

Klasifikasi Genre Musik Menggunakan Metode Hibridasi Particle Swarm Optimization dan Backpropagation

I Kadek Agus Candra Widnyana^{a1}, I Made Widiartha^{a2}, I Gede Arta Wibawa^{a3}, I Gusti Ngurah Anom Cahyadi Putra^{a4}

^aProgram Studi Informatika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Udayana
Badung, Bali, Indonesia

¹widnyana.2208561129@student.unud.ac.id

²madewidiartha@unud.ac.id

³gede.artha@unud.ac.id

⁴anom.cp@unud.ac.id

Abstrak

Klasifikasi genre musik secara otomatis merupakan salah satu tantangan utama dalam bidang Music Information Retrieval (MIR) akibat tingginya kemiripan karakteristik akustik antar-genre musik. Penelitian ini mengusulkan pendekatan hibrid yang menggabungkan Mel-Frequency Cepstral Coefficients (MFCC) sebagai metode ekstraksi fitur akustik dengan Backpropagation Neural Network (BPNN), yang hyperparameternya dioptimasi menggunakan Particle Swarm Optimization (PSO). Dataset GTZAN yang terdiri atas 1.000 berkas audio dari 10 genre musik digunakan sebagai data penelitian. Proses optimasi PSO secara otomatis mencari nilai optimal lima hyperparameter BPNN, yaitu jumlah neuron pada tiga lapisan tersembunyi, learning rate, dan dropout rate, menggunakan skema 5-fold cross-validation. Hasil evaluasi parameter kendali PSO, yaitu koefisien kognitif ($c1$), koefisien sosial ($c2$), dan inertia weight (w), menunjukkan bahwa kombinasi $c1 = 1,2$, $c2 = 1,5$, dan $w = 0,7$ menghasilkan rata-rata validation loss terendah sebesar 0,9463 sehingga dipilih sebagai konfigurasi PSO terbaik. Berdasarkan konfigurasi tersebut, PSO memperoleh hyperparameter BPNN optimal berupa 192 neuron pada lapisan tersembunyi pertama, 192 neuron pada lapisan tersembunyi kedua, 32 neuron pada lapisan tersembunyi ketiga, learning rate sebesar 0,000185, dan dropout rate sebesar 0,379. Model BPNN-PSO yang dihasilkan memperoleh rata-rata accuracy sebesar 84,40%, precision sebesar 85,09%, recall sebesar 84,40%, dan F1-score sebesar 84,34%, lebih tinggi dibandingkan model BPNN tanpa optimasi yang memperoleh accuracy sebesar 82,80%, precision sebesar 83,60%, recall sebesar 82,80%, dan F1-score sebesar 82,80%. Berdasarkan evaluasi per genre, genre classical menunjukkan performa klasifikasi terbaik, diikuti oleh metal, blues, dan jazz, sedangkan genre rock, disco, dan reggae memiliki performa klasifikasi yang relatif lebih rendah. Hasil penelitian menunjukkan bahwa optimasi hyperparameter menggunakan PSO mampu meningkatkan performa BPNN pada klasifikasi genre musik berbasis MFCC, meskipun peningkatan yang diperoleh relatif kecil secara absolut.

Kata Kunci: Klasifikasi Genre Musik, Jaringan Saraf Tiruan BPNN, Particle Swarm Optimization, MFCC, GTZAN.

1. Pendahuluan

Industri musik digital telah berkembang sangat pesat dalam satu dekade terakhir. Platform streaming seperti Spotify, Apple Music, dan YouTube Music kini menyimpan puluhan juta lagu dari berbagai genre yang dapat diakses kapan saja oleh pengguna di seluruh dunia. Di balik kemudahan akses tersebut, terdapat tantangan teknis yang krusial, yaitu kebutuhan akan sistem yang mampu secara otomatis mengidentifikasi dan mengorganisasi konten musik berdasarkan genre-nya. Genre musik merupakan kategori konvensional yang mengidentifikasi sekelompok karya musik sebagai bagian dari tradisi atau seperangkat konvensi yang sama [1], dan kemampuan mengklasifikasikan genre secara otomatis menjadi fondasi penting bagi sistem rekomendasi musik, pencarian konten, dan pengelolaan perpustakaan audio berskala besar.

Secara tradisional, klasifikasi genre dilakukan secara manual dengan mendengarkan setiap lagu dan mengelompokkannya berdasarkan karakteristik yang didengar. Pendekatan ini menghasilkan penilaian yang akurat, namun sama sekali tidak skalabel ketika berhadapan dengan ribuan hingga jutaan berkas audio [2]. Di sisi lain, manusia pun memiliki keterbatasan dalam proses ini karena konsep genre bersifat subjektif, yaitu interpretasinya dipengaruhi oleh budaya, preferensi pribadi, serta pengetahuan musikologis masing-masing individu [3]. Oleh karena itu, pengembangan sistem klasifikasi genre musik otomatis berbasis komputer menjadi kebutuhan mendesak dalam bidang Music Information Retrieval (MIR).

