



Pembangunan *Dataset* Gestur Angka (1-5) yang Terkontekstualisasi Budaya sebagai inisiasi Basis Data Pengenalan untuk Alat Bantu Pembelajaran Matematika

Khawaritzmi Abdallah Ahmad^{1*}, Arian Nurrifqhi², Fettyana³, Nur Fadilah⁴

^{1,2,3,4} Jurusan Matematika, Fakultas MIPA, Universitas Negeri Makassar, South Sulawesi 90222, Indonesia

* Penulis Korespondensi. Email: khawaritzmi.abdallah@unm.ac.id

ABSTRAK

Perkembangan pesat Kecerdasan Buatan (*Artificial Intelligence*/AI) telah mendorong evolusi interaksi manusia-komputer menuju sistem berbasis gestur yang lebih natural dan intuitif. *Hand Gesture Recognition* (HGR) menjadi komponen penting dalam pengembangan media pembelajaran interaktif, khususnya untuk pengenalan angka 1-5 menggunakan gestur jari. Namun, sistem HGR konvensional masih didominasi bias budaya tertentu, terutama berbasis dataset *American Sign Language*, sehingga berisiko menimbulkan kesalahan interpretasi dan ketidaksesuaian secara normatif maupun kognitif. Penelitian ini menginisiasi pengembangan dataset gestur angka 1-5 yang lebih inklusif dengan mempertimbangkan keberagaman latar belakang budaya. Ekstraksi fitur dilakukan menggunakan MediaPipe Hands untuk memperoleh *hand landmarks* secara efisien dan akurat. *Dataset* yang dibangun kemudian digunakan untuk melatih model klasifikasi gestur tangan sebagai media pembelajaran matematika interaktif. Hasil pengujian menunjukkan bahwa model mampu mengenali kelas tertentu dengan baik, khususnya kelas 1 dan 5, dengan recall mencapai 1.00 serta *f1-score* tertinggi sebesar 0.85. Namun, performa pada kelas 3 dan 4 masih rendah akibat ketidakseimbangan data. Meskipun akurasi keseluruhan mencapai 66%, hasil ini menunjukkan perlunya optimalisasi lebih lanjut dalam strategi pelatihan untuk meningkatkan generalisasi model. *Dataset* yang dihasilkan diharapkan dapat menjadi fondasi pengembangan sistem HGR yang lebih aman, akurat, dan inklusif secara budaya untuk mendukung pembelajaran matematika interaktif.

Kata Kunci:

Pengenalan Gestur Tangan; Keberagaman Budaya; Berhitung dengan Jari; Media Pembelajaran Interaktif; Matematika Terapan.

ABSTRACT

The rapid advancement of Artificial Intelligence (AI) has transformed human-computer interaction toward more natural and intuitive vision-based gesture systems. Hand Gesture Recognition (HGR) plays a crucial role in the development of interactive learning media, particularly for recognizing finger gestures representing numbers 1-5. However, conventional HGR systems are predominantly built upon culturally biased datasets, especially those derived from American Sign Language, which may lead to misinterpretation and normative mismatch across different cultural contexts. This study initiates the development of a culturally inclusive hand gesture dataset for numbers 1-5 by incorporating diverse cultural finger-counting habits. Feature extraction is performed using the MediaPipe Hands framework to obtain accurate and efficient hand landmarks. The constructed dataset is then utilized to train a gesture classification model for interactive mathematics learning applications. Experimental results

show that the model achieves strong performance in certain classes, particularly for classes 1 and 5, with a recall of 1.00 and the highest f1-score of 0.85. However, challenges remain in classifying classes 3 and 4 due to class imbalance, resulting in lower recall and f1-scores. Although the overall accuracy reaches 66%, this imbalance indicates the need for further optimization in training strategies to improve model generalization. The proposed dataset is expected to serve as a more relevant and culturally inclusive foundation for developing safe, accurate, and adaptive HGR-based interactive mathematics learning systems.

Keywords:

Hand Gesture Recognition, Cultural Diversity, Finger Counting, Interactive Learning Media, Mathematics Application

1. Pendahuluan

Perkembangan pesat Kecerdasan Buatan (AI) telah mendorong interaksi manusia-komputer (HCI) melampaui antarmuka tradisional, menuju sistem berbasis gestur bebas-tangan (*vision-based*) yang lebih intuitif dan alami [1]. Dalam konteks ini, Pengenalan Gestur Tangan (*Hand Gesture Recognition* - HGR) telah menjadi bidang penelitian penting dengan aplikasi yang luas, termasuk dalam sistem otomotif, realitas virtual, dan terutama alat bantu pembelajaran interaktif [2]. Gestur jari yang merepresentasikan angka diskrit, khususnya angka 1 hingga 5, sangat fundamental dan berfungsi sebagai primitif fisik untuk perintah dasar seperti pemilihan menu, navigasi, atau masukan numerik [1].

Meskipun akurasi kinematic, yaitu kemampuan melacak pose tangan dan 21 *hand landmarks* secara *precision*, telah mencapai kematangan teknis yang tinggi [11-13], keberhasilan HGR secara global terhambat oleh kesenjangan kultural dan bias *US-centric* [3, 4]. Penelitian HGR konvensional cenderung memperlakukan gestur angka (1-5) sebagai entitas universal yang semata-mata bersifat geometris [3, 4], mengabaikan bahwa gestur tersebut memiliki polysemi dan muatan simbolis yang tinggi yang bervariasi antar budaya.

Keterbatasan ini berasal dari penggunaan ulang *dataset* fundamental yang memiliki bias geografis, seperti *American Sign Language* (ASL), yang secara tidak sengaja telah membentuk bias *default* dalam *baseline* AI saat ini [4]. Gestur yang termasuk dalam kategori 1-5 membawa risiko ofensif yang signifikan; sebagai contoh, gestur "jempol ke atas" (yang merupakan representasi pose "1" dalam beberapa sistem) adalah hinaan yang sangat berat di Iran, Yunani, atau Afghanistan. Kegagalan dalam mengatasi konteks normatif kultural ini berisiko menyebabkan misinterpretasi, menyinggung pengguna, atau bahkan menghasilkan kegagalan fungsional yang serius [3, 5].

Selain masalah normatif, variasi dalam kebiasaan berhitung jari (*finger-counting habits*) menggunakan lima jari sangat dipengaruhi oleh faktor budaya, praktik kognitif, serta kebiasaan yang diajarkan sejak dini [6, 7]. Kebiasaan berhitung jari ini terbukti bervariasi secara signifikan antarbudaya dan membentuk sistem representasi numerik yang berbeda-beda. Pada sistem berhitung Eropa kontinental seperti di Jerman, Prancis, dan Italia, hitungan umumnya dimulai dari ibu jari sebagai representasi angka satu, kemudian diikuti oleh jari telunjuk hingga kelingking untuk mencapai angka lima. Sebaliknya, di Amerika Serikat dan beberapa komunitas penutur bahasa Spanyol, hitungan dimulai dari jari telunjuk sebagai angka satu dan ibu jari baru digunakan sebagai representasi angka lima [6, 7, 10].

Sementara itu, beberapa budaya seperti Tiongkok menggunakan variasi pose tangan yang lebih kompleks untuk memungkinkan perhitungan hingga sepuluh hanya dengan

satu tangan. Perbedaan ini terbukti memengaruhi pemrosesan numerik secara kognitif, di mana pengguna dari budaya Tiongkok tidak menunjukkan fenomena *five-break effect* sebagaimana yang dijumpai pada partisipan dari Jerman [1, 6].

