

PENGEMBANGAN MODEL CONVLSTM UNTUK KLASIFIKASI AUDIO GAMELAN BALI

I Gede Risva Darma Sentana^{1*}, Ketut Agus Seputra², I Gede Mahendra Darmawiguna³, Putu Zasya Eka Satya Nugraha⁴, Helmi Maulana Hadiwinata⁵

^{1,2,3} Fakultas Teknik dan Kejuruan, Universitas Pendidikan Ganesha; Jalan Udayana; 81116.

^{4,5} PT. Dago Engineering; Bukit Dago Selatan, Bandung;

Keywords:

ConvLSTM, Klasifikasi Audio, Gamelan Bali, *Deep Learning*.

Correspondent Email:

risva@undiksha.ac.id



Copyright © [JITET](http://www.jitet.org) (Jurnal Informatika dan Teknik Elektro Terapan). This article is an open access article distributed under terms and conditions of the Creative Commons Attribution (CC BY NC)

Abstrak. Musik tradisional seperti Gamelan Bali menghadapi tantangan pelestarian di era digital. Proses identifikasi manual yang lambat dan subjektif menghambat upaya pengarsipan skala besar. Penelitian ini mengusulkan solusi melalui pengembangan model klasifikasi audio otomatis menggunakan arsitektur *Convolutional Long Short-Term Memory*. Tujuan penelitian adalah untuk merancang, melatih, dan mengevaluasi model yang mampu mengklasifikasikan lima Gamelan Bali: Angklung, Baleganjur, Gong Gede, Gong Kebyar, dan Semar Pegulingan. Metode penelitian ini mencakup beberapa tahapan, mulai dari pengumpulan dan validasi dataset oleh ahli, pra-pemrosesan audio dengan segmentasi durasi tetap dan *denoising*, hingga ekstraksi fitur gabungan (MFCC, Chroma, Spectral Contrast). Beberapa skenario pelatihan dieksperimentasikan untuk menemukan konfigurasi optimal. Hasil penelitian menunjukkan bahwa model terbaik, yang dilatih menggunakan audio 10 detik dengan teknik transformasi data, berhasil mencapai akurasi validasi sebesar 75%. Analisis lebih lanjut mengungkap bahwa model menunjukkan performa baik pada kelas dengan karakteristik unik seperti Baleganjur (F1-Score 0.91), namun menghadapi tantangan pada kelas dengan tumpang tindih akustik seperti Gong Kebyar (F1-Score 0.50). Disimpulkan bahwa arsitektur ConvLSTM cukup dalam klasifikasi audio Gamelan Bali namun tetap perlu ditingkatkan.

Abstract. Traditional music such as Balinese Gamelan faces preservation challenges in the digital age. Slow and subjective manual identification processes hinder large-scale archiving efforts. This study proposes a solution through the development of an automatic audio classification model using a Convolutional Long Short-Term Memory architecture. The objective of this study is to design, train, and evaluate a model capable of classifying five types of Balinese Gamelan: Angklung, Baleganjur, Gong Gede, Gong Kebyar, and Semar Pegulingan. The research method includes several stages, starting from data collection and validation by experts, audio pre-processing with fixed duration segmentation and denoising, to combined feature extraction (MFCC, Chroma, Spectral Contrast). Several training scenarios were experimented with to find the optimal configuration. The results show that the best model, trained using 10-second audio clips with data transformation techniques, achieved a validation accuracy of 75%. Further analysis revealed that the model performed well on classes with unique characteristics such as Baleganjur (F1-Score 0.91), but faced challenges on classes with acoustic overlap such as Gong Kebyar (F1-Score 0.50). It was concluded that the ConvLSTM architecture is sufficient for Balinese Gamelan audio classification but still needs improvement.

