

## Pengenalan Kata Kolok Secara Real Time Menggunakan Mediapipe dan Algoritma Long Short-Term Memory (LSTM)

I Made Panji Prayoga<sup>1</sup>, I Gusti Ayu Agung Diatri Indradewi<sup>2</sup>, Putu Hendra Suputra<sup>3</sup>

<sup>1,2,3</sup>Sistem Informasi, Universitas Pendidikan Ganesha  
Jl. Udayana No.11, Banjar Tegal, Singaraja, Kabupaten Buleleng, Bali, Indonesia

e-mail: [panji.prayoga@undiksha.ac.id](mailto:panji.prayoga@undiksha.ac.id)<sup>1</sup>, [indradewi@undiksha.ac.id](mailto:indradewi@undiksha.ac.id)<sup>2</sup>, [hendra.suputra@undiksha.ac.id](mailto:hendra.suputra@undiksha.ac.id)<sup>3</sup>

Received : August, 2025

Accepted : August, 2025

Published : August, 2025

### Abstract

Sign language is the primary form of communication for individuals who are deaf or mute. In the Bengkulu community, there exists a unique sign language known as Kolok. Modern technology enables this local knowledge to be integrated into recognition systems. Building on this opportunity, the present study aims to develop a recognition system for Kolok Sign Language using the Long Short-Term Memory (LSTM) algorithm. The dataset was obtained directly from the Bengkulu Village community, collected through two methods: manual recording using a smartphone camera and real-time frame capture with OpenCV assisted by a webcam. Ten vocabulary categories were utilized in a dataset comprising 300 time-series videos, each consisting of 30 frames. MediaPipe was employed to analyze each frame and extract keypoint coordinates of body and hand positions. The features used for modeling were the x, y, and z coordinates of body poses and hand landmarks. The extracted data were then processed using a two-layer LSTM architecture. The training results demonstrated highly consistent performance. At epochs 500 and 550, the model achieved accuracies of 0.91 and 0.93, along with F1-scores of 0.92 and 0.93, respectively. The highest performance was recorded at epoch 450, with a precision of 0.95, recall of 0.96, and an F1-score of 0.94.

**Keywords:** Kolok Sign Language, MediaPipe, Long Short-Term Memory (LSTM), Deep Learning.

### Abstrak

Bahasa isyarat merupakan bentuk komunikasi utama bagi penyandang tuna rungu atau bisu. Di masyarakat Bengkulu, terdapat bahasa isyarat unik yang disebut Kolok. Teknologi modern memungkinkan pengetahuan lokal ini untuk diintegrasikan ke dalam sistem pengenalan teknologi. Penelitian ini, dengan memanfaatkan peluang ini, bertujuan untuk mengembangkan sistem pengenalan bahasa isyarat Kolok menggunakan algoritma Long Short-Term Memory (LSTM). Dataset diperoleh langsung dari masyarakat Desa Bengkulu, yang dikumpulkan melalui dua metode, yaitu perekaman manual menggunakan kamera smartphone dan pengambilan frame secara real-time dengan OpenCV berbantuan webcam. Sepuluh kategori kosakata digunakan dalam dataset yang berisi 300 video deret waktu, yang masing-masing terdiri dari 30 bingkai. MediaPipe digunakan untuk menganalisis setiap bingkai dan memperoleh koordinat kunci posisi tubuh dan tangan. Fitur yang digunakan untuk pemodelan adalah koordinat x, y, dan z dari postur dan landmark tangan. Data yang diekstraksi kemudian diproses menggunakan arsitektur MLTP dua lapis. Hasil pelatihan menunjukkan kinerja yang sangat konsisten. Pada epoch 500 dan 550, model mencapai akurasi 0,91 dan 0,93, bersama dengan skor F1 masing-masing 0,92 dan 0,93. Kinerja tertinggi tercatat pada epoch 450, dengan presisi 0,95, recall 0,96, dan skor F1 0,94.

**Kata Kunci:** Bahasa Isyarat Kolok, MediaPipe, Long Short-Term Memory (LSTM), Deep Learning.

## 1. PENDAHULUAN

Bahasa merupakan sarana penting dalam menyampaikan informasi dan menjalin interaksi sosial antarmanusia. Namun, bagi mereka yang menghadapi hambatan komunikasi, seperti penyandang tuna rungu atau sulit mendengar, bahasa isyarat menjadi metode komunikasi yang paling penting.

Bahasa isyarat tidak memiliki standar universal dan bervariasi di tiap negara dan tiap daerah. Strukturnya juga dipengaruhi konteks geografis, tidak selalu berurutan seperti bahasa lisan [1]. Dalam konteks ini, pengenalan kosakata bahasa isyarat memegang peranan penting dalam mendukung terciptanya sistem komunikasi yang inklusif bagi mereka.

Gangguan bicara digambarkan sebagai situasi di mana seseorang menghadapi tantangan dalam mengungkapkan gagasannya secara verbal, baik karena keterbatasan fisik maupun keterlambatan perkembangan bahasa. Kondisi ini dapat disebabkan oleh berbagai penyebab, seperti masalah sensorik, kelainan genetik, atau komplikasi neurologis. Menurut data [2] Kementerian Koordinator Bidang Pembangunan Manusia dan Kebudayaan (Kemendikbud), jumlah penyandang disabilitas di Indonesia diproyeksikan mencapai 22,97 juta jiwa pada tahun 2023, setara dengan hampir 8,5% dari total populasi. Berdasarkan data tersebut menunjukkan pentingnya memberikan perhatian dan dukungan khusus terhadap kebutuhan komunikasi bagi mereka, termasuk melalui pengembangan teknologi dan layanan yang memadai.

