

DETEKSI BAHASA ISYARAT BERDASARKAN ABJAD MENGGUNAKAN METODE LSTM (*LONG SHORT-TERM MEMORY*)

Syanti Irviantina, Dela Agustri Wijaya, Desiana R. Situmorang,
Nazhiifah Mawaddah Juliyanda Nasution✉

Fakultas Informatika, Universitas Mikroskil, Medan, Indonesia

Email: nazhiifahn@gmail.com

DOI: <https://doi.org/10.46880/methoda.Vol14No3.pp371-376>

ABSTRACT

The LSTM based sign language detection system combined with the use of mediapipe can recognize hand gestures in real time with high accuracy. Alphabet based sign language can use this model to collect temporal patterns of hand gestures. The data used in this study are 30 sample for each alphabet based on American Sign Language (ASL). The data is processed through landmark detection on the hand using mediapipe and opencv, keypoints extraction, folder creation and pre – processing with 80% data divisor for training data and 20 % data for testing data. Using Adam optimization, categorical crossentropy loss function and evaluation metric. The model was trained with 100 epochs and evaluated using confusion matrix. The evaluation results showed that the LSTM model performed well with an accuracy of 98%.

Keyword: LSTM, Sign Language, Mediapipe, ASL.

ABSTRAK

Sistem deteksi bahasa isyarat berbasis LSTM yang dikombinasikan dengan penggunaan mediapipe dapat mengenali gerakan tangan secara real – time dengan akurasi yang tinggi. Bahasa Isyarat berbasis alfabet dapat menggunakan model ini untuk mengumpulkan pola temporal dari gerakan tangan. Data yang digunakan pada penelitian ini ada 30 sampel untuk setiap abjad yang didasarkan pada bahasa isyarat American Sign Language (ASL). Data diproses melalui pendeteksian landmark pada tangan menggunakan mediapipe dan opencv, ekstraksi poin kunci, pembuatan folder dan pre – processing dengan pembagian data 80% untuk data training dan 20% untuk data testing. Menggunakan optimasi Adam, fungsi kerugian crossentropy dan metrik evaluasi. Model dilatih dengan 100 epoch dan dievaluasi menggunakan confusion matrix. Hasil evaluasi menunjukkan bahwa model LSTM memiliki performa yang baik dengan akurasi sebesar 98%.

Kata Kunci: LSTM, Bahasa Isyarat, Mediapipe, ASL.

PENDAHULUAN

Bagi orang dengan gangguan pendengaran atau tuna rungu, bahasa isyarat merupakan bahasa yang sangat penting untuk berkomunikasi (Fahrurrozi, 2023). Menggerakkan anggota tubuh, mulut dan ekspresi wajah merupakan semua cara untuk melakukan komunikasi dalam bahasa isyarat. WHO mengemukakan bahwa pada tahun 2020 terdapat 466 juta orang yang ada di seluruh dunia yang mengalami gangguan pendengaran (dr. Novita

Tirtaprawita, 2020). Bahasa isyarat di Indonesia terdiri dari dua jenis, yang pertama adalah Sistem Bahasa Isyarat Indonesia (SIBI) yang distandardisasi dari *American Sign Language* (ASL) dengan awalan dan akhiran, yang kedua adalah Bahasa Isyarat Indonesia (BISINDO) dimana menjadi bahasa pertama yang digunakan oleh komunitas tuna rungu setiap harinya (Dani et al., 2024).

