

## BAB II

### TINJAUAN PUSTAKA

#### 2.1 Penelitian Terdahulu

##### 2.1.1 Pengenalan Gerakan Olahraga Berbasis *Long Short-Term Memory* Menggunakan *Mediapipe* [6]

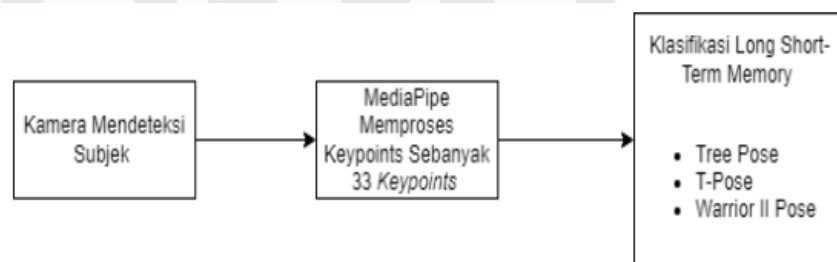
Pengenalan gerakan manusia sudah pernah diidentifikasi menggunakan CCTV, Robot, dan IMK (Interaksi Manusia Komputer). Seiring berjalannya kemajuan teknologi, pengenalan tersebut diterapkan menggunakan keilmuan bidang Computer Vision dan Machine Learning. Computer Vision dan Machine Learning dapat mendeteksi gerakan manusia dengan akurat yang biasanya dikenal dengan istilah HAR (Human Activity Recognition). HAR dapat mendeteksi gerakan yang dilakukan manusia di dalam ruangan seperti berjalan, berdiri, duduk dan berbicara. Namun, HAR belum dapat mendeteksi gerakan spesifik dari Olahraga tertentu.

Yoga merupakan salah satu gerakan olahraga yang banyak diikuti kaum wanita. Jenis gerakan yoga dianggap dapat memberikan keseimbangan dan kesehatan bagi tubuh serta mental. Dengan adanya teknologi HAR pada Computer Vision, gerakan Yoga dapat dipantau agar tidak salah gerakannya, sehingga tidak perlu didampingi oleh tutor Yoga. Selama pandemi, sanggar Olahraga ditutup dan dihimbau untuk olahraga di rumah. Karena gerakan Yoga dianggap lumayan sulit bagi pemula, sehingga dibutuhkan adanya teknologi HAR untuk membantu mengenalkan gerakan tersebut. Agar gerakan Yoga yang dilakukan benar. Pada penelitian sebelumnya yang dilakukan oleh (Sun et al., 2018) melakukan pengenalan gerakan manusia dengan CNN, dengan cara melakukan tracking kemudian ditampilkan secara visual dengan kamera. Namun pengenalan ini dianggap tingkat klasifikasinya kurang akurat. CNN dianggap bekerja dengan baik pada deteksi

gambar dan ucapan saja yang tidak real time.

Oleh sebab itu, pada penelitian ini diangkat metode lain yaitu, LSTM (Long Short-Term Memory) yang berbasis RNN yang dapat mendeteksi suatu gerakan secara real time. Pada artikel ini membahas mengenai sistem pengenalan dan klasifikasi gerakan Yoga yaitu T-Pose, Warrior II Pose, dan Tree Pose menggunakan Mediapipe dan LSTM. Gerakan Yoga yang benar dapat dengan mudah dideteksi pada sistem ini. Sistem ini sangat tepat diterapkan pada saat pandemi Covid-19, menghindari keramaian dan seseorang tidak perlu keluar rumah untuk berolahraga Yoga.

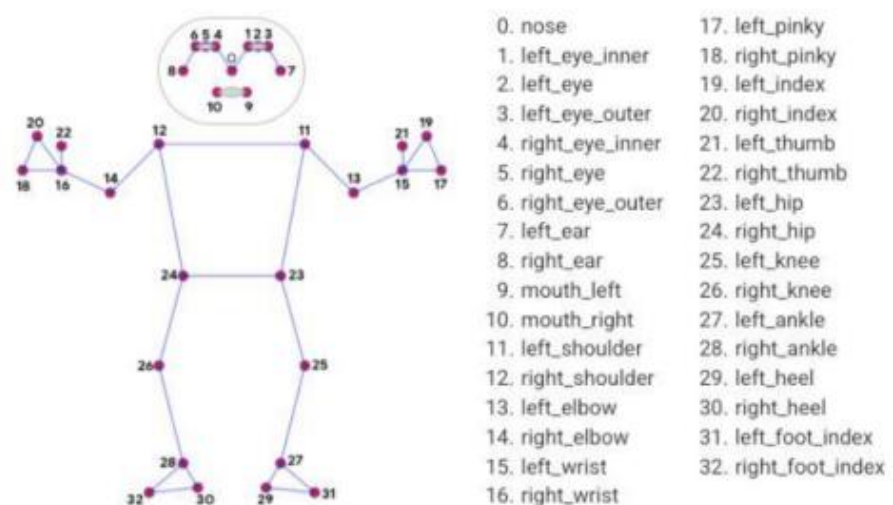
Secara prosedur teori pada HAR, aktifitas gerakan tubuh manusia di setiap Frame dapat diklasifikasikan kemudian dilanjutkan dengan analisis bagian tubuh dari gerakan tersebut berdasarkan pengaturan waktu dari klasifikasiya. Prosedur pertama gerakan direkam menggunakan kamera, kemudian dilanjutkan dengan deteksi 33 Keypoints dari postur badan manusia. Setelah itu, video dengan 33 keypoint pada frame video diamati apakah ada yang tidak terdeteksi. Prosedur kedua menganalisis gerakan tubuh dengan 33 keypoints secara real time dan membuat klasifikasi gerakan dengan LSTM secara real time. Prosedur penelitian ini digambarkan pada Gambar 2.1.



