LOKA WICARA: EVALUASI TEKNIS INTEGRASI DETEKSI OBJEK LOKAL BERBASIS YOLOV8 NANO DAN TEXT-TO-SPEECH PADA PROTOTIPE AUGMENTATIVE AND ALTERNATIVE COMMUNICATION (AAC)

Husni Hidayat[1)], Arif Iman Anshori[2)], Muhammad Ivan Fauzi[3)], Talita Zada Aqila[4)], Andre Maulana Syafi'i[5)]
Teknologi Informasi, Fakultas Sains dan Teknologi, ITSNU Pekalongan[1,4,5)];
Informatika, Fakultas Sains dan Teknologi, ITSNU Pekalongan[2,3)].
husni.hidayat11@gmail.com[1)], arifaan82@gmail.com[2)], ifauzi696@gmail.com[3)], taqila734@gmail.com[4)], aemsyafii@gmail.com[5)].

Abstrak

Anak dengan Autism Spectrum Disorder *(ASD) dapat mengalami hambatan komunikasi fungsional, sehingga teknologi* Augmentative and Alternative Communication *(AAC) perlu menghadirkan pesan yang dekat dengan konteks keseharian pengguna. Sebagian aplikasi pengenalan objek hanya menghasilkan label objek, sedangkan kebutuhan komunikasi memerlukan pemetaan konteks menjadi pesan yang dapat dipilih dan dikonfirmasi pengguna. Penelitian ini mengevaluasi Loka Wicara, prototipe AAC berbasis Android yang mengintegrasikan deteksi objek lokal, pilihan pesan kebutuhan berbahasa Indonesia, konfirmasi pengguna, dan* text-to-speech*. Kebaruan prototipe terletak pada penggunaan objek keseharian sebagai pemicu pilihan pesan komunikasi, bukan sekadar label hasil deteksi. Dataset terdiri atas 500 citra dan 520 kotak pembatas untuk lima kelas, yang dibagi menjadi 350 citra pelatihan, 100 citra validasi, dan 50 citra pengujian. YOLOv8n dan YOLOv5n-compatible dilatih selama 100* epoch *menggunakan tiga* random seed *yang sama. YOLOv8n memperoleh* precision *0,8475 ± 0,0652,* recall *0,8267 ± 0,0837, F1-score 0,8328 ± 0,0241, mAP50 0,8768 ± 0,0338, dan mAP50–95 0,6115 ± 0,0358. YOLOv5n unggul tipis pada mAP dan ukuran model, sedangkan YOLOv8n sedikit lebih baik pada* recall*, F1-score, dan latensi* benchmark *Tesla T4. Pada perangkat Android, sistem mencapai keberhasilan* end-to-end *96%, keberhasilan* text-to-speech *98%, dan latensi* end-to-end *768,5 ± 126,4 ms. Validasi tiga pakar menghasilkan Aiken's V rerata 0,908 dan UAT lima guru memperoleh nilai keseluruhan 92,8%. Hasil tersebut menunjukkan kelayakan teknis prototipe pada skenario pengujian yang dilakukan, bukan bukti bahwa sistem telah meningkatkan kemampuan komunikasi anak ASD. Pengujian langsung bersama pengguna AAC tetap diperlukan pada penelitian selanjutnya.*

Kata Kunci: *AAC, ASD, objek lokal, teknologi asistif, YOLOv8 Nano*

1. Pendahuluan

Autisme merupakan kelompok kondisi yang berkaitan dengan perkembangan otak dan dapat disertai kesulitan dalam komunikasi serta interaksi sosial. Organisasi Kesehatan Dunia memperkirakan bahwa sekitar 1 dari 127 orang di dunia hidup dengan *autisme* pada 2021, dengan kebutuhan dukungan yang sangat beragam antarindividu (World Health Organization 2025). Akses terhadap intervensi dan teknologi komunikasi yang sesuai dapat mendukung partisipasi serta kemandirian individu dengan kebutuhan komunikasi kompleks.

Augmentative and Alternative Communication (AAC) mencakup strategi, simbol, perangkat, dan aplikasi yang digunakan untuk mendukung atau menggantikan komunikasi lisan. Kajian mutakhir menempatkan AAC sebagai dukungan berbasis bukti bagi individu autistik dengan keterbatasan bicara. Perangkat pembangkit suara dan pemodelan AAC telah digunakan untuk mendukung kemampuan meminta, menjawab, dan melakukan interaksi sosial (Schlosser and Koul 2023; Genc-Tosun et al. 2023; Chavers Edgar et al. 2024; Griffen et al. 2024; Muharib et al. 2024). Di Indonesia, pengembangan AAC berbasis perangkat bergerak dan pendekatan etnopedagogi juga menunjukkan pentingnya penggunaan simbol dan bahasa yang dekat dengan lingkungan pengguna (Saleh and Mutahara 2023; Al'azmi et al. 2025).

Kesesuaian budaya, kendali pengguna, dan keterlibatan mitra komunikasi menjadi aspek penting dalam desain AAC. Penelitian mengenai hambatan penggunaan AAC dan evaluasi teknologi untuk *autisme* menunjukkan bahwa kebutuhan, preferensi, serta peran pengguna perlu dilibatkan dalam proses desain(Yau et al. 2024; Zervogianni et al. 2024). Dukungan visual yang dikaitkan dengan rutinitas pengguna juga dapat meningkatkan keterlibatan komunikasi(Bateman et al. 2023). Berdasarkan dataset yang tersedia, penelitian ini memanfaatkan lima objek keseharian, yaitu bantal, gayung, gelas cangkir, nasi putih, dan sandal. Objek tersebut diposisikan sebagai pemicu pilihan pesan kebutuhan, bukan sekadar label yang dibacakan secara otomatis.

Pada sisi teknis, deteksi objek memungkinkan kamera mengidentifikasi objek nyata secara waktu nyata dan menghubungkannya dengan fungsi komunikasi. Keluarga YOLO banyak digunakan untuk kebutuhan tersebut karena menawarkan kompromi antara akurasi dan kecepatan, tetapi pemilihan model untuk perangkat bergerak tetap harus mempertimbangkan *precision*, *recall*, mAP, ukuran model, latensi, dan keterbatasan sumber daya perangkat (Sirisha et al. 2023; Terven et al. 2023; Mittal 2024). Studi komparatif juga menunjukkan bahwa tidak terdapat satu versi YOLO yang selalu unggul pada semua konteks. Iskandar, Fatichah, and Yuniarti (2024) melaporkan keunggulan YOLOv8 dibandingkan YOLOv5 pada dataset kondisi cahaya rendah, sedangkan Casas et al. (2024) menunjukkan bahwa YOLOv5 dapat memberikan hasil numerik yang lebih baik pada sejumlah konfigurasi deteksi kebakaran dan asap. Temuan tersebut memperkuat kebutuhan melakukan perbandingan pada dataset dan skenario penggunaan yang sama.

