

BAB II

TINJAUAN PUSTAKA

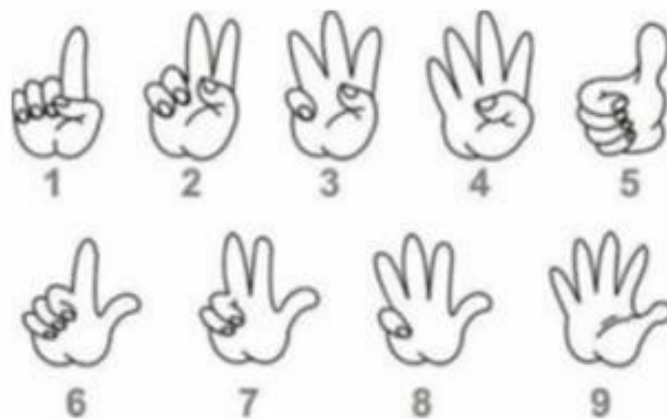
2.1 Landasan Teori

2.1.1 Jarimatika

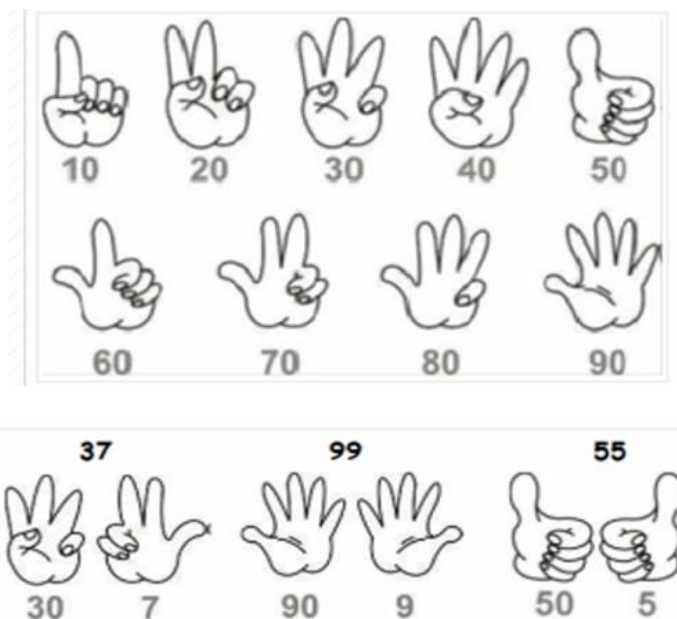
Jarimatika merupakan metode pembelajaran matematika dasar yang memanfaatkan jari tangan sebagai alat bantu untuk menghitung konsep penjumlahan, pengurangan, perkalian, dan pembagian tanpa menggunakan kalkulator atau mesin berhitung lainnya. Metode jarimatika telah diterapkan dalam berbagai sistem pendidikan di Indonesia, sebagai alat bantu siswa memahami operasi dasar matematika dengan lebih mudah dan menyenangkan (Pratikno dkk., 2023). Secara umum, jarimatika memanfaatkan posisi dan gerakan jari untuk menunjukkan angka atau nilai matematika. Terdapat 18 simbol dasar dalam jarimatika, jari pada tangan kanan berfungsi untuk mewakili satuan, sedangkan tangan kiri digunakan untuk mewakili puluhan. Dengan menggerakkan jari, siswa dapat memvisualisasikan proses perhitungan yang sedang berlangsung. Pendekatan ini tidak hanya membuat proses pembelajaran menjadi lebih interaktif, tetapi juga efektif dalam membantu anak-anak yang mengalami kesulitan dalam belajar matematika (Pratikno dkk., 2023).

Dalam bidang teknologi, terdapat beberapa penelitian yang menghubungkan jarimatika dengan aplikasi berbasis visi komputer untuk menciptakan interaksi yang lebih dinamis. Beberapa studi terbaru telah memanfaatkan teknologi seperti *Convulation Neural Network* (CNN) dan MediaPipe untuk meningkatkan

pengalaman belajar jarimatika (Pratikno dkk., 2023). Dengan menggunakan teknologi ini, proses pengenalan dan penerapan jarimatika menjadi lebih bervariasi dan menyenangkan (Pratikno dkk., 2023). Untuk pengenalan angka-angka menggunakan jari-jari tangan dengan metode jarimatika ditunjukkan pada gambar 2.1 dan gambar 2.2. Jari tangan kanan sebagai satuan dan jari tangan kiri sebagai puluhan. (Tahir & Halim, 2021).



Gambar 2. 1 Gestur Jarimatika Tangan Kanan (Nautica, 2022)

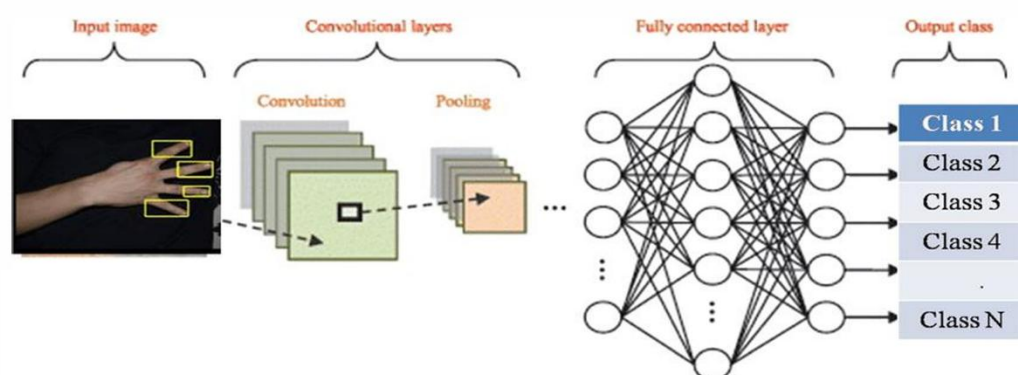


Gambar 2. 2 Gestur Jarimatika Tangan Kiri(Nautica, 2022)

2.1.2 Convolution Neural Network (CNN)

Convolution Neural Network (CNN) merupakan jenis jaringan syaraf tiruan yang banyak digunakan dalam pengolahan citra dan pengenalan pola. CNN memiliki lapisan konvolusi dan *pooling* yang memungkinkan mengidentifikasi fitur spasial dalam data yang kompleks. Dengan adanya CNN, cara sistem komputer dapat mengenali objek termasuk gestur tangan yang telah mengalami perubahan (Naufal dkk., 2021).

