

BAB I

PENDAHULUAN

Bab ini memaparkan latar belakang, rumusan masalah, tujuan, manfaat, ruang lingkup, serta sistematika penulisan skripsi mengenai analisis perbandingan integrasi citra asli dan landmark tangan pada Sistem Isyarat Bahasa Indonesia (SIBI) menggunakan EfficientNetB0.

1.1 Latar Belakang

Komunikasi merupakan sarana utama bagi manusia untuk berinteraksi, bertukar informasi, serta membangun hubungan sosial yang mendukung kelangsungan hidup. Akan tetapi, tidak semua individu memiliki kemampuan komunikasi verbal dan nonverbal yang memadai, akibat adanya berbagai hambatan. Berbagai hambatan, seperti perbedaan bahasa, gerakan tubuh, ekspresi wajah serta sinyal verbal lainnya, dapat menyebabkan terjadinya kesalahan interpretasi dan kesalahpahaman dalam berkomunikasi (Abu-Arqoub dan Alserhan, 2020).

Salah satu kelompok yang secara nyata menghadapi hambatan dalam berkomunikasi adalah penyandang disabilitas pendengaran dan wicara, yang sering disebut sebagai komunitas tuli (*Deaf Community*). Berdasarkan data dari Organisasi Kesehatan Dunia (*World Health Organization*) pada tahun 2021, lebih dari 1,5 miliar orang di seluruh dunia mengalami gangguan pendengaran, dan sekitar 430 juta di antaranya memerlukan rehabilitasi. Di Indonesia sendiri, berdasarkan Survei Kesehatan Indonesia (SKI) 2023, prevalensi disabilitas pendengaran pada usia ≥ 1 tahun mencapai sekitar 1,1 juta orang, yang menggambarkan besarnya kebutuhan akan akses komunikasi yang lebih inklusif bagi komunitas Tuli.

Bahasa isyarat merupakan media komunikasi utama bagi komunitas Tuli. Bahasa ini memanfaatkan gestur visual tangan, ekspresi wajah, serta gerakan tubuh sebagai unsur linguistik utama. Di Indonesia, terdapat dua sistem bahasa isyarat yang dikenal, yaitu Bahasa Isyarat Indonesia (BISINDO) dan Sistem Isyarat Bahasa Indonesia (SIBI). BISINDO berkembang secara alami dalam komunitas Tuli sebagai bagian dari identitas sosial yang melekat pada budaya lokal, dan lebih umum digunakan dalam interaksi sehari-hari

(Asriandhini dan Rahmawati, 2021). Sementara itu, SIBI merupakan sistem isyarat terstandarisasi yang digunakan secara formal di Sekolah Luar Biasa (SLB) sebagai bahasa pengantar dalam pembelajaran (Murwati dan Syefriani, 2024).

Selain digunakan dalam situasi formal dan pendidikan, struktur SIBI mengikuti tata bahasa Indonesia lisan dengan penggunaan imbuhan serta susunan kalimat yang sistematis. Hal ini menjadikan SIBI lebih mudah dipelajari oleh masyarakat non-Tuli atau pemula yang baru mengenal bahasa isyarat. Penelitian yang dilakukan oleh Irchamna dkk., (2025) mengungkapkan bahwa meskipun bahasa isyarat diakui sebagai media komunikasi yang penting, pengetahuan masyarakat terhadap bahasa isyarat masih sangat rendah. Akibatnya, interaksi dengan penyandang tunarungu sering kali menjadi sulit dan kurang optimal. Kondisi ini menimbulkan kesenjangan komunikasi yang signifikan, terutama pada Sistem Isyarat Bahasa Indonesia (SIBI) yang digunakan sebagai standar dalam bidang pendidikan, transportasi, kesehatan, dan layanan publik. Oleh karena itu, penting untuk mengembangkan solusi yang dapat mengurangi kesenjangan komunikasi ini.