Dalam MIR, proses klasifikasi diawali dengan ekstraksi fitur audio untuk mengubah sinyal mentah menjadi representasi numerik yang dapat diproses oleh algoritma. Salah satu metode ekstraksi fitur yang paling mapan adalah Mel-Frequency Cepstral Coefficients (MFCC), yang bekerja dengan merepresentasikan karakteristik spektral sinyal audio menggunakan skala mel yang mencerminkan sensitivitas sistem pendengaran manusia [4]. Dalam konteks klasifikasi genre musik, MFCC unggul dalam menangkap informasi timbre, yaitu warna atau karakter bunyi yang membedakan satu instrumen dari instrumen lain, yang menjadi pembeda utama antargenre [5]. Keunggulan ini dibuktikan secara empiris melalui perbandingan model berbasis MFCC dan Chroma Feature pada dataset GTZAN, di mana MFCC menghasilkan akurasi 80,2% sementara Chroma Feature hanya 53% [6].

Kualitas representasi fitur yang baik perlu didukung oleh model klasifikasi yang mampu mempelajari pola kompleks dari data audio yang beragam. Backpropagation Neural Network (BPNN) merupakan salah satu model yang banyak digunakan karena kemampuannya mendekati fungsi non-linear, memproses informasi secara paralel, dan beradaptasi terhadap data sampel yang bervariasi. Meskipun demikian, BPNN memiliki kelemahan mendasar, yaitu bobot jaringan diinisialisasi secara acak sehingga proses pelatihan rentan terjebak pada local optimum dan mengalami konvergensi yang lambat. Kelemahan ini berdampak langsung pada konsistensi dan akurasi klasifikasi, khususnya pada tugas multi-class seperti klasifikasi 10 genre musik yang memiliki kemiripan karakteristik akustik satu sama lain. Di sinilah letak kesenjangan penelitian yang perlu diatasi: diperlukan mekanisme optimasi yang mampu mengarahkan BPNN menemukan konfigurasi hyperparameter yang lebih optimal secara sistematis.

Penelitian ini mengusulkan hibridasi antara Particle Swarm Optimization (PSO) dan BPNN. PSO merupakan algoritma optimasi berbasis populasi yang terinspirasi dari perilaku kawanan burung atau ikan dalam mencari sumber makanan secara kolektif. Setiap partikel dalam PSO merepresentasikan satu kandidat solusi, yaitu satu konfigurasi hyperparameter BPNN, dan bergerak dalam ruang pencarian dengan mempertimbangkan pengalaman terbaik dirinya sendiri (pBest) dan pengalaman terbaik seluruh kawanan (gBest). Dengan cara ini, PSO memungkinkan pencarian hyperparameter BPNN, meliputi jumlah neuron pada setiap lapisan tersembunyi, learning rate, dan dropout rate, dilakukan secara otomatis dan efisien tanpa bergantung pada nilai inisialisasi yang acak. Efektivitas hibridasi PSO-BPNN pada domain klasifikasi audio berbasis musik telah dilaporkan mampu menekan nilai loss pelatihan dibandingkan BPNN konvensional [7], namun sebagian besar penelitian klasifikasi genre musik yang ada masih mengandalkan metode tunggal seperti SVM, k-NN, atau BPNN dengan hyperparameter yang ditetapkan secara manual berdasarkan referensi, tanpa proses pencarian otomatis. Kebaruan penelitian ini terletak pada penerapan PSO-BPNN secara khusus untuk klasifikasi genre musik dengan ekstraksi fitur MFCC pada dataset GTZAN yang mencakup 10 genre dan 1.000 berkas audio, disertai analisis mendalam mengenai pengaruh masing-masing parameter kendali PSO (cognitive, social, dan inertia) terhadap kualitas hyperparameter BPNN yang dihasilkan.

Penelitian ini bertujuan untuk menentukan nilai optimal hyperparameter BPNN melalui optimasi PSO yang menghasilkan akurasi terbaik dalam klasifikasi genre musik, serta membuktikan bahwa BPNN yang dioptimasi menggunakan PSO menghasilkan akurasi klasifikasi yang lebih tinggi dibandingkan BPNN tanpa optimasi pada dataset GTZAN.

2. Metodologi Penelitian

2.1 Data Penelitian

Data yang digunakan dalam penelitian ini merupakan data sekunder berupa berkas audio berformat .wav yang diperoleh dari platform Kaggle, yaitu dataset GTZAN Genre Collection yang

pertama kali diperkenalkan oleh Tzanetakis dan Cook [12] dan telah menjadi benchmark standar dalam bidang Music Information Retrieval. Dataset ini terdiri atas 1.000 klip audio berdurasi 30 detik yang terbagi ke dalam 10 genre musik, yaitu Blues, Classical, Country, Disco, Hip Hop, Jazz, Metal, Pop, Reggae, dan Rock, dengan masing-masing genre memiliki 100 berkas audio sehingga distribusi data bersifat seimbang antarkelas. Seluruh berkas audio memiliki sampling rate sebesar 22.050 Hz.

