Nabil Hasan Fayyadh El-Rasyid

INSTEK_Revisi1

INSTEK, 2026 - Adellia, Hasan

Document Details

Submission ID
trn:oid:::3618:144969654

Submission Date
Jul 6, 2026, 10:06 AM GMT+7

Download Date
Jul 6, 2026, 10:08 AM GMT+7

File Name
INSTEK_Revisi1.docx

File Size
8.5 MB

14 Pages

5,151 Words

32,931 Characters

7% Overall Similarity

The combined total of all matches, including overlapping sources, for each database.

Filtered from the Report

- Bibliography

Top Sources

3% Internet sources
3% Publications
5% Submitted works (Student Papers)

Integrity Flags

0 Integrity Flags for Review

Our system's algorithms look deeply at a document for any inconsistencies that would set it apart from a normal submission. If we notice something strange, we flag it for you to review.

A Flag is not necessarily an indicator of a problem. However, we'd recommend you focus your attention there for further review.

Top Sources

3% Internet sources
3% Publications
5% Submitted works (Student Papers)

Top Sources

The sources with the highest number of matches within the submission. Overlapping sources will not be displayed.

1 Student papers
Telkom University on 2026-06-18 <1%

2 Student papers
Arab Academy for Science, Technology & Maritime Transport CAIRO on 2024-05-15 <1%

3 Internet
jpti.journals.id <1%

4 Student papers
Universitas Muhammadiyah Makassar on 2026-06-25 <1%

5 Student papers
Universitas Islam Indonesia on 2026-05-12 <1%

6 Internet
e-jurnal.lppmunsera.org <1%

7 Internet
ejournal.itn.ac.id <1%

8 Student papers
Universitas Pancasila on 2026-06-26 <1%

9 Student papers
Universitas Pancasila on 2026-06-30 <1%

10 Internet
ejurnal.seminar-id.com <1%

11 Internet
jurnal-tarbiyah.iainsorong.ac.id <1%

12 Publication
Deepa Javiya, Gordhan Jethava, Himani Parmar, Sweta Jethava. "Enhancing deepf... <1%

13 Internet
diajeng.lldikti6.id <1%

14 Internet
jurnal.pustakagalerimandiri.co.id <1%

15 Student papers
Universitas Raharja on 2024-08-04 <1%

16 Student papers
University of East London on 2026-06-26 <1%

17 Internet
e-journal.hamzanwadi.ac.id <1%

18 Student papers
Telkom University on 2026-06-26 <1%

19 Student papers
Universitas Bengkulu on 2025-10-17 <1%

20 Student papers
Universitas Borneo Tarakan on 2025-06-25 <1%

21 Publication
Wesley Ferreira Maia de Souza. "End-to-End Sign Language Translation Pipeline U... <1%

22 Internet
kc.umn.ac.id <1%

23 Internet
text-id.123dok.com <1%

24 Publication
Aldin Fathiray, Joni Maulindar, Wiji Lestari. "Pengembangan Sistem Penerjemah ... <1%

25 Student papers
Universitas Pamulang on 2026-07-01 <1%

26 Internet

journal.ipm2kpe.or.id <1%

Vol. 10, No. 1, April 2025, Page. 1-250
https://journal.uin-alauddin.ac.id/index.php/instek
p-ISSN: 2541-1179
e-ISSN: 2581-1711
JURNAL INSTEK
INFORMATIKA SAINS DAN TEKNOLOGI

Pengenalan BISINDO Berbasis MediaPipe dengan Transformer Encoder, Long Short-Term Memory, Connectionist Temporal Classification, dan Text-to-Speech

Elvina Adellia [1], Firman Noor Hasan [*2]

[1,2] Teknik Informatika, Universitas Muhammadiyah Prof. Dr. Hamka, Indonesia
Email: [1]elvinadellia01@email.ac.id, [2]firman.noorhasan@uhamka.ac.id

Nomor Handphone : 081311406664

Abstrak

Perbedaan pemahaman antara pengguna Bahasa Isyarat Indonesia (BISINDO) dan masyarakat yang belum menguasai bahasa isyarat masih sering menghambat komunikasi, terutama di lingkungan pendidikan dan layanan publik. Penelitian ini bertujuan mengembangkan prototipe penerjemah BISINDO real-time yang mengubah gerakan isyarat menjadi teks dan suara. Sistem menggunakan MediaPipe Holistic untuk mengekstraksi landmark tangan dan tubuh, sedangkan landmark wajah tidak digunakan agar fitur lebih ringan dan sistem lebih fokus pada gerakan utama. Koordinat landmark digabungkan dengan velocity untuk membaca perubahan gerakan antar-frame, lalu disusun menjadi sequence 30 frame. Model hybrid Transformer Encoder dan Long Short-Term Memory digunakan untuk mengenali 15 kosakata atau frasa dasar BISINDO dari 4.500 sequence. Hasil prediksi diproses menggunakan Connectionist Temporal Classification dan Levenshtein Distance untuk mengurangi blank output, prediksi berulang, dan kesalahan penulisan sebelum dikonversi menjadi suara melalui Text-to-Speech. Hasil pengujian memperoleh akurasi 95,56%, precision 95,53%, recall 95,57%, F1-score 95,55%, dan Character Error Rate 4,00%. Pengujian real-time menunjukkan performa terbaik pada jarak ±100 cm sebesar 92,00%, pencahayaan terang 92,67%, dan posisi kamera lurus 92,67%. Hasil ini menunjukkan sistem dapat berjalan baik pada kondisi pengujian yang sesuai, tetapi masih terbatas pada 15 kosakata dasar.

Kata kunci: BISINDO, Connectionist Temporal Classification, Long Short-Term Memory, MediaPipe Holistic, Text-to-Speech, Transformer Encoder.

Abstract

Communication between Indonesian Sign Language (BISINDO) users and people who do not understand sign language is still often limited, especially in education and public service settings. This study aims to develop a real-time BISINDO translation prototype that converts sign gestures into Indonesian text and speech. The system uses MediaPipe Holistic to extract hand and body landmarks, while facial landmarks are not used so the features remain lighter and focused on the main gesture movements. Landmark coordinates are combined with velocity features to capture motion changes between frames and arranged into 30-frame sequences. A hybrid Transformer Encoder and Long Short-Term Memory model is used to recognize 15 basic BISINDO words or short phrases from 4,500 sequences. The prediction results are processed using Connectionist Temporal Classification and Levenshtein Distance to reduce blank outputs, repeated predictions, and spelling errors before being converted into speech using Text-to-Speech. The evaluation results show 95.56% accuracy, 95.53% precision, 95.57% recall, 95.55% F1-score, and 4.00% Character Error Rate. Real-time testing showed the best performance at a distance of ±100 cm with 92.00% accuracy, bright lighting with 92.67%, and a straight camera position with 92.67%.

Keywords: *BISINDO, Connectionist Temporal Classification, Long Short-Term Memory, MediaPipe Holistic, Text-to-Speech, Transformer Encoder.*

1. PENDAHULUAN

Komunikasi merupakan kebutuhan utama agar setiap orang dapat saling memahami, termasuk teman Tuli yang menggunakan bahasa isyarat dalam aktivitas sehari-hari. Bahasa Isyarat Indonesia (BISINDO) banyak digunakan oleh komunitas Tuli sebagai media komunikasi berbasis gerakan tangan, tubuh, dan ekspresi. BISINDO berbeda dengan Sistem Isyarat Bahasa Indonesia (SIBI) karena BISINDO berkembang dari komunitas pengguna, sedangkan SIBI cenderung mengikuti struktur bahasa Indonesia formal [1]-[3].

Permasalahan muncul ketika lawan bicara tidak memahami bahasa isyarat. Kondisi ini dapat membuat komunikasi menjadi lambat dan berpotensi menimbulkan salah paham, terutama pada

turnitin Page 7 of 19 - Integrity Submission Submission ID trn:oid:::3618:144969654
Vol. 10, No. 1, April 2025, Page. 1-250
https://journal.uin-alauddin.ac.id/index.php/instek
p-ISSN: 2541-1179
e-ISSN: 2581-1711
JURNAL INSTEK
INFORMATIKA SAINS DAN TEKNOLOGI

pendidikan, layanan publik, dan situasi sosial yang membutuhkan respons cepat. Sistem berbasis computer vision mulai banyak digunakan karena kamera dapat membaca pola gerakan tanpa perangkat sensor tambahan[1], [2], [4], [5]. Namun, pengenalan BISINDO secara real-time masih menghadapi kendala, seperti pencahayaan, jarak pengguna, sudut kamera, kestabilan landmark, dan kemiripan gerakan antar kosakata [1], [2], [6].

