

PENERAPAN CONVOLUTIONAL NEURAL NETWORK UNTUK PENGENALAN BAHASA ISYARAT INDONESIA: STUDI KASUS DAN TINJAUAN FILSAFAT SAINS

**Hernalom Sitorus¹, Ucu Nugraha^{2*}, Sri Titi Handayani³, Agus Nursikuwagus⁴, Usep Mohamad Ishaq⁵,
Andrias Darmayadi⁶**

Program Doktor Sistem Informasi, Fakultas Pascasarjana^{1,2,3,4,5,6}
Universitas Komputer Indonesia^{1,2,3,4,5,6}

Jl. Dipatiukur No. 102-118, Kota Bandung, Jawa Barat, Indonesia^{1,2,3,4,5,6}

hernalom.75725022@mahasiswa.unikom.ac.id¹, ucu.75725016@mahasiswa.unikom.ac.id^{2*},
sri.75725014@mahasiswa.unikom.ac.id³, agusnursikuwagus@email.unikom.ac.id⁴,
usep.mohamad.ishaq@email.unikom.ac.id⁵, andrias.darmayadi@email.unikom.ac.id⁶

Abstrak

Rendahnya literasi Bahasa Isyarat Indonesia (BISINDO) di masyarakat umum masih menjadi hambatan komunikasi dengan komunitas Tuli, sementara riset pengenalan bahasa isyarat berbasis AI umumnya berfokus pada capaian teknis semata. Penelitian ini bertujuan mengembangkan sistem pengenalan alfabet BISINDO berbasis Convolutional Neural Network (CNN) sekaligus mengevaluasinya melalui perspektif filsafat sains. Metode yang digunakan meliputi pengumpulan dataset citra tangan alfabet BISINDO, augmentasi data, serta pelatihan CNN berbasis transfer learning dengan arsitektur MobileNetV2 dan skema pelatihan bertahap, kemudian dideploy ke Android menggunakan TensorFlow Lite. Hasil pengujian menunjukkan sistem mampu mencapai akurasi sekitar 93% pada data uji terkontrol dengan performa inferensi real-time yang stabil. Kontribusi ilmiah penelitian ini tidak hanya pada pengembangan sistem AI terapan, tetapi juga pada penyediaan kerangka reflektif ontologis, epistemologis, dan aksiologis untuk menilai validitas dan implikasi sosial teknologi pengenalan BISINDO.

Kata kunci:

Convolutional Neural Network, Bahasa Isyarat Indonesia, Mobile Computing, Filsafat Sains, Teknologi Inklusif

Abstract

The low literacy of Indonesian Sign Language (BISINDO) in the general public remains a barrier to communication with the Deaf community, while research on AI-based sign language recognition generally focuses solely on technical achievements. This study aims to develop a BISINDO alphabet recognition system based on Convolutional Neural Network (CNN) and evaluate it through a philosophy of science perspective. The methods used include collecting a BISINDO alphabet hand image dataset, data augmentation, and transfer learning-based CNN training with the MobileNetV2 architecture and a stepwise training scheme, then deployed to Android using TensorFlow Lite. Test results show the system is able to achieve an accuracy of around 93% on controlled test data with stable real-time inference performance. The scientific contribution of this research is not only in the development of applied AI systems, but also in providing a reflective ontological, epistemological, and axiological framework to assess the validity and social implications of BISINDO recognition technology.

Keywords:

Convolutional Neural Network, Indonesian Sign Language, Mobile Computing, Philosophy of Science, Inclusive Technology

I. PENDAHULUAN

Bahasa Isyarat Indonesia (BISINDO) merupakan bahasa gestural yang berkembang dari praktik komunitas Tuli dan dipakai luas dalam interaksi sehari-hari. Kendati demikian, ketimpangan literasi BISINDO di masyarakat umum menyebabkan hambatan komunikasi di ruang layanan publik, pendidikan, dan aktivitas sosial. Teknologi pengenalan gestur berbasis visi komputer menawarkan jalur mediasi—bukan untuk menggantikan BISINDO, melainkan sebagai teknologi bantu yang dapat mengurangi friksi komunikasi (Huroniyah & Martita, 2023).

Sebagian besar publikasi pengenalan bahasa isyarat cenderung menekankan aspek rekayasa: desain dataset, pemilihan arsitektur deep learning, serta peningkatan akurasi. Pendekatan tersebut penting, tetapi pada konteks BISINDO (yang beragam secara regional, kaya ekspresi non-manual, dan berakar pada relasi sosial), penilaian semata berdasarkan metrik klasifikasi dapat mengabaikan pertanyaan ilmiah yang lebih mendasar: apa yang sebenarnya direpresentasikan model, apa yang dimaksud ‘pengetahuan’ ketika model menghasilkan probabilitas kelas, dan nilai apa yang harus dijaga ketika teknologi menyentuh komunitas rentan (Budiman et al., 2023; Sholawati et al., 2022). Untuk menghadirkan kontribusi ilmiah yang lebih komprehensif, artikel ini tidak hanya menampilkan rancangan teknis aplikasi CNN, tetapi juga memberikan kerangka analisis dari sudut pandang filsafat sains guna memahami dampak teknologis dan kemanusiaan dari penelitian ini.

Perkembangan teknologi kecerdasan buatan, khususnya computer vision dan deep learning, membuka peluang pengembangan sistem penerjemah bahasa isyarat. Convolutional Neural Network (CNN) memiliki kemampuan ekstraksi fitur otomatis yang efisien untuk mengenali pola visual, termasuk gerakan tangan. Arsitektur ringan seperti MobileNetV2 dipilih karena efisien dijalankan di perangkat mobile dengan keterbatasan komputasi, tanpa mengorbankan akurasi. Penelitian ini bertujuan membangun sistem penerjemah BISINDO berbasis Android yang mampu mengenali huruf secara real-time dari kamera, menyusunnya menjadi kata, dan mengonversinya menjadi suara melalui teknologi text to speech (TTS). Dengan demikian, sistem diharapkan dapat menjadi sarana komunikasi praktis sekaligus media edukasi bahasa isyarat bagi masyarakat.

Perlu ditegaskan bahwa kontribusi utama penelitian ini tidak terletak pada pengajuan arsitektur Convolutional Neural Network yang sepenuhnya baru, melainkan pada integrasi antara studi kasus pengembangan sistem pengenalan alfabet BISINDO berbasis MobileNetV2 dengan evaluasi kritis melalui perspektif filsafat sains. Dengan demikian, kebaruan artikel ini berada pada upaya membaca capaian teknis model tidak hanya sebagai persoalan akurasi klasifikasi, tetapi juga sebagai persoalan representasi realitas bahasa, validitas pengetahuan yang dihasilkan model, dan implikasi etis-sosial dari teknologi inklusif yang dikembangkan.

Sejumlah penelitian terdahulu pada pengenalan bahasa isyarat umumnya berfokus pada peningkatan performa klasifikasi, pemilihan arsitektur model, dan efisiensi implementasi sistem. Namun, dalam konteks BISINDO, pendekatan tersebut belum sepenuhnya menjawab persoalan yang lebih mendasar, yaitu bagaimana sistem AI merepresentasikan bahasa isyarat sebagai fenomena sosial-linguistik, seberapa sah klaim pengetahuan yang dihasilkan model di luar data uji terkontrol, dan bagaimana konsekuensi etik serta inklusivitasnya ketika diterapkan pada komunitas pengguna nyata. Oleh karena itu, artikel ini memosisikan diri bukan sekadar sebagai laporan rekayasa sistem, tetapi sebagai studi yang menghubungkan performa teknis dengan evaluasi ontologis, epistemologis, dan aksiologis.

