Journal of Artificial Intelligence and Technology Information
Website : https://ejournal.techcart-press.com/index.php/jaiti
E-ISSN 2985-6396
P-ISSN 2985-5306
Volume 4, Number 2, Juni 2026
https://doi.org/10.58602/jaiti.v4i2.253
Page 92-106

Klasifikasi Gerakan Tari Bali Perempuan Menggunakan Metode Spatial-Temporal Graph Convolutional Network (ST-GCN)

Daniel Sande Bona
Desain Komunikasi Visual, Institut Seni Budaya Indonesia Tanah Papua, Jayapura, Indonesia
daniel.sande.bona@gmail.com

Abstrak: Pengenalan gerakan tari Bali berpotensi mendukung dokumentasi warisan budaya, media pembelajaran, dan sistem umpan balik gerak berbasis komputer. Namun, model klasifikasi video RGB dapat mempelajari latar belakang, kostum, pencahayaan, atau identitas penari sebagai shortcut, bukan struktur gerakan. Penelitian ini bertujuan menyusun baseline Spatial-Temporal Graph Convolutional Network (ST-GCN) berbasis skeleton untuk klasifikasi 13 gerakan dasar tari Bali perempuan menggunakan MediaPipe Pose. Setiap video diproses menjadi 33 landmark tubuh dengan kanal koordinat x, koordinat y, dan visibility, kemudian distandarkan menjadi 64 frame. Folder train dan validation asli digabungkan hanya untuk validasi silang 5-fold berstratifikasi, sedangkan folder test resmi dipertahankan sebagai holdout akhir. Model menggunakan graf 33 landmark MediaPipe, backbone ST-GCN, dan GCNHead dengan global pooling serta linear classifier. Hasil validasi silang memperoleh top-1 accuracy 99,55% +/- 0,49%, top-5 accuracy 99,92% +/- 0,17%, dan macro F1 99,49% +/- 0,55%. Evaluasi holdout akhir satu kali menghasilkan top-1 accuracy 99,39%, top-5 accuracy 100,00%, dan macro F1 99,39%. Audit duplikasi identifier dan overlap SHA-256 tidak menemukan kebocoran data train-test. Strategi evaluasi ini menegaskan pemisahan antara validasi model dan pengujian akhir. Hasil ini menunjukkan bahwa ST-GCN berbasis skeleton menjadi baseline within-dataset yang kuat untuk pengenalan gerak tari Bali, meskipun generalisasi subject-independent belum dapat diklaim karena dataset tidak menyediakan identitas penari.
Kata Kunci: ST-GCN; tari Bali; skeleton; pengenalan aksi; deep learning; MediaPipe Pose;

Abstract: Balinese dance movement recognition can support cultural heritage documentation, learning media, and computer-based motion feedback systems. However, RGB video classifiers may learn background, costume, lighting, or dancer identity as shortcuts rather than the movement structure itself. This study aims to establish a skeleton-based Spatial-Temporal Graph Convolutional Network (ST-GCN) baseline for classifying 13 basic Balinese women dance movements using MediaPipe Pose. Each video was processed into 33 body landmarks with x-coordinate, y-coordinate, and visibility channels, then standardized to 64 frames. The original train and validation folders were

Daniel Sande Bona: * Penulis Korespondensi
Copyright © 2026, Daniel Sande Bona.

Journal of Artificial Intelligence and Technology Information
Website : https://ejournal.techcart-press.com/index.php/jaiti
E-ISSN 2985-6396
P-ISSN 2985-5306
Volume 4, Number 2, Juni 2026
https://doi.org/10.58602/jaiti.v4i2.253
Page 92-106

combined only for stratified 5-fold cross-validation, while the official test folder was preserved as the final holdout. The model used a MediaPipe 33-landmark graph, an ST-GCN backbone, and a GCNHead with global pooling and a linear classifier. Cross-validation achieved 99.55% +/- 0.49% top-1 accuracy, 99.92% +/- 0.17% top-5 accuracy, and 99.49% +/- 0.55% macro F1. A single final holdout evaluation achieved 99.39% top-1 accuracy, 100.00% top-5 accuracy, and 99.39% macro F1. Duplicate identifier and SHA-256 overlap audits found no train-test data leakage. This protocol clarifies the separation between model validation and final testing. These results show that skeleton-based ST-GCN is a strong within-dataset baseline for Balinese dance movement recognition, although subject-independent generalization cannot yet be claimed because dancer identity metadata are unavailable.

Keywords: ST-GCN; Balinese dance; skeleton; action recognition; deep learning; MediaPipe Pose;

1. PENDAHULUAN

Tari Bali merupakan praktik budaya tradisional yang memiliki karakter gerak khas, antara lain postur tubuh, gestur tangan, arah pandang mata, kontrol torso, serta koordinasi anggota tubuh. Pengenalan otomatis terhadap unit gerak dasar dapat membantu dokumentasi budaya digital, pengembangan media pembelajaran interaktif, dan sistem evaluasi gerak berbasis komputer. Dalam konteks preservasi budaya, sistem pengenalan gerakan tidak dimaksudkan menggantikan guru tari, tetapi menjadi alat bantu dokumentasi dan pembanding pola gerak yang dapat diperiksa ulang.

Dataset dan penelitian terkini tentang tari Bali serta tari tradisional Indonesia menunjukkan bahwa gerakan tari dapat dianalisis secara komputasional melalui video dan teknik pengenalan pola [1]-[4]. Lantara et al. menyediakan dataset gerakan dasar tari Bali perempuan dan melaporkan baseline RGB dengan VGG16-LSTM dan I3D [1]. Dataset publik versi Mendeley kemudian menyediakan struktur train, validation, dan test untuk 13 kelas gerak dasar [2]. Studi tahun 2025 juga mengusulkan sistem CNN untuk klasifikasi tari Bali dengan hyperparameter optimization dan penerapan aplikasi web [3]. Di luar tari Bali, klasifikasi tari tradisional Indonesia berbasis CNN-LSTM dan estimasi pose menunjukkan bahwa informasi pose dapat membantu pemodelan temporal gerak [4].

Sebagian besar sistem pengenalan aksi bekerja langsung pada video RGB. Model CNN, recurrent neural network, dan 3D convolutional network dapat mempelajari pola diskriminatif dari tampilan dan perubahan gerak. Namun pada dataset tari budaya yang relatif kecil atau direkam dalam kondisi terkontrol, model berbasis piksel dapat memanfaatkan shortcut dari latar, kostum, pencahayaan, atau identitas penari. Shortcut tersebut dapat menaikkan akurasi di dalam dataset, tetapi belum tentu mencerminkan pemahaman terhadap struktur gerakan. Oleh karena itu, representasi skeleton menjadi alternatif penting karena mengurangi ketergantungan terhadap tampilan piksel dan lebih menekankan lintasan sendi serta konfigurasi tubuh.

Spatial-Temporal Graph Convolutional Network (ST-GCN) memodelkan rangka manusia sebagai graf spasial-temporal. Node merepresentasikan sendi atau landmark tubuh, sedangkan edge merepresentasikan hubungan anatomis antar landmark dalam satu frame dan hubungan temporal landmark yang sama pada frame berurutan [5]. Formulasi ini sesuai untuk pengenalan gerakan tari karena kelas gerakan sering ditentukan oleh kombinasi orientasi badan, posisi tungkai, penempatan lengan, timing transisi, serta pola gestur berulang.