Hasil wawancara awal dengan dua siswa tunarungu kelas XI dan satu guru di SLB Negeri 5 Jakarta menunjukkan bahwa BISINDO lebih sering dipilih karena dianggap mudah dipahami. Kendala utama muncul ketika lawan bicara tidak memahami isyarat, sehingga komunikasi biasanya dibantu dengan membaca gerak bibir atau mengetik pesan melalui ponsel. Guru juga menekankan bahwa ketepatan terjemahan lebih penting daripada kecepatan karena kesalahan terjemahan dapat mengubah makna pesan. Temuan ini menguatkan perlunya sistem penerjemah BISINDO yang dapat berjalan secara langsung, tetapi tetap memperhatikan kejelasan hasil terjemahan.

Beberapa penelitian sebelumnya telah mengembangkan pengenalan bahasa isyarat menggunakan MediaPipe, CNN, LSTM, dan metode machine learning lain. Fauzi et al. mengembangkan pengenalan BISINDO menjadi suara [7]. Kartaputra et al. mendeteksi alfabet BISINDO dengan MediaPipe Holistic [8]. Dan beberapa penelitian menggunakan LSTM untuk membaca gerakan isyarat berurutan [9]-[12]. Penelitian pada Kata Kolok juga menunjukkan bahwa MediaPipe dan LSTM dapat bekerja cukup baik pada kosakata terbatas, tetapi kesalahan masih muncul pada isyarat yang memiliki pola gerakan mirip [4].

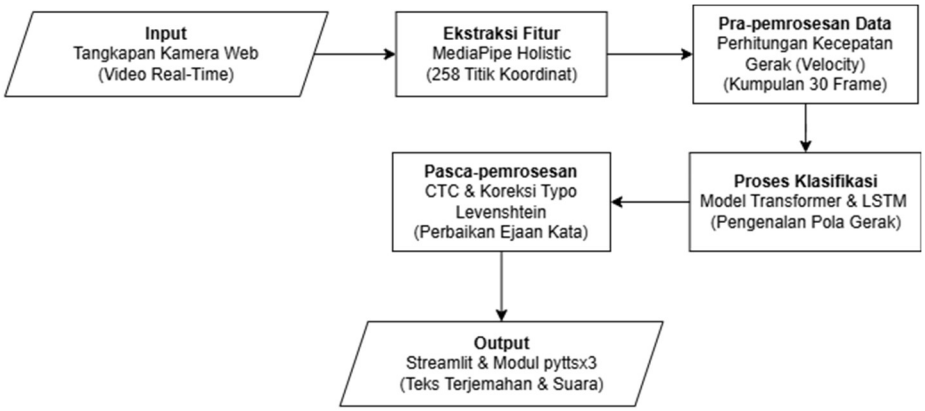
Dari penelitian-penelitian tersebut, masih terlihat beberapa celah yang perlu dikembangkan. Penelitian sebelumnya sudah banyak membahas pengenalan isyarat menggunakan MediaPipe, CNN, atau LSTM, tetapi sebagian besar masih berfokus pada klasifikasi gerakan atau hasil teks [1], [2], [4], [7]-[12]. Belum banyak penelitian yang menggabungkan fitur koordinat dan velocity, pemodelan temporal menggunakan Transformer Encoder dan LSTM, perbaikan teks dengan CTC dan Levenshtein Distance, serta keluaran suara dalam satu alur BISINDO real- time [13]-[20]. Selain itu, pengujian terhadap kondisi penggunaan seperti jarak, pencahayaan, dan sudut kamera juga masih perlu diperjelas karena kondisi kamera dapat memengaruhi kualitas landmark[1], [2], [21]. Oleh karena itu, penelitian ini mengembangkan alur penerjemah BISINDO yang tidak hanya mengenali gerakan, tetapi juga menstabilkan hasil teks dan mengubahnya menjadi suara.

Kebaruan penelitian ini terletak pada penggabungan MediaPipe Holistic, fitur koordinat dan velocity, Transformer Encoder, LSTM, CTC, Levenshtein Distance, Streamlit, dan Text-to-Speech dalam satu sistem penerjemah BISINDO real-time [13]-[19], dan menguji sistem pada beberapa kondisi yang tidak hanya menguji pada jarak dan pencahayaan tetapi juga menguji pada kondisi sudut kamera. Kontribusi penelitian ini meliputi empat hal, yaitu: (1) menyusun dataset 15 kosakata BISINDO sebanyak 4.500 sequence, (2) membentuk input 516 fitur per frame dari gabungan koordinat dan velocity, (3) membangun alur sistem dari input kamera sampai keluaran teks dan suara, serta (4) menguji sistem pada variasi jarak, pencahayaan, dan sudut kamera. Dengan batasan tersebut, penelitian ini diarahkan sebagai prototipe awal untuk penerjemahan kosakata dasar BISINDO, sehingga belum ditujukan sebagai penerjemah penuh untuk percakapan bebas.

2. METODE PENELITIAN

Penelitian ini menggunakan pendekatan kuantitatif dengan menggunakan metode eksperimental dan pengembangan sistem. Pendekatan kuantitatif digunakan karena performa sistem diukur menggunakan accuracy, F1-score, precision, recall, Character Error Rate (CER), dan pengujian real-time [1], [22]. Metode eksperimental digunakan karena sistem diuji pada beberapa kondisi penggunaan, yaitu jarak pengguna, pencahayaan, dan sudut kamera [1], [2], [6]. Secara umum, alur penelitian dimulai dari input video menggunakan webcam, ekstraksi landmark menggunakan MediaPipe Holistic, pembentukan fitur koordinat dan velocity, klasifikasi menggunakan model hybrid Transformer Encoder dan LSTM, post-processing menggunakan CTC dan Levenshtein Distance, kemudian hasil akhir ditampilkan dalam bentuk teks dan suara menggunakan Text-to-Speech melalui antarmuka Streamlit [5], [13]-[21]. Gambar 1 menunjukkan alur pelaksanaan penelitian.

turnitin Page 8 of 19 - Integrity Submission Submission ID trn:oid:::3618:144969654
Vol. 10, No. 1, April 2025, Page. 1-250
https://journal.uin-alauddin.ac.id/index.php/instek
p-ISSN: 2541-1179
e-ISSN: 2581-1711
JURNAL INSTEK
INFORMATIKA SAINS DAN TEKNOLOGI



Gambar 1. Alur sistem penerjemah BISINDO

Gambar 1 menunjukkan bahwa sistem menerima input video real-time melalui webcam. Setiap frame video diproses untuk mengambil landmark tangan dan tubuh menggunakan MediaPipe Holistic. Hasil landmark kemudian diubah menjadi fitur koordinat dan velocity dalam bentuk sequence 30 frame. Sequence tersebut diproses oleh model Transformer Encoder dan LSTM untuk menghasilkan prediksi kosakata. Hasil prediksi kemudian dirapikan menggunakan CTC dan Levenshtein Distance sebelum ditampilkan sebagai teks dan diubah menjadi suara.

2.1. Dataset dan Struktur Penyimpanan

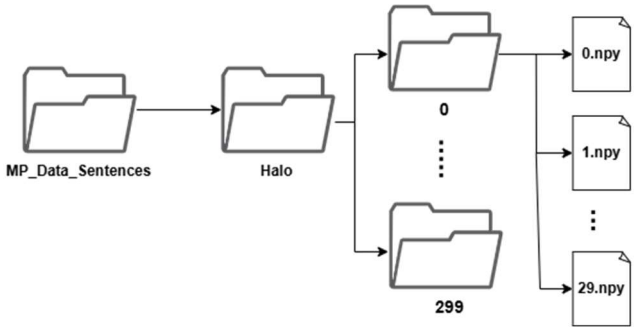
Dataset dikumpulkan dengan merekam 15 kosakata atau frasa dasar BISINDO menggunakan webcam. Kosakata yang digunakan adalah “apa”, “dimana”, “halo”, “hati-hati”, “kamu”, “maaf”, “nama”, “rumah”, “sama-sama”, “saya”, “sekarang”, “siapa”, “teman”, “terimakasih”, dan “tolong”. Setiap kosakata direkam sebanyak 300 sequence, dan setiap sequence terdiri dari 30 frame. Dengan demikian, total dataset yang digunakan berjumlah 4.500 sequence atau 135.000 frame. Jumlah kosakata ini digunakan sebagai batasan penelitian karena sistem difokuskan terlebih dahulu pada pengenalan kosakata dasar BISINDO [1], [4], [7]-[12], [23].

