinkan deteksi dan penerjemahan bahasa isyarat secara langsung, sehingga interaksi antara pengguna dan sistem menjadi lebih alami dan efisien [7]. Streamlit dipilih sebagai *framework* untuk mengembangkan aplikasi ini karena kemampuannya yang memudahkan pembuatan aplikasi web interaktif dengan cepat dan tanpa memerlukan keahlian pengembangan web yang mendalam. Dengan sintaksis yang sederhana dan integrasi yang kuat dengan ekosistem Python, Streamlit memungkinkan pengembangan sistem yang responsif dan mudah diakses oleh pengguna [19].

Berdasarkan penjelasan tersebut, maka dibuatlah penelitian dengan judul “**Implementasi YOLOv8 untuk Deteksi Objek Bahasa Isyarat secara *Real-Time***”. Penelitian ini akan mengeksplorasi proses pengembangan penerjemah bahasa isyarat menggunakan model YOLOv8 dengan *framework* Streamlit. Dalam penelitian ini akan dilakukan pemilihan *hyperparameter* untuk menghasilkan nilai metrik yang paling optimal pada model YOLOv8. Penelitian ini juga akan memaparkan mengenai hasil kinerja nilai metrik *precision*, *recall*, *F1-Score*, dan *mAP⁵⁰* dari model penerjemah bahasa isyarat yang dikembangkan.

1.2 Perumusan Masalah

Berdasarkan uraian latar belakang, diperoleh rumusan masalah yaitu terkait *language barrier* atau hambatan bahasa dalam komunikasi sehari-hari antara individu tunarungu atau tuna wicara yang menggunakan bahasa isyarat, dan masyarakat umum yang tidak memahaminya.

1.3 Pertanyaan Penelitian

Berdasarkan perumusan masalah di atas, muncul pertanyaan penelitian berikut:

1. Bagaimana penerapan model penerjemah bahasa isyarat menggunakan model YOLOv8?
2. Bagaimana kinerja dari nilai metrik *precision*, *recall*, *F1-Score*, dan *mAP⁵⁰* pada model penerjemah bahasa isyarat yang dikembangkan?

1.4 Batasan Masalah

Untuk mewujudkan penelitian yang terfokus pada masalah yang ada, maka terdapat batasan-batasan masalah sebagai berikut:

1. Hasil dari penelitian ini berupa *prototype* web yang mampu menjalankan model pelatihan yang telah dikembangkan.
2. Web hanya mampu mendeteksi dan menerjemahkan 26 abjad A hingga Z dan 10 kata dalam Bahasa Isyarat Indonesia (BISINDO), antara lain ayah, halo, kakak, minum, rumah, sama-sama, sehat, teman, terima kasih, dan tidur.
3. Fokus penelitian terbatas pada penerjemahan bahasa isyarat melalui media gambar dan/atau *real-time* video.
4. Deteksi objek terbatas pada gestur tangan.
5. Model YOLOv8 yang diujicobakan adalah versi YOLOv8m (*Medium*).
6. Objek dalam proses pelatihan menggunakan gambar pribadi (data primer) serta gambar yang diambil dari Kaggle (data sekunder).

7. Objek dalam tahap pengujian menggunakan gambar pribadi, termasuk penggunaan perangkat *webcam* serta gambar yang diambil dari situs internet.

1.5 Tujuan Penelitian

Penelitian ini bertujuan untuk mengembangkan model penerjemah bahasa isyarat menggunakan model YOLOv8 dengan mempertimbangkan pemilihan *hyperparameter* untuk mencapai nilai metrik optimal seperti *precision*, *recall*, *F1-Score*, dan *mAP⁵⁰*.

1.6 Manfaat Penelitian

Manfaat yang dapat diperoleh dari penelitian ini, baik dari segi teoritis maupun praktis, meliputi:

1. Manfaat Teoritis:
 - a. Bagi penulis, dapat menambah pengetahuan mengenai implementasi YOLOv8 dalam deteksi objek bahasa isyarat.
 - b. Bagi pembaca, dapat menambah pemahaman yang lebih mendalam tentang prinsip dan konsep yang mendasari teknologi kecerdasan buatan dalam konteks bahasa isyarat.
2. Manfaat Praktis:
 - a. Bagi penulis, dapat menghasilkan solusi praktis untuk masalah komunikasi masyarakat difabel.
 - b. Bagi pembaca, dapat memberikan pemahaman mendalam tentang potensi teknologi kecerdasan buatan dalam meningkatkan aksesibilitas.
 - c. Bagi masyarakat, dapat memberikan kontribusi positif pada peningkatan layanan masyarakat secara menyeluruh melalui pemanfaatan teknologi kecerdasan buatan.