1. PENDAHULUAN

Kemajuan teknologi kecerdasan buatan (AI) telah membuka paradigma baru dalam berbagai disiplin ilmu, termasuk dalam upaya pelestarian warisan budaya. Pemanfaatan AI untuk melindungi, menganalisis, dan menyebarluaskan budaya tradisional kini menjadi arah penting dalam konservasi warisan budaya takbenda (*intangible cultural heritage*) [1]. Musik tradisional, sebagai salah satu bentuk warisan budaya yang kaya akan nilai historis dan estetis, menghadapi tantangan pelestarian di era digital. Proses identifikasi dan klasifikasi musik secara manual oleh para ahli tidak hanya membutuhkan waktu dan biaya yang besar, tetapi juga bersifat subjektif dan sulit untuk diaplikasikan pada arsip berskala besar. Oleh karena itu, penerapan metode *deep learning* untuk otomatisasi tugas-tugas ini menawarkan solusi yang efisien, objektif, dan terukur, memungkinkan analisis yang lebih dalam terhadap kekayaan musik dunia [2]. Salah satu warisan musik yang paling kompleks dan signifikan di Indonesia adalah Gamelan Bali. Lebih dari sekadar alunan musik, Gamelan Bali merupakan bagian integral dari kehidupan masyarakat, ritual keagamaan, dan ekspresi seni di Bali. Keunikan Gamelan Bali tidak hanya terletak pada ragam ansambelnya, tetapi juga pada karakteristik akustiknya yang sangat khas. Fenomena ombak atau gelombang suara yang bergetar, dihasilkan dari sistem penyetelan berpasangan (*paired tuning*) di mana dua instrumen identik disetel dengan frekuensi yang sedikit berbeda, menciptakan tekstur audio yang kaya dan hidup [3]. Kompleksitas ini, ditambah dengan adanya respons akustik nonlinier pada instrumen seperti gong, menjadikan audio Gamelan Bali sebagai objek penelitian yang menarik sekaligus menantang dari perspektif komputasional. Setiap ansambel seperti Gong Kebyar yang dinamis dan meledak-ledak, Semar Pegulingan yang halus dan meditatif untuk suasana istana, atau Baleganjur yang ritmis untuk prosesi [4], memiliki sidik jari yang unik, yang merepresentasikan fungsi sosial dan estetikanya masing-masing. Meskipun potensi *deep learning* untuk analisis musik sangat besar, penerapannya pada domain musik Gamelan masih terbatas dan menyisakan celah penelitian yang signifikan. Beberapa penelitian yang ada telah berhasil menggunakan arsitektur *deep*

learning seperti CNN-LSTM pada Gamelan Jawa, namun dengan fokus pada tugas generasi notasi musik (*note generation*) dengan tingkat akurasi mencapai 91,9% [5]. Hingga saat ini, penelitian yang secara spesifik menangani masalah klasifikasi audio untuk membedakan berbagai jenis ansambel Gamelan Bali secara otomatis masih belum banyak dieksplorasi. Untuk mengatasi tantangan ini, diperlukan sebuah model yang mampu menganalisis sinyal audio secara komprehensif. Ketika sinyal audio direpresentasikan sebagai spektrogram, ia memiliki dua dimensi informasi krusial: fitur spasial (distribusi frekuensi pada satu waktu) dan fitur temporal (evolusi distribusi frekuensi dari waktu ke waktu). Arsitektur Convolutional Long Short-Term Memory (ConvLSTM) dirancang secara khusus untuk menangani data spasial-temporal seperti ini, karena kemampuannya untuk mengekstrak fitur spasial melalui lapisan konvolusional sambil memodelkan dependensi temporal jangka panjang melalui struktur LSTM [6]. Oleh karena itu, penelitian ini bertujuan untuk merancang dan mengimplementasikan sebuah model *deep learning* menggunakan arsitektur ConvLSTM yang mampu mengklasifikasikan lima jenis ansambel Gamelan Bali diantaranya Angklung, Gong Kebyar, Gong Gede, Baleganjur, dan Semar Pegulingan secara otomatis berdasarkan rekaman audionya. Keberhasilan penelitian ini diharapkan dapat memberikan beberapa kontribusi penting. Secara teknis, penelitian ini akan menjadi studi kasus penerapan arsitektur ConvLSTM pada data audio musik tradisional yang kompleks. Secara praktis, model yang dihasilkan dapat menjadi alat bantu (*assistive tool*) bagi para etnomusikolog untuk mengidentifikasi dan mengkatalogkan arsip audio dalam jumlah besar secara efisien. Dalam jangka panjang, penelitian ini merupakan langkah fundamental menuju pembangunan sistem pengarsipan digital yang cerdas dan dapat ditelusuri (*searchable*), yang tidak hanya mengamankan warisan budaya Bali dari ancaman degradasi [7], tetapi juga membuatnya lebih mudah diakses untuk tujuan pendidikan, penelitian, dan apresiasi budaya di tingkat global.

2. TINJAUAN PUSTAKA

2.1. *Game Edukasi dan Metode Agile*

Game edukasi didefinisikan sebagai media pembelajaran interaktif yang menggabungkan unsur hiburan dan pendidikan untuk meningkatkan daya ingat serta pemahaman materi melalui keterlibatan aktif [8]. Dalam pengembangannya, metode Agile digunakan karena fleksibilitasnya dalam menghadapi perubahan melalui proses iteratif yang mencakup fase *pre-production*, *production*, dan *post-production* [9]. Mekanisme seperti *staging*, *rewarding*, dan *scoring* menjadi elemen kunci untuk memacu motivasi dan interaksi pemain dalam game kuis [10].

