BAB II
TINJAUAN PUSTAKA

2.1 Penelitian Terdahulu

Penelitian dalam bidang pengenalan bahasa isyarat berbasis computer vision telah mengalami perkembangan signifikan seiring kemajuan teknologi kecerdasan buatan. Beragam pendekatan telah diterapkan untuk mengenali isyarat tangan, mulai dari metode konvensional seperti pemrosesan citra dua dimensi hingga implementasi model pembelajaran mendalam yang kompleks. Berikut adalah beberapa studi sebelumnya yang relevan dengan fokus penelitian ini.

2.1.1. Penggunaan MediaPipe dan LSTM untuk ASL

Sundar dan Bagyammal (2022) mengembangkan sistem pengenalan alfabet American Sign Language (ASL) berbasis pengolahan citra dengan menggabungkan MediaPipe sebagai ekstraktor landmark tangan dan LSTM sebagai model klasifikasinya. Dalam studi tersebut, sistem diuji menggunakan dataset khusus berisi 26 huruf alfabet yang diperoleh dari empat individu berbeda dengan variasi usia dan jenis kelamin. Hasil penelitian menunjukkan tingkat akurasi mencapai 99%, serta ketangguhan dalam mengenali gestur baik statis maupun dinamis. Sistem ini juga dirancang untuk dapat berjalan secara real-time pada perangkat tanpa memerlukan daya komputasi tinggi.

2.1.2. Pengenalan Bahasa Isyarat Menggunakan CNN

Jefri (2023) dari Universitas Harapan Bangsa mengimplementasikan arsitektur Convolutional Neural Network (CNN) untuk mengklasifikasikan gambar alfabet ASL dari dataset Kaggle. Model CNN yang dibangun berhasil mencapai akurasi sebesar 99,53% dalam mengklasifikasi 29 kelas gestur tangan, dan pengujian dilakukan terhadap citra statis yang telah dinormalisasi sebelumnya. Penelitian ini menekankan pentingnya augmentasi data untuk meningkatkan generalisasi model terhadap data baru.

2.1.3. Integrasi MediaPipe dan CNN

Sutrisno dkk. (2023) melakukan penelitian dengan menggabungkan MediaPipe dan CNN untuk mengklasifikasikan lima jenis gestur dasar tangan dalam Bahasa Isyarat Indonesia (BISINDO). Proses ekstraksi ciri dilakukan dengan memanfaatkan koordinat landmark tangan dari MediaPipe, kemudian diteruskan ke model CNN untuk klasifikasi. Sistem diuji dalam lingkungan laboratorium dengan latar belakang netral, menghasilkan akurasi rata-rata sebesar 95%. Penelitian ini menunjukkan potensi MediaPipe sebagai solusi deteksi cepat dan akurat untuk pengenalan isyarat secara langsung dari kamera.

2.1.4. Sistem Real-Time untuk Bahasa Isyarat Daerah

Sutaryo dan Ayu (2023) mengembangkan sistem penerjemah bahasa isyarat daerah (vernacular sign language) secara waktu nyata menggunakan kombinasi MediaPipe, KNN, dan SVM. Sistem dirancang untuk mengenali isyarat dalam video langsung menggunakan webcam, dengan hasil klasifikasi ditampilkan dalam bentuk teks. Dataset yang digunakan berasal dari rekaman pengguna lokal, dan akurasi mencapai sekitar 96% untuk gestur tunggal. Penelitian ini menyoroti pentingnya penggunaan dataset lokal agar sistem mampu beradaptasi dengan variasi bahasa isyarat regional.

2.1.5. MediaPipe dan CNN untuk Sistem Deteksi Adaptif

Kusuma et al. (2023) merancang sistem deteksi bahasa isyarat berbasis CNN yang diawali dengan ekstraksi ciri tangan melalui MediaPipe. Penelitian ini menekankan pentingnya preprocessing, seperti segmentasi tangan dan normalisasi posisi landmark, untuk meningkatkan akurasi klasifikasi. Arsitektur CNN yang digunakan terdiri dari beberapa lapisan konvolusi dan dense layer yang dikustomisasi. Hasil evaluasi menunjukkan model mampu mengenali gestur dengan presisi di atas 95%, terutama untuk gestur statis dengan pencahayaan dan posisi tangan yang stabil.

