

Analisis Kinerja dan Tantangan Implementasi MobileNetV2 Pada Penerjemahan BISINDO Real-Time

Randi Maulana¹, Muhammad Rifqi Maulana², Natanael Vica Eungaggalion³, Giatika Chrisnawati⁴

^{1,2,3,4} Fakultas Teknik Informatika, Universitas Bina Sarana Informatika, Bekasi, Indonesia

Article Info

Article history:

Received Januari 1, 2026
Revised Januari 7, 2026
Accepted Januari 12, 2026

Kata Kunci:

BISINDO,
MobileNetV2,
MediaPipe,
Tantangan Implementasi,
Penerjemahan Waktu Nyata.

Keywords:

*BISINDO,
MobileNetV2,
Analisis Kinerja,
Implementation Challenges,
Real-Time Translation.*

ABSTRAK

Bahasa Isyarat Indonesia (BISINDO) merupakan media komunikasi vital bagi komunitas Tuli, namun kesenjangan interaksi dengan masyarakat umum yang tidak memahami bahasa isyarat masih menjadi tantangan signifikan. Penelitian ini mengembangkan sistem penerjemah abjad BISINDO berbasis Computer Vision menggunakan arsitektur MobileNetV2 dan MediaPipe Hands dengan algoritma Unified Bounding Box untuk menangani input dua tangan. Pelatihan model dilakukan melalui pendekatan transfer learning pada dataset yang memuat 26 kelas abjad, terdiri dari total 9.169 citra. Berdasarkan hasil eksperimen, model berhasil mencapai akurasi validasi tertinggi sebesar 99,23% pada epoch ke-8. Namun, pengujian pada skenario real-time menunjukkan adanya penurunan kinerja yang signifikan dibandingkan hasil pelatihan. Evaluasi sistem memperlihatkan bahwa meskipun model memiliki konfidensi tinggi pada data uji statis, implementasi langsung menghadapi tantangan berat terkait stabilitas deteksi, sensitivitas terhadap perubahan latar belakang, serta variabilitas pencahayaan. Penelitian ini menyoroti disparitas kinerja antara lingkungan pelatihan yang terkontrol dengan kondisi penggunaan nyata, serta mengidentifikasi faktor-faktor teknis yang menyebabkan inkonsistensi responsivitas sistem saat diterapkan secara langsung di lapangan.

ABSTRACT

Indonesian Sign Language (BISINDO) serves as a vital communication medium for the Deaf community, yet the interaction gap with the general public who do not understand sign language remains a significant challenge. This study develops a computer vision-based BISINDO alphabet translation system utilizing the MobileNetV2 architecture and MediaPipe Hands, integrated with a Unified Bounding Box algorithm to specifically handle dual-hand inputs. Model training was conducted through a transfer learning approach on a dataset containing 26 alphabet classes, comprising a total of 9,169 images. Experimental results demonstrate that the model achieved a highest validation accuracy of 99.23% at the 8th epoch. However, testing in real-time scenarios revealed a significant performance degradation compared to training results. System evaluation indicates that although the model exhibits high confidence on static test data, direct implementation faces severe challenges regarding detection stability, sensitivity to background changes, and lighting variability. This research highlights the performance disparity between controlled training environments and real-world usage conditions, while identifying technical factors causing the inconsistency in system responsiveness when deployed in the field.

This is an open access article under the [CC BY](https://creativecommons.org/licenses/by/4.0/) license.



Corresponding Author:

Randi Maulana
Fakultas Teknik Informatia, Universitas Bina Sarana Informatika
Bekasi, Indonesia
Email: 17230227@bsi.ac.id

1. PENDAHULUAN

Bahasa isyarat merupakan modalitas komunikasi utama bagi penyandang disabilitas rungu (Tuli) untuk berinteraksi dan menyampaikan informasi. Di Indonesia, terdapat dua sistem isyarat yang umum digunakan, yaitu Sistem Isyarat Bahasa Indonesia (SIBI) yang merupakan sistem tata bahasa baku adopsi pemerintah, dan Bahasa Isyarat Indonesia (BISINDO) yang tumbuh secara alami dari budaya komunitas Tuli [1]. Namun, kesenjangan komunikasi antara komunitas Tuli dan masyarakat dengar (non-tuli) masih menjadi hambatan sosial yang signifikan karena mayoritas masyarakat umum tidak memahami bahasa isyarat. Keterbatasan ini sering kali membatasi akses penyandang Tuli terhadap layanan publik, pendidikan, dan peluang kerja yang setara [2]. Oleh karena itu, pengembangan teknologi penerjemah bahasa isyarat otomatis menjadi solusi krusial untuk menjembatani hambatan komunikasi tersebut tanpa memerlukan kehadiran penerjemah manusia secara fisik.

