

BAB I

PENDAHULUAN

1.1. Latar Belakang

Manusia sebagai makhluk sosial memiliki naluri untuk terhubung dengan sesamanya, memahami dunia di sekitar mereka, dan mengeksplorasi apa yang terjadi pada diri mereka (Iffah & Fitri Yasni, 2022). Rasa ingin tahu itulah yang kemudian mendorong seseorang untuk berkomunikasi. Komunikasi yang baik adalah yang dapat diterima dan dipahami oleh lawan bicara (Mahadi, 2021). Melalui komunikasi, seseorang dapat mengekspresikan diri dan menyampaikan ide serta gagasan. Namun, bagi penyandang disabilitas sensoris, komunikasi verbal menjadi hambatan yang signifikan. Data dari Badan Pusat Statistik (BPS) tahun 2022 menunjukkan bahwa terdapat sekitar dua juta lebih penyandang disabilitas rungu dan wicara di Indonesia. Mereka cenderung menggunakan bahasa isyarat sebagai medium utama dalam komunikasi.

Bahasa isyarat adalah bentuk komunikasi visual yang mengandalkan gestur tubuh, gerak bibir, dan ekspresi wajah. Setiap negara memiliki standar bahasa isyarat sendiri, termasuk Indonesia yang memiliki Bahasa Isyarat Indonesia (Bisindo) dan Sistem Isyarat Bahasa Indonesia (SIBI) (Gumelar et al., 2018). Di lembaga pendidikan Sekolah Luar Biasa (SLB), SIBI lebih sering digunakan dalam proses pembelajaran. Meski demikian, pemahaman umum terhadap SIBI masih terbatas. Sebagian masyarakat yang tidak memahami SIBI lebih condong menggunakan tulisan sebagai sarana interaksi terhadap teman rungu dan wicara. Dampaknya, komunikasi yang terjadi menjadi kurang efektif. Di samping itu, komunikasi dengan tulisan dianggap kurang relevan untuk situasi yang membutuhkan respons cepat seperti interaksi medis, bencana alam, dan keadaan darurat lainnya. Oleh karena itu, diperlukan sebuah metode yang mampu mengenali gerakan isyarat SIBI melalui gerak tangan dan postur tubuh secara efisien. Metode yang diusulkan mampu menerjemahkan gerakan isyarat SIBI ke dalam teks dan audio. Tujuannya adalah untuk memfasilitasi komunikasi yang lebih efektif terhadap komunitas rungu dan wicara di Indonesia.

Selanjutnya, dalam pengembangan ide ini memunculkan kebutuhan akan penggunaan teknologi visi komputer dan machine learning. Deep learning, sebagai cabang dari machine learning telah digunakan secara luas dalam praktek pengenalan pola. Algoritma deep learning memungkinkan sistem untuk mengenali dan mempelajari pola-pola rumit dalam gerakan isyarat. Sejumlah penelitian di banyak negara telah menyorot penggunaan teknologi deep learning dalam pengembangan proyek di bidang pengenalan gerakan isyarat.

Studi yang dilakukan oleh (Alawwad et al., 2021), menerapkan algoritma Faster R-CNN untuk pengenalan Bahasa Isyarat Arab (ArSL). Arsitekturnya terdiri dari Convolutional Neural Network (CNN) dan Region Proposal Network (RPN) yang dirancang untuk menghasilkan proposal wilayah objek pada citra. Dengan menggunakan model VGG-16 dan ResNet-18, sistem ini berhasil mencapai tingkat akurasi sebesar 93% dalam pengujian real-time.

Studi terbaru oleh (Khartheesvar et al., 2023), mengusulkan penggunaan algoritma Long Short-Term Memory (LSTM) dengan kerangka MediaPipe Holistic Pipeline untuk mengenali kata-kata terisolasi dalam Bahasa Isyarat India (ISL). Algoritma ini menggunakan 75 estimasi koordinat (*landmarks*) yang diekstraksi dari bingkai (*frames*) video RGB. Sistem yang dikembangkan berhasil mencapai tingkat akurasi sebesar 94,8% pada dataset INCLUDE-50 dan 87,4% pada dataset INCLUDE.

Berdasarkan penelitian-penelitian tersebut, didapati bahwa model CNN dan LSTM secara independen mampu menghasilkan nilai akurasi yang tinggi dalam mengenali gerakan isyarat. Pada penelitian ini, dua algoritma tersebut diadopsi dalam model gabungan (hybrid) CNN-LSTM. Kombinasi kedua model ini memungkinkan sistem memanfaatkan dua informasi sekaligus (spasial dan temporal) dari data, sehingga model dapat menjalankan tugas pengenalan bahasa isyarat dengan lebih baik. Penelitian sebelumnya oleh (Sinha et al., 2023) telah mengonfirmasi keefektifan penggunaan model CNN-LSTM dalam konteks Pengenalan Bahasa Isyarat. Dalam penelitian tersebut, model CNN-LSTM berhasil mencapai tingkat akurasi sebesar 95.85% pada dataset publik yang berisi 26 gerakan abjad dalam bahasa isyarat.

