Journal of Information Technology and Computer Science (INTECOMS)
Volume 9 Nomor 5, Tahun 2026
e-ISSN : 2614-1574
p-ISSN : 2621-3249

OPTIMASI KLASIFIKASI BAHASA ISYARAT INDONESIA (BISINDO) BERBASIS VISION TRANSFORMER DENGAN CLAHE DAN MIXUP

OPTIMIZATION OF INDONESIAN SIGN LANGUAGE (BISINDO) CLASSIFICATION BASED ON VISION TRANSFORMER USING CLAHE AND MIXUP

Moh. Hatta[1], Muhamad Akbar[2], Muhammad Irvai[3], Joni Karman[4]
Fakultas Ilmu Teknik, Program Studi Informatika, Universitas Bina Insan, Lubuklinggau[1,2,3,4]
mohhatta440@gmail.com[1], muhammad.akbar@univbinainsan.ac.id[2],
muhammadirvai@univbinainsan.ac.id[3], jonikarman@univbinainsan.ac.id[4]

ABSTRACT

Indonesian Sign Language (BISINDO) serves as a primary means of communication for the deaf community; therefore, an image classification system capable of accurately recognizing gestures is essential. This study aims to enhance BISINDO classification performance by utilizing the Vision Transformer (ViT) architecture combined with Contrast Limited Adaptive Histogram Equalization (CLAHE) and MixUp data augmentation. The research employs an experimental quantitative approach involving data acquisition, image preprocessing (224×224 pixel resizing, Gaussian Blur, and CLAHE), the application of MixUp to the training data, and model training and evaluation using metrics such as accuracy, precision, recall, F1-score, and a confusion matrix. The dataset comprises 900 images across 10 classes, split into 50% training data, 40% validation data, and 10% test data. The results demonstrate an accuracy of 99%, with average precision, recall, and F1-score values of 0.99. These findings indicate that integrating CLAHE and MixUp effectively improves feature quality and the model's generalization capabilities for BISINDO classification.

Keywords: *BISINDO, Vision Transformer, CLAHE, MixUp*

ABSTRAK

Bahasa Isyarat Indonesia (BISINDO) merupakan sarana komunikasi utama bagi penyandang tunarungu sehingga diperlukan sistem klasifikasi citra yang mampu mengenali gestur secara akurat. Penelitian ini bertujuan meningkatkan performa klasifikasi BISINDO menggunakan arsitektur Vision Transformer (ViT) yang dikombinasikan dengan teknik Contrast Limited Adaptive Histogram Equalization (CLAHE) dan augmentasi data MixUp. Metode penelitian menggunakan pendekatan kuantitatif eksperimental melalui tahapan akuisisi data, preprocessing citra (ukuran 224×224 piksel, Gaussian Blur, dan CLAHE), penerapan MixUp pada data latih, serta pelatihan dan evaluasi model menggunakan metrik accuracy, precision, recall, F1-score, dan confusion matrix. Dataset terdiri atas 900 citra dari 10 kelas dengan pembagian 50% data latih, 40% data validasi, dan 10% data uji. Hasil penelitian menunjukkan akurasi 99% dengan nilai rata-rata precision, recall, dan F1-score sebesar 0,99. Temuan ini menunjukkan bahwa integrasi CLAHE dan MixUp efektif meningkatkan kualitas fitur dan kemampuan generalisasi model dalam klasifikasi BISINDO.

Kata Kunci: BISINDO, Vision Transformer, CLAHE, MixUp

PENDAHULUAN

Bahasa Isyarat Indonesia (BISINDO) merupakan media komunikasi utama yang digunakan oleh komunitas Tuli di Indonesia. Sebagai bahasa berbasis visual, BISINDO memanfaatkan gerakan tangan, ekspresi wajah, dan orientasi tubuh sebagai elemen utama penyampaian makna sehingga proses interpretasinya memerlukan kemampuan analisis visual yang kompleks [1]. Peran BISINDO sangat penting dalam mendukung komunikasi, interaksi sosial, serta peningkatan aksesibilitas bagi penyandang disabilitas rungu di berbagai aspek kehidupan. Namun, tingkat pemahaman masyarakat terhadap BISINDO masih relatif rendah sehingga kesenjangan komunikasi antara komunitas Tuli dan masyarakat umum masih sering terjadi [2]. Kondisi tersebut menunjukkan perlunya pengembangan teknologi yang mampu mendukung proses pengenalan dan penerjemahan bahasa isyarat secara otomatis, akurat, dan efisien.

