

# APLIKASI PENGIDENTIFIKASI BAHASA ISYARAT BERDASARKAN GERAK TUBUH SECARA REAL TIME MENGGUNAKAN YOLO

Hizkia Halim<sup>\*1</sup>, Lina<sup>2</sup>

<sup>1,2</sup>Fakultas Teknologi Informasi Universitas Tarumanagara,  
e-mail: <sup>1</sup>hizkiahalim2412@gmail.com, <sup>2</sup>lina@untar.ac.id

*Bahasa isyarat merupakan bahasa yang umumnya digunakan oleh para tunarungu dan tunawicara untuk berkomunikasi. Bahasa isyarat menggunakan tangan sebagai sarana untuk memberikan informasi kepada lawan bicaranya. Cara berkomunikasi yang berbeda dengan masyarakat pada umumnya membuat masyarakat yang tidak mengerti bahasa isyarat kesulitan untuk mengerti informasi apa yang sedang disampaikan dalam bahasa isyarat. Penelitian ini bertujuan untuk membuat sebuah aplikasi yang dapat mengidentifikasi dan memprediksi arti dari 50 gerakan isyarat dalam Bahasa Isyarat Indonesia (BISINDO) secara langsung dari video yang ditangkap webcam dengan menggunakan metode You Only Look Once (YOLO). Dalam penelitian ini, arsitektur yang akan digunakan adalah arsitektur YOLOv5 dengan model YOLOv5s. Penelitian ini akan melakukan pelatihan dengan jumlah variasi pelatihan sebanyak 3 variasi. Berdasarkan penelitian yang sudah dilakukan, diperoleh akurasi validasi tertinggi sebesar 1. Sedangkan, berdasarkan pengujian yang sudah dilakukan, diperoleh akurasi tertinggi sebesar 67% untuk pengujian pertama dan 51,95% untuk pengujian kedua.*

**Kata Kunci:** Aplikasi, Bahasa Isyarat, Computer Vision, YOLO.

## I. PENDAHULUAN

Bahasa isyarat merupakan bahasa yang umum digunakan oleh orang – orang tunarungu dan tunawicara untuk berkomunikasi. Komunikasi ini terjadi dengan menggunakan tangan sebagai alat untuk menyampaikan informasi ke lawan bicara mereka. [1] Seperti bahasa pada umumnya, bahasa isyarat bukanlah bahasa yang universal. Setiap jenis bahasa isyarat memiliki beberapa perbedaan pada kosa katanya. Di Indonesia sendiri, terdapat 2 jenis bahasa isyarat yang biasa digunakan yaitu Bahasa Isyarat Indonesia (BISINDO) dan Sistem Bahasa Isyarat Indonesia (SIBI). [2]

Bahasa Isyarat Indonesia (BISINDO), merupakan bahasa isyarat yang digalakkan Gerakan Kesejahteraan Tunarungu Indonesia (GERKATIN) dan dikembangkan oleh masyarakat tunarungu sendiri. Berbeda dengan SIBI,

BISINDO merupakan bahasa isyarat yang lebih populer dan lebih sering digunakan oleh teman – teman tuli. Sayangnya, cara berkomunikasi yang sangat berbeda, tingkat kesulitan yang cukup tinggi, dan sedikitnya pengetahuan dasar masyarakat mengenai cara berinteraksi dengan menggunakan bahasa isyarat membuat para penyandang tunarungu dan tunawicara kesulitan untuk berinteraksi dengan masyarakat umum.

Computer Vision adalah salah satu cabang dari ilmu teknik informatika yang memungkinkan komputer melihat, mengamati, dan menganalisa objek di hadapannya. Selain melihat, mengamati, dan menganalisa objek, komputer juga bisa memproses lebih lanjut data yang sudah didapatkan dengan menggunakan deep learning. Dengan menggunakan Computer Vision, gerakan – gerakan yang dilakukan saat memperagakan bahasa isyarat bisa ditangkap dan diproses secara otomatis. Dengan menggunakan teknologi ini, diharapkan gerakan peragaan bahasa isyarat bisa diterjemahkan dan kedepannya tidak ada lagi kesulitan komunikasi antar pengguna bahasa isyarat dengan masyarakat yang tidak mengerti bahasa isyarat.