2.2. *Gamelan Bali sebagai Konten Edukasi*

Gamelan Bali merupakan warisan budaya yang memiliki nilai spiritual dan sosial yang mendalam bagi masyarakat Bali [11]. Penelitian ini secara khusus berfokus pada lima jenis genre gamelan Bali, yaitu Angklung, Balaganjur, Semar Pegulingan, Gong Kebyar, dan Gong Gede. Penggunaan teknologi digital dalam pengenalan gamelan menjadi solusi inovatif untuk mengatasi keterbatasan akses fisik dan menurunnya minat generasi muda terhadap musik tradisional [12].

2.3. *Convolutional Long Short-Term Memory (ConvLSTM)*

ConvLSTM adalah arsitektur jaringan saraf hibrida yang menggabungkan kekuatan Convolutional Neural Networks (CNN) untuk ekstraksi fitur spasial, sebuah pendekatan yang telah terbukti tingkat akurasi dalam berbagai klasifikasi citra [13], dan Long Short-Term Memory (LSTM) untuk analisis fitur temporal. Dalam klasifikasi audio, ConvLSTM bekerja dengan menganalisis sekuens representasi visual audio (matriks 2D) untuk mengenali pola ritmis, melodi, dan tekstur musik yang dinamis secara simultan [14].

2.4. *Ekstraksi Fitur Audio*

Untuk mengonversi sinyal audio menjadi input yang dapat diproses oleh model deep learning, dilakukan ekstraksi tiga fitur utama:

- Mel-Frequency Cepstral Coefficients (MFCC): Menangkap karakteristik

timbre atau "warna suara" yang meniru persepsi pendengaran manusia [15].

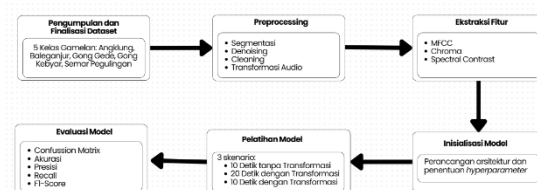
- Chroma Features: Menganalisis konten harmonik dan distribusi energi nada dalam tangga nada kromatik [16].
- Spectral Contrast: Mengukur tekstur musikal dengan menghitung perbedaan antara puncak dan lembah energi dalam spektrum frekuensi [17].

2.5. *Pengujian Perangkat Lunak dan Evaluasi Model*

Kualitas aplikasi diverifikasi menggunakan *Blackbox Testing* untuk memastikan fungsionalitas fitur sesuai kebutuhan pengguna tanpa melihat struktur internal kode [18]. Sementara itu, performa model klasifikasi dievaluasi menggunakan metrik standar seperti Akurasi, *Confusion Matrix*, Precision, Recall, dan F1-Score untuk menjamin keandalan prediksi dalam skenario nyata [19].

3. METODE PENELITIAN

Penelitian ini menerapkan model Research and Development (R&D) untuk membangun sebuah sistem klasifikasi audio Gamelan Bali. Fokus utama dari tahap pengembangan adalah merancang, melatih, dan mengevaluasi sebuah model *deep learning* dengan arsitektur *Convolutional Long Short-Term Memory* (ConvLSTM). Model ini dirancang secara spesifik untuk dapat mengidentifikasi dan membedakan lima kelas Gamelan Bali berdasarkan rekaman audionya. Pendekatan ini dipilih karena kemampuan arsitektur ConvLSTM yang unggul dalam memproses data yang memiliki dimensi spasial dan temporal secara bersamaan, seperti representasi visual dari sinyal audio. Proses pengembangan model ConvLSTM dilakukan secara sistematis melalui serangkaian langkah yang terstruktur, mulai dari pengumpulan dataset hingga evaluasi akhir performa model. Tahapan-tahapan tersebut diuraikan pada Gambar 1.



Gambar 1. Tahapan Penelitian

3.1. Pengumpulan dan Finalisasi Dataset

Dataset yang digunakan dalam penelitian ini terdiri dari kumpulan audio yang merepresentasikan 5 kelas Gamelan Bali, yaitu Angklung, Baleganjur, Semar Pegulingan, Gong Kebyar, dan Gong Gede. Data audio diperoleh dari dua sumber utama: rekaman langsung dari Sanggar Seni Anglocita Suara Singaraja dan koleksi audio dari platform YouTube. Untuk memastikan validitas dan relevansi setiap sampel, keseluruhan dataset divalidasi oleh seorang ahli musik gamelan dengan pemilihan dataset berfokus pada karakteristik pembeda pada Tabel 1.

