

BAB I

PENDAHULUAN

1.1 Latar Belakang

Bahasa isyarat merupakan bahasa yang memanfaatkan gabungan gerakan tangan, lengan tubuh atau ekspresi wajah untuk menciptakan makna dan arti spesifik yang dapat dipahami oleh pihak lain, mirip dengan kalimat verbal pada umumnya (Masood dkk., 2018). Bahasa ini umumnya dijadikan sebagai alat komunikasi bagi penyandang tunarungu.

Tunarungu merujuk pada situasi seseorang yang mengalami hambatan dalam fungsi pendengaran mereka, sehingga mengalami kesulitan dalam menangkap suara. Meskipun bantuan alat pendengaran digunakan, kondisi ini tetap menghambat kemampuan berkomunikasi secara lisan (Rahmah, 2018). Hal ini berlaku ketika individu yang mengalami tunarungu berusaha berkomunikasi dengan orang-orang di sekitarnya yang umumnya tidak mengerti bahasa isyarat. Bahasa isyarat pada awalnya dirancang sebagai media komunikasi bagi penyandang tunarungu atau komunitas Tuli (PSIBK, 2018). Agar bisa memahami bahasa isyarat, masyarakat perlu mempelajari “bahasa baru” yang terdiri dari beragam gerakan dan gestur tangan yang memiliki variasi untuk setiap huruf. Namun, solusi untuk masalah ini salah satunya dapat ditemukan melalui pendekatan teknologi komputerisasi oleh para peneliti, seperti sistem pendeteksi bahasa isyarat secara langsung (*realtime*).

Beberapa upaya telah dilakukan untuk mengatasi permasalahan ini, dengan sebagian besar berfokus pada penerjemahan abjad bahasa isyarat sebagai input yang diteliti. Salah satu contoh sistem yang telah dikembangkan mencakup sistem yang berhasil diimplementasikan dalam mengenali abjad bahasa isyarat Italia (LIS) (Schmalz, 2022), abjad bahasa isyarat Arab (ArSLA) (Alsaadi dkk., 2022) dan paling banyak diimplementasikan untuk mendeteksi abjad bahasa isyarat Amerika Serikat (ASL).

Metode yang diaplikasikan dalam waktu terdekat pada sistem ini masih dominan mengadopsi pendekatan secara *deep learning (vision-based)* (Rastgoo, Kiani dan Escalera, 2021). Salah satu model arsitektur *deep learning* yang paling sering digunakan dalam sistem deteksi ini adalah *Convolutional Neural Network* (CNN), yang dapat diambil dari rangkaian arsitektur siap pakai maupun dirancang dari dasar. Beberapa contoh model CNN yang telah diimplementasikan mencakup VGG19 (Schmalz, 2022), *AlexNet* (Alsaadi dkk., 2022), *Google Inception V3* (Malakan dan Albaqami, 2021) dan YOLO-LSTM (Rivera-Acosta dkk., 2021). Masing-masing model ini berhasil mencapai tingkat hasil deteksi yang beragam, mulai dari 77,2% hingga 99,81%.

Upaya juga telah dilakukan dengan mengusulkan pendekatan alternatif, yaitu melibatkan penggunaan perangkat sensor khusus (*sensor-based*), serta menggunakan algoritma *machine learning* umum untuk mengklasifikasikan gestur tangan. Contoh dari pendekatan ini pernah dilakukan melalui penggunaan *Leap Motion Camera* (Humaira dkk., 2018), *Delsys Trigno* (Yuan dkk., 2020) dan modul akselerometer triaksial (Na, Yang dan Woo, 2021). Meskipun hasil yang

diperoleh juga sangat baik, tingkat akurasinya sedikit lebih rendah sehingga berkisar antara 78% hingga 89,6% saja.

