

Klasifikasi Sub-Genre Musik Dangdut Menggunakan Jaringan Saraf Tiruan Long Short-Term Memory

Jevan Bernard Kaloko^{a1}, I Made Widiartha.^{a2}

^aProgram Studi Informatika, Fakultas Matematika dan Ilmu Pengetahuan Alam,
Universitas Udayana
Jalan Raya Kampus Udayana, Bukit Jimbaran, Kuta Selatan, Badung, Bali, Indonesia
¹kaloko.2308561078@student.unud.ac.id
²madewidiartha@unud.ac.id

Abstract

The automatic classification of Dangdut music sub-genres (Klasik, Koplo, and Campursari) presents a significant challenge in the field of Music Information Retrieval (MIR) due to their overlapping yet distinct musical characteristics. This research proposes a classification system based on a Recurrent Neural Network (RNN) with a Long Short-Term Memory (LSTM) architecture to address this problem. The model is trained using Mel-Frequency Cepstral Coefficients (MFCC) audio features to represent the spectral and timbral characteristics of each sub-genre. The LSTM architecture was chosen for its superior ability to learn temporal dependencies from the sequence of MFCC features. By modeling the evolution of timbre over time, the system can recognize the distinctive patterns that differentiate between Dangdut Klasik, Koplo, and Campursari. The proposed system aims to provide an accurate and efficient classification method, contributing to practical applications such as music recommendation and digital archiving.

Keywords: Music Genre Classification, Dangdut, Koplo, Campursari, Recurrent Neural Network, Long Short-Term Memory, Audio Feature Extraction, MFCC, MIR

1. Pendahuluan

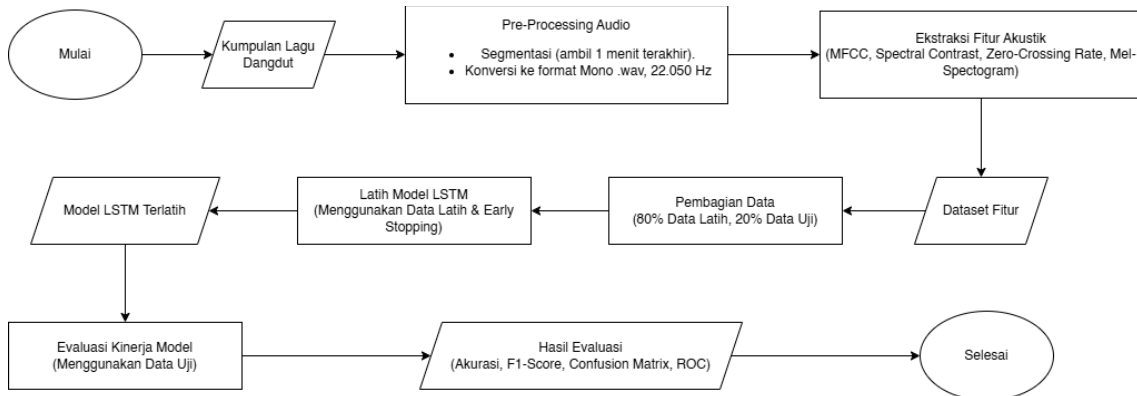
Dangdut, sebagai bagian dari warisan budaya Indonesia, telah berkembang menjadi beberapa sub-genre yang memiliki karakteristik musikal unik, seperti dangdut klasik, koplo, dan campursari. Dangdut klasik umumnya mengusung tempo dan instrumen tradisional yang lebih lambat, sedangkan koplo menonjolkan beat cepat dan dominasi instrumen perkusi [1]. Campursari, di sisi lain, merupakan perpaduan antara unsur gamelan Jawa dan musik modern sehingga mencerminkan kompleksitas audio yang berbeda [1]. Karakteristik temporal dan ritmis khusus ini menimbulkan tantangan tersendiri dalam upaya klasifikasi otomatis sub-genre dangdut, yang memerlukan analisis fitur sekuensial secara mendalam.

