

Klasifikasi Audio Gamelan Bali Berdasarkan Periode Waktu Menggunakan Metode Convolutional Neural Network

I Kadek Dwi Gandika Supartha^{1*}, Adi Panca Saputra Iskandar², Ni Kadek Sri Nopiani³,
I Made Agus Wirya Pratama⁴

^{1*, 2, 3, 4} Institut Bisnis dan Teknologi Indonesia, Denpasar, Indonesia

^{1*} gandika.supartha@instiki.ac.id, ² adi.panca@instiki.ac.id, ³ kd.sri.nopiani@gmail.com, ⁴

aguswiryapratama@gmail.com

INFO ARTIKEL

Article history:

Received Juni 2026

Published Juli 2026

ABSTRAK

Gamelan Bali merupakan warisan budaya Indonesia yang diklasifikasikan berdasarkan periode waktu menjadi gamelan wayah, madya, dan anyar. Penelitian ini bertujuan mengembangkan sistem klasifikasi otomatis menggunakan metode Layer convolutional Neural Network (CNN) dengan ekstraksi fitur Mel Frequency Cepstral Coefficients (MFCC). Dataset terdiri dari 600 audio berdurasi 30 detik dari 200 lagu Gamelan Bali, dibagi ke dalam tiga kelas periode. Proses pra-pemrosesan meliputi normalisasi, pemotongan durasi, dan ekstraksi fitur MFCC beserta turunannya, yaitu delta dan delta-delta. Model CNN diuji dengan variasi arsitektur 2, 3 dan 4 lapisan konvolusi, epoch 100 dan 200, serta patience 10 dan 15, menggunakan skema 10-fold Cross Validation. Hasil menunjukkan bahwa arsitektur CNN 2 lapisan memberikan performa terbaik dengan akurasi rata-rata 77,67% pada konfigurasi patience 15 dan epoch 200. Temuan ini membuktikan bahwa model CNN sederhana lebih efektif untuk dataset audio terbatas, sekaligus membuka peluang penerapan kecerdasan buatan dalam pelestarian budaya tradisional.

Kata Kunci: Gamelan Bali, MFCC, Convolutional Neural Network, K-Fold Cross Validation.

ABSTRAK

Balinese Gamelan is an Indonesian cultural heritage classified by time period into gamelan wayah, madya, and anyar. This study aims to develop an automatic classification system using the Layer convolutional Neural Network (CNN) method with Mel-Frequency Cepstral Coefficients (MFCC) feature extraction. The dataset consists of 600 audio clips, each 30 seconds long, derived from 200 Balinese Gamelan songs, divided into three period-based classes. Preprocessing includes normalization, duration trimming, and MFCC feature extraction along with its derivatives delta and delta-delta. The CNN model was evaluated using variations of 2, 3, and 4 layer convolutional layers, epochs of 100 and 200, and patience values of 10 and 15, under a 10-fold cross-validation scheme. The results show that the 2 layer CNN architecture achieved the best performance, with an average accuracy of 77.67% under the configuration of patience 15 and 200 epochs. These findings demonstrate that a simpler CNN model is more effective for limited audio datasets and open opportunities for applying artificial intelligence in preserving traditional cultural heritage.

Keywords: Balinese Gamelan, MFCC, Layer convolutional Neural Network, K-Fold Cross Validation.

1. Pendahuluan

Gamelan Bali merupakan seni musik tradisional khas Provinsi Bali yang didominasi oleh bunyi perkusi seperti metalofon, gong, dan simbal. Kesenian ini memiliki peran penting dalam kehidupan masyarakat Bali, khususnya dalam konteks adat dan keagamaan Hindu, serta berfungsi sebagai media komunikasi, hiburan, ekspresi estetika, dan pariwisata (Suharta, 2023). Seiring perkembangannya, Gamelan Bali memiliki beragam jenis yang mencerminkan perjalanan sejarah dan perubahan sosial budaya masyarakat Bali (Bandem, 2013).

