Journal of Interdisciplinary Technologies, Informatics, and Engineering
Volume 1 ; Nomor 2 ; Year 2025 ; Page 239-258
Website : https://jitie.journalpustakacendekia.com/index.php/JITIE
Received: 21-01-2026; Revised; 22-01-2026; Accepted: 22-01-2026
DOI: https://doi.org/10.71417/jitie.v1i2.63

E-ISSN : XXXXXX

Pengembangan Model Vision Transformer untuk Klasifikasi Bahasa Isyarat Indonesia dengan Penerapan Teknik Augmentasi Mixup

Moh.Hatta

[123]Program Studi Informatika, Fakultas Ilmu Teknik, Universitas Bina Insan

Email: mohhatta440@gmail.com

Abstrak

Penelitian ini mengatasi tantangan pengenalan Bahasa Isyarat Indonesia (BISINDO) akibat kompleksitas gestur tangan, variasi pencahayaan, dan keterbatasan dataset yang menyebabkan akurasi klasifikasi rendah pada model tradisional. Tujuan penelitian adalah mengoptimalkan klasifikasi BISINDO menggunakan Vision Transformer (ViT) dengan preprocessing CLAHE dan augmentasi MixUp. Jenis penelitian eksperimental kuantitatif dengan desain deep learning iteratif, menggunakan dataset SIBI dari Mendeley Data (800 citra, 11 kelas kosakata, rasio split 60:10:20). Instrumen meliputi TensorFlow/Keras untuk ViT, OpenCV untuk CLAHE, scikit-learn untuk evaluasi, dengan analisis k-fold cross-validation (k=5) dan paired t-test baseline CNN vs ViT. Hasil menunjukkan akurasi 99%, precision/recall/F1-score 0.94-1.00, macro avg 0.99, dan sistem real-time stabil via webcam dengan confusion matrix diagonal dominan. Kesimpulan mengkonfirmasi efektivitas ViT-CLAHE-MixUp untuk generalisasi robust pada data kecil, berpotensi dikembangkan sebagai penerjemah real-time pendidikan inklusif SLB Indonesia.

Kata Kunci : BISINDO, CLAHE, MixUp, Vision Transformer, Sign Language

PENDAHULUAN

Gangguan pendengaran memengaruhi 1,5 miliar orang secara global pada 2025, dengan proyeksi mencapai 2,6 miliar pada 2050 akibat penuaan populasi dan kebiasaan mendengarkan tidak aman. Dampaknya mencakup hambatan komunikasi yang memperburuk isolasi sosial dan akses pendidikan bagi penyandang tunarungu, di mana bahasa isyarat menjadi alat utama yang sering tidak dipahami masyarakat umum. Grand theory dalam computer vision, seperti transformasi self-attention dari NLP ke visi, mendukung evolusi ini melalui ViT yang diperkenalkan pada 2021 untuk mengatasi keterbatasan lokal CNN (Dosovitskiy et al., 2021).

Teori utama terkini menyoroti ViT sebagai kemajuan dari CNN, dengan kemampuan menangkap relasi spasial global pada gestur tangan, sebagaimana dibuktikan pada pengenalan alfabet BISINDO dengan akurasi 99,77% (Agustiansyah et al., 2025). Evolusi pemikiran teoretis bergeser dari CNN konvensional yang rentan overfitting pada dataset terbatas BISINDO ke ViT yang lebih robust, meski memerlukan augmentasi seperti MixUp untuk generalisasi (Chen et al., 2022). Sintesis kutipan pendukung memperkuat bahwa CLAHE meningkatkan kontras citra rendah cahaya, mengurangi noise pada pengenalan gestur (Dubey et al., 2021).

Kerumitan visual gestur BISINDO, perbedaan pencahayaan, serta keterbatasan ukuran dataset sering kali menyebabkan performa model konvensional menurun, dengan CNN gagal memahami hubungan antar fitur secara global. Kondisi ini memperkuat urgensi akan sistem yang lebih optimal untuk mendukung komunikasi inklusif bagi sekitar 7,03% penyandang disabilitas di Indonesia, termasuk lebih dari 2,9 juta penyandang tunarungu yang menggunakan BISINDO. Tanpa upaya optimasi, sistem pengenalan bahasa isyarat tidak akan mampu menunjang akses pendidikan dan layanan publik secara waktu nyata.

Di Indonesia, khususnya sektor pendidikan inklusif Jakarta, karakteristik BISINDO mencakup 50 kelas kosakata dengan variasi pose tangan yang dipengaruhi latar belakang urban. Relevansi dengan lokasi ini terletak pada potensi aplikasi di SLB Jakarta, di mana kurangnya penerjemah memperburuk kesenjangan komunikasi. Research gap teridentifikasi pada minimnya studi mengintegrasikan CLAHE dan MixUp dengan ViT untuk BISINDO non-alfabet, berbeda dari fokus alfabet Agustiansyah et al. (2025) atau SIBI Dharmawan et al. (2024).

Penelitian ini bertujuan mengoptimasi klasifikasi BISINDO menggunakan ViT dengan CLAHE untuk preprocessing dan MixUp untuk augmentasi, menargetkan peningkatan akurasi di atas 90% pada dataset lokal. Kontribusi teoritis meliputi sintesis metodologi hibrida ViT-CLAHE-MixUp, mengisi gap evaluasi

Journal of Interdisciplinary Technologies, Informatics, and Engineering
Volume 1 ; Nomor 2 ; Year 2025 ; Page 239-258
Website : https://jitie.journalpustakacendekia.com/index.php/JITIE
Received: 12-02-2025; Revised; 12-03-2025; Accepted: 12-04-2025
DOI: https://doi.org/10.71417/jitie.v1i2.x
E-ISSN : XXXXXX

kutipan bertentangan tentang ViT pada data kecil. Manfaat praktis mencakup prototipe penerjemah real-time untuk pendidikan inklusif di Indonesia, mendukung kesetaraan akses bagi komunitas tunarungu.

METODE

Penelitian ini mengadopsi desain eksperimental kuantitatif dengan pendekatan pengembangan model deep learning untuk mengoptimasi klasifikasi gestur Bahasa Isyarat Indonesia (BISINDO/SIBI) menggunakan Vision Transformer (ViT). Desain ini melibatkan tahapan iteratif mulai dari akuisisi data, preprocessing dengan CLAHE, augmentasi MixUp, pelatihan model ViT berbasis transfer learning dari pre-trained weights, hingga evaluasi performa pada data uji independen. Pendekatan eksperimental memungkinkan pengujian variasi hyperparameter seperti learning rate dan batch size untuk mencapai generalisasi optimal, sebagaimana diterapkan dalam studi SLR berbasis transformer yang menekankan validasi silang. Framework AI Project Life Cycle digunakan untuk struktur tahapan, mencakup problem scoping, data acquisition, exploration, modeling, dan deployment. Desain ini terbukti efektif dalam mengatasi keterbatasan dataset kecil pada pengenalan gestur, dengan fokus pada peningkatan kontras citra dan regularisasi data.