Perkembangan computer vision dan deep learning telah mendorong munculnya

berbagai metode klasifikasi citra untuk pengenalan bahasa isyarat [3]. Salah satu pendekatan yang banyak digunakan adalah Convolutional Neural Network (CNN) karena memiliki kemampuan yang baik dalam mengekstraksi fitur visual [4][5]. Meskipun demikian, CNN cenderung berfokus pada informasi lokal sehingga kurang optimal dalam menangkap hubungan spasial global yang diperlukan untuk mengenali variasi gestur tangan yang kompleks [6]. Keterbatasan tersebut mendorong pemanfaatan arsitektur yang mampu mempelajari representasi visual secara lebih menyeluruh.

Salah satu arsitektur yang menunjukkan kinerja baik pada klasifikasi citra adalah Vision Transformer (ViT). Berbeda dengan CNN, ViT memanfaatkan mekanisme self-attention untuk memodelkan hubungan global antarbagian citra sehingga lebih efektif dalam memahami pola visual yang kompleks [7]. Meskipun demikian, performa ViT sangat dipengaruhi oleh kualitas citra masukan dan jumlah data pelatihan. Variasi pencahayaan, kontras yang rendah, serta keberadaan noise dapat mengurangi kualitas fitur yang diekstraksi dan berdampak pada penurunan akurasi klasifikasi [8]. Selain itu, keterbatasan data pelatihan juga dapat meningkatkan risiko overfitting sehingga kemampuan generalisasi model menjadi kurang optimal.

Untuk mengatasi permasalahan tersebut, penelitian ini mengintegrasikan teknik Contrast Limited Adaptive Histogram Equalization (CLAHE) sebagai tahap preprocessing dan MixUp sebagai teknik augmentasi data pada arsitektur Vision Transformer [9]. CLAHE digunakan untuk meningkatkan kontras citra sehingga detail visual pada area tangan menjadi lebih jelas dan mudah dipelajari oleh model [10].

Sementara itu, MixUp diterapkan untuk meningkatkan variasi data pelatihan melalui pembentukan kombinasi linier antar sampel sehingga mampu mengurangi overfitting dan meningkatkan kemampuan generalisasi model [11].

Berdasarkan uraian tersebut, penelitian ini bertujuan menganalisis performa model Vision Transformer yang dipadukan dengan CLAHE dan *MixUp* dalam klasifikasi citra Bahasa Isyarat Indonesia. Hasil penelitian diharapkan dapat memberikan kontribusi terhadap pengembangan sistem pengenalan bahasa isyarat yang lebih akurat serta menjadi referensi bagi penelitian selanjutnya di bidang computer vision, deep learning, dan teknologi aksesibilitas bagi penyandang disabilitas rungu.

METODE

Penelitian ini menerapkan metode kuantitatif dengan desain eksperimen untuk mengevaluasi kinerja Vision Transformer dalam mengklasifikasikan Bahasa Isyarat Indonesia (BISINDO).

Pengumpulan Dataset

Dataset penelitian diambil dari "SIBI: Official Indonesian Sign Language" yang tersedia pada Mendeley Data yang dapat di unduh melalui website https://data.mendeley.com/datasets/rmmdntvpft/1.