Keanekaragaman jenis Gamelan Bali menjadi kekayaan budaya yang menarik perhatian masyarakat dan wisatawan. Namun, membedakan jenis gamelan bukanlah hal yang mudah karena membutuhkan pemahaman mengenai karakteristik, fungsi, dan sejarah masing-masing gamelan (Riski dkk., 2024). Di Bali sendiri telah teridentifikasi sekitar 30 jenis perangkat gamelan yang diklasifikasikan ke dalam beberapa kategori, salah satunya berdasarkan periode waktu kemunculannya, yaitu gamelan wayah, madya, dan anyar (Pradnyani, 2023). Kompleksitas tersebut menunjukkan perlunya pendekatan berbasis teknologi untuk membantu proses identifikasi dan klasifikasi gamelan secara lebih efisien, sekaligus mendukung pembelajaran dan pelestarian budaya.

Penelitian ini mengklasifikasikan otomatis berdasarkan bertujuan untuk Gamelan Bali periode secara waktu kemunculannya agar mudah dipahami oleh masyarakat umum (Astita, 2019). Tahapan penelitian meliputi pengumpulan dataset audio, pengolahan sinyal, ekstraksi fitur, pelatihan model, dan evaluasi performa. Beberapa metode ekstraksi fitur audio telah banyak digunakan, seperti MFCC, Spectral Centroid, Zero-Crossing Rate (ZCR), dan Chroma Features. MFCC dipilih dalam penelitian ini karena kemampuannya merepresentasikan karakteristik spektral audio yang menyerupai sistem pendengaran manusia (Jitendra, dkk., 2020). Penelitian (Toleu dkk., 2024) menunjukkan bahwa MFCC merupakan fitur paling efektif dalam mendeteksi perubahan pembicara dengan nilai MAX P(Speaker Change | Feature Change) sebesar 0,534 (Toleu dkk., 2024).

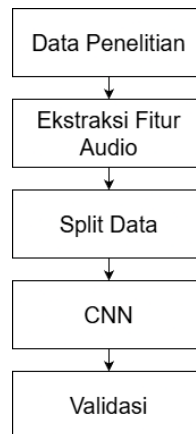
Dalam proses klasifikasi, berbagai model deep learning dapat digunakan, seperti Convolutional Neural Network (CNN), Recurrent Neural Network (RNN), dan Long Short-Term Memory (LSTM). CNN dipilih karena kemampuannya mengenali pola secara efisien dari representasi visual audio dua dimensi serta performanya yang baik dalam klasifikasi musik dan suara (Prince dkk., 2022). Penelitian Massoudi et al. (2021) membuktikan bahwa CNN mampu mencapai akurasi hingga 91% dalam klasifikasi suara lingkungan. Oleh karena itu, penelitian ini mengombinasikan MFCC sebagai metode ekstraksi fitur dan CNN sebagai metode klasifikasi untuk mengelompokkan Gamelan Bali ke dalam kategori gamelan wayah, madya, dan anyar. Kombinasi MFCC dan CNN diharapkan mampu menghasilkan sistem klasifikasi yang akurat dan efektif (Zaman dkk., 2023).

2. Metode Penelitian

Bagian ini memuat penjelasan tentang tahap tahap penelitian yang menggambarkan urutan logis untuk mendapatkan output penelitian sesuai dengan harapan.

2.1 Gambaran Umum Sistem

Sistem klasifikasi yang dikembangkan pada penelitian ini terdiri dari beberapa tahapan utama, yaitu pengumpulan dataset audio gamelan, praproses dan ekstraksi fitur, pelatihan model CNN, serta evaluasi performa model. Data audio gamelan dari tiga periode waktu diproses untuk menghasilkan fitur MFCC yang selanjutnya digunakan sebagai input bagi model CNN. Proses pelatihan dan pengujian dilakukan menggunakan teknik K-Fold Cross Validation untuk memperoleh hasil evaluasi yang lebih stabil. Alur kerja sistem secara umum ditunjukkan pada Gambar 1.



Gambar 1. Gambaran Umum Sistem
Sumber: dokumentasi penulis

2.2 Data Penelitian

Dataset yang digunakan dalam penelitian ini diperoleh dari berbagai platform digital seperti YouTube, Spotify, dan Joox. Total dataset terdiri dari 200 lagu gamelan Bali yang terbagi ke dalam tiga kategori, yaitu gamelan wayah sebanyak 65 lagu, gamelan madya sebanyak 69 lagu, dan gamelan anyar sebanyak 66 lagu. Setiap rekaman audio kemudian disegmentasi menjadi tiga bagian berdurasi 30 detik sehingga menghasilkan total 600 data audio. Seluruh data diseragamkan dalam format file .wav untuk memastikan konsistensi pada tahap pengolahan selanjutnya. Table 1. Menunjukkan jumlah audio dari masing-masing periode waktu setelah mengalami proses pemotongan.

