
IMPLEMENTASI YOLOv11 UNTUK DETEKSI KATA BAHASA ISYARAT BISINDO DAN SIBI

Ahmad Syaefudin, Hidayatus Sibyan, Rina Mahmudati, M. Fuat Asnawi, Nur Hasanah
Teknik Informatika, Fakultas Teknik dan Ilmu Komputer, Universitas Sains Al Qur'an
Email : saefudin.ua@gmail.com

ABSTRAK

Komunikasi merupakan hak dasar manusia, namun penyandang disabilitas rungu sering menghadapi hambatan dalam berinteraksi dengan masyarakat umum yang minim pemahaman bahasa isyarat. Penelitian ini bertujuan untuk mengimplementasikan dan mengevaluasi kinerja algoritma YOLOv11 dalam mendeteksi kata pada Bahasa Isyarat Indonesia (BISINDO) dan Sistem Isyarat Bahasa Indonesia (SIBI) secara real-time. Metode yang digunakan meliputi pengumpulan dataset sebanyak 1.745 citra yang mencakup 12 kelas kata, preprocessing menggunakan Roboflow untuk anotasi dan augmentasi, serta pelatihan model menggunakan Google Colab. Hasil penelitian menunjukkan bahwa model YOLOv11 mampu mendeteksi bahasa isyarat dengan sangat baik, mencapai nilai mean Average Precision (mAP@50) sebesar 98%, Precision 96.8%, dan Recall 96.4% pada tahap validasi. Implementasi sistem dilakukan berbasis web menggunakan framework Flask, memungkinkan deteksi interaktif melalui kamera. Secara keseluruhan, penelitian ini menunjukkan bahwa penerapan model YOLOv11 efektif dalam mendukung penerjemahan bahasa isyarat secara real-time. Implikasi dari penelitian ini adalah meningkatnya aksesibilitas komunikasi bagi penyandang disabilitas rungu, khususnya dalam interaksi sehari-hari dengan masyarakat umum, serta membuka peluang pengembangan sistem penerjemah bahasa isyarat yang lebih luas, adaptif, dan terintegrasi pada berbagai platform digital.

Kata Kunci : Bahasa Isyarat, BISINDO, SIBI, YOLOv11, Deteksi Objek.

ABSTRACT

Communication is a fundamental human right, yet deaf individuals often face barriers interacting with the general public due to a lack of sign language understanding. This study aims to implement and evaluate the performance of the YOLOv11 algorithm in detecting words in Indonesian Sign Language (BISINDO) and the Indonesian Sign Language System (SIBI) in real-time. The methods included collecting a dataset of 1,745 images covering 12 word classes, preprocessing using Roboflow for annotation and augmentation, and training the model using Google Colab. The results showed that the YOLOv11 model detected sign language effectively, achieving a mean Average Precision (mAP@50) of 98%, Precision of 96.8%, and Recall of 96.4% during validation. The system was implemented as a web-based application using the Flask framework, allowing interactive detection via camera.

Keywords : Sign Language, BISINDO, SIBI, YOLOv11, Object Detection.

1. PENDAHULUAN

Bahasa memegang peranan krusial dalam kehidupan manusia sebagai alat komunikasi yang efektif untuk menyampaikan maksud dan tujuan. Dalam situasi formal maupun informal, keberhasilan komunikasi ditentukan oleh kemampuan bahasa dalam menyampaikan pesan sesuai tujuan penutur (Mailani dkk., 2022). Namun, proses ini menghadapi hambatan signifikan bagi penyandang disabilitas sensorik, khususnya tunarungu. Berdasarkan data Badan Pusat Statistik, jumlah penduduk Indonesia yang mengalami gangguan pendengaran total mencapai 225.161 jiwa. Bagi komunitas ini, bahasa isyarat menjadi modalitas utama yang bersifat visual-gestural, berbeda dengan bahasa lisan yang bersifat oral-auditoris (Isma, 2018).

Di Indonesia, terdapat dua sistem isyarat yang dominan, yaitu Bahasa Isyarat Indonesia (BISINDO) dan Sistem Isyarat Bahasa Indonesia (SIBI). Gumelar dkk. (2018) menjelaskan bahwa BISINDO merupakan bahasa isyarat alami budaya asli Indonesia yang tumbuh secara natural di komunitas Tuli, sehingga memiliki variasi unik di setiap daerah. Sebaliknya, SIBI adalah sistem yang dibakukan oleh pemerintah sebagai standar pengantar di Sekolah Luar Biasa (SLB), yang strukturnya merepresentasikan tata bahasa Indonesia lisan (Utami, 2016). Perbedaan karakteristik keduanya menimbulkan tantangan tersendiri dalam pengembangan sistem deteksi bahasa isyarat yang bersifat umum dan adaptif.

Kesenjangan komunikasi antara masyarakat dengar dan teman Tuli sering kali menimbulkan dampak psikososial. Octaviani (2020) dalam penelitiannya menemukan bahwa keterbatasan komunikasi membuat penyandang tunarungu merasa rendah diri dalam interaksi sosial, sementara keluarga sering merasa khawatir akan terjadinya diskriminasi. Selain itu, rendahnya literasi masyarakat terhadap bahasa isyarat juga disebabkan oleh keterbatasan media pembelajaran dan distribusi informasi yang belum merata (Farah dkk., 2025). Hal ini menunjukkan perlunya solusi teknologi yang tidak hanya akurat, tetapi juga mudah diakses oleh masyarakat luas.