Namun, bahasa isyarat yang akan digunakan dalam penelitian ini adalah American

Sign Language (ASL), di Amerika Utara orang yang menderita tuna rungu sering menggunakan ASL sebagai bahasa utama saat berkomunikasi (Peling et al., 2024). ASL adalah salah satu bahasa yang paling banyak digunakan di dunia, jika seseorang tidak terbiasa menggunakannya sulit bagi orang tersebut untuk memahaminya. Dengan kemajuan teknologi saat ini, banyak peneliti yang berfokus pada penggunaan pembelajaran Bahasa isyarat dengan *machine learning* seperti mendeteksi bahasa isyarat Indonesia secara *real - time* menggunakan LSTM menghasilkan akurasi sebanyak 92% untuk model 10 kelas dengan lapisan LSTM *bidirectional*, *epoch* 1000, *hidden layer* 64 dan *batch size* 32. Pada penelitian selanjutnya, dalam menggunakan *mediapipe* dan LSTM menghasilkan akurasi sebesar 91% menggunakan metode *single layer*, *epoch* 500, *hidden layer* 32, *batch size* 64 (Budi Utomo, 2024).

Penelitian ini menggunakan dataset bahasa isyarat untuk abjad berdasarkan dengan ASL. Gerakan tangan akan dideteksi secara *real - time*.

TINJAUAN PUSTAKA

Bahasa Isyarat

American *Sign Language* (ASL) adalah bahasa isyarat yang digunakan oleh masyarakat Amerika Utara, khususnya oleh individu yang tuna rungu dan tuna wicara, sebagai sarana komunikasi. Meskipun ASL merupakan bahasa yang sangat penting bagi komunitas ini, tidak semua anggota masyarakat umum memahami atau mengetahui bahasa isyarat. Hal ini menyebabkan penggunaan ASL, terutama dalam bentuk *fingerspelling* (penyebutan huruf dengan jari), menjadi jarang. Kurangnya pengetahuan dan pemahaman tentang bahasa isyarat di kalangan masyarakat umum dapat mengakibatkan kesulitan dalam komunikasi antara individu yang mendengar dan mereka yang tuna rungu atau tuna wicara, sehingga diperlukan suatu sistem yang dapat mengenali huruf alfabet dalam bahasa isyarat untuk membantu masyarakat memahami *fingerspelling* ASL (Putri et al., 2022).

Di era teknologi yang semakin maju, masalah ini dapat diatasi dengan mengembangkan sistem yang mampu mengenali bahasa isyarat. Sistem ini memanfaatkan teknologi *machine learning* untuk meningkatkan akurasi dan efisiensi dalam pengenalan gerakan tangan yang mewakili huruf dan kata dalam ASL (Yulian & Suhartono, 2017).

Kebutuhan akan sistem pengenalan ASL sangat penting, terutama untuk meningkatkan komunikasi antara individu yang mendengar dan mereka yang tuna rungu. Dengan adanya sistem ini, diharapkan akan ada peningkatan interaksi sosial dan pemahaman antara kedua kelompok tersebut. Penelitian-penelitian sebelumnya menunjukkan bahwa meskipun beberapa model telah dikembangkan dengan hasil yang menjanjikan, masih ada ruang untuk perbaikan dalam hal akurasi dan kecepatan deteksi (Aisyah et al., 2022).

LSTM (*Long Short Term Memory*)

Long Short Term Memory (LSTM) merupakan salah satu jenis jaringan saraf tiruan yang dirancang khusus untuk menangani dan memprediksi data yang bentuknya berurutan. Keunggulan utama algoritma ini adalah kemampuannya dalam memahami ketergantungan jangka panjang pada data sekuensial, sehingga sangat cocok digunakan dalam pengenalan gerakan sistem bahasa isyarat (Jasmine, 2014). Sistem deteksi bahasa isyarat berbasis LSTM melibatkan beberapa langkah, termasuk pengambilan citra, pra-pemrosesan, ekstraksi fitur, dan pelatihan model (Altarika & Sari, 2023). LSTM memiliki struktur dasar, terdiri dari lapisan input, lapisan tersembunyi, dan lapisan output. LSTM juga memiliki sel memori dan gerbang, yang terdiri dari tiga gerbang: gerbang lupa, gerbang *input*, dan gerbang *output*. Persamaannya seperti berikut ini:

