

BAB II

TINJAUAN PUSTAKA

2.1 Landasan Teori

2.1.1 SIBI (Sistem Bahasa Isyarat Indonesia)

2.1.1.1 Komunikasi Total

Suatu falsafah yang dikenal sebagai "komunikasi total" menggambarkan cara berkomunikasi dengan menggunakan tiga cara yaitu aural, manual, dan lisan. Tujuannya adalah untuk memastikan bahwa orang tunarungu dapat berkomunikasi dengan baik dengan orang lain dan agar penerima dan pengirim pesan saling memahami satu sama lain, sehingga tidak ada ketegangan atau salah paham (Somad, 2009).

Suatu sistem isyarat diperlukan untuk penerapan komtal, pada tahun 1978, SLB-B Zinnia di Jakarta dan SLB-B Karya Mulya di Surabaya mulai menggunakan komunikasi total. Pada awalnya, SLB-B Zinnia menggunakan isyarat spontan, tetapi pada tahun 1978, SLB-B Karya Mulya mulai menggunakan isyarat ASL yang diperkenalkan oleh Ibu Baron Sutadisastra. Kemudian Pedoman Isyarat Bahasa Indonesia, yang dibuat oleh SLB-B Mulya pada tahun 1989, adalah awal dari Kamus Isyarat tunarungu Indonesia. Pada tahun 1990, Kamus Dasar Bahasa Isyarat Indonesia, yang dibuat oleh SLB-B Zinnia, muncul dan terus dikembangkan pada tahun-tahun berikutnya. Pada tahun yang sama, KKPLB juga membuat kamus isyarat. Kamus ini terdiri dari isyarat yang berkembang di sebelas tempat di Indonesia, yang disebut isyarat lokal berasal isyarat yang berkembang di negara

lain. Pada tahun 1993, Direktorat Pendidikan Dasar dan Direktorat Jenderal Pendidikan Dasar dan Menengah Departemen Pendidikan dan Kebudayaan memutuskan untuk menggabungkan empat temuan tersebut menjadi Sistem Isyarat Nasional (Lembaga Penelitian dan Pengembangan Sistem Isyarat Bahasa Indonesia, n.d.).

Sistem Isyarat Bahasa Indonesia (SIBI) yang dibakukan itu adalah salah satu media yang membantu kaum tunarungu berkomunikasi di masyarakat yang lebih luas. Susunannya terdiri dari rangkaian isyarat jari, tangan, dan berbagai gerak yang berfungsi sebagai representasi kosa kata bahasa Indonesia (Nugraheni et al., 2023).

2. 1. 1. 2 Komponen Pembeda Makna

Dalam sistem isyarat terdapat dua jenis komponen. Yang satu berfungsi sebagai penentu atau pembeda makna, sedangkan yang lain berfungsi sebagai penunjang. Semua bersifat visual sehingga dapat dilihat. Menurut (Haliza et al., 2020) komponen komponen itu adalah sebagai berikut:

1. Penampil, yaitu tangan atau bagian tangan yang digunakan untuk membentuk isyarat
2. Posisi, yaitu kedudukan tangan atau kedua tangan terhadap pengisyarat
3. Tempat, yaitu bagian badan yang menjadi tumpuan awal isyarat dibentuk atau arah akhir isyarat
4. Arah, yaitu gerak penampil ketika membentuk isyarat
5. Frekuensi, yaitu jumlah gerak penampil ketika isyarat dibentuk.

Dalam sistem isyarat, penunjang berfungsi untuk menekankan atau memperjelas makna. Dalam hal ini berupa mimik muka yang mengisyaratkan rasa senang, sedih, atau ceria. Selanjutnya adalah gerakan tubuh, seperti gerakan bahu, memberikan kesan tambahan tentang pesan, seperti isyarat tidak tahu. Menaikkan kedua bahu menunjukkan bahwa tidak tahu. Lalu terakhir ada kecepatan gerak, dan kelenturan gerak.

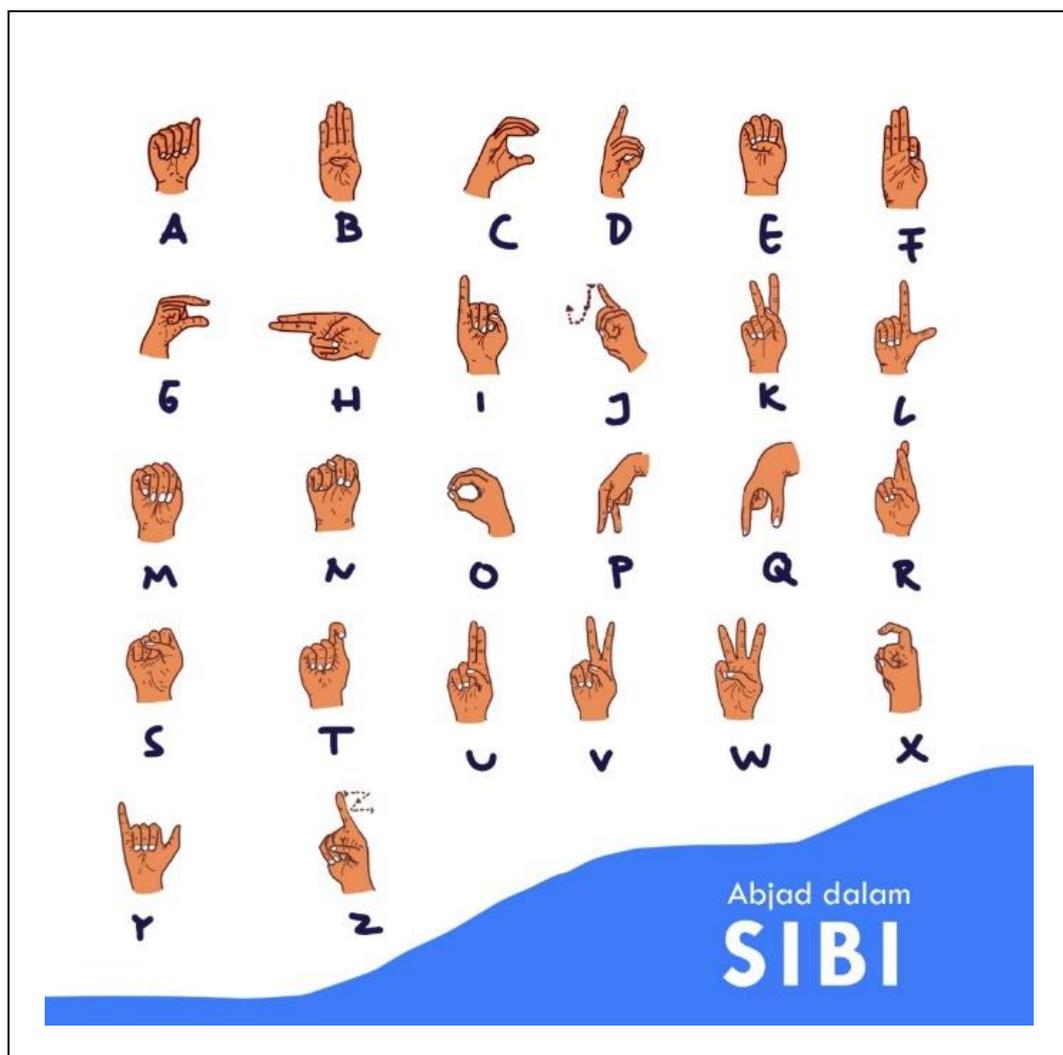
2. 1. 1. 3 Lingkup SIBI

Berdasarkan (Legowo & Surya Sumpeno, 2017) ada beberapa lingkup Bahasa Isyarat berdasarkan SIBI:

1. Isyarat pokok: Merupakan isyarat yang maknanya mewakili suatu kata atau konsep.
2. Isyarat tambahan:
 - a. Isyarat awalan: Dibentuk dengan tangan kanan sebagai penampil utama dan tangan kiri sebagai penampil pendamping. Dibentuk sebelum isyarat pokok.
 - b. Isyarat akhiran dan partikel: Dibentuk dengan tangan kanan sebagai penampil, ditempatkan di depan dada, dan digerakkan mendatar ke kanan.
3. Isyarat bentukan: Isyarat yang dibentuk dari penggabungan isyarat pokok dengan isyarat tambahan atau penggabungan dua atau lebih isyarat pokok.
4. Abjad jari: Isyarat yang dibentuk dengan jari-jari tangan kanan atau kiri untuk mengeja gerakan huruf dan angka. Abjad jari digunakan untuk

mengisyaratkan nama diri, mengisyaratkan singkatan atau akronim, dan mengisyaratkan kata yang belum ada isyaratnya.

Dalam abjad jari setiap huruf dalam alfabet digambarkan dengan gerakan atau posisi khusus dari jari-jari tangan seperti yang terlihat pada Gambar 2.1.



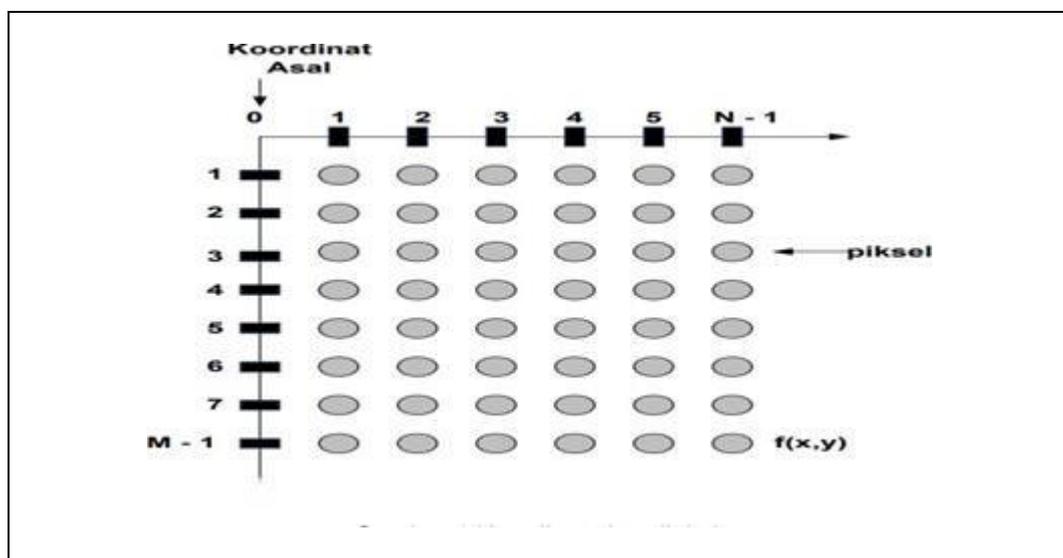
Gambar 2. 1 Ejaan abjad jari dalam SIBI (Habibiati Bestari & Alwiyah Maulidiyah, 2018)

Gambar 2.1 menunjukkan ejaan abjad jari dalam SIBI dari huruf A sampai Z. Terlihat pada huruf J dan Z terjadi pergerakan tangan, sementara untuk huruf

lainnya, hanya terjadi perubahan dalam susunan jari-jari tangan tanpa pergerakan tangan.

2. 1. 2 Citra Digital

Citra digital diwakili oleh matriks yang terdiri dari M kolom dan N baris, yang mana perpotongan antara kolom dan baris tersebut disebut dengan piksel yang merupakan elemen terkecil dari sebuah citra. Setiap piksel memiliki dua parameter yaitu koordinat (x, y) dan intensitas atau warnanya (Darma Putra, 2010). Nilai yang ada pada koordinat (x, y) disebut $f(x, y)$, yang mewakili besar intensitas atau warna dari piksel di titik tersebut. Dengan cara ini, setiap titik dalam citra digital memiliki nilai yang mewakili tingkat kecerahan atau warna pada posisi yang sesuai dalam citra, koordinat pada citra dapat dilihat pada Gambar 2.2.