Artikel ini menyajikan studi kasus penerapan CNN untuk pengenalan alfabet BISINDO real-time, sekaligus membangun pembacaan filsafat sains untuk menilai (i) status ontologis data dan representasi, (ii) legitimasi epistemik klaim performa model, dan (iii) dimensi aksiologis etika, keadilan, dan kemanfaatan sosial. Dengan kerangka ini, kontribusi artikel tidak hanya berupa ‘cara membangun sistem’, tetapi juga ‘cara memahami’ apa yang dikerjakan sains terapan AI dalam domain bahasa dan disabilitas (Chollet, 2021; Sandler et al., 2018; TensorFlow, 2023).

Berdasarkan latar belakang dan tujuan artikel ini yakni mempertahankan studi kasus rekayasa sistem sekaligus menilai konsekuensi ilmiahnya. Penelitian ini tidak terletak pada pengembangan arsitektur Convolutional Neural Network yang sepenuhnya baru, melainkan pada pendekatan evaluatif yang mengintegrasikan studi kasus AI terapan dengan kajian filsafat sains. Berbeda dari penelitian pengenalan bahasa isyarat yang umumnya berfokus pada peningkatan akurasi dan performa teknis, artikel

ini secara eksplisit menganalisis implikasi ontologis, epistemologis, dan aksiologis dari penggunaan CNN dalam konteks Bahasa Isyarat Indonesia (BISINDO). Dengan demikian, penelitian ini menawarkan kontribusi konseptual berupa kerangka reflektif untuk menilai klaim keberhasilan sistem AI tidak hanya berdasarkan metrik klasifikasi, tetapi juga berdasarkan validitas pengetahuan yang dihasilkan serta konsekuensi etis dan sosialnya dalam teknologi inklusif.

II. METODE PENELITIAN

Secara metodologis, pengembangan sistem ini dapat dibaca melalui pendekatan design science/rekayasa perangkat lunak: membangun artefak (model + aplikasi), lalu mengevaluasi kegunaannya. Namun dalam perspektif filsafat sains, setiap langkah metodologis merupakan ‘komitmen epistemik’—peneliti memutuskan apa yang menjadi bukti (evidence) dan bagaimana bukti itu dinilai (Kuhn, 1996; Popper, 1959; Pressman & Maxim, 2019).

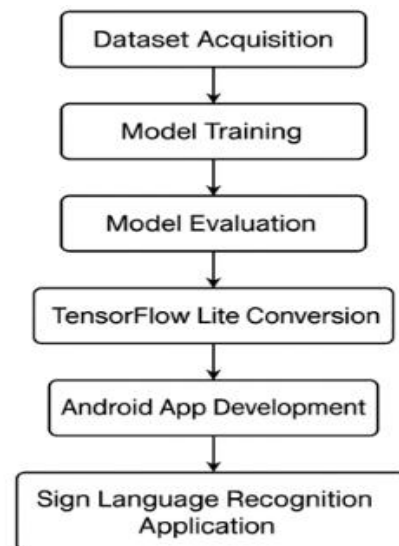
Pertama, operasionalisasi fenomena: bahasa isyarat diproyeksikan menjadi citra statis tangan dan label huruf. Ini memudahkan pengukuran, tetapi menyisakan residu semantik yang tidak terukur (intonasi gestural, koartikulasinya, ekspresi non-manual).

Kedua, konstruksi data: augmentasi (mis. perubahan latar, rotasi, pencahayaan) bukan sekadar trik teknis, melainkan intervensi yang membentuk ‘dunia’ tempat model belajar. Validitas eksternal karenanya bergantung pada kedekatan dunia data dengan dunia penggunaan (Chollet, 2021).

Ketiga, validasi: metrik akurasi/precision/recall memberi indikator performa, tetapi tidak cukup untuk menyimpulkan keterandalan lintas-konteks. Evaluasi yang lebih kuat seharusnya juga memeriksa (i) ketahanan terhadap out-of-distribution (pencahayaan ekstrem, latar ramai, variasi bentuk tangan), (ii) latensi dan stabilitas inferensi real-time, dan (iii) dampak kesalahan (cost of error) bagi pengguna (Mittelstadt et al., 2016; Selbst et al., 2019).

Dalam penelitian ini beberapa tahapan dilakukan (gambar 1), Pertama, dilakukan pengumpulan dataset (Dataset Acquisition) berupa citra tangan yang merepresentasikan huruf-huruf dalam Bahasa Isyarat Indonesia. Selanjutnya, dataset digunakan dalam

proses pelatihan model (Model Training) dengan arsitektur MobileNetV2 untuk memperoleh representasi fitur dan menghasilkan model klasifikasi. Model yang telah dilatih kemudian melalui tahap evaluasi (Model Evaluation) menggunakan metrik akurasi, precision, recall, serta confusion matrix untuk mengukur kinerjanya. Setelah diperoleh model dengan performa terbaik, dilakukan konversi ke TensorFlow Lite (TensorFlow Lite Conversion) agar dapat dijalankan secara efisien di perangkat Android. Tahap berikutnya adalah pengembangan aplikasi Android (Android App Development) dengan mengintegrasikan model TFLite, kamera, serta modul text-to-speech (TTS). Hasil akhir adalah sebuah aplikasi pengenalan bahasa isyarat (Sign Language Recognition Application) yang mampu mengenali pola gambar menjadi huruf secara real-time, menyusunnya menjadi kata, dan mengonversinya menjadi suara.



Gambar 1. Kerangka Kerja Pengembangan Aplikasi

Arsitektur Model

Arsitektur CNN pada penelitian ini terdiri dari beberapa lapisan utama seperti pada tabel 1.

Penelitian ini menggunakan pendekatan transfer learning bertahap pada MobileNetV2, namun dokumentasi rinci mengenai parameter pelatihan seperti optimizer, learning rate, batch size, dan jumlah epoch belum tersaji secara lengkap dalam naskah awal. Oleh karena itu, aspek tersebut dinyatakan sebagai keterbatasan pelaporan metodologis yang perlu diperbaiki pada pengembangan artikel

selanjutnya, terutama untuk mendukung reproduksibilitas eksperimen secara lebih kuat.

Tabel 1. Arsitektur Lapisan dan Fungsi CNN

No	Lapisan	Fungsi
1	<i>Input Layer</i>	Menerima input citra berukuran 128 x 128 piksel dengan 3 channel warna.
2	<i>Base Model: MobileNetV2</i>	Sebagai ekstraktor fitur visual. Pada tahap awal, layer ini dibekukan (non-trainable).
3	<i>Global Average Pooling 2D</i>	Mereduksi dimensi output dari <i>feature extractor</i> menjadi 1D.
4	<i>Dense Layer (128 unit)</i>	<i>Layer fully connected</i> dengan aktivasi ReLU untuk pembelajaran fitur lanjut.
5	<i>Dropout (rate 0.3)</i>	Untuk mengurangi <i>overfitting</i> dengan cara mengabaikan 30% neuron saat training.
6	<i>Output Layer</i>	<i>Dense layer</i> dengan 26 neuron dan aktivasi Softmax, menghasilkan probabilitas untuk masing-masing huruf A-Z.