Penelitian ini berlangsung di kota Bengala, yang terletak di distrik Kubutambahan, wilayah Buleleng, provinsi Bali. Daerah ini terkenal sebagai rumah bagi komunitas Kolok, yang terdiri dari penyandang disabilitas pendengaran dan wicara. Komunitas ini menggunakan bahasa isyarat khusus yang disebut "Bahasa Kolok". Sebagai warisan budaya, bahasa ini berbeda dari BISINDO maupun SIBI, dengan ciri utama pada struktur gerakan yang unik.

Hasil wawancara dengan tokoh masyarakat dan tenaga pengajar di desa menunjukkan bahwa para penyandang disabilitas menghadapi kendala komunikasi ketika berinteraksi dengan orang luar. Untuk itu, desa menetapkan kebijakan menyediakan pemandu komunikasi. Hal ini menunjukkan bahwa meskipun Bahasa Kolok efektif dalam komunitasnya, tetap

diperlukan solusi untuk menjembatani komunikasi lintas kelompok.

Untuk menjawab tantangan ini, teknologi pengenalan bahasa isyarat (sign language recognition) menjadi salah satu pendekatan potensial [3]. Teknologi ini memungkinkan sistem mengenali dan memahami gerakan tubuh, termasuk tangan sehingga komunikasi dapat dilakukan secara lebih efektif.

Dalam beberapa tahun ke belakang, teknologi di bidang pengenalan Bahasa isyarat sudah mengalami kemajuan, karena integrasi teknik pemrosesan dan deep learning. Berdasarkan hal tersebut penelitian [4] mengembangkan model CNN-BiLSTM end-to-end untuk pengenalan dan terjemahan bahasa isyarat dengan akurasi 95%. Penelitian ini menunjukkan bahwa penggunaan deep learning sangat mendukung komunikasi bagi penyandang disabilitas.

Karya ini berkontribusi pada pengembangan sistem yang mengidentifikasi bahasa isyarat lokal, khususnya bahasa isyarat Kolok yang digunakan oleh penduduk Desa Bengkala di Bali. Berbeda dengan studi sebelumnya seperti yang dilakukan oleh [5] yang menitikberatkan pada pengenalan Bahasa Isyarat Indonesia (SIBI) berskala nasional, penelitian ini lebih mengedepankan pelestarian dan pemanfaatan bahasa isyarat khas daerah yang hanya dikenal dan digunakan dalam komunitas tertentu. Pendekatan ini tidak hanya mendorong penerapan teknologi yang inklusif bagi penyandang disabilitas, tetapi juga memperkuat upaya pelestarian budaya lokal dengan mengintegrasikan teknologi modern dan kearifan lokal dalam satu kesatuan.

Pada penelitian [6] mengklasifikasikan 30 kosakata BISINDO menggunakan CNN dan MobileNetV2 dengan akurasi 54,8%, namun disarankan penggunaan LSTM untuk deteksi isyarat dinamis. Sementara itu, Penelitian yang dilakukan [7] mendeteksi 5 kata ASL menggunakan MediaPipe dan LSTM, menghasilkan akurasi yang cukup tinggi.

Dalam penelitian serupa yang menggunakan algoritma LSTM [8] dalam pengenalan gerakan yoga menunjukkan bahwa kombinasi MediaPipe dan LSTM memberikan hasil akurasi tinggi dalam memahami urutan gerak. Dengan kemampuan LSTM dalam memahami dinamika gerakan dan dukungan ekstraksi fitur dari MediaPipe, diharapkan sistem yang dikembangkan mampu mengenali kosakata dalam Bahasa Kolok secara akurat dan real-

time. Oleh karena itu, algoritma LSTM sangat cocok digunakan untuk mengenali data sekuensial, karena mampu mempertahankan konteks temporal dari gerakan secara berurutan.

Peneliti berencana untuk melakukan penelitian berjudul "Pengenalan Kata Kolok Real-Time secara real time Menggunakan MediaPipe dan Algoritma Long Short Term Memory (LSTM)." Tujuan penelitian ini adalah menciptakan sistem yang mampu mengidentifikasi dan memproses kata kolok dengan akurasi tinggi secara real-time, serta mengevaluasi efektivitas modelnya.

## 2. KAJIAN TEORI

### 2.1 Tuna Rungu dan Tuna wicara

Tuna rungu merupakan sebutan bagi kondisi ketika seseorang mengalami hambatan pendengaran, yang dapat terjadi secara parsial ataupun total [9]. Hal ini dapat terjadi karena berbagai alasan, seperti cacat lahir, cedera fisik, atau penyakit tertentu yang memengaruhi kemampuan pendengaran. Sedangkan tuna wicara adalah kondisi di mana seseorang menghadapi kesulitan untuk mengungkapkan pikiran atau ide-ide mereka melalui bahasa lisan atau verbal. Hal ini menyebabkan kesulitan bagi mereka untuk dipahami oleh orang lain, sehingga mereka cenderung menggunakan bahasa non-verbal, seperti gerakan tubuh, mimik wajah, atau gambar, untuk berkomunikasi [10]. Dalam kondisi ini, komunikasi seringkali memerlukan pendekatan alternatif yang lebih visual.

### 2.2 Bahasa Isyarat (Bahasa Kolok)

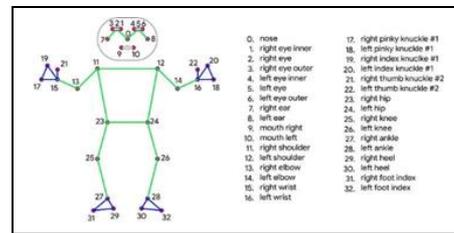
Beberapa faktor dapat menyebabkan hal ini, seperti cacat lahir, trauma, atau beberapa penyakit yang memengaruhi kemampuan mendengar. Keberadaannya membedakan diri dari SIBI dan BISINDO karena berkembang secara organik dalam komunitas, tanpa mengikuti standar formal.