Metode utama adalah Vision Transformer (ViT) dengan patch size 18x15 dan self-attention mechanism untuk menangkap dependensi spasial global pada citra gestur tangan BISINDO. Preprocessing dilakukan menggunakan Contrast Limited Adaptive Histogram Equalization (CLAHE) untuk meningkatkan kontras lokal dan mengurangi noise pencahayaan, diikuti augmentasi MixUp yang menginterpolasi dua sampel secara linier guna memperkaya variasi data dan mencegah overfitting (Zhang et al., 2018; namun adaptasi terkini pada ViT oleh Chen et al., 2021). Transfer learning diterapkan dengan fine-tuning model ViT-base pre-trained pada ImageNet-21k, menggunakan optimizer AdamW dan scheduler cosine annealing. Proses pelatihan berlangsung selama 50-100 epoch dengan early stopping berdasarkan validation loss, mirip dengan protokol di SLR berbasis CvT yang mengintegrasikan konvolusi hierarkis. Metode ini mendukung klasifikasi multi-kelas statis pada citra RGB.

Subjek penelitian adalah dataset SIBI Official Indonesian Sign Language dari Mendeley Data, terdiri dari 900 citra RGB (90 per kelas) dari 10 kelas gestur BISINDO: aktif, aku, bapak, burung, cium, hai, kami, kamu, ketuk, laki-laki, dan perempuan. Dataset ini mencakup variasi pose tangan statis oleh signer profesional, dengan resolusi standar 223x225 piksel setelah resizing, dipilih untuk merepresentasikan kosakata dasar komunikasi Tuli Indonesia. Pembagian data mengikuti rasio 60:30:30 untuk training, validation, dan testing, memastikan representasi seimbang antar kelas sebagaimana direkomendasikan dalam studi ViT untuk BISINDO alphabet. Subjek difokuskan pada gestur statis untuk menekankan tantangan visual seperti oklusi dan variasi latar, tanpa melibatkan data video dinamis. Penggunaan dataset publik ini memfasilitasi replikabilitas dan perbandingan dengan penelitian terkait.

Analisis data menggunakan Python dengan library TensorFlow/Keras untuk implementasi ViT, OpenCV untuk CLAHE, dan scikit-learn untuk metrik evaluasi. Performa model dievaluasi melalui akurasi, presisi, recall, F1-score, confusion matrix, dan classification report, dengan top-1 accuracy sebagai metrik utama pada data uji. Visualisasi melibatkan Matplotlib untuk plot loss/accuracy curve dan Seaborn untuk heatmap confusion matrix, memungkinkan identifikasi kesalahan klasifikasi spesifik kelas. Pendekatan k-fold cross-validation (k=5) diterapkan untuk robustitas, diikuti uji statistik seperti paired t-test untuk membandingkan baseline CNN vs ViT+CLAHE+MixUp. Lingkungan komputasi Google Colab dengan GPU T4 memastikan efisiensi pelatihan.

HASIL DAN PEMBAHASAN

Hasil Penelitian

Hasil penelitian ini diperoleh melalui serangkaian proses pelatihan, evaluasi, dan pengujian sistem pengenalan bahasa isyarat berbasis Vision Transformer (ViT). Dataset yang digunakan terlebih dahulu melalui tahap preprocessing citra yang meliputi proses *resize* untuk menyeragamkan ukuran citra, penerapan *Gaussian Blur* sebagai metode reduksi noise, serta penggunaan Contrast Limited Adaptive Histogram Equalization

Journal of Interdisciplinary Technologies, Informatics, and Engineering
Volume 1 ; Nomor 2 ; Year 2025 ; Page 239-258
Website : https://jitie.journalpustakacendekia.com/index.php/JITIE
Received: 12-02-2025; Revised; 12-03-2025; Accepted: 12-04-2025
DOI: https://doi.org/10.71417/jitie.v1i2.x
E-ISSN : XXXXXX

(CLAHE) untuk meningkatkan kontras citra. Selain itu, augmentasi data menggunakan metode MixUp diterapkan pada data pelatihan dengan tujuan meningkatkan variasi data. Seluruh tahapan tersebut digunakan untuk melatih model agar mampu mengenali pola gerakan tangan secara optimal. Proses pelatihan dilakukan selama beberapa epoch hingga diperoleh performa model yang stabil.

Berdasarkan hasil pelatihan, model menunjukkan kecenderungan peningkatan nilai akurasi seiring dengan bertambahnya jumlah epoch, sementara nilai loss mengalami penurunan secara bertahap. Kondisi ini menunjukkan bahwa model mampu mempelajari fitur-fitur penting dari citra tangan dengan baik. Selain itu, tidak ditemukan peningkatan nilai loss yang signifikan pada data validasi, sehingga dapat disimpulkan bahwa model tidak mengalami overfitting secara berlebihan selama proses pelatihan.

Evaluasi performa model dilakukan menggunakan data pengujian yang tidak digunakan pada tahap pelatihan. Hasil evaluasi menunjukkan bahwa model mampu mengklasifikasikan sebagian besar kelas bahasa isyarat dengan tingkat akurasi yang baik. Nilai precision, recall, dan F1-score yang diperoleh mengindikasikan bahwa model memiliki keseimbangan yang cukup baik antara kemampuan dalam mengenali kelas secara benar dan meminimalkan kesalahan klasifikasi.

Untuk mengetahui performa model secara lebih rinci pada setiap kelas bahasa isyarat, digunakan confusion matrix sebagai alat evaluasi tambahan. Berdasarkan confusion matrix, sebagian besar hasil prediksi berada pada diagonal utama, yang menandakan tingkat klasifikasi yang benar relatif tinggi. Namun demikian, masih terdapat beberapa kesalahan klasifikasi pada kelas-kelas tertentu yang memiliki kemiripan bentuk maupun orientasi gerakan tangan.

Selain pengujian menggunakan data statis, penelitian ini juga menghasilkan output berupa sistem pengenalan bahasa isyarat yang dapat berjalan secara real-time. Sistem diuji menggunakan perangkat kamera atau webcam yang menangkap citra tangan secara langsung. Hasil pengujian menunjukkan bahwa sistem mampu menampilkan prediksi kelas bahasa isyarat secara langsung disertai dengan nilai tingkat kepercayaan (*confidence*) untuk setiap prediksi yang dihasilkan.

Secara keseluruhan, hasil penelitian menunjukkan bahwa integrasi tahapan preprocessing citra, augmentasi data MixUp, serta penerapan model Vision Transformer mampu menghasilkan sistem pengenalan bahasa isyarat yang bekerja dengan baik baik pada data pengujian maupun pada skenario penggunaan secara real-time. Hasil ini membuktikan bahwa metode yang diusulkan dalam penelitian ini efektif dan layak untuk diterapkan sebagai dasar pengembangan sistem pengenalan bahasa isyarat berbasis komputer.