Daniel Sande Bona: * Penulis Korespondensi
Copyright © 2026, Daniel Sande Bona.

Journal of Artificial Intelligence and Technology Information
Website : https://ejournal.techcart-press.com/index.php/jaiti
E-ISSN 2985-6396
P-ISSN 2985-5306
Volume 4, Number 2, Juni 2026
https://doi.org/10.58602/jaiti.v4i2.253
Page 92-106

Metode graf setelah ST-GCN, seperti PoseC3D dan model graph convolution yang memanfaatkan perhatian, semantik, atau fitur pusat gravitasi, menunjukkan bahwa representasi skeleton tetap relevan dalam pengenalan aksi modern [6]-[10]. Pada domain seni pertunjukan, penelitian motion capture untuk gerak tari, klasifikasi tari klasik India, klasifikasi raga Hindustani berbasis pose, serta klasifikasi gerak Muay Thai berbasis sekuens pose menguatkan bahwa gerakan budaya memiliki struktur spasial-temporal yang dapat dipelajari secara komputasional [11], [12], [13]-[15]. Walaupun demikian, belum ditemukan implementasi ST-GCN yang secara khusus diposisikan sebagai baseline untuk tari tradisional Indonesia.

Dalam konteks pengembangan sistem cerdas yang lebih umum, artikel SKANIKA juga memperlihatkan pemanfaatan deep learning untuk klasifikasi citra kesehatan dan deteksi objek secara real-time [16], [17]. Dua contoh tersebut tidak langsung membahas tari, tetapi relevan untuk menunjukkan bahwa pendekatan pembelajaran mendalam telah menjadi metode yang lazim pada persoalan visual. Perbedaan utama penelitian ini terletak pada pemilihan representasi skeleton agar fokus model lebih dekat ke struktur gerak tubuh.

Tujuan penelitian ini adalah membangun pipeline ST-GCN berbasis skeleton yang dapat direproduksi untuk klasifikasi 13 gerakan dasar tari Bali perempuan menggunakan MediaPipe Pose, kemudian melaporkan konfigurasi data secara transparan, dan menyertakan audit kebocoran data untuk menafsirkan akurasi model secara hati-hati. Setelah tujuan tersebut ditetapkan, penelitian ini menyajikan baseline ST-GCN berbasis skeleton untuk klasifikasi 13 gerakan dasar tari Bali perempuan menggunakan MediaPipe Pose. Implementasi menggunakan backbone ST-GCN dengan graf 33 landmark MediaPipe dan kepala klasifikasi GCNHead dari MMAction2. Berbeda dari klasifikasi akhir ST-GCN asli yang menggunakan konvolusi 1 x 1, implementasi ini menggunakan global pooling pada dimensi temporal dan landmark, kemudian linear classifier untuk 13 kelas.

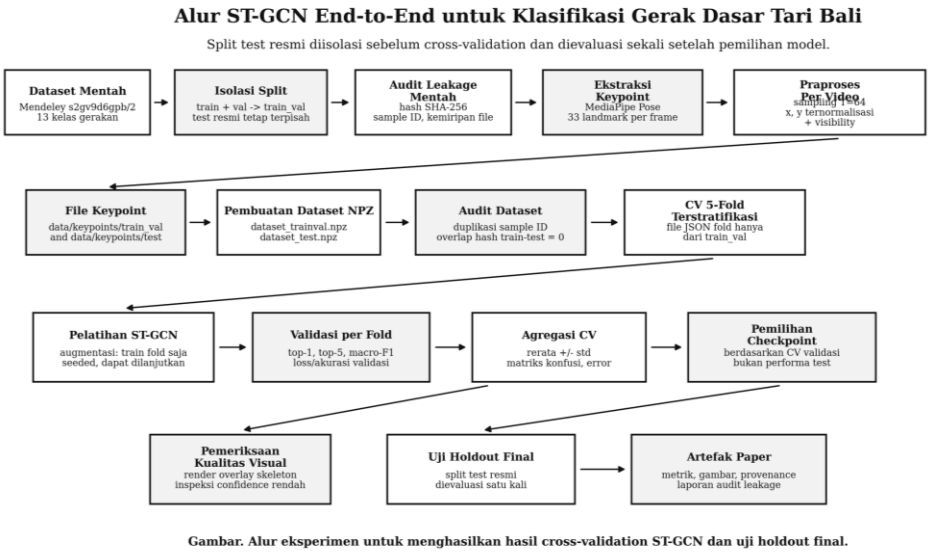
Kontribusi penelitian ini adalah sebagai berikut. Pertama, penelitian ini menyusun pipeline pengenalan gerakan tari Bali berbasis MediaPipe Pose dan ST-GCN. Kedua, penelitian ini melaporkan validasi silang 5-fold berstratifikasi dan evaluasi holdout akhir yang dipisahkan dari proses pemilihan model. Ketiga, penelitian ini menyertakan audit duplikasi identifier dan overlap SHA-256 untuk mengurangi risiko kebocoran data. Keempat, penelitian ini menempatkan hasil sebagai baseline within-dataset karena dataset tidak menyediakan label identitas penari, sehingga generalisasi subject-independent melalui leave-one-subject-out belum dapat diklaim.

2. METODE PENELITIAN

Metode penelitian terdiri atas beberapa tahap: pengumpulan data publik, ekstraksi pose, pra-pemrosesan landmark, audit kualitas data, pembentukan split evaluasi, pelatihan model ST-GCN, dan evaluasi performa. Tahapan ini dirancang agar hasil akurasi tidak hanya dilaporkan sebagai angka, tetapi juga disertai informasi tentang representasi data, kualitas pose, dan perlindungan terhadap kebocoran data. Secara berurutan, video mentah diubah menjadi landmark MediaPipe, diseragamkan ke panjang 64 frame, diaudit kualitasnya, lalu digunakan sebagai input model graph convolution.

Daniel Sande Bona: * Penulis Korespondensi
Copyright © 2026, Daniel Sande Bona.

Journal of Artificial Intelligence and Technology Information
Website : https://ejournal.techcart-press.com/index.php/jaiti
E-ISSN 2985-6396
P-ISSN 2985-5306
Volume 4, Number 2, Juni 2026
https://doi.org/10.58602/jaiti.v4i2.253
Page 92-106



Gambar 1. Alur penelitian klasifikasi gerakan tari Bali berbasis skeleton

Gambar 1 memperlihatkan alur penelitian mulai dari pengumpulan video, ekstraksi pose MediaPipe, pembentukan dataset skeleton, audit kebocoran data, pelatihan ST-GCN, hingga evaluasi. Alur ini mendukung penelitian dengan menunjukkan bahwa proses eksperimen dilakukan secara berurutan dan terpisah antara validasi model dan pengujian akhir.

2.1. Dataset Gerakan Dasar Tari Bali Perempuan

Dataset yang digunakan adalah dataset publik Mendeley Data berisi video gerakan dasar tari Bali perempuan [2]. Dataset memiliki 13 kelas, yaitu Agem Kanan, Agem Kiri, Ngeed, Ngegol, Ngelo, Ngelung, Ngeseh, Ngumbang, Nyalud, Nyeregseg, Seledet, Tapak Sirangpada, dan Ulap-Ulap. Folder train dan validation asli digabungkan menjadi train_val untuk validasi silang berstratifikasi. Folder test resmi tidak digunakan untuk tuning hyperparameter, pemilihan epoch, atau early stopping, melainkan disimpan sebagai holdout akhir.