Perbedaan cara memulai hitungan, orientasi telapak tangan, serta penggunaan satu tangan atau dua tangan dalam merepresentasikan angka terbukti memengaruhi pemrosesan numerik secara kognitif. Hal ini menunjukkan bahwa sistem *Hand Gesture Recognition* (HGR) dalam konteks pendidikan tidak cukup hanya mengandalkan akurasi pengenalan pose, tetapi juga harus mampu mengakomodasi keberagaman kognitif dan normatif pengguna. Oleh karena itu, dataset HGR yang ada saat ini masih belum memadai untuk mendukung pengembangan alat bantu pembelajaran matematika interaktif yang aman dan relevan secara kognitif.

Dalam konteks tersebut, penelitian ini menginisiasi pembangunan dataset gestur angka 1–5 yang lebih natural dengan melibatkan keberagaman latar belakang budaya. Untuk mendukung proses ekstraksi fitur yang konsisten dan efisien, digunakan framework MediaPipe Hands sebagai alat utama dalam memperoleh *hand keypoints*, mengingat performanya yang telah terbukti akurat, stabil, dan ringan untuk pemrosesan *real-time* [11–13]. *Dataset* yang dikembangkan selanjutnya akan diinvestigasi pemanfaatannya dalam sistem pengenalan gestur tangan sebagai media pembelajaran matematika yang inklusif dan adaptif.

Dataset ini diharapkan dapat menjadi basis data yang lebih relevan dan inklusif dalam pengembangan sistem HGR, khususnya pada alat bantu pembelajaran matematika interaktif. Dengan memperhatikan perbedaan budaya dalam berhitung serta interpretasi simbolik gestur, penelitian ini bertujuan untuk membangun sistem yang lebih aman, akurat, dan responsif terhadap kebutuhan pengguna global, sehingga keberagaman budaya dapat menjadi fondasi dalam pengembangan media pembelajaran yang lebih efektif dan inklusif.

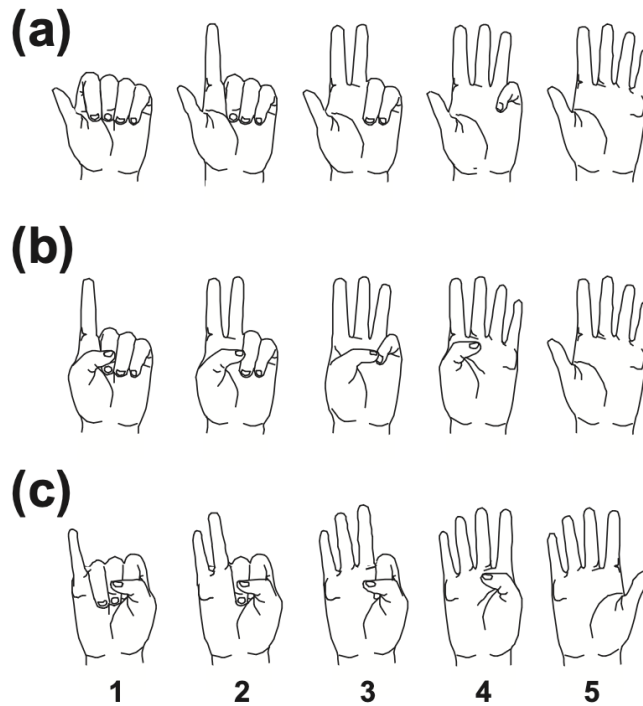
2. Metode atau Model

2.1. Karakteristik Partisipan dan Keberagaman Budaya

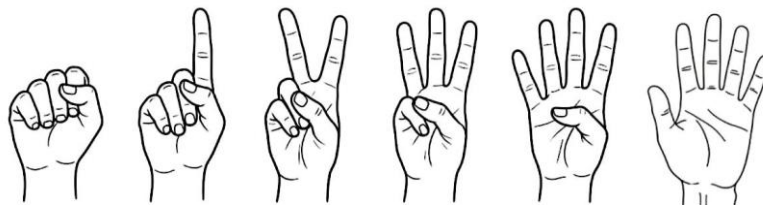
Dataset ini dikembangkan dengan melibatkan total 5 partisipan yang berasal dari latar belakang mahasiswa dalam lingkungan kampus. Partisipan diberikan arahan dan pengetahuan mengenai representasi budaya dengan kebiasaan berhitung jari yang berbeda, yaitu misalnya: Eropa (dengan penggunaan ibu jari sebagai awal hitungan), China dan Amerika Serikat (dengan penggunaan jari telunjuk sebagai awal hitungan), serta Iran (dengan penggunaan jari kelingking sebagai awal hitungan) seperti pada Gambar 1. Keberagaman ini bertujuan untuk menangkap variasi alami dalam representasi gestur angka 1–5 yang dipengaruhi oleh kebiasaan kognitif dan budaya masing-masing subjek.

Secara umum, *dataset finger counting* yang banyak digunakan dalam penelitian pengenalan gestur tangan masih merepresentasikan angka sebagai pose statis tunggal yang bersifat universal [2, 4, 8, 11, 12, 13, 15], seperti yang ditunjukkan pada Gambar 2, yaitu satu jari untuk angka 1, dua jari untuk angka 2, hingga lima jari untuk angka 5. Representasi ini mengasumsikan bahwa cara berhitung dengan jari bersifat seragam di seluruh dunia, tanpa mempertimbangkan variasi budaya dalam memulai hitungan, orientasi telapak tangan, maupun makna simbolik dari suatu gestur seperti pada HaGRID, IPN Hands, dan Fingers Dataset dari Kaggle [2, 8, 15]. Akibatnya, dataset semacam ini cenderung hanya menangkap

aspek geometris dari pose tangan, namun belum merepresentasikan kompleksitas kognitif dan kultural yang sebenarnya melekat pada praktik finger counting di berbagai budaya [14 - 15].



Gambar 1. Variasi sistem lima jari, yang tercatat di sebagian besar Eropa (a), China dan Amerika Serikat (b), serta Iran (c). Ketiga variasi ini berbeda dalam urutan jari yang digunakan: ibu jari pada (a), jari telunjuk pada (b), dan jari kelingking pada (c) [15]

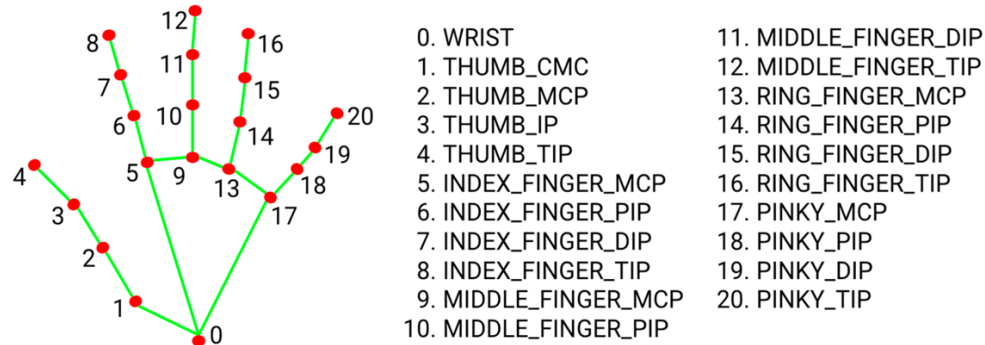


Gambar 2. Contoh umum representasi finger counting angka 0-5 yang sering digunakan dalam dataset HGR konvensional, di mana setiap angka direpresentasikan sebagai pose statis berbasis jumlah jari yang terangkat tanpa mempertimbangkan variasi budaya [4,15]

