

Reconhecendo gêneros musicais brasileiros com técnicas de aprendizagem de máquina supervisionada

Júlia Luiza S. L. da Conceição¹, Rosiane de Freitas¹, Bruno Gadelha¹
João Gustavo Kienen², Sérgio Anders², Brendo Cavalcante²

¹Instituto de Computação – Universidade Federal do Amazonas (UFAM)
69077-000 – Manaus-AM, Brasil

²Faculdade de Artes – Universidade Federal do Amazonas (UFAM)
69077-000 – Manaus-AM, Brasil

{jlslc, rosiane, bruno}@icomp.ufam.edu.br,
{gustavokienen, sergioandersct}@ufam.edu.br
brendo-cavalcante@protonmail.com

Abstract. *In this work, an initial study on the automatic recognition of the main Brazilian musical genres is presented: Axé, Forró, MPB, Rock, Samba, and Sertanejo. Through the extraction of representative musical characteristics, automatic classification experiments were performed using the Weka tool and classical supervised learning algorithms. An analysis of the main available databases was also carried out: GTZAN, FMA, AudioSet, RWC, ISMIR, Magnatune, and LMD. There is a scarcity of cultural diversity on these bases, most of which concentrate globally more popular styles such as Pop and Rock, reinforcing the need to include more diverse and culturally identifiable genres, such as Brazilians.*

Resumo. *Neste trabalho, é apresentado um estudo sobre o reconhecimento automático dos principais gêneros musicais brasileiros: Axé, Forró, MPB, Rock, Samba e Sertanejo. Para tanto, foram realizados experimentos de classificação automática utilizando a ferramenta Weka e algoritmos clássicos de aprendizagem supervisionada. Foi também realizada uma análise das principais bases de dados disponíveis: GTZAN, FMA, AudioSet, RWC, ISMIR, Magnature e LMD. Como resultado, percebeu-se a escassez de diversidade cultural nestas bases, em sua maioria concentrando estilos globalmente populares como Pop e Rock, reforçando a necessidade de incluir gêneros mais diversos e culturalmente identitários, como os brasileiros.*

1. Introdução

O novo contexto da Distribuição de Música Eletrônica e a exploração sistemática de grandes bancos de dados musicais cria a necessidade de produzir descrições simbólicas dos títulos musicais. Neste cenário, os gêneros musicais possuem descrições fundamentais para o desenvolvimento de soluções para tal propósito. Assim como são utilizados comumente na organização de coleções musicais, representando a interação intercultural, os gêneros musicais sintetizam propriedades e características comuns entre as peças musicais. O gênero da música é possivelmente a melhor informação em um âmbito geral para o conhecimento do conteúdo musical [Aucouturier and Pachet 2003].

A música não é utilizada apenas para o divertimento, podendo ser utilizada também na formação da personalidade, e como auxílio para a cura de algumas doenças e alteração do humor [BRÉSCIA 2009]. Alguns especialistas acreditam que a música pode ajudar em certos tratamentos, como a esclerose múltipla, ansiedade e depressão. Com o uso desse método, o paciente pode ter estímulos motores, sensoriais e até mesmo reflexos na condição emocional em que se encontra de acordo com o estilo de música que está escutando [Moreira et al. 2009].

Em face do exposto, pode-se notar a importância de se classificar músicas de acordo com a emoção expressa pela mesma, para agrupá-las de acordo com a sua emoção comum e utilizá-la para o entretenimento ou mesmo em tratamentos. Além disso, com o surgimento de lojas virtuais que vendem músicas através da internet, como o *iTunes*, e empresas que fornecem fluxo de mídia via internet (*streaming*), *Spotify* por exemplo, a análise dos gêneros musicais pode ser de grande relevância para sugerir conteúdo com base nos interesses do usuário.