$$ft = \sigma (wf \cdot [ht-1, xt] + bf) \quad (1)$$

$$it = \sigma (wi \cdot [ht-1, xt] + bi) \quad (2)$$

$$Ct\sim = \tanh (wc \cdot [ht-1, xt] + bc) \quad (3)$$

$$Ct = ft * Ct-1 + it * Ct\sim \quad (4)$$

$$ot = \sigma (wt \cdot [ht-1, xt] + bo) \quad (5)$$

$$ht = ot * \tanh (Ct) \quad (6)$$

matriks w berfungsi sebagai bobot, b berfungsi sebagai nilai bias, dan f_t , i_t , $C\sim t$, o_t , dan h_t merupakan keluaran dari *forget gate*, *input gate*, *cell state*, *output gate*, dan nilai *output* selama periode waktu (t) (Putri et al., 2022).

METODE PENELITIAN

Pengumpulan data

Data diambil dari *Kaggle* berdasarkan gerakan tangan menurut ketentuan ASL. Dengan mengambil 30 sampel per abjadnya. Total keseluruhan *dataset* ada 780 data citra dengan format .jpg.

Pengolahan data

Pengolahan data pada penelitian ini melewati beberapa proses yaitu :

1. Melakukan pendeteksian *landmarks* pada tangan : akan membuat model *landmarks* menggunakan *mediapipe* dan *opencv* dengan menempatkan 21 titik *landmarks* pada setiap gerakan tangan dari ujung jari, pergelangan tangan dan titik lainnya.
2. Melakukan ekstraksi *keypoints* : *keypoints* yang telah dikumpulkan akan diekstraksi ke dalam *array Numpy*.
3. Pembuatan *Folder*
Folder ini berguna sebagai menyimpan data gerakan tangan yang telah dideteksi sebelumnya menggunakan *keypoints mediapipe*.

Pre-process Data

Pada tahapan ini, label akan diinisialisasikan kemudian data dibagi sebanyak 80% data *training* dan 20% data *testing* menggunakan *library scikit learn*.

Pembagian data *pre- procees* diatas untuk memastikan bahwa model memiliki cukup data untuk belajar dan mengoptimalkan parameter yang digunakan serta untuk menghindari *bias* yang dapat mempengaruhi hasil evaluasi model dan untuk menjaga keseimbangan antar kelas.

Pemodelan dan Pelatihan

Pemodelan dilakukan dengan menggunakan arsitektur LSTM yang dirancang untuk menangkap pola temporal pada data yang

sequential. Model ini terdiri dari 3 lapisan LSTM dan 3 lapisan *Dense*. Pada setiap lapisan LSTM menggunakan fungsi aktivasi ReLu. Lapisan pertama pada LSTM memiliki 64 unit dengan mengembalikan seluruh urutan dari output. Lapisan yang kedua pada LSTM memiliki 128 unit dan juga mengembalikan seluruh urutan. Dan lapisan yang ketiga memiliki 64 unit dan hanya mengembalikan output yang terakhir dari urutan.

Model ini dilengkapi dengan dua lapisan *Dense* setelah lapisan LSTM dan masing – masing memiliki 64 dan 32 unit dengan menggunakan fungsi aktivasi ReLu. Pada lapisan output terdiri dari unit yang sesuai dengan jumlah kelas tindakan yang diidentifikasi untuk menghasilkan probabilitas dalam klasifikasi maka menggunakan fungsi aktivasi *softmax*. Model ini juga dibangun dengan melakukan pengoptimalan Adam dan fungsi kerugian *categorical crossentropy* dimana cocok untuk mengatasi masalah klasifikasi dengan multi kelas dengan metrik evaluasi berupa akurasi kategorikal.

Dengan 100 *epoch* total, model dilatih menggunakan data *training* terpisah dan *callback TensorBoard* digunakan untuk memantau proses *training*. Setelah proses *training* selesai, model dievaluasi pada data *testing* untuk mengevaluasi akurasi dan kerugiannya dimana akan menunjukkan seberapa baik model dapat mengklasifikasikan tindakan yang sedang diamati.