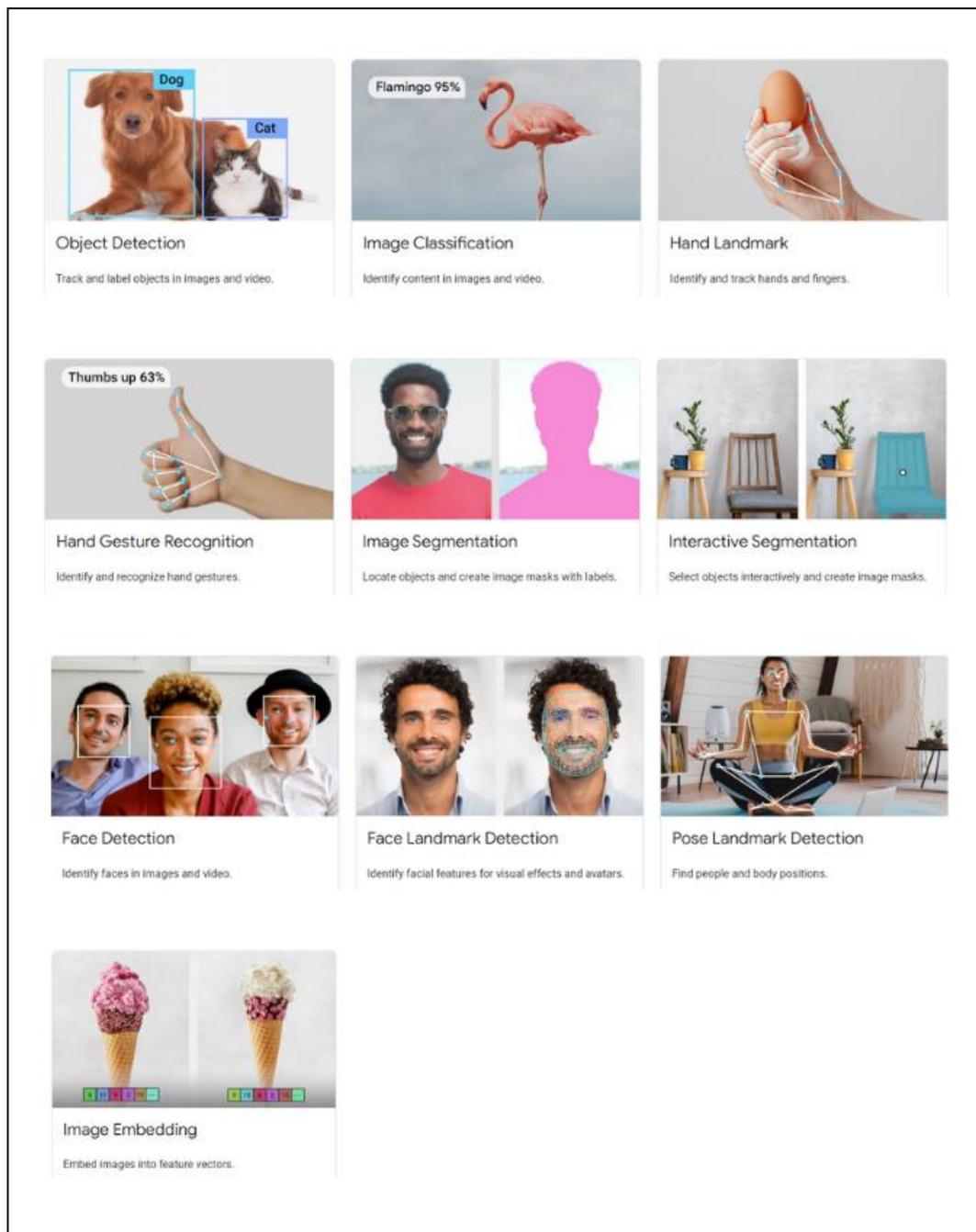
Gambar 2. 2 Koordinat pada gambar (Munantri et al., 2020)

2. 1. 3 Ekstraksi Fitur

Ekstraksi fitur adalah proses di mana fitur-fitur penting dalam sebuah gambar dideteksi dan direpresentasikan dalam bentuk yang lebih mudah diproses. Proses ini sangat penting dalam komputer vision dan pengolahan citra digital karena membantu mengubah data gambar menjadi data numerik atau simbolik yang bisa dianalisis lebih lanjut. Dengan representasi ini, bisa digunakan teknik pengenalan pola dan klasifikasi untuk memahami atau mengidentifikasi isi dari gambar. Jadi, ekstraksi fitur membantu komputer untuk 'melihat' dan 'memahami' gambar dengan cara yang lebih terstruktur dan terorganisir (Wahiddin, 2020).

2. 1. 4 Mediapipe

Mediapipe adalah sebuah *framework* yang dikembangkan oleh Google untuk memfasilitasi pembuatan aplikasi visi komputer dan pembelajaran mesin. *Mediapipe* memanfaatkan konsep *pipeline* untuk membangun dan menjalankan model *machine learning* (Lugaresi et al., 2019). *Mediapipe* ini dirancang untuk memproses data dari berbagai sumber, seperti video, audio, dan sensor. Pada *computer vision* *mediapipe* dapat digunakan dalam berbagai aplikasi seperti yang terlihat pada Gambar 2.3.

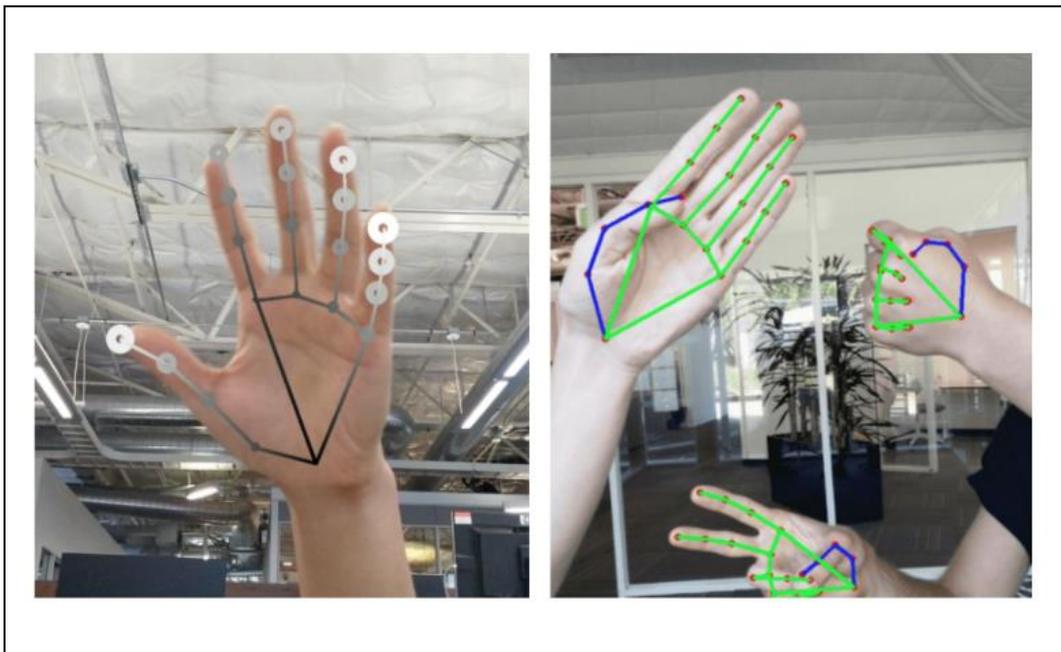


Gambar 2. 3 Penerapan mediapipe pada *computer vision*

2. 1. 4. 1 Mediapipe Hands

Mediapipe hands adalah sebuah komponen dari *platform mediapipe* yang dikembangkan oleh Google. Komponen ini bertujuan untuk mendeteksi dan

melacak tangan dalam citra atau video secara *realtime* seperti yang terlihat pada Gambar 2.5. Pada prosesnya *mediapipe hands* menggunakan dua paket model, yaitu, *palm detection model* dan *hand landmark detection model* (Zhang et al., 2020).



