

BAB I PENDAHULUAN

1.1 Latar Belakang

Bahasa isyarat adalah cara utama bagi teman tuli untuk mengungkapkan ide, emosi, dan keinginan mereka kepada orang lain. Di Indonesia, terdapat dua jenis bahasa isyarat yang sering dipakai, yakni Bahasa Isyarat Indonesia (BISINDO) dan Sistem Isyarat Bahasa Indonesia (SIBI). SIBI secara resmi dikembangkan oleh pemerintah sebagai sistem bahasa isyarat yang menggunakan struktur tata bahasa Bahasa Indonesia. Namun, hingga sekarang, pemakaiannya masih terbatas di antara masyarakat luas. Banyak orang yang belum mengetahui bentuk dan makna isyarat dalam SIBI, sehingga hal ini menciptakan kesulitan dalam berkomunikasi antara teman tuli dan lingkungan sosial mereka. Kurangnya pendidikan, pelatihan, serta dukungan teknologi adalah beberapa faktor yang menghalangi penyebaran dan penerapan SIBI dengan baik di kalangan masyarakat.

Banyak upaya telah dilakukan untuk mengatasi hambatan komunikasi ini, salah satunya dengan memanfaatkan teknologi komputer dan kecerdasan buatan. Dalam bidang pengolahan citra, dan pengenalan gerakan tangan menjadi salah satu topik yang cukup sering diteliti. Pendekatannya bermacam-macam, mulai dari pengambilan data secara manual, penggunaan metode klasifikasi, hingga pemakaian kamera untuk mengenali gerakan tangan. Akhir-akhir ini, penggunaan model deep learning berbasis *Convolutional Neural Network* (CNN) mulai memberikan hasil yang baik. Meski masih ada kendala seperti terbatasnya data yang mewakili variasi bahasa isyarat SIBI. Selain itu, sebagian besar penelitian yang sudah ada mampu mengenali huruf atau angka secara terpisah, belum pada tahap membentuk kata atau kalimat.

Untuk mengatasi kekurangan tersebut, penelitian ini mengembangkan model pengenalan bahasa isyarat dengan memanfaatkan *deep learning* melalui kombinasi CNN dan LSTM, menggunakan dataset khusus yang merepresentasikan gerakan tangan statis pada Sistem Isyarat Bahasa Indonesia (SIBI). Dataset ini mencakup 26 huruf alfabet, 10 angka, serta 5 kata umum yang sering digunakan dalam percakapan sehari-hari, yaitu “saya”, “kamu”, “dia”, “makan”, dan “tidur”. Model dibangun dengan *frame work TensorFlow*, didukung oleh *OpenCV* dan *MediaPipe* untuk mendeteksi tangan secara langsung melalui kamera. Keunggulan pendekatan ini terletak pada kemampuannya mengenali lebih banyak kategori isyarat secara efisien, bahkan pada kondisi pencahayaan dan sudut pandang yang bervariasi. Dari hasil pengujian, model mampu mencapai akurasi hingga 96% untuk huruf dan angka, serta 94% untuk kata.

Beberapa penelitian terdahulu telah menerapkan CNN atau kombinasi CNN-LSTM untuk mengenali gesture huruf atau angka menggunakan dataset Bahasa Isyarat Indonesia (BISINDO), seperti studi oleh Aljabar & Suharjito [1], sedangkan penelitian lain seperti yang dilakukan oleh Limantara & Tristianito [2] hanya mengenali huruf alfabet dalam SIBI

menggunakan CNN saja. Selain itu, riset yang menggunakan *pipeline MediaPipe* dan LSTM seperti oleh Kurniawan [3] masih terbatas pada beberapa kosakata pertanyaan saja. Hingga saat ini, belum ada sistem yang menggabungkan pengenalan huruf, angka, dan kata dalam SIBI melalui mode *real-time* menggunakan *pipeline MediaPipe* dan CNN-LSTM terpisah untuk tiap kategori, seperti yang dikembangkan dalam penelitian ini. Dengan demikian, model ini diharapkan dapat menjadi solusi teknologi yang implementatif dalam menjembatani komunikasi antara teman tuli dengan masyarakat umum, serta mendorong adopsi lebih luas terhadap SIBI di Indonesia.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah dijelaskan, maka rumusan masalah dalam Tugas Akhir ini adalah sebagai berikut:

1. Bagaimana pembuatan sistem bahasa isyarat menggunakan metode *Convolutional Neural Network* (CNN) untuk deteksi dan klasifikasi?
2. Seberapa akurat model dalam mendeteksi 26 huruf alfabet, angka, dan 5 kata dalam bahasa isyarat?

