

Penerapan *Graph Neural Network* dalam Pengenalan Alfabet BISINDO dengan Fokus pada Gerakan Dinamis

Febri Damatraseta¹, Muhammad Alfian², Yuliandi³

^{1,2}Program Studi Teknologi Informatika, Fakultas Informatika dan Pariwisata, Jawa Barat, Indonesia

³Program Studi Manajemen, Fakultas Bisnis, Institut Bisnis dan Informatika Kesatuan, Bogor, Indonesia

Email: ¹febrid@ibik.ac.id, ²212310017@student.ibik.ac.id, ³yuliandi@ibik.ac.id

Abstract. *A Graph Neural Network Approach to BISINDO Alphabet Recognition with Emphasis on Dynamic Movements.* Most studies on Indonesian Sign Language (BISINDO) recognition focus solely on static hand gestures, despite several letters, such as R and J, requiring dynamic movement patterns that static representations cannot capture. This study employs MediaPipe to extract 21 hand keypoints as input features. These keypoints are modeled as graph structures and processed using Graph Neural Networks (GNNs) to recognize the alphabet and dynamic gestures simultaneously. The model was trained and evaluated using K-fold cross-validation to ensure consistency. The GNN model achieved 96% recognition accuracy on the test dataset. A prototype web-based system successfully recognized all 26 BISINDO letters dynamically, achieving a prediction accuracy of up to 91%, demonstrating the model's effectiveness and potential to enhance accessible communication through sign language technology.

Keywords: BISINDO, hand gesture recognition, Graph Neural Networks, MediaPipe, Dynamic Sign Language Recognition

Abstrak. Sebagian besar studi pengenalan alfabet Bahasa Isyarat Indonesia (BISINDO) masih terbatas pada gesture statik, meskipun beberapa huruf seperti R dan J memiliki karakteristik gerakan dinamis yang tidak dapat direpresentasikan secara statis. Penelitian ini menggunakan MediaPipe untuk mendeteksi 21 keypoints tangan sebagai input fitur. Titik-titik ini dimodelkan dalam bentuk graf dan diproses menggunakan Graph Neural Networks (GNNs) guna mengenali alfabet secara simultan, termasuk huruf-huruf dinamis. Proses pelatihan menggunakan K-Fold Cross Validation untuk menguji konsistensi performa model. Model GNN menghasilkan akurasi sebesar 96% pada pengujian data alfabet BISINDO. Prototipe sistem dalam bentuk aplikasi web berhasil mengenali 26 huruf BISINDO secara dinamis dengan tingkat akurasi prediksi mencapai 91%, menunjukkan potensi implementasi nyata dari pendekatan GNN dalam mendukung aksesibilitas komunikasi inklusif.

Kata Kunci: BISINDO, pengenalan gerakan tangan, Graph Neural Networks, MediaPipe, pengenalan bahasa isyarat dinamis

1. Pendahuluan

Bahasa Isyarat digunakan sebagai alat komunikasi utama bagi penyandang tunarungu dan tunawicara yang memiliki keterbatasan dalam kemampuan mendengar dan berbicara. Untuk berkomunikasi dengan masyarakat umum, penyandang tunarungu memerlukan penerjemah yang fasih dalam bahasa isyarat [1]. Bahasa isyarat menjadi media penting bagi teman tuli dalam berkomunikasi dengan orang lain. Bahasa ini melibatkan gerakan tubuh dan bibir untuk menyampaikan pesan, tanpa menggunakan bunyi ucapan atau tulisan [2]. Bahasa Isyarat Indonesia (BISINDO) merupakan salah satu alat komunikasi utama bagi disabilitas tunarungu. Dengan jumlah penyandang tunarungu di Indonesia sebanyak 6.952.797 jiwa dari total populasi 207.839.035 jiwa, terdapat ratusan ribu kata dalam bahasa Indonesia yang dapat diwakili oleh bahasa isyarat [3]. Keberadaan dan perkembangan BISINDO tidak lepas dari peran GERKATIN (Gerakan untuk Kesejahteraan Tuna Rungu Indonesia), satu-satunya organisasi penyandang tunarungu di Indonesia yang dikelola sepenuhnya oleh orang tunarungu [4]. Namun, keterbatasan pemahaman masyarakat umum terhadap bahasa isyarat Indonesia sering kali menjadi kendala dalam interaksi sehari-hari.

Dengan berkembangnya teknologi, kecerdasan buatan memberikan dampak signifikan di berbagai bidang, termasuk penerjemahan efisiensi komunikasi. Beragamnya bentuk dan gerakan tangan untuk membuat sebuah bentuk huruf alfabet, terkadang menjadi kendala dalam mendeteksi huruf tersebut. Seperti huruf “R” dan “J” dalam BISINDO yang memiliki pola gerakan tangan yang dinamis dan kompleks. Tantangan ini menunjukkan perlunya teknologi yang lebih optimal untuk menangani variasi gerakan lebih baik sehingga sistem penerjemah bahasa isyarat dapat bekerja secara lebih akurat. Hal ini menunjukkan pentingnya penelitian lebih lanjut untuk menemukan metode yang mampu mengatasi kendala tersebut dengan lebih efisien dan dapat meningkatkan kualitas hidup penyandang tunarungu.