Penelitian ini memiliki 10 kelas bahasa isyarat dengan total 900 citra. Pada setiap kelasnya terdiri dari 90 citra, dengan rincian sebagai berikut :

Tabel 1. Kelas Dataset dan Jumlah

No	Kelas Bahasa Isyarat	Jumlah Citra
1	Aktif	90
2	Aku	90
3	Bapak	90
4	Burung	90
5	Cium	90
6	Hai	90
7	Kami	90
8	Ketuk	90
9	Laki-laki	90
10	Perempuan	90
Total		**900**

Split Data

Dataset dibagi menjadi tiga bagian yaitu, 50 % Data latih, 40 % Data validasi dan 10 % Data uji. Pembagian ini dilakukan untuk memastikan model dapat dilatih secara optimal sekaligus dievaluasi secara objektif terhadap data yang tidak digunakan pada saat melakukan training.

Berikut beberapa contoh dataset berdasarkan kelasnya :



Gambar 1. Contoh Dataset Kelas Kami



Gambar 2. Contoh Dataset Kelas Ketuk



Gambar 3. Contoh Dataset Kelas Aku

Pre-processing Data

Sebelum proses pelatihan, seluruh citra melalui tahap preprocessing untuk meningkatkan kualitas dan keseragaman data. Tahapan yang dilakukan meliputi resize menjadi 224×224 piksel agar sesuai dengan masukan Vision Transformer, penerapan Gaussian Blur untuk mengurangi noise, serta CLAHE (Contrast Limited Adaptive Histogram Equalization) untuk meningkatkan kontras sehingga detail tangan tetap terlihat jelas meskipun terjadi variasi pencahayaan.

Augmentasi Data

Augmentasi Data digunakan untuk mengurangi risiko overfitting akibat keterbatasan jumlah dataset, diterapkan teknik augmentasi Mixup pada data pelatihan. Mixup bekerja dengan menggabungkan dua citra beserta labelnya secara linear untuk menghasilkan sampel baru. Proses ini dapat membantu model mempelajari batas keputusan yang lebih halus dan meningkatkan kemampuan generalisasi terhadap data baru.

Arsitektur Model

Penelitian ini menggunakan Vision Transformer (ViT) sebagai model utama. Citra yang telah diproses dibagi menjadi beberapa patch, kemudian diubah menjadi embedding dan dipelajari menggunakan mekanisme self-attention untuk menangkap hubungan global antarbagian citra. Proses pelatihan dilakukan menggunakan TensorFlow dan Keras, sedangkan evaluasi menggunakan data validasi untuk memantau performa model dan mengurangi risiko overfitting.

Evaluasi Model

Evaluasi model dilakukan menggunakan data uji dengan metrik accuracy, precision, recall, F1-score, dan confusion matrix untuk mengukur ketepatan klasifikasi, kemampuan mengenali setiap kelas, serta menganalisis kesalahan prediksi model.

HASIL DAN PEMBAHASAN

Pada tahap hasil dan pembahasan, penelitian ini berfokus pada proses preprocessing data pelatihan yang bertujuan meningkatkan kualitas citra sebelum digunakan dalam pelatihan model Vision Transformer. Selain itu, pada tahap ini juga disajikan hasil pelatihan dan evaluasi model untuk menganalisis pengaruh penerapan preprocessing, CLAHE, dan MixUp terhadap performa klasifikasi Bahasa Isyarat Indonesia (BISINDO).

Proses Training Model



Gambar 4. Training Model VIT

Hasil pelatihan pada gambar 4 memperlihatkan bahwa model Vision Transformer (ViT) mengalami konvergensi yang baik, ditandai dengan nilai loss pelatihan dan validasi yang rendah serta nilai akurasi dan validasi yang tinggi dan relatif seimbang. Kondisi ini menunjukkan bahwa model mampu mempelajari representasi fitur citra secara optimal serta memiliki kemampuan generalisasi yang baik terhadap data yang tidak digunakan dalam proses pelatihan, yang menunjukkan bahwa model tidak mengalami overfitting secara berlebihan.