1.3 Tujuan

Adapun tujuan dari Tugas Akhir ini adalah sebagai berikut:

1. Merancang dan membangun sistem pengenalan bahasa isyarat berbasis CNN yang mampu melakukan deteksi dan klasifikasi terhadap 26 huruf alfabet, 10 angka, dan 5 kata dalam bahasa isyarat SIBI secara *real-time*.
2. Mengukur tingkat akurasi model dalam mengenali kategori isyarat SIBI menggunakan dataset yang telah dibangun secara khusus.
3. Mengintegrasikan metode CNN dan LSTM dengan *MediaPipe* untuk memastikan sistem bekerja stabil dalam berbagai sudut pandang serta mendukung deteksi gerakan tangan yang konsisten dan efisien.

1.4 Batasan Masalah

Dalam Proyek Akhir ini, dilakukan pembatasan masalah sebagai berikut:

1. Penelitian ini hanya berfokus pada Sistem Isyarat Bahasa Indonesia (SIBI) dan tidak mencakup Bahasa Isyarat Indonesia (BISINDO) atau bahasa isyarat lainnya.
2. Model hanya akan mendeteksi 26 huruf alfabet, angka (0-9), dan 5 kata umum seperti; saya, kamu, dia, makan, tidur. Kata-kata lain di luar cakupan ini tidak termasuk dalam penelitian.
3. Sistem deteksi gestur diimplementasikan dalam tiga model terpisah untuk masing-masing kategori (huruf, angka, dan kata). Setelah kamera aktif, pengguna dapat memilih kategori deteksi keseluruhan dan gabungan. Namun ada ketidak

sesuaian dalam gabungan karena faktor kesamaan Gerakan dan CPU yang membuat deteksi model kurang akurat.

1.5 Cakupan Pengerjaan

Cakupan pekerjaan dalam Tugas Akhir ini mencakup perancangan, pelaksanaan, dan pengujian sistem yang dapat mengenali bahasa isyarat SIBI dengan menggunakan *model deep learning* yang didasarkan pada *Convolutional Neural Network (CNN)* dan *Long Short-Term Memory (LSTM)*. Sistem ini dirancang untuk mendeteksi serta mengidentifikasi gerakan tangan yang tidak bergerak secara langsung dalam waktu nyata, yang ditangkap oleh kamera dan kemudian diolah menggunakan pustaka *MediaPipe* untuk mendapatkan titik-titik koordinat tangan sebagai data masukan pada model. Sistem ini mampu mengenali 26 huruf alfabet, 10 angka, serta 5 kata umum dalam SIBI, yang terdiri dari: “saya”, “kamu”, “dia”, “makan”, dan “tidur”.

Data dikumpulkan lewat perekaman langsung menggunakan kamera laptop. Setiap gerakan diubah menjadi rangkaian titik koordinat tangan dengan bantuan *MediaPipe*. Satu gerakan terdiri dari 30 *frame*, lalu datanya diperbanyak menggunakan teknik augmentasi supaya variasinya lebih beragam dan seimbang antar kategori. Ada beberapa keterbatasan pada sistem ini, seperti jumlah data di beberapa kategori yang masih sedikit, kemampuan mengenali hanya satu tangan dalam satu waktu, dan hanya bisa mendeteksi gerakan statis, belum bisa membaca gerakan berurutan membentuk kalimat.