Seiring kemajuan teknologi, kecerdasan buatan atau *Artificial Intelligence* (AI) telah banyak dimanfaatkan untuk mengatasi hambatan ini. Penelitian terdahulu banyak

menerapkan algoritma *Convolutional Neural Network* (CNN). Namun, Alfikri dkk. (2022) mencatat kelemahan metode ini, di mana akurasi model mencapai 97% saat *training*, "tetapi pada saat melakukan *deployment* akurasi menurun hingga 73% khususnya dalam memprediksi bahasa isyarat huruf L". Hal ini menunjukkan bahwa CNN standar kurang optimal untuk kondisi nyata yang dinamis. Alternatif lain seperti *Faster R-CNN* memiliki akurasi tinggi namun lambat (12 FPS), sedangkan SSD (*Single Shot MultiBox Detector*) lebih cepat namun akurasi menurun pada objek kecil (Li dkk., 2020). Keterbatasan ini menunjukkan adanya kebutuhan akan metode yang mampu menyeimbangkan akurasi dan kecepatan dalam skenario real-time.

Sebagai solusi yang lebih efisien, algoritma YOLO (*You Only Look Once*) menawarkan pendekatan *single-stage detection* yang menyeimbangkan kecepatan dan akurasi. YOLOv11 menghadirkan peningkatan pada arsitektur dan proses pelatihan sehingga mampu menghasilkan performa yang lebih stabil dan cepat dibandingkan generasi sebelumnya (Jocher dkk., 2025; Jegham dkk., 2024). Keunggulan ini menjadikan YOLOv11 berpotensi kuat untuk diterapkan dalam sistem deteksi bahasa isyarat berbasis real-time.

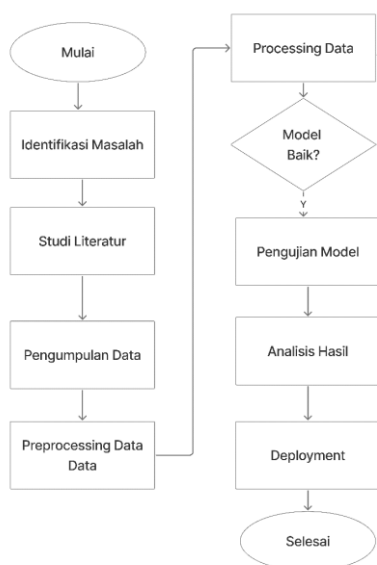
Penelitian ini mengusulkan pendekatan deteksi bahasa isyarat dengan memanfaatkan algoritma YOLOv11 yang diterapkan pada dua sistem yang digunakan di Indonesia, yaitu BISINDO dan SIBI. Berbeda dengan penelitian sebelumnya yang umumnya berfokus pada pengenalan huruf, pendekatan ini diarahkan pada deteksi berbasis kata agar lebih kontekstual dalam mendukung komunikasi sehari-hari. Sistem yang dikembangkan juga dirancang berbasis web dan berjalan secara real-time, sehingga memberikan kemudahan akses serta interaksi yang lebih praktis bagi pengguna.

Berdasarkan hal tersebut, penelitian ini bertujuan untuk mengimplementasikan dan mengevaluasi kinerja model YOLOv11 dalam mendeteksi kata bahasa isyarat menggunakan metrik precision, recall, dan mean Average Precision (mAP), serta mengembangkan sistem deteksi yang dapat digunakan secara real-time melalui kamera. Hasil yang diperoleh diharapkan dapat meningkatkan aksesibilitas

komunikasi bagi penyandang tunarungu dan mendukung pemanfaatan kecerdasan buatan dalam konteks inklusi sosial.

2. METODE

Penelitian ini merupakan penelitian rekayasa (*engineering research*) dengan pendekatan kuantitatif eksperimental yang berfokus pada pengembangan dan pengujian kinerja sistem deteksi objek. Tahapan penelitian meliputi pengumpulan data, pra-pemrosesan, pelatihan model, dan evaluasi kinerja seperti terlihat pada Gambar 1.



Gambar 1. Alur Penelitian

2.1. Pengumpulan Data

Pengumpulan Data Objek penelitian ini adalah citra gerakan tangan yang merepresentasikan kata dalam Bahasa Isyarat Indonesia (BISINDO) dan Sistem Isyarat Bahasa Indonesia (SIBI). Pengumpulan data dilakukan melalui tiga metode utama: studi literatur, wawancara dengan guru bahasa isyarat, dan observasi langsung di SLB Negeri Banjarnegara. Dataset yang dikumpulkan terdiri dari 1.745 citra yang terbagi ke dalam 12 kelas kata, yaitu: *bantu*, *tidur*, *saya*, *rumah*, *mau*, *ibu*, *bapak*, *maaf*, *tolong*, *makan*, *kamu*, dan *sehat*. Pemilihan kata pada Tabel 1 didasarkan pada kesederhanaan pola gerakan yang dapat dideteksi secara visual oleh algoritma.

Tabel 1. Pemilihan Kata

Kata	BISINDO	SIBI	Keterangan
Bantu	✓		
Tidur	✓	✓	Sama
Saya		✓	
Rumah	✓	✓	Sama
Mau	✓		
Ibu	✓	✓	Sama
Bapak	✓	✓	Sama
Maaf	✓	✓	Sama
Tolong	✓		
Makan	✓	✓	Sama
Kamu	✓		
Sehat	✓	✓	Sama

Dataset melibatkan beberapa subjek/peraga dengan variasi kondisi pengambilan data, meliputi variasi pencahayaan (indoor dan semi-outdoor), latar belakang (kelas, ruang terbuka, dan latar sederhana), jarak kamera ($\pm 0,5-1,5$ meter), serta sudut pengambilan yang beragam. Variasi ini bertujuan untuk meningkatkan kemampuan generalisasi model terhadap kondisi nyata.