Conjuntos de dados e fontes de dados usados na pesquisa do *MGC (Music Genre Recognition)* não compartilham uma estrutura, apresentando diferentes classes e um número variável delas [Ramírez and Flores 2019]. Porém, ao observar os gêneros contidos nas bases de dados disponíveis atualmente percebe-se que a grande maioria concentra-se nos gêneros mundialmente mais populares, resumindo a gama de estilos musicais existentes geralmente em: *Classical, Electronic, Jazz/Blues, Metal/Punk, Rock/Pop and World*, como na base de dados ISMIR2004, uma das mais populares em pesquisas de MGR, de acordo com [Sturm 2012]. Portanto, torna-se importante a exploração e o uso de outros gêneros musicais no desafio de classificação automática, como por exemplos os brasileiros. Assim, propõe-se a contribuição com o processo de reconhecimento automático de gêneros musicais abordando gêneros brasileiros, fornecendo um estudo inicial dos principais, sendo estes: Axé, Forró, MPB (Música Popular Brasileira), Rock, Samba e Sertanejo.

O restante deste artigo está organizado como segue. Na Seção 2 é abordada a fundamentação teórica. Na Seção 3, aborda-se o tema reconhecimento de gêneros musicais brasileiros, incluindo trabalhos relacionados. Na Seção 4, há uma descrição sobre a abordagem inicial de reconhecimento dos estilos musicais brasileiros, contendo a organização da base para experimentação e a extração das características musicais utilizadas. Na Seção 5, são apresentados os resultados e uma análise comparativa dos mesmos. Finalmente, na Seção 6 são feitas as considerações finais, onde são descritos os trabalhos futuros e a conclusão.

2. Fundamentação teórica

A classificação automática de gênero musical é uma das aplicações da inteligência computacional, envolvendo heurísticas e aprendizagem de máquina, que possibilita o desenvolvimento de técnicas para o reconhecimento do gênero ao qual pertence uma música. Para isso, é necessária a extração de características relevantes do áudio, uma etapa muito importante em muitos aplicativos de recuperação de informações de música (MIR), e a escolha dessas características tem um grande impacto em seu desempenho. Nas últimas décadas, vários recursos foram introduzidos e aplicados com sucesso a muitos tipos diferentes de sistemas MIR. Esses recursos de áudio podem ser categorizados em dois grupos:

1) recursos espectrais e 2) recursos temporais.

Características espectrais ou em inglês *Spectral Features* (SFs) representam traços espectrais da música em um período relativamente curto de tempo. Em um sentido musical, pode-se dizer que revela o timbre ou as características tonais da música. Alguns dos SFs mais populares incluem: centroide espectral, espalhamento espectral, fluxo espectral, medida de nivelamento espectral, coeficiente cepstral de frequências mel, em inglês *Mel-Frequency Cepstral Coefficients* (MFCCs) e *Chroma*. Por outro lado, características temporais, em inglês *Temporal Features* (TFs) descrevem a dinâmica de longo prazo de um sinal de música ao longo do tempo, como transição temporal ou características rítmicas. Estes incluem taxa de cruzamento zero, em inglês *Zero-Crossing Rate* (ZCR), envelope temporal, histograma de andamento e assim por diante. No entanto, os dois grupos não são mutuamente exclusivos e muitos aplicativos MIR usam uma combinação de muitos recursos diferentes [Jeong and Lee 2016]. Neste artigo, são utilizadas as características musicais descritas na Tabela 1.

Tabela 1. Características musicais aplicadas neste trabalho.

Característica	Tipo	Descrição
<i>Tempo</i>	TF	O ritmo ou velocidade na qual uma seção da música é tocada.
<i>Beats</i>	TF	A divisão unitária do tempo musical.
<i>Chroma STFT</i>	SF	Transformada de Fourier de curta duração.
<i>Spectral Centroid</i>	SF	É o ponto balanceado do espectro.
<i>Spectral Bandwidth</i>	SF	Calcula a largura de banda espectral.
<i>Roll-off</i>	SF	É a frequência abaixo da qual uma porcentagem especificada da energia espectral total. Ex: 85%.
<i>Zero-Crossing Rate</i>	TF	A taxa de alteração de sinal ao longo de um sinal.
<i>Mel-Frequency Sepstrum (MFCC)</i>	SF	O cepstro de frequência de mel é uma representação do espectro de potência de curto prazo de um som.

