

Klasifikasi Frame Bahasa Isyarat Huruf Sibi Menggunakan Fitur Landmark dan Long Short-Term Memory (LSTM)

¹Ilham Maulana Mahmud

¹Program Studi Informatika, Fakultas Sains dan Informatika, Universitas Jenderal Achmad Yani

¹ilham.m.m324@gmail.com

Abstract - Sign language is the primary means of communication for people with hearing impairments. In Indonesia, one of the official systems used is the Indonesian Sign Language System (SIBI). This study aims to evaluate a SIBI letter classification approach based on hand landmark frame sequences obtained from MediaPipe Hands. The extracted sequential data is used as input for a Long Short-Term Memory (LSTM) model to recognize temporal patterns of hand movements. The research stages include data acquisition, preprocessing, feature extraction, training, and evaluation. The test results show that the LSTM model is able to classify SIBI letters with a validation accuracy of more than 95%. Evaluation using precision, recall, and F1-score shows stable performance. This research has the potential to become the basis for developing responsive, efficient, and inclusive sign language recognition applications.

Keywords - SIBI, Sign Language, LSTM, MediaPipe, Gesture Classification.

Abstrak - Bahasa isyarat merupakan sarana komunikasi utama bagi penyandang tunarungu. Di Indonesia, salah satu sistem resmi yang digunakan adalah Sistem Isyarat Bahasa Indonesia (SIBI). Penelitian ini bertujuan mengevaluasi pendekatan klasifikasi huruf SIBI berbasis urutan frame landmark tangan yang diperoleh dari MediaPipe Hands. Data sekuensial hasil ekstraksi digunakan sebagai input model Long Short-Term Memory (LSTM) untuk mengenali pola temporal dari pergerakan tangan. Tahapan penelitian meliputi akuisisi data, preprocessing, ekstraksi fitur, pelatihan, dan evaluasi. Hasil pengujian menunjukkan bahwa model LSTM mampu mengklasifikasikan huruf SIBI dengan akurasi validasi lebih dari 95%. Evaluasi menggunakan precision, recall, dan F1-score menunjukkan kinerja yang stabil. Penelitian ini berpotensi menjadi dasar pengembangan aplikasi pengenalan bahasa isyarat yang responsif, efisien, dan inklusif.

Kata Kunci - SIBI, Bahasa Isyarat, LSTM, MediaPipe, Klasifikasi Gestur.

I. Pendahuluan

Komunikasi merupakan aspek fundamental dalam interaksi sosial manusia. Bagi penyandang tunarungu, hambatan komunikasi dapat membatasi akses terhadap pendidikan, pekerjaan, dan interaksi sosial. Pemerintah Indonesia melalui Keputusan Mendikbud No. 0161/U/1994 telah membakukan Sistem Isyarat Bahasa Indonesia (SIBI) sebagai standar komunikasi resmi di sekolah luar biasa (SLB) [1]. SIBI berfungsi sebagai media komunikasi formal yang mengikuti struktur bahasa Indonesia baku, berbeda dengan BISINDO yang lebih bersifat alami dan digunakan di komunitas tuli sehari-hari.

Seiring perkembangan teknologi, penelitian pengenalan otomatis bahasa isyarat telah banyak dilakukan pada American Sign Language (ASL), dengan dominasi pendekatan berbasis *computer vision* dan *deep learning*. Metode berbasis Convolutional Neural Network (CNN) cukup efektif dalam mengenali gesture statis, namun memiliki keterbatasan dalam menangani gesture dinamis yang membutuhkan analisis temporal, seperti huruf 'J' dan 'Z' pada SIBI [2]. Penelitian terbaru menunjukkan bahwa model sekuensial seperti Long Short-Term Memory (LSTM) lebih unggul dalam mengenali pola urutan karena memiliki mekanisme memori jangka panjang yang mampu mengatasi masalah vanishing gradient pada Recurrent Neural Network (RNN) [3], [4], [5], [6].

Selain itu, penggunaan fitur *landmark* tangan dari *Media Pipe Hands* menjadi pendekatan yang lebih efisien dibandingkan dengan pemrosesan video penuh, karena hanya memanfaatkan koordinat spasial (x, y, z) dari 21 titik kunci tangan. Landmark ini memungkinkan representasi gesture secara numerik yang ringan namun tetap akurat [7], [8]. Beberapa studi menunjukkan bahwa kombinasi landmark tangan dan model LSTM dapat meningkatkan akurasi klasifikasi gesture hingga lebih dari 90% [9], [6].