Perkembangan teknologi kecerdasan buatan (*Artificial Intelligence*), khususnya dalam bidang visi komputer (*computer vision*), telah membuka peluang besar untuk menciptakan sistem penerjemah bahasa isyarat otomatis. Salah satu pendekatan yang paling menjanjikan dalam pengolahan citra adalah *Convolutional Neural Network* (CNN). CNN memiliki kemampuan untuk mengenali pola visual yang kompleks dalam citra, yang membuatnya sangat efektif dalam tugas-tugas pengenalan pola, termasuk pengenalan gestur tangan. Namun, penggunaan citra Red, Green, Blue (RGB) murni dalam pengenalan bahasa isyarat memiliki beberapa keterbatasan, seperti kesulitan dalam membedakan posisi jari yang saling tumpang tindih dan ketergantungan pada variasi pencahayaan serta latar belakang (Li dkk., 2022; Zhang dkk., 2023). Hal ini berpotensi menurunkan stabilitas dan akurasi model pengenalan bahasa isyarat.

Pendekatan berbasis citra murni, meskipun efektif dalam mempertahankan informasi visual yang lengkap, cenderung sangat sensitif terhadap perubahan pencahayaan, latar belakang, dan warna kulit, yang dapat memengaruhi stabilitas dan akurasi model pengenalan. Untuk mengatasi tantangan tersebut, pendekatan multimodal yang menggabungkan citra visual dengan data sensorik seperti data sarung tangan (*glove*) atau informasi kerangka tangan (*hand skeleton*) mulai banyak dikembangkan. Penggabungan

berbagai modalitas ini diharapkan dapat meningkatkan representasi gerakan secara lebih komprehensif serta meningkatkan akurasi pengenalan gestur. Studi yang menggabungkan data sudut sendi jari dari sensor sarung tangan dengan informasi titik kunci tangan yang diekstraksi dari kamera dan diproses menggunakan model hibrida seperti CNN-BiLSTM menunjukkan hasil yang lebih baik dibandingkan dengan penggunaan satu modalitas saja (Ong dan Ranganath, 2021; Chen dkk., 2022). Namun, pendekatan ini memerlukan perangkat sensor tambahan yang relatif mahal dan kurang praktis untuk diterapkan dalam sistem penerjemahan bahasa isyarat yang efisien.

Salah satu pendekatan alternatif yang saat ini banyak digunakan adalah pendekatan berbasis *keypoint* tangan. Pendekatan ini menggunakan titik-titik *landmark* untuk merepresentasikan posisi jari, sendi, dan pergelangan tangan dalam citra digital. Keuntungan dari representasi berbasis *landmark* ini adalah stabilitasnya terhadap perubahan kondisi lingkungan, karena hanya memanfaatkan informasi *geometris* tangan yang tidak terpengaruh oleh faktor eksternal seperti pencahayaan dan latar belakang (Utomo dkk., 2025; Rais dkk., 2025). Pemanfaatan fitur *keypoint* sebagai masukan CNN juga mampu mengurangi kompleksitas komputasi sekaligus meningkatkan ketahanan model terhadap gangguan latar belakang dan pencahayaan (Zhang dkk., 2023; Liu dkk., 2024). Dengan memanfaatkan *MediaPipe Hands* yang dikembangkan oleh Google, titik-titik *landmark* pada tangan dapat dideteksi secara *real-time* dengan tingkat akurasi yang tinggi. *MediaPipe Hands* mampu mendeteksi hingga 21 titik *landmark* tangan, dan memungkinkan pemanfaatan citra tangan dalam berbagai bentuk representasi data, seperti citra tangan murni, citra dengan *overlay landmark*, atau hanya menggunakan koordinat *landmark* tangan.