2.2 Desain Sistem

Dataset dibagi menjadi tiga himpunan yang saling independen menggunakan rasio 70:15:15 dengan metode stratified split, yaitu 700 berkas untuk data latih, 150 berkas untuk data validasi, dan 150 berkas untuk data uji, sehingga proporsi setiap genre tetap seimbang pada masing-masing subset. Pada data latih diterapkan lebih lanjut skema 5-fold cross-validation untuk mengurangi risiko overfitting dan memastikan hasil evaluasi lebih andal, dengan hasil akhir performa ditentukan dari rata-rata kelima fold. Pembagian data dilakukan sejak awal untuk mencegah kebocoran data (data leakage), sehingga seluruh parameter preprocessing, augmentasi, dan ekstraksi fitur dipelajari (fit) secara eksklusif dari data latih, kemudian diterapkan (apply) secara identik pada data validasi dan data uji.

Tahap preprocessing terdiri atas empat proses berurutan. Pertama, audio signal standardization, yaitu penyeragaman seluruh berkas audio menjadi kanal mono dengan sampling rate 22.050 Hz sesuai spesifikasi asli dataset GTZAN. Kedua, normalisasi amplitudo menggunakan metode peak normalization, yaitu membagi setiap sampel dengan nilai amplitudo absolut tertinggi dalam rekaman sehingga setiap berkas memiliki puncak amplitudo sebesar 1,0. Ketiga, padding dan truncation untuk menyeragamkan panjang sinyal, dengan truncation memotong sinyal yang lebih panjang dari target dan zero-padding menambahkan nol pada sinyal yang lebih pendek. Keempat, segmentasi audio, yaitu setiap berkas berdurasi 30 detik dipotong menjadi segmen-segmen berdurasi 6 detik menggunakan teknik overlapping segmentation dengan tingkat tumpang tindih 50% (3 detik), menghasilkan 9 segmen per berkas audio sehingga 700 berkas data latih menghasilkan 6.300 segmen sebelum augmentasi.

Augmentasi data diterapkan secara eksklusif pada data latih untuk memperbanyak variasi sampel tanpa mengumpulkan data baru, menggunakan tiga teknik yaitu pitch shifting sebesar ± 1 semitone, time stretching dengan faktor peregangan 0,95–1,05, dan SpecAugment yang diterapkan langsung pada matriks fitur MFCC berupa time masking (lebar maksimum 25 frame) dan frequency masking (1 mask, lebar maksimum 5).

Ekstraksi fitur menggunakan Mel-Frequency Cepstral Coefficients (MFCC) sebagai satu-satunya representasi fitur. Proses ekstraksi diawali dengan zero-padding pada kedua ujung sinyal, dilanjutkan dengan framing berukuran 2.048 sampel dan hop length 256 sampel, kemudian ditransformasikan ke domain frekuensi menggunakan Fast Fourier Transform (FFT) berukuran $N_{FFT} = 2.048$ untuk menghasilkan power spectrogram. Power spectrogram selanjutnya dipetakan ke skala mel menggunakan filter bank mel dengan 128 filter segitiga, dilogaritmakan, kemudian diterapkan Discrete Cosine Transform (DCT) untuk menghasilkan 40 koefisien MFCC per frame. Turunan pertama (Delta MFCC) dan turunan kedua (Delta-Delta MFCC) turut dihitung untuk memperkaya representasi fitur, lalu seluruh vektor fitur dinormalisasi menggunakan z-score normalization dengan parameter mean dan standar deviasi yang dihitung eksklusif dari data latih.

Arsitektur metode merupakan kombinasi dua algoritma: Particle Swarm Optimization (PSO) sebagai mekanisme pencari hyperparameter optimal, dan Backpropagation Neural Network (BPNN) sebagai model klasifikasi utama. Arsitektur BPNN terdiri atas lapisan masukan yang menerima vektor fitur MFCC, tiga lapisan tersembunyi dengan jumlah neuron h_1 , h_2 , dan h_3 , serta lapisan keluaran dengan 10 unit yang merepresentasikan 10 kelas genre. Setiap lapisan tersembunyi menggunakan fungsi aktivasi GELU dan dropout dengan laju yang mengecil secara bertahap, sementara bobot jaringan diinisialisasi menggunakan metode Kaiming Normal. Pada blok PSO, setiap partikel merepresentasikan satu kandidat konfigurasi hyperparameter BPNN dalam ruang pencarian 5 dimensi, yaitu jumlah neuron pada tiga lapisan tersembunyi (h_1 , h_2 , h_3), learning rate dalam skala logaritmik, dan dropout rate. Fungsi fitness setiap partikel dihitung dengan melatih kandidat konfigurasi selama 10 epoch proksi lalu mengukur validation loss yang dihasilkan; semakin kecil validation loss, semakin baik kualitas partikel tersebut.