Distribusi data antara BISINDO dan SIBI dianalisis pada setiap kelas. Beberapa kata memiliki representasi pada kedua sistem (misalnya: tidur, rumah, ibu, bapak), sementara sebagian lainnya hanya terdapat pada salah satu sistem. Ketidakseimbangan jumlah data antar kelas ditangani melalui teknik augmentasi pada tahap preprocessing.

Selain itu, pemilihan 12 kata juga mempertimbangkan frekuensi penggunaan dalam komunikasi sehari-hari, representasi jenis kata (kata kerja, kata benda, dan kata ganti), serta ketersediaan padanan dalam BISINDO dan/atau SIBI, sehingga lebih relevan dalam konteks komunikasi dasar.

2.2. Preprocessing Data

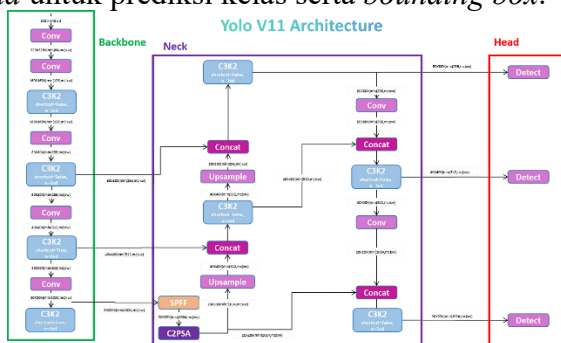
Proses *Preprocessing* dilakukan menggunakan platform Roboflow. Tahapan ini mencakup: (1) Ekstraksi frame video menjadi gambar statis; (2) Anotasi data (*labeling*) dengan membuat *bounding box* pada area tangan; (3) Normalisasi ukuran citra (*resizing*) menjadi 640x640 piksel agar sesuai dengan *input layer* model; dan (4) Penyesuaian kontras (*histogram equalization*) untuk variasi pencahayaan. Untuk meningkatkan keragaman data dan mencegah *overfitting*, dilakukan augmentasi data yang meliputi rotasi (*rotation*)

± 6 derajat, distorsi (*shear*) ± 6 derajat, dan penyesuaian kecerahan (*brightness*) $\pm 17\%$. Dataset kemudian dibagi secara otomatis menjadi tiga subset: 70% untuk data pelatihan (*training*), 20% untuk validasi (*validation*), dan 10% untuk pengujian (*testing*).

Selain pembagian tersebut, penelitian ini juga menerapkan k-fold cross validation ($k=5$). Dataset dibagi menjadi 5 fold, di mana pada setiap iterasi 4 fold digunakan untuk pelatihan dan 1 fold untuk validasi. Untuk menghindari data leakage, pembagian data dilakukan berdasarkan subjek (*subject-wise split*), sehingga data dari peraga yang sama tidak muncul pada data pelatihan dan pengujian/validasi secara bersamaan.

2.3. Training Model

Pengembangan dan Pelatihan Model Algoritma yang digunakan adalah YOLOv11 (*You Only Look Once* versi 11) dengan model dasar `yolo11n.pt`. Pelatihan model dilakukan menggunakan lingkungan komputasi Google Colab yang didukung oleh GPU T4. Konfigurasi *hyperparameter* yang diterapkan selama proses pelatihan adalah 100 *epochs*, *batch size* 16, dan *optimizer* diatur secara otomatis (*auto*). Arsitektur model terlihat seperti pada Gambar 2 yang terdiri dari *backbone* untuk ekstraksi fitur, *neck* untuk penggabungan fitur multi-skala, dan *head* untuk prediksi kelas serta *bounding box*.



Gambar 2. Arsitektur YOLOv11

Proses pelatihan dilakukan pada setiap fold dalam skema cross-validation, sehingga performa model tidak bergantung pada satu pembagian data saja.

2.4. Evaluasi Model

Evaluasi kinerja model dilakukan menggunakan metrik standar deteksi objek yang dihitung berdasarkan *Confusion Matrix*, yaitu *True Positive* (TP), *False Positive* (FP), dan *False Negative* (FN). Metrik utama yang digunakan meliputi *Precision* (tingkat ketepatan), *Recall* (sensitivitas), *F1-Score* (rata-

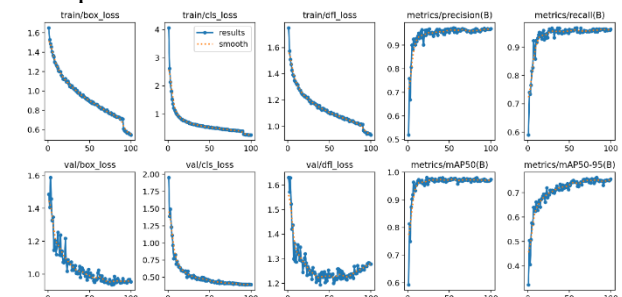
rata harmonis presisi dan recall), serta *mean Average Precision* (mAP) pada *threshold* 0.5 (*mAP@50*) dan rentang 0.5-0.95 (*mAP@50-95*). Model terbaik hasil pelatihan kemudian diekspor ke format ONNX untuk diimplementasikan ke dalam aplikasi berbasis web menggunakan *framework* Flask.

