VOL. 3. NO. 02 MARET 2026

JOURNAL OF COMPUTER SCIENCE AND ARTIFICIAL INTELLIGENCE (JCSAI)

IMPLEMENTASI SISTEM PENGENALAN BISINDO SECARA REAL-TIME BERBASIS MEDIAPIPE DENGAN NORMALISASI GEOMETRIK UNTUK REDUKSI OVERFITTING

Ramshal Hussein[1], Martanto[2], Yudhistira Arie Wijaya[3], Ahmad Rifa'i[4].

Program Studi Teknik Informatika[14]
Program Studi Manajemen Informatika[2]
Program Studi Sistem Informasi[3]

STMIK IKMI Cirebon
https://ikmi.ac.id/page/18/?lang=de
ramshalhussein15@gmail.com

(*) Corresponding Author : ramshalhussein15@gmail.com
Published : 30 Maret 2026

Abstract—*This study develops a computer-vision–based recognition system for static handshape letters of Indonesian Sign language (BISINDO) to support inclusive communication between signers and non-signers. The system utilizes MediaPipe Hands to extract hand landmarks, which are then enhanced through geometric normalization and relational feature construction to improve cross-user shape consistency. The dataset consists of 24 letters (A–Y excluding J and Z) with an 80:20 stratified split for training and testing. The normalized features are fed into a lightweight Multilayer Perceptron (MLP) equipped with Batch Normalization, Dropout, and label smoothing to reduce overfitting. Experimental results show that the proposed approach increases accuracy from 0.9270 to 0.9928 after applying geometric normalization and relational features, achieving a macro-precision of 0.9937, macro-recall of 0.9927, and macro-F1 score of 0.9930. The confusion matrix indicates strong class separation with a dominant diagonal, while real-time performance is maintained at 12–15 FPS without additional GPU or TensorRT optimization. Overall, the integration of geometric normalization and relational features within the MediaPipe pipeline, combined with an efficient MLP architecture, effectively enhances handshape consistency, improves class separability, and enables highly accurate real-time BISINDO recognition. This approach contributes to the development of lightweight and stable sign recognition systems ready for deployment in communication-assistive applications.*

Keywords: *BISINDO, MediaPipe Hands, geometric normalization, relational features, MLP, real-time sign recognition.*

Abstrak—Penelitian ini mengembangkan sistem pengenalan huruf statis Bahasa Isyarat Indonesia (BISINDO) berbasis visi komputer untuk mendukung komunikasi inklusif antara pengguna bahasa isyarat dan masyarakat umum. Sistem dibangun dengan memanfaatkan MediaPipe Hands sebagai ekstraktor titik landmark tangan, yang kemudian diproses melalui normalisasi geometrik serta pembentukan fitur relasional untuk meningkatkan konsistensi bentuk tangan antar pengguna. Dataset mencakup 24 huruf (A–Y tanpa J dan Z) dengan pembagian berstratifikasi 80:20 untuk data latih dan uji. Fitur yang telah dinormalisasi menjadi masukan bagi Multilayer Perceptron (MLP) ringan yang dilengkapi Batch Normalization, Dropout, dan label smoothing untuk mencegah overfitting. Hasil pengujian menunjukkan peningkatan akurasi dari 0,9270 menjadi 0,9928 setelah penerapan normalisasi geometrik dan fitur relasional, dengan *precision*-makro 0,9937, *recall*-makro 0,9927, dan F1-makro 0,9930. Confusion matrix memperlihatkan pemisahan kelas yang dominan pada diagonal utama, sedangkan pengujian kecepatan menampilkan performa real-time pada 12-15 FPS tanpa optimasi tambahan GPU atau TensorRT. Secara keseluruhan, integrasi normalisasi geometrik dan fitur relasional dalam pipeline MediaPipe serta arsitektur MLP yang efisien terbukti mampu meningkatkan konsistensi bentuk tangan, memperbaiki separasi kelas, dan menghasilkan pengenalan isyarat BISINDO yang akurat serta dapat dijalankan secara real-time. Pendekatan ini memberikan kontribusi pada pengembangan sistem pengenalan isyarat yang ringan, stabil antar pengguna, dan siap diintegrasikan ke aplikasi asisten komunikasi.

Kata Kunci: BISINDO, MediaPipe Hands, normalisasi geometrik, fitur relasional, MLP, pengenalan isyarat real-time.