Tabel 1. Hasil Jumlah Pemotongan Dataset.

No	Periode Waktu	Jumlah Audio
1	Gamelan Wayah	195
2	Gamelan Madya	207
3	Gamelan Anyar	198

2.3 Pemrosesan Data Dan Ekstraksi Fitur

Tahap praproses dan ekstraksi fitur dilakukan untuk memperoleh representasi numerik dari sinyal audio yang dapat diproses oleh model CNN. Fitur yang digunakan pada penelitian ini adalah Mel-Frequency Cepstral Coefficients (MFCC) beserta fitur turunannya, yaitu delta dan delta-delta. MFCC dipilih karena kemampuannya dalam menangkap karakteristik spektral sinyal audio secara efektif. Proses ekstraksi fitur meliputi pemuatan sinyal audio, penghilangan bagian hening, normalisasi amplitudo, serta penyesuaian durasi audio menjadi 30 detik. Selanjutnya, koefisien MFCC, delta, dan delta-delta digabungkan menjadi satu representasi fitur berdimensi $39 \times \text{time steps}$ yang digunakan sebagai input model.

2.4 Split Data

Pada penelitian ini digunakan metode K-Fold Cross Validation dengan jumlah fold sebanyak 10. Teknik ini membagi dataset menjadi sepuluh bagian yang relatif seimbang, di mana pada setiap iterasi satu fold digunakan sebagai data uji dan sembilan fold lainnya sebagai data latih. Proses ini diulang hingga setiap fold memperoleh kesempatan yang sama sebagai data uji. Penggunaan K-Fold Cross Validation bertujuan untuk meningkatkan reliabilitas evaluasi model serta mengurangi bias akibat pembagian data secara acak.

2.5 Convolutional Neural Network

Metode klasifikasi yang digunakan dalam penelitian ini adalah Convolutional Neural Network (CNN). CNN dipilih karena kemampuannya dalam mengekstraksi pola spasial dan temporal dari data dua dimensi, seperti representasi MFCC. Arsitektur CNN yang digunakan terdiri dari lapisan input, convolutional layer, pooling layer, dan fully connected layer. Pada penelitian ini dilakukan pengujian

terhadap tiga skenario arsitektur CNN dengan jumlah lapisan konvolusi yang berbeda, yaitu dua, tiga, dan empat layer, untuk menganalisis pengaruh kedalaman jaringan terhadap performa klasifikasi.

2.6 Early Stopping

Untuk mencegah terjadinya overfitting selama proses pelatihan, penelitian ini menerapkan teknik Early Stopping. Mekanisme ini memantau nilai validation loss dan menghentikan proses pelatihan apabila tidak terjadi peningkatan performa dalam sejumlah epoch tertentu. Dengan demikian, Early Stopping membantu menjaga keseimbangan antara akurasi pelatihan dan kemampuan generalisasi model terhadap data uji (Saptadi, dkk., 2025).

3.1 Skenario Pengujian

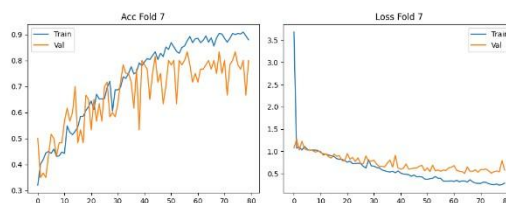
Pengujian dilakukan menggunakan tiga skenario arsitektur CNN yang berbeda berdasarkan jumlah lapisan konvolusi, yaitu dua, tiga, dan empat layer. Seluruh skenario menggunakan parameter pelatihan yang sama, meliputi optimizer Stochastic Gradient Descent (SGD) dengan learning rate 0,001, jumlah epoch 100, dan batch size 128. Penggunaan SGD dipilih karena kemampuannya dalam menjaga stabilitas konvergensi dan menghasilkan performa yang konsisten pada klasifikasi audio berbasis CNN.