Beberapa penelitian yang terkait dengan penelitian ini adalah penelitian dengan judul “Penerjemahan Bahasa Isyarat Indonesia Menggunakan Kamera pada Telepon Genggam Android” yang melakukan deteksi gestur bahasa isyarat dan mengklasifikasikannya dengan metode SVM [3], penelitian dengan judul “Short Communication: Detecting Heavy Goods Vehicles in Rest Areas in Winter Conditions Using YOLOv5” yang membahas sebuah studi kasus penggunaan YOLOv5 untuk mendeteksi kendaraan barang berat saat musim salju [4], penelitian dengan judul “Pendeteksian Sel Darah Putih Dari Citra Preparat Dengan You Only Look Once (YOLO)” yang menggunakan YOLO untuk melakukan pendeteksian sel darah putih [5], dan penelitian dengan judul “Implementation of YOLO-V5 for a Real Time Social Distancing Detection” yang mengimplementasikan YOLOv5 untuk mendeteksi dan memperingatkan pelanggaran social distancing [6].

II. LANDASAN TEORI

A. Bahasa Isyarat Indonesia (BISINDO)

Bahasa Isyarat Indonesia (BISINDO) adalah salah satu media komunikasi yang membantu komunikasi sesama tunarungu. BISINDO adalah bahasa isyarat dari Indonesia yang berkembang secara alami di kalangan masyarakat tunarungu. BISINDO menggunakan isyarat jari, tangan, dan gerak untuk melambangkan kosa kata bahasa Indonesia. Untuk aplikasi yang dibangun ini, jumlah kata BISINDO yang akan dilatih berjumlah 50 kata. Berdasarkan hasil wawancara dengan Hambali yang merupakan salah satu teman tuli dan Muhammad Naufal seorang operation manager dari Sunyi Cafe yang mampu berkomunikasi dengan teman – teman tuli, perbedaan antara BISINDO dan SIBI terlihat dari gerakan tangannya. BISINDO menggunakan kedua tangan dalam mengisyaratkan abjad, sedangkan SIBI hanya menggunakan satu tangan saja. Sebagai salah satu teman tuli, Hambali lebih senang menggunakan BISINDO ketimbang SIBI karena lebih nyaman untuk digunakan.