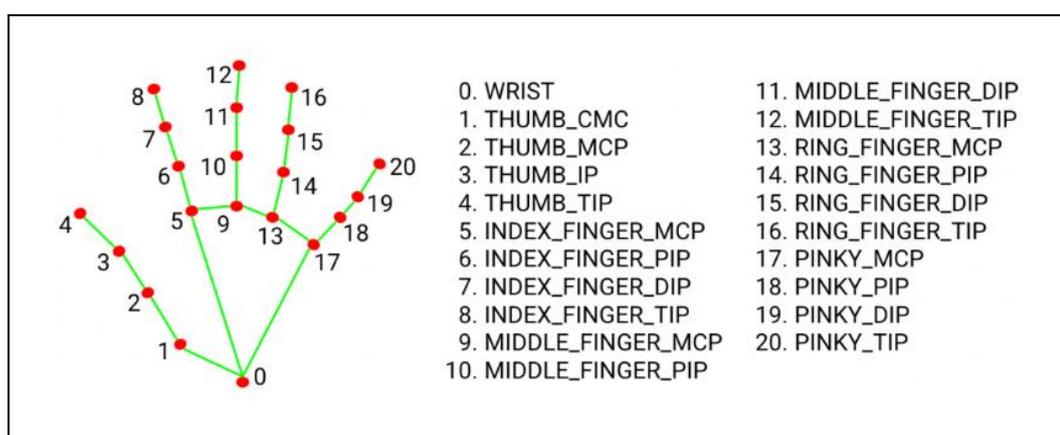
Gambar 2. 4 *Hand landmark detection* (Zhang et al., 2020)

Palm detection model bertugas untuk memproses setiap *frame* per detik dan menghasilkan kotak pembatas tangan yang mengikuti posisi tangan dalam gambar. Tujuannya adalah untuk mendeteksi apakah sebuah telapak tangan ada dalam gambar dan menentukan letaknya, seperti yang terlihat pada Gambar 2.5. Untuk melakukan hal tersebut, *mediapipe* menggunakan model *single shot detector* (SSD).



Gambar 2. 5 Penerapan *palm detection model* (Jonietz et al., 2015)

Setelah dilakukan pendeteksian pada telapak tangan, langkah selanjutnya adalah menggunakan model deteksi *landmark* tangan untuk menentukan posisi titik-titik kunci pada tangan, seperti ujung jari dan pergelangan tangan. Model ini bekerja dengan mengidentifikasi lokasi titik-titik *landmark* penting pada telapak tangan yang telah terdeteksi sebelumnya. Pada Gambar 2.6 terlihat *landmark* tangan terdiri dari total 21 titik *landmark* yang memberikan koordinat x, y, dan z. Koordinat x menunjukkan posisi horizontal, y menunjukkan posisi vertikal, dan z menunjukkan jauh dekatnya titik landmark dari kamera.

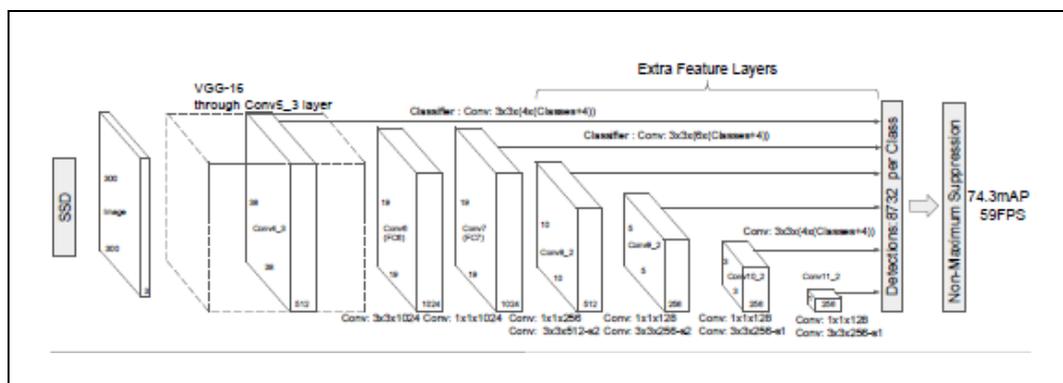


Gambar 2. 6 *21 Hand landmark* (Google, 2020)

2. 1. 5 Single Shot Detector

Pada penerapannya, *MediaPipe* menggunakan *Single Shot Detector (SSD)* sebagai bagian dari teknologi deteksi telapak tangan. *SSD* merupakan metode yang efisien dan cepat dalam mendeteksi objek, di mana proses deteksi dilakukan hanya dalam satu kali pengambilan gambar. Dalam proses ini, *SSD* mengandalkan *Convolutional Neural Network (CNN)* untuk ekstraksi fitur. *Pre-trained CNN* seperti *VGG*, *ResNet*, atau *MobileNet* sering digunakan karena telah dilatih pada *dataset* besar seperti *ImageNet*, sehingga mampu menghasilkan representasi fitur yang sangat baik dari gambar (Google, 2020).

Bagian awal dari *CNN* ini, yang dikenal sebagai "*base network*" atau "*backbone*," bertugas untuk mengidentifikasi pola-pola dasar dan fitur-fitur umum dari gambar input. *SSD* kemudian menggunakan beberapa lapisan konvolusional dalam *CNN* untuk menghasilkan peta fitur dengan resolusi yang berbeda-beda, yang disebut *Multi-scale Feature Maps*. Peta fitur ini memungkinkan *SSD* untuk mendeteksi objek dalam berbagai skala dan lokasi dalam gambar (Liu et al., 2016).



Gambar 2. 7 Arsitektur dari SSD

Gambar 2. 7 merupakan arsitektur dari single shot detector, menurut (Liu et al., 2016) penjelasan dari tiap layer arsitektur tersebut adalah sebagai berikut.

- a. *Feature Extraction*: *CNN* digunakan sebagai bagian dari proses ekstraksi fitur. *SSD* menggunakan *CNN* pra-pelatihan (*pre-trained CNN*) seperti *VGG*, *ResNet*, atau *MobileNet* sebagai bagian dari ekstraksi fitur. *CNN* ini telah dilatih pada *dataset* besar (seperti *ImageNet*) untuk menghasilkan representasi fitur yang sangat baik dari gambar. Fitur-fitur ini kemudian digunakan untuk mendeteksi objek.
- b. *Base Network*: Bagian awal dari *CNN* digunakan sebagai "*backbone*" atau "*base network*" untuk ekstraksi fitur. Ini bertanggung jawab atas identifikasi pola-pola dasar dan fitur-fitur umum dari gambar input.
- c. *Multi-scale Feature Maps*: *SSD* menggunakan beberapa lapisan konvolusional dalam *CNN* untuk menghasilkan peta fitur dengan resolusi yang berbeda. Peta fitur ini memungkinkan *SSD* untuk mendeteksi objek dalam berbagai skala dan lokasi dalam gambar.
- d. *Predictions*: Pada setiap titik dalam peta fitur, *SSD* menggunakan lapisan konvolusional khusus untuk melakukan prediksi terkait kelas objek dan penyesuaian lokasi (*offset*) untuk setiap *anchor box* (kotak baku). Prediksi ini mencakup probabilitas kelas objek dan *offset* untuk mengubah ukuran dan posisi *anchor box* agar sesuai dengan objek yang sebenarnya dalam gambar.
- e. *Non-Maximum Suppression (NMS)*: Setelah prediksi dilakukan, *SSD* menerapkan teknik *non-maximum suppression* untuk mengurangi tumpang tindih antara kotak prediksi.