3.2 Performa CNN Dua Layer

Model CNN dengan dua lapisan konvolusi menghasilkan performa terbaik dengan akurasi rata-rata sebesar 77,66%, presisi 77.53%, recall 76.57%, dan F1-score 76.77%. Performa tertinggi ini diperoleh pada konfigurasi pelatihan dengan jumlah epoch sebanyak 200 dan patience sebesar 15 pada mekanisme early stopping, yang memungkinkan model mencapai konvergensi optimal tanpa mengalami overfitting. Hasil evaluasi akurasi pada setiap fold dapat dilihat pada Tabel 2. Grafik akurasi dan loss pada fold dengan performa terbaik ditunjukkan pada Gambar 2, yang memperlihatkan proses pelatihan yang stabil. Sementara itu, kemampuan model dalam mengklasifikasikan ketiga kelas Gamelan Bali secara detail ditunjukkan melalui confusion matrix pada Gambar 3.

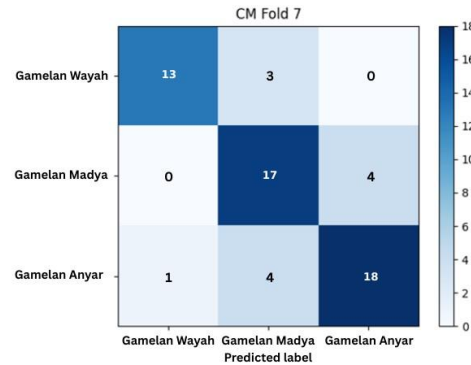
Tabel 2. Hasil Akurasi Per-Fold CNN 2 Layer dengan patience 15 dan epoch 200 [Nopiani,2026]

No	Akurasi	Presisi	Recall	F1-Score
1	80.00	61.15	77.63	76.02
2	83.33	63.62	83.62	82.92
3	71.66	65.64	71.44	71.90
4	73.33	65.18	74.22	71.73
5	80.00	46.73	80.11	79.37
6	69.99	55.53	67.56	65.18
7	80.00	50.24	81.83	80.15
8	85.00	57.86	85.00	84.00
9	69.99	61.61	70.03	71.09
10	83.33	49.83	83.90	83.29



Gambar 2. Grafik Akurasi dan Loss pada fold ke-7 CNN 2 Layer dengan patience 15 dan epoch 200

Sumber: dokumentasi penulis



Gambar 3. Confusion Matrix pada fold ke-7 CNN 2 Layer dengan patience 15 dan epoch 200

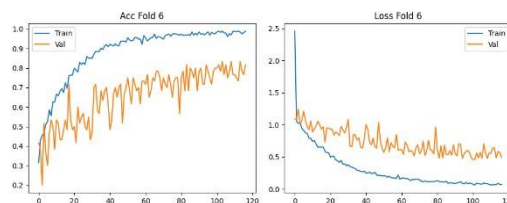
Sumber: dokumentasi penulis

3.3 Performa CNN Tiga Layer

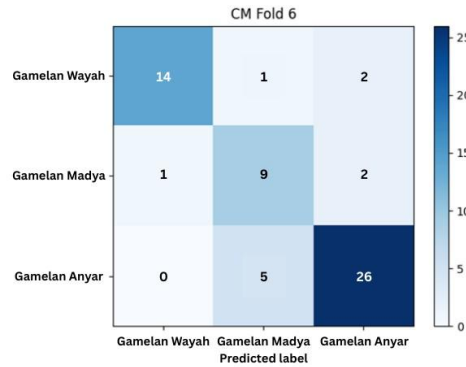
Pada skenario CNN dengan tiga lapisan konvolusi, diperoleh akurasi rata-rata sebesar 75,83%, presisi 76,31%, recall 75,15%, dan F1score 75,04%. Performa tertinggi ini diperoleh pada konfigurasi pelatihan dengan jumlah epoch sebanyak 200 dan patience sebesar 15. Peningkatan performa ini menunjukkan bahwa penambahan satu lapisan konvolusi mampu meningkatkan kemampuan model dalam mengekstraksi fitur akustik gamelan secara lebih representatif, tanpa menyebabkan degradasi performa yang signifikan. Nilai akurasi pada setiap fold pengujian ditampilkan pada Tabel 3, sedangkan grafik akurasi dan loss selama proses pelatihan dapat dilihat pada Gambar 4. Selain itu, distribusi hasil klasifikasi pada masing-masing kelas gamelan divisualisasikan melalui confusion matrix yang ditunjukkan pada Gambar 5.