Dalam aspek pengenalan gestur, beberapa penelitian telah menunjukkan bahwa penerapan CNN sangat efektif, yang mendukung pendapat bahwa metode ini sangat cocok untuk klasifikasi citra secara *real-time* (Amartama dkk, 2024). Sebagai contoh, penelitian menunjukkan bahwa analisis citra yang dilakukan menggunakan CNN menghasilkan akurasi yang tinggi dalam pengenalan pola (Amartama dkk, 2024). Mengingat tingginya kebutuhan pada sistem yang cepat dan akurat, CNN dapat menjadi pilihan utama aplikasi pengenalan gestur tangan.



Gambar 2. 3 Arsitektur CNN (Narayana & Praveena, 2023)

Berdasarkan Gambar 2.3 CNN (Narayana & Praveena, 2023) memiliki arsitektur yang terdiri dari beberapa komponen utama yang masing-masing memiliki tugas spesifik dalam proses klasifikasi gambar.

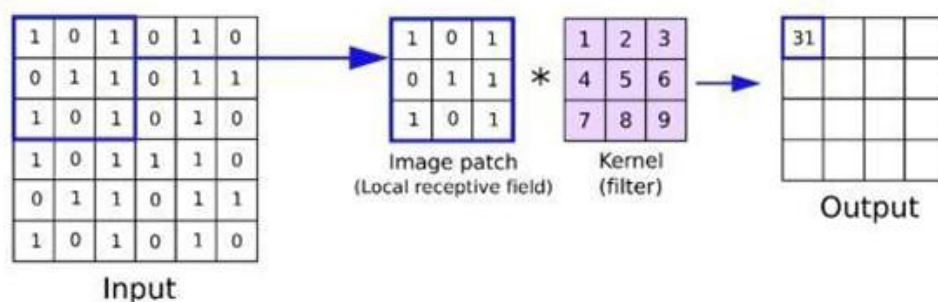
1. *Input Layer*

Input pada umumnya berupa gambar RGB dengan dimensi tertentu, contohnya $64 \times 64 \times 3$, di mana 3 mewakili tiga saluran warna merah, hijau, dan biru. Untuk memastikan bahwa data berada dalam rentang yang tepat, dilakukan proses normalisasi piksel (Oruh dkk., 2022).

2. *Convolution Layer*

Proses konvolusi dilakukan dengan menggunakan kernel atau filter yang bertujuan untuk menghasilkan peta fitur (*feature map*). Dalam proses ini terdapat beberapa parameter penting, seperti *stride* dan *padding* yang mempengaruhi interpretasi tepi gambar. Lapisan ini sangat penting dalam mengekstraksi fitur lokal, seperti tepi dan pola yang terdapat pada gambar (Dhawan & Agrawal, 2023); (Guo, 2022). *Feature map* dapat ditentukan dengan menggunakan persamaan berikut (Neethu dkk., 2020).

$$Y_i = f(W_i * I) \quad (2.1)$$



Gambar 2. 4 Ilustrasi *Feature Map* (Fawwaz dkk., 2020)

2. Fungsi Aktivasi : ReLU (*Rectified Linear Units*)

Secara umum fungsi aktivasi yang diterapkan dalam CNN adalah ReLU. Penggunaan ReLU memiliki beberapa keuntungan, antara lain dapat mempercepat proses pelatihan model dan mengatasi masalah *vanishing gradient* yang sering muncul pada fungsi aktivasi lain, seperti sigmoid dan tanh (Mukhtar dkk, 2022). Persamaan ReLU (Verma dkk., 2024) sebagai berikut :

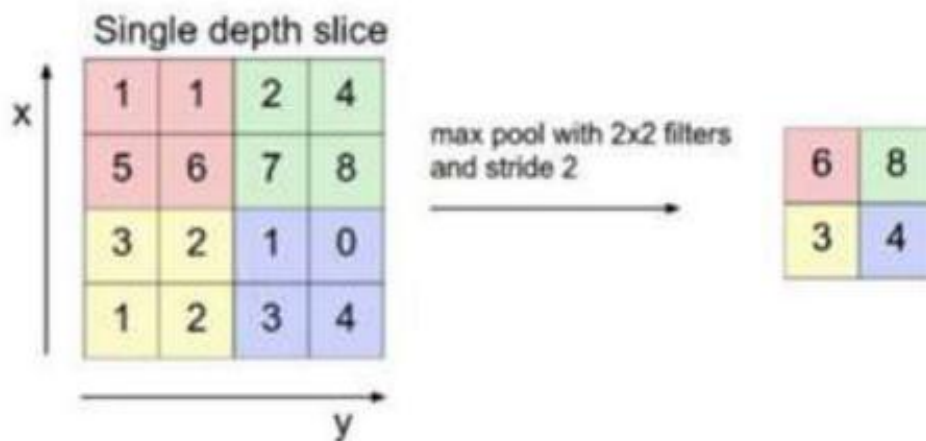
$$ReLU(x) = \max(0, x) \quad (2. 2)$$

3. *Pooling*

Pooling adalah teknik yang sering menggunakan metode *Max Pooling* dan *Average Pooling*. Teknik ini berfungsi untuk mengurangi dimensi data (*downsampling*) namun tetap mempertahankan fitur-fitur penting. Penggunaan *pooling* sangat berguna untuk mengurangi kompleksitas komputasi yang diperlukan selama proses pelatihan (Hernandez dkk, 2022).