Tabel 1. Karakteristik Dataset

Kelas Gamelan	Sistem Laras (Tuning)	Karakteristik Pembeda Utama
Angklung	Slendro (4 nada)	Laras 4 nada yang khas, suara ringan & jernih dari perunggu/besi. Identik dengan upacara kematian.
Baleganjur	Pelag (4 nada)	Tempo cepat & heroik, untuk mengiringi prosesi/arak-arakan. Dominasi suara simbal Ceng-ceng Kopyak.
Gong Gede	Pelag (5 nada)	Tempo sangat lambat (slow), megah, dan agung. Suara gong besar sangat dominan. Bernuansa kuno dan sakral untuk upacara besar di pura.
Gong Kebyar	Pelag (5 nada)	Sangat dinamis, kontras, dan eksplosif (kebyar). Tempo bisa berubah drastis dari lambat ke sangat cepat. Permainan gangsa

		(metalofon) sangat dominan.
Semar Pegulingan	Pelag (7 nada)	Alunan melodi yang cenderung lembut, manis, dan romantis. Instrumen Trompong seringkali memegang peranan sebagai pembawa melodi utama.

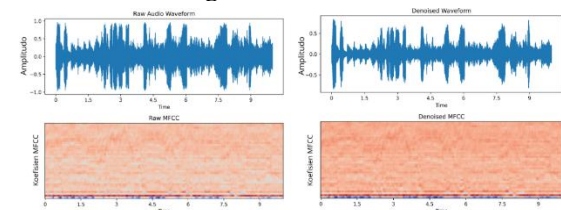
3.2. Preprocessing dan Ekstraksi Fitur

Sebelum data audio dapat diolah oleh model, serangkaian langkah pra-pemrosesan dan ekstraksi fitur dilakukan untuk mengubah sinyal audio mentah menjadi tensor numerik yang terstruktur dan kaya informasi. Alur *preprocessing* dan ekstraksi fitur sebagai Berikut:

3.2.1. Segmentasi Durasi Tetap (Fixed-length Segmentation)

Sinyal audio yang telah bersih dipotong menjadi segmen-segmen dengan durasi yang seragam (10 detik dan 20 detik) untuk mengidentifikasi batas waktu dari suara [20]. Langkah ini krusial karena model deep learning seperti ConvLSTM memerlukan input dengan dimensi temporal yang konsisten untuk setiap sampel.

3.2.2. Denoising



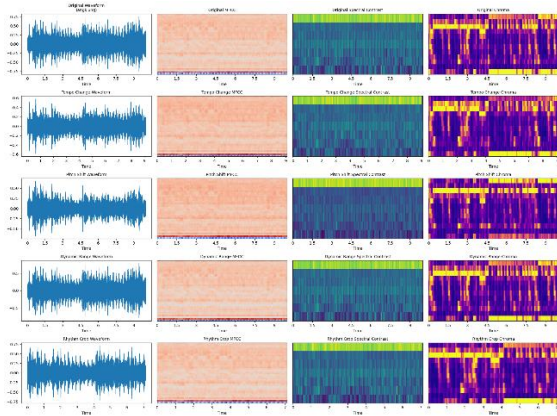
Gambar 2. Denoising

Sinyal audio dibersihkan dari derau (noise) latar yang tidak diinginkan untuk memastikan fitur yang diekstraksi murni merepresentasikan karakteristik musik gamelan.

3.2.3. Transformasi Audio

Khusus untuk data latih (*training set*), diterapkan serangkaian transformasi seperti *Pitch Shifting* dan *Time Stretching*. Teknik ini,

yang terbukti efektif dalam riset klasifikasi audio [21], bertujuan untuk meningkatkan ketangguhan (*robustness*) model agar lebih toleran terhadap variasi alami dalam musik gamelan, seperti perbedaan laras atau tempo permainan.



Gambar 3. Transformasi Data

3.2.4. Ekstraksi Fitur Gabungan

Dari setiap segmen audio, tiga jenis fitur akustik diekstraksi secara paralel:

- Mel-Frequency Cepstral Coefficients (MFCC): 40 koefisien diekstraksi untuk menangkap karakteristik timbre atau “warna suara” instrumen, yang terbukti meningkatkan akurasi dalam sistem pengenalan berbasis deep learning [15].
- Chroma Features: 12 bin diekstraksi untuk menganalisis konten harmonik dan melodi, yang efektif dalam membedakan antara sinyal musik dan ucapan [16].
- Spectral Contrast: 7 pita frekuensi diekstraksi untuk mengukur tekstur dan dinamika musikal, yang berguna untuk klasifikasi tipe musik [17].