Namun demikian, penelitian terhadap SIBI masih sangat terbatas jika dibandingkan dengan ASL maupun BISINDO. Kebanyakan penelitian terdahulu hanya berfokus pada citra statis atau menggunakan video penuh dengan kebutuhan komputasi tinggi [10]. Gap penelitian ini membuka peluang untuk mengeksplorasi penggunaan sekuens landmark tangan dari frame JPEG sebagai input LSTM, yang lebih hemat sumber daya dan berpotensi diimplementasikan pada perangkat mobile maupun sistem pembelajaran inklusif di SLB.

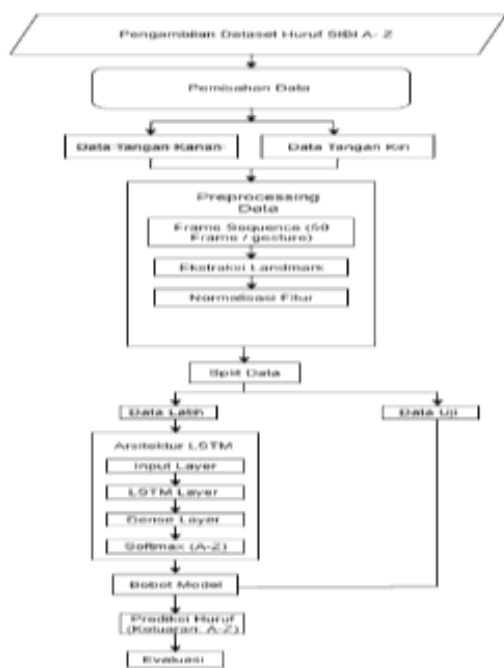
Dengan demikian, penelitian ini tidak hanya berkontribusi pada ranah akademik dengan memperkenalkan pendekatan baru dalam klasifikasi huruf SIBI, tetapi juga memberikan manfaat

praktis dalam mendukung teknologi asistif lokal yang inklusif, efisien, dan kontekstual sesuai kebutuhan masyarakat Indonesia [11], [12], [13].

II. Metode Penelitian

Tahapan penelitian ini dilakukan secara sistematis untuk memastikan akurasi dan reliabilitas hasil klasifikasi gesture huruf SIBI. Pertama, dilakukan akuisisi data dengan merekam video gesture huruf A hingga Z, yang kemudian diekstraksi menjadi 50 *frame* per gesture. Setiap frame diproses menggunakan MediaPipe Hands untuk mengekstraksi 21 titik landmark tangan dalam koordinat tiga dimensi (x, y, z). Selanjutnya, data landmark yang diperoleh dinormalisasi menggunakan metode Z-score normalization, sehingga nilai koordinat berada dalam skala seragam. Setiap gesture kemudian direpresentasikan dalam bentuk sekuens berukuran 50 frame $\times$ 63 fitur, yang mencerminkan pola pergerakan tangan secara temporal. Dataset yang dihasilkan dibagi menjadi data latih (70%) dan data uji (30%) dengan metode stratified split untuk menjaga distribusi kelas tetap seimbang.

Proses pemodelan dilakukan menggunakan arsitektur Long Short-Term Memory (LSTM) dengan susunan Bidirectional LSTM (128 unit), LSTM (64 unit), lapisan dropout untuk mencegah overfitting, dense layer, dan softmax sebagai fungsi aktivasi output untuk mengklasifikasikan 52 kelas (26 huruf $\times$ 2 sisi tangan). Evaluasi kinerja model dilakukan dengan menghitung metrik akurasi, precision, recall, F1-score, serta analisis confusion matrix untuk mengetahui tingkat kesalahan klasifikasi antar huruf.



Gambar 1. Langkah Penelitian

A. Akuisisi Data

Video gesture huruf A–Z direkam, Dataset yang digunakan merupakan kumpulan video gesture tangan yang merepresentasikan 26 huruf dalam Bahasa Isyarat Indonesia (SIBI). Setiap 50 Frame menampilkan satu gesture statis atau dinamis untuk masing-masing huruf dari A hingga Z. Gesture dilakukan menggunakan tangan kanan maupun tangan kiri, tergantung pada skenario pengambilan data. Dataset ini menjadi dasar utama untuk pelatihan dan pengujian model klasifikasi [14],[15].

Preprocessing Ekstraksi Frame



Gambar 2. Proses Ekstraksi

Pada tahap ini, data gambar yang merekam gesture huruf dikonversi menjadi rangkaian frame citra statis yang merepresentasikan pergerakan tangan secara sekuensial. Proses ini dikenal sebagai ekstraksi frame, Ekstraksi ini dilakukan agar informasi temporal dari setiap gesture dapat dipelajari lebih lanjut oleh model sekuensial seperti Long Short-Term Memory (LSTM).

