

BAB II

TINJAUAN PUSTAKA

2.1 Landasan Teori

2.1.1 Tunarungu dan Tuli

Tunarungu adalah orang yang gangguan pendengarannya sedemikian rupa sehingga mengganggu proses komunikasi bahasa melalui pendengarannya, meskipun menggunakan alat bantu dengar. Gangguan pendengaran bisa dikategorikan menjadi beberapa derajat mulai dari ringan, sedang, berat, atau mendalam (Rahmah, 2018).

Biasanya mereka tidak ingin dipanggil ‘tunarungu’ dan lebih memilih dengan panggilan ‘Tuli’ karena menurut mereka dengan panggilan ini lebih nyaman dan tidak terkesan diskriminatif (PSIBK, 2018).

Tabel 2.1 Perbedaan Tunarungu dan Tuli (PSIBK, 2018)

‘Tunarungu’	‘Tuli’
Ketidakmampuan mendengar	Kemampuan berbahasa isyarat
Kerusakan indera pendengaran	Dapat digelarkan kepada siapapun
Diagnosa medis	Sebuah identitas sosial
Terkesan lebih diskriminatif	Lebih nyaman terdengar

Hal ini karena penulisan Tuli dengan huruf kapital ‘T’ sekaligus sapaan Tuli, menunjukkan identitas orang Tuli sebagai sebuah kelompok masyarakat yang mempunyai identitas, memiliki bahasa, dan budayanya tersendiri. Manakala tunarungu dianggap sebagai sebuah keharusan untuk mengoptimalkan kemampuan pendengarannya dengan berbagai cara agar menyerupai orang-orang yang dapat mendengar (PSIBK, 2018).

2.1.2 Bahasa Isyarat

Bahasa isyarat adalah seperangkat gerakan tangan yang terstruktur dengan makna tertentu yang digunakan oleh para penyandang tuna rungu untuk berkomunikasi dalam kehidupan sehari-hari (Konstantinidis, Dimitropoulos dan Daras, 2018). Bahasa ini biasanya digunakan sebagai media komunikasi bagi individu-individu yang memiliki keterbatasan kemampuan fungsional indra seperti para Tuli. Selain itu, bahasa ini dapat dipelajari dan digunakan oleh masyarakat pada umumnya untuk bisa berkomunikasi dengan para Tuli.

Bahasa isyarat hampir dimiliki oleh seluruh negara di dunia, dan setiap bahasa isyarat memiliki gerakan dan gestur khasnya masing-masing bersesuaian dengan bahasa negaranya. Salah satu contoh bahasa isyarat adalah *American Sign Language* (ASL) dari Amerika Serikat, *Русский Жестовый Язык* (*Russkiy Zhestovyi Yazyk*) dari Rusia, *Türk İşaret Dili* dari Turki dan banyak lagi bahasa isyarat lainnya (Kimmelman, 2019; Halder dan Tayade, 2021).

Indonesia juga termasuk negara yang memiliki bahasa isyaratnya tersendiri. Terdapat dua bahasa isyarat yang umum digunakan di Indonesia yaitu salah satunya adalah Sistem Isyarat Bahasa Indonesia (SIBI).

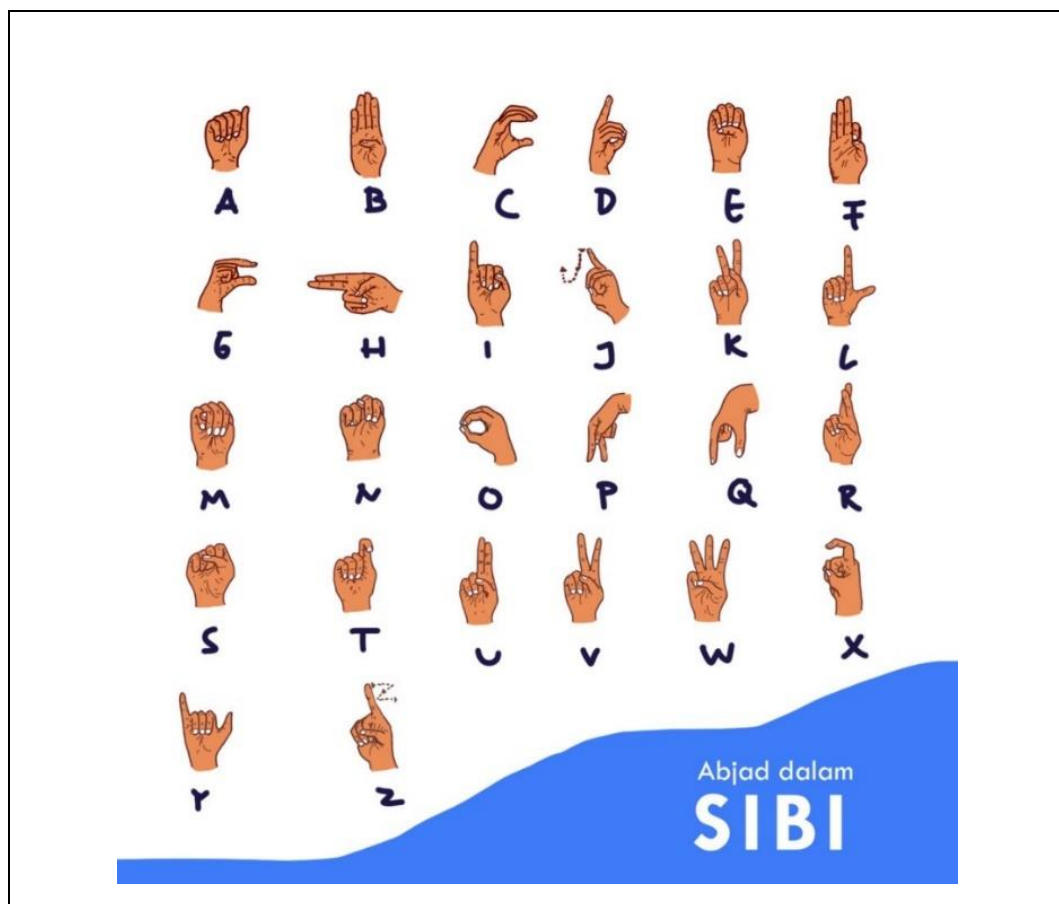
2.1.2.1 *Fingerspelling*

Fingerspelling atau ejaan jari merupakan salah satu bagian dari sebuah sistem bahasa isyarat. *Fingerspelling* di mana kata-kata dihasilkan oleh serangkaian bentuk tangan atau lintasan yang sesuai dengan huruf tunggal (menggunakan abjad atau alfabet Latin). *Fingerspelling* digunakan terutamanya untuk objek yang tidak memiliki tanda bahasa isyaratnya sendiri, seperti kata benda yang tepat atau istilah teknis, yang sering kali merupakan kata-kata konten yang penting (Shi dkk., 2018).

2.1.3 SIBI

Sistem Isyarat Bahasa Indonesia (SIBI) merupakan media komunikasi yang umumnya digunakan untuk para Tuli di Indonesia dan menggabungkan bahasa lisan, isyarat, mimik, dan gerak lainnya, telah ditetapkan oleh pemerintah sebagai bahasa isyarat resmi yang digunakan di Sekolah Luar Biasa (SLB) (Nugraheni, Husain dan Unayah, 2021). Hal ini dapat dilihat dari aturan yang telah dikeluarkan oleh pemerintah pada tanggal 30 Juni 1994 silam, dengan keputusan Mendikbud No. 0161/U/2994 berisi tentang pembakuan SIBI telah meresmikan Kamus SIBI yang digunakan dalam institusi atau pendidikan formal untuk para Tuli.

