

CONVOLUTIONAL NEURAL NETWORK DENGAN MEDIAPIPE HAND UNTUK PENGENALAN HURUF BAHASA ISYARAT INDONESIA (SIBI)

Ilham Wahyu Saputro, Budi Hartono

Teknik Informatika, Universitas Stikubank Semarang
Jalan Kendeng V, Bendan Ngisor, Kota Semarang, Jawa Tengah., Indonesia
Ilhamwahyu0987@gmail.com

ABSTRAK

Sistem Isyarat Bahasa Indonesia (SIBI) merupakan bahasa isyarat baku yang dikembangkan pemerintah Indonesia untuk mendukung komunikasi penyandang tuli dan wicara. Permasalahannya, pemahaman masyarakat umum terhadap SIBI masih sangat terbatas sehingga menyulitkan interaksi sehari-hari. Tujuan penelitian ini adalah mengembangkan prototipe sistem pengenalan huruf SIBI (24 huruf alfabet A–Y tanpa J dan Z) secara *real-time* dari video *webcam* menggunakan *Convolutional Neural Network (CNN)*. Metode yang digunakan meliputi: pengumpulan dataset sebanyak 5.280 gambar; *preprocessing* berupa konversi *grayscale*, ekstraksi *Region of Interest (ROI)* dengan MediaPipe, penyeragaman ukuran 64×64 piksel, serta normalisasi nilai piksel; arsitektur CNN terdiri dari tiga blok konvolusi, dua lapisan *fully connected*, dan *dropout*, dilatih dengan optimizer Adam (learning rate 0,0001) serta *early stopping*. Hasil penelitian menunjukkan: akurasi model pada data uji mencapai 85,51%; kecepatan inferensi 14 frame per detik pada laptop tanpa GPU; akurasi uji coba langsung oleh tiga partisipan sebesar 90%; selain itu, ditemukan bahwa augmentasi data (rotasi, pergeseran, perubahan kecerahan) justru menyebabkan ketidakstabilan gradien sehingga tidak digunakan dalam model final. Sistem prototipe yang dihasilkan mampu mendeteksi dan mengklasifikasikan huruf SIBI secara *real-time* dilengkapi visualisasi landmark tangan dan teks prediksi.

Kata kunci : CNN, Computer Vision, Deep Learning , MediaPipe, SIBI

1. PENDAHULUAN

Bahasa isyarat merupakan salah satu bentuk komunikasi utama yang digunakan oleh penyandang tuli dan wicara untuk menyampaikan informasi dalam kehidupan sehari-hari. Di Indonesia, Sistem Isyarat Bahasa Indonesia (SIBI) digunakan sebagai bahasa isyarat formal yang telah distandarisasi oleh pemerintah. Perkembangan teknologi *computer vision* dan *deep learning* memungkinkan proses pengenalan gerakan tangan dilakukan secara otomatis menggunakan kamera digital. Salah satu metode *deep learning* yang banyak digunakan dalam pengolahan citra adalah *Convolutional Neural Network (CNN)* karena memiliki kemampuan yang baik dalam proses ekstraksi fitur citra dan klasifikasi objek. Selain CNN, teknologi deteksi landmark tangan seperti MediaPipe Hands juga berkembang pesat dan mampu melakukan deteksi posisi tangan secara *real-time*. MediaPipe dan CNN banyak digunakan dalam pengenalan bahasa isyarat karena mampu meningkatkan akurasi deteksi tangan secara otomatis dan *real-time* [1].

Akan tetapi, pemahaman masyarakat umum terhadap bahasa isyarat masih tergolong rendah sehingga komunikasi antara penyandang tuli dan masyarakat umum masih mengalami kesulitan. Penelitian mengenai pengenalan alfabet SIBI menggunakan metode CNN menunjukkan bahwa CNN mampu memberikan performa klasifikasi yang baik pada pengenalan huruf bahasa isyarat Indonesia [2]. Namun, masih diperlukan pengembangan sistem yang lebih responsif dan akurat dalam kondisi *real-time*. Penelitian lain menunjukkan bahwa arsitektur MobileNet dapat digunakan untuk pengenalan SIBI

dengan kompleksitas model yang lebih ringan dan cocok untuk implementasi *real-time* [3]. Penelitian lainnya mengembangkan sistem pembelajaran SIBI berbasis CNN yang mampu mengenali gestur tangan sebagai media pembelajaran interaktif [4]. Sementara itu, penerapan model *hybrid CNN* dan *Long Short-Term Memory (LSTM)* untuk pengenalan SIBI *real-time* memperoleh hasil yang baik pada pengolahan urutan gerakan tangan [5]. Meskipun demikian, belum banyak sistem yang menggabungkan MediaPipe Hands dan CNN secara terpadu untuk pengenalan huruf statis SIBI dengan mempertimbangkan aspek *preprocessing* yang optimal serta ketahanan terhadap variasi kondisi.