Nilai evaluasi yang dilaporkan merupakan rata-rata hasil dari seluruh fold pada cross-validation untuk meningkatkan reliabilitas dan mengurangi bias evaluasi.

3. HASIL DAN PEMBAHASAN

3.1. Analisis Proses Training

Analisis Proses Pelatihan (Training) Proses pelatihan model YOLOv11 dilakukan selama 100 epoch. Berdasarkan visualisasi grafik pelatihan pada Gambar 3, terlihat tren penurunan yang konsisten pada nilai Box Loss, Class Loss, dan DFL Loss baik pada data latih maupun validasi. Tidak adanya divergensi yang signifikan antara grafik training dan validation mengindikasikan bahwa model memiliki kemampuan generalisasi yang baik dan tidak mengalami overfitting. Stabilitas ini penting untuk memastikan model dapat bekerja dengan baik pada data baru.



Gambar 3. Proses Training Model

3.2. Analisis Evaluasi Model

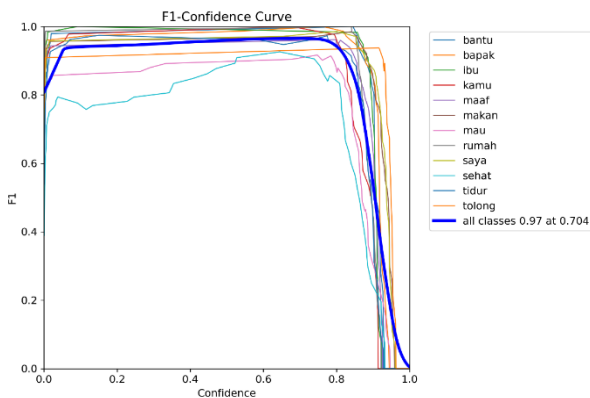
Berdasarkan pengujian terhadap data validasi pada Tabel 2, model menunjukkan performa yang sangat tinggi. Secara keseluruhan, model mencapai rata-rata Precision 96.8%, Recall 96.4%, dan *mAP@50* sebesar 98%. Namun, ketika ambang batas IoU diperketat (*mAP@50-95*), skor rata-rata turun menjadi 76.3%. Hal ini wajar dalam deteksi objek, menandakan bahwa meskipun model sangat akurat dalam mengklasifikasikan jenis isyarat, presisi kotak pembatas (*bounding box*) terhadap objek asli masih memiliki variasi ketepatan posisi.

Tabel 2. Hasil Validasi Model

Class	Images	Instances	Precision (P)	Recall (R)	mAP50	mAP50-95
all	349	360	0.968	0.964	0.98	0.763
bantu	24	24	0.993	1	0.995	0.721
bapak	36	36	0.993	1	0.995	0.797
ibu	30	30	1	0.973	0.995	0.858
kamu	22	22	1	0.989	0.995	0.755
maaf	26	26	0.957	0.962	0.971	0.741
makan	34	34	0.969	0.971	0.981	0.783
mau	36	36	0.901	0.917	0.927	0.638
rumah	36	36	0.996	1	0.995	0.764
saya	34	34	0.954	1	0.982	0.875
sehat	17	28	0.961	0.871	0.947	0.628
tidur	38	38	0.965	0.947	0.993	0.652
tolong	16	16	0.925	0.938	0.981	0.944

Analisis per kelas menunjukkan disparitas performa. Kelas dengan gestur tangan yang distingtif seperti "Tolong", "Saya", dan "Ibu" memiliki mAP@50-95 tertinggi (>0.85). Sebaliknya, kelas "Sehat" dan "Mau" memiliki skor terendah (sekitar 0.62–0.63). Rendahnya performa pada kelas ini disebabkan oleh kompleksitas visual dan kemiripan pola gerakan dengan gestur lain atau latar belakang.

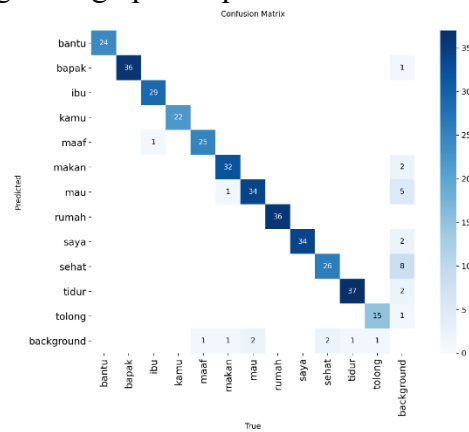
Analisis Kurva Performa dan Penentuan Threshold untuk menentukan titik kerja optimal model, dilakukan analisis terhadap kurva F1-Confidence. Gambar 4 menunjukkan bahwa nilai F1-Score tertinggi (0.97) dicapai pada confidence threshold sebesar 0.704. Titik ini merepresentasikan keseimbangan terbaik antara precision dan recall, di mana model mampu meminimalkan false positive tanpa kehilangan terlalu banyak true positive. Angka 0.704 ini kemudian direkomendasikan sebagai batas minimal confidence saat implementasi sistem. Selain itu, kurva Precision-Recall menunjukkan area di bawah kurva yang sangat luas (mAP 0.980), menegaskan reliabilitas model yang tinggi.