Gambar 2. 1 Prodedur Deteksi dan Klasifikasi Gerakan

MediaPipe membangun sebuah “pipeline” dan akan mengolah data yang masuk sembarangan. Mediapipe merupakan sebuah framework yang dirancang untuk kecerdasan buatan dalam sebuah aplikasi. Mediapipe juga mengandung Tensor Flow yang

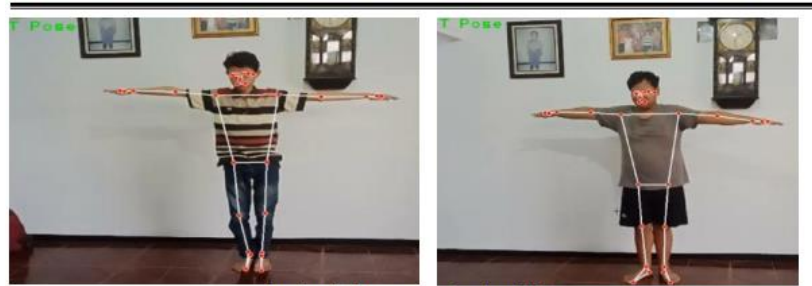
mendukung akselerasi GPU (Grapihal Processing Unit) dan CPU (Central Processing Unit Flow) dari sebuah perangkat. Langkah pertama dicapai dengan menggunakan kamera atau webcam yang digunakan, lalu MediaPipe menampilkan kerangka mewakili dtubuh badan manusia. postur tubuh atau body skeletons sebanyak (33 keypoints). Setelah itu mengamati satu frame dalam sebuah video atau gambar. Kemudian menganalisis gerakan dari tubuh secara berubah dan membuat prediksi itu. dilakukan dengan menggunakan Long Short-Term Memory (LSTM). Keypoints dari urutan frame dikirim ke LSTM untuk klasifikasi memprediksi aktivitas yang sesuai. Pada penelitian ini, digunakan CPU dengan prosessor yang tinggi agar komputasi dari proses deteksi dan klasifikasi lebih efisien serta lebih cepat. Khususnya untuk deteksi 33 Keypoints pada inputan vidio, yang ditunjukkan pada Gambar 2.2.



Gambar 2.2 Index Keypoints pada Mediapipe

Hasil pengujian menunjukkan bahwa LSTM dengan 400 epoch dan optimizer "Adam" menghasilkan akurasi training sebesar 91% dan loss sebesar 29%. Parameter validasi menunjukkan akurasi sebesar 89%. Pengujian real-time keypoints menggunakan kamera handphone dengan aplikasi DroidCam menunjukkan hasil deteksi

yang baik pada 33 keypoints. Pengujian klasifikasi gerakan melibatkan empat subjek dengan lima kali percobaan pada setiap subjek menunjukkan akurasi pengenalan gerakan T-Pose sebesar 100%, Warrior II Pose sebesar 85%, dan Tree Pose sebesar 80%. Faktor yang mempengaruhi hasil pengujian termasuk peletakan kamera, pencahayaan, dan peniruan gerakan yang tepat.



Gambar 7. Pengenalan Gerakan T-Pose



Gambar 8. Pengenalan Gerakan Tree-Pose



Gambar 2. 3 Pengenalan Gerakan Yoga dari Hasil Pengujian

UNIVERSITAS  
MULTIMEDIA  
NUSANTARA

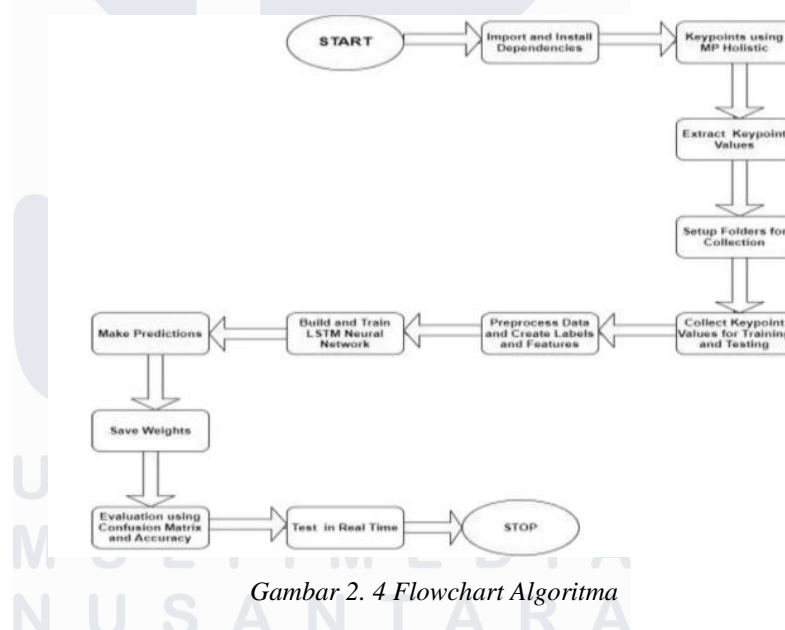
Akurasi Training Dataset LSTM dari gerakan yang telah dilakukan adalah 91% . Parameter validasi menghasilkan nilai akurasi sebesar 89%, dengan nilai loss sebesar 0.29 dan validasi loss sebesar 0.35 dengan 400 Epocs. Klasifikasi gerakan “T-Pose” sudah berhasil terklasifikasi 100%. Klasifikasi gerakan “Warrior II Pose berhasil terklasifikasi 85%. Penyebab dari gerakan yang tidak berhasil dikenali diakrenakan pencahayaan pada arena sekitar gerakan kurang merata, posisi kedua tangan tidak sejajar dengan tepat dan belum searah sebesar 180° derajat. Sedangkan klasifikasi “Tree Pose” didapatkan nilai akurasi sebesar 80%.

Subject	Warrior II Pose	T-Pose	Tree Pose
1	✓	✓	✓
	✓	✓	✓
	✓	✓	✓
	✓	✓	-
2	-	✓	✓
	-	✓	✓
	✓	✓	✓
	✓	✓	✓
3	-	✓	✓
	✓	✓	-
	✓	✓	✓
	✓	✓	✓
4	✓	✓	✓
	✓	✓	✓
	✓	✓	-
	✓	✓	✓
Rata-Rata	85%	100%	80%

Tabel 2. 1 Tabel Keberhasilan Klasifikasi Gerakan

## 2.1.2 Sign Language Detection Using Action Recognition with Python [7]

Penelitian oleh Dutta et al. (2023) berjudul "Sign Language Detection Using Action Recognition with Python". Penelitian ini bertujuan untuk mengatasi hambatan komunikasi antara individu dengan gangguan pendengaran dan masyarakat umum melalui sistem yang dapat mengenali dan menafsirkan gerakan bahasa isyarat secara real-time. Penelitian ini membahas tentang deteksi bahasa isyarat menggunakan pengenalan aksi dengan model deep learning LSTM yang diimplementasikan dalam Python. Penelitian ini berfokus pada pengembangan sistem yang mampu mengenali dan menafsirkan gerakan bahasa isyarat secara real-time dengan akurat. Sistem yang diusulkan terdiri dari beberapa tahap: akuisisi video, pra-pemrosesan, ekstraksi fitur, dan pelatihan model LSTM.