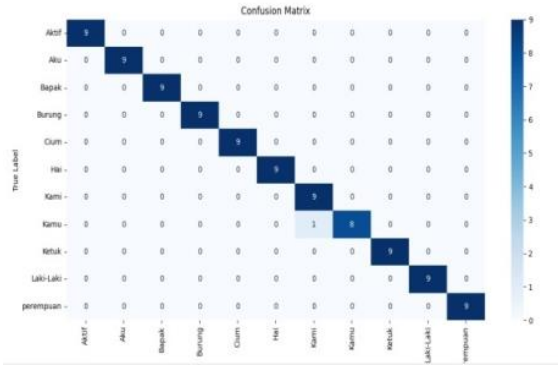
Evaluasi Hasil Kinerja Model

Evaluasi dilakukan menggunakan data uji sebanyak 10% dari total dataset. Berikut tabel hasil evaluasi kinerja model :

Tabel 2. Hasil Evaluasi Kinerja Model

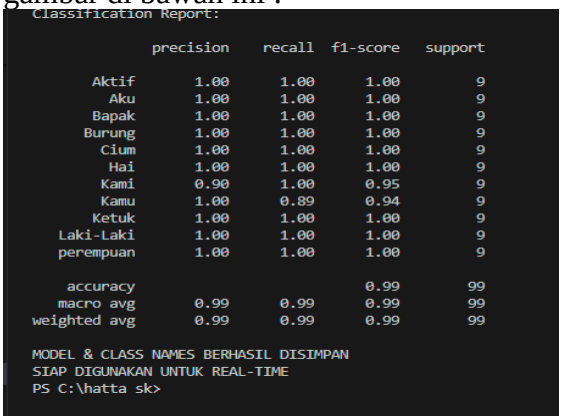
No	Metrik Evaluasi	Nilai
1.	*Accuracy*	99%
2.	*Precision* (rata-rata)	0,99
3.	*Recall* (rata-rata)	0,99
4.	F1-*score* (rata-rata)	0,99

Berdasarkan hasil evaluasi pada data uji, model mencapai tingkat akurasi sebesar 99% dengan nilai *precision*, *recall*, dan F1-*score* rata-rata sebesar 0,99. Hasil ini menunjukkan bahwa model memiliki kemampuan klasifikasi yang sangat baik serta konsisten pada seluruh kelas.



Gambar 5. Confusion Matrix

Confusion matrix menunjukkan bahwa sebagian besar prediksi berada pada diagonal utama, yang menandakan tingkat klasifikasi benar sangat tinggi. Selanjutnya classification report, dapat dilihat pada gambar di bawah ini :



Classification Report:

	precision	recall	f1-score	support
Aktif	1.00	1.00	1.00	9
Aku	1.00	1.00	1.00	9
Bapak	1.00	1.00	1.00	9
Burung	1.00	1.00	1.00	9
Cium	1.00	1.00	1.00	9
Hai	1.00	1.00	1.00	9
Kami	0.90	1.00	0.95	9
Kamu	1.00	0.89	0.94	9
Ketuk	1.00	1.00	1.00	9
Laki-Laki	1.00	1.00	1.00	9
perempuan	1.00	1.00	1.00	9
accuracy			0.99	99
macro avg	0.99	0.99	0.99	99
weighted avg	0.99	0.99	0.99	99

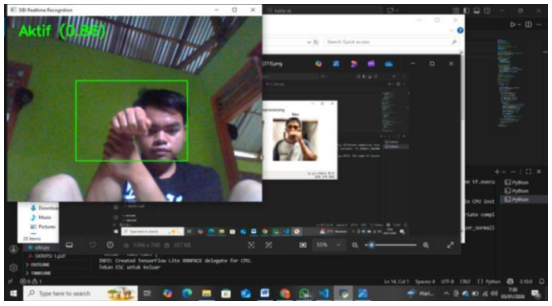
MODEL & CLASS NAMES BERHASIL DISIMPAN
SIAP DIGUNAKAN UNTUK REAL-TIME
PS C:\hatta sk>