Pooling layer berfungsi pada setiap tumpukan *feature map* dan secara efektif mengurangi dimensinya. Umumnya, proses *pooling* dilakukan menggunakan filter berukuran 2 x 2 dengan langkah (*stride*) sebesar dua yang diterapkan disetiap bagian dari input. Hal ini dapat diilustrasikan dengan persamaan (Neethu dkk., 2020) dan pada Gambar 2.10 (Fawwaz dkk., 2020) berikut :

$$P_i = \max(Y_i) \quad (2. 3)$$

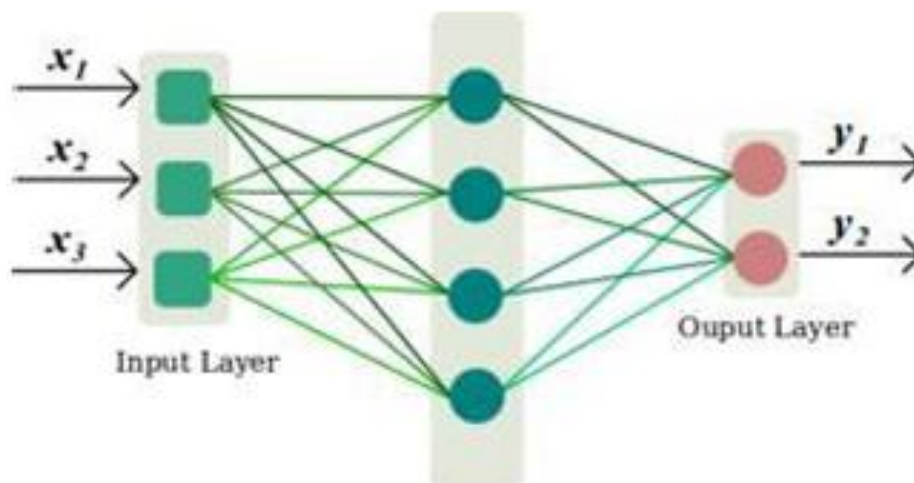


Gambar 2. 5 Ilustrasi Operasi *Max Pooling* (Fawwaz dkk., 2020)

4. *Fully Connected Layer*

Lapisan *fully connected* adalah lapisan di mana setiap neuron dari lapisan sebelumnya terhubung sepenuhnya dengan semua neuron di lapisan berikutnya, mirip dengan struktur *neural network* pada umumnya. Pada lapisan ini, fitur-fitur yang telah diekstraksi akan dihubungkan untuk menghasilkan *output* klasifikasi yang diinginkan. Dalam arsitektur CNN umumnya lapisan ini terletak di akhir, setelah beberapa lapisan konvolusi dan pooling (Chen dkk, 2022; Yan dkk, 2020).

Fully dapat diilustrasikan pada Gambar 2.11 (Bakti dkk., 2023) sebagai berikut :



Gambar 2. 6 Ilustrasi *Fully Connected* (Bakti dkk., 2023)

5. *Output Layer*

Dalam klasifikasi multi-kelas, fungsi *softmax* sering digunakan. Fungsi ini bertugas untuk menghasilkan prediksi akhir berdasarkan fitur yang diambil dari lapisan sebelumnya. Selain itu, fungsi ini juga memberikan probabilitas untuk setiap kelas yang terdapat dalam dataset, sehingga dapat membantu dalam mengoptimalkan keputusan akhir dari model (Abboud dkk, 2022). Fungsi *softmax* dinyatakan dalam persamaan berikut:

$$\text{Softmax}(z_i) = \frac{e^{z_i}}{\sum_{j=1}^K e^{z_j}} \quad (2.4)$$

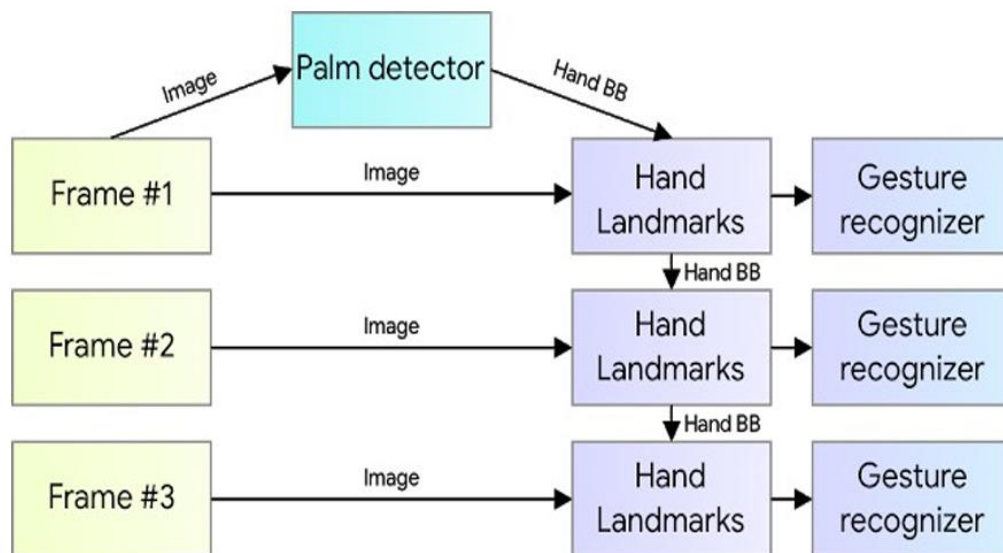
Pada sistem deteksi gestur tangan jarimatika, *output layer* digunakan untuk mengklasifikasikan berbagai gestur yang dapat dikenali oleh model. Sebagai contoh, jika ingin mendeteksi enam gestur dasar dalam jarimatika (seperti “1”, “2”, “3”, “4”, “5”, dan “6”), maka *output layer* akan memiliki enam neuron, di mana masing-masing neuron mewakili satu gestur.

2.1.3 *MediaPipe Hands*

MediaPipe *Hands* merupakan salah satu modul khusus dalam *framework* MediaPipe yang dirancang untuk mendeteksi dan melacak gestur tangan. MediaPipe *Hands* memungkinkan pengguna untuk membangun *pipeline* yang dapat melakukan *inferensi* pada dataset. Selain itu, MediaPipe *Hands* juga telah menyediakan kode sumber untuk keperluan penelitian dan pengembangan *prototype* teknologi. (Suryaperdana Agoes, 2021). *Pipeline* yang diimplementasikan oleh MediaPipe dapat dilihat pada Gambar 2.12 (Suryaperdana Agoes, 2021).