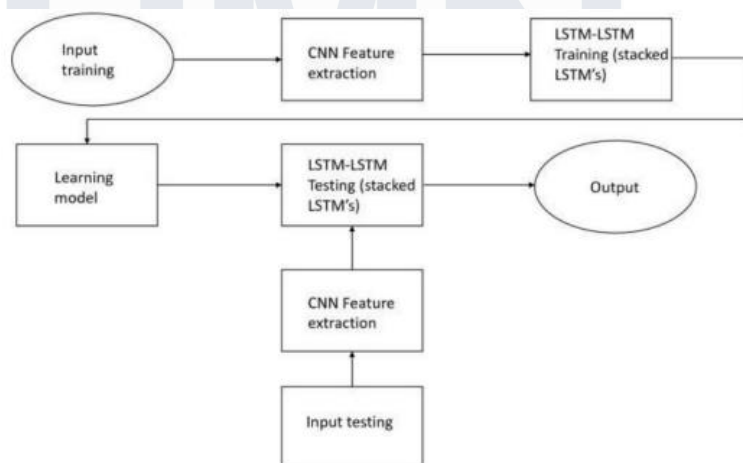
Gambar 2. 4 Flowchart Algoritma

Pada tahap akuisisi video, sistem mengambil input video yang berisi gerakan bahasa isyarat. Video ini kemudian dipecah menjadi frame individu. Dalam tahap pra-pemrosesan, frame-frame tersebut diproses untuk meningkatkan kualitas visual dan menghilangkan noise yang dapat mengganggu proses pengenalan.

Teknik pemrosesan gambar seperti normalisasi dan augmentasi gambar digunakan untuk memastikan bahwa data yang diumpungkan ke model adalah data yang bersih dan terstruktur.

Ekstraksi fitur merupakan tahap krusial berikutnya di mana fitur-fitur penting dari setiap frame diidentifikasi dan diekstraksi. Dalam penelitian ini, fitur-fitur ini diekstraksi menggunakan metode seperti optical flow dan model Convolutional Neural Network (CNN) yang telah dilatih sebelumnya. Optical flow membantu dalam menangkap gerakan antara frame berturut-turut, sedangkan CNN digunakan untuk mengekstrak fitur spasial yang penting dari setiap frame.

Setelah fitur diekstraksi, model LSTM dilatih untuk mengenali pola temporal dalam data tersebut. LSTM dipilih karena kemampuannya untuk menangkap ketergantungan jangka panjang dalam urutan data, yang sangat penting untuk memahami gerakan yang terjadi dalam bahasa isyarat. Model LSTM dilatih menggunakan dataset yang terdiri dari berbagai gerakan bahasa isyarat, memastikan bahwa model ini dapat mengenali dan menafsirkan berbagai jenis isyarat dengan tingkat akurasi yang tinggi.



Gambar 2. 5 Arsitektur LSTM

Arsitektur model LSTM untuk deteksi bahasa isyarat biasanya terdiri dari beberapa lapisan unit LSTM bersama dengan komponen tambahan. Jaringan LSTM digunakan untuk menangkap ketergantungan temporal dalam urutan video, dan menggabungkan CNN dengan LSTM membantu dalam mengenali aktivitas yang melibatkan gerakan berurutan.



Gambar 2. 6 Hasil Pengujian Secara Real Time

Model LSTM yang dilatih menunjukkan akurasi dan performa tinggi selama evaluasi, sebagaimana dibuktikan dengan metrik seperti akurasi, presisi, perolehan, dan skor F1. Hasil eksperimen menunjukkan bahwa model LSTM yang diusulkan efektif dalam mengenali gerakan bahasa isyarat dengan tingkat akurasi yang tinggi, memperlihatkan performa real-time yang dapat diterapkan dalam berbagai aplikasi seperti alat interpretasi real-time atau perangkat bantu untuk individu dengan gangguan pendengaran.

### 2.1.3 Aplikasi Penerjemah Bahasa Isyarat Indonesia “Hear Me” [8]

*Hear Me* adalah aplikasi *mobile* yang dirancang untuk membantu orang berkomunikasi dengan lebih mudah, terutama bagi mereka yang tunarungu atau mengalami kesulitan pendengaran.



Aplikasi ini dikembangkan oleh PT. Inovasi Disabilitas Indonesia dan diluncurkan pada tahun 2021. Hear Me menawarkan dua fitur utama:

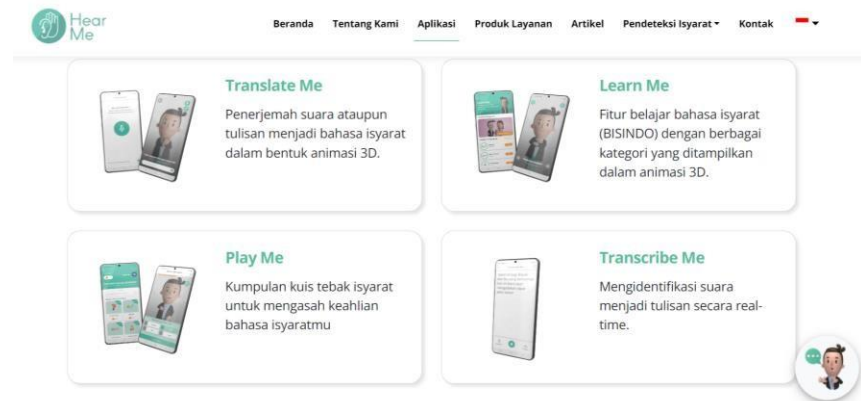
- **Pendeteksi Isyarat Angka:** Fitur ini memungkinkan pengguna untuk berkomunikasi dengan menggunakan bahasa isyarat angka. Pengguna cukup menggerakkan tangan mereka di depan kamera ponsel, dan aplikasi akan menerjemahkan gerakan tersebut menjadi teks.
- **Pendeteksi Isyarat Huruf:** Fitur ini memungkinkan pengguna untuk berkomunikasi dengan menggunakan bahasa isyarat huruf. Pengguna cukup menggerakkan tangan mereka di depan kamera ponsel, dan aplikasi akan menerjemahkan gerakan tersebut menjadi teks.