Untuk menjaga konsistensi dan efisiensi proses pelatihan model, jumlah maksimal frame yang diambil dari setiap gesture dibatasi sebanyak 50 frame. Jika jumlah frame dalam gambar gesture melebihi 50, maka hanya 50 frame pertama yang dipilih. Sebaliknya, jika jumlah frame kurang dari 50, maka akan dilakukan padding agar panjang sekuens tetap seragam. Pemilihan panjang sekuens ini didasarkan pada eksperimen awal yang menunjukkan bahwa 50 frame sudah cukup untuk merepresentasikan satu gesture secara utuh tanpa kehilangan informasi penting.

Setelah frame berhasil diekstrak, setiap frame dianalisis menggunakan pustaka MediaPipe, yang bertugas mengekstraksi 21 titik koordinat landmark tangan (*keypoints*) dari setiap gambar. Informasi landmark ini mencakup koordinat x,y,zx, y, zx,y,z dari masing-masing titik pada tangan, seperti ujung jari, pangkal jari, dan pergelangan. Data koordinat tersebut tidak disimpan dalam bentuk citra RGB, melainkan dikonversi menjadi struktur data numerik berbentuk JSON, sehingga menghasilkan sequence of keypoints yang efisien dan mudah diproses secara komputasional.

Dengan mengubah data gesture dari bentuk video menjadi sequence of keypoints yang merepresentasikan pergerakan secara spasial dan temporal, sistem ini mampu menangkap dinamika gesture secara menyeluruh. Representasi berbasis frame dan koordinat landmark ini sangat penting untuk model

LSTM, karena model tersebut dirancang untuk mengenali pola dalam data urutan dan menangkap hubungan antar frame yang merepresentasikan gerakan tangan secara bertahap.

Secara keseluruhan, proses preprocessing ini bertujuan untuk menyederhanakan, menstandarkan, dan memformat ulang data agar sesuai dengan kebutuhan input model deep learning, khususnya LSTM, dalam melakukan klasifikasi gesture bahasa isyarat huruf.

B. Skema Eksperimen Data

Setelah ekstraksi frame dilakukan, data kemudian dipisahkan berdasarkan jenis tangan yang digunakan dalam gesture, yaitu:

1. Data Tangan Kanan
2. Data Tangan Kiri

Pemisahan ini dilakukan agar sistem dapat menganalisis pengaruh jenis tangan terhadap performa model, serta untuk melakukan eksperimen klasifikasi secara independen maupun gabungan.

C. Preprocessing Data

Frame Sequence

Setiap gesture disusun dalam bentuk urutan 50 frame secara konsisten agar dapat diolah sebagai data sekuensial. Setiap urutan frame mewakili satu gesture huruf dari bahasa isyarat, dan menjadi input utama dalam proses ekstraksi fitur serta pelatihan model Long Short-Term Memory (LSTM). Dengan pendekatan ini, model mampu mempelajari dinamika pergerakan gesture secara temporal.

Tabel 1. Frame Sequence

Huruf	Jumlah Frame
A Kanan	50 Frame
A Kiri	50 Frame
B Kanan	50 Frame
B kiri	50 Frame
....	
Z Kanan	50 Frame
Z Kiri	50 Frame

Ekstraksi Landmark

Setiap frame dianalisis menggunakan pustaka MediaPipe, yang mengekstraksi 21 titik landmark tangan dalam bentuk koordinat tiga dimensi (x, y, z). Landmark ini mencerminkan posisi spasial dari tulang-tulang tangan dan jari, yang sangat penting untuk merepresentasikan gesture secara akurat.

Model deteksi tangan dari MediaPipe telah dilatih pada berbagai kondisi, termasuk variasi posisi tangan, sudut pandang, latar belakang, dan pencahayaan. Hal ini memungkinkan MediaPipe untuk mengenali bentuk tangan secara akurat dan konsisten, bahkan dalam lingkungan yang kompleks. Secara total, terdapat $21 \text{ landmark} \times 3 \text{ koordinat} = 63$ fitur numerik per frame. Landmark tersebut mencakup.

Tabel 2. Ekstraksi Landmark

Landmark	Posisi
0	Pergelangan tangan (wrist)
1–4	Ibu jari (thumb): MCP, MCP, IP, Tip
5–8	Jari telunjuk (index): MCP, PIP, DIP, Tip
9–12	Jari tengah (middle): MCP, PIP, DIP, Tip
13–16	Jari manis (ring): MCP, PIP, DIP, Tip
17–20	Jari kelingking (pinky): MCP, PIP, DIP, Tip

Struktur landmark ini dirancang sedemikian rupa agar mampu menangkap konfigurasi tangan yang kompleks, seperti jari yang terlipat, membuka, menyilang, atau membentuk pola tertentu yang khas pada gesture alfabet dalam bahasa isyarat.

