



Pengembangan Model CNN-LSTM Berbasis MobileNetV2 untuk Pengenalan Kosakata Bahasa Isyarat Indonesia (BISINDO) Tingkat Kata pada Sistem AIGLASS

Ahmad Riki Andreas¹, Dewi Purnamasari^{*2}, Kurniawati³, Ahmad Bais⁴

Universitas Ivet

ahmadrikiandreas@unisvet.ac.id¹, dewipurnamasari@unisvet.ac.id^{*2}, kurniawati1092@gmail.com³,

ahmdbais0118@gmail.com⁴

Informasi Artikel

Dikirim :26-02-2026

Direview :02-03-2026

Diterbitkan :30-05-2026

Kata Kunci

Bahasa Isyarat
Indonesia, CNN-LSTM,
MobileNetV2,
Pengenalan Bahasa
Isyarat, AIGLASS

Abstrak

Komunikasi antara penyandang tunarungu dan masyarakat umum masih menghadapi hambatan akibat rendahnya pemahaman terhadap Bahasa Isyarat Indonesia (BISINDO). Penelitian ini bertujuan mengembangkan model pengenalan kosakata BISINDO tingkat kata menggunakan arsitektur CNN-LSTM berbasis MobileNetV2 sebagai fondasi kecerdasan buatan pada sistem AIGLASS. Penelitian menggunakan metode CRISP-DM yang meliputi *Business Understanding*, *Data Understanding*, *Data Preparation*, *Modeling*, dan *Evaluation*. Dataset yang digunakan adalah *Word-Level Bahasa Isyarat Indonesia (WL-BISINDO)* yang terdiri atas 1.600 video dengan 32 kosakata. Hasil penelitian menunjukkan model memperoleh *training accuracy* sebesar 99,22%, *validation accuracy* terbaik sebesar 96,48%, serta *accuracy*, *precision*, *recall*, dan *F1-score* pada data pengujian masing-masing sebesar 94,69%, 95,25%, 94,69%, dan 94,66%. Hasil tersebut menunjukkan bahwa model mampu mengenali kosakata BISINDO tingkat kata secara efektif dan berpotensi menjadi fondasi sistem AIGLASS untuk mendukung penerjemahan bahasa isyarat menjadi audio secara real-time.

1. PENDAHULUAN

Komunikasi merupakan kebutuhan *fundamental* yang berperan penting dalam kehidupan sehari-hari. Namun, proses komunikasi tidak selalu berlangsung secara efektif, khususnya penyandang tunarungu yang mengalami gangguan pendengaran sehingga menghambat kemampuan berbahasa, baik secara reseptif (memahami bahasa) maupun ekspresif (menggunakan bahasa) (Damayanti et al., 2025). Penyandang tunarungu sering menghadapi kesulitan saat berinteraksi dengan masyarakat akibat keterbatasan penguasaan bahasa isyarat serta rendahnya pemahaman lingkungan terhadap bahasa isyarat (Hidayat et al., 2025).

Bahasa isyarat merupakan salah satu media komunikasi bagi penyandang tunarungu dalam menyampaikan informasi dan berinteraksi dengan lingkungan sekitarnya (Irchamna et al., 2025). Di Indonesia, bahasa isyarat yang banyak digunakan oleh komunitas tunarungu adalah Bahasa Isyarat Indonesia (BISINDO) (Therry et al., 2024). Berbeda dengan Sistem Isyarat Bahasa Indonesia (SIBI) yang disusun mengikuti struktur bahasa Indonesia formal

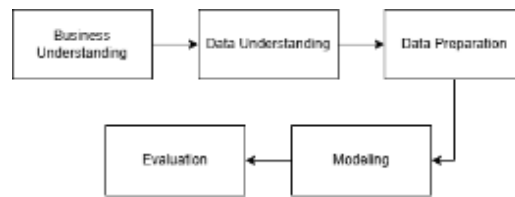
(Ulfah & Ubaidah, 2023), BISINDO berkembang secara alami di komunitas tunarungu dan digunakan sebagai sarana komunikasi sehari-hari (Therry et al., 2024)(Khoeriyah et al., 2026). Akan tetapi, rendahnya pemahaman bahasa isyarat di kalangan masyarakat umum masih menjadi hambatan dalam membangun komunikasi yang efektif (Savitri et al., 2026). Akibatnya, penyandang tunarungu kerap mengalami kesalahpahaman informasi, keterbatasan interaksi sosial, serta kesulitan dalam mengakses berbagai layanan yang memerlukan komunikasi dua arah. Permasalahan ini semakin penting mengingat jumlah penyandang tunarungu di Indonesia masih relatif tinggi. Survei Kesehatan Indonesia (SKI) tahun 2023 mencatat jumlah penyandang tunarungu mencapai sekitar 0,4% dari total populasi Indonesia, dengan prevalensi tertinggi di Provinsi Papua Pegunungan sebesar 1,1% dan Daerah Istimewa Yogyakarta sebesar 0,8% (Santika, 2024). Sementara itu, di Provinsi Jawa Tengah tercatat sebanyak 13.591 penyandang tunarungu (Andanitya & Azizah, 2024). Data tersebut menunjukkan perlunya inovasi teknologi yang mampu menjembatani komunikasi antara penyandang tunarungu dan masyarakat secara efektif.