Gambar 4. F1-Score Confidence Curve

Analisis Kesalahan (Confusion Matrix) Evaluasi menggunakan Confusion Matrix pada

Gambar 5 memperlihatkan bahwa model sangat dominan dalam memprediksi kelas dengan benar. Namun, matriks juga mengungkap kelemahan spesifik model dalam membedakan objek isyarat dari latar belakang (background). Kesalahan terbesar terjadi pada kelas "Sehat" (8 kejadian) dan "Mau" (5 kejadian) yang salah diklasifikasikan sebagai background, serta sebaliknya. Hal ini menunjukkan bahwa pada kondisi tertentu, fitur tangan pada kelas tersebut gagal diekstraksi dengan sempurna oleh model, kemungkinan akibat pencahayaan atau kontras yang kurang optimal pada data validasi.

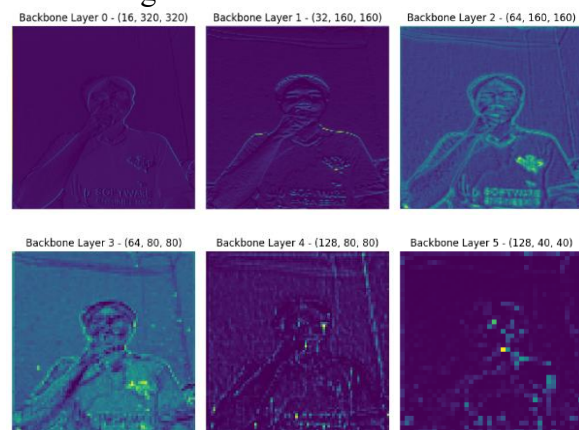


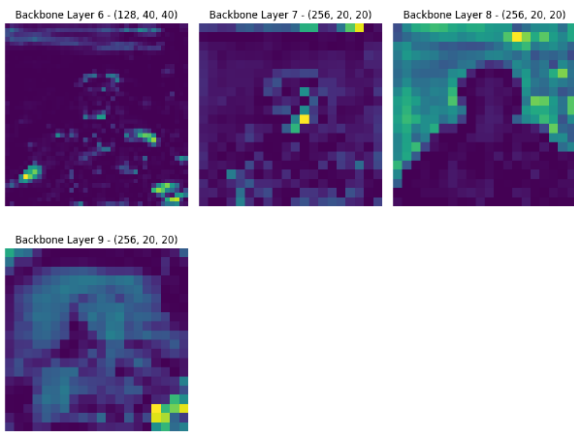
Gambar 5. Confusion Matrix

3.3 Visualisasi Aktivasi Layer

Visualisasi Aktivasi Layer (Model Explainability) Untuk memahami bagaimana YOLOv11 memproses citra, dilakukan analisis heatmap pada lapisan Backbone, Neck, dan Head.

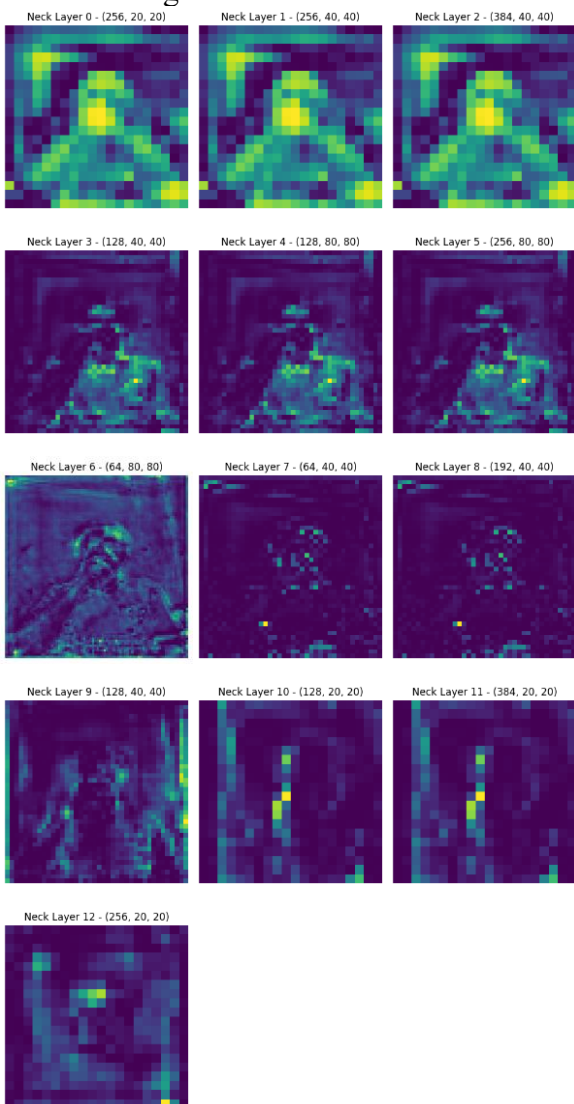
Pada tahap awal ini, visualisasi backbone pada Gambar 6 menunjukkan bahwa model fokus mengekstraksi fitur dasar seperti tepi dan tekstur tangan.





Gambar 6. Backbone

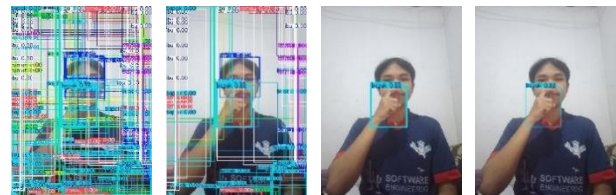
Gambar 7 memperlihatkan bagian neck, di mana fitur multi-skala digabungkan pada lapisan ini. Heatmap memperlihatkan bahwa atensi model mulai terpusat secara spesifik pada area tangan dan jari, memisahkan objek dari latar belakang.