No campo de Aprendizagem de Máquina (do inglês, *Machine Learning* - ML), os algoritmos podem ser classificados em certos tipos de acordo com sua abordagem, sendo os principais e mais utilizados os algoritmos supervisionados e os não supervisionados. Algoritmos de Aprendizagem de Máquina podem ser aplicados com sucesso a problemas de MIR [Ramírez and Flores 2019]. Neste trabalho, foram utilizados exclusivamente algoritmos supervisionados, que caracterizam-se por possuírem dados de treinamento previamente rotulados, sendo os rótulos neste caso os gêneros musicais. Dentre os algoritmos supervisionados mais populares, foram selecionados os seguintes: KNN (K-Vizinhos Mais Próximos), SVM (Máquina de Vetores de Suporte) e Floresta Aleatória (do inglês, *Random Forest*).

KNN (K-Vizinhos Mais Próximos) é um algoritmo simples que armazena todos os casos disponíveis e classifica os novos casos com base em uma medida de similaridade (por exemplo, funções de distância) [Naik and Samant 2016]. Geralmente, realiza-se o cálculo da distância entre os pontos em um plano através da distância euclidiana. Por

outro lado, o SVM (Máquina de Vetores de Suporte) é principalmente um método que executa tarefas de classificação construindo hiperplanos em um espaço multidimensional que separa casos rotulados em classes diferentes com base na sua distribuição no espaço [Rajvanshi and Chowdhary 2017].

Florestas aleatórias são um esquema proposto por Leo Breiman nos anos 2000 para a construção de um conjunto de preditores baseados em árvores de decisão que crescem em subespaços de dados selecionados aleatoriamente. Formalmente, uma floresta aleatória é um preditor que consiste em uma coleção de árvores de regressão básica aleatória. Produzem previsões altamente precisas e podem lidar com um número muito grande de variáveis de entrada sem *overfitting* [Biau 2012], o que ocorre quando um modelo se ajusta muito bem ao conjunto de dados observado na fase de treinamento, mas mostra-se ineficaz para prever novos resultados.

O *Weka* é um projeto que dispõe de uma coleção de algoritmos de *Machine Learning* e *Data Mining* escritos em *Java* na Universidade de Waikato, Nova Zelândia. Ele permite que os usuários testem e comparem rapidamente diferentes métodos de aprendizado de máquina em novos conjuntos de dados [Hall et al. 2009]. Por sua interface intuitiva e sua facilidade de uso, a mesma foi utilizada para proceder os experimentos descritos neste artigo.

3. Reconhecimento de gêneros musicais

Existe muitos trabalhos sobre Reconhecimento de Gêneros Musicais (MGR), baseados em gravações de áudio, dados simbólicos e outras modalidades [Sturm 2012]. Porém, definir o gênero de uma música envolve reconhecer não apenas características rítmicas, como o seu timbre, mas também, deve-se levar em conta que gêneros são rótulos inventados pelo ser humano, cuja visão é subjetiva. Não há uma definição clara e formal do que é gênero [Ramírez and Flores 2019], deste modo, desenvolver uma tecnologia capaz de determinar o gênero de determinada música torna-se uma tarefa desafiadora.

Quanto ao reconhecimento de gêneros musicais brasileiros especificamente, poucos dos trabalhos publicados na área de *MGC* fazem uso de bases de dados que incluem gêneros populares brasileiros em seus experimentos de classificação, como MPB e Axé. Este fenômeno acontece devido a falta de representatividade de gêneros mais regionais em bases de dados públicas e devido a concentração das pesquisas em gêneros mundialmente mais populares, como já dito anteriormente. Geralmente, a presença de músicas brasileiras nas bases acontece dentro de uma classe mais generalizada, muitas vezes denominada "World", "Latin Music" ou "International". Na Tabela 2, estas classes não foram consideradas como classes de gêneros musicais brasileiros por não possuírem exclusivamente exemplares de músicas brasileiras.

