

# Extração e Seleção de Características para Classificação de Gêneros Musicais

Matheus Augusto Monte Silva, Antônio Apolinário Barbosa, Mariana Melo dos Santos

Engenharia da Computação

Centro de Informática - UFPE

Email: mams4@cin.ufpe.br, aab2@cin.ufpe.br, mms11@cin.ufpe.br

**Abstract**—Este trabalho apresenta a extração e seleção de características de sinais de áudio para a classificação automática de músicas por gênero musical. Utilizando o dataset GTZAN, foi aplicada a técnica de Random Forest para identificar as características mais relevantes. Após um pré-processamento envolvendo a segmentação das faixas de áudio, foram extraídas características dos sinais, e o modelo Random Forest foi treinado e avaliado. O modelo atingiu uma acurácia de 72,26% sem perturbações no sinal e 70,88% quando ruído e eco foram adicionados. Além disso, foram analisadas as características mais influentes para a predição, permitindo um melhor entendimento dos padrões acústicos.

**Index Terms**—Processamento de áudio, Classificação de Gênero Musical, Machine Learning, Random Forest.

## I. INTRODUÇÃO

A classificação automática de músicas por gênero é um problema relevante no campo do processamento de sinais de áudio e aprendizado de máquina. Neste trabalho, utilizamos o dataset GTZAN para a extração e seleção de características relevantes para essa tarefa, empregando a técnica Random Forest. A abordagem incluiu segmentação de áudio, seleção de características e treinamento de um modelo de aprendizado de máquina.

## II. METODOLOGIA

### A. Base de Dados

Foi utilizado o dataset GTZAN, que contém 1000 faixas de áudio distribuídas igualmente entre 10 gêneros musicais.

### B. Pré-processamento

Cada faixa de áudio (30 segundos) foi dividida em segmentos de 3 segundos, e foram selecionados aleatoriamente 400 segmentos de cada gênero, totalizando 4000 amostras. Para garantir a preservação de informações relevantes e evitar perdas entre segmentos, foi utilizada a técnica de sobreposição de janela (overlap), permitindo maior continuidade temporal e melhor captura das características acústicas.

Além disso, as características extraídas foram representadas utilizando a escala de frequência Mel, através do Mel Spectrograma, que aproxima a percepção auditiva humana. Essa técnica reduz a dimensionalidade dos dados e enfatiza frequências mais relevantes para a distinção entre gêneros musicais.

### C. Seleção de Características

A seleção das características foi realizada utilizando um arquivo CSV do próprio dataset GTZAN, contendo diversas características previamente extraídas. O algoritmo Random Forest foi aplicado para gerar um ranking de importância das características para a predição.

A Tabela I apresenta as dez características mais relevantes identificadas pelo modelo Random Forest.

TABLE I  
CARACTERÍSTICAS MAIS RELEVANTES SEGUNDO O RANDOM FOREST

Característica	Importância
perceptr_var	0.049480
chroma_stft_mean	0.036692
rms_var	0.036440
rms_mean	0.032749
mfcc4_mean	0.030803
rolloff_mean	0.029650
spectral_bandwidth_mean	0.029431
perceptr_mean	0.027238
mfcc1_mean	0.025630
harmony_var	0.025062

### D. Extração de Características

A extração de características foi realizada para capturar informações representativas da estrutura harmônica e espectral dos sinais de áudio. Foram utilizadas métricas como:

- **Energia perceptual (perceptr\_var e perceptr\_mean):** Mede a distribuição de energia ao longo do espectro de frequência, permitindo identificar padrões de intensidade sonora característicos de cada gênero musical.
- **Coefficientes Cepstrais de Frequência Mel (MFCCs):** Esses coeficientes extraem representações da envoltória espectral do sinal, sendo amplamente utilizados em tarefas de classificação de áudio.
- **Espectro de Fourier (fft\_mean e fft\_var):** A Transformada de Fourier permite decompor o sinal em suas componentes espectrais, auxiliando na distinção entre gêneros musicais que possuem padrões rítmicos e harmônicos distintos.

O uso dessas características melhora a capacidade do modelo de capturar padrões específicos de diferentes gêneros musicais, resultando em uma classificação mais eficiente.

### E. Avaliação com Matrizes de Confusão

Para melhor visualização do desempenho do modelo, foram geradas matrizes de confusão comparando os resultados sem ruído e com ruído. As Figuras 1 e 2 mostram essas matrizes.

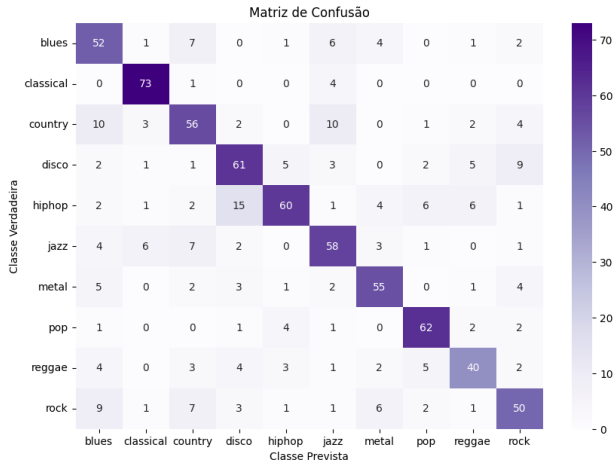


Fig. 1. Matriz de confusão do modelo sem ruído.

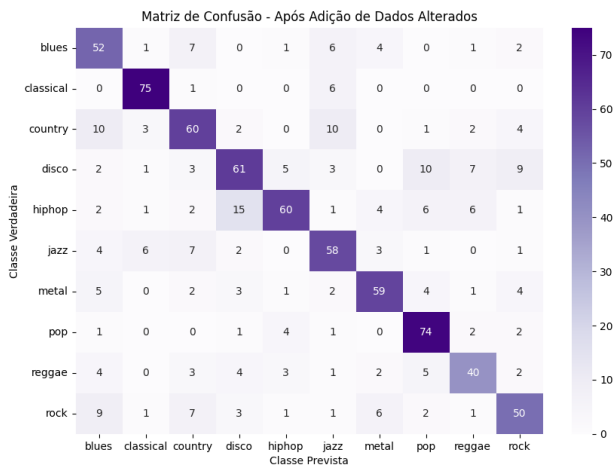


Fig. 2. Matriz de confusão do modelo com ruído.

### F. Treinamento e Avaliação do Modelo

Utilizou-se o modelo Random Forest para a classificação das amostras, com uma divisão de 80% para treinamento e 20% para teste. A acurácia obtida foi de 72,26%.

### G. Teste com Ruído e Eco

Para avaliar a robustez do modelo, foram adicionados ruído e eco a 10 amostras de áudio do conjunto de teste. Após a extração de características e nova avaliação com o Random Forest, a acurácia obtida foi de 70,88%.

## III. CONCLUSÃO

O estudo demonstrou a viabilidade da classificação de gêneros musicais a partir da extração e seleção de características de sinais de áudio. A técnica Random Forest apresentou um desempenho satisfatório, mesmo com a adição

de perturbações no sinal. A identificação das características mais relevantes permitiu um melhor entendimento dos aspectos que influenciam a predição do modelo. Além disso, o uso de pré-processamento, incluindo a separação por janelas, sobreposição de janelas e a escala de frequência Mel, foi essencial para melhorar a captura de padrões musicais. Futuras melhorias podem incluir otimizações no tempo de processamento e a exploração de outras abordagens de aprendizado de máquina.