Evaluasi

Pada proses evaluasi menggunakan *confusion matrix* untuk mengevaluasi kinerja model.

HASIL DAN PEMBAHASAN

Pengujian

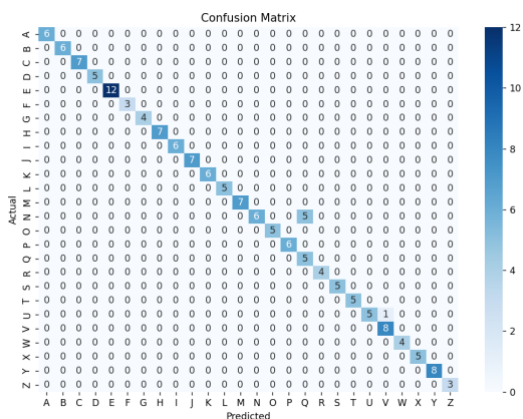
Setelah melakukan proses training, model yang sudah melewati tahap training akan dievaluasi menggunakan *confusion matrix* untuk mengukur kinerja yang dilakukan pada data testing. Hasil dari evaluasi ini ditunjukkan pada tabel dibawah.

Tabel 1. Hasil Akurasi Evaluasi

Kelas	Model LSTM	Epoch	Batch Size		
			16	32	64
26 Kelas	1 Layer LSTM	100	0.98	0.98	0.98
		500	0.97	0.97	0.57
		1000	0.57	0.57	0.57
	2 Layer LSTM	100	0.57	0.57	0.57
		500	0.57	0.57	0.57
		1000	0.57	0.57	0.57

Hasil dari evaluasi kinerja model LSTM diatas dengan mengevaluasi kinerja model LSTM dalam konteks klasifikasi dengan mempertimbangkan perbedaan dalam jumlah lapisan LSTM, jumlah *epoch*, dan *batch size*. Pengujian ini dilakukan dengan setiap konfigurasi dari model dievaluasi berdasarkan akurasi dan *loss* pada data *testing* setelah proses *training*. Ini dilakukan untuk mengidentifikasi *hyperparameter* yang optimal dan memahami pengaruh konfigurasi model terhadap kinerja dan mendeteksi potensi terjadinya *overfitting* selama peningkatan jumlah *epoch*.

Seperti yang dilihat dari Tabel 1 dengan menggunakan 1 *layer* memiliki performa terbaik pada 100 *epoch* dengan akurasi sebesar 0.98 untuk semua *batch size*, tetapi performanya mulai menurun secara signifikan pada *epoch* ke 500 dan 1000. Pada model LSTM dengan 2 *layer* menunjukkan performa yang konsisten di semua *epoch* dan *batch size* dengan akurasi 0.57 ini bisa menunjukkan jika kemungkinan selama prosesnya model mengalami *underfitting* atau model yang terlalu sederhana untuk 26 kelas.



Gambar 1. Hasil Confusion Matrix

Pada Gambar 1 menunjukkan bahwa model berhasil mengklasifikasikan karakter alfabet tetapi, ada beberapa alfabet yang berhasil diprediksi dengan tidak baik seperti alfabet Q dan N serta V dan U. Model memiliki kecenderungan untuk salah dalam mengklasifikasikannya dikarenakan visual dari karakter tersebut mirip.