Pada penelitian sebelumnya yang telah dilakukan mengenai penerjemahan bahasa isyarat terdapat perbedaan pendekatan yang digunakan dalam mendeteksi huruf-huruf alfabet. Salah satu penelitian yang ada menunjukkan hasil bahwa huruf seperti B, D, J, K, dan P sulit untuk dikenali atau prediksi, walau saat proses testing dengan *still-image* mendapatkan hasil benar 76% dan hasil prediksi ketika dilakukan uji secara real-time didapatkan sebesar 60% [5]. Lalu pada penelitian lainnya berhasil mendapatkan tingkat akurasi keseluruhan sistem mencapai 92,08% meskipun beberapa huruf seperti B, D, P, M, N, F, K, dan Z memiliki akurasi dibawah 90% karena bentuk tangan yang hampir mirip atau sulit dibentuk. Namun huruf 'R' dan 'J' tidak dapat diterjemahkan karena algoritma yang digunakan tidak dapat mendeteksi gerakan tangan [1]. Adapun penelitian lain mendapatkan hasil bahwa sistem mampu mencapai akurasi sebesar 90,10% untuk mendeteksi objek tangan, dan 97,33% untuk klasifikasi alfabet BISINDO setelah 50 epoch pelatihan [6]. Beberapa huruf J, L, dan R memiliki tingkat kesalahan yang lebih tinggi karena perbedaan jarak atau posisi tangan dalam gambar. Dan juga penelitian mengenali alfabet Bahasa Isyarat Amerika (ASL) dengan pendekatan *Depth-based Geometrical Sign Language Recognition* (DGSLR). ANN digunakan untuk meningkatkan akurasi pengenalan, mencapai tingkat akurasi hingga 96,78[7].

Solusi dari masalah yang ada, terdapat penelitian yang melakukan analisis terhadap penerapan *Graph Neural Networks* (GNNs) di bidang Computer Vision. GNNs digunakan untuk berbagai tugas seperti deteksi perilaku, pengenalan wajah, pose tubuh, deteksi objek, dan analisis video [8]. Penelitian tersebut mendapatkan hasil bahwa secara efektif dapat menangkap hubungan antar node dalam bentuk data berbentuk graf, yang mana ini memungkinkan sistem dapat mendeteksi gerakan sebuah objek. Lalu pada penelitian lain yang berfokus pada penggunaan GNNs untuk menganalisis gambar dan video [9]. Dari penelitian tersebut dapat disimpulkan bahwa GNN mampu menangani data gambar yang memiliki hubungan kompleks antar objek. Dan hasilnya menunjukkan bahwa model berbasis GNN yang diusulkan mencapai akurasi 96,63% pada dataset ImageNet. GNN tidak hanya memberikan akurasi yang tinggi tetapi juga nilai kesalahan yang lebih rendah, yaitu 0,32 dibandingkan dengan metode seperti CNN, ANN, dan pendekatan heuristik lainnya.

Berdasarkan latar belakang tersebut, pada penelitian ini bertujuan untuk menganalisis dan membandingkan efektivitas metode deteksi statis dan dinamis dalam mengenali gerakan tangan, serta sebuah sistem yang mampu menerjemahkan BISINDO secara akurat terutama pada huruf 'R' dan 'J' dengan menggunakan algoritma GNNs. Sistem yang dibangun ini diharapkan dapat memberikan pengetahuan mengenai efektivitas metode statis dan dinamis, yang dapat membantu dalam memilih metode yang lebih tepat untuk penerjemahan bahasa isyarat secara otomatis.

2. Tinjauan Pustaka

2.1. Literatur Terkait

Penelitian terkait pengenalan BISINDO terus berkembang seiring kemajuan pengolahan citra dan kecerdasan buatan. Sebagian besar studi berfokus pada gerakan statis, sementara pengenalan gesture dinamis, khususnya huruf 'R' dan 'J', masih menjadi tantangan. Oleh karena itu, bagian ini mengulas penelitian terdahulu terkait teknik ekstraksi fitur tangan dan metode pembelajaran mesin yang digunakan. Penelitian [5] membangun sistem penerjemah alfabet BISINDO secara real-time dengan membandingkan arsitektur LeNet-5 dan AlexNet. Hasilnya, AlexNet mencapai akurasi validasi 76% (lebih tinggi dari LeNet-5 sebesar 19%), namun turun

menjadi 60% saat diuji dalam kondisi real-time. Hal ini menunjukkan bahwa meskipun AlexNet lebih unggul, pengembangan lanjutan masih diperlukan agar sistem mampu bekerja optimal, khususnya untuk gesture dinamis.

Penelitian yang dirancang untuk mengembangkan sistem deteksi isyarat tangan kosong dalam Bahasa Isyarat Indonesia (BISINDO) menggunakan metode *Adaptive Neuro-Fuzzy Inference System* (ANFIS) [4]. Data citra tangan diperoleh dari satu peraga melalui segmentasi warna untuk memisahkan objek tangan dari latar belakang, lalu diekstraksi fitur menggunakan metode *Histogram Oriented Gradient* (HOG). Proses klasifikasi menggunakan ANFIS menunjukkan hasil yang memuaskan dengan akurasi sebesar 78,31% pada dataset yang tersedia. Penelitian ini memperkuat potensi ANFIS untuk tugas pengenalan gestur BISINDO berbasis visi komputer dan merekomendasikan untuk penelitian lanjutan mencakup peningkatan jumlah dataset, penggunaan peraga lebih banyak, dan implementasi dalam sistem real-time agar performa model dapat ditingkatkan secara signifikan.

Penelitian lain mengembangkan sistem pengenalan alfabet BISINDO (A–Z) dengan metode Fourier Descriptor untuk mengekstraksi ciri bentuk tangan [10]. Hasil eksperimen menunjukkan akurasi rata-rata 57,43% pada citra berotasi dan 34,36% pada citra translasi, meskipun huruf 'F' mencapai akurasi hingga 99,49%. Temuan ini menegaskan bahwa Fourier Descriptor efektif pada citra statis dengan orientasi konsisten, tetapi kurang robust terhadap transformasi geometris, sehingga dibutuhkan pendekatan tambahan untuk aplikasi dunia nyata. Penelitian-penelitian tersebut umumnya masih berfokus pada gesture statis dan menghadapi tantangan dalam mengenali gesture dinamis seperti huruf 'R' dan 'J'. Oleh karena itu, dibutuhkan pendekatan yang mampu menangkap hubungan spasial antar titik tangan dan mengenali gesture secara simultan dan adaptif dalam berbagai kondisi nyata.