Keunikannya terletak pada gerakan yang lebih fleksibel dan ekspresif, dipengaruhi oleh budaya lokal, serta penggunaan ekspresi wajah dan tubuh yang intens. Hal ini menjadikan Bahasa Kolok tidak hanya fungsional, tetapi juga kaya akan nilai historis dan sosial.

### 2.3 Mediapipe

MediaPipe adalah sistem sumber terbuka yang dibuat oleh Google. Tujuannya adalah untuk membantu menciptakan aplikasi yang memproses multimedia, termasuk video, audio,

dan gambar, menggunakan pembelajaran mesin dan visi komputer. Contoh MediaPipe dapat dilihat pada Gambar 1.



Gambar 1. Pose Landmark

Kelebihan Mediapipe Menurut penelitian [11], MediaPipe terbukti mampu mendeteksi gerakan tangan yang kompleks dengan akurasi tinggi secara real-time.

Dalam beberapa tahun terakhir mediapipe menjadi pilihan dalam penelitian pengenalan Bahasa isyarat untuk ekstraksi fitur, kemudian dikombinasikan dengan beberapa metode pendukung [12]. Contohnya pada penelitian [13] dengan menggabungkan MediaPipe dan LSTM, penelitian ini berhasil mengenali alfabet ASL secara akurat melalui deteksi landmark real-time dan analisis temporal dari LSTM.

### 2.4 Open CV

OpenCV, singkatan dari Open Computer Vision, adalah pustaka kode sumber yang mudah diakses dan dirancang untuk pemrosesan gambar instan. Pustaka ini menawarkan lebih dari 2.500 algoritma yang dapat diterapkan pada program visi komputer dan pembelajaran mesin. Algoritma OpenCV dirancang untuk tugas-tugas spesifik, seperti mendeteksi dan mengenali wajah, mengklasifikasikan jenis kelamin, mengidentifikasi gestur tangan, mengenali objek, dan banyak lagi [14]. OpenCV ditulis dalam C++ untuk machine learning, namun memiliki pembungkus Python, oleh karena itu ia memiliki API untuk Python untuk machine learning [15].

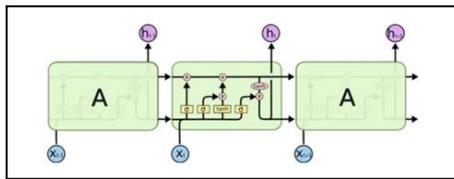
### 2.5 Long Short-Term Memory

Memori jangka pendek panjang (LSTM) merupakan penyempurnaan model jaringan saraf berulang (RNN), yang dirancang untuk menyimpan data dalam jangka waktu lama. Pada model ini, node pada lapisan tersembunyi RNN digantikan oleh sel LSTM yang berfungsi menyimpan dan mengelola informasi dari waktu sebelumnya.

Kemampuan LSTM dalam menangkap dependensi temporal menjadikannya unggul untuk memodelkan urutan gerakan bahasa

isyarat [16]. Meski demikian, kekurangannya pada aspek fitur spasial sering memerlukan dukungan model lain atau tahap pra-pemrosesan [17].

Menurut penelitian [18], LSTM memiliki cell state, yaitu memori jaringan yang mampu menyimpan informasi untuk jangka waktu panjang maupun pendek. Selain itu, LSTM juga memiliki memory block yang berfungsi untuk memilih nilai yang paling relevan sebagai keluaran berdasarkan masukan yang diterima. Inilah salah satu keunggulan utama dari LSTM. Gambar 2 menunjukkan diagram LSTM.



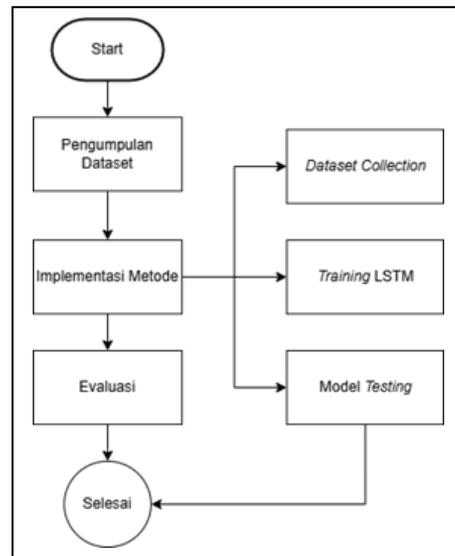
Gambar 2. Skema LSTM

### 2.6 Evaluasi

Evaluasi adalah proses untuk mengukur performa model dalam mengklasifikasikan data secara akurat. Penelitian ini memanfaatkan confusion matrix sebagai instrumen evaluasi. Confusion matrix merupakan tabel yang digunakan dalam sistem klasifikasi untuk membandingkan kelas aktual dengan kelas yang diprediksi oleh model. Melalui tabel ini dapat diketahui jumlah prediksi benar maupun salah pada setiap kelas, sehingga kinerja model dapat dianalisis secara lebih detail.

## 3. METODE PENELITIAN

Penelitian ini dilakukan dalam beberapa tahap yang berfokus pada identifikasi bahasa isyarat Kolok menggunakan pustaka OpenCV dan MediaPipe. Pustaka ini digunakan untuk menemukan dan mengambil titik kunci dari setiap frame dalam video postur dan tangan. Data yang diperoleh diproses menggunakan algoritma memori jangka panjang (LSTM) untuk mengembangkan model yang dapat mengenali gestur secara akurat. Proses penelitian ditunjukkan secara lebih rinci pada Gambar 3.



Gambar 3. Metode Penelitian

### 3.1 Dataset Collection

Dataset untuk Tugas Akhir ini dikumpulkan dari masyarakat Desa Bengkala dan terdiri dari berbagai pose. Pengambilan data dilakukan dengan dua metode, yaitu menggunakan kamera smartphone dan melalui proses pengambilan secara langsung menggunakan OpenCV dengan bantuan webcam. Kamera smartphone digunakan untuk merekam foto dari berbagai pose secara manual. OpenCV digunakan untuk pengambilan frame secara real-time menggunakan webcam. Pendekatan ini digunakan oleh para peneliti untuk menghasilkan data video terorganisir yang memfasilitasi analisis posisi secara mendalam. Posisi bahasa isyarat yang digunakan dalam penelitian ini mencakup sepuluh kata: "minum", "terima kasih", "telepon", "ayah", "ibu", "uang", "lapar", "teman", "makan", dan "maaf". Pengambilan dataset dapat dilihat pada gambar 4.