Perkembangan teknologi digital, berbagai pendekatan telah dikembangkan untuk meningkatkan aksesibilitas komunikasi bagi penyandang tunarungu. serta penerapan *Artificial Intelligence* (AI), khususnya *Deep Learning* membuka peluang pengembangan sistem pengenalan bahasa isyarat secara otomatis. Penelitian oleh Agata dkk. mengembangkan sistem pengenalan Bahasa Isyarat Indonesia (BISINDO) menggunakan kombinasi metode *Scale Invariant Feature Transform* (SIFT) dan *Convolutional Neural Network* (CNN). Hasil penelitian menunjukkan bahwa metode tersebut mampu melakukan klasifikasi alfabet BISINDO dengan tingkat akurasi yang tinggi (Agata et al., 2024). Sementara itu, Lemmuela dkk. menerapkan kombinasi CNN-LSTM untuk pengenalan bahasa isyarat dinamis berbasis video pada 15 kosakata yang berkaitan dengan aktivitas keagamaan. kombinasi CNN-LSTM mampu meningkatkan performa pengenalan gerakan dibandingkan penggunaan CNN secara tunggal sehingga lebih efektif dalam memahami urutan gerakan kata (Lemmuela et al., 2025). Meskipun demikian, penelitian terdahulu masih didominasi pada pengenalan alfabet, *gesture* sederhana, atau jumlah kosakata yang terbatas, sehingga belum mampu merepresentasikan kompleksitas pengenalan BISINDO tingkat kata yang digunakan dalam komunikasi sehari-hari.

Berdasarkan kesenjangan tersebut, penelitian ini memanfaatkan kombinasi *Convolutional Neural Network* (CNN) dan *Long Short-Term Memory* (LSTM) untuk membangun model klasifikasi kosakata BISINDO tingkat kata. Kebaruan penelitian ini terletak pada kombinasi MobileNetV2 sebagai *feature extractor* dengan LSTM untuk mengenali 32 kosakata BISINDO menggunakan dataset video *Word-Level* Bahasa Isyarat Indonesia (WL-BISINDO). Berbeda dengan penelitian terdahulu yang masih berfokus pada alfabet atau kosakata terbatas, model yang dikembangkan dirancang sebagai fondasi kecerdasan buatan pada sistem AIGLASS berbasis *wearable device*. Kontribusi penelitian ini bertujuan membangun dan mengevaluasi performa model dengan mengkombinasikan CNN berbasis MobileNetV2 dengan LSTM dalam mengenali kosakata BISINDO tingkat kata, serta penyediaan fondasi kecerdasan buatan yang berpotensi diintegrasikan pada sistem AIGLASS sebagai perangkat bantu komunikasi bagi penyandang tunarungu.

2. METODOLOGI

Penelitian ini menggunakan metode CRISP-DM (*Cross Industry Standard Process for Data Mining*) yang terdiri dari tahapan *Business Understanding*, *Data Understanding*, *Data Preparation*, *Modeling*, dan *Evaluation* (Rianti et al., 2023). Metode ini dipilih karena menyediakan alur kerja yang sistematis dan banyak digunakan dalam penelitian berbasis Artificial Intelligence dan machine learning (Rohmah et al., 2025). Gambar 1 menunjukkan alur penelitian menggunakan metode CRISP-DM.



Gambar 1. Diagram Alur Penelitian Metode CRISP-DM

2.2 Business Understanding

Tahap *Business Understanding* bertujuan mengidentifikasi permasalahan komunikasi yang dihadapi penyandang tunarungu akibat keterbatasan pemahaman masyarakat terhadap Bahasa Isyarat Indonesia (BISINDO). Berdasarkan permasalahan tersebut, penelitian ini mengembangkan model *Deep Learning* berbasis *Convolutional Neural Network* (CNN) dan *Long Short-Term Memory* (LSTM) untuk mengenali kosakata BISINDO tingkat kata sebagai fondasi kecerdasan buatan pada sistem AIGLASS.