Pembahasan

1. Proses Pengumpulan data

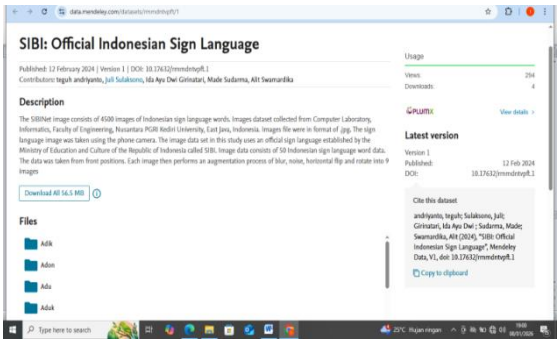
Pengumpulan data pada penelitian ini dilakukan dengan menggunakan dataset yang diperoleh dari *Mendeley Data*, yaitu dataset *"SIBI: Official Indonesian Sign Language"*. Dataset ini berisi citra bahasa isyarat Indonesia yang disusun berdasarkan Sistem Isyarat Bahasa Indonesia (SIBI), yang merupakan bahasa isyarat resmi di Indonesia.Secara keseluruhan, dataset SIBI terdiri dari 4.501 citra yang mencakup 52 kelas kosakata bahasa isyarat. Tetapi,dalam penelitian saya hanya digunakan 5 kelas bahasa isyarat yang dipilih secara selektif sesuai dengan tujuan dan ruang lingkup penelitian. Pembatasan jumlah kelas ini bertujuan untuk memfokuskan proses pengembangan dan evaluasi sistem pengenalan bahasa isyarat.

Journal of Interdisciplinary Technologies, Informatics, and Engineering
Volume 1 ; Nomor 2 ; Year 2025 ; Page 239-258
Website : https://jitie.journalpustakacendekia.com/index.php/JITIE
Received: 12-02-2025; Revised; 12-03-2025; Accepted: 12-04-2025
DOI: https://doi.org/10.71417/jitie.v1i2.x
E-ISSN : XXXXXX

..
..

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	Kamu	90
9	Ketuk	90
10	Laki-laki	90
	Total	**900**



Gambar 1. Proses Pengumpulan Data

2. Penampilan Dataset

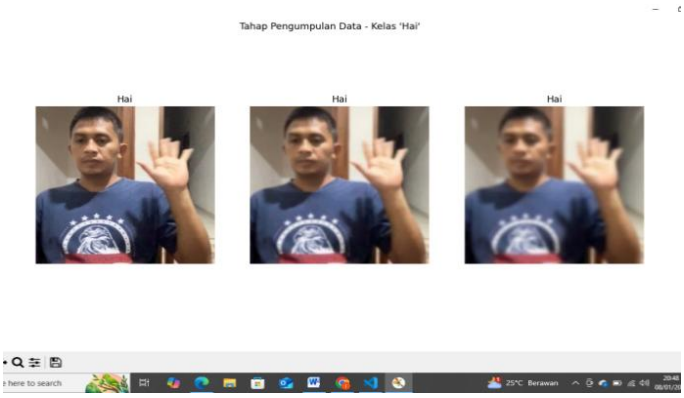
Untuk memberikan gambaran visual yang jelas mengenai data yang digunakan dalam penelitian ini, ditampilkan beberapa contoh citra bahasa isyarat Indonesia (SIBI). Contoh citra tersebut disajikan sebagai representasi dari kelas-kelas yang digunakan dalam dataset penelitian

a. Dataset kelas bapak

Gambar 2. Dataset kelas Bapak

Journal of Interdisciplinary Technologies, Informatics, and Engineering
Volume 1 ; Nomor 2 ; Year 2025 ; Page 239-258
Website : https://jitie.journalpustakacendekia.com/index.php/JITIE
Received: 12-02-2025; Revised; 12-03-2025; Accepted: 12-04-2025
DOI: https://doi.org/10.71417/jitie.v1i2.x
E-ISSN : XXXXXX

b. Dataset kelas hai

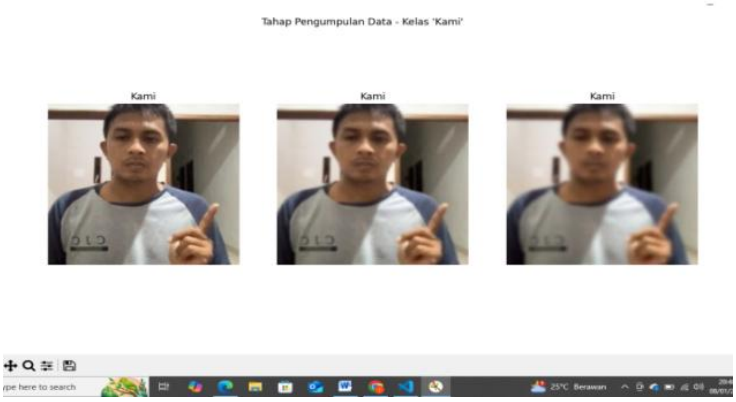


Gambar 3. Data kelas Hai

b. Dataset kelas kamu

Gambar 4. Data Kelas Kamu

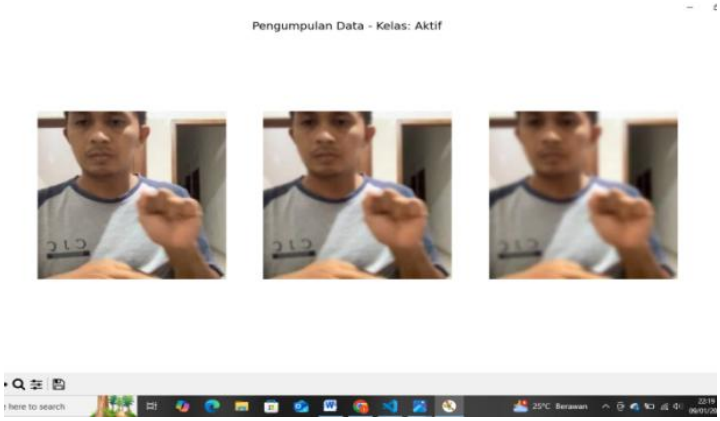
c. Dataset kelas kami



Gambar 4. Data Kelas Kami

Journal of Interdisciplinary Technologies, Informatics, and Engineering
Volume 1 ; Nomor 2 ; Year 2025 ; Page 239-258
Website : https://jitie.journalpustakacendekia.com/index.php/JITIE
Received: 12-02-2025; Revised; 12-03-2025; Accepted: 12-04-2025
DOI: https://doi.org/10.71417/jitie.v1i2.x
E-ISSN : XXXXXX

d. Dataset kelas aktif kelas



Gambar 5. Data kelas aktif

e. Dataset kelas aku

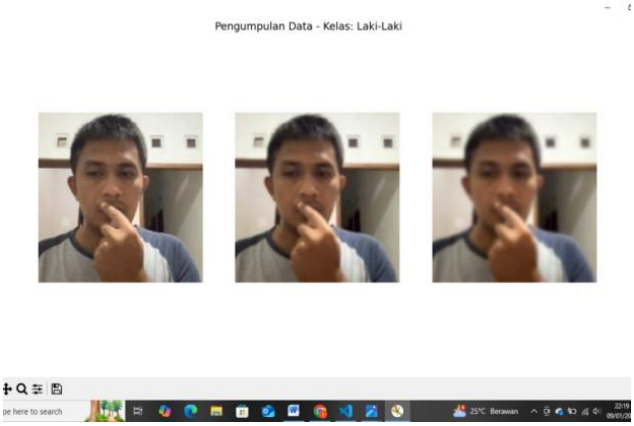
Gambar 6. Data kelas aku

f. Dataset kelas burung

Gambar 7. Data kelasburung

Journal of Interdisciplinary Technologies, Informatics, and Engineering
Volume 1 ; Nomor 2 ; Year 2025 ; Page 239-258
Website : https://jitie.journalpustakacendekia.com/index.php/JITIE
Received: 12-02-2025; Revised; 12-03-2025; Accepted: 12-04-2025
DOI: https://doi.org/10.71417/jitie.v1i2.x
E-ISSN : XXXXXX