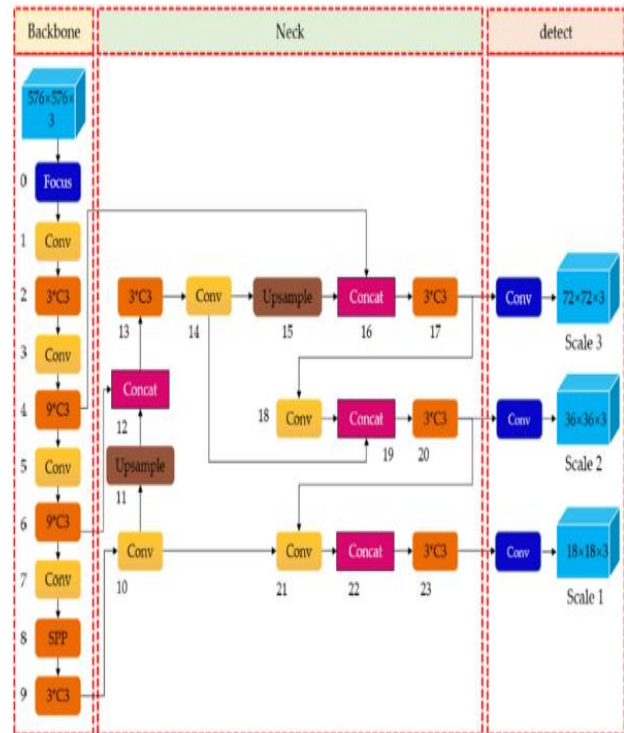
B. Dataset

Dataset pada penelitian ini merupakan gambar gerakan bahasa isyarat. Gambar ini diperoleh dengan melakukan wawancara dan perekaman video dengan teman – teman tuli (orang tunarungu) yang penulis temui di salah satu kafe di daerah Tangerang. Pengambilan video dilakukan dengan menggunakan kamera telepon genggam dengan pengaturan 1080p 30fps dan 720p 30fps. Setelah seluruh video diperoleh, maka akan dilakukan sebuah ekstraksi citra. Proses ekstraksi ini akan dilakukan kepada video yang akan dijadikan sebagai sumber data latih. Video – video ini memiliki durasi paling singkat 1 detik dan paling panjang 3 detik. Proses ekstraksi citra ini akan menghasilkan 30 citra untuk setiap videonya. Untuk video dengan durasi 1 detik, maka akan diambil citra sebanyak 30 citra dalam 1 detik. Untuk video yang berdurasi lebih 1 detik, dalam setiap detiknya akan diambil citra sebanyak 30 dibagi dengan durasi video. Setelah seluruh proses ekstraksi citra ini dilakukan, didapatkan data dengan jumlah 1500 citra untuk data latih. Untuk data uji akan digunakan data dalam bentuk video. Data uji ini akan memiliki 2 tipe yaitu data uji tipe pertama dengan 1 gerakan kata BISINDO dalam setiap videonya dan data uji tipe kedua dengan 2 gerakan kata BISINDO dalam setiap videonya. Jumlah data uji untuk tipe pertama akan berjumlah 50 data uji dan untuk video data uji tipe kedua akan berjumlah 20 data uji.

C. You Only Look Once (YOLO)

Dalam penelitian ini digunakan YOLOv5 sebagai model pendeteksi. YOLOv5 adalah model pendeteksi objek yang dirilis pada April 2020. YOLOv5 memiliki arsitektur yang mirip dengan generasi YOLO sebelumnya namun YOLOv5 menggunakan PyTorch sebagai pengganti Darknet. YOLOv5 dipilih karena YOLOv5 memiliki kecepatan dan akurasi yang lebih dibandingkan dengan YOLOv4. [7] Arsitektur YOLOv5 terbagi

menjadi tiga bagian yaitu backbone, neck, dan head (detect) dapat dilihat pada Gambar 1.



Gambar 1. Arsitektur YOLOv5

YOLOv5 memiliki head yang identik dengan head YOLOv3 [8]. YOLOv5 menghasilkan sebuah feature map yang akan berisikan confidence score, probabilitas kelas, dan koordinat serta ukuran dari bounding box (x, y, w, h).