Gambar 7. Neck

Head merupakan lapisan akhir yang menghasilkan prediksi numerik. Visualisasi post-processing menunjukkan bahwa sistem mampu menghilangkan deteksi ganda dan menampilkan satu bounding box dengan *confidence* score tertinggi pada objek yang sesuai, sebagaimana ditunjukkan pada Gambar 8.

Head Layer 0 - (1, 16, 8400)



Gambar 8. Head


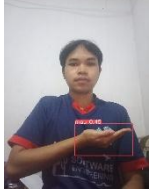


Head Layer 0 menghasilkan output dengan dimensi berukuran (1, 16, 8400), yang berarti bahwa model memproduksi 16 kanal fitur untuk setiap 8.400 titik prediksi pada satu batch inferensi. Nilai tersebut merepresentasikan distribusi prediksi bounding box dan probabilitas kelas pada seluruh grid fitur yang digunakan untuk mendeteksi objek pada skala resolusi tersebut.

3.4. Pengujian Model

Pengujian dilakukan melalui pengujian real-time menggunakan 12 citra uji baru yang diambil di lingkungan berbeda dari data latih seperti tersaji pada Tabel 3. Hasilnya menunjukkan penurunan performa dibandingkan data validasi: 7 citra terdeteksi valid, 1 citra salah klasifikasi, dan 4 citra gagal terdeteksi (no detection). Kegagalan deteksi ini mengindikasikan bahwa model masih sensitif terhadap variasi pencahayaan ekstrem dan latar belakang yang kompleks di dunia nyata.

Tabel 3. Pengujian Model

Kelas	Hasil Prediksi	Keterangan
Bantu		Valid

Tidur		Valid
Saya		Valid
Rumah		Valid
Mau		Valid
Ibu		Valid
Bapak		Valid
Maaf		Tidak Valid
Tolong		Tidak Terdeteksi

Makan		Tidak Terdeteksi
Kamu		Tidak Terdeteksi
Sehat		Tidak Terdeteksi

3.4.1 Pengujian Cross Validation

Pengujian performa model pada penelitian ini dilakukan menggunakan metode *cross validation* dengan skema *3-fold* tersaji pada Tabel 4.

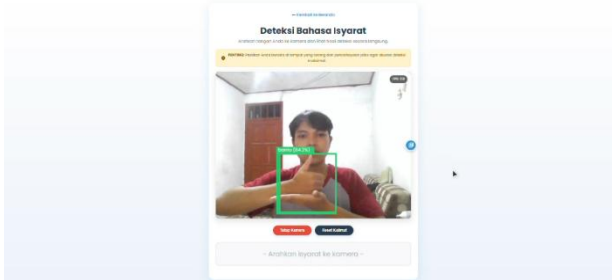
Tabel 4 Hasil Cross Validation

Fold	Precision	Recall	mAP-50	mAP-50-95
Fold 1	0.982	0.995	0.993	0.769
Fold 2	0.987	0.987	0.992	0.777
Fold 3	0.984	0.989	0.99	0.774
Rata-Rata	0.984	0.990	0.992	0.773

3.5. Implementasi Sistem

Implementasi Sistem Untuk kebutuhan deployment ke dalam lingkungan produksi, model YOLOv11 yang telah dilatih diekspor ke dalam format ONNX (Open Neural Network Exchange). Format ini dipilih karena interoperabilitasnya yang tinggi, memungkinkan model dijalankan secara efisien di berbagai platform tanpa bergantung pada framework pelatihan awal. Sistem kemudian diimplementasikan dalam bentuk aplikasi berbasis web menggunakan framework Flask yang menghubungkan model deteksi (back-end)

dengan antarmuka pengguna (front-end). Alur kerja sistem dimulai dari pengguna mengakses halaman deteksi, mengaktifkan kamera, dan melakukan gerakan isyarat. Sistem secara otomatis memproses feed video secara real-time, kemudian menampilkan visualisasi bounding box pada area tangan beserta label teks hasil terjemahan isyarat tersebut pada layer seperti terlihat pada Gambar 9.



Gambar 9. Tampilan Deteksi

3.5.1. Pengujian Realtime

Lingkungan pengujian menggunakan spesifikasi sebagai berikut:

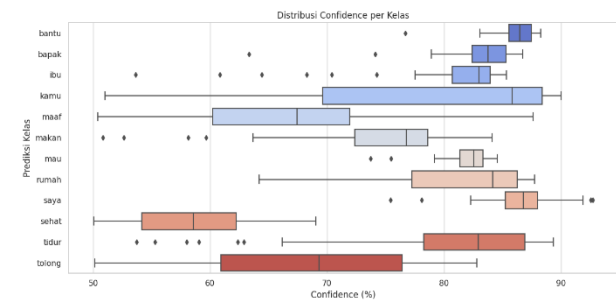
Processor	: Intel Core i3 GEN 7th
RAM	: 4GB
Browser	: Google chrome
Framework web	: Flask
Model Type	: ONNX
Total Waktu	: 5 Menit 47 Detik (00:57:49 – 01:03:36)