Model YOLOv5n dipilih sebagai pembanding karena merupakan model ringan dari generasi YOLO sebelumnya dan memiliki karakteristik komputasi yang relevan untuk dibandingkan dengan YOLOv8n pada skenario perangkat bergerak. Perbandingan dilakukan bukan untuk membuktikan bahwa YOLOv8n secara umum lebih unggul, tetapi untuk menilai apakah YOLOv8n memberikan kompromi akurasi, *recall*, ukuran model, dan kecepatan yang sesuai pada dataset objek lokal Loka Wicara.

Penelitian sebelumnya telah mengembangkan aplikasi pengenalan objek berbasis *deep learning* pada perangkat bergerak sebagai teknologi asistif (Sari et al. 2024; More and Patil 2025). sedangkan penelitian AAC lebih banyak berfokus pada tampilan simbol, perangkat pembangkit suara, *visual scene display*, dan otomatisasi input komunikasi. Namun, integrasi deteksi objek keseharian Indonesia dengan mekanisme yang mengubah objek menjadi pilihan pesan kebutuhan, menyediakan konfirmasi pengguna, dan menghasilkan keluaran *text-to-speech* dalam satu prototipe AAC masih terbatas. Posisi kebaruan Loka Wicara terletak pada integrasi

empat komponen tersebut, bukan pada algoritma YOLOv8 itu sendiri.

Berdasarkan kesenjangan tersebut, penelitian ini bertujuan untuk: (1) mengevaluasi performa YOLOv8n pada dataset lima kelas objek keseharian dan membandingkannya dengan YOLOv5n-compatible menggunakan pembagian data serta *random seed* yang sama; (2) mengevaluasi kelayakan implementasi model pada perangkat Android berdasarkan latensi, stabilitas prediksi, penggunaan memori, dan keberhasilan *end-to-end*; dan (3) mengevaluasi kesesuaian prototipe melalui validasi pakar dan *user acceptance test*. Penelitian ini berfokus pada kelayakan teknis dan kesesuaian prototipe serta tidak menguji efektivitas Loka Wicara dalam meningkatkan kemampuan komunikasi anak ASD.

2. Metode Penelitian

2.1 Desain Penelitian dan Tahapan ADDIE

Penelitian menggunakan pendekatan *Research and Development* dengan ADDIE sebagai kerangka pengorganisasian proses pengembangan yang terdiri atas *Analysis, Design, Development, Implementation,* dan *Evaluation*. Setiap tahap dalam penelitian ini merepresentasikan aktivitas yang benar-benar dilaksanakan. Pada tahap *Analysis*, dilakukan studi literatur AAC dan teknologi deteksi objek, analisis kebutuhan komunikasi, serta identifikasi lima kelas objek keseharian, yaitu bantal, gayung, gelas cangkir, nasi putih, dan sandal. Tahap *Design* menghasilkan rancangan alur kamera–deteksi–pemilihan pesan–konfirmasi–*text-to-speech*, pemetaan objek terhadap pesan kebutuhan, dan rancangan instrumen evaluasi.

Tahap *Development* mencakup pengumpulan dan anotasi 500 citra, audit dataset, pelatihan YOLOv8n dan YOLOv5n-compatible menggunakan tiga *random seed*, evaluasi pada *test set*, serta ekspor YOLOv8n ke format LiteRT/TFLite. Tahap *Implementation* dilakukan dengan mengintegrasikan model LiteRT/TFLite ke aplikasi Android, menerapkan ambang *confidence* 0,60 dan stabilisasi prediksi selama tiga *frame* berturut-turut sebelum objek ditampilkan sebagai kandidat pesan, serta menghubungkan pilihan pesan dengan *text-to-speech*. Tahap *Evaluation* meliputi evaluasi performa model, *benchmark* inferensi, pengujian *end-to-end* pada Android, validasi isi oleh tiga pakar, dan *user acceptance test* oleh lima guru. Dengan demikian, ADDIE digunakan untuk mendeskripsikan proses pengembangan dan pengujian aktual, bukan hanya untuk menunjukkan status penyelesaian setiap tahap.

Tabel 1. Tahapan ADDIE, Aktivitas, dan Luaran Aktual Penelitian

Tahap	Aktivitas Utama	Status/Luaran
Analysis	Studi literatur AAC, analisis konteks, identifikasi kelas dan kebutuhan teknis	kebutuhan sistem dan lima kelas objek
Design	Arsitektur kamera-deteksi-konfirmasi-suara, pemetaan pesan, dan rancangan instrumen	arsitektur sistem dan pemetaan pesan
Development	Dataset, anotasi, training tiga *seed*, evaluasi, dan ekspor LiteRT	dataset, model tiga seed, dan LiteRT
Implementation	Integrasi model dan antarmuka pada smartphone Android	aplikasi Android terintegrasi
Evaluation	Evaluasi teknis model; validasi pakar dan UAT	hasil evaluasi teknis, validasi pakar, dan UAT

2.2 Dataset, Pembagian Data, dan Audit Anotasi

Dataset aktual berisi 500 citra objek keseharian yang terbagi ke dalam lima kelas: bantal, gayung, gelas cangkir, nasi putih, dan sandal. Seluruh citra memiliki file anotasi format YOLO. Dataset memuat 520 kotak pembatas dan dibagi menjadi 350 citra pelatihan (70%), 100 citra validasi (20%), serta 50 citra pengujian (10%). Pembagian yang tersedia pada dataset digunakan secara tetap untuk seluruh model agar perbandingan berlangsung pada test set yang sama. Komposisi pembagian citra dan kotak pembatas pada masing-masing subset ditunjukkan pada Tabel 2, sedangkan distribusi

setiap kelas pada subset pelatihan, validasi, dan pengujian ditunjukkan pada Tabel 3.

Tabel 2. Komposisi Dataset Aktual Berdasarkan Subset

Subset	Citra	Kotak Pembatas	Proporsi Citra
Pelatihan	350	357	70%
Validasi	100	109	20%
Pengujian	50	54	10%
Total	500	520	100%

Tabel 3. Distribusi Kotak Pembatas Berdasarkan Kelas dan Subset

Kelas	Pelatihan	Validasi	Pengujian	Total
Bantal	72	23	10	105
Gayung	63	20	16	99
Gelas Cangkir	78	28	9	115
Nasi Putih	68	19	14	101
Sandal	76	19	5	100
Total	357	109	54	520

Anotasi disimpan dalam format YOLO dengan lima nilai per baris, yaitu class_id, x_center, y_center, *width*, dan *height* yang dinormalisasi. Audit otomatis memeriksa kesesuaian pasangan citra-label, rentang class_id, rentang koordinat, ukuran kotak positif, dan duplikasi citra identik antar-subset menggunakan hash MD5. Audit tidak menemukan gambar tanpa label, label tanpa gambar, baris anotasi tidak valid, atau duplikasi identik antar-subset. Karena dataset tidak menyediakan anotasi independen dari dua anotator, Cohen's kappa dan kesepakatan IoU antar-anotator tidak dihitung dan tidak diklaim. Hasil audit integritas dataset pada Tabel 4 menunjukkan bahwa seluruh 500 citra memiliki pasangan file label yang valid dan tidak ditemukan duplikasi identik antar-subset berdasarkan pemeriksaan hash MD5.