Media Pipe Hands menggunakan model berbasis deep learning yang telah dilatih secara ekstensif pada dataset besar berisi berbagai variasi tangan: warna kulit, ukuran tangan, sudut kamera, pencahayaan, serta latar belakang yang beragam. Hal ini membuat proses ekstraksi landmark menjadi akurat dan stabil, bahkan pada kondisi lingkungan yang tidak ideal.

Data landmark yang diperoleh kemudian disusun dalam urutan frame dan disimpan dalam format .json atau .npy untuk digunakan sebagai input ke model klasifikasi LSTM. Setiap urutan gesture dengan 50 frame akan menghasilkan $50 \times 63 = 3.150$ data numerik, yang mewakili dinamika spasial dan temporal gesture tersebut.

Pendekatan ini tidak hanya efisien secara komputasional, tetapi juga memungkinkan pemodelan yang presisi terhadap gesture berbasis waktu, karena memperhitungkan posisi, gerakan, dan kedalaman setiap bagian tangan dalam tiap frame.

D. Normalisasi Fitur

Setelah tahap ekstraksi landmark dari setiap frame selesai dilakukan, langkah selanjutnya adalah normalisasi fitur, yaitu proses transformasi nilai-nilai koordinat (x,y,z)(x, y, z)(x,y,z) agar berada dalam skala yang seragam. Normalisasi merupakan bagian penting dalam preprocessing data, khususnya ketika data akan digunakan dalam pelatihan model berbasis deep learning seperti Long Short-Term Memory (LSTM).

Setelah normalisasi, setiap frame menghasilkan vektor fitur sepanjang 63 nilai (21 titik landmark $\times$ 3 koordinat), yang semuanya berada pada skala yang seimbang. Vektor-vektor ini kemudian disusun dalam urutan waktu (sequence) untuk setiap gesture, yang akan digunakan sebagai input bagi model LSTM. Dengan demikian, model dapat mempelajari pola pergerakan tangan tanpa terganggu oleh perbedaan skala numerik antar fitur.

E. Split Data

Data hasil preprocessing dibagi menjadi dua subset, yaitu:

1. Data Latih (Training Set): Digunakan untuk melatih model.
2. Data Uji (Testing Set): Digunakan untuk mengevaluasi performa model.

Pembagian ini dilakukan dengan rasio umum 70:30, bergantung pada konfigurasi eksperimen. Pemisahan dilakukan secara acak namun tetap menjaga proporsi antar kelas agar seimbang.

F. Pelatihan LSTM

Arsitektur terdiri dari Bidirectional LSTM (128 unit), LSTM (64 unit), dropout, dense layer, dan softmax untuk klasifikasi 52 kelas (26 huruf × 2 sisi tangan).

G. Evaluasi

Performa model dikaji menggunakan berbagai metrik evaluasi untuk mengetahui sejauh mana akurasi dan ketepatan klasifikasi gesture huruf SIBI. Metrik yang digunakan antara lain:

1. Akurasi (*Accuracy*): Persentase prediksi yang benar terhadap total data uji.
2. *Precision*: Kemampuan model dalam memprediksi kelas atau mengukur proporsi prediksi benar dari seluruh prediksi positif yang dihasilkan oleh model. Metrik ini berguna untuk melihat sejauh mana model menghasilkan prediksi yang relevan.

$$Precision = \frac{TP}{(TP+FP)} \quad (1)$$

Keterangan :

TP (True Positive): Jumlah prediksi benar pada kelas positif, yaitu ketika model memprediksi benar suatu gesture sesuai dengan label aslinya.

FP (False Positive): Jumlah prediksi salah ketika model memprediksi gesture sebagai kelas tertentu padahal seharusnya bukan.

$$Recall = \frac{TP}{(TP+FN)} \quad (2)$$

Keterangan :

Recall: Kemampuan model dalam menemukan seluruh contoh relevan dari suatu kelas. Mengukur proporsi prediksi benar dari seluruh data aktual yang seharusnya diklasifikasikan sebagai kelas tertentu. Recall penting dalam melihat kemampuan model mendeteksi gesture tertentu secara konsisten.

FN (False Negative): Jumlah prediksi salah ketika model gagal memprediksi gesture sebagai kelas tertentu padahal seharusnya iya.

$$F1 = 2 \frac{(Precision \times Recall)}{(Precision+Recall)} \quad (3)$$

Keterangan:

F1-Score: Harmonik rata-rata dari precision dan recall, menggambarkan keseimbangan performa. Metrik ini berguna ketika diperlukan keseimbangan antara precision dan recall, terutama pada dataset yang tidak seimbang.