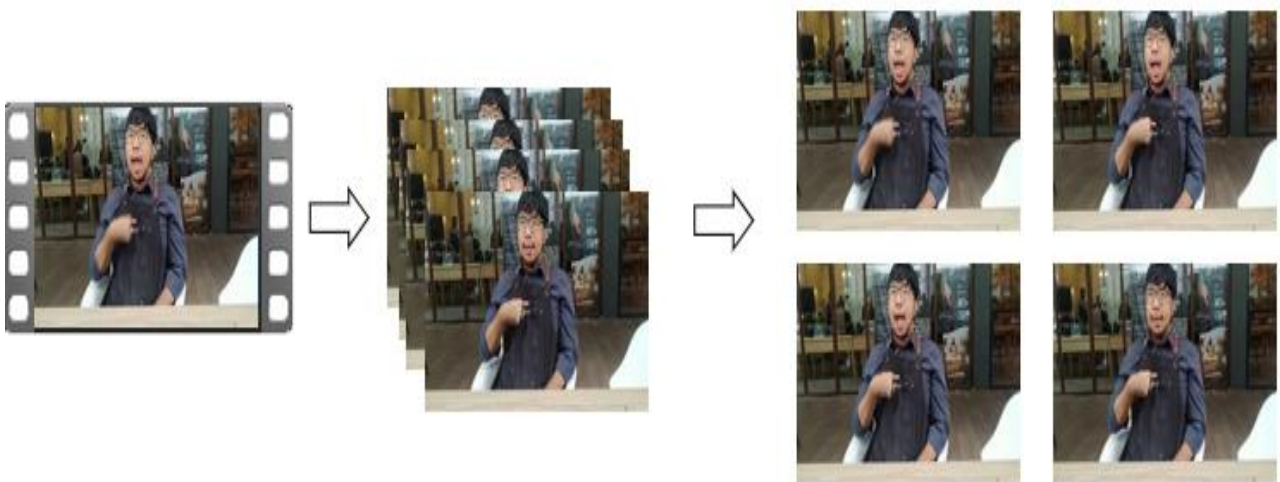
III. METODE PENELITIAN

Aplikasi penerjemah ini dibangun dengan menggunakan metode You Only Look Once (YOLO). YOLO akan digunakan dalam melakukan deteksi dan pengenalan gerakan isyarat yang ditangkap oleh kamera. Bahasa isyarat yang digunakan sebagai data dari penelitian ini merupakan video gerakan Bahasa Isyarat Indonesia (BISINDO). Video – video peragaan ini terbagi menjadi dua tipe. Tipe pertama merupakan video yang berisikan peragaan bahasa isyarat 1 kata, dan tipe kedua merupakan video yang berisikan peragaan bahasa isyarat 2 kata. Video yang akan digunakan sebagai sumber pelatihan/data latih adalah video tipe pertama.

Setiap video data latih diambil dengan bantuan dari teman – teman tuli. Video ini direkam dengan 30 frame per second, dan akan berisikan satu kata di tiap video. Setiap video yang diambil akan digunakan dalam penelitian ini apabila memenuhi beberapa kriteria yaitu sebagai berikut:

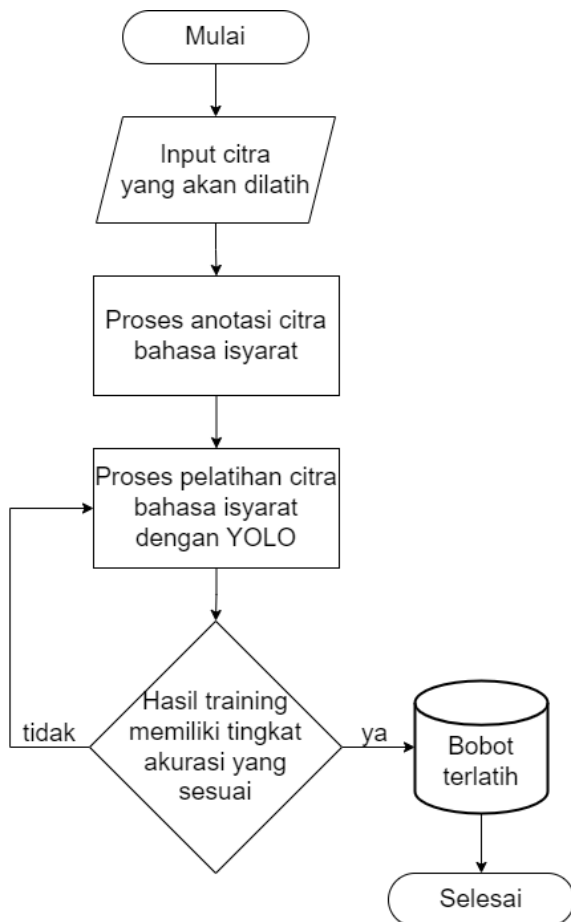
1. Peraga harus terlihat dari ujung kepala sampai ke perut atau pinggang.
2. Video harus memiliki pencahayaan yang baik.
3. Peragaan harus dilakukan mengarah ke arah kamera

Setelah data dikumpulkan, akan dilakukan pengambilan frame dari video dengan menggunakan library OpenCV sebanyak 30 frame untuk setiap videonya.