Berdasarkan permasalahan tersebut, penelitian ini bertujuan untuk mengembangkan sistem pengenalan huruf bahasa isyarat Indonesia (SIBI) menggunakan kombinasi MediaPipe Hands dan CNN berbasis *webcam* secara *real-time*. Sistem ini diharapkan mampu membantu proses pembelajaran bahasa isyarat serta meningkatkan komunikasi antara penyandang tuli dan masyarakat umum.

Rencana penyelesaian masalah dalam penelitian ini meliputi beberapa tahapan. Pertama, pengumpulan data dilakukan dengan merekam video partisipan yang menunjukkan 24 huruf SIBI (A–Y tanpa J dan Z) dengan latar belakang netral dan pencahayaan terkendali, kemudian mengekstrak *frame* menjadi 5.280 gambar. Kedua, tahap *preprocessing* meliputi konversi gambar ke *grayscale*, ekstraksi *Region of Interest (ROI)* menggunakan MediaPipe Hands, penyeragaman ukuran menjadi 64×64 piksel, serta normalisasi nilai piksel. Ketiga, dirancang arsitektur

CNN dengan tiga blok konvolusi, dua lapisan *fully connected*, dan *dropout* untuk mencegah *overfitting*. Keempat, model dilatih menggunakan *optimizer Adam* dengan *learning rate* 0,0001 serta *early stopping*, sementara itu *augmentasi data* (rotasi, pergeseran, perubahan kecerahan) dievaluasi dan ditemukan menyebabkan ketidakstabilan gradien sehingga tidak digunakan dalam model final. Kelima, evaluasi sistem dilakukan melalui pengujian akurasi pada data uji, pengukuran kecepatan inferensi dalam *frame per detik* (FPS), serta uji coba langsung oleh partisipan. Keenam, implementasi prototipe berupa sistem *real-time* dari *webcam* yang dilengkapi visualisasi *landmark* tangan dan teks prediksi.

2. TINJAUAN PUSTAKA

2.1. Sistem Isyarat Bahasa Indonesia (SIBI)

Sistem Isyarat Bahasa Indonesia (SIBI) merupakan bahasa isyarat resmi yang digunakan di Indonesia sebagai media komunikasi bagi penyandang tuli dan wicara. SIBI menggunakan representasi gerakan tangan untuk menyampaikan huruf, kata, dan kalimat tertentu. Dataset publik yang digunakan dalam penelitian ini diperoleh dari repositori Kaggle dengan nama dataset "Sistem Isyarat Bahasa Indonesia (SIBI)" [6]. Dataset tersebut berisi kumpulan citra alfabet bahasa isyarat yang digunakan sebagai data pelatihan dan pengujian model CNN.

2.2. Convolutional Neural Network (CNN)

Convolutional Neural Network (CNN) merupakan salah satu metode *deep learning* yang banyak digunakan dalam pengolahan citra digital. CNN memiliki kemampuan untuk melakukan ekstraksi fitur otomatis melalui *convolution layer* dan *pooling layer*. Metode CNN memiliki kemampuan yang baik dalam proses ekstraksi fitur citra dan klasifikasi objek secara otomatis sehingga banyak digunakan pada berbagai bidang pengolahan citra digital, termasuk klasifikasi citra medis [3]. Penelitian mengenai pengenalan alfabet SIBI menggunakan CNN memperoleh performa klasifikasi yang baik pada pengenalan huruf bahasa isyarat Indonesia [2].

Arsitektur CNN pada penelitian ini terdiri dari tiga *convolutional layer*, *pooling layer*, *dropout layer*, dan *fully connected layer*. Penggunaan *dropout* bertujuan untuk mengurangi *overfitting* selama proses pelatihan model.

2.3. MediaPipe Hands

MediaPipe Hands merupakan *framework* deteksi tangan berbasis *machine learning* yang dikembangkan oleh Google untuk mendeteksi *landmark* tangan secara *real-time*. *Framework* ini mampu mendeteksi posisi tangan dan titik *landmark* dengan cepat sehingga cocok digunakan pada aplikasi *real-time* berbasis *webcam*.