Precision: Kemampuan model dalam memprediksi kelas atau mengukur proporsi prediksi benar dari seluruh prediksi positif yang dihasilkan oleh model. Metrik ini berguna untuk melihat sejauh mana model menghasilkan prediksi yang relevan.

Recall: Kemampuan model dalam menemukan seluruh contoh relevan dari suatu kelas. Mengukur proporsi prediksi benar dari seluruh data aktual yang seharusnya diklasifikasikan sebagai kelas tertentu. Recall penting dalam melihat kemampuan model mendeteksi gesture tertentu secara konsisten.

III. Hasil dan Pembahasan

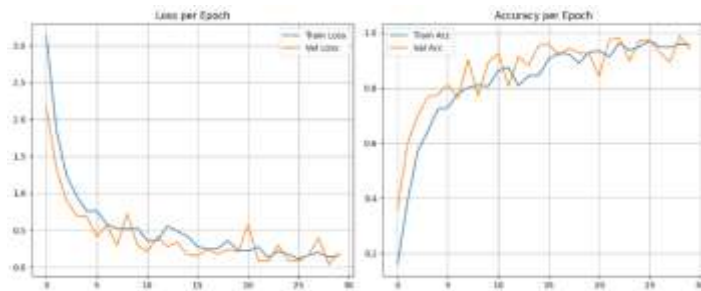
A. Hasil

Dataset yang digunakan dalam penelitian ini merupakan kumpulan video gesture huruf alfabet Sistem Isyarat Bahasa Indonesia (SIBI) untuk huruf A–Z, direkam untuk kedua sisi tangan (kanan dan kiri) sehingga total kelas yang diklasifikasikan adalah 52 kelas (26 huruf × 2 sisi). Setelah proses ekstraksi frame, setiap gesture direpresentasikan sebagai urutan tetap sebanyak 50 frame per sampel, dan setiap frame menghasilkan 63 fitur numerik (21 landmark × 3 koordinat: x, y, z) menggunakan pustaka MediaPipe Hands. Secara keseluruhan dataset akhir yang digunakan untuk pelatihan dan pengujian berjumlah 1.108 sampel (support total pada evaluasi: 1108), dengan pembagian data latih dan uji menggunakan stratified split sebesar 70% untuk pelatihan dan 30% untuk pengujian/validasi, sebagaimana dijelaskan pada bagian metode.

Dataset memiliki distribusi kelas yang relatif seimbang dengan rata-rata ≈ 21 sampel per kelas (1.108 / 52 ≈ 21.3). Namun, analisis metrik per kelas memperlihatkan variasi performa pada beberapa kelas (mis. O_Kanan, D_Kanan, R_Kiri), yang mengindikasikan adanya perbedaan representasi atau tantangan diskriminatif pada kelas-kelas tersebut. Oleh karena itu, perlu dicatat bahwa walaupun jumlah rata-rata per kelas cukup, variasi internal antar sampel (mis. subjek, sudut, pencahayaan) memengaruhi kemampuan model membedakan kelas tertentu.

Tabel 3. Ringkasan dataset

Item	Nilai / Keterangan
Jumlah kelas	52 (26 huruf × 2 sisi tangan)
Sampel total (support)	1.108
Rata-rata sampel per kelas	≈ 21.3
Frame per sampel	50 frame
Fitur per frame	63 (21 landmark × 3 koordinat)
Ekstraksi landmark	MediaPipe Hands
Normalisasi fitur	Z-score
Pembagian data	Training 70% / Testing 30% (stratified split)
Perangkat perekam	(tidak dilaporkan — tambahkan merk/model & resolusi)
Jumlah subjek	(tidak dilaporkan — tambahkan N jika tersedia)
Variasi kondisi	(tidak dilaporkan rinci — rekomendasi: latar, pencahayaan, sudut, jarak)

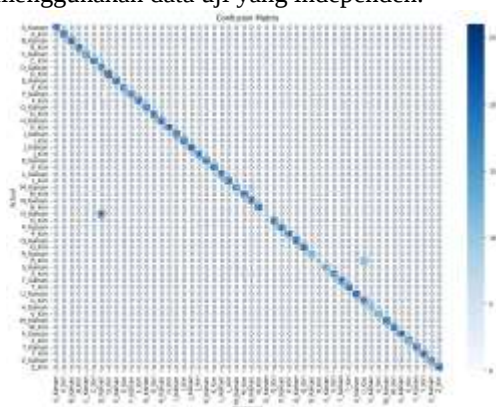


Gambar 3. Grafik Loss per Epoch dan Accuracy per Epoch

Grafik hasil pelatihan model menunjukkan bahwa nilai loss pada data latih maupun data validasi mengalami penurunan yang signifikan sejak awal *epoch* hingga mendekati *epoch* ke-30. Penurunan ini menandakan bahwa model mampu melakukan proses pembelajaran dengan baik, di mana perbedaan antara train loss dan validation loss relatif kecil sehingga tidak menunjukkan adanya indikasi *overfitting* yang serius. Meskipun terdapat sedikit fluktuasi pada validation loss, pola penurunan tetap konsisten sehingga menegaskan stabilitas proses pelatihan.