2.2. Mediapipe Hands

MediaPipe Hands digunakan dalam penelitian ini sebagai framework untuk memperoleh fitur spasial dari gestur tangan. *Framework* yang dikembangkan oleh Google LLC ini mampu mendeteksi serta melacak posisi tangan dan menghasilkan 21 titik *landmark* pada jari dan telapak tangan, sebagaimana terlihat pada Gambar 3 [9]. Proses deteksi telapak tangan pada MediaPipe Hands didukung oleh model yang dilatih menggunakan tiga jenis dataset dengan total lebih dari 116.000 sampel. Dataset tersebut terdiri atas sekitar 6.000 sampel dari In-the-wild dataset, 10.000 sampel yang

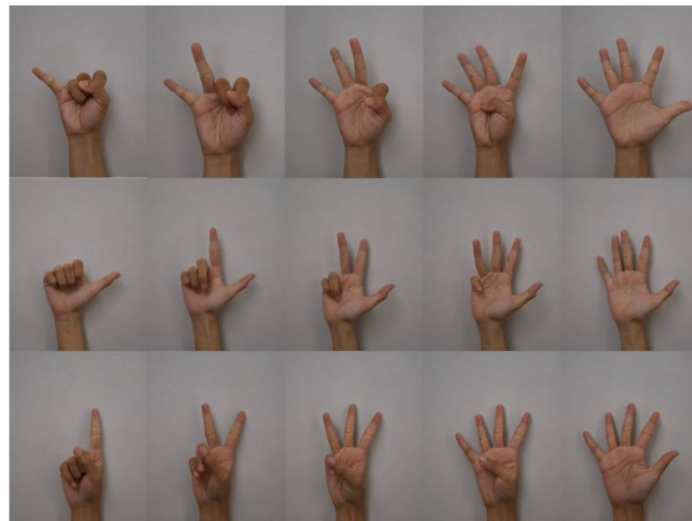
dikumpulkan secara internal, dan 100.000 sampel dari data sintetis gerakan tangan [9]. Setiap nilai z pada landmark menunjukkan kedalaman relatif terhadap titik pergelangan tangan. Sementara itu, ketiga koordinat x , y , dan z telah melalui proses normalisasi ke rentang $[0.0, 1.0]$ untuk memastikan konsistensi skala data [9].



Gambar 3. Keypoint pada *framework* Mediapipe Hands [9]

2.3. Akuisisi Dataset

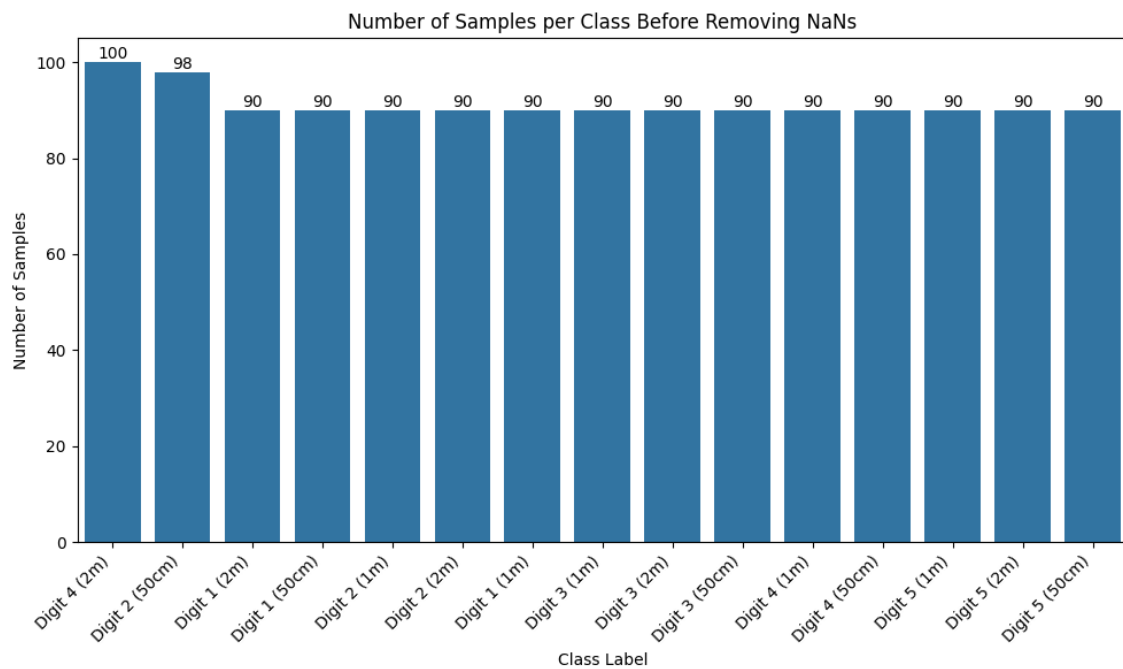
Penelitian ini menggunakan pendekatan eksperimental-pengembangan (*experimental-developmental*) dengan fokus utama pada pembangunan basis data (*dataset construction*). Tujuan rancangan ini adalah menghasilkan *Natural Hand Digit Dataset* yang robust dan spesifik, yang berfungsi sebagai basis data pengenalan (*recognition database*) untuk aplikasi *Human-Computer Interaction* (HCI) dalam konteks pembelajaran matematika interaktif. Dataset yang digunakan dalam proyek ini adalah *Natural Hand Digit Dataset*, yang dikembangkan secara mandiri (*proprietary*) dan didesain khusus untuk mendukung pengenalan gestur tangan berupa angka 1 hingga 5. Dataset ini diakuisisi sebagai video sebanyak 45 video dan kemudian dipecah menjadi *frame* statis yang diseleksi untuk merepresentasikan gestur tangan dalam berbagai skenario dunia nyata seperti terlihat pada Gambar 4.



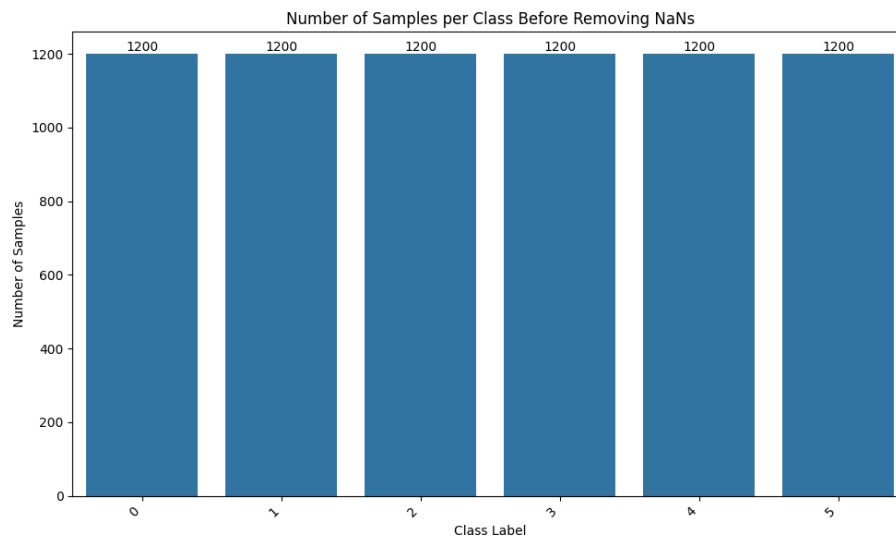
Gambar 4. *Natural Hand Digit Dataset* yang dikoleksi