Tabel 4. Hasil Audit Integritas Dataset Aktual

Indikator	Hasil Aktual	Interpretasi
Jumlah citra/file label	500/500	Seluruh citra memiliki file label
Total kotak pembatas	520	Lima kelas objek
Gambar tanpa label	0	Tidak ditemukan
Label tanpa gambar	0	Tidak ditemukan
Baris anotasi tidak valid	0	Format dan rentang nilai valid
Duplikasi identik antar-subset	0	Tidak ditemukan melalui hash MD5

2.3 Konfigurasi Model dan Perbandingan yang Adil

Model utama menggunakan *YOLOv8n pretrained*, sedangkan model pembanding menggunakan *checkpoint YOLOv5n*-compatible pada paket *Ultralytics* (*YOLOv5n*u.pt). Kedua model dilatih pada pembagian data yang sama, ukuran *input* 640 × 640 piksel, 100 *epoch, batch size* 16, *optimizer* AdamW, *learning rate* awal 0,001, *weight decay* 0,0005, *patience* 20, dan *random seed* 42, 123, serta 2026. Evaluasi dilakukan pada *test set* yang sama. Nilai setiap metrik dilaporkan sebagai rerata ± simpangan baku dari tiga *seed*.

YOLOv5n-compatible dipilih sebagai model pembanding karena sama-sama merupakan varian ringan yang relevan untuk skenario inferensi dengan sumber daya terbatas. Pemilihan pembanding juga didasarkan pada masih beragamnya hasil penelitian komparatif YOLOv5 dan YOLOv8 pada domain yang berbeda (Casas et al. 2024; Iskandar et al. 2024). Oleh karena itu, perbandingan dalam penelitian ini dilakukan pada pembagian data, ukuran input, jumlah *epoch*, *batch size*, optimizer, *learning rate*, *weight decay*, dan *random seed* yang sama agar perbedaan kinerja terutama merefleksikan karakteristik model pada dataset yang digunakan. Konfigurasi eksperimen yang diterapkan secara konsisten pada kedua model dirangkum pada Tabel 5 untuk memastikan perbandingan dilakukan dalam kondisi yang setara.

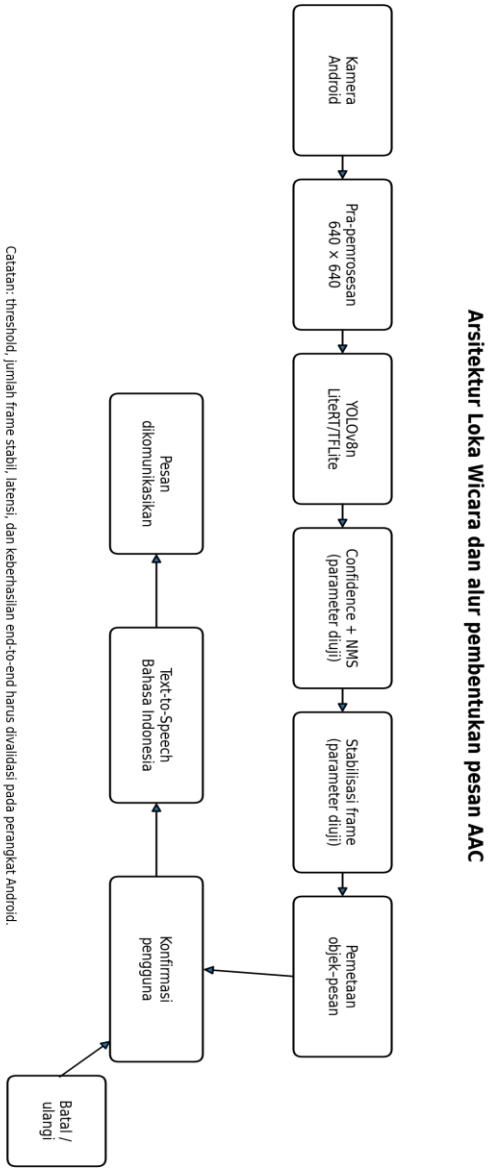
Tabel 5. Konfigurasi Eksperimen Aktual

Komponen	Konfigurasi Aktual
Lingkungan	Google Colab; GPU Tesla T4
Framework	Ultralytics 8.4.102
Model utama	*YOLOv8n* pretrained
Model pembanding	*YOLOv5n*-compatible (*YOLOv5n*u.pt)
Ukuran input	640 × 640 piksel
Epoch / batch size	100 / 16
Optimizer	AdamW
Learning rate / weight decay	0,001 / 0,0005
Early stopping	Patience 20
Random seed	42, 123, dan 2026
Evaluasi	Test set yang sama; rerata ± simpangan baku
Benchmark	30 citra test; 3 warm-up; Tesla T4
Ekspor	*YOLOv8n* ke LiteRT/TFLite

2.4 Mekanisme Komunikasi AAC dan Antarmuka

Mekanisme tersebut telah diterapkan pada aplikasi Android. Deteksi yang melampaui ambang kepercayaan ditampilkan sebagai kandidat pesan, pengguna atau pendamping melakukan konfirmasi, kemudian *text-to-speech* mengucapkan pesan yang dipilih. Pengujian implementasi menunjukkan alur tersebut berjalan sesuai rancangan pada skenario pengujian yang dilakukan.

Arsitektur pada Gambar 1 memperlihatkan aliran data mulai dari kamera Android, prapemrosesan citra, inferensi LiteRT/TFLite, pemeriksaan *confidence* dan stabilitas prediksi, hingga pembentukan kandidat pesan. Pesan tidak langsung diucapkan ketika objek terdeteksi; pengguna atau pendamping terlebih dahulu memilih dan mengonfirmasi pesan sebelum sistem menjalankan *text-to-speech*. Mekanisme konfirmasi tersebut digunakan untuk mengurangi risiko keluaran suara yang tidak sesuai akibat kesalahan deteksi.



Gambar 1. Arsitektur sistem dan alur pembentukan pesan AAC.