Penelitian lain menggunakan pendekatan LSTM-*Holistic* untuk mendeteksi gerakan tangan pada bahasa isyarat SIBI dan menunjukkan bahwa kombinasi *landmark* tangan dan *deep learning* dapat meningkatkan kemampuan pengenalan gestur [7]. Selain itu, penerapan EfficientNet V2 dan MediaPipe pada pengenalan bahasa isyarat berbasis citra menunjukkan bahwa proses segmentasi citra dan ekstraksi *landmark* berpengaruh terhadap performa klasifikasi [1].

2.4. Penelitian Terdahulu

Penelitian lain mengembangkan sistem pembelajaran SIBI menggunakan CNN untuk mengenali gestur tangan secara interaktif [4]. Penelitian lainnya menggabungkan metode CNN dan LSTM untuk pengenalan bahasa isyarat secara *real-time* [5]. Metode YOLO dan CNN digunakan untuk mendeteksi tangan otomatis pada video bahasa isyarat Indonesia dan menunjukkan hasil yang baik pada deteksi objek tangan secara *real-time* [8]. Kemampuan CNN dalam ekstraksi fitur visual telah dibuktikan pada berbagai domain, termasuk klasifikasi citra medis [9]. Penelitian lain menyebutkan bahwa *augmentasi data* seperti rotasi, pergeseran, dan perubahan kecerahan dapat meningkatkan akurasi model pada dataset terbatas, namun dalam penelitian ini *augmentasi* justru menyebabkan ketidakstabilan gradien [10]. Berdasarkan penelitian-penelitian tersebut, metode CNN dan MediaPipe memiliki potensi yang baik untuk diterapkan pada pengenalan huruf SIBI berbasis *webcam*.

3. METODE PENELITIAN

Penelitian ini menggunakan metode *experimental research* dengan pendekatan kuantitatif untuk membangun sistem pengenalan huruf Sistem Isyarat Bahasa Indonesia (SIBI) secara *real-time* menggunakan metode *Convolutional Neural Network (CNN)* dan MediaPipe Hands. Penelitian dilakukan dengan memanfaatkan teknologi *computer vision* untuk mendeteksi area tangan pengguna dan melakukan klasifikasi huruf berdasarkan citra tangan yang diperoleh melalui *webcam*.

Tujuan utama penelitian ini adalah menghasilkan sistem pengenalan huruf bahasa isyarat yang mampu bekerja secara *real-time* dengan tingkat akurasi yang baik dan dapat dijalankan pada perangkat dengan spesifikasi menengah tanpa menggunakan GPU tambahan. Penelitian dilakukan melalui beberapa tahapan utama mulai dari pengumpulan dataset, *preprocessing data*, pembangunan model CNN, pelatihan model, evaluasi performa model, hingga implementasi sistem secara *real-time* menggunakan *webcam*.

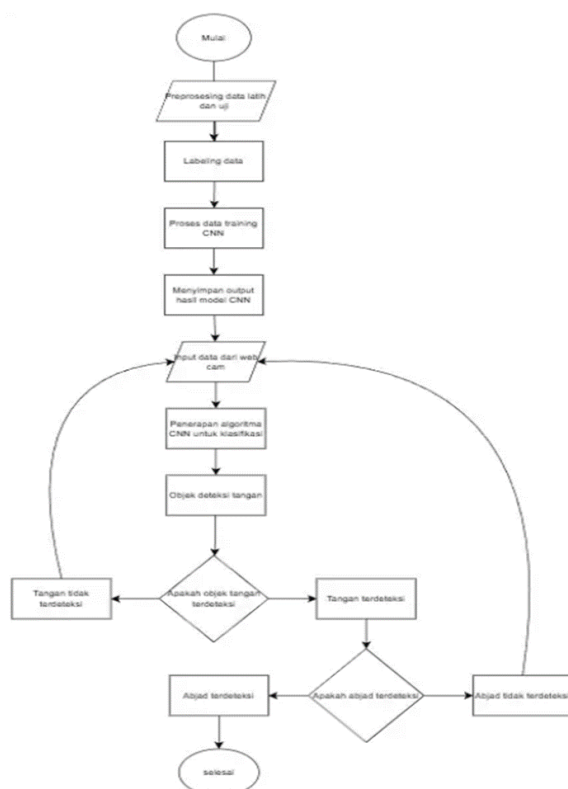
3.1. Tahapan Penelitian

Pada tahap awal dilakukan studi literatur terkait penelitian pengenalan bahasa isyarat menggunakan metode *deep learning* dan *computer vision*. Studi

literatur dilakukan untuk memahami metode yang paling sesuai digunakan pada penelitian serta mengetahui kelemahan dan kelebihan penelitian sebelumnya.