Selain memiliki gerakan isyarat yang melibatkan pergerakan kedua belah tangan, SIBI juga memiliki bahasa isyarat individual untuk abjad (*fingerspelling*) seperti yang bisa dilihat pada Gambar 2.1.

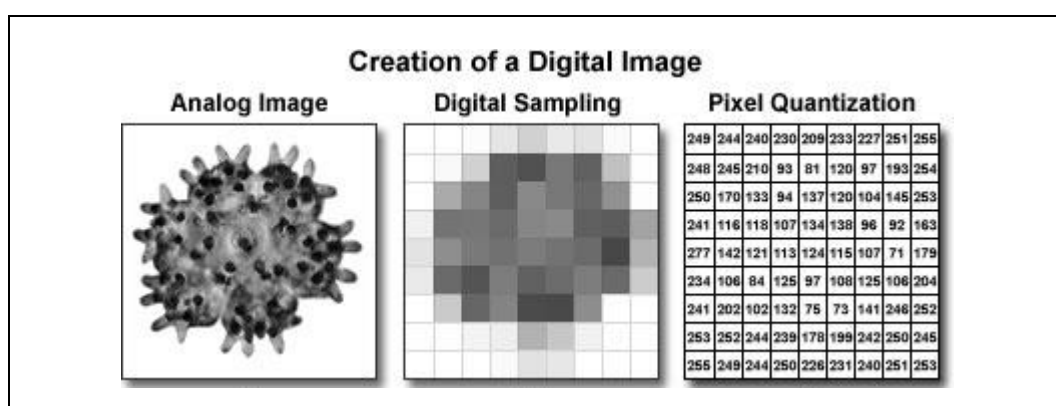


Gambar 2.1 *Fingerspelling* dalam SIBI (Lestari, 2018)

Gambar tersebut menunjukkan gestur tangan yang masing-masing memiliki arti abjadnya tersendiri. Terdapat 26 abjad yang terdiri daripada 24 gestur statis dan 2 gestur yang bergerak yaitu di J dan Z.

2.1.4 Citra dan Citra Digital

Citra adalah gambar tunggal yang mewakili sesuatu. Ini bisa berupa gambar seseorang, orang, atau hewan, pemandangan di luar ruangan, mikrofotograf komponen elektronik, atau hasil pencitraan medis. Bahkan jika gambarnya tidak langsung dikenali, gambar tersebut tidaklah berakhir dengan sebuah gambar *blur* yang *random* atau tiada artinya (McAndrew, 2015).

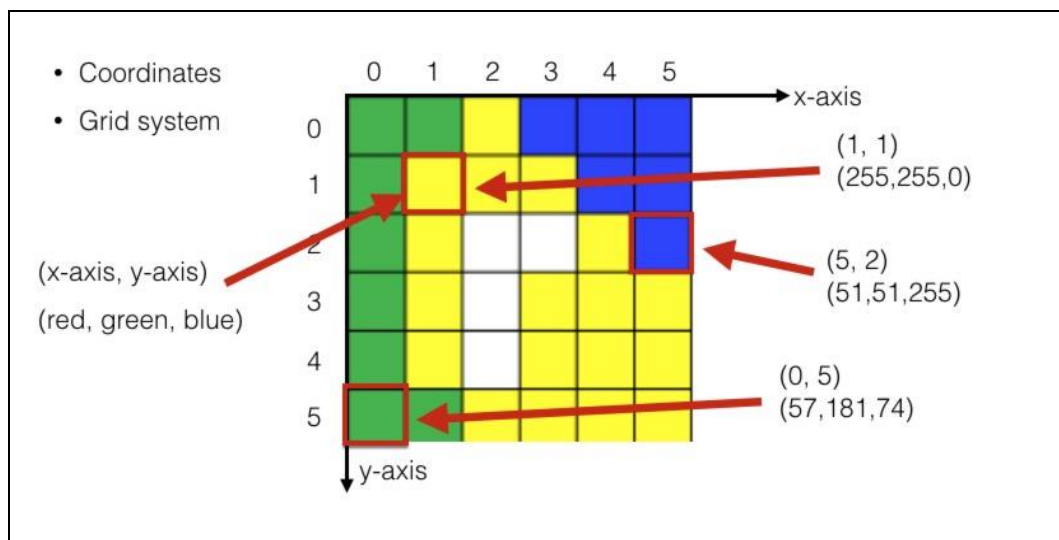


Gambar 2.2 Proses pengubahan citra ke citra digital (Spring dan Davidson, 2012)

Gambar digital (citra digital) dan video digital masing-masing adalah gambar dan filem yang telah dikonversi ke dalam format biner yang dapat dibaca komputer, yang terdiri atas logika 0 dan 1. Biasanya, yang dimaksud dengan gambar adalah gambar statis yang tidak berubah seiring dengan waktu, sedangkan video berkembang seiring dengan waktu dan umumnya berisi objek yang bergerak dan/atau berubah (Bovik, 2005).

Menurut sudut pandang matematis, gambar dapat didefinisikan sebagai fungsi dua dimensi, di mana x dan y adalah koordinat spasial (bidang), dan amplitudo pada setiap pasangan koordinat (x, y) disebut intensitas atau tingkat

keabuan gambar pada titik itu. Titik tersebut juga bisa dikenal dengan piksel. Apabila x , y , dan nilai amplitudo semuanya terbatas, kuantitas diskrit, gambar tersebut dapat dikenali sebagai citra digital (Gonzales dan Woods, 2003).



Gambar 2.3 Representasi citra digital (Parlante dan Stanford University, 2014)

Sehubungan dengan itu, citra digital juga direpresentasikan dalam bentuk peta yang terdiri daripada grid dan elemen piksel yang membentuk sebuah matriks dua dimensi. Setiap piksel mewakili intensitas kecerahan titik dan memiliki nilai yang mewakili posisi piksel dan nilai yang mewakili komponen warna yang biasanya dapat disebut sebagai *channel*. Bilangan channel adalah berbeda, tergantung dengan jenis citra yang didapatkan. Citra *grayscale* memiliki 1 channel warna manakala citra berwarna atau bisa dikenal dengan citra RGB memiliki 3 channel warna yang terdiri daripada merah (R), hijau (G) dan biru (B) untuk setiap pikselnya (Shawal, Shoyab dan Begum, 2014).

2.1.5 *Machine Learning*

Machine Learning (Pembelajaran Mesin) adalah bagian dari ilmu kecerdasan buatan (Faggella, Bengio dan Faggella, 2019), yang membuat komputer belajar dan bertindak seperti yang dilakukan manusia, dan meningkatkan pembelajaran mereka dari waktu ke waktu secara otomatis, dengan menerima data dan informasi dalam bentuk pengamatan dan interaksi dunia nyata (Faggella, 2018).

Machine learning secara algoritmis adalah proses komputasi yang menggunakan data input untuk mencapai tugas yang diinginkan tanpa diprogram secara manual, menghasilkan hasil tertentu melalui algoritma “*soft coded*” yang secara otomatis mengubah atau mengadaptasi arsitekturnya melalui pengulangan untuk menjadi sistem yang semakin baik dalam mencapai tugas yang diinginkan. Proses adaptasi ini, disebut pelatihan (training), menyediakan sampel data input bersama dengan hasil yang diinginkan, memungkinkan algoritma untuk mengkonfigurasi dirinya sendiri secara optimal sehingga tidak hanya dapat menghasilkan hasil yang diinginkan ketika disajikan dengan input pelatihan, tetapi juga dapat menggeneralisasi untuk menghasilkan hasil yang diinginkan dari data baru yang sebelumnya tidak terlihat (Naqa dan Murphy, 2015).