Implementasi Sistem Deteksi Secara *Real-Time*


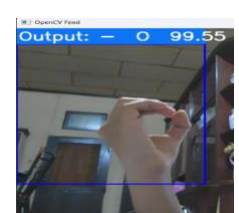
Sistem deteksi bahasa isyarat dirancang untuk mengenali pola dalam gerakan tangan secara langsung menggunakan kamera sebagai input utamanya. Proses implementasi sistem meliputi beberapa tahapan yaitu :

1. Pengambilan gerakan secara real – time
Sistem memanfaatkan kamera yang telah di integrasikan dengan *library OpenCV* agar dapat membaca setiap gerakan tangan yang sedang dilakukan untuk mendeteksi gerakan tangan apa yang sedang dibuat.
2. Deteksi tangan
Dengan menggunakan *mediapipe hands* sistem dapat mendeteksi gerakan tangan dari setiap frame yang diambil.
3. Klasifikasi gerakan dengan model LSTM
Data *landmark* dari *frame* yang terdeteksi akan diteruskan ke model LSTM untuk mengenali pola – pola dari setiap gerakan tangan. Model ini dirancang untuk menangkap pola yang temporal dari setiap gerakan tangan yang dilakukan
4. Tampilan *output*

Berikut adalah tampilan hasil dari deteksi bahasa isyarat :


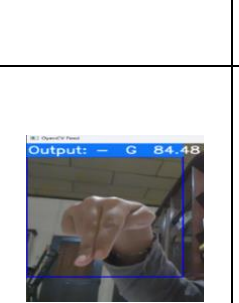
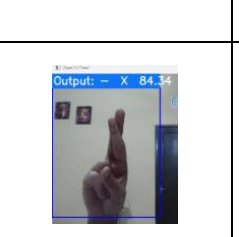
Tabel 2. Hasil Output yang Berhasil Dideteksi

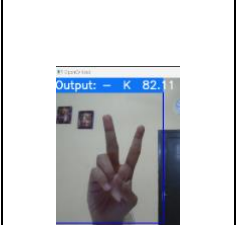
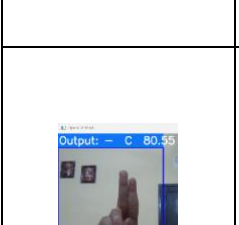
Gambar	Isyarat	Hasil Deteksi
	B	Berhasil dalam mendeteksi abjad “B” dengan akurasi 83.41%

	W	Berhasil dalam mendeteksi abjad “W” dengan akurasi 94.55%
	O	Berhasil dalam mendeteksi abjad “O” dengan akurasi 99.55%

Hasil *output* pada Tabel 2 menunjukkan bahwa sistem deteksi abjad memiliki Tingkat akurasi yang berbeda, dimana tingkat akurasi tertinggi adalah abjad “O” sebesar 99.55%. Selain itu, untuk memberikan gambaran yang lebih baik tentang masalah yang dihadapi sistem deteksi, tabel di bawah ini menampilkan hasil deteksi yang tidak dapat di prediksi dengan baik.

Tabel 3. Hasil Output yang Tidak Berhasil Dideteksi

Gambar	Isyarat	Hasil Deteksi
	D	Abjad “D” tidak berhasil di deteksi dikarenakan seharusnya abjad yang ditampilkan adalah abjad “J”
	G	Abjad “G” tidak berhasil di deteksi dikarenakan harusnya abjad yang ditampilkan adalah “N”
	R	Abjad “R” tidak berhasil di deteksi dikarenakan harusnya abjad

		yang ditampilkan adalah “X”
	V	Abjad “V” tidak berhasil di deteksi dikarenakan gerakan tangan pada dataset abjad “V” dan “K” serupa.
	U	Abjad “C” tidak berhasil di deteksi dikarenakan harusnya abjad yang ditampilkan adalah “C”

Hasil *output* diatas menunjukkan meskipun tingkat akurasi rata – rata sebesar 80%, masih terdapat beberapa kesalahan yang signifikan dalam melakukan pendeteksian abjad pada gerakan tangan serta sistem masih mengalami kesulitan saat akan membedakan gerakan tangan antara isyarat yang memiliki karakteristik yang serupa.