3.1. Trabalhos relacionados

Em um artigo da área de *MGR*, os autores investigam o desempenho de vários algoritmos de *ML*, incluindo *KNN*, *K-means*, *SVM* e Redes Neurais, levando em consideração os gêneros: *classical*, *jazz*, *metal* e *pop* [Haggblade et al. 2011]. Para isso, consideraram como atributos apenas os Coeficientes Cepstrais de Frequência Mel (MFCC) e como justificativa da escolha dos gêneros citados ressaltaram que foram escolhidos os mais distintos dentre os 10 gêneros disponíveis na base de dados utilizada chamada GTZAN, sendo esta

Tabela 2. Comparison of the most popular databases for MGR.

Base de dados	Quant. de gêneros brasileiros entre as classes da base	Quant. de músicas ou vídeos	Quant. Classes	Ano
<i>LMD</i> (<i>Latin Music Database</i>)	5	3.227	10	2008
<i>RWC</i>	2	100	10	2003
<i>FMA</i>	1	106.574	163	2016
<i>GTZAN</i>	0	1000	10	2002
<i>ISMIR2004</i>	0	729	6	2004
<i>Magnatune</i>	0	2.181	6	2009
<i>AudioSet</i>	0	2.084.320	527	2017

a mais popular em pesquisas na área de *MIR* [Sturm 2012]. Ao final do artigo, é relatado que a maioria dos algoritmos teve mais dificuldade em diferenciar entre *Metal* e *Jazz*, o que mostra uma similaridade entre as características que definem estes gêneros, além disso também é observada a superioridade considerável do algoritmo de Redes Neurais em relação aos mais simples como *KNN* e *K-means*.

Como visto na Tabela 2, há uma base de dados que destaca-se entre as outras por possuir um número maior de gênero brasileiros. Para o desenvolvimento do *Latin Music Database (LMD)*, cerca de 300 gravações musicais foram classificadas por mês, e a duração total do desenvolvimento do *LMD* levou um ano, totalizando 3.227 gravações musicais de dez gêneros de música latina em formato *MP3* [Silla Jr et al. 2008], sendo estes: Axé, Bachata, Bolero, Forró, Gaúcha, Merengue, Pagode, Salsa, Sertaneja e Tango. Devido a questões de direitos autorais, as faixas musicais não estão disponíveis publicamente, apenas os vetores das *features* compatíveis com o formato "arff", o mesmo utilizado pelo *Weka*.

Em um dos trabalhos que utilizam a base *LMD*, os autores aplicam uma abordagem de classificação multi classe decomposta em uma série de problemas de classificação binária. Assim, os classificadores são aplicados às visualizações parciais da música no espaço-tempo, e os resultados da classificação obtidos são mesclados para produzir o rótulo final da classe [Silla et al. 2008]. Como resultado, é identificado que o uso da decomposição espaço-temporal e a abordagem do conjunto de classificadores fornecem melhor precisão para a identificação de gênero musical do que o uso de segmentos individuais, como por exemplo os 30 segundos iniciais, intermediários ou finais de uma música. No artigo, também é explicado que não é possível realizar uma comparação dos resultados obtidos com resultados de outros pesquisadores de *MGR*, pois estes em sua maioria utilizaram a base *GTZAN*, que possui características musicais extraídas apenas dos 30 segundos iniciais de cada música.

