



Identification of Indonesian Sign Language System Using Deep Learning in Yolo-based

Arahmad Taupiq

Department of Informatics, Dinamika Bangsa University, Jambi, Indonesia

E-mail: arahmadtauoq612@gmail.com

*Corresponding Author

Muhammad Wildan Fajri

Department of Informatics, Dinamika Bangsa University, Jambi, Indonesia

E-mail: wildanfjari2020@gmail.com

Dannylee

Department of Informatics, Dinamika Bangsa University, Jambi, Indonesia

E-mail: hayagura181@gmail.com

Received: 17 May, 2024; Accepted: 12 June, 2024; Published: 30 June, 2024

Abstract: Deafness, or hearing impairment, refers to the loss of auditory capability in one or both ears. Deaf communities often develop sign languages to facilitate communication. Sign language, which employs hand movements, is commonly adopted by individuals with hearing impairments. In Indonesia, two primary sign languages are used: BISINDO (Bahasa Isyarat Indonesia) and SIBI (Sistem Isyarat Bahasa Indonesia). The main distinction between these languages is that BISINDO employs both hands for signing, whereas SIBI uses only one hand. Individuals with hearing impairments face significant communication challenges. This study focuses on the detection of alphabets in the Indonesian Sign Language System (SIBI) using YOLO v5. The objective is to recognize alphabetic characters through hand gesture signals. Experimental results indicate a detection success rate of 95.38%, accurately identifying 23 out of the 24 tested letters.

Keywords: SIBI, Computer Vision, Yolo, Sign Language Recognition, Deep Learning

I. Introduction

Bahasa merupakan pesan yang disampaikan dalam bentuk ekspresi sebagai alat komunikasi pada situasi tertentu dalam berbagai aktivitas. Dalam hal ini ekspresi berkaitan unsur segmental dan suprasegmental baik itu lisan atau kinesik sehingga sebuah kalimat akan bisa berfungsi sebagai alat komunikasi dengan pesan yang berbeda apabila disampaikan dengan ekspresi yang berbeda[1]. Komunikasi menjadi kendala bagi sebagian orang, terutama bagi mereka penyandang disabilitas dengan keterbatasan di indera pendengaran yaitu tunarungu.

Tunarungu adalah istilah yang mengacu pada seseorang yang mengalami gangguan pendengaran sehingga tidak dapat mendengar suara atau memiliki gangguan pendengaran yang signifikan[2]. Istilah ini sering kali digunakan untuk menggambarkan mereka yang lahir dengan gangguan pendengaran atau mengalami kehilangan pendengaran pada tahap perkembangan tertentu dalam hidup mereka. Keterbatasan dalam pendengaran menyebabkan penyandang tunarungu tidak mampu berkomunikasi dengan baik. Adanya gangguan komunikasi secara tidak langsung menjadi kesulitan dalam berinteraksi sehingga penyandang tunarungu menjadi terisolasi[3]. Oleh karena itu, komunitas tunarungu sering kali mengembangkan bahasa isyarat untuk berkomunikasi satu sama lain salah satunya menggunakan Sistem Isyarat Bahasa Indonesia(SIBI).

Sistem Isyarat Bahasa Indonesia(SIBI) merupakan sistem gestur isyarat yang digunakan dalam Bahasa Isyarat Indonesia. Dirancang untuk menerjemahkan bahasa isyarat menjadi teks atau suara, SIBI memiliki katalog gestur yang merepresentasikan abjad, kata, frasa, dan kalimat dalam Bahasa Isyarat Indonesia. Sistem ini penting dalam meningkatkan aksesibilitas komunikasi bagi komunitas tunarungu di Indonesia, memungkinkan interaksi dengan non-tunarungu, serta membuka jalan untuk integrasi teknologi guna memperluas cakupan akses informasi. SIBI telah dibakukan oleh pemerintah melalui Keputusan Mendikbud No. 0161/U/2994 tanggal 30 Juni 1994 tentang Pembakuan Sistem Isyarat Bahasa Indonesia[4].