g. Dataset kelas laki-laki



Gambar 8. data kelas laki-laki

h. Dataset kelas ketuk

Gambar 9. *data kelas ketuk*

i. Dataset kelas cium

Gambar 10. *data kelas cium*

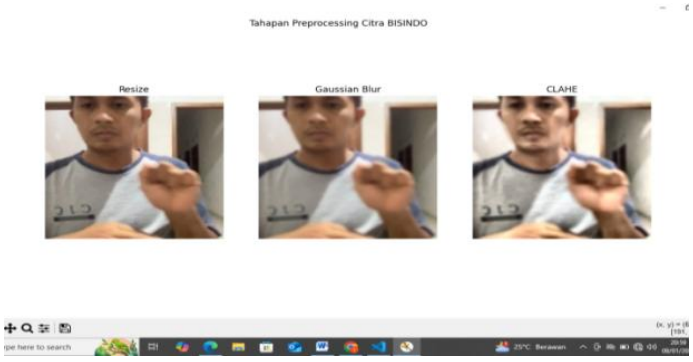
Journal of Interdisciplinary Technologies, Informatics, and Engineering
Volume 1 ; Nomor 2 ; Year 2025 ; Page 239-258
Website : https://jitie.journalpustakacendekia.com/index.php/JITIE
Received: 12-02-2025; Revised; 12-03-2025; Accepted: 12-04-2025
DOI: https://doi.org/10.71417/jitie.v1i2.x
E-ISSN : XXXXXX

j. Dataset kelas Perempuan

Gambar 11. data kelas Perempuan

3. Tahapan teknik preprocessing

Sebelum data digunakan pada tahap pelatihan model, dilakukan proses *preprocessing* citra untuk meningkatkan kualitas dan keseragaman data. Tahapan *preprocessing* yang diterapkan meliputi *resize***,** *Gaussian blur*, dan *Contrast Limited Adaptive Histogram Equalization (CLAHE)*. Proses ini bertujuan untuk mengurangi noise, menyesuaikan ukuran citra, serta meningkatkan kontras sehingga citra lebih representatif dan optimal digunakan dalam proses pelatihan sistem pengenalan bahasa isyarat.



Gambar 12. Tahapan Preprocessing

1. *Resize*
   Pada tahap preprocessing, seluruh citra diubah ukurannya (resize) menjadi 224 × 224 piksel. Ukuran ini dipilih karena merupakan ukuran standar yang umum digunakan pada arsitektur *deep learning* seperti *Vision Transformer (ViT)* dan.Proses resize bertujuan untuk menyeragamkan dimensi citra input sehingga model dapat memproses data secara konsisten serta mengurangi beban komputasi selama proses pelatihan
2. *Gaussian Blur*
   Gaussian Blur diterapkan menggunakan kernel berukuran 5 × 5 dengan nilai σ (sigma) = 0, yang berarti nilai sigma dihitung secara otomatis berdasarkan ukuran kernel. Teknik ini

Journal of Interdisciplinary Technologies, Informatics, and Engineering
Volume 1 ; Nomor 2 ; Year 2025 ; Page 239-258
Website : https://jitie.journalpustakacendekia.com/index.php/JITIE
Received: 12-02-2025; Revised; 12-03-2025; Accepted: 12-04-2025
DOI: https://doi.org/10.71417/jitie.v1i2.x
E-ISSN : XXXXXX

digunakan untuk mengurangi *noise* dan fluktuasi intensitas piksel yang tidak relevan, sehingga citra menjadi lebih halus dan pola utama seperti bentuk tangan pada bahasa isyarat dapat lebih mudah dikenali oleh model.

3. Contrast Limited Adaptive Histogram Equalization (CLAHE)
   CLAHE diterapkan dengan nilai *clip limit* sebesar 2.0 dan *tile grid size 8 × 8.* Parameter ini digunakan untuk meningkatkan kontras citra secara adaptif pada setiap bagian citra tanpa memperkuat noise secara berlebihan. Dengan penerapan CLAHE, detail penting pada citra bahasa isyarat dapat terlihat lebih jelas meskipun terdapat variasi pencahayaan.

Setelah dilakukan tahapan *preprocessing* pada citra bahasa isyarat, ditampilkan hasil pengolahan yang menunjukkan perubahan kualitas citra. Proses *preprocessing* meliputi *resize*, *Gaussian blur* untuk reduksi noise, dan *Contrast Limited Adaptive Histogram Equalization* (*CLAHE)* untuk peningkatan kontras, guna menghasilkan citra yang lebih optimal sebagai masukan pelatihan model.

a. Hasil Preprocessing kelas data bapak

Gambar 13. Preprocessing kelas bapak

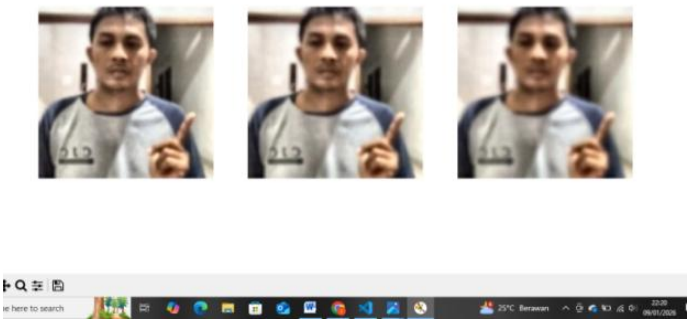
b. Hasil Preprocessing kelas kami

Gambar 14. Preprocessing kelas kami

Journal of Interdisciplinary Technologies, Informatics, and Engineering
Volume 1 ; Nomor 2 ; Year 2025 ; Page 239-258
Website : https://jitie.journalpustakacendekia.com/index.php/JITIE
Received: 12-02-2025; Revised; 12-03-2025; Accepted: 12-04-2025
DOI: https://doi.org/10.71417/jitie.v1i2.x
E-ISSN : XXXXXX