Selain fitur utama tersebut, *Hear Me* juga menawarkan beberapa fitur lainnya, seperti:

- **Pendeteksi wajah:** Fitur ini memungkinkan aplikasi untuk mendeteksi wajah pengguna dan melacak gerakan mereka.
- **Pengenalan suara:** Fitur ini memungkinkan aplikasi untuk mengenali suara pengguna dan mengubahnya menjadi teks.
- **Terjemahan:** Fitur ini memungkinkan aplikasi untuk menerjemahkan teks dari satu bahasa ke bahasa lain.



Gambar 2. 7 Tampilan Home Website Hear Me (Sumber: <https://hearme.id/beranda>)



Gambar 2. 8 Website Aplikasi Hear Me (Sumber: <https://hearme.id/app>)

Aplikasi *Hear Me* dapat diakses secara gratis melalui *Google Play Store* dan *App Store*. *Hear Me* bertujuan untuk memberikan bantuan dalam berkomunikasi bagi individu yang tunarungu atau memiliki masalah pendengaran, sehingga mempermudah interaksi dengan orang lain.

Berdasarkan uraian di atas, aplikasi *Hear Me* merupakan aplikasi *mobile* yang dirancang untuk membantu orang berkomunikasi dengan lebih mudah, terutama bagi mereka yang tunarungu atau mengalami kesulitan pendengaran. Aplikasi *Hear Me* dijadikan acuan penelitian terdahulu karena berkaitan dengan isi penelitian penulis, yaitu mengenai *computer vision* untuk translasi bahasa isyarat. Aplikasi *Hear Me* menggunakan teknologi *computer vision* untuk merubah teks menjadi gerakan bahasa isyarat yang diperagakan oleh karakter AI. Hal ini menjadikannya sebagai referensi yang baik untuk penelitian penulis.

## 2.2 Kesimpulan Penelitian Terdahulu

Dari beberapa penelitian terdahulu yang telah diulas, terdapat poin penting dari penelitian terdahulu yang memberikan landasan teoritis dan kontribusi untuk pengembangan sistem penerjemah bahasa isyarat menggunakan LSTM yang lebih efektif dan efisien.

### 1. Pengenalan Gerakan Olahraga Berbasis Long Short-Term Memory Menggunakan Mediapipe

- Teknologi Computer Vision dan Machine Learning telah diterapkan untuk mendeteksi gerakan manusia dengan akurat melalui metode Human Activity Recognition (HAR). Namun, penelitian ini mengidentifikasi bahwa penggunaan CNN saja dalam pengenalan gerakan masih kurang akurat dan membuktikan efektivitas LSTM dalam mendeteksi gerakan yang akurat dan real-time, yang mana dapat diterapkan dalam pengenalan gerakan tangan pada bahasa isyarat.
- Metode ini memberikan landasan bagi penggunaan algoritma pembelajaran mesin untuk mengenali pola gerakan tangan secara tepat dan efisien. Relevansi jurnal tersebut yaitu teknik LSTM yang digunakan dalam penelitian ini dapat diterapkan dalam sistem penerjemahan bahasa isyarat untuk meningkatkan akurasi pengenalan gerakan tangan pengguna.

### 2. Penerapan Computer Vision pada Aplikasi Hear Me

- Integrasi fitur pendeteksi wajah dan pengenalan suara dalam aplikasi penerjemah menunjukkan keunggulan teknologi yang dapat membantu memperkaya interaksi sosial bagi tunarungu, mendukung penelitian ini untuk mengembangkan solusi yang lebih holistik.
- Aplikasi ini bertujuan membantu komunikasi bagi

individu tunarungu atau yang memiliki masalah pendengaran, meningkatkan efisiensi dan efektivitas interaksi sosial mereka

### **3. Pengenalan dan Klasifikasi Gerakan Menggunakan CNN dan LSTM**

- Penelitian ini menunjukkan bahwa kombinasi CNN (Convolutional Neural Network) dan LSTM dapat secara efektif mengenali aktivitas manusia dari bingkai video. CNN digunakan untuk mempelajari fitur visual sementara LSTM menangkap ketergantungan temporal. Kombinasi CNN dan LSTM dapat diimplementasikan dalam sistem penerjemahan bahasa isyarat untuk meningkatkan akurasi dan efisiensi pengenalan gerakan tangan.
- Penelitian ini menunjukkan bahwa untuk gerakan yang lebih spesifik dan kompleks, metode yang lebih adaptif seperti LSTM diperlukan untuk meningkatkan akurasi dan keefektifan pengenalan gerakan.

Penelitian terdahulu menunjukkan berbagai pendekatan dalam pengenalan gerakan manusia menggunakan teknologi Computer Vision dan Machine Learning, khususnya CNN dan LSTM. Meskipun CNN telah diterapkan dengan beberapa keberhasilan, akurasinya masih kurang dalam pengenalan gerakan spesifik seperti yoga. Di sisi lain, aplikasi seperti Hear Me telah berhasil memanfaatkan teknologi ini untuk meningkatkan komunikasi bagi penyandang tunarungu. Penelitian-penelitian ini memberikan dasar kuat untuk pengembangan lebih lanjut, khususnya dalam mengadopsi LSTM untuk meningkatkan akurasi pengenalan gerakan yang kompleks.

### 2.3 Hipotesis

Hipotesis dalam penelitian ini dirumuskan untuk mengevaluasi kinerja dari sistem penerjemahan bahasa isyarat yang telah dikembangkan. Hipotesis yang diajukan adalah sebagai berikut; "Penggunaan teknologi *computer vision* dalam sistem penerjemahan bahasa isyarat ke dalam teks bahasa Indonesia akan meningkatkan efisiensi dan efektivitas komunikasi bagi pengguna bahasa isyarat, terutama komunitas tunarungu, dengan mengurangi hambatan dalam interaksi sosial dan memperluas akses mereka ke layanan dan informasi.