Dataset dibagi ke dalam data pelatihan dan data pengujian dengan perbandingan 90:10. Dari total 4.500 sequence, sebanyak 4.050 sequence digunakan sebagai data yang akan dilatih dan 450 sequence digunakan untuk data yang akan uji. Pembagian akan dilakukan pada setiap label kosakata agar jumlah data tiap kelas tetap seimbang. Setiap kosakata memiliki 300 sequence, sehingga 270 sequence digunakan melatih data dan 30 sequence digunakan menguji data. Pembagian ini dilakukan berdasarkan sequence, bukan berdasarkan frame, karena data bahasa isyarat berbentuk urutan gerakan yang saling berkaitan dari frame awal sampai frame akhir [5], [8]-[12], [24]. Cara ini digunakan supaya frame dari satu gerakan yang sama tidak masuk ke data latih dan data uji sekaligus. Jika pembagian dilakukan per-frame, model dapat mengenali potongan gerakan yang masih berasal dari sequence yang sama, sehingga hasil evaluasi berisiko terlihat lebih tinggi dari kondisi sebenarnya. Dengan pembagian berbasis sequence, data uji berisi gerakan utuh yang tidak digunakan pada proses pelatihan.

Data disimpan berdasarkan label kosakata agar proses pelatihan dan pengujian lebih terstruktur. Setiap folder label berisi 300 folder sequence, mulai dari sequence 0 sampai 299. Di dalam setiap sequence terdapat 30 file frame dalam format NumPy, yaitu 0.npy sampai 29.npy. Struktur penyimpanan dataset ditunjukkan pada Gambar 2.

turnitin Page 8 of 19 - Integrity Submission Submission ID trn:oid:::3618:144969654

Vol. 10, No. 1, April 2025, Page. 1-250
https://journal.uin-alauddin.ac.id/index.php/instek
p-ISSN: 2541-1179
e-ISSN: 2581-1711
JURNAL INSTEK
INFORMATIKA SAINS DAN TEKNOLOGI



Gambar 2. Struktur penyimpanan dataset

Gambar 2 menunjukkan dataset disusun berdasarkan label kosakata BISINDO. Setiap label memiliki folder sequence, kemudian setiap sequence berisi file frame hasil ekstraksi fitur. Struktur ini membantu menjaga keteraturan dataset dan mengurangi risiko kesalahan pemanggilan label saat proses pelatihan model.

Dataset pada penelitian ini masih terbatas pada 15 kosakata atau frasa dasar BISINDO. Jumlah tersebut cukup untuk menguji prototipe awal, tetapi belum mewakili percakapan BISINDO yang lebih beragam dan kompleks. Karena itu, hasil penelitian ini belum dapat digeneralisasi untuk semua bentuk komunikasi BISINDO. Pengembangan berikutnya dapat dilakukan dengan menambah jumlah kosakata, melibatkan lebih banyak subjek, menggunakan variasi lingkungan perekaman, serta menguji bentuk kalimat yang lebih panjang.

2.2. Ekstraksi Fitur dan Pra-pemrosesan

MediaPipe Holistic digunakan untuk mengekstraksi landmark tangan dan tubuh sebagai fitur utama dalam pengenalan gerakan BISINDO [5], [21]. Landmark wajah tidak digunakan karena penelitian ini berfokus pada 15 kosakata dasar yang lebih banyak dibedakan melalui gerakan tangan dan pose tubuh. Selain itu, penggunaan landmark wajah akan menambah jumlah fitur yang diproses, sehingga data menjadi lebih besar dan proses real-time menjadi lebih berat. Dengan menghilangkan landmark wajah, sistem dapat lebih fokus pada bagian tangan dan tubuh yang menjadi sumber gerakan utama. Namun, keputusan ini tetap menjadi keterbatasan karena sistem belum dirancang untuk mengenali isyarat yang sangat bergantung pada ekspresi wajah atau gerak bibir [6], [21]. Setiap tangan memiliki 21 titik landmark dengan tiga koordinat, yaitu x, y, dan z. Karena sistem membaca tangan kanan dan kiri, jumlah fitur tangan dihitung menggunakan Persamaan (1) dan Persamaan (2) [21].

$$F_{kanan} = 21\ Landmark \times 3\ koordinat\,(x, y, z) = 63 \quad (1)$$

$$F_{kiri} = 21\ Landmark \times 3\ koordinat\,(x, y, z) = 63 \quad (2)$$

Selain tangan, sistem menggunakan 33 titik pose tubuh. Setiap titik pose memiliki nilai x, y, z, dan visibility [21]. Oleh karena itu, jumlah fitur pose tubuh dihitung menggunakan Persamaan (3).

$$F_{pose} = 33\ Landmark \times (3\ Koordinat \times 1\ Visibility)(x, y, z, v) = 132 \quad (3)$$

Fitur koordinat tangan dan tubuh kemudian digabungkan menjadi 258 fitur koordinat. Jumlah tersebut dihitung menggunakan Persamaan (4).

$$Fitur\ koordinat\ =\ 126\ fitur\ tangan\ +\ 132\ fitur\ pose\ =\ 258 \quad (4)$$

Penelitian ini juga menggunakan fitur velocity agar model tidak hanya membaca posisi landmark pada satu frame. Velocity dihitung dari perubahan posisi landmark antar-frame, sehingga model dapat memperoleh informasi arah dan perubahan gerakan. Fitur ini dibutuhkan karena beberapa isyarat dapat memiliki bentuk tangan yang mirip, tetapi berbeda pada lintasan atau arah geraknya. Pada penelitian berbasis pose, joint velocity dan bone velocity digunakan untuk menangkap dinamika gerakan yang tidak terlihat dari koordinat statis saja [20]. Dengan demikian, penggunaan velocity pada penelitian ini

Vol. 10, No. 1, April 2025, Page. 1-250
https://journal.uin-alauddin.ac.id/index.php/instek
p-ISSN: 2541-1179
e-ISSN: 2581-1711
JURNAL INSTEK
INFORMATIKA SAINS DAN TEKNOLOGI

bertujuan membantu model membaca pola gerakan secara lebih lengkap. Setelah fitur koordinat dan velocity digabungkan, jumlah fitur pada setiap frame menjadi 516 fitur, seperti ditunjukkan pada Persamaan (5) [20].

$$Fitur\ per\ frame\ =\ 258\ fitur\ koordinat\ +\ 258\ fitur\ velocity\ =\ 516 \quad (5)$$

Setiap input model terdiri dari 30 frame, sehingga bentuk input sequence yang masuk ke model ditunjukkan pada Persamaan (6).