Proses optimasi PSO dilaksanakan dalam dua tahap. Tahap pertama adalah grid search terhadap parameter kendali PSO (c_1 , c_2 , w) mengacu pada syarat konvergensi Clerc dan Kennedy, yaitu $c_1 + c_2 < 4$ dan $w < 1$, dengan $c_1 \in \{1,2; 1,5; 1,8\}$, $c_2 \in \{1,2; 1,5; 1,8\}$, dan $w \in \{0,4; 0,7; 0,9\}$, menghasilkan 7 kombinasi unik yang dievaluasi menggunakan skema 5-fold cross-validation [11]. Tahap kedua adalah PSO final yang menggunakan kombinasi (c_1 , c_2 , w) terbaik dari grid search untuk mencari hyperparameter BPNN optimal (h_1 , h_2 , h_3 , learning rate, dropout) secara penuh, menggunakan 30 partikel dan 50 iterasi. Konfigurasi hyperparameter terbaik (gBest) yang ditemukan selanjutnya digunakan untuk membangun dan melatih model BPNN final menggunakan CrossEntropyLoss dengan label smoothing 0,05 dan optimizer Adam.

2.3 Perancangan Implementasi

Pengembangan dan pelatihan model dilakukan menggunakan bahasa pemrograman Python versi 3.11.0 pada dua lingkungan komputasi secara bertahap: eksplorasi awal dan pelatihan model di Google Colaboratory yang menyediakan akses GPU, serta pengembangan antarmuka dan integrasi model secara lokal menggunakan Visual Studio Code. Seluruh tahap pemrosesan sinyal audio, meliputi preprocessing, augmentasi, dan ekstraksi fitur MFCC beserta turunannya, dibangun menggunakan library Librosa, dengan bantuan SciPy untuk penghitungan statistik agregasi, NumPy untuk manipulasi data numerik, dan Pandas untuk manipulasi tabel data. Arsitektur BPNN diimplementasikan menggunakan framework PyTorch melalui modul torch.nn, torch.optim, serta TensorDataset dan DataLoader dari torch.utils.data, dan diekspor dalam format .pth. Algoritma PSO untuk optimasi hyperparameter diimplementasikan menggunakan library PySwarms yang menyediakan implementasi PSO Global-Best terstandarisasi dan kompatibel dengan ekosistem NumPy. Pembagian dataset menggunakan train_test_split dari scikit-learn, sementara evaluasi performa model (precision, recall, F1-score, confusion matrix) menggunakan modul sklearn.metrics, dan visualisasi kurva pelatihan menggunakan Matplotlib serta Seaborn. Model BPNN-PSO yang telah dilatih selanjutnya diintegrasikan ke dalam prototipe aplikasi web menggunakan framework Streamlit.

2.4 Desain Evaluasi Sistem

Pengujian metode dibagi menjadi dua tahapan berurutan. Tahapan pertama adalah pengujian optimasi hyperparameter PSO yang bertujuan menentukan kombinasi parameter kendali c_1 , c_2 , dan w yang menghasilkan validation loss rata-rata terkecil, dengan ketiga parameter divariasikan satu per satu menggunakan pendekatan one-factor-at-a-time sementara dua parameter lainnya ditetapkan sebagai variabel kontrol. Variabel kontrol lain yang ditetapkan konstan meliputi jumlah partikel sebesar 30, jumlah iterasi PSO sebesar 50, jumlah epoch proksi sebesar 10, dan ukuran batch sebesar 64. Pengujian tahap ini dilakukan pada data latih menggunakan skema 5-fold cross-validation.

Tahapan kedua adalah pengujian ukuran evaluasi model klasifikasi, yang dilakukan setelah model BPNN final dilatih menggunakan hyperparameter optimal hasil PSO. Evaluasi dilakukan pada data uji yang sejak awal diisolasi dan tidak pernah dilihat model selama pelatihan maupun pencarian hyperparameter, direpresentasikan dalam bentuk confusion matrix berukuran 10×10 yang menjadi dasar penghitungan metrik accuracy, precision, recall, dan F1-score untuk dua skenario perbandingan, yaitu model BPNN-PSO dan model BPNN tanpa optimasi PSO. Model BPNN tanpa optimasi menggunakan hyperparameter yang ditetapkan berdasarkan referensi penelitian terdahulu sebagai baseline pembandingan yang wajar dan dapat dipertanggungjawabkan secara akademis, yaitu arsitektur piramida $h_1 = 256$, $h_2 = 128$, $h_3 = 64$ [8], learning rate sebesar 1×10^{-3} sesuai rekomendasi default algoritma Adam [9], dan dropout rate sebesar 0,5 [10].

3. Hasil dan Pembahasan

3.1. Pengaruh Perubahan Parameter

Pengujian pengaruh parameter kendali PSO dilakukan untuk menentukan kombinasi nilai cognitive coefficient (c_1), social coefficient (c_2), dan inertia weight (w) yang menghasilkan rata-

rata validation loss paling rendah selama proses optimasi hyperparameter BPNN. Ketiga parameter diuji menggunakan pendekatan one-factor-at-a-time, yaitu hanya satu parameter yang divariasikan pada setiap eksperimen sementara dua parameter lainnya ditetapkan sebagai variabel kontrol, menghasilkan tujuh kombinasi parameter yang masing-masing dievaluasi menggunakan skema 5-fold cross-validation pada data latih. Ringkasan hasil ketujuh kombinasi tersebut disajikan pada Tabel 1.