Hipotesis ini didasarkan pada asumsi bahwa dengan mengintegrasikan teknologi pengenalan gerakan tangan yang canggih dan algoritma pembelajaran mesin, sistem yang dikembangkan dapat secara akurat menerjemahkan bahasa isyarat ke dalam teks. Hal ini diharapkan dapat membantu individu tunarungu berkomunikasi lebih mudah dengan orang-orang yang tidak mengerti bahasa isyarat, sehingga memperbaiki kualitas interaksi sosial mereka dan memungkinkan integrasi yang lebih baik dalam berbagai aspek kehidupan sosial dan profesional.

Penerapan sistem penerjemahan yang mengkonversi Sistem Isyarat Bahasa Indonesia (SIBI) menjadi teks bahasa Indonesia tertulis akan secara signifikan meningkatkan kemudahan dan efektivitas komunikasi antara komunitas tuna rungu dan populasi pendengar. Sistem ini akan menjembatani kesenjangan komunikasi dengan menyediakan terjemahan bahasa isyarat yang jelas dan mudah dipahami ke dalam teks tertulis, sehingga meningkatkan pemahaman bersama dan integrasi sosial.

## 2.4 Tinjauan Teori

### 2.4.1 *Deep Learning*

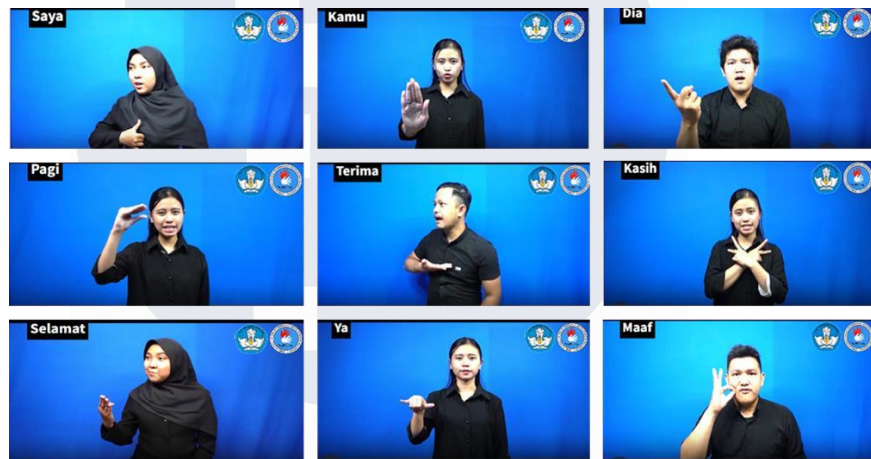
*Deep learning* merupakan metode dalam bidang kecerdasan buatan (AI) yang memungkinkan komputer untuk memproses data dengan cara terinspirasi dari fungsi otak manusia. Dengan menggunakan model *deep learning*, komputer dapat mengenali pola-pola kompleks dalam berbagai jenis data seperti gambar, teks, suara, dan data lainnya, untuk menghasilkan wawasan dan prediksi yang akurat. Kemampuan ini memungkinkan otomatisasi tugas-tugas yang sebelumnya memerlukan campur tangan manusia, seperti deskripsi citra atau konversi file suara menjadi teks.

*Deep learning* adalah evolusi dari konsep *machine learning* yang mengadopsi penggunaan tiga lapisan atau lebih dalam modelnya. Perbedaan mendasar antara *deep learning* dan *machine learning* terletak pada kemampuan *deep learning* untuk memahami karakteristik data tanpa perlu pengawasan manual manusia. Hal ini dicapai melalui penggunaan banyak lapisan yang dikenal sebagai *hidden layer*, yang terletak antara lapisan *input* dan *output*. *Hidden layer* ini penting dalam proses pembelajaran, karena membantu model untuk memahami data secara mendalam dan menghasilkan prediksi yang lebih akurat. Dengan demikian, *deep learning* menawarkan pendekatan yang lebih canggih dan kompleks dalam pembelajaran mesin, memungkinkan sistem untuk memahami data lebih baik dan menghasilkan hasil yang lebih akurat secara otomatis.[9][10]

### 2.4.2 Bahasa Isyarat (SIBI)

Sistem Isyarat Bahasa Indonesia (SIBI) adalah sistem bahasa isyarat yang dirancang untuk memfasilitasi komunikasi dalam pendidikan dan kehidupan sehari-hari bagi individu dengan gangguan pendengaran di Indonesia. SIBI resmi ditetapkan oleh

pemerintah Indonesia pada tanggal 30 Juni 1994 melalui Keputusan Menteri Pendidikan dan Kebudayaan. Hal ini dilakukan dengan maksud untuk mengambil standar bahasa Indonesia yang tepat dan baku, yang kemudian diterapkan dalam proses pengajaran di sekolah-sekolah khusus. Pengembangan ini merupakan bagian dari respons terhadap perkembangan teknologi informasi, yang memungkinkan penyediaan sumber daya pendidikan yang lebih efektif dan efisien untuk komunitas tunarungu. Berikut beberapa contoh gerakan bahasa isyarat SIBI yang diambil dari Kamus SIBI Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi:



Gambar 2. 9 Contoh Gerakan Bahasa Isyarat SIBI (Sumber: Kamus SIBI Kemendikbud, 2024)

Sebelum adopsi SIBI, pendekatan pendidikan untuk anak-anak tunarungu berfokus pada pengembangan kemampuan berbicara dan penguasaan bahasa. Namun, sekitar tahun 1970-an, metode komunikasi total yang memanfaatkan berbagai media komunikasi mulai diadopsi, yang juga meliputi penggunaan isyarat alami, abjad jari, dan isyarat yang dibakukan, termasuk adaptasi dari American Sign Language (ASL). Pendekatan ini memungkinkan anak tunarungu untuk menggunakan semua media komunikasi yang tersedia untuk mendukung pembelajaran mereka, termasuk isyarat.[11]

Dengan adanya SIBI, anak-anak tunarungu diharapkan bisa lebih efektif mengintegrasikan diri dalam pendidikan inklusi dan masyarakat secara luas, menyerap kurikulum dan memperoleh kemampuan literasi yang lebih baik.