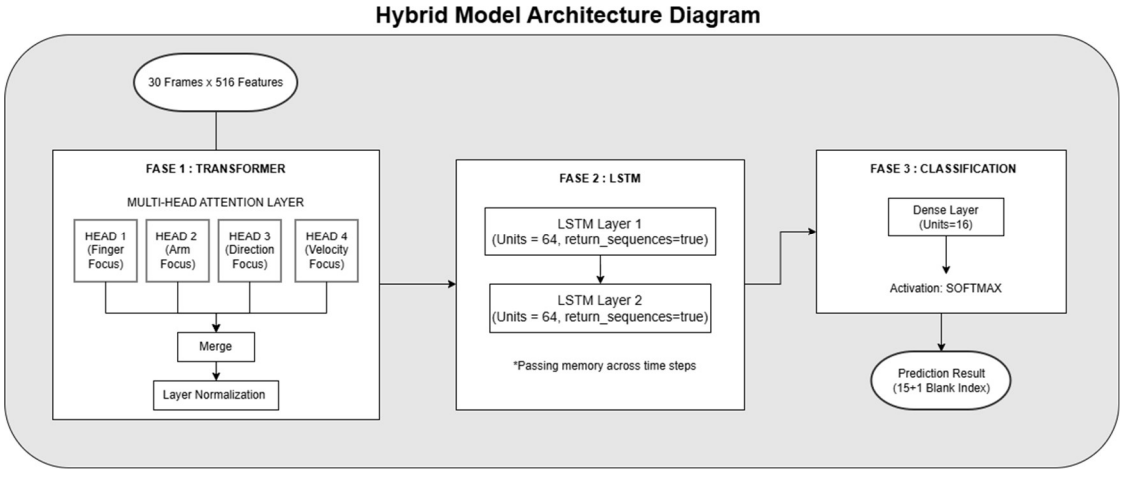
$$Input\ sequence = 30\ frame \times 516\ fitur \quad (6)$$

2.3. Pengembangan Model dan Post-processing

Model pada penelitian ini dikembangkan secara bertahap. Pada tahap awal, sistem menggunakan MediaPipe Holistic untuk mengambil landmark dan LSTM sebagai model utama. Tahap ini digunakan sebagai baseline awal karena LSTM banyak digunakan untuk membaca data gerakan yang berbentuk sequence [4], [9], [10]-[12], [21]. Namun, dari pengujian awal, model MediaPipe dan LSTM saja masih kurang stabil pada beberapa kosakata yang memiliki pola gerakan mirip. Kesalahan sering muncul ketika bentuk tangan hampir sama, tetapi arah atau urutan gerakannya berbeda.

Setelah itu, Transformer Encoder ditambahkan sebelum LSTM. Transformer Encoder digunakan untuk membantu model memperhatikan bagian frame yang lebih penting melalui mekanisme self-attention [13], [14]. Dengan cara ini, model tidak hanya membaca urutan gerakan melalui LSTM, tetapi juga dapat memberi perhatian lebih pada bagian gerakan yang paling berpengaruh terhadap prediksi. Selain itu, fitur velocity juga ditambahkan agar model tidak hanya membaca posisi landmark, tetapi juga perubahan gerakan antar-frame [20].

Pengujian awal juga dilakukan pada keluaran model sebelum menggunakan post-processing. Pada tahap tersebut, model sudah dapat mengenali kosakata, tetapi hasil teks masih dapat muncul berulang, kurang stabil, atau menghasilkan blank output. Oleh karena itu, CTC digunakan untuk merapikan keluaran model dengan menghapus blank dan menggabungkan prediksi yang berulang [15], [16]. Levenshtein Distance kemudian digunakan sebagai pemeriksa akhir agar teks hasil prediksi tetap sesuai dengan daftar kosakata BISINDO yang digunakan [17]. Dengan demikian, baseline awal pada penelitian ini adalah MediaPipe Holistic dan LSTM, sedangkan sistem akhir menggunakan tambahan Transformer Encoder, fitur velocity, CTC, dan Levenshtein Distance. Arsitektur model ditunjukkan pada Gambar 3.

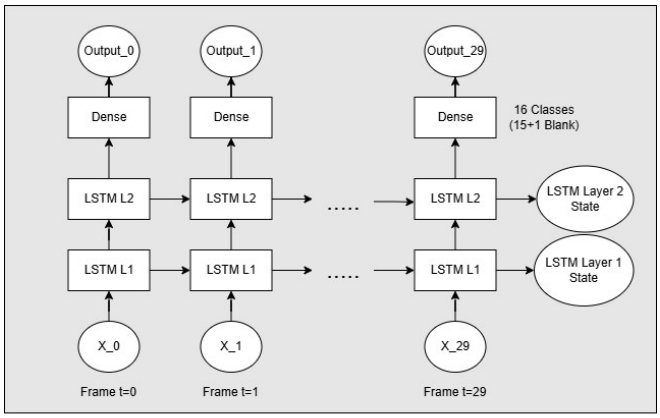


Gambar 3. Arsitektur model Transformer Encoder dan LSTM

Gambar 3 menunjukkan bahwa input sequence diproses melalui tiga tahap utama, yaitu Transformer Encoder, LSTM, dan classification layer. Pada tahap Transformer Encoder, fitur gerakan diproses menggunakan multi-head attention agar model dapat membaca pola gerakan dari beberapa bagian, seperti posisi tangan, arah gerakan, dan perubahan gerakan antar-frame [13], [14], kemudian diteruskan ke dua layer LSTM [4], [9]-[12]. Setelah itu, hasilnya masuk ke dense layer dengan aktivasi softmax

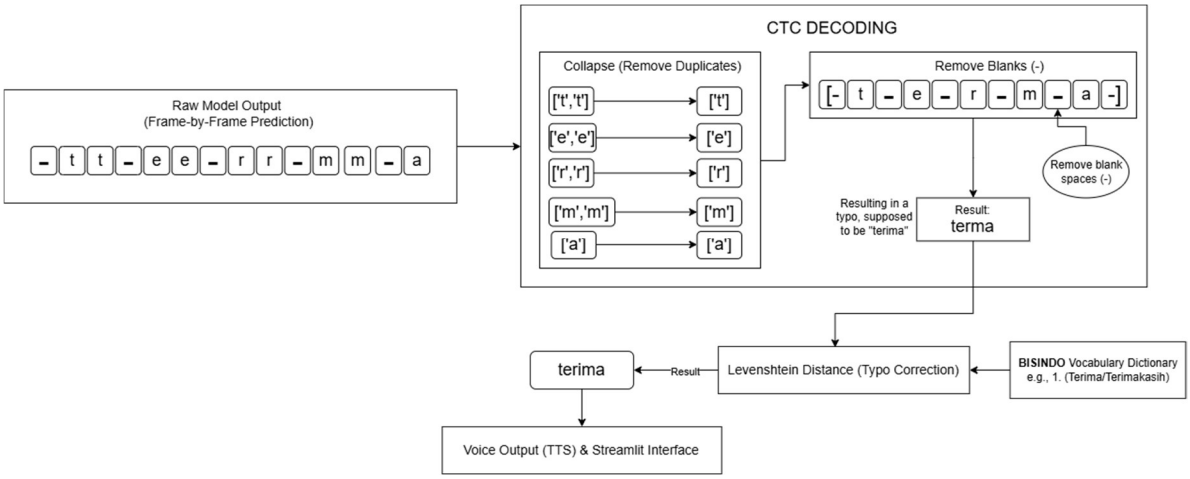
Vol. 10, No. 1, April 2025, Page. 1-250
https://journal.uin-alauddin.ac.id/index.php/instek
p-ISSN: 2541-1179
e-ISSN: 2581-1711
JURNAL INSTEK
INFORMATIKA SAINS DAN TEKNOLOGI

untuk menghasilkan prediksi terhadap 16 kelas, yaitu 15 kosakata BISINDO dan satu kelas blank untuk kebutuhan CTC.



Gambar 4. Proses LSTM dalam membaca sequence

Gambar 4 menunjukkan proses LSTM dalam membaca sequence dari frame ke-0 sampai frame ke-29. Setiap frame direpresentasikan sebagai vektor fitur X_0, X_1, hingga X_{29}, dengan masing-masing berisi 516 fitur. Pada Layer 1, setiap frame diproses secara berurutan dan menghasilkan hidden state. Hasil dari Layer 1 kemudian diteruskan sebagai input ke Layer 2 untuk memperkuat pola gerakan. Setelah itu, hasil akhir dari LSTM masuk ke dense layer. Dense layer digunakan untuk mengubah fitur hasil pembelajaran menjadi nilai prediksi kelas. Pada penelitian ini, dense layer menggunakan aktivasi softmax untuk menghasilkan prediksi terhadap 16 kelas, yaitu 15 kosakata BISINDO dan satu kelas blank untuk kebutuhan CTC [4], [9]-[12], [15].



