

BAB II

TINJAUAN PUSTAKA

2.1. Landasan Teori

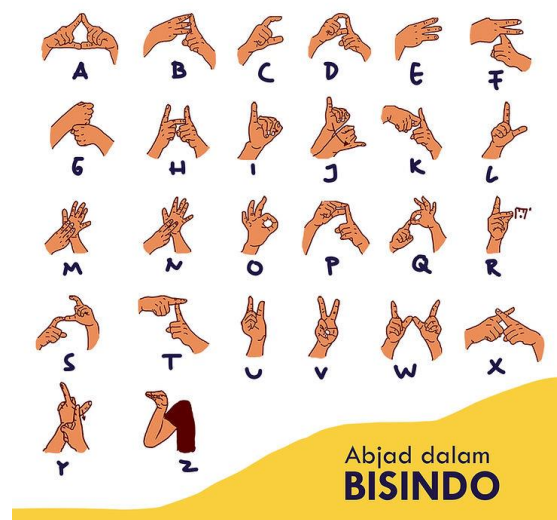
2.1.1. Bahasa Isyarat

Dalam kehidupan sehari-hari, berkomunikasi adalah suatu hal yang esensial. Namun, bagi komunitas tuna rungu di Indonesia, komunikasi seringkali menghadapi tantangan karena keterbatasan dalam pemahaman bahasa lisan. Oleh karena itu, muncullah bahasa isyarat sebagai sarana komunikasi yang efektif bagi mereka. Di Indonesia, bahasa isyarat ini dikenal sebagai Bahasa Isyarat Indonesia atau lebih dikenal dengan singkatan BISINDO.

Di Indonesia, dua bahasa isyarat yang digunakan secara umum adalah Sistem Isyarat Bahasa Indonesia (SIBI) dan Bahasa Isyarat Indonesia (BISINDO). Meskipun SIBI diakui sebagai bahasa isyarat resmi di sekolah, fakta menunjukkan bahwa SIBI jarang digunakan oleh penyandang tunarungu dalam kehidupan sehari-hari mereka. Sebaliknya, BISINDO menjadi lebih *dominan* dan relevan dalam interaksi sehari-hari komunitas tuna rungu di Indonesia (Handhika dkk., 2018).

Dengan adanya model pengenalan isyarat yang berbasis MediaPipe dan CNN, komunitas tuna rungu di Indonesia diharapkan dapat memiliki alat bantu komunikasi yang lebih efektif. Meskipun akurasi real-time model ini masih perlu ditingkatkan, keberhasilannya dalam mengenali isyarat BISINDO dengan tingkat akurasi yang cukup baik menunjukkan potensi besar untuk membantu penyandang

tunarungu dalam berkomunikasi lebih lancar dengan masyarakat umum. Teknologi ini tidak hanya memfasilitasi komunikasi, tetapi juga dapat meningkatkan kesadaran dan pemahaman masyarakat tentang bahasa isyarat, sehingga mendukung inklusi dan partisipasi aktif komunitas tuna rungu dalam berbagai aspek kehidupan.



Gambar 2. 1 Abjad dalam BISINDO

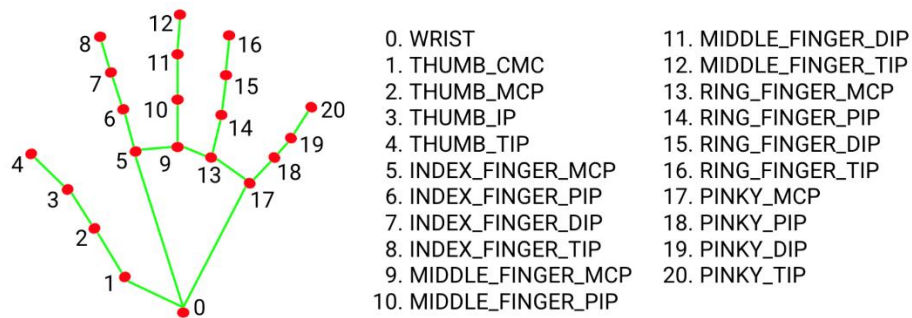
Sumber : (Klobility, 2019)

Penulis memilih Bahasa Isyarat Indonesia (BISINDO) sebagai fokus utama dalam pembahasan ini, dan pilihan ini didasarkan pada beberapa alasan yang kuat. Salah satu alasan utama adalah kekayaan BISINDO dalam varian gerakan tangan, termasuk kombinasi dua tangan dan satu tangan.