Tahap berikutnya adalah pengumpulan dataset huruf SIBI yang akan digunakan sebagai data pelatihan dan pengujian model CNN. Setelah dataset terkumpul, dilakukan preprocessing data untuk meningkatkan kualitas citra sebelum masuk ke proses pelatihan model.

Model CNN kemudian dibangun menggunakan beberapa *convolutional layer* dan *pooling layer* untuk melakukan ekstraksi fitur citra tangan. Setelah model selesai dilatih, dilakukan pengujian menggunakan data testing dan implementasi sistem secara *real-time* menggunakan *webcam*.



Gambar 1. Flowchart

Gambar 1. menampilkan cara kerja program dari awal hingga kita dapat mendeteksi bentuk tangan.

3.2. Pengumpulan Dataset

Dataset yang digunakan pada penelitian ini terdiri dari citra huruf alfabet Sistem Isyarat Bahasa Indonesia (SIBI) sebanyak 24 kelas huruf yaitu A sampai Y tanpa huruf J dan Z. Huruf J dan Z tidak digunakan karena kedua huruf tersebut menggunakan gerakan dinamis sehingga penelitian difokuskan pada huruf statis. Dataset diperoleh dari repositori Kaggle dengan nama dataset Sistem Isyarat Bahasa Indonesia (SIBI)[6]. Selain menggunakan dataset public.

Jumlah keseluruhan dataset yang digunakan sebanyak 5.280 citra dengan distribusi data yang relatif

seimbang pada setiap kelas huruf. Data kemudian dibagi menjadi data training dan data testing menggunakan rasio 80:20. Pembagian data dilakukan secara acak agar model dapat mempelajari variasi citra tangan dengan lebih baik

Tabel 1. Komposisi dataset penelitian

No	Huruf	Jumlah Gambar Latih	Jumlah Gambar Uji
1	A	176	44
2	B	176	44
3	C	176	44
4	D	176	44
5	E	176	44
6	F	176	44
7	G	176	44
8	H	176	44
9	I	176	44
10	K	176	44
11	L	176	44
12	M	176	44
13	N	176	44
14	O	176	44
15	P	176	44
16	Q	176	44
17	R	176	44
18	S	176	44
19	T	176	44
20	U	176	44
21	V	176	44
22	W	176	44
23	X	176	44
24	Y	176	44

Pada Tabel 1 menyajikan contoh komposisi data untuk beberapa huruf. Semua huruf memiliki distribusi yang sama: 176 gambar untuk latih dan 44 gambar untuk uji.

3.3. Preprocessing Data

Tahap preprocessing dilakukan untuk meningkatkan kualitas citra dan menyeragamkan bentuk input sebelum digunakan pada proses pelatihan model CNN. Tahap preprocessing merupakan salah satu tahapan penting karena kualitas data input sangat mempengaruhi performa model *deep learning*.

Proses preprocessing pada penelitian ini meliputi beberapa tahapan yaitu deteksi tangan menggunakan *MediaPipe Hands*, ekstraksi *Region of Interest (ROI)*, konversi citra menjadi *grayscale*, *resize* citra, dan normalisasi nilai piksel.

MediaPipe Hands digunakan untuk mendeteksi area tangan secara otomatis menggunakan *webcam*. *Framework* ini mampu mendeteksi *landmark* tangan secara *real-time* dengan tingkat akurasi yang baik. Setelah area tangan berhasil dideteksi, sistem melakukan cropping pada area tangan menggunakan *Region of Interest (ROI)* sehingga model hanya fokus pada area tangan dan mengurangi gangguan dari latar belakang gambar.

Tahap berikutnya adalah konversi citra RGB menjadi *grayscale*. Konversi *grayscale* dilakukan

untuk mengurangi kompleksitas data dan mempercepat proses komputasi selama pelatihan model. Setelah itu, citra diubah ukurannya menjadi 64×64 piksel agar seluruh data memiliki ukuran yang seragam.