Gambar 5. Proses CTC decoding dan koreksi teks

Gambar 5 menunjukan hasil keluaran model tidak langsung digunakan sebagai teks akhir karena prediksi dari data sequence masih dapat menghasilkan label berulang atau blank output. Kondisi ini terlihat pada pengujian awal tanpa post-processing, ketika model sudah dapat mengenali kosakata, tetapi teks yang keluar belum selalu stabil. Oleh karena itu, CTC decoding digunakan untuk merapikan hasil prediksi dengan menghapus blank output dan menggabungkan prediksi yang berulang [15], [16]. Setelah tahap CTC, hasil teks sebenarnya sudah lebih rapi, tetapi Levenshtein Distance tetap digunakan sebagai pemeriksa akhir terhadap daftar kosakata BISINDO. Tahap ini membantu mengurangi kesalahan kecil pada penulisan atau perbedaan karakter yang masih mungkin muncul pada hasil prediksi [17]. Dengan demikian, CTC berfungsi untuk menstabilkan urutan prediksi, sedangkan Levenshtein Distance berfungsi sebagai penjaga agar teks akhir tetap sesuai dengan kosakata yang digunakan dalam penelitian. Teks akhir kemudian ditampilkan pada antarmuka Streamlit dan diubah menjadi suara menggunakan

turnitin Page 12 of 19 - Integrity Submission Submission ID trn:oid:::3618:144969654
Vol. 10, No. 1, April 2025, Page. 1-250
https://journal.uin-alauddin.ac.id/index.php/instek
p-ISSN: 2541-1179
e-ISSN: 2581-1711
JURNAL INSTEK
INFORMATIKA SAINS DAN TEKNOLOGI

Text-to-Speech agar lawan bicara yang tidak memahami BISINDO dapat menerima hasil terjemahan dalam bentuk suara [18], [19].

2.4. Pengujian Real-Time dan Metrik Evaluasi

Pengujian real-time dilakukan menggunakan 15 kosakata BISINDO yang telah dilatih pada model. Setiap kosakata diuji sebanyak 10 kali pada setiap kondisi, sehingga jumlah pengujian pada setiap kondisi adalah 150 percobaan. Skenario pengujian meliputi variasi jarak, pencahayaan, dan sudut kamera karena kondisi tersebut dapat memengaruhi kualitas deteksi gerakan pada sistem [1], [2], [6], [21]. Rincian skenario ditunjukkan pada Tabel 1.

Tabel 1. Skenario pengujian real-time

Skenario	Kondisi	Total Test
Variasi Jarak	±50 cm, ±100 cm, ±150 cm	150 tests
Variasi Pencahayaan	Terang dan Redup	150 tests
Variasi Posisi	Lurus dan Serong	150 tests

Evaluasi model dilakukan menggunakan confusion matrix, accuracy, F1-score, recall, precision, dan CER. Confusion matrix digunakan untuk mengidentifikasi jumlah prediksi yang benar maupun salah pada masing-masing kelas, Adapun accuracy, recall, F1-score, dan precision dimanfaatkan untuk menilai performa model dalam melakukan klasifikasi [1], [11], [12]. CER digunakan untuk mengukur tingkat kesalahan karakter pada teks hasil prediksi, terutama setelah proses CTC decoding dan koreksi teks [15], [16], [17], [22]. Rumus evaluasi ditunjukkan pada Persamaan (7) sampai Persamaan (11).

$$Accuracy = = \frac{TP+TN}{TP+TN+FP+FN} \times 100\% \quad (7)$$

$$Precision = = \frac{TP}{TP+FP} \quad (8)$$

$$Recall = = \frac{TP}{TP+FN} \quad (9)$$

$$F1\text{-}score = = 2 \times \frac{Recall \times Precision}{Recall+Precision} \quad (10)$$

$$CER = \frac{S+D+I}{N} \quad (11)$$

Pada Persamaan (7) sampai Persamaan (10), TP menunjukkan True Positive, TN menunjukkan True Negative, FP menunjukkan False Positive, sedangkan FN menunjukkan False Negative. Pada Persamaan (11), S adalah substitution, D adalah deletion, I adalah insertion, dan N adalah jumlah karakter pada teks referensi. Untuk pengujian real-time, accuracy dihitung dari jumlah prediksi benar dibandingkan dengan total percobaan. Misalnya, pada jarak normal terdapat 138 prediksi benar dari 150 percobaan, sehingga perhitungannya ditunjukkan pada Persamaan (12) [1], [22].

$$Accuracy\ real\text{-}time = \frac{138}{150} \text{ x } 100\% = 92{,}00\% \quad (12)$$

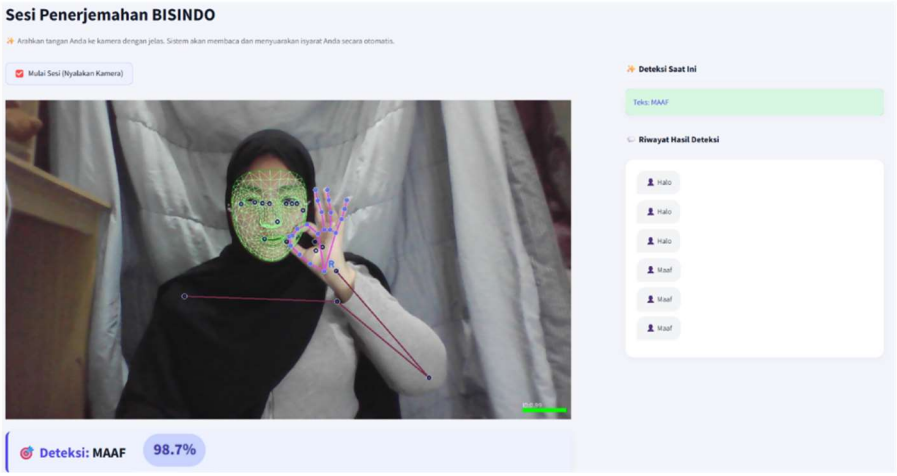
3. HASIL DAN PEMBAHASAN

3.1 Hasil Implementasi Sistem

Sistem penerjemah BISINDO berhasil dibangun dengan alur input kamera, ekstraksi landmark MediaPipe, pra-pemrosesan fitur, klasifikasi menggunakan Transformer Encoder dan LSTM, CTC, koreksi Levenshtein Distance, antarmuka Streamlit, dan Text-to-Speech. Tampilan sistem dan contoh hasil ekstraksi landmark ditunjukkan pada Gambar 6.

turnitin Page 12 of 19 - Integrity Submission Submission ID trn:oid:::3618:144969654

turnitin Page 13 of 19 - Integrity Submission Submission ID trn:oid:::3618:144969654
Vol. 10, No. 1, April 2025, Page. 1-250
https://journal.uin-alauddin.ac.id/index.php/instek
p-ISSN: 2541-1179
e-ISSN: 2581-1711
JURNAL INSTEK
INFORMATIKA SAINS DAN TEKNOLOGI



Gambar 6. Tampilan sistem dan contoh ekstraksi landmark

Gambar 6 menunjukkan bahwa sistem dapat menerima gerakan secara real-time, menampilkan hasil prediksi dalam bentuk teks, menyimpan riwayat hasil deteksi, dan menghasilkan keluaran suara. Hasil ini menunjukkan bahwa alur sistem sudah berjalan dari input visual sampai output teks dan suara. Namun, hasil implementasi ini belum digunakan untuk menyatakan bahwa metode yang digunakan lebih unggul dari semua metode lain. Implementasi ini lebih tepat dibaca sebagai bukti bahwa pipeline yang diusulkan dapat berjalan dan diuji pada kosakata serta kondisi yang sudah ditentukan.

3.2 Hasil Pengujian Berdasarkan Jarak

Pengujian jarak dilakukan untuk mengetahui pengaruh posisi pengguna terhadap kualitas deteksi landmark. Kondisi yang diuji adalah jarak dekat ±50 cm, jarak normal ±100 cm, dan jarak jauh ±150 cm. Contoh pengujian berdasarkan jarak ditunjukkan pada Gambar 7.



Gambar 7. Contoh pengujian kata ”Tolong” berdasarkan variasi jarak

Gambar 7 menunjukkan bahwa jarak pengguna memengaruhi ukuran tangan dan tubuh yang terbaca oleh kamera. Pada jarak dekat, tangan terlihat lebih besar, tetapi beberapa bagian gerakan dapat terlalu dekat dengan kamera. Pada jarak normal, tangan dan tubuh terlihat lebih proporsional sehingga landmark dapat terbaca lebih stabil. Pada jarak jauh, ukuran tangan terlihat lebih kecil dan beberapa detail gerakan menjadi lebih sulit ditangkap oleh sistem.