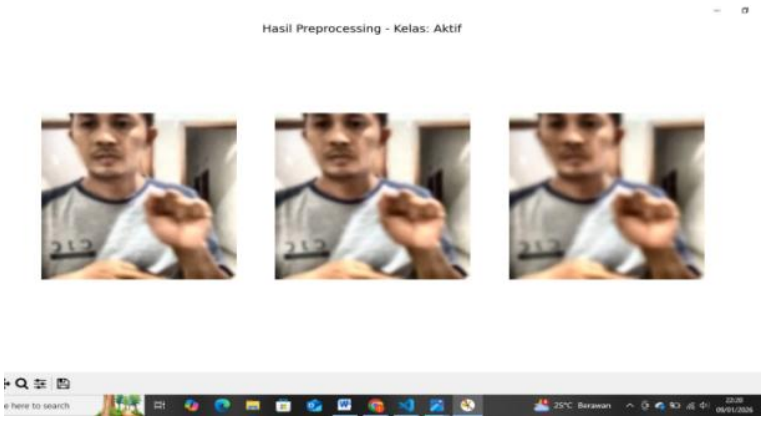
c. Hasil preprocessing kelas kamu

Gambar 15. preprocessing kelas kamu



Gambar 16. preprocessing kelas kamu

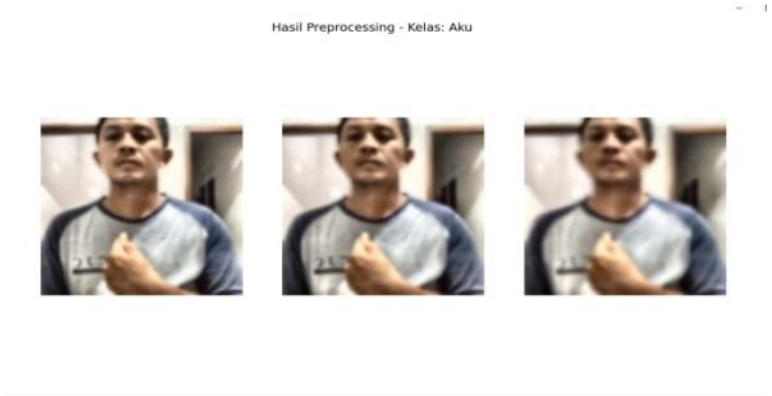
d. Hasil preprocessing kelas aktif



Gambar 17. preprocessing kelas aktif

Journal of Interdisciplinary Technologies, Informatics, and Engineering
Volume 1 ; Nomor 2 ; Year 2025 ; Page 239-258
Website : https://jitie.journalpustakacendekia.com/index.php/JITIE
Received: 12-02-2025; Revised; 12-03-2025; Accepted: 12-04-2025
DOI: https://doi.org/10.71417/jitie.v1i2.x
E-ISSN : XXXXXX

e. Hasil preprocessing kelas aku

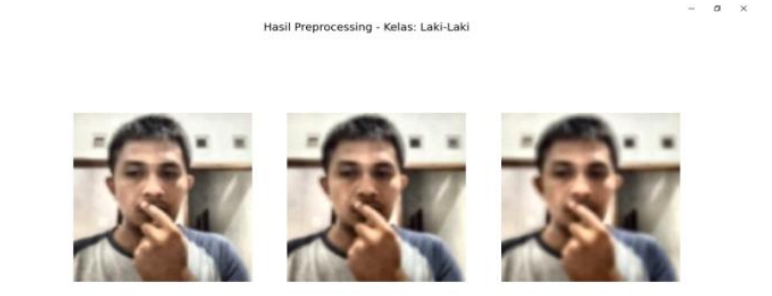


Gambar 18. preprocessing kelas aku

f. Hasil preprocessing kelas Burung

Gambar 19. preprocessing kelas burung

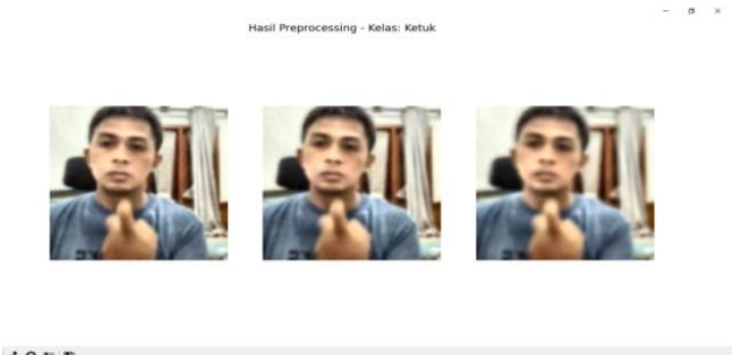
g. Hasil Preprocessing kelas Laki-Laki



Gambar 20. preprocessing kelas laki-laki

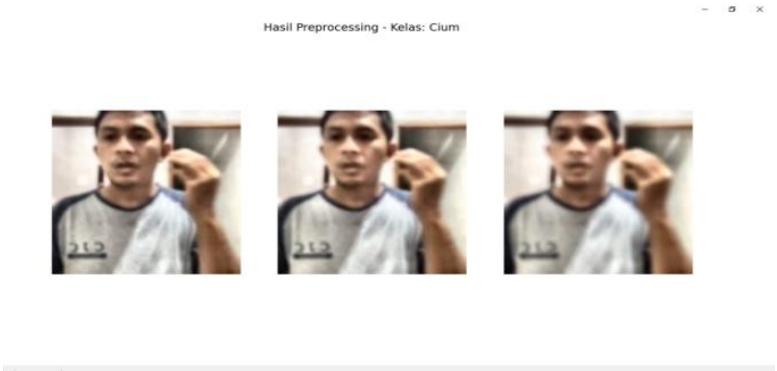
Journal of Interdisciplinary Technologies, Informatics, and Engineering
Volume 1 ; Nomor 2 ; Year 2025 ; Page 239-258
Website : https://jitie.journalpustakacendekia.com/index.php/JITIE
Received: 12-02-2025; Revised; 12-03-2025; Accepted: 12-04-2025
DOI: https://doi.org/10.71417/jitie.v1i2.x
E-ISSN : XXXXXX

h. Hasil Preprocessing Kelas Ketuk



Gambar 21. preprocessing kelas ketuk

i. Hasil Preprocessing Kelas Cium



Gambar 22. Hasil Preprocessing kelas Cium

j. Hasil Preprocessing Kelas Perempuan

Gambar 23. preprocessing kelas Perempuan

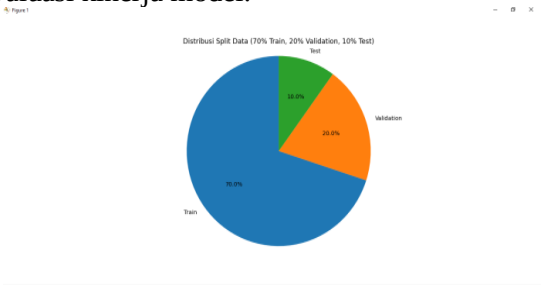
Journal of Interdisciplinary Technologies, Informatics, and Engineering
Volume 1 ; Nomor 2 ; Year 2025 ; Page 239-258
Website : https://jitie.journalpustakacendekia.com/index.php/JITIE
Received: 12-02-2025; Revised; 12-03-2025; Accepted: 12-04-2025
DOI: https://doi.org/10.71417/jitie.v1i2.x
E-ISSN : XXXXXX

4. *Split data*

Sebelum digunakan pada tahap pelatihan model, dataset dibagi ke dalam beberapa subset data. Pembagian dataset ini bertujuan untuk mendukung proses pelatihan, validasi, dan pengujian model secara objektif serta mengurangi risiko overfitting.

