

**SISTEM PENERJEMAH BAHASA ISYARAT INDONESIA
(BISINDO) REAL-TIME BERBASIS WEB MENGGUNAKAN
METODE DEEP LEARNING CNN-LSTM DAN
TEXT-TO-SPEECH**

SKRIPSI

DISUSUN OLEH

NABIEL MUHAMMAD IMJAUZANANSYAH

NPM. 2209010141



UMSU

Unggul | Cerdas | Terpercaya

**PROGRAM STUDI SISTEM INFORMASI
FAKULTAS ILMU KOMPUTER DAN TEKNOLOGI INFORMASI
UNIVERSITAS MUHAMMADIYAH SUMATERA UTARA**

MEDAN

2026

Sistem Penerjemah Bahasa Isyarat Indonesia (BISINDO) Real-Time Berbasis Web Menggunakan Metode Deep Learning CNN-LSTM dan Text-to-Speech.

SKRIPSI

Diajukan sebagai salah satu syarat untuk memperoleh gelar Sarjana Komputer (S.Kom) dalam Program Studi Sistem Informasi pada Fakultas Ilmu Komputer dan Teknologi Informasi, Universitas Muhammadiyah Sumatera Utara

NABIEL MUHAMMAD IMJAUZANANSYAH

NPM 2209010141

**PROGRAM STUDI SISTEM INFORMASI
FAKULTAS ILMU KOMPUTER DAN TEKNOLOGI INFORMASI
UNIVERSITAS MUHAMMADIYAH SUMATERA UTARA**

MEDAN

2026

LEMBAR PENGESAHAN

Judul Skripsi : Sistem Penerjemah Bahasa Isyarat Indonesia (BISINDO)
Real-Time Berbasis Web Menggunakan Metode Deep
Learning CNN-LSTM dan Text-to-Speech
Nama Mahasiswa : NABIEL MUHAMMAD IMJAUZANANSYAH
NPM : 2209010141
Program Studi : SISTEM INFORMASI

Menyetujui
Komisi Pembimbing



(Hevlie Winda Nazry Simbolon, S.Pd., M.Si)
NIDN. 0129079301

Ketua Program Studi



(Mahardika Abdi Prawira Tanjung, S.Kom, M.Kom)
NIDN. 0117088902

Dekan



(A. Iskandar, S.Kom., M.Kom.)
NIDN. 0127099201

PERNYATAAN ORISINALITAS

**Sistem Penerjemah Bahasa Isyarat Indonesia (BISINDO) Real-Time
Berbasis Web Menggunakan Metode Deep Learning CNN-LSTM dan Text-
to-Speech**

SKRIPSI

Saya menyatakan bahwa karya tulis ini adalah hasil karya sendiri, kecuali beberapa kutipan dan ringkasan yang masing-masing disebutkan sumbernya.

Medan, 11 April 2026

Yang membuat pernyataan



Nabiel Muhammad Imjauzanansyah

NPM. 2209010141

PERNYATAAN PERSETUJUAN PUBLIKASI
KARYA ILMIAH UNTUK KEPENTINGAN
AKADEMIS

Sebagai sivitas akademika Universitas Muhammadiyah Sumatera Utara, saya bertanda tangan dibawah ini:

Nama : Nabel Muhammad Imjauzanansyah
NPM : 2209010141
Program Studi : Sistem Informasi
Karya Ilmiah : Skripsi

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Muhammadiyah Sumatera Utara Hak Bedas Royalti Non-Eksekutif (*Non-Exclusive Royalty free Right*) atas penelitian skripsi saya yang berjudul:

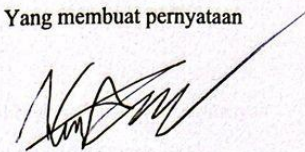
**Sistem Penerjemah Bahasa Isyarat Indonesia (BISINDO) Real-Time
Berbasis Web Menggunakan Metode Deep Learning CNN-LSTM dan Text-
to-Speech**

Beserta perangkat yang ada (jika diperlukan). Dengan Hak Bebas Royalti Non-Eksekutif ini, Universitas Muhammadiyah Sumatera Utara berhak menyimpan, mengalih media, memformat, mengelola dalam bentuk database, merawat dan mempublikasikan Skripsi saya ini tanpa meminta izin dari saya selama tetap mencantumkan nama saya sebagai penulis dan sebagai pemegang dan atau sebagai pemilik hak cipta.

Demikian pernyataan ini dibuat dengan sebenarnya.

Medan, 11 April 2026

Yang membuat pernyataan



Nabel Muhammad Imjauzanansyah

NPM. 2209010141

RIWAYAT HIDUP

DATA PRIBADI

Nama Lengkap : Nabel Muhammad Imjauzanansyah
Tempat dan Tanggal Lahir : Meulaboh / 14 January 2004
Alamat Rumah : Jl T Hasan Dek Lr Gelatik No 32
Telepon/Faks/HP : 085211479415
E-mail : nabel.zmuhammad11@gmail.com
Instansi Tempat Kerja : -
Alamat Kantor : -

DATA PENDIDIKAN

SD : MIN Teladan Banda Aceh (MIN 7) TAMAT: 2016
SMP : SMPN 2 Banda Aceh TAMAT: 2019
SMA : SMAN 3 Banda Aceh TAMAT: 2022

KATA PENGANTAR



Assalamu 'alaikum warahmatullahi wabarakatuh

Alhamdulillahirabbil'alamin. Puji syukur penulis panjatkan ke hadirat Allah SWT atas segala rahmat, karunia, dan hidayah-Nya, sehingga penulis dapat menyelesaikan tugas akhir ini dengan baik dan tepat waktu.

Shalawat serta salam semoga senantiasa tercurah kepada Nabi Muhammad SAW, beserta keluarga, sahabat, dan seluruh umatnya hingga akhir zaman.

Tugas akhir ini berjudul " SISTEM PENERJEMAH BAHASA ISYARAT INDONESIA (BISINDO) REAL-TIME BERBASIS WEB MENGGUNAKAN METODE DEEP LEARNING CNN-LSTM DAN TEXT-TO-SPEECH " yang disusun sebagai salah satu syarat untuk menyelesaikan pendidikan Strata Satu (S1) pada Program Studi Sistem Informasi, Fakultas Ilmu Komputer dan Teknologi Informasi, Universitas Muhammadiyah Sumatera Utara. Penulis telah berupaya semaksimal mungkin dalam menyusun skripsi ini, namun masih terdapat keterbatasan baik dari segi isi maupun penyajian.

Dalam proses penyusunan skripsi ini, penulis juga menyadari bahwa banyak pihak yang telah memberikan bantuan, bimbingan, serta dukungan.

Dengan segala kerendahan hati, penulis mengucapkan terima kasih yang sebesar-besarnya kepada:

1. Bapak Prof. Dr. Akrim, M.Pd. Rektor Universitas Muhammadiyah Sumatera Utara (UMSU)
2. Bapak Dr. Al-Khowarizmi, S.Kom., M.Kom. Dekan Fakultas Ilmu Komputer dan Teknologi Informasi (FIKTI) UMSU.
3. Ibu Dr. Firahmi Rizky, S.Kom., M.Kom. selaku Wakil Dekan I Bidang Akademik FIKTI UMSU.
4. Bapak Mhd. Basri, S.Si., M.Kom. selaku Wakil Dekan III Bidang Kemahasiswaan FIKTI UMSU.
5. Bapak Mahardika Abdi Prawira Tanjung, S.Kom., M.Kom. selaku Ketua Program Studi Sistem Informasi FIKTI UMSU.
6. Bapak Mulkan Azhari, S.Kom., M.Kom. selaku Sekretaris Program Studi Sistem Informasi FIKTI UMSU.
7. Ibu Hevlie Winda Nazry Simbolon, S.Pd., M.Si., selaku Dosen Pembimbing yang dengan penuh kesabaran telah meluangkan waktu, tenaga, dan pikiran untuk memberikan bimbingan, arahan, serta masukan yang sangat berharga dari awal hingga selesainya skripsi ini.
8. Kepada Mama dan Ayah, penulis mengucapkan terima kasih yang sebesar-besarnya kepada Mama dan Ayah tercinta atas segala doa, dukungan, kasih sayang, serta pengorbanan yang tiada henti. Terima kasih atas setiap perjuangan, kesabaran, dan kepercayaan yang selalu diberikan kepada penulis dalam setiap langkah kehidupan, khususnya selama proses penyusunan tugas akhir ini. Dalam setiap kesulitan yang dihadapi, doa dan dukungan dari Mama dan Ayah menjadi kekuatan terbesar bagi penulis untuk tetap bertahan dan terus berusaha hingga mampu menyelesaikan tugas ini dengan baik. Penulis menyadari bahwa tanpa doa dan dukungan dari Mama dan Ayah, pencapaian ini tidak akan terwujud. Semoga segala kebaikan dan pengorbanan yang telah diberikan senantiasa dibalas dengan keberkahan dan kebahagiaan oleh Tuhan Yang Maha Esa.

9. Kepada kakak tercinta, Nabila AUFARAIHAN, dan adik tersayang, Sultan JOHANSYAH, terima kasih atas segala dukungan, doa, dan kehangatan persaudaraan yang tak pernah putus. Terima kasih atas setiap canda tawa yang menjadi pelipur lara di saat penulis merasa lelah dan penat dalam mengerjakan tugas akhir ini. Kehadiran kalian senantiasa memberikan warna, motivasi, dan keceriaan tersendiri yang membuat penulis kembali bersemangat untuk menyelesaikan tanggung jawab ini.
10. Kepada seluruh teman teman seperjuangan , yaitu Said, Noval, Adet, Fajri, dan Hafiz yang telah menjadi bagian tak terpisahkan dalam perjalanan akademik penulis. Terima kasih telah menjadi rekan tumbuh sejak awal masa perkuliahan, melewati setiap semester dengan segala suka dan dukanya, hingga akhirnya kita sampai di titik penyelesaian skripsi ini bersama-sama. Terima kasih atas solidaritas, diskusi yang tak kenal waktu, semangat yang saling menguatkan di kala jenuh, serta persahabatan tulus yang membuat perjuangan ini terasa lebih ringan.
11. Kepada seluruh teman teman seperjuangan semasa SMA yang tergabung dalam grup "Pejuang Danantara" dan "P A K D E 🍷", terima kasih atas persahabatan yang tetap erat terjaga hingga saat ini. Kehadiran, dukungan moril, kekompakan, serta canda tawa dari kalian selalu menjadi penyegar dan penyemangat tersendiri bagi penulis, dari bangku sekolah hingga di tahap penyelesaian tugas akhir ini.
12. Semua pihak yang tidak dapat disebutkan satu per satu, yang telah memberikan bantuan baik secara langsung maupun tidak langsung dalam penyelesaian skripsi ini

13. Last but not least, anak kedua, Nabel Muhammad Imjauzanansyah. Ya, diri saya sendiri. Apresiasi sebesar-besarnya untuk diri sendiri yang telah diam-diam berjuang tanpa henti. Terima kasih telah bekerja keras dan bertahan sejauh ini. Untuk setiap malam yang dihabiskan dalam kelelahan, setiap pagi yang disambut dengan kekhawatiran, namun tetap dijalani dan berhasil dilalui. Terima kasih kepada hati yang tetap ikhlas, meski banyak hal yang terjadi di luar prediksi. Terima kasih kepada jiwa yang tetap kuat, meski berkali-kali hampir menyerah karena kondisi dan tekanan penyusunan skripsi. Terima kasih kepada raga yang terus melangkah, meski lelah menatap baris kode tapi tetap teguh diperjuangkan. Teruslah belajar dan mensyukuri nikmat yang Tuhan berikan. Tetap semangat untuk terus berusaha, berbahagialah di mana pun kamu berada. Penulis berdoa agar langkah kecilmu selalu diperkuat, dikelilingi orang-orang baik dan hebat, serta mimpimu satu per satu akan terjawab. Aamiin.

**SISTEM PENERJEMAH BAHASA ISYARAT INDONESIA (BISINDO)
REAL-TIME BERBASIS WEB MENGGUNAKAN METODE DEEP
LEARNING CNN-LSTM DAN
TEXT-TO-SPEECH**

ABSTRAK

Bahasa Isyarat Indonesia (BISINDO) merupakan media komunikasi utama bagi penyandang disabilitas rungu, namun rendahnya pemahaman masyarakat umum sering memicu hambatan komunikasi. Penelitian terkait pengenalan bahasa isyarat sebelumnya sebagian besar masih berjalan secara *offline*, belum terintegrasi di lingkungan *web real-time*, dan hanya menghasilkan keluaran berupa teks. Oleh karena itu, penelitian ini bertujuan merancang dan membangun sistem penerjemah BISINDO berbasis *web real-time* menggunakan metode *hybrid Convolutional Neural Network (CNN)* dan *Long Short-Term Memory (LSTM)*, serta mengintegrasikannya dengan fitur *Text-to-Speech (TTS)*. Dataset yang digunakan berupa data video primer dari 3 orang peraga, mencakup 11 kelas kategori (10 kata isyarat dan 1 kelas diam) dengan total 1.000 *frame* per kelas berformat *grayscale* (100x89 piksel). Arsitektur model membagi tugas komputasi di mana CNN mengekstraksi fitur spasial statis dan LSTM memodelkan kesinambungan temporal dari sekuens 30 *frame*. Model ini kemudian diintegrasikan ke dalam antarmuka aplikasi berbasis React.js dan Node.js. Hasil pengujian menunjukkan model hibrida ini mencapai tingkat akurasi statis sebesar 96% berdasarkan evaluasi *Confusion Matrix*. Pada pengujian fungsional *real-time*, penerapan aturan *Confidence Threshold* sebesar 80% terbukti efektif menyaring gerakan tidak tepat, sehingga sistem mampu menerjemahkan gerakan isyarat yang valid menjadi teks

dan luaran suara secara akurat, meskipun masih terdapat misklasifikasi minor pada gestur dengan kemiripan spasial.

Kata Kunci: BISINDO; Deep Learning; CNN-LSTM; Text-to-Speech; Web Real-Time.