2.2 Data Understanding

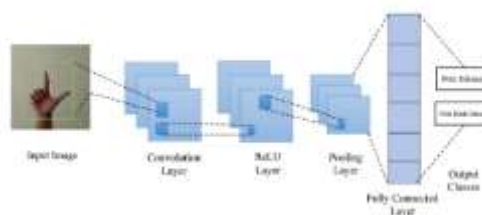
Dimulai dengan menganalisis karakteristik dataset yang digunakan. Dataset penelitian ini berupa *Word-Level* Bahasa Isyarat Indonesia (WL-BISINDO), yaitu dataset video BISINDO yang terdiri dari 1.600 video RGB dengan 32 kosakata tingkat kata yang diperagakan oleh lima *signer* dan disimpan dalam format video MP4 (Kindy et al., 2025). Analisis meliputi jumlah kelas kosakata, distribusi data, format data, variasi gerakan bahasa isyarat pada setiap kelas, serta kemungkinan ketidakseimbangan jumlah data antar kelas yang dapat memengaruhi proses pelatihan model.

2.3 Data Preparation

Dataset berupa video bahasa isyarat dipersiapkan agar sesuai dengan kebutuhan model CNN-LSTM. Setiap video diekstraksi menjadi 20 frame menggunakan uniform sampling, kemudian setiap frame diubah menjadi ukuran 224×224 piksel dan ditransformasikan ke rentang -1 hingga 1 sesuai preprocessing MobileNetV2. Selanjutnya, setiap video diberi label sesuai kosakata BISINDO yang direpresentasikan. Dataset kemudian dibagi menjadi data training (64%), validation (16%), dan testing (20%) dengan *stratified split* sehingga proporsi setiap kelas tetap seimbang pada seluruh subset.

2.4 Modelling

Pada tahap *modeling*, dikembangkan model CNN-LSTM untuk melakukan klasifikasi kosakata BISINDO tingkat kata. *Convolutional Neural Network* (CNN) diimplementasikan menggunakan MobileNetV2 sebagai *feature extractor* untuk mengekstraksi pola visual pada setiap frame video bahasa isyarat, seperti bentuk tangan, posisi jari, dan konfigurasi gesture (Arifuddin et al., 2025). Gambar 2 menunjukkan arsitektur *Convolutional Neural Network* yang digunakan.

Gambar 2. Arsitektur *Convolutional Neural Network*

Fitur yang dihasilkan MobileNetV2 selanjutnya diproses oleh *Long Short-Term Memory* (LSTM) untuk mempelajari hubungan temporal antar frame sehingga model mampu menangkap informasi gerakan secara berurutan (Mahaputra et al., 2025). Kombinasi MobileNetV2 dan LSTM diharapkan mampu meningkatkan akurasi klasifikasi dengan memanfaatkan informasi spasial dan temporal secara bersamaan.

Implementasi model dilakukan menggunakan Python dengan dukungan pustaka *TensorFlow/Keras*, *OpenCV*, dan *NumPy*. Pelatihan dirancang selama 100 *epoch* menggunakan optimizer Adam (*learning rate* 0,001), *batch size* 8, serta fungsi loss *categorical crossentropy*. Untuk menjaga reproduktibilitas, proses pelatihan menggunakan random seed 42 dan dijalankan pada Google Colab dengan GPU NVIDIA A100.

2.5 Evaluation

Evaluasi dilakukan menggunakan data testing yang tidak terlibat selama proses pelatihan maupun validasi untuk mengukur kemampuan generalisasi model. Kinerja model dianalisis menggunakan *Confusion Matrix*, serta metrik *Accuracy*, *Precision*, *Recall*, dan *F1-Score* (Saraswati & Muljono, 2026). Keempat metrik tersebut digunakan untuk mengevaluasi kemampuan klasifikasi model terhadap kosakata BISINDO tingkat kata sekaligus menilai potensinya sebagai fondasi sistem AIGLASS.