2.1.2. Mediapipe

MediaPipe adalah platform pengembangan *open-source* yang dikembangkan oleh Google untuk analisis dan pemrosesan media berbasis AI. Platform ini

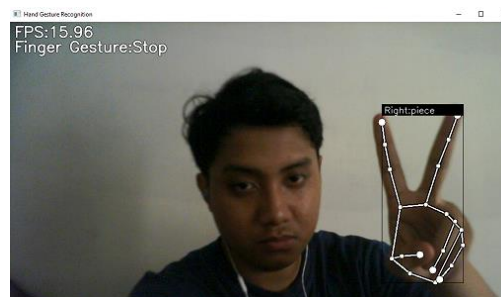
menyediakan berbagai alat dan model untuk mendeteksi, melacak, dan menginterpretasi berbagai elemen dalam data visual.



Gambar 2. 2 *Hand Landmark* Mediapipe

Sumber : (*Hand Landmarks Detection Guide | MediaPipe | Google for Developers*, t.t.)

Salah satu contohnya deteksi tangan, MediaPipe *Hand Landmark* adalah tugas yang memungkinkan untuk mendeteksi *landmark* (titik kunci) pada tangan dalam gambar 2.3. Terdapat 21 titik *landmark* pada setiap tangan yang dapat dikenali oleh model ini. Setiap titik *landmark* diwakili oleh koordinat X dan Y dalam sistem koordinat tangan. Model ini dapat memberikan informasi tentang posisi relatif dan orientasi dari setiap *landmark*, dan dapat digunakan untuk berbagai aplikasi, termasuk pengenalan gestur tangan dan interaksi manusia dan mesin lainnya.



Gambar 2. 3 *Finger Gesture* Mediapipe

Dalam *Computer vision*, teknologi ini dapat digunakan untuk berbagai tujuan, termasuk mengenali objek dalam gambar, mendeteksi wajah, memahami konteks dalam adegan visual, dan bahkan membuat model tiga dimensi dari objek atau lingkungan. Aplikasi *Computer vision* melibatkan pengenalan pola, pengolahan citra, dan integrasi teknologi dalam berbagai industri, seperti otomotif, keamanan, kesehatan, dan manufaktur.

Dengan kemajuan dalam algoritma dan kekuatan komputasi, *Computer vision* terus berkembang dan memiliki dampak besar dalam mendorong inovasi teknologi di berbagai sektor.

2.1.4. OpenCV



Gambar 2. 5 *OpenCV*

OpenCV (Open Source Computer Vision Library) adalah pustaka perangkat lunak open-source yang populer untuk visi komputer dan pembelajaran mesin, dirancang untuk efisiensi komputasi dan aplikasi *real-time*. OpenCV menyediakan alat untuk membaca, menulis, dan memanipulasi gambar dan video, termasuk deteksi tepi, transformasi geometris, dan pengenalan objek seperti wajah dan plat nomor. Mendukung berbagai algoritma *machine learning*, OpenCV dapat

diintegrasikan dengan pustaka *deep learning* seperti TensorFlow dan PyTorch. OpenCV kompatibel dengan berbagai bahasa pemrograman (C++, Python, Java, MATLAB) dan platform (Windows, Linux, macOS, Android, iOS), memudahkan pengguna dari berbagai latar belakang untuk memanfaatkannya dalam proyek mereka.

2.1.5. *Deep learning*

Deep learning adalah pendekatan dalam machine learning yang menggunakan jaringan saraf kompleks dengan banyak lapisan untuk memahami dan memproses data. Konsep utamanya adalah memberikan kemampuan komputer untuk secara otomatis memahami pola rumit dari data tanpa pemrograman manual.

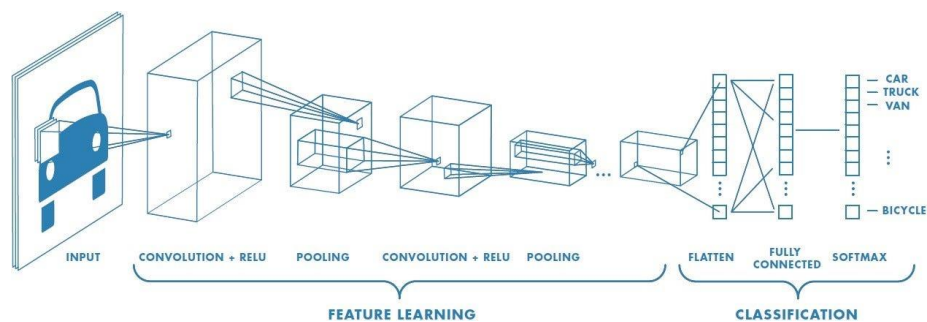
Dalam *Deep learning*, jaringan saraf mendalam menghasilkan hierarki fitur dari data, memungkinkan komputer "belajar" melalui informasi yang diberikan. Keunggulan *Deep learning* melibatkan ekstraksi fitur otomatis, mengurangi kebutuhan akan ekstraksi manual.