**REAL-TIME WEB-BASED INDONESIAN SIGN LANGUAGE (BISINDO)
TRANSLATOR SYSTEM USING CNN-LSTM DEEP LEARNING AND
TEXT-TO-SPEECH**

ABSTRACT

Indonesian Sign Language (BISINDO) is the primary communication medium for the deaf community, yet low public understanding often causes communication barriers . Previous sign language recognition studies mostly operated offline, lacked real-time web integration, and only produced text output . Therefore, this study aims to design and develop a real-time web-based BISINDO translator system using a hybrid Convolutional Neural Network (CNN) and Long Short-Term Memory (LSTM) method, integrated with a Text-to-Speech (TTS) feature . The dataset consists of primary video data from 3 different subjects, covering 11 category classes (10 sign words and 1 idle class) with a total of 1,000 frames per class in grayscale format (100x89 pixels) . The model architecture divides computational tasks where CNN extracts static spatial features and LSTM models the temporal continuity of 30-frame sequences . This model is then integrated into an application interface based on React.js and Node.js . The test results demonstrate that the hybrid model achieved a static accuracy rate of 96% based on Confusion Matrix evaluation . In real-time functional testing, the implementation of an 80% Confidence Threshold rule proved effective in filtering out incorrect gestures, enabling the system to accurately translate valid sign gestures into text and voice output, despite minor misclassifications on gestures with spatial similarities.

Keywords: *BISINDO; Deep Learning; CNN-LSTM; Text-to-Speech; Real-Time Web.*

DAFTAR ISI

LEMBAR PENGESAHAN	Error! Bookmark not defined.
PERNYATAAN ORISINALITAS	Error! Bookmark not defined.
PERNYATAAN PERSETUJUAN PUBLIKASI	i
RIWAYAT HIDUP	iv
KATA PENGANTAR	v
ABSTRAK	ix
ABSTRACT	xi
DAFTAR ISI	xiii
DAFTAR TABEL	xvi
DAFTAR GAMBAR	xvii
BAB I PENDAHULUAN	1
1.1 Latar Belakang Masalah	1
1.2 Rumusan Masalah	4
1.3 Batasan Masalah.....	4
1.4 Tujuan Penelitian.....	5
1.5 Manfaat Penelitian.....	6
BAB II LANDASAN TEORI	7
2.1 Bahasa Isyarat Indonesia (BISINDO)	7
2.2 Deep Learning	7
2.3 Convolutional Neural Network (CNN)	8

2.4 Long Short-Term Memory (LSTM).....	9
2.5 Model Hybrid CNN-LSTM.....	10
2.6 Text-to-Speech (TTS).....	11
2.7 Pengembangan Sistem Berbasis Web	12
2.7.1 React.js.....	12
2.7.2 Node.js	13
2.8 Tinjauan Penelitian Terdahulu dan Kebaruan Penelitian	13
BAB III METODOLOGI PENELITIAN	16
3.1 Jenis dan Pendekatan Penelitian.....	16
3.2 Dataset dan Teknik Pengumpulan Data	16
3.3 Metode Penelitian Sistem.....	18
3.4 Perangkat dan Alat Penelitian	20
3.4.1 Perangkat Keras (Hardware).....	20
3.4.2 Perangkat Lunak (Software)	21
3.5 Arsitektur Sistem NzSignify	22
3.5.1 Perancangan Model Deep Learning (CNN-LSTM).....	22
3.5.2 Flowchart Sistem	25
3.5.3 Use Case Diagram	26
3.5.4 Activity Diagram	26
3.5.5 Mockup Antarmuka Aplikasi Web.....	28
3.6 Implementasi Sistem	29

3.7 Teknik Analisis Data dan Pengujian Sistem	29
3.7.1 Pengujian Probabilitas Prediksi	31
3.8 Jadwal Penelitian	31
BAB IV IMPLEMENTASI DAN PENGUJIAN SISTEM	34
4.1 Analisis Kebutuhan Sistem	34
4.2 Perancangan Prototype Sistem	35
4.2.1 Dataset & Pra-Pemrosesan.....	35
4.2.2 Arsitektur Model.....	36
4.3 Pengembangan dan Pelatihan Model CNN-LSTM	37
4.4 Implementasi Sistem Berbasis Web	38
4.5 Integrasi Modul <i>Text-to-Speech</i> (TTS).....	40
4.6 Evaluasi dan Penyempurnaan Prototype	41
4.7 Analisis Performa dan Kekuatan Model CNN-LSTM.....	45
4.7.1 Analisis Keluaran Probabilitas.....	45
4.7.2 Penanganan Gerakan Tidak Tepat	45
4.7.3 Evaluasi Misklasifikasi Minor pada Sistem Web	46
BAB V PENUTUP.....	47
5.1 Kesimpulan.....	47
5.2 Saran	48
DAFTAR PUSTAKA	50

DAFTAR TABEL

Table 2. 1 Perbandingan Sistem	14
Table 3. 1 Jadwal Penelitian	32
Table 4. 1 Distribusi Dataset Mentah BISINDO	35
Table 4.2 Hasil Classification Report.....	43

DAFTAR GAMBAR

Gambar 3. 1 Alur Sistem NzSignify.....	22
Gambar 3. 2 Rancangan Model	23
Gambar 3. 3 Flowchart sistem.....	25
Gambar 3. 4 Use Case Diagram	26
Gambar 3. 5 Activity Diagram	27
Gambar 3. 6 Halaman Utama Aplikasi NzSignify	28
Gambar 3. 7 Halaman Penerjemahan Bahasa Isyarat Indonesia (BISINDO)	28
Gambar 4.1 Potongan Kode Pra-Pemrosesan Dinamis	36
Gambar 4.2 Potongan Kode Implementasi Arsitektur CNN-LSTM	36
Gambar 4.3 Grafik Akurasi dan Loss Training	38
Gambar 4.4 Proses Permintaan Izin Akses Perangkat Kamera pada Antarmuka Web	39
Gambar 4.5 Screenshot Halaman Utama/Dashboard Web NzSignify	39
Gambar 4.6 Halaman Penerjemahan Web NzSignify	40
Gambar 4.7 Potongan Kode Eksekusi Suara	40
Gambar 4.8 Confusion Matrix dari hasil evaluasi model.....	41
Gambar 4.9 Bukti Keberhasilan Pengujian Gerakan "Halo" secara Instan (Live Log: 100%)	44
Gambar 4.10 Temuan Ambiguitas Model pada Gerakan Transisi (Live Log: 37.9%)	45

BAB I

PENDAHULUAN

1.1 Latar Belakang Masalah

Bahasa isyarat merupakan bentuk komunikasi visual yang digunakan oleh penyandang disabilitas rungu dalam menyampaikan pesan melalui gerakan tangan, ekspresi wajah, dan posisi tubuh (World Health Organization, 2025). Di Indonesia, Bahasa Isyarat Indonesia (BISINDO) menjadi salah satu sistem bahasa isyarat yang digunakan dalam kehidupan sehari-hari. Namun demikian, tingkat pemahaman masyarakat terhadap BISINDO masih rendah, sehingga menimbulkan hambatan komunikasi antara penyandang disabilitas rungu dan masyarakat umum. Kondisi ini sering memunculkan ketidaknyamanan dan keterbatasan akses, terutama dalam bidang pelayanan publik, pendidikan, dan interaksi sosial.

Berbagai penelitian terdahulu menunjukkan bahwa pendekatan deep learning memiliki kinerja yang signifikan dalam pengenalan bahasa isyarat berbasis citra dan video. Model *Convolutional Neural Network* (CNN) terbukti efektif dalam mengekstraksi fitur spasial dari gerakan tangan dan bentuk isyarat, sementara *Long Short-Term Memory* (LSTM) unggul dalam memodelkan hubungan temporal antarframe pada data sekuens. Kombinasi CNN–LSTM banyak digunakan dalam penelitian pengenalan bahasa isyarat karena mampu mengintegrasikan informasi spasial dan temporal secara komprehensif, sehingga menghasilkan tingkat akurasi yang lebih tinggi dibandingkan penggunaan model tunggal (Aljabar, 2020).

Meskipun hasil penelitian sebelumnya menunjukkan performa yang baik dalam pengenalan gesture bahasa isyarat, sebagian besar sistem yang dikembangkan masih bersifat terbatas pada pengujian offline dan belum terintegrasi dalam sistem penerjemahan real-time yang aplikatif. Selain itu, keluaran sistem umumnya masih berupa teks atau label kelas, tanpa konversi langsung ke bentuk suara (Aljabar, 2020). Oleh karena itu, diperlukan pengembangan lebih lanjut berupa sistem penerjemah Bahasa Isyarat Indonesia (BISINDO) berbasis web yang mampu bekerja secara real-time dengan integrasi *text-to-speech*, sehingga dapat menjembatani kesenjangan komunikasi antara penyandang disabilitas rungu dan masyarakat umum secara lebih efektif dan inklusif.

Secara spesifik, metode CNN–LSTM dalam penelitian ini dirancang untuk memproses data video bahasa isyarat secara real-time, di mana CNN berperan mengekstraksi fitur spasial pada setiap frame video, sementara LSTM digunakan untuk mempelajari hubungan temporal antarframe dalam satu rangkaian gerakan. Pendekatan ini memungkinkan sistem mengenali makna bahasa isyarat berdasarkan konteks urutan gerakan secara menyeluruh, bukan hanya dari satu frame tunggal. Hasil klasifikasi dari model CNN–LSTM kemudian diterjemahkan ke dalam bentuk teks dan selanjutnya dikonversi menjadi keluaran suara menggunakan modul Text-to-Speech (TTS), sehingga sistem tidak hanya berfokus pada akurasi pengenalan gestur, tetapi juga pada kebermanfaatan praktis dalam mendukung komunikasi secara langsung dan inklusif.

Di sisi lain, perkembangan teknologi dalam bidang kecerdasan buatan *Artificial Intelligence* telah menghadirkan peluang untuk menciptakan solusi inovatif dalam mengatasi hambatan komunikasi tersebut (Russell & Norvig, 2021). Metode *deep learning* seperti *Convolutional Neural Network* (CNN) mampu mengekstraksi fitur visual dari gambar atau video secara efektif, sementara *Long Short-Term Memory* (LSTM) unggul dalam memahami pola urutan gerakan tangan pada video. Kombinasi CNN-LSTM memungkinkan sistem menganalisis rangkaian frame video untuk mengenali dan mengklasifikasikan gerakan BISINDO secara lebih akurat (Altairika & Sari, 2023).

Seiring dengan perkembangan tersebut, kemajuan teknologi *web development* turut mendukung implementasi sistem pengenalan gerakan secara real-time. React JS banyak digunakan dalam pengembangan antarmuka web modern karena kemampuannya membangun tampilan yang responsif dan dinamis berbasis komponen, sedangkan Node JS mendukung pengolahan *back-end* secara efisien melalui arsitektur *event-driven* dan *non-blocking I/O*. Integrasi React JS dan Node JS memungkinkan pengembangan aplikasi web real-time yang interaktif, ringan, serta mudah diakses oleh pengguna tanpa instalasi tambahan (Axza et al., 2023; Ramadhani et al., 2025).

Selain itu, implementasi sistem secara real-time berbasis web menjadikannya mudah diakses tanpa instalasi khusus dan dapat berjalan pada berbagai perangkat. Fitur *Text-to-Speech* (TTS) memberikan keluaran suara dari hasil terjemahan, sehingga meningkatkan kebermanfaatan sistem dalam interaksi langsung.

Berdasarkan urgensi permasalahan komunikasi yang dihadapi penyandang disabilitas rungu serta peluang pemanfaatan teknologi *deep learning* dan aplikasi web *real-time*, penelitian ini difokuskan pada pengembangan “**Sistem Penerjemah Bahasa Isyarat Indonesia (BISINDO) Real-Time Berbasis Web Menggunakan Metode Deep Learning CNN-LSTM dan Text-to-Speech**” sebagai solusi teknologi untuk menjembatani kesenjangan komunikasi, sehingga mampu meningkatkan aksesibilitas dan inklusivitas dalam interaksi sehari-hari.

1.2 Rumusan Masalah

Berdasarkan uraian latar belakang yang telah di sampaikan, maka permasalahan dalam penelitian ini dapat di rumuskan sebagai berikut:

1. Bagaimana menerapkan metode CNN-LSTM untuk mendeteksi dan mengklasifikasikan gerakan bahasa isyarat Indonesia?
2. Bagaimana tingkat akurasi model CNN-LSTM dalam mengklasifikasikan gerakan bahasa isyarat berdasarkan dataset yang digunakan?
3. Bagaimana mengintegrasikan fitur Text-to-Speech pada hasil terjemahan agar dapat digunakan secara langsung oleh pengguna?
4. Bagaimana merancang dan membangun sistem penerjemah bahasa isyarat berbasis web yang mampu bekerja secara *real-time* menggunakan antarmuka React.js dan backend Node.js?

1.3 Batasan Masalah

Untuk menjaga fokus dan ruang lingkup penelitian ini, penelitian ini dibatasi pada hal-hal berikut ini:

1. Penelitian ini hanya membahas penerjemahan Bahasa Isyarat Indonesia (BISINDO) dan tidak mencakup bahasa isyarat lain seperti SIBI atau ASL, dengan dataset yang digunakan terbatas pada video BISINDO sesuai dengan kebutuhan penelitian.
2. Sistem yang dikembangkan melakukan penerjemahan bahasa isyarat pada level kata dengan memanfaatkan data video sekuensial sebagai masukan.
3. Metode pengenalan gerakan bahasa isyarat yang digunakan adalah arsitektur deep learning hybrid CNN–LSTM.
4. Implementasi sistem dilakukan dalam bentuk aplikasi web real-time dengan menggunakan React JS sebagai frontend dan Node JS sebagai backend.
5. Keluaran sistem dibatasi pada hasil terjemahan dalam bentuk teks dan suara melalui fitur Text-to-Speech (TTS).

1.4 Tujuan Penelitian

Berdasarkan Tujuan Penelitian:

1. Menerapkan metode CNN-LSTM dalam mendeteksi serta mengklasifikasikan gerakan Bahasa Isyarat Indonesia (BISINDO) berdasarkan data video.
2. Menganalisis kinerja dan tingkat akurasi model CNN-LSTM dalam mengklasifikasikan gerakan Bahasa Isyarat Indonesia
3. Mengintegrasikan fitur *Text-to-Speech* (TTS) pada hasil terjemahan bahasa isyarat agar output sistem dapat disampaikan dalam bentuk suara secara langsung.

4. Merancang dan membangun sistem penerjemah Bahasa Isyarat Indonesia (BISINDO) berbasis web yang mampu bekerja secara real-time dengan memanfaatkan antarmuka React JS dan backend Node JS.