Tabel 5. Uji Real-time

Kelas	Jumlah Deteksi	Rata-rata Confidence (%)	Min Conf (%)	Max Conf (%)	Avg Latency (ms)	Avg FPS	Total waktu (s)
None (idle)	364	-	-	-	-	-	100.25
Bantu	34	86.15	76.69	88.27	256.50	3.60	9.29
Tidur	95	80.95	53.72	89.38	257.20	3.70	25.89
Saya	72	86.66	75.42	92.73	260.16	3.67	19.62
Rumah	60	80.84	64.19	87.76	258.05	3.70	16.31
Mau	63	82.21	73.72	84.56	259.93	3.64	17.23
Ibu	51	80.70	53.62	85.35	267.46	3.50	14.68
Bapak	46	83.10	63.31	86.76	275.84	3.48	13.44
Maaf	72	66.52	50.38	87.60	303.45	3.19	22.99
Tolong	83	68.51	50.14	82.79	254.07	3.79	22.44
Makan	105	75.12	50.83	84.11	258.84	3.69	28.65
Kamu	114	79.37	51.00	90.00	281.41	3.42	34.85
Sehat	80	58.61	50.02	69.06	256.82	3.70	21.76

Berdasarkan Tabel 5 nilai rata-rata *confidence* tertinggi diperoleh pada kelas “Saya” sebesar 86,66%, diikuti oleh kelas

“Bantu” sebesar 86,15%, dan “Bapak” sebesar 83,10%. Hasil ini menunjukkan bahwa sistem memiliki tingkat keyakinan yang tinggi dalam mengenali gerakan isyarat pada kelas-kelas tersebut.



Gambar 10. Distribusi Confidence per kelas

Berdasarkan grafik pada Gambar 10 tersebut, kelas 'bantu' dan 'bapak' menunjukkan karakteristik deteksi yang sangat baik dan stabil. Hal ini ditandai dengan kotak (*Interquartile Range*) yang pendek dan posisi median yang tinggi (di atas 85%), yang berarti model secara konsisten memberikan nilai kepercayaan tinggi dengan variasi yang minim pada setiap frame. Pada kelas 'sehat', selain memiliki *median* terendah (di bawah 60%), seluruh sebaran datanya berada pada rentang nilai yang rendah (50% - 70%). Hal ini menunjukkan bahwa model secara sistematis kesulitan mengenali isyarat 'sehat' di hampir seluruh sampel pengujian, bukan hanya pada beberapa frame tertentu saja.

3.5. Implementasi Sistem

Hasil penelitian ini menunjukkan performa yang sangat tinggi pada data validasi dengan $mAP@50$ sebesar 98% dan *precision* mencapai 96.8%. Jika dibandingkan dengan penelitian terdahulu pada deteksi bahasa isyarat berbasis YOLO, performa ini termasuk dalam kategori sangat tinggi. Sebagai contoh, penelitian menggunakan YOLOv5 pada dataset BISINDO alfabet hanya mencapai mAP sebesar 87.1% dengan *precision* 85.2% (Munandar, 2024). Hal ini menunjukkan adanya peningkatan performa yang signifikan pada penelitian ini, terutama dalam aspek klasifikasi objek.

Namun demikian, hasil ini perlu dianalisis lebih kritis. Penelitian lain berbasis YOLOv7 bahkan melaporkan nilai $mAP@50$ hingga 0.995 pada kondisi eksperimen tertentu, tetapi performa tersebut menurun ketika diterapkan pada skenario real-time (Renaningtias, 2025). Fenomena ini sejalan dengan temuan pada

penelitian ini, di mana performa tinggi pada data validasi tidak sepenuhnya tercermin pada pengujian data baru (hanya 7 dari 12 citra terdeteksi dengan benar). Hal ini mengindikasikan adanya masalah generalisasi model.

Selain itu, studi komparatif terbaru terhadap berbagai varian YOLO (v8–v12) menunjukkan bahwa meskipun model dapat mencapai $mAP@50$ di atas 0.98, terdapat trade-off antara akurasi dan kecepatan inferensi (Tawfik, 2025). Temuan ini konsisten dengan hasil penelitian ini, di mana sistem hanya mampu berjalan pada 3–4 FPS. Dengan demikian, meskipun akurasi tinggi, performa real-time masih terbatas, terutama pada perangkat dengan spesifikasi rendah.

Lebih lanjut, penelitian sebelumnya juga menunjukkan bahwa performa tinggi pada dataset terkontrol sering kali tidak mencerminkan performa pada kondisi dunia nyata. Sebagai contoh, penelitian YOLO berbasis video pada BISINDO menunjukkan penurunan akurasi signifikan saat diuji pada lingkungan yang lebih dinamis, meskipun hasil pelatihan sangat baik (Daniels, 2020). Hal ini memperkuat indikasi bahwa perbedaan distribusi data (domain shift) menjadi faktor utama dalam penurunan performa.

Dari sisi kompleksitas tugas, gesture recognition memiliki tantangan tersendiri dibandingkan object detection umum, terutama karena adanya kemiripan antar kelas dan variasi pose tangan. Penelitian lain juga menekankan pentingnya augmentasi data dan variasi kondisi lingkungan untuk meningkatkan robustness model (Li, 2025). Hal ini relevan dengan temuan pada kelas “Sehat” dan “Mau” dalam penelitian ini yang memiliki performa rendah akibat kemiripan visual dan sensitivitas terhadap kondisi lingkungan.

Secara keseluruhan, jika dibandingkan dengan penelitian sebelumnya, model yang diusulkan menunjukkan keunggulan pada performa validasi, namun masih memiliki keterbatasan pada aspek generalisasi dan kecepatan inferensi. Oleh karena itu, kontribusi utama penelitian ini terletak pada peningkatan akurasi deteksi berbasis kata (bukan huruf), sementara tantangan ke depan adalah

meningkatkan robustness model pada kondisi dunia nyata dan optimasi performa real-time.