Normalisasi nilai piksel dilakukan dengan membagi nilai piksel menjadi rentang 0 sampai 1. Proses ini bertujuan untuk mempercepat proses konvergensi model dan meningkatkan stabilitas pelatihan.

3.4. Perancangan Arsitektur CNN

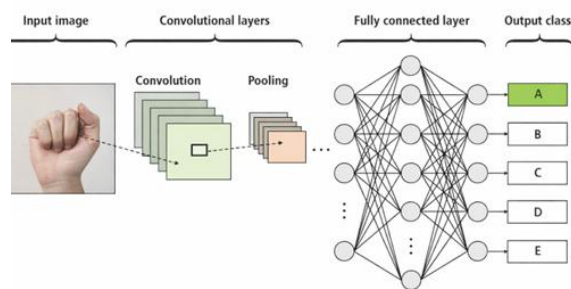
Model CNN yang digunakan pada penelitian ini terdiri dari tiga *convolutional layer*, *max pooling layer*, *dropout layer*, *flatten layer*, dan *fully connected layer*. Arsitektur model dirancang untuk mampu melakukan ekstraksi fitur citra tangan secara optimal namun tetap ringan sehingga dapat dijalankan pada perangkat tanpa GPU tambahan.

Convolutional layer berfungsi untuk mengekstraksi fitur penting dari citra input seperti bentuk tangan, pola jari, dan kontur tangan. Setiap *convolutional layer* menggunakan fungsi aktivasi ReLU untuk membantu model mempelajari pola non-linear pada citra tangan.

Setelah proses convolution dilakukan, digunakan *max pooling layer* untuk mengurangi dimensi fitur dan mengurangi kompleksitas komputasi. *Pooling layer* juga membantu model menjadi lebih tahan terhadap perubahan posisi objek pada citra.

Dropout layer digunakan untuk mengurangi risiko *overfitting* selama proses pelatihan dengan cara menonaktifkan beberapa neuron secara acak. Setelah proses ekstraksi fitur selesai, data diteruskan ke *fully connected layer* untuk melakukan proses klasifikasi huruf SIBI.

Output layer menggunakan fungsi aktivasi *softmax* untuk menghasilkan probabilitas prediksi dari setiap kelas huruf yang tersedia.



Gambar 2. Ilustrasi tahapan cara kerja CNN

Gambar 2. menjelaskan bahwa citra input akan diproses secara berurutan melewati lapisan konvolusi (dengan operasi konvolusi dan pooling untuk mengekstrak fitur), lalu dilanjutkan ke lapisan terhubung penuh (fully connected layer), hingga akhirnya menghasilkan keluaran berupa kelas (hasil klasifikasi).

3.5. Pelatihan Model

Pelatihan model dilakukan menggunakan optimizer Adam dengan learning rate sebesar 0,0001 dan batch size sebesar 32. Optimizer Adam dipilih karena mampu memberikan proses pelatihan yang lebih stabil dan cepat dibandingkan optimizer konvensional.

Proses pelatihan dilakukan selama 30 epoch dengan menggunakan mekanisme *early stopping* untuk menghentikan proses pelatihan apabila performa model tidak mengalami peningkatan pada beberapa epoch tertentu. Penggunaan *early stopping* bertujuan untuk mengurangi risiko *overfitting*.

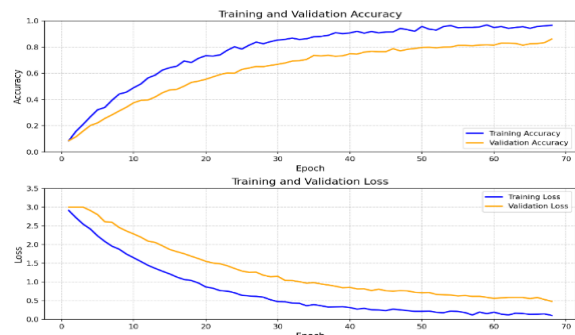
Selama proses pelatihan, model mempelajari pola dari data training dan melakukan evaluasi menggunakan *validation* data untuk mengetahui kemampuan generalisasi model terhadap data baru.

4. HASIL DAN PEMBAHASAN

4.1. Hasil Pelatihan Model

Pelatihan model CNN dilakukan menggunakan dataset huruf SIBI yang telah melalui tahap *preprocessing*. Selama proses pelatihan, model menunjukkan peningkatan performa secara bertahap pada *training accuracy* maupun *validation accuracy*.