Selain itu, dalam pengembangan model CNN untuk tugas pengenalan citra, pemilihan arsitektur yang efisien secara komputasi sangat penting untuk mencapai keseimbangan antara akurasi dan kebutuhan sumber daya. Arsitektur CNN klasik seperti *VGG16* dan *ResNet50* telah banyak digunakan dalam berbagai aplikasi *computer vision* karena stabilitas dan kinerjanya yang baik. Namun, arsitektur tersebut cenderung memerlukan jumlah parameter dan operasi komputasi yang tinggi, sehingga kurang ideal untuk diterapkan pada lingkungan dengan sumber daya terbatas (Amien dkk., 2025). Sebagai alternatif, arsitektur *EfficientNet* yang dirancang untuk mengurangi kompleksitas model sekaligus tetap mempertahankan akurasi yang kompetitif, telah menunjukkan hasil yang baik dalam berbagai aplikasi pengenalan citra. Dengan menggunakan teknik *compound scaling*,

EfficientNet mengoptimalkan kedalaman, lebar, dan resolusi jaringan secara bersamaan (Amien dkk., 2025; MDPI Applied Sciences, 2024). Beberapa penelitian komparatif menunjukkan bahwa *EfficientNetB0* memiliki kinerja yang baik dalam pengenalan gestur, bahkan pada *dataset* yang terbatas, meskipun hasilnya sangat bergantung pada strategi pelatihan dan ukuran *dataset* yang digunakan (Pu dan Yi, 2024).

Oleh karena itu, penelitian ini bertujuan untuk menganalisis perbandingan integrasi citra asli dan *landmark* tangan dalam sistem pengenalan bahasa isyarat SIBI menggunakan *EfficientNetB0*. Dengan mengevaluasi pengaruh variasi *praproses* masukan terhadap kinerja model, diharapkan dapat diperoleh informasi mengenai pendekatan yang lebih efisien. Selain arsitektur model, representasi masukan juga berperan penting dalam menentukan kinerja sistem pengenalan gestur. Representasi masukan dapat memengaruhi kemampuan model dalam mengekstraksi fitur visual dan spasial yang penting untuk meningkatkan akurasi klasifikasi (Akdag dan Baykan, 2024). Namun, integrasi *landmark* dalam *pipeline* CNN juga menghadapi tantangan, terutama terkait dengan variasi bentuk dan kompleksitas dinamika gestur yang dapat memengaruhi kinerja model secara signifikan (Akdag dan Baykan, 2024).

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah dipaparkan, maka rumusan masalah dalam penelitian ini adalah sebagai berikut:

1. Bagaimana kinerja model *deep learning* *EfficientNetB0* dalam melakukan klasifikasi gestur alfabet Sistem Isyarat Bahasa Indonesia (SIBI) menggunakan empat representasi citra yang berbeda, yaitu citra asli, citra berbasis *landmark* tangan, citra *overlay* (citra asli + *landmark*), dan citra yang bervariasi atau heterogen sebagai masukan?
2. Representasi citra manakah yang memberikan kinerja terbaik dalam meningkatkan akurasi klasifikasi gestur alfabet SIBI?

1.3 Tujuan Penelitian

Tujuan penelitian yang ingin dicapai dalam penelitian ini adalah sebagai berikut:

1. Menganalisis dan membandingkan kinerja model *deep learning* dalam melakukan klasifikasi gestur alfabet Sistem Isyarat Bahasa Indonesia (SIBI) menggunakan berbagai representasi citra masukan.
2. Mengevaluasi pengaruh penggunaan yaitu citra asli, citra berbasis *landmark* tangan, citra *overlay* (citra asli + *landmark*), dan citra yang bervariasi atau heterogen terhadap kemampuan model EfficientNetB0 dalam mengenali dan mengklasifikasikan gestur bahasa isyarat secara akurat.
3. Mengidentifikasi konfigurasi pelatihan yang optimal melalui penerapan variasi *preprocessing* data dan pengaturan *hyperparameter*, sehingga dapat diketahui faktor-faktor yang memengaruhi peningkatan kinerja dan akurasi model klasifikasi.