Tabel 2. Hasil pengujian berdasarkan jarak

Kondisi Jarak	Prediksi Benar	Total Test	Accuracy
Jarak Dekat (±50 cm)	130	150	86,67%
Jarak Normal (±100 cm)	138	150	92,00%
Jarak Jauh (±150 cm)	98	150	65,33%

Berdasarkan Tabel 2, hasil terbaik diperoleh pada jarak normal ±100 cm dengan akurasi 92,00%. Kondisi ini menunjukkan bahwa posisi pengguna pada jarak normal membuat tangan dan tubuh terlihat

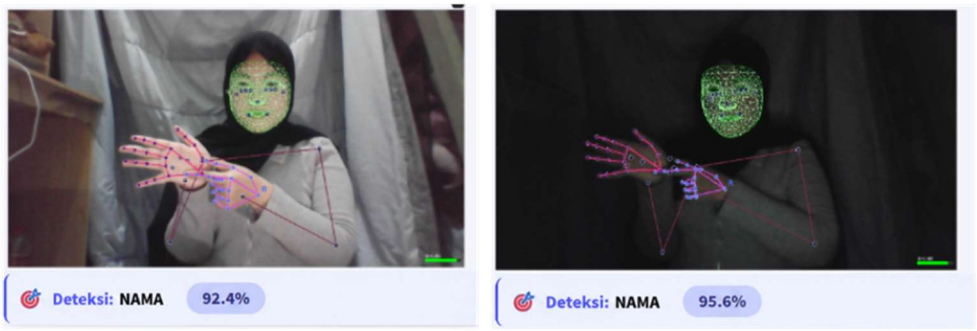
turnitin Page 13 of 19 - Integrity Submission Submission ID trn:oid:::3618:144969654

turnitin Page 14 of 19 - Integrity Submission Submission ID trn:oid:::3618:144969654
Vol. 10, No. 1, April 2025, Page. 1-250
https://journal.uin-alauddin.ac.id/index.php/instek
p-ISSN: 2541-1179
e-ISSN: 2581-1711
JURNAL INSTEK
INFORMATIKA SAINS DAN TEKNOLOGI

lebih proporsional oleh kamera. Pada jarak dekat, akurasi turun menjadi 86,67% karena beberapa gerakan tangan dapat terlalu dekat dengan kamera sehingga posisi landmark tidak selalu terbaca stabil. Penurunan paling besar terjadi pada jarak jauh ±150 cm dengan akurasi 65,33%. Hal ini terjadi karena ukuran tangan dan tubuh menjadi lebih kecil pada frame, sehingga detail gerakan lebih sulit ditangkap. Dengan demikian, jarak pengguna berpengaruh langsung terhadap kualitas landmark yang diterima model.

3.3 Hasil Pengujian Berdasarkan Pencahayaan

Pengujian pencahayaan dilakukan untuk mengetahui pengaruh cahaya terhadap hasil deteksi gerakan. Kondisi yang diuji adalah pencahayaan terang dan redup. Contoh pengujian pencahayaan ditunjukkan pada Gambar 8.



Kondisi Terang "Nama" Kondisi Redup "Nama"

Gambar 8. Contoh pengujian kata ”ama” berdasarkan pencahayaan

Gambar 8 menunjukkan bahwa pada pencahayaan terang, bentuk tangan dan tubuh terlihat lebih jelas sehingga titik landmark dapat terbaca lebih stabil. Pada pencahayaan redup, kualitas gambar menurun dan beberapa titik landmark menjadi kurang konsisten. Kondisi ini berpengaruh pada hasil prediksi karena model bergantung pada fitur landmark yang diambil dari kamera.

Tabel 3. Hasil pengujian berdasarkan pencahayaan

Kondisi Pencahayaan	Prediksi Benar	Total Test	Accuracy
Kondisi Terang	139	150	92,67%
Kondisi Redup	122	150	81,33%

Berdasarkan Tabel 3, sistem memperoleh akurasi 92,67% pada pencahayaan terang dan 81,33% pada pencahayaan redup. Perbedaan ini menunjukkan bahwa pencahayaan berpengaruh terhadap kestabilan pembacaan landmark. Pada kondisi terang, bentuk tangan dan tubuh terlihat lebih jelas sehingga titik landmark lebih mudah dikenali. Sebaliknya, pada kondisi redup, sebagian detail tangan kurang terlihat dan posisi landmark menjadi kurang stabil. Kondisi ini membuat fitur yang masuk ke model tidak sejelas pada pencahayaan terang, sehingga hasil prediksi ikut menurun.

3.4 Hasil Pengujian Berdasarkan Sudut Kamera

Pengujian sudut kamera dilakukan untuk melihat pengaruh posisi tubuh terhadap hasil deteksi. Kondisi yang diuji adalah posisi lurus menghadap kamera dan posisi serong. Contoh pengujian sudut kamera ditunjukkan pada Gambar 9.

turnitin Page 14 of 19 - Integrity Submission Submission ID trn:oid:::3618:144969654

Vol. 10, No. 1, April 2025, Page. 1-250
https://journal.uin-alauddin.ac.id/index.php/instek
p-ISSN: 2541-1179
e-ISSN: 2581-1711
JURNAL INSTEK
INFORMATIKA SAINS DAN TEKNOLOGI



Posisi Lurus "Saya" Posisi Serong "Saya"

Gambar 9. Contoh pengujian kata ”Saya” berdasarkan sudut kamera

Gambar 9 menunjukkan bahwa pada posisi lurus, tangan dan tubuh terlihat lebih lengkap oleh kamera. Pada posisi serong, sebagian tangan atau tubuh dapat tertutup, sehingga fitur yang diterima model menjadi kurang lengkap. Kondisi ini dapat menyebabkan hasil prediksi menurun karena sistem membutuhkan landmark tangan dan tubuh yang terlihat jelas.

Tabel 4. Hasil pengujian berdasarkan sudut kamera

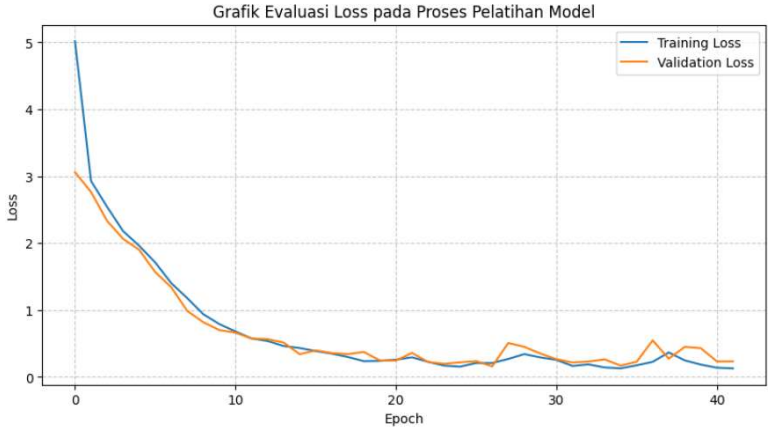
Sudut Kamera	Prediksi Benar	Total Test	Accuracy
Posisi Lurus	139	150	92,67%
Posisi Serong	117	150	78,00%

Berdasarkan Tabel 4, sistem memperoleh akurasi 92,67% pada posisi lurus dan 78,00% pada posisi serong. Penurunan pada posisi serong terjadi karena sebagian tangan atau tubuh tidak selalu terlihat utuh oleh kamera. Selain itu, arah gerakan dapat terlihat berbeda ketika pengguna tidak menghadap kamera secara langsung. Kondisi ini membuat fitur landmark yang diterima model menjadi kurang lengkap, sehingga prediksi lebih mudah bergeser ke kelas lain.

Dari pengujian jarak, pencahayaan, dan sudut kamera, terlihat bahwa sistem bekerja paling baik ketika tangan dan tubuh terlihat jelas oleh kamera. Kondisi terbaik muncul pada jarak normal, pencahayaan terang, dan posisi lurus. Sebaliknya, performa menurun ketika gerakan sulit dibaca, seperti pada jarak jauh, pencahayaan redup, dan posisi serong. Hasil tersebut menunjukkan bahwa performa sistem turut dipengaruhi oleh kualitas input dari kamera dan posisi pengguna, bukan hanya oleh kemampuan model.