Gambar 2. Ilustrasi Proses Ekstraksi Citra

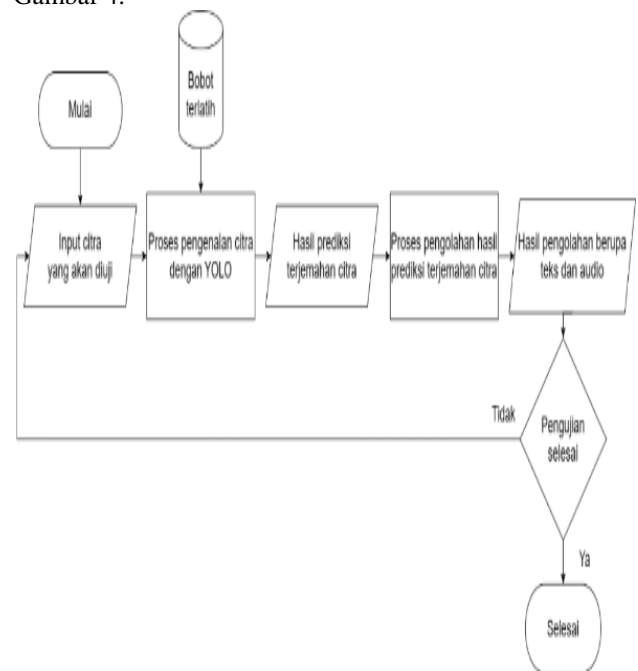
Setelah proses ekstraksi citra selesai dilakukan, akan dilakukan proses anotasi citra dengan menggunakan aplikasi labelImg. Dalam melakukan proses anotasi ini, bounding box akan diberikan kepada setiap citra. Hasil dari anotasi akan digunakan untuk melatih model YOLO dan hasil dari pelatihan ini akan ditimbang apakah akurasi sesuai. Diagram alur pelatihan dapat dilihat pada Gambar 3.



Gambar 3. Diagram Alur Proses Pelatihan

Setelah seluruh tahap pelatihan selesai dilakukan, maka akan dilakukan pengujian terhadap bobot terlatih. Pengujian ini akan menggunakan video selain dari data uji dan video tersebut akan dimasukkan ke dalam aplikasi

untuk selanjutnya diprediksi arti dari peragaan bahasa isyarat yang sedang dilakukan. Hasil prediksi dari model akan ditampilkan dan kemudian dibandingkan dengan arti sebenarnya untuk mengetahui kebenaran dari hasil prediksi. Diagram alur proses pengujian dapat dilihat pada Gambar 4.



Gambar 4. Diagram Alur Proses Pengujian

#### IV. HASIL DAN PEMBAHASAN

Dalam penelitian ini dilakukan 3 kali pelatihan model YOLOv5 dengan parameter epoch yang berbeda di setiap pelatihan. Ketiga jenis pelatihan itu adalah pelatihan dengan epoch 100, 200, dan 400. Parameter selain dari epoch dibuat sama untuk setiap epoch nya yaitu batch\_size 32 dan img 640.

##### A. Hasil Validasi

Setelah proses pelatihan model selesai, bobot – bobot terlatih akan melalui proses validasi. Proses validasi ini dilakukan dengan menggunakan keseluruhan citra data latih. Hasil Validasi dapat dilihat pada Tabel 1.

Tabel 1. Tabel Hasil Validasi

No	Epoch	Precision	Recall	F1-score
1	100	0.978	0.947	0.952
2	200	1	1	1
3	400	0.963	0.96	0.94

Berdasarkan hasil validasi pada Tabel 1, epoch 200 memiliki hasil terbaik dengan hasil precision, recall, dan F1-score sebesar 1.