2.1.6 *Computer Vision*

Computer Vision atau komputer visi adalah analisis otomatis melalui media gambar dan video oleh komputer untuk mendapatkan suatu pemahaman dunia melalui media tersebut dan terinspirasi dari kemampuan sistem penglihatan manusia (Dawson-Howe, 2014).

Visi komputer adalah ilmu yang memberikan komputer atau mesin lain dengan visi, atau kemampuan untuk melihat dan mengeksekusi perintah jika mendeteksi objek tertentu, sehingga mampu menampilkan objek digital dan mengumpulkan data. Fokus utama dari banyak bidang dalam visi komputer adalah pengembangan aplikasi yang dapat digunakan di dunia nyata, termasuk kontrol kualitas di bidang manufaktur, pengenalan karakter optik, sistem bantuan pengemudi, pengawasan, fotografi dan hiburan (Learned-Miller, 2013).



Gambar 2.4 Implementasi *Computer vision*

Secara visual komputer visi bisa melakukan beberapa pekerjaan yang tidak bisa dilakukan oleh manusia (Purno dan Wibowo, 2016), diantaranya:

- a. Visi komputer memiliki kemampuan untuk melihat data dalam bentuk piksel bahkan jika diantara piksel tersebut mempunyai perbedaan yang kecil.

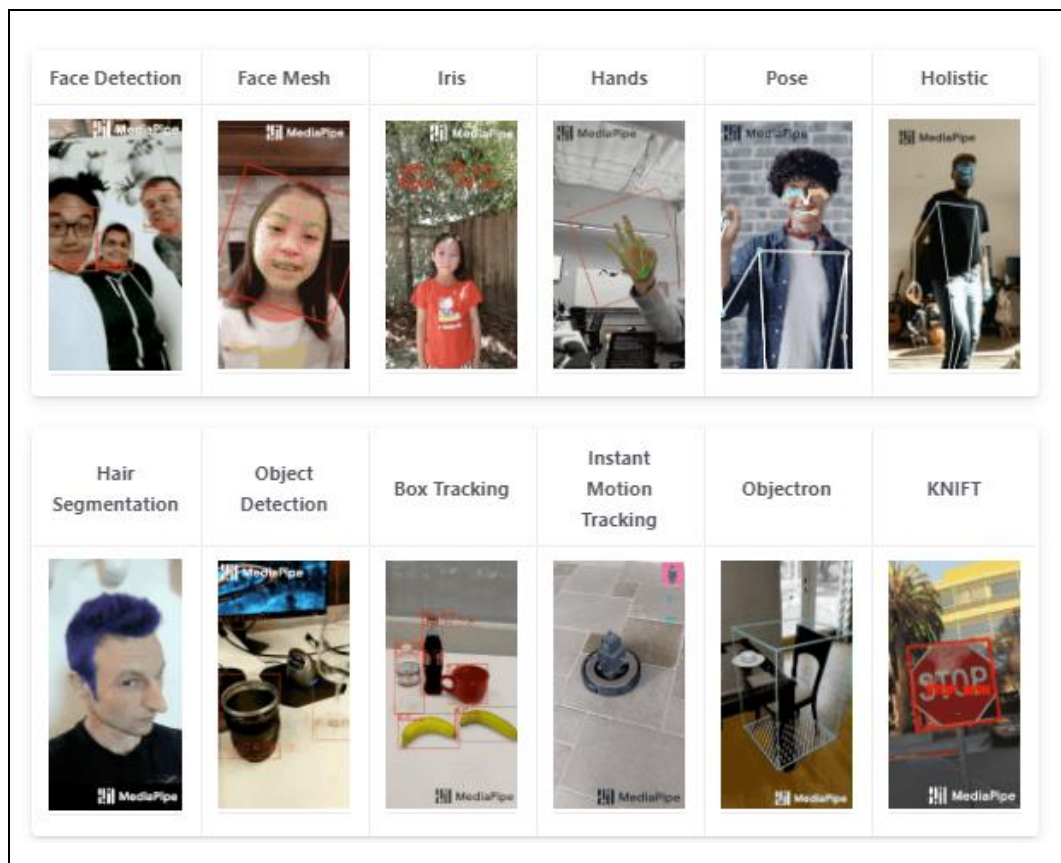
- b. Visi komputer mampu membandingkan dua buah objek gambar yang sama persis jika dilihat dari sudut pandang mata manusia.
- c. Visi komputer dapat melakukan penglihatan objek selama berjam-jam atau sehari-hari tanpa henti dengan proses yang sangat cepat. Visi sendiri merupakan suatu proses ekstraksi data objek yang berasal dari kamera dengan menggunakan algoritma tertentu.

Penelitian dan pengembangan *computer vision* telah berkembang sangat pesat dalam beberapa tahun terakhir. Hal ini dibuktikan dengan semakin banyaknya *library* yang mendukung program *computer vision* gratis seperti *OpenCV* dan *Mediapipe*. Penelitian dan pengembangan tersebut telah pun diterapkan dalam pelbagai aspek bidang, dimulai dari prototipe sehingga dalam penggunaan sehari-hari.

2.1.7 Mediapipe

Mediapipe adalah sebuah *framework computer vision* yang dikembangkan oleh *Google* yang dikhususkan untuk pengembangan multimedia (Lugaresi dkk., 2019). *Mediapipe* dirancang untuk praktisi *machine learning*, termasuk peneliti, mahasiswa dan pengembang perangkat lunak yang mengimplementasikan aplikasi *machine learning* siap produksi, menerbitkan kode yang terkait dengan masing-masing penelitian dan pembangunan prototipe proyek. Kasus penggunaan utama untuk *Mediapipe* adalah pembuatan prototipe cepat dari *pipeline* persepsi dengan model inferensi dan komponen lain yang dapat digunakan kembali (Lugaresi dkk., 2019).

Model *machine learning* siap produksi yang disediakan dalam *Mediapipe* di antara lainnya dimulai dari model untuk pendeteksian wajah, pergerakan mata, pendeteksian pergerakan sendi tangan, pose tubuh, deteksi objek dan masih banyak lagi. Adapun model *machine learning* yang lengkap dapat dilihat pada Gambar 2.5.

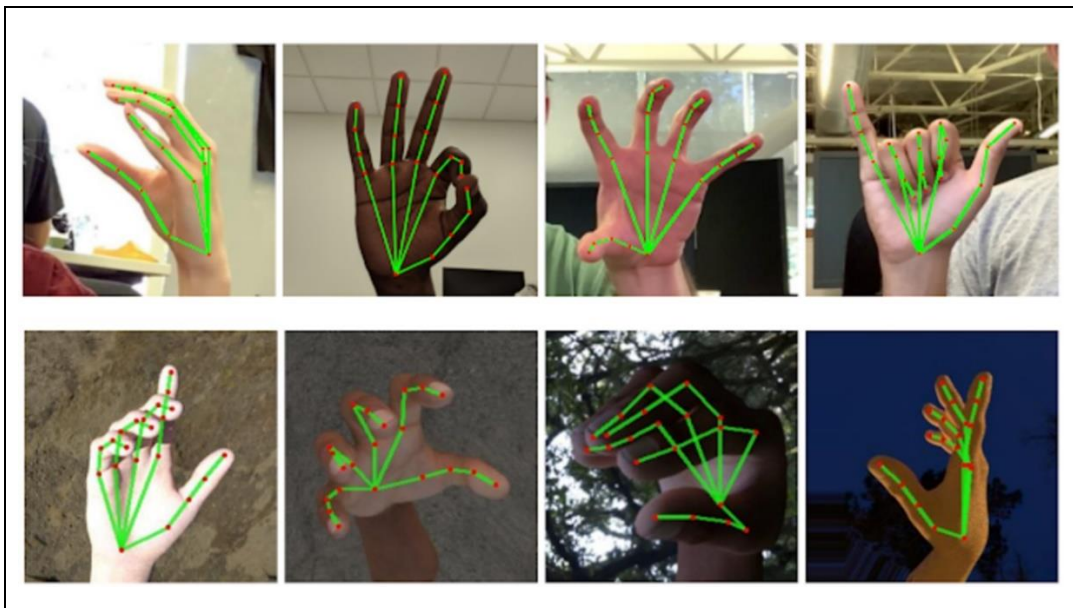


Gambar 2.5 Daftar solusi *framework Mediapipe* (Google, 2021)

Sejalur dengan penelitian ini, model *machine learning* yang digunakan adalah dengan model *Hands*, atau bisa disebut dengan *Mediapipe Hands*.