Beberapa video memiliki framing yang berbeda antar kelas. Sebagian kelas menampilkan gerakan tubuh penuh, sedangkan kelas lain lebih menonjolkan bagian tubuh tertentu. Kondisi ini penting karena MediaPipe Pose dapat mengembalikan representasi 33 landmark tubuh penuh meskipun sebagian tubuh tidak sepenuhnya terlihat.

Video dengan kualitas deteksi pose rendah diberi tanda peringatan pada tahap pra-pemrosesan. Satu frame dianggap berhasil terdeteksi apabila MediaPipe mengembalikan koordinat landmark non-nol. Apabila rasio frame gagal deteksi mencapai 20% atau lebih, video diberi label warning_low_detection. Pada eksperimen ini, video dengan peringatan deteksi rendah dikeluarkan dari dataset train_val dan test agar model tidak dilatih pada sampel yang kualitas skeleton-nya sangat lemah. Distribusi sampel setelah pra-pemrosesan disajikan pada Tabel 1.

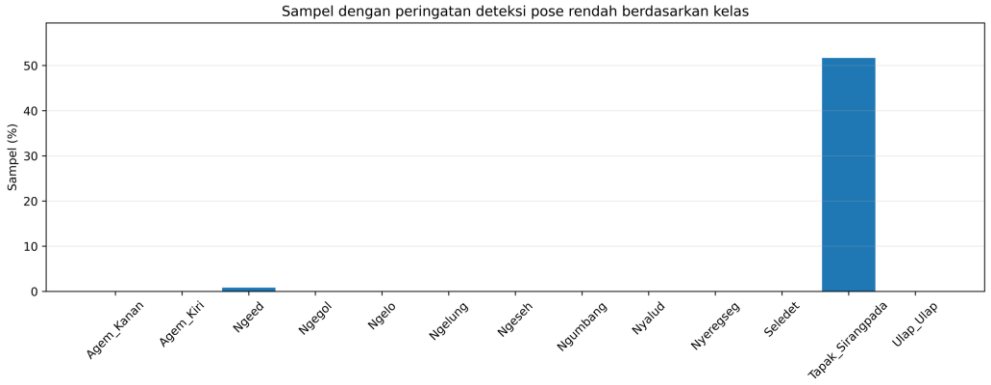
Tabel 1. Distribusi Sampel Dataset Setelah Pra-Pemrosesan

Kelas	Train+Val	Test
Agem Kanan	96	24
Agem Kiri	96	24

Daniel Sande Bona: * Penulis Korespondensi
Copyright © 2026, Daniel Sande Bona.

Journal of Artificial Intelligence and Technology Information
Website : https://ejournal.techcart-press.com/index.php/jaiti
E-ISSN 2985-6396
P-ISSN 2985-5306
Volume 4, Number 2, Juni 2026
https://doi.org/10.58602/jaiti.v4i2.253
Page 92-106

Ngeed	119	30
Ngegol	96	24
Ngelo	118	30
Ngelung	120	30
Ngeseh	120	29
Ngumbang	120	30
Nyalud	96	24
Nyeregseg	96	24
Seledet	96	24
Tapak Sirangpada	58	12
Ulap-Ulap	96	24
Total	1327	329



Gambar 2. Persentase sampel dengan peringatan deteksi pose rendah per kelas

Gambar 2 menunjukkan persentase sampel yang diberi peringatan deteksi pose rendah pada setiap kelas. Tahap ini diperlukan karena membantu menilai kualitas input skeleton sebelum data digunakan untuk pelatihan dan evaluasi model.