Pada era modern saat ini, teknologi dapat menjadi solusi dalam berbagai sektor kehidupan manusia, salah satunya dalam hal berkomunikasi, masih banyak orang awam yang tidak paham dengan bahasa isyarat yang menjadi kendala komunikasi antara penyandang tunarungu dengan manusia lainnya. Computer vision dapat menjadi solusi untuk menjembatani hal tersebut, yaitu dengan otomisasi penerjemahan bahasa isyarat tangan oleh komputer sehingga orang awam-pun dapat mengerti apa yang dikatakan oleh penyandang tunarungu. Oleh karna itu penelitian ini berfokus untuk menerapkan computer vision pada pendeteksi Alfabet dengan menggunakan YOLO.

Computer Vision merupakan teknologi yang membuat komputer dapat melihat, mendeteksi dan memproses gambar layaknya penglihatan manusia, kemudian komputer akan menampilkan hasil yang sesuai dengan input yang diberikan. Computer Vision dapat diterapkan di bagian pengawasan, menghitung objek, melacak objek, mendeteksi objek, bahkan sampai pada pengenalan objek[5]. Sebelumnya telah dilakukan penelitian sejenis terkait dengan pengenalan bahasa isyarat tangan menggunakan beberapa metode diantaranya yaitu menggunakan metode klasifikasi objek yaitu Haar Classifier pada alfabet Bahasa Isyarat Indonesia(Bisindo) alphabet yang diterjemahkan pada penelitian ini yaitu huruf A, B, C, D, E, F, G, H, I, K, L, M, N, O, P, Q, S, T, U, V, W, X, Y, dan Z. Huruf J dan R tidak dapat diterjemahkan karena berbentuk gerakan tangan. Dari hasil pengujian tersebut didapatkan rata-rata hasil akhir yang di dapat adalah 91,8 %[6]. Pada penelitian yang lain telah dilakukan pengklasifikasian bahasa isyarat indonesia menggunakan metode Convolutional Neural Network dan arsitektur MobilenetV2 menggunakan tensorflow. Berdasarkan pengujian model pada penelitian tersebut, tingkat akurasi yang dihasilkan mencapai 54,8% dalam mengklasifikasi 30 bahasa isyarat [7].

Selain itu terdapat juga sebuah penelitian Deteksi Tangan Otomatis Pada Video Percakapan Bahasa Isyarat Indonesia Menggunakan Metode Deep Gated Recurrent Unit (GRU). Penelitian tersebut menggunakan 3 class untuk percakapan menggunakan bahasa isyarat yaitu ucapan hallo, terimakasih, dan sama sama, pada penelitian tersebut menggunakan Pengujian terhadap 45 video data training dan 36 video data testing. Pengujian tersebut yang dilakukan terhadap data testing menghasilkan nilai akurasi sebesar 88%[8]. Yolo menjadi salah satu algoritma pendeteksi objek yang sedang trend saat ini, oleh karna itu pada penelitian ini menggunakan algoritma YOLO v5 sebagai algoritma pendeteksi alfabet pada Sistem Isyarat Bahasa Indonesia(SIBI).

YOLO (You Only Look Once) merupakan Salah satu metode atau pendekatan pada Computer Vision yaitu merupakan algoritma yang dapat memproses pendeteksi objek[9], pengawasan video, atau analisis citra. YOLO membagi gambar menjadi grid dan melakukan prediksi objek serta kotak pembatas (bounding box) serta probabilitas kelas objek dalam setiap grid secara simultan. Hal ini memungkinkan YOLO untuk bekerja secara cepat dan memberikan deteksi objek yang akurat secara real time. Pada penelitian ini memanfaatkan algoritma YOLO V5 sebagai training dataset. Sistematika penulisan pada penelitan ini selanjutnya adalah pada bagian kedua berisi metode yang disusulkan. Ketiga adalah hasil dan diskusi, serta terakhir adalah kesimpulan.

2. Research Method

2.1. Experiment Setup

Penelitian ini bertujuan untuk mendeteksi huruf alfabet dengan gestur tangan pada SIBI dengan menggunakan YOLO v5. Adapun beberapa tahapan penelitian yang harus dilakukan untuk dapat mencapai tujuan dari penelitian ini, maka perlu dirancang alur penelitian yang ditampilkan pada Fig 1.