2.1.7.1 *Mediapipe Hands*

Mediapipe Hands adalah salah satu model *machine learning* yang terdapat dalam *framework Mediapipe* yang berfokus pada pendeteksian gestur tangan dengan menggunakan sebuah kamera RGB untuk pengaplikasian *augmented reality/virtual reality* dalam sistem *real-time* yang memprediksi sendi tangan manusia (Zhang dkk., 2020).



Gambar 2.6 Penggunaan *Mediapipe Hands* (Zhang dkk., 2020)

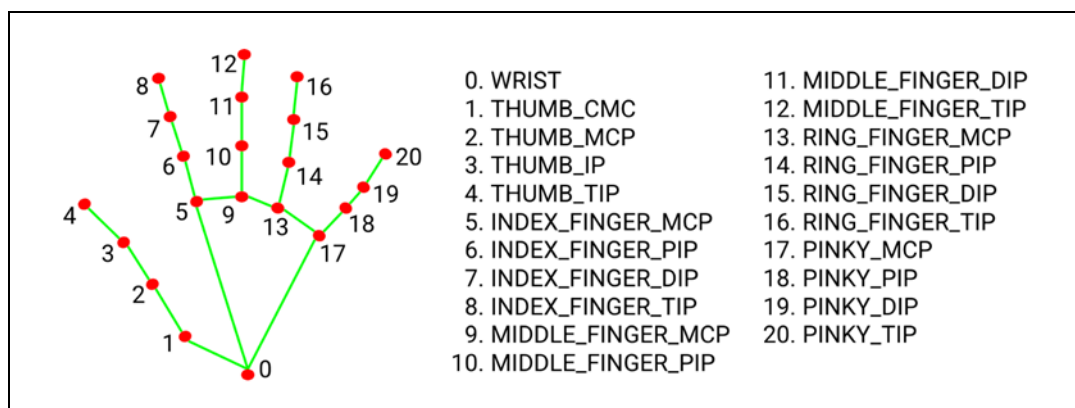
Secara jalan kerja *pipeline*-nya, *Mediapipe Hands* mengimplementasikan gabungan dua jenis model pendeteksian gestur tangan yang terdiri daripada sebagai berikut sesuai dengan urutan tahapan proses deteksi tangan *Mediapipe Hands*.

a. *BlazePalm Detector Model*

Model pendeteksi telapak tangan yang dikenal dengan *BlazePalm* untuk memproses tiap *frame* per detik dan mengubah setiap *frame* tersebut dengan kotak pembatas tangan yang mengikuti posisi tangan.

b. *Hand Landmark Detector Model*

Model pendeteksi *landmark* tangan hasil dari *frame* tangan yang telah diberi kotak pembatas sebelumnya. *Landmark* tangan ini akan memberikan output berupa koordinat tiga dimensi setiap 21 titik yang masing-masing terdiri dari koordinat x (lebar), y (tinggi) dan z (kedalaman) *landmark* seperti yang ditunjukkan dalam Gambar 2.7. Nilai dari koordinat tersebut juga dinormalisasikan menjadi 0.0 hingga 1.0 berdasarkan posisi relatif dengan ukuran citra digital.



Gambar 2.7 Landmark pada *Mediapipe Hands* (Google, 2020)

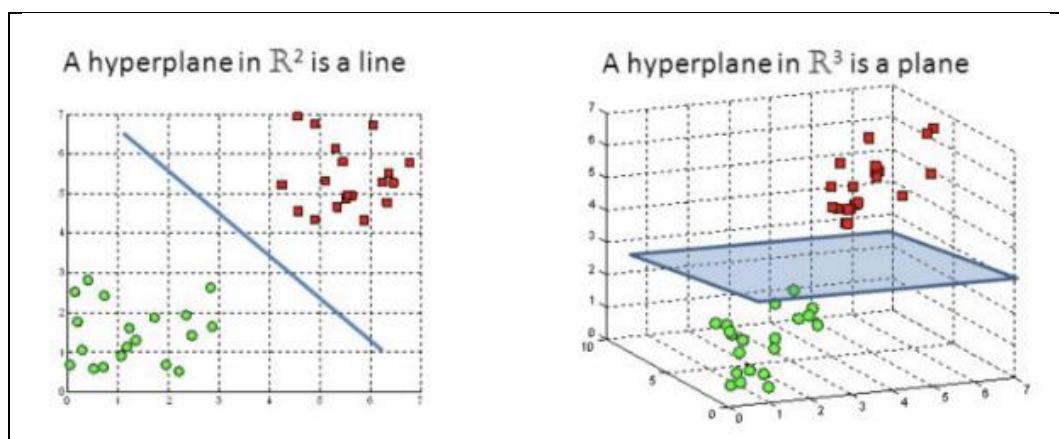
2.1.8 *Support-vector Machine (SVM)*

Support-vector Machine atau dapat disingkat dengan SVM adalah algoritma *machine learning* bertipe *supervised* yang menggunakan metode klasifikasi untuk permasalahan klasifikasi berkelompokan (Cervantes dkk., 2020).

SVM umum digunakan karena menghasilkan akurasi yang signifikan dengan daya komputasi yang lebih sedikit. SVM juga dapat menjalankan operasi regresi dan klasifikasi serta dapat memecahkan masalah praktis bersifat linear dan nonlinear dengan kinerja yang cukup baik. Namun begitu, algoritma ini secara luas digunakan dalam klasifikasi (Pisner dan Schnyer, 2019).