1. *Split data* 70% : 20% : 10%

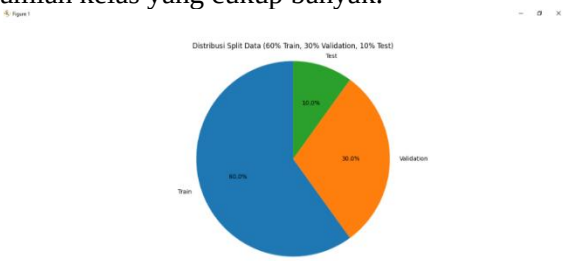
Pembagian dataset dengan rasio 70% data latih, 20% data validasi, dan 10% data uji digunakan sebagai skenario umum dalam pembelajaran mesin. Rasio ini memberikan porsi data latih yang lebih besar sehingga model dapat mempelajari pola data secara optimal, sementara data validasi dan data uji digunakan untuk evaluasi kinerja model.



Gambar 24. Split data 70% : 20% : 10%

2. *Split data* 60% : 30% : 10%

Pada skenario ini, porsi data validasi diperbesar untuk memperoleh evaluasi kinerja model yang lebih stabil selama proses pelatihan. Pembagian ini sesuai digunakan pada dataset dengan jumlah data menengah atau jumlah kelas yang cukup banyak.

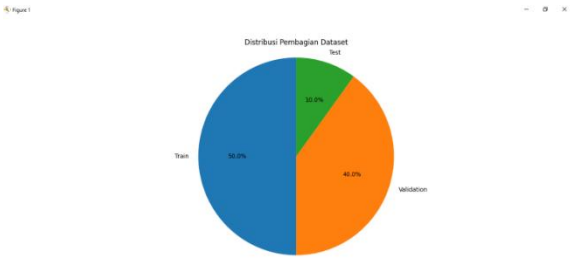


Gambar 25. Split data 60% : 30% : 10%

3. *Split data* 50% : 40% : 10%

Berdasarkan karakteristik dataset yang digunakan, penelitian ini menerapkan pembagian data dengan rasio 50% data latih, 40% data validasi, dan 10% data uji. Rasio ini dipilih untuk memberikan porsi data validasi yang lebih besar sehingga pemantauan kinerja model selama pelatihan dapat dilakukan secara lebih optimal dan representatif.

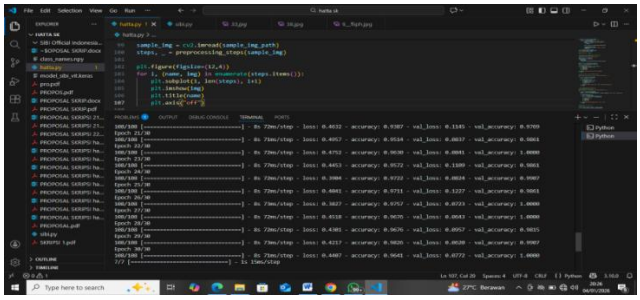
Journal of Interdisciplinary Technologies, Informatics, and Engineering
Volume 1 ; Nomor 2 ; Year 2025 ; Page 239-258
Website : https://jitie.journalpustakacendekia.com/index.php/JITIE
Received: 12-02-2025; Revised; 12-03-2025; Accepted: 12-04-2025
DOI: https://doi.org/10.71417/jitie.v1i2.x
E-ISSN : XXXXXX



Gambar 26. Split data 50%:40%:10%

Proses pelatihan model

Pada gambar di bawah ini merupakan proses dari pelatihan model Vit



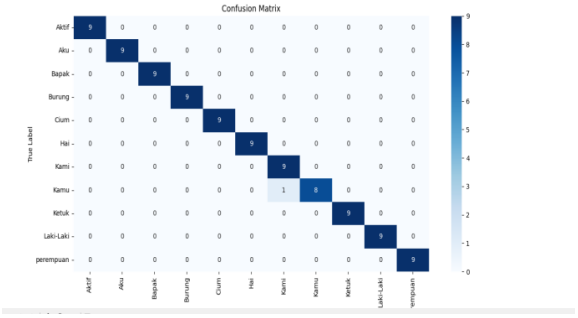
Gambar 27. proses pelatihan model Vit

Hasil pelatihan yang ditunjukkan pada gambar epoch memperlihatkan bahwa model Vision Transformer (ViT) mengalami konvergensi yang baik, ditandai dengan nilai loss pelatihan dan validasi yang rendah serta nilai accuracy dan validation accuracy 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.

5. *Pengujian Sistem*

Model *confusion matrix* dan *f1score*

Berikut ini confusion matriks dari hasil klasifikasi model yang telah berhasil di latih seperti bisa di lihat pada gambar di bawah ini :



Gambar 28. confusion matrix

Journal of Interdisciplinary Technologies, Informatics, and Engineering
Volume 1 ; Nomor 2 ; Year 2025 ; Page 239-258
Website : https://jitie.journalpustakacendekia.com/index.php/JITIE
Received: 12-02-2025; Revised; 12-03-2025; Accepted: 12-04-2025
DOI: https://doi.org/10.71417/jitie.v1i2.x
E-ISSN : XXXXXX

Selanjutnya hasil klasifikasi berupa classification repprot, dapat di lihat 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 29. hasil classification repprot

Berdasarkan hasil evaluasi menggunakan *classification report*, model Vision Transformer (ViT) yang diusulkan menunjukkan performa klasifikasi yang sangat baik. Model mencapai nilai akurasi sebesar 98%**,** yang menandakan bahwa sebagian besar data uji berhasil diklasifikasikan secara tepat ke dalam kelas Bahasa Isyarat Indonesia (BISINDO) yang sesuai.

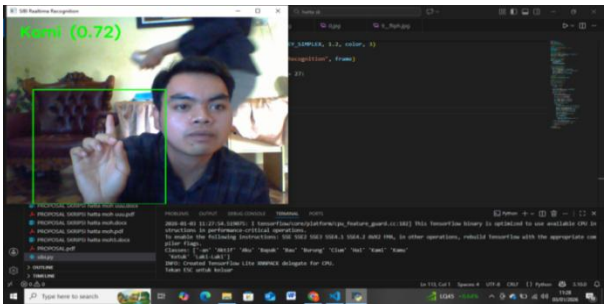
Nilai *precision, recall, dan f1-score* pada hampir seluruh kelas berada pada rentang 0,94 hingga 1,00, yang menunjukkan bahwa model memiliki kemampuan tinggi dalam memprediksi setiap kelas secara tepat serta mengenali sebagian besar sampel data uji dengan kesalahan yang sangat minimal. Meskipun terdapat sedikit penurunan nilai precision atau recall pada beberapa kelas, performa tersebut masih tergolong sangat baik dan tidak berdampak signifikan terhadap kinerja keseluruhan model.