Penggunaan *Deep learning* membutuhkan pelatihan dengan dataset besar, seringkali menggunakan GPU untuk percepatan. *Deep learning* dapat diterapkan pada berbagai tugas, termasuk pengenalan gambar dan suara.

Berbagai arsitektur *Deep learning*, seperti *Convolutional Neural Networks* (CNN) dan *Recurrent Neural Networks* (RNN), telah mengubah teknologi kecerdasan buatan dan mendominasi aplikasi di berbagai bidang.

2.1.6. Convolutional Neural Networks (CNN)

CNN adalah jenis arsitektur dalam jaringan saraf yang sangat efektif dalam mengolah data visual, seperti gambar dan video. Dengan lapisan konvolusi, CNN dapat mengekstrak fitur dan pola visual dari data, Convolutional-Neural-Network (CNN) adalah model DL (*Deep learning*) yang sangat terkenal yang digunakan dalam aplikasi berbasis gambar membuatnya sangat cocok untuk pengenalan isyarat tangan yang melibatkan analisis gerakan tangan dalam citra atau video (Ur Rehman dkk., 2022).



Gambar 2. 6 Tampilan *layer* CNN

Sumber : (Saha, 2022)

Berikut adalah penjelasan lebih rinci untuk setiap *layer* dalam *Convolutional Neural Networks* (CNN):

1. *Input Layer*:

Deskripsi: Menerima *input* dalam bentuk gambar atau data spasial lainnya.

Fungsi: Memasukkan citra atau data *input* ke dalam jaringan.

2. *Convolutional Layer*:

Deskripsi: Terdiri dari filter atau kernel yang bergerak melintasi *input* untuk mengekstrak fitur-fitur lokal.

Fungsi: Mendeteksi pola seperti tepi, tekstur, atau bentuk pada citra.

3. *Activation Layer (ReLU - Rectified Linear Unit)*:

Deskripsi: Menerapkan fungsi aktivasi ReLU untuk memperkenalkan non-linearitas ke dalam model.

Fungsi: Menonaktifkan nilai-nilai negatif dan mempertahankan nilai positif, memungkinkan model mempelajari hubungan yang lebih kompleks.

4. *Pooling Layer*:

Deskripsi: Melakukan subsampling atau downsampling untuk mengurangi dimensi data dengan mempertahankan fitur paling penting.

Fungsi: Mengurangi ukuran representasi spasial untuk efisiensi komputasi.

5. *Flattening Layer*:

Deskripsi: Mengubah data dari matriks 2D menjadi vektor 1D.

Fungsi: Menyiapkan data untuk lapisan-lapisan sepenuhnya terhubung.

6. *Fully Connected (Dense) Layer*:

Deskripsi: Lapisan yang terdiri dari neuron-neuron yang terhubung sepenuhnya, menerima *input* dari vektor yang dihasilkan sebelumnya.

Fungsi: Memproses informasi secara global dan menghasilkan representasi fitur tingkat tinggi.

7. *Output Layer*:

Deskripsi: Menyediakan *output* dari model, biasanya dalam bentuk probabilitas untuk setiap kelas pada tugas klasifikasi.

Fungsi: Menghasilkan prediksi atau *output* model.

Setiap lapisan ini memainkan peran penting dalam mengekstrak dan memahami fitur-fitur dari data *input*, yang kemudian digunakan untuk membuat prediksi. Dalam beberapa arsitektur CNN yang lebih kompleks, ada juga lapisan-lapisan tambahan seperti *normalization*, *dropout*, atau lapisan konvolusi yang dilapis (seperti dalam arsitektur ResNet). Kombinasi lapisan-lapisan ini membantu CNN menghasilkan representasi fitur yang lebih abstrak dan kompleks..

2.1.7. *Confusion matrix*

Sebuah *Confusion matrix* adalah alat pengukuran kinerja untuk algoritma klasifikasi pembelajaran mesin. Ini adalah tabel yang sering digunakan untuk menggambarkan kinerja model klasifikasi pada satu set data uji di mana nilai-nilai sebenarnya diketahui.

Berikut adalah pembagian *Confusion matrix*:

- True Positives (TP): Ini adalah kasus ketika kelas aktual dari titik data adalah positif, dan model memprediksinya sebagai positif.
- True Negatives (TN): Ini adalah kasus ketika kelas aktual dari titik data adalah negatif, dan model memprediksinya sebagai negatif.