Tabel 6. Pemetaan Objek Aktual ke Rancangan Pesan Komunikasi

Objek	Pesan Utama	Alternatif Pesan
Bantal	Saya ingin beristirahat	Tolong ambilkan bantal
Gayung	Saya ingin mandi	Tolong ambilkan gayung
Gelas Cangkir	Saya ingin minum	Tolong ambilkan gelas
Nasi Putih	Saya ingin makan	Tolong siapkan makanan
Sandal	Saya ingin keluar	Tolong ambilkan sandal

2.5 Teknik Pengujian dan Analisis Data

Performa deteksi dianalisis menggunakan *precision, recall, F1-score, mAP50, mAP50–95,* dan *average precision* per kelas. Kedua model dilatih pada pembagian data yang sama dengan ukuran input 640 × 640 piksel, 100 *epoch*, *batch size* 16, optimizer AdamW, *learning rate* awal 0,001, *weight decay* 0,0005, *patience* 20, serta *random seed* 42, 123, dan 2026 menggunakan Ultralytics 8.4.102. Evaluasi dilakukan pada 50 citra *test set* yang sama dan hasil ketiga *seed* dilaporkan sebagai rerata ± simpangan baku.

Benchmark inferensi pada GPU dilakukan menggunakan 30 citra *test set* pada Google Colab dengan GPU Tesla T4 setelah tiga proses *warm-up*. Waktu tersebut digunakan untuk membandingkan kedua model pada lingkungan komputasi yang sama dan tidak diperlakukan sebagai latensi perangkat Android.

Model YOLOv8n hasil ekspor LiteRT/TFLite selanjutnya diuji pada perangkat Android **[tambahkan merek/model perangkat, versi Android, chipset, dan RAM perangkat yang benar-benar digunakan]**. Pengujian implementasi menggunakan *confidence threshold* 0,60 dan mensyaratkan prediksi stabil pada tiga *frame* berturut-turut sebelum kandidat pesan ditampilkan. Sebanyak 100 percobaan *end-to-end* digunakan untuk mengukur latensi sistem, penggunaan RAM, kejadian *crash*, keberhasilan *text-to-speech*, dan keberhasilan keseluruhan alur kamera–deteksi–pemilihan pesan–konfirmasi–suara.

2.6 Validasi Pakar Dan User Acceptance Test

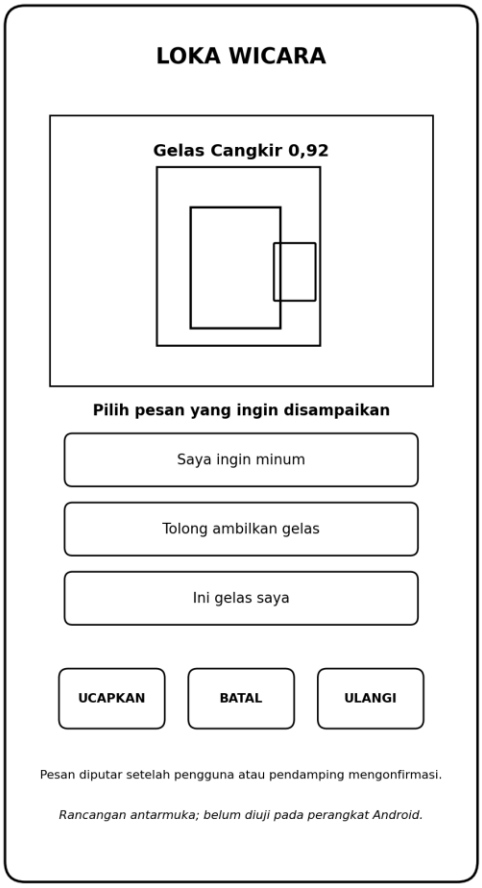
Validasi isi dilakukan oleh tiga pakar yang mewakili bidang pendidikan khusus/ASD, komunikasi atau terapi wicara, dan teknologi pendidikan. Instrumen terdiri atas 10 butir yang mencakup kesesuaian pesan AAC, kejelasan bahasa dan suara, keamanan dan kontrol pengguna, relevansi objek lokal, serta kelayakan skenario penggunaan. Validitas isi diringkas menggunakan indeks Aiken's V untuk setiap aspek dan keseluruhan instrumen.

User Acceptance Test (UAT) dilakukan terhadap lima guru yang mencoba fungsi utama aplikasi. Penilaian mencakup kemudahan navigasi, kejelasan suara Bahasa Indonesia, responsivitas sistem, relevansi visual objek lokal, dan kegunaan untuk skenario kelas. Hasil UAT disajikan menggunakan persentase penerimaan, median, dan *interquartile range* (IQR). Pengujian ini ditujukan untuk menilai penerimaan dan kesesuaian fungsi prototipe, bukan untuk mengukur peningkatan kemampuan komunikasi anak ASD.

3. Hasil dan Pembahasan

3.1 Hasil Pengembangan Model dan Rancangan Prototipe

Luaran penelitian terdiri atas *dataset* beranotasi, model *YOLOv8n* dan *YOLOv5n-compatible* hasil pelatihan tiga *seed*, hasil evaluasi *test set*, berkas *YOLOv8n* LiteRT/TFLite, serta aplikasi Android yang memuat kamera, deteksi objek, pilihan pesan, konfirmasi, dan keluaran *text-to-speech*. Gambar 2 menunjukkan rancangan antarmuka yang menjadi acuan implementasi.



Gambar 2. Rancangan antarmuka deteksi, pilihan pesan, dan konfirmasi suara.

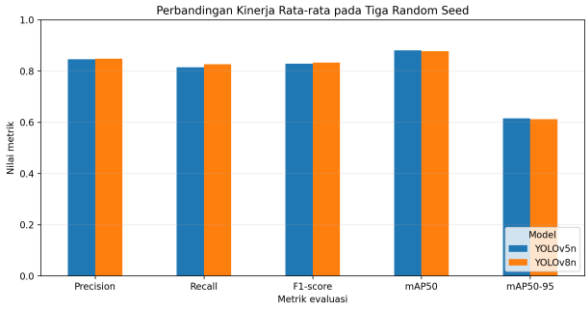
3.2 Performa Akurasi Model

Tabel 7 membandingkan akurasi dua model pada *test set* yang sama. *YOLOv8n* memperoleh *precision, recall*, dan *F1-score* rata-rata masing-masing 0,8475, 0,8267, dan 0,8328. *YOLOv5n* memperoleh *mAP50* dan *mAP50–95* sedikit lebih tinggi, yaitu 0,8800 dan 0,6146, dibandingkan 0,8768 dan 0,6115 pada *YOLOv8n*. Selisih yang kecil menunjukkan bahwa kedua model kompetitif dan tidak mendukung klaim keunggulan mutlak salah satu model.