Fig.1. Experiment setup

Fig.1 adalah diagram eksperimen yang dirancang khusus untuk penelitian ini. Oleh karena itu, proses penelitian dibagi menjadi beberapa tahap terpisah.

- Sebagai langkah awal, melakukan pengumpulan dataset berbentuk image yang nantinya dataset tersebut akan ditraining dengan menggunakan *YOLO v5*.
- Langkah kedua adalah melakukan labelling image berdasarkan kelas objek tersebut dari dataset yang telah dikumpulkan.
- Langkah ketiga yaitu melakukan training dengan menggunakan *YOLO v5* pada dataset yang sudah dilabelli.
- Pada Langkah terakhir adalah melakukan pengujian.

2.2. Dataset

Penelitian ini memanfaatkan dataset Sistem Isyarat Bahasa Indonesia(SIBI) yang terdapat pada website *Kaggle* sebagai sumber data. Terdapat 5280 image tangan dengan gesture alfabet SIBI yang diklasifikasikan menjadi 24 kategori yang berbeda yaitu A, B, C, D, E, F, G, H, I, K, L, M, N, O, P, Q, R, S, T, U, V, W, X, Y, pada Fig 2 merupakan beberapa contoh dari isi dataset *SIBI* yang didapatkan dari website *Kaggle* yang belum dilakukan pelabellan.



Fig.2.Contoh Dataset

2.3. Labelling Image

Pada penelitian ini memanfaatkan sebuah aplikasi berbasis web yaitu *Roboflow* yang digunakan sebagai alat untuk melabeli seluruh image yang ada di dataset sesuai dengan kategorinya masing masing. Pada tahap ini juga dilakukan dataset split yang membagi 5280 image dataset menjadi 3 bagian yaitu : 3696 (70%) Train Set, 1056 (20%) Valid Set, Serta 528 (10%) Test set, seperti yang dapat dilihat pada Fig 3 berikut:



Fig.3. Labelling Image

2.3. Proposed Method

Training dataset bertujuan untuk menghasilkan sebuah bobot yang akan digunakan dalam proses deteksi alfabet *SIBI*. Dataset akan diproses oleh program untuk menggunakan *YOLO v5* sebagai deteksi. Metode *YOLO* memproses pendeteksian dan pengenalan objek menggunakan sebuah jaringan syaraf tunggal (single neural network), yang memprediksi koordinat dengan kotak pembatas (*Bounding Box*) pada objek dan probabilitas kelas secara langsung dalam satu deteksi. *YOLO* memiliki arsitektur yang terdiri atas lapisan konvolusional (convolutional layer) sebanyak 24 lapisan dengan 4 lapisan max pooling, 2 lapisan yang terhubung penuh (fully connected later). Lapisan reduksi 1x1 yang diterapkan di beberapa lapisan konvolusi sebagai alternatif untuk mengurangi kedalaman feature maps [10]. Arsitektur dari *YOLO* pertama kali diperkenalkan oleh Joseph Redmon ditunjukkan seperti pada Fig 4s di bawah ini.