### 2.4.3 Python

*Python* merupakan bahasa pemrograman tingkat tinggi yang ditandai dengan kemudahan pembacaan kode-kodenya. Dirancang oleh Guido van Rossum dan pertama kali dirilis pada tahun 1991, *Python* mendukung berbagai paradigma pemrograman, termasuk pemrograman objek-orientasi, terstruktur, dan fungsional. Kode *Python* biasanya diketahui karena sintaksisnya yang jelas dan mudah dipahami, yang memungkinkan pengembang untuk mengekspresikan konsep dalam jumlah baris kode yang lebih sedikit dibandingkan bahasa pemrograman lainnya. *Python* sangat populer dalam pengembangan perangkat lunak, analisis data, kecerdasan buatan, dan pembelajaran mesin karena perpustakaan standarnya yang kaya dan dukungan komunitas yang luas.

Python berperan penting dalam pengolahan citra dan pembelajaran mesin, yang merupakan dasar dari sistem translasi bahasa isyarat menjadi teks. Dengan menggunakan pustaka seperti *OpenCV* untuk pengolahan citra dan *TensorFlow* atau *PyTorch* untuk pembelajaran mesin, *Python* memungkinkan pengembangan algoritma yang dapat mendeteksi, menganalisis, dan menerjemahkan isyarat tangan yang ditangkap oleh kamera menjadi teks yang dapat dipahami. *Python* menyediakan alat yang diperlukan untuk membentuk jembatan antara bahasa isyarat visual dan komunikasi teks, yang sangat berguna dalam meningkatkan aksesibilitas bagi individu yang tuna rungu atau memiliki keterbatasan pendengaran.[12]



*OpenCV*, singkatan dari *Open Source Computer Vision Library*, adalah sebuah perangkat lunak yang dirancang untuk fokus pada pengembangan visi komputer, pemrosesan citra, dan algoritma pembelajaran mesin. Tujuan utamanya adalah untuk menyediakan infrastruktur yang luas bagi aplikasi visi komputer dan untuk mempercepat integrasi teknologi persepsi mesin dalam produk-produk komersial. Sebagai sebuah pustaka *open-source*, *OpenCV* tersedia di bawah lisensi BSD, memungkinkan penggunaan baik untuk proyek akademis maupun komersial.

*OpenCV* memfasilitasi penciptaan sistem yang bisa mendeteksi dan mengenali tangan, mengidentifikasi objek, mengklasifikasikan tindakan manusia dalam video, melacak pergerakan tangan, mengikuti objek yang bergerak, dan menggabungkan citra untuk menghasilkan gambar beresolusi tinggi dari seluruh adegan, yang sangat penting dalam proyek-proyek seperti translasi bahasa isyarat. Dengan adanya modul khusus yang dirancang untuk pemrosesan citra dan pembelajaran mesin, *OpenCV* menjadi alat yang tak ternilai untuk penulis dan pengembang di bidang visi komputer.[13]

#### **2.4.4 Long Short-Term Memory (LSTM)**

*Long short-term memory* (LSTM) telah menjadi alat yang ampuh dalam pembelajaran mendalam, menunjukkan performa luar biasa dalam berbagai tugas pemrosesan urutan data, seperti pengenalan suara, terjemahan mesin, dan pengenalan teks. Potensi LSTM dalam pengenalan aktivitas (AR) juga telah menarik perhatian para peneliti, menawarkan solusi yang menjanjikan untuk memahami dan menganalisis perilaku manusia. LSTM terdiri dari sel-sel khusus yang dirancang untuk mengatasi masalah menghilang dan meledaknya gradien yang sering dihadapi oleh jaringan saraf berulang (RNN) tradisional.

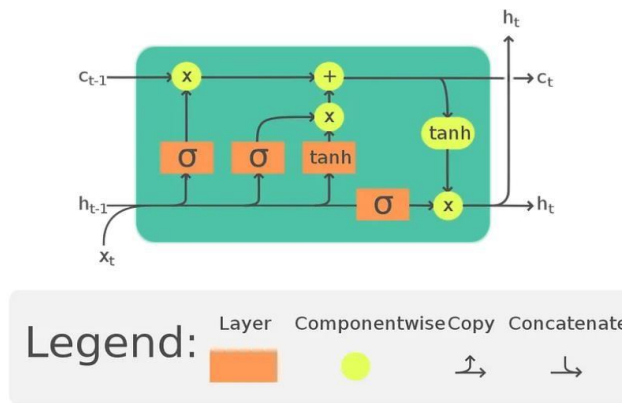
Dalam konteks *Action Recognition* LSTM dapat digunakan untuk menganalisis urutan data sensor, seperti akselerometer dan giroskop, untuk mengklasifikasikan berbagai jenis aktivitas fisik, seperti berjalan, berlari, duduk, dan berdiri. Jaringan LSTM dapat menangkap pola temporal yang kompleks dalam data sensor, memungkinkan mereka untuk membedakan dengan jelas antara aktivitas yang berbeda.

LSTM dirancang khusus untuk mengatasi dependensi jangka panjang dalam data urutan. Bahasa isyarat itu sendiri adalah rangkaian gerakan yang memiliki konteks temporal, sehingga LSTM sangat efektif dalam memahami dan mengingat gerakan yang terjadi dalam urutan tertentu. LSTM unggul dalam memproses data urutan dan time-series, menjadikannya cocok untuk memahami dan menerjemahkan bahasa isyarat yang tergantung pada konteks temporal dari satu gerakan ke gerakan berikutnya. Dalam bahasa isyarat, arti dari suatu gerakan sering kali dipengaruhi oleh gerakan sebelumnya atau gerakan yang datang setelahnya. LSTM memiliki arsitektur yang memungkinkan jaringan untuk mempertahankan informasi penting dari langkah-langkah sebelumnya dalam urutan, yang membantu dalam menerjemahkan bahasa isyarat dengan lebih akurat.