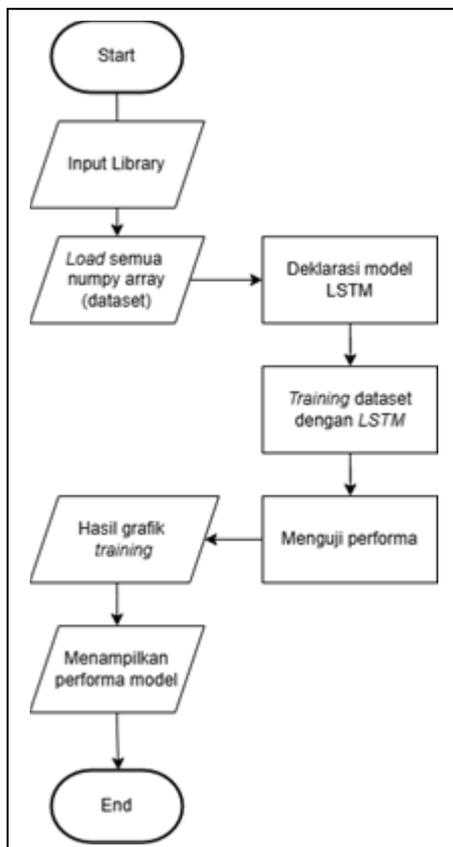


Gambar 4. Pengambilan dataset menggunakan smartphone dan Open CV

### 3.2 Training Model

Dalam penelitian ini, dua percobaan dilakukan untuk mengidentifikasi struktur jaringan memori jangka panjang (LSTM) yang akan diterapkan selama pelatihan. Percobaan pertama menggunakan arsitektur dengan tiga lapisan LSTM, sedangkan percobaan kedua menggunakan dua lapisan LSTM. Kedua konfigurasi ini dirancang untuk dibandingkan performanya dalam mengklasifikasikan pose berdasarkan dataset yang telah dikumpulkan. Arsitektur dengan hasil terbaik dari kedua percobaan tersebut akan digunakan sebagai model utama dalam penelitian ini.

Berikut merupakan alur dari proses training menggunakan metode Long Sort Term Memory dapat dilihat pada gambar 5:



Gambar 5. Alur proses training

Proses dimulai dengan memuat library yang diperlukan, seperti TensorFlow atau Keras untuk implementasi model LSTM, serta NumPy untuk manipulasi data. Langkah berikutnya adalah memuat dataset yang telah dikumpulkan dan disimpan dalam format array NumPy. Dataset ini berisi data landmark dari

pose yang akan digunakan untuk melatih model. Setelah dataset dimuat, deklarasi model LSTM dilakukan, yang mencakup konfigurasi arsitektur jaringan, termasuk jumlah lapisan LSTM, neuron per lapisan, dan fungsi aktivasi yang digunakan. Proses pelatihan dilakukan menggunakan 80% dataset, sementara 20% sisanya digunakan untuk pengujian guna mengevaluasi performa model berdasarkan variabel uji seperti akurasi dan F1-score.

### 3.3 Proses Testing

Proses testing merupakan tahapan pengujian sistem atau model pengenalan bahasa isyarat secara real-time. Dalam proses ini, webcam berfungsi sebagai perangkat input yang menangkap gambar atau video dari pose tubuh pengguna yang melakukan bahasa isyarat. Berikut ini adalah tahapan proses pengujian yang diilustrasikan pada Gambar 6:



Gambar 6. Alur proses testing

### 3.4 Evaluasi

Setelah proses training selesai, tahapan evaluasi dengan menggunakan Confusion Matrix dilakukan pada tahap akhir. Evaluasi ini bertujuan untuk menilai kinerja model dalam mengenali dan mengklasifikasikan pose bahasa isyarat Kolok. Prediksi kelas dilakukan di mana hasil prediksi model terhadap dataset pengujian dikumpulkan dan dibandingkan

dengan label sebenarnya dari data pengujian. Kemudian, Confusion Matrix dibangun menggunakan hasil prediksi dan label sebenarnya. Selanjutnya, berbagai metrik evaluasi dihitung berdasarkan Confusion Matrix untuk menilai performa model.

#### 4. HASILDAN PEMBAHASAN

##### 4.1 Dataset Collection

Selama fase pengumpulan data, data dikumpulkan dan diorganisasikan ke dalam 10 kategori: “minum,” “terima kasih,” “telepon,” “ayah,” “ibu,” “uang,” “lapar,” “teman,” “makan,” dan “maaf.” Video pose direkam menggunakan webcam dengan bantuan library OpenCV secara real-time. Ekstraksi keypoint dari pose tubuh dan tangan dilakukan menggunakan MediaPipe, yang menghasilkan total 258 keypoint per frame. Penelitian ini hanya menggunakan data primer berupa pose bahasa isyarat yang telah disepakati. Dengan 10 kelas dan masing-masing 30 video berdurasi 30 frame, total data yang diperoleh adalah 300 video atau 9.000 frame.

Hasil dari proses ekstraksi keypoint disimpan dalam format numpy array yang ditunjukkan pada gambar 7. Data hasil ekstraksi ini, merepresentasikan koordinat berbagai titik tubuh dan tangan dalam satu frame, yang dideteksi menggunakan MediaPipe. Setiap titik memiliki empat komponen utama, yaitu koordinat tiga dimensi (x, y, dan z) serta nilai visibility.