**B. Hasil Pengujian**

Setelah melalui proses validasi, pengujian dilakukan dengan menggunakan data uji. Pengujian pertama dilakukan dengan menggunakan video yang berisikan 1 gerakan BISINDO untuk setiap video. Dari hasil pengujian pertama dengan data uji terhadap ketiga model, didapatkan hasil seperti pada Tabel 2.

Tabel 2. Tabel Hasil Pengujian Pertama

No	Epoch	Jumlah Prediksi Berhasil	Jumlah Prediksi Parsial	Akurasi
1	100	22	13	56.32%
2	200	29	9	67%
3	400	23	9	56.32%

Selanjutnya untuk pengujian kedua, video yang akan digunakan akan berisikan 2 gerakan BISINDO yang digerakan secara berkelanjutan untuk setiap video. Hasil dari pengujian kedua dapat dilihat pada Tabel 3.

Tabel 1. Tabel Hasil Pengujian Kedua

No	Jumlah Epoch	Jumlah Data	Akurasi Rata - Rata
1	100	20	39,15%
2	200	20	51,95%
3	400	20	35,5%

Dapat dilihat pada Tabel 2 dan Tabel 3, akurasi terbaik berhasil didapatkan oleh epoch 200 dengan akurasi 67% untuk pengujian pertama dan 51,95% untuk pengujian kedua. Angka pada kolom “Jumlah Prediksi Parsial” pada Tabel 4.2 merupakan jumlah prediksi benar yang mengalami penambahan pada bagian sebelum atau sesudah kata prediksi yang tepat. Penambahan kata ini terjadi dikarenakan adanya kesalahan klasifikasi. Setelah dilakukan pengamatan, kesalahan klasifikasi ini terjadi ketika model mengidentifikasi gerakan yang dilakukan peraga saat melakukan transisi sebagai sebuah kata. Kesalahan klasifikasi ini juga bisa terjadi ketika sebuah gerakan isyarat memiliki unsur gerakan yang identik dan

penempatan gerakan yang mirip dengan gerakan lainnya seperti gerakan foto dan bulan yang dapat dilihat pada Gambar 5.



Gambar 5. Gerakan Foto (kiri) dan Gerakan Bulan (kanan)

Kedua gerakan ini memiliki kemiripan pada bagian bentuk tangan yang seperti membuat angka tujuh atau seperti menunjuk ke arah atas dan perbedaan dari kedua foto diatas adalah gerakan kata “foto” menggunakan kedua tangan sedangkan gerakan “bulan” menggunakan satu tangan saja. Hal ini membuat gerakan kata “foto” berpotensi dikenali oleh komputer sebagai kata “bulan” dan menghasilkan sebuah prediksi yang memiliki tambahan kata di bagian sebelum atau sesudah kata yang tepat. Tentunya hasil prediksi yang tidak sempurna ini akan berdampak pada perhitungan akurasi yang dilakukan. Dalam melakukan perhitungan akurasi penelitian ini menggunakan sebuah persamaan yaitu sebagai berikut:

$$Akurasi = \left(\frac{1}{n} \sum_{i=1}^n x_i\right) * 100\% \tag{1}$$

Keterangan:

n = jumlah data uji

xi = nilai ketepatan setiap pengujian

Nilai xi pada Persamaan 1 didapatkan dengan membandingkan apakah prediksi dari model sudah sesuai dengan ground truth. Apabila hasil prediksi benar maka nilai ketepatan adalah 1, apabila hasil prediksi salah atau tidak mengandung kata yang benar maka nilai ketepatan adalah 0, dan apabila hasil prediksi mengandung kata yang benar namun terdapat kesalahan klasifikasi pada hasil prediksinya maka nilai ketepatan akan dihitung dengan cara membagi jumlah kata pada ground truth dengan jumlah kata dalam prediksi.