Sejumlah sistem yang telah dibuat memiliki keunggulan masing-masing dalam mengidentifikasi berbagai abjad bahasa isyarat. Namun demikian, dalam konteks ini setiap kelompok sistem juga menunjukkan keterbatasan. Pendekatan berbasis *deep learning* umumnya memerlukan sumber daya komputasi yang cukup besar dan sering kali memerlukan penggunaan unit pemrosesan grafis (GPU) yang bertenaga tinggi untuk mempercepat proses pembelajaran dan deteksi (O'Mahony dkk., 2020; Moshayedi dkk., 2022). Kendala ini seringkali menjadi tantangan dalam penelitian yang menerapkan pendekatan *deep learning*, terutama pada perangkat dengan spesifikasi yang terbatas (Shin dkk., 2021). Di sisi lain, penggunaan sistem yang bergantung pada perangkat sensor khusus jarang digunakan, terutama karena keterbatasan finansial untuk memperoleh peralatan tersebut serta bilangan unit yang sedikit dan setiap penelitian belum tentu menggunakan peralatan serupa (Shin dkk., 2021; Alsaadi dkk., 2022). Oleh karena itu, banyak peneliti cenderung memilih pendekatan *deep learning* sebagai solusi yang lebih ekonomis dan praktis sehingga tidak harus menggunakan alat yang berlainan di setiap penelitian berbeda.

Solusi yang berpotensi untuk mengatasi kedua-dua permasalahan ini adalah melalui penggunaan teknologi *hand gesture recognition* bersifat *open source* yang dikeluarkan oleh para peneliti *Google* pada tahun 2020 silam, yaitu *Mediapipe Hands* (Zhang dkk., 2020). *Mediapipe Hands* adalah sebuah *framework* pelacakan tangan dan pengenalan gerakan yang dapat dilakukan secara

realtime dan bersifat sangat ringan dibandingkan dengan *framework* CNN lainnya (Yasumuro dan Jin'no, 2022). Ia menggunakan model *machine learning* untuk mendeteksi dan melacak posisi 21 *keypoints* pada setiap tangan, termasuk ujung jari, buku jari, dan pergelangan tangan dengan tingkat akurasi yang tinggi. Keunggulan lainnya adalah bahwa penggunaan *Mediapipe Hands* hanya memerlukan CPU dan kamera konvensional, sehingga dapat diimplementasikan pada perangkat dengan spesifikasi rendah (Zhang dkk., 2020). Secara ringkasnya, *Mediapipe Hands* merupakan sistem *vision-based* dengan prinsip kerja *sensor-based*.

Beberapa peneliti telah mulai menggunakan *Mediapipe Hands* dalam penelitiannya semenjak ia dikeluarkan. *Framework* ini diuji sebagai pengekstrak fitur gestur tangan abjad beberapa bahasa isyarat asing seperti ASL, India, LIS serta Turkiye dan membuktikan bahwa *Mediapipe Hands* juga dapat digunakan sebagai metode ekstraksi fitur untuk sistem deteksi bahasa isyarat serta dapat diklasifikasikan dengan pelbagai macam algoritma *machine learning* (Halder dan Tayade, 2021). Selain itu, pendekatan *Mediapipe Hands* untuk sistem deteksi bahasa isyarat terbukti memakan biaya komputasi 70% lebih rendah daripada *deep learning* dalam proses ekstraksi fitur dan pelatihan (Yasumuro dan Jin'no, 2022). Hal ini memungkinkan sistem deteksi bahasa isyarat digunakan dalam perangkat dengan rentang spesifikasi yang lebih luas mulai dari *low-end* dan tentunya hingga *high-end*.

Berdasarkan latar belakang tersebut, penelitian ini mengusulkan sebuah prototipe sistem deteksi abjad bahasa isyarat dengan proses *realtime* yang

dihasilkan melalui implementasi teknologi *Mediapipe Hands* sebagai upaya kontribusi untuk mengatasi masalah ini. Sistem yang diusulkan diperuntukkan untuk mendeteksi 26 abjad SIBI. Adapun keterbaruan dari penelitian ini berupa sebuah sistem yang selain mampu mendeteksi tangan kanan, juga bisa mendeteksi tangan kiri secara *realtime* dan dapat dideteksi dari pelbagai posisi dan jarak. Sistem ini juga akan menggunakan algoritma SVM untuk mengklasifikasikan abjad SIBI. Algoritma ini digunakan karena kombinasi *Mediapipe Hands* dan SVM menghasilkan performa yang paling terunggul dalam proses pelatihan (Halder dan Tayade, 2021; Shin dkk., 2021; Yasumuro dan Jin'no, 2022). Maka dari itu, penelitian ini juga akan menguji kebenaran kombinasi tersebut melalui proses pelatihan yang sama. Kemudian sistem yang dibuat akan diuji daya deteksi abjadnya secara *realtime* dan membandingkannya dengan hasil proses pelatihan. Selain itu, pengujian deteksi gestur abjad juga dilakukan di pelbagai posisi yang ditentukan. Hasil keseluruhan dari pengujian ini akan dibandingkan dengan sistem-sistem sebelumnya.