2. 1. 6 Klasifikasi

Klasifikasi merupakan suatu proses dalam *machine learning* di mana data dibagi ke dalam kategori atau kelas yang telah ditentukan berdasarkan atribut-atribut atau fitur-fiturnya. Proses ini melibatkan pembuatan model atau algoritma yang dapat mempelajari pola-pola dalam data pelatihan dan menggunakan pengetahuan tersebut untuk membuat prediksi terhadap data baru (Rani, 2016).

2. 1. 7 Random Forest

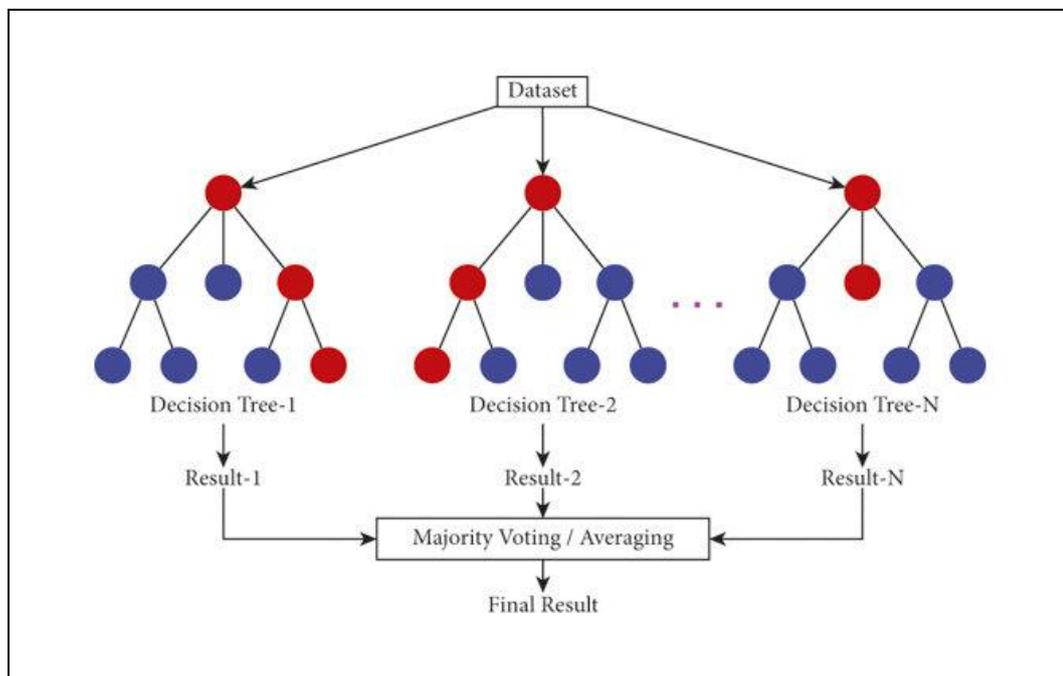
Random forest merupakan salah satu algoritma pembelajaran mesin yang digunakan untuk tugas-tugas klasifikasi dan regresi. Algoritma ini berupa *ensemble learning* yang menggabungkan *decision tree* menjadi satu model. *Decision tree* dalam *random forest* dibangun secara acak, yang berarti setiap pohon dibangun menggunakan sampel data yang dipilih secara acak dari *dataset training*, serta subset acak dari fitur-fitur yang tersedia. Kemudian, hasil dari banyak pohon keputusan diambil untuk melakukan prediksi (Biau & Scornet, 2016).

Adapun menurut (Xu et al., 2017) mekanisme dari *random forest* adalah terdiri dari tahapan sebagai berikut:

1. Menghasilkan sampel acak: titik data K dipilih secara acak dari *training set*. Titik ini digunakan sebagai sampel dalam membangun setiap *decision tree*.
2. Membangun *decision tree*: ketika titik data yang dipilih secara acak telah ditentukan, sebuah *decision tree* dibangun menggunakan *subset* tersebut.
3. Penentuan hasil akhir: langkah terakhir dalam proses pembangunan *random forest* adalah menentukan hasil akhir berdasarkan prediksi dari semua

decision tree yang telah dibangun. Terdapat dua metode yang digunakan untuk menentukan hasil akhir ini yaitu, untuk kasus klasifikasi menggunakan *majority voting* sedangkan untuk kasus regresi dengan pengambilan nilai rata-rata prediksi.

Dengan menggabungkan hasil prediksi dari berbagai pohon, *random forest* menunjukkan tingkat ketahanan yang tinggi terhadap *overfitting* dan kinerja yang baik dalam berbagai tugas klasifikasi dan regresi. Secara praktis, kelebihan *random forest* terletak pada kemampuannya untuk menangani sejumlah besar fitur dan data dengan efisien, sementara parameter seperti jumlah pohon, kedalaman pohon, dan jumlah fitur yang digunakan dapat disesuaikan untuk meningkatkan kinerja dan fleksibilitas model (Biau & Scornet, 2016). Gambar 2.7 menunjukkan gambaran dari bagaimana cara *random forest* bekerja.



Gambar 2. 8 Algoritma *random forest* (Khan et al., 2021)

2. 1. 8 Confusion Matrix

Confusion matrix merupakan metode evaluasi kinerja yang penting dalam konteks klasifikasi, memberikan gambaran terperinci tentang bagaimana model klasifikasi mengklasifikasikan data ke dalam kategori-kategori yang berbeda. Terdiri dari empat kategori, yaitu *True Positive* (TP), *False Positive* (FP), *True Negative* (TN), dan *False Negative* (FN). Beberapa persamaan pada parameter evaluasi *confusion matrix* yaitu *accuracy*, *precision*, *recall*, dan *f1-score*:

$$Accuracy = \frac{(TP + TN)}{(TP + FP + FN + TN)} \quad (2.1)$$

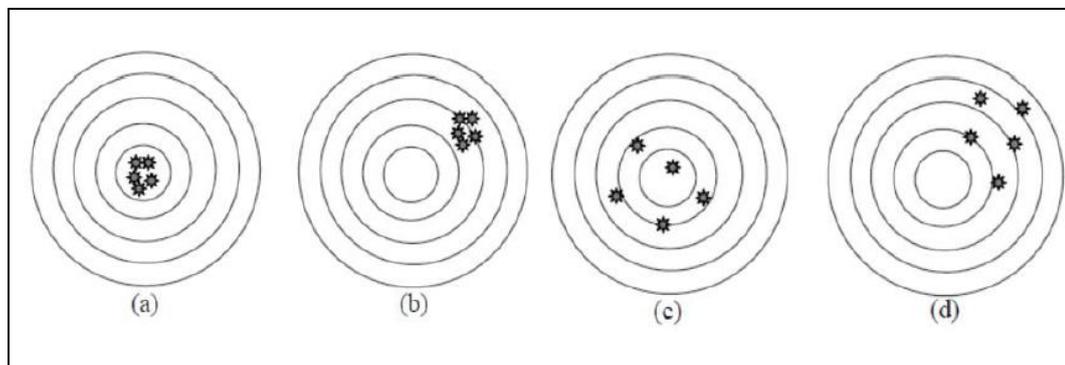
$$Precision = \frac{TP}{TP + FP} \quad (2.2)$$

$$Recall = \frac{TP}{TP + FN} \quad (2.3)$$

$$F1 - Score = 2 \times \frac{precision \times recall}{precision + recall} \quad (2.4)$$

- a. *True Positive (TP)*: Jumlah prediksi positif yang benar (model memprediksi positif dan sebenarnya memang positif).
- b. *True Negative (TN)*: Jumlah prediksi negatif yang benar (model memprediksi negatif dan sebenarnya memang negatif).
- c. *False Positive (FP)*: Jumlah prediksi positif yang salah (model memprediksi positif tapi sebenarnya negatif).
- d. *False Negative (FN)*: Jumlah prediksi negatif yang salah (model memprediksi negatif tapi sebenarnya positif).

Akurasi, presisi, *recall*, dan *F1-score* adalah metrik evaluasi yang digunakan untuk mengukur kinerja model klasifikasi. Akurasi adalah ukuran yang menentukan tingkat kemiripan antara hasil pengukuran dengan nilai yang sebenarnya diukur atau ukuran seberapa sering sesuatu benar atau tepat sasaran. Presisi adalah ukuran seberapa sering prediksi positif model benar atau ukuran seberapa konsisten hasil pengukuran saat dilakukan berulang kali dalam kondisi yang sama. *Recall* adalah ukuran seberapa banyak hasil positif yang benar-benar diidentifikasi sebagai positif. *F1 score* adalah ukuran yang menggabungkan dua metrik evaluasi klasifikasi yaitu presisi dan *recall* (Caelen, 2017). Gambaran dari akurasi dan presisi dapat dilihat pada Gambar 2. 9. (a) menunjukkan hasil akurat dan presisi, (b) menunjukkan hasil presisi tetapi tidak akurat, (c) menunjukkan hasil akurat tetapi tidak presisi, (d) menunjukkan hasil tidak akurat dan tidak presisi.



Gambar 2. 9 Gambaran akurasi dan presisi.

2.2 State-of-the-art

State of the art berisi penelitian yang sudah dilakukan sebelumnya mengenai deteksi bahasa isyarat dengan metode dan data yang berbeda-beda. *State of the art* dari penelitian ini disajikan pada Tabel 2.1.

Tabel 2. 1 *State of the art penelitian*

No	Peneliti	Metode Ekstraksi Fitur	Metode Klasifikasi	Hasil Penelitian
1	(Wiraswendro & Soetanto, 2022)	<i>Mediapipe Holistic</i>	<i>Random Forest</i>	Penggunaan algoritma <i>random forest classifier</i> dengan bantuan <i>mediapipe holistic</i> dan OpenCV dalam bahasa pemrograman Python. <i>Dataset</i> yang dibuat terdiri dari 10 class simbol yang mewakili kata dalam SIBI “Hai”, “Aku”, “Kamu”, “Terima”, “Kasih”, “Maaf”, “Cinta”, “Makan”, “Minum”, dan “Tidur”, dengan total 8734 baris data. <i>Dataset</i> ini kemudian dibagi menjadi 70% data latih dan 30% data uji. Tahapan penelitian mencakup

No	Peneliti	Metode Ekstraksi Fitur	Metode Klasifikasi	Hasil Penelitian
				<p>pembuatan deteksi, pembuatan <i>dataset</i>, pelatihan model klasifikasi, dan pengujian. Pengujian dilakukan menggunakan <i>confusion matrix</i>, dan hasilnya menunjukkan tingkat akurasi, presisi, dan <i>recall</i> masing-masing sebesar 98.6%.</p>
2	(Naufal & Kusuma, 2023)	<i>CNN</i>	<i>KNN, SVM, Deep Learning.</i>	<p>Penelitian ini membandingkan algoritma K-Nearest Neighbors (KNN), Support Vector Machine (SVM), dan Convolutional Neural Network (CNN) dengan arsitektur transfer learning seperti Xception, ResNet50, VGG15, dan MobileNetV2. Hasilnya, CNN dengan Xception memiliki F1 Score tertinggi (99,57%) tetapi membutuhkan waktu komputasi lama (1.387 detik), sementara KNN memiliki waktu komputasi tercepat (0,03 detik) tetapi F1 Score</p>

No	Peneliti	Metode Ekstraksi Fitur	Metode Klasifikasi	Hasil Penelitian
				<p>lebih rendah (86,95%). Pilihan metode tergantung pada kebutuhan; machine learning (KNN atau SVM) cocok untuk kondisi dengan keterbatasan spesifikasi komputer karena cepat dan cukup akurat, sedangkan deep learning (Xception) cocok untuk kebutuhan performa maksimal tanpa batasan spesifikasi. Penelitian ini juga menyarankan penggunaan fitur lain seperti edge detection dan PCA untuk meningkatkan performa klasifikasi dengan waktu komputasi lebih rendah.</p>
3	(Limantara & Trisianto, 2024)	<i>CNN</i>	<i>CNN</i>	<p>Penelitian tersebut membahas implementasi dan pengujian sistem deteksi bahasa isyarat abjad SIBI menggunakan metode Convolutional Neural Network (CNN). Tujuan utamanya adalah untuk mengembangkan sistem yang</p>

No	Peneliti	Metode Ekstraksi Fitur	Metode Klasifikasi	Hasil Penelitian
				<p>mampu mengenali huruf-huruf dalam bahasa isyarat SIBI secara real-time berdasarkan demonstrasi pengguna. Dalam penelitian ini, proses pelatihan dataset dilakukan menggunakan 2.600 gambar dengan 5 epoch dan ukuran batch 10, menghasilkan akurasi pelatihan sebesar 99%.</p>
4	(Sholawati et al., 2022)	<i>Convolutional Neural Network (CNN)</i>	<i>Convolutional Neural Network (CNN)</i>	<p>Mengembangkan sebuah sistem klasifikasi peragaan bahasa isyarat abjad SIBI <i>secara real-time</i> menggunakan teknologi <i>Convolutional Neural Network (CNN)</i> berbasis <i>web</i>. Dalam penelitian ini, data citra digital peragaan abjad SIBI dikumpulkan dari guru dan murid penyandang tunarungu, dengan total 416 citra. <i>Dataset</i> ini meliputi 384 citra hasil pengambilan gambar dan 32 citra hasil konversi video peragaan abjad J dan Z, yang merupakan gerakan</p>