3.2.5. Pembentukan Tensor Input

Ketiga set fitur tersebut kemudian digabungkan dengan cara ditumpuk secara vertikal (*vertical stacking*), menghasilkan matriks fitur tunggal berukuran 59 fitur. Matriks 2D inilah yang kemudian diubah bentuknya menjadi tensor 4D sebagai input untuk model ConvLSTM.

3.3. Arsitektur Model

Model yang digunakan adalah Convolutional Long Short-Term Memory (ConvLSTM). Arsitektur ini dirancang untuk secara efektif mempelajari pola spasial (fitur dalam setiap segmen waktu) dan pola temporal (bagaimana fitur tersebut berkembang seiring waktu), memberikan potensi besar dalam meningkatkan akurasi klasifikasi suara gamelan [14]. Model ini terdiri dari beberapa blok utama, termasuk lapisan ConvLSTM2D sebagai ekstraktor fitur spasial-temporal, lapisan konvolusi (Conv2D), lapisan pooling untuk reduksi dimensi, dan diakhiri dengan lapisan *fully connected* (Dense) dengan fungsi aktivasi softmax untuk melakukan klasifikasi akhir ke dalam lima kelas gamelan.

3.4. Pelatihan Model

Proses pelatihan model melibatkan penyesuaian bobot jaringan menggunakan algoritma optimasi Adam untuk meminimalkan fungsi *loss Categorical Cross-Entropy*. Pelatihan dilakukan menggunakan maksimal 500 *epoch* dengan ukuran *batch* maksimal 32. Beberapa *callbacks* penting diaktifkan untuk mengelola proses pelatihan, seperti *EarlyStopping* untuk mencegah *overfitting*, *ModelCheckpoint* untuk menyimpan model dengan performa terbaik, dan *ReduceLROnPlateau* untuk menyesuaikan *learning rate* secara dinamis.

3.5. Evaluasi Model

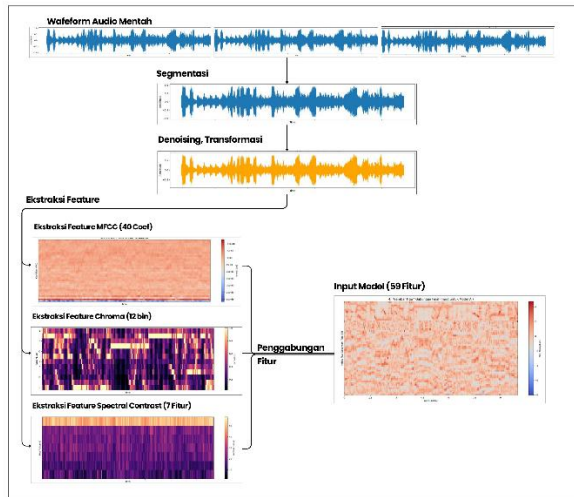
Setelah proses pelatihan selesai, performa model dievaluasi pada data validasi yang belum pernah dilihat sebelumnya. Evaluasi dilakukan menggunakan beberapa metrik standar untuk tugas klasifikasi, di antaranya adalah Confusion Matrix untuk memvisualisasikan performa per kelas [19], serta metrik kuantitatif seperti Akurasi, Presisi, Recall, dan F1-Score untuk mengukur kinerja model secara keseluruhan.

4. HASIL DAN PEMBAHASAN

4.1. Hasil Persiapan Data, Preprocessing, dan Ekstraksi Fitur

Model ConvLSTM dilatih menggunakan representasi numerik yang diekstraksi dari sinyal audio. Proses ini mengubah gelombang suara menjadi “gambar fitur” yang kaya informasi statistik. Alur kerja ini, mulai dari

pembersihan sinyal audio (*denoising*) hingga penggabungan tiga set fitur utama (MFCC, Chroma, dan Spectral Contrast) menjadi satu matriks input berukuran 59 fitur. Alur proses divisualisasikan pada Gambar 3.