INTRODUCTION

Bahasa Isyarat Indonesia (BISINDO) merupakan bahasa visual-gestural utama komunitas Tuli di Indonesia dan berperan penting dalam pendidikan, layanan publik, serta interaksi sosial. Namun, pengembangan teknologi pengenalan BISINDO masih tertinggal dibanding bahasa isyarat yang lebih banyak diteliti seperti ASL dan BSL. Keterbatasan *dataset* berskala besar, keragaman penutur yang terbatas, serta ketiadaan standar anotasi nasional menyebabkan model sulit digeneralisasikan dan rentan mengalami *overfitting*, terutama ketika data tidak seimbang dan kurang representatif [1], [2], [3]. Kondisi ini menghambat lahirnya sistem pengenalan BISINDO yang akurat dan dapat dioperasikan secara *real-time* pada perangkat komputasi umum.

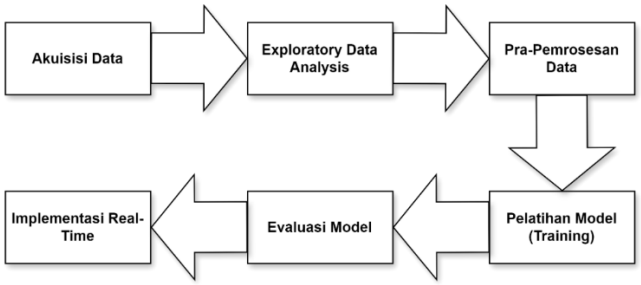
Tantangan teknis dalam pengenalan bahasa isyarat *real-time* meliputi variabilitas penutur, oklusi tangan, latar belakang kompleks, serta perubahan sudut pandang kamera yang berdampak langsung pada kestabilan prediksi [2], [4]. Sebagian besar penelitian BISINDO yang ada masih berbasis citra RGB dan model CNN berkapasitas besar yang sensitif terhadap pencahayaan dan latar belakang, serta jarang memanfaatkan representasi *landmark* yang dinormalisasi. Di sisi lain, studi pada *gesture recognition* menunjukkan bahwa normalisasi geometrik dan fitur relasional antartangan dapat mengurangi variansi yang tidak relevan dan menurunkan risiko *overfitting*, terutama pada *dataset* kecil atau penutur tunggal [5], [6], [7]. *Gap* utama yang muncul adalah minimnya *pipeline* BISINDO berbasis *landmark MediaPipe* yang secara eksplisit mengintegrasikan normalisasi geometrik dan fitur relasional untuk mendukung pengenalan huruf statis secara *real-time*.

Berdasarkan *gap* tersebut, penelitian ini mengusulkan sistem pengenalan huruf statis BISINDO berbasis *MediaPipe* Hands yang mengekstraksi 21 titik kunci tangan dan kemudian menerapkan normalisasi geometrik agar skala, orientasi, dan posisi tangan menjadi selaras. Fitur relasional dua tangan, seperti selisih koordinat pusat tangan dan rasio skala, dikonstruksi untuk memperkuat representasi spasial huruf yang memiliki bentuk serupa. Vektor fitur hasil normalisasi dan rekayasa relasional diklasifikasikan menggunakan model *Multilayer Perceptron* (*MLP*) ringan yang diperkaya teknik regularisasi seperti *Batch Normalization*, *Dropout*, dan *label smoothing* guna mereduksi *overfitting*. Model *Multilayer Perceptron* (*MLP*) dipilih karena arsitekturnya ringan, jumlah parameternya terukur, dan dapat memberikan *trade-off* yang baik antara akurasi dan latensi pada perangkat berbasis *CPU* tanpa memerlukan *GPU* khusus. Sistem selanjutnya diuji secara *real-time* dengan *temporal smoothing* dan *confidence threshold* untuk memastikan prediksi yang stabil. Kebaruan penelitian terletak pada integrasi normalisasi geometrik dan fitur relasional dua tangan dalam *pipeline* BISINDO berbasis *landmark MediaPipe* yang ringan serta dioptimalkan untuk pengenalan huruf statis secara *real-time* pada perangkat berbasis *CPU[8], [9], [10]*.

Pada eksperimen awal menggunakan *landmark* mentah tanpa normalisasi geometrik, model *MLP* menunjukkan gejala *overfitting*, ditandai dengan selisih akurasi train dan validasi yang cukup besar serta klaster fitur yang saling tumpang tindih pada proyeksi *PCA*. Kondisi ini menguatkan kebutuhan akan strategi normalisasi dan rekayasa fitur yang lebih sistematis.