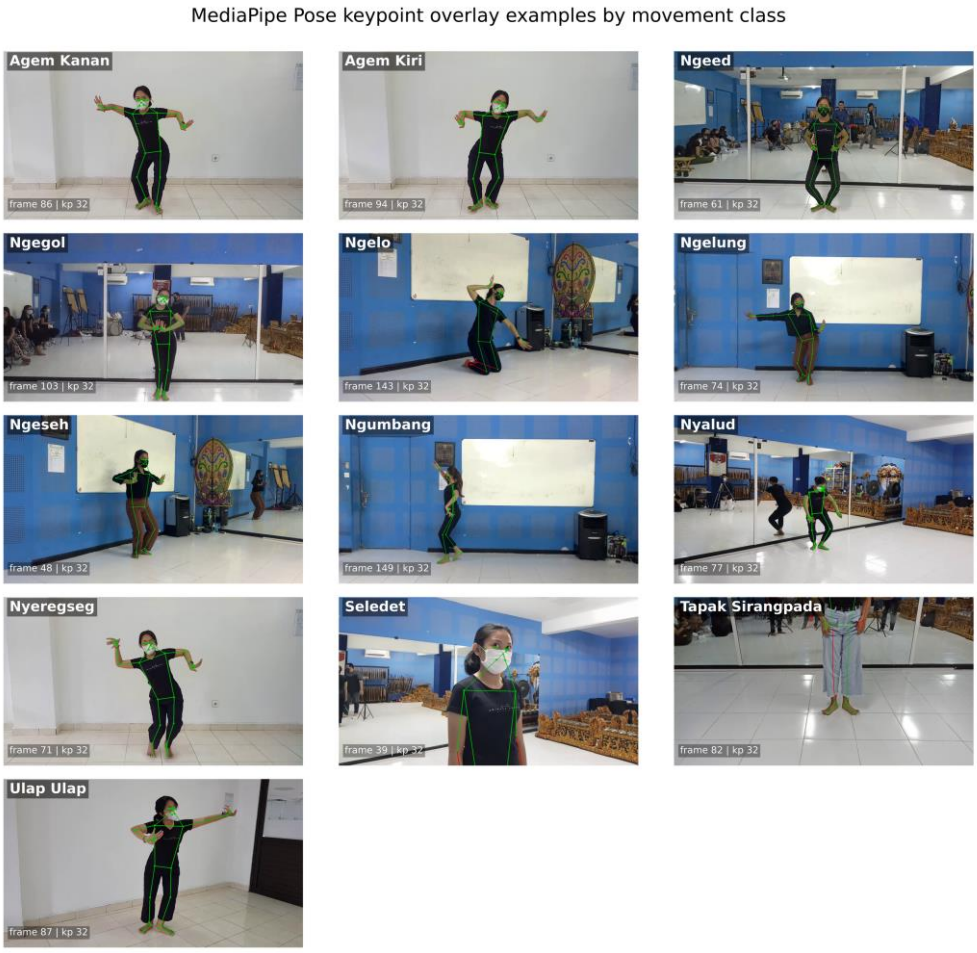
2.2. Ekstraksi Pose dan Pra-pemrosesan

Setiap video diproses menggunakan MediaPipe Pose untuk memperoleh 33 landmark tubuh [18]. Untuk setiap frame yang disampling, tiga kanal fitur disimpan, yaitu koordinat x, koordinat y, dan skor visibility. Setiap sampel direpresentasikan sebagai tensor dengan bentuk T x V x C, dengan T = 64 frame, V = 33 landmark, dan C = 3 kanal. Kanal koordinat dinormalisasi per video sehingga tidak ada global scaler yang di-fit pada seluruh dataset. Keputusan ini penting untuk menghindari penggunaan statistik global yang dapat mencampurkan informasi antar subset evaluasi.
Panjang temporal distandarkan menjadi 64 frame. Jika video memiliki lebih dari 64 frame, frame disampling secara merata. Jika video memiliki kurang dari 64 frame, sekuens dipadatkan melalui padding atau pengulangan sesuai routine pra-pemrosesan. Representasi tetap ini memungkinkan model ST-GCN menerima input dengan dimensi temporal yang konsisten. Kanal visibility dipertahankan karena dalam gerakan tari, kostum, arah tubuh, dan oklusi tangan atau kaki dapat memengaruhi reliabilitas landmark tertentu.
Normalisasi dilakukan pada tingkat video, bukan pada keseluruhan dataset. Dengan cara ini, koordinat tubuh dipusatkan dan diskalakan berdasarkan informasi dari sampel yang

Daniel Sande Bona: * Penulis Korespondensi
Copyright © 2026, Daniel Sande Bona.

Journal of Artificial Intelligence and Technology Information
Website : https://ejournal.techcart-press.com/index.php/jaiti
E-ISSN 2985-6396
P-ISSN 2985-5306
Volume 4, Number 2, Juni 2026
https://doi.org/10.58602/jaiti.v4i2.253
Page 92-106

sama. Strategi tersebut membuat input lebih stabil terhadap variasi posisi penari dalam frame, tetapi tidak menggunakan statistik dari validation fold atau test.



Gambar 3. Contoh overlay keypoint MediaPipe Pose pada 13 kelas Gerakan

Gambar 3 menampilkan contoh overlay landmark MediaPipe Pose pada video dari tiap kelas gerakan. Visualisasi ini mendukung penelitian dengan memperlihatkan bahwa pola tubuh penari dapat direpresentasikan sebagai skeleton yang menjadi input utama ST-GCN.

2.3. Model ST-GCN

Graf ST-GCN mengikuti topologi 33 landmark MediaPipe Pose. Edge spasial menghubungkan pasangan landmark yang secara anatomis berdekatan, sedangkan edge temporal menghubungkan landmark yang sama pada frame berurutan. Input backbone diorganisasikan sebagai C x T x V x M, dengan C = 3, T = 64, V = 33, dan M = 1 orang terdeteksi. Backbone terdiri dari 10 blok ST-GCN dengan progresi kanal 3-64-64-64-64-128-128-128-256-256-256. Downsampling temporal dilakukan pada tahap 128 kanal dan 256 kanal melalui stride temporal.
Pada setiap blok, unit_gcn melakukan agregasi spasial menggunakan adjacency matrix MediaPipe, sedangkan unit_tcn melakukan konvolusi temporal untuk menangkap dinamika gerak antar frame. Secara ringkas, graph convolution menggabungkan fitur dari node bertetangga berdasarkan partisi graf. Penelitian ini menggunakan tiga partisi, yaitu self,

Daniel Sande Bona: * Penulis Korespondensi
Copyright © 2026, Daniel Sande Bona.

Journal of Artificial Intelligence and Technology Information
Website : https://ejournal.techcart-press.com/index.php/jaiti
E-ISSN 2985-6396
P-ISSN 2985-5306
Volume 4, Number 2, Juni 2026
https://doi.org/10.58602/jaiti.v4i2.253
Page 92-106

centripetal, dan centrifugal. Partisi tersebut mengikuti gagasan ST-GCN asli bahwa tetangga graf tidak selalu memiliki kontribusi yang sama terhadap node pusat.
Setelah backbone, GCNHead melakukan adaptive average pooling pada dimensi temporal dan landmark, merata-ratakan dimensi person, dan meneruskan vektor fitur ke linear classifier 13 kelas. Implementasi ini tetap mempertahankan inti ST-GCN, tetapi berbeda dari kepala klasifikasi ST-GCN asli yang memakai final 1 x 1 convolution. Perbedaan ini dicatat sebagai adaptasi implementasi, bukan kontribusi arsitektur baru.
Rincian arsitektur implementasi disajikan pada Tabel 2, sedangkan Tabel 3 membedah isi ST-GCN Block menjadi spatial graph convolution, temporal convolution, residual branch, dan aktivasi akhir. Notasi N adalah batch size, M=1 person, T=64 frame, V=33 landmark, dan C jumlah kanal fitur.

Tabel 2. Arsitektur ST-GCN keseluruhan yang digunakan

No.	Layer/Stage	Input Shape	Output Shape	Jumlah Parameter
1	Input skeleton	(N, 1, 64, 33, 3)	(N, 1, 64, 33, 3)	0
2	Data BN (VC)	(N, 1, 64, 33, 3)	(N*M, 3, 64, 33)	198
3	ST-GCN Block 1 (3->64, s=1)	(N*M, 3, 64, 33)	(N*M, 64, 64, 33)	41.219
4	ST-GCN Block 2 (64->64, s=1)	(N*M, 64, 64, 33)	(N*M, 64, 64, 33)	52.931
5	ST-GCN Block 3 (64->64, s=1)	(N*M, 64, 64, 33)	(N*M, 64, 64, 33)	52.931
6	ST-GCN Block 4 (64->64, s=1)	(N*M, 64, 64, 33)	(N*M, 64, 64, 33)	52.931
7	ST-GCN Block 5 (64->128, s=2)	(N*M, 64, 64, 33)	(N*M, 128, 32, 33)	184.899
8	ST-GCN Block 6 (128->128, s=1)	(N*M, 128, 32, 33)	(N*M, 128, 32, 33)	200.899
9	ST-GCN Block 7 (128->128, s=1)	(N*M, 128, 32, 33)	(N*M, 128, 32, 33)	200.899
10	ST-GCN Block 8 (128->256, s=2)	(N*M, 128, 32, 33)	(N*M, 256, 16, 33)	726.979
11	ST-GCN Block 9 (256->256, s=1)	(N*M, 256, 16, 33)	(N*M, 256, 16, 33)	791.747
12	ST-GCN Block 10 (256->256, s=1)	(N*M, 256, 16, 33)	(N*M, 256, 16, 33)	791.747
13	Backbone reshape	(N*M, 256, 16, 33)	(N, 1, 256, 16, 33)	0
14	GCNHead (avg pool + FC)	(N, 1, 256, 16, 33)	(N, 13)	3.341
	Total Parameter			3.100.721

Tabel 3. Rincian layer internal pada ST-GCN Block

No.	Bagian	Layer/Operasi di Dalam Blok	Input Shape	Output Shape
1	Spatial graph convolution	Adjacency importance A x PA untuk 3 partisi graf: self, centripetal, centrifugal	(N*M, C_in, T, V)	(3, V, V)

Daniel Sande Bona: * Penulis Korespondensi
Copyright © 2026, Daniel Sande Bona.