Pengambilan data dilakukan dengan menggunakan perangkat smartphone yang memiliki spesifikasi teknis tinggi untuk menghasilkan rekaman berkualitas. Perangkat ini menjalankan sistem operasi Android 14, dilengkapi dengan chipset Exynos 1480, dan mendukung perekaman video Ultra HD 4K (3840 x 2160 piksel) pada 30 *frames per second* (fps). Protokol pengambilan data dirancang untuk menciptakan variasi yang kaya dan mencerminkan skenario aplikasi dunia nyata, dengan data dikumpulkan pada tiga variasi jarak berbeda dari kamera, yaitu 50 cm, 1 m, dan 2 meter, guna melatih model agar tahan terhadap perubahan skala. Proses pengambilan juga dilakukan dalam kondisi pencahayaan rendah untuk meniru lingkungan yang tidak ideal, meningkatkan ketahanan model terhadap noise dan bayangan. Latar belakang yang digunakan adalah putih netral untuk meminimalkan gangguan visual dan memastikan fokus pada gestur tangan. Kamera dipasang pada ketinggian 125 cm menggunakan tripod, mencerminkan tinggi rata-rata bahu atau dada manusia, sehingga menghasilkan perspektif pengambilan data yang konsisten, mengimitasi sudut pandang manusia.

Proses anotasi dan validasi data dilakukan secara manual oleh satu orang anotator yang memiliki pemahaman mendalam tentang gestur numerik dan konteks budaya, tanpa adanya pengukuran lain. Anotator bertanggung jawab memberikan label kelas yang akurat pada data untuk digunakan dalam pelatihan model, dengan memastikan validitas dan reliabilitas label yang dihasilkan, serta memperhatikan interpretasi yang tepat berdasarkan pengetahuan domain. Tahapan selanjutnya adalah pra-pemrosesan data, dimulai dengan ekstraksi fitur hand landmark menggunakan MediaPipe Hands. Setelah ekstraksi, data keypoint yang diperoleh akan melalui proses verifikasi manual oleh anotator. Verifikasi ini memastikan akurasi keypoint serta pengklasifikasian dan koreksi representasi gestur sesuai dengan konteks budaya yang ditentukan dalam penelitian. Proses ini bertujuan untuk memastikan dataset yang dihasilkan tidak hanya akurat secara teknis, tetapi juga inklusif dan relevan dalam berbagai budaya. Dataset yang telah divalidasi dan dikoreksi kemudian disusun dalam format terstruktur (CSV), siap digunakan untuk melatih model klasifikasi pada alat bantu pembelajaran matematika interaktif.

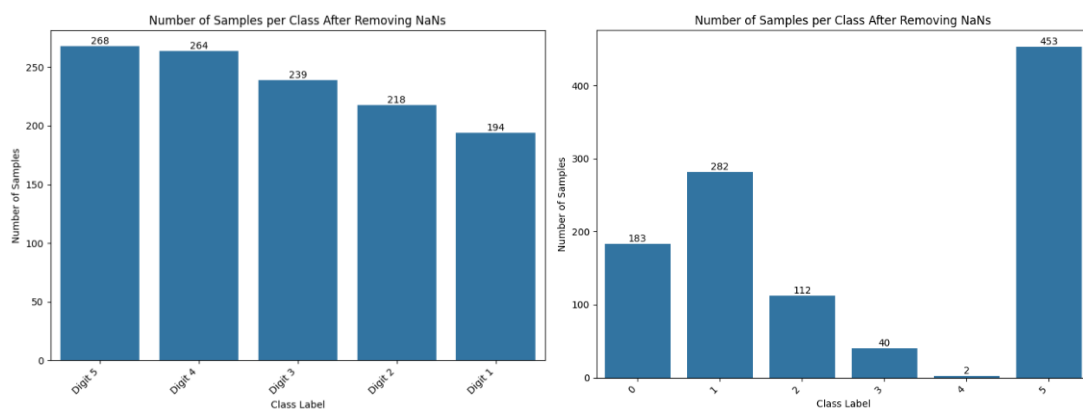


(a). Jumlah sample sebelum *preprocessing* pada *Natural Hand Digit Dataset*



(b). Jumlah sample sebelum *preprocessing* pada *Fingers Dataset* Kaggle
Gambar 5. Jumlah sample sebelum *preprocessing*

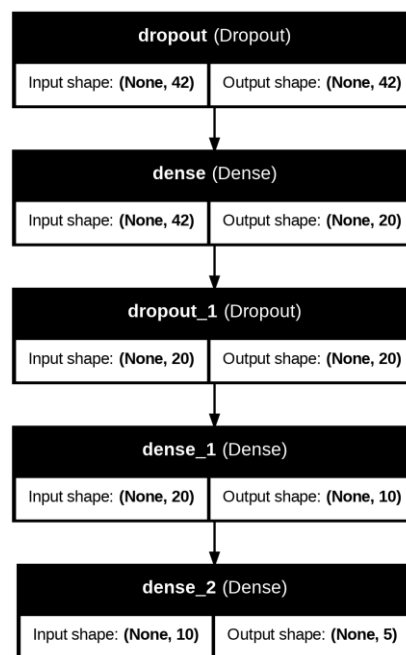
Gambar 5 merupakan grafik distribusi jumlah citra pada setiap kelas dan variasi jarak pengambilan (1 m, 2 m, dan 50 cm) menunjukkan bahwa pada tahap awal seluruh kelas digit 1-5 dikumpulkan secara relatif seimbang, masing-masing dengan jumlah citra yang hampir sama untuk setiap variasi jarak. Hal ini mengindikasikan bahwa desain pengambilan data awal telah dirancang secara konsisten untuk meminimalkan bias jumlah sampel antar kelas. Namun, pada Gambar 6 setelah citra diproses menggunakan MediaPipe Hands, sebagian data gagal terdeteksi keypoint-nya dan menghasilkan nilai (*Not a Number*) NaN, sehingga harus dieliminasi. Dampaknya terlihat pada grafik sebelumnya, di mana jumlah sampel valid menjadi tidak seimbang, dengan Digit 5 dan Digit 4 memiliki sampel terbanyak, sementara Digit 1 menjadi kelas dengan jumlah sampel paling sedikit. Ketidakseimbangan ini menunjukkan bahwa meskipun distribusi citra awal sudah merata, faktor teknis seperti kegagalan deteksi akibat jarak, pencahayaan, oklusi jari, dan posisi tangan tetap memengaruhi kualitas data akhir dan berpotensi berdampak pada kinerja model klasifikasi, khususnya pada kelas dengan jumlah sampel yang lebih sedikit.



Gambar 6. Jumlah sample setelah *preprocessing*, *Natural Hand Digit Dataset* (kiri) dan *Fingers Dataset* Kaggle (kanan)

2.4. Desain Model

Model yang digunakan dalam penelitian ini adalah jaringan saraf tiruan berbasis Keras dengan arsitektur *Sequential*, yang terdiri dari beberapa lapisan untuk memproses data secara bertahap. Pada Gambar 7, model ini menerima input berupa data dengan bentuk $(21 * 2)$, yang kemungkinan besar merujuk pada fitur *hand landmark* yang diekstraksi dari gestur tangan. Lapisan pertama adalah *Dropout* 0.2, yang berfungsi untuk mengurangi overfitting dengan membuang 20% secara acak dari neuron selama pelatihan. Selanjutnya, terdapat lapisan *Dense* dengan 20 neuron dan fungsi aktivasi ReLU yang digunakan untuk memodelkan hubungan non-linear antara fitur input.