Seluruh tahap dalam Tugas Akhir ini dilakukan secara mandiri oleh penulis, mulai dari pengumpulan dan pengolahan dataset, desain arsitektur model, pelatihan serta evaluasi model, sampai implementasi sistem pengenalan secara *real-time*. Di harapkan sistem yang dibangun ini dapat menjadi solusi praktis untuk mendukung aksesibilitas dalam komunikasi bagi teman tuli dan mendorong penggunaan SIBI yang lebih luas di masyarakat.

1.6 Tahapan Pengerjaan

1. Analisis Kebutuhan

Di tahap ini, ditentukan kebutuhan untuk sistem yang akan dibuat, baik dari sisi fungsional maupun *non-fungsional*. Fokus utamanya adalah menilai masalah komunikasi antara teman tuli dan masyarakat, termasuk kesulitan memahami bahasa isyarat SIBI. Dari hasil penilaian, disimpulkan bahwa sistem perlu bisa mengenali 26 huruf, 10 angka, dan 5 kata umum dalam bahasa isyarat SIBI secara langsung. Proses ini dibantu dengan dokumen desain kebutuhan serta referensi dari jurnal dan kajian literatur yang relevan.

2. Desain Sistem dan Arsitektur Model

Tahap ini berfokus pada perancangan struktur sistem dan desain model *deep learning*. Arsitektur gabungan CNN-LSTM dibuat untuk mengolah data koordinat tangan (*landmark*) yang dihasilkan oleh *MediaPipe*. Desainnya mencakup alur kerja langsung dari input kamera, proses ekstraksi fitur, hingga tahap klasifikasi. Pengembangan model dilakukan dengan *TensorFlow Keras*, sedangkan diagram alur dibuat menggunakan *draw.io* untuk menggambarkan langkah-langkah sistem.

3. Pengumpulan dan Pengolahan Dataset

Dataset diperoleh dengan merekam gerakan tangan yang diam menggunakan kamera laptop dan kemudian diproses dengan *MediaPipe* untuk mendapatkan 21 titik koordinat tangan di setiap *frame*, yang selanjutnya diubah menjadi vektor berdimensi 126. Setiap gerakan terdiri dari 30 *frame* dan diulang sebanyak 30 kali untuk setiap label. Untuk meningkatkan kinerja model, data diperbanyak dengan melakukan augmentasi dua kali menggunakan *noise Gaussian* dan transformasi skala acak untuk meningkatkan variasi. Alat yang digunakan mencakup *OpenCV* untuk pengambilan video, dan *Python* untuk pemrosesan serta penyimpanan dataset dalam format *.npy*.

4. Implementasi dan Pelatihan Model

Tahap ini mencakup pembuatan model CNN-LSTM menggunakan *TensorFlow* dan *Keras*, yang dilatih dengan dataset yang sudah diproses sebelumnya. Model terdiri dari beberapa lapisan *Conv1D*, *LSTM*, *Dense*, *BatchNormalization*, dan *Dropout*. Proses pelatihan berlangsung selama 100 *epoch* dengan pengaturan *early stopping* untuk menghindari *overfitting*. Setelah selesai, model disimpan dalam format *.h5* agar bisa digunakan pada aplikasi secara *real-time*. Alat yang dipakai meliputi *TensorFlow*, *Keras*, dan *Numpy*.

5. Pengujian dan Evaluasi Sistem

Setelah model selesai dibuat, tahap berikutnya adalah menguji seberapa akurat kinerjanya menggunakan data uji dan pengujian langsung lewat kamera. Pengujian ini mencakup penghitungan akurasi, presisi, *recall*, dan *f1-score*. Selain itu, sistem juga diuji secara *real-time* pada pengguna untuk melihat kemampuannya dalam mengenali gerakan secara langsung. Untuk analisis statistik digunakan *scikit-learn*, sedangkan antarmuka *real-time* dibuat dengan bantuan *OpenCV*.

6. Dokumentasi dan Penyusunan Laporan

Tahap akhir mencakup penyusunan laporan Tugas Akhir yang memuat seluruh proses, mulai dari analisis hingga evaluasi hasil. Dokumentasi ditulis menggunakan Microsoft Word, serta visualisasi data dan grafik menggunakan *Matplotlib*.