Journal of Artificial Intelligence and Technology Information
Website : https://ejournal.techcart-press.com/index.php/jaiti
E-ISSN 2985-6396
P-ISSN 2985-5306
Volume 4, Number 2, Juni 2026
https://doi.org/10.58602/jaiti.v4i2.253
Page 92-106

2	Spatial graph convolution	Conv2D 1x1: C_in -> 3*C_out, lalu reshape per partisi graf	(N*M, C_in, T, V)	(N*M, 3, C_out, T, V)
3	Spatial graph convolution	Graph aggregation/einsum terhadap adjacency MediaPipe	(N*M, 3, C_out, T, V)	(N*M, C_out, T, V)
4	Spatial graph convolution	BatchNorm2D + ReLU	(N*M, C_out, T, V)	(N*M, C_out, T, V)
5	Temporal convolution	Conv2D temporal kernel 9x1, stride s pada dimensi T	(N*M, C_out, T, V)	(N*M, C_out, T/s, V)
6	Temporal convolution	BatchNorm2D + Dropout (dropout=0 pada konfigurasi blok default)	(N*M, C_out, T/s, V)	(N*M, C_out, T/s, V)
7	Residual branch	Identity, projection 1x1 + BN, atau disabled pada blok pertama	(N*M, C_in, T, V)	(N*M, C_out, T/s, V)
8	Output blok	Penjumlahan output TCN dan residual, lalu ReLU	(N*M, C_out, T/s, V)	(N*M, C_out, T/s, V)

2.4. Konfigurasi Pelatihan

Pelatihan dilakukan menggunakan PyTorch dan MMAction2 dengan custom dataset adapter dan wrapper BaliRecognizerGCN [19]. Seed deterministik diterapkan pada random, NumPy, PyTorch, dan CUDA sejauh memungkinkan. Optimizer yang digunakan adalah SGD dengan learning rate awal 0,05, momentum 0,9, weight decay 0,0005, batch size 16, dan dropout 0,5. Ringkasan hyperparameter ditampilkan pada Tabel 4.

Tabel 4. Hyperparameter pelatihan baseline ST-GCN

Parameter	Nilai
Representasi input	Landmark MediaPipe Pose
Tensor input	C x T x V = 3 x 64 x 33
Jumlah kelas	13
Backbone	ST-GCN baseline
Graf spasial	Topologi 33 landmark MediaPipe
Panjang temporal	64 frame
Optimizer	SGD
Learning rate awal	0,05
Momentum	0,9
Weight decay	0,0005
Batch size	16
Dropout	0,5
Learning-rate schedule	MultiStepLR
Epoch maksimum	65
Early stopping	Berdasarkan validation loss
Metrik	Top-1, Top-5, Macro F1, confusion matrix

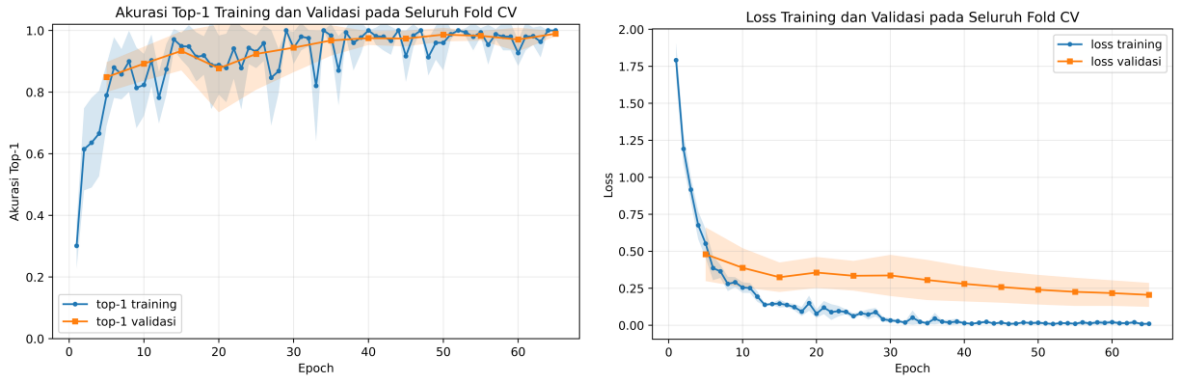
Daniel Sande Bona: * Penulis Korespondensi
Copyright © 2026, Daniel Sande Bona.

Journal of Artificial Intelligence and Technology Information
Website : https://ejournal.techcart-press.com/index.php/jaiti
E-ISSN 2985-6396
P-ISSN 2985-5306
Volume 4, Number 2, Juni 2026
https://doi.org/10.58602/jaiti.v4i2.253
Page 92-106

2.5. Protokol Evaluasi dan Audit Kebocoran Data

Folder train dan validation asli digabungkan menjadi train_val, kemudian dibagi menjadi 5 fold berstratifikasi. Evaluasi validasi silang melaporkan top-1 accuracy, top-5 accuracy, dan macro F1. Fold digunakan untuk memperkirakan stabilitas model pada data train_val. Test resmi tidak digunakan pada tahap ini dan hanya dievaluasi satu kali setelah checkpoint dipilih.
Audit kebocoran data dilakukan melalui dua cara. Pertama, identifier sampel diperiksa untuk memastikan tidak ada nama sampel yang sama muncul pada train_val dan test. Kedua, hash SHA-256 file dicek untuk mendeteksi overlap byte-level antar subset. Audit ini penting karena akurasi yang sangat tinggi pada dataset video kecil dapat terjadi akibat duplikasi klip, near-duplicate, atau penggunaan test data selama tuning. Hasil test holdout tidak digunakan untuk mengubah hyperparameter.
Batasan evaluasi juga dicatat sejak awal. Dataset tidak menyediakan identitas penari, sehingga penelitian ini tidak dapat melakukan leave-one-subject-out cross-validation. Oleh karena itu, hasil tidak ditafsirkan sebagai bukti generalisasi terhadap penari baru. Klaim yang lebih tepat adalah baseline within-dataset yang transparan untuk dataset publik yang sama.

3. HASIL DAN PEMBAHASAN



Gambar 4. Kurva akurasi top-1 dan loss training-validasi pada seluruh fold CV

Gambar 4 menampilkan peningkatan akurasi dan penurunan loss selama pelatihan dan validasi pada seluruh fold. Grafik ini mendukung analisis performa dengan memperlihatkan pola konvergensi model serta perbedaan perilaku antara data pelatihan dan validasi dan tidak menunjukan trend overfitting.

3.1. Analisis Kurva Akurasi dan Loss

Gambar 4 menunjukkan akurasi training naik cepat pada epoch awal dan stabil mendekati 1,0, sedangkan akurasi validasi meningkat lebih halus namun tetap tinggi. Loss training turun tajam, sementara loss validasi menurun lebih lambat dan konsisten lebih besar daripada loss training. Pola ini menunjukkan model cepat mempelajari representasi train_val, tetapi hasil tetap perlu dibaca bersama audit kebocoran data dan evaluasi *holdout*.

Daniel Sande Bona: * Penulis Korespondensi
Copyright © 2026, Daniel Sande Bona.