Dataset

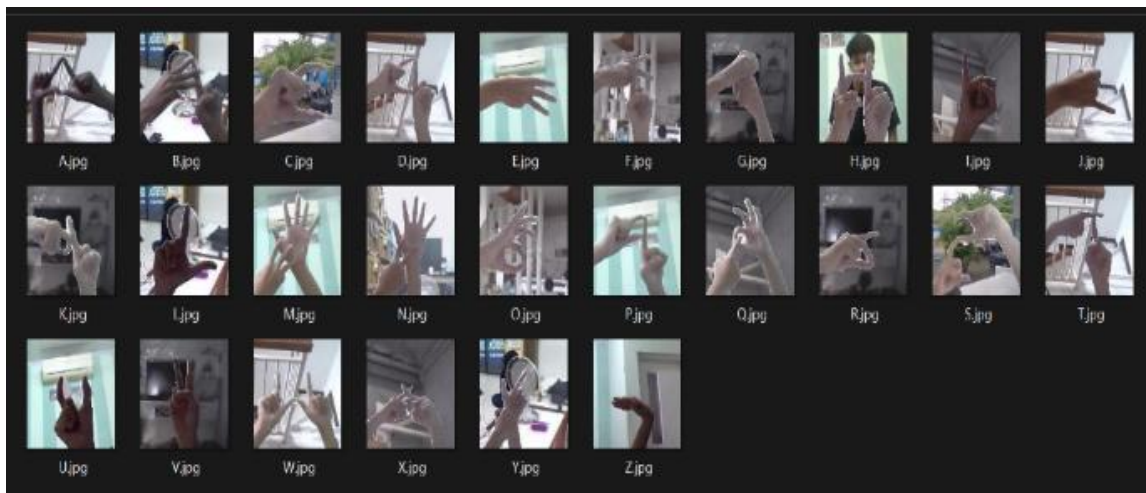
Dataset merupakan komponen penting dalam pengembangan sistem pengenalan bahasa isyarat berbasis deep learning. Dataset yang digunakan dalam penelitian ini berupa gambar citra tangan yang merepresentasikan alfabet dalam Bahasa Isyarat Indonesia (BISINDO), dengan masing-masing huruf

sebagai satu kelas (A sampai Z). Dataset original diambil secara mandiri melalui sebuah program python dimana gambar diambil melalui sebuah webcam. Total dataset original adalah 10490 gambar yang terdiri dari 26 kelas.

Dataset yang digunakan terdiri atas 10.490 citra tangan alfabet BISINDO untuk 26 kelas huruf A–Z yang dikumpulkan melalui webcam. Dalam penelitian ini, dokumentasi dataset yang tersedia telah memuat jumlah total citra dan struktur kelas, namun belum mencantumkan secara rinci jumlah partisipan, distribusi karakteristik pengguna, serta variasi kondisi pencahayaan dan latar belakang pada saat akuisisi data. Keterbatasan dokumentasi tersebut penting dicatat karena faktor-faktor tersebut berpotensi memengaruhi kemampuan generalisasi model pada situasi penggunaan nyata.

Augmentasi Background dengan Mediapipe Selfie Segmentation

Untuk meningkatkan variasi dataset dan memperkuat kemampuan model dalam menghadapi kondisi nyata, dilakukan proses augmentasi citra dengan cara mengganti background asli dari gambar tangan dengan latar belakang baru. Teknik ini menggunakan MediaPipe Selfie Segmentation, yaitu metode segmentasi citra berbasis machine learning yang dapat memisahkan objek manusia (termasuk tangan) dari latar belakang secara otomatis.

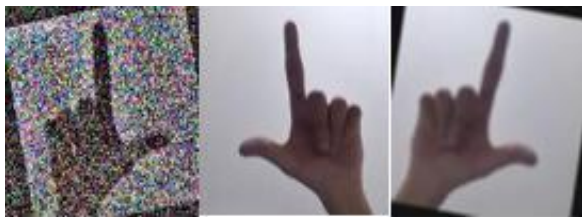


Gambar 2. Augmentasi Background

Augmentasi Gambar (*Albumentations*)

Augmentasi lanjutan dilakukan pada dataset yang telah memiliki latar belakang baru beserta gambar originalnya, proses ini bertujuan untuk meningkatkan variasi dan kekayaan fitur visual dalam data pelatihan tanpa menambah jumlah data asli. Proses augmentasi ini terdiri dari empat variasi untuk setiap gambar: gambar asli, hasil flip horizontal (cermin), augmentasi ringan (rotate, brightness/contrast, noise, blur), serta kombinasi augmentasi kompleks (shift, scale, rotate, motion blur, CLAHE).

Dari sudut pandang metodologis, augmentasi tidak hanya berfungsi untuk memperbanyak variasi data pelatihan, tetapi juga untuk mensimulasikan sebagian kondisi visual yang mungkin muncul pada penggunaan nyata, seperti perubahan orientasi tangan, pencahayaan, noise, dan kompleksitas latar. Meskipun demikian, augmentasi tetap merupakan konstruksi sintetik yang tidak sepenuhnya menggantikan keberagaman kondisi lapangan, sehingga hasil pelatihan tetap perlu dibaca dalam batas distribusi data yang berhasil direpresentasikan.



Gambar 3. Augmentasi Gambar

Pembagian Dataset

Dataset yang telah dikumpulkan dan diaugmentasi dibagi menjadi empat bagian utama untuk mendukung proses pelatihan dan evaluasi model, yaitu: train set, val set, test set, dan eval set. Adapun persentase pembagiannya adalah sebagai berikut:

- a. Train DataSet (60%)
Dataset ini digunakan untuk proses pelatihan model Convolutional Neural Network (CNN). Model akan mempelajari pola-pola citra dari data ini untuk mengklasifikasikan gesture bahasa isyarat secara akurat.
- b. Validation DataSet (15%)
Digunakan selama proses pelatihan untuk memantau performa model dan menghindari overfitting. Data ini tidak dilihat oleh model

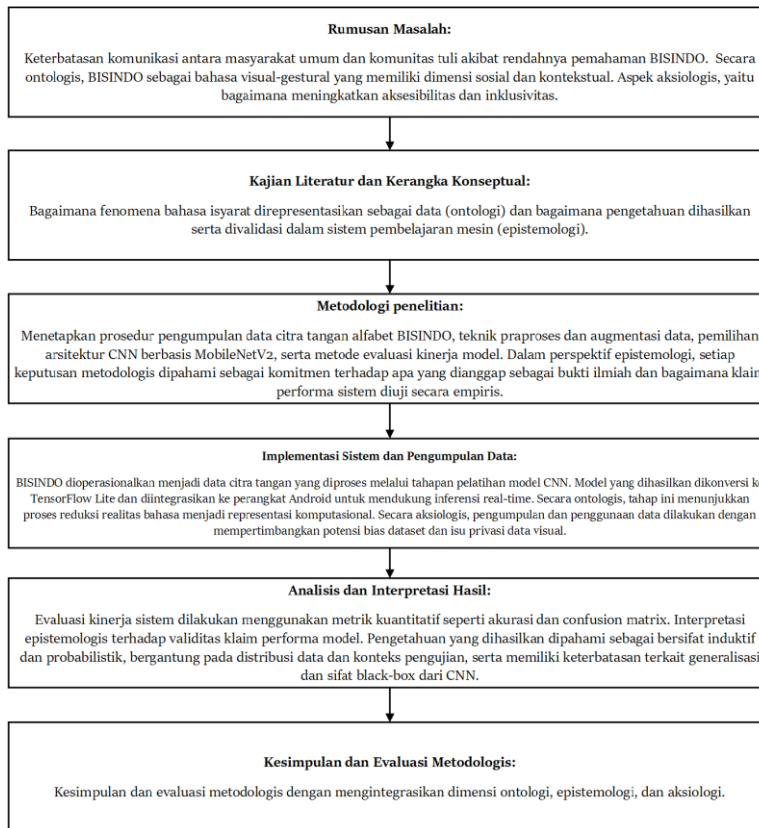
selama training, sehingga dapat memberikan gambaran performa model terhadap data yang belum pernah dilatih.