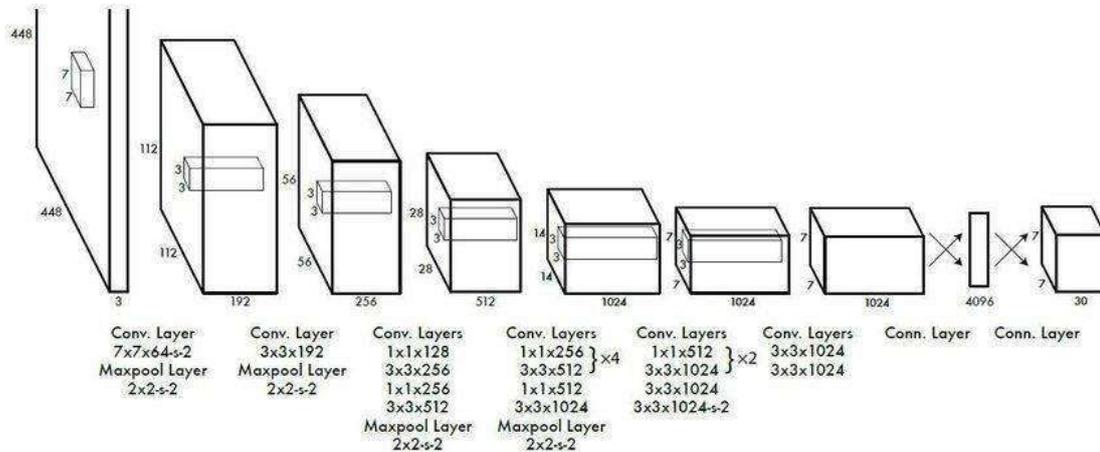


Fig.4. Arsitektur *YOLO*

Program akan mendeteksi tangan dengan setiap gestur yang telah diberikan di dataset dengan menggunakan algoritma *YOLO* yang berada pada citra tersebut. Lalu citra akan diberi bounding box dan text ditampilkan sebagai keluaran. Proses training data dilakukan menggunakan *YOLO v5* dengan *pytorch*. Training dilakukan menggunakan *Google Colaboratory*. Proses training berlangsung selama 2 jam 30 menit dengan training menggunakan 200 epochs, dengan img size 416 dan batch size 16 dengan jumlah 24 kelas

2.4. Environment Setup

Eksperimen dalam penelitian ini dilaksanakan pada sebuah komputer dengan spesifikasi yaitu sistem operasi Windows 10 Pro 64-bit, prosesor AMD Ryzen 5 PRO 2500U, Radeon Vega Mobile Gfx (dengan 8 CPU), 2.0 GHz, serta RAM berkapasitas 16 GB. Selain itu, tool analisis yang diperlukan dalam penelitian meliputi Python, roboflow, scikit-learn.

3. Results and Discussion

Bagian ini mencakup hasil eksperimen yang telah dilaksanakan dari beberapa tahapan di atas, termasuk laporan hasil dari training dataset yang dilakukan dengan menggunakan *YOLO v5* (*You Only Look Once Versi 5*) sebagai algoritma pendeteksian objek, serta melakukan uji coba dari hasil.

3.1. Hasil Training Dataset

Dataset yang telah ditraining akan menghasilkan dua buah file bernama “best.pt” dan “last.pt”, yang merupakan output dari training dataset dengan *YOLO* yang berupa file bobot(weights) dari model yang telah ditraining sebelumnya (pretrained model). Ini berisi nilai-nilai dari parameter-parameter model yang telah diatur selama proses training.

Anda dapat melihat Fig 5 di bawah sebagai kurva hasil dari training yang telah di lakukan, terdapat 2 sumbu yaitu sumbu X dan Sumbu Y, sumbu X mewakili banyak nya training(*epochs*) yang dilakukan, kurva ‘train’ merupakan hasil dari training yang dilakukan, sedangkan ‘val’ merupakan hasil validasi atau pengujian dari training yang dilakukan. Jika nilai dari ‘box_loss’, ‘obj_loss’, dan ‘cls_loss’ semakin lama semakin mendekati nilai ‘0’ maka semakin bagus, sebaliknya pada kurva ‘metrics’ jika nilainya semakin lama mendekati ‘1’ maka akan semakin bagus.