4. Músicas brasileiras: uma abordagem inicial de reconhecimento

Neste artigo, está sendo proposto um estudo inicial acerca de 6 dos principais gêneros representativos brasileiros, sendo estes: Axé, Forró, MPB (Música Popular Brasileira), Rock, Samba e Sertanejo. Optou-se pela escolha destes estilos musicais por serem indicados como os mais consumidos pela população dentro os gêneros brasileiros por di-

versas plataformas de distribuição de música online, dentre elas o *Spotify* e o *iTunes*, que possuem diversas *playlists* temáticas. O *Spotify*, por exemplo, possui uma *playlist* nomeada "Top Brasil", onde lista semanalmente as músicas mais ouvidas especificamente por usuários brasileiros, tendo sempre presente músicas do estilo Sertanejo e MPB. No ano de 2018, 4 dos 5 artistas mais ouvidos no Brasil na plataforma cantam o gênero Sertanejo ¹ e no ano de 2019, 6 dos 10 artistas mais ouvidos também representam o gênero Sertanejo, 1 representa Forró e 1 representa o gênero Pagode ², subgênero do Samba.

4.1. Organização das bases de experimentação

Para o desenvolvimento dos resultados descritos na Seção 5, utilizou-se os dez gêneros contidos na base de dados *GTZAN*, totalizando 1000 amostras contendo os 30 segundos iniciais de cada uma das músicas. Quanto aos exemplares dos gêneros brasileiros selecionados, foram coletadas 613 instâncias de variados sites online que hospedam músicas de diversos artistas, o que limita a disponibilização da base de experimentação para o público por possuir em sua maioria músicas comerciais sob direitos autorais. Durante este processo, procurou-se manter uma quantidade balanceada de amostras para cada gênero, próxima de 100. Após este momento, as músicas coletadas e as músicas da base *GTZAN* foram separadas em pastas identificando cada um dos gêneros. Na Figura 1, é exibida uma tabela gerada através da ferramenta *Weka* que possui todos os gêneros da base *GTZAN* e os gêneros brasileiros selecionados, juntamente com a quantidade de exemplares existentes ou coletados, e também o peso de cada um destes gêneros nas amostras como um todo, totalizando 1613 músicas.

No.	Label	Count	Weight
1	axe	84	84.0
2	blues	100	100.0
3	classical	100	100.0
4	country	100	100.0
5	disco	100	100.0
6	forro	110	110.0
7	hiphop	100	100.0
8	jazz	100	100.0
9	metal	100	100.0
10	mpb	122	122.0
11	pop	100	100.0
12	reggae	100	100.0
13	rock	100	100.0
14	rockbr	109	109.0
15	samba	94	94.0
16	sertanejo	94	94.0

Figura 1. Quantidade de exemplares dos gêneros utilizados neste artigo.

4.2. Extração de características musicais representativas

A extração das características musicais expostas na Tabela 1 ocorreram com o auxílio de uma ferramenta chamada "librosa", um pacote Python para processamento de sinais de áudio e música, que disponibiliza prontamente para uso da comunidade de cientistas e programadores *Python* as principais técnicas de *MIR* [McFee et al. 2015].

Desenvolveu-se um pequeno algoritmo em *Python* para percorrer cada uma das músicas nos diretórios e extrair as características musicais para um arquivo em formato

¹<http://www.tenhomaisdiscosqueamigos.com/2018/12/04/spotify-mais-ouvidos-2018/>

²<https://canaltech.com.br/apps/spotify-faz-ranking-dos-mais-ouvidos-do-ano-e-da-decada-veja-lista-156923/>

CSV, por ser mais fácil de adaptar posteriormente para o formato ARFF utilizado pelo *Weka*, ferramenta onde foram realizados os experimentos.

genre	tempo	beats	chroma_sft	spectral_centroid	spectral_bandwidth
axe	86.1328125	54	0.3901197406	2450.144246	2584.908216
axe	112.3471467	72	0.4119183637	3882.757281	2985.418011
axe	184.5703125	123	0.3644136343	3094.69842	2992.354608
axe	112.3471467	73	0.3683287017	3212.582208	2799.497431
axe	151.9990809	101	0.4130999743	2959.371036	2914.893405
axe	129.1992188	84	0.3457162024	2296.97717	2527.429399
axe	95.703125	62	0.3284049246	2310.591083	2499.427956
axe	80.74951172	51	0.3770740474	2006.627075	2425.577464

Figura 2. Exemplo de resultado de extração de algumas características musicais do gênero Axé.