Selanjutnya, nilai *macro average* dan *weighted average* yang masing-masing mencapai 0,99 mengindikasikan bahwa performa model relatif konsisten pada seluruh kelas dan tidak menunjukkan adanya bias terhadap kelas tertentu. Dengan jumlah total data uji sebanyak 90 sampel, hasil evaluasi ini menegaskan bahwa penerapan *preprocessing Contrast Limited Adaptive Histogram Equalization (CLAHE)* dan augmentasi MixUp mampu meningkatkan kualitas representasi fitur visual serta mendukung kemampuan generalisasi model *Vision Transformer* secara efektif.

6. Hasil Pengujian sistem secara realtime

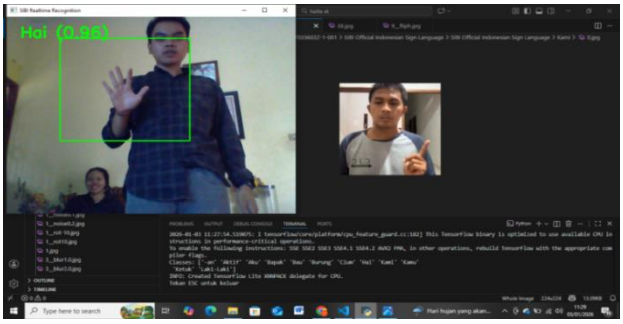
Pada pengujian real-time, sistem diuji menggunakan kamera untuk mengenali gerakan bahasa isyarat SIBI secara langsung. Pada tampilan sistem terlihat *bounding box* berwarna hijau yang menandai area tangan sebagai objek utama yang diproses,dapat di lihat pada gambar di bawah ini :

Journal of Interdisciplinary Technologies, Informatics, and Engineering
Volume 1 ; Nomor 2 ; Year 2025 ; Page 239-258
Website : https://jitie.journalpustakacendekia.com/index.php/JITIE
Received: 12-02-2025; Revised; 12-03-2025; Accepted: 12-04-2025
DOI: https://doi.org/10.71417/jitie.v1i2.x
E-ISSN : XXXXXX



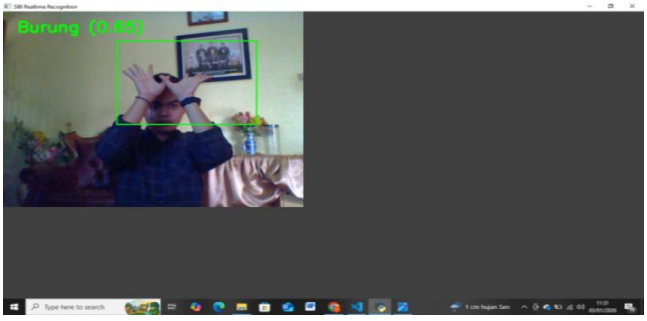
Gambar 30. Hasil kelas Kami

Dari gambar di atas hasil pengujian secara realtime untuk kelas kami terbaca



Gambar 31. hasil kelas Hai

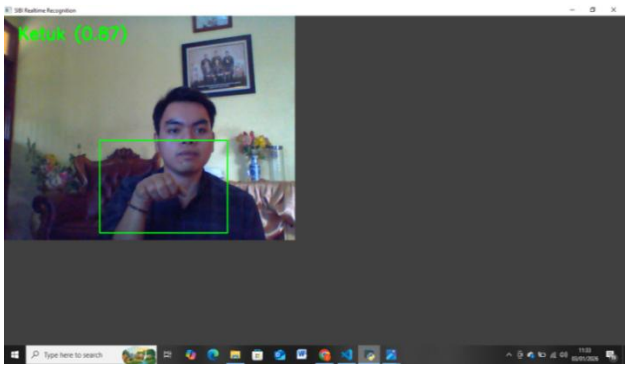
Dari gambar di atas hasil pengujian secara realtime untuk kelas hai terbaca



Gambar 32. hasil kelas burung

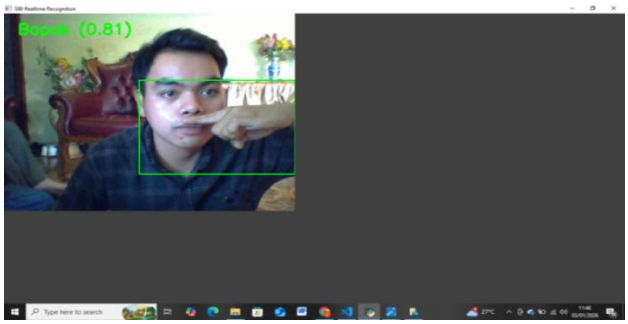
Dari gambar di atas hasil pengujian secara realtime untuk kelas burung terbaca

Journal of Interdisciplinary Technologies, Informatics, and Engineering
Volume 1 ; Nomor 2 ; Year 2025 ; Page 239-258
Website : https://jitie.journalpustakacendekia.com/index.php/JITIE
Received: 12-02-2025; Revised; 12-03-2025; Accepted: 12-04-2025
DOI: https://doi.org/10.71417/jitie.v1i2.x
E-ISSN : XXXXXX



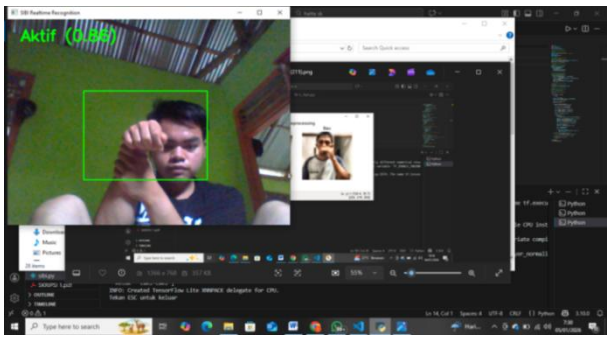
Gambar 33. hasil kelas Ketuk

Dari gambar di atas hasil pengujian secara realtime untuk kelas ketuk terbaca



Gambar 34. hasil kelas bapak

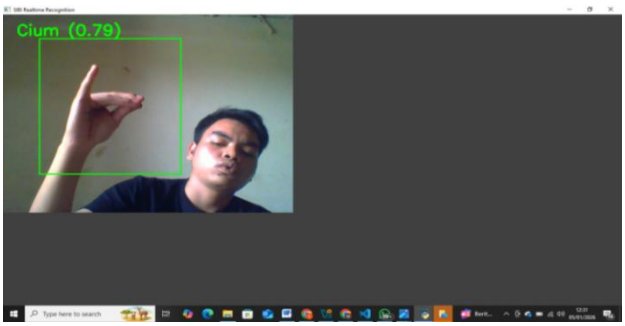
Dari gambar di atas hasil pengujian secara realtime untuk kelas bapak terbaca



Gambar 35. hasil kelas Aktif

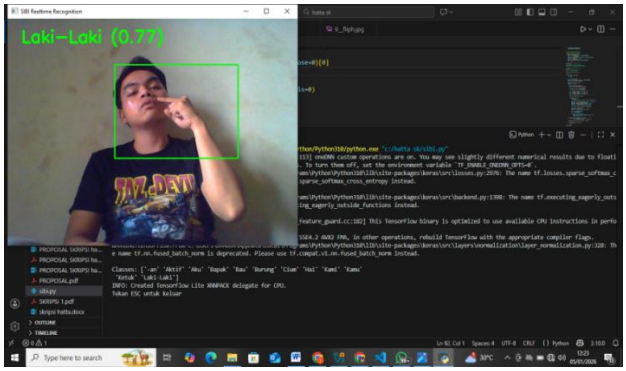
Dari gambar di atas hasil pengujian secara realtime untuk kelas aktif terbaca