- False Positives (FP): Ini adalah kasus ketika kelas aktual dari titik data adalah negatif, tetapi model memprediksinya sebagai positif (kesalahan Tipe I).
- False Negatives (FN): Ini adalah kasus ketika kelas aktual dari titik data adalah positif, tetapi model memprediksinya sebagai negatif (kesalahan Tipe II).

Sebuah *Confusion matrix* terlihat seperti ini:

| | | | |
|------------------|--------------|---------------|--------------|
| | | Actual Values | |
| | | Positive (1) | Negative (0) |
| Predicted Values | Positive (1) | TP | FP |
| | Negative (0) | FN | TN |

Gambar 2. 7 *Confusion matrix*

Dari matriks ini, berbagai metrik kinerja dapat dihitung, seperti akurasi, presisi, recall (sensitivitas), spesifisitas, skor F1, dll. Metrik-metrik ini memberikan wawasan tentang seberapa baik model melakukan klasifikasi titik data ke dalam berbagai kelas.

2.2. Kajian Penelitian

| No. | Judul | Peneliti /Tahun | Jurnal | Hasil/Informasi yang diperoleh |
|-----|--|-----------------------|----------------------|--|
| 1. | SIBI (Sistem Bahasa Isyarat Indonesia) | (Budiman dkk., 2023b) | Jurnal Teknologi dan | Hasil Penelitian: <ul style="list-style-type: none"> • Model sequential dengan <i>optimizer</i> ADAM menghasilkan |

| | | | | |
|----|--|--------------------------|--|--|
| | berbasis <i>Machine Learning</i> dan <i>Computer vision</i> untuk Membantu Komunikasi Tuna Rungu dan Tuna Wicara | | Manajemen Informatika, Vol. 9 No.2 | <p>loss akhir sebesar 0.0188 dan akurasi sebesar 99.72%.</p> <ul style="list-style-type: none"> • Akurasi pengujian pada jarak 40 cm mencapai 87.50%. <p>Kelemahan:</p> <ul style="list-style-type: none"> • Alfabet J dan Z tidak dapat dikenali karena membutuhkan pola gerakan tangan (gesture), sedangkan pada penelitian ini hanya dapat membaca pola tangan saja. • Pada jarak 60 cm dari webcam, akurasi menurun menjadi 79.17%. |
| 2. | Implementasi Deteksi Bahasa Isyarat Tangan Menggunakan OpenCV dan MediaPipe | (Nur'az izan dkk., 2024) | Prosiding Seminar Nasional Teknologi Dan Sains Tahun | <p>Hasil Penelitian:</p> <ul style="list-style-type: none"> • Menggunakan data sebanyak 4 jenis: "Aku Sayang Kamu," "Halo," "Iya," dan "Tidak." • Model mencapai akurasi tinggi dengan f1 score 0.9875, recall 0.9875, dan precision 0.9875. |

| | | | | |
|----|--|---------------------|--|---|
| | | | 2024, Vol. 3. | <p>Kelemahan:</p> <ul style="list-style-type: none"> • Jumlah jenis data yang terbatas (4 kategori) mungkin tidak mencakup keanekaragaman isyarat bahasa isyarat yang lebih luas. • Bergantung pada kehandalan MediaPipe untuk ekstraksi <i>landmark</i>, dan performa dapat bervariasi tergantung pada kondisi pencahayaan dan sudut kamera. |
| 3. | Deteksi Tangan Otomatis Pada Video Percakapan Bahasa Isyarat Indonesia Menggunakan Metode YOLO Dan CNN | (Arifah dkk., 2022) | <i>Journal of Applied Informatics and Computing</i> (JAIC) Vol.6, No.2 | <p>Hasil Penelitian:</p> <ul style="list-style-type: none"> • Metode yang digunakan untuk mendeteksi dan membedakan objek gambar tangan adalah hasil dari penggabungan YOLOv5 dan <i>Convolutional Neural Networks</i> (CNN). |

| | | | | |
|----|--|------------------------|----------------------------------|---|
| | | | | <ul style="list-style-type: none"> • Kelima objek yang dibedakan: tangan berbentuk angka satu, dua, tiga, empat, dan lima. <p>Kelemahan:</p> <ul style="list-style-type: none"> • Beberapa data tangan tidak terdeteksi dengan benar disebabkan oleh pengaruh pencahayaan. • Perubahan background dapat mempengaruhi hasil pendeteksian dan memunculkan kesalahan. |
| 4. | Data Augmentasi Untuk Mengatasi Keterbatasan Data Pada Model Penerjemah Bahasa Isyarat Indonesia (BISINDO) | (Fadilla h dkk., 2021) | Jurnal Informatika, Vol. 8 No. 2 | <p>Hasil Penelitian:</p> <ul style="list-style-type: none"> • Menggunakan algoritma <i>Convolutional Neural Networks</i> (CNN). • Akurasi model mencapai 94.38%. • Implementasi data augmentasi berhasil meningkatkan akurasi, sedangkan tanpa data augmentasi hanya mencapai 30%. |