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN}$$

$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

3. HASIL DAN PEMBAHASAN

3.1 Dataset BISINDO

Penelitian ini menggunakan dataset *Word-Level* Bahasa Isyarat Indonesia (WL-BISINDO) yang dikembangkan oleh Kindy dkk. dan tersedia di Kaggle (Kindy et al., 2025). Dataset terdiri atas 1.600 video RGB yang merepresentasikan 32 kosakata BISINDO tingkat kata, dengan setiap kosakata diperagakan oleh lima *signer* dan sepuluh sampel video untuk setiap kombinasi signer dan kosakata. Seluruh video disimpan dalam format MP4 dan merepresentasikan satu kosakata bahasa isyarat (*isolated sign*). Dataset ini dipilih karena memiliki distribusi sampel yang seimbang pada setiap kelas serta variasi gerakan dari beberapa *signer*, sehingga sesuai untuk pengembangan dan evaluasi model *Deep Learning*. Contoh frame dari dataset disajikan pada Gambar 3.



Gambar 3. Contoh Frame Dataset WL-BISINDO

Sebelum proses pelatihan, dataset dibagi menjadi *data training*, *validation*, dan *testing* dengan proporsi 64%, 16%, dan 20%, Tabel 1 berikut menunjukkan rincian distribusi dataset.

Tabel 1. Rincian Distribusi Dataset

Dataset	Jumlah Video	Persentase
Training	1024	64%
Validation	256	16%
Testing	320	20%
Total	1600	100%

Distribusi tersebut menjaga jumlah sampel tetap seimbang pada setiap kelas sehingga dapat mengurangi potensi *class imbalance* selama proses pelatihan dan evaluasi model.

3.2 Data Preparation

Dataset video BISINDO dipersiapkan supaya sesuai dengan kebutuhan model CNN-LSTM. Proses yang dilakukan meliputi ekstraksi frame, *image resizing*, normalisasi, ekstraksi fitur, pelabelan data, dan membagi dataset menjadi data *training*, *validation*, dan *testing*. Hasil dalam tahap ini digunakan sebagai masukan pada proses pelatihan model.

3.2.1. Frame Extraction

Setiap video pada dataset WL-BISINDO diekstraksi menjadi 20 frame menggunakan teknik *uniform sampling* agar informasi temporal tetap terwakili dan setiap sampel memiliki jumlah frame yang konsisten. Seluruh frame kemudian diubah ukurannya menjadi 224×224 piksel, sehingga setiap video direpresentasikan dalam bentuk data berdimensi (20, 224, 224, 3). Secara keseluruhan, dataset memiliki dimensi (1600, 20, 224, 224, 3). Gambar 4 menunjukkan Hasil *frame extraction*.



Gambar 4. Ekstraksi Frame dari Video BISINDO

3.2.2. Normalization

Tahap normalisasi dilakukan menggunakan fungsi *preprocess_input()* pada MobileNetV2 yang mengubah nilai piksel ke rentang -1 hingga 1 serta mengonversi data menjadi tipe *float32* agar sesuai dengan masukan model. Proses ini bertujuan meningkatkan stabilitas pembelajaran dan mempercepat konvergensi selama pelatihan.

3.2.3. Feature Extraction

Ekstraksi fitur dilakukan menggunakan MobileNetV2 *pre-trained* sebagai *feature extractor*. Setiap frame berukuran 224×224 piksel diproses tanpa *top layer* sehingga menghasilkan vektor fitur berdimensi 1280. Selanjutnya, setiap video direpresentasikan sebagai matriks fitur berukuran (20, 1280) yang digunakan sebagai masukan model LSTM untuk mempelajari hubungan temporal antar frame dalam proses klasifikasi kosakata.

3.2.4. Label Encoding dan Data Splitting

Setiap sampel video diberi label numerik (*label encoding*) berdasarkan kosakata BISINDO yang direpresentasikan sehingga dapat diproses oleh model klasifikasi. Representasi label bisa dilihat pada Tabel 2.

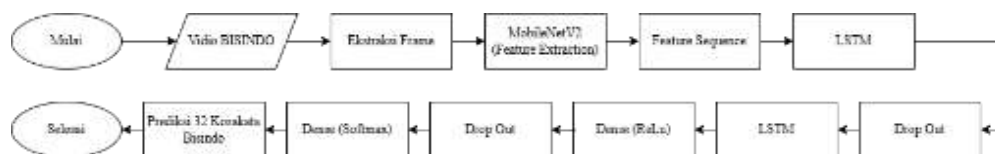
Tabel 2. Representasi Label Kosakata Bisindo