Tabel 1. Hasil dari kombinasi parameter

c1	c2	w	Fold-1	Fold-2	Fold-3	Fold-4	Fold-5
1.2	1.5	0.7	1.026051	1.081089	0.946287	1.106751	1.066217
1.5	1.5	0.9	1.036673	1.082529	0.951088	1.087258	1.063954
1.5	1.5	0.4	1.037256	1.096221	0.952829	1.079758	1.067190
1.5	1.2	0.7	1.037608	1.089907	0.950289	1.090869	1.047657
1.5	1.5	0.7	1.040355	1.097346	0.949519	1.104492	1.056042
1.8	1.5	0.7	1.042325	1.052149	0.954977	1.104505	1.062837
1.5	1.8	0.7	1.046561	1.084494	0.964770	1.070945	1.063300

Pengujian parameter cognitive (c1) dengan c2 dan w ditetapkan tetap sebesar 1,5 dan 0,7 menunjukkan bahwa nilai c1 = 1,2 menghasilkan rata-rata validation loss terendah sebesar 0,9463, diikuti c1 = 1,5 sebesar 0,9495, dan c1 = 1,8 sebesar 0,9550. Pola ini mengindikasikan bahwa nilai c1 yang terlalu besar membuat partikel terlalu berfokus pada pengalaman lokalnya sendiri (pBest) sehingga kemampuan eksplorasi global populasi menjadi terbatas, sementara nilai c1 yang lebih kecil memberikan keseimbangan yang lebih baik antara eksplorasi lokal dan global dalam ruang pencarian hyperparameter BPNN.

Pengujian parameter social (c2) dengan c1 dan w ditetapkan tetap sebesar 1,5 dan 0,7 menunjukkan bahwa nilai c2 = 1,5 menghasilkan rata-rata validation loss terendah sebesar 0,9495, sedikit lebih rendah dibandingkan c2 = 1,2 sebesar 0,9503, dan jauh lebih rendah dibandingkan c2 = 1,8 sebesar 0,9648. Tren ini menunjukkan bahwa pengaruh sosial yang terlalu kuat menyebabkan partikel terlalu cepat berkonvergensi menuju satu titik gBest tanpa cukup mengeksplorasi ruang pencarian hyperparameter secara menyeluruh, sehingga nilai c2 yang moderat memberikan keseimbangan terbaik antara konvergensi dan eksplorasi.

Pengujian parameter inertia (w) dengan c1 dan c2 ditetapkan tetap sebesar 1,5 menunjukkan bahwa nilai w = 0,7 menghasilkan rata-rata validation loss terendah sebesar 0,9495, sementara w = 0,4 dan w = 0,9 masing-masing menghasilkan 0,9528 dan 0,9511. Bobot inersia yang moderat memberikan kemampuan eksplorasi yang cukup luas kepada partikel sehingga populasi tidak terlalu cepat terkonvergensi pada satu wilayah ruang pencarian, namun tetap mampu menetap pada solusi yang optimal dalam jumlah iterasi yang tersedia.

Berdasarkan ketiga pengujian tersebut, kombinasi parameter optimal yang ditetapkan untuk optimasi hyperparameter BPNN adalah c1 = 1,2, c2 = 1,5, dan w = 0,7, yang menghasilkan rata-rata validation loss terendah sebesar 0,9463 dari kelima fold. Menggunakan konfigurasi ini, PSO menemukan hyperparameter BPNN optimal berupa jumlah neuron pada lapisan tersembunyi pertama (h1) sebesar 192, lapisan tersembunyi kedua (h2) sebesar 192, lapisan tersembunyi ketiga (h3) sebesar 32, learning rate sebesar 0,000185, dan dropout rate sebesar 0,379. Konfigurasi ini mencerminkan karakteristik PSO yang lebih mengutamakan eksplorasi global dengan bobot inersia moderat dan pengaruh kognitif yang relatif kecil, terbukti lebih efektif dalam menjelajahi ruang pencarian hyperparameter BPNN berdimensi tinggi dibandingkan konfigurasi yang lebih berorientasi pada eksploitasi lokal.

3.2. Perbandingan Model BPNN-PSO dan BPNN

Model BPNN-PSO dievaluasi menggunakan konfigurasi hyperparameter hasil optimasi PSO (h1 = 192, h2 = 192, h3 = 32, learning rate = 0,000185, dropout rate = 0,379), lalu dibandingkan dengan model BPNN tanpa optimasi yang hyperparameternya ditetapkan berdasarkan referensi penelitian terdahulu. Evaluasi dilakukan pada data uji yang sejak awal diisolasi dan tidak pernah

dilihat model selama pelatihan maupun pencarian hyperparameter, menggunakan skema 5-fold cross-validation. Ringkasan perbandingan kedua model disajikan pada Tabel 2.