Tabel 2.1 Perbandingan Penelitian Terdahulu

No	Peneliti & Tahun	Judul Penelitian	Metode/Algoritma	Data set / Lingkup Kelas	Akurasi	Kelebihan	Keterbatasan
1	Sundar & Bagyammal (2022)	ASL Alphabet Recognition using MediaPipe and LSTM	MediaPipe + LSTM	26 huruf ASL dari 4 pengguna	99%	Real-time, robust untuk gestur statis dan dinamis	Dataset terbatas pada alfabet; variasi regional belum diakomodasi
2	Jefri (2023)	Pengenalan Huruf ASL Menggunakan CNN	CNN	29 kelas alfabet ASL (Kaggle)	99,53%	Akurasi sangat tinggi, proses augmentasi mendalam	Hanya mengolah citra statis, belum mendukung input video langsung
3	Sutrisno et al. (2023)	Klasifikasi Gestur BISINDO Menggunakan MediaPipe dan CNN	MediaPipe + CNN	5 kelas BISINDO	95%	Komputasi ringan, cocok untuk implementasi cepat	Kelas terbatas, pengujian di lingkungan terkendali

4	Sutaryo & Ayu (2023)	Real-Time Vernacular Sign Language Recognition using MediaPipe and ML	MediaPipe + KNN/SVM	Bahasa isyarat daerah (dataset lokal)	±96%	Real-time, menggunakan bahasa lokal, klasifikasi cepat	Rentan terhadap perubahan latar dan pencahayaan
5	Kusuma et al. (2023)	Enhancing Sign Language Detection with MediaPipe and CNN	MediaPipe + Custom CNN	Gestur statis (variasi posisi dan cahaya)	≥95%	Akurasi stabil, preprocessing kuat, fleksibel untuk ekspansi ke video	Terbatas pada kondisi ideal dan belum mencakup gestur dinamis

2.2 Tantangan dalam Penelitian Sebelumnya

Meskipun berbagai penelitian terkait pengenalan bahasa isyarat telah menunjukkan hasil yang menjanjikan, masih terdapat sejumlah tantangan teknis dan praktis yang perlu diatasi agar sistem yang dikembangkan dapat diimplementasikan secara luas dan efektif di dunia nyata. Tantangan-tantangan tersebut dapat dikelompokkan dalam beberapa aspek berikut:

2.2.1 Keterbatasan Dataset

Salah satu kendala paling umum adalah keterbatasan dataset, baik dari segi ukuran, keberagaman, maupun kualitas. Banyak penelitian masih bergantung pada dataset publik seperti *ASL Alphabet Dataset* yang hanya

mencakup gestur statis dan kurang mencerminkan keragaman pengguna dari segi usia, jenis kelamin, warna kulit, serta sudut pandang kamera. Minimnya data gerakan dinamis, variasi ekspresi, dan kondisi pencahayaan juga membuat model sulit untuk melakukan generalisasi saat diuji di lingkungan berbeda dari data latihnya.

2.2.2 Ketergantungan pada Kondisi Lingkungan

Beberapa studi yang mengandalkan citra kamera langsung atau pemrosesan real-time, seperti yang menggunakan MediaPipe, menunjukkan bahwa performa sistem sangat bergantung pada pencahayaan, latar belakang, dan kestabilan posisi tangan. MediaPipe, meskipun efisien dan ringan, masih menghadapi kesulitan dalam mendeteksi landmark secara akurat pada kondisi pencahayaan rendah atau latar belakang kompleks. Hal ini dapat mempengaruhi kualitas ekstraksi fitur yang akan digunakan dalam tahap klasifikasi.

2.2.3 Keterbatasan dalam Mengenali Gestur Dinamis

Sebagian besar model yang dilatih untuk klasifikasi gestur hanya memproses citra statis dan belum mendukung pengenalan gestur yang melibatkan gerakan kontinu atau rangkaian isyarat seperti dalam kalimat penuh bahasa isyarat. Penanganan gestur dinamis memerlukan pendekatan berbasis data sekuensial seperti LSTM atau RNN, yang dalam praktiknya menuntut sumber daya komputasi lebih besar dan pelabelan data yang lebih kompleks.

2.2.4 Kompleksitas Model dan Kinerja Real-Time

Meskipun model seperti CNN dan transfer learning telah memberikan akurasi tinggi, kompleksitas arsitektur yang digunakan terkadang menyebabkan waktu pemrosesan lambat, terutama ketika dijalankan di perangkat dengan kemampuan terbatas seperti ponsel pintar. Hal ini bertolak belakang dengan kebutuhan aplikasi asistif berbasis bahasa isyarat yang menuntut kecepatan dan efisiensi tinggi dalam klasifikasi isyarat secara langsung.