Gambar 4. Alur Persiapan Data, Preprocessing, dan Ekstraksi Fitur

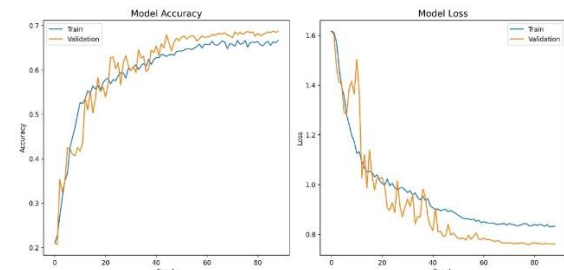
Dari proses tersebut didapatkan dataset final dengan jumlah 5 kelas dengan Setiap varian dataset 10 dan 20 detik yang dibagi dengan rasio 80% untuk data pelatihan (*training*) dan 20% untuk data validasi (*validation*), dengan jumlah total data yang didapatkan setelah dilakukan segmentasi dan pembersihan data diantaranya sebanyak 6070 file audio untuk dataset 10 detik dan 1275 file audio 20 detik. Data tersebut didistribusikan dengan seimbang untuk setiap kelasnya. Data audio 10 detik untuk *training* berjumlah 1050 untuk setiap kelasnya dan data *validation* berjumlah 210 untuk setiap kelasnya. Sedangkan data audio 20 detik untuk *training* berjumlah 250 untuk setiap kelasnya dan data *validation* berjumlah 50 untuk setiap kelasnya

4.2. Hasil Pelatihan dan Komparasi Hasil Model

Untuk mendapatkan performa terbaik, dilakukan tiga skenario eksperimen pelatihan yang berbeda. Hasil dari setiap eksperimen dievaluasi berdasarkan akurasi pada data validasi untuk memilih model final yang akan digunakan.

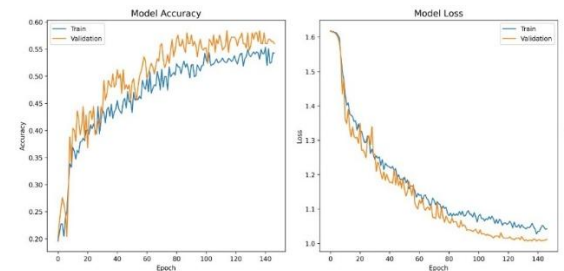
Model pertama dilatih menggunakan dataset dengan segmen audio 10 detik tanpa melalui proses transformasi *data*. Pelatihan ini

berfungsi sebagai baseline untuk mengukur dampak dari teknik yang diterapkan selanjutnya. Model ini mencapai akurasi validasi keseluruhan sebesar 69%.



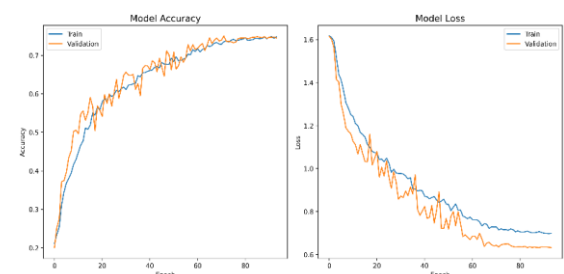
Gambar 5. Training Plot 10 Detik Tanpa Transformasi Data

Untuk menguji hipotesis bahwa durasi segmen yang lebih panjang dapat menangkap lebih banyak informasi musikal, model dilatih menggunakan segmen audio 20 detik yang telah melalui transformasi data. Hasilnya, performa model justru menurun secara signifikan, dengan akurasi validasi hanya mencapai 58%.



Gambar 6. Training Plot 20 Detik dengan Transformasi Data

Skenario ketiga menggunakan segmen audio 10 detik dengan menerapkan teknik transformasi data seperti *pitch shifting* dan *time stretching* pada data latih. Hasil pelatihan ini menghasilkan performa terbaik, di mana model mencapai akurasi validasi sebesar 75%. Proses pelatihan untuk model ini divisualisasikan pada Gambar 7, yang menunjukkan kurva akurasi dan *loss* yang stabil tanpa indikasi *overfitting*.



Gambar 7. Training Plot 10 Detik dengan Transformasi Data

Perbandingan hasil training secara kuantitatif dapat dilihat pada Tabel 3.

Tabel 2. Komparasi Hasil Training

Skenario Training	Akurasi Validasi	F1-Score (Macro Avg)
10 Detik tanpa Transformasi	69%	0.67
20 Detik dengan Transformasi	58%	0.58
10 Detik dengan Transformasi	75%	0.74

Berdasarkan perbandingan hasil akurasi pada Tabel 3, model dari dataset 10 detik dengan transformasi data dipilih sebagai model final karena menunjukkan akurasi tertinggi yang kedepannya dapat dikembangkan ke bidang lainnya.

4.3. Evaluasi Performa Model Terbaik

Analisis selanjutnya difokuskan secara eksklusif pada model terbaik yang terpilih (10 detik dengan transformasi data).