Label	Kosakata	Label	Kosakata	Label	Kosakata	Label	Kosakata
0	Air	8	Motor	16	Mengapa	24	Datang
1	Belajar	9	Saya	17	Bagaimana	25	Teman
2	Cari	10	Terimakasih	18	Merah	26	Keluarga
3	Hari	11	Tuli	19	Kuning	27	Rumah
4	Ingat	12	Apa	20	Hijau	28	Pagi
5	Lagi	13	Siapa	21	Hitam	29	Siang
6	Maaf	14	Kapan	22	Dengar	30	Sore
7	Makan	15	Dimana	23	Berangkat	31	Malam

Dataset selanjutnya dibagi menjadi *data training* (64%), *validation* (16%), dan *testing* (20%) dengan mempertahankan proporsi pada masing-masing kelas sehingga distribusi data pada setiap subset seimbang selama proses pelatihan dan pengujian model.

3.3 Arsitektur Model CNN-LSTM

Model CNN-LSTM dikembangkan untuk mengklasifikasikan kosakata Bahasa Isyarat Indonesia (BISINDO) tingkat kata dengan menggabungkan *Convolutional Neural Network* (CNN) menggunakan arsitektur MobileNetV2 untuk mengekstraksi fitur spasial dan *Long Short-Term Memory* (LSTM) dalam mempelajari hubungan temporal antar frame. MobileNetV2 menghasilkan representasi fitur berdimensi 1280 pada setiap frame sehingga setiap video direpresentasikan sebagai urutan fitur berukuran (20, 1280). Alur Arsitektur model ditunjukkan pada Gambar 5.



Gambar 5. Alur Arsitektur Model CNN-LSTM

Urutan fitur kemudian diproses oleh dua lapisan LSTM dan output LSTM diteruskan ke lapisan *Dense*, *Dropout*, serta *Softmax* untuk mengklasifikasikan 32 kosakata BISINDO. Konfigurasi model disajikan pada Tabel 3.

Tabel 3. Konfigurasi Arsitektur CNN-LSTM

Layer	Konfigurasi	Output
MobileNet V2	<i>Pre-trained, include_top=False</i>	1280 fitur/frame
LSTM 1	<i>256 unit, return_sequences=True</i>	(20,256)
Dropout	0.5	(20,256)
LSTM 2	<i>128 unit</i>	(128)
Dropout	0.5	(128)
Dense	<i>128 neuron, ReLU</i>	(128)
Dropout	0.5	(128)
Dense Output	<i>32 neuron, Softmax</i>	(32)

3.4 Hasil Pelatihan Model

Proses pelatihan model CNN-LSTM dilakukan menggunakan 1.024 *data training* dan 256 *data validation*. Model dilatih selama 100 *epoch* menggunakan *optimizer* Adam, *learning rate* 0,001, *batch size* 8, dan fungsi *loss* *Categorical Crossentropy*. Selama pelatihan diterapkan *ModelCheckpoint* untuk menyimpan model dengan nilai *validation accuracy* terbaik. Konfigurasi pelatihan ditunjukkan pada Tabel 4.

Tabel 4. Konfigurasi Pelatihan Model CNN-LSTM

Parameter	Nilai
<i>Epoch</i>	100
<i>Batch Size</i>	8
<i>Learning Rate</i>	0,001
<i>Optimizer</i>	Adam
<i>Loss Function</i>	<i>Categorical Crossentropy</i>
<i>Input Frame Size</i>	224×224 piksel
<i>Feature Dimension</i>	1280

Performa model dipantau pada setiap epoch menggunakan metrik *accuracy* dan *loss* pada *data training* maupun *validation*. Monitoring kedua metrik bertujuan untuk mengevaluasi proses konvergensi model, mengamati perkembangan kemampuan model dalam mempelajari pola bahasa isyarat, serta mendeteksi terjadinya *overfitting*.

Gambar 6. Grafik *Training Accuracy* dan *Validation Accuracy*Gambar 7. Grafik *Training Loss* dan *Validation Loss*

Gambar 6 dan Gambar 7 menunjukkan perkembangan *accuracy* dan *loss* selama proses pelatihan. *Training accuracy* meningkat hingga 99,22%, sedangkan *validation accuracy*

terbaik mencapai 96,48% pada *epoch* ke-57. Selisih nilai *training accuracy* dan *validation accuracy* yang bedekatan berdekatan menunjukkan model mampu melakukan generalisasi dengan baik tanpa indikasi *overfitting* yang signifikan. Pada saat yang sama, *training loss* menurun dari 3,46 menjadi 0,03, sedangkan *validation loss* terbaik mencapai 0,14. Pola tersebut mengindikasikan bahwa model dapat mempelajari karakteristik data secara baik.