Perkembangan teknologi *Computer Vision* dan *Deep Learning* telah membuka peluang besar dalam pengembangan Sistem Pengenalan Bahasa Isyarat (*Sign Language Recognition System*). Pendekatan berbasis visi (*vision-based*) kini lebih diminati dibandingkan pendekatan berbasis sensor (*wearable sensors*) karena lebih alami, ekonomis, dan tidak mengganggu kenyamanan pengguna [3]. Berbagai arsitektur *Convolutional Neural Network* (CNN) telah diterapkan untuk mengklasifikasikan citra isyarat tangan dengan akurasi yang tinggi. Namun, tantangan utama dalam implementasi sistem penerjemah pada perangkat pengguna adalah tingginya beban komputasi model CNN konvensional, yang sering kali menyebabkan latensi tinggi saat dijalankan secara *real-time* [4].

Untuk mengatasi kendala komputasi tersebut, arsitektur *MobileNetV2* diusulkan sebagai solusi yang efisien. *MobileNetV2* menggunakan *inverted residual blocks* dan *linear bottlenecks* yang dirancang khusus untuk mempertahankan akurasi tinggi dengan parameter yang jauh lebih sedikit dibandingkan arsitektur standar seperti VGG16 atau ResNet [5]. Efisiensi ini menjadikan *MobileNetV2* sangat relevan untuk diterapkan pada skenario penerjemahan portabel yang membutuhkan responsivitas tinggi [6]. Selain itu, akurasi pengenalan isyarat tangan dapat ditingkatkan secara signifikan melalui integrasi dengan *MediaPipe Hands*, sebuah kerangka kerja pelacakan tangan yang mampu mengekstraksi 21 titik *landmark* kerangka tangan secara *real-time* dengan presisi tinggi, bahkan pada latar belakang yang kompleks [7], [8].

Meskipun banyak penelitian telah melaporkan akurasi pelatihan yang tinggi pada dataset statis, implementasi sistem penerjemahan bahasa isyarat pada kondisi dunia nyata (*real-world*) sering kali menghadapi degradasi kinerja yang drastis. Faktor-faktor seperti variasi pencahayaan, *noise* latar belakang, dan ketidakstabilan deteksi *frame-per-frame* menjadi tantangan yang belum sepenuhnya teratasi hanya dengan memaksimalkan akurasi model pada data latih [9]. Penelitian sebelumnya oleh Agustiansyah et al. juga menunjukkan bahwa meskipun model modern seperti *Vision Transformer* mencapai akurasi mendekati sempurna pada dataset, tantangan komputasi dan responsivitas tetap menjadi isu kritis untuk penerapan langsung [10].

Penelitian ini bertujuan untuk menganalisis kinerja dan tantangan implementasi model *MobileNetV2* yang diintegrasikan dengan *MediaPipe* untuk penerjemahan abjad BISINDO secara *real-time*. Kebaruan (*novelty*) dari penelitian ini terletak pada penerapan algoritma *Unified Bounding Box* yang dirancang untuk menangani variabilitas input satu atau dua tangan secara dinamis, serta analisis mendalam mengenai

kesenjangan antara akurasi pelatihan teoretis dan stabilitas performa pada pengujian langsung. Melalui pendekatan ini, diharapkan dapat dihasilkan evaluasi komprehensif yang bermanfaat bagi pengembangan sistem komunikasi inklusif yang lebih tangguh dan responsif