| | | | | |
|----|---|---------------------------|--|---|
| | | | | <p>Kelemahan:</p> <ul style="list-style-type: none"> • Penelitian hanya fokus pada menerjemahkan alfabet Bisindo. Penerjemahan kata dan kalimat belum diakomodasi sepenuhnya. • Tanpa data augmentasi, model hanya mencapai 30% akurasi, menunjukkan keterbatasan dalam jumlah dataset. |
| 5. | <p>Pendeteksian Bahasa Isyarat Indonesia Secara Real-Time Menggunakan Long Short-Term Memory (Lstm)</p> | <p>(Putri dkk., 2022)</p> | <p>Jurnal Teknologi Terapan and Sains 4.0. Vol 3</p> | <p>Hasil Penelitian:</p> <ul style="list-style-type: none"> • Dengan menggunakan model BiLSTM, ditemukan tingkat akurasi sebesar 92% setelah melalui 1000 iterasi <i>epoch</i>. Konfigurasi model ini melibatkan 46 lapisan tersembunyi dan batch size sebesar 64, fokus pada deteksi 10 isyarat. • Hasil dari model 2 <i>layer</i> LSTM menunjukkan tingkat akurasi sebesar 65% setelah menjalani 500 iterasi <i>epoch</i>. Struktur model |

| | | | | |
|----|---|--------------------------|--|---|
| | | | | <p>ini mencakup 64 lapisan tersembunyi dan batch size sebesar 64, difokuskan pada pengujian <i>real-time</i> untuk mendeteksi 30 isyarat.</p> <p>Kelemahan:</p> <ul style="list-style-type: none"> • Klasifikasi gerakan pada deteksi <i>real-time</i> menghadapi tantangan dalam mengidentifikasi gerakan yang serupa dengan isyarat lain. • Tidak meratanya pencahayaan dapat mempengaruhi deteksi <i>mediapipe holistic landmarks</i> pada tangan, wajah, dan badan objek. |
| 6. | Implementasi Deteksi Bahasa Isyarat Tangan Menggunakan OpenCV dan MediaPipe | (Ardian syah dkk., 2024) | Prosiding Seminar Nasional Teknologi Dan Sains Tahun | <p>Hasil Penelitian:</p> <ul style="list-style-type: none"> • Metode <i>Support Vector Machine</i> (SVM) digunakan untuk klasifikasi. • Model mencapai akurasi tinggi dengan <i>f1 score</i> 0.9875, <i>recall</i> 0.9875, dan <i>precision</i> 0.9875. |

| | | | | |
|----|---|-----------------------|--------------------|---|
| | | | 2024, Vol. 3. | <p>Kelemahan:</p> <ul style="list-style-type: none"> • Proses pengambilan gambar secara realtime memerlukan waktu 15 detik per frame, yang mungkin terlalu lama untuk beberapa aplikasi real-time. |
| 7. | <p>Pengenalan Alfabet Bahasa Isyarat Tangan Secara <i>RealTime</i> dengan Menggunakan Metode <i>Convolutional Neural Networks</i> dan <i>Recurrent Neural Network</i></p> | (Yolan da dkk., 2020) | Jurnal Infra Vol 8 | <p>Hasil Penelitian:</p> <ul style="list-style-type: none"> • Metode <i>Convolutional Neural Networks</i> (CNN) dan <i>Recurrent Neural Network</i> (RNN) untuk pengenalan alfabet bahasa isyarat tangan mendapatkan akurasi keseluruhan sebesar 60.58% untuk seluruh huruf. <p>Kelemahan:</p> <ul style="list-style-type: none"> • Keterbatasan akurasi pada keseluruhan huruf menunjukkan bahwa terdapat ruang untuk peningkatan, terutama pada kasus huruf tertentu. |