Tabel 3. Classification Report

	Precisi	Recall	F1-Score	Support
Angklung	0,79	0,83	0,81	210
Baleganjur	0,89	0,92	0,91	210
Gong Gede	0,80	0,85	0,83	210
Gong Kebyar	0,57	0,44	0,50	210
Semar Pegulingan	0,64	0,69	0,66	210
Accuracy			0,75	1050
Macro Avg	0,74	0,75	0,74	1050
Weighted Avg	0,74	0,75	0,74	1050



Gambar 8. Confusion Matrix

Data evaluasi menunjukkan bahwa model memiliki performa yang sangat baik pada beberapa kelas (misalnya, Baleganjur dengan F1-Score 0.91) namun menghadapi tantangan pada kelas lainnya, terutama Gong Kebyar (F1-Score 0.50).

5. KESIMPULAN DAN SARAN

Hasil penelitian memberikan beberapa wawasan penting. Peningkatan akurasi dari 69% menjadi 75% setelah penerapan transformasi data secara kuantitatif membuktikan bahwa melatih model pada data yang lebih bervariasi dapat meningkatkan kemampuan generalisasinya. Pendekatan ini “memaksa” model untuk mempelajari representasi fitur yang lebih tangguh terhadap variasi alami dalam musik gamelan, sebuah metode yang terbukti efektif untuk mengurangi overfitting dalam klasifikasi audio [21]. Sebaliknya, penurunan performa pada segmen 20 detik (58%) mengindikasikan bahwa durasi yang lebih panjang justru memasukkan variasi atau bagian transisi yang dapat “membingungkan” model, sehingga durasi 10 detik merupakan kompromi yang lebih baik.

Analisis lebih dalam pada performa model terbaik menunjukkan tantangan signifikan dalam membedakan kelas Gong Kebyar dan Semar Pegulingan. Hal ini sejalan dengan tantangan dalam riset klasifikasi audio,

di mana batasan segmen yang artifisial tidak selalu selaras dengan struktur musik yang sebenarnya, yang dapat membatasi tingkat akurasi yang bisa dicapai [22]. Gong Kebyar adalah pengembangan historis yang lebih dinamis dari Semar Pegulingan, dan keduanya berbagi inti instrumentasi (terutama gangsa) serta laras Pelog. Perbedaan sonik yang halus kemungkinan tertutupi oleh fenomena oklusi akustik, di mana suara instrumen yang sangat dominan “menutupi” timbre instrumen pembeda yang lebih halus, membuat sinyal audio keduanya sangat mirip bagi model [23]. Keterbatasan ini, bersama dengan sifat segmen audio berdurasi tetap yang tidak selalu selaras dengan struktur musik alami, menjadi area utama untuk perbaikan di masa depan. Meskipun demikian, akurasi 75% dinilai cukup andal untuk tujuan penelitian ini, yaitu sebagai sistem pendukung yang memberikan rekomendasi klasifikasi audio kepada pengguna ahli.

Untuk pengembangan selanjutnya, disarankan untuk 1) Mengeksplorasi arsitektur deep learning lain yang lebih canggih; 2) Memperbesar dan melakukan kurasi dataset secara lebih ketat dengan bantuan ahli musik untuk memastikan kemurnian sampel; dan 3) Mengimplementasikan teknik augmentasi data yang lebih beragam untuk meningkatkan jumlah dan variasi data latih.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih yang sebesar-besarnya kepada Ir. Ketut Agus Seputra, S.ST., M.T., selaku dosen pembimbing I, dan Ir. I Gede Mahendra Darmawiguna, S.Kom., M.Sc., selaku dosen pembimbing II. Ucapan terima kasih juga ditujukan kepada PT. Dago Engineering serta Bapak Putu Aldi Philberta Harta Celuk dan Bapak I Wayan Suarsana sebagai ahli musik yang telah membantu dalam validasi konten dan dataset.

DAFTAR PUSTAKA

[1] H. Guo, J. Wang, X. Xu, C. C. Hung, dan S. An, “A Hybrid Deep Learning Model for Automatic Classification of Cantonese Opera,” *Sustainability*, vol. 14, no. 5, hlm. 2923, Mar. 2022.

[2] A. M. Tarsitani dan G. G. Acosta, “Artificial intelligence in music,” *Revista de la Asociación Argentina de Musicología*, vol. 27, no. 1, hlm. 1-22, 2021.

[3] K. L. Gee, “The Gamelan,” *BYU Physics and Astronomy*, 2022. [Online]. Tersedia: <https://physics.byu.edu/faculty/gee/gamelan>.

[4] W. Vitale, “Balinese Gamelan Semar Pagulingan,” *Gamelan.org*, 2002. [Online]. Tersedia: <https://www.gamelan.org/ensembles/saih-pitu>.