Gambar 7. Desain model *sequential* yang digunakan

Untuk lebih mengurangi *overfitting*, digunakan *Dropout* 0.4 pada lapisan berikutnya, yang membuang 40% secara acak dari neuron selama pelatihan. Setelah itu, ada lagi lapisan *Dense* dengan 10 neuron dan fungsi aktivasi ReLU, yang memungkinkan model untuk memproses informasi lebih lanjut sebelum mencapai lapisan output. Lapisan output menggunakan *Dense* 5 neuron dengan fungsi aktivasi *Softmax*, yang dirancang untuk menghasilkan probabilitas yang sesuai dengan lima kelas output, sesuai dengan tugas klasifikasi.

Table 1. *Hyperparameter* dalam pelatihan model

Parameter	Nilai
<i>Learning Rate</i>	0.1
<i>Batch Size</i>	128
<i>Epoch</i>	300

<i>Optimizer</i>	Adam
<i>Loss Function</i>	<i>sparse_categorical_crossentropy</i>
<i>Activation Function</i>	<i>Softmax</i>
<i>Dropouts</i>	20% and 40%
<i>Total Parameters</i>	1,125

Model ini memiliki total 1.125 parameter, yang semuanya dapat dilatih selama proses pelatihan. Untuk mengoptimalkan kinerja model, digunakan optimizer “adam” dan *loss function sparse_categorical_crossentropy*, yang sesuai untuk klasifikasi multikelas. Selama pelatihan, dua *callback* digunakan: *ModelCheckpoint* untuk menyimpan model terbaik dan *EarlyStopping* untuk menghentikan pelatihan lebih awal jika tidak ada peningkatan yang signifikan pada *loss validasi*. Dengan menggunakan teknik *dropout* dan lapisan *dense*, model ini dirancang untuk menghindari *overfitting* dan meningkatkan kemampuan generalisasi. Untuk *hyperparameter* yang digunakan pada pelatihan model dapat dilihat pada Tabel 1.

2.5. Metode Evaluasi

Metodologi klasifikasi yang diterapkan dalam penelitian ini menggunakan metode klasifikasi multiclass, yang melibatkan lebih dari dua kelas dalam proses klasifikasi. Pada kasus ini, dataset yang digunakan memiliki 5 kelas, yang cukup representatif untuk tugas klasifikasi. Akurasi, *precision*, *recall*, *f1-score*, dan *loss* digunakan sebagai ukuran untuk menilai apakah kinerja model setelah pelatihan memenuhi standar yang diinginkan. Akurasi (dilihat pada Persamaan 3) didefinisikan sebagai rasio antara jumlah prediksi yang benar dengan total prediksi, tanpa memperhatikan kelas. Sebaliknya, *f1-score* (dilihat pada Persamaan 4) merupakan rata-rata harmonis dari *precision* dan *recall* model, di mana *precision* (dilihat pada Persamaan 1) mengacu pada proporsi item relevan yang berhasil diprediksi, sementara *recall* (dilihat pada Persamaan 2) menunjukkan rasio jumlah item relevan yang berhasil diprediksi. Nilai variabel untuk *precision*, *recall*, akurasi, dan *f1-score* dapat diperoleh dari matriks kebingungan yang disajikan pada Tabel 2.

Table 2. Tabel *confusion matrix*

Kelas Aktual	Kelas Prediksi			
	Classes	<i>a</i>	<i>b</i>	<i>c</i>
	<i>a</i>	TP_a	$F_{b a}$	$F_{c a}$
	<i>b</i>	$F_{a b}$	TP_b	$F_{c b}$
	<i>c</i>	$F_{a c}$	$F_{b c}$	TP_c

Sumber: [11]

Dalam hal perhitungan *loss*, digunakan *Categorical Cross-Entropy* (*cross-entropy kategorikal*), yang merupakan fungsi untuk menghitung *loss* yang umum digunakan dalam tugas klasifikasi multiclass. Fungsi ini sering juga dikenal dengan sebutan *softmax*

loss (lihat pada Persamaan 5). Akurasi, tingkat *loss*, dan *f1-score* menjadi parameter utama dalam mengevaluasi apakah model yang dihasilkan telah memenuhi kualitas yang diharapkan. Semakin tinggi nilai akurasi, yaitu ketika nilai antara akurasi pada data pelatihan dan validasi semakin mendekati, maka model dapat dianggap semakin baik. Sedangkan *f1-score* yang tinggi menunjukkan bahwa model mampu mengklasifikasikan setiap kelas dengan baik. Dalam kondisi overfitting, performa pada data validasi akan jauh lebih rendah dibandingkan dengan akurasi pada data pelatihan, sementara pada nilai *loss*, semakin rendah nilai *loss* (yang mendekati 0), semakin baik model tersebut, dengan catatan bahwa nilai *loss* pada validasi dan pelatihan harus relatif sama dan tidak memiliki perbedaan yang signifikan.

$$Precision_a = \frac{TP_a}{TP_a + \sum_X F_{a|X}} \quad (1)$$

$$Recall_a = \frac{TP_a}{TP_a + \sum_X F_{X|a}} \quad (2)$$

$$Accuracy = \frac{\sum_a TP_a}{\sum_a TP_a + \sum_a F_{X|a} + \sum_a F_{a|X}} = \frac{\sum_a TP_a}{N} \quad (3)$$

$$F1 - Score = 2 \cdot \frac{precision \cdot recall}{precision + recall} \quad (4)$$

$$Categorical CE loss = -\log\left(\frac{e^{s_i}}{\sum_j^C e^{s_i}}\right) \quad (5)$$

2.6. Metode Pelatihan

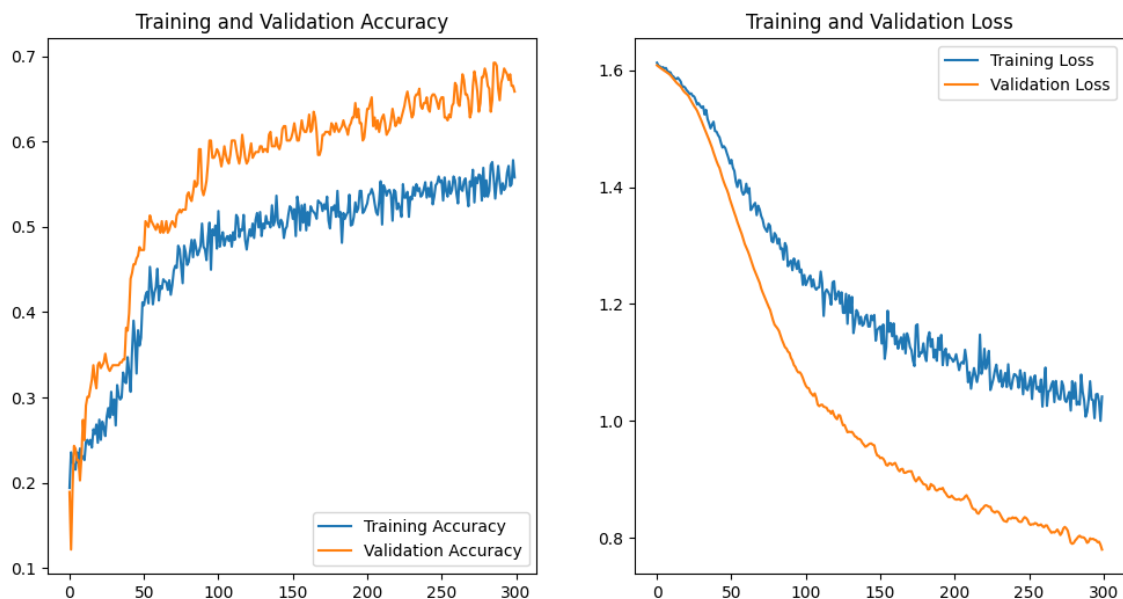
Pada penelitian ini, proses pelatihan model dilakukan untuk membandingkan kinerja sistem klasifikasi finger counting pada dua dataset, yaitu dataset yang diajukan dalam penelitian ini (*Natural Hand Digit Dataset*) dan *Fingers Kaggle Dataset* sebagai dataset pembanding. Untuk menjaga konsistensi dan objektivitas evaluasi, kedua dataset dibagi menggunakan skema yang sama, yaitu 75% sebagai data latih (*training set*) dan 25% sebagai data uji/validasi (*validation set*). Pembagian data dilakukan secara acak dengan tetap menjaga keseimbangan jumlah sampel pada setiap kelas.