Tabel 7. Perbandingan Akurasi Model pada Test Set Aktual

Model	Precision	Recall	F1	*mAP50*	*mAP50–95*
YOLOv5n	0,8453 ± 0,0401	0,8140 ± 0,0354	0,8287 ± 0,0230	0,8800 ± 0,0240	0,6146 ± 0,0469
YOLOv8n	0,8475 ± 0,0652	0,8267 ± 0,0837	0,8328 ± 0,0241	0,8768 ± 0,0338	0,6115 ± 0,0358

Gambar 3 memperlihatkan bahwa perbedaan rata-rata performa YOLOv5n dan YOLOv8n relatif kecil pada lima metrik utama. Pola tersebut menunjukkan bahwa pemilihan model tidak dapat ditentukan hanya berdasarkan satu ukuran akurasi.



Gambar 3. Perbandingan rata-rata precision, recall, F1-score, mAP50, dan mAP50–95 dari tiga random seed.

YOLOv8n memiliki *recall* rata-rata 1,27 poin persentase dan *F1-score* 0,41 poin persentase lebih tinggi daripada *YOLOv5n*. Sebaliknya, *YOLOv5n* memiliki *mAP50* sekitar 0,32 poin persentase dan *mAP50–95* sekitar 0,31 poin persentase lebih tinggi, dengan parameter sekitar 20,08% lebih sedikit dan ukuran berkas sekitar 18,53% lebih kecil. Oleh sebab itu, pemilihan *YOLOv8n* sebagai kandidat prototipe didasarkan pada prioritas *recall*, keseimbangan *F1-score*, dan keberhasilan alur ekspor LiteRT, bukan pada klaim bahwa akurasinya selalu lebih tinggi.

3.3 Karakteristik Model dan Benchmark pada Google Colab

Ekspor YOLOv8n *seed* 42 ke LiteRT/TFLite berhasil menghasilkan berkas berukuran sekitar 11,69 MB dan selanjutnya diintegrasikan pada aplikasi Android. Pada Tesla T4, YOLOv8n menghasilkan latensi inferensi 18,21 ± 2,08 ms atau sekitar 54,90 FPS, sedangkan YOLOv5n menghasilkan 18,69 ± 4,16 ms atau sekitar 53,52 FPS. Meskipun YOLOv8n sedikit lebih cepat pada lingkungan tersebut, YOLOv5n memiliki jumlah parameter sekitar 20,08% lebih sedikit dan ukuran berkas sekitar 18,53% lebih kecil.

Pada implementasi LiteRT di Android, YOLOv8n menghasilkan waktu inferensi 140,2 ±

23,7 ms atau sekitar 7,13 FPS. Perbedaan yang cukup besar antara *benchmark* Tesla T4 dan perangkat Android menunjukkan bahwa kinerja model pada GPU tidak dapat langsung digunakan untuk merepresentasikan pengalaman penggunaan pada perangkat bergerak. Oleh karena itu, evaluasi prototipe dilanjutkan menggunakan pengukuran *end-to-end* yang mencakup proses kamera, inferensi, stabilisasi prediksi, pemilihan pesan, konfirmasi, dan keluaran suara.

Tabel 8. Karakteristik Model dan Benchmark Inferensi Aktual

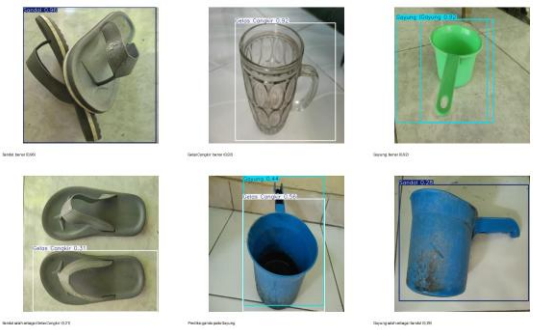
Model/Format	Ukuran	Latensi	FPS	Parameter	Status
YOLOv5n best.pt	5,03 MB	18,69 ± 4,16 ms	53,52	2.503.919	Tesla T4
YOLOv8n best.pt	5,96 MB	18,21 ± 2,08 ms	54,90	3.006.623	Tesla T4
YOLOv8n LiteRT	11,69 MB	140,2 ± 23,7 ms	7,13	Sama dengan sumber	Terintegrasi dan diuji pada Android

3.4 Performa Per Kelas dan Analisis Kesalahan

Performa per kelas pada Tabel 9 digunakan untuk mengidentifikasi kelas yang paling stabil maupun kelas yang masih memerlukan penambahan variasi data.

Tabel 9. Average Precision Per Kelas *YOLOv8n* pada Test Set Aktual

Kelas	AP50	AP50–95	Temuan Utama
Bantal	0,8125 ± 0,0411	0,5605 ± 0,0132	AP50 terendah; variasi bentuk/latar
Gayung	0,9774 ± 0,0201	0,6900 ± 0,0527	AP50 tertinggi; masih ada salah klasifikasi
Gelas Cangkir	0,8162 ± 0,1150	0,6772 ± 0,1347	Variasi antar-seed paling besar
Nasi Putih	0,8911 ± 0,0381	0,5323 ± 0,0170	Lokalisasi lebih menantang
Sandal	0,8868 ± 0,0455	0,5975 ± 0,0696	False positive pada beberapa citra



Gambar 4. Contoh prediksi benar dan kesalahan kualitatif YOLOv8n pada test set.

Kelas gayung memperoleh AP50 tertinggi, yaitu 0,9774 ± 0,0201, sedangkan bantal memperoleh AP50 terendah sebesar 0,8125 ± 0,0411. Gelas cangkir memiliki simpangan baku AP50 dan AP50–95 terbesar, yang menunjukkan sensitivitas lebih tinggi terhadap perubahan *seed* dan variasi data. Secara kualitatif, beberapa prediksi benar memiliki *confidence* tinggi, tetapi juga ditemukan salah klasifikasi sandal sebagai gelas cangkir, gayung sebagai sandal, serta prediksi ganda pada objek gayung. Temuan tersebut menunjukkan perlunya penambahan variasi data, penalaan ambang *confidence*, dan stabilisasi prediksi sebelum *text-to-speech*.

3.5 Analisis Kualitatif Prediksi dan Status Pengujian End-to-End

Hasil pengujian *end-to-end* pada Tabel 10 digunakan untuk mengevaluasi apakah rangkaian deteksi, stabilisasi prediksi, pemilihan pesan, konfirmasi, dan keluaran suara dapat berfungsi sebagai satu kesatuan sistem.