1.5 Manfaat Penelitian

Penelitian ini diharapkan memberikan manfaat dalam pengembangan solusi teknologi berbasis kecerdasan buatan untuk mendukung penerjemahan Bahasa Isyarat Indonesia (BISINDO) secara real-time. Hasil penelitian ini dapat berkontribusi dalam peningkatan aksesibilitas komunikasi melalui pemanfaatan metode deep learning CNN-LSTM dan integrasi fitur Text-to-Speech, sehingga membantu mengurangi hambatan interaksi antara penyandang disabilitas rungu dan masyarakat umum serta mendorong terciptanya lingkungan komunikasi yang lebih inklusif.

BAB II

LANDASAN TEORI

2.1 Bahasa Isyarat Indonesia (BISINDO)

Bahasa Isyarat Indonesia (BISINDO) merupakan sistem komunikasi visual yang digunakan oleh komunitas Tuli di Indonesia dalam kegiatan sehari-hari. BISINDO disampaikan melalui kombinasi gerakan tangan, ekspresi wajah, dan gerakan tubuh sebagai media utama untuk menyampaikan pesan. Bahasa ini berkembang secara alami di komunitas Tuli dan digunakan dalam berbagai konteks sosial dan pendidikan (Yardi et al., 2023).

Penelitian terkait komunikasi di kelas menunjukkan bahwa BISINDO lebih mudah dipahami oleh siswa Tuli dibandingkan dengan SIBI, karena struktur BISINDO lebih komunikatif dan mengikuti pola alami bahasa visual (Isnaniah et al., 2023). Selain itu, BISINDO memiliki variasi regional yang berbeda antar wilayah, mencerminkan dinamika budaya lokal dari komunitas Tuli di Indonesia.

Karakteristik visual dan temporal BISINDO yang berbasis gerakan menjadikannya sangat relevan untuk diproses menggunakan teknologi pengolahan video dan model deep learning berbasis sekuens.

2.2 Deep Learning

Deep Learning merupakan bagian dari *machine learning* yang menggunakan jaringan saraf tiruan dengan banyak lapisan untuk mempelajari pola kompleks dari data. Pendekatan ini memungkinkan sistem untuk mengekstraksi fitur secara otomatis tanpa memerlukan perancangan fitur manual. Deep learning telah banyak

digunakan dalam pengolahan citra, suara, dan bahasa alami karena kemampuannya dalam menangani data berdimensi tinggi.

Dalam konteks pengolahan data visual, deep learning mampu mempelajari karakteristik spasial dan temporal secara bersamaan. Hal ini menjadikan deep learning sangat sesuai untuk aplikasi yang melibatkan video, termasuk pengenalan gerakan dan aktivitas manusia. Keunggulan utama deep learning terletak pada kemampuannya dalam memodelkan hubungan non-linear yang kompleks, sehingga menghasilkan performa yang lebih baik dibandingkan metode konvensional.

Perkembangan deep learning dalam beberapa tahun terakhir semakin pesat seiring dengan peningkatan daya komputasi dan ketersediaan data dalam jumlah besar. Model deep learning modern juga banyak diintegrasikan dengan aplikasi berbasis web untuk menghasilkan sistem cerdas yang dapat diakses secara real-time (Alzahrani & Alshamrani, 2024; Mienye & Swart, 2024).

2.3 Convolutional Neural Network (CNN)

Convolutional Neural Network (CNN) merupakan arsitektur jaringan saraf yang dirancang untuk memproses data berbentuk *grid*, seperti citra dan video. CNN digunakan untuk mengekstraksi fitur spasial dari setiap *frame* video bahasa isyarat yang menjadi input sistem.

Dalam proses kerjanya, CNN menerima frame video satu per satu sebagai masukan. Setiap frame diproses melalui lapisan konvolusi, di mana *kernel* digeser pada citra untuk mendeteksi fitur visual dasar seperti tepi, bentuk tangan, dan pola gerakan. Hasil konvolusi ini menghasilkan *feature map* yang kemudian dilewatkan ke fungsi aktivasi untuk mempertahankan informasi non-linear.

Selanjutnya, *feature map* diproses oleh pooling layer untuk mereduksi dimensi data dan meningkatkan efisiensi komputasi tanpa menghilangkan informasi penting. Fitur hasil pooling kemudian diratakan (*flatten*) menjadi vektor satu dimensi sebagai representasi visual dari setiap frame. Representasi ini selanjutnya diteruskan ke tahap pemodelan temporal menggunakan LSTM.

Dalam sistem penerjemahan bahasa isyarat, CNN berperan sebagai pengekstraksi fitur visual utama pada setiap frame video sebelum informasi tersebut diproses secara sekuensial. Kemampuan CNN dalam mengekstraksi fitur visual secara otomatis menjadikannya komponen penting dalam pengolahan citra dan video berbasis *deep learning* (Khan et al., 2021).

2.4 Long Short-Term Memory (LSTM)

Long Short-Term Memory (LSTM) merupakan pengembangan dari *Recurrent Neural Network* (RNN) yang dirancang untuk mengatasi permasalahan hilangnya gradien pada pemrosesan data sekuens. LSTM memiliki mekanisme memori yang memungkinkan jaringan menyimpan informasi jangka panjang melalui struktur *gate*, yaitu *input gate*, *forget gate*, dan *output gate*.

Dalam proses kerjanya, LSTM menerima urutan fitur dari setiap frame video secara bertahap. Pada setiap langkah waktu, *input gate* menentukan informasi baru yang akan dimasukkan ke dalam memori sel, *forget gate* mengatur informasi lama yang perlu dipertahankan atau dihapus, sedangkan *output gate* menghasilkan keluaran berdasarkan kondisi memori terkini. Mekanisme ini memungkinkan LSTM mempertahankan informasi penting dari frame sebelumnya saat memproses frame berikutnya.

Keunggulan LSTM terletak pada kemampuannya dalam mempelajari ketergantungan temporal antar data secara berurutan. Dalam penelitian ini, LSTM memproses urutan fitur hasil ekstraksi CNN untuk memahami hubungan antar gerakan tangan dari waktu ke waktu, sehingga pola gerakan bahasa isyarat dapat dikenali secara lebih akurat (Alzubi et al., 2024; Udurume et al., 2024).

2.5 Model Hybrid CNN-LSTM

Model hybrid CNN–LSTM merupakan arsitektur *deep learning* yang menggabungkan kemampuan *Convolutional Neural Network* (CNN) dalam mengekstraksi fitur spasial dan *Long Short-Term Memory* (LSTM) dalam memodelkan ketergantungan temporal dari data video. Pendekatan ini dirancang untuk menangkap informasi visual dan dinamika gerakan secara simultan, sehingga sesuai untuk pemrosesan video bahasa isyarat yang bersifat dinamis (Myagila & Nyambo, 2025).

Dalam pemodelannya, setiap video bahasa isyarat terlebih dahulu dipecah menjadi sejumlah frame berurutan. Setiap frame diproses oleh CNN sebagai *feature extractor* untuk menghasilkan representasi fitur visual, seperti bentuk tangan dan orientasi gerakan. Fitur dari seluruh frame kemudian disusun menjadi sebuah sekuens fitur yang merepresentasikan urutan gerakan dalam satu video. *Sekuens* fitur ini selanjutnya menjadi masukan bagi LSTM untuk mempelajari hubungan antarframe dari waktu ke waktu sehingga konteks gerakan dapat dipahami secara menyeluruh.

LSTM memanfaatkan mekanisme memori jangka panjang untuk mempertahankan informasi penting dari gerakan sebelumnya saat memproses gerakan berikutnya. Dengan cara ini, model CNN–LSTM tidak hanya memahami

informasi statis pada setiap frame, tetapi juga mampu menangkap dinamika temporal gerakan bahasa isyarat yang kompleks. Pendekatan hybrid ini terbukti lebih unggul dibandingkan penggunaan CNN atau LSTM secara terpisah dalam pengenalan gestur berbasis video (Kumari & Anand, 2024).

Beberapa penelitian menunjukkan bahwa model CNN–LSTM tetap relevan dan efektif dalam tugas pengenalan gestur karena kemampuan LSTM dalam menangani konteks urutan panjang dan integrasinya dengan CNN dalam ekstraksi fitur spasial. Hal ini mendukung pemilihan model hybrid CNN–LSTM dalam pengembangan sistem penerjemah Bahasa Isyarat Indonesia berbasis web yang membutuhkan keseimbangan antara akurasi pengenalan dan efisiensi komputasi (Myagila & Nyambo, 2025).

2.6 Text-to-Speech (TTS)

Text-to-Speech (TTS) merupakan teknologi yang digunakan untuk mengubah teks menjadi keluaran suara. Sistem TTS memungkinkan komputer menyampaikan informasi dalam bentuk audio yang dapat dipahami oleh pengguna. TTS umumnya terdiri dari tahap pemrosesan teks dan tahap sintesis suara.

Dalam proses kerjanya, teks hasil penerjemahan bahasa isyarat terlebih dahulu melalui tahap pemrosesan teks, yang mencakup normalisasi teks dan penyesuaian struktur kalimat agar sesuai untuk diucapkan. Selanjutnya, teks yang telah diproses dikonversi menjadi representasi fonetik atau unit suara yang dapat dipahami oleh mesin sintesis. Pada tahap sintesis suara, sistem menghasilkan sinyal audio berdasarkan representasi tersebut sehingga suara yang dihasilkan dapat terdengar jelas dan natural.

Dalam aplikasi penerjemahan bahasa isyarat, TTS berperan sebagai media keluaran yang menyampaikan hasil terjemahan teks kepada pengguna non-Tuli secara real-time. Integrasi TTS dengan sistem pemrosesan bahasa memungkinkan keluaran suara yang lebih terstruktur dan mudah dipahami, sehingga meningkatkan aksesibilitas dan efektivitas komunikasi antara pengguna bahasa isyarat dan masyarakat umum (Ext et al., 2020; Tan et al., 2023).

2.7 Pengembangan Sistem Berbasis Web

2.7.1 React.js

React.js merupakan pustaka (*library*) JavaScript bersifat *open-source* yang dikembangkan oleh Facebook untuk membangun antarmuka pengguna (*User Interface*), khususnya untuk aplikasi satu halaman (*Single Page Application*). React memungkinkan pengembang membangun antarmuka dalam bentuk komponen yang dapat digunakan kembali (*reusable components*) serta mengelola *state* aplikasi secara efisien (Axza et al., 2023).

Dalam proses penerapannya pada sistem ini, React.js digunakan untuk membangun halaman antarmuka utama, mengelola interaksi pengguna, dan mengontrol alur data antarkomponen. Aplikasi web dijalankan sebagai *Single Page Application*, di mana React memuat tampilan awal dan selanjutnya melakukan pembaruan antarmuka secara dinamis berdasarkan perubahan *state*.

Keunggulan utama React.js terletak pada penggunaan *Virtual DOM* (*Document Object Model*). Pada saat terjadi perubahan data, React membandingkan *Virtual DOM* dengan *DOM* sebelumnya dan hanya memperbarui bagian antarmuka yang mengalami perubahan. Mekanisme ini memungkinkan sistem menampilkan stream video dari kamera pengguna, memproses permintaan

penerjemahan, serta menampilkan hasil terjemahan teks dan keluaran suara secara real-time tanpa melakukan pemuatan ulang halaman, sehingga meningkatkan responsivitas dan efisiensi aplikasi (Ramadhan & Imanda, 2025).

2.7.2 Node.js

Node.js adalah lingkungan eksekusi (*runtime environment*) untuk JavaScript yang dibangun di atas mesin JavaScript V8 milik Chrome. Node.js memungkinkan JavaScript dijalankan di sisi server (*server-side*), yang sebelumnya hanya dapat berjalan di sisi klien (*browser*) (Kurniawan & Rozi, 2020).

Dalam proses kerjanya pada sistem ini, Node.js berperan sebagai lapisan backend yang menerima permintaan (*request*) dari antarmuka React.js melalui *Application Programming Interface* (API). Permintaan tersebut meliputi pengiriman data video atau hasil ekstraksi frame yang dikirim dari sisi klien untuk diproses lebih lanjut.

Node.js kemudian mengelola alur komunikasi dengan modul pemrosesan *Deep Learning* yang berjalan pada lingkungan Python. Setelah menerima hasil prediksi dari model CNN-LSTM, Node.js meneruskan respons tersebut kembali ke antarmuka React.js dalam bentuk data terstruktur. Seluruh proses komunikasi ini dijalankan menggunakan mekanisme event-driven dan non-blocking I/O, sehingga sistem mampu menangani pertukaran data secara real-time dengan efisien dan responsif (Ramadhani et al., 2025).

2.8 Tinjauan Penelitian Terdahulu dan Kebaruan Penelitian

Untuk memastikan kebaruan (*novelty*) dan posisi penelitian ini dibandingkan dengan pengembangan sebelumnya, dilakukan analisis komparasi terhadap

beberapa penelitian terkait. Berikut adalah ringkasan perbandingan antara penelitian terdahulu dengan sistem NzSignify yang dikembangkan dalam skripsi ini:

Table 2. 1 Perbandingan Sistem

No	Peneliti & Tahun	Metode / Fokus Penelitian	Kelemahan / Keterbatasan (Gap)	Kebaruan pada Skripsi Ini
1.	Altiarika & Sari (2023)	Deteksi <i>real-time</i> BISINDO menggunakan metode CNN dan LSTM (Altiarika & Sari, 2023).	Fokus hanya pada akurasi deteksi dan output sebatas label teks, belum dikemas menjadi aplikasi interaktif.	Mengintegrasikan model ke dalam Aplikasi Web interaktif dan menambahkan fitur Text-to-Speech (TTS) sebagai output suara.
2.	Deleviar, Oktaviani, & Permatasari (2025)	Website penerjemah Speech-to-Video BISINDO berbasis LSTM (Informatika et al., 2025)	Arah penerjemahan terbalik (Suara ke Video isyarat) yang ditujukan untuk orang dengar belajar isyarat.	Fokus pada arah sebaliknya (Video Isyarat ke Suara) untuk membantu Teman Tuli berkomunikasi secara lisan.
3.	Fadillah, Irawan, & Susanty (2022)	Penerjemah BISINDO dengan pendekatan Transfer Learning (dari model ASL) (Fadillah et al., 2022).	Akurasi rendah (30%) karena memaksakan model bahasa isyarat asing (ASL) pada karakteristik gestur BISINDO.	Membangun dan melatih model CNN-LSTM dari awal (<i>scratch</i>) menggunakan dataset video BISINDO mandiri untuk akurasi lebih tinggi.