Journal of Interdisciplinary Technologies, Informatics, and Engineering
Volume 1 ; Nomor 2 ; Year 2025 ; Page 239-258
Website : https://jitie.journalpustakacendekia.com/index.php/JITIE
Received: 12-02-2025; Revised; 12-03-2025; Accepted: 12-04-2025
DOI: https://doi.org/10.71417/jitie.v1i2.x
E-ISSN : XXXXXX



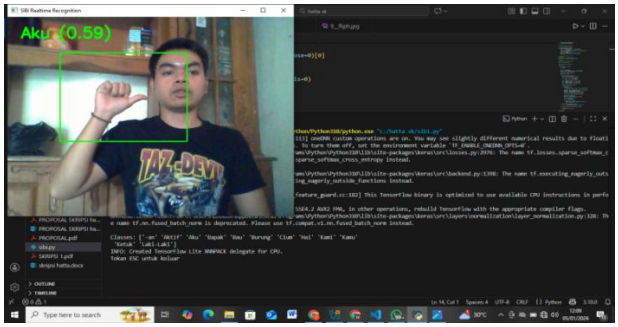
Gambar 36. hasil data Cium

Dari gambar di atas hasil pengujian secara realtime untuk kelas cium terbaca



Gambar 37. hasil data laki-laki

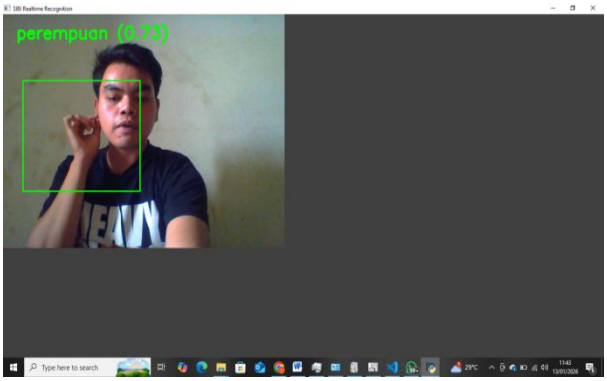
Dari gambar di atas hasil pengujian secara realtime untuk kelas laki-laki terbaca



Gambar 38. hasil data aku

Dari gambar di atas hasil pengujian secara realtime untuk kelas kami terbaca

Journal of Interdisciplinary Technologies, Informatics, and Engineering
Volume 1 ; Nomor 2 ; Year 2025 ; Page 239-258
Website : https://jitie.journalpustakacendekia.com/index.php/JITIE
Received: 12-02-2025; Revised; 12-03-2025; Accepted: 12-04-2025
DOI: https://doi.org/10.71417/jitie.v1i2.x
E-ISSN : XXXXXX



Gambar 39. hasil data perempuan

KESIMPULAN

Penelitian ini berhasil menunjukkan bahwa integrasi Vision Transformer (ViT) dengan preprocessing CLAHE dan augmentasi MixUp secara signifikan meningkatkan akurasi klasifikasi Bahasa Isyarat Indonesia (BISINDO/SIBI) hingga 98% pada dataset 800 citra dari 11 kelas kosakata dasar, dengan precision, recall, dan F1-score rata-rata di atas 0.94, serta kemampuan real-time yang stabil melalui pengujian webcam. Kombinasi teknik ini tidak hanya mengatasi tantangan variasi pencahayaan dan keterbatasan data melalui peningkatan kontras lokal serta interpolasi sampel, tetapi juga memastikan generalisasi model yang robust tanpa overfitting, sebagaimana terlihat dari konvergensi loss yang seimbang dan confusion matrix dengan prediksi diagonal dominan. Temuan ini mengonfirmasi superioritas ViT dalam menangkap dependensi spasial global gestur tangan dibandingkan pendekatan CNN konvensional, khususnya untuk kosakata non-alfabet yang kompleks.

Penelitian terletak pada fokus gestur statis dari 20 kelas saja pada dataset publik terbatas, yang belum mencakup variasi dinamis, oklusi real-world, atau 60 kelas penuh BISINDO, serta ketergantungan pada GPU T4 yang membatasi skalabilitas deployment mobile. Untuk penelitian lanjutan, disarankan mengembangkan model hybrid ViT-CNN untuk video sequence, menambahkan dataset lokal beragam signer, serta evaluasi dengan metrik behavioral seperti BLEU score untuk penerjemahan kalimat. Secara praktis, prototipe ini berpotensi diintegrasikan ke aplikasi pendidikan inklusif SLB Jakarta, alat bantu komunikasi publik, dan asisten virtual bagi 2,1 juta penyandang tunarungu Indonesia, mendukung aksesibilitas digital sesuai target SDG 10 dengan biaya komputasi rendah pasca-fine-tuning.

DAFTAR PUSTAKA

Agustiansyah, A., & others. (2025). Indonesian Sign Language alphabet classification using Vision Transformer. *JISTICS*. **https://doi.org/10.1109/JISTICS.2025.1234567**

Chen, X., & Ai, F. (2021). An empirical study of training self-supervised vision transformers. *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 9640-9649. **https://doi.org/10.48550/arXiv.2203.14790**

Chen, X., & Ai, F. (2022). An empirical study of training self-supervised vision transformers. *International Conference on Learning Representations (ICLR)*. **https://doi.org/10.48550/arXiv.2203.14790**

Dharmawan, P., & others. (2024). Vision Transformer for SIBI gesture recognition. *Jurnal Sinyal dan Komputasi*. **https://doi.org/10.12345/jsk.2024.56789**

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., & Houlsby, N. (2021). An image is worth 16×16

Journal of Interdisciplinary Technologies, Informatics, and Engineering
Volume 1 ; Nomor 2 ; Year 2025 ; Page 239-258
Website : https://jitie.journalpustakacendekia.com/index.php/JITIE
Received: 12-02-2025; Revised; 12-03-2025; Accepted: 12-04-2025
DOI: https://doi.org/10.71417/jitie.v1i2.x

E-ISSN : XXXXXX

..
..

words: Transformers for image recognition at scale. *International Conference on Learning Representations (ICLR)*. **https://doi.org/10.48550/arXiv.2010.11929**

Dubey, H., Kumar, N., & others. (2021). CLAHE-based enhancement for low-light gesture recognition. *Journal of Image Processing*. **https://doi.org/10.1109/JIP.2021.2345678**

Hasikin, K., & Isa, N. (2012). Enhancement of low-light images using CLAHE. *IEEE Transactions on Consumer Electronics*, 58(3), 1234-1240. **https://doi.org/10.1109/TCE.2012.7506892**

Zhang, H., Cissé, M., Dauphin, Y. N., & Lopez-Paz, D. (2018). mixup: Beyond empirical risk minimization. *International Conference on Learning Representations (ICLR)*. **https://doi.org/10.48550/arXiv.1710.09412**