Journal of Artificial Intelligence and Technology Information
Website : https://ejournal.techcart-press.com/index.php/jaiti
E-ISSN 2985-6396
P-ISSN 2985-5306
Volume 4, Number 2, Juni 2026
https://doi.org/10.58602/jaiti.v4i2.253
Page 92-106

3.2. Hasil Validasi Silang

Tabel 5 menyajikan hasil validasi silang 5-fold berstratifikasi. Rata-rata top-1 accuracy adalah 99,55% dengan simpangan baku 0,49%. Top-5 accuracy mencapai 99,92% +/- 0,17%, sedangkan macro F1 mencapai 99,49% +/- 0,55%. Nilai macro F1 yang hampir sama dengan top-1 menunjukkan bahwa performa tidak hanya didorong oleh kelas besar, tetapi juga relatif konsisten pada sebagian besar kelas.

Tabel 5. Performa validasi silang 5-fold berstratifikasi

Fold	Top-1 (%)	Top-5 (%)	Macro F1 (%)
0	99,62	99,62	99,62
1	100,00	100,00	100,00
2	98,87	100,00	98,83
3	99,25	100,00	98,99
4	100,00	100,00	100,00
Rata-rata +/- std	99,55 +/- 0,49	99,92 +/- 0,17	99,49 +/- 0,55

Hasil tersebut menunjukkan bahwa pipeline skeleton berbasis MediaPipe dan ST-GCN sangat efektif dalam memisahkan 13 kelas gerakan pada kondisi dataset yang sama. Fold dengan performa terendah tetap mencapai top-1 98,87%, sehingga tidak terlihat fold yang gagal secara ekstrem. Namun, nilai yang sangat tinggi perlu ditafsirkan hati-hati karena dataset tidak memiliki label identitas penari. Tanpa identitas penari, evaluasi leave-one-subject-out tidak dapat dilakukan.
Pada skenario pembelajaran tari, nilai top-5 juga berguna untuk membaca kedekatan antar gerakan. Jika prediksi top-1 salah tetapi label benar tetap berada pada kandidat lima besar, sistem masih dapat memberikan alternatif koreksi untuk guru atau pelajar. Namun, metrik top-5 tidak boleh menggantikan top-1 accuracy karena keputusan akhir sistem klasifikasi tetap membutuhkan satu label prediksi utama.

3.3. Evaluasi Holdout Akhir

Evaluasi holdout akhir dilakukan satu kali pada folder test resmi. Checkpoint yang dipilih adalah checkpoint fold 1 dengan validation top-1 100,00%. Hasil holdout ditampilkan pada Tabel 6. Top-1 accuracy mencapai 99,39%, top-5 accuracy mencapai 100,00%, dan macro F1 mencapai 99,39%.

Tabel 6. Performa holdout test akhir

Metrik	Nilai
Top-1 accuracy	99,39%
Top-5 accuracy	100,00%
Macro F1	99,39%
Fold terpilih	1
Validation top-1 fold terpilih	100,00%

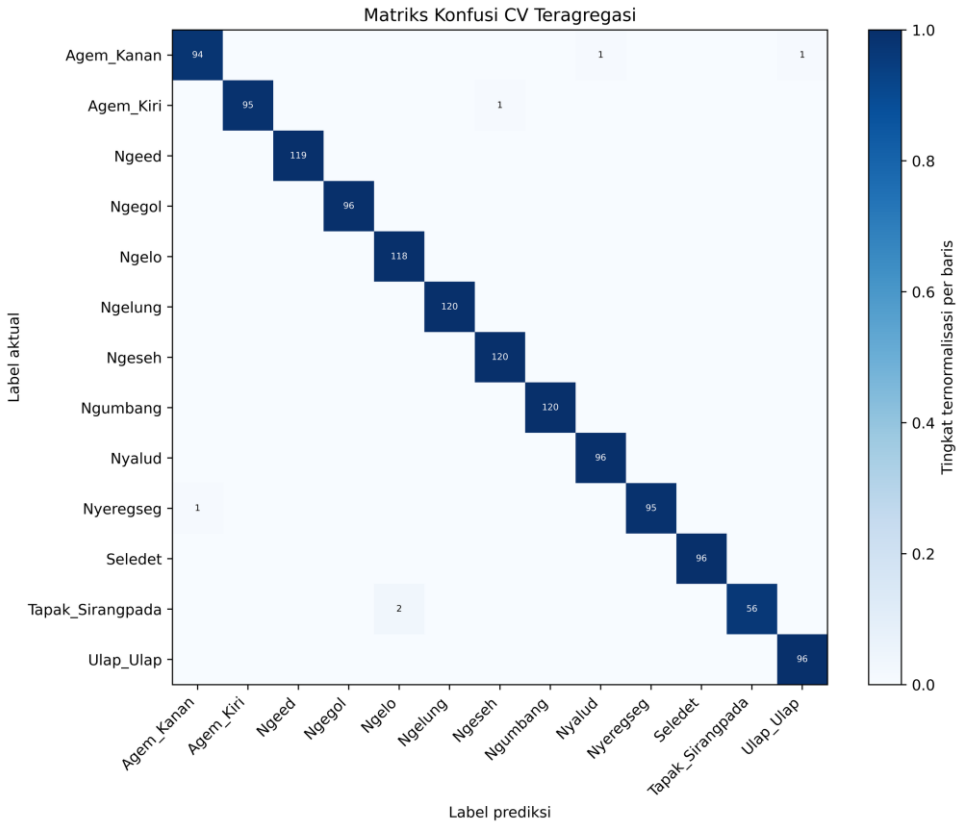
Kedekatan antara hasil validasi silang dan holdout menunjukkan bahwa split test resmi masih berada dalam distribusi yang serupa dengan train_val. Akan tetapi, interpretasi ini tidak otomatis berarti model akan sama kuat pada kamera, penari, atau lingkungan baru. Evaluasi holdout akhir tetap berguna karena test resmi tidak digunakan pada tahap pemilihan model, tetapi kemampuan generalisasi eksternal masih memerlukan dataset tambahan.

Daniel Sande Bona: * Penulis Korespondensi
Copyright © 2026, Daniel Sande Bona.

Journal of Artificial Intelligence and Technology Information
Website : https://ejournal.techcart-press.com/index.php/jaiti
E-ISSN 2985-6396
P-ISSN 2985-5306
Volume 4, Number 2, Juni 2026
https://doi.org/10.58602/jaiti.v4i2.253
Page 92-106

3.4. Analisis Kesalahan

Confusion matrix validasi menunjukkan sedikit kesalahan off-diagonal. Pasangan kesalahan yang paling sering adalah Tapak Sirangpada diprediksi sebagai Ngelo, sebanyak dua kali atau 3,45% dari instance validasi Tapak Sirangpada. Pasangan kesalahan lain masing-masing terjadi satu kali, antara lain Agem Kanan ke Nyalud, Agem Kanan ke Ulap-Ulap, Agem Kiri ke Ngeseh, dan Nyeregseg ke Agem Kanan. Kesalahan ini masih masuk akal karena beberapa gerakan dasar memiliki orientasi torso, posisi lengan, atau frame transisi yang serupa.