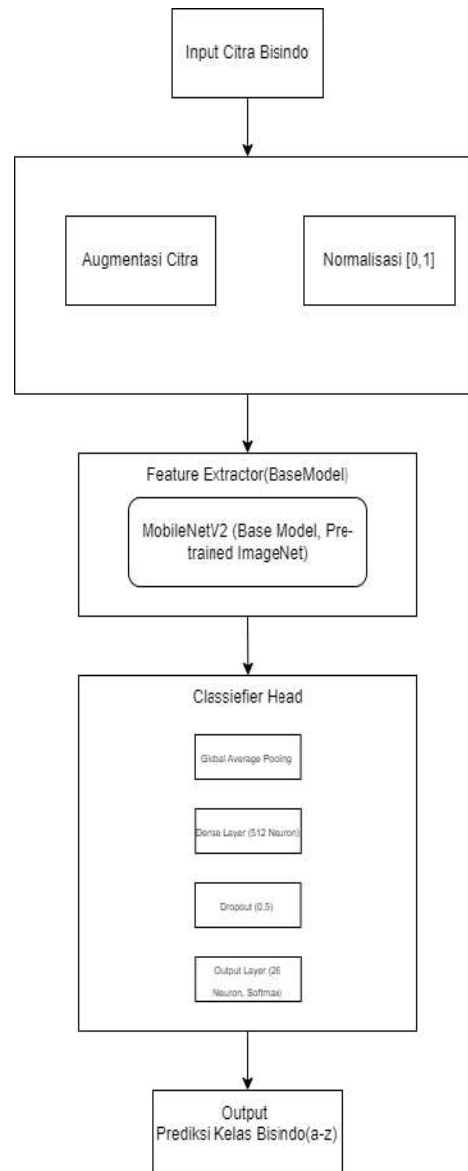
2. METODE

Penelitian ini menggunakan desain eksperimental untuk mengembangkan sistem penerjemahan Bahasa Isyarat Indonesia (BISINDO) secara *real-time*. Kronologis penelitian diawali dengan tahap pengumpulan data, pra-pemrosesan (*preprocessing*), perancangan arsitektur model, hingga implementasi algoritma deteksi pada antarmuka pengguna. Dataset yang digunakan terdiri dari citra isyarat tangan untuk 26 kelas abjad (A-Z). Total data dibagi dengan rasio 80:20 untuk memisahkan data pelatihan (*training*) dan validasi. Mengingat input kamera pengguna bisa sangat bervariasi, teknik augmentasi data diterapkan menggunakan *ImageDataGenerator* untuk memperkaya variasi dataset dan mencegah *overfitting*. Parameter augmentasi yang digunakan dapat dilihat pada Tabel 1.

Tabel 1. Parameter Augmentasi Data

Parameter	Nilai Konfigurasi
Rescale	1./255
Rotation Range	20°
Width Shift	0.2
Height Shift	0.2
Shear Range	0.2
Zoom Range	0.2

Setelah augmentasi, seluruh citra dinormalisasi agar nilai piksel berada dalam rentang [0,1]. Hal ini dilakukan untuk mempercepat proses konvergensi gradien selama pelatihan model berlangsung. Arsitektur yang diadopsi adalah MobileNetV2 sebagai ekstraktor fitur utama (base model) dengan bobot pra-latih (pre-trained) dari ImageNet . Modifikasi dilakukan pada bagian klasifikasi (head) dengan menambahkan lapisan Global Average Pooling dan Dense Layer (512 neuron) serta Dropout (0.5) untuk menyesuaikan dengan 26 kelas keluaran BISINDO. Alur arsitektur ini divisualisasikan pada Gambar 1.



Gambar 1. Arsitektur Sistem Penerjemahan BISINDO

Selain arsitektur model, sistem ini menerapkan mekanisme deteksi objek berbasis MediaPipe Hands untuk pengambilan input isyarat secara real-time. Berbeda dengan metode konvensional yang memproses seluruh bingkai kamera, sistem ini menggunakan algoritma Unified Bounding Box yang dirancang khusus untuk menangani variabilitas input satu atau dua tangan. Algoritma ini bekerja dengan menghitung koordinat ekstrem (x_{min} , y_{min}) dan (x_{max} , y_{max}) dari seluruh landmark tangan yang terdeteksi, lalu menggabungkannya menjadi satu wilayah kepentingan (Region of Interest/ROI) tunggal. Sebuah padding sebesar 20% ditambahkan pada area pemotongan (cropping) untuk memastikan seluruh bagian tangan tertangkap dengan utuh sebelum citra diubah ukurannya (resize) ke dimensi 224x224 piksel untuk proses klasifikasi.