MATERIALS AND METHODS

Penelitian ini dilakukan melalui enam tahapan utama yang dirancang untuk membangun dan mengevaluasi sistem pengenalan BISINDO berbasis *MediaPipe* secara komprehensif. Setiap tahapan dilakukan secara sistematis dengan memastikan keterkaitan logis antara proses, data, dan hasil evaluasi. Tahapan ini mencakup alur kerja dari pengumpulan data hingga implementasi *real-time.*



Gambar 1 Diagram Alur Tahapan

Alur Tahapan Penelitian:

Akuisisi Data: Proses pengumpulan data BISINDO dua tangan dengan variasi posisi tangan dan pencahayaan. Hasil berupa koordinat *landmark* 3D dalam format *.npy* yang diekstraksi langsung melalui *MediaPipe* menggunakan *data_capture.py*.

Eksplorasi Data dan Analisis: Dilakukan untuk menganalisis distribusi kelas, keseimbangan data, potensi *outlier*, serta visualisasi awal seperti *PCA* dan *confusion matrix baseline* untuk memahami karakteristik data mentah.

Pra-Pemrosesan Data: Normalisasi geometrik diterapkan untuk menghilangkan pengaruh rotasi, translasi, dan skala, serta menambahkan fitur

relasional dua tangan untuk memperkuat representasi spasial.

Pelatihan Model: Model *MLP* dilatih menggunakan data yang telah dinormalisasi dengan penyesuaian *hyperparameter* dan augmentasi data untuk meningkatkan generalisasi.

Evaluasi Model: Performa model diukur menggunakan metrik akurasi, presisi, *recall* dan *F1-score*. Hasil dibandingkan antara model *baseline* dan model akhir.

Implementasi *Real-time*: Model terbaik diintegrasikan dengan kamera dan diuji dalam kondisi nyata untuk memastikan performa stabil dan efisien.

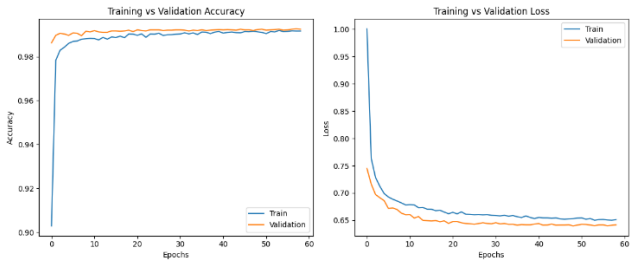
Dataset yang digunakan dalam penelitian ini diperoleh dari proses akuisisi data primer dengan menggunakan satu subjek (peneliti sendiri) yang mengeksekusi 24 huruf statis BISINDO dua tangan (A-Y, tanpa J dan Z). Total diperoleh 6.429 sampel mentah hasil ekstraksi *landmark MediaPipe Hands*, dengan jumlah sampel per kelas berada pada rentang 263-274 sampel. Distribusi rinci per huruf disajikan pada Tabel 4.3 Jumlah *Dataset*.Setelah dilakukan proses normalisasi geometrik, penambahan fitur relasional, dan augmentasi (rotasi acak, skala acak, *jitter noise*), jumlah sampel meningkat menjadi 32.145 sampel pada berkas *processed_dataset.npz*. *Dataset* ini kemudian dibagi menjadi 80% data latih dan 20% data validasi menggunakan *stratified split* per *frame* berdasarkan label huruf, sehingga proporsi setiap kelas pada set latih dan validasi tetap seimbang. Setiap sampel pada berkas *.npy* diberi label huruf BISINDO secara eksplisit berdasarkan huruf yang sedang dieksekusi ketika *frame* tersebut diambil. Proses anotasi dilakukan secara *online* di dalam skrip *data_capture.py*, yaitu dengan memilih huruf target terlebih dahulu, kemudian merekam beberapa *frame* berturut-turut dan menyimpan koordinat *landmark* beserta label huruf ke dalam direktori *data_npy/.* Dengan demikian, setiap baris data pada *dataset* memiliki pasangan (fitur, label) yang konsisten dengan huruf yang ditampilkan pada saat perekaman.