Tabel 2. Perbandingan performa model BPNN-PSO dan BPNN

Model	Accuracy	Precision	Recall	F1-Score
BPNN-PSO	84,40% ± 3,59%	85,09%	84,40%	84,34%
BPNN (tanpa optimasi)	82,80% ± 3,19%	83,60%	82,80%	82,80%

Model BPNN-PSO menghasilkan rata-rata accuracy sebesar 84,40%, lebih tinggi dibandingkan model BPNN sebesar 82,80%. Peningkatan yang sama juga teramati pada precision (85,09% berbanding 83,60%), recall (84,40% berbanding 82,80%), dan F1-score (84,34% berbanding 82,80%). Hasil ini menunjukkan bahwa optimasi PSO meningkatkan kemampuan model dalam mendeteksi sampel dari genre yang benar secara lebih merata di seluruh kelas dibandingkan penetapan hyperparameter berdasarkan referensi umum.

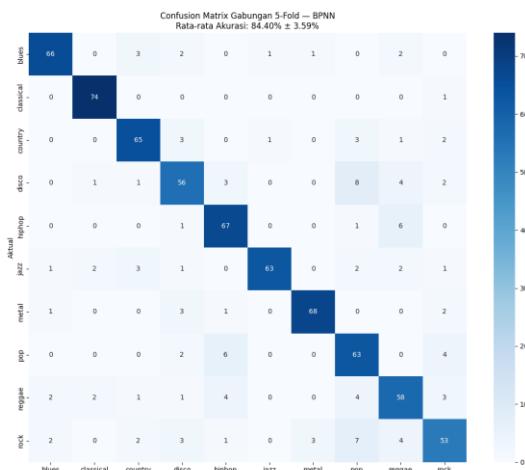
Analisis performa per kelas genre pada model BPNN-PSO menunjukkan bahwa genre jazz dan metal mencatatkan nilai precision tertinggi masing-masing sebesar 0,967 dan 0,944, yang mengindikasikan bahwa model sangat jarang salah mengklasifikasikan genre lain sebagai kedua genre tersebut. Sementara itu, berdasarkan nilai F1-score, genre classical memperoleh nilai tertinggi sebesar 0,961, diikuti oleh metal sebesar 0,925, blues sebesar 0,900, dan jazz sebesar 0,898. Sebaliknya, genre rock mencatatkan nilai F1-score terendah sebesar 0,733, diikuti oleh disco sebesar 0,761 dan reggae sebesar 0,766. Pola yang relatif serupa juga teramati pada model BPNN tanpa optimasi, dengan genre classical (F1-score 0,967) dan metal (0,917) sebagai genre dengan performa tertinggi, serta rock (0,703) dan disco (0,720) sebagai genre dengan performa terendah.

Peningkatan akurasi dari 82,80% menjadi 84,40% menunjukkan kenaikan sebesar 1,60 poin persentase setelah optimasi hyperparameter menggunakan Particle Swarm Optimization (PSO). Meskipun demikian, peningkatan tersebut relatif kecil karena selisih rata-rata akurasi lebih rendah dibandingkan standar deviasi hasil 5-fold cross-validation, yaitu 3,59% pada BPNN-PSO dan 3,19% pada BPNN. Oleh karena itu, tanpa pengujian statistik formal, seperti paired t-test atau Wilcoxon signed-rank test, belum dapat disimpulkan bahwa perbedaan performa kedua model signifikan secara statistik. Hal ini menjadi salah satu keterbatasan penelitian.

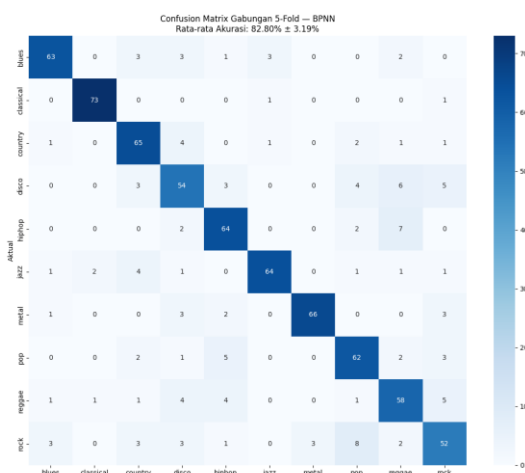
Di sisi lain, penerapan PSO meningkatkan kompleksitas komputasi karena memerlukan evaluasi beberapa kombinasi parameter kendali dan proses optimasi sebanyak 1.500 evaluasi fitness (30 partikel × 50 iterasi), sehingga membutuhkan waktu pelatihan yang lebih besar dibandingkan penentuan hyperparameter secara manual. Meskipun demikian, peningkatan yang konsisten pada seluruh metrik evaluasi menunjukkan bahwa PSO mampu mengarahkan pencarian menuju konfigurasi hyperparameter yang lebih baik secara sistematis. Oleh karena itu, penelitian ini menunjukkan potensi PSO sebagai metode optimasi hyperparameter, meskipun validasi melalui pengujian signifikansi statistik masih diperlukan pada penelitian selanjutnya.