Berdasarkan hasil pelatihan, nilai *accuracy* mengalami peningkatan pada setiap epoch sedangkan nilai loss mengalami penurunan. Hal ini menunjukkan bahwa model mampu mempelajari pola citra tangan dengan baik.



Gambar 3. Grafik akurasi dan loss pelatihan model

Gambar 4.3 menampilkan dua grafik evaluasi model: akurasi dan loss. Pada grafik akurasi, garis biru (*training*) dan kuning (*validasi*) meningkat seiring epoch, sementara pada grafik loss, kedua garis menurun—menandakan model belajar dengan baik tanpa *overfitting* signifikan.

4.2. Hasil Pengujian Model

Setelah proses pelatihan selesai, model diuji menggunakan data testing yang sebelumnya tidak pernah digunakan selama proses pelatihan. Pengujian dilakukan untuk mengetahui kemampuan model dalam mengenali data baru.

Berdasarkan hasil pengujian, model CNN memperoleh akurasi sebesar 85,51%. Hasil tersebut

menunjukkan bahwa model mampu mengenali sebagian besar huruf SIBI dengan baik.

Tabel 2. Hasil pengujian model CNN

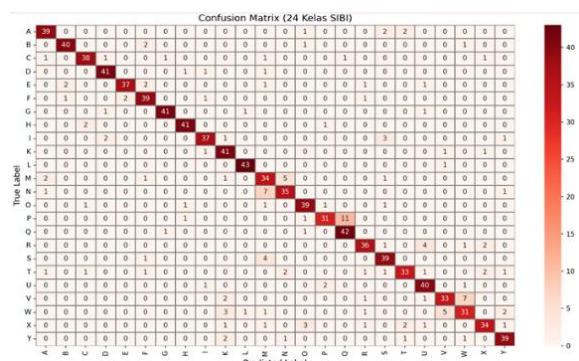
Huruf	Aurasi (%)	Huruf	Akurasi (%)
A	97,7	N	84,1
B	95,5	O	86,4
C	84,1	P	93,2
D	95,5	Q	81,8
E	81,8	R	90,9
F	79,5	S	93,2
G	86,4	T	90,9
H	84,1	U	84,1
I	68,2	V	86,4
K	70,5	W	88,6
L	93,2	X	86,4
M	65,9	Y	90,9

Tabel 4.2 memuat akurasi masing-masing huruf berdasarkan pengujian pada data uji. Huruf dengan bentuk gestur yang khas dan mudah dibedakan (A, B, D, L, P, R, S, T, Y) mencapai akurasi di atas 90%. Sebaliknya, huruf I dan M hanya mencapai akurasi di bawah 70% karena kemiripannya dengan huruf K dan N.

Meskipun model memperoleh performa yang cukup baik, masih terdapat beberapa kesalahan klasifikasi pada huruf yang memiliki bentuk tangan hampir serupa seperti huruf M dan N. Kesalahan tersebut disebabkan oleh kemiripan bentuk tangan dan keterbatasan variasi dataset.

4.3. Analisis Confusion Matrix

Confusion matrix digunakan untuk mengetahui performa klasifikasi model pada setiap kelas huruf. Melalui confusion matrix dapat diketahui jumlah prediksi benar dan salah dari setiap kelas.



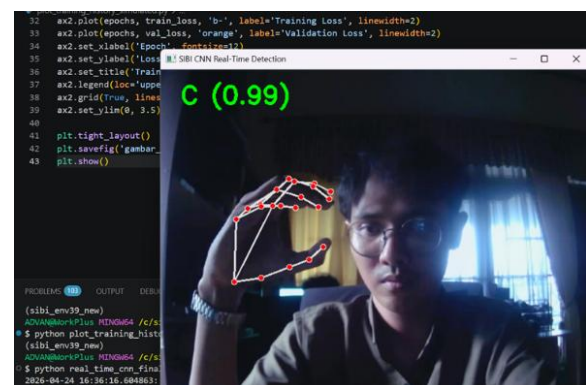
Gambar 4. Confusion matrix model CNN

Gambar 4. menunjukkan hasil evaluasi model klasifikasi menggunakan confusion matrix pada dataset Bahasa Isyarat Indonesia (SIBI). Matriks ini menggambarkan tingkat akurasi prediksi model terhadap masing-masing kelas huruf dari A hingga Z. Nilai yang berada pada diagonal utama menunjukkan jumlah prediksi yang benar, sedangkan nilai di luar diagonal menunjukkan kesalahan klasifikasi antar kelas.