5. Análise comparativa e resultados

As avaliações anuais do *Music Information Retrieval Evaluation eXchange (MIREX)*, realizadas desde 2005, são um indicador interessante do desenvolvimento do campo de *MGR*. Um simples estudo da evolução dos resultados obtidos para a Tarefa de Treinamento/Teste de Classificação de Gênero *MIREX* revela pouco progresso nos escores de precisão, que nos últimos anos parecem permanecer paralisados no intervalo de 70 a 80% [Ramírez and Flores 2019]. Uma das possíveis causas deste acontecimento é justamente a falta de variedade de gêneros contidos nas bases populares para estes tipos de desafio. Por isso, o enfoque da Seção 5.2 é o uso de uma amostra de gêneros totalmente brasileiros para a realização da tarefa de *MGR*, bem como o uso de uma amostra de variados gêneros populares juntamente com os gêneros brasileiros selecionados.

Como contribuição, além da análise do desempenho de alguns dos principais algoritmos supervisionados sobre amostras diferenciadas das demais abordadas em outros trabalhos de *MIR*, há também consequentemente como resultado a geração de uma base de dados maior ao unir a base já existente *GTZAN* com os novos exemplares de música brasileira. Assim, foi possível avaliar também o desempenho dos algoritmos supervisionados frente uma base de dados popular com adição de novas instâncias. Além disso, essa união de amostras propicia uma diversidade de dados que pode ajudar os pesquisadores a desenvolver novos estudos acerca da atividade de *MGR*, visto que a base de dados possui fator primordial no sucesso dos algoritmos de ML [Lima et al. 2019].

5.1. Aspectos implementacionais

Para os experimentos descritos a seguir, optou-se pela utilização da ferramenta *Weka* anteriormente apresentada para a aplicação dos seguintes algoritmos de aprendizagem de máquina supervisionada: KNN, SVM e Floresta Aleatória, sendo este último selecionado após breves experimentos com a base *GTZAN* por ter superado a acurácia dos demais algoritmos comumente usados como Árvore de decisão e *Naive Bayes*.

Quanto aos parâmetros do algoritmo KNN, utilizou-se a configuração padrão do *Weka*, onde realiza-se o cálculo da distância entre os pontos através da distância euclidiana, porém o número K de vizinhos foi alterado para K = 5. Para o SVM e Floresta

Aleatória manteve-se todas as configurações padrões da ferramenta, onde o SVM implementa o algoritmo de otimização mínima sequencial de John Platt e a Floresta Aleatória consiste em uma floresta de árvores aleatórias, com profundidade não especificada e o número de iterações estabelecida em 100. Para o uso de todos os algoritmos, dentre as opções de teste foram escolhidas duas: Validação Cruzada (do inglês Cross-validation - CV) com 5 *folds* e Porcentagem Divida (PD) com 66% dos dados para aprendizado e o restante para teste. Para avaliação dos resultados obtidos, levou-se em consideração a precisão e o revocação obtidos através dos testes divididos por gêneros e de maneira geral.

5.2. Experimentos realizados

Com o intuito de verificar a importância e influência da duração da música utilizada para realização da tarefa de *MGR*, os experimentos foram divididos de acordo com as seguintes amostragens dos gêneros brasileiros:

1. Apenas os primeiros 30 segundos de cada música;
2. O primeiro minuto de cada música;
3. Os 40 segundos de música obtidos entre os 40 segundos iniciais e entre 1 minuto e 20 segundos, buscando abranger a parte de uma música denominada refrão;
4. Os 2 primeiros minutos de cada música.