Model dilatih menggunakan pengoptimal (*optimizer*) Adam dengan laju pembelajaran awal (*learning rate*) sebesar 0,001. Fungsi kerugian yang diterapkan adalah *Categorical Crossentropy* mengingat studi ini merupakan klasifikasi multi-kelas. Proses pelatihan dijalankan sebanyak 50 *epoch* dengan ukuran *batch* 32. Untuk menjaga performa pelatihan, diterapkan mekanisme *callback* berupa ModelCheckpoint untuk menyimpan bobot terbaik, ReduceLROnPlateau untuk menurunkan laju pembelajaran jika akurasi stagnan, serta EarlyStopping untuk menghentikan pelatihan secara otomatis apabila tidak terjadi peningkatan performa yang signifikan guna mencegah *overfitting*.

3. HASIL DAN PEMBAHASAN

Bagian ini menguraikan hasil eksperimen pelatihan model MobileNetV2 dan analisis implementasinya pada sistem penerjemahan BISINDO *real-time*. Evaluasi dilakukan berdasarkan metrik kuantitatif selama proses pelatihan serta pengamatan kualitatif saat model diuji coba secara langsung.

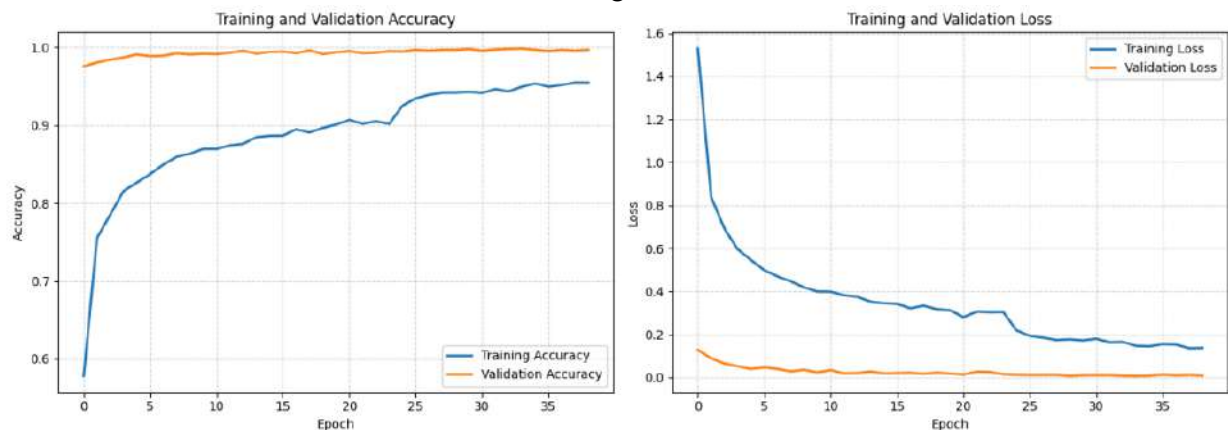
3.1. Evaluasi Kinerja Model

Proses pelatihan dilakukan menggunakan dataset yang telah diaugmentasi sebanyak 9.169 citra. Evaluasi utama difokuskan pada kemampuan model dalam mempelajari fitur isyarat tangan dari data latih dan melakukan generalisasi pada data validasi yang belum pernah dilihat sebelumnya.

3.1.1 Analisis Metrik Pelatihan dan Validasi

Berdasarkan grafik riwayat pelatihan (*training history*), model menunjukkan konvergensi yang cepat dan stabil. Model dilatih selama 50 *epoch* dengan ukuran *batch* 32. Hasil pemantauan menunjukkan bahwa performa terbaik model tercapai pada Epoch ke-8. Pada titik ini, akurasi validasi (*validation accuracy*) mencapai angka tertinggi sebesar 99,23% dengan nilai *loss* validasi yang sangat rendah, yaitu 0,0280. Dinamika pergerakan akurasi dan *loss* selama proses pelatihan dapat dilihat pada Gambar 2.