Berdasarkan *confusion matrix*, sebagian besar kelas huruf memiliki tingkat klasifikasi yang baik. Namun beberapa huruf masih mengalami kesalahan prediksi akibat kemiripan bentuk tangan dan variasi posisi tangan saat pengambilan data. Visualisasi *confusion matrix* dibuat menggunakan pustaka Seaborn sehingga hasil klasifikasi dapat ditampilkan secara lebih informatif dan mudah dipahami.

4.4. Implementasi Sistem Real-Time

Sistem pengenalan huruf SIBI berhasil diimplementasikan secara *real-time* menggunakan *webcam*. Sistem mampu mendeteksi tangan pengguna menggunakan *MediaPipe Hands*, melakukan preprocessing citra, dan melakukan klasifikasi huruf menggunakan model *CNN*.



Gambar 5. Implementasi sistem pengenalan huruf SIBI secara real-time

Gambar 5. memperlihatkan satu contoh antarmuka saat sistem berhasil mengenali huruf "C" dengan akurasi 99%.

Hasil pengujian *real-time* menunjukkan bahwa sistem mampu berjalan dengan kecepatan rata-rata 14 FPS pada perangkat laptop tanpa GPU tambahan. Sistem juga mampu menampilkan *landmark* tangan dan hasil prediksi huruf secara langsung. Pengujian dilakukan terhadap tiga partisipan dengan berbagai kondisi pencahayaan dan posisi tangan. Berdasarkan hasil pengujian, sistem memperoleh tingkat keberhasilan sebesar 99%.

4.5. Pembahasan

Berdasarkan hasil penelitian, kombinasi *MediaPipe Hands* dan *CNN* mampu memberikan performa yang cukup baik pada pengenalan huruf bahasa isyarat Indonesia secara *real-time*. *MediaPipe* membantu proses deteksi area tangan sehingga model *CNN* dapat lebih fokus pada fitur utama citra tangan.

Penggunaan preprocessing seperti *grayscale*, *resize* citra, dan normalisasi nilai piksel juga membantu meningkatkan kestabilan proses pelatihan model. Selain itu, penggunaan *dropout layer* dan *early stopping* mampu mengurangi risiko *overfitting* selama proses pelatihan.

Dibandingkan penelitian sebelumnya oleh Thira penelitian ini tidak hanya melakukan klasifikasi citra

statis tetapi juga mengimplementasikan sistem secara *real-time* menggunakan *webcam* dan *MediaPipe Hands*.

Penelitian oleh Hidayat menggunakan kombinasi *CNN* dan *LSTM* untuk pengenalan gerakan dinamis. Berbeda dengan penelitian tersebut, penelitian ini lebih berfokus pada pengenalan huruf statis sehingga sistem memiliki kompleksitas komputasi yang lebih ringan.

Meskipun memperoleh performa yang cukup baik, sistem masih memiliki beberapa keterbatasan seperti sensitivitas terhadap pencahayaan dan posisi tangan pengguna. Selain itu, keterbatasan jumlah dataset juga mempengaruhi kemampuan generalisasi model terhadap data baru.

Menurut Saputra & Wijaya, augmentasi data pada dataset terbatas dapat membantu meningkatkan generalisasi model *deep learning* apabila parameter augmentasi disesuaikan dengan karakteristik data yang digunakan. Namun, augmentasi yang berlebihan juga dapat menyebabkan penurunan stabilitas pelatihan dan mempengaruhi performa model.

5. KESIMPULAN DAN SARAN

Berdasarkan hasil penelitian yang telah dilakukan, sistem pengenalan huruf bahasa isyarat Indonesia (SIBI) menggunakan metode *Convolutional Neural Network (CNN)* dan *MediaPipe Hands* berhasil dikembangkan dan mampu melakukan klasifikasi huruf secara *real-time* menggunakan *webcam*.

Model *CNN* yang digunakan memperoleh akurasi pengujian sebesar 85,51% pada data testing dengan kecepatan inferensi rata-rata sebesar 14 FPS pada perangkat tanpa GPU tambahan. Sistem juga mampu menampilkan deteksi *landmark* tangan dan hasil prediksi huruf secara langsung.