2.2.5 Ketergantungan pada Fitur Ekstraksi Eksternal

Beberapa penelitian masih mengandalkan framework eksternal seperti MediaPipe untuk ekstraksi fitur. Walaupun efektif, pendekatan ini membuat sistem sangat tergantung pada keluaran MediaPipe yang bersifat numerik dan belum mengandung makna isyarat. Akibatnya, dibutuhkan integrasi yang cermat antara sistem ekstraksi dan model klasifikasi agar hasil akhir benar-benar mencerminkan maksud dari gestur yang ditampilkan pengguna.

2.3 Arah Penelitian Selanjutnya

Berdasarkan evaluasi terhadap sejumlah studi terdahulu, masih terdapat ruang pengembangan yang signifikan dalam implementasi sistem penerjemah bahasa isyarat berbasis teknologi computer vision. Meskipun berbagai pendekatan seperti penggunaan CNN, RNN, maupun integrasi MediaPipe telah diterapkan, sebagian besar penelitian tersebut masih menghadapi kendala teknis, seperti keterbatasan dataset, sensitivitas terhadap kondisi lingkungan, dan kinerja sistem dalam pengolahan real-time.

Penelitian ini diarahkan untuk mengatasi sebagian tantangan tersebut dengan beberapa pendekatan berikut:

2.3.1 Fokus pada Huruf dan Angka dalam ASL

Berbeda dari studi sebelumnya yang lebih banyak meneliti pengenalan gestur umum atau alfabet statis, penelitian ini secara khusus menargetkan **34 kelas gestur** dalam bahasa isyarat ASL, yaitu huruf A–Y (tanpa J dan Z karena gerakannya bersifat dinamis) serta angka 1–10. Pemilihan fokus ini didasarkan pada pentingnya huruf dan angka dalam komunikasi sehari-hari, terutama untuk pengejaan nama, alamat, atau istilah yang tidak memiliki simbol isyarat khusus.

2.3.2 Pemanfaatan MediaPipe sebagai Ekstraktor Fitur Ringan

Framework **MediaPipe** dipilih sebagai alat utama untuk mendeteksi dan mengekstraksi landmark tangan secara real-time. Keunggulan

MediaPipe yang mampu berjalan di perangkat dengan sumber daya terbatas menjadikannya solusi praktis untuk pengembangan sistem klasifikasi gestur yang cepat dan efisien.

2.3.3 Penerpan CNN untuk Klasifikasi Gestur

Koordinat landmark hasil ekstraksi MediaPipe akan digunakan sebagai input pada **model CNN** yang dirancang khusus untuk mengenali pola spasial dari gestur tangan. Model ini dilatih menggunakan dataset terstruktur untuk membedakan antar kelas gestur dengan akurasi tinggi, sekaligus mempertahankan kecepatan komputasi yang dibutuhkan dalam pengoperasian real-time.

2.3.4 Penerapan Sistem Secara Waktu Nyata (Real-Time)

Penelitian ini dirancang agar sistem yang dikembangkan mampu mengenali dan mengklasifikasikan gestur secara langsung melalui input kamera, tanpa harus melalui proses perekaman atau pemrosesan berkas eksternal. Hal ini diharapkan dapat meningkatkan kemudahan akses dan efektivitas penggunaannya dalam konteks dunia nyata.

2.3.5 Pengembangan Sistem Penerjemah yang Inklusif

Dengan mengintegrasikan teknologi visual dan kecerdasan buatan, penelitian ini diharapkan dapat memberikan kontribusi nyata terhadap **pembangunan teknologi asistif** yang inklusif bagi komunitas tunarungu. Sistem yang dikembangkan berpotensi menjadi prototipe aplikasi penerjemah bahasa isyarat ringan yang dapat digunakan oleh masyarakat luas maupun dalam bidang pendidikan dan layanan publik.

2.4 Teori Terkait

Teori yang dikemukakan dalam subbab ini berfungsi sebagai landasan konseptual yang menjadi pijakan dalam merancang sistem penerjemah bahasa isyarat ASL. Teori-teori yang dipilih merupakan hasil sintesis dari literatur ilmiah yang relevan dan mendukung proses identifikasi, ekstraksi, serta klasifikasi gestur tangan dalam konteks pemrosesan citra berbasis computer vision.