RESULTS AND DISCUSSION

Tahap pelatihan model dilakukan menggunakan skrip *train_model.py* pada *dataset* berdimensi 133 yang telah dinormalisasi dan di-augmentasi, dengan total sekitar 32.145 sampel. Arsitektur dan hyperparameter model mengikuti rancangan pada bagian 3.2.4, yaitu *MLP* tiga lapis tersembunyi (256-256-128 *neuron*) dengan *Batch Normalization*, *Dropout, label smoothing*, *optimizer Adam*, 100 *epoch* maksimum, *batch size* 32, serta *early stopping (patience = 10).* Fokus pada bagian ini adalah hasil pelatihan. Gambar 4.5 dan 4.7 memperlihatkan perbandingan kurva pelatihan antara model *baseline* dan model final. Pada model *baseline*, *gap* antara akurasi pelatihan dan validasi cukup besar, mengindikasikan *overfitting*. Sebaliknya, pada model final, kurva pelatihan dan validasi berhimpitan sangat stabil di atas 0,99, dengan konvergensi tercapai dalam sekitar 60 *epoch* sehingga menunjukkan kondisi *Goodfit* serta generalisasi yang lebih baik.

Perbandingan Kurva Pelatihan Efektivitas tahap pra-pemrosesan dan arsitektur model final ini dibuktikan dengan membandingkan kurva pelatihan *baseline* (model dari tahap EDA) dengan kurva pelatihan final (model hasil eksperimen).

Kurva pelatihan pada model *baseline* menunjukkan generalization *gap* yang jelas. Akurasi pelatihan meningkat mendekati 1,00, sedangkan akurasi validasi stabil di kisaran 0,92 sampai 0,94. Sejalan dengan itu, *training loss* menurun mendekati 0, sementara *validation loss* masih lebih tinggi dari *training loss* di akhir pelatihan. Pola ini merupakan indikasi *overfitting* pada model *baseline*. memperlihatkan hasil yang jauh lebih baik. Kurva *Training* dan *Validation* bergerak rapat serta berhimpitan sangat stabil di atas 0,99, dengan konvergensi tercapai dalam sekitar 60 epoch, menandakan bahwa model final berhasil mencapai keseimbangan antara akurasi pelatihan dan validasi. Pola ini menunjukkan bahwa model final berada pada kondisi *goodfit* dan tidak mengalami *overfitting*.



Gambar 2 *Training Curve Goodfit*

Model *MLP* final berhasil memanfaatkan fitur 133-dimensi hasil normalisasi dan augmentasi dengan efisien. Penerapan *Batch Normalization*, *Dropout*, dan *label smoothing* meningkatkan stabilitas pelatihan dan mengurangi risiko *overfitting*. Hasil ini menunjukkan bahwa pra-pemrosesan dan arsitektur model yang diusulkan memberikan kinerja yang konsisten serta siap untuk tahap evaluasi lanjutan pada bagian berikutnya.

Tahap evaluasi model dilakukan untuk mengukur kinerja kuantitatif model *MLP* final pada *test set*. Evaluasi dilakukan menggunakan skrip *evaluate_model.py* dengan data pengujian sebesar 20% dari total *dataset* (6.429 sampel uji dari total 32.145 data) yang tidak pernah dilihat model selama pelatihan.

Tujuan evaluasi ini adalah untuk memvalidasi secara objektif apakah tahapan pra-pemrosesan (4.1.3) dan arsitektur model final (4.1.4) benar-benar berhasil mengatasi dua masalah utama yang ditemukan pada tahap EDA (4.1.2), yaitu separabilitas kelas yang buruk dan indikasi *overfitting*.

Metrik Evaluasi Metrik utama yang digunakan untuk menilai performa model adalah sebagai berikut:

Accuracy: Persentase prediksi yang benar dari seluruh sampel pengujian.

Precision (Macro): Rata-rata ketepatan model dalam memprediksi setiap kelas tanpa bias terhadap kelas dominan.

Recall (Macro): Kemampuan model dalam mendeteksi seluruh sampel positif pada setiap kelas.

F1-score (Macro): Rata-rata harmonis antara *Precision* dan *Recall*, digunakan sebagai metrik utama untuk menilai keseimbangan performa antar kelas pada kasus klasifikasi multi-kelas BISINDO.

Metrik-metrik tersebut diekstraksi langsung dari file laporan evaluasi untuk model final, serta dibandingkan dengan laporan yang digunakan pada tahap EDA.