Berbagai metode pembelajaran mesin, seperti K-Nearest Neighbor (K-NN) dan Support Vector Machine (SVM), telah digunakan untuk klasifikasi genre musik dengan mengandalkan fitur seperti MFCC, zero-crossing rate, serta spectral roll-off [2],[3],[4],[5]. Namun, metode-metode tersebut kurang mempertimbangkan dinamika temporal dari musik, yang sebetulnya sangat relevan dalam membedakan sub-genre dangdut. Sebagai alternatif, arsitektur Long Short-Term Memory (LSTM) telah terbukti efektif dalam mengklasifikasikan genre musik melalui pemodelan sekuensial, sebagaimana dicontohkan oleh Yi et al. yang berhasil meraih akurasi 93% menggunakan LSTM dengan fitur MFCC dan ZCR [6]. Penelitian ini bertujuan mengaplikasikan LSTM untuk klasifikasi sub-genre klasik, koplo, dan campursari, dengan harapan mampu mengungguli metode prior pada analisis ritme dan urutan fitur audio.

2. Metode Penelitian

Penelitian ini dilakukan melalui serangkaian tahapan yang sistematis untuk membangun dan mengevaluasi model klasifikasi sub-genre musik dangdut. Alur penelitian dirancang secara logis,

dimulai dari pengumpulan data hingga evaluasi performa model. Gambaran umum dari metodologi penelitian ini diilustrasikan pada Gambar 1.



Gambar 1. Flowchart Metode Penelitian

2.1. Pengumpulan dan Pre-processing Data

Tahap pertama dalam penelitian ini adalah pengumpulan data audio. Dataset dibangun secara manual untuk memastikan relevansi dan kekhasan setiap sub-genre. Tiga sub-genre dangdut yaitu klasik, koplo, dan campursari didefinisikan dan direpresentasikan oleh tiga artis ikonik: Rhoma Irama (klasik), Denny Caknan (koplo), dan Didi Kempot (campursari).

Untuk setiap lagu, dilakukan proses pra-pemrosesan sebagai berikut:

- a. Segmentasi: Sampel audio dengan durasi satu menit terakhir dari setiap lagu diambil. Pemilihan segmen akhir ini bertujuan untuk menangkap bagian klimaks atau aransemen penutup yang seringkali kaya akan ciri khas musikal [6].
- b. Konversi Format: Seluruh sampel audio dikonversi ke dalam format .mp3 dengan mode mono dan sample rate 22.050 Hz. Standardisasi ini diperlukan untuk memastikan konsistensi selama proses ekstraksi fitur [6].

2.2. Ekstraksi Fitur Akustik

Setelah pra-pemrosesan, fitur-fitur akustik diekstrak dari setiap sampel audio menggunakan library Librosa pada Python. Fitur-fitur ini dipilih karena kemampuannya dalam merepresentasikan karakteristik suara yang berbeda, mulai dari timbre, harmoni, hingga ritme. Deskripsi dari setiap fitur yang digunakan disajikan pada Tabel 1.

Tabel 1. Deskripsi Fitur Akustik yang Digunakan

Fitur	Deskripsi
MFCC	Merepresentasikan "warna" atau tekstur suara (timbre) yang sangat penting untuk membedakan instrumen dan vokal [4].
Spectral Contrast	Mengukur perbedaan antara puncak dan lembah energi pada spektrum frekuensi, berkaitan dengan kejernihan tekstur musik.
Zero-Crossing Rate	Menghitung laju perubahan sinyal dari positif ke negatif, sering berkorelasi dengan suara perkusi atau <i>noise</i> .
Mel-Spectrogram	Representasi spektrum frekuensi dari waktu ke waktu dengan skala Mel yang meniru persepsi pendengaran manusia.

Seluruh fitur ini digabungkan untuk membentuk sebuah matriks fitur sekuensial (sequence feature matrix) untuk setiap sampel lagu, yang kemudian menjadi masukan bagi model LSTM.

2.3. Arsitektur Model Jaringan Saraf Tiruan

Penelitian ini menggunakan Jaringan Saraf Tiruan (JST) dengan arsitektur *Long Short-Term Memory* (LSTM), yang merupakan jenis dari *Recurrent Neural Network* (RNN). Arsitektur ini dipilih karena kemampuannya yang unggul dalam memproses data sekuensial seperti audio [6]. Model dibangun menggunakan framework TensorFlow dengan Keras API.