2.2. MediaPipe Hands

MediaPipe Hands adalah solusi *real-time hand tracking* berbasis *computer vision* dari Google yang bekerja melalui dua tahap: deteksi telapak tangan menggunakan *BlazePalm*, lalu prediksi 21 titik kunci (keypoint) 3D pada sendi-sendi jari. Untuk analisis pose dan klasifikasi *gesture*, digunakan metode matematis seperti perhitungan sudut antar sendi dengan rumus *dot product* pada Persamaan 1.

$$\theta = \cos^{-1} \left(\frac{(p_a - p_b) \cdot (p_c - p_b)}{\|p_a - p_b\| \cdot \|p_c - p_b\|} \right) \quad (1)$$

Persamaan (1) digunakan proses normalisasi citra tangan sebelum dilakukan ekstraksi titik landmark dilakukan melalui serangkaian transformasi geometris, salah satunya adalah rotasi. Sudut rotasi (θ) diperoleh dengan menganalisis orientasi tangan berdasarkan titik-titik kunci (*key landmarks*) yang ditandai dengan variable (p_a, p_b, p_c), seperti pergelangan tangan (*wrist*), sendi dasar jari telunjuk (*index finger MCP*), serta sendi dasar jari kelingking (*pinky MCP*). Untuk menghitung besarnya sudut tersebut, MediaPipe menggunakan formulasi matematis berbasis perkalian titik *dot product* (\cos^{-1}) antar vektor yang dibentuk dari pasangan titik landmark tersebut.

Dalam perhitungan jarak *Euclidean* pada *MediaPipe Hands*, setiap titik *landmark* tangan direpresentasikan dalam koordinat tiga dimensi (x, y, z). Misalnya, untuk menghitung jarak antara ujung jari telunjuk (*index fingertip, landmark 8*) dengan ujung ibu jari (*thumb tip, landmark 4*), digunakan Persamaan 2.

$$d_{ij} = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2 + (z_i - z_j)^2} \quad (2)$$

Setiap *landmark* memiliki nilai koordinat (x, y, z). Namun, nilai z sering kali relatif terhadap kamera dan bisa berbeda tergantung posisi tangan terhadap sensor. Oleh karena itu,

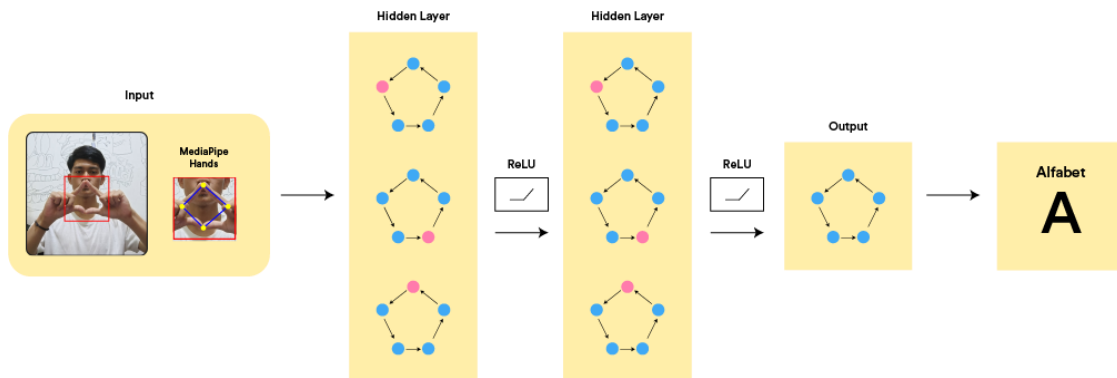
dilakukan proses *normalized* dengan menjadikan titik *wrist* sebagai referensi dengan Persamaan 3.

$$Z_{normalized} = Z_i - Z_{wrist} \tag{3}$$

Pendekatan ini memungkinkan sistem mengenali gesture secara efisien dan konsisten, bahkan dalam kondisi latar belakang dan pencahayaan yang kompleks. Beberapa jurnal [11][12] menunjukkan bahwa sistem ini sangat cocok untuk integrasi dalam pengenalan bahasa isyarat seperti BISINDO, karena mampu menangkap hubungan spasial antar titik tangan secara akurat dan responsif.

2.3. Graph Neural Networks

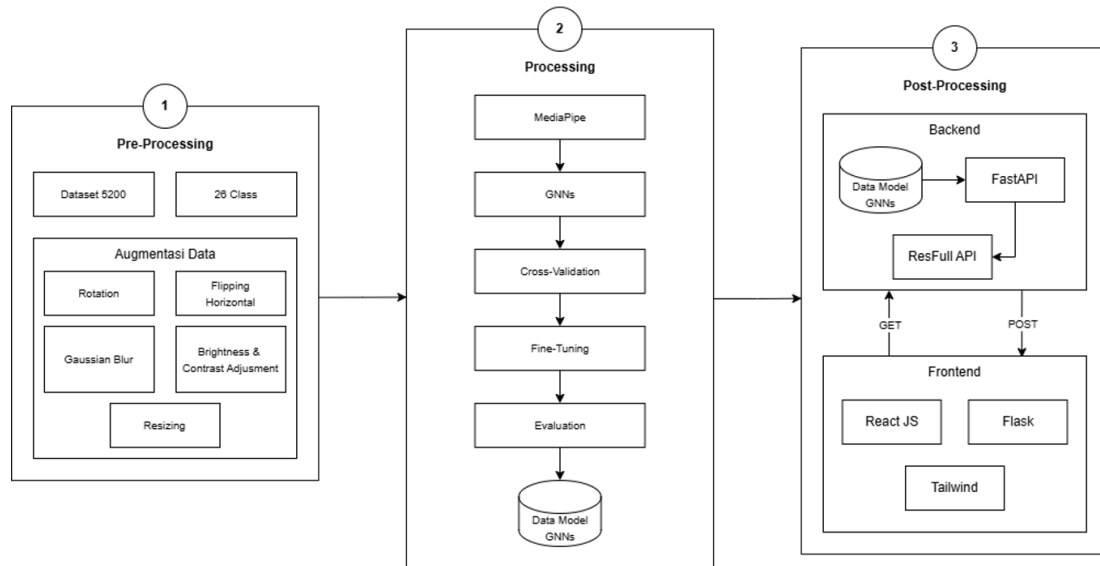
Graph Neural Network (GNN) merupakan arsitektur pembelajaran mesin yang secara khusus dirancang untuk bekerja pada data berbentuk graf, yang terdiri dari *node* dan *edge* untuk mewakili entitas dan hubungan antarentitas tersebut. Pada dasarnya, GNN memanfaatkan struktur data graf untuk menangkap hubungan atau pola antara node-node, yang umumnya tidak dapat diwakili dengan baik oleh model pembelajaran mesin tradisional. Model ini banyak diterapkan dalam berbagai bidang, seperti analisis jaringan sosial, sistem rekomendasi, dan terutama dalam bidang *computer vision* yaitu pengenalan pose tubuh dan klasifikasi objek pada gambar atau video [8].