[5] L. A. Kurniawati, A. T. Hidayatno, dan I. P. H. S. Putra, “Deep Learning for Multi-Structured Javanese Gamelan Note Generator,” *Jurnal Ilmiah Teknologi dan Informasi Asia*, vol. 17, no. 2, hlm. 119-128, Agu. 2023.

[6] Z. Shi, H. Wang, dan Y. Li, “An Attention-based Conv-LSTM for Spatiotemporal Flow Prediction,” *arXiv preprint arXiv:2006.08520*, 2020.

[7] A. Kaipainen, “Challenges of Music in the Digital Era,” *THEORIA: An International Journal for Theory, History, and Foundations of Science*, vol. 34, no. 1, hlm. 121-125, Jan. 2019.

[8] F. M. Dewadi, “Pembelajaran dan pengenalan musik dalam melatih daya ingat anak,” *JECED: Journal of Early Childhood Education and Development*, vol. 3, no. 1, hlm. 15-23, 2021.

[9] J. S. Atmadja, “Membangun game menggunakan metode agile development,” *Medium*, 30 Apr. 2020. [Online]. Tersedia: <https://medium.com/javanlabs/membangun-game-menggunakan-metode-agile-development-a661cb48fe81>.

[10] A. Haddid, M. Q. Aufar, A. Ramadhan, dan Y. Pratama, “Perancangan game edukasi menebak lirik lagu-lagu daerah berbasis Unity 2D untuk platform Android,” *TIN: Terapan Informatika Nusantara*, vol. 4, no. 2, hlm. 140-151, 2023.

[11] I. N. R. Artana, “Balinese Gamelan Learning Process During The Covid-19 Pandemic,” *LULIK: Jurnal Ilmiah Kreatifitas, Seni, dan Budaya*, vol. 11, no. 2, hlm. 106-115, 2022.

[12] I. G. G. Surya, I. N. D. Putra, dan N. M. R. A. Gelgel, “The Balinese gamelan phenomenon in digital simulation,” *E-Journal of Cultural Studies*, vol. 17, no. 3, hlm. 1-10, 2024.

[13] G. A. Pratama, E. Y. Puspaningrum, dan H. Maulana, “Convolutional Neural Network dan Faster Region Convolutional Neural Network Untuk Klasifikasi Kualitas Biji Kopi Arabika,” *Jurnal Informatika dan Teknik Elektro Terapan (JITET)*, vol. 12, no. 3, 2024.

[14] A. Shvets, I. Vankov, dan D. Serokutov, “Spiking ConvLSTM for semantic music generation,” *arXiv preprint arXiv:1911.11447*, 2019.

- [15] A. Telmem, "Speech recognition based on convolutional neural networks and MFCC algorithm," *International Journal of Advanced Computer Science and Applications*, vol. 12, no. 3, 2021.
- [16] G. K. Birajdar dan M. Patil, "Speech/music classification using visual and spectral chromagram features," *Journal of Ambient Intelligence and Humanized Computing*, vol. 11, no. 1, pp. 329–347, 2020.
- [17] D. N. Jiang, L. Lu, H. J. Zhang, J. H. Tao, dan L. H. Cai, "Music type classification by spectral contrast feature," dalam *Proc. IEEE Int. Conf. Multimedia and Expo*, 2002, vol. 1, pp. 113–116.
- [18] A. Fahrezi, F. N. Salam, G. M. Ibrahim, R. Rahman, dan A. Saifudin, "Pengujian black box testing pada aplikasi inventori barang berbasis web di PT. AINO Indonesia," *Jurnal Teknologi dan Sistem Informasi (JTISI)*, vol. 1, no. 1, hlm. 61-66, 2022.
- [19] Ž. Vujović, "Classification model evaluation metrics," *International Journal of Advanced Computer Science and Applications*, vol. 12, no. 6, 2021.
- [20] S. Venkatesh, D. Moffat, dan E. R. Miranda, "You only hear once: A YOLO like algorithm for audio segmentation and sound event detection," *Applied Sciences*, vol. 12, no. 7, p. 3293, 2022.
- [21] J. Salamon dan J. P. Bello, "Deep convolutional neural networks and data augmentation for environmental sound classification," *IEEE Signal Processing Letters*, vol. 24, no. 3, pp. 279-283, 2017.
- [22] K. West dan S. Cox, "Finding an optimal segmentation for audio genre classification," dalam *Proc. 25th Int. Society for Music Information Retrieval Conf., ISMIR 2024*, 2024.
- [23] J. Zheng, M. Cao, dan C. Zhang, "ICKAN: A deep musical instrument classification model incorporating Kolmogorov-Arnold network," *Scientific Reports*, vol. 14, no. 1, p. 9493, 2024.