Pada grafik akurasi terlihat peningkatan yang cukup tajam pada epoch awal, di mana akurasi model meningkat dari sekitar 20% menjadi lebih dari 80% hanya dalam beberapa *epoch* pertama. Selanjutnya, baik akurasi data latih maupun data validasi terus mengalami peningkatan hingga mendekati 100% pada *epoch* ke-20 hingga ke-30. Kondisi ini menunjukkan bahwa model tidak hanya mampu belajar dengan baik dari data latih, tetapi juga mampu melakukan generalisasi yang cukup baik pada data validasi.

Secara keseluruhan, grafik ini memperlihatkan bahwa model yang digunakan memiliki performa yang optimal. Proses pembelajaran berlangsung efektif dengan loss yang menurun stabil serta akurasi yang meningkat hingga mencapai tingkat yang sangat tinggi, sehingga dapat disimpulkan bahwa model ini sangat potensial untuk diterapkan dalam pengujian lebih lanjut menggunakan data uji yang independen.



Gambar 4. Confusion matrix

Confusion matrix pada gambar menunjukkan hasil performa klasifikasi model terhadap data uji dengan jumlah kelas yang

cukup banyak (A hingga Z). Secara umum, distribusi prediksi model terlihat dominan pada diagonal utama, yang berarti sebagian besar data uji berhasil diprediksi dengan benar oleh model. Intensitas warna biru pada diagonal utama cukup kuat dan konsisten dari kelas pertama hingga kelas terakhir, sehingga menunjukkan tingkat akurasi yang tinggi pada hampir semua kelas.

Meskipun demikian, masih terdapat beberapa kesalahan klasifikasi yang tampak dari adanya nilai di luar diagonal utama. Kesalahan ini relatif kecil dan hanya terjadi pada sebagian kecil kelas, sehingga tidak memberikan pengaruh signifikan terhadap performa keseluruhan model. Dengan kata lain, error yang terjadi lebih bersifat sporadis daripada sistematis.

Secara keseluruhan, *confusion matrix* ini memperlihatkan bahwa model memiliki kemampuan klasifikasi yang sangat baik dalam mengenali pola dari masing-masing kelas. Dominasi prediksi yang benar pada diagonal utama menjadi indikasi bahwa model mampu melakukan generalisasi dengan baik, serta memiliki potensi tinggi untuk digunakan pada aplikasi nyata dengan tingkat keandalan yang memadai.

Tabel 4. tabel metrik evaluasi (precision, recall, f1-score, dan support)

Kelas	Precision	Recall	F1-Score	Support
A_Kanan	1.00	1.00	1.00	21
A_Kiri	1.00	1.00	1.00	21
B_Kanan	1.00	1.00	1.00	22
B_Kiri	1.00	1.00	1.00	21
C_Kanan	1.00	0.81	0.89	21
C_Kiri	1.00	1.00	1.00	21
D_Kanan	0.45	0.95	0.62	21
D_Kiri	1.00	1.00	1.00	26
E_Kanan	1.00	1.00	1.00	21
E_Kiri	1.00	0.90	0.95	21
F_Kanan	1.00	0.95	0.98	22
F_Kiri	1.00	1.00	1.00	21
G_Kanan	1.00	1.00	1.00	20
G_Kiri	1.00	1.00	1.00	22
H_Kanan	1.00	1.00	1.00	22
H_Kiri	0.96	1.00	0.98	22
I_Kanan	1.00	1.00	1.00	21
I_Kiri	1.00	1.00	1.00	21
J_Kanan	1.00	1.00	1.00	23
J_Kiri	1.00	1.00	1.00	21
K_Kanan	1.00	1.00	1.00	21
K_Kiri	1.00	1.00	1.00	21
L_Kanan	1.00	1.00	1.00	21
L_Kiri	1.00	1.00	1.00	21
M_Kanan	0.86	0.78	0.82	23
M_Kiri	0.85	1.00	0.92	22
N_Kanan	0.76	1.00	0.86	22
N_Kiri	1.00	1.00	1.00	21
O_Kanan	0.40	0.09	0.14	23
O_Kiri	1.00	1.00	1.00	22