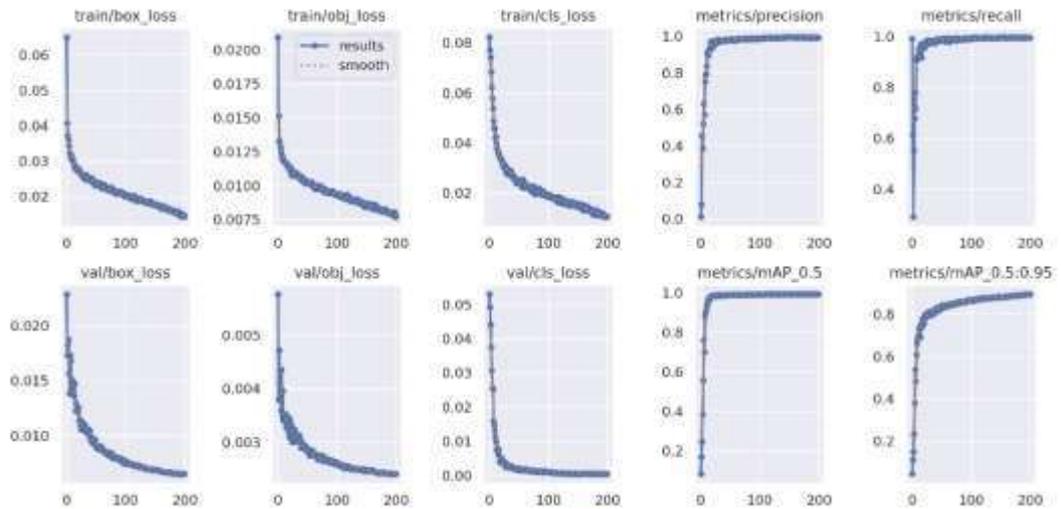


Fig.5. Kurva Hasil

Pada Fig 6 berikut terdapat ‘Confusion Matrix’ yang dimana merupakan nilai error atau kekeliruan deteksi dari training yang dilakukan, hasil Confussion Matrix dinilai bagus jika membentuk sebuah garis lurus diagonal dan berisi nilai ‘1’ yang dimana berarti objek yang dideteksi memang benar, sedangkan nilai yang berada di luar garis diagonal merupakan pendeteksian error dari sebuah objek yang dikenali sebagai objek yang lain.