- c. Test DataSet (15%)
Dataset ini digunakan untuk melakukan evaluasi akhir terhadap performa model dalam format .h5 sebelum dikonversi ke TensorFlow Lite. Evaluasi dilakukan menggunakan metrik seperti akurasi, confusion matrix, dan classification report.
- d. Evaluation DataSet (10%)
Evaluation set ini digunakan untuk mengukur kinerja model setelah dikonversi ke format TensorFlow Lite (.tflite). Pengujian dilakukan untuk memastikan bahwa performa model tetap konsisten saat dijalankan di lingkungan aplikasi mobile.

Pembagian ini bertujuan untuk memastikan bahwa model dapat belajar secara optimal dari data pelatihan, dievaluasi secara objektif, dan divalidasi ulang performanya setelah deployment ke perangkat Android.

Metode Penelitian Berbasis Kajian Filsafat Sains

Kajian filsafat sains dalam penelitian ini tidak diposisikan sebagai refleksi normatif yang berdiri di luar eksperimen, melainkan sebagai kerangka analisis yang diterapkan secara langsung pada setiap tahap penelitian. Dimensi ontologi digunakan untuk menelaah bagaimana fenomena bahasa isyarat direduksi menjadi citra, label, dan fitur visual yang dapat diproses model. Dimensi epistemologi digunakan untuk menilai status klaim pengetahuan yang dihasilkan model, terutama terkait akurasi, generalisasi, dan keterbatasan inferensi probabilistik. Sementara itu, dimensi aksiologi digunakan untuk menilai implikasi nilai dari sistem yang dibangun, termasuk isu aksesibilitas, bias dataset, privasi data visual, dan tanggung jawab sosial dalam pengembangan teknologi bantu berbasis AI. Dengan kerangka ini, analisis filsafat sains menjadi bagian evaluatif yang terintegrasi dengan hasil eksperimen, bukan sekadar pelengkap konseptual.



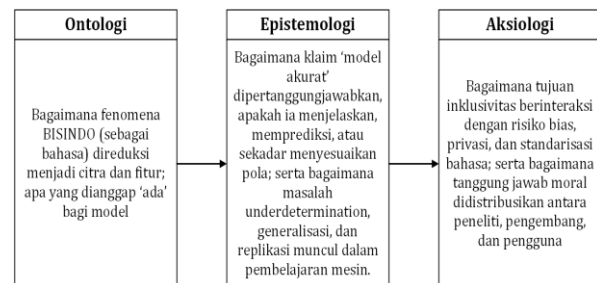
Gambar 4. Tahapan Penelitian

Kerangka Filsafat Sains Untuk Riset Ai Terapan

Dalam filsafat sains, sains tidak hanya dipahami sebagai kumpulan hasil, tetapi sebagai praktik produksi pengetahuan. Pada riset AI, praktik ini mencakup: memilih apa yang diukur, mendefinisikan label, membangun prosedur observasi (akuisisi data), serta menentukan kriteria keberhasilan (metrik). Dengan kata lain, model AI adalah artefak teknosains yang menggabungkan asumsi teoritis, keputusan metodologis, dan nilai-nilai sosial (Hempel, 1965; Kuhn, 1996; Lakatos, 1970; Popper, 1959; van Fraassen, 1980).

Terdapat tiga lensa yang digunakan pada artikel ini. (i) ontologi: bagaimana fenomena BISINDO (sebagai bahasa) direduksi menjadi citra dan fitur; apa yang dianggap ‘ada’ bagi model. (ii) epistemologi: bagaimana klaim ‘model akurat’ dipertanggungjawabkan apakah ia menjelaskan, memprediksi, atau sekadar menyesuaikan pola; serta bagaimana masalah underdetermination, generalisasi, dan replikasi muncul dalam pembelajaran mesin. (iii) aksiologi: bagaimana tujuan inklusivitas berinteraksi dengan risiko bias, privasi, dan standarisasi bahasa;

serta bagaimana tanggung jawab moral didistribusikan antara peneliti, pengembang, dan pengguna (Costanza-Chock, 2020; Mittelstadt et al., 2016; Selbst et al., 2019; van Fraassen, 1980; Wachter et al., 2017).



Gambar 5. Kerangka Filsafat Sains

Kerangka ini memungkinkan evaluasi yang lebih ‘substansial’ daripada laporan teknis murni: model dapat unggul secara statistik namun tetap rapuh secara epistemik (misalnya gagal pada konteks pencahayaan/skin-tone tertentu), atau bermasalah

secara aksiologis (misalnya pengumpulan data tanpa consent yang memadai) (Crawford, 2021; Mittelstadt et al., 2016; Selbst et al., 2019).

III. ANALISIS DAN PERANCANGAN

Studi Kasus Teknis: Aplikasi Pengenalan Bahasa Isyarat

Studi kasus difokuskan pada pengenalan alfabet A–Z dalam BISINDO berbasis citra tangan. Cakupan alfabet dipilih sebagai tahap awal yang terukur, mengingat sistem bahasa isyarat yang utuh melibatkan dinamika temporal, ekspresi wajah, dan konteks linguistik yang lebih kompleks (Budiman et al., 2023; Sholawati et al., 2022).

Alur sistem: (i) *akuisisi citra tangan*, (ii) *praproses dan augmentasi* untuk meningkatkan keragaman kondisi, (iii) *pelatihan model CNN* berbasis transfer learning (MobileNetV2), (iv) *evaluasi menggunakan data uji terpisah*, dan (v) *deployment* ke Android menggunakan TensorFlow Lite untuk inferensi real-time dan integrasi text-to-speech (Chollet, 2021; Sandler et al., 2018; TensorFlow, 2023).