Gambar 1. Ilustrasi penerapan arsitektur GNN's pada BISINDO

3. Metodologi Penelitian

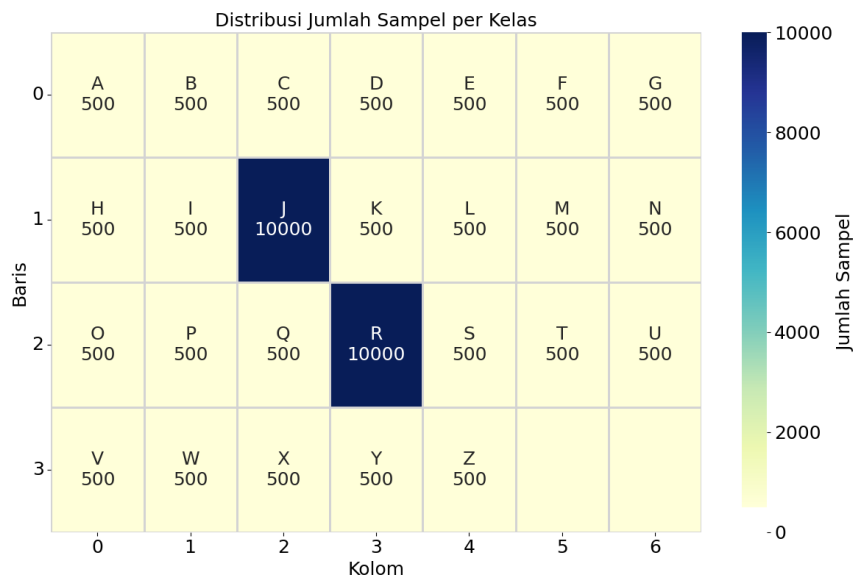
Bagian ini menjelaskan metode penelitian yang digunakan dengan mengadopsi pendekatan SDLC tipe *fountain* berdasarkan tahap *logical design* seperti pada Gambar 2, yang terdiri dari tiga fase utama: *pre-processing*, *processing*, dan *post-processing*. Tahap *pre-processing* berfokus pada persiapan dan pembersihan data agar siap digunakan dalam proses inti, termasuk normalisasi, segmentasi, dan penghilangan noise yang dapat memengaruhi hasil pengenalan. Selanjutnya, tahap *processing* merupakan fase utama yang melibatkan penerapan algoritma inti untuk pengolahan data, seperti ekstraksi fitur dan proses klasifikasi, sehingga informasi yang relevan dapat diperoleh secara optimal. Terakhir, tahap *post-processing* dilakukan untuk mengevaluasi, memvalidasi, dan menyajikan hasil pengolahan data agar dapat diinterpretasikan dengan baik, termasuk analisis performa sistem melalui metrik evaluasi yang telah ditentukan.



Gambar 2. Struktur Alur Penelitian Logical Design Berbasis Fountain

3.1. Pre-Processing

Pada tahap *pre-processing* ini mencakup tahapan awal dalam pengolahan data untuk memastikan bahwa data yang telah dikumpulkan siap digunakan dalam pelatihan model. Tahapan ini melibatkan pemilihan dataset, pemecahan data, serta ekstraksi frame data agar model dapat mengenali pola dengan baik. Pada Tabel 1 merupakan hasil observasi peneliti mengenai jenis gestur Bahasa Isyarat Indonesia (BISINDO) berdasarkan jumlah alfabet Bahasa Indonesia yaitu 26 huruf sebagai bahan pelatihan. Dengan huruf alfabet yang berjumlah 26 gestur tersebut. Masing-masing data huruf statis yang berjumlah 24 huruf memiliki 5 variasi gambar (.jpg) untuk setiap huruf. Lalu untuk huruf dinamis yaitu R dan J masing-masing memiliki 5 data berupa video (.mp4) berdurasi kurang lebih 5-6 detik dengan 30 fps, yang selanjutnya diekstrak per frame.



Gambar 3. Distribusi jumlah sample per-kelas

Selanjutnya melakukan memperluas dan memodifikasi dataset dengan menambah variasi data. Tujuannya yaitu untuk meningkatkan generalisasi model agar dapat mudah terdeteksi. Teknik augmentasi yang dilakukan ialah *brightness and contrast* (perubahan pencahayaan), *resizing* (merubah ukuran), *rotate* (rotasi), *flipping horizontal* dan *gaussian blur*.

Sehingga diperoleh jumlah dataset perkelas seperti pada Gambar 3 yang merupakan hasil dari perluasan menggunakan teknik augmentasi tersebut.

3.2. Processing

Pada tahap *processing*, Teknik Penggunaan Sistem Pengenalan BISINDO memanfaatkan *MediaPipe* untuk mengekstraksi dan mengidentifikasi 21 *keypoints* (titik koordinat) yang merepresentasikan struktur dan gerakan tangan, sehingga mampu memberikan informasi spasial yang penting untuk tahap pengenalan isyarat. Selanjutnya, Metode Klasifikasi menggunakan GNNs, penelitian ini membahas pemanfaatan *Graph Neural Networks* (GNNs) sebagai algoritma klasifikasi utama, seperti yang diilustrasikan pada Gambar 1. GNNs dipilih karena kemampuannya dalam memproses data berbasis graf yang sesuai dengan representasi hubungan antar-*keypoints*. Untuk memastikan keandalan hasil klasifikasi, digunakan teknik *K-fold cross-validation* yang membagi dataset ke dalam beberapa lipatan secara seimbang. Pendekatan ini memungkinkan evaluasi performa model secara lebih objektif dan mengurangi potensi *overfitting* pada data pelatihan.