Gambar 6. Hasil Classification Report

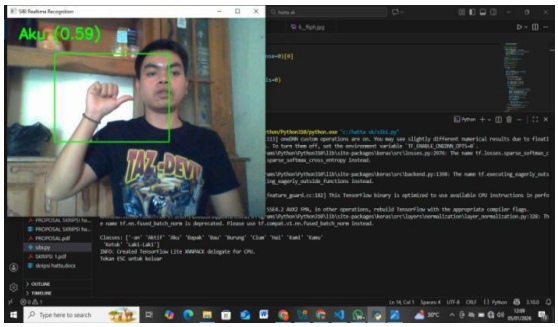
Hasil Classification Report menunjukkan bahwa model klasifikasi BISINDO mencapai akurasi sebesar 99%. Hampir seluruh kelas memiliki nilai precision, recall, dan f1-score mendekati atau mencapai 1.00, sehingga model dapat dikatakan memiliki performa yang sangat baik dan siap digunakan untuk implementasi real-time.

Pengujian Secara Real-Time

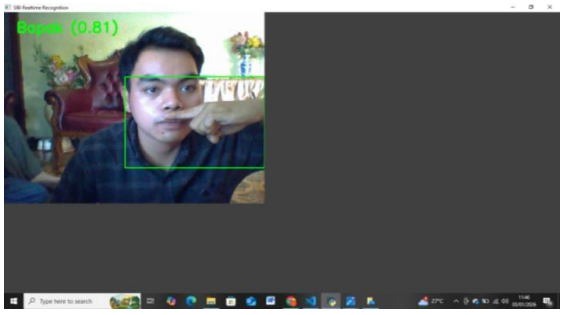
Model yang telah dilatih kemudian diimplementasikan dalam sistem real-time menggunakan kamera. Sistem mampu mendeteksi area tangan dan menampilkan label prediksi beserta tingkat kepercayaan (confidence score). Berikut ini adalah hasil dari pengujian sistem secara real-time:



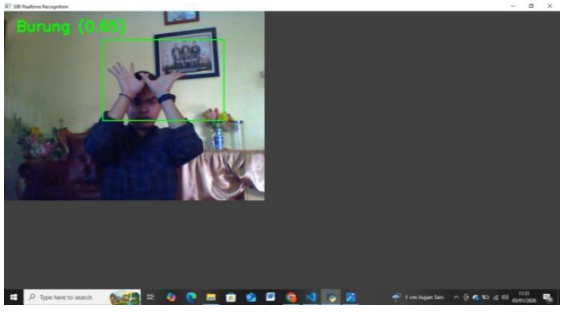
Gambar 7. Hasil Kelas Aktif



Gambar 8. Hasil Kelas Aku



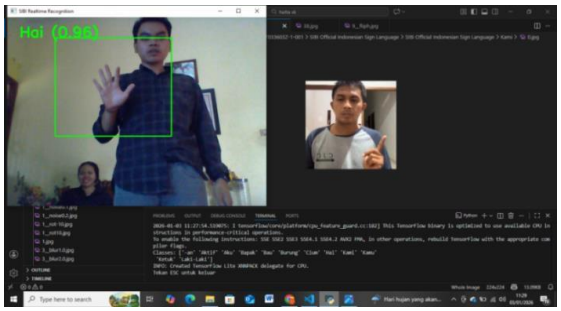
Gambar 9. Hasil Kelas Bapak



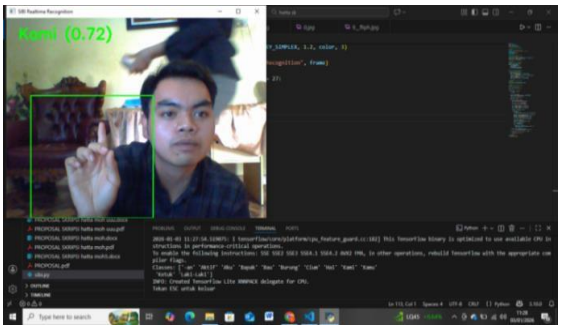
Gambar 10. Hasil Kelas Burung



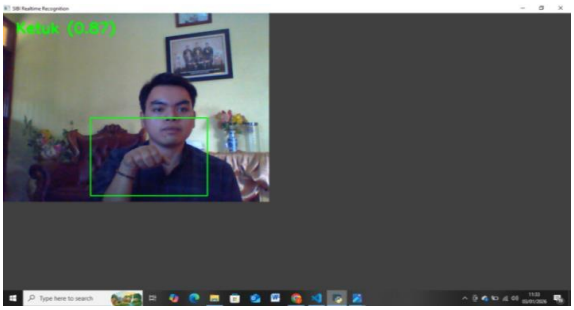
Gambar 11. Gambar Kelas Cium



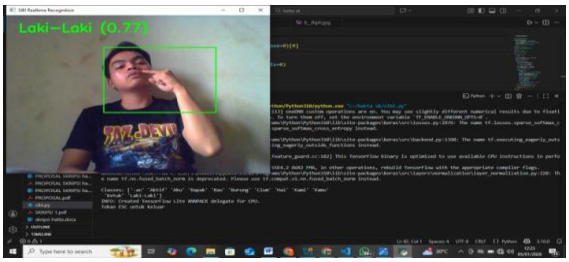
Gambar 12. Hasil Kelas Hai



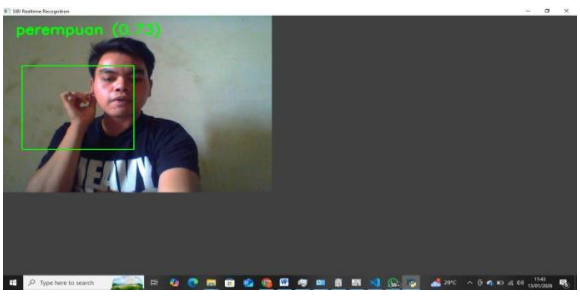
Gambar 13. Hasil Kelas Kami



Gambar 14. Hasil Kelas Ketuk



Gambar 15. Hasil Gambar Kelas Laki-Laki



Gambar 16. Hasil Kelas Perempuan

Berdasarkan gambar-gambar yang ditampilkan, seluruhnya merupakan hasil pengujian (testing) sistem klasifikasi Bahasa Isyarat Indonesia (BISINDO) secara real-time menggunakan kamera (webcam) dengan model Vision Transformer (ViT). Sistem terlebih dahulu mendeteksi area tangan menggunakan bounding box berwarna hijau, kemudian melakukan klasifikasi terhadap gestur yang ditampilkan.

Hasil prediksi ditampilkan pada bagian kiri atas berupa nama kelas beserta nilai confidence score yang menunjukkan tingkat keyakinan model terhadap prediksi tersebut. Pada pengujian ini, model berhasil mengenali beberapa gestur BISINDO, yaitu Aktif (0,86), Aku (0,59), Bapak (0,81), Burung (0,69), Cium (0,79), Hai (0,96), Kamu (0,72), Ketuk (0,87), Laki-laki (0,77), dan Perempuan (0,73). Nilai confidence yang diperoleh bervariasi antara 0,59 hingga 0,96.

Prediksi dengan nilai confidence tinggi, seperti pada kelas Hai dan Ketuk, menunjukkan bahwa model mampu mengenali pola gestur dengan sangat baik. Sementara itu, kelas seperti Aku dan Burung memiliki nilai confidence yang lebih rendah, yang mengindikasikan bahwa bentuk gestur, posisi tangan, pencahayaan, atau latar belakang masih memengaruhi tingkat keyakinan model dalam melakukan klasifikasi.