2.4.1 Teori Computer Vision

Menurut Szeliski (2011), computer vision adalah bidang interdisipliner yang berfokus pada bagaimana komputer dapat memperoleh pemahaman tingkat tinggi dari informasi visual dalam bentuk gambar atau video. Dalam implementasinya, computer vision digunakan untuk mengenali pola, objek, atau gerakan secara otomatis dari input visual. Dalam konteks bahasa isyarat, teknologi ini dimanfaatkan untuk mendeteksi dan menafsirkan gestur tangan pengguna yang ditangkap oleh kamera secara real-time.

2.4.2 Convolutional Neural Network (CNN)

CNN adalah arsitektur jaringan saraf tiruan yang dirancang untuk mengenali pola spasial dalam data dua dimensi seperti gambar. LeCun et al. (1998) menjelaskan bahwa CNN terdiri dari lapisan konvolusi, pooling, dan fully connected yang bekerja secara bertahap dalam mengekstraksi fitur hingga melakukan klasifikasi. CNN sangat efektif dalam pengenalan pola visual karena mampu mengurangi kompleksitas data tanpa kehilangan informasi penting. Dalam penelitian ini, CNN digunakan untuk mengklasifikasi fitur gestur yang telah diekstraksi dari koordinat tangan.

2.4.3 MediaPipe Hands

MediaPipe adalah kerangka kerja pemrosesan media lintas platform yang dikembangkan oleh Google. Solusi MediaPipe Hands secara khusus dirancang untuk mendeteksi hingga 21 titik landmark tangan dari citra input secara real-time dengan akurasi tinggi. Model ini menggunakan dua tahap deteksi: *palm detection* untuk mendeteksi keberadaan tangan, dan *hand landmark model* untuk melokalisasi posisi titik-titik utama pada tangan. MediaPipe bekerja secara efisien di perangkat dengan spesifikasi terbatas, menjadikannya sangat ideal untuk aplikasi berbasis gestur.

2.4.4 American Sign Language (ASL)

ASL adalah sistem bahasa isyarat visual-gestural yang digunakan oleh komunitas tunarungu, terutama di Amerika Serikat. ASL memiliki struktur linguistik tersendiri, dengan kombinasi gerakan tangan, ekspresi wajah, dan

postur tubuh sebagai penyampai makna. Penelitian ini secara khusus fokus pada pengenalan alfabet (A–Y, tanpa J dan Z) dan angka (1–10) dalam ASL, yang berbasis pada gestur tangan statis. Studi oleh Wilbur (2000) menunjukkan bahwa pengenalan elemen fonologis dalam ASL secara visual dapat menjadi dasar dalam pengembangan sistem otomatis penerjemah bahasa isyarat.

2.4.5 Alat dan Teknologi Pendukung

Beberapa perangkat keras dan perangkat lunak yang digunakan dalam penelitian ini antara lain:

- **Kamera/Webcam**: Digunakan sebagai input visual untuk merekam gerakan tangan pengguna secara langsung.
- **Python**: Bahasa pemrograman utama dalam pengembangan sistem, didukung oleh pustaka seperti TensorFlow, Keras, dan OpenCV.
- **Jupyter Notebook/Google Colab**: Lingkungan pengembangan yang digunakan untuk melatih dan menguji model.
- **NumPy, Pandas, Scikit-learn**: Digunakan untuk manipulasi data, pemrosesan numerik, dan evaluasi model.

Teknologi ini berperan penting dalam proses pengumpulan data, pelatihan model, serta pengujian sistem penerjemah bahasa isyarat yang dikembangkan.

2.4.6 Kerangka Teoritis Penelitian

Berdasarkan teori-teori di atas, dapat disusun kerangka berpikir sebagai berikut:

Input berupa citra gestur tangan diambil melalui kamera dan diproses menggunakan **MediaPipe Hands** untuk mengekstrak koordinat landmark. Vektor koordinat ini kemudian digunakan sebagai **fitur masukan** untuk

model **CNN** yang dilatih untuk mengenali pola spasial dari gestur tertentu. Seluruh proses dianalisis dalam kerangka **computer vision** untuk mengotomatisasi penerjemahan **ASL huruf dan angka**, dengan target penerapan dalam sistem real-time yang efisien dan responsif.