3.5 Hasil Evaluasi Model

Evaluasi model dilakukan untuk mengetahui kemampuan model dalam mengenali data uji yang belum digunakan pada tahap pelatihan. Grafik training loss dan validation loss digunakan untuk mengamati proses pembelajaran model selama proses pelatihan. Hasil grafik loss ditampilkan pada Gambar 10.

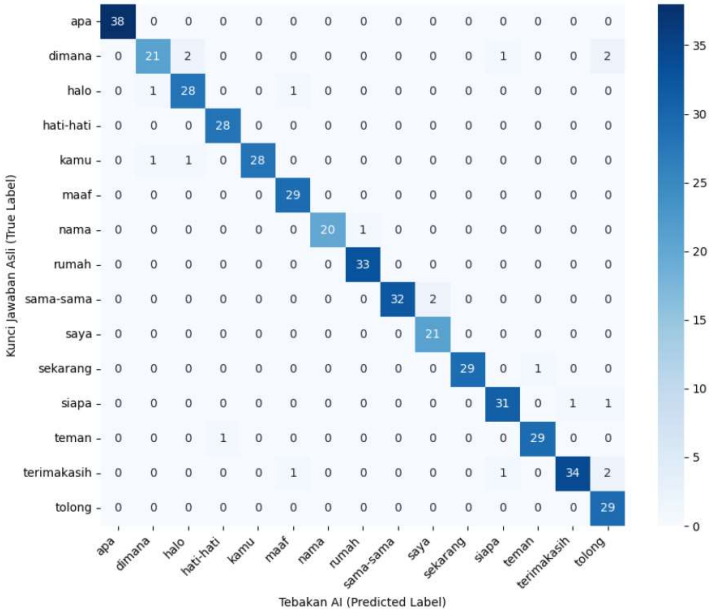


Gambar 10. Hasil training loss dan validation loss

Vol. 10, No. 1, April 2025, Page. 1-250
https://journal.uin-alauddin.ac.id/index.php/instek
p-ISSN: 2541-1179
e-ISSN: 2581-1711
JURNAL INSTEK
INFORMATIKA SAINS DAN TEKNOLOGI

Gambar 10 menunjukkan bahwa nilai training loss dan validation loss menurun pada epoch awal, kemudian mulai stabil pada nilai kesalahan yang rendah. Jarak antara training loss dan validation loss juga tidak terlalu besar, sehingga model tidak menunjukkan tanda overfitting yang kuat pada data uji yang digunakan. Hasil pada data uji menunjukkan bahwa model cukup mampu mengenali dan mempelajari pola gerakan yang terdapat dalam dataset penelitian. Namun, hasil ini tetap perlu dibatasi karena dataset masih terdiri dari 15 kosakata dasar dan belum dibandingkan secara penuh dengan model lain melalui ablation study.

Confusion matrix digunakan untuk mengevaluasi hasil prediksi model pada masing-masing kelas. Hasil confusion matrix disajikan pada Gambar 11.



Gambar 11. Hasil confusion matrix

Gambar 11 menunjukkan bahwa sebagian besar kosakata berhasil dikenali dengan benar. Dari 450 data uji, terdapat 430 prediksi benar. Beberapa kosakata seperti “apa”, “hati-hati”, “rumah”, dan “saya” memperoleh hasil yang baik karena pola gerakannya lebih mudah dibedakan dari kelas lain.

Namun, confusion matrix juga menunjukkan bahwa masih ada kesalahan pada beberapa kosakata yang memiliki pola gerakan mirip. Misalnya, kosakata “dimana” beberapa kali diprediksi sebagai “halo” atau “tolong”. Kesalahan seperti ini dapat terjadi karena kemiripan bentuk tangan, posisi awal gerakan, atau arah gerakan pada beberapa frame. Jika landmark tidak terbaca stabil, model dapat menerima fitur yang kurang lengkap dan memilih kelas lain yang dianggap paling mirip. Pola error ini menunjukkan bahwa batas kemampuan model masih terlihat pada kosakata yang memiliki bentuk tangan atau lintasan gerak yang hampir sama.

Tabel 5. Hasil evaluasi model

Metrik	Hasil
Accuracy	95,56%
F1-score	95,55%
Recall	95,57%
Precision	95,53%
Character Error Rate (CER)	4,00%

Berdasarkan Tabel 5, model memperoleh akurasi 95,56%, F1-score 95,55%, recall 95,57%, precision 95,53%, dan Character Error Rate (CER) 4,00%. Hasil ini menunjukkan bahwa pipeline yang dikembangkan mampu mengenali sebagian besar kosakata BISINDO pada data uji. Nilai CER yang rendah juga menunjukkan bahwa kesalahan karakter pada teks hasil prediksi relatif kecil setelah melalui CTC dan Levenshtein Distance. Namun, hasil akurasi ini belum dapat digunakan untuk menyatakan bahwa metode yang digunakan lebih unggul dari semua metode lain, karena penelitian ini belum

Vol. 10, No. 1, April 2025, Page. 1-250 p-ISSN: 2541-1179
https://journal.uin-alauddin.ac.id/index.php/instek e-ISSN: 2581-1711
JURNAL INSTEK INFORMATIKA SAINS DAN TEKNOLOGI

menyajikan pembanding lengkap seperti model tanpa Transformer, tanpa velocity, atau tanpa post-processing. Oleh karena itu, hasil ini lebih tepat dibaca sebagai performa sistem yang diusulkan pada dataset dan skenario pengujian yang digunakan.

4. KESIMPULAN

Penelitian ini berhasil membangun sistem penerjemah BISINDO real-time yang dapat menerjemahkan gerakan isyarat ke dalam bentuk teks dan suara. Sistem dibangun dengan MediaPipe Holistic untuk mengekstraksi landmark tangan dan tubuh, kemudian fitur koordinat digabungkan dengan velocity agar model dapat membaca posisi sekaligus perubahan gerakan antar-frame. Data sequence tersebut diproses menggunakan model hybrid Transformer Encoder dan LSTM, lalu hasil prediksi dirapikan dengan CTC dan Levenshtein Distance sebelum ditampilkan melalui Streamlit dan diubah menjadi suara menggunakan Text-to-Speech.

Kontribusi utama penelitian ini terletak pada penggabungan beberapa komponen dalam satu alur sistem, yaitu ekstraksi landmark, fitur velocity, model hybrid Transformer Encoder dan LSTM, post-processing teks, serta keluaran suara. Dengan alur tersebut, sistem tidak hanya berhenti pada klasifikasi gerakan, tetapi juga menghasilkan teks yang lebih stabil dan suara yang dapat digunakan oleh lawan bicara yang tidak memahami BISINDO. Hasil evaluasi menunjukkan bahwa model memperoleh akurasi 95,56%, precision 95,53%, recall 95,57%, F1-score 95,55%, dan Character Error Rate 4,00%. Pada pengujian real-time, performa terbaik diperoleh pada jarak ±100 cm dengan akurasi 92,00%, pencahayaan terang dengan akurasi 92,67%, dan posisi kamera lurus dengan akurasi 92,67%.

Meskipun hasilnya cukup baik, klaim keberhasilan sistem tetap dibatasi pada ruang lingkup penelitian ini. Dataset yang digunakan masih terdiri dari 15 kosakata atau frasa dasar BISINDO, sehingga hasil penelitian belum dapat digeneralisasi untuk percakapan BISINDO yang lebih luas dan kompleks. Sistem juga masih mengalami penurunan pada kondisi jarak jauh, pencahayaan redup, dan sudut kamera serong. Selain itu, kesalahan klasifikasi masih muncul pada kosakata yang memiliki pola gerakan mirip. Oleh karena itu, pengembangan berikutnya dapat dilakukan dengan menambah jumlah kosakata, melibatkan lebih banyak subjek, memperluas variasi kondisi perekaman, menguji model pembanding secara lebih lengkap, meningkatkan kualitas Text-to-Speech melalui layanan berbasis cloud, serta mengembangkan sistem agar dapat membaca kalimat BISINDO yang lebih panjang.

5. DAFTAR PUSTAKA

[1] R. W. L. Therry, A. Junaidi, and A. N. Sihananto, “Program Penerjemah Bahasa Isyarat Indonesia (BISINDO) Secara Real Time Menggunakan Convolutional Neural Network Dan Mediapipe,” *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 8, no. 6, pp. 11642–11649, 2024, doi: 10.36040/jati.v8i6.11582.