3.3. Analisis Confusion Matrix dan Pola Kesalahan Antar-Genre

Analisis confusion matrix pada Gambar 1 dan Gambar 2 menunjukkan bahwa pada model BPNN-PSO, kesalahan klasifikasi terbesar masih terjadi pada genre rock. Sebanyak 7 sampel salah diklasifikasikan sebagai pop, diikuti 4 sampel sebagai reggae, 3 sampel sebagai disco, 3 sampel sebagai metal, serta masing-masing 2 sampel sebagai blues dan country. Genre disco juga menunjukkan tingkat kesalahan klasifikasi yang cukup tinggi, dengan 8 sampel salah diklasifikasikan sebagai pop, diikuti 4 sampel sebagai reggae dan 3 sampel sebagai hiphop. Sementara itu, genre reggae masih cukup sering diklasifikasikan sebagai hiphop sebanyak 4 sampel dan sebagai rock sebanyak 3 sampel. Sebaliknya, genre classical mencatatkan jumlah kesalahan klasifikasi paling sedikit, dengan hanya 1 sampel yang salah diklasifikasikan.



Gambar 1. Confusion Matrix BPNN-PSO



Gambar 2. Confusion Matrix BPNN tanpa Optimasi

Perbandingan pola kesalahan antara kedua model menunjukkan beberapa perbaikan pada model BPNN-PSO. Pada model BPNN, genre hiphop salah diklasifikasikan sebagai reggae sebanyak 7 sampel, sedangkan pada model BPNN-PSO jumlah tersebut menurun menjadi 6 sampel. Genre rock pada model BPNN-PSO juga menunjukkan sedikit perbaikan, dengan jumlah sampel yang salah diklasifikasikan sebagai pop menurun dari 8 sampel pada model BPNN menjadi 7 sampel. Meskipun demikian, pada pasangan genre tertentu, seperti disco dan pop, jumlah kesalahan klasifikasi masih relatif tinggi. Secara keseluruhan, hasil ini menunjukkan bahwa optimasi PSO belum mampu menghilangkan pola kesalahan klasifikasi yang dipengaruhi oleh kemiripan karakteristik akustik antar-genre, namun mampu meningkatkan performa model secara keseluruhan.

Pola kesalahan klasifikasi pada beberapa pasangan genre mengindikasikan bahwa representasi fitur yang digunakan masih memiliki keterbatasan dalam membedakan genre dengan karakteristik akustik yang serupa. Penelitian ini berfokus pada penggunaan MFCC sebagai fitur utama, sehingga informasi yang direpresentasikan terutama berupa karakteristik timbral atau amplop spektral. Sementara itu, karakteristik lain seperti pola ritmis, struktur harmonis, dan hubungan tonal belum sepenuhnya terwakili. Kondisi tersebut diduga berkontribusi terhadap masih tingginya kesalahan klasifikasi pada pasangan genre seperti *rock-pop*, *disco-pop*, dan *reggae-hiphop*, yang memiliki kemiripan karakteristik akustik.

Dengan demikian, keterbatasan representasi fitur MFCC menjadi salah satu batasan penelitian ini. Meskipun optimasi hyperparameter menggunakan PSO mampu meningkatkan performa model secara keseluruhan, optimasi tersebut belum dapat mengatasi keterbatasan informasi yang terkandung pada fitur masukan. Oleh karena itu, penelitian selanjutnya dapat

mempertimbangkan penggunaan kombinasi MFCC dengan fitur lain, seperti Chroma Feature, Spectral Contrast, Tonnetz, atau *tempogram*, untuk menghasilkan representasi karakteristik musik yang lebih komprehensif.

3.4. Implementasi Web

Sebagai tahap validasi praktis, model BPNN-PSO dengan konfigurasi optimal diintegrasikan ke dalam prototipe aplikasi web berbasis Streamlit yang memungkinkan pengguna mengunggah berkas audio berformat WAV maupun MP3 untuk memperoleh prediksi genre beserta nilai confidence score secara langsung. Aplikasi juga menampilkan tahapan preprocessing dan ekstraksi fitur MFCC yang dilakukan oleh sistem sehingga proses klasifikasi dapat diamati secara lebih transparan.

4. Kesimpulan

Berdasarkan hasil penelitian yang telah dilakukan, dapat ditarik beberapa kesimpulan. Pertama, optimasi hyperparameter Backpropagation Neural Network (BPNN) menggunakan Particle Swarm Optimization (PSO) dengan konfigurasi parameter kendali $c1 = 1,2$, $c2 = 1,5$, dan $w = 0,7$ menghasilkan konfigurasi optimal berupa jumlah neuron pada lapisan tersembunyi pertama ($h1$) sebanyak 192, lapisan tersembunyi kedua ($h2$) sebanyak 192, lapisan tersembunyi ketiga ($h3$) sebanyak 32, learning rate sebesar 0,000185, dan dropout rate sebesar 0,379, dengan rata-rata validation loss terendah sebesar 0,9463 berdasarkan evaluasi menggunakan 5-fold cross-validation.