Arsitektur yang diusulkan terdiri dari beberapa lapisan utama, seperti yang diilustrasikan pada Gambar 2.

Model: "functional"

Layer (type)	Output Shape	Param #
input_layer (InputLayer)	(None, 1292, 85)	0
lstm (LSTM)	(None, 1292, 64)	38,400
dropout (Dropout)	(None, 1292, 64)	0
lstm_1 (LSTM)	(None, 64)	33,024
dropout_1 (Dropout)	(None, 64)	0
feature_extractor_layer (Dense)	(None, 64)	4,160
dropout_2 (Dropout)	(None, 64)	0
dense (Dense)	(None, 3)	195

Total params: 75,779 (296.01 KB)
 Trainable params: 75,779 (296.01 KB)
 Non-trainable params: 0 (0.00 B)

Gambar 2. Arsitektur Model LSTM yang Diusulkan

Model ini terdiri dari lapisan input, dua lapisan LSTM dengan 64 unit, lapisan Dense dengan 64 unit sebagai feature extractor internal, dan lapisan output dengan 3 unit menggunakan fungsi aktivasi Softmax untuk klasifikasi multi-kelas. Lapisan Dropout dengan laju 0.3 disisipkan di antara lapisan untuk regularisasi dan mencegah overfitting.

2.4. Pelatihan dan Validasi Model

Dataset yang telah berisi vektor fitur dibagi menjadi dua bagian secara acak: 80% untuk data pelatihan dan 20% untuk data pengujian. Proses pelatihan menggunakan *optimizer Adam* dan *loss function Categorical Crossentropy*.

Untuk meningkatkan efisiensi training dan mengatasi potensi *overfitting* akibat dataset yang terbatas, diterapkan mekanisme *Early Stopping*. Mekanisme ini akan menghentikan proses pelatihan secara otomatis jika metrik *validation_loss* tidak menunjukkan perbaikan setelah 10 epoch berturut-turut, dan akan mengembalikan bobot model dari epoch terbaik.

2.5. Evaluasi Model

Performa model yang telah dilatih dievaluasi menggunakan data pengujian. Metrik yang digunakan untuk mengukur kinerja model adalah:

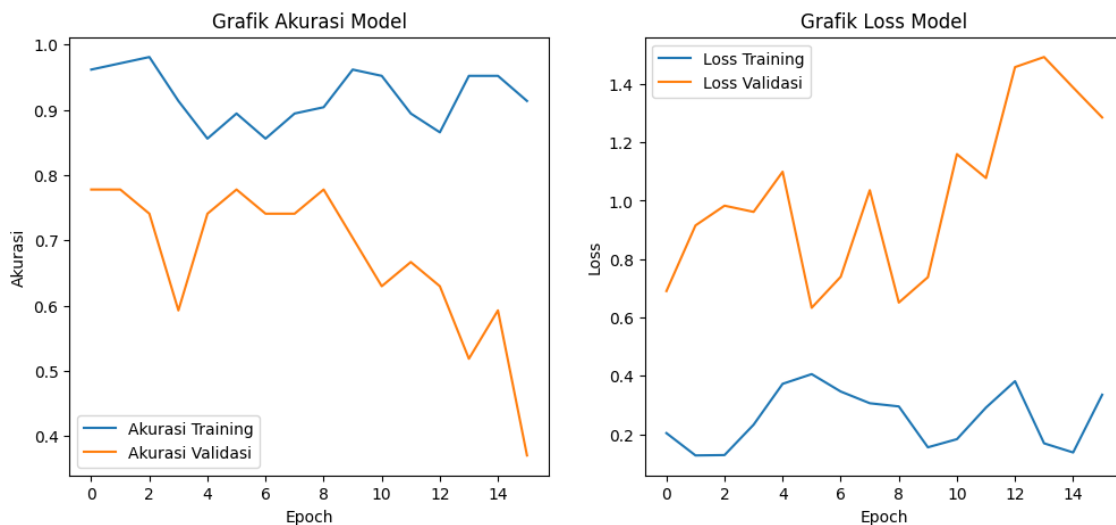
- Akurasi: Persentase prediksi yang benar secara keseluruhan.
- F1-Score: Rata-rata harmonik dari presisi dan recall, memberikan gambaran performa yang seimbang untuk setiap kelas.
- Confusion Matrix: Matriks untuk memvisualisasikan performa klasifikasi secara detail, menunjukkan kesalahan antara kelas-kelas yang berbeda.
- Kurva ROC dan AUC: Grafik untuk mengukur kemampuan diskriminatif model untuk setiap kelas.