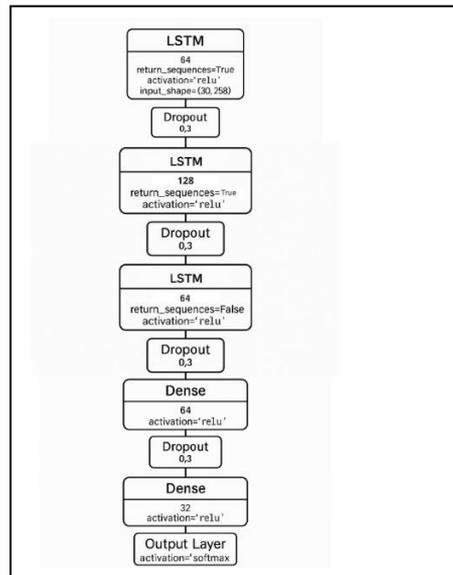
```
[ 0.52273214, 0.25789657, -0.37912351, 0.99981523, 0.53118181, 0.22549878, -0.36116987,
0.99939543, 0.53753718, 0.22468945, -0.36104311, 0.99943483, 0.54353696, 0.22381473,
-0.36111097, 0.99929488, 0.50778524, 0.22290817, -0.36031313, 0.99903602, 0.49656298,
0.22483905, -0.36796832, 0.99968964, 0.48823366, 0.22356528, -0.36828029, 0.99967134,
0.54951918, 0.23112176, -0.17887637, 0.99949938, 0.47273856, 0.23337586, -0.21232869,
0.99981832, 0.53518387, 0.28959826, -0.38456448, 0.99988508, 0.56438357, 0.28918958,
-0.31379834, 0.99992973, 0.68874945, 0.36482725, -0.86194474, 0.99994688, 0.48584479,
0.38593688, -0.88528461, 0.99993718, 0.78925615, 0.42643284, -0.17974725, 0.99973661,
0.37322769, 0.62659723, 0.84145425, 0.99336329, 0.64118688, 0.27778823, -0.38562717,
0.99797744, 0.35876359, 0.84697338, -0.11856639, 0.97816521, 0.66453185, 0.24591637,
-0.35542423, 0.98375976, 0.33429322, 0.91807646, -0.13463558, 0.94468489, 0.58884947,
0.24951471, -0.20369238, 0.30873848, 0.34659528, 0.92282999, -0.18249144, 0.94882578,
0.59173477, 0.26165384, -0.28882584, 0.98968417, 0.35613766, 0.98311396, -0.13624386,
0.94496888, 0.56782591, 0.84923488, 0.81137843, 0.99138642, 0.44787179, 0.84527238,
-0.81875437, 0.99919551, 0.55619693, 1.18318026, 0.81818455, 0.29238697, 0.42178853,
1.16827534, 0.81621189, 0.38284366, 0.55591691, 1.45751178, 0.32633948, 0.82313265,
0.43861961, 1.44299543, 0.23931785, 0.82967538, 0.55395329, 1.49369498, 0.34959155,
0.82198488, 0.43682367, 1.47884811, 0.25651386, 0.82152521, 0.54848496, 1.55487621,
0.16588699, 0.82251597, 0.46479811, 1.54668499, 0.82325438, 0.81472774, 0.65883122,
0.26768519, -0.88808817, 0.63888371, 0.24128419, -0.88808491, 0.62084262, 0.21888973,
-0.88226583, 0.66495284, 0.28814675, -0.88518842, 0.59757578, 0.18286872, -0.88755596,
0.58418489, 0.22546175, -0.88646584, 0.58883729, 0.24617551, -0.81822831, 0.57688417,
0.25697401, -0.81876774, 0.58695287, 0.25687882, -0.81888771, 0.59352529, 0.24499196,
-0.88899664, 0.57872983, 0.26474264, -0.81335883, 0.58125639, 0.27177447, -0.81317496,
0.59855275, 0.26921861, -0.81311488, 0.59588856, 0.26528824, -0.81118788, 0.57188883,
0.28223777, -0.81569466, 0.58166875, 0.28698153, -0.81421884, 0.59134888, 0.28345868,
-0.81319427, 0.59344288, 0.28426488, -0.81388496, 0.57986444, 0.38191255, -0.81753978,
0.58926686, 0.31548489, -0.81893988, 0.55916448, 0.32512671, -0.81869576, 0.34238833,
0.85585786, -0.88808884, 0.35629484, 0.87486196, -0.82187344, 0.35485529, 0.98545988,
-0.82571512, 0.34957436, 0.93444718, -0.82468246, 0.34677753, 0.95538892, -0.82181196,
0.33116588, 0.92883288, -0.81887781, 0.33755445, 0.95872235, -0.81883137, 0.34974387,
0.97376784, -0.82138917, 0.35992175, 0.97948242, -0.82164376, 0.32759696, 0.91998334,
-0.88513896, 0.33613244, 0.95568627, -0.88529761, 0.34866741, 0.97332889, -0.88684443,
0.36145523, 0.98111463, -0.88843867, 0.32988823, 0.91732955, 0.88858483, 0.31681467,
0.94955665, 0.88547817, 0.34887783, 0.96283853, 0.88284232, 0.35856855, 0.97813927,
0.88676429, 0.33245224, 0.91588955, 0.81698983, 0.33942418, 0.93838669, 0.81451257,
0.34782639, 0.94819138, 0.81818181, 0.35882522, 0.95417588, 0.81818181]
```