Sebelum proses pelatihan dilakukan, seluruh data melewati tahap pra-pemrosesan yang sama, meliputi normalisasi nilai fitur dan penyelarasan format data agar sesuai dengan kebutuhan input model. Dengan perlakuan yang seragam ini, perbandingan kinerja model pada kedua dataset dapat dilakukan secara adil (*fair comparison*). Proses pelatihan model dilakukan menggunakan dataset latih masing-masing, kemudian evaluasi dilakukan menggunakan data uji yang tidak terlibat dalam proses pembelajaran. Kinerja model dianalisis menggunakan beberapa metrik evaluasi, terutama *confusion matrix*, untuk melihat distribusi prediksi benar dan kesalahan klasifikasi pada setiap kelas.

3. Hasil dan Pembahasan

Gambar 8 memperlihatkan dua grafik yang menggambarkan hasil pelatihan model dalam hal akurasi dan loss selama proses pelatihan. Grafik pertama menunjukkan hubungan antara epoch dan akurasi untuk data pelatihan (*training*) dan data validasi (*validation*). Terlihat bahwa training accuracy (garis biru) meningkat secara konsisten seiring berjalannya waktu, meskipun dengan fluktuasi yang menunjukkan ketidakstabilan dalam proses pelatihan. Di sisi lain, validation accuracy (garis oranye) juga menunjukkan peningkatan, namun tidak secepat akurasi pada data pelatihan, yang dapat mengindikasikan adanya *overfitting*, di mana model cenderung lebih baik dalam mengklasifikasikan data pelatihan dibandingkan data validasi.

Grafik kedua menggambarkan *loss* (kerugian) untuk kedua dataset tersebut. *Training loss* (garis biru) menunjukkan penurunan yang konsisten, mencerminkan perbaikan model dalam mengklasifikasikan data pelatihan. Namun, *validation loss* (garis oranye) juga menurun, meskipun tidak secepat penurunan pada data pelatihan. Perbedaan antara kedua *loss* ini menunjukkan adanya kecenderungan *overfitting*, di mana model menyesuaikan diri dengan baik pada data pelatihan tetapi tidak mampu menggeneralisasi dengan baik pada data validasi. Secara keseluruhan, grafik ini mengindikasikan bahwa meskipun model menunjukkan kinerja yang sangat baik pada data pelatihan, masih ada ruang untuk perbaikan agar model dapat lebih baik dalam mengklasifikasikan data yang tidak terlihat sebelumnya.



Gambar 8. Akurasi dan *loss* dalam proses pelatihan model

Table 3. Hasil pelatihan model *Natural Hand Digit Dataset*

Kelas	<i>Precision</i>	<i>Recall</i>	<i>F1-Score</i>	Support
Digit 1	0.74	1.00	0.85	46
Digit 2	0.89	0.60	0.72	57
Digit 3	0.88	0.23	0.37	64
Digit 4	0.53	0.57	0.55	67
Digit 5	0.58	1.00	0.73	62
Accuracy	0.66			296
Macro avg	0.73	0.68	0.64	296
Weighted avg	0.72	0.66	0.63	296

Hasil dari *classification report* pada Tabel 3 memberikan gambaran yang lebih mendalam mengenai kinerja model dalam mengklasifikasikan gestur jari yang merepresentasikan angka 1 hingga 5. Berdasarkan metrik *precision*, *recall*, dan *f1-score*, model menunjukkan kinerja yang bervariasi untuk masing-masing kelas. Untuk kelas 0, model menunjukkan kinerja terbaik dengan *precision* 0.74 dan *recall* 1.00, yang berarti model sangat akurat dalam memprediksi kelas ini dan berhasil mengenali semua label yang benar. Kelas 1 memiliki *precision* tinggi (0.89), namun *recall* yang lebih rendah (0.60), menunjukkan bahwa meskipun model sering memprediksi kelas ini dengan benar, masih ada beberapa prediksi yang salah. Kelas 3 memiliki *precision* 0.88, namun *recall* yang sangat rendah (0.23), yang menunjukkan bahwa model kesulitan dalam mengenali gestur pada kelas ini dengan akurat. Kelas 4 menunjukkan *precision* 0.53 dan *recall* 0.57, dengan *f1-score* 0.55, yang menunjukkan kinerja yang kurang optimal namun masih cukup seimbang antara *precision* dan *recall*. Sementara itu, kelas 5 memiliki *precision* 0.58 dan *recall* 1.00, dengan *f1-score* 0.73, yang berarti meskipun ada beberapa kesalahan dalam prediksi, model dapat mengenali kelas ini dengan sangat baik.

Table 4. Perbandingan hasil pelatihan model

<i>Metric</i>	<i>Model 1</i> (<i>Fingers</i>)	<i>Model 2</i> (<i>Natural</i> <i>Hand</i> <i>Digit</i>)
<i>Accuracy</i>	0.95	0.66
<i>Macro Avg</i>	0.75	0.73
<i>Precision</i>		
<i>Macro Avg</i>	0.79	0.68
<i>Recall</i>		
<i>Macro Avg F1-Score</i>	0.77	0.64
<i>Weighted Avg</i>	0.91	0.72
<i>Precision</i>		
<i>Weighted Avg</i>	0.95	0.66
<i>Recall</i>		
<i>Weighted Avg</i>	0.93	0.63
<i>F1-Score</i>		

Pada Tabel 4 Secara keseluruhan, Model 2 memiliki akurasi 66% (0.66), yang menunjukkan bahwa sekitar dua pertiga dari prediksi model adalah benar. Meskipun demikian, macro average menunjukkan bahwa meskipun model ini memiliki precision yang relatif baik (0.73), recall (0.68), dan F1-score (0.64), terdapat ketidakseimbangan dalam kinerja antar kelas. Hal ini tercermin pada nilai weighted average yang lebih rendah (F1-score 0.63), yang menunjukkan bahwa kelas-kelas dengan jumlah sampel lebih banyak, seperti kelas 1 dan 5, memberikan kontribusi yang lebih besar terhadap kinerja model secara keseluruhan. Oleh karena itu, meskipun Model 2 efektif dalam beberapa kelas, masih ada tantangan dalam meningkatkan kinerja, khususnya untuk kelas 3 dan 4, yang memerlukan perhatian lebih dalam pengolahan data dan perbaikan model untuk mencapai klasifikasi yang lebih seimbang dan akurat.