4. PENUTUP

4.1. Kesimpulan

Penelitian ini menunjukkan bahwa algoritma YOLOv11 mampu digunakan untuk mendeteksi kata dalam bahasa isyarat BISINDO dan SIBI dengan kinerja yang sangat baik pada tahap validasi. Model mencapai $mAP@0.5$ sebesar 98% dengan Precision 96,8%, Recall 96,4%, dan F1-Score 97%, serta menunjukkan stabilitas melalui pengujian cross-validation (rata-rata $mAP@0.5$ sebesar 99,2%). Implementasi sistem berbasis Flask juga berhasil mendukung deteksi secara real-time dengan performa yang konsisten.

Namun, pengujian pada data baru menunjukkan adanya keterbatasan dalam generalisasi model pada kondisi nyata, terutama terhadap variasi pencahayaan dan latar belakang, yang menyebabkan sebagian data tidak terdeteksi atau mengalami kesalahan prediksi.

Implikasinya, model ini berpotensi mendukung komunikasi berbasis bahasa isyarat secara real-time, tetapi masih memerlukan peningkatan robustness agar dapat digunakan secara lebih andal dalam lingkungan yang beragam.

4.2. Saran

Penelitian selanjutnya disarankan untuk meningkatkan jumlah dan keragaman dataset, baik dari sisi variasi pencahayaan, latar belakang, maupun jumlah subjek (signer) agar model lebih robust terhadap kondisi nyata. Selain itu, pengujian langsung pada pengguna Tuli perlu dilakukan untuk memperoleh evaluasi yang lebih kontekstual terhadap performa sistem dalam penggunaan sehari-hari. Pengembangan aplikasi ke platform mobile serta penambahan fitur text-to-speech juga penting untuk meningkatkan aksesibilitas dan kemudahan penggunaan. Di sisi teknis, penelitian lanjutan dapat melakukan eksperimen komparatif dengan beberapa varian YOLOv11, seperti YOLOv11s, YOLOv11m, dan YOLOv11l, serta melakukan fine-tuning hyperparameter guna meningkatkan akurasi dan kemampuan generalisasi model.

5. DAFTAR PUSTAKA

- Daniels, S., Suciati, N., & Fathichah, C. (2021, February). Indonesian sign language recognition using yolo method. In IOP Conference Series: Materials Science and Engineering (Vol. 1077, No. 1, p. 012029). IOP Publishing.
- Farah, S., Husna, D., Desintha, S., & Mustikawan, D. A. (2025). Perancangan Buku “Jari Berbicara” Sebagai Pengenalan Bahasa Isyarat Bagi Kalangan Generasi Z. *Jurnal Desain*, 12(1).
- Gumelar, G., Hafiar, H., & Subekti, P. (2018). Konstruksi Makna Bisindo Sebagai Budaya Tuli Bagi Anggota Gerkatin. *Informasi*, 48(1), 65.
- Alfikri, H. R., Siswo Utomo, M., Februriyanti, H., & Nurwahyudi, E. (2022). Pembangunan Aplikasi Penerjemah Bahasa Isyarat Dengan Metode CNN Berbasis Android. *Jurnal Teknoinfo*, 16(2).
- Isma, S. T. (2018). Meneliti Bahasa Isyarat Dalam Perspektif Variasi Bahasa. (Naskah Publikasi/Tesis).
- Jegham, N., Koh, C. Y., Abdelatti, M., & Hendawi, A. (2024). YOLO Evolution: A Comprehensive Benchmark and Architectural Review of YOLOv12, YOLO11, and Their Previous Versions. arXiv preprint arXiv:2411.00201.
- Joher, G., et al. (2025). Ultralytics YOLO. Tersedia di: <https://github.com/ultralytics/ultralytics>.
- Li, X., Jettanasen, C., & Chiradeja, P. (2025). Exploration of Sign Language Recognition Methods Based on Improved YOLOv5s. *Computation*, 13(3), 59.
- Mailani, O., Nuraeni, I., Syakila, S. A., Lazuardi, J., & Komunikasi, P. I. (2022). Bahasa Sebagai Alat Komunikasi Dalam Kehidupan Manusia. *Jurnal Kampret*, 1(2).
- Munandar, A., Yunizar, Z., & Retno, S. (2024, December). Indonesian Sign Language (BISINDO) Alphabet Detection System Using YOLO (You Only Look Once) Algorithm. In *Proceedings of Malikussaleh International Conference on Multidisciplinary Studies (MICoMS)* (Vol. 4, pp. 00001-00001).
- Octaviani, Y. (2020). Kemampuan Interaksi Sosial Tunarungu di Kelurahan Batununggal Kota Bandung..
- Renaningtias, N., Utama, F. P., & Sobri, A. N. A. (2025). Detection System Indonesian Sign Language (BISINDO) in Video with YOLOv7. *JSAI (Journal Scientific and Applied Informatics)*, 8(1), 1-8.
- Tawfik, M., & Badawy, W. (2025, November). Performance Evaluation of YOLO Models for Sign Language Detection: A Comparative Study. In *2025 Twelfth International Conference on Intelligent Computing and Information Systems (ICICIS)* (pp. 120-127). IEEE.
- Utami, I. S. (2016). Sikap Siswa Tunarungu Terhadap SIBI (Sistem Isyarat Bahasa Indonesia).