

ABSTRAK

Klasifikasi genre musik dibutuhkan untuk membantu pengelolaan koleksi audio digital dan meningkatkan proses pencarian maupun rekomendasi musik. Penelitian ini membangun sistem klasifikasi genre musik berbasis spektrum audio menggunakan Audio Spectrogram Transformer (AST) dengan masukan log Mel-spectrogram. Pengujian dilakukan dalam tiga skenario, yaitu 10 genre sebagai baseline, 5 genre (classical, hiphop, jazz, metal, pop) sebagai skenario utama, serta 3 genre sulit (pop-reggae-rock) untuk menguji tantangan pada genre yang memiliki kemiripan karakteristik. Dataset yang digunakan berasal dari GTZAN, dengan subset utama 5 genre dan pembagian data secara stratified menjadi sekitar 70% data latih, 10% validasi, dan 20% data uji.

Tahap prapemrosesan meliputi standarisasi audio (mono, 16 kHz, durasi 15 detik), ekstraksi log Mel-spectrogram (128 mel), normalisasi, serta augmentasi untuk meningkatkan generalisasi. Representasi log Mel-spectrogram kemudian diproses oleh AST melalui pembentukan patch (patch embedding) dan Transformer encoder untuk menghasilkan prediksi kelas. Evaluasi kinerja dilakukan menggunakan confusion matrix serta metrik accuracy, precision (macro), recall (macro), dan F1-score (macro).

Hasil menunjukkan bahwa skenario 5 genre memberikan performa terbaik dengan accuracy 86,00%, precision (macro) 87,19%, recall (macro) 86,00%, dan F1-score (macro) 86,25%. Pada skenario 10 genre, performa menurun menjadi accuracy 66,50%, sedangkan pada skenario 3 genre sulit diperoleh accuracy 66,67%; penurunan ini dipengaruhi oleh kompleksitas multikelas dan kemiripan pola spektral antar genre. Selain itu, pada skenario 10 genre, capaian AST lebih tinggi dibanding pendekatan tradisional berbasis MFCC+CNN ResNet101 yang dilaporkan mencapai akurasi terbaik 58%, sehingga menunjukkan peningkatan sekitar 8,50 poin persentase.

Kata Kunci: Klasifikasi Genre Musik, Audio Spectrogram, Transformer, Deep Learning, Spektrum Audio

ABSTRACT

Music genre classification is needed to assist in managing digital audio collections and improve music search and recommendation processes. This study develops an audio Spectrogram-based music genre classification system using the Audio Spectrogram Transformer (AST) with log Mel-spectrogram input. Testing was conducted in three scenarios: 10 genres as a baseline, 5 genres (classical, hip-hop, jazz, metal, and pop) as the main scenario, and 3 challenging genres (pop–reggae–rock) to assess the challenges of genres with similar characteristics. The dataset used comes from GTZAN, with a primary subset of 5 genres and a stratified data division into approximately 70% training data, 10% validation data, and 20% test data.

The preprocessing stage includes audio standardization (mono, 16 kHz, 15 seconds duration), log Mel-spectrogram extraction (128 mel), normalization, and augmentation to improve generalization. The log Mel-spectrogram representation is then processed by the AST through patch embedding and a Transformer encoder to generate class predictions. Performance evaluation was conducted using a confusion matrix and the metrics accuracy, precision (macro), recall (macro), and F1-score (macro).

The results showed that the 5-genre scenario performed best with 86.00% accuracy, 87.19% precision (macro), 86.00% recall (macro), and 86.25% F1-score (macro). In the 10-genre scenario, performance decreased to 66.50% accuracy, while in the 3-genre scenario, accuracy of 66.67% was difficult to achieve. This decrease was influenced by multi-class complexity and similarity in spectral patterns between genres. Furthermore, in the 10-genre scenario, AST achieved higher performance than the traditional MFCC+CNN ResNet101 approach, which reportedly achieved the best accuracy of 58%, representing an improvement of approximately 8.50 percentage points.

Keywords: Music Genre Classification, Audio Spectrogram, Transformer, Deep Learning, Audio Spectrogram.