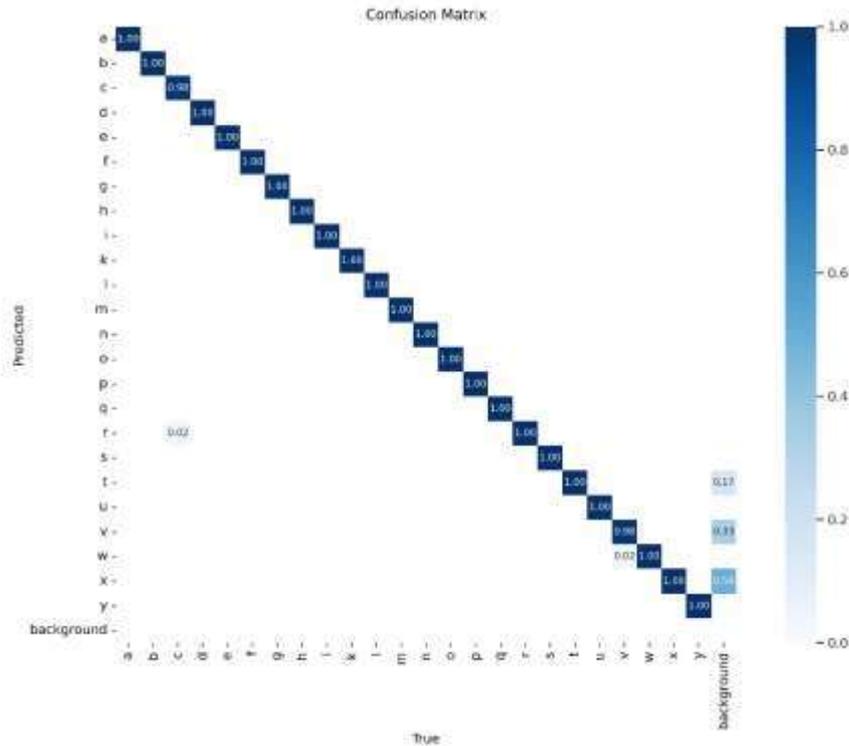


Fig.6. Confussion Matrix



Fig.7. Train Batch 0

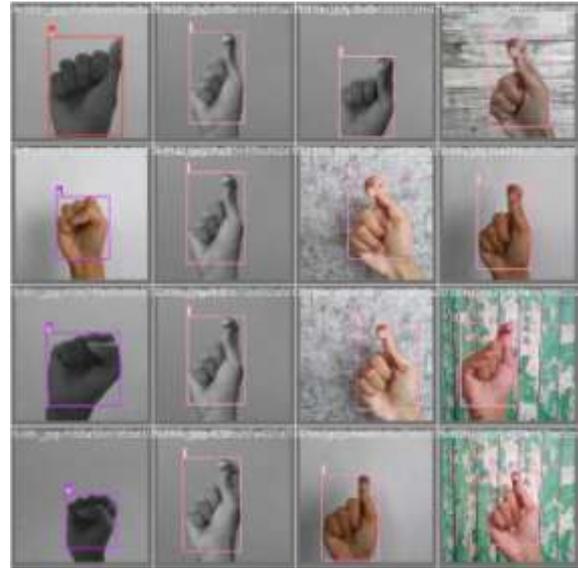


Fig.8.Val Batch 0

Pada Fig 7 merupakan proses pelatihan batch ke 0 yang dimana model akan mempelajari setiap pola gestur tangan dari *Sistem Isyarat Bahasa Indonesia(SIBI)*, sedangkan pada Fig 8 merupakan hasil batch ke 0 proses validasi model dari training yang telah dilakukan sebelumnya

3.2. Hasil Deteksi Yolo

Pada tahap ini penelitian akan menguji coba file “best.pt” yang merupakan hasil dari training menggunakan *YOLO* yang telah dilakukan untuk mendeteksi gestur isyarat tangan alfabet dari *SIBI*, pengujian ini akan menggunakan kamera webcam secara real-time yang akan mendeteksi gestur dari tangan penguji, hasil dari uji coba dapat dilihat pada Table 1

Tabel 1. Hasil pengujian deteksi Random forest

Huruf	Pengujian	Berhasil	Tidak berhasil	Ket	Huruf	Pengujian	Berhasil	Tidak berhasil	Ket
A		✓			B		✓		
C		✓			D		✓		
E		✓			F		✓		
G		✓			H		✓		
I		✓			K		✓		

L		✓		M		✓	
N		✓		O		✓	
P			✓	Huruf 'P' dideteksi sebagai huruf lain	Q		✓
R		✓		S		✓	
T		✓		U		✓	
V		✓		W		✓	
X		✓		Y		✓	

Pada uji coba hampir semua huruf dapat terdeteksi dengan baik, kecuali huruf “P” yang dikenali oleh model training sebagai huruf lain, serta muka penguji dikenali sebagai huruf. Hasil pengujian memberikan hasil sebanyak 23 alfabet dapat dideteksi dari 24 yang artinya keberhasilan model training untuk mendeteksi alfabet dari *Sistem Isyarat Bahasa Indonesia(SIBI)* adalah 95,83%. Tentunya pengujian dapat mengalami hasil yang berbeda terkait dengan parameter-parameter saat pengujian dilakukan seperti resolusi kamera, pencahayaan, background dan lain sebagainya. Model training ini masih belum dapat mendeteksi huruf ‘J’ dan ‘Z’ dikarenakan kedua huruf tersebut isyarat-nya menggunakan sebuah gerakan, serta model training ini masih kesulitan dalam mendeteksi jika ada objek lain di belakangnya

4. Conclusion

Berdasarkan hasil seluruh tahapan dalam penelitian yang telah dikembangkan dengan Metode *You Only Look Once versi 5 (YOLOv5)* dapat diimplementasikan dalam mendeteksi objek gambar dari sebuah gesture tangan serta membedakan 24 objek yang meliputi gambar tangan berbentuk isyarat alfabet *SIBI* yang terdiri dari huruf A, B, C, D, E, F, G, H, I, K, L, M, N, O, P, Q, R, S, T, U, V, W, X, dan Y, dengan menggunakan dataset yang terdiri dari 5280 image dataset dibagi menjadi 3 bagian yaitu : 3696 (70%) Train Set, 1056 (20%) Valid Set, Serta 528 (10%) test set , dengan melakukan training pada *YOLO v5* dengan parameter 200 epochs, img size 416 dan batch size 16 dengan jumlah 24 kelas. Hasil dari model training diuji coba dengan menggunakan kamera webcam secara real-time, tingkat keberhasilan pendeteksian alfabet isyarat adalah 95,38% sebanyak 23 huruf yang dideteksi benar dari 24 huruf. Namun masih terdapat beberapa kekurangan di penelitian ini di antaranya model training masih belum dapat mendeteksi huruf ‘J’ dan ‘Z’ dikarenakan kedua huruf tersebut menggunakan sebuah gerakan sebagai isyarat-nya, serta model training ini masih kesulitan dalam mendeteksi jika ada objek lain di belakangnya