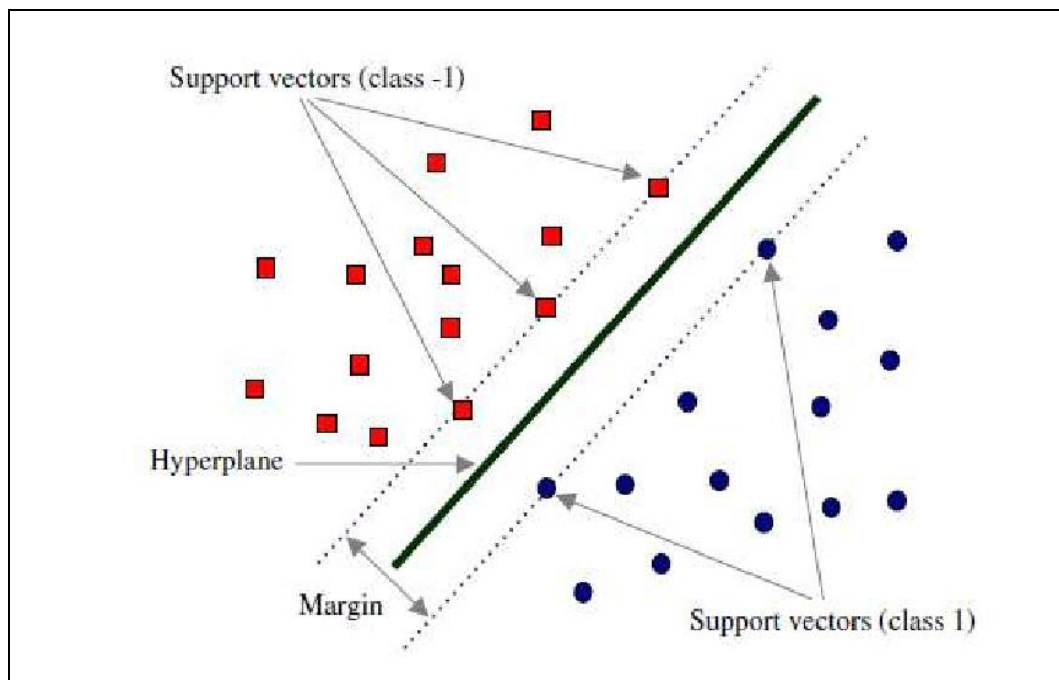
Secara umum, *Support Vector Machine* (SVM) berfungsi dengan mencari *hyperplane* pemisah terbaik antara data dari dua kelas yang berbeda dan berperan sebagai batas keputusan yang dapat mengklasifikasikan titik data, dimana titik-titik data yang jatuh di kedua sisi *hyperplane* dikaitkan dengan kelas yang berbeda (Cervantes dkk., 2020).

Selain itu, dimensi *hyperplane* tergantung pada jumlah fitur. Jika jumlah fitur input adalah 2, maka *hyperplane* hanyalah sebuah garis. Jika jumlah fitur input adalah 3, maka *hyperplane* menjadi bidang dua dimensi. Fitur input ini dapat berjumlah lebih daripada 3 tetapi membuat visualisasi menjadi lebih sulit (Pisner dan Schnyer, 2019).



Gambar 2.8 Bentuk *hyperplane* dalam dimensi 2 (kiri) dan 3 (kanan)

Hyperplane dalam SVM bekerja mengoptimalkan pemisahan dengan cara membuat batas margin dari *hyperplane* tersebut. Tahap ini dilakukan dengan cara menemukan titik-titik yang paling dekat dengan garis dari kedua kelas yang disebut sebagai *support vectors*. Selain itu, *support vectors* mempengaruhi posisi dan orientasi *hyperplane*. Dengan menggunakan vektor-vektor pendukung ini, margin dari *classifier* dapat dimaksimalkan. Menghapus vektor-vektor pendukung akan mengubah posisi *hyperplane*. Ini adalah titik-titik yang membantu membangun SVM (Pisner dan Schnyer, 2019).



Gambar 2.9 *Hyperplane* yang telah diberikan margin

2.2 *State-of-the-Art*

Seiring dengan kemajuan teknologi pengenalan gerakan dan *machine learning*, metode deteksi bahasa isyarat telah mengalami perkembangan yang signifikan, tak kira diperuntukkan bahasa asing ataupun SIBI. Secara umum pembahasan metode deteksi abjad bahasa isyarat dalam penelitian-penelitian waktu terdekat dapat dibagi menjadi dua kategori, yaitu berbasis alat (*sensor-based*) dan berbasis komputer visi (*vision-based*). Sehingga kini kumpulan penelitian dalam waktu terdekat telah didominasi oleh sistem deteksi yang menggunakan proses secara *vision-based*, dengan tujuan untuk meneliti hasil performa *machine learning* dari data yang dihasilkan melalui metode ekstraksi fitur gestur tangan abjad bahasa isyarat setelah dideteksi. Setiap penelitian memiliki metode masing-masing dalam mengekstraksi fitur gestur tangan dan mampu mendeteksi hampir keseluruhan gestur abjad berbagai bahasa yang ada dengan baik.

Penelitian ke-1 (Anwar dkk., 2018) dengan judul “*Feature Extraction for Indonesian Sign Language (SIBI) Using Leap Motion Controller*” dalam *21st International Computer Science and Engineering Conference (ICSEC) 2017, Vol. 6, Pg. 196-200*, membahas tentang sebuah metode ekstraksi fitur menggunakan *Leap Motion Controller* untuk mendeteksi data abjad SIBI yang diklasifikasikan dengan algoritma KNN dalam mendapatkan data gestur yang lebih akurat. Metode ini mengubah input menjadi yang diterima dari *Leap Motion Controller* menjadi sebuah kumpulan data koordinat 3D terdiri daripada nilai x, y dan z dari 16 titik buku jari dan tangan. Sistem ini menggunakan *point referencing* dengan telapak

tangan (titik 1) sebagai sumber *referencing*-nya agar setiap perubahan titik tidak dipengaruhi oleh pergerakan gestur di sekitar sensor. Keseluruhan data 26 abjad SIBI diambil melalui *Leap Motion Controller*. Sistem yang diusulkan berhasil mencapai tingkat akurasi rata-rata sebesar 95,15% untuk pengenalan bahasa isyarat menggunakan klasifikasi KNN dan 93,85% menggunakan klasifikasi SVM. Algoritma KNN mencapai akurasi tertinggi sebesar 95,15% dengan $k=1$, sementara algoritma SVM dengan kernel *Radial Basis Function* (RBF) menghasilkan akurasi sebesar 93,85%. Baik gerakan statis maupun dinamis dari gestur tangan dikenali dengan baik oleh sistem. Namun, terdapat beberapa kesalahan klasifikasi pada beberapa fitur dari gestur-gestur tersebut, khususnya untuk gerakan A, E, M, N, dan S.

Penelitian ke-2 (Rahagiyanto, 2019) dengan judul “Identifikasi Ekstraksi Fitur untuk Gerakan Tangan dalam Bahasa Isyarat (SIBI) Menggunakan Sensor *MYO Armband*” dalam Jurnal MATRIK, Vol. 19 No.1 (Nov) 2019, Hal 127-137, membahas tentang identifikasi dan ekstraksi fitur dari data mentah yang dihasilkan oleh sensor *MYO Armband* untuk gerak tangan dalam bahasa isyarat SIBI. Penelitian ini menggunakan lima buah sensor (akselerometer, giroskop, orientasi, orientasi-euler, dan EMG) dan menerapkan metode min-max untuk normalisasi data dan metode *Moment Invariant* untuk ekstraksi fitur. Pengujian dilakukan menggunakan gerakan bahasa isyarat statis (mulai dari abjad A hingga F) dan dinamis (gerakan *forward-backward*, *right-left* dan *up-down*). Hasil penelitian menunjukkan bahwa perbedaan antara gerakan statis dan dinamis dapat ditentukan dengan baik, menunjukkan keberhasilan metode yang diusulkan dalam

mengenali gerakan tangan dalam bahasa isyarat SIBI dengan akurasi sebesar 70%. Berdasarkan hal ini, menggunakan keseluruhan sensor yang terdapat pada *MYO Armband* dapat menghasilkan akurasi yang lebih baik dibandingkan dengan menggunakan subset sensor.