2.1.4 Justifikasi Pemilihan Metode

Penelitian ini memilih menggunakan kombinasi CNN dan MediaPipe *Hands* dibanding dengan metode lain. CNN digunakan pada metode klasifikasi konvensional (SVM dan KNN) karena mampu mengekstraksi fitur citra secara otomatis tanpa rekayasa fitur manual, dan dibandingkan dengan model *pretrained* (ResNet, MobileNet) karena tugas klasifikasi 18 kelas gestur pada penelitian ini relatif sederhana. Sehingga model CNN yang digunakan sudah cukup efisien untuk *real-time* pada CPU tanpa kompleksitas tambahan. MediaPipe *Hands* yang digunakan dibandingkan dengan metode deteksi tangan lain (deteksi warna kulit, *Haar Cascade*) karena sudah menyediakan model terlatih (*BlazePalm*) yang ringan dan langsung menghasilkan 21 titik *landmark* beserta informasi *handedness* tanpa perlu pelatihan ulang model deteksi. Integrasi kedua metode ini juga didukung oleh penelitian (Nautica, 2022) yang menunjukkan MediaPipe lebih unggul untuk deteksi tangan (90,99%) dibandingkan CNN saja (38,67%), sehingga kombinasi keduanya sesuai fungsinya masing-masing menghasilkan klasifikasi gestur Jarimatika yang optimal.

Gambar 2. 7 Ilustrasi Alur *Pipeline*

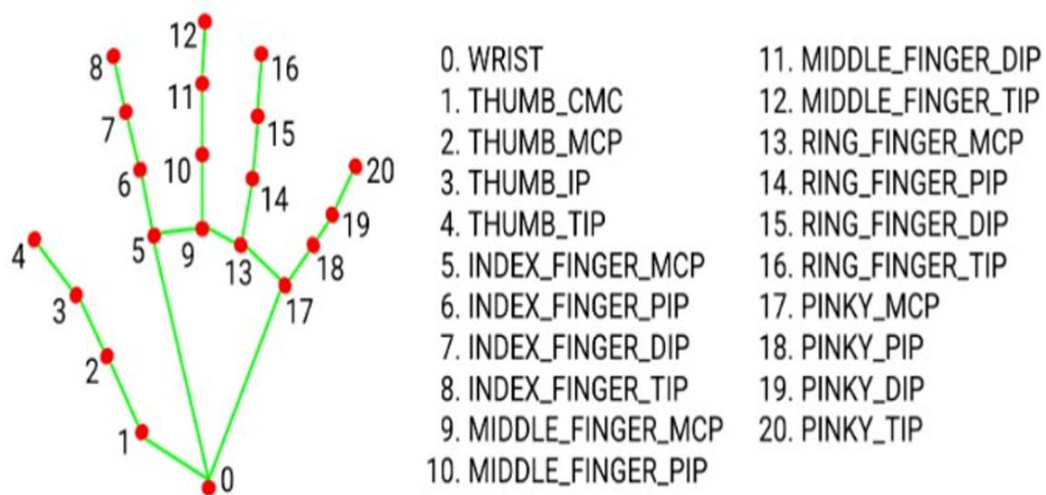
1. Model Deteksi Telapak Tangan (*Palm Detector*)

Framework MediaPipe telah mengembangkan detektor awal telapak tangan yang disebut *BlazePalm*. Langkah pertama adalah melatih telapak tangan sebagai pengganti detektor tangan. Kemudian algoritma *non-maximum suppression* diterapkan pada telapak tangan, di mana telapak tangan dimodelkan menggunakan *bounding box* persegi untuk menghindari rasio aspek lainnya dan mengurangi jumlah *anchors* dengan faktor 3-5 (Suryaperdana Agoes, 2021). Selanjutnya, *encoder-decoder* digunakan untuk fitur yang memungkinkan kesadaran konteks adegan lebih besar bahkan objek kecil. Terakhir, meminimalkan *focal loss* selama pelatihan yang mendukung sejumlah besar *anchors* yang dihasilkan dari varian skala yang tinggi (Suryaperdana Agoes, 2021).

2. *Hand Landmark*

Model ini mencapai lokalisasi titik kunci yang tepat dari 21 titik kunci dengan menggunakan koordinat buku jari tangan 3D yang dilakukan di dalam area tangan

yang terdeteksi melalui regresi, yang akan menghasilkan prediksi koordinat secara langsung. Maka dari itu, model ini berfungsi sebagai representasi *landmark* tangan dalam MediaPipe (Suryaperdana Agoes, 2021). *Landmark* tangan pada MediaPipe dapat diilustrasikan pada gambar 2.13 (Suryaperdana Agoes, 2021) berikut :



Gambar 2. 8 *Landmark* Tangan pada MediaPipe (Verma dkk., 2024)

Setiap ruas jari pada *landmark* memiliki koordinat yang terdiri dari x,y, dan z. Di mana nilai x dan y dinormalisasi dalam rentang [0.0, 1.0] berdasarkan lebar dan tinggi gambar, sedangkan z menunjukkan kedalaman *landmark*. Kedalaman *landmark* ini diukur dari pergelangan tangan yang dijadikan acuan. Semakin dekat *landmark* dengan kamera, semakin kecil pula nilai kedalamannya.

2.2 *Gestur Recognizer*

Dalam pengenalan gestur untuk mencapai tingkat akurasi yang diinginkan sistem harus dapat menangkap gerakan melalui kamera. Kualitas gambar memiliki dampak yang sangat signifikan terhadap hasil pengenalan gestur. Setiap *frame* vidio yang diambil harus cukup jelas agar dapat memastikan bahwa semua detail penting

dari gerakan tangan dapat terdeteksi dengan baik (Prof. C.D. Sawarkar dkk, 2024). Gambar 2.13 menunjukkan 21 titik 3D yang terdapat di setiap bagian telapak tangan, nilai koordinat dari titik-titik (4, 8, 12, 16, 20) yang mewakili ujung jari. Sementara itu, nilai koordinat dari titik-titik (3, 5, 9, 13, 17) mewakili titik tengah relatif terhadap sumbu x dan y (Wameed dkk, 2023).