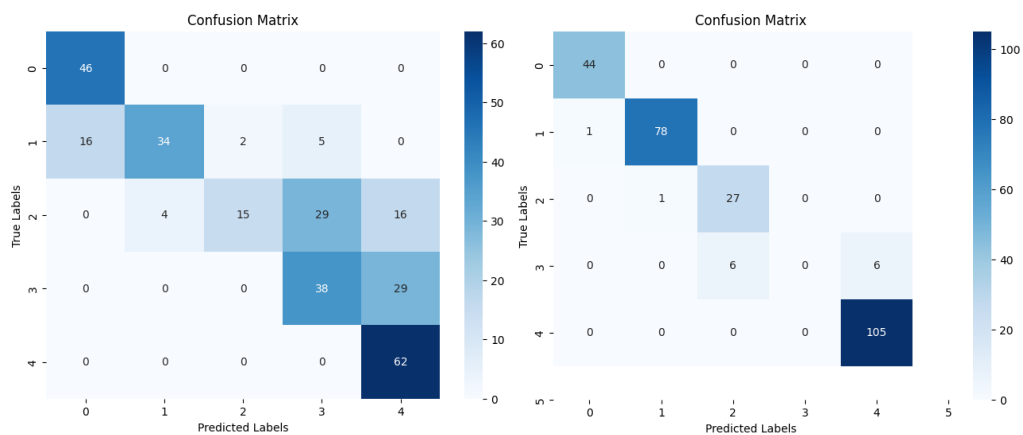
Di sisi lain, Model 1 menunjukkan akurasi yang jauh lebih tinggi, yaitu 95% (0.95), yang pada pandangan pertama menunjukkan kinerja yang lebih unggul. Namun, ketika dilihat lebih mendalam dalam konteks distribusi kelas pada dataset, kita menemukan adanya ketidakseimbangan yang sangat mencolok, terutama karena kelas 5 memiliki jumlah sampel yang jauh lebih banyak dibandingkan kelas lainnya. Ketidakseimbangan ini menyebabkan Model 1 cenderung mengutamakan klasifikasi kelas dominan, menghasilkan performa yang sangat baik untuk kelas 5 namun buruk untuk kelas-kelas dengan sampel lebih sedikit, seperti kelas 3 dan kelas 4. Dalam hal ini, Model 1 mengalami bias terhadap kelas yang lebih besar, yang dikenal dengan istilah class imbalance. Bias ini terjadi karena model akan meminimalkan kesalahan keseluruhan (accuracy) dengan memprioritaskan klasifikasi pada kelas dominan, meskipun hal ini mengorbankan kinerja pada kelas-kelas minoritas.

Meskipun Model 2 memiliki akurasi yang lebih rendah, yaitu 66%, model ini memiliki keunggulan dalam hal keseimbangan data. Distribusi data pada Model 2 lebih seimbang, dengan tidak ada kelas yang jauh lebih dominan dibandingkan kelas lainnya. Dalam kondisi ini, Model 2 tidak terpengaruh oleh bias yang kuat terhadap kelas tertentu, yang berarti model ini memberikan performa yang lebih adil dan konsisten di seluruh kelas. Keunggulan dalam keseimbangan data memungkinkan Model 2 untuk menghindari overfitting pada kelas dominan dan tetap memberikan hasil yang lebih baik pada kelas-kelas minoritas. Dalam aplikasi dunia nyata, di mana penting untuk memastikan bahwa model dapat mengklasifikasikan semua kelas dengan adil, Model 2 menawarkan hasil yang lebih dapat diandalkan meskipun akurasinya lebih rendah.

Sebaliknya, Model 1, meskipun unggul dalam hal akurasi total (95%), menunjukkan bias yang lebih besar terhadap kelas dengan jumlah sampel yang lebih banyak, yang dapat menyebabkan overfitting. Model ini "terlalu belajar" untuk memprediksi kelas dominan dengan sangat baik namun mengabaikan kelas minoritas yang lebih kecil. Hal ini menjadi masalah ketika aplikasi memerlukan performa yang konsisten di seluruh kelas, terlepas dari ketidakseimbangan sampel. Model 2, yang lebih seimbang, tidak menghadapi masalah ini dan dapat memberikan hasil yang lebih adil dan konsisten di seluruh kelas. Dengan demikian, meskipun Model 1 lebih unggul dalam hal akurasi total, Model 2 lebih disarankan untuk aplikasi yang memerlukan keseimbangan yang lebih baik dalam klasifikasi seluruh kelas. Model 2 lebih cocok digunakan ketika tujuan utama adalah menghindari bias terhadap kelas minoritas dan memastikan bahwa model memberikan hasil yang lebih adil dan konsisten, tanpa kecenderungan untuk mengabaikan kelas-kelas yang kurang terwakili.

Selain itu, tingginya akurasi yang dicapai oleh Model 1 juga mengindikasikan adanya

keterbatasan variasi morfologi tangan dan fitur geometris antar sampel dalam setiap kelas. Kondisi ini menunjukkan bahwa dataset yang digunakan, khususnya yang bersumber dari *Fingers Kaggle Dataset*, cenderung merepresentasikan data citra secara umum dalam lingkungan yang relatif homogen, baik dari segi bentuk tangan, orientasi pose, maupun latar belakang subjek. Dataset tersebut belum secara eksplisit melibatkan keberagaman budaya dalam praktik *finger counting*, sehingga variasi representasi numerik yang dipengaruhi oleh kebiasaan kultural tidak tertangkap secara memadai. Akibatnya, model dapat mencapai akurasi yang sangat tinggi karena pola geometris yang dipelajari bersifat seragam dan mudah dikenali, namun memiliki keterbatasan dalam kemampuan generalisasi ketika dihadapkan pada variasi gestur angka yang berbeda secara budaya. Hal ini semakin menegaskan bahwa akurasi yang tinggi tidak selalu mencerminkan kemampuan generalisasi yang baik, terutama pada sistem HGR yang ditujukan untuk aplikasi global dan lintas budaya.



Gambar 9. Hasil *confusion matrix* setelah pelatihan model, *Natural Hand Digit Dataset* (kiri) dan *Fingers Dataset Kaggle* (kanan)

Hasil confusion matrix pada Gambar 9 menunjukkan perbedaan karakteristik yang jelas antara dataset yang diajukan (kiri) dan *Fingers Kaggle Dataset* (kanan). Pada dataset yang diajukan, terlihat masih terjadi kesalahan klasifikasi yang cukup signifikan, khususnya pada kelas 3, 4, dan 5, yang saling tertukar satu sama lain. Hal ini menunjukkan bahwa variasi morfologi tangan, orientasi jari, serta perbedaan representasi gestur akibat latar belakang budaya yang beragam menyebabkan batas antar kelas menjadi lebih kompleks, sehingga menuntut model untuk melakukan generalisasi yang lebih kuat. Sebaliknya, pada *Fingers Kaggle Dataset*, hampir seluruh prediksi terpusat pada diagonal utama, menandakan tingkat klasifikasi yang sangat tinggi dan minim kesalahan antarkelas. Namun, performa yang sangat tinggi ini berkorelasi dengan rendahnya variasi fitur geometris dan homogenitas pose tangan pada setiap kelas, yang mengindikasikan bahwa dataset tersebut tidak merepresentasikan keberagaman budaya dalam praktik *finger counting*. Dengan demikian, meskipun *Fingers Kaggle Dataset* menghasilkan akurasi yang lebih tinggi, dataset yang diajukan justru lebih menantang secara ilmiah dan lebih relevan untuk pengembangan sistem HGR yang inklusif, adaptif, serta mampu menghadapi variasi gestur nyata di lingkungan multikultural, khususnya dalam konteks media pembelajaran matematika interaktif.