KESIMPULAN

Berdasarkan hasil dari pengujian diatas maka dapat disimpulkan bahwa :

1. Sistem deteksi bahasa isyarat berbasis LSTM yang dikombinasikan dengan penggunaan *medaipipe* mampu mengenali gerakan tangan *real – time* dengan akurasi yang tinggi.
2. Model mampu menangkap pola yang temporal pada gerakan tangan, sehingga cocok untuk bahasa isyarat berbasis abjad.
3. Tantangan seperti variasi pencahayaan dan posisi tangan masih menjadi kendala yang perlu diperbaiki untuk meningkatkan performa dan hasil deteksi yang tepat.

Dari kesimpulan yang telah diberikan adapun saran yang dapat diberikan yaitu dengan

meningkatkan jumlah data pada dataset, variasi pencahayaannya serta gerakan tangan yang lebih jelas agar dapat di deteksi dengan baik serta bisa melakukan pengoptimalan model dengan menggunakan metode *hybrid* seperti kombinasi antara CNN dan LSTM.

DAFTAR PUSTAKA

- Aisyah, N., Amin, M., & Pribadi, F. (2022). Urgensi Bahasa Isyarat dalam Pendidikan Formal sebagai Media Komunikasi dan Transmisi Informasi Penyandang Disabilitas Rungu dan Wicara. *Jurnal Hasil Pemikiran, Penelitian, Dan Pengembangan Keilmuan Sosiologi Pendidikan*, 9(1), 77–86.
- Altiarika, 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. *Jurnal Teknologi Informatika Dan Komputer*, 9(1), 1–13. <https://doi.org/10.37012/jtik.v9i1.1272>
- Budi Utomo, P. (2024). Deteksi Gerak Tangan sebagai Pengenal Bahasa Isyarat menggunakan Mediapipe dan Long-Short Term Memory. *Jurnal SIMETRIS*, 15(1), 121–136. <https://doi.org/10.24176/simet.v15i1.10505>
- Dan, R., Wicara, T., Yolov, M., Fresmanda, M. M., & Iriananda, S. W. (2024). *Deteksi Objek Video Bahasa Isyarat Untuk Anak Tuna*. 2, 1–9.
- dr. Novita Tirtaprawita. (2020). *Epidemiologi Tuli*. <https://www.alomedika.com/penyakit/telinga-hidung-tenggorokan/tuli/epidemiologi>
- Fahrurozi, R. N. S. (2023). Pendeteksian Bahasa Isyarat Menjadi Suara berbasis Artificial Intelligence METODE BLAZEPOSE (PEBISI) Dengan Output Suara Dan Text. *Informatik : Jurnal Ilmu Komputer*, 2(1), 29–38. <https://doi.org/10.37802/candrarupa.v2i1.307>
- JASMINE, K. (2014). 濟無No Title No Title No Title. *Penambahan Natrium Benzoat Dan Kalium Sorbat (Antiinversi) Dan Kecepatan Pengadukan Sebagai Upaya Penghambatan Reaksi Inversi Pada Nira Tebu*.
- Peling, I. B. A., Ariawan, I. M. P. A., & Subiksa, G. B. (2024). Deteksi Bahasa Isyarat Menggunakan Tensorflow Lite dan American Sign Language (ASL). *Jurnal Krisnadana*, 3(2), 90–100. <https://doi.org/10.58982/krisnadana.v3i2.534>
- Putri, H. M., Fadlisyah, F., & Fuadi, W. (2022). Pendeteksian Bahasa Isyarat Indonesia Secara Real-Time Menggunakan Long Short-Term Memory (Lstm). *Jurnal Teknologi Terapan and Sains 4.0*, 3(1), 663. <https://doi.org/10.29103/tts.v3i1.6853>
- Yulian, S. R., & Suhartono, S. (2017). Pengenalan Bahasa Isyarat Huruf Abjad Menggunakan Metode Learning Vector Quantization (LVQ). *Jurnal Masyarakat Informatika*, 8(1), 1–8. <https://doi.org/10.14710/jmasif.8.1.31450>