Penelitian ke-3 (Jiang dan Ahmad, 2019) dengan judul “*Hand Gesture Detection Based Real-Time American Sign Language Letters Recognition Using Support Vector Machine*” dalam *IEEE Intl. Conf. on Dependable, Automatic and Secure Computing* membahas tentang performa metode deteksi abjad bahasa isyarat ASL yang diambil secara *realtime* menggunakan SVM. Penelitian ini menggunakan pendekatan algoritma deteksi warna kulit dalam ruang warna HSV untuk menemukan *Region of Interest (ROI)* di mana gerakan tangan berada. Fitur-fitur diekstraksi menggunakan *Principal Component Analysis (PCA)* setelah melalui tahap *preprocessing*. Penelitian ini juga membahas metode penentuan rotasi, penentuan rentang warna kulit yang adaptif, penghilangan *noise* dan penggunaan klasifikasi SVM secara *one vs. one*. Hasil dari penelitian ini menunjukkan bahwa sistem berhasil membedakan abjad-abjad ASL tertentu seperti B, D, F, L dan U dengan tingkat keberhasilan sekitar 99,4%.

Penelitian ke-4 (Darmatasia, 2021) dengan judul “Pengenalan Sistem Isyarat Bahasa Indonesia (SIBI) Menggunakan *Gradient-Convolutional Neural Network*” dalam *Jurnal INSTEK (Informatika Sains dan Teknologi)* Vol. 6, No. 1, April 2021 membahas tentang performa *Convolutional Neural Network* yang dimodifikasi dengan filter *gradient* atau bisa dikenal dengan *Gradient-Convolutional Neural Network* dalam mendeteksi abjad SIBI. Filter ini digunakan

untuk memperjelas fitur-fitur pada objek. Arsitektur CNN yang diusulkan terdiri dari lima lapisan, termasuk dua pasang lapisan konvolusi dan *sub-sampling*, serta lapisan terhubung penuh sebagai pengklasifikasi. Objek input adalah data video abjad SIBI yang dikonversi menjadi 600 gambar yang terdiri daripada 5 abjad yaitu A hingga E. Hasil penelitian menunjukkan bahwa pemberian filter dapat meningkatkan akurasi deteksi yaitu sekitar 85% dibandingkan dengan CNN murni yang hanya memperoleh akurasi sebesar 25%. Akurasi terbaik yang diperoleh yaitu sebesar 98% dengan meningkatkan jumlah iterasi. Metode ini juga diuji menggunakan tiga *benchmark* dataset dan menunjukkan peningkatan akurasi pengenalan pada dataset yang memiliki *background* yang kompleks.

Penelitian ke-5 (Zhang dkk., 2020) dengan judul “*Mediapipe Hands: On-device Real-time Hand Tracking*” dalam *Google Research* (2020) merupakan penelitian terpenting dalam perkembangan metode deteksi abjad bahasa isyarat seterusnya. Penelitian ini membahas tentang pengembangan solusi komputer visi baru, *Mediapipe Hands*, untuk mendeteksi tangan secara *realtime* menggunakan *machine learning* dan kamera RGB. Solusi ini menggunakan dua model yang berjalan secara bersamaan. Model pertama mendeteksi telapak tangan dan memangkas citra berdasarkan ukuran tangan yang terdeteksi. Model kedua mendeteksi buku jari dan menempatkan 21 titik *landmark* berdasarkan posisi jari ini. Proses pelathina menggunakan dataset citra gambar dari dunia nyata dan sintetis. Penelitian ini juga menjelaskan proses pengujian model secara *realtime* dan menunjukkan bahwa *framework* ini memiliki tingkat kekokohan model yang cukup baik.

Penelitian ke-6 (Halder dan Tayade, 2021) dengan judul “*Real-time Vernacular Sign Language Recognition using Mediapipe and Machine Learning*” dalam *Internation Journal of Research Publication and Reviews Vol (2) Issue (5) (2021) Page 9 – 17* merupakan salah satu penelitian terawal yang menggunakan *Mediapipe* sebagai metode ekstraksi fitur untuk gestur abjad bahasa isyarat. Penelitian ini merancang metode deteksi bahasa isyarat yang lebih efisien menggunakan algoritma seperti *Support-vector Machine, K-Nearest Neighbours, Random Forest, Decision Tree, Naïve Bayes, ANN* dan *MLP. Mediapipe Hands* digunakan untuk mendeteksi gestur tangan dan mengambil data gambar tangan. Jumlah gambar yang diambil berbeda untuk setiap bahasa isyarat. Hasil penelitian menunjukkan bahwa algoritma SVM unggul dengan rata-rata nilai akurasi sebesar 99% untuk semua bahasa isyarat, lebih baik dibandingkan dengan algoritma lainnya dan hasil penelitian terkait sebelumnya.

Penelitian ke-7 (Alvin dkk., 2021) dengan judul “*Hand Gesture Detection for Sign Language using K-Nearest Neighbor with Mediapipe*” dalam *Ultima Computing, Jurnal Sistem Komputer, Vol. 13, No.2, Desember 2021* membahas tentang *framework Mediapipe Hands* yang memungkinkan pengguna berinteraksi dengan komputer menggunakan abjad bahasa isyarat ASL. Penelitian ini menggunakan *library Scikit-learn* untuk membuat model klasifikasi KNN dan menggunakan 3 *neighbours* dengan algoritma otomatis, fungsi bobot seragam dan metrik *Minkowski* sebagai pengaturan *default*. Dataset terdiri dari 87.000 gambar dengan 21 landmark points yang diperoleh dari *Mediapipe Hands*. Dataset ini dibagi menjadi dua set, yaitu set pelatihan dan pengujian dengan rasio 80:20. Data

diukur menggunakan metode *StandardScaler* dari *library Scikit-learn* untuk meningkatkan kinerja algoritma KNN. Hasil yang didapatkan mencapai sebesar 94,4% pada skala 0-1, menunjukkan akurasi yang baik untuk model deteksi isyarat. Namun, beberapa abjad seperti J dan Z sulit dideteksi dengan akurat, mengindikasikan adanya keterbatasan dalam mendeteksi dengan akurat beberapa isyarat tertentu.

Penelitian ke-8 (Shin dkk., 2021) dengan judul “*American Sign Language Alphabet Recognition By Extracting Feature From Hand Pose Estimation*” dalam *MDPI Sensors* 2021, 21, 5856 membahas tentang pengenalan abjad bahasa isyarat ASL menggunakan citra tangan yang diperoleh dari kamera *webcam*. Penelitian ini menggunakan *framework Mediapipe Hands* untuk mendeteksi tangan dan menghasilkan dua jenis fitur yaitu jarak antara titik sendi dan sudut antara vektor dan sumbu 3D. Penelitian ini membandingkan pendekatan *vision-based* dan *sensor-based* untuk deteksi bahasa isyarat. Meskipun pendekatan berbasis sensor mencapai tingkat deteksi yang lebih tinggi, pendekatan tersebut memakan biaya yang cukup besar, ruang pergerakan yang terbatas dan sensitive terhadap *noise*. Oleh karena itu, penelitian ini berfokus pada pendekatan berbasis *Mediapipe Hands* untuk estimasi tangan melalui *webcam*. Algoritma klasifikasi yang digunakan adalah SVM dan GBM. Tiga dataset digunakan yaitu dataset *ASL Alphabets*, dataset *Massey* dan dataset *Finger Spelling A*. Hasil penelitian menunjukkan akurasi tinggi untuk dataset *Massey* (99,39%), akurasi sedang untuk dataset *ASL Alphabet* (87,6%) dan akurasi tinggi untuk dataset *Finger Spelling A*

(98,45%). Metode yang diusulkan terbukti hemat biaya, tidak mahal secara komputasional dan tidak memerlukan sensor atau perangkat khusus.