1.2 Rumusan Masalah

Berdasarkan latar belakang di atas, maka dapat dirumuskan masalah dalam penelitian ini sebagai berikut:

- a. Bagaimana implementasi teknologi *Mediapipe Hands* dan algoritma SVM pada aplikasi prototipe sistem deteksi abjad bahasa isyarat SIBI secara *realtime* dilakukan?
- b. Berapakah tingkat akurasi yang didapatkan setelah proses pelatihan menggunakan algoritma SVM terhadap data SIBI yang dihasilkan dari *Mediapipe Hands*?
- c. Berapakah tingkat akurasi yang didapatkan dari aplikasi prototipe saat mendeteksi abjad SIBI setelah diuji secara *realtime*?
- d. Berapakah tingkat akurasi yang didapatkan dari aplikasi prototipe saat mendeteksi abjad SIBI setelah diuji di berbagai posisi?

1.3 Tujuan Penelitian

Berdasarkan latar belakang dan rumusan masalah di atas, adapun tujuan yang didapatkan untuk penelitian ini adalah sebagai berikut:

- a. Mengimplementasikan teknologi *Mediapipe Hands* dan algoritma SVM pada aplikasi prototipe sistem deteksi abjad bahasa isyarat SIBI.
- b. Mengukur berapakah tingkat akurasi setelah proses pelatihan menggunakan algoritma SVM terhadap data SIBI yang dihasilkan dari *Mediapipe Hands*.
- c. Menguji tingkat akurasi yang didapatkan dari aplikasi prototipe saat mendeteksi abjad SIBI secara *realtime*.

- d. Menguji tingkat akurasi yang didapatkan dari aplikasi prototipe saat mendeteksi abjad SIBI di pelbagai posisi.

1.4 Manfaat Penelitian

Adapun manfaat yang dapat diperoleh dari penelitian ini adalah sebagai berikut:

- a. Dapat membuat sistem pengenalan abjad bahasa isyarat SIBI kedua belah tangan dan tidak melibatkan alat khusus tambahan sehingga penerapan dan penggunaannya menjadi lebih praktis.
- b. Sistem yang mampu mendeteksi abjad gestur tangan di berbagai posisi memudahkan *user* dalam penggunaan.
- c. Sistem dapat digunakan dengan perangkat bersifat ringan atau *low-end*.
- d. Penelitian ini memberikan kontribusi berupa sistem deteksi bahasa isyarat praktis yang bisa digunakan pada bahasa isyarat SIBI dan bahasa lainnya.

1.5 Batasan Masalah

Adapun batasan masalah yang terdapat dalam penelitian ini adalah sebagai berikut :

- a. Implementasi dilakukan dalam bentuk sebuah sistem deteksi 26 abjad bahasa isyarat SIBI secara *realtime* berbasis *desktop* yang bersifat prototipe.
- b. *Framework* yang digunakan untuk proses deteksi dan ekstraksi fitur gestur tangan adalah *Mediapipe Hands*.

- c. Penelitian ini hanya menggunakan bahasa pemrograman Python 3.10 saja karena periode penelitian ini dilakukan hanya bahasa ini yang di-*support* oleh *Mediapipe Hands*.
- d. Algoritma klasifikasi yang digunakan adalah SVM linear dengan parameter *default* yang disediakan oleh *Scikit-learn*.
- e. Kamera *webcam* yang digunakan sebagai input deteksi adalah *Logitech C920*.
- f. Tangan kanan dan kiri tidak dideteksi secara bersamaan.
- g. Sistem hanya mampu mendeteksi kedua tangan dari satu individu saja.
- h. Parameter SVM yang digunakan adalah dari *Scikit-learn* dengan parameter *default*.
- i. Pengujian dilakukan secara *realtime* dalam ruangan terkontrol yang dibatasi sejauh 2 meter.
- j. Sumber cahaya yang digunakan adalah lampu LED dengan daya terik 800 lumens yang dipasang 300 cm di atas dari gestur tangan dan kamera *webcam*.