3. Hasil dan Diskusi

Pada bab ini, akan dipaparkan hasil dari proses pelatihan dan evaluasi model Jaringan Saraf Tiruan (JST) Long Short-Term Memory (LSTM) yang telah dibangun. Analisis dilakukan terhadap serangkaian metrik evaluasi untuk mengukur kinerja model dalam mengklasifikasikan tiga sub-genre musik dangdut.

3.1. Hasil Pelatihan Model

Proses pelatihan model dijalankan selama 15 epoch sebelum dihentikan (kemungkinan oleh mekanisme Early Stopping atau batasan manual). Gambar 3 menyajikan grafik akurasi dan loss selama proses pelatihan.



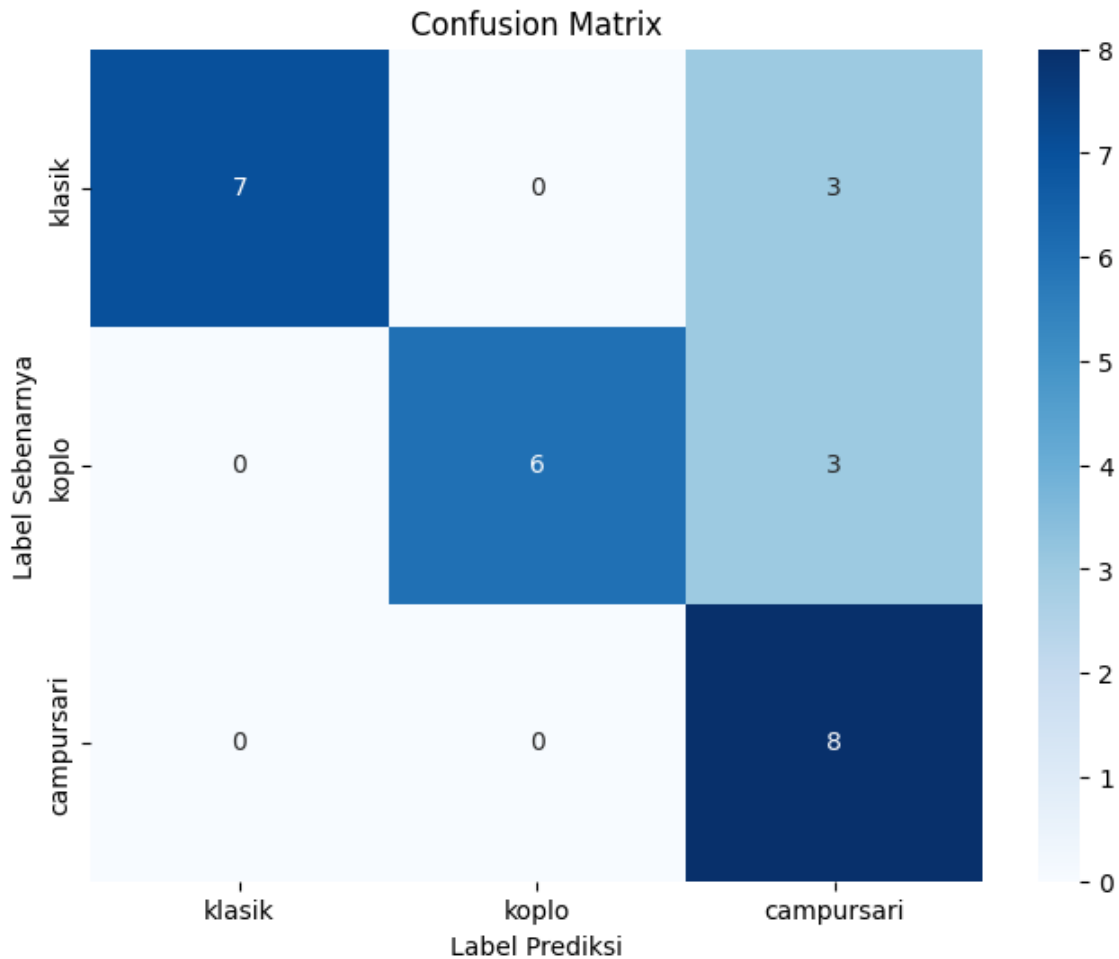
Gambar 3. Grafik Akurasi dan Loss Model Selama Pelatihan