Tabel 3. Hasil Akurasi Per-Fold CNN 3 Layer dengan patience 15 dan epoch 200 [Nopiani,2026]

No	Akurasi	Presisi	Recall	F1-Score
1	76.66	74.58	75.82	74.69
2	76.66	79.49	77.90	78.18
3	76.66	73.53	73.52	73.49
4	75.00	78.64	72.17	74.40
5	75.00	74.52	74.52	74.40
6	81.66	80.00	80.40	79.80
7	80.00	82.03	81.06	80.55
8	75.00	72.98	72.85	72.77
9	64.99	66.47	64.73	64.68
10	76.66	80.86	78.52	77.46



Gambar 4. Grafik Akurasi dan Loss pada fold ke-6 CNN 3 Layer dengan patience 15 dan epoch 200
Sumber: Nopiani, 2026



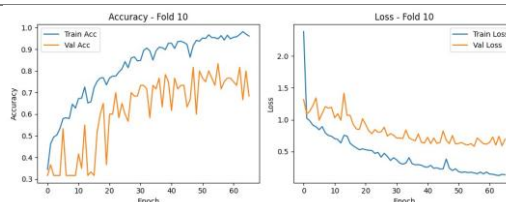
Gambar 5. Confusion Matrix pada fold ke-6 CNN 3 Layer dengan patience 15 dan epoch 200
Sumber: dokumentasi penulis

3.4 Performa CNN Empat Layer

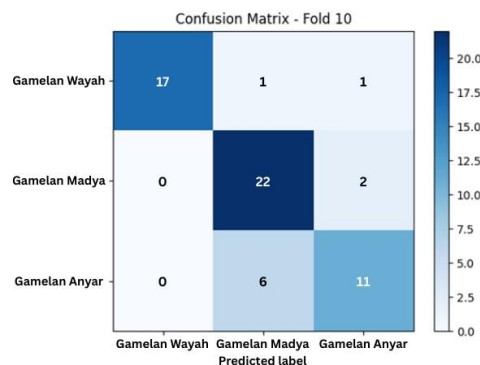
Model CNN dengan empat lapisan konvolusi menghasilkan akurasi rata-rata sebesar 71,83%, dengan nilai precision 74,26%, recall 71,04%, dan F1-score 70,94%. Meskipun pada beberapa fold diperoleh akurasi yang tinggi, secara keseluruhan performa model cenderung menurun. Hal ini menunjukkan bahwa penambahan jumlah lapisan konvolusi yang berlebihan dapat mengurangi kemampuan generalisasi model, terutama ketika jumlah dataset terbatas. Rincian akurasi pada setiap fold untuk skenario CNN empat layer dapat dilihat pada Tabel 4. Visualisasi grafik akurasi dan loss selama proses pelatihan ditampilkan pada Gambar 6, sedangkan confusion matrix hasil klasifikasi model pada fold terbaik ditunjukkan pada Gambar 7, yang menggambarkan pola kesalahan klasifikasi antar kelas gamelan Bali.

Tabel 4. Hasil Akurasi Per-Fold CNN 4 Layer dengan patience 10 dan epoch 100 [Nopiani,2026]

No	Akurasi	Presisi	Recall	F1-Score
1	76.66	58.15	74.61	75.58
2	75.00	63.62	74.83	75.54
3	80.00	57.56	77.72	76.41
4	66.66	68.89	67.43	67.85
5	64.99	67.22	66.22	64.42
6	61.66	89.08	64.26	64.16
7	78.33	50.29	79.45	78.92
8	78.33	57.29	79.91	75.67
9	68.33	61.25	70.55	69.70
10	83.33	58.12	84.81	81.94



Gambar 6. Grafik Akurasi dan Loss pada fold ke-10 CNN 4 Layer dengan patience 10 dan epoch 100
Sumber: dokumentasi penulis