Berdasarkan tabel di atas, sistem NzSignify yang diusulkan memiliki keunggulan kompetitif berupa integrasi Aplikasi Web Real-Time yang mudah diakses tanpa perangkat khusus (hanya webcam), penggunaan metode Hybrid CNN-LSTM yang dilatih khusus dengan dataset BISINDO, serta fitur Text-to-Speech yang belum banyak diimplementasikan pada penelitian sejenis sebelumnya.

BAB III

METODOLOGI PENELITIAN

3.1 Jenis dan Pendekatan Penelitian

Penelitian ini merupakan penelitian kuantitatif dengan pendekatan eksperimen, karena data yang digunakan dan dianalisis berupa data numerik untuk mengukur kinerja model *deep learning* secara objektif melalui metrik evaluasi seperti akurasi, presisi, recall, dan nilai *loss* yang direpresentasikan menggunakan *confusion matrix* (Helmiyah et al., 2025; Ovalle & Tecnológica, 2023). Pendekatan eksperimen diterapkan melalui serangkaian proses perancangan, pelatihan, dan pengujian model CNN–LSTM serta penyesuaian parameter untuk mengetahui pengaruhnya terhadap kemampuan sistem dalam mengenali gerakan Bahasa Isyarat Indonesia (Creswell, 2021). Penelitian ini juga bersifat aplikatif, karena hasil pengujian metode diimplementasikan ke dalam sebuah sistem penerjemah bahasa isyarat berbasis web yang dapat digunakan secara real-time oleh pengguna.

3.2 Dataset dan Teknik Pengumpulan Data

Dataset yang digunakan dalam penelitian ini merupakan data primer berupa citra digital sekuensial yang merepresentasikan gerakan Bahasa Isyarat Indonesia (BISINDO). Data dikumpulkan secara mandiri melalui proses perekaman kontinu menggunakan kamera web (webcam) untuk menangkap rangkaian gerakan secara utuh dari awal hingga akhir isyarat. Adapun karakteristik dataset dan teknik pengumpulan data dijelaskan sebagai berikut:

1. Jenis dan Sumber Dataset

Dataset berupa data primer hasil perekaman kontinu (*continuous capturing*) yang melibatkan 3 (tiga) orang peraga yang berbeda. Penggunaan 3 orang peraga ini bertujuan untuk mendapatkan variasi karakteristik gerakan tangan, postur tubuh, dan ritme gerakan, sehingga model dapat mengenali bahasa isyarat secara lebih umum dan adaptif.

2. Kelas Dataset

Dataset terdiri dari total 11 kelas, yang meliputi 10 kelas kata isyarat dan 1 kelas idle (diam). Kelas idle digunakan untuk merepresentasikan kondisi tanpa gerakan tangan yang signifikan guna mengurangi kesalahan prediksi (*false positive*). Sepuluh kelas kata isyarat yang digunakan meliputi: Halo, Nama, Saya, Nabel, Terima Kasih, Maaf, Software Engineering, Makan, Minum, dan Tidur.

3. Jumlah Data dan Format Citra

Setiap kelas isyarat direkam hingga menghasilkan sekitar 1.000 frame citra, sehingga mampu merepresentasikan variasi gerakan secara memadai. Seluruh citra diproses menjadi format grayscale dengan ukuran 100×89 piksel untuk menjaga konsistensi input pada model CNN-LSTM serta mengurangi kompleksitas komputasi.

4. Teknik dan Perangkat Perekaman

Proses perekaman data dilakukan menggunakan kamera web (webcam) yang terintegrasi pada perangkat komputer. Akuisisi citra dilakukan dengan bantuan platform berbasis web Teachable Machine sebagai alat bantu perekaman, yang

memungkinkan pengambilan data secara otomatis dan kontinu selama peraga melakukan gerakan isyarat secara natural.

5. Kondisi Lingkungan Perekaman

Pengambilan data dilakukan di dalam ruangan (*indoor*) dengan pencahayaan buatan yang cukup dan merata untuk meminimalkan noise pada citra. Objek direkam menggunakan latar belakang polos dengan warna kontras terhadap warna kulit guna memudahkan proses pemisahan objek tangan dari latar belakang pada tahap pra-pemrosesan.

6. Pelabelan Data (Labeling)

Pelabelan data dilakukan menggunakan metode pelabelan berbasis direktori, di mana setiap kelas isyarat disimpan dalam folder terpisah sesuai dengan nama kelasnya. Pada tahap pra-pemrosesan, sistem membaca nama folder sebagai label kategori, kemudian mengonversinya ke dalam format numerik menggunakan teknik One-Hot Encoding agar dapat diproses oleh algoritma *deep learning*.

7. Pembagian Dataset

Seluruh dataset yang telah melalui proses pra-pemrosesan dan pelabelan dibagi menjadi tiga bagian, yaitu 80% data latih (training set), 10% data validasi (validation set), dan 10% data uji (testing set). Pembagian ini bertujuan untuk melatih model, memantau kinerja selama pelatihan, serta mengukur akurasi akhir model secara objektif.

3.3 Metode Penelitian Sistem

Metode pengembangan sistem yang digunakan dalam penelitian ini adalah Model Prototype. Model Prototype memungkinkan pengembangan sistem dilakukan secara bertahap melalui pembuatan purwarupa awal yang dapat diuji,

dievaluasi, dan disempurnakan secara berulang. Pemilihan model ini dinilai sesuai karena sistem NzSignify bersifat interaktif, berbasis web, serta melibatkan integrasi teknologi deep learning CNN–LSTM dan Text-to-Speech, sehingga memerlukan proses evaluasi dan penyempurnaan secara bertahap agar sistem yang dihasilkan optimal dan sesuai dengan kebutuhan pengguna (Irawan & Mu, 2024)

1. Analisis Kebutuhan Sistem

Tahap awal dilakukan dengan menganalisis kebutuhan sistem berdasarkan permasalahan komunikasi antara penyandang disabilitas rungu dan pengguna lainnya. Kebutuhan sistem meliputi kemampuan menangkap video bahasa isyarat secara real-time melalui kamera, melakukan klasifikasi gerakan BISINDO menggunakan model CNN–LSTM, serta menghasilkan keluaran berupa teks dan suara melalui fitur Text-to-Speech pada aplikasi berbasis web.

2. Perancangan Prototype Sistem

Tahap ini mencakup perancangan arsitektur sistem, alur kerja aplikasi, dan rancangan antarmuka pengguna. Perancangan disesuaikan dengan flowchart sistem, mulai dari proses pengguna mengakses aplikasi web, pemberian izin kamera, hingga proses penerjemahan bahasa isyarat. Prototype awal digunakan sebagai gambaran fungsional sistem sebelum dilakukan implementasi penuh.

3. Pengembangan dan Pelatihan Model CNN–LSTM

Pada tahap ini dilakukan pengembangan model deep learning menggunakan arsitektur CNN–LSTM. CNN digunakan untuk mengekstraksi fitur spasial dari setiap frame video bahasa isyarat, sedangkan LSTM digunakan untuk mempelajari hubungan temporal antar frame. Model dilatih menggunakan dataset video BISINDO yang telah melalui proses preprocessing dan pelabelan.

4. Implementasi Sistem Berbasis Web

Model CNN–LSTM yang telah dilatih kemudian diintegrasikan ke dalam aplikasi berbasis web. Aplikasi berfungsi untuk menangkap video dari kamera pengguna, memproses data menggunakan model deep learning, serta menampilkan hasil klasifikasi dalam bentuk teks secara real-time sesuai dengan alur kerja sistem.

5. Integrasi Modul Text-to-Speech (TTS)

Hasil klasifikasi berupa teks selanjutnya diproses oleh modul Text-to-Speech untuk menghasilkan keluaran suara. Integrasi TTS dilakukan setelah proses klasifikasi selesai sehingga hasil terjemahan dapat langsung disampaikan kepada pengguna dalam bentuk audio.

6. Evaluasi dan Penyempurnaan Prototype

Tahap akhir dilakukan evaluasi terhadap kinerja model CNN–LSTM dan fungsional sistem secara keseluruhan. Evaluasi dilakukan menggunakan metrik seperti akurasi dan confusion matrix, serta pengujian sistem secara fungsional. Hasil evaluasi digunakan sebagai dasar penyempurnaan prototype hingga sistem NzSignify mencapai kinerja yang optimal.

3.4 Perangkat dan Alat Penelitian

3.4.1 Perangkat Keras (Hardware)

Perangkat keras yang digunakan dalam penelitian ini adalah sebagai berikut:

1. Laptop/Komputer

Digunakan sebagai perangkat utama untuk perekaman dataset, preprocessing

data, pelatihan model CNN–LSTM, serta pengujian sistem. Spesifikasi prosesor Ryzen 5 6600H, RAM 16 GB, dan GPU mendukung proses komputasi intensif pada pengembangan model deep learning.

2. Kamera

Digunakan sebagai perangkat input untuk menangkap video bahasa isyarat BISINDO baik pada tahap pengumpulan dataset maupun implementasi sistem penerjemah secara real-time.

3. Perangkat Audio

Digunakan sebagai media keluaran suara hasil konversi teks melalui fitur Text-to-Speech untuk mendukung komunikasi pengguna.

3.4.2 Perangkat Lunak (Software)

Perangkat lunak yang digunakan dalam penelitian ini dikelompokkan menjadi dua kategori utama, yaitu untuk pengembangan model kecerdasan buatan dan pengembangan sistem aplikasi web:

A. Pengembangan Model *Deep Learning* (CNN-LSTM)

Pengembangan model dilakukan menggunakan Google Colaboratory sebagai lingkungan komputasi berbasis cloud dengan dukungan GPU untuk mempercepat proses pelatihan. Bahasa pemrograman Python digunakan sebagai bahasa utama dengan dukungan library TensorFlow dan Keras untuk membangun arsitektur CNN–LSTM. Proses perekaman awal data dibantu menggunakan Teachable Machine, sedangkan OpenCV dan NumPy digunakan untuk preprocessing citra dan manipulasi data numerik.

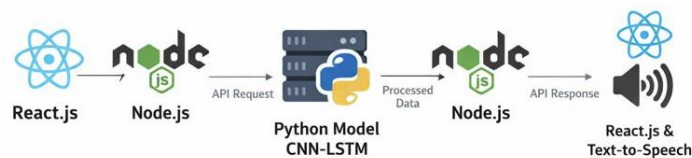
B. Pengembangan Sistem Web

Pengembangan aplikasi web dilakukan menggunakan Visual Studio Code sebagai code editor, React.js sebagai framework frontend untuk membangun antarmuka dan menangkap video dari webcam, serta Node.js sebagai backend untuk mengelola komunikasi data antara aplikasi web dan model deep learning. Sistem dijalankan dan diuji menggunakan web browser, serta dilengkapi modul Text-to-Speech untuk mengonversi hasil teks terjemahan menjadi suara. dan menguji sistem penerjemah secara *real-time*.

3.5 Arsitektur Sistem NzSignify

Sistem *NzSignify* dirancang menggunakan arsitektur terpisah antara aplikasi web dan model deep learning. Alur arsitektur sistem adalah sebagai berikut:

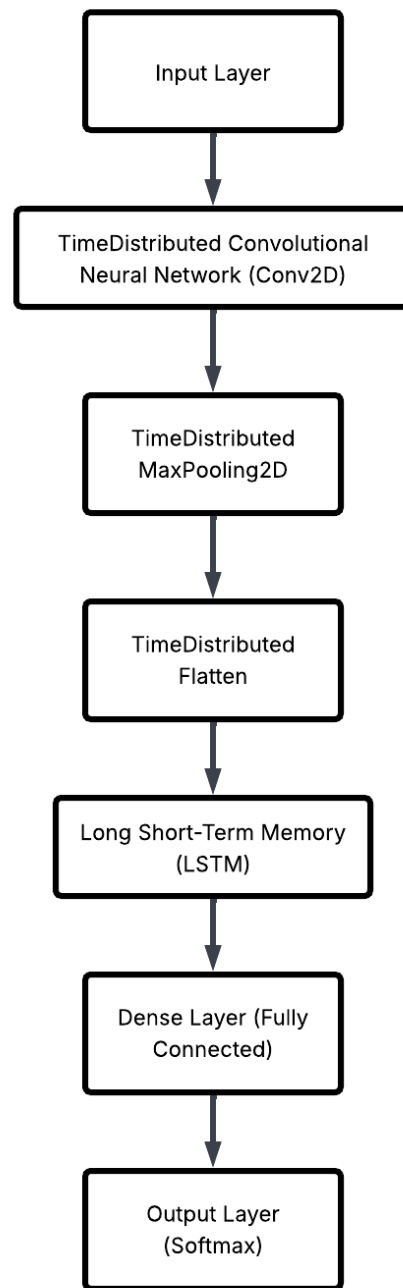
Arsitektur ini memungkinkan pengelolaan model dan aplikasi web secara terpisah sehingga meningkatkan skalabilitas dan kemudahan pengembangan sistem.



Gambar 3. 1 Alur Sistem NzSignify

3.5.1 Perancangan Model Deep Learning (CNN-LSTM)

Model kecerdasan buatan yang dibangun menggunakan arsitektur *Hybrid* yang menggabungkan *Convolutional Neural Network* (CNN) untuk ekstraksi fitur spasial dan *Long Short-Term Memory* (LSTM) untuk ekstraksi fitur temporal.