**V. KESIMPULAN**

Penelitian ini dilakukan dengan tujuan untuk membantu orang – orang yang tidak mengerti bahasa isyarat agar mampu mengerti arti dari gerakan bahasa isyarat yang sedang diperagakan sehingga pengguna aplikasi ini dapat memahami informasi apa yang disampaikan dalam Bahasa Isyarat Indonesia (BISINDO). Algoritma YOLO dipilih untuk digunakan dalam penelitian ini karena kecepatannya yang mampu melebihi

beberapa algoritma pendeteksian serupa. Berdasarkan hasil pengujian terhadap epoch 100, 200, dan 400, hasil akurasi terbaik diraih oleh model yang dilatih dengan 200 epoch dengan akurasi sebesar 67% pada pengujian pertama dan 51,95% pada pengujian kedua.

Dengan hasil penelitian ini, saran untuk penelitian selanjutnya adalah:

1. Penelitian selanjutnya dapat mencoba untuk menambahkan jumlah data dan variasi untuk setiap katanya.
2. Penelitian selanjutnya dapat mencoba memperbesar batch\_size agar normalisasi batch YOLOv5 lebih baik.

Mencoba untuk mengubah model selain YOLOv5 atau tambahkan/gabungkan metode lainnya seperti CNN untuk mengamati setiap detil bentuk tangan untuk meminimalisir kesalahan klasifikasi.

#### DAFTAR PUSTAKA

- [1] R. I. Borman and B. Priyopradono, "Implementasi Penerjemah Bahasa Isyarat Pada Bahasa Isyarat Indonesia (BISINDO) Dengan Metode Principal Component Analysis (PCA)," *Jurnal Informatika: Jurnal Pengembangan IT (JPIT)*, vol. 3, no. 1, pp. 103-108, Januari 2018.
- [2] A. S. Nugraheni, A. P. Husain and H. Unayah, "OPTIMALISASI PENGGUNAAN BAHASA ISYARAT DENGAN SIBI DAN BISINDO PADA MAHASISWA DIFABEL TUNARUNGU DI PRODI PGMI UIN SUNAN KALIJAGA," *HOLISTIKA : Jurnal Ilmiah PGSD*, vol. V, no. 1, pp. 28-33, Mei 2021.
- [3] M. Y. Andrian, D. Purwanto and R. Mardiyanto, "Penerjemahan Bahasa Isyarat Indonesia Menggunakan Kamera pada Telpn Genggam Android," *Jurnal Teknik ITS*, vol. 6, no. 1, pp. 174-179, Maret 2017.
- [4] M. Kasper-Eulaers, N. Hahn, S. Berger, T. Sebulonsen, Ø. Myrland and P. E. Kummervold, "Short Communication: Detecting Heavy Goods Vehicles in Rest Areas in Winter Conditions Using YOLOv5," *Algorithms*, vol. 14, no. 4, March 2021.
- [5] F. Andrianson, Lina and A. Chris, "Pendeteksian Sel Darah Putih Dari Citra Preparat Dengan You Only Look Once (YOLO)," *Jurnal Ilmu Komputer & Sistem Informasi (JIKSI)*, vol. 9, no. 1, Januari 2021.
- [6] I. H. A. Amin and F. H. Arby, "Implementation of YOLO-V5 for a Real Time Social Distancing Detection," *Journal of Applied Informatics and Computing*, vol. 6, no. 1, April 2022.
- [7] Z. Chen, R. Wu, Y. Lin, C. Li, S. Chen, Z. Yuan, S. C. and X. Zou, "Plant Disease Recognition Model Based on Improved YOLOv5," *Agronomy*, vol. 12, no. 2, p. 365, 2022.
- [8] R. V. Iyer, P. S. Ringe and K. P. Bhensdadiya, "Comparison of YOLOv3, YOLOv5s and MobileNet-SSD V2 for Real-Time," *International Research Journal of Engineering and Technology (IRJET)*, vol. 8, no. 7, pp. 1156-1160, July 2021.