

BAB I

PENDAHULUAN

1.1 Latar Belakang

Bahasa isyarat merupakan salah satu sarana komunikasi utama yang digunakan oleh individu dengan keterbatasan pendengaran maupun kemampuan berbicara di berbagai negara. Sebagai bentuk komunikasi non-verbal, sistem ini mengandalkan kombinasi gerakan tangan, mimik wajah, dan bahasa tubuh untuk menyampaikan pesan, melakukan interaksi, serta membangun hubungan sosial antarindividu. Di Indonesia, Bahasa Isyarat Indonesia (BISINDO) telah memperoleh pengakuan resmi dan menjadi bahasa utama yang digunakan secara luas oleh komunitas penyandang tunarungu.

Namun, meskipun pentingnya bahasa isyarat tersebut, tingkat pemahaman masyarakat umum terhadap bahasa ini masih sangat rendah. Hal ini terutama terjadi pada kalangan pendidik, tenaga kesehatan, keluarga penyandang tunarungu, dan masyarakat luas yang tidak memiliki akses atau kesempatan belajar bahasa isyarat secara memadai. Akibatnya, terjadi hambatan komunikasi yang cukup signifikan antara penyandang tunarungu dan masyarakat luas. Hambatan komunikasi ini bukan hanya berdampak pada interaksi sosial sehari-hari, melainkan juga menimbulkan keterbatasan dalam akses pendidikan, kesempatan kerja, dan partisipasi sosial secara keseluruhan.

Selain itu, perkembangan teknologi informasi dan komunikasi belum sepenuhnya menjangkau komunitas penyandang tunarungu dalam konteks komunikasi yang inklusif. Oleh karena itu, terdapat kebutuhan mendesak untuk menjembatani kesenjangan komunikasi ini dan meningkatkan aksesibilitas informasi serta interaksi sosial bagi penyandang tunarungu. Salah satu solusi yang menjanjikan adalah pengembangan sistem penerjemah bahasa isyarat otomatis yang mampu mengonversi suara menjadi bentuk gerakan bahasa isyarat secara *real-time*. Sistem ini diharapkan tidak hanya dapat membantu penyandang tunarungu berkomunikasi dengan masyarakat umum, tetapi juga dapat meningkatkan

pemahaman masyarakat umum terhadap bahasa isyarat.

Dalam beberapa tahun terakhir, kemajuan pesat dalam bidang kecerdasan buatan dan khususnya teknologi *deep learning* membuka peluang baru dalam pengembangan sistem penerjemah bahasa isyarat otomatis. *Deep learning*, yang merupakan bagian dari *machine learning*, memiliki kemampuan luar biasa dalam mengenali pola kompleks dan fitur tersembunyi dari data visual dan temporal, yang merupakan karakteristik utama dari bahasa isyarat.

Model *Convolutional Neural Network* merupakan arsitektur *neural network* yang sangat efektif dalam mengekstraksi fitur spasial dari citra atau video. CNN dapat mengidentifikasi bentuk tangan, posisi jari, dan pose tubuh secara akurat melalui lapisan konvolusi yang berfungsi menangkap pola-pola visual yang relevan. Keunggulan CNN adalah kemampuannya untuk mengenali fitur lokal dan menggeneralisasi informasi dari berbagai sudut pandang atau kondisi pencahayaan.

Sementara itu, *Long Short-Term Memory* adalah jenis *Recurrent Neural Network* yang dirancang khusus untuk mengolah data sekuensial dan temporal. Bahasa isyarat tidak hanya terdiri dari gerakan statis, tetapi juga perubahan gerakan yang dinamis dan urutan tanda yang bermakna. LSTM mampu mengingat informasi jangka panjang dan mengolah konteks temporal dari urutan frame video, sehingga dapat memahami pola gerakan tangan dan ekspresi secara berkelanjutan. Kombinasi antara CNN dan LSTM memberikan pendekatan yang sangat kuat, di mana CNN mengekstraksi fitur spasial dari setiap frame, dan LSTM memodelkan hubungan temporal antar frame dalam video bahasa isyarat.