Jempol dikatakan terbuka jika nilai koordinat dari titik (4), yang mewakili ujung jari lebih besar dibandingkan dengan nilai koordinat dari titik (3) yang mewakili titik tengah, keduanya relatif terhadap sumbu X. Sebaliknya, jika nilai koordinat dari titik (4) lebih kecil, maka jempol dianggap tertutup (Wameed dkk, 2023).

Keadaan semua jari, kecuali jempol di mana jari-jari (telunjuk, jari tengah, jari manis, dan kelingking) dianggap terbuka jika nilai koordinat dari titik-titik (8, 12, 16, 20) yang mewakili ujung jari lebih besar dibandingkan dengan nilai koordinat dari titik-titik (5, 9, 13, 17) yang mewakili titik tengah, pada sumbu Y. Sebaliknya, jika nilai koordinat tersebut lebih kecil, maka titik-titik tersebut dianggap tertutup (Wameed dkk, 2023).

2.1.5 Open Source Computer Vision Library (OpenCV)

OpenCV merupakan *library* perangkat lunak yang khusus dirancang untuk aplikasi visi komputer dan pengolahan citra. *OpenCV* memfasilitasi pengguna untuk mengambil gambar secara *real-time* dan melakukan pemrosesan citra kompleks yang sangat penting dalam berbagai aplikasi, mulai dari deteksi objek hingga pengenalan wajah (Supiyandi dkk., 2024). Pada penelitian ini, *OpenCV* digunakan secara spesifik sebagai pembuka akses ke kamera dan menangkap video

yang kemudian akan diproses lebih lanjut menggunakan MediaPipe *Hands*. Penggunaan fungsi seperti *cv2.VideoCapture()* untuk membuka kamera dan *cv2.imshow()* untuk menampilkan video secara langsung (Divanny dkk., 2024). Pada *OpenCV* merupakan *open source computer vision library* untuk bahasa pemrograman C/C++, dan telah dikembangkan ke Python, Java, dan Matlab (Susim dkk., 2021).

2.1.6 State Of The Art

State Of The Art yang disajikan pada tabel 2.1 berisi mengenai hasil penelitian yang sudah dilakukan sebelumnya yang berkaitan dengan penerapan judul penelitian yang dilakukan yaitu “Pengenal Gestur Tangan untuk Jarimatika Menggunakan CNN dan Mediapipe *Hands* secara *Real-time*”.

Tabel 2. 1 *State-of-the-Art* Penelitian

No	Nama Peneliti/Journal		Hasil Penelitian
1	Peneliti (tahun)	(Huse dkk., 2025)	Penelitian ini mengusulkan sistem pengenalan Bahasa Isyarat India (ISL) secara <i>real-time</i> dengan memanfaatkan <i>Convolutional Neural Network</i> (CNN) dan MediaPipe. Sistem ini mampu mengenali 36 gestur tangan berupa huruf dan angka dengan akurasi tinggi, serta dapat dijalankan hanya dengan menggunakan webcam. Hasil uji coba menunjukkan sistem bekerja efisien dan stabil untuk berbagai aplikasi, seperti komunikasi daring, ruang kelas virtual, dan interaksi manusia–komputer. Kelebihan sistem ini adalah dapat beroperasi secara <i>real-time</i>, mudah diimplementasikan pada perangkat desktop maupun aplikasi web serta fleksibel untuk dikembangkan lebih lanjut misalnya menambah variasi gestur atau bahasa isyarat lain. Namun, sistem ini masih memiliki keterbatasan karena kinerjanya sangat dipengaruhi kondisi pencahayaan, latar belakang, dan posisi tangan di depan kamera. Selain itu, performa juga bergantung pada spesifikasi perangkat keras yang digunakan.
	Judul	Real-Time ISL Recognition Using CNN and MediaPipe	
	Algoritma/ Metode	MediaPipe, CNN, Akselerasi GPU	
2	Peneliti	(Kumar & Bajpai, 2024)	

No	Nama Peneliti/Journal		Hasil Penelitian
	Judul	MediaPipe and CNNs for Real-Time ASL Gestur Recognition	Penelitian ini membuktikan bahwa kombinasi MediaPipe dan <i>Convolutional Neural Network</i> (CNN) mampu mengenali huruf alfabet Bahasa Isyarat Amerika (ASL) secara <i>real-time</i> dengan akurasi sangat tinggi yaitu 99,95% serta nilai presisi, <i>recall</i>, dan F1-Score hingga 100%. Kelebihan sistem ini adalah modelnya ringan, tidak memerlukan perangkat keras tambahan dan mudah diimplementasikan pada berbagai perangkat, serta fleksibel untuk diperluas ke bahasa isyarat lain. Namun, sistem masih terbatas pada pengenalan huruf tunggal dan belum dapat mengenali rangkaian gestur menjadi kata atau kalimat, sehingga belum sepenuhnya mendukung percakapan nyata. Selain itu, penelitian ini ini masih memiliki kinerja yang dipengaruhi oleh variasi pencahayaan dan latar belakang, sehingga perlu pengembangan lebih lanjut agar lebih adaptif pada kondisi dunia nyata.
	Algoritma/ Metode	MediaPipe, CNN, Augmentasi Data	