Secara keseluruhan, hasil pengujian real-time menunjukkan bahwa model mampu melakukan klasifikasi gestur BISINDO dengan baik pada berbagai variasi pose, jarak, dan pencahayaan. Meskipun masih terdapat perbedaan nilai confidence pada beberapa kelas, sistem tetap dapat memberikan prediksi yang sesuai dengan gestur yang ditampilkan. Hasil ini menunjukkan bahwa pendekatan Vision Transformer yang didukung teknik CLAHE dan MixUp efektif untuk diterapkan pada sistem pengenalan Bahasa Isyarat Indonesia berbasis citra secara real-time.

SIMPULAN

Penelitian ini berhasil mengoptimalkan klasifikasi Bahasa Isyarat Indonesia (BISINDO) menggunakan model Vision Transformer yang dipadukan dengan teknik Contrast Limited Adaptive Histogram Equalization (CLAHE) dan augmentasi data MixUp. Model yang diusulkan memperoleh akurasi sebesar 99%, dengan nilai rata-rata precision, recall, dan F1-score sebesar 0,99, yang menunjukkan kemampuan klasifikasi yang sangat baik pada seluruh kelas. Pengujian secara real-time juga menunjukkan bahwa sistem mampu mengenali berbagai gestur BISINDO dengan baik pada berbagai kondisi. Hasil penelitian membuktikan bahwa penerapan CLAHE dan MixUp efektif dalam meningkatkan kualitas fitur, kemampuan generalisasi, serta performa Vision Transformer, sehingga berpotensi diterapkan sebagai sistem pengenalan Bahasa Isyarat Indonesia berbasis citra secara real-time.

DAFTAR PUSTAKA

[1] D. Mahmud, A., Prasetyo, B., & Lestari, “Understanding Indonesian Sign Language (BISINDO) for Inclusive Communication,” *J. Disabil. Stud.*, vol. 5, pp. 282–287, 2020, doi: 10.25046/aj050535.

[2] M. Saputri and. Murfi, “Analisis Pemahaman Bahasa Isyarat di Kalangan Masyarakat Umum,” *urnal Pendidik. Inklusif Indones.*, 2021.

[3] and S. L. J. Khan, M. Hussain, “A Survey on Sign Language Recognition Using Computer Vision and Deep Learning,” *IEEE Access*, vol. 11, pp. 4730–4749, 2023, doi: https://doi.org/10.1109/ACCESS.2023.3234569.

[4] A. Budiman et al., “Sistem pembelajaran isyarat Bahasa Indonesia menggunakan CNN,” *J. Sist. Cerdas*, 2023.

[5] D. Rakhmadi et al, "CNN-RNN untuk pengenalan bahasa isyarat global," *J. Comput. Vis. Res.*, 2025.

[6] and R. B. O. Koller, S. Zargaran, N. C. Camgoz, "Weakly Supervised Learning for Sign Language RecognitionNo Title," *IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2019, doi: 10.1109/CVPR.2019.01083.

[7] et al A. Dosovitskiy, L. Beyer, A. Kolesnikov, "An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale," *Int. Conf. Learn. Represent.*, 2021.

[8] L. Wijaya, B., & Nurhayati, "Pengaruh pencahayaan pada sistem pengenalan citra," *urnal Inf. Multimed.*, 2021.

[9] and H. J. H. Touvron, M. Cord, M. Douze, F. Massa, A. Sablayrolles, "Training data-efficient image transformers & distillation through attention," *Int. Conf. Mach. Learn.*, pp. 10256–10266, 2021, doi: 10.48550/arXiv.2012.12877.

[10] M. Kelana, E. L., Riko, M., Prasetya, A., & Zulfadhilah, "Integrating the CNN model with the web for Indonesian Sign Language (BISINDO) recognition," vol. 9, 2025.

[11] & S. Mahendra, R., Laksana, E. A., "Pendekatan Transfer Learning dengan InceptionResNetV2 dan Augmentasi MixUp untuk Peningkatan Klasifikasi Tumor Otak," *J. Algoritm.*, vol. 22, pp. 161–172, 2025, doi: https://doi.org/10.33364/algoritma/v.22-1.2194.