Selain itu, beberapa penelitian terbaru juga mengintegrasikan mekanisme perhatian dalam arsitektur *hybrid* ini untuk meningkatkan fokus model pada fitur penting yang relevan dengan konteks gerakan isyarat, sehingga meningkatkan akurasi dan ketahanan terhadap variasi lingkungan seperti pencahayaan dan latar belakang yang kompleks.

Berbagai penelitian telah mengadopsi model *hybrid* CNN-LSTM dalam pengembangan sistem pengenalan bahasa isyarat, dengan tingkat akurasi yang berbeda-beda tergantung pada jenis dataset, kondisi pengujian, serta teknik yang digunakan. Elhagry & Gla mengembangkan sistem pengenalan bahasa isyarat

Mesir dengan mengombinasikan CNN dan LSTM [1]. Sistem ini berhasil mencapai akurasi sebesar 72%, yang meskipun tergolong moderat, menunjukkan adanya tantangan signifikan dalam mengenali bahasa isyarat yang kompleks dan dinamis. Faktor utama yang membatasi performa sistem ini adalah keterbatasan data serta keragaman gestur yang harus dikenali.

Selanjutnya, Kumari & Anand, mengembangkan *framework hybrid* CNN-LSTM yang dilengkapi dengan mekanisme perhatian untuk mengenali bahasa isyarat dari video isolasi [2]. Studi mereka melaporkan akurasi rata-rata sebesar 84,65% pada dataset *benchmark* WLASL yang terdiri dari 100 kelas isyarat. Namun, penelitian ini juga mengidentifikasi bahwa kondisi nyata seperti pencahayaan yang minim dan latar belakang yang padat berpengaruh negatif terhadap performa model, menyoroti tantangan utama dalam penerapan sistem secara *real-time*.

Di Indonesia, penelitian oleh Sihananto et al. pada Bahasa Isyarat Indonesia (BISINDO) menunjukkan akurasi awal sekitar 82,56%[3]. Hasil ini mengindikasikan potensi besar teknologi pengenalan bahasa isyarat di Indonesia, namun juga mempertegas adanya kendala yang disebabkan oleh variasi gestur tangan dan kompleksitas gerakan yang masih perlu diatasi melalui pengembangan algoritma yang lebih canggih dan dataset yang lebih representatif.

Selain itu, Summon et al. mengaplikasikan model MobileNetV2-LSTM-SelfMLP untuk pengenalan bahasa isyarat Arab dalam kondisi lingkungan nyata yang bervariasi, termasuk perubahan pencahayaan dan latar belakang [4]. Model ini berhasil memperoleh akurasi sebesar 87,69%, yang memperlihatkan kemajuan teknologi dalam menghadapi tantangan kondisi dunia nyata yang beragam dan sulit.

Secara keseluruhan, hasil-hasil penelitian tersebut menegaskan bahwa meskipun terdapat kemajuan pesat dalam teknologi pengenalan bahasa isyarat, sistem *real-time* masih menghadapi tantangan besar, khususnya terkait dengan kualitas *input* video, kecepatan pemrosesan, dan keragaman gerakan yang harus diakomodasi oleh model.

Implementasi sistem penerjemah bahasa isyarat berbasis web memberikan kemudahan akses bagi berbagai lapisan masyarakat. Sistem berbasis web dapat

diakses melalui perangkat sederhana seperti *smartphone*, tablet, atau komputer tanpa perlu perangkat keras khusus yang mahal. Hal ini mendukung pemerataan teknologi dan inklusivitas bagi komunitas tunarungu dan masyarakat luas.

Penggunaan metode *deep learning* seperti CNN-LSTM dalam sistem ini tidak hanya memungkinkan pengenalan yang akurat, tetapi juga adaptif terhadap kondisi lingkungan yang beragam. Dengan optimalisasi model dan penggunaan algoritma efisien, kecepatan pemrosesan dapat ditingkatkan agar aplikasi berjalan lancar dalam kondisi *real-time*.