Gambar 3. 2 Rancangan Model

Berikut adalah rancangan lapisan (*layer*) model yang digunakan:

keterangan:

a. Input Layer

Input berupa sekumpulan frame video bahasa isyarat dengan dimensi *width*, *height*, dan *channels* yang membentuk data sekuens.

b. TimeDistributed Convolutional Neural Network (Conv2D)

Lapisan ini berfungsi mengekstraksi fitur spasial dari setiap frame video secara independen, seperti bentuk tangan dan pola visual.

c. TimeDistributed MaxPooling2D

Digunakan untuk mereduksi dimensi fitur hasil konvolusi agar proses komputasi menjadi lebih efisien tanpa menghilangkan informasi penting.

d. TimeDistributed Flatten

Mengubah hasil ekstraksi fitur berbentuk matriks menjadi vektor satu dimensi sebagai input bagi lapisan sekuensial.

e. Long Short-Term Memory (LSTM)

LSTM digunakan untuk mempelajari hubungan temporal antar frame sehingga sistem dapat memahami urutan gerakan bahasa isyarat.

f. Dense Layer (Fully Connected)

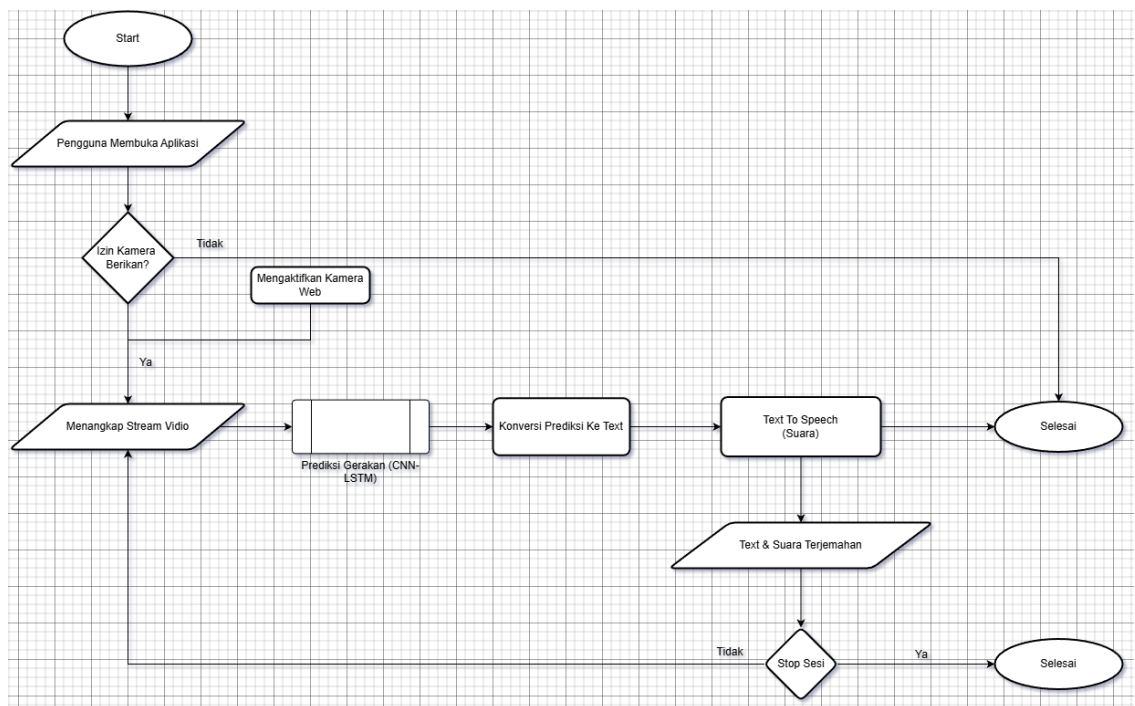
Lapisan ini melakukan proses klasifikasi berdasarkan fitur spasial dan temporal yang telah dipelajari.

g. Output Layer (Softmax)

Lapisan keluaran menggunakan fungsi aktivasi Softmax dengan 11 neuron untuk menghasilkan probabilitas masing-masing kelas bahasa isyarat.

3.5.2 Flowchart Sistem

Flowchart sistem NzSignify menggambarkan alur kerja aplikasi penerjemah Bahasa Isyarat Indonesia (BISINDO) berbasis web secara real-time. Proses dimulai saat pengguna membuka aplikasi dan memberikan izin akses kamera. Sistem kemudian menangkap video bahasa isyarat dan memprosesnya menggunakan model CNN-LSTM untuk menghasilkan prediksi kelas gerakan. Hasil prediksi dikonversi menjadi teks dan selanjutnya diubah menjadi suara melalui modul Text-to-Speech (TTS). Apabila izin kamera tidak diberikan, proses dihentikan dan sistem tidak melakukan penerjemahan.

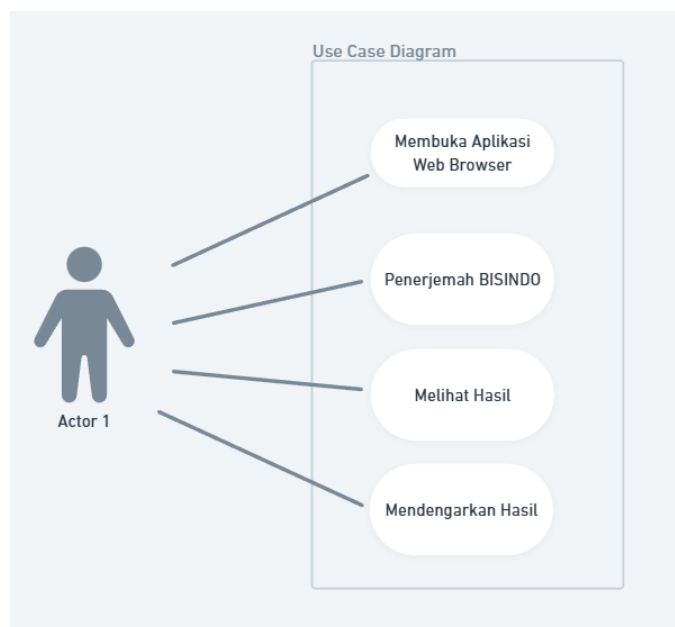


Gambar 3. 3 Flowchart sistem

Flowchart sistem NzSignify

3.5.3 Use Case Diagram

Use Case Diagram sistem NzSignify menggambarkan interaksi antara pengguna sebagai aktor utama dengan sistem penerjemah BISINDO berbasis web. Pengguna dapat membuka aplikasi, mengaktifkan kamera, dan menjalankan proses penerjemahan bahasa isyarat. Sistem memproses video menggunakan model CNN–LSTM untuk menghasilkan keluaran berupa teks yang selanjutnya dapat dikonversi menjadi suara melalui fitur Text-to-Speech secara



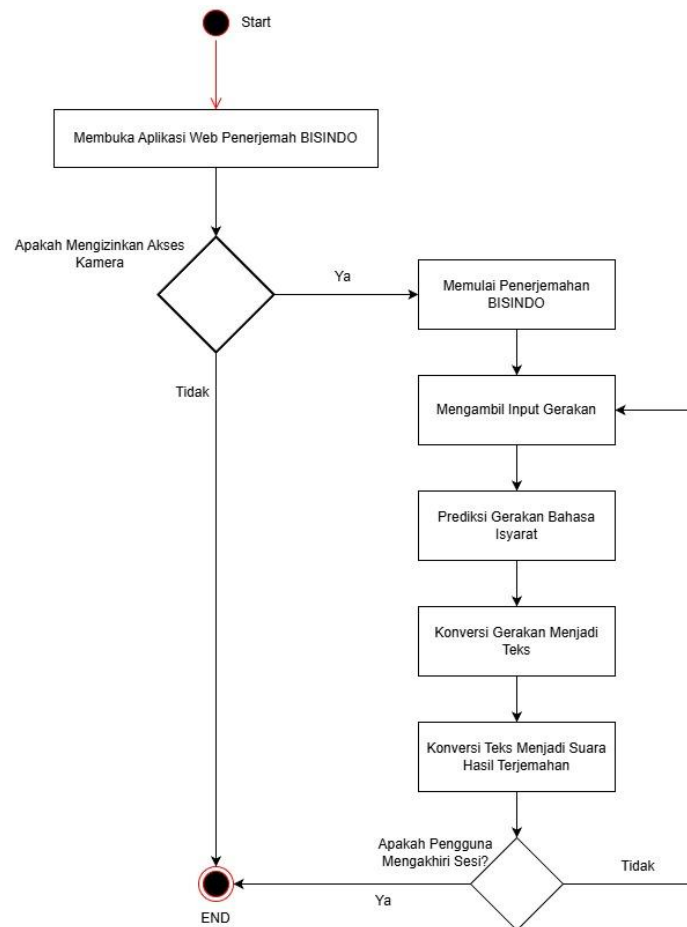
Gambar 3. 4 Use Case Diagram

real-time.

3.5.4 Activity Diagram

Activity Diagram digunakan untuk menggambarkan alur aktivitas dan interaksi antara pengguna dan sistem dalam aplikasi penerjemah Bahasa Isyarat Indonesia (BISINDO) berbasis web. Diagram ini menunjukkan urutan aktivitas yang terjadi mulai dari pengguna membuka aplikasi, sistem meminta izin dan

mengaktifkan kamera, hingga proses penerjemahan bahasa isyarat menjadi teks

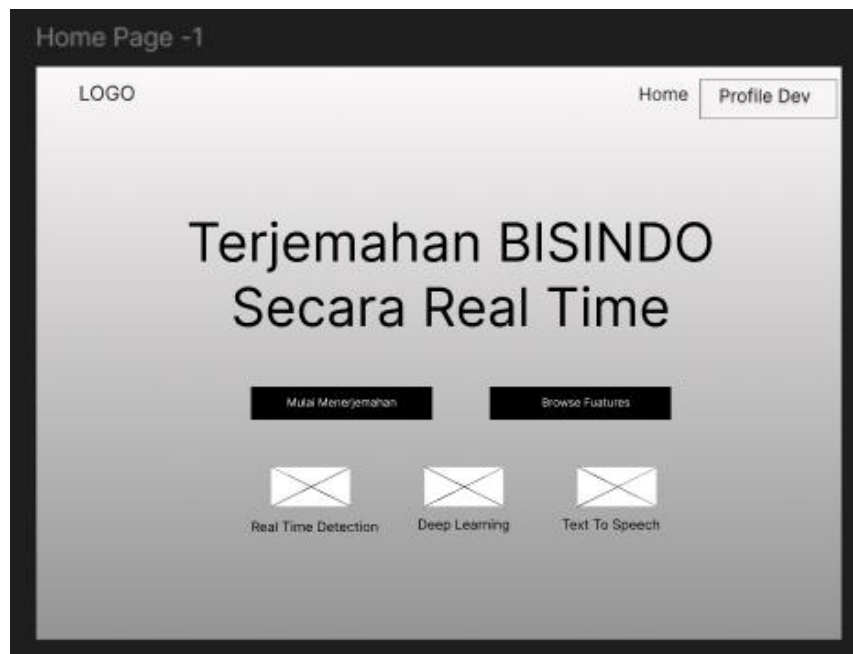


Gambar 3.5 Activity Diagram

dan suara. Activity Diagram berfokus pada perilaku sistem (system behavior) dan aliran kontrol aktivitas berdasarkan keputusan pengguna dan respons sistem.

3.5.5 Mockup Antarmuka Aplikasi Web

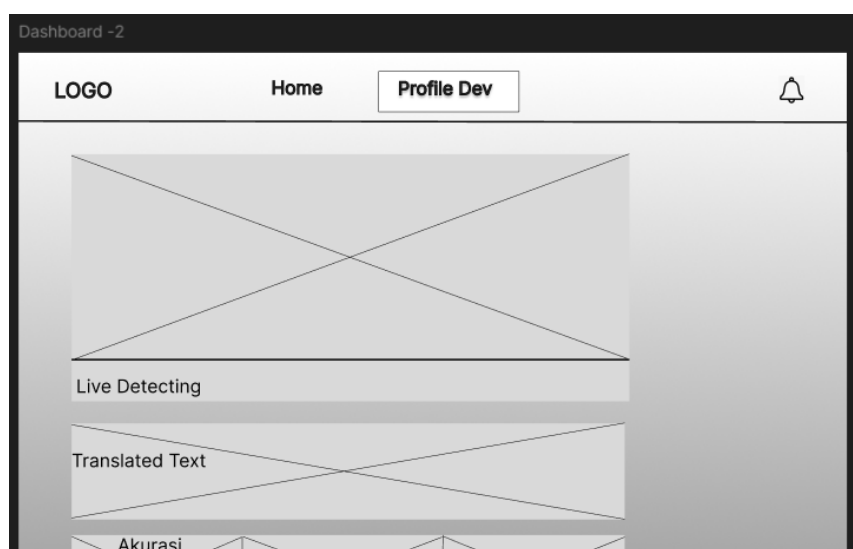
Gambar ini menampilkan rancangan halaman utama aplikasi NzSignify yang berfungsi sebagai pintu masuk pengguna sebelum memulai proses penerjemahan



Gambar 3. 6 Halaman Utama Aplikasi NzSignify

Bahasa Isyarat Indonesia (BISINDO) secara real-time.

Gambar ini menunjukkan rancangan antarmuka halaman penerjemahan yang menampilkan proses deteksi bahasa isyarat secara real-time menggunakan model



Gambar 3. 7 Halaman Penerjemahan Bahasa Isyarat Indonesia (BISINDO)

CNN-LSTM, hasil terjemahan dalam bentuk teks, serta keluaran suara melalui fitur Text-to-Speech.

3.6 Implementasi Sistem

Implementasi sistem dilakukan dengan mengintegrasikan model CNN-LSTM ke dalam aplikasi web. Frontend dikembangkan menggunakan React.js, backend menggunakan Node.js, dan model deep learning dijalankan pada server Python terpisah.

3.7 Teknik Analisis Data dan Pengujian Sistem

Teknik analisis data pada penelitian ini bertujuan untuk mengevaluasi kinerja model CNN-LSTM dalam mengenali dan menerjemahkan Bahasa Isyarat Indonesia (BISINDO). Evaluasi dilakukan secara kuantitatif dengan membandingkan hasil prediksi model terhadap label aktual pada data uji yang tidak dilibatkan dalam proses pelatihan.

Pengujian kinerja model menggunakan Confusion Matrix sebagai dasar perhitungan metrik evaluasi, meliputi akurasi, presisi, recall, dan F1-score, serta analisis nilai loss untuk menilai proses pembelajaran model selama pelatihan. Confusion Matrix memvisualisasikan performa klasifikasi pada 11 kelas bahasa isyarat berdasarkan nilai True Positive (TP), True Negative (TN), False Positive (FP), dan False Negative (FN).

a. Akurasi

digunakan untuk menunjukkan tingkat keberhasilan model dalam mengklasifikasikan data uji secara keseluruhan berdasarkan perbandingan jumlah prediksi benar terhadap total data uji.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \times 100\% \dots\dots\dots(1)$$

b. Precision dan Recall

Presisi mengukur ketepatan model dalam memprediksi suatu kelas isyarat, sedangkan recall mengukur kemampuan model dalam mengenali seluruh data yang termasuk dalam kelas tersebut.

$$Precision = \frac{TP}{TP+FP} \dots\dots\dots (2)$$

$$Recall = \frac{TP}{TP+FN} \dots\dots\dots (3)$$

c. F1-Score

F1-Score merupakan rata-rata harmonis antara presisi dan recall yang digunakan untuk memberikan evaluasi kinerja model secara seimbang pada klasifikasi multi-kelas.

$$F1-Score = \frac{2 \times Precision \times Recall}{Precision + Recall} \dots\dots\dots (4)$$

d. Support

Support merepresentasikan jumlah kemunculan data aktual (sampel uji sesungguhnya) untuk masing-masing kelas dalam dataset pengujian. Metrik ini berfungsi sebagai indikator struktural untuk mengonfirmasi distribusi data yang dievaluasi, sehingga memastikan bahwa perhitungan performa model (seperti akurasi dan presisi) didasarkan pada proporsi sampel uji yang valid dan representatif.

e. Loss Function

Penelitian ini menggunakan Categorical Crossentropy sebagai fungsi kerugian karena bersifat klasifikasi multi-kelas dengan 11 kelas. Nilai loss digunakan untuk mengukur selisih antara probabilitas prediksi model dan label sebenarnya selama proses pelatihan. Nilai loss yang semakin kecil

menunjukkan bahwa model semakin optimal dalam mempelajari pola data bahasa isyarat.