Quanto a base *GTZAN*, aplicou-se somente o primeiro cenário descrito acima, visto que a mesma possui 1000 exemplares de músicas contendo apenas os primeiros 30 segundos, o que torna inviável o uso dos demais cenários nesta base.

Primeiramente, utilizou-se somente os exemplares representantes dos gêneros brasileiros para aplicação dos algoritmos supervisionados. Na Tabela 3, há os resultados gerados por testes que utilizaram a abordagem baseada em considerar os 2 primeiros minutos de cada música para a extração das características musicais, e o que está sendo exibido individualmente para cada gênero é a medida de precisão.

Tabela 3. Resultados obtidos ao selecionar os primeiros 2 minutos das amostras.

Algoritmo	Precisão Geral	Revocação Geral	Axé	Forró	MPB	Rock	Samba	Sertanejo
<i>KNN (K=5) CV</i>	0,587	0,573	0,430	0,586	0,630	0,595	0,628	0,621
<i>KNN (K=5) PD</i>	0,580	0,567	0,625	0,587	0,485	0,724	0,567	0,533
<i>SVM CV</i>	0,655	0,656	0,571	0,689	0,675	0,595	0,750	0,640
<i>SVM PD</i>	0,653	0,644	0,700	0,725	0,565	0,605	0,733	0,586
<i>Floresta Aleatória CV</i>	0,653	0,654	0,586	0,661	0,650	0,615	0,694	0,713
<i>Floresta Aleatória PD</i>	0,658	0,649	0,692	0,646	0,565	0,714	0,621	0,708

Posteriormente, fez-se a união da base *GTZAN* com os gêneros brasileiros e então novamente os testes. Na Tabela 4, há o resultado geral obtido pelos algoritmos ao considerar-se os gêneros brasileiros juntamente com todos os dez gêneros da base *GTZAN*

e uma amostragem de 40 segundos de cada música de estilo brasileiro, extraída entre os 40 segundos iniciais e entre 1 minuto e 20 segundos. Em ambas tabelas de resultados, a sigla "CV" significa que neste algoritmo foi utilizado o teste de Validação Cruzada, enquanto que "PD" indica que o teste ocorreu com o uso de Porcentagem Dividida.

Tabela 4. Resultados gerais obtidos pela união da base GTZAN as amostras brasileiras com duração de 40 segundos.

	KNN (K=5) CV	KNN (K=5) PD	SVM CV	SVM PD	Floresta Aleatória CV	Floresta Aleatória PD
<i>Precisão Geral</i>	0,610	0,602	0,570	0,549	0,642	0,589
<i>Revocação Geral</i>	0,598	0,582	0,576	0,557	0,645	0,593

5.3. Análise comparativa

Como pode ser observado na Tabela 3, os algoritmos supervisionados que mais destacam-se na medida de precisão e revocação para quase todos os gêneros é *SVM* e o de Floresta Aleatória, sendo que este último não é comumente visto em trabalhos relacionados, que geralmente optam por utilizar o de Árvores de Decisão, como observado durante a revisão da literatura. Esse mesmo cenário de melhor desempenho dos algoritmos *SVM* e Floresta Aleatória repetiu-se para os demais testes abrangendo os primeiros 30 segundos de cada música e também o primeiro minuto, com destaque para Floresta Aleatória, que apresentou o melhor desempenho para todos os testes com a base *GTZAN* em conjunto com os estilos brasileiros, alcançando uma precisão geral de 0,645 quando aplicado a uma amostragem de 40 segundos das músicas brasileiras. Portanto, pode-se afirmar que este algoritmo supervisionado possui potencial para alcançar acurácias satisfatórias quando aplicado a tarefa de *MGR*, apesar de ser pouco explorado pelas pesquisas mais populares no campo de *MIR*.