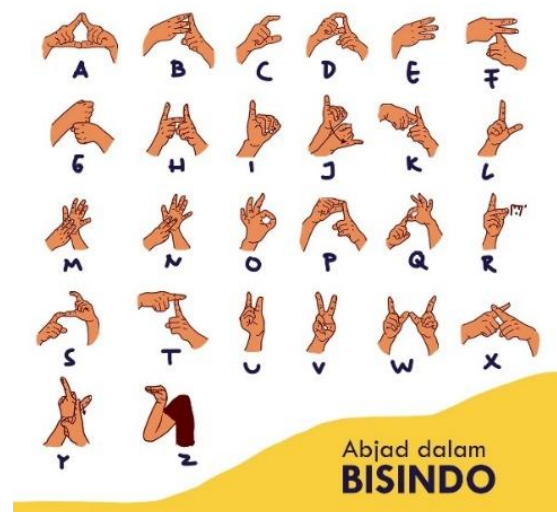
Tabel 2. Ringkasan Arsitektur Aplikasi dan Implikasi Filosofisnya

Komponen	Peran	Catatan Filosofis
Input citra 128×128 RGB	Representasi observasi gestur	Mengubah fenomena komunikasi menjadi data visual terkuantisasi (reduksi).
MobileNetV2 (feature extractor) (Sandler et al., 2018)	Menyarikan pola visual	Ciri yang dipelajari adalah konstruksi statistik; tidak identik dengan makna linguistik.
Global Average Pooling + Dense	Mengubah peta fitur menjadi keputusan kelas	Keputusan adalah inferensi probabilistik; memerlukan interpretasi kehati-hatian.
Softmax 26 kelas (A–Z)	Probabilitas huruf	Operasionalisasi bahasa dibatasi pada alfabet; konteks, morfologi, dan sintaks BISINDO belum tercakup.
TensorFlow Lite (Android) (TensorFlow, 2023)	Deployment real-time	Kecepatan/latensi menjadi bagian dari ‘kriteria kebenaran praktis’ pada teknologi terapan.

Pembatasan penelitian pada alfabet statis A–Z perlu dipahami sebagai keputusan ruang lingkup agar eksperimen dapat dilakukan secara terukur pada tahap awal. Konsekuensinya, sistem yang dihasilkan belum merepresentasikan kompleksitas penuh BISINDO, karena bahasa isyarat dalam praktik nyata juga melibatkan gestur dinamis, transisi gerak, ekspresi wajah, dan konteks interaksi. Dengan demikian, hasil penelitian ini sebaiknya dibaca sebagai fondasi awal untuk pengembangan sistem yang lebih komprehensif, bukan sebagai representasi utuh dari seluruh praktik bahasa isyarat Indonesia.

Bahasa Isyarat Indonesia

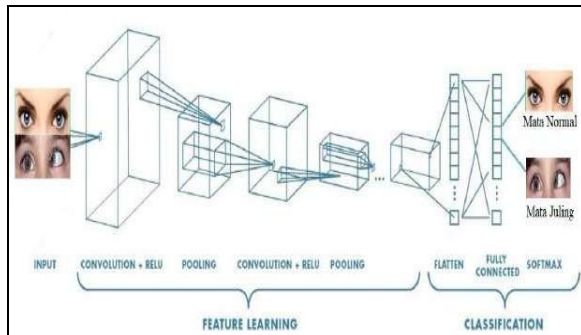
Bahasa isyarat adalah sistem komunikasi visual yang menggunakan gerakan tangan, ekspresi wajah, dan postur tubuh untuk menyampaikan informasi, khususnya digunakan oleh penyandang tunarungu atau gangguan bicara. Setiap negara atau daerah memiliki sistem bahasa isyarat yang berbeda, seperti American Sign Language (ASL), British Sign Language (BSL). Di Indonesia terdapat 2 jenis bahasa isyarat, yaitu SIBI dan BISINDO (Gracita, 2024). SIBI sendiri merupakan bahasa isyarat yang diciptakan oleh orang dengar dan diakui oleh pemerintah, sementara BISINDO lahir dari para penyandang tunarungu itu sendiri. (Huronyah & Martita, 2023) Dalam konteks penelitian ini, bahasa isyarat BISINDO (Gracita, 2024) digunakan sebagai objek utama yang akan dikenali dan diterjemahkan ke dalam bentuk teks atau suara.



Gambar 6. Bahasa Isyarat BISINDO

Convolutional Neural Network

Convolutional Neural Network (CNN) merupakan salah satu jenis algoritma neural network yang didesain untuk memproses data citra. Algoritma CNN digunakan untuk mempelajari serta mendeteksi feature pada citra digital. (Sholawati et al., 2022) CNN terdiri dari beberapa lapisan utama seperti convolutional layer, pooling layer, dan fully connected layer yang bekerja bersama untuk mengekstraksi dan mempelajari fitur penting dari sebuah gambar.



Gambar 7. Arsitektur CNN

Konvolusi: Tahap pertama dalam arsitektur CNN adalah konvolusi. Proses ini menggunakan kernel dengan ukuran tertentu untuk mengekstraksi fitur dari citra. Jumlah kernel yang digunakan ditentukan berdasarkan jumlah fitur yang ingin dihasilkan.

Fungsi Aktivasi: Setelah melewati tahap konvolusi, hasilnya kemudian dimasukkan ke dalam fungsi aktivasi, yang umumnya menggunakan ReLU (Rectified Linear Unit). Fungsi aktivasi ini memperkenalkan non-linearitas ke dalam model, memungkinkan CNN untuk mempelajari representasi yang lebih kompleks dari data.

Pooling: Setelah melalui fungsi aktivasi, fitur-fitur yang dihasilkan menjalani proses pooling. Pooling bertujuan untuk mengurangi dimensi spasial dari fitur-fitur tersebut, sehingga mengurangi jumlah parameter dan komputasi yang dibutuhkan dalam jaringan. Operasi pooling seperti max-pooling atau average-pooling sering digunakan. Proses *konvolusi*, *fungsi aktivasi*, dan *pooling* diulang beberapa kali untuk menghasilkan peta fitur yang cukup untuk kemudian dimasukkan ke dalam fully connected neural network. Dari fully connected network, CNN menghasilkan output dalam bentuk kelas-kelas yang direpresentasikan oleh data citra tersebut.

Dalam penelitian ini, algoritma Convolutional Neural Network (CNN) berperan sebagai inti dari sistem klasifikasi untuk mengenali bahasa isyarat. Model CNN membutuhkan proses pelatihan (training) menggunakan dataset yang berisi contoh gambar bahasa isyarat. Selama pelatihan, model akan belajar mengenali pola-pola visual pada setiap kelas isyarat. Setelah proses pelatihan selesai, model akan dievaluasi menggunakan metrik seperti akurasi, presisi, recall, dan F1-score untuk mengetahui seberapa baik model mengenali isyarat pada data uji. Proses ini penting untuk menjamin kinerja aplikasi deteksi bahasa isyarat yang dibangun. CNN digunakan untuk memproses citra tangan yang diambil melalui kamera dan mengidentifikasi bentuk atau gerakan tangan yang merepresentasikan huruf atau kata dalam bahasa isyarat.

MobileNetV2

MobileNetV2 merupakan arsitektur Convolutional Neural Network (CNN) yang dikembangkan Google AI pada 2018 untuk mendukung pembelajaran mendalam pada perangkat dengan keterbatasan komputasi, seperti smartphone dan IoT (Ridhovan & Suharso, 2022). Arsitektur ini menyeimbangkan akurasi dan efisiensi melalui penggunaan depthwise separable convolution, yaitu pemisahan ekstraksi fitur spasial (depthwise) dan kombinasi fitur (pointwise 1x1 convolution). Inovasi ini mampu mengurangi jumlah parameter dan beban komputasi secara signifikan tanpa mengurangi kemampuan dalam mengenali pola kompleks (Aufar & Kaloka, 2022).