Kelas	Precision	Recall	F1-Score	Support
P_Kanan	1.00	1.00	1.00	21
P_Kiri	1.00	1.00	1.00	21
Q_Kanan	1.00	1.00	1.00	23
Q_Kiri	1.00	1.00	1.00	22
R_Kanan	1.00	0.83	0.91	18
R_Kiri	1.00	0.22	0.36	18
S_Kanan	1.00	0.80	0.89	20
S_Kiri	1.00	0.90	0.95	21
T_Kanan	1.00	0.95	0.98	22
T_Kiri	1.00	1.00	1.00	22
U_Kanan	0.88	1.00	0.94	23
U_Kiri	0.61	1.00	0.76	22
V_Kanan	1.00	1.00	1.00	14
V_Kiri	1.00	1.00	1.00	15
W_Kanan	1.00	1.00	1.00	22
W_Kiri	1.00	1.00	1.00	22
X_Kanan	0.88	1.00	0.94	22
X_Kiri	1.00	0.82	0.90	22
Y_Kanan	1.00	1.00	1.00	21
Y_Kiri	1.00	1.00	1.00	22
Z_Kanan	1.00	0.87	0.93	23
Z_Kiri	0.89	1.00	0.94	24
Akurasi			0.94	1108
Total				
Rata-rata Makro	0.95	0.94	0.94	1108
Rata-rata Tertimbang	0.95	0.94	0.94	1108

B. Pembahasan

Hasil evaluasi kinerja model klasifikasi menunjukkan bahwa secara keseluruhan model mampu mencapai akurasi total sebesar 94% dengan nilai rata-rata makro dan tertimbang untuk precision, recall, dan f1-score berada pada kisaran 0.94–0.95. Hal ini menandakan bahwa model memiliki performa yang sangat baik dalam mengenali pola pada sebagian besar kelas.

Sebagian besar kelas, baik untuk posisi kanan maupun kiri, menunjukkan nilai precision, recall, dan f1-score yang sempurna (1.00), misalnya pada kelas A_Kanan, A_Kiri, B_Kanan, B_Kiri, C_Kiri, D_Kiri, F_Kiri, G_Kiri, H_Kanan, I_Kanan, J_Kanan, K_Kiri, L_Kiri, N_Kiri, O_Kiri, P_Kanan, Q_Kanan, T_Kiri, V_Kanan, W_Kanan, Y_Kanan, Y_Kiri, dan beberapa kelas lainnya. Kondisi ini mengindikasikan bahwa model dapat mengklasifikasikan kelas-kelas tersebut tanpa adanya kesalahan.

Namun demikian, masih terdapat beberapa kelas yang memiliki performa relatif lebih rendah. Sebagai contoh, kelas O_Kanan memiliki precision 0.40, recall 0.09, dan f1-score 0.14, yang menunjukkan kesulitan model dalam mengenali kelas ini. Hal yang sama terlihat pada kelas D_Kanan (f1-score 0.62), M_Kanan (0.82), M_Kiri (0.92), N_Kanan (0.86), R_Kanan (0.91), R_Kiri (0.36), U_Kiri (0.76), X_Kiri (0.90), dan Z_Kanan (0.93). Rendahnya nilai pada kelas-kelas tersebut bisa

disebabkan oleh kesamaan pola dengan kelas lain, distribusi data yang kurang seimbang, atau karakteristik fitur yang sulit dibedakan oleh model.

Meskipun terdapat kelemahan pada beberapa kelas, secara umum model tetap menunjukkan tingkat generalisasi yang tinggi. Nilai rata-rata metrik evaluasi yang konsisten di atas 0.90 menegaskan bahwa model ini sangat layak untuk digunakan pada aplikasi nyata, dengan catatan perlu adanya perbaikan atau augmentasi data pada kelas-kelas yang memiliki performa rendah agar model lebih seimbang.

IV. Kesimpulan

Berdasarkan hasil penelitian, penulis menarik beberapa kesimpulan sebagai berikut:

1. Penelitian ini membuktikan bahwa klasifikasi huruf SIBI dapat dilakukan secara efektif dengan memanfaatkan fitur landmark tangan dan model LSTM. Hasil evaluasi menunjukkan akurasi tinggi dan performa stabil pada berbagai metrik.
2. Pendekatan ini lebih efisien dibanding penggunaan video penuh, sehingga berpotensi diimplementasikan pada aplikasi mobile atau sistem pembelajaran inklusif. Penelitian selanjutnya dapat diarahkan pada klasifikasi kata atau kalimat berbasis SIBI dengan pendekatan Transformer atau multimodal learning.