3.3. Post-Processing

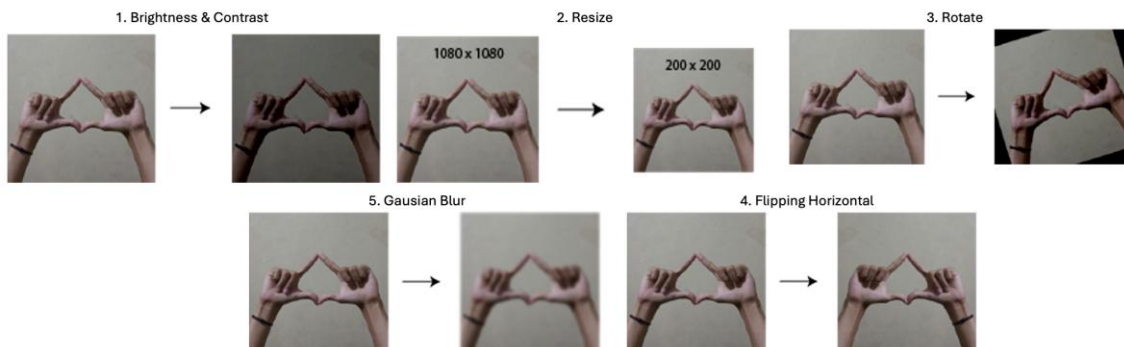
Dalam arsitektur *post-processing* pada Gambar 2, ini menghubungkan model pengenalan bahasa isyarat BISINDO berbasis *Graph Neural Network* (GNNs) di *Backend* dengan UI di *Frontend*, sehingga hasil klasifikasi bahasa isyarat dapat diakses melalui *framework FastAPI* untuk pengelolaan *RESTfull API*, memungkinkan *backend* berfungsi sebagai penyedia data dan UI interaktif, lalu *Flask* yang sebagai *web server* dan *TailwindCss* untuk desain tampilan yang responsif. Arsitektur ini memastikan alur komunikasi yang terstruktur antara *Backend* dan *Frontend*, sehingga hasil pengenalan bahasa isyarat dapat diambil dan ditampilkan secara *real-time*.

4. Hasil dan Diskusi

Bagian ini menyajikan hasil dan diskusi dari setiap tahapan penelitian yang telah dilakukan berdasarkan pendekatan *Fountain* berdasarkan tahap *logical design*, yang terdiri dari tiga fase utama: *pre-processing*, *processing*, dan *post-processing*. Pendekatan ini digunakan untuk memastikan alur pengolahan data berlangsung secara sistematis dan terstruktur mulai dari tahap awal hingga akhir.

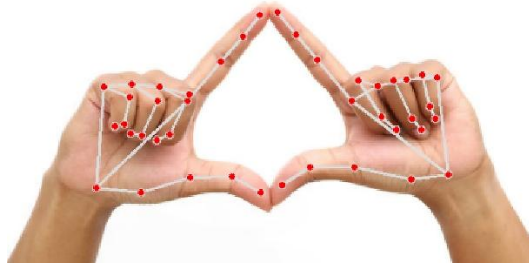
4.1. Pre-Processing

Pada tahap *pre-processing*, dilakukan serangkaian transformasi terhadap dataset untuk meningkatkan kualitas data dan memperkaya variasi sampel yang digunakan dalam pelatihan model. Augmentasi dataset diterapkan melalui lima teknik utama, yaitu perubahan pencahayaan (*brightness* dan *contrast*), *resizing*, *rotasi*, *flipping horizontal*, serta *Gaussian blur*. Selanjutnya melakukan proses pembuatan dataset keypoints berdasarkan Gambar 4, tahap ini diawali dengan merubah warna gambar dari BGR ke RGB lalu diekstrak keypoints menggunakan *MediaPipe Hands* untuk menghasilkan 21 titik koordinat (x, y, z) dari setiap tangan pada gambar, dengan 42 keypoint untuk huruf yang polanya menggunakan 2 tangan.



Gambar 4. Teknik Augmentasi Dataset

Data hasil bentuk 21 keypoint seperti pada Gambar 5 dilakukan ekstrak keypoints tersebut perlu dikonversi menjadi struktur graf agar dapat dikenali oleh model GNN. Proses konversi dilakukan dengan membentuk node. Proses konversi dilakukan dengan membentuk node dari setiap titik keypoints dan menghubungkannya sesuai struktur anatomi tangan ke dalam edge list, sehingga membentuk sebuah objek graf.



Gambar 5. Tahap Ekstraksi 21 Keypoint

4.2. Processing

Tahap ini bertujuan mengukur performa sistem dalam mengenali huruf alfabet BISINDO menggunakan dataset citra berjumlah 32.000 data, yang dibagi menjadi 25.600 data latih dan 6.400 data uji. Proses pelatihan menggunakan teknik *early stopping* untuk memperoleh akurasi terbaik sekaligus mencegah *overfitting*.

Model yang digunakan, yaitu BisindoGNN, merupakan arsitektur *Graph Neural Network* yang dirancang khusus untuk klasifikasi alfabet BISINDO berdasarkan 21 titik keypoint tangan dari MediaPipe. Model ini memanfaatkan kombinasi *Graph Attention Networks* (GAT) dan *Graph Convolutional Networks* (GCN) secara berlapis untuk mengolah data graf hasil ekstraksi keypoints.