Kedua, model BPNN-PSO yang dibangun menggunakan ekstraksi fitur MFCC menghasilkan rata-rata accuracy sebesar 84,40%, precision sebesar 85,09%, recall sebesar 84,40%, dan F1-score sebesar 84,34%, lebih tinggi dibandingkan model BPNN tanpa optimasi. Hasil evaluasi per kelas menunjukkan bahwa genre classical memiliki performa klasifikasi terbaik, sedangkan genre rock, disco, dan reggae memiliki performa yang relatif lebih rendah. Temuan ini menunjukkan bahwa penggunaan MFCC sebagai fitur utama masih memiliki keterbatasan dalam merepresentasikan seluruh karakteristik musik, sehingga menjadi salah satu batasan dalam penelitian ini.

Ketiga, optimasi hyperparameter menggunakan PSO mampu meningkatkan performa model BPNN secara konsisten pada seluruh metrik evaluasi, dengan peningkatan accuracy sebesar 1,60 poin persentase dibandingkan BPNN tanpa optimasi. Namun, peningkatan tersebut relatif kecil dan signifikansi statistiknya belum dapat dipastikan karena belum dilakukan pengujian statistik formal. Selain itu, penerapan PSO meningkatkan kompleksitas komputasi akibat proses optimasi hyperparameter yang lebih intensif. Oleh karena itu, PSO menunjukkan potensi sebagai metode optimasi hyperparameter, meskipun efektivitasnya masih perlu dikaji lebih lanjut melalui pengujian signifikansi statistik dan evaluasi terhadap biaya komputasi yang diperlukan.

Berdasarkan hasil tersebut, penelitian selanjutnya disarankan untuk membandingkan PSO dengan metode optimasi hyperparameter lain, seperti Genetic Algorithm atau Bayesian Optimization, serta melakukan pengujian signifikansi statistik formal terhadap performa antar-model. Selain itu, representasi fitur dapat ditingkatkan dengan mengombinasikan MFCC dengan fitur audio lain, seperti Chroma Feature, Spectral Contrast, Tonnetz, atau fitur berbasis tempogram, sehingga karakteristik harmonis, tonal, dan ritmis musik dapat direpresentasikan secara lebih komprehensif.

Referensi

- [1] R. S. Bhat, R. B. R, and M. K. R, "Music genre classification," International Journal of Communication and Media Science, vol. 7, no. 1, pp. 8–13, 2020, doi: 10.14445/2349641x/ijcms-v7i1p102.
- [2] G. Z. Dhamara and Sucipto, "Klasifikasi genre musik menggunakan machine learning," Bulletin of Information Technology (BIT), vol. 6, no. 3, pp. 206–217, 2025, doi: 10.47065/bit.v6i3.2021.

- [3] E. V. Epure, A. Khlif, and R. Hennequin, "Leveraging knowledge bases and parallel annotations for music genre translation," arXiv preprint arXiv:1907.08698, Jul. 2019. [Online]. Available: <https://arxiv.org/abs/1907.08698>
- [4] Z. Kh. Abdul and A. K. Al-Talabani, "Mel Frequency Cepstral Coefficient and its applications: A review," IEEE Access, vol. 10, pp. 122136–122158, 2022, doi: 10.1109/access.2022.3223444.
- [5] G. Kour and N. Mehan, "Music genre classification using MFCC, SVM and BPNN," International Journal of Computer Applications, vol. 112, no. 6, pp. 12–14, 2015.
- [6] R. Refianti and F. Mahardi, "Comparison of music genre classification results using multilayer perceptron with chroma feature and mel frequency cepstral coefficients extraction features," International Journal of Engineering, Science and Information Technology, vol. 3, no. 2, pp. 53–59, 2023, doi: 10.52088/ijesty.v3i2.444.
- [7] Y. Liu, B. Wang, and Y. Zang, "Music genre classification of BP neural network based on particle swarm optimization," in 2021 6th International Symposium on Computer and Information Processing Technology (ISCIPIT), 2021, pp. 78–82, doi: 10.1109/iscipt53667.2021.00022.
- [8] S. K. Mahanta, A. F. U. Rahman Khilji, and P. Pakray, "Deep neural network for musical instrument recognition using MFCCs," Computación y Sistemas, vol. 25, no. 2, 2021, doi: 10.13053/cys-25-2-3946.
- [9] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," arXiv preprint arXiv:1412.6980, Dec. 2014. [Online]. Available: <https://arxiv.org/abs/1412.6980>
- [10] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," Journal of Machine Learning Research, vol. 15, no. 56, pp. 1929–1958, 2014.
- [11] A. Rezaee Jordehi and J. Jasni, "Parameter selection in particle swarm optimisation: A survey," Journal of Experimental & Theoretical Artificial Intelligence, vol. 25, no. 4, pp. 527–542, 2013, doi: 10.1080/0952813x.2013.782348.
- [12] G. Tzanetakis and P. Cook, "Musical genre classification of audio signals," IEEE Transactions on Speech and Audio Processing, vol. 10, no. 5, pp. 293–302, 2002, doi: 10.1109/tsa.2002.800560.