Gambar 7. Hasil ekstraksi keypoint

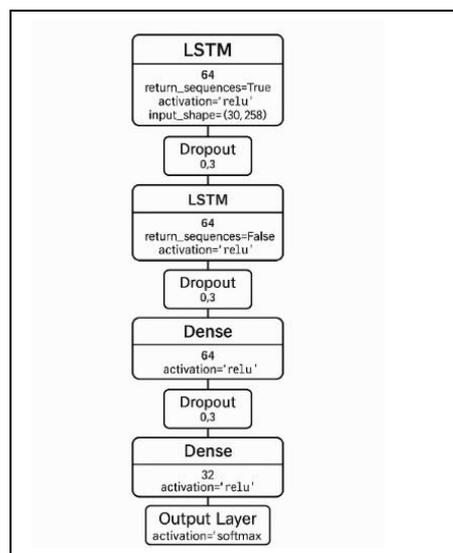
##### 4.2 Training Model

Dalam penelitian ini, model memori jangka pendek panjang (LSTM) yang dikembangkan dengan pustaka TensorFlow oleh para peneliti digunakan. Sebagai bagian dari proses

eksplorasi model, peneliti juga melakukan perbandingan antara dua konfigurasi arsitektur jaringan, yaitu arsitektur dengan tiga lapisan LSTM dan arsitektur dengan dua lapisan LSTM yang sama sama dilengkapi dengan dropout [19]. Dalam proses pelatihan model LSTM pada kedua arsitektur tersebut, diterapkan teknik dropout. Untuk arsitekturnya dapat dilihat pada gambar 8 dan 9.



Gambar 8. Arsitektur 3 layer



Gambar 9. Arsitektur 2 layer

Selanjutnya proses training menggunakan arsitektur LSTM yang telah dirancang. Para ilmuwan melakukan proses pelatihan tiga kali

untuk mencapai hasil yang lebih stabil dan representatif. Set data dibagi menjadi rasio 80:20, dengan 80% digunakan untuk pelatihan dan 20% untuk pengujian. Dalam penelitian ini, hanya menggunakan parameter epoch sebagai waktu training model.

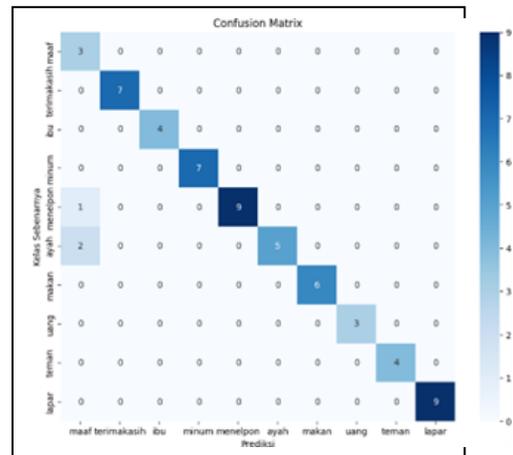
### 4.3 Evaluasi

Setelah proses pelatihan model selesai dilakukan, langkah selanjutnya adalah melakukan evaluasi terhadap model tersebut menggunakan confusion matrix. Evaluasi ini bertujuan untuk mengukur performa model dalam mengklasifikasikan data pada tahap pengujian.

Percobaan yang dilakukan mengikuti tahapan serupa seperti yang telah dijelaskan sebelumnya, yaitu dimulai dengan menggunakan arsitektur tiga lapisan LSTM serta variasi jumlah epoch sebanyak 450, 500, dan 550 kemudian menggunakan 2 layer LSTM dengan jumlah epoch yang saya untuk menentukan jumlah epoch yang memberikan hasil paling optimal selama proses pelatihan model. Selanjutnya, peneliti juga menerapkan arsitektur dengan dua lapisan LSTM untuk membandingkan performanya, guna mengetahui konfigurasi arsitektur yang menghasilkan kinerja terbaik dalam konteks penelitian ini. Hasil evaluasi ini menunjukkan tingkat akurasi model sebagaimana disajikan pada tabel 1 berikut.

Tabel 1: Hasil training

Jumlah layer	Jumlah epoch	acc	precis ion	recall	F1 Score
3 layer	450	0.82	0.84	0.84	0.82
	500	0.55	0.51	0.48	0.46
	550	0.68	0.78	0.75	0.66
2 layer	450	0.95	0.95	0.96	0.94
	500	0.91	0.93	0.93	0.92
	550	0.93	0.94	0.95	0.93



Gambar 10. Hasil evaluasi tertinggi dengan epoch 450 dengan 2 layer LSTM dengan ak

Berdasarkan hasil akurasi dari evaluasi diatas, penelitian ini mampu mencapai akurasi sebesar 95%. Hasil ini menunjukkan bahwa versi dengan 450 epoch dan arsitektur LSTM dua lapis menghasilkan hasil terbaik dibandingkan konfigurasi lainnya. Lihat Tabel 1 untuk evaluasi keseluruhan model yang dihasilkan

Jika dibandingkan dengan penelitian yang dilakukan oleh [20]. Dari sisi akurasi, memperoleh hasil sebesar 65%, menggunakan 2 layer LSTM namun hasil masih lebih rendah dibandingkan pada penelitian ini. Perbedaan ini dapat disebabkan oleh jumlah data latih, kompleksitas gerakan pada bahasa isyarat Kolok, serta variasi ekspresi gerakan tangan. Namun demikian, kedua penelitian menunjukkan bahwa LSTM efektif dalam memproses data sekuensial untuk sistem pengenalan bahasa isyarat secara real-time.