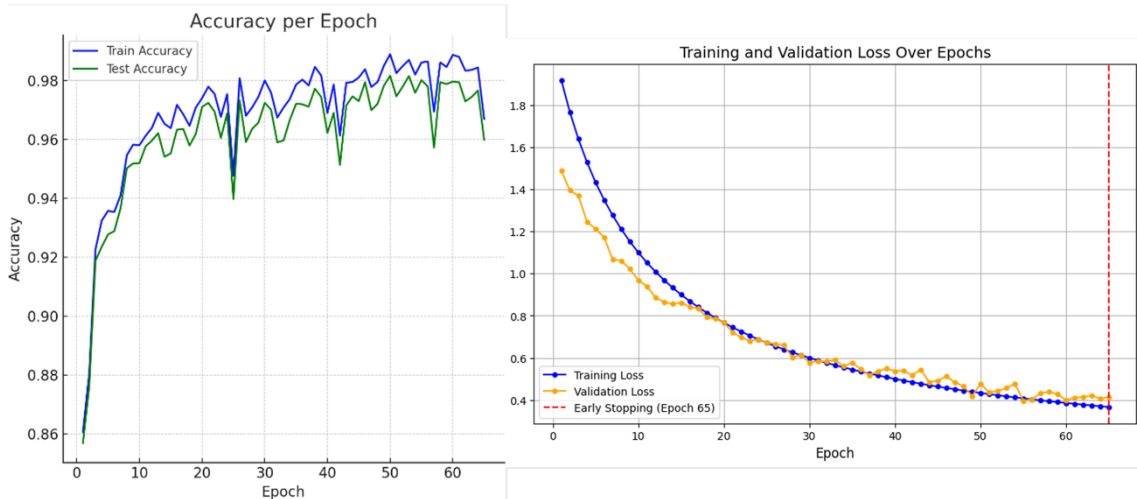
Tabel 1. Arsitektur Lapisan GNN model

Layer	Type	Parameters	Details
gat1	GATConv	3072	In channels=3, out channels=128, heads=4
gat2	GATConv	65920	In channels=512, out channels=128, heads=1
gcn3	GCNConv	16512	In channels=128, out channels=128,
gcn4	GCNConv	16512	In channels=128, out channels=128,
fc	Linear	3354	In features=128, out features=26
dropout	Dropout	0	P=0.6
bn1	BatchNorm1d	1024	num features=512
bn2	BatchNorm1d	256	num features=128
bn3	BatchNorm1d	256	num features=128

Model ini memiliki total 106.906 parameter *trainable* yang dilatih selama proses *training*. Aktivasi yang digunakan pada model adalah fungsi *ReLU*, dan proses agregasi fitur node dilakukan dengan menggunakan metode *global mean pooling*, yaitu meratakan semua *node feature* dalam satu *graph* sebelum masuk ke *fully connected layer* untuk klasifikasi akhir. Arsitektur GNN model yang digunakan dalam penelitian ini terdiri dari penggunaan *epoch*: 100, *patience*: 15, *conv*: 4, *hidden_channels*: 128, *batch_size*: 32, dan *learning_rate*: 0.001, didapatkan hasil yang cukup baik dengan akurasi sebesar 98%, hasil tersebut didapatkan ketika selama proses *training* berhenti di epoch 65 yang menandakan bahwa epoch itulah yang merupakan nilai akurasi terbaik.

Berdasarkan simulasi grafik pada Gambar 6, *training loss*, *validation loss*, serta *training* dan *validation accuracy* dalam penelitian ini menunjukkan tren penurunan yang relatif stabil, sejalan dengan *training loss*. Hal ini mengindikasikan bahwa model mampu melakukan generalisasi dengan baik pada data yang tidak pernah dilihat sebelumnya. Perbedaan antara nilai *training loss* dan *validation loss* yang tidak terlalu besar memperlihatkan bahwa model tidak mengalami *overfitting* yang signifikan. Selain itu, pola penurunan yang konsisten hingga mendekati titik konvergensi menunjukkan bahwa proses pelatihan berjalan optimal. Dengan

adanya penggunaan *early stopping* pada *patience* 15, proses pelatihan dapat dihentikan secara otomatis ketika tidak terjadi peningkatan performa yang berarti, sehingga menghindari pelatihan berlebihan dan menghemat waktu komputasi. Secara keseluruhan, visualisasi *validation loss* memberikan gambaran bahwa model yang dibangun tidak hanya efektif dalam mempelajari data latih, tetapi juga memiliki kemampuan prediksi yang baik pada data uji.



Gambar 6. Grafik Kinerja Model Selama Proses Pelatihan

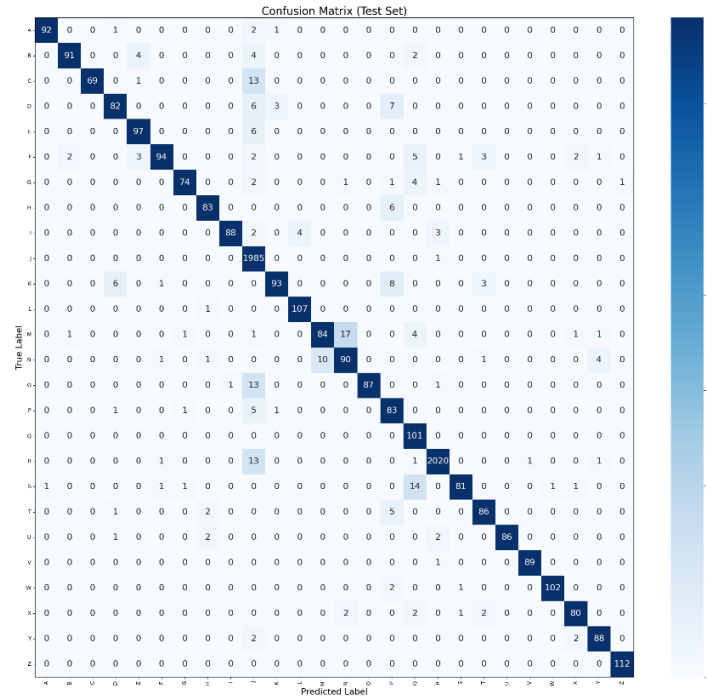
Berdasarkan Gambar 7, model pengenalan alfabet BISINDO menunjukkan performa cukup baik dengan mayoritas prediksi benar berada pada diagonal *confusion matrix*, khususnya pada huruf dengan jumlah data besar. Huruf 'J' (1985 benar), 'R' (2020 benar), dan 'I' (88 benar) menjadi contoh dengan akurasi tinggi. Namun, beberapa huruf masih mengalami kesalahan prediksi, seperti 'M' (84 benar, 17 salah), 'C' (69 benar, 13 salah ke 'J'), dan 'F' yang juga keliru diprediksi ke huruf lain. Kesalahan ini dapat disebabkan oleh kemiripan bentuk tangan antarhuruf atau kualitas citra. Perbedaan jumlah data antar kelas turut memengaruhi hasil, sehingga diperlukan strategi seperti data *augmentation*, penyeimbangan dataset, atau teknik ekstraksi fitur lanjutan untuk meningkatkan performa model, khususnya pada kelas minoritas dan huruf dengan bentuk mirip.