Gambar 5. Matriks konfusi agregat validasi silang 5-fold

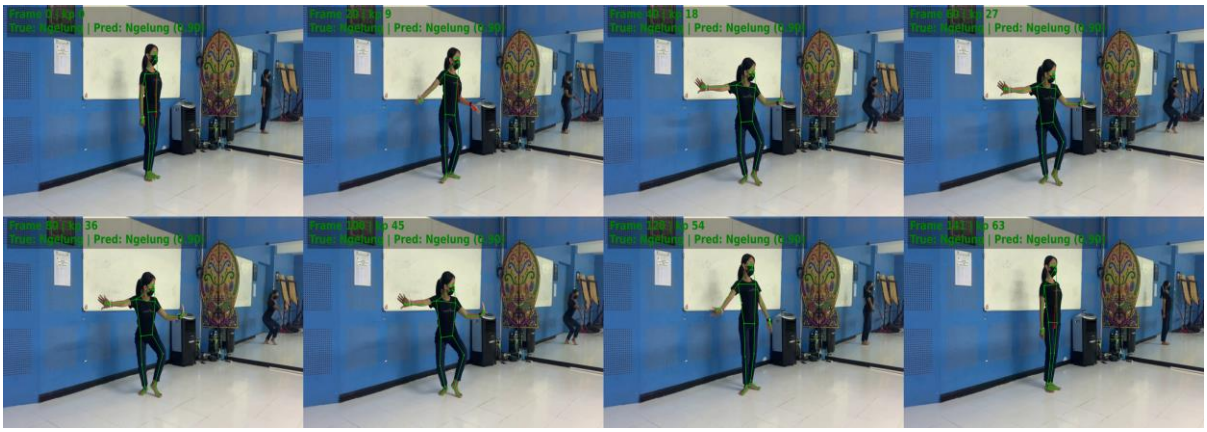
Gambar 5 memperlihatkan matriks konfusi agregat dari validasi silang 5-fold. Visualisasi ini mendukung analisis kesalahan karena menunjukkan kelas mana yang telah dikenali dengan baik dan kelas mana yang masih memiliki kemungkinan tertukar.
Tabel 7 menampilkan contoh sampel *visual-check* dengan *confidence* terendah tetapi tetap benar. Sampel ini penting karena prediksi benar dengan *confidence* rendah dapat menunjukkan bagian dataset yang paling dekat dengan batas keputusan model. Pemeriksaan visual terhadap kasus seperti ini berguna untuk memahami apakah model ragu karena *pose estimator*, kemiripan gerakan, atau variasi framing.

Tabel 7. Sampel visual-check dengan confidence rendah

Confidence	True	Predicted	Video
0,8971	Ngelung	Ngelung	Ngelung_30.mp4
0,9646	Nyeregseg	Nyeregseg	Nyeregseg_26.mp4
0,9727	Tapak Sirangpada	Tapak Sirangpada	Tapak_Sirangpada_122.mp4

Daniel Sande Bona: * Penulis Korespondensi
Copyright © 2026, Daniel Sande Bona.

Journal of Artificial Intelligence and Technology Information
Website : https://ejournal.techcart-press.com/index.php/jaiti
E-ISSN 2985-6396
P-ISSN 2985-5306
Volume 4, Number 2, Juni 2026
https://doi.org/10.58602/jaiti.v4i2.253
Page 92-106



Gambar 6. Urutan frame visual-check sampel Ngelung untuk pemeriksaan konsistensi prediksi

Gambar 6 menampilkan urutan frame visual-check pada salah satu sampel Ngelung. Rangkaian frame ini mendukung pemeriksaan kualitatif dengan memperlihatkan konsistensi pose dan konteks visual dari sampel yang digunakan untuk menilai prediksi model.

3.5. Interpretasi Akurasi Tinggi

Akurasi yang sangat tinggi tetap perlu dibatasi klaimnya. Audit menemukan nol duplikasi identifier dan nol overlap SHA-256 antara train_val dan test, tetapi dataset tidak memiliki metadata identitas penari. Karena itu, hasil ini menunjukkan baseline kuat pada dataset publik yang sama, bukan bukti bahwa model sudah general untuk penari, kamera, kostum, atau lingkungan baru.

3.6. Perbandingan dengan Lantara et al. dan Implementasi GitHub Hendrawan et al.

Baseline terdekat adalah studi Lantara et al. yang melaporkan klasifikasi RGB untuk gerakan tari Bali perempuan [1]. Namun, perbandingan langsung tidak sepenuhnya setara karena jumlah kelas, modalitas input, dan implementasi berbeda. Studi tersebut menggunakan pengaturan RGB 6 kelas dengan 30 frame berukuran 224 x 224 per video. Pada eksperimen kedua, I3D menjadi model terbaik dengan accuracy, precision, recall, dan F1 sebesar 0,98, sedangkan VGG16-LSTM memperoleh accuracy dan F1 sebesar 0,93. Tabel 8 merangkum perbandingan tersebut.

Tabel 8. Perbandingan dengan paper Lantara dan implementasi GitHub Hendrawan

Model/sumber	Input dan kelas	Parameter	Performa
Lantara et al. VGG16-LSTM [1]	RGB, 6 kelas	~15,77M total; ~1,06M trainable	Test accuracy/F1 = 0,93
Lantara et al. I3D [1]	RGB, 6 kelas	~13,35M total; ~1,06M trainable	Test accuracy/F1 = 0,98
Hendrawan GitHub VGG16-LSTM [20]	RGB, 7 kelas	15,77M total; 1,06M trainable	Best observed accuracy = 0,78

Daniel Sande Bona: * Penulis Korespondensi
Copyright © 2026, Daniel Sande Bona.

Journal of Artificial Intelligence and Technology Information
Website : https://ejournal.techcart-press.com/index.php/jaiti
E-ISSN 2985-6396
P-ISSN 2985-5306
Volume 4, Number 2, Juni 2026
https://doi.org/10.58602/jaiti.v4i2.253
Page 92-106

Hendrawan GitHub I3D [20]	RGB, 7 atau 11 kelas	12,82M total; 0,53M trainable	Best observed 7-class accuracy = 0,803; 11-class accuracy = 0,438

Perbandingan ini menunjukkan bahwa skeleton-based ST-GCN merupakan baseline yang ringkas dan kompetitif untuk dataset tari Bali perempuan. Nilai utama penelitian ini adalah penyajian pipeline yang dapat direproduksi, pemisahan holdout test, dan audit kebocoran data. Untuk klaim generalisasi yang lebih kuat, dataset perlu dilengkapi metadata identitas penari agar evaluasi subject-independent dapat dilakukan.

4. KESIMPULAN

Penelitian ini melaporkan baseline ST-GCN untuk klasifikasi 13 gerakan dasar tari Bali perempuan menggunakan skeleton MediaPipe Pose. Dengan validasi silang 5-fold berstratifikasi pada subset train_val, model memperoleh top-1 accuracy 99,55% +/- 0,49%, top-5 accuracy 99,92% +/- 0,17%, dan macro F1 99,49% +/- 0,55%. Pada holdout test resmi, model mencapai top-1 accuracy 99,39%, top-5 accuracy 100,00%, dan macro F1 99,39%. Audit duplikasi identifier dan overlap SHA-256 tidak menemukan kebocoran train-test.

Hasil tersebut menunjukkan bahwa representasi skeleton berbasis MediaPipe dan ST-GCN mampu mengenali pola gerakan dasar tari Bali dengan sangat baik pada kondisi dataset yang sama. Kelebihan utama pendekatan ini adalah input yang lebih ringkas, lebih mudah diinterpretasikan, dan lebih sedikit bergantung pada latar belakang visual dibandingkan video RGB. Kekurangannya adalah performa masih bergantung pada kualitas pose estimator dan belum dapat mengklaim generalisasi subject-independent karena dataset tidak menyediakan label identitas penari.