4. Kesimpulan

Penelitian ini menunjukkan bahwa dataset finger counting yang diajukan memiliki karakteristik yang jauh lebih kompleks dibandingkan dengan Fingers Kaggle Dataset. Berdasarkan analisis *confusion matrix*, terlihat bahwa pada dataset yang diajukan masih terjadi kesalahan klasifikasi yang cukup signifikan, khususnya pada kelas tengah (3, 4, dan 5). Hal ini disebabkan oleh tingginya variasi pose tangan, morfologi jari, sudut pengambilan gambar, serta latar belakang budaya pengguna yang beragam. Temuan ini mengindikasikan bahwa dataset yang diajukan lebih merepresentasikan kondisi nyata (*real-world scenario*) dan menuntut kemampuan generalisasi model yang lebih kuat.

Sebaliknya, *Fingers Kaggle Dataset* menunjukkan performa klasifikasi yang sangat tinggi dengan tingkat kesalahan yang minimal. Namun, tingginya akurasi tersebut berkorelasi dengan rendahnya variasi gestur dan homogenitas bentuk pose tangan pada setiap kelas, sehingga kurang merepresentasikan keberagaman praktik finger counting secara global. Dengan demikian, meskipun dataset standar menghasilkan performa yang lebih baik secara kuantitatif, dataset yang diajukan dalam penelitian ini memiliki kontribusi yang lebih signifikan secara ilmiah karena menawarkan tantangan yang lebih realistis dan relevan untuk pengembangan sistem *Hand Gesture Recognition* (HGR) yang adaptif, robust, dan inklusif, khususnya untuk aplikasi pembelajaran matematika interaktif berbasis budaya.

Sebagai arah penelitian selanjutnya (*future work*), diperlukan pengembangan teknik penanganan ketidakseimbangan data seperti *data augmentation*, *resampling*, dan optimasi arsitektur model untuk meningkatkan generalisasi pada seluruh kelas. Selain itu, penelitian lanjutan juga perlu menginvestigasi integrasi sistem HGR ini secara langsung dalam lingkungan pembelajaran nyata untuk menguji efektivitas, aspek *usability*, serta keamanan kultural dalam skala pengguna yang lebih luas.

Ucapan Terima Kasih

Kami menyampaikan apresiasi dan ucapan terima kasih yang tulus kepada individu-individu yang kontribusinya sangat vital dalam fase akuisisi data penelitian ini. Apresiasi khusus ditujukan kepada para mahasiswa Program Studi Sistem Informasi Unhas angkatan 2022 yang telah membantu secara langsung dalam proses pengambilan data untuk pembangunan *Natural Hand Digit Dataset* dengan mengikuti protokol yang ketat. Terima kasih sebesar-besarnya kami sampaikan kepada: Nurul Alya, Rezqia Nurqalbi, dan Andi Muthia Mulia Putri. Dukungan dan dedikasi Anda sangat menentukan kualitas dan *robustness* data yang menjadi fondasi utama penelitian ini.

Referensi

- [1] R. Aigner *et al.*, "Understanding mid-air hand gestures: A study of human preferences in usage of gesture types for HCI," *Microsoft Research Tech. Rep.*, no. MSR-TR-2012-111, pp. 1-30, 2012.
- [2] Kapitanov, K. Kvanchiani, A. Nagaev, R. Kraynov, and A. Makhliarchuk, "HaGRID—Hand Gesture Recognition Image Dataset," in *Proc. IEEE/CVF Winter Conf. Appl. Comput. Vis.*, 2024, pp. 4572-4581.

- [3] Yerukola, S. Gabriel, N. Peng, and M. Sap, "Mind the Gesture: Evaluating AI Sensitivity to Culturally Offensive Non-Verbal Gestures," *arXiv preprint arXiv:2502.17710*, 2025.
- [4] N. P. A. Vithanage, R. L. A. S. Ayesha, and L. B. H. M. Vijayarathna, "Number Gestures 1-5: Hand Landmark Dataset," Kaggle, Dataset, 2024. doi: 10.34740/KAGGLE/DSV/8941407.
- [5] Hashi, S. Z. M. Hashim, and A. B. Asamah, "A systematic review of hand gesture recognition: An update from 2018 to 2024," *IEEE Access*, vol. 12, pp. 143599–143626, 2024.
- [6] K. R. Morrissey, M. Liu, J. Kang, D. Hallett, and Q. Wang, "Cross-cultural and intra-cultural differences in finger-counting habits and number magnitude processing: Embodied numerosity in Canadian and Chinese university students," *J. Numer. Cogn.*, vol. 2, no. 1, pp. 1–19, 2016.
- [7] Yaseen, O. J. Kwon, J. Kim, J. Lee, and F. Ullah, "Evaluation of benchmark datasets and deep learning models with pre-trained weights for vision-based dynamic hand gesture recognition," *Appl. Sci.*, vol. 15, no. 11, p. 6045, 2025.
- [8] G. Benitez-Garcia, J. Olivares-Mercado, G. Sanchez-Perez, and H. Takahashi, "IPN HandS: Efficient annotation tool and dataset for skeleton-based hand gesture recognition," *Appl. Sci.*, vol. 15, no. 11, p. 6321, 2025.
- [9] F. Zhang, V. Bazarevsky, A. Tkachenka, G. Sung, C. L. Chang, and M. Grundmann, "MediaPipe Hands: On-device real-time hand tracking," *arXiv preprint arXiv:2006.10214*, 2020.
- [10] "Finger-counting across cultures," *Multilingual Families*, [Online]. Available: <https://multilingual-families.com/finger-counting-across-cultures/>. [Accessed: Nov. 30, 2025].
- [11] K. A. Ahmad, D. C. Silpani, and K. Yoshida, "The impact of large sample datasets on hand gesture recognition by hand landmark classification," *Int. J. Affective Eng.*, vol. 22, no. 3, pp. 253–259, 2023.
- [12] K. A. Ahmad, T. Higashi, and K. Yoshida, "Dynamic hand gesture recognition by hand landmark classification using long short-term memory," *Pertanika J. Sci. Technol.*, vol. 73, 2025.
- [13] K. A. Ahmad, D. C. Silpani, and K. Yoshida, "Hand gesture recognition by hand landmark classification," in *Proc. Int. Symp. Affective Sci. Eng. (ISASE 2022)*, Japan Society of Kansei Engineering, 2022, pp. 1–4.
- [14] A. Bender and S. Beller, "Nature and culture of finger counting: Diversity and representational effects of an embodied cognitive tool," *Cognition*, vol. 124, no. 2, pp. 156–182, 2012.
- [15] P. Koryakin, "Fingers: Classify images of 0,1,2,3,4 or 5 fingers," Kaggle, 2018. [Online]. Available: <https://www.kaggle.com/datasets/koryakinp/fingers>. [Accessed: Dec. 6, 2025].