| | | | | |
|----|---|------------------------|--|--|
| 8. | <i>Dynamic Hand Gesture Recognition Using 3D-CNN and LSTM Networks</i> | (Ur Rehman dkk., 2022) | <i>Computers, Materials & Continua</i> Vol 70 | <p>Hasil Penelitian:</p> <ul style="list-style-type: none"> • Kombinasi 3DCNN dengan LSTM memberikan hasil yang lebih baik dibandingkan dengan MobileNetv2 + LSTM. • Menggunakan normalisasi batch dan regularisasi L2 untuk mencegah <i>overfitting</i>. • Model yang diusulkan mencapai akurasi pelatihan 99%, validasi 97.5%, dan akurasi prediksi 97% selama pengujian real-time. |
| 9. | <i>A Study on Real-time Hand Gesture Recognition Technology by Machine Learning-based MediaPipe</i> | (Hand dkk., 2022) | <i>Journal of System and Management Sciences</i> Vol 12 | <p>Hasil Penelitian:</p> <ul style="list-style-type: none"> • Model kurang mempertimbangkan karakteristik anak-anak yang bergerak bebas dalam ruang luas. • Diperlukan pengembangan algoritma pengenalan tangan yang mempertimbangkan ukuran anak-anak dan jarak kamera. <p>Kelemahan:</p> |

| | | | | |
|-----|---|-----------------|--|--|
| | | | | <ul style="list-style-type: none"> • Kurangnya pertimbangan ukuran tangan anak-anak dan kurangnya adaptasi model untuk pergerakan anak-anak. • Model sulit diterapkan secara langsung dalam pendidikan anak usia tiga hingga lima tahun. |
| 10. | <i>Recognition Of Hand Gestures Using Mediapipe Hands</i> | (KM & NR, 2022) | <i>International Research Journal of Modernization in Engineering Technology and Science</i> Vol 04 | <p>Hasil Penelitian:</p> <ul style="list-style-type: none"> • Metodologi menunjukkan bahwa MediaPipe dapat digunakan secara efisien sebagai alat untuk mendeteksi gestur tangan kompleks dengan tepat. • Metodologi yang diusulkan mencapai akurasi rata-rata 99% pada sebagian besar dataset bahasa isyarat menggunakan teknologi Media Pipe dan machine learning. • Pelatihan dan pengujian dengan berbagai dataset bahasa isyarat menunjukkan bahwa kerangka |

| | | | | |
|--|--|--|--|--|
| | | | | <p>kerja ini dapat efektif diadaptasi untuk setiap dataset bahasa isyarat regional, mencapai akurasi maksimum.</p> |
|--|--|--|--|--|

Tabel 2. 1 Teori penelitian

1. SIBI (Sistem Bahasa Isyarat Indonesia) berbasis Machine Learning dan *Computer vision* untuk Membantu Komunikasi Tuna Rungu dan Tuna Wicara

Dalam penelitian yang dilakukan oleh Budiman, Lestanti, dan Yuana terkait pengembangan Sistem Bahasa Isyarat Indonesia (SIBI), dapat disimpulkan bahwa penerapan teknologi *machine learning* (ML) dan *Computer vision* (CV) memberikan kontribusi positif dalam meningkatkan komunikasi di kalangan komunitas tuli. Penggunaan *Convolutional Neural Networks* (CNN) untuk mengenali pola gerakan tangan dan MediaPipe untuk deteksi *landmark* tangan memungkinkan SIBI melakukan terjemahan bahasa isyarat secara real-time. Meskipun terdapat keterbatasan pada dataset, hasil penelitian menunjukkan akurasi yang baik, terutama pada jarak 40 cm dengan akurasi 87.50% dan 60 cm dengan akurasi 79.17%.

2. Implementasi Deteksi Bahasa Isyarat Tangan Menggunakan OpenCV dan MediaPipe

Melalui implementasi deteksi Bahasa Isyarat Tangan menggunakan OpenCV dan MediaPipe, penelitian ini berhasil menciptakan sebuah sistem pembelajaran Bahasa Isyarat yang efektif. Dengan memanfaatkan teknologi

OpenCV, penelitian ini memungkinkan pembelajaran Bahasa Isyarat secara langsung, membantu individu yang belum mengenal bahasa isyarat untuk belajar dengan lebih interaktif. Penggunaan MediaPipe sebagai library pelatihan pengenalan pola tangan dengan metode *Support Vector Machine* (SVM) memberikan hasil yang mengesankan.

Dalam pengujian secara *real-time*, sistem ini mampu mencapai f1 score sebesar 0.9875, recall sebesar 0.9875, dan precision sebesar 0.9875. Hasil ini menunjukkan tingkat akurasi dan ketepatan yang tinggi dalam mengenali pola tangan untuk klasifikasi Bahasa Isyarat. Dengan menggunakan 400 dataset citra.