Hasil rancangan dan pengujian menghasilkan dataset augmentasi yang diperoleh dari pengolahan dataset asli dan dataset latar belakang menggunakan teknik seperti rotasi, penambahan noise, dan perubahan kecerahan acak. Dataset tersebut digunakan dalam pelatihan model dengan pendekatan transfer learning menggunakan arsitektur MobileNetV2 melalui dua tahap, yaitu transfer learning dan fine-tuning. Model hasil pelatihan disimpan dalam format .h5 sebagai model utama, kemudian dikonversi ke format TensorFlow Lite (.tflite) agar dapat diimplementasikan pada aplikasi Android untuk mendukung klasifikasi bahasa isyarat secara real time.

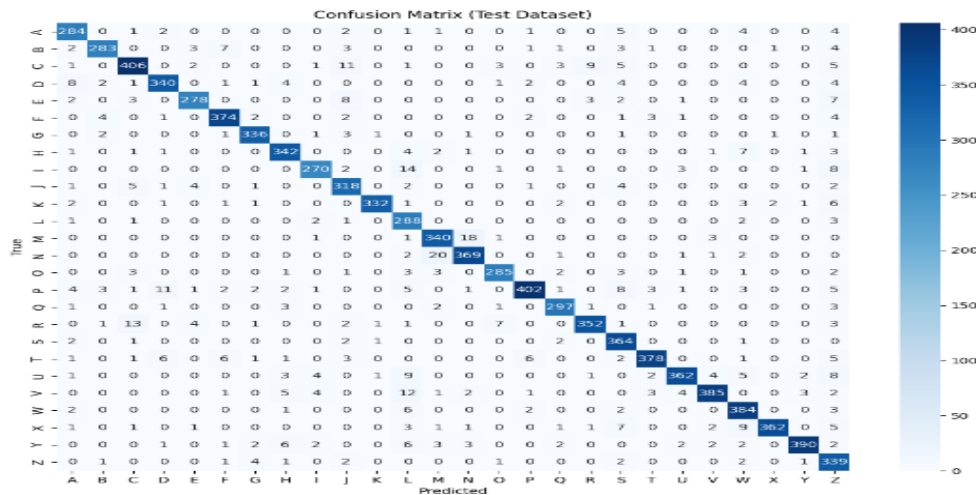
Evaluasi Model dengan Testing Set

Setelah model selesai dilatih, tahap selanjutnya adalah mengevaluasi performa model menggunakan data uji. Evaluasi ini dilakukan untuk mengukur akurasi dan loss model pada data yang belum pernah dilihat oleh model selama pelatihan. Semua proses evaluasi dilakukan dengan menggunakan bahasa pemrograman Python dengan Visual Studio Code sebagai code editor. Hasil evaluasi model menggunakan testing dataset, yang menghasilkan akurasi sebesar 92.96% dan loss sebesar 22%. Evaluasi ini bertujuan untuk mengukur kinerja model terhadap data yang belum pernah dilihat selama pelatihan, sehingga memberikan gambaran objektif terhadap kemampuan generalisasi model.

Jika dibandingkan dengan penelitian-penelitian pengenalan bahasa isyarat yang menggunakan CNN pada alfabet atau gestur statis, capaian akurasi 92,96% menunjukkan bahwa model yang dikembangkan berada pada tingkat performa yang kompetitif untuk skenario klasifikasi terkontrol. Namun demikian, artikel ini tidak menempatkan nilai akurasi tersebut sebagai klaim superioritas teknis terhadap seluruh studi terdahulu, karena adanya perbedaan karakteristik dataset, skenario akuisisi, jumlah kelas, strategi augmentasi, serta lingkungan deployment yang digunakan oleh masing-masing penelitian. Oleh karena itu, kontribusi penelitian ini lebih tepat diletakkan pada integrasi antara implementasi model ringan untuk perangkat bergerak dan pembacaan kritis terhadap makna ilmiah dan sosial dari performa yang dihasilkan.

Hasil dari evaluasi lanjutan ini ditunjukkan melalui dua komponen utama, yaitu confusion matrix dan classification report.

1. Confusion matrix: Menampilkan jumlah prediksi benar dan salah untuk masing-masing kelas (huruf), sehingga memudahkan identifikasi huruf-huruf yang paling sering tertukar. Confusion matrix hasil dari evaluasi lanjutan dapat dilihat pada gambar 8.



Gambar 8. Confusion Matrix

Kesalahan tersebut kemungkinan disebabkan oleh variasi posisi tangan, pencahayaan, serta keterbatasan jumlah data pada kelas tertentu. Secara keseluruhan, hasil ini mengindikasikan bahwa model memiliki generalisasi yang baik dengan rata-rata akurasi tinggi.

- Classification report : Menyajikan metrik yang lebih mendetail seperti precision, recall, f1-score, dan support untuk setiap kelas. Classification report hasil evaluasi lanjutan dapat dilihat pada gambar 9.

	precision	recall	f1-score	support
A	0.90	0.93	0.92	305
B	0.96	0.92	0.94	309
C	0.93	0.91	0.92	447
D	0.93	0.91	0.92	372
E	0.95	0.91	0.93	304
F	0.95	0.95	0.95	394
G	0.96	0.97	0.96	348
H	0.93	0.94	0.93	364
I	0.94	0.90	0.92	300
J	0.88	0.94	0.91	339
K	0.99	0.94	0.97	352
L	0.80	0.97	0.88	298
M	0.91	0.93	0.92	364
N	0.93	0.93	0.93	396
O	0.95	0.93	0.94	305
P	0.96	0.88	0.92	456
Q	0.95	0.96	0.95	310
R	0.96	0.91	0.93	386
S	0.88	0.98	0.93	373
T	0.97	0.92	0.94	411
U	0.96	0.90	0.93	402
V	0.97	0.91	0.94	423
W	0.89	0.96	0.93	400
X	0.99	0.92	0.95	395
Y	0.98	0.92	0.95	424
Z	0.79	0.96	0.87	354
accuracy			0.93	9531
macro avg	0.93	0.93	0.93	9531
weighted avg	0.93	0.93	0.93	9531

Gambar 9. Classification Report

Untuk setiap huruf (A sampai Z), tertera empat metrik:

- Precision: Mengukur seberapa akurat prediksi untuk suatu kelas. Misalnya Kelas A precision = 0.90, artinya 90% prediksi “A” memang benar A.
- Recall: Mengukur seberapa banyak data kelas tersebut yang berhasil dideteksi model. Misalnya Kelas A recall = 0.93, berarti 93% data A teridentifikasi dengan benar.
- F1-score: Harmonic mean dari precision dan recall. Mengukur keseimbangan antara keduanya. Misalnya Kelas A F1 = 0.92, menunjukkan performa keseluruhan yang baik.
- Support: Jumlah data sebenarnya untuk kelas tersebut dalam dataset uji. Support kelas A = 305, berarti ada 305 sampel huruf A.

Secara keseluruhan model memiliki performa sangat baik dengan accuracy 93%, dimana hampir

semua kelas memiliki f1-score tinggi (0.90 ke atas), hanya kelas tertentu seperti Z yang sedikit lebih lemah, dataset terlihat seimbang, terlihat dari support yang seragam. Secara keseluruhan model sudah bekerja stabil dan kuat untuk pengenalan huruf A-Z.