### 4.4 Testing real Time

Model yang telah selesai melalui proses training disimpan dalam format file dengan ekstensi .h5, yang merupakan format umum untuk menyimpan model dalam Keras. Model tersebut kemudian digunakan untuk melakukan pengujian atau prediksi waktu nyata. Peneliti memilih jumlah epoch dengan akurasi tertinggi untuk digunakan dalam proses pengujian. Berikut merupakan hasil dari testing real time:

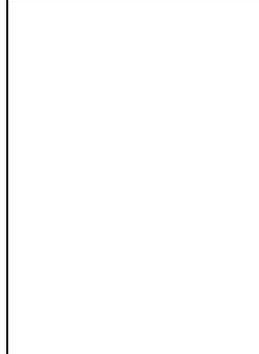
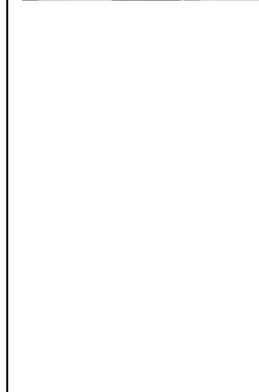
Tabel 2: Pose berhasil terdeteksi

Gambar	Isyarat	Hasil deteksi
	Terimakasih	Berhasil mendeteksi kosakata "terimakasih".
	Ibu	Berhasil mendeteksi kosakata "ibu".
	Lapar	Berhasil mendeteksi kosakata "lapar".
	Teman	Berhasil mendeteksi kosakata "teman".
	Uang	Berhasil mendeteksi kosakata "uang".
	Makan	Berhasil mendeteksi kosakata "makan".

Temuan deteksi yang disajikan pada Tabel 2 menunjukkan bahwa sistem mengidentifikasi dan menyajikan postur Bahasa Isyarat Kolok secara akurat. Hal ini menunjukkan bahwa model yang dibuat beroperasi secara efisien dalam kategori yang ditentukan.

Tabel 3: Pose tidak berhasil terdeteksi

Gambar	Isyarat	Hasil deteksi
	Maaf	Berhasil mendeteksi kosakata "maaf", namun juga terjadi

		kesalahan deteksi terhadap kosakata "ayah" akibat kemiripan pose antara kedua gerakan tersebut.
	Menelpon	Tidak berhasil mendeteksi.
	Minum	Berhasil mendeteksi kosakata "minum", namun juga terjadi kesalahan deteksi terhadap kosakata "ayah, makan" akibat kemiripan pose antara kedua gerakan tersebut.
	Ayah	Berhasil mendeteksi kosakata "ayah", namun juga terjadi kesalahan deteksi terhadap kosakata "maaf" akibat kemiripan pose antara kedua gerakan tersebut.

Tabel 3 menunjukkan bahwa terdapat beberapa pose yang tidak berhasil terdeteksi

dengan baik oleh sistem. Selain itu, terdapat pula beberapa pose yang berhasil terdeteksi namun menunjukkan kemiripan dengan pose lain, sehingga menyebabkan kesalahan dalam proses klasifikasi. Hal ini mengindikasikan bahwa model masih menghadapi tantangan dalam membedakan pose-pose yang memiliki pola atau bentuk yang mirip.

#### 4.5 Pembahasan

Hasil pengujian menunjukkan bahwa arsitektur tiga lapisan LSTM menghasilkan performa yang kurang stabil. Hal ini disebabkan oleh kompleksitas model yang terlalu tinggi sehingga memicu overfitting, membuat model kesulitan melakukan generalisasi pada data baru. Sebaliknya, arsitektur dua lapisan LSTM memberikan performa yang lebih stabil dan konsisten, dengan akurasi terbaik mencapai 0,95 pada epoch ke-450. Kondisi ini mengindikasikan bahwa penggunaan arsitektur dengan kompleksitas moderat lebih sesuai untuk dataset yang terbatas, karena mampu menyeimbangkan antara kedalaman model dan kemampuan generalisasi.

Implikasi dari temuan ini adalah bahwa kombinasi MediaPipe dan LSTM dapat digunakan secara efektif untuk mendeteksi bahasa isyarat Kolok secara real-time. Selain itu, hasil ini menunjukkan potensi pengembangan sistem penerjemah berbasis deep learning yang dapat membantu meningkatkan aksesibilitas komunikasi bagi penyandang tuna rungu dan tuna wicara.

## 5. KESIMPULAN

Penelitian ini berhasil menciptakan sistem yang mengidentifikasi kata secara real-time dalam bahasa Kolok, menggunakan MediaPipe dan algoritma memori jangka pendek (LSTM). Sistem yang dikembangkan menggunakan arsitektur dua lapis LSTM dengan lapisan dropout dan dense, serta data input berupa keypoint tubuh hasil ekstraksi dari MediaPipe. Dengan dataset sebanyak 300 video dan konfigurasi pelatihan terbaik pada 450 epoch, model mampu mencapai akurasi tinggi. Hasil evaluasi menunjukkan bahwa sistem berkinerja baik dalam klasifikasi pose, mencapai akurasi 0,95, recall 0,96, dan skor F1 0,94. Evaluasi kinerja menggunakan matriks konfusi memverifikasi efektivitas pendekatan ini dalam mengidentifikasi kata-kata Kolok secara akurat dan efisien.

## DAFTAR PUSTAKA

- [1] D. K. Jain, A. Kumar, and S. R. Sangwan, "TANA: The amalgam neural architecture for sarcasm detection in Indian indigenous language combining LSTM and SVM with word-emoji embeddings," *Pattern Recognition Letters*, vol. 160, pp. 11–18, Aug. 2022, doi: 10.1016/j.patrec.2022.05.026.
- [2] Kementerian Koordinator Bidang Pembangunan Manusia dan Kebudayaan, "Pemerintah Penuhi Hak Penyandang Disabilitas di Indonesia," Kementerian Koordinator Bidang Pembangunan Manusia dan Kebudayaan. [Daring]. Tersedia: <https://www.kemenkopmk.go.id/pemerintah-penuhi-hak-penyandang-disabilitas-di-indonesia>. [Diakses: 24 Juni 2024].
- [3] A. Aljabar and Suharjo, "Bisindo (Bahasa Isyarat Indonesia) Sign Language Recognition Using CNN and LSTM," *Advances in Science, Technology and Engineering Systems*, vol. 5, no. 5, pp. 282–287, 2020, doi: 10.25046/aj050535.
- [4] B. Natarajan et al., "Development of an End-to-End Deep Learning Framework for Sign Language Recognition, Translation, and Video Generation," in *IEEE Access*, vol. 10, pp. 104358–104374, 2022, doi: 10.1109/ACCESS.2022.3210543.
- [5] L. Arisandi and B. Satya, "Sistem klarifikasi bahasa isyarat Indonesia (Bisindo) dengan menggunakan algoritma convolutional neural network," *Jurnal Sistem Cerdas*, vol. 5, no. 3, pp. 135–146, 2022, doi: 10.37396/jsc.v5i3.262.
- [6] N. H. A. E. and M. I. Zul, "Aplikasi Penerjemah Bahasa Isyarat Indonesia menjadi Suara berbasis Android menggunakan Tensorflow," *Jurnal Komputer Terapan*, vol. 7, no. 1, pp. 74–83, 2021. doi: 10.35143/jkt.v7i1.4629.
- [7] S. Ghosh, "Proposal of a Real-time American Sign Language Detector using MediaPipe and Recurrent Neural Network," *International Journal of Computer Sciences and Engineering*, vol. 9, no. 7, pp. 46–52, 2021, doi: 10.26438/ijcse/v9i7.4652.
- [8] F. Daniel Tanugraha, H. Pratikno, M. Musayanah, and W. Indah Kusumawati, "Pengenalan Gerakan Olahraga Berbasis (Long Short-Term Memory) Menggunakan