Evaluasi lebih lanjut dilakukan dengan *K-Fold Cross Validation* sebanyak lima lipatan untuk mengukur generalisasi model pada data baru dan menilai stabilitas pelatihan. Dalam skema ini, dataset dibagi menjadi lima bagian proporsional, di mana setiap bagian bergantian digunakan sebagai data uji, sementara sisanya digunakan untuk pelatihan. Proses ini memastikan bahwa performa model tidak hanya baik pada data latih, tetapi juga konsisten pada data uji, dengan konfigurasi evaluasi *fold: 5, epoch: 50, Total Dataset: 32.000, Train: 80%* dan *Test: 20%*.

Pada Tabel 2 merupakan hasil evaluasi dari setiap *fold* yang menunjukkan performa model berdasarkan metrik evaluasi utama, proses *K-Fold Cross Validation* dilakukan sebanyak 5 kali, dimana setiap *fold* mewakili kombinasi unik antara data pelatihan dan data pengujian. Evaluasi dilakukan dengan menghitung nilai *loss, accuracy, precision, dan recall*. Hasil ini menunjukkan bahwa model mampu mencapai akurasi dan *f1-score* di atas 96% pada semua *fold*.

Tabel 2. Hasil Evaluasi Cross Validation

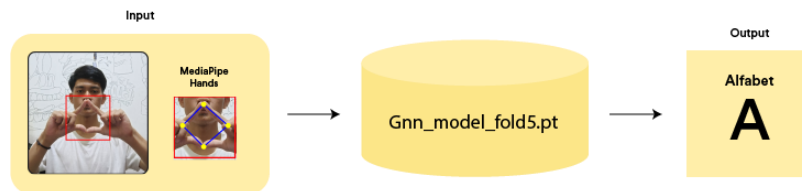
<i>Fold</i>	<i>Eval Loss</i>	<i>Accuracy</i>	<i>Precision</i>	<i>Recall</i>	<i>F1-Score</i>
1	0.10990	0.96766	0.96925	0.96766	0.96792
2	0.12814	0.95813	0.96194	0.95813	0.95896
3	0.11397	0.96594	0.96784	0.96594	0.96628
4	0.14124	0.96016	0.96304	0.96016	0.96070
5	0.10129	0.96766	0.97003	0.96766	0.96816



Gambar 7. Hasil Evaluasi Model Menggunakan Confusion Matrix

4.3. Post-Processing

Post-Processing adalah tahapan untuk mengintegrasikan model yang telah dilatih dengan aplikasi berbasis web agar mudah diakses pengguna dari berbagai platform. Alurnya dimulai dari input gambar tangan (via kamera/webcam). Gambar diproses menggunakan MediaPipe Hands untuk mendapatkan 21 keypoints tangan. Titik-titik ini kemudian dikonversi menjadi struktur graph (node adalah titik, edge adalah hubungan alami tangan). Graph ini menjadi masukan bagi model GNN yang terdiri dari lapisan graph, batch normalization, aktivasi ReLu, dan dropout.



Gambar 8. Ilustrasi integrasi model GNN's dengan Web Apps

Pengujian prototipe melibatkan tiga partisipan dengan variasi bentuk tangan berbeda, yaitu gempal, kurus, dan standar, seperti yang ditampilkan pada Gambar 9. Seluruh uji coba dilakukan pada kondisi pencahayaan seragam dengan jarak pengambilan citra 30 cm untuk menjaga konsistensi. Nilai akurasi ditentukan dari hasil prediksi pada percobaan pertama tanpa pengulangan.

Hasil pengujian pada Tabel 3 menunjukkan partisipan pertama memiliki tingkat akurasi pengenalan huruf BISINDO lebih rendah, terutama pada huruf F, G, H, Q, R, T, dan U yang berada di bawah 96%. Sementara itu, dua partisipan lainnya lebih konsisten dengan rata-rata akurasi 94,5%. Temuan ini mengindikasikan bahwa variasi bentuk tangan, konsistensi gerakan, dan gaya individu memengaruhi performa sistem, sehingga meskipun hasil umum cukup baik, tantangan tetap ada dalam menjaga konsistensi pengenalan pada setiap pengguna.



Gambar 9. Hasil Pengujian Secara Simultan

Tabel 3. Performa Sistem (*Accuracy*) dalam Pengenalan Gestur BISINDO Dinamis

Alfabet	Partisipan 1	Partisipan 2	Partisipan 3	Alfabet	Partisipan 1	Partisipan 2	Partisipan 3
A	94	100	99	N	85	80	100
B	95	99	94	O	92	95	97
C	92	98	86	P	97	97	100
D	85	94	99	Q	56	97	98
E	90	96	99	R	70	98	100
F	70	90	92	S	90	99	99
G	52	97	92	T	75	97	85
H	50	98	85	U	45	99	86
I	99	96	100	V	99	99	92
J	70	95	72	W	99	100	100
K	92	98	99	X	97	98	100
L	99	99	73	Y	97	99	86
M	98	86	80	Z	99	97	100
Accuracy	Partisipan 1	84 %	Partisipan 2	96 %	Partisipan 3	93 %	
Total	91 %						