Salah satu tantangan utama dalam menerapkan LSTM untuk pengenalan aktivitas adalah ketersediaan data yang cukup untuk melatih model. Mengumpulkan dan memberilabel data sensor secara manual dapat memakan waktu dan mahal. Untuk mengatasi masalah ini, beberapa teknik telah dikembangkan, seperti sintesis data dan transfer belajar, untuk meningkatkan jumlah data pelatihan yang tersedia. Tantangan lain adalah menangani variasi dalam data sensor yang disebabkan oleh faktor-faktor seperti kondisi lingkungan, perangkat keras sensor, dan gaya individu. Untuk mengatasi variasi ini, teknik normalisasi dan augmentasi data dapat digunakan untuk meningkatkan ketahanan model terhadap variasi

data.[14][15]

Jaringan LSTM menawarkan solusi yang menjanjikan untuk pengenalan aktivitas, memungkinkan analisis yang akurat dan terperinci tentang perilaku manusia. Dengan mengatasi tantangan yang ada saat ini, LSTM berpotensi merevolusi cara kita memahami dan memantau aktivitas fisik, membuka jalan bagi aplikasi baru dalam bidang kesehatan, kebugaran, dan teknologi yang dapat dikenakan.



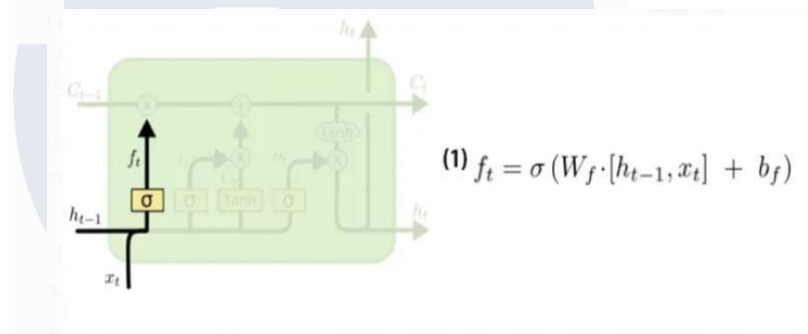
Gambar 2. 10 Arsitektur LSTM (Sumber: Artikel LSTM (Long Short Term Memory), 2023)

Sel LSTM memiliki struktur internal yang kompleks, dan LSTM terdiri dari komponen utama yaitu *Input gate*, *Forget gate*, dan *Output gate*, yang memungkinkan mereka untuk menyimpan dan memproses informasi secara efektif dari urutan data yang panjang. Karakteristik pada LSTM network terdapat pada layer yang tersembunyi dari apa yang disebut sel memori. Gerbang pada LSTM bekerja sebagai filter, dan masing-masing memiliki tujuan yang berbeda, Berikut adalah tahapan proses dalam LSTM:

1. *Forget Gate* (Gerbang Lupakan)

Forget gate berperan sebagai penanda informasi yang relevan untuk dipertahankan atau dihapus dalam memori sel berikutnya. Proses ini dimulai dengan menerima input dari keadaan tersembunyi sebelumnya serta informasi baru saat ini. Menggunakan fungsi sigmoid, informasi tersebut dinilai

dalam rentang antara 0 hingga 1, di mana nilai yang cenderung mendekati 0 menunjukkan informasi yang dihapus, sementara nilai yang cenderung mendekati 1 menunjukkan informasi yang dipertahankan. Pada langkah ini, unit LSTM menentukan informasi mana yang harus dihapus dari sel memori sebelumnya ( $h(t-1)$ ). Karenanya, nilai aktivasi  $f_t$  pada Forget gate  $t$  dihitung berdasarkan keluaran saat ini ( $x_t$ ), keadaan tersembunyi sebelumnya ( $h(t-1)$ ), dan keadaan sel memori dari langkah waktu sebelumnya ( $t-1$ ).[17]

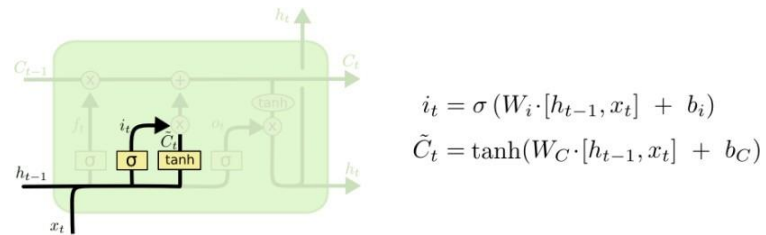


Gambar 2. 11 Fungsi Forget Gate (Sumber: Artikel Understanding LSTM Networks, 2015)

## 2. Input Gate (Gerbang Masukan)

*Input gate* memiliki tanggung jawab dalam menggabungkan informasi baru ke dalam sel memori. Proses ini melibatkan input dari keadaan tersembunyi sebelumnya dan informasi saat ini. Informasi tersebut kemudian diproses melalui fungsi *sigmoid* dan *tanh*. Fungsi *sigmoid* mengatur tingkat kepentingan informasi yang akan diperbarui, sementara fungsi *tanh* membantu sel memori untuk mengintegrasikan informasi baru dengan lebih efektif. Pada tahap ini, fokus utamanya adalah untuk memperbarui komponen blok input ( $x_t$ ) yang menyatukan input saat ini dengan keluaran dari unit LSTM sebelumnya ( $h(t-1)$ ). *Input*

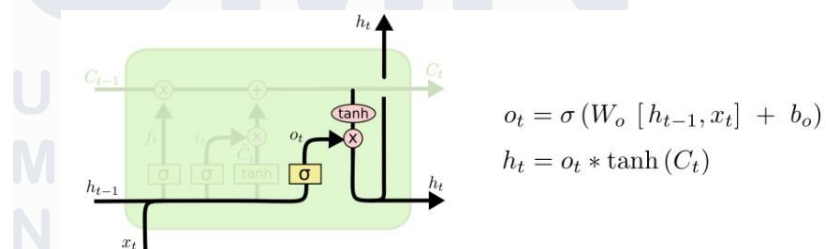
gate yang memadukan nilai sel saat ini ( $x_t$ ), keluaran dari unit LSTM sebelumnya ( $h(t-1)$ ), dan nilai sel sebelumnya ( $h(t-1)$ ) akan diperbarui.[18]



Gambar 2. 12 Fungsi Input Gate (Sumber: Artikel Understanding LSTM Networks, 2015)

### 3. Output Gate (Gerbang Keluaran)

*Output gate* bertugas menentukan bagaimana informasi yang tersimpan dalam sel memori akan dialirkan ke selanjutnya dalam rangkaian. Proses ini bergantung pada *input* dari keadaan tersembunyi sebelumnya dan informasi saat ini. *Input* tersebut kemudian diolah melalui fungsi *sigmoid*. Selanjutnya, sel memori baru dijalani melalui fungsi *tanh*. Hasil dari kedua proses tersebut dikalikan bersama untuk menghasilkan informasi yang akan disimpan dalam keadaan tersembunyi baru, yang akan diteruskan ke sel berikutnya dalam rangkaian. Pada langkah ini, nilai sel dihitung dengan menggabungkan input saat ini ( $x_t$ ), hasil dari LSTM sebelumnya ( $h(t-1)$ ), dan nilai sel pada iterasi sebelumnya ( $h_t$ ).