Meskipun evaluasi menunjukkan performa yang baik pada testing set, hasil ini terutama merepresentasikan kemampuan generalisasi internal terhadap distribusi data yang masih dekat dengan data pelatihan. Evaluasi dalam penelitian ini belum secara khusus menguji robustness model pada kondisi out-of-distribution, seperti pencahayaan ekstrem, latar belakang yang sangat kompleks, jarak tangan yang bervariasi, occlusion, atau variasi bentuk tangan pengguna yang lebih beragam. Karena itu, akurasi yang diperoleh perlu dipahami sebagai indikator performa awal, bukan sebagai jaminan penuh terhadap kestabilan sistem pada seluruh konteks penggunaan nyata.

Kajian Filsafat

Pada tataran teknis, prototipe dibangun melalui alur: akuisisi citra tangan, praproses dan augmentasi, pelatihan CNN berbasis transfer learning dengan MobileNetV2 sebagai pengekstraksi fitur, evaluasi pada data uji terpisah, lalu konversi dan integrasi model ke Android menggunakan TensorFlow Lite untuk inferensi real-time. Pada pengujian terkontrol, sistem menunjukkan performa klasifikasi alfabet yang tinggi dan latensi inferensi yang memadai untuk penggunaan on-device. (Chollet, 2021; Sandler et al., 2018; TensorFlow, 2023).

Namun, penilaian filsafat sains atas capaian tersebut menuntut pembacaan yang lebih ketat daripada sekadar metrik. Secara ontologis, keberhasilan model dicapai melalui reduksi fenomena BISINDO menjadi objek yang ‘dapat dihitung’ (piksel/tensor/label), sehingga makna linguistik dan konteks sosial tidak otomatis terwakili. Secara epistemologis, keluaran softmax adalah pengetahuan induktif probabilistik yang valid dalam batas distribusi data; karenanya klaim ‘akurat’ perlu dibatasi oleh isu black-box, underdetermination, dan generalisasi lintas-konteks. Secara aksiologis, manfaat aksesibilitas harus ditimbang terhadap risiko bias, privasi/biometrik, dan potensi standardisasi varian BISINDO; konsekuensinya, evaluasi dan tata kelola perlu melibatkan komunitas Tuli serta mekanisme akuntabilitas saat sistem gagal.

Ontologi: Apa yang ‘Ada’ bagi Model?

Hasil eksperimen menunjukkan bahwa model mampu mengenali alfabet BISINDO dengan akurasi tinggi pada data uji terkontrol. Namun, dari sudut pandang ontologi, keberhasilan tersebut diperoleh melalui proses reduksi fenomena bahasa isyarat menjadi entitas visual yang dapat dihitung, yaitu piksel, tensor, dan label huruf A–Z. Artinya, yang dihadapi model bukanlah bahasa isyarat sebagai praktik sosial-linguistik yang utuh, melainkan representasi visual terbatas dari sebagian aspek bahasa isyarat. Dengan demikian, keberhasilan klasifikasi tidak identik dengan keberhasilan memahami bahasa secara menyeluruh (Huronyah & Martita, 2023).

Konsekuensi ontologisnya adalah reduksionisme representasional: model dapat ‘benar’ secara klasifikasi sambil tetap ‘buta’ terhadap aspek bahasa yang esensial. Selain itu, proses segmentasi/augmentasi menciptakan realitas sintetik: latar diganti, pencahayaan dimodifikasi, dan variasi dibuat algoritmik. Ini memperlihatkan bahwa data dalam AI bukan cermin pasif dunia, melainkan hasil konstruksi dunia yang disusun agar dapat dipelajari mesin (Kuhn, 1996; van Fraassen, 1980).

Epistemologi: Status Pengetahuan yang Dihasilkan CNN

Nilai akurasi 92,96% dan hasil classification report memperlihatkan bahwa model memiliki kemampuan prediktif yang kuat pada data uji yang digunakan. Namun secara epistemologis, pengetahuan yang dihasilkan model bersifat probabilistik dan bergantung pada kedekatan antara data baru dengan distribusi data pelatihan. Oleh sebab itu, keluaran model lebih tepat dipahami sebagai inferensi statistik berbasis pola visual daripada pengetahuan eksplanatoris mengenai makna bahasa isyarat itu sendiri. Dalam konteks ini, klaim “model akurat” perlu dibatasi secara hati-hati agar tidak melampaui validitas bukti yang tersedia.

Pengetahuan yang dihasilkan CNN bersifat induktif dan probabilistik. Model ‘mengetahui’ huruf A–Z sejauh ia mampu menggeneralisasi dari contoh pelatihan ke contoh baru dalam distribusi yang serupa. Ini berbeda dari pengetahuan eksplanatoris (mis. alasan kausal mengapa suatu gestur bermakna), sehingga klaim ilmiahnya lebih dekat pada prediksi instrumental daripada penjelasan teoritis (Chollet, 2021; van Fraassen, 1980).

Secara epistemik, beberapa risiko perlu dicatat. Pertama, underdetermination: banyak model dan konfigurasi dapat mencapai akurasi serupa tanpa menjamin representasi internal yang sama. Kedua, black-box problem: keputusan model sulit dijelaskan kepada pengguna, padahal teknologi bantu komunikasi menuntut transparansi dan kemampuan audit. Ketiga, masalah generalisasi: akurasi tinggi pada data uji tidak serta-merta menjamin performa pada pengguna baru atau lingkungan baru (Mittelstadt et al., 2016; van Fraassen, 1980; Wachter et al., 2017).

Implikasinya, validitas sebaiknya dibangun melalui triangulasi bukti: (i) evaluasi lintas-skenario (out-of-distribution testing), (ii) analisis kesalahan yang memperhitungkan dampak praktis, (iii) pelaporan ketidakpastian (mis. threshold confidence), dan (iv) praktik reproduksibilitas (versioning dataset, dokumentasi preprocessing, dan keterlacakan eksperimen) (Chollet, 2021; Mittelstadt et al., 2016).

Aksiologi: Nilai, Etika, dan Tanggung Jawab

Implementasi model ke perangkat Android melalui TensorFlow Lite menunjukkan potensi praktis yang penting bagi pengembangan teknologi bantu komunikasi yang lebih mudah diakses. Akan tetapi, dari perspektif aksiologi, kebermanfaatannya tersebut harus dibaca bersamaan dengan tanggung jawab etik yang menyertainya. Sistem yang dibangun dapat memberi dukungan pada aksesibilitas komunikasi, tetapi juga membawa risiko bias akibat keterbatasan dataset, risiko miskomunikasi akibat prediksi yang salah, serta risiko privasi karena pemrosesan citra tangan merupakan bagian dari data visual personal. Oleh sebab itu, nilai guna sistem harus diseimbangkan dengan prinsip kehati-hatian, transparansi, dan pelibatan pengguna yang bermakna.

Aksiologi menempatkan teknologi ini dalam horizon nilai. Manfaat utamanya adalah aksesibilitas: memperluas kesempatan interaksi bagi komunitas Tuli dan masyarakat luas. Namun, manfaat tersebut harus ditimbang terhadap risiko: bias dataset (mis. variasi warna kulit, bentuk tangan, atau latar tertentu), kesalahan interpretasi yang dapat memicu miskomunikasi, serta isu privasi karena data video/citra berpotensi mengandung informasi biometrik (Crawford, 2021; Mittelstadt et al., 2016; Selbst et al., 2019).