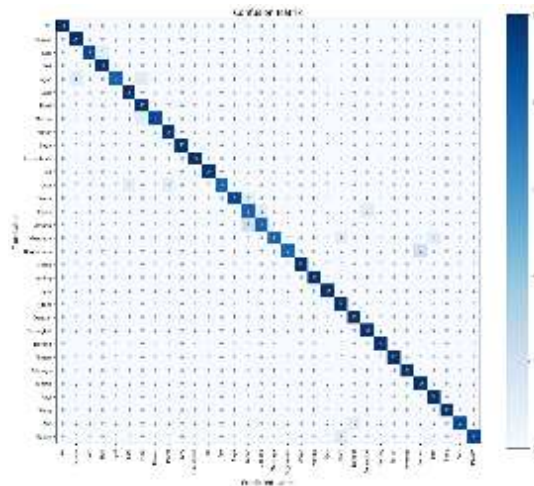
3.5 Evaluasi Performa Model

Evaluasi performa dilakukan menggunakan 320 data testing yang tidak digunakan selama proses pelatihan maupun validasi. Pengukuran performa menggunakan metrik *accuracy*, *precision*, *recall*, dan *F1-score* terhadap 32 kosakata BISINDO tingkat kata.

Tabel 5. Hasil Evaluasi Model CNN-LSTM

<i>Accuracy</i>	<i>Precision</i>	<i>Recall</i>	<i>F1-Score</i>
94,69%	95,25%	94,69%	94,66%

Berdasarkan hasil evaluasi pada Tabel 5, model CNN-LSTM memperoleh nilai *accuracy* sebesar 94,69%, *precision* sebesar 95,25%, *recall* sebesar 94,69%, dan *F1-score* sebesar 94,66% yang menunjukkan kemampuan klasifikasi kosakata BISINDO dengan tingkat ketepatan yang tinggi.



Gambar 8. Confusion Matrix Model CNN-LSTM

Confusion matrix pada Gambar 8 menunjukkan sebagian besar prediksi berada pada diagonal utama, yang menandakan mayoritas sampel berhasil diklasifikasikan dengan benar. Beberapa kelas, Ingat, Apa, Dimana, dan Kapan masih memiliki nilai *recall* 80%, sedangkan sebagian besar kelas memperoleh *precision* dan *recall* di atas 90%. Beberapa kosakata, seperti Air, Hijau, Datang, Teman, Keluarga, dan Siang berhasil dikenali dengan *precision* dan *recall* sebesar 100%. Berdasarkan hasil evaluasi, menunjukkan model CNN-LSTM dapat mengidentifikasi kosakata BISINDO tingkat kata secara efektif.

3.6 Potensi Integrasi Model CNN-LSTM pada Sistem AIGLASS

Model CNN-LSTM berbasis MobileNetV2 yang dikembangkan pada penelitian ini memiliki potensi sebagai dasar kecerdasan buatan pada sistem AIGLASS. Model ini mampu mengenali kosakata BISINDO tingkat kata berdasarkan urutan frame video sehingga dapat diarahkan untuk mendukung proses penerjemahan bahasa isyarat menjadi audio pada perangkat bantu komunikasi tunarungu.



Gambar 9. Diagram alur Sistem AIGLASS

Gambar 9 memperlihatkan alur konseptual integrasi model CNN-LSTM pada sistem AIGLASS. kamera menangkap gerakan bahasa isyarat, kemudian data visual diproses menggunakan model CNN-LSTM berbasis MobileNetV2 untuk mengenali kosakata BISINDO. Hasil prediksi selanjutnya diteruskan ke modul *text-to-speech* sehingga dapat dikonversi menjadi audio. Berdasarkan hasil evaluasi, model berpotensi sebagai fondasi kecerdasan buatan pada sistem AIGLASS. Namun, integrasi pada penelitian ini masih berada pada tahap konseptual, pengujian langsung pada perangkat *wearable* perlu dilakukan pada penelitian selanjutnya untuk mengevaluasi performa model pada kondisi nyata.