Penelitian ke-9 (Yasumuro dan Jin'no, 2022) dengan judul “*Japanese Fingerspelling Identification By Using Mediapipe*” dalam *Nonlinear Theory and Its Applications, IEICE, Vol. 13, No. 2, pp. 288-293* mengusulkan sebuah sistem pengenalan abjad bahasa isyarat *hiragana* dari Jepang menggunakan informasi titik tangan dua dimensi yang diperoleh oleh *MediaPipe Hands* dan diidentifikasi oleh sebuah SVM linear. Sistem ini membutuhkan jumlah data masukan yang lebih sedikit dan mencapai akurasi pengenalan yang tinggi untuk ejaan isyarat tangan yang diam. Metode ini mengurangi jumlah komputasi dengan hanya menggunakan dua dimensi dari koordinat 3D yang diperoleh dari *Mediapipe* dan melakukan pengenalan dengan SVM linear. Kinerja sistem dievaluasi menggunakan 41 abjad *hiragana* dengan lebih dari 1200 contoh dari setiap karakter yang disiapkan untuk percobaan. Data dibagi menjadi set pelatihan dan validasi, dengan 80% digunakan untuk pelatihan dan 20% untuk validasi. Metode yang diusulkan menggunakan SVM mencapai akurasi pengenalan sebesar 100 untuk ejaan abjad jari dan 99,9% untuk ejaan abjad *hiragana*, hanya dengan sedikit kesalahan pengenalan.

Berdasarkan penelitian – penelitian ini, metode ekstraksi yang dibahas sangatlah bermanfaat sebagai elemen proses deteksi gestur tangan dalam sebuah aplikasi sistem deteksi bahasa isyarat secara *realtime* sehingga dapat digunakan oleh *user* secara langsung. Terdapat beberapa penelitian seterusnya yang telah melakukan pembuatan sistem ini dengan memanfaatkan berbagai macam metode

ekstraksi fitur sebagai elemen proses deteksinya dan dapat mendeteksi berbagai macam abjad bahasa isyarat dengan baik secara *realtime*.

Penelitian ke-10 (Humaira dkk., 2018) dengan judul “*Real Time SIBI Sign Language Recognition Based on K-Nearest Neighbor*” dalam *Proceeding of EECSI 2018*, Malang-Indonesia, 16-18 Oct. 2018 membahas tentang sebuah sistem pengenalan bahasa isyarat untuk Sistem Isyarat Bahasa Indonesia (SIBI) menggunakan pengontrol *Leap Motion Controller* dan algoritma *K-Nearest Neighbor* (KNN). Tujuannya adalah untuk meningkatkan komunikasi antara orang-orang dengan disabilitas, khususnya mereka yang menggunakan bahasa isyarat. *Leap Motion Controller* mendeteksi titik koordinat pada setiap tulang di tangan, yang digunakan sebagai fitur masukan. Jarak antara posisi telapak tangan dan falang distal diukur menggunakan jarak *euclidean*. Algoritma KNN digunakan untuk data pelatihan dan pengujian, dengan nilai K diatur sebagai 5. Hasil eksperimen menunjukkan bahwa sistem mencapai akurasi sebesar 0,78% dan tingkat kesalahan sebesar 0,22% dengan parameter $K = 5$. Namun terdapat beberapa huruf yang memiliki akurasi lebih rendah karena kontroler *leap motion* tidak mampu mendeteksi koordinat saat mereka terlindungi atau tertutup oleh jari-jari lain. Selain itu, ketika huruf-huruf memiliki data yang mirip, sistem menghadapi kesulitan dalam membedakannya.

Penelitian ke-11 (Yuan dkk., 2020) dengan judul “*Chinese Sign Language Alphabet Recognition Based on Random Forest Algorithm*” dalam jurnal IEEE di tahun 2020 membahas tentang sebuah sistem deteksi abjad Tionghoa menggunakan sinyal Elektromiografi Permukaan (sEMG) yang dikumpulkan dari

lengan bawah kanan subjek. Algoritma *random forest*, sebuah metode *machine learning*, digunakan untuk mengklasifikasikan data sEMG setelah proses penyaringan dan ekstraksi fitur. Data sEMG yang terkumpul kemudian digunakan untuk pengenalan alfabet Bahasa Isyarat Tiongkok secara *realtime*. Algoritma hutan acak diterapkan untuk mengklasifikasikan data dan mencapai akurasi pengenalan sebesar 95,48% untuk setiap abjad. Namun, beberapa isyarat yang mirip satu sama lain, seperti huruf P, R, T, dan Z, sulit untuk dikategorikan dengan akurat. Para peneliti juga menyebutkan dampak potensial dari *noise* eksternal, seperti artefak gerakan, lingkungan, dan garis listrik, terhadap peningkatan tingkat kesalahan dalam perhitungan vektor fitur.

Penelitian ke-12 (Anshary, Hidayat dan Amalia, 2020) dengan judul “*Prototype Program Hand Gesture Recognize Using the Convex Hull Method and Convexity Defect on Android*” dalam *Jurnal Online Informatika (JOIN)*, Vol. 5, No. 2, December 2020: 205-211 membahas tentang pengembangan pengenalan gerakan tangan pada ponsel pintar menggunakan sistem operasi *Android*. Algoritma *convex* digunakan untuk mengidentifikasi batas luar dari gerakan tangan, sementara algoritma *convexity defect* digunakan untuk mendeteksi wilayah cekung di dalam gerakan tangan. Data masukan terdiri dari gambar *realtime* dari satu tangan dengan posisi terbuka, yang diambil menggunakan ruang warna RGB dari kamera ponsel pintar. Pengujian dilakukan dalam berbagai skenario, termasuk pencahayaan yang bervariasi, warna latar belakang, dan kondisi dalam ruangan/luar ruangan. Hasilnya menunjukkan bahwa akurasi pengenalan gerakan tangan dipengaruhi oleh lingkungan pengujian. Untuk

penggunaan di luar ruangan, warna latar belakang hijau dengan intensitas cahaya sebesar 1725 lux menghasilkan akurasi sebesar 76,7%. Untuk penggunaan di dalam ruangan, warna latar belakang merah dengan intensitas cahaya sebesar 300 lux memberikan akurasi tertinggi sebesar 83,3%.

Penelitian ke-13 (Na, Yang dan Woo, 2021) dengan judul “*Classification of the Korean Sign Language Alphabet Using an Accelerometer with a Support Vector Machine*” dalam *Hindawi, Journal of Sensors, Vol. 2021, Article ID 9304925* membahas tentang suatu metode untuk mengenali abjad bahasa isyarat Korea menggunakan accelerometer yang dipasang pada jari telunjuk tangan dominan. Sinyal *accelerometer* digunakan untuk membagi gerakan isyarat dan mengklasifikasikan 31 huruf bahasa isyarat Korea. Algoritma SVM digunakan dalam penelitian ini dan mencapai tingkat akurasi tinggi sebesar 92,2% dalam mengklasifikasikan 31 abjad bahasa isyarat Korea. Beberapa pasangan huruf bahasa isyarat seringkali disalahartikan karena kesamaan dalam posisi tangan, seperti huruf ke-4 dan ke-12, ke-7 dan ke-9, serta ke-21 dan ke-22. Penelitian ini melibatkan 15 subjek (7 pria dan 8 wanita) tanpa gangguan muskuloskeletal di tangan dan tidak memiliki gangguan pendengaran. Subjek-subjek ini diminta untuk mengenakan sarung tangan dengan modul sensor gerakan yang ditempatkan pada jari telunjuk dan melakukan 31 huruf bahasa isyarat Korea. Para partisipan diminta untuk secara acak mengisaratkan 31 huruf bahasa isyarat Korea sebanyak 10 kali dalam dua sesi, menghasilkan total 620 gerakan untuk setiap subjek.