3.7.1 Pengujian Probabilitas Prediksi

Selain menggunakan metrik Confusion Matrix untuk pengujian statis, pengujian sistem juga dilakukan dengan mengukur tingkat keyakinan model (Confidence Score) saat memproses data sekuensial. Nilai ini didapatkan dari distribusi probabilitas yang dihasilkan oleh fungsi aktivasi Softmax pada Output Layer.

Untuk menangani kondisi di mana pengguna melakukan gerakan yang tidak tepat, mengambang, atau sedang dalam fase transisi antar isyarat, sistem menerapkan mekanisme *Confidence Threshold* (Ambang Batas Keyakinan). Mekanisme ini dievaluasi dengan aturan matematis berikut:

- a. Jika probabilitas kelas prediksi $\geq 80\%$, maka sistem akan menganggap gerakan valid dan meneruskannya menjadi keluaran teks serta suara (*Text-to-Speech*).
- b. Jika probabilitas kelas prediksi $< 80\%$, maka sistem mengklasifikasikan gerakan tersebut sebagai "tidak tepat" atau ambigu. Sistem tidak akan memicu keluaran teks maupun suara hingga pengguna memperbaiki pola gerakannya.

3.8 Jadwal Penelitian

Penelitian ini dilaksanakan pada bulan Desember 2025 hingga April 2026. Seluruh kegiatan penelitian dilakukan secara mandiri di kediaman peneliti yang berlokasi di Medan. Mengingat penelitian ini berbasis pengembangan perangkat lunak dan eksperimen komputasi, pelaksanaan penelitian tidak memerlukan lokasi

laboratorium khusus dan difokuskan pada pengembangan sistem, pelatihan model, serta evaluasi kinerja sistem (Sugiyono, 2021).

Table 3. 1 Jadwal Penelitian

No	Kegiatan Penelitian	Des 2025	Jan 2026	Feb 2026	Mar 2026	Apr 2026	Jun 2026
1.	Studi Literatur dan Pencarian Referensi	✓	✓				
2.	Penyusunan Proposal (Bab I, II, dan III)	✓	✓				
3.	Bimbingan dan Revisi Proposal	✓	✓				
4.	Pendaftaran dan Administrasi Sempro		✓				
5.	Pelaksanaan Seminar Proposal		✓				
6.	Pengumpulan Dataset & Pembuatan <i>Prototype</i>			✓			
7.	Pelatihan Model & Integrasi Sistem Web			✓			
8.	Pelaksanaan Exhibition / Demo Sistem Awal				✓		
9.	Pengujian Sistem & Penyusunan Laporan Skripsi			✓	✓		
10	Sidang Skripsi & Penyelesaian Revisi Akhir					✓	

BAB IV

IMPLEMENTASI DAN PENGUJIAN SISTEM

4.1 Analisis Kebutuhan Sistem

Langkah pertama dalam merealisasikan sistem penerjemah Bahasa Isyarat Indonesia (BISINDO) adalah mendefinisikan dan mempersiapkan spesifikasi infrastruktur penunjang, baik dari segi perangkat keras (*hardware*) maupun perangkat lunak (*software*).

a. Infrastruktur Perangkat Keras

Mengingat proses pelatihan model *deep learning* membutuhkan daya komputasi yang masif, penelitian ini memanfaatkan layanan komputasi awan (*cloud*) Google Colaboratory. Sementara itu, proses implementasi aplikasi web dan pengujian fungsional secara *real-time* dieksekusi menggunakan perangkat komputer/laptop berspesifikasi prosesor AMD Ryzen 5 6600H yang didukung oleh RAM berkapasitas 16 GB, serta webcam standar sebagai media penangkap visual.

b. Infrastruktur Perangkat Lunak

Algoritma kecerdasan buatan dibangun dan dilatih menggunakan bahasa pemrograman Python 3, dengan memanfaatkan library mutakhir seperti TensorFlow dan Keras. Untuk menunjang sisi antarmuka, aplikasi web dikembangkan dengan framework React.js, yang selanjutnya dihubungkan ke backend Node.js untuk menangani pertukaran data antara frontend, modul AI, dan fitur *Text-to-Speech* (TTS)

4.2 Perancangan Prototype Sistem

Pada fase ini, rancangan konseptual yang sebelumnya telah dijabarkan pada Bab III mulai direalisasikan ke dalam bentuk purwarupa awal. Fokus utama perancangan adalah membangun alur interaksi antarmuka yang logis, mulai dari otorisasi akses kamera oleh pengguna, penangkapan frame video secara dinamis, hingga tampilan area penerjemahan gerakan isyarat. Purwarupa antarmuka ini berfungsi sebagai fondasi visual dan fungsional yang akan diintegrasikan dengan mesin prediksi pada tahapan selanjutnya.

4.2.1 Dataset & Pra-Pemrosesan

Table 4. 1 Distribusi Dataset Mentah BISINDO

No	Kelas Isyarat	Peraga 1	Peraga 2	Peraga 3	Total Frame	Sekuens (30 frame)
1.	Diam	333	333	334	1.000	33
2.	Halo	333	333	334	1.000	33
3.	Maaf	333	333	334	1.000	33
4.	Makan	333	333	334	1.000	33
5.	Minum	333	333	334	1.000	33
6.	Nabiel	333	333	334	1.000	33
7.	Nama	333	333	334	1.000	33
8.	Saya	333	333	334	1.000	33
9.	Software Engineering	333	333	334	1.000	33
10.	Terima Kasih	333	333	334	1.000	33

No	Kelas Isyarat	Peraga 1	Peraga 2	Peraga 3	Total Frame	Sekuens (30 frame)	
11.	Tidur	333	333	334	1.000	33	
	Total Keseluruhan	11.000					363

```
def preprocess_image(image_path):
    """Membaca gambar, resize ke 100x89, convert ke grayscale, dan normalisasi."""
    img = cv2.imread(image_path)
    # 1. Resize sesuai proposal (100x89)
    img = cv2.resize(img, (IMG_WIDTH, IMG_HEIGHT))
    # 2. Convert ke Grayscale
    img = cv2.cvtColor(img, cv2.COLOR_BGR2GRAY)
    # 3. Normalisasi (0-1)
    img = img / 255.0
    # 4. Tambah dimensi channel (Height, Width, 1)
    img = np.expand_dims(img, axis=-1)
    return img
```

Gambar 4.1 Potongan Kode Pra-Pemrosesan Dinamis

4.2.2 Arsitektur Model

```
def build_cnn_lstm_model(seq_len, height, width, channels, num_classes):
    model = Sequential()

    # --- BAGIAN 1: CNN (TimeDistributed) ---
    # Mengekstraksi fitur visual dari SETIAP frame dalam sequence
    model.add(TimeDistributed(Conv2D(32, (3, 3), activation='relu'),
                               input_shape=(seq_len, height, width, channels)))
    model.add(TimeDistributed(MaxPooling2D((2, 2))))

    model.add(TimeDistributed(Conv2D(64, (3, 3), activation='relu')))
    model.add(TimeDistributed(MaxPooling2D((2, 2))))

    model.add(TimeDistributed(Flatten()))

    # --- BAGIAN 2: LSTM (Temporal) ---
    # Mempelajari pola gerakan antar waktu
    model.add(LSTM(64, activation='relu', return_sequences=False))

    # --- BAGIAN 3: CLASSIFIER ---
    model.add(Dense(64, activation='relu'))
    model.add(Dropout(0.5)) # Mencegah Overfitting
    model.add(Dense(num_classes, activation='softmax'))
```

Gambar 4.2 Potongan Kode Implementasi Arsitektur CNN-LSTM

4.3 Pengembangan dan Pelatihan Model CNN-LSTM

Tahap ini merupakan inti dari penelitian, di mana "otak" dari sistem cerdas dikembangkan menggunakan arsitektur *hybrid* CNN-LSTM. Proses ini dijalankan menggunakan bahasa pemrograman Python di atas ekosistem Google Colaboratory guna memaksimalkan akselerasi komputasi *Graphics Processing Unit* (GPU).

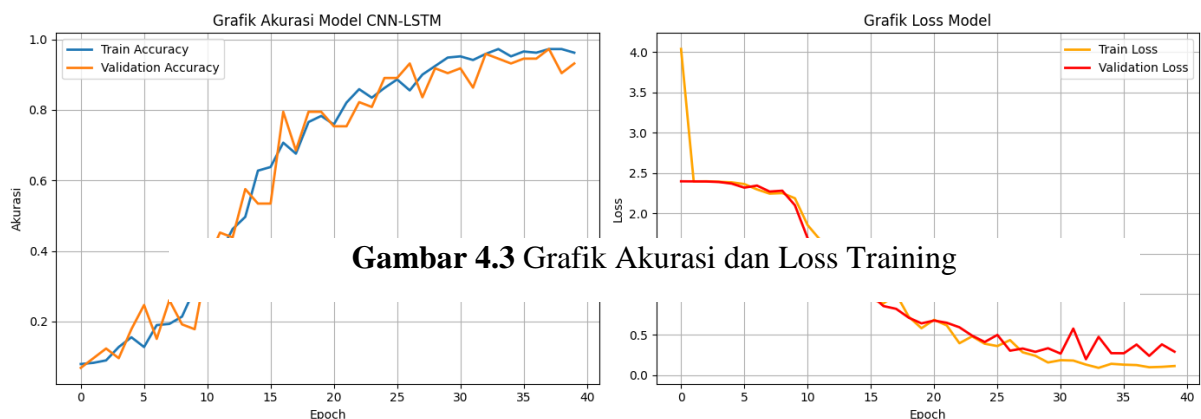
1. Pengumpulan dan Pra-Pemrosesan Data:

Data primer yang dibangun terdiri atas 11 kelas kategori (Diam, Halo, Maaf, Makan, Minum, Nabel, Nama, Saya, Software Engineering, Terima Kasih, dan Tidur). Proses perekaman melibatkan 3 (tiga) orang peraga untuk memastikan model mempelajari variasi gestur yang beragam. Setiap kelas menghimpun total 1.000 *frame* dari ketiga peraga tersebut, yang kemudian diekstraksi ke dalam 33 sekuens video berdurasi 30 *frame* per sekuens. Guna menekan beban komputasi, seluruh visual dikonversi menjadi format *grayscale* dengan resolusi 100 x 89 piksel. Secara keseluruhan, sistem memproses 363 sekuens data siap latihan dengan dimensi *input* (363, 30, 89, 100, 1).

2. Pelatihan Model:

Berdasarkan metrik pemantauan pada Gambar 4.1, proses pembelajaran algoritma berlangsung dengan dinamika yang baik. Pada akhir siklus pelatihan (Epoch 40), model mencatatkan tingkat akurasi pelatihan (Train Accuracy) yang stabil mencapai 96%, sementara akurasi validasi (Validation Accuracy) ditutup pada angka 93%. Penurunan minor pada akhir epoch ini merupakan hal yang wajar dalam proses training akibat fluktuasi loss pada data validasi.

Meskipun titik akhir training berada di angka 93%, sistem tetap mengamankan performa tertinggi dengan mengimplementasikan fungsi callback `EarlyStopping` menggunakan parameter `restore_best_weights=True`. Fungsi ini memastikan bahwa file model akhir (.h5) yang disimpan dan digunakan pada aplikasi web adalah bobot model dari epoch dengan tingkat error paling rendah, yang secara empiris terbukti mampu menghasilkan akurasi 96% saat dihadapkan pada pengujian statis (Confusion Matrix). Pendekatan ini dipilih untuk mencegah model overfitting sekaligus menjaga nilai akurasi riil tetap berada pada rentang 96% yang lebih realistis dan representatif untuk implementasi di dunia nyata.