No	Peneliti	Metode Ekstraksi Fitur	Metode Klasifikasi	Hasil Penelitian
				<p>aktif. Aplikasi ini berbasis <i>web</i> dan menampilkan label kelas abjad serta nilai probabilitas hasil klasifikasi pada <i>webcam</i> di dalam <i>website</i>. Penelitian tersebut menghasilkan akurasi hasil <i>training</i> sebesar 90,05% dan akurasi hasil melakukan uji coba sebanyak 52 kali menggunakan model klasifikasi hasil <i>training</i> sebesar 80,76%.</p>
5	(Kuswardhana et al., 2020)	<i>Hue Saturation Value (HSV)</i>	<i>Convolutional Neural Network (CNN)</i>	<p>Melakukan pembuatan dataset SIBI (Sistem Isyarat Bahasa Indonesia) melalui gerakan tangan. Dataset SIBI berhasil dibuat dengan isyarat yang terdiri dari 8 kata statis. Berdasarkan <i>confusion matrix</i>, semua isyarat yang telah dibuat memiliki presisi 100%. Sementara itu, berdasarkan hasil pengujian dengan kondisi pencahayaan</p>

No	Peneliti	Metode Ekstraksi Fitur	Metode Klasifikasi	Hasil Penelitian
				<p>yang cukup dan latar belakang yang polos, dataset SIBI memiliki tingkat keberhasilan prediksi sebesar 97,5%. Metode ekstraksi fitur yang digunakan adalah <i>HSV (Hue, Saturation, Value)</i> yang merupakan representasi warna dalam citra. Digunakan arsitektur <i>Convolutional Neural Network (CNN)</i> yang dibangun menggunakan <i>Keras</i> untuk mengklasifikasikan citra berdasarkan fitur-fitur <i>HSV</i> yang diekstraksi sebelumnya.</p>
6	(Nurhayati et al., 2022)	<i>Hue Saturation Value (HSV)</i>	<i>Convolutional Neural Network (CNN)</i>	<p>Penelitian ini menggunakan dataset berupa 29 citra isyarat tangan, termasuk 26 huruf alfabet dan 3 isyarat tambahan seperti spasi, hapus, dan tidak terklasifikasi. Proses <i>pre-processing</i> dilakukan dengan mengubah citra ke dalam format <i>HSV</i>, kemudian melakukan <i>cropping</i> dan</p>

No	Peneliti	Metode Ekstraksi Fitur	Metode Klasifikasi	Hasil Penelitian
				<p><i>thresholding</i> untuk memudahkan pengenalan. Metode <i>Convolutional Neural Network (CNN)</i> digunakan sebagai alat untuk pembelajaran fitur (<i>feature learning</i>) dan pengklasifikasi isyarat tangan. Pengujian dilakukan dengan memvariasikan tingkat pencahayaan berkisar antara 10 hingga 200 lux, serta jarak tangan ke webcam berkisar antara 50 hingga 200 cm. Hasil penelitian menunjukkan bahwa penggunaan metode <i>CNN</i> pada citra isyarat tangan memberikan akurasi sebesar 97,2%, presisi 91,96%, sensitivitas 91,9%, spesivitas 91,96%, dan f1 score 91,9%. Selain itu hasil akurasi model menghasilkan skor 91,6%</p>

No	Peneliti	Metode Ekstraksi Fitur	Metode Klasifikasi	Hasil Penelitian
7	(Thira et al., 2023)	<i>Hue Saturation Value (HSV)</i>	<i>Convolutional Neural Network (CNN)</i>	<p>Mengklasifikasikan alfabet pada Bahasa Isyarat Bahasa Indonesia (SIBI) menggunakan <i>Convolutional Neural Network (CNN)</i>, dengan tujuan untuk menemukan arsitektur yang paling optimal. Penelitian tersebut melibatkan tiga arsitektur <i>CNN MobileNet</i>, yaitu <i>MobileNetV2</i>, <i>MobileNetV3Small</i>, dan <i>MobileNetV3Large</i>, untuk dibandingkan kinerjanya dalam mengklasifikasikan alfabet SIBI. Hasil penelitian menunjukkan bahwa <i>MobileNetV3Small</i> menghasilkan model yang paling baik. Model ini mencapai akurasi sebesar 98,81% pada data tes dengan menggunakan <i>batchsize</i> 32 dan menjalankan proses pelatihan selama 30 <i>epoch</i>. Dari hasil perbandingan, <i>MobileNetV3Small</i></p>

No	Peneliti	Metode Ekstraksi Fitur	Metode Klasifikasi	Hasil Penelitian
				dengan jumlah <i>epoch</i> 30 menjadi model yang paling baik di antara yang lain, dengan akurasi pelatihan 100% dan validasi 97,62%. Pengujian terhadap model terbaik dilakukan pada 84 data tes dan menghasilkan akurasi sebesar 98,81%, dengan hanya satu citra yang diprediksi salah. Evaluasi model terbaik menunjukkan nilai <i>recall</i> sebesar 0,99, nilai presisi sebesar 0,99, dan nilai <i>f1-score</i> sebesar 0,99.
8	(Larasati et al., 2021)	<i>Histogram of Oriented Gradients (HOG)</i>	<i>Random Forest</i>	Penggunaan ekstraksi fitur <i>Histogram of Oriented Gradient (HOG)</i> dan metode klasifikasi <i>random forest</i> untuk pengenalan <i>American Sign Language (ASL)</i> berdasarkan citra <i>handshape</i> alfabet ASL mendapatkan hasil pengujian yang menunjukkan tingkat akurasi yang

No	Peneliti	Metode Ekstraksi Fitur	Metode Klasifikasi	Hasil Penelitian
				<p>baik, dengan nilai akurasi keseluruhan sebesar 99.10%. Berdasarkan penelitian tersebut, dapat disimpulkan bahwa penggunaan HOG dengan klasifikasi <i>random forest</i> dapat meningkatkan nilai akurasi dalam pengenalan ASL dibandingkan dengan pendekatan sebelumnya yang menggunakan fitur HOG-LDA dan klasifikasi KNN. Pendekatan tersebut sebelumnya memberikan hasil akurasi keseluruhan sebesar 72.43%. Selain itu, dapat disimpulkan bahwa penggunaan HOG dengan klasifikasi <i>random forest</i> menghasilkan akurasi yang lebih baik daripada menggunakan HOG-LDA dengan klasifikasi <i>random forest</i>. Tingkat akurasi dengan menggunakan klasifikasi <i>random forest</i> dan HOG tanpa LDA adalah</p>