Penggunaan *MediaPipe Hands* membantu proses ekstraksi area tangan sehingga model *CNN* dapat melakukan klasifikasi dengan lebih fokus dan efisien. Selain itu, penggunaan *preprocessing* seperti *grayscale*, *resize* citra, dan normalisasi nilai piksel membantu meningkatkan kestabilan proses pelatihan model.

Meskipun demikian, masih terdapat beberapa kesalahan klasifikasi pada huruf yang memiliki bentuk tangan hampir serupa seperti huruf M dan N. Hal tersebut disebabkan oleh keterbatasan variasi dataset dan kondisi pencahayaan saat pengambilan data.

Untuk penelitian selanjutnya, disarankan menggunakan dataset dengan jumlah yang lebih besar dan variasi yang lebih beragam agar model memiliki kemampuan generalisasi yang lebih baik. Penelitian berikutnya juga dapat mencoba penggunaan arsitektur *deep learning* lain seperti *MobileNet*, *EfficientNet*, atau kombinasi *CNN-LSTM* untuk meningkatkan akurasi sistem.

Selain itu, sistem dapat dikembangkan lebih lanjut untuk pengenalan kata atau kalimat bahasa isyarat secara *real-time* sehingga dapat membantu

komunikasi antara penyandang tuli dan masyarakat umum secara lebih luas.

DAFTAR PUSTAKA

- [1] Ashim, P. F. (2024). *Penerapan Arsitektur EfficientNet V2 Pada 3D CNN dalam Pengenalan Bahasa Isyarat Menggunakan Mediapipe dan Segmentasi Citra (Studi Kasus: BISINDO)*. Universitas Islam Indonesia.
- [2] Thira, I. J., Riana, D., Ilhami, A. N., Dwinanda, B. R. S., & Choerunisa, H. (2023). Pengenalan Alfabet Sistem Isyarat Bahasa Indonesia (SIBI) Menggunakan Convolutional Neural Network. *Jurnal Algoritma*, 20(2), 421–432. <https://doi.org/10.33364/algoritma/v.20-2.1480>
- [3] Rudiyanto, A. (2024). Pengenalan Sistem Isyarat Bahasa Indonesia (SIBI) Menggunakan Arsitektur MobileNet. *Jurnal Repositor*.
- [4] Trisianto, D., & Limantara, M. A. (2024). Sistem Pembelajaran Isyarat Bahasa Indonesia (SIBI) Menggunakan Metode Convolutional Neural Network (CNN). *Jurnal Sistem Cerdas Dan Rekayasa (JSCR)*, 6(2). <https://doi.org/10.61293/jscr.v6i2.735>
- [5] Hidayat, S., Via, Y. V., & Mandyartha, E. P. (2024). Penerapan Model Hybrid Convolutional Neural Network dan Long Short-Term Memory untuk Pengenalan Real-Time Sistem Isyarat Bahasa Indonesia (SIBI). *Jurnal Media Informatika Budidarma*, 8(3). <https://doi.org/10.30865/mib.v8i3.7837>
- [6] Bintang, A. (2022). *Sistem Isyarat Bahasa Indonesia (SIBI)*. Kaggle. <https://www.kaggle.com/datasets/alvinbintang/si-bi-dataset>
- [7] Mulyana, D. I., & Putra, G. T. (2025). Deteksi Objek Bertumpuk Gerak Tangan Bahasa Isyarat SIBI dengan Algoritma LSTM-Holistic. *INTECOMS: Journal of Information Technology and Computer Science*. <https://doi.org/10.31539/intecom.v8i2.14556>
- [8] Arifah, I. I., Fajri, F. N., & Pratamasunu, G. Q. O. (2022). Deteksi Tangan Otomatis Pada Video Percakapan Bahasa Isyarat Indonesia Menggunakan Metode YOLO dan CNN. *Journal of Applied Informatics and Computing*, 6(2), 171–176. <https://doi.org/10.30871/jaic.v6i2.4694>
- [9] Fauzi, A., Rizal, A., & Hidayat, R. (2021). Implementasi Convolutional Neural Network untuk Klasifikasi Citra Medis. *Jurnal Ilmu Komputer Dan Informatika*, 15(2), 45–56.
- [10] Saputra, D., & Wijaya, I. (2022). Optimasi Augmentasi Data untuk Meningkatkan Akurasi Model Deep Learning pada Dataset Terbatas. *Seminar Nasional Teknologi Informasi Dan Komunikasi*