Model *hybrid* CNN-LSTM merupakan pendekatan efektif dalam pengembangan sistem penerjemah bahasa isyarat otomatis dengan kemampuan mengekstraksi fitur spasial dan temporal secara simultan. Penelitian terdahulu menunjukkan hasil yang menjanjikan meskipun tantangan seperti pencahayaan, keragaman gestur, dan kondisi lingkungan nyata masih perlu diatasi.

Pengembangan sistem berbasis web dan integrasi *speech recognition* menjadi kunci untuk mewujudkan teknologi yang inklusif dan mudah diakses. Ke depan, diperlukan pengembangan dataset yang lebih luas dan representatif serta algoritma yang lebih adaptif dan efisien. Selain itu, kolaborasi erat dengan komunitas tunarungu, pendidik, dan pemangku kepentingan akan memastikan teknologi ini sesuai dengan kebutuhan nyata dan berdampak positif secara sosial.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah diuraikan, maka rumusan masalah yang disusun adalah sebagai berikut :

1. Bagaimana merancang sistem berbasis web yang dapat menerjemahkan *input* suara menjadi teks dan visualisasi bahasa isyarat dalam bentuk gambar secara *real-time*?
2. Bagaimana penerapan metode CNN-LSTM dalam mengenali suara dan mengklasifikasikannya menjadi teks serta label isyarat yang sesuai?
3. Bagaimana performa model CNN-LSTM dalam mengenali kata dari suara dan bagaimana hasil prediksi model tersebut dibandingkan dengan hasil transkripsi dari *Speech Recognition*?

1.3 Tujuan Penelitian

Adapun tujuan dari penelitian ini sebagai berikut:

1. Mengembangkan sistem berbasis web yang mampu menerjemahkan *input* suara menjadi teks dan visualisasi bahasa isyarat dalam bentuk gambar secara *real-time*.
2. Menerapkan metode CNN-LSTM untuk mengenali dan mengklasifikasikan suara menjadi label yang sesuai dengan kata atau frasa tertentu.
3. Mengevaluasi performa model CNN-LSTM dan membandingkan hasil prediksinya dengan hasil transkripsi dari teknologi *Speech Recognition*.

1.4 Batasan Masalah

Terdapat beberapa batasan masalah yang penulis lakukan terhadap penelitian yang akan diteliti, diantaranya sebagai berikut:

1. Sistem hanya mengenali tiga jenis kata atau frasa tertentu yang telah dilatih dalam dataset, yaitu “halo”, “selamat_pagi”, dan “selamat_siang”, dengan total 100 file audio untuk setiap kelas.
2. *Output* bahasa isyarat yang dihasilkan berupa gambar statis huruf A–Z Bahasa Isyarat Indonesia (BISINDO).
3. Sistem hanya menerima *input* suara dalam bahasa Indonesia dengan durasi rekaman 8 detik dan format file WAV.

1.5 Manfaat Penelitian

Penelitian ini diharapkan dapat memberikan manfaat sebagai berikut:

1. Memberikan kontribusi terhadap pengembangan ilmu pengetahuan di bidang kecerdasan buatan dan pengolahan sinyal suara, khususnya dalam penerapan metode *hybrid* CNN–LSTM untuk sistem penerjemah bahasa isyarat.
2. Menjadi referensi bagi penelitian selanjutnya yang berfokus pada penerjemahan *multimodal* (suara ke isyarat).
3. Menyediakan solusi teknologi berbasis web yang dapat membantu masyarakat umum berkomunikasi dengan penyandang tunarungu melalui konversi suara menjadi teks dan bahasa isyarat, sekaligus meningkatkan

pemahaman masyarakat terhadap bahasa isyarat dasar (huruf A–Z) secara interaktif dan inklusif.