Saran untuk penelitian selanjutnya adalah menambahkan atau menganotasi identitas penari agar leave-one-subject-out cross-validation dapat dilakukan. Eksperimen berikutnya juga perlu membandingkan ST-GCN dengan 2s-AGCN, MS-G3D, PoseC3D, dan model pose/action recognition yang lebih baru. Selain itu, perlu dilakukan pengujian pada variasi kamera, kostum, pencahayaan, dan lingkungan perekaman yang berbeda agar sistem tidak hanya kuat pada dataset publik, tetapi juga bermanfaat untuk dokumentasi budaya dan pembelajaran tari dalam kondisi nyata.

5. REFERENCES

[1] I. P. P. B. Lantara, I. P. D. Payana, G. A. S. Pratama, W. E. A. Munayana, K. S. Nopiani, and I. N. R. Hendrawan, "GETARI: Dataset untuk klasifikasi gerakan dasar Tari Bali perempuan," Jurnal Nasional Pendidikan Teknik Informatika: JANAPATI, vol. 11, no. 3, pp. 290-300, 2022, doi: 10.23887/janapati.v11i3.52598.

[2] I. N. R. Hendrawan, P. Setyarini, A. T. Tedja, I. P. P. B. Lantara, I. P. D. Payana, G. A. S. Pratama, W. E. A. Munayana, and K. S. Nopiani, "Video Dataset of Woman Basic Balinese Dance Movement for Action Recognition," Mendeley Data, version 2, 2023, doi: 10.17632/s2gv9d6gpb.2.

[3] N. W. Utami, M. A. P. Putra, I. G. J. E. Putra, and G. A. Sampedro, "A CNN-based information system for Balinese dance classification with hyperparameter optimization," International Journal of Advances in Data and Information Systems, vol. 6, no. 2, pp. 376-390, 2025, doi: 10.59395/ijadis.v6i2.1402.

[4] C. Irawan, H. P. Hadi, C. Jatmoko, and M. Doheir, "Video classification of Indonesian traditional dance using a hybrid CNN-LSTM model with pose estimation," Bulletin of

Daniel Sande Bona: * Penulis Korespondensi
Copyright © 2026, Daniel Sande Bona.

Electrical Engineering and Informatics, vol. 15, no. 1, pp. 787-798, 2026, doi: 10.11591/eei.v15i1.11093.
[5] S. Yan, Y. Xiong, and D. Lin, "Spatial temporal graph convolutional networks for skeleton-based action recognition," in Proc. AAAI Conference on Artificial Intelligence, vol. 32, no. 1, 2018, doi: 10.1609/aaai.v32i1.12328.
[6] H. Duan, Y. Zhao, K. Chen, D. Lin, and B. Dai, "Revisiting skeleton-based action recognition," in Proc. IEEE/CVF Conf. Computer Vision and Pattern Recognition (CVPR), 2022, pp. 2969-2978.
[7] K. Guo, P. Wang, P. Shi, C. He, and C. Wei, "A new partitioned spatial-temporal graph attention convolution network for human motion recognition," Applied Sciences, vol. 13, no. 3, Art. no. 1647, 2023, doi: 10.3390/app13031647.
[8] M. H. Al-Hakimi, I. Ahmed, M. Haseeb, T. H. Rassem, F. H. Quradaa, and R. S. Almoqbily, "An enhanced spatial-temporal graph convolution network with high order features for skeleton-based action recognition," PLOS ONE, vol. 20, no. 10, Art. no. e0332815, 2025, doi: 10.1371/journal.pone.0332815.
[9] H. Hu, Y. Cao, Y. Fang, and Z. Meng, "Semantics-assisted training graph convolution network for skeleton-based action recognition," Sensors, vol. 25, no. 6, Art. no. 1841, 2025, doi: 10.3390/s25061841.
[10] P. Jin, S. Guo, and C. Li, "Center-of-gravity-aware graph convolution for unsafe behavior recognition of construction workers," Sensors, vol. 25, no. 17, Art. no. 5493, 2025, doi: 10.3390/s25175493.
[11] C. Lu, "Feature data analysis of dance movements by motion capture," Journal of Measurements in Engineering, vol. 13, no. 3, pp. 701-713, 2025, doi: 10.21595/jme.2025.24742.
[12] X. Yan, J. Yang, and T. Salami, "Classification of Indian classical dances using MnasNet architecture with advanced polar fox optimization for hyperparameter optimization," Scientific Reports, vol. 15, Art. no. 3054, 2025.
[13] M. Clayton, J. Li, A. Clarke, and M. Weinzierl, "Hindustani raga and singer classification using 2D and 3D pose estimation from video recordings," Journal of New Music Research, vol. 52, no. 4, pp. 285-300, 2024, doi: 10.1080/09298215.2024.2331788.
[14] P. Malavath and N. Devarakonda, "Natya Shastra: Deep learning for automatic classification of hand mudra in Indian classical dance videos," Revue d Intelligence Artificielle, vol. 37, no. 3, pp. 251-260, 2023, doi: 10.18280/ria.370317.
[15] T. Yoddamnern and P. Riyamongkol, "Deep learning for two-person Mae Mai Muay Thai classification: Leveraging normalized human pose sequences and CNN-LSTM," Scientific Culture, vol. 11, no. 4, pp. 333-349, 2025, doi: 10.5281/zenodo.11042528.
[16] M. Muslih and E. H. Rachmawanto, "Convolutional Neural Network (CNN) untuk klasifikasi citra penyakit diabetes retinopathy," SKANIKA: Sistem Komputer dan Teknik Informatika, vol. 5, no. 2, pp. 167-176, 2022, doi: 10.36080/skanika.v5i2.2945.
[17] A. S. Kusuma, A. I. Pradana, and B. W. Pamekas, "Pengembangan sistem perhitungan jumlah kendaraan berdasarkan jenis kendaraan menggunakan algoritma YOLO secara realtime," SKANIKA: Sistem Komputer dan Teknik Informatika, vol. 7, no. 2, pp. 166-179, 2024.
[18] C. Lugaresi et al., "MediaPipe: A framework for building perception pipelines," arXiv:1906.08172, 2019.
[19] MMAction2 Contributors, "OpenMMLab action understanding toolbox and benchmark," 2020. [Online]. Available: https://github.com/open-mmlab/mmaction2

Daniel Sande Bona: * Penulis Korespondensi
Copyright © 2026, Daniel Sande Bona.

Journal of Artificial Intelligence and Technology Information
Website : https://ejournal.techcart-press.com/index.php/jaiti
E-ISSN 2985-6396
P-ISSN 2985-5306
Volume 4, Number 2, Juni 2026
https://doi.org/10.58602/jaiti.v4i2.253
Page 92-106

[20] R. Hendrawan, D. Payana, D. Payana, and A. Tedja, "rudyhendrawn/traditional-dance- videoclassification: v1.0.0," Zenodo, Mar. 13, 2023, doi: 10.5281/zenodo.7726950.

Daniel Sande Bona: * Penulis Korespondensi
Copyright © 2026, Daniel Sande Bona.