Hasil Evaluasi Kuantitatif

Tabel 1 Hasil Perbandingan Evaluasi Model *Baseline* dan Model *Final*

Metrik	Model *Baseline*	Model Eksperimen
Accuracy	0,9270	0,9928
***Precision* (*macro*)**	0,9383	0,9937
***Recall* (*macro*)**	0,9271	0,9927
***F1-score* (*macro*)**	0,9245	0,9930

Seluruh metrik pada Tabel 1 diperoleh dari satu kali proses pelatihan dengan *seed* acak terkontrol; oleh karena itu, ukuran ketidakpastian antar-run (misalnya standar deviasi atau interval kepercayaan) belum disajikan dan menjadi salah satu keterbatasan penelitian ini

Tabel 2 Ringkasan *Precision*, *Recall*, dan *F1-score* per Kelas Huruf BISINDO

Huruf	*Precision*	*recall*	*F1-score*	*support*
A	1,0000	0,9886	0,9943	1320
B	1,0000	0,9970	0,9985	1355
C	0,9970	0,9955	0,9962	1330
D	1,0000	0,9926	0,9963	1350
E	0,9970	0,9925	0,9947	1330
F	1,0000	0,9888	0,9944	1340
G	1,0000	0,9924	0,9962	1315
H	1,0000	0,9910	0,9955	1340
I	1,0000	0,9924	0,9962	1315
K	0,9985	0,9924	0,9954	1315
L	0,8726	1,0000	0,9320	1370
M	0,9993	1,0000	0,9996	1340
N	1,0000	0,9993	0,9996	1350
O	0,9955	0,9925	0,9940	1330
P	1,0000	0,9926	0,9963	1350
Q	1,0000	0,9881	0,9940	1350
R	0,9910	0,9851	0,9880	1340
S	1,0000	0,9925	0,9962	1330
T	1,0000	0,9888	0,9944	1345
U	1,0000	0,9926	0,9963	1355
V	1,0000	0,9889	0,9944	1355
W	0,9970	0,9962	0,9966	1330
X	1,0000	0,9926	0,9963	1350
Y	1,0000	0,9933	0,9966	1340

Detail metrik per kelas huruf disajikan pada Tabel 2 untuk mengidentifikasi kelas yang masih relatif sulit diklasifikasikan. Peningkatan akurasi dari 0,9270 menjadi 0,9928 serta *F1-macro* dari 0,9245 menjadi 0,9930 menunjukkan bahwa *pipeline* yang diusulkan tidak hanya lebih tepat dalam mengenali huruf-huruf BISINDO, tetapi juga lebih seimbang performanya antar kelas. Bagi skenario *real-time*, peningkatan ini berarti sistem mampu memberikan prediksi huruf yang lebih konsisten dengan tingkat kesalahan yang rendah meskipun inferensi dijalankan sepenuhnya di *CPU*.

CONCLUSION

Penelitian ini berkontribusi pada pengembangan sistem pengenalan huruf statis Bahasa Isyarat Indonesia (BISINDO) berbasis *landmark MediaPipe* dengan integrasi normalisasi geometrik dan fitur relasional dua tangan dalam arsitektur *MLP* ringan yang dapat berjalan secara *real-time* di perangkat berbasis *CPU*. *Pipeline* yang diusulkan berhasil meningkatkan separabilitas fitur, mengurangi indikasi *overfitting*, serta memberikan peningkatan akurasi dan *F1-macro* yang signifikan dibandingkan model *baseline*.

Penelitian ini telah dilakukan untuk merancang, mengembangkan, dan mengevaluasi sistem pengenalan huruf statis Bahasa Isyarat Indonesia (BISINDO) berbasis *landmark MediaPipe* dengan pendekatan normalisasi geometrik dan penambahan fitur relasional antartangan. Pada kondisi eksperimen yang melibatkan satu penutur dan pengujian internal ini, sistem belum diuji secara luas, dan pengukuran latensi *end-to-end* secara numerik baru dilakukan pada satu perangkat komputer dengan satu skenario uji. Oleh karena itu, generalisasi kinerja *real-time* ke beragam pengguna dan konfigurasi perangkat lain masih perlu divalidasi lebih lanjut. Berdasarkan hasil analisis dan eksperimen pada skenario tersebut, kesimpulan utama penelitian ini adalah sebagai berikut:

Pipeline pengenalan huruf BISINDO yang dirancang, mencakup ekstraksi *landmark MediaPipe* dalam format *.npy*, normalisasi fitur, klasifikasi menggunakan *MLP*, penerapan *confidence threshold* sebesar 0,7, serta *temporal smoothing* menggunakan *deque(maxlen=7)*, dapat berjalan secara terpadu pada perangkat berbasis *CPU* dan menghasilkan prediksi yang stabil dengan kecepatan sekitar 12-15 *FPS*. Pada kondisi eksperimen ini, hal tersebut sudah cukup untuk memenuhi kebutuhan pengenalan huruf statis BISINDO secara *real-time* pada perangkat yang diuji.