5. Kesimpulan dan Saran

Penelitian ini berhasil mengimplementasikan sistem klasifikasi alfabet Bahasa Isyarat Indonesia (BISINDO) berbasis citra menggunakan *Graph Neural Networks* (GNNs) dengan ekstraksi fitur keypoints dari *MediaPipe Hands*. Citra tangan pengguna diubah menjadi representasi graf spasial yang kemudian diproses oleh arsitektur GNN untuk mengenali huruf BISINDO. Evaluasi dengan *K-Fold Cross Validation* menunjukkan model mencapai akurasi rata-rata 96,39% dan F1-Score 96,44%, menandakan performa yang sangat baik dalam pengenalan huruf statis. Namun, hasil pengujian prototipe menunjukkan akurasi bervariasi berdasarkan morfologi tangan partisipan. Partisipan dengan tangan gempal cenderung menghasilkan akurasi lebih rendah pada huruf tertentu, sementara partisipan dengan tangan kurus dan standar mencapai akurasi mendekati hasil pelatihan. Secara keseluruhan, prototipe mencatat akurasi rata-rata 91% seperti pada Tabel 3, yang menegaskan sistem cukup baik tetapi masih memiliki keterbatasan.

Tantangan utama ditemukan pada huruf-huruf dinamis seperti J dan R yang membutuhkan deteksi temporal, serta huruf kompleks seperti Z yang melibatkan pergerakan

hingga bahu dan sulit ditangkap oleh MediaPipe Hands. Untuk pengembangan lebih lanjut, disarankan perluasan dataset dengan variasi partisipan, latar belakang, pencahayaan, serta perangkat kamera agar model lebih robust. Selain itu, kombinasi MediaPipe Hands dengan MediaPipe Pose dapat dipertimbangkan untuk mendeteksi keypoints tambahan pada lengan dan bahu. Pengujian sistem di lingkungan nyata seperti komunitas tunarungu serta integrasi ke platform implementasi yang lebih luas, misalnya aplikasi mobile, akan semakin memperkuat validitas dan pemanfaatan sistem ini.

6. Ucapan Terima Kasih

Penulis mengucapkan terima kasih kepada seluruh pihak yang telah memberikan dukungan dan kontribusi dalam pelaksanaan penelitian ini. Terima kasih khusus disampaikan kepada Lembaga Penelitian dan Pengabdian Masyarakat (LPPM), para dosen dan mahasiswa IBIK yang telah memberikan arahan serta masukan konstruktif selama proses penelitian berlangsung. Serta kontribusi literatur dan inspirasi dari penelitian-penelitian terdahulu terkait pengenalan bahasa isyarat.

Referensi

- [1] A. Rahman, and S. Informatika, “*Aplikasi Penerjemah Bahasa Isyarat Menggunakan Metode K-NN (K-Nearest Neighbour)*,” *Jurnal Teknologi Pintar*, vol. 2, no. 4, pp. 1–12, 2022. [Online]. Available: <https://ejurnal.stmik-budidarma.ac.id/mib/article/view/7714>
- [2] 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, 2023, doi: 10.24853/holistika.5.1.28-33. [Online]. Available: <https://jurnal.umj.ac.id/index.php/holistika/article/view/9355>
- [3] T. Tao and Y. Zhao, “*Sign language recognition: A comprehensive review of traditional and deep learning approaches, datasets, and challenges*,” *IEEE Access*, vol. 12, pp. 111216–111234, 2024, doi: 10.1109/ACCESS.2024.3387279
- [4] Y. E. Mulyanto, “*Sign Language Recognition Based on Geometric Features Using Deep Learning*,” *Jurnal Nasional Pendidikan Teknik Informatika: JANAPATI*, vol. 13, no. 2, pp. 338–348, Jul. 2024, doi: 10.23887/janapati.v13i2.82103.
- [5] F. Damatraseta, R. Novariany, and M. A. Ridhani, “*Real-time BISINDO Hand Gesture Detection and Recognition with Deep Learning CNN*,” *Jurnal Informatika Kesatuan*, vol. 1, no. 1, pp. 71–80, Jul. 2021, doi: 10.37641/jikes.v1i1.774.
- [6] Y. V. Via, W. S. J. Saputra, M. I. Fachrurrozi, E. Y. Puspaningrum, F. T. Anggraeny, and S. R. Nudin, “*Object Localization and Detecting Alphabet in Sign Language BISINDO Using Convolution Neural Network*,” *Technium: Romanian Journal of Applied Sciences and Technology*, vol. 5, no. 12, pp. 143–149, 2023. [Online]. Available: <https://ideas.repec.org/s/tec/techni.html>
- [7] H. Kolivand, S. Joudaki, M. S. Sunar, and D. Tully, “*A new framework for sign language alphabet hand posture recognition using geometrical features through artificial neural network (part 1)*,” *Neural Computing and Applications*, vol. 33, no. 10, pp. 4945–4963, May 2021, doi: 10.1007/s00521-020-05279-7.
- [8] B. Khemani, S. Qamar, and S. Anwar, “*A review of graph neural networks: Concepts, architectures, and applications*,” *Journal of Big Data*, vol. 11, no. 1, pp. 1–38, 2024, doi: 10.1186/s40537-024-00815-4
- [9] A. Sharma, A. Sharma, A. Tselykh, A. Bozhenyuk, and B. G. Kim, “*Image and video analysis using graph neural network for Internet of Medical Things and computer vision applications*,” *CAAI Transactions on Intelligent Technology*, 2024, doi: 10.1049/cit2.12306.
- [10] S. Dey, S. Samanta, and S. Banerjee, “*Sign language recognition using convolutional neural networks: A review and open research challenges*,” *International Journal of Advanced Computer Science and Applications (IJACSA)*, vol. 14, no. 5, pp. 123–131, 2023, doi: 10.14569/IJACSA.2023.0140515

- [11] W. Chen, J. Liu, and Z. Luo, “*A survey on hand pose estimation with wearable sensors and computer-vision-based methods,*” *Sensors*, vol. 20, no. 3, pp. 1–25, Feb. 2020, doi: 10.3390/s20030856
- [12] H. Ansar, M. Ali, and S. Ahmed, “*Hand gesture recognition based on auto-landmark localization,*” *Sustainability*, vol. 13, no. 17, pp. 1–18, 2021, doi: 10.3390/su13179902