3.7 Pembahasan

Tabel 6. Perbandingan Hasil Penelitian dengan Penelitian Terdahulu

Peneliti	Metode	Jumlah Kelas	Jenis Data	Accuracy
Agata dkk.	CNN+SIFT	26 Alfabet	Citra Statis	99,78%
Lemmuela dkk.	CNN+LSTM	15 Kata	Video	95,24%
Penelitian ini	MobileNetV2+LSTM	32 Kata	Video	94,69%

Berdasarkan perbandingan pada Tabel 6, penelitian ini memperoleh *accuracy* sebesar 94,69% menggunakan arsitektur MobileNetV2-LSTM pada dataset video 32 kosakata BISINDO. Meskipun nilai *accuracy* sedikit lebih rendah dibandingkan penelitian Agata dkk. dan Lemmuela dkk., penelitian ini melakukan klasifikasi terhadap 32 kosakata dibandingkan 15 kosakata sehingga tingkat kompleksitas klasifikasi menjadi lebih tinggi.

Hasil *classification report* menunjukkan sebagian besar kelas memperoleh nilai *precision* dan *recall* di atas 90%. Beberapa kosakata seperti Air, Hijau, Datang, Teman, Keluarga, dan Siang berhasil dikenali dengan nilai *precision* dan *recall* sebesar 100%, sedangkan kosakata Ingat, Apa, Dimana, dan Kapan masih memiliki *recall* sebesar 80% akibat kemiripan pola gerakan dan variasi *gesture* antar *signer*. Secara keseluruhan, kombinasi MobileNetV2 sebagai ekstraktor fitur spasial dan LSTM sebagai pemodel hubungan temporal mampu mengenali kosakata BISINDO tingkat kata dengan baik, sehingga berpotensi menjadi fondasi kecerdasan buatan pada sistem AIGLASS.

4. KESIMPULAN

Penelitian ini berhasil mengembangkan model pengenalan kosakata Bahasa Isyarat Indonesia (BISINDO) tingkat kata menggunakan arsitektur CNN-LSTM dengan MobileNetV2 sebagai *feature extractor*. Model dikembangkan menggunakan dataset Word-Level BISINDO (WL-BISINDO) yang terdiri atas 1.600 video dengan 32 kosakata. Tahap preprocessing mencakup ekstraksi 20 frame menggunakan teknik *uniform sampling*, *image resizing* menjadi 224×224 piksel, normalisasi nilai piksel, ekstraksi fitur menggunakan MobileNetV2, serta pelabelan data sebelum proses pelatihan.

Hasil penelitian menunjukkan bahwa model memperoleh *training accuracy* sebesar 99,22%, *validation accuracy* terbaik sebesar 96,48%, serta *accuracy*, *precision*, *recall*, dan *F1-score* pada *data testing* masing-masing sebesar 94,69%, 95,25%, 94,69%, dan 94,66%. Hasil tersebut menunjukkan bahwa kombinasi MobileNetV2 dan LSTM mampu mengenali kosakata BISINDO tingkat kata secara efektif. Model yang dikembangkan berpotensi menjadi fondasi kecerdasan buatan pada sistem AIGLASS sebagai perangkat bantu komunikasi berbasis *wearable*. Pengembangan selanjutnya dapat dilakukan melalui penambahan variasi dataset, pengujian pada kondisi nyata, serta implementasi model secara langsung pada perangkat *wearable*.