3. Deteksi Tangan Otomatis Pada Video Percakapan Bahasa Isyarat Indonesia

Menggunakan Metode YOLO Dan CNN

Penelitian ini bertujuan untuk mengembangkan sistem otomatis deteksi tangan dalam video percakapan menggunakan Bahasa Isyarat Indonesia (BISINDO), dengan menerapkan metode *You Only Look Once* (YOLO) dan *Convolutional Neural Networks* (CNN). YOLO digunakan untuk mengenali objek tangan, sedangkan CNN digunakan untuk mengklasifikasi berbagai jenis gerakan tangan. Penggunaan dataset yang terdiri dari gambar dan video BISINDO memberikan keberagaman gerakan yang dapat diidentifikasi oleh sistem. Dalam proses pengumpulan data dan pra-pemrosesan, penelitian ini berhasil mencapai tingkat akurasi sekitar 89%.

4. Data Augmentasi Untuk Mengatasi Keterbatasan Data Pada Model Penerjemah Bahasa Isyarat Indonesia (BISINDO)

Penelitian menggunakan pendekatan machine learning dengan arsitektur *Convolutional Neural Networks* (CNN) untuk mengembangkan model penerjemah Bisindo. Keterbatasan dataset Bisindo diatasi dengan metode data augmentation, memperluas variasi gestur bahasa isyarat yang dapat dikenali oleh model. Upaya ini berhasil meningkatkan performa model, mencapai tingkat akurasi sebesar 94.38%.

Kesimpulan dari penelitian ini menunjukkan bahwa penggunaan teknik data augmentation efektif dalam mengatasi keterbatasan data pada model penerjemah Bisindo. Dengan akurasi yang tinggi, model ini dapat menjadi alat yang berharga untuk menerjemahkan gestur bahasa isyarat menjadi teks, membantu dalam peningkatan aksesibilitas bagi komunitas Tuli.

5. Pendeteksian Bahasa Isyarat Indonesia Secara Real-Time Menggunakan Long Short-Term Memory (Lstm)

Metode klasifikasi BISINDO pada penelitian ini menggunakan *Long Short-Term Memory* (LSTM), yang merupakan jenis arsitektur jaringan saraf rekuren (RNN). MediaPipe Holistic digunakan untuk mendeteksi kerangka pada tangan, wajah, dan badan. Penelitian fokus pada 30 kosakata isyarat BISINDO yang sering digunakan oleh teman Tuli.

Hasil evaluasi deteksi *real-time* menunjukkan bahwa model dengan 10 kelas mendapatkan akurasi sebesar 92%, menggunakan bidirectional *layer* LSTM dengan *epoch* 1000, hidden *layer* 64, dan *batch size* 32. Sementara itu,

model dengan 30 kelas mendapatkan akurasi sebesar 65%, menggunakan 2 *layer* LSTM dengan *epoch* 500, *hidden layer* 64, dan *batch size* 64.

Kesimpulan dari penelitian ini adalah bahwa sistem yang dikembangkan mampu secara efektif mendeteksi dan mengklasifikasi gestur bahasa isyarat BISINDO secara real-time, dengan tingkat akurasi yang cukup baik.

6. Implementasi Deteksi Bahasa Isyarat Tangan Menggunakan OpenCV dan MediaPipe

Dalam penelitian ini, implementasi deteksi bahasa isyarat tangan menggunakan OpenCV dan MediaPipe telah berhasil memberikan hasil yang memuaskan. Sistem yang dikembangkan mampu mengenali gestur tangan secara *real-time* dengan tingkat akurasi yang tinggi, terbukti dari *f1 score*, *recall*, dan *precision* mencapai 0.9875. Penggunaan OpenCV dalam kombinasi dengan MediaPipe membuktikan keefektifan teknologi dalam pembelajaran bahasa isyarat, memberikan solusi praktis untuk mereka yang ingin belajar dan berlatih bahasa isyarat. Jumlah data yang mencakup 400 dataset citra juga memberikan keandalan pada hasil pengujian secara real-time. Dengan demikian, sistem ini dapat dianggap sebagai alat pembelajaran yang efisien dan dapat berpotensi menjadi kontribusi positif dalam memfasilitasi pemahaman dan praktik bahasa isyarat bagi pengguna yang membutuhkannya.

7. Pengenalan Alfabet Bahasa Isyarat Tangan Secara RealTime dengan Menggunakan Metode *Convolutional Neural Networks* dan *Recurrent Neural Network*

Penelitian ini bertujuan untuk mengembangkan sistem pengenalan alfabet bahasa isyarat tangan secara waktu nyata dengan menggunakan kombinasi metode *Convolutional Neural Networks* (CNN) dan *Recurrent Neural Network* (RNN). CNN diterapkan untuk mengekstraksi fitur spasial dari video bahasa isyarat tangan, sementara RNN digunakan untuk mengkorelasikan frame secara temporal dan menghasilkan teks alfabet sebagai output. Hasil pengujian menunjukkan tingkat akurasi rata-rata sebesar 60.58% dalam pengenalan huruf-huruf. Meskipun demikian, pengujian real-time menghadapi kendala teknologi yang tidak mampu mendukung arsitektur yang telah dibangun.