Penelitian ke-14 (Alsaadi dkk., 2022) dengan judul “*A Real Time Arabic Sign Language Alphabets (ArSLA) Recognition Model Using Deep Learning*”

Architecture dalam MDPI *Computers* 2022, Volume 11, Issue 78 menjelaskan tentang sebuah aplikasi sistem deteksi abjad bahasa isyarat Arab secara *realtime* menggunakan metode arsitektur *deep learning*. Dataset publik 32 abjad Arab (ArSLA) yang terdiri dari 54.049 citra *grayscale* berformat .jpg dengan ukuran 64x64 piksel digunakan sebagai data untuk proses pelatihan model. Salah satu varian algoritma CNN yang dikenali sebagai *AlexNet* telah dipilih sebagai metode *deep learning* terbaik dalam penelitian ini karena keunggulannya dibandingkan dengan algoritma CNN lainnya dengan hasil sebesar 99,75%. Varian algoritma ini dikompilasi menjadi sebuah model untuk diterapkan ke dalam sebuah sistem deteksi bahasa isyarat. Aplikasi sistem yang dibuat bekerja melalui tiga tahap yaitu diawali dengan penangkapan citra dari *webcam* secara *realtime*, proses ekstraksi *Region of Interest* (ROI) untuk menentukan batas ukuran gestur tangan melalui *contour* citra serta *convex hull*, penemuan *defects* berdasarkan jumlah kecacatan *confexity* dan diakhiri dengan penentuan gestur melalui algoritma dipilih. Hasil pengujian validasi secara *realtime* adalah sebesar 94.81% yang dimana sistem berhasil mendeteksi keseluruhan abjad dengan baik.

Penelitian ke-15 dari (Schmalz, 2022) dengan berjudul “*Real – time Italian Language Recognition with Deep Learning*” dalam *CEUR Workshop Proceedings* 2022, Volume 3078, Issue 17 menjelaskan tentang sebuah aplikasi sistem deteksi abjad bahasa isyarat Italia secara *realtime* yang juga menggunakan metode *deep learning*. Tujuan dari penelitian ini adalah untuk membuat sistem deteksi abjad bahasa isyarat Italia berdasarkan gestur – gestur abjad statis secara *realtime*. ataset yang digunakan dalam penelitian ini adalah dataset LIS yang

terdiri dari 22 gestur abjad. Metode *deep learning* yang digunakan adalah model VGG19.4 (*fine-tuned*) dengan hasil pelatihan sebesar 99% dari *F1-score*, *precision* dan *recall*. Kemudian model tersebut diterapkan ke dalam sebuah sistem deteksi *realtime*. Sistem ini menempatkan *bounding box* berukuran 28x28 piksel sebagai lokasi deteksi gestur tangan dan diproses secara *realtime*. Setelah pengujian, didapati akurasi keseluruhannya sebesar 77,2% yang ditentukan dengan rata-rata 3,46 detik untuk pengenalan yang benar dari setiap abjad.

Penelitian ke-16 (Rivera-Acosta dkk., 2021) dengan judul “*Spelling Correction Real-Time American Sign Language Alphabet Translation System Based On Yolo Network And LSTM*” dalam MDPI, *Electronics* 2021, 10, 1035 membahas tentang pengenalan abjad bahasa isyarat Amerika secara *realtime* menggunakan *Deep Neural Network* (DNN) dan jaringan *Long Short-Term Memory* (LSTM) dua arah. Dataset abjad ASL *open source* yang terdiri dari 87.000 gambar RGB dari 29 kelas digunakan untuk melatih model DNN. Model jaringan saraf dalam mendalam (DNN) yang dilatih mencapai *mean average precision* sebesar 81,74% dan laju pemrosesan sebesar 61,35 *frame-per-detik*. Arsitektur jaringan YOLO A2 (352 x 352) dipilih sebagai yang paling sesuai untuk sistem ASLAT, menawarkan waktu respons waktu nyata dan *mean average precision* yang lebih tinggi dibandingkan dengan model lain. DNN dan LSTM yang diimplementasikan untuk koreksi ejaan mencapai akurasi pelatihan sebesar 98,07%.

2.3 Matriks Penelitian

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

Tabel 2.2 Matriks Penelitian

No	Peneliti	Ruang lingkup																										
		Bahasa		Alat	Metode Ekstraksi Fitur						Metode Klasifikasi					Parameter Pengujian (Akurasi)												
		Bahasa Asing	Bahasa Indonesia (SIBI)	<i>Vision-based (Webcam)</i>	<i>Sensor-based</i>	<i>Mediapipe Hands</i>	<i>Myo Armband</i>	<i>Accelerometer Triaxial</i>	<i>Delys Trigno</i>	HSV	<i>Leap Motion Controller</i>	CNN / YOLO	<i>Min-max + moment invariant</i>	KNN	<i>SVM (Linear/RBF/GBM)</i>	<i>Random Forest</i>	<i>Decision Tree</i>	<i>Naive Bayes</i>	<i>ANN / MLP / Deep Learning</i>	<i>Convex Hull / Convexity Defect</i>	LSTM	Evaluasi pelatihan data	<i>Offtime</i>	<i>Realtime</i>	Satu posisi	Berbagai posisi	Tangan kanan	Tangan kiri
1.	(Anwar dkk., 2018)	✓		✓						✓			✓								✓	✓					✓	

2.	(Rahagiyanto, 2019)	✓			✓		✓						✓								✓	✓				✓			
3.	(Jiang dan Ahmad, 2019)		✓	✓					✓					✓								✓	✓		✓		✓		
4.	(Darmatasia, 2021)	✓		✓																	✓	✓		✓		✓			
5.	(Zhang dkk., 2020)			✓		✓																	✓		✓	✓	✓	✓	✓
6.	(Halder dan Tayade, 2021)		✓	✓		✓							✓	✓	✓	✓	✓	✓					✓	✓		✓		✓	
7.	(Alvin dkk., 2021)		✓	✓		✓							✓										✓	✓		✓		✓	
8.	(Shin dkk., 2021)		✓	✓		✓								✓									✓	✓		✓		✓	
9.	(Yasumuro dan Jin'no, 2022)		✓	✓		✓								✓									✓	✓		✓		✓	
10	(Humaira dkk., 2018)	✓			✓									✓									✓		✓	✓		✓	
11	(Yuan dkk., 2020)		✓		✓				✓															✓		✓	✓		✓

12	(Anshary, Hidayat dan Amalia, 2020)	✓		✓						✓								✓				✓	✓		✓	
13	(Na, Yang dan Woo, 2021)		✓		✓															✓	✓			✓		✓
14	(Alsaadi dkk., 2022)		✓	✓							✓							✓			✓		✓	✓		✓
15	(Schmalz, 2022)		✓	✓							✓							✓			✓		✓	✓		✓
16	(Rivera-Acosta dkk., 2021)		✓	✓							✓							✓	✓			✓	✓			✓
17	Penelitian ini	✓		✓		✓					✓									✓		✓	✓	✓	✓	✓

Berdasarkan tabel di atas, dapat disimpulkan bahwa penelitian ini menggunakan beberapa parameter penelitian sebelumnya yaitu bahasa, alat, metode ekstraksi fitur, metode klasifikasi dan jenis pengujian. Penelitian ini menggunakan abjad SIBI sebagai bahasa yang dipilih, menggunakan alat berbasis *vision-based (webcam)*, *Mediapipe Hands* sebagai metode ekstraksi fiturnya dan *Support Vector Machine (SVM)* sebagai metode untuk mengklasifikasi data SIBI. Adapun keterbaruan yang ada dalam penelitian ini

terletak pada jenis pengujiannya, yaitu pengujian akurasi deteksi tangan kanan dan tangan kiri secara *realtime* dan juga dilakukan dalam berbagai posisi.