DAFTAR PUSTAKA

- Agata, A. W., Saputra, W. S. J., & Putra, C. A. (2024). Pengenalan Bahasa Isyarat Indonesia (Bisindo) Menggunakan Algoritma Scale Invariant Feature Transform (SIFT) Dan Convolutional Neural Network (CNN). In *Jurnal Mahasiswa Teknik Informatika* (Vol. 8, Number 1). <https://doi.org/https://doi.org/10.36040/jati.v8i1.8917>
- Andanitya, A. P., & Azizah, R. (2024). *Deaf Culture Center Sebagai Pusat Pelatihan Kreativitas Dan Terapi Untuk Penyandang Tunarungu Kota Surakarta Dengan Pendekatan Deaf Space Design Guide (DSDG)*. <http://eprints.ums.ac.id/id/eprint/128523>
- Arifuddin, N. A., Capri, H., Setiawan, D., Amalia, R. D., & Gusti, K. W. (2025). *Optimizing Road Safety with MobileNet-Based Classification of Over-dimensioned Trucks*. 10(2), 2477–5126. <https://doi.org/10.30591/jpit.v9ix.xxx>
- Damayanti, M. M., Putri, K., & Yolla, Y. (2025). Hambatan yang Dihadapi Anak Tunarungu dalam Proses Pembelajaran Anak Tunarungu. In *JOECES Journal of Early Childhood Education Studies* (Vol. 2, Number 2). <https://doi.org/https://doi.org/10.54180/joeces.v5i1.481>
- Hidayat, N., Fernanda, Y., & Uman, K. (2025). Dampak Keterlambatan Bicara dan Kurangnya Penguasaan Bahasa Isyarat Pada Anak Tuna Rungu. *Jurnal Studi Pendidikan Dasar*, 3. <https://doi.org/https://doi.org/10.54180/jsped.v3i1.550>
- Irchamna, A. M., Amira, V. N. F., Oktavianti, N., & Mintowati, M. (2025). *Pengaruh Bahasa Isyarat Terhadap Pembelajaran Komunikasi Lisan Tunaganda (Tunarungu dan Tunawicara)*. 5(1), 40–49. <https://doi.org/http://dx.doi.org/10.30821/eunoia.v5i1.4200>
- Khoeriyah, I. M., Winarsih, M., & Kasirah, I. (2026). *Persepsi Guru terhadap Penggunaan SIBI dan BISINDO di SLB*. <https://doi.org/https://doi.org/10.21009/jpk.v3i2.58518>
- Kindy, G. O., Leonali, G., & Lucky, H. (2025). Word-Level BISINDO: A Novel Video Indonesian Sign Language Dataset and Baseline Methods. *Procedia Computer Science*, 269, 249–258. <https://doi.org/10.1016/j.procs.2025.08.277>
- Lemmuela, I. V., Ayub, M., & Karnalim, O. (2025). Dynamic Sign Language Recognition in Bahasa using MediaPipe, Long Short-Term Memory, and Convolutional Neural Network. *Journal of Information Systems Engineering and Business Intelligence*, 11(1), 17–29. <https://doi.org/10.20473/jisebi.11.1.17-29>
- Mahaputra, I. G. A. M. Y., Santiary, P. A. W., & Sawardika, I. K. (2025). Analisis Temporal Gerakan Kata BISINDO Menggunakan Landmark Tangan dan LSTM dengan Keluaran Suara Berbasis ESP32 Secara Real-time., 18(2). <https://journal.stekom.ac.id/index.php/elkom>
- Rianti, A., Majid, N. W. A., & Fauzi, A. (2023). *CRISP-DM: Metodologi Proyek Data Science*. <https://ojs.uib.ac.id/index.php/Senatib/article/view/3015>
- Rohmah, T. N., Purnamasari, D., Kurniawati, Herlinudinkhaji, D., Arif, Y. M., & Criollo-C, S. (2025). *Deepfake Image Detection Using Transfer Learning Method*. <https://doi.org/https://doi.org/10.18860/ijeie.v1i2.40796>

- Santika, E. F. (2024, July 16). *Ini Prevalensi Disabilitas Indonesia pada 2023*. <https://Databoks.Katadata.Co.Id/Layanan-Konsumen-Kesehatan/Statistik/66a13e9eb02df/Ini-Prevalensi-Disabilitas-Indonesia-Pada-2023>.
- Saraswati, E. R. A., & Muljono. (2026). Deteksi Emosi Pada Twitter Berbasis Fasttext: Evaluasi Performa Arsitektur Cnn Dan Gru. *Rabit: Jurnal Teknologi Dan Sistem Informasi Univrab*, 11(1), 1175–1186. <https://doi.org/10.36341/rabit.v11i1.7252>
- Savitri, D. I., Bua, M. T., Adiasti, N., Mas'an, S. M. Al, & Rhamdan, D. (2026). *Menjembatani Sunyi: Bahasa Isyarat Sebagai Upaya Membangun Komunikasi Inklusif Penyandang Tuli*. 8(1). <https://doi.org/10.70095/dimasejati.51.000>
- Therry, R. W. L., Junaidi, A., & Sihananto, A. N. (2024). Program Penerjemah Bahasa Isyarat Indonesia (Bisindo) Secara Real Time Menggunakan Convolutional Neural Network Dan Mediapipe. In *Jurnal Mahasiswa Teknik Informatika* (Vol. 8, Number 6). <https://doi.org/https://doi.org/10.36040/jati.v8i6.11582>
- Ulfah, S. M., & Ubaidah, S. (2023). Penerapan Bahasa Isyarat dalam Pembelajaran bagi Anak Berkebutuhan Khusus Tuna Rungu. In *Journal Of Disability Studies And Research* (Vol. 2023, Number 1). JDSR. <https://doi.org/DOI:https://doi.org/10.30631/jdsr.v2i1.1764>