V. Daftar Pustaka

- [1] Kementerian Pendidikan dan Kebudayaan, *Keputusan Mendikbud No. 0161/U/1994 tentang Pembakuan SIBI*. 1994.
- [2] K. Halim and E. Rakun, "Sign Language System for Bahasa Indonesia (Known as SIBI) Recognizer using TensorFlow and Long Short-Term Memory," in *2018 International Conference on Advanced Computer Science and Information Systems (ICACSIS)*, 2018, pp. 403–407. doi: 10.1109/ICACSIS.2018.8618134.
- [3] V. Yolanda and W. Silfianti, "Prediksi Gerakan secara Real-Time untuk Pengenalan Bahasa Isyarat menggunakan LSTM," *J. Inf. Technol. Comput. Sci. (JOINTECS)*, vol. 8, no. 3, 2023, doi: <https://doi.org/10.31328/jointecs.v8i3.4607>.
- [4] H. M. Putri, F. Fadlisyah, and W. Fuadi, "PENDETEKSIAN BAHASA ISYARAT INDONESIA SECARA REAL-TIME MENGGUNAKAN LONG SHORT-TERM MEMORY (LSTM)," *J. Teknol. Terap. Sains 4.0 (TTS 4.0)*, vol. 3, no. 1, 2022, doi: <https://doi.org/10.29103/tts.v3i1.6853>.
- [5] E. Altirika and W. P. Sari, "Pengembangan Deteksi Realtime untuk Bahasa Isyarat Indonesia dengan Menggunakan Metode Deep Learning Long Short Term Memory dan Convolutional Neural Network," *J. Teknol. Inform. dan Komput. /*, vol. 3, no. 1, 2023, doi:

-
- <https://doi.org/10.37012/jtik.v9i1.1272>.
- [6] M. Sholawati, K. Auliasari, and F. Ariwibisono, "PENGEMBANGAN APLIKASI PENGENALAN BAHASA ISYARAT ABJAD SIBI MENGGUNAKAN METODE CONVOLUTIONAL NEURAL NETWORK (CNN)," *J. Mhs. Tek. Inform.*, vol. 6, no. 1, 2022, doi: <https://doi.org/10.36040/jati.v6i1.4507>.
- [7] M. H. N. Fauzan, E. Rakun, and D. Hardianto, "Feature Extraction from Smartphone Images by Using Elliptical Fourier Descriptor, Centroid and Area for Recognizing Indonesian Sign Language SIBI (Sistem Isyarat Bahasa Indonesia)," in *2019 2nd International Conference on Intelligent Autonomous Systems (ICoIAS)*, 2019, pp. 8–14. doi: 10.1109/ICoIAS.2019.00008.
- [8] O. D. Nurhayati, D. Eridani, and M. H. Tsalavin, "Sistem Isyarat Bahasa Indonesia (SIBI) Metode Convolutional Neural Network Sequential secara Real Time," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 9, no. 4, 2022, doi: <https://doi.org/10.25126/jtiik.2022944787>.
- [9] I. J. Thira, D. Riana, A. N. Ilhami, B. R. S. Dwinanda, and H. Choerunisya, "Pengenalan Alfabet Sistem Isyarat Bahasa Indonesia (SIBI) Menggunakan Convolutional Neural Network," *J. Algoritma*, vol. 20, no. 2, pp. 421–432, 2023, doi: 10.33364/algoritma/v.20-2.1480.
- [10] R. Ridwang, "Pengenalan Bahasa Isyarat Indonesia (SIBI) Menggunakan Leap Motion Controller dan Algoritma Data Mining Naïve Bayes," *J. Insypro (Information Syst. Process.)*, vol. 2, no. 2, 2017, doi: 10.24252/insypro.v2i2.4070.
- [11] S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," *Neural Comput.*, vol. 9, pp. 1735–1780, Nov. 1997, doi: 10.1162/neco.1997.9.8.1735.
- [12] F. Zhang *et al.*, *MediaPipe Hands: On-device Real-time Hand Tracking*. 2020. doi: 10.48550/arXiv.2006.10214.
- [13] A. Graves, *Supervised Sequence Labelling with Recurrent Neural Networks*, vol. 385. 2012. doi: 10.1007/978-3-642-24797-2.
- [14] S. Dwijayanti, Hermawati, S. I. Taqiyyah, H. Hikmarika, and B. Y. Suprpto, "Indonesia Sign Language Recognition using Convolutional Neural Network," *Int. J. Adv. Comput. Sci. Appl.*, vol. 12, no. 10, pp. 415–422, 2021.
- [15] M. F. Naufal and S. F. Kusuma, "Analisis Perbandingan Algoritma Machine Learning dan Deep Learning untuk Klasifikasi Citra Sistem Isyarat Bahasa Indonesia (SIBI)," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 10, no. 4, 2023, doi: <https://doi.org/10.25126/jtiik.2024106823>.