Selanjutnya, integrasi suara dilakukan dengan memanfaatkan API dari Google *Text-to-Speech* (gTTS). API ini menawarkan kualitas yang baik dalam mengonversi teks ke audio. gTTS mendukung banyak bahasa, termasuk Bahasa Indonesia, sehingga memungkinkan model menghasilkan output suara dengan logat Indonesia yang lebih alami. Dengan demikian, penelitian ini diharapkan dapat memberikan kontribusi yang signifikan dalam pengembangan teknologi pengenalan bahasa isyarat yang lebih canggih serta mampu memfasilitasi komunikasi yang efektif terhadap komunitas runtu dan wicara di Indonesia.

1.2. Rumusan Masalah

Berdasarkan latar belakang di atas, rumusan masalahnya sebagai berikut :

1. Bagaimana penerapan model hybrid CNN-LSTM dalam pengenalan gerakan isyarat SIBI?
2. Bagaimana gTTS diintegrasikan untuk menerjemahkan gerakan isyarat SIBI ke dalam teks suara?
3. Bagaimana performa sistem setelah menerapkan model hybrid CNN-LSTM dengan integrasi gTTS?

1.3. Tujuan

Berdasarkan rumusan masalah di atas, tujuan penelitian sebagai berikut :

1. Meneliti penerapan model hybrid CNN-LSTM dalam pengenalan gerakan isyarat SIBI.
2. Mengembangkan fitur untuk menerjemahkan gerakan isyarat SIBI ke dalam teks suara menggunakan gTTS.
3. Mengevaluasi performa sistem setelah menerapkan model hybrid CNN-LSTM dengan integrasi gTTS.

1.4. Manfaat

Berdasarkan penelitian yang dilakukan, manfaat dari penelitian ini adalah :

1. Meningkatkan pemahaman tentang efektivitas penerapan model hybrid CNN-LSTM dalam pengenalan gerakan isyarat SIBI.

2. Menghasilkan fitur yang mampu menerjemahkan gerakan isyarat SIBI ke dalam teks suara menggunakan layanan gTTS, untuk memfasilitasi komunikasi yang efektif terhadap komunitas runtu dan wicara di Indonesia.
3. Memberikan pemahaman tentang performa sistem setelah menerapkan model CNN-LSTM dengan integrasi gTTS, untuk mengevaluasi kualitas model dalam mengenali gerakan isyarat SIBI ke dalam teks suara.

1.5. Batasan Masalah

Dalam penelitian ini, terdapat sejumlah batasan masalah sebagai berikut :

1. Fokus penelitian terbatas pada penerapan model hybrid CNN-LSTM untuk mengenali gerakan isyarat SIBI, dan integrasi gTTS untuk menerjemahkan hasil pengenalan model ke dalam teks suara.
2. Dataset SIBI yang digunakan terbatas pada:
 - 13 Kosakata : ‘Aku’, ‘Kamu’, ‘Main’, ‘Tunggu’, ‘Bola’, ‘Ajar’, ‘Lapang’, ‘Panjang’, ‘Hari’, ‘Harga’, ‘Senyum’, ‘Sama’, ‘Netral’.
 - 1 Kata hubung : ‘dan’.
 - 4 Prefiks (awalan) : ‘ber-’, ‘di-’, ‘me-’, ‘ter-’.
 - 4 Sufiks (akhiran) : ‘-an’, ‘-nya’, ‘-kan’, ‘-i’.
 - 3 Partikel : ‘-pun’, ‘-lah’, ‘-kah’.
3. Dataset yang digunakan terbatas pada data milik pribadi yang dikumpulkan secara mandiri oleh penulis dengan merekam gerakan isyarat SIBI menggunakan webcam yang terhubung dengan perangkat laptop.
4. Pengembangan model terbatas pada pembuatan program pengenalan gerakan isyarat SIBI yang terintegrasi dengan gTTS untuk menghasilkan teks suara, tanpa mencakup pengembangan perangkat keras khusus.
5. Fitur *Text-to-Speech* hanya mendukung teks suara dalam Bahasa Indonesia.
6. Pengenalan gerakan isyarat terbatas pada 3-10 kata dalam 1 kalimat.
7. Pengujian real-time hanya dilakukan melalui perangkat laptop.
8. Evaluasi performa hanya difokuskan pada pengukuran akurasi, presisi, recall, dan f1-score, serta kecepatan komputasi pada pemrosesan model.