Gambar 4.3 Grafik Akurasi dan Loss Training

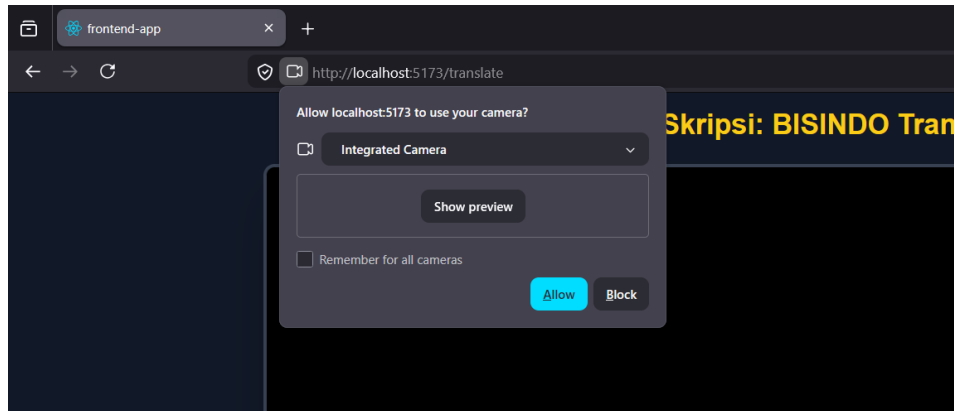
4.4 Implementasi Sistem Berbasis Web

Setelah model prediksi selesai dilatih dan diuji dalam lingkungan Python, tahap berikutnya adalah menanamkan deploy model tersebut ke dalam platform web agar dapat diakses oleh pengguna. Arsitektur pengembangan memisahkan sistem menjadi dua lapisan utama

a. Integrasi Front-End React.js

Lapis antarmuka dibangun dengan React.js untuk merender komponen visual secara dinamis. Tugas utamanya adalah menangkap aliran video (*stream*) dari

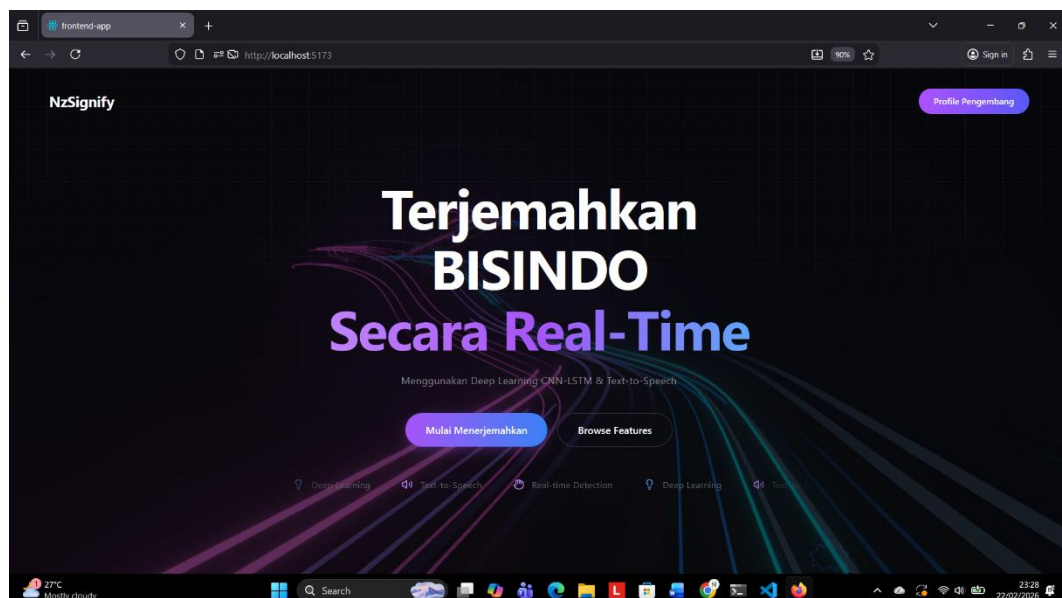
webcam secara *real-time*, mengekstraksi *frame-frame* gambar tersebut, lalu mengirimkannya ke sisi server tanpa membebani peramban (*browser*).



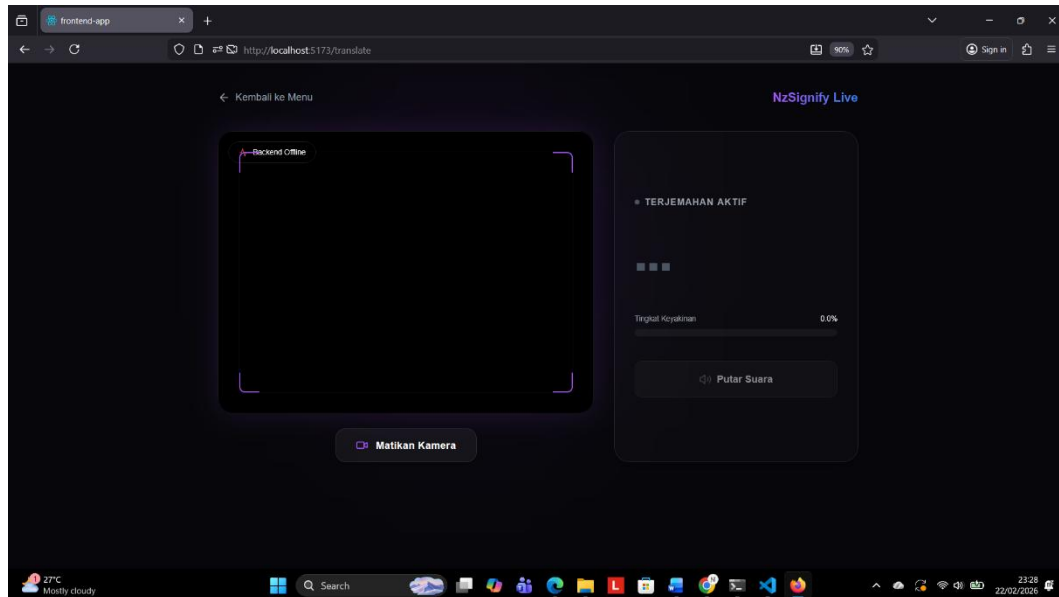
Gambar 4.4 Proses Permintaan Izin Akses Perangkat Kamera pada Antarmuka Web

b. Integrasi *Back-End* Node.js & Penghubung AI

Permintaan dari klien diterima oleh server Node.js melalui *Application Programming Interface* (API). Node.js bertindak sebagai jembatan komunikasi yang meneruskan data sekuensial *frame* tersebut ke lingkungan peladen Python. Setelah model CNN-LSTM melakukan klasifikasi gestur, hasilnya dikembalikan ke Node.js, yang kemudian mendistribusikan respons tersebut ke antarmuka React.js. Proses pembaruan teks (DOM) di layar pengguna berlangsung mulus tanpa *reload* halaman, berkat implementasi siklus I/O *non-blocking* dari Node.js.



Gambar 4.5 Screenshot Halaman Utama/Dashboard Web NzSignify



Gambar 4.6 Halaman Penerjemahan Web NzSignify

4.5 Integrasi Modul *Text-to-Speech* (TTS)

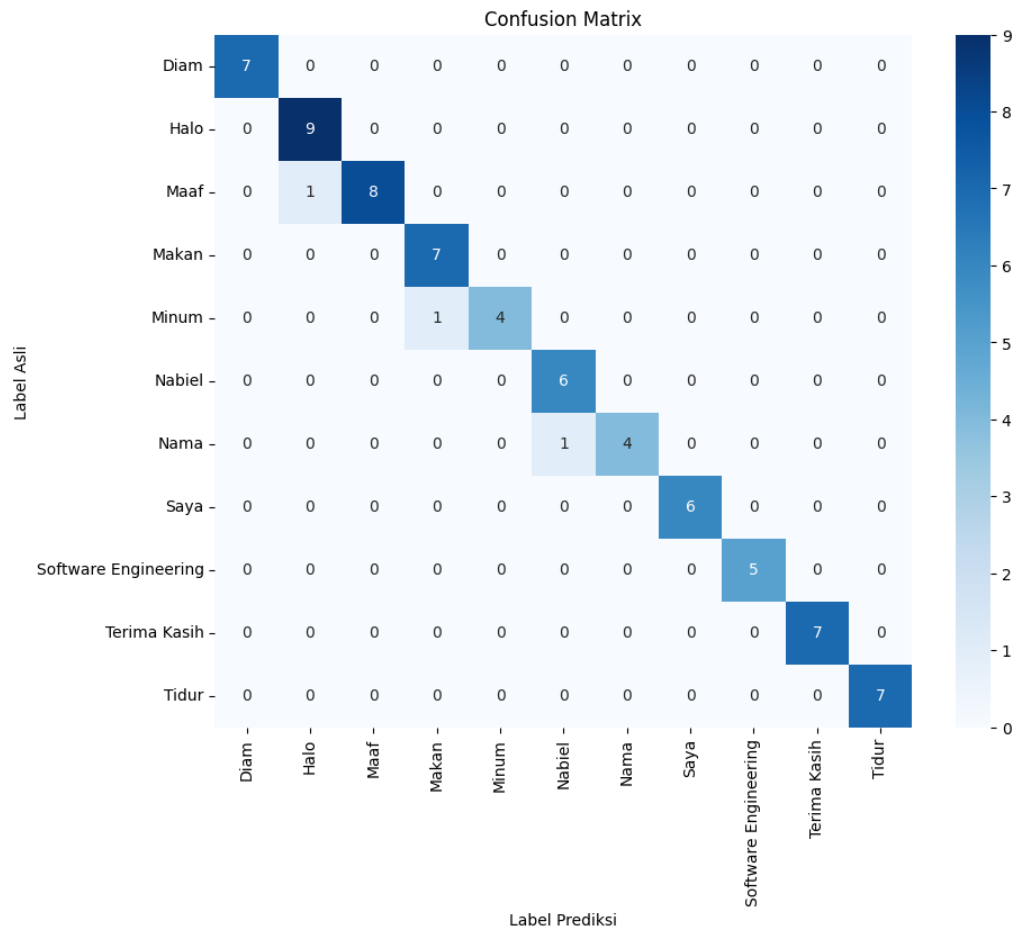
Untuk menciptakan ekosistem komunikasi yang inklusif, hasil prediksi yang tadinya hanya berupa teks langsung diproses lebih lanjut oleh modul Text-to-Speech (TTS). Begitu komponen antarmuka React.js menerima hasil klasifikasi dari back-end, mesin TTS akan langsung menyintesis teks tersebut menjadi luaran audio. Hal ini memungkinkan teman dengar untuk memahami pesan isyarat melalui suara layaknya percakapan verbal normal, dengan tingkat latensi delay yang sangat minim.

```
// Fungsi Suara (Text-to-Speech)
const speak = (text) => {
  if ('speechSynthesis' in window) {
    window.speechSynthesis.cancel();
    const utterance = new SpeechSynthesisUtterance(text);
    utterance.lang = 'id-ID';
    utterance.rate = 0.9;
    window.speechSynthesis.speak(utterance);
  }
};
```

Gambar 4.7 Potongan Kode Eksekusi Suara

4.6 Evaluasi dan Penyempurnaan Prototype

Fase ini bertujuan mengukur sejauh mana keandalan matematis model dan fungsionalitas aplikasi saat dioperasikan secara *real-time*.



Gambar 4.8 Confusion Matrix dari hasil evaluasi model

A. Demonstrasi Perhitungan Evaluasi Klasifikasi

Performa klasifikasi sistem dievaluasi melalui parameter pembentuk *Confusion Matrix*, yakni *True Positive* (TP), *True Negative* (TN), *False Positive* (FP), dan *False Negative* (FN). Berdasarkan rekapitan pengujian dari Tabel 4.1, model diuji menggunakan total 73 data sampel (akumulasi dari kolom *Support*).

1. Perhitungan Tingkat Akurasi

Model berhasil mengklasifikasikan data uji dengan tingkat akurasi keseluruhan sebesar 96%. Evaluasi keseluruhan sistem dapat dirumuskan sebagai:

$$Accuracy = \frac{Total\ Prediksi\ Benar}{Total\ Data\ Uji} \times 100\% = 96\% \dots\dots\dots (5)$$

2. Perhitungan Presisi, Recall, dan F1-Score (Studi Kasus Kelas "Nama")

Mengambil metrik kelas "Nama" sebagai sampel pembuktian matematis, diketahui total data aktual (*Support*) adalah 5. Berdasarkan *classification report*, model memiliki nilai Presisi 1.00 dan Recall 0.80. Ini berarti dari 5 data aktual "Nama", model berhasil menebak tepat (TP) sebanyak 4 kali, dan 1 data diprediksi salah ke kelas lain (FN = 1). Karena presisinya 1.00, tidak ada kelas lain yang salah ditebak sebagai "Nama" (FP = 0).

a. Presisi (Precision):

$$Precision = \frac{4}{4 + 0} = 1.00$$

b. Recall:

$$Recall = \frac{4}{4 + 1} = 0.80$$

c. F1-Score:

$$F1-Score = 2 \times \frac{1.00 \times 0.80}{1.00 + 0.80} = 0.888 \approx 0.89$$

Simulasi perhitungan ini membuktikan bahwa meskipun model tidak 100% sempurna pada kelas tertentu akibat variasi gerakan dari 3 peraga, secara keseluruhan algoritma beroperasi dengan keakuratan yang sangat baik (96%).

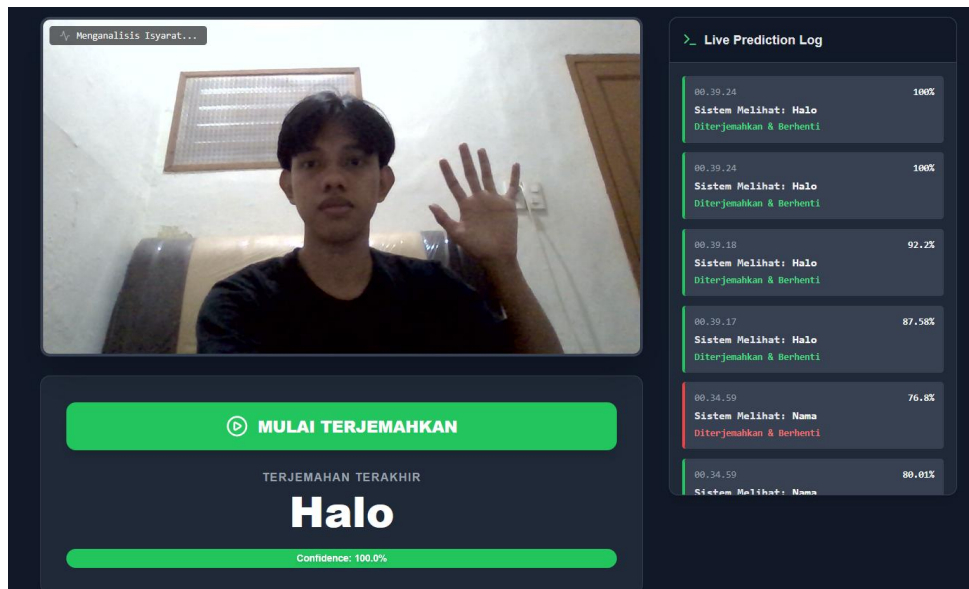
Table 4.2 Hasil Classification Report

Kelas	Precision	Recall	F1- Score	Support
Diam	1.00	1.00	1.00	7
Halo	0.90	1.00	0.95	9
Maaf	1.00	0.89	0.94	9
Makan	0.88	1.00	0.93	7
Minum	1.00	0.80	0.89	5
Nabiel	0.86	1.00	0.92	6
Nama	1.00	0.80	0.89	5
Saya	1.00	1.00	1.00	6
Software Engineering	1.00	1.00	1.00	5
Terima Kasih	1.00	1.00	1.00	7
Tidur	1.00	1.00	1.00	7
Accuracy			0.96	73