Grafik Akurasi Model (kiri) menunjukkan bahwa akurasi pada data training (biru) secara konsisten meningkat hingga mencapai hampir 100%, yang menandakan model mampu mempelajari data latih dengan baik. Namun, akurasi pada data validasi (orange) menunjukkan fluktuasi yang signifikan dan tren menurun setelah mencapai puncaknya di awal, yang merupakan indikasi awal dari overfitting.

Indikasi ini diperkuat oleh Grafik Loss Model (kanan). Terlihat jelas adanya divergensi antara kurva loss training dan validasi. Sementara training loss terus menurun, *validation loss* cenderung meningkat seiring berjalannya waktu. Fenomena ini mengonfirmasi bahwa model mulai "menghafal" data training dan kesulitan untuk menggeneralisasi polanya pada data baru yang tidak terlihat sebelumnya.

3.2. Evaluasi Kinerja Klasifikasi

Kinerja model pada data uji dievaluasi secara kuantitatif menggunakan Confusion Matrix, F1-Score, dan Kurva ROC. Confusion matrix pada Gambar 4 menyajikan rincian hasil prediksi model. Dari total 27 data uji, model berhasil memprediksi 21 data dengan benar, menghasilkan akurasi keseluruhan sekitar 77.8%.

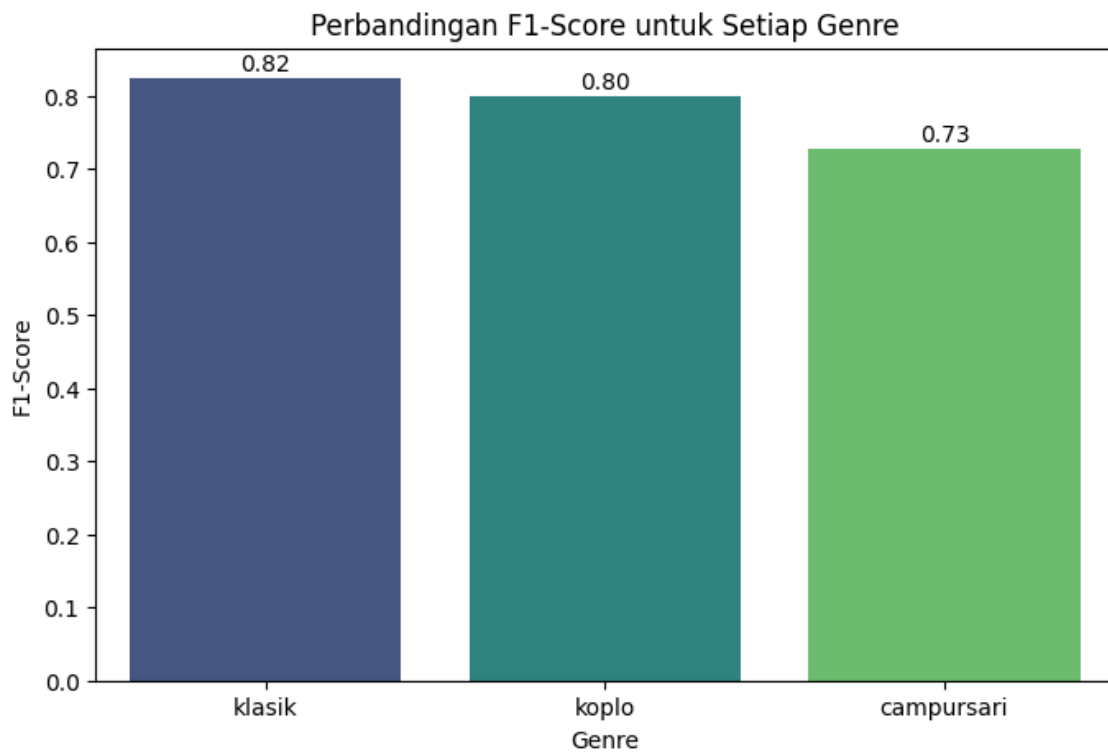


Gambar 4. Confusion Matrix Hasil Prediksi pada Data Uji

Dari matriks tersebut, dapat dianalisis sebagai berikut:

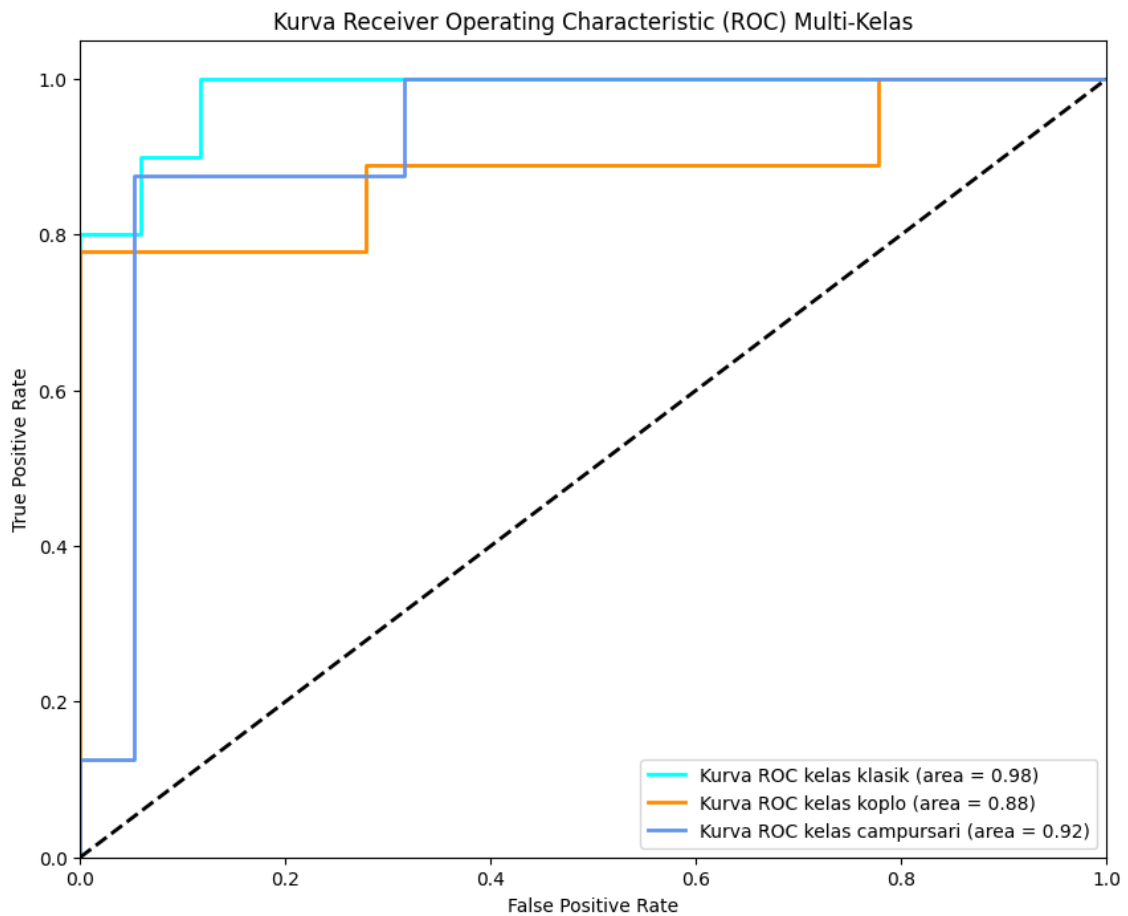
- Model menunjukkan performa terbaik pada kelas campursari, dengan 8 dari 8 sampel berhasil diklasifikasikan dengan benar.
- Kesalahan klasifikasi paling signifikan terjadi pada kelas klasik dan koplo yang sering keliru diidentifikasi sebagai campursari. Sebanyak 3 sampel klasik dan 3 sampel koplo salah diprediksi sebagai campursari. Hal ini menunjukkan adanya kemiripan fitur akustik dari sudut pandang model antara genre klasik/koplo dengan campursari.