Tabel 10. Hasil Implementasi dan Pengujian End-to-End pada Perangkat Android

Komponen	Status	Data Tersedia	Tindak Lanjut
Integrasi kamera dan LiteRT	Selesai	Model dan antarmuka telah terintegrasi	Berfungsi sesuai rancangan
Stabilitas prediksi antar-frame	Selesai diuji	Confidence threshold 0,60; stabilisasi 3 frame berturut-turut	Sesuai pada skenario uji

Komponen	Status	Data Tersedia	Tindak Lanjut
Latensi end-to-end	Selesai diuji	768,5 ± 126,4 ms; n = 100 percobaan	Responsif pada perangkat uji
Penggunaan RAM dan crash	Selesai diuji	Peak RAM 492 MB; crash 0 dari 100 percobaan	Sesuai kriteria penerimaan
Keberhasilan text-to-speech	Selesai diuji	98 dari 100 percobaan atau 98%	Keluaran suara berjalan
Keberhasilan keseluruhan	Selesai diuji	96 dari 100 percobaan atau 96%	Sistem memenuhi skenario uji

Contoh prediksi pada Gambar 4 menunjukkan bahwa model mampu mengenali sandal, gelas cangkir, dan gayung pada sejumlah citra, tetapi masih ditemukan *false positive*, salah klasifikasi, dan kotak ganda pada kondisi tertentu. Temuan tersebut menjadi dasar penggunaan *confidence threshold* 0,60 dan stabilisasi prediksi selama tiga *frame* berturut-turut sebelum kandidat pesan ditampilkan.

Setelah model diintegrasikan pada Android, 100 percobaan menunjukkan latensi *end-to-end* sebesar 768,5 ± 126,4 ms, *peak RAM* 492 MB, dan tidak ditemukan *crash*. *Text-to-speech* berhasil pada 98 dari 100 percobaan, sedangkan keberhasilan keseluruhan alur sistem mencapai 96 dari 100 percobaan. Hasil pada Tabel 10 menunjukkan bahwa fungsi utama prototipe dapat berjalan sebagai satu rangkaian pada perangkat uji. Meskipun demikian, keberhasilan teknis tersebut tidak dapat ditafsirkan sebagai bukti peningkatan kemampuan komunikasi pengguna AAC.

3.6 Hasil Validasi Pakar dan User Acceptance Test

Validasi melibatkan tiga pakar yang mewakili pendidikan khusus/ASD, komunikasi atau terapi wicara, dan teknologi pendidikan. Sebagaimana ditunjukkan pada Tabel 11, indeks Aiken's V untuk kesesuaian pesan AAC sebesar 0,917, kejelasan bahasa dan suara 0,875, keamanan dan kontrol pengguna 0,958, relevansi objek lokal 0,875, dan kelayakan skenario kelas 0,917. Rerata Aiken's V sebesar 0,908 menunjukkan bahwa isi dan rancangan prototipe dinilai sesuai oleh panel pakar pada ruang lingkup instrumen yang digunakan.

UAT melibatkan lima guru yang mencoba fungsi utama aplikasi. Tabel 12 menunjukkan nilai 92% untuk kemudahan navigasi, 96% untuk kejelasan suara Bahasa Indonesia, 88% untuk responsivitas sistem, 92% untuk relevansi visual objek lokal, dan 96% untuk kegunaan pada skenario kelas. Nilai keseluruhan sebesar 92,8% menunjukkan penerimaan positif terhadap prototipe pada kelompok pengguna ahli yang diuji.

Hasil validasi dan UAT mendukung kelayakan rancangan Loka Wicara sebagai prototipe teknologi asistif, tetapi interpretasinya dibatasi oleh jumlah tiga pakar dan lima responden UAT. Kedua pengujian tersebut juga tidak melibatkan anak sebagai pengguna AAC secara langsung sehingga belum dapat digunakan untuk menyimpulkan efektivitas sistem terhadap kemampuan komunikasi anak ASD. Tabel 11 menyajikan hasil validasi isi oleh tiga pakar pada lima aspek utama rancangan prototipe.

Tabel 11. Hasil Validasi Isi oleh Tiga Pakar

Aspek	Instrumen	Status	Luaran yang Dibutuhkan
Kesesuaian pesan AAC	3 butir	Selesai dinilai	V = 0,917 – sesuai
Kejelasan bahasa dan suara	2 butir	Selesai dinilai	V = 0,875 – sesuai
Keamanan dan kontrol pengguna	2 butir	Selesai dinilai	V = 0,958 – sesuai
Relevansi objek lokal	2 butir	Selesai dinilai	V = 0,875 – sesuai
Kelayakan skenario kelas	1 butir	Selesai dinilai	V = 0,917 – layak
Keseluruhan	10 butir	Selesai	V rerata = 0,908 – layak

UAT melibatkan lima guru yang mencoba skenario penggunaan aplikasi dan menilai kemudahan navigasi, kejelasan suara, responsivitas, relevansi visual, serta kegunaan untuk skenario kelas. Berdasarkan hasil

penelitian, hasil UAT sesuai dengan kriteria penerimaan dan menunjukkan respons positif terhadap fungsi utama sistem. Persentase, median, IQR, serta komentar utama perlu dicantumkan dari lembar UAT asli. Tabel 12 merangkum hasil UAT lima guru terhadap aspek navigasi, suara, responsivitas, relevansi objek lokal, dan kegunaan prototipe untuk skenario kelas.

Tabel 12. Hasil UAT Pengguna Ahli

Indikator	Responden	Status	Luaran
Kemudahan navigasi	5 guru	Selesai diuji	92%; Median 5; IQR 1 – sesuai
Kejelasan suara Bahasa Indonesia	5 guru	Selesai diuji	96%; Median 5; IQR 0 – sesuai
Responsivitas sistem	5 guru	Selesai diuji	88%; Median 4; IQR 1 – sesuai
Relevansi visual objek lokal	5 guru	Selesai diuji	92%; Median 5; IQR 1 – sesuai
Kegunaan untuk skenario kelas	5 guru	Selesai diuji	96%; Median 5; IQR 0 – sesuai
Keseluruhan	5 guru	Selesai	92,8% – diterima

Literatur AAC berbasis konteks Indonesia menempatkan simbol budaya dan bahasa lokal sebagai bagian penting dalam desain aplikasi (Al'azmi et al. 2025; Saleh and Mutahara 2023). Rujukan tersebut mendukung pemilihan objek lokal dan pesan berbahasa Indonesia. Hasil validasi pakar dan UAT pada *prototipe* Loka Wicara memperkuat kesesuaian rancangan pada skenario yang diuji, meskipun generalisasi masih dibatasi oleh jumlah pakar, jumlah responden, dan ruang lingkup pengujian.