Acknowledgment

Penelitian ini didukung oleh Universitas Dinamika Bangsa, Jambi, Indonesia

References

- [1] Noermanzah, "Bahasa sebagai Alat Komunikasi, Citra Pikiran, dan Kepribadian," in *Prosiding Seminar Nasional Bulan Bahasa (Semiba)*, 2019, pp. 306–319. [Online]. Available: <https://ejournal.unib.ac.id/index.php/semiba>
- [2] M. B. S. Bakti and Y. M. Pranoto, "Pengenalan Angka Sistem Isyarat Bahasa Indonesia Dengan Menggunakan Metode Convolutional Neural Network," *Semin. Nas. Inov. Teknol.*, vol. 1, no. 1, pp. 11–16, 2019.
- [3] Y. U. Solikhatusun, "Penyesuaian Sosial Pada Penyandang Tunarungu di SLB Negeri Semarang," *Educ. Psychol. J.*, vol. 1, no. 1, pp. 65–72, 2013.
- [4] Dr. Samto, "Kamus SIBI Kerjasama antara Kementerian Pendidikan dan Kebudayaan dengan Lembaga Penelitian dan Pengembangan Sistem Isyarat Bahasa Indonesia," *Kamus SIBI - Direktorat PMPK*.
- [5] V. M. Lumabiang, C. Andris, L. Manaha, and A. T. Liem, "Prototipe Pendeteksi Object Menggunakan Computer Vision dan Raspberry Pi," in *SENSITif 2019*. Manado, 2020, pp. 1341–1351.
- [6] R. I. Borman, B. Priyopradono, and A. R. Syah, "Klasifikasi Objek Kode Tangan pada Pengenalan Isyarat Alphabet Bahasa Isyarat Indonesia (BISINDO)," in *Seminar Nasional Informatika dan Aplikasinya (SNIA)*, 2018, no. September, pp. 1–4.
- [7] Nasha Hikmatia A.E. and M. I. Zul, "Aplikasi Penerjemah Bahasa Isyarat Indonesia menjadi Suara berbasis Android menggunakan Tensorflow," *J. Komput. Terap.*, vol. 7, no. 1, pp. 74–83, 2021, doi: 10.35143/jkt.v7i1.4629.
- [8] P. Kurnia Sari, G. Qorik Oktagalu Pratamasunu, and F. Nur Fajri, "Deteksi Tangan Otomatis Pada Video Percakapan Bahasa Isyarat Indonesia Menggunakan Metode Deep Gated Recurrent Unit (GRU)," *J. Komput. Terap.*, vol. 8, no. 1, pp. 186–193, 2022, doi: 10.35143/jkt.v8i1.4901.
- [9] A. A. B, A. Amin, and M. W. Kasrani, "Penerapan Metode Yolo Object Detection V1 Terhadap Proses Pendeteksian Jenis Kendaraan Di Parkiran," *J. Tek. Elektro Uniba (JTE UNIBA)*, vol. 6, no. 1, pp. 194–199, 2021, doi: 10.36277/jteuniba.v6i1.130.
- [10] J. H. Sri Wisna *et al.*, "Deteksi Kendaraan Secara Real Time Menggunakan Metode YOLO Berbasis Android," vol. 09, no. 01, pp. 8–14, 2020.

Authors' Profiles



Arahmad Taupiq lahir di Jambi, Indonesia. Saat ini dia tengah menempuh Program Sarjanah Informatika di Universitas Dinamika Bangsa, Indonesia. Dia memiliki fokus penelitian pada machine learning dan software engineering.



M Wildan Fajri lahir di Jambi, Indonesia. Saat ini dia tengah menempuh Program Sarjanah Informatika di Universitas Dinamika Bangsa, Indonesia. Dia memiliki fokus penelitian pada machine learning dan software engineering.



DANNYLEE lahir di Batam, Indonesia. Saat ini dia tengah menempuh Program Sarjanah Informatika di Universitas Dinamika Bangsa, Indonesia. Dia memiliki fokus penelitian pada machine learning dan software engineering.