8. Dynamic Hand Gesture Recognition Using 3D-CNN and LSTM Networks

Penyakit Daun Tomat

Penelitian ini bertujuan untuk mengatasi tantangan dalam pengenalan gestur tangan dinamis secara real-time dengan mengusulkan arsitektur *Deep learning* yang menggabungkan *Convolutional Neural Networks* 3D (3D-CNN) dan *Long Short-Term Memory* (LSTM). Arsitektur ini dirancang untuk mengekstraksi informasi spasial-temporal dari urutan video dengan menghindari perhitungan yang intensif. 3D-CNN digunakan untuk mengekstrak fitur spektral dan spasial, yang kemudian diberikan kepada jaringan LSTM untuk klasifikasi. Model ini memiliki arsitektur ringan dengan hanya 3,7 juta parameter pelatihan.

Evaluasi dilakukan menggunakan dataset 20BN-jester dengan 15 kelas, dan model ini dilatih pada 2000 klip video per kelas. Model mencapai akurasi sebesar 99% dan 97% pada data pelatihan dan pengujian, secara berturut-turut.

Temuan ini menunjukkan bahwa penggabungan 3D-CNN dengan LSTM memberikan hasil yang lebih unggul dibandingkan dengan MobileNetv2 + LSTM.

9. A Study on Real-time Hand Gesture Recognition Technology by Machine Learning-based MediaPipe

Penelitian ini bertujuan untuk mengenali gerakan tangan secara real-time melalui kamera yang terhubung ke komputer, dengan tujuan digunakan dalam kegiatan edukatif anak-anak. Kementerian Pendidikan Republik Korea telah mengembangkan kurikulum untuk anak-anak usia tiga hingga lima tahun, dengan fokus pada lima area, termasuk olah tubuh, kesehatan, komunikasi, hubungan sosial, pengalaman seni, dan eksplorasi alam. Dalam konteks ini, penelitian mencoba membantu anak-anak belajar menjelajahi alam dalam ruangan dengan menggunakan MediaPipe untuk mengenali gerakan tangan anak-anak dan mengoperasikan tetikus virtual melalui gerakan tangan tersebut. Meskipun situasi sulit karena pandemi COVID-19, pendekatan ini dapat membantu mendukung pendidikan anak-anak dalam menghadapi keterbatasan aktivitas eksternal.

10. Recognition Of Hand Gestures Using Mediapipe Hands

Penelitian ini membahas pengenalan gestur tangan menggunakan MediaPipe Hands untuk interaksi tanpa sentuhan, dengan potensi aplikasi mulai dari sistem medis hingga permainan. Pengenalan bahasa isyarat juga menjadi fokus, memberikan manfaat bagi orang dengan kesulitan pendengaran. Solusi

pelacakan tangan waktu nyata ini dapat memprediksi kerangka tangan manusia menggunakan kamera RGB tunggal. Menggunakan perpustakaan Mediapipe yang disediakan oleh Google, penelitian ini memanfaatkan model yang terlatih dengan baik untuk mencapai kinerja tinggi. Dalam penelitian ini, Mediapipe Hands menggunakan dua model: pendetektor telapak tangan dan model *landmark* tangan. Hasilnya menunjukkan bahwa pengguna dapat berinteraksi dengan komputer yang dilengkapi kamera RGB menggunakan gestur tangan.

Dalam penelitian yang dilakukan oleh Budiman, Lestanti, dan Yuana tentang pengembangan Sistem Bahasa Isyarat Indonesia (SIBI), ditemukan bahwa penggunaan teknologi machine learning (ML) dan *computer vision* (CV) berdampak positif dalam meningkatkan komunikasi di kalangan komunitas tuli. Melalui penerapan *Convolutional Neural Networks* (CNN) untuk mengenali pola gerakan tangan dan MediaPipe untuk mendeteksi *landmark* tangan, SIBI dapat menerjemahkan bahasa isyarat secara real-time. Meskipun terdapat keterbatasan pada dataset, hasil penelitian menunjukkan akurasi yang baik, terutama pada jarak 40 cm dengan akurasi 87.50% dan 60 cm dengan akurasi 79.17%.

Penelitian ini sangat relevan dengan penelitian saya yang juga menggunakan CNN untuk pengenalan bahasa isyarat, namun saya berencana mengatasi keterbatasan dataset dengan menggunakan data vektor yang memiliki ukuran lebih kecil.