No	Peneliti	Metode Ekstraksi Fitur	Metode Klasifikasi	Hasil Penelitian
				99.10%, sementara menggunakan fitur HOG dengan LDA menghasilkan akurasi sebesar 96.21%.
9	(Saputro & Sumantri, 2022)	<i>Hue Saturation Value (HSV)</i>	<i>Fuzzy K-Nearest Neighbor (FKNN)</i>	<p>Penelitian ini mengembangkan sistem pengenalan Sistem Isyarat Bahasa Indonesia (SIBI) untuk membantu komunikasi antara orang tuli atau bisu dengan orang normal, yang terdiri dari tiga tahap utama: akuisisi citra, preprocessing, dan pengenalan. Metode klasifikasi yang digunakan adalah Fuzzy K-Nearest Neighbor (FKNN), yang menghasilkan akurasi 88,8% berdasarkan lima kali percobaan dengan 130 citra, di mana 104 citra digunakan sebagai data latih dan 26 sebagai data uji. Beberapa huruf seperti C, E, L, U, dan V masih sulit dikenali karena kemiripan tinggi. Untuk meningkatkan akurasi,</p>

No	Peneliti	Metode Ekstraksi Fitur	Metode Klasifikasi	Hasil Penelitian
				<p>disarankan untuk mengembangkan metode atau ekstraksi fitur lain yang relevan, serta mengembangkan pengenalan SIBI secara real-time di masa mendatang.</p>
10	(Yuan et al., 2020)	<p><i>surface Electromyography (sEMG)</i></p>	<p><i>Random Forest, ANN, SVM</i></p>	<p>Membahas pengenalan Bahasa Isyarat Tionghoa (<i>chinese sign language</i>) melalui penggunaan sinyal <i>surface Electromyography (sEMG)</i>, dipasang sebanyak 8 sensor <i>sEMG</i> pada lengan bawah kanan subjek dan mengumpulkan sinyal <i>sEMG</i> saat dilakukan peragaan 30 huruf alfabet. Algoritma <i>random forest, ANN, SVM</i> digunakan untuk mengklasifikasikan data setelah proses ekstraksi fitur. Hasil pengujian menunjukkan bahwa algoritma random forest mencapai tingkat pengenalan rata-rata sebesar 95.48%, yang lebih tinggi daripada</p>

No	Peneliti	Metode Ekstraksi Fitur	Metode Klasifikasi	Hasil Penelitian
				<i>artificial neural networks</i> dan <i>support vector machine</i> yang memiliki hasil masing-masing 92,56% dan 94,68.
11	(Yasumuro & Jin'no, 2022)	<i>Mediapipe Hands</i>	<i>SVM</i>	Mengembangkan sistem pengenalan abjad bahasa isyarat hiragana dari Jepang dengan menggunakan informasi titik tangan dua dimensi yang diperoleh dari <i>mediapipe hands</i> . Proses pengenalan dilakukan dengan mengurangi jumlah komputasi menggunakan hanya dua dimensi dari koordinat 3D yang diperoleh dari <i>mediapipe</i> , dan kemudian dilanjutkan dengan pengenalan menggunakan <i>SVM linear</i> . Kinerja sistem dievaluasi menggunakan 41 abjad hiragana dengan lebih dari 1200 citra dari setiap karakter yang disiapkan untuk percobaan. Data dibagi menjadi set pelatihan dan validasi, di mana 80%

No	Peneliti	Metode Ekstraksi Fitur	Metode Klasifikasi	Hasil Penelitian
				digunakan untuk pelatihan dan 20% untuk validasi. Hasil eksperimen menunjukkan bahwa metode yang diusulkan menggunakan <i>SVM</i> mencapai akurasi pengenalan sebesar 100% untuk ejaan alfabet dan 99,9% untuk ejaan abjad hiragana.
12	(Alvin et al., 2021)	<i>Mediapipe Hands</i>	<i>K-NN</i>	Metode yang digunakan adalah metode klasifikasi <i>K-Nearest Neighbor (KNN)</i> , yang diimplementasikan dengan menggunakan <i>library scikit-learn</i> . Model <i>KNN</i> dibuat dengan menggunakan 3 tetangga terdekat. <i>Dataset</i> yang digunakan dalam penelitian ini terdiri dari 87.000 gambar dengan 21 <i>landmark points</i> yang diperoleh dari <i>mediapipe hands</i> . <i>Dataset</i> tersebut dibagi menjadi dua set, yaitu set pelatihan dan pengujian, dengan rasio pembagian 80:20.

No	Peneliti	Metode Ekstraksi Fitur	Metode Klasifikasi	Hasil Penelitian
				Untuk meningkatkan kinerja algoritma KNN, data diukur menggunakan metode <i>StandardScaler</i> dari <i>library scikit-learn</i> . Hasil dari eksperimen menunjukkan bahwa model <i>KNN</i> mencapai akurasi sebesar 94,4% pada skala 0-1, yang menunjukkan kualitas yang baik dalam mendeteksi isyarat tangan.

2.3 Matriks Penelitian

Berdasarkan penelitian terkait yang telah dirangkum sebelumnya, dapat disimpulkan dalam bentuk matriks penelitian seperti yang ditunjukkan dalam Tabel 2. 2.

Tabel 2. 2 Matriks penelitian

No	Penelitian	Bahasa		Metode Ekstraksi Fitur					Metode klasifikasi						
		Indonesia	Asing	Mediapipe holistic	CNN	HSV	HOG	sEMG	Mediapipe hands	Random forest	KNN	SVM	CNN	ANN	Transfer learning
1	(Wiraswendr o & Soetanto, 2022)	✓		✓						✓					

8	(Larasati et al., 2021)		✓				✓			✓					
9	(Saputro & Sumantri, 2022)	✓				✓					✓				
10	(Yuan et al., 2020)		✓					✓		✓		✓		✓	
11	(Yasumuro & Jin'no, 2022)		✓						✓						
12	(Alvin et al., 2021)		✓						✓		✓	✓			
13	Penelitian ini		✓						✓	✓					

Penelitian ini mengacu pada beberapa penelitian sebelumnya yang menyarankan untuk menggunakan metode ekstraksi fitur dengan pemilihan fitur selain warna dan pemilihan algoritma klasifikasi yang relevan. Pada penelitian ini digunakan *random forest* sebagai algoritma klasifikasi dan *mediapipe hands* sebagai metode ekstraksi fitur dengan harapan bisa meningkatkan performa model dari penelitian sebelumnya.