1.4 Manfaat Penelitian

Penelitian ini diharapkan dapat memberikan kontribusi signifikan dalam pengembangan teknologi penerjemah bahasa isyarat otomatis, khususnya untuk Bahasa Isyarat Indonesia (SIBI). Secara akademis, penelitian ini dapat memperkaya pemahaman mengenai pengaruh representasi citra, seperti citra asli dan *landmark* tangan, terhadap akurasi pengenalan gestur. Secara praktis, penelitian ini berpotensi menghasilkan solusi penerjemah bahasa isyarat yang lebih efisien dan dapat diimplementasikan secara luas, mengurangi kesenjangan komunikasi antara penyandang disabilitas pendengaran dan masyarakat umum.

1.5 Ruang Lingkup

Ruang lingkup penelitian ini dibatasi untuk menjaga fokus pada perancangan sistem pengenalan bahasa isyarat berbasis citra dan *landmark* tangan, dengan ketentuan sebagai berikut:

1. Dataset yang digunakan merupakan *dataset* citra gestur alfabet Sistem Isyarat Bahasa Indonesia (SIBI) yang diperoleh dari Kaggle. *Dataset* tersebut terdiri dari 24 kelas huruf alfabet dengan pengecualian huruf J dan Z karena kedua huruf tersebut merupakan gestur dinamis yang memerlukan representasi gerakan.
2. Model *deep learning* yang digunakan adalah EfficientNetB0 dengan pendekatan *transfer learning*.
3. Fokus penelitian adalah klasifikasi isyarat tangan abjad satu tangan (*one-handed Signs*), tanpa mencakup frasa atau kalimat dalam SIBI.

4. Penelitian ini menerapkan empat scenario sebagai representasi citra masukan dataset yang diperlakukan secara terpisah sebelum digunakan sebagai input model, yaitu variasi Dataset A (*overlay landmark* dan gambar asli), variasi Dataset B (hanya gambar asli), variasi Dataset C (hanya *landmark* dengan canvas hitam), dan variasi Dataset D (heterogen, *overlay landmark* dan gambar asli).
5. Evaluasi kinerja model menggunakan metrik akurasi, presisi, *recall*, dan *F1-Score*.

1.6 Sistematika Penulisan

Sistematika penulisan skripsi ini terbagi ke dalam lima bab dengan pokok bahasan sebagai berikut.

BAB I PENDAHULUAN

Bab ini menyajikan latar belakang masalah, rumusan masalah, tujuan dan manfaat penelitian, ruang lingkup, serta sistematika penulisan dalam penyusunan skripsi mengenai analisis perbandingan integrasi citra asli dan *landmark* tangan pada sistem isyarat bahasa Indonesia (SIBI) menggunakan EfficientNetB0.

BAB II TINJAUAN PUSTAKA

Bab ini menguraikan teori-teori yang relevan dengan pelaksanaan dan penyusunan penelitian, termasuk teknologi pengenalan bahasa isyarat, model CNN, serta teknik *preprocessing* dan representasi citra.

BAB III METODOLOGI PENELITIAN

Bab ini menjelaskan langkah-langkah penelitian yang digunakan, meliputi pemilihan *dataset*, tahapan *preprocessing*, visualisasi *landmark*, perancangan dan pelatihan model EfficientNetB0, serta evaluasi kinerja model.

BAB IV HASIL DAN PEMBAHASAN

Bab ini memaparkan hasil eksperimen dan evaluasi model pengenalan isyarat tangan yang telah dikembangkan. Selain itu, disajikan analisis dan interpretasi hasil yang diperoleh, termasuk perbandingan antara skenario yang diuji, serta analisis terhadap tingkat akurasi dan efektivitas eksperimen yang dijalankan.

BAB V PENUTUP

Bab ini berisi kesimpulan dari keseluruhan uraian pada bab-bab sebelumnya serta saran dari penulis untuk pengembangan penelitian lebih lanjut.