Penerapan normalisasi geometrik yang dipadukan dengan lima fitur relasional antartangan (Δx, Δy, Δz, jarak antar pusat tangan, dan rasio skala) meningkatkan kualitas representasi fitur, yang terlihat dari klaster antar kelas yang lebih terpisah pada hasil *PCA* serta kurva pelatihan yang lebih seimbang antara data pelatihan dan validasi. Kondisi ini menunjukkan bahwa kecenderungan *overfitting* pada fitur mentah dapat dikurangi pada skenario eksperimen penelitian ini.

Model *Multilayer Perceptron* (*MLP*) yang dilatih menggunakan fitur hasil normalisasi 133 dimensi menunjukkan kinerja yang lebih baik dibandingkan model *baseline* berbasis fitur mentah, dengan peningkatan akurasi dari 0,9270 menjadi 0,9928 dan *F1-macro* dari 0,9245 menjadi 0,9930 pada data uji signer-dependent dalam penelitian ini. Hasil *classification report* dan *confusion matrix* memperlihatkan bahwa sebagian besar huruf dapat dikenali dengan baik dengan tingkat kesalahan yang rendah, sehingga model dinilai mampu melakukan generalisasi secara stabil pada kondisi eksperimen yang digunakan.

REFERENCE

[1] I. Papastratis, K. Dimitropoulos, And P. Daras, “Continuous Sign Language Recognition Through A Context-Aware Generative Adversarial Network,” *Sensors*, Vol. 21, No. 7, P. 2437, 2021, Doi: 10.3390/S21072437.

[2] K. Kozyra, K. Trzyniec, E. Popardowski, And M. Stachurska, “Application For Recognizing Sign Language Gestures Based On An Artificial Neural Network,” *Sensors*, Vol. 22, No. 24, P. 9864, 2022, Doi: 10.3390/S22249864.

[3] M. S. Amin, S. T. H. Rizvi, And M. M. Hossain, “A Comparative Review On Applications Of Different Sensors For Sign Language Recognition,” *J. Imaging*, Vol. 8, No. 4, P. 98, 2022, Doi: 10.3390/Jimaging8040098.

[4] N. Amangeldy, S. Kudubayeva, A. Kassymova, A. Karipzhanova, B. Razakhova, And S. Kuralov, “Continuous Sign Language Recognition And Its Translation Into Intonation-Colored Speech,” *Sensors*, Vol. 23, No. 14, P. 6383, 2023, Doi: 10.3390/S23146383.

[5] A. Ben Haj Amor, S. Jlassi, And A. Kachouri, “Sign Language Recognition Using The Electromyographic Signal: A Comprehensive Study,” *Sensors*, Vol. 23, No. 19, P. 8343, 2023, Doi: 10.3390/S23198343.

[6] Y. Li And P. Zhang, “Static Hand Gesture Recognition Based On Hierarchical Decision And Classification Of Finger Features,” *Sci. Prog.*, Vol. 105, No. 1, P. 00368504221086362, 2022, Doi: 10.1177/00368504221086362.

[7] Y. Meng, S. Chen, Z. Li, And K. Yan, “Real-Time Hand Gesture Monitoring Model Based On Mediapipe’s Registerable System,” *Sensors*, Vol. 24, No. 19, P. 6262, 2024, Doi: 10.3390/S24196262.

[8] N. Wu, D. Jia, C. Zhang, And Z. Li, “Cervical Cell Classification Based On Strong Feature Cnn-Lsvm Network Using Adaboost

Optimization," Vol. 44, No. 3, Pp. 4335–4355, 2023, Doi: 10.3233/Jifs-221604.

[9] D. J. Chaudhari And K. Malathi, "Detection And Prediction Of Rice Leaf Disease Using A Hybrid Cnn-Svm Model," *Optical Memory And Neural Networks (Information Optics)*, Vol. 32, No. 1, Pp. 39–57, 2023, Doi: 10.3103/S1060992x2301006x.

[10] B. Karatay, D. Beştepe, K. Sailunaz, T. Özyer, And R. Alhajj, "Cnn-Transformer Based Emotion Classification From Facial Expressions And Body Gestures," *Multimed. Tools Appl.*, Vol. 83, No. 8, Pp. 23129–23171, 2024, Doi: 10.1007/S11042-023-16342-5.