Gambar 7. Confusion Matrix pada fold ke-10 CNN 4 Layer dengan patience 10 dan epoch 100
Sumber: Nopiani, 2026

3. Kesimpulan

Performa model CNN dalam mengklasifikasikan Gamelan Bali menunjukkan hasil yang cukup baik, terutama ketika menggunakan pendekatan 10-fold Cross Validation. Hasil pengujian menunjukkan bahwa arsitektur CNN 2 layer memberikan performa terbaik secara keseluruhan, dengan rata-rata akurasi tertinggi mencapai 77,67% pada konfigurasi patience 15 dan epoch 200. Meskipun CNN 4 layer mencatat akurasi maksimal per-fold sebesar 83,33%, rata-rata akurasinya lebih rendah, yaitu 71,83% dan menunjukkan kecenderungan overfitting pada beberapa skenario. Hal ini mengindikasikan bahwa kompleksitas arsitektur yang lebih dalam tidak selalu meningkatkan generalisasi, terutama pada dataset audio berukuran terbatas seperti dalam penelitian ini, yaitu dengan 600 sampel.

Berdasarkan hasil penelitian yang telah dilakukan, terdapat beberapa saran yang dapat dijadikan acuan untuk pengembangan penelitian selanjutnya. Pertama, penelitian selanjutnya dapat menggunakan dataset yang lebih banyak dan lebih beragam, dari segi variasi, tempo ataupun kualitas rekaman. Kedua, pengembangan metode ekstraksi fitur dapat dilakukan dengan mengkombinasikan MFCCC dengan fitur audio lain seperti Chroma, dan Spectral Contrast, untuk menangkap informasi musikal yang lebih kaya.

5. Ucapan Terima Kasih

Terimakasih kepada bapak Rektok INSTIKI, bagian DRPM INSTIKI serta seluruh rekan dosen dan staff,

Daftar Pustaka

- Astita 2019. "Tradisi Gamelan Anyar di Banjar Tengah, Sesetan, Kota Denpasar".
- Mukkamala S N V Jitendra, D. Y. R. 2020. "A Review: Music Feature Extraction from an Audio Signal". **E3S Web of Conferences**, 4(2), 15–21.
- Norbertus Tri Suswanto Saptadi, Agung Yuliyanto Nugroho, Hedi Kristiawan, N. R. 2025. "Pengantar Deep Learning Teori, Algoritma, dan Aplikasi dalam Kecerdasan Buatan Penerbit".
- Pradnyani, K. D. 2023. "Sistem Klasifikasi Gamelan Bali Berbasis Web a Website-Based Balinese Classification System of the Balinese gamelan", 16(2), 82–90.
- Prince, S., Thomas, J. J., dkk. 2022. "Music Genre Classification using Deep learning - A review". **6th IEEE International Conference on Computational System and Information Technology for Sustainable Solutions, CSITSS 2022**.
<https://doi.org/10.1109/CSITSS57437.2022.10026394>.
- Riski, I. K., Putra, A., dkk. 2024. "Klasifikasi Gamelan Bali Menggunakan Support Vector Machine (SVM)", 13(2), 447–458.
- Suharta, I. W. 2023. **BALAGANJUR MUSIK PROSESI BALI**.
- Supartha, I. K. D. G., Yuniarti, A., dkk. 2025. "Enhancing Balinese Gamelan Classification using a CNN with Multi Feature Fusion". **2025 15th International Conference on Information**

- & Communication Technology and System (ICTS)**, (DI), 1–6.
<https://doi.org/10.1109/ICTS67612.2025.11369498>.
- Toleu, A., Tolegen, G., dkk. 2024. "Comparative Analysis of Distance Measures for Unsupervised Speaker Change Detection". **Proceedings - 2024 20th International Asian School-Seminar on Optimization Problems of Complex Systems, OPCS 2024**, 28–32.
<https://doi.org/10.1109/OPCS63516.2024.10720386>.
- Zaman, K., Sah, M., dkk. 2023. "A Survey of Audio Classification Using Deep Learning". **IEEE Access**, 11, 106620–106649. <https://doi.org/10.1109/ACCESS.2023.3318015>.