Gambar 2. Grafik Perbandingan Akurasi Pelatihan dan Validasi



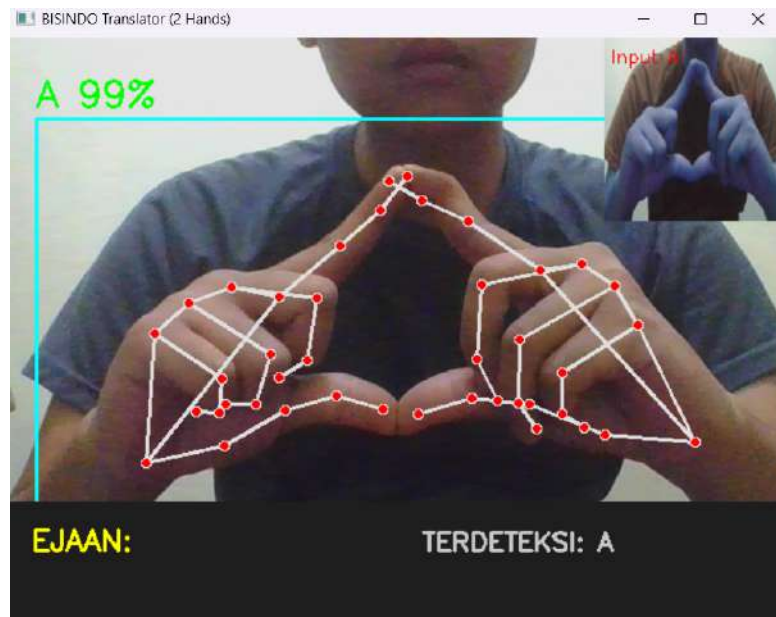
Seperti terlihat pada Gambar 2, kesenjangan (gap) antara garis akurasi pelatihan dan validasi sangat tipis. Hal ini mengindikasikan bahwa penggunaan teknik transfer learning dengan bobot ImageNet serta mekanisme callback ReduceLROnPlateau sangat efektif dalam mencegah overfitting. Model mampu mengekstraksi fitur spasial tangan yang kompleks tanpa menghafal data latih secara berlebihan.

3.2. Implementasi dan Pengujian Real-Time

Pengujian tahap lanjut dilakukan dengan mengintegrasikan model ke dalam sistem deteksi berbasis kamera menggunakan algoritma *Unified Bounding Box*. Berbeda dengan pengujian statis, implementasi dinamis menghadapi tantangan variabilitas lingkungan seperti pencahayaan dan latar belakang yang tidak terkontrol.

3.2.1 Keberhasilan Deteksi pada Isyarat Jelas

Pada kondisi ideal dengan pencahayaan cukup, sistem mampu mendeteksi isyarat tangan dengan sangat baik. Gambar 3 memperlihatkan keberhasilan sistem dalam menerjemahkan isyarat huruf 'A' yang melibatkan penggunaan dua tangan.



Gambar 3. Deteksi Huruf 'A' dengan Unified Bounding Box

Berdasarkan Gambar 3, algoritma berhasil memetakan landmark dari kedua tangan dan menggabungkannya menjadi satu input. Model memprediksi huruf 'A' dengan tingkat keyakinan (confidence) sebesar 99%. Hal ini membuktikan bahwa metode penggabungan ROI (Region of Interest) efektif untuk menangani isyarat BISINDO yang kompleks.

3.2.2. Variabilitas Kinerja dan Faktor Lingkungan

Meskipun akurasi tinggi tercapai pada huruf dengan bentuk distingtif seperti 'A', pengujian lapangan menunjukkan adanya disparitas kinerja pada kelas huruf lainnya. Beberapa kendala (yang Anda sebut "kacau") dianalisis sebagai berikut:

1. Kemiripan Bentuk Antar-Kelas: Beberapa huruf memiliki pola isyarat yang sangat mirip (seperti E, M, dan N) jika dilihat dari sudut pandang 2D kamera *webcam*. Hal ini menyebabkan model terkadang mengalami fluktuasi prediksi (*flickering*) antara huruf-huruf tersebut.
2. Ketergantungan pada MediaPipe: Pada gerakan tangan yang cepat atau posisi tangan yang saling menutupi (*occlusion*), MediaPipe terkadang gagal mendeteksi *landmark* jari secara utuh. Kegagalan deteksi *landmark* ini menyebabkan input gambar yang masuk ke model menjadi terpotong atau tidak valid, sehingga prediksi menjadi tidak akurat.
3. Sensitivitas Pencahayaan: "Kekacauan" deteksi meningkat signifikan pada kondisi pencahayaan yang tidak merata (terlalu gelap atau *backlight*). Hal ini menegaskan bahwa performa sistem *real-time* sangat bergantung pada kualitas input visual dari kamera pengguna.