[2] D. Ahwani, S. Budiman, and M. Rahmat, “Penerapan Mediapipe dalam Pengenalan Bisindo Berbasis Deep Learning dan Computer Vision (Studi Kasus: SLB-C B Yayasan Pendidikan Luar Biasa (YPLB) Blitar),” *INNOVATIVE: Journal Of Social Science Research*, vol. 4, no. 5, pp. 3608–3628, 2024, doi: 10.31004/innovative.v4i5.15199.

[3] A. Setiawan and F. N. Hasan, “ANALISIS SENTIMEN TANGGAPAN PENGGUNA APLIKASI BALE BY BTN MENGGUNAKAN METODE SUPPORT VECTOR MACHINE (SVM),” *Jurnal Ilmiah Teknik dan Ilmu Komputer*, vol. 4, no. 4, pp. 315–326, 2025, doi: 10.55123.

[4] M. P. Prayoga, G. A. A. D. Indradewi, and P. H. Suputra, “Pengenalan Kata Kolok Secara Real Time Menggunakan Mediapipe dan Algoritma Long Short-Term Memory (LSTM),” *Sintech Journal*, vol. 8, no. 2, pp. 178–187, 2025, doi: 10.31598/sintechjournal.v8i2.1945.

[5] S. Nur Budiman, S. Lestanti, S. Marselius Evvandri, and R. Kartika Putri, “Pengenalan Gestur Gerakan Jari Untuk Mengontrol Volume Di Komputer Menggunakan Library

Vol. 10, No. 1, April 2025, Page. 1-250
https://journal.uin-alauddin.ac.id/index.php/instek
p-ISSN: 2541-1179
e-ISSN: 2581-1711
JURNAL INSTEK
INFORMATIKA SAINS DAN TEKNOLOGI

Opencv Dan Mediapipe," *Antivirus : Jurnal Ilmiah Teknik Informatika*, vol. 16, no. 2, pp. 223–232, Nov. 2022, doi: 10.35457/antivirus.v16i2.2508.

[6] Hajrah, A. azis, and I. Rahim, "Analisis Penggunaan Bahasa Isyarat Indonesia (BISINDO) pada Siswa SLB," *Jurnal Onoma: Pendidikan, Bahasa dan Sastra*, vol. 9, no. 2, pp. 1396–1402, 2023, doi: 10.30605/onoma.v9i2.2717.

[7] M. Z. Fauzi, R. Sarno, and S. C. Hidayati, "Recognition of Real-Time BISINDO Sign Language-to-Speech using Machine Learning Methods," International Conference on Computer Science, Information Technology and Engineering (ICCoSITE), 2023, pp. 986–991. doi: 10.1109/ICCoSITE57641.2023.10127743.

[8] D. P. Kartaputra *et al.*, "Deteksi Alfabet BISINDO Menggunakan Mediapipe Holistic Secara Real-Time," *Jurnal Teknologi Informasi dan Komunikasi*, vol. 12, no. 1, pp. 46–52, Jun. 2023, doi: 10.58761/jurtikstmikbandung.v12.i1.113.

[9] H. M. Putri, Fadlisyah, and W. Fuadi, "Pendeteksian Bahasa Isyarat Indonesia Secara Real-Time Menggunakan Long Short-Term Memory (Lstm)," *Jurnal Teknologi Terapan & Sains*, vol. 3, no. 1, pp. 663–675, 2022, doi: 10.29103/tts.v3i1.6853.

[10] E. Altiarika and W. P. Sari, "Pengembangan Deteksi Realtime untuk Bahasa Isyarat Indonesia dengan Menggunakan Metode Deep Learning Long Short Term Memory dan Convolutional Neural Network," *Jurnal Teknologi Informatika dan Komputer*, vol. 9, no. 1, pp. 1–13, Mar. 2023, doi: https://doi.org/10.37012/jtik.v9i1.1272.

[11] C. A. Sari, E. H. Rachmawanto, Z. Saifullah, C. Jatmoko, and D. Sinaga, "Real-time detection of indonesian sign language (ISL) gestures based on long short-term memory," *Journal of Soft Computing Exploration*, vol. 5, no. 3, pp. 251–262, Sep. 2024, doi: 10.52465/joscex.v5i3.452.

[12] P. B. Utomo, R. A. Ramadhani, and H. Kurniawan, "Deteksi Gerak Tangan sebagai Pengenal Bahasa Isyarat menggunakan Mediapipe dan Long-Short Term Memory," *Jurnal SIMETRIS*, vol. 15, no. 1, pp. 121–136, 2024, doi: 10.24176/simet.v15i1.10505.

[13] A. Bahari and K. E. Dewi, "Peringkasan Teks Otomatis Abstraktif Menggunakan Transformer Pada Teks Bahasa Indonesia," *KOMPUTA : Jurnal Ilmiah Komputer dan Informatika*, vol. 13, no. 1, pp. 83–91, 2024, doi: 10.34010/komputa.v13i1.11197.

[14] M. Julkarnain, E. Mardinata, and R. Susilowati, "Penerapan Transformer-Based Neural Machine Translation untuk Bahasa Bima," *Jurnal Algoritma*, vol. 22, no. 2, pp. 793–802, Nov. 2025, doi: 10.33364/algoritma/v.22-2.2949.

[15] S. Tan, T. Miyazaki, N. Khan, and K. Nakadai, "Improvement in Sign Language Translation Using Text CTC Alignment," in *ACL Anthology*, Abu Dhabi, UAE: International Committee on Computational Linguistics (ICCL), 2025, pp. 3255–3266. Accessed: Dec. 15, 2025. [Online]. Available: https://aclanthology.org/2025.coling-main.219/

[16] B. A. Firmansyah and F. A. Bachtiar, "Automatic Speech Recognition Bahasa Indonesia menggunakan Unidirectional Gated Recurrent Unit," *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 5, no. 12, pp. 5180–5187, 2021, doi: 10.21512/jptiik.v5i12.10200.

[17] I. F. Akbar, Mugiarso, K. F. Ramdhania, and Rasim, "Algoritma Levenshtein Distance sebagai Solusi Efisiensi Pencarian pada Training Registration System," *Journal of Students' Research in Computer Science*, vol. 5, no. 2, pp. 147–154, Nov. 2024, doi: 10.31599/mzr20n84.

[18] D. Bernadisman and R. Lesdiansyah, "Aplikasi Text To Speech Berbasis Android Menggunakan Xml Java," *Jurnal SIBERNETIKA*, vol. 8, no. 2, pp. 62–72, 2023, doi: 10.56459/sibernetika.v8i2.68.

Vol. 10, No. 1, April 2025, Page. 1-250
https://journal.uin-alauddin.ac.id/index.php/instek
p-ISSN: 2541-1179
e-ISSN: 2581-1711
JURNAL INSTEK
INFORMATIKA SAINS DAN TEKNOLOGI

[19] A. Tholib M.Kom, *Implementasi Algoritma Machine Learning Berbasis Web dengan Framework Streamlit*, 1st ed. Pustaka Nurja, 2023. [Online]. Available: https://pustakanurja.unuja.ac.id

[20] N. Trivedi and R. K. Sarvadevabhatla, "PSUMNet: Unified Modality Part Streams are All You Need for Efficient Pose-based Action Recognition," *ArXiv*, Aug. 2022, doi: 10.48550/arXiv.2208.05775.

[21] G. H. Samaan *et al.*, "MediaPipe's Landmarks with RNN for Dynamic Sign Language Recognition," *Electronics (Switzerland)*, vol. 11, no. 19, pp. 1–15, Oct. 2022, doi: 10.3390/electronics11193228.

[22] K. B. Yoga and C. Pramartha, "Pengenalan Tulisan Tangan Aksara Bali Menggunakan Algoritma CRNN," *JNATIA*, vol. 3, no. 4, pp. 845–854, 2025, doi: 10.24843/JNATIA.2025.v03.i04.p14.

[23] E. Ismanto *et al.*, *Revolusi Natural Language Processing: Ketika Manusia Mesin Memahami Bahasa*. PT. Penerbit Qriset Indonesia, 2025.

[24] G. Maulani *et al.*, *Machine Learning*. CV. Mega Press Nusantara, 2025.