No	Nama Peneliti/Journal		Hasil Penelitian
3	Peneliti	(Gopal dkk., 2025)	Penelitian ini mengembangkan sistem pengenalan gestur tangan bahasa isyarat dengan menggunakan <i>Convolutional Neural Network</i> (CNN) dan MediaPipe. Sistem ini mencapai akurasi lebih dari 90% pada dataset khusus, sehingga efektif untuk deteksi bahasa isyarat secara <i>real-time</i>. Kelebihannya terletak pada efisiensi, karena MediaPipe memungkinkan deteksi tangan berlangsung cepat tanpa pra-pemrosesan rumit. Sehingga sistem ringan dan dapat dijalankan di perangkat dengan daya terbatas seperti <i>smartphone</i>. CNN juga membuat sistem mampu bekerja lintas pengguna dengan hasil yang cukup konsisten. Namun, sistem masih sensitif terhadap kondisi lingkungan seperti pencahayaan, latar belakang yang kompleks, serta perbedaan ukuran tangan. Selain itu, kinerjanya lebih baik untuk gestur statis, sementara pengenalan gerakan dinamis masih terbatas.
	Judul	Hand Gestur Detection For Sign Language Using CNN and MediaPipe	
	Algoritma/ Metode	CNN, MediaPipe, Data Augmentation	
4	Peneliti (tahun)	(Ahwani dkk., 2024)	Penelitian ini, mengembangkan sistem yang dapat mendeteksi huruf BISINDO dengan menggunakan MediaPipe dan CNN. Sistem ini mencapai akurasi 94% ketika pelatihan dan 78,46% pada pengujian secara <i>real-time</i>. Sistem ini bekerja secara langsung menggunakan kamera biasa serta
	Judul	Penerapan MediaPipe dalam Pengenalan Bisindo Berbasis Deep Learning dan Computer	

No	Nama Peneliti/Journal		Hasil Penelitian
	Algoritma/ Metode	CNN, MediaPipe	menunjukkan efisiensi dalam klasifikasi berkat penggunaan titik-titik <i>landmark</i> . Meskipun demikian, akurasi sistem masih tergolong rendah untuk beberapa huruf tertentu seperti B, H, K, S, dan X. Hal ini disebabkan oleh faktor pencahayaan, latar belakang, dan kualitas kamera, serta penurunan akurasi saat dilakukan pengujian langsung yang menunjukkan perlunya perbaikan dalam aspek generalisasi.
5	Peneliti (tahun)	(Nautica, 2022)	Penelitian ini menunjukkan bahwa MediaPipe mampu mendeteksi gestur tangan dengan akurasi 90,99% dan kecepatan stabil 20–25 FPS, sehingga efisien untuk aplikasi <i>real-time</i> . Sebaliknya, CNN hanya mencapai akurasi 38,67% dengan kecepatan lebih rendah (10–12 FPS). Keunggulan MediaPipe yaitu pada efisiensi komputasi dan akurasi tinggi, sehingga bisa dijalankan di perangkat dengan spesifikasi terbatas tanpa proses tambahan yang rumit. Metode ini juga fleksibel, misalnya digunakan dalam pendidikan sebagai alat bantu interaktif bagi anak usia dini. Namun, CNN masih bermasalah dengan overfitting karena meski akurasinya tinggi saat pelatihan, performanya buruk di kondisi nyata. MediaPipe memiliki keterbatasan, terutama pada variasi pencahayaan, latar belakang yang
	Judul	<i>Hand Gestur Detection Sebagai Alat Bantu Ajar Berhitung Menggunakan MediaPipe dan Convolutional Neural Network Secara Real-time</i>	
	Algoritma/ Metode	MediaPipe, CNN	

No	Nama Peneliti/Journal		Hasil Penelitian
			kompleks, serta perbedaan bentuk atau ukuran tangan yang dapat memengaruhi akurasi sistem.
6	Peneliti (tahun)	(Narayana & Praveena, 2023)	Penelitian ini mengusulkan sistem pengenalan gestur tangan <i>real-time</i> berbasis <i>Convolutional Neural Network</i> (CNN). Model terdiri dari dua bagian utama, yaitu ekstraksi fitur dan klasifikasi, sehingga mampu mengenali pola gestur tanpa rekayasa fitur manual. Hasil pengujian menunjukkan sistem tetap bekerja baik pada kondisi pencahayaan dan sudut kamera yang berbeda. Kelebihannya adalah mampu mengenali gestur dengan akurasi tinggi dan respons cepat karena CNN dapat langsung mengekstraksi bentuk, orientasi, dan tekstur tangan dari citra sehingga lebih efisien dibanding metode tradisional. Sistem juga dilengkapi keluaran berupa teks dan suara untuk memudahkan komunikasi bagi penyandang tunarungu dan tunawicara. Namun, keterbatasannya masih terletak pada pengenalan gestur statis saja, sehingga belum mendukung rangkaian gestur dinamis atau kalimat, serta menghadapi tantangan penerapan di kondisi nyata dengan latar belakang kompleks dan variasi ukuran tangan.
	Judul	Real-Time Hand Gestur Recognition System Using CNN	
	Algoritma/ Metode	CNN, SqueezeNet	

No	Nama Peneliti/Journal		Hasil Penelitian
7	Peneliti (tahun)	(Ageishi dkk., 2021)	Penelitian ini mengembangkan sistem deteksi gerakan tangan <i>secara real-time</i> dengan <i>Deep Neural Network</i> (DNN) berbasis kamera, tanpa memerlukan sensor tambahan seperti EEG. Sistem mampu mengenali 10 angka Bahasa Isyarat Amerika (ASL) dengan akurasi validasi 97,31% dan kecepatan rata-rata 18 FPS, sehingga responsif dan dapat menyesuaikan dengan variasi ukuran tangan pengguna. Kelebihannya adalah sistem efisien dan mudah digunakan karena hanya membutuhkan kamera standar, serta mampu mengekstraksi fitur visual otomatis tanpa proses kalibrasi yang rumit. Kinerja <i>real-time</i> ini membuatnya relevan untuk aplikasi interaktif, terutama komunikasi berbasis bahasa isyarat. Namun, akurasi sistem menurun pada pencahayaan buruk atau latar belakang yang kompleks, dan saat ini masih terbatas pada 10 angka statis tanpa mendukung gestur dinamis. Oleh karena itu, pengembangan lebih lanjut diperlukan agar sistem lebih tangguh dan siap digunakan dalam komunikasi nyata.
	Judul	Real-time Hand-Gestur Recognition based on Deep Neural Network	
	Algoritma/ Metode	Deep Neural Network (DNN), CNN, Image Preprocessing, Augmentation	