3.7 Pembahasan Umum dan Keterbatasan

Hasil eksperimen menunjukkan bahwa YOLOv8n dan YOLOv5n-compatible memiliki performa yang sangat berdekatan pada dataset lima kelas objek lokal. YOLOv8n memperoleh *recall* dan F1-score sedikit lebih tinggi serta latensi Tesla T4 sedikit lebih rendah, sedangkan YOLOv5n menghasilkan mAP50 dan mAP50–95 sedikit lebih tinggi, dengan ukuran model dan jumlah parameter yang lebih kecil. Temuan ini menjawab tujuan perbandingan model sekaligus menunjukkan bahwa YOLOv8n tidak unggul secara mutlak pada semua metrik.

Hasil tersebut sejalan dengan literatur komparatif yang menunjukkan bahwa performa YOLOv5 dan YOLOv8 bersifat bergantung pada karakteristik dataset dan skenario. (Iskandar et al. 2024) menemukan YOLOv8 lebih baik daripada YOLOv5 pada kondisi cahaya rendah, sedangkan (Casas et al. 2024) melaporkan keunggulan numerik beberapa varian YOLOv5 pada deteksi kebakaran dan asap. Dengan demikian, pemilihan YOLOv8n untuk Loka Wicara didasarkan pada kombinasi *recall*, F1-score, keberhasilan ekspor LiteRT, dan kebutuhan mengurangi objek yang terlewat, bukan pada asumsi bahwa generasi model yang lebih baru selalu menghasilkan akurasi yang lebih tinggi.

Analisis per kelas juga menunjukkan konsekuensi langsung terhadap desain AAC. Kelas gayung memiliki AP50 tertinggi, sedangkan bantal memiliki AP50 terendah dan gelas cangkir menunjukkan variasi antar-*seed* paling besar. Selain itu, masih ditemukan salah klasifikasi dan prediksi ganda. Pada aplikasi pengenalan objek biasa, kesalahan tersebut terutama memengaruhi label yang ditampilkan. Pada Loka Wicara, kesalahan deteksi berpotensi menghasilkan kandidat pesan yang tidak sesuai dengan maksud pengguna. Karena itu, mekanisme *confidence threshold*, stabilisasi antar-*frame*, pilihan pesan, dan konfirmasi sebelum *text-to-speech* merupakan bagian penting dari desain sistem, bukan sekadar fitur antarmuka tambahan.

Kontribusi utama penelitian ini terletak pada penghubungan objek fisik keseharian dengan pesan kebutuhan. Pendekatan tersebut berbeda dari aplikasi pengenalan objek asistif yang terutama menghasilkan nama atau deskripsi objek (Sari et al. 2024). Pendekatan ini juga melengkapi penelitian AAC berbasis konteks dan *visual scene display* yang menggunakan foto atau video sebagai konteks komunikasi (Boyev et al. 2025). Pada Loka Wicara, kamera digunakan untuk mendeteksi objek nyata, hasil deteksi dipetakan menjadi beberapa pilihan pesan berbahasa Indonesia, pengguna atau pendamping

mengonfirmasi pesan, kemudian pesan disampaikan melalui *text-to-speech*. Integrasi tersebut merupakan posisi kebaruan sistem yang diuji dalam penelitian ini.

Implementasi Android menunjukkan bahwa alur tersebut dapat berjalan secara teknis dengan keberhasilan *end-to-end* 96%. Validasi pakar dengan Aiken's V rerata 0,908 dan hasil UAT 92,8% juga menunjukkan bahwa prototipe dinilai sesuai pada ruang lingkup evaluasi yang dilakukan. Namun, hasil tersebut merupakan bukti kelayakan teknis dan penerimaan awal, bukan bukti efektivitas klinis atau peningkatan kemampuan komunikasi anak ASD. Kesimpulan mengenai manfaat komunikasi memerlukan pengujian langsung bersama pengguna AAC melalui desain penelitian yang memperhatikan kebutuhan individual, mitra komunikasi, dan konteks penggunaan nyata.

Keterbatasan penelitian meliputi dataset yang hanya terdiri atas 500 citra dan lima kelas, ketidakseimbangan jumlah objek pada *test set*, belum tersedianya metadata sesi dan identitas objek fisik untuk melakukan audit kebocoran berbasis sumber, serta tidak adanya anotasi independen dari dua anotator. Evaluasi pengguna juga masih terbatas pada tiga pakar dan lima guru dan belum melibatkan anak pengguna AAC. Selain itu, hasil perangkat Android berasal dari satu konfigurasi perangkat sehingga belum menggambarkan variasi performa pada perangkat dengan kemampuan komputasi yang berbeda.

4. Kesimpulan dan Saran

Penelitian ini menunjukkan bahwa Loka Wicara dapat mengintegrasikan deteksi objek lokal, pilihan pesan kebutuhan, konfirmasi pengguna, dan *text-to-speech* dalam satu prototipe AAC berbasis Android. Pada dataset 500 citra dan lima kelas objek, YOLOv8n memperoleh *precision* 0,8475 ± 0,0652, *recall* 0,8267 ± 0,0837, F1-score 0,8328 ± 0,0241, mAP50 0,8768 ± 0,0338, dan mAP50–95 0,6115 ± 0,0358. Dibandingkan YOLOv8n, YOLOv5n unggul tipis pada mAP serta efisiensi ukuran dan parameter, sedangkan YOLOv8n menghasilkan *recall* dan F1-score sedikit lebih tinggi serta *benchmark* Tesla T4 sedikit lebih cepat. Hasil tersebut menunjukkan bahwa kedua model kompetitif dan pemilihan model perlu mempertimbangkan beberapa indikator secara bersamaan.

YOLOv8n berhasil diekspor ke LiteRT/TFLite dan diterapkan pada Android. Pengujian 100 percobaan menghasilkan latensi *end-to-end* 768,5 ± 126,4 ms, tidak ditemukan *crash*, keberhasilan *text-to-speech* sebesar 98%, dan keberhasilan keseluruhan sistem sebesar 96%. Validasi tiga pakar menghasilkan Aiken's V rerata 0,908 dan UAT lima guru menghasilkan nilai keseluruhan 92,8%. Dengan demikian, tujuan penelitian untuk mengevaluasi performa deteksi, implementasi Android, serta kesesuaian awal prototipe telah tercapai.

Kontribusi utama Loka Wicara bukan sekadar mengenali objek melalui kamera, tetapi menghubungkan objek keseharian dengan beberapa pilihan pesan kebutuhan berbahasa Indonesia yang dikonfirmasi sebelum disampaikan melalui suara. Meskipun demikian, penelitian ini hanya menunjukkan keberhasilan teknis dan penerimaan awal prototipe dan belum membuktikan bahwa Loka Wicara dapat meningkatkan kemampuan komunikasi anak ASD.