Analisis F1-Score pada Gambar 5 memberikan pandangan yang seimbang mengenai performa per kelas. Kelas klasik (0.82) dan koplo (0.80) menunjukkan skor yang sangat baik, diikuti oleh campursari (0.73). Skor yang relatif tinggi ini menandakan bahwa meskipun ada kesalahan klasifikasi, model tetap memiliki keseimbangan presisi dan recall yang baik untuk setiap kelasnya.



Gambar 5. Perbandingan F1-Score untuk Setiap Genre

Kemampuan diskriminatif model untuk setiap kelas divisualisasikan melalui Kurva ROC pada Gambar 6. Semua kelas menunjukkan nilai Area Under the Curve (AUC) yang sangat tinggi yaitu klasik (0.98), koplo (0.88), dan campursari (0.92). Nilai AUC yang mendekati 1.0 ini mengindikasikan bahwa model memiliki kemampuan yang sangat baik untuk membedakan probabilitas antara satu kelas dengan kelas lainnya, bahkan jika keputusan klasifikasi akhirnya (berdasarkan probabilitas tertinggi) terkadang keliru.



Gambar 6. Kurva ROC Multi-Kelas

4. Kesimpulan

Penelitian ini berhasil menunjukkan bahwa model Jaringan Saraf Tiruan Long Short-Term Memory (LSTM) mampu mengklasifikasikan sub-genre dangdut klasik, koplo, dan campursari dengan performa yang menjanjikan, seperti yang divalidasi oleh nilai F1-Score dan AUC yang tinggi untuk setiap kelas. Meskipun demikian, model menghadapi tantangan signifikan dalam membedakan sub-genre dengan karakteristik akustik yang tumpang tindih, terutama antara koplo dan campursari, serta menunjukkan gejala overfitting akibat keterbatasan jumlah data. Oleh karena itu, penelitian di masa depan disarankan untuk fokus pada perluasan dataset dengan variasi lagu dan artis yang lebih banyak serta melakukan rekayasa fitur yang lebih mendalam untuk menangkap nuansa ritmis dan instrumental yang lebih spesifik guna meningkatkan akurasi dan kemampuan generalisasi model.

Daftar Pustaka

- [1] A. N. Weintraub, "The sound and spectacle of dangdut koplo: Genre and counter-genre in East Java, Indonesia," *Asian Music*, vol. 44, no. 2, hlm. 120-156, 2013.
- [2] I. G. Harsemadi, "Perbandingan kinerja algoritma K-NN dan SVM dalam sistem klasifikasi genre musik Gamelan Bali," *Informatics for Educators and Professional: Journal of Informatics*, vol. 8, no. 1, hlm. 1–10, 2023.
- [3] G. A. V. M. Giri, "Klasifikasi musik berdasarkan genre dengan metode K-Nearest Neighbor," *Jurnal Ilmu Komputer*, vol. 11, no. 2, hlm. 103–108, 2018.
- [4] H. Abdulbar, P. P. Adikara, dan S. Adinugroho, "Klasifikasi genre lagu dengan fitur akustik menggunakan metode K-Nearest Neighbor," *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 3, no. 8, hlm. 8259–8268, 2019.
- [5] T. Pratiwi, A. Sunyoto, dan D. Ariatmanto, "Music genre classification using K-Nearest Neighbors and Mel-Frequency Cepstral Coefficients," *Sinkron: Jurnal dan Penelitian Teknik Informatika*, vol. 8, no. 2, hlm. [nomor halaman], Apr. 2024.
- [6] Y. Yi, X. Zhu, Y. Yue, and W. Wang, "Music genre classification with LSTM based on time and frequency domain features," in *Proc. 2021 IEEE 6th Int. Conf. Comput. Commun. Syst. (ICCCS)*, Chengdu, China, Apr. 2021, hlm. 678–682. doi: 10.1109/ICCCS52626.2021.9449177.