Lebih jauh, terdapat risiko standarisasi yang tidak sensitif budaya: BISINDO memiliki variasi regional; apabila sistem hanya melatih satu varian,

maka ia dapat mendorong homogenisasi dan menggeser praktik lokal. Prinsip etika yang relevan mencakup partisipasi bermakna komunitas Tuli dalam desain (co-design), persetujuan dan tata kelola data yang jelas, serta mekanisme akuntabilitas ketika sistem gagal (mis. log kesalahan, fallback ke input manual, dan pemberian peringatan ketidakpastian) (Costanza-Chock, 2020; Huroniyah & Martita, 2023).

Implikasi: Dari ‘Akurasi’ ke ‘Kelayakan Ilmiah dan Sosial’

Jika dibaca sebagai riset filsafat sains terapan, keberhasilan sistem tidak boleh direduksi pada akurasi. Kriteria yang lebih mendekati kelayakan ilmiah dan sosial meliputi: (i) robustitas lintas-konteks, (ii) kejelasan batas penggunaan (scope), (iii) mekanisme interpretabilitas dan pelaporan ketidakpastian, (iv) evaluasi berbasis pengguna (usability dan beban kognitif), serta (v) kesesuaian dengan prinsip etika dan keadilan (Mittelstadt et al., 2016; Selbst et al., 2019; Wachter et al., 2017).

Dengan demikian, kontribusi riset AI untuk BISINDO sebaiknya dipahami sebagai proyek sosio-teknis: kualitas ilmiahnya ditentukan oleh integrasi antara model, data, protokol evaluasi, dan tata kelola nilai. Kerangka ontologi–epistemologi–aksiologi memberi perangkat untuk menilai integrasi tersebut secara eksplisit (Costanza-Chock, 2020; Crawford, 2021; Selbst et al., 2019).

IV. KESIMPULAN

Sejalan dengan tujuan penelitian yang dirumuskan pada bagian pendahuluan, penelitian ini menunjukkan bahwa penerapan Convolutional Neural Network berbasis MobileNetV2 yang dideploy melalui TensorFlow Lite mampu mengenali alfabet Bahasa Isyarat Indonesia (BISINDO) secara real-time pada perangkat Android dengan performa yang baik pada kondisi pengujian terkontrol.

Dengan demikian, kontribusi penelitian ini tidak hanya terletak pada implementasi model CNN ringan berbasis MobileNetV2 untuk pengenalan alfabet BISINDO pada perangkat Android, tetapi juga pada penyediaan kerangka evaluasi yang menempatkan hasil eksperimen dalam horizon ontologis, epistemologis, dan aksiologis. Di sisi lain, penelitian ini masih memiliki keterbatasan, antara lain ruang lingkup yang terbatas pada alfabet statis A–Z, belum adanya pengujian robustness yang lebih luas pada

kondisi penggunaan nyata, serta belum lengkapnya pelaporan beberapa parameter pelatihan dan karakteristik akuisisi dataset. Oleh karena itu, penelitian lanjutan perlu diarahkan pada pengenalan gestur dinamis, integrasi ekspresi non-manual, penguatan dokumentasi eksperimen, dan pelibatan komunitas pengguna secara lebih partisipatif (Costanza-Chock, 2020; Mittelstadt et al., 2016; Sandler et al., 2018; TensorFlow, 2023).

REFERENSI

- Huroniyah, F., & Martita, L. (2023). Implementasi Metode Bahasa Isyarat Indonesia (BISINDO) untuk Meningkatkan Kepercayaan Diri Anak Tunarungu di SLBN Tompokersan Lumajang. *Indonesian Journal of Disability Research*, 1(2), 123–128.
- Budiman, S. N., Lestanti, S., Yuana, H., & Awwalin, B. N. (2023). SIBI berbasis Machine Learning dan Computer Vision untuk Membantu Komunikasi Tuna Rungu dan Tuna Wicara. *Jurnal Teknologi dan Manajemen Informatika*, 9(2), 119–128. <https://doi.org/10.26905/jtmi.v9i2.10993>
- Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., & Chen, L.-C. (2018). MobileNetV2: Inverted residuals and linear bottlenecks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 4510–4520).
- TensorFlow. (2023). TensorFlow Lite (Documentation). Retrieved December 13, 2025, from <https://www.tensorflow.org/lite>
- Popper, K. R. (1959). *The logic of scientific discovery*. Routledge.
- Kuhn, T. S. (1996). *The structure of scientific revolutions* (3rd ed.). University of Chicago Press.
- Lakatos, I. (1970). Falsification and the methodology of scientific research programmes. In I. Lakatos & A. Musgrave (Eds.), *Criticism and the growth of knowledge* (pp. 91–196). Cambridge University Press.
- van Fraassen, B. C. (1980). *The scientific image*. Clarendon Press.
- Hempel, C. G. (1965). *Aspects of scientific explanation and other essays in the philosophy of science*. Free Press.

- Floridi, L. (2020). AI and its new winter: From myths to realities. *Philosophy & Technology*, 33, 1–3.
- Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S., & Floridi, L. (2016). The ethics of algorithms: Mapping the debate. *Big Data & Society*, 3(2).
- Selbst, A. D., Boyd, D., Friedler, S., Venkatasubramanian, S., & Vertesi, J. (2019). Fairness and abstraction in sociotechnical systems. In *Proceedings of the ACM Conference on Fairness, Accountability, and Transparency (FAT*)* (pp. 59–68). ACM.
- Costanza-Chock, S. (2020). *Design justice: Community-led practices to build the worlds we need*. MIT Press.
- Google. (2024). CameraX overview (Android Developers documentation). Retrieved December 13, 2025, from <https://developer.android.com/training/camera>
- Chollet, F. (2021). *Deep learning with Python* (2nd ed.). Manning.
- Pressman, R. S., & Maxim, B. R. (2019). *Software engineering: A practitioner's approach* (9th ed.). McGraw-Hill.
- Sholawati, M., Auliasari, K., & Ariwibisono, F. (2022). Pengembangan aplikasi pengenalan bahasa isyarat abjad SIBI menggunakan CNN. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 6(1), 134–144. <https://doi.org/10.36040/jati.v6i1.4507>
- Aufar, Y., & Kaloka, T. P. (2022). Robusta coffee leaf diseases detection based on MobileNetV2 model. *International Journal of Electrical and Computer Engineering*, 12(6), 6675–6683. <https://doi.org/10.11591/ijece.v12i6.pp6675-6683>
- Wachter, S., Mittelstadt, B., & Floridi, L. (2017). Transparent, explainable, and accountable AI for robotics. *Science Robotics*, 2(6).
- Crawford, K. (2021). *Atlas of AI: Power, politics, and the planetary costs of artificial intelligence*. Yale University Press.
- Ridhovan, A., & Suharso, A. (2022). Metode deteksi pokok pohon secara otomatis pada citra perkebunan sawit menggunakan model Convolutional Neural Network (CNN) pada perangkat lunak sistem informasi geografis. *Seminar Nasional Teknologi dan Rekayasa Informasi (SENTRIN) 2022*, 9(7), Special Issue.
- Gracita, M. (2024). Klobility—BISINDO dan SIBI: Apa bedanya? Retrieved from <https://id.pinterest.com/gracitra/bisindo/> (Klobility).