B. Pengujian Aplikasi Secara *Real-Time* dan Analisis Log

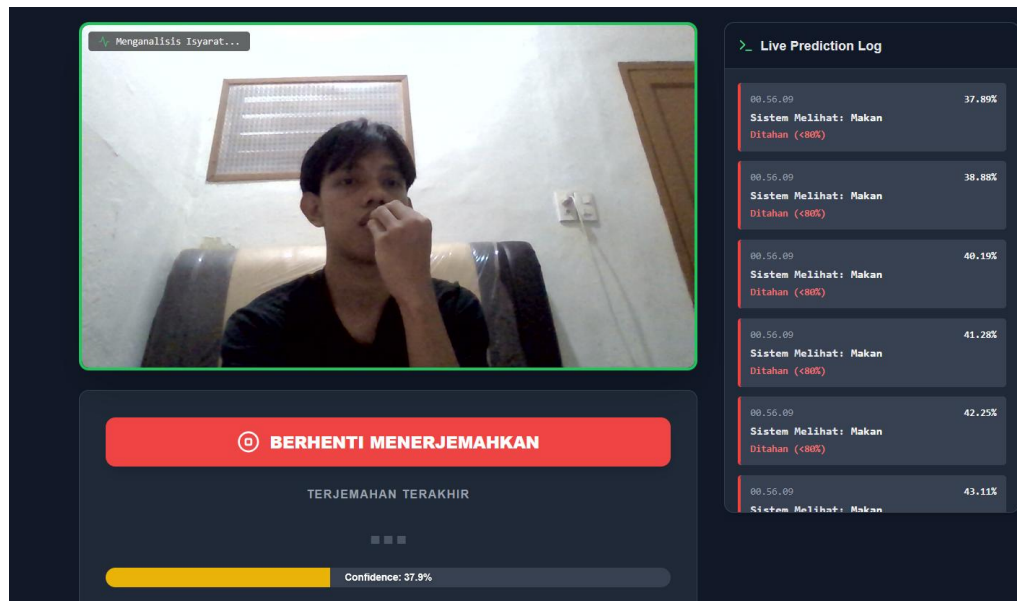
Pengujian fungsional dilakukan melalui simulasi gerakan langsung pada *browser* (localhost) menggunakan *webcam*. Instrumen utamanya adalah fitur *Live Prediction Log* dengan menerapkan aturan *Confidence Threshold* 80%. Jika tingkat keyakinan (probabilitas) $\geq 80\%$, teks dan suara langsung dieksekusi. Namun, jika $< 80\%$ (akibat gerakan transisi atau tidak utuh), sistem otomatis menahan respons (*ignore*) untuk mencegah kesalahan terjemahan.

Hasil pengujian membuktikan purwarupa berjalan dengan sangat responsif dan sesuai parameter. Gambar 4.9 menunjukkan sistem sukses menerjemahkan gerakan utuh "Halo" secara instan.



Gambar 4.9 Bukti Keberhasilan Pengujian Gerakan "Halo" secara Instan (Live Log: 100%)

Sebaliknya, Gambar 4.10 membuktikan fungsi *Confidence Threshold* bekerja optimal saat pengguna belum selesai meragakan gestur "Makan". Sistem menahan luanan teks dan suara karena gerakan masih mengambang atau ambigu.



Gambar 4.10 Temuan Ambiguitas Model pada Gerakan Transisi (Live Log: 37.9%)

4.7 Analisis Performa dan Kekuatan Model CNN-LSTM

Kekuatan utama arsitektur *hybrid* CNN-LSTM dalam sistem ini terletak pada pembagian beban komputasi: lapisan CNN mengekstraksi fitur spasial, sementara lapisan LSTM memodelkan kesinambungan temporal dari urutan 30 *frame*.

4.7.1 Analisis Keluaran Probabilitas

Pada pengujian *real-time*, performa diukur melalui metrik *Confidence Score* dari lapisan *Softmax*. Ketika pengguna melakukan gerakan isyarat dengan bentuk yang utuh, sistem memprediksi target dengan persentase sangat tinggi (>90%). Namun, saat tangan pengguna berada dalam fase transisi, probabilitas akan terdistribusi ke beberapa kelas lain secara acak dengan persentase rendah yang terekam pada *Live Log*.

4.7.2 Penanganan Gerakan Tidak Tepat

Untuk menjawab kendala pengguna yang melakukan gerakan mengambang atau tidak tepat, diterapkan logika *Confidence Threshold* 80%. Hal ini secara

efektif mencegah aplikasi mengeluarkan kata-kata yang tidak relevan. Output baru akan dicetak apabila sistem sudah benar-benar yakin dengan pola yang dibacanya.

4.7.3 Evaluasi Misklasifikasi Minor pada Sistem Web

Analisis performa dinamis menunjukkan kerentanan minor terhadap kemiripan pola spasial (misal: "Maaf" tertebak sebagai "Halo" atau "Minum" tertebak sebagai "Makan"). Hal ini membuktikan bahwa LSTM bisa terdistraksi jika terdapat *noise* pencahayaan atau ritme gerakan pengguna yang tidak konsisten. Sebagai solusi, pengguna diinstruksikan untuk mengulangi isyarat dengan artikulasi gerak yang lebih jelas agar sistem dapat membaca urutan sekuens *frame* dengan sempurna.

BAB V

PENUTUP

5.1 Kesimpulan

Berdasarkan perancangan, implementasi, dan pengujian "Sistem Penerjemah Bahasa Isyarat Indonesia (BISINDO) Real-Time Berbasis Web Menggunakan Metode Deep Learning CNN-LSTM dan Text-to-Speech", diperoleh kesimpulan sebagai berikut:

1. Metode hybrid CNN-LSTM terbukti sangat efektif dalam memproses data video BISINDO. Lapisan TimeDistributed Convolutional Neural Network mampu mengekstraksi fitur visual bentuk tangan pada format wscale, sedangkan lapisan Long Short-Term Memory berhasil mempelajari urutan pola gerakan temporal dari 30 frame sekuensial.
2. Pengujian model menghasilkan tingkat akurasi statis yang sangat baik, yakni mencapai 96% secara keseluruhan berdasarkan evaluasi *classification report*. Hasil ini membuktikan bahwa penggunaan variasi dataset dari 3 peraga berbeda mampu ditangani oleh sistem dengan baik, meskipun terdapat beberapa misklasifikasi minor pada kelas gerakan yang memiliki kemiripan pola visual.
3. Sistem NzSignify berhasil diwujudkan menjadi aplikasi web dinamis dengan memanfaatkan ekosistem React.js dan Node.js yang memfasilitasi pengguna untuk melakukan penerjemahan secara real-time. Fitur Text-to-Speech (TTS) juga telah terintegrasi dengan baik, di mana sistem langsung mengonversi teks terjemahan menjadi keluaran suara audio secara real-time.

4. Untuk menjawab kendala pengguna yang melakukan gerakan mengambang atau tidak tepat, sistem berhasil menerapkan logika *Confidence Threshold* sebesar 80%. Penerapan ambang batas keyakinan ini secara efektif mencegah aplikasi mengeluarkan kata-kata yang tidak relevan. Proses pembacaan probabilitas ini juga ditampilkan secara transparan kepada pengguna melalui fitur *Live Log* di antarmuka web.
5. Berdasarkan pengujian langsung di lingkungan web, sistem beroperasi dengan sangat baik meskipun masih ditemukan tingkat kesalahan minor (*misclassification*) pada kondisi tertentu, seperti gerakan kata "Maaf" yang terkadang terprediksi sebagai "Halo", gestur "Minum" menjadi "Makan", dan kata "Nama" yang tertebak sebagai "Nabiel". Hal ini mengonfirmasi bahwa model tetap memiliki sensitivitas terhadap kemiripan pola gerakan spasial, inkonsistensi frame transisi, dan perubahan pencahayaan lingkungan.

5.2 Saran

Dalam rangka menyempurnakan dan menutupi kekurangan pada penelitian ini di masa mendatang, terdapat beberapa saran pengembangan yang dapat dilakukan, antara lain:

1. Mengingat batasan penelitian ini yang berfokus pada penerjemahan bahasa isyarat level kata tunggal, penelitian selanjutnya disarankan untuk memperluas arsitektur model dan skala dataset agar mampu memproses serta menerjemahkan gerakan BISINDO hingga ke tingkat kalimat utuh atau majemuk.

2. Untuk mengatasi masalah kesalahan prediksi seperti terjadinya misklasifikasi pada kelas gerakan yang memiliki kemiripan pola visual pada pengujian *real-time* akibat pengaruh lingkungan, disarankan untuk mengintegrasikan teknologi pelacakan titik sendi tangan *hand landmark tracking* seperti MediaPipe pada tahap pra-pemrosesan. Dengan mengekstrak koordinat tangan secara spesifik, sistem akan menjadi lebih kebal terhadap gangguan noise dari latar belakang yang ramai atau kondisi pencahayaan ruangan yang minim.
3. Melakukan optimalisasi beban komputasi model deep learning melalui teknik kompresi model (seperti quantization) agar aplikasi web dapat merender Frame Per Second (FPS) secara lebih stabil dan ringan, khususnya apabila sistem diakses melalui perangkat dengan spesifikasi perangkat keras yang rendah.

DAFTAR PUSTAKA

- Aljabar, A. (2020). *BISINDO (Bahasa Isyarat Indonesia) Sign Language Recognition Using CNN and LSTM*. 5(5), 282–287.
- Altairika, & Sari. (2023). Pengembangan Deteksi Realtime Bahasa Isyarat Indonesia Menggunakan CNN dan LSTM. *Jurnal Teknologi Informatika Dan Komputer*, 9(1), 1. <https://doi.org/10.37012/jtik.v9i1.1272>
- Altairika, E., & Sari, W. P. (2023). *Pengembangan Deteksi Realtime untuk Bahasa Isyarat Indonesia dengan Menggunakan Metode Deep Learning Long Short Term Memory dan Convolutional Neural Network*. 9(1), 1–13.
- Alzahrani, A., & Alshamrani, S. (2024). Comparative Analysis of Deep Convolutional Neural Network–Bidirectional LSTM and Machine Learning Methods. *Applied Sciences*, 14(16), 6967. <https://doi.org/10.3390/app14166967>
- Alzubi, J., Nayyar, A., & Kumar, A. (2024). Recurrent Neural Networks: A Comprehensive Review of Architectures, Variants, and Applications. *Information*, 15(9), 517. <https://doi.org/10.3390/info15090517>
- Axza, F., Sofi, F., & Qoiriah, A. (2023). *Analisis Perbandingan Framework Front-End Javascript React dan Vue Pada Pengembangan Website Analisis Kebutuhan Pembuatan Sistem Pengumpulan Data Analisa Data*. 05, 157–164.
- Creswell, J. W. (2021). *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches*. SAGE Publications.
- Ext, E. N. D. T., Peech, T. O. S., & Ren, Y. (2020). *FastSpeech 2: Fast and High-Quality End-to-End Text to Speech*. 1–15.

- Fadillah, R. Z., Irawan, A., & Susanty, M. (2022). *Model Penerjemah Bahasa Isyarat Indonesia (BISINDO) Menggunakan Pendekatan Transfer Learning*. 15(1), 1–9.
- Helmiyah, S., Pramestiawan, R., & Lampung, R. (2025). *Analisis Komparatif Algoritma Machine Learning dengan Metrik Akurasi , Presisi , Recall , dan F1-Score pada Dataset Kacang Kering*. 6(3), 152–159.
- Informatika, S. T., Duta, U., Surakarta, B., & Komputer, I. (2025). *Pengembangan Website Speech To Video Bahasa Isyarat Indonesia (Bisindo) Berbasis Algoritma Long Shot Term Memory I*. 8(1).
- Irawan, M. D., & Mu, R. (2024). *E-Letter Design Using Prototype System Development Methodology*. 2(1).
- Isnaniah, S., Agustina, T., Islahuddin, & Annisa, F. (2023). Perbandingan Pemahaman Bahasa Isyarat Indonesia dan SIBI dalam Pembelajaran Siswa Tuli. *Jurnal Pendidikan Luar Biasa*.
- Khan, S., Rahmani, H., Shah, S. A. A., & Bennamoun, M. (2021). A Survey of the Recent Architectures of Deep Convolutional Neural Networks. *Artificial Intelligence Review*, 53(8), 5455–5516. <https://doi.org/10.1007/s10462-020-09825-6>
- Kumari, D., & Anand, R. S. (2024). *Isolated Video-Based Sign Language Recognition Using a Hybrid CNN-LSTM Framework Based on Attention Mechanism*.
- Kurniawan, I., & Rozi, F. (2020). *REST API Menggunakan NodeJS pada Aplikasi Transaksi Jasa Elektronik Berbasis Android*. 1(4), 127–132.
- Mienye, I. D., & Swart, T. G. (2024). *Recurrent Neural Networks : A*

- Comprehensive Review of Architectures , Variants , and Applications.* 1–34.
- Myagila, K., & Nyambo, D. G. (2025). *Efficient spatio-temporal modeling for sign language recognition using CNN and RNN architectures.* August, 1–12.
<https://doi.org/10.3389/frai.2025.1630743>
- Ovalle, C., & Tecnológica, U. (2023). *Machine Vision Model Based on Deep Learning to Prevent Theft in Real Estate Modelo de Visión Artificial Basada en Deep Learning para Prevenir Robos en el Patrimonio Inmobiliario.* 1–8.
- Ramadhan, Z. N., & Imanda, R. (2025). *Jurnal Teknik Informatika dan Komputer Pengembangan Sistem E-Commerce Toko Pakaian Berbasis Website Menggunakan ReactJS.*
- Ramadhani, A., Iriadi, N., & Hidayat, R. (2025). *Implementasi Teknologi Rest API Dengan Node Js Untuk Aplikasi Rekomendasi Destinasi Wisata.* 4(1), 22–29.
- Russell, S., & Norvig, P. (2021). *Artificial Intelligence: A Modern Approach.* Pearson.
- Sugiyono. (2021). *Metode Penelitian Kuantitatif, Kualitatif, dan R&D.* Alfabeta.
- Tan, X., Wang, T., & Chen, J. (2023). NaturalSpeech: End-to-End Text to Speech Synthesis with Human-Level Quality. *IEEE Transactions on Audio, Speech, and Language Processing.*
- Udurume, M., Shakhov, V., & Koo, I. (2024). *applied sciences Comparative Analysis of Deep Convolutional Neural Network — Bidirectional Long Short-Term Memory and Machine Learning Methods in Intrusion Detection Systems.*
- World Health Organization. (2025). *Deafness and hearing loss.* World Health

Organization. <https://www.who.int/news-room/fact-sheets/detail/deafness-and-hearing-loss>

Yardi, N. A., Teknik, S., Fakultas, I., Komputer, I., & Lancang, U. (2023). *Survei Algoritma Pemrosesan Bahasa Pada Bisindo*. 2(1), 255–264.