Gambar 2.13 Fungsi Output Gate (Sumber: Artikel Understanding LSTM Networks)

#### 2.4.5 *MediaPipe*

*MediaPipe* adalah sebuah kerangka kerja yang digunakan untuk membangun *pipeline* yang mampu melakukan inferensi terhadap data sensorik secara bebas. Dengan *MediaPipe*, *pipeline* dapat dibangun sebagai graf komponen modular, mencakup model inferensi, algoritma pemrosesan media, dan transformasi data. Data sensorik seperti data audio atau video yang masuk, akan diproses melalui *MediaPipe*, misalnya dengan localisasi objek untuk menghasilkan *output* seperti *landmark* tangan dari graf. *MediaPipe* menyediakan berbagai solusi sumber terbuka yang dapat digunakan, salah satunya adalah *MediaPipe Hands* yang memudahkan.[19]

#### 2.4.6 *Tensorflow*

*TensorFlow* merupakan sebuah kerangka kerja sumber terbuka untuk pembelajaran mendalam yang dikembangkan oleh *Google*. Selain itu, *TensorFlow* juga merupakan sebuah pustaka matematika simbolik yang digunakan untuk jaringan saraf dan sangat sesuai untuk pemrograman aliran data dalam berbagai bidang. *TensorFlow* menyediakan berbagai tingkat abstraksi untuk membangun dan melatih model. Selain itu, *TensorFlow* juga mampu mengonversi model menjadi format *tflite* dengan teknik *Quantization*, yang mengubah nilai-nilai *float* berbasis 64 pada model menjadi nilai-nilai integer. Model yang telah dilatih dan dikonversi menjadi *tflite* seharusnya dapat langsung digunakan untuk melakukan inferensi pada perangkat *smartphone*, karena format *tflite* memang ditujukan untuk melakukan inferensi pada perangkat *mobile*. [20]

#### 2.4.7 **Keras**

*Keras* merupakan sebuah antarmuka pemrograman tingkat tinggi (*high-level*) untuk jaringan saraf yang ditulis menggunakan bahasa pemrograman *Python*. *Library* sumber terbuka ini dirancang

untuk menyediakan lingkungan yang memungkinkan eksperimen cepat terhadap jaringan saraf mendalam, serta dapat dijalankan di atas berbagai *platform* seperti *Microsoft Cognitive Toolkit (CNTK)*, *TensorFlow*, dan *Theano*.

Dalam pembangunan model klasifikasi gerakan tangan, *Keras* digunakan dengan model *sequential* sebagai struktur utama modelnya. Model *sequential* ini menumpuk lapisan (*layer*), di mana setiap lapisan memiliki satu tensor masukan dan satu tensor keluaran.[20]

#### **2.4.8 Computer Vision**

*Computer vision* adalah cabang ilmu komputer yang fokus pada pengembangan sistem yang memungkinkan komputer untuk "melihat" dan memahami isi dari data visual. Proses ini melibatkan pengolahan dan analisis citra serta video untuk mendeteksi, mengklasifikasikan, dan melacak objek atau atribut tertentu. Dengan mengintegrasikan metode dari pengolahan citra, kecerdasan buatan, dan pembelajaran mesin, terutama melalui penggunaan Jaringan Saraf Tiruan dan Convolutional Neural Networks, Computer Vision berupaya untuk memberikan mesin kemampuan penglihatan yang serupa dengan manusia. Aplikasi dari teknologi ini sangat luas, mencakup dari pengenalan wajah, pengawasan video, hingga kendaraan otonom dan diagnosa medis. Meskipun telah mencapai kemajuan yang signifikan, Computer Vision masih terus berkembang dan menghadapi tantangan dalam memahami konteks kompleks dan beradaptasi dengan keadaan lingkungan yang dinamis. Kemampuannya untuk menginterpretasikan dan memahami dunia visual membuka cakrawala baru dalam inovasi teknologi dan aplikasi praktis di berbagai bidang.[21][22]

### 2.4.9 Dataset SIBI

Untuk mendukung penelitian dalam pengembangan sistem *computer vision* yang mampu menerjemahkan bahasa isyarat SIBI menjadi teks tulisan melalui kamera, diperlukan dataset yang sudah disesuaikan dengan batasan penelitian. Dataset ini berisi sekumpulan gambar yang menampilkan berbagai kata dalam bahasa isyarat SIBI.[23] Berikut adalah *dataset* SIBI yang digunakan;

1		“Saya” (bahasa isyarat SIBI)
2		“Kamu” (bahasa isyarat SIBI)
3		“Dia” (bahasa isyarat SIBI)
4		“Selamat” (bahasa isyarat SIBI)
5		“Pagi” (bahasa isyarat SIBI)



6	<p>Terima</p>	<p>“Terima” (bahasa isyarat SIBI)</p>
7	<p>Kasih</p>	<p>“Kasih” (bahasa isyarat SIBI)</p>
8	<p>Ya</p>	<p>“Ya” (bahasa isyarat SIBI)</p>
9	<p>Tidak</p>	<p>“Tidak” (bahasa isyarat SIBI)</p>
10	<p>Makan</p>	<p>“Makan” (bahasa isyarat SIBI)</p>
11	<p>Minum</p>	<p>“Minum” (bahasa isyarat SIBI)</p>
12	<p>Tidur</p>	<p>“Tidur” (bahasa isyarat SIBI)</p>

13		<p>“Mandi” (bahasa isyarat SIBI)</p>
14		<p>“Maaf” (bahasa isyarat SIBI)</p>

*Tabel 2. 2 Dataset SIBI yang Digunakan untuk Penelitian (Sumber: Olahan Penulis, 2024)*