No	Nama Peneliti/Journal		Hasil Penelitian
8	Peneliti (tahun)	(Raju dkk., 2025)	Penelitian ini mengembangkan sistem pengenalan gestur tangan statis secara <i>real-time</i> menggunakan <i>Convolutional Neural Network</i> (CNN). Sistem ini mencapai akurasi lebih dari 96% dengan kecepatan hingga 30 fps, sehingga layak digunakan untuk aplikasi nyata seperti penerjemahan bahasa isyarat. Keunggulan CNN adalah mampu mengekstraksi fitur otomatis tanpa rekayasa manual, lebih efisien dibandingkan metode tradisional. Serta tetap tangguh terhadap variasi posisi tangan dan pencahayaan. Sistem juga responsif dengan latensi yang sangat rendah, sehingga berpotensi besar diterapkan pada teknologi bantu komunikasi maupun interaksi manusia–komputer. Namun, sistem ini masih terbatas pada gestur statis, rentan mengalami kesalahan klasifikasi pada gestur dengan bentuk yang serupa, dan belum dapat mengenali rangkaian gerakan dinamis. Karena itu, penelitian selanjutnya menyarankan penggunaan model temporal seperti 3D CNN atau LSTM, serta integrasi data tambahan seperti sensor kedalaman agar sistem lebih kuat di kondisi nyata.
	Judul	Real Time Hand Gestur Recognition Using CNN	
	Algoritma/ Metode	CNN, Preprocessing gambar	

No	Nama Peneliti/Journal		Hasil Penelitian
9	Peneliti (tahun)	(Reddy dkk., 2024)	Penelitian ini mengembangkan sistem pengenalan gestur tangan berbasis <i>Convolutional Neural Network</i> (CNN) yang dapat bekerja tanpa segmentasi manual maupun ekstraksi fitur rumit. Sistem ini efisien dengan akurasi sangat tinggi yaitu 99,96% pada dataset HaGrid, melampaui hasil terbaik sebelumnya pada dataset ASL sebesar 99,92%. Pada penelitian ini terdapat keunggulan pada arsitektur yang sederhana namun efektif, serta tetap stabil meski latar belakang kompleks atau pencahayaan kurang ideal, sehingga berpotensi digunakan pada aplikasi nyata seperti penerjemahan bahasa isyarat maupun manajemen lalu lintas. Namun, keterbatasannya adalah hanya mendukung gestur statis dan belum mampu mengenali gerakan dinamis atau multi-pengguna, sehingga masih perlu pengembangan lanjutan dengan integrasi model temporal agar lebih adaptif di dunia nyata.
	Judul	Hand Gestur Recognition Using Convolutional Neural Networks	
	Algoritma/ Metode	CNN, Preprocessing Gambar	
10	Peneliti (tahun)	(Singh dkk., 2022)	Penelitian ini mengembangkan sistem pengenalan gestur tangan statis berbasis <i>Convolutional Neural Network</i> (CNN) menggunakan dataset American Sign Language (ASL) dari Kaggle. Sistem ini mencapai akurasi
	Judul	Hand Gestur Identification for Improving Accuracy Using	

No	Nama Peneliti/Journal		Hasil Penelitian
		convolution Neural Network (CNN)	93,61% pada pengujian dan validasi hingga 100%, sehingga lebih baik dibandingkan beberapa metode sebelumnya. Keunggulannya terletak pada akurasi tinggi, pemrosesan yang efisien dengan pra-pemrosesan sederhana, serta kemudahan penggunaan kamera biasa tanpa perangkat tambahan khusus. CNN juga mampu mengekstraksi fitur otomatis sehingga mempercepat pelatihan. Namun, sistem ini masih terbatas pada gestur statis akurasinya bisa menurun pada kondisi pencahayaan atau latar belakang yang bervariasi dan belum diuji untuk banyak pengguna. Karena itu, penelitian selanjutnya disarankan untuk mengembangkan pengenalan gestur dinamis dan mengoptimalkan kinerja pada kondisi nyata.
	Algoritma/ Metode	CNN, Image Preprocessing	
11	Peneliti (tahun)	(Stella dkk., 2024)	Penelitian ini mengembangkan sistem pengenalan gerakan <i>tangan real-time</i> dengan menggabungkan MediaPipe untuk deteksi tangan dan OpenCV untuk pemrosesan citra. Sistem ini mampu mengenali berbagai gestur dan menerjemahkannya menjadi perintah, sehingga bisa digunakan untuk mengendalikan perangkat elektronik, aplikasi VR/AR, maupun penerjemahan bahasa isyarat. Hasil penelitian menunjukkan bahwa kombinasi kedua teknologi ini mampu memproses citra dengan cepat dan
	Judul	Hand Gestur Recognition Using MediaPipe and OpenCV	
	Algoritma/ Metode	MediaPipe, OpenCV	

No	Nama Peneliti/Journal		Hasil Penelitian
			akurat tanpa membutuhkan perangkat keras khusus. Keunggulannya terletak pada penggunaan model pra-latih yang efisien, kompatibilitas lintas platform, serta fleksibilitas pada perangkat dengan sumber daya terbatas. Namun, sistem ini masih memiliki kelemahan, seperti akurasi yang menurun pada pencahayaan buruk, kesulitan mendeteksi gestur kompleks atau interaksi banyak pengguna, serta terbatas pada gestur statis. Oleh karena itu, diperlukan pengembangan lebih lanjut dengan integrasi model <i>deep learning</i> dan optimasi agar dapat digunakan pada kondisi nyata yang lebih beragam.
12	Peneliti (tahun)	(Rheiner dkk., 2025)	Penelitian ini mengembangkan sistem pengenalan Bahasa Isyarat Amerika (ASL) menggunakan arsitektur MobileNetV3 yang dirancang agar dapat berjalan <i>real-time</i> di CPU tanpa membutuhkan GPU. Prosesnya meliputi prapemrosesan data, deteksi tangan dengan MediaPipe, dan klasifikasi menggunakan <i>transfer learning</i> dua tahap. Hasil uji menunjukkan akurasi sangat tinggi yaitu 99,98% pada dataset Kaggle dan 99,93% pada dataset Londhe dengan waktu inferensi kurang dari 500 <i>ms</i> per <i>frame</i> . Keunggulan sistem ini adalah efisiensi komputasi, kemampuan generalisasi yang baik,
	Judul	From Pixels to Letters: A High-Accuracy CPU-Real-Time American Sign Language Detection Pipeline	
	Algoritma/ Metode	CNN, MobileNetV3, Transfer Learning, Grad-CAM	