4. KESIMPULAN

Penelitian ini berhasil mengembangkan purwarupa sistem penerjemah Bahasa Isyarat Indonesia (BISINDO) yang mengintegrasikan arsitektur MobileNetV2 dengan teknologi pelacakan tangan MediaPipe. Berdasarkan hasil eksperimen, model klasifikasi terbukti sangat efektif dengan pencapaian akurasi validasi tertinggi sebesar 99,23% pada epoch ke-8. Penerapan algoritma Unified Bounding Box juga berhasil memberikan solusi teknis untuk menangani variabilitas input isyarat satu maupun dua tangan dalam satu kerangka kerja deteksi.

Meskipun demikian, pengujian implementasi real-time menunjukkan bahwa kinerja sistem masih sangat dipengaruhi oleh faktor lingkungan, khususnya stabilitas pencahayaan dan kemiripan pola antar-abjad yang menyebabkan fluktuasi prediksi. Oleh karena itu, penelitian selanjutnya disarankan untuk

memfokuskan pengembangan pada pengayaan dataset dengan variasi latar belakang ekstrem serta penerapan metode smoothing temporal untuk meningkatkan konsistensi deteksi di kondisi nyata.

REFERENSI

- [1] A. S. Nugraheni, A. P. Husain, and H. Unayah, "Optimalisasi Penggunaan Bahasa Isyarat Dengan Sibi Dan Bisindo Pada Mahasiswa Difabel Tunarungu Di Prodi Pgmi Uin Sunan Kalijaga," *Jurnal Holistika*, vol. 5, no. 1, pp. 28–33, Feb. 2021, doi: 10.24853/HOLISTIKA.5.1.28-33.
- [2] Meysa Malika and Berlianti Berlianti, "Penggunaan Bahasa Isyarat SIBI dan BISINDO untuk Siswa Tuli di Sekolah Luar Biasa Negeri Pembina Medan," *Concept: Journal of Social Humanities and Education*, vol. 4, no. 3, pp. 78–87, Sep. 2025, doi: 10.55606/concept.v4i3.2178.
- [3] J. Rong, "A Comparative Analysis of Glove-Based and Image-Based Sign Language Recognition Systems," *INSTICC*, May 2025, pp. 408–412. doi: 10.5220/0013337000004558.
- [4] G. Kazbekova *et al.*, "Real-Time Lightweight Sign Language Recognition on Hybrid Deep CNN-BiLSTM Neural Network with Attention Mechanism," 2025. [Online]. Available: www.ijacsa.thesai.org
- [5] I. Qutab, L. Po, F. Rollo, and W. Naqvi, "Optimizing Sign Language Recognition Through a Tailored MobileNet Self-Attention Framework," *Applied Sciences* 2025, Vol. 15, Page 12622, vol. 15, no. 23, p. 12622, Nov. 2025, doi: 10.3390/APP152312622.
- [6] D. Jollyta, P. Prihandoko, J. Johan, W. Ramdhan, and E. Santoso, "Transfer Learning Model Evaluation on CNN Algorithm: Indonesian Sign Language System (SIBI)," *Journal of Applied Business and Technology*, vol. 6, no. 2, pp. 83–92, May 2025, doi: 10.35145/jabt.v6i2.213.
- [7] F. Zhang *et al.*, "MediaPipe Hands: On-device Real-time Hand Tracking," Jun. 2020, [Online]. Available: <http://arxiv.org/abs/2006.10214>
- [8] S. Jiang, B. Sun, L. Wang, Y. Bai, K. Li, and Y. Fu, "Skeleton Aware Multi-modal Sign Language Recognition," May 2021, [Online]. Available: <http://arxiv.org/abs/2103.08833>
- [9] Y. Meng, H. Jiang, N. Duan, and H. Wen, "Real-Time Hand Gesture Monitoring Model Based on MediaPipe's Registerable System," *Sensors*, vol. 24, no. 19, Oct. 2024, doi: 10.3390/s24196262.
- [10] Y. Agustiansyah and D. Kurniadi, "Yoga Agustiansyah: Indonesian Sign Language Alphabet Indonesian Sign Language Alphabet Image Classification using Vision Transformer," 2025. [Online]. Available: <https://journal.aptika.org/index.php/jistics>