Penelitian selanjutnya disarankan memperluas dataset dengan variasi objek fisik, latar, pencahayaan, sudut pengambilan, skala, dan oklusi; menggunakan pemisahan data berbasis sesi atau objek fisik untuk mencegah kebocoran sumber; melibatkan dua anotator independen; serta membandingkan performa LiteRT pada beberapa kelas perangkat Android. Pengujian langsung bersama pengguna AAC perlu dilakukan secara bertahap dengan persetujuan etik, persetujuan orang tua atau wali, pendamping yang sesuai, mekanisme penghentian partisipasi, dan ukuran hasil komunikasi yang telah ditentukan sebelumnya.

DAFTAR PUSTAKA

Al'azmi, N. U., A. Wijiastuti, and W. Wagino. 2025. "Pengembangan Aplikasi Augmentative and Alternative Communication (AAC) Untuk Meningkatkan Komunikasi Ekspresif Anak Dengan Autisme Berbasis Etnopedagogi." *Jurnal ORTOPEDAGOGIA* 11 (2): 397–401. https://doi.org/10.17977/um031v11i22025p397-401.

Bateman, K. J., S. E. Wilson, A. Gauvreau, et al. 2023. "Visual Supports to Increase Conversation Engagement for Preschoolers with Autism Spectrum Disorder during Mealtimes: An Initial Investigation." *Journal of Early Intervention* 45 (2): 163–84. https://doi.org/10.1177/10538151221111762.

Boyev, Alexander, Irina Rabaev, Nitzan Cohen, and Hadas Chassidim. 2025. "Using a Video Visual Scene Display Application to Improve Communication for Individuals with Autism Disorder." *Journal of Developmental and Physical Disabilities* 37 (2): 199–215. https://doi.org/10.1007/s10882-024-10000-w.

Casas, Edmundo, Leo Ramos, Eduardo Bendek, and Francklin Rivas-Echeverria. 2024. "YOLOv5 vs. YOLOv8: Performance Benchmarking in Wildfire and Smoke Detection Scenarios." *Journal of Image and Graphics* 12 (2): 127–36. https://doi.org/10.18178/joig.12.2.127-136.

Chavers Edgar, T., R. W. Schlosser, and R. Koul. 2024. "Effects of an Augmentative and Alternative Communication Intervention Package on Socio-Communicative Behaviors between Minimally Speaking Autistic Children and Their Peers." *American Journal of Speech-Language Pathology* 33 (4): 1619–38. https://doi.org/10.1044/2024_AJSLP-23-00313.

Genc-Tosun, D., O. Kurt, Z. Cevher, and E. V. Gregori. 2023. "Teaching Children with Autism Spectrum Disorder to Answer Questions Using an iPad-Based Speech-Generating Device." *Journal of Autism and Developmental Disorders* 53 (9): 3724–39. https://doi.org/10.1007/s10803-022-05683-z.

Griffen, B., C. Holyfield, E. R. Lorah, and N. Caldwell. 2024. "Increasing Linguistic and Prelinguistic Communication for Social Closeness during Naturalistic AAC Instruction with Young Children on the Autism Spectrum." *Augmentative and Alternative Communication* 40 (3): 168–81. https://doi.org/10.1080/07434618.2023.2283846.

Iskandar, Riyadi Jimmy, Chastine Fatichah, and Anny Yuniarti. 2024. "Object Detection in Low-Light Conditions: A Comparison Using YOLOv5 and YOLOv8." 558–63. https://doi.org/10.1109/ICSINTESA62455.2024.10748090.

Mittal, P. 2024. "A Comprehensive Survey of Deep Learning-Based Lightweight Object Detection Models for Edge Devices." *Artificial Intelligence Review* 57: Article 242. https://doi.org/10.1007/s10462-024-10877-1.

More, S. S., and N. Patil. 2025. "Empowering the Visually Impaired: YOLOv8-Based Object Detection in Android Applications." *Procedia Computer Science* 252: 457–69. https://doi.org/10.1016/j.procs.2025.01.005.

Muharib, R., V. L. Walker, and W. Dunn. 2024. "Effects of Interventions Involving Tablet-Based Speech-Generating Devices for Individuals with ASD: A Meta-Analysis." *Journal of Autism and*

Developmental Disorders 54 (12): 4496–514. https://doi.org/10.1007/s10803-023-06173-6.

Saleh, W. A., and N. M. Mutahara. 2023. "Augmentatif Dan Alternatif Komunikasi Dalam Pengembangan Kemampuan Komunikasi Anak Autis." *SPEED Journal: Journal of Special Education* 6 (2): 177–82. https://doi.org/10.31537/speed.v6i2.936.

Sari, B. P., R. W. Sholikah, and R. V. H. Ginardi. 2024. "The Development of Mobile Application for Object Recognition Based on Deep Learning to Assist People with Visually Impaired." 222–27. https://doi.org/10.1109/ISITDI62380.2024.10796984.

Schlosser, R. W., and R. Koul. 2023. "Advances in Augmentative and Alternative Communication Research for Individuals with Autism Spectrum Disorder: Moving Research and Practice Forward." *Augmentative and Alternative Communication* 39 (1): 2–6. https://doi.org/10.1080/07434618.2023.2181214.

Sirisha, U., S. Phani Praveen, P. Naga Srinivasu, P. Barsocchi, and A. K. Bhoi. 2023. "Statistical Analysis of Design Aspects of Various YOLO-Based Deep Learning Models for Object Detection." *International Journal of Computational Intelligence Systems* 16: Article 126. https://doi.org/10.1007/s44196-023-00302-w.

Terven, J., D. M. Cordova-Esparza, and J. A. Romero-Gonzalez. 2023. "A Comprehensive Review of YOLO Architectures in Computer Vision: From YOLOv1 to YOLOv8 and YOLO-NAS." *Machine Learning and Knowledge Extraction* 5 (4): 1680–716. https://doi.org/10.3390/make5040083.

World Health Organization. 2025. "Autism." September 17. https://www.who.int/news-room/fact-sheets/detail/autism-spectrum-disorders.

Yau, S. H., K. Choo, J. Tan, O. Monson, and S. Bovell. 2024. "Comparing and Contrasting Barriers in Augmentative Alternative Communication Use in Nonspeaking Autism and Complex Communication Needs: Multi-Stakeholder Perspectives." *Frontiers in Psychiatry* 15: Article 1393017. https://doi.org/10.3389/fpsyt.2024.1393017.

Zervogianni, V., S. Fletcher-Watson, G. Herrera, et al. 2024. "A User-Based Information Rating Scale to Evaluate the Design of Technology-Based Supports for Autism." *Universal Access in the Information Society* 23: 1783–94. https://doi.org/10.1007/s10209-023-00997-y.