No	Nama Peneliti/Journal		Hasil Penelitian
			serta transparansi prediksi melalui integrasi Grad-CAM. Selain itu, sistem tetap dapat dijalankan pada perangkat dengan spesifikasi terbatas. Namun, kelemahannya adalah pengujian hanya dilakukan pada gestur statis di lingkungan terkontrol, sehingga belum teruji pada kondisi nyata dengan pencahayaan, latar belakang, dan interaksi pengguna yang beragam. Oleh karena itu, penelitian selanjutnya perlu diarahkan pada pengenalan gestur dinamis dan pengujian di dunia nyata agar sistem lebih adaptif.

2.1.7 Matriks Penelitian

Tabel 2.2 merupakan matriks penelitian yang berisi perbandingan berbagai penelitian terkait deteksi gestur. Kolom-kolom yang disertakan memberikan gambaran tentang metode yang diterapkan.

Tabel 2. 2 Matriks penelitian

No	Peneliti/ Tahun	Ruang Lingkup														
		Judul	Alat		Metode Ekstraksi Fitur				Metode Klasifikasi			Dataset				
			WebCam	Kamera	MediaPipe	CNN	DNN	OpenCv	CNN	MobileNetV3	DNN	ISL	ASL	BISINDO	Angka	Lainnya
1	(Huse dkk, 2025)	Real-Time ISL Recognition Using CNN and MediaPipe	✓		✓				✓			✓				

2	(Kumar dkk., 2024)	MediaPipe and CNNs for Real-Time ASL Gestur recognition	✓		✓				✓				✓			
3	(Gopal dkk., 2025)	Hand Gestur Detection For Sign Language Using CNN and MediaPipe	✓		✓				✓				✓			✓
4	(Ahwani dkk., 2024)	Penerapan MediaPipe dalam Pengenalan Bisindo Berbasis Deep Learning dan Computer Vision (Studi Kasus: SLB-C B Yayasan Pendidikan Luar Biasa (YPLB) Blitar)	✓		✓				✓					✓		
5	(Nautica, 2022)	Hand Gestur Detection Sebagai Alat Bantu Ajar Berhitung Menggunakan MediaPipe dan Convolutional Neural Network Secara Real-time	✓		✓				✓						✓	
6	(Narayana & Praveena, 2023)	Real-Time Hand Gestur Recognition System Using CNN	✓			✓			✓							✓
7	(Naoto Ageishi., 2021)	Real-time Hand-Gestur Recognition based on Deep Neural Network	✓			✓					✓		✓			

8	(Raju dkk., 2025)	Real Time Hand Gestur Recognition Using CNN	✓			✓			✓							✓
9	(Reddy dkk., 2024)	Hand Gestur Recognition Using Convolutional Neural Networks	✓			✓			✓							✓
10	(Singh dkk., 2022a)	Hand Gestur Identification for Improving Accuracy Using convolution Neural Network (CNN)	✓			✓			✓				✓			
11	(Stella dkk., 2024)	Hand Gestur Recognition Using MediaPipe and OpenCV	✓		✓				✓							✓
12	(Rheiner dkk., 2025)	From Pixels to Letters: A High-Accuracy CPU-Real-Time American Sign Language Detection Pipeline	✓			✓		✓	✓	✓			✓			✓
13	(Ghaitsa Zahira Shofa, 2026)	Pengenalan Gestur Tangan untuk Jarimatika Menggunakan CNN dan Mediapipe Hands secara Realtime	✓		✓				✓				✓			✓

Berdasarkan Tabel 2.2 terlihat bahwa dari seluruh penelitian yang dianalisis, sebagian besar menggunakan CNN sebagai metode klasifikasi dan MediaPipe sebagai metode ekstraksi fitur. Hampir seluruhnya berfokus pada pengenalan bahasa isyarat seperti ASL, ISL, dan BISINDO. Hanya penelitian (Nautica, 2022) yang memiliki kemiripan konteks dengan penelitian ini, yaitu menggunakan MediaPipe dan CNN untuk deteksi gestur angka dalam pembelajaran Jarimatika. Namun penelitian tersebut belum mengintegrasikan informasi *handedness* untuk membedakan tangan kanan dan tangan kiri secara otomatis, serta belum mengevaluasi kecepatan sistem menggunakan metrik FPS secara eksplisit.

Belum ditemukan penelitian yang secara khusus mengembangkan sistem deteksi gestur Jarimatika dengan dataset yang mencakup 18 kelas gestur yaitu sembilan gestur satuan pada tangan kanan dan sembilan gestur puluhan pada tangan kiri, dan menggabungkan hasil klasifikasi kedua tangan untuk membentuk bilangan dua digit secara otomatis. Keterbatasan dataset publik yang spesifik untuk gestur Jarimatika juga menjadi hambatan yang belum diatasi oleh penelitian-penelitian sebelumnya.

Berdasarkan kesenjangan tersebut, penelitian ini dirancang untuk mengisi kekosongan yang ada dengan mengembangkan sistem pengenalan gestur tangan Jarimatika secara *real-time* menggunakan CNN dan MediaPipe *Hands*. Dengan evaluasi ketepatan klasifikasi menggunakan *accuracy*, *precision*, *recall*, dan *F1-Score*, serta pengujian kecepatan sistem menggunakan metrik FPS pada perangkat berbasis CPU.