- Mediapipe," *Journal of Advances in Information and Industrial Technology*, vol. 4, no. 1, pp. 37–45, 2022, doi: 10.52435/jaiit.v4i1.182.
- [9] A. Nurmansyah, N. R. Rhamadhani, S. A. N. Hakim, S. A. Agustin, and S. Hamidah, "Permasalahan komunikasi yang kerap terjadi pada penyandang disabilitas," *Jurnal Pendidikan, Bahasa dan Budaya*, vol. 2, no. 2, pp. 200–210, 2023, doi: 10.55606/jpbb.v2i2.1515.
- [10] A. Widiarti and H. Rahayu, "Pemberdayaan Penyandang Disabilitas Dalam Proses Pembangunan Di Indonesia Ditinjau Dari Perspektif Perundang-Undangan," *Jurnal Ilmu Hukum*, vol. 3, no. 2, pp. 274–283, 2020, doi: 10.32493/rjih.v3i2.8095.
- [11] A. Halder and A. Tayade, "Real-Time Vernacular Sign Language Recognition Using Mediapipe And Machine Learning," *International Journal of Research Publication and Reviews*, vol. 2, pp. 9–17, 2021, doi: 10.13140/RG.2.2.32364.03203.
- [12] F. M. Dewanto, H. A. Santoso, G. F. Shidik, and Purwanto, "Scoping Review of Sign Language Recognition: An Analysis of MediaPipe Framework and Deep Learning Integration," in *Proc. IEEE Int. Seminar on Application for Technology of Information and Communication (iSemantic)*, Semarang, Indonesia, Sep. 21–22, 2024, pp. 451–458, doi: 10.1109/iSemantic63362.2024.10761983.
- [13] B. Sundar and T. Bagyammal, "American sign language recognition for alphabets using MediaPipe and LSTM," *Procedia Computer Science*, vol. 215, pp. 642–651, 2022, doi: 10.1016/j.procs.2022.12.066.
- [14] A. N. Ramdhon and F. Febriya, "Penerapan Face Recognition Pada Sistem Presensi," *Journal of Applied Computer Science and Technology*, vol. 2, no. 1, pp. 12–17, 2021, doi: 10.52158/jacost.v2i1.121.
- [15] S. B. Abdullahi and K. Chamnongthai, "American Sign Language words recognition using spatio-temporal prosodic and angle features: A sequential learning approach," *IEEE Access*, vol. 10, pp. 15911–15923, 2022, doi: 10.1109/ACCESS.2022.3148132.
- [16] J. Huang, J. Chaijaruwanich, and V. Chouvatut, "Video-based sign language recognition with R(2+1)D and LSTM networks," in *Proc. IEEE Int. Conf. on Knowledge and Smart Technology (KST)*, Krabi, Thailand, Feb. 28–Mar. 2, 2024, pp. 214–219, doi: 10.1109/KST61284.2024.10499646.
- [17] I. Suyudi, S. Sudadio, and S. Suherman, "Pengenalan Bahasa Isyarat Indonesia Menggunakan Mediapipe Dengan Model Random Forest Dan Multinomial Logistic Regression," *Jurnal Ilmu Siber dan Teknologi Digital (JISTED)*, vol. 1, no. 1, pp. 65–80, 2022, doi: 10.35912/jisted.v1i1.1899.
- [18] L. Wiranda, M. Sadikin, and Jurusan Teknik Informatika, Fakultas Ilmu Komputer, Universitas Mercu Buana, "Penerapan Long Short Term Memory pada Data Time Series untuk Memprediksi Penjualan Produk PT. Metiska Farma," *Jurnal Nasional Pendidikan Teknik Informatika*, no. 3, p. 184, 2019, doi: 10.23887/janapati.v8i3.19139.
- [19] W. A. F. B. Nordin, R. Alfred, C. P. Yee, S. H. Tanalol, R. V. Loudin, and Z. Iswandono, "Malay fake news classification using a deep learning approach," in *Proc. 9th Int. Conf. on Computational Science and Technology (ICCST 2022)*, Lecture Notes in Electrical Engineering, vol. 983, D. K. Kang, R. Alfred, Z. I. B. A. Ismail, A. Baharum, and V. Thiruchelvam, Eds. Singapore: Springer, 2023, doi: 10.1007/978-981-19-8406-8\_2.
- [20] H. M. Putri, Fadlisyah, and W. Fuadi, "Pendeteksian Bahasa Isyarat Indonesia secara Real-Time Menggunakan Long Short-Term Memory (LSTM)," *Skripsi, Fakultas Teknik, Universitas Malikussaleh*, 2021, doi: 10.29103/tts.v3i1.6853.