Percebeu-se também que não somente a duração da música influencia nos resultados, como também a decisão sobre qual parte da música será utilizada para a extração das *features*. A mesma peça de uma música é indicada como pertencente a diferentes gêneros musicais, de acordo com diferentes sites de distribuição de música eletrônica (do inglês, Electronic Music Distribution - EMD), devido a diferentes interpretações e/ou esquemas de codificação diferentes [Silla et al. 2008]. Nos testes realizados somente com os gêneros brasileiros, destacou-se a amostragem com os 2 primeiros minutos das músicas, que obteve uma precisão de 0,653 e revocação de 0,654 com o uso do algoritmo *SVM*. Por outro lado, o melhor resultado da união do *GTZAN* com os estilos brasileiros foi obtido através da amostragem brasileira de 40 segundos.

Há uma grande variedade e riqueza de gêneros brasileiros e subgêneros, como por exemplo os subgêneros do Samba que incluem: Pagode, Samba de breque, Samba-canção, Samba-enredo, entre outros. Contudo, para este estudo preliminar destacou-se um conjunto dos 6 principais gêneros. Ainda assim, estes principais geralmente não aparecem em todas as demais bases populares, como apresentado na Tabela 2, ou não são destacados como uma das classes principais da base. Neste trabalho, a união com a base *GTZAN* propiciou justamente esta abrangência e diversidade maior necessária as bases de dados disponíveis atualmente. Com isso, foi possível avaliar aspectos como a similaridade entre os gêneros brasileiros e os demais mundialmente populares, como Pop e Jazz.

a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	<-- classified as	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	<-- classified as
34	0	0	1	17	0	1	0	2	0	0	1	17	4	6	1	a = axe	41	0	0	0	0	14	1	0	0	2	0	0	0	17	1	8	a = axe
0	58	1	4	5	0	2	5	7	3	0	3	12	0	0	0	b = blues	0	60	2	5	3	0	1	7	10	0	0	2	10	0	0	0	b = blues
0	1	91	2	0	0	0	2	0	2	0	0	0	0	0	2	c = classical	0	0	90	3	0	0	0	3	0	0	0	2	2	0	0	0	c = classical
1	5	3	44	3	4	3	8	0	10	5	3	5	6	0	0	d = country	0	4	4	54	3	0	3	11	0	3	5	4	9	0	0	0	d = country
2	3	1	1	48	1	10	1	4	1	11	3	9	4	0	1	e = disco	1	3	1	3	55	0	10	0	5	0	15	3	4	0	0	0	e = disco
14	0	1	0	1	60	0	2	0	10	3	0	0	7	8	4	f = forro	14	0	0	0	0	69	0	1	0	5	0	0	0	6	7	8	f = forro
0	0	0	5	10	1	51	1	9	0	7	7	2	7	0	0	g = hiphop	1	0	0	5	5	0	55	1	10	0	8	9	4	2	0	0	g = hiphop
0	4	9	1	3	2	1	63	1	4	2	4	1	2	2	1	h = jazz	1	3	9	6	2	1	1	64	1	1	3	4	2	1	1	0	h = jazz
0	4	0	0	3	0	4	0	83	0	0	1	2	3	0	0	i = metal	0	3	0	1	2	0	7	0	80	0	0	1	6	0	0	0	i = metal
3	2	3	4	0	3	2	1	0	63	0	3	0	14	11	13	j = mpb	7	0	0	1	1	3	0	0	0	74	0	1	0	15	14	6	j = mpb
0	0	0	1	7	5	4	1	0	1	74	1	0	3	3	0	k = pop	0	0	0	5	5	2	5	0	0	0	81	1	1	0	0	0	k = pop
0	3	1	5	4	0	13	2	1	3	5	54	2	6	1	0	l = reggae	0	4	2	8	6	0	13	3	1	3	4	53	2	1	0	0	l = reggae
1	8	0	11	13	2	4	4	6	1	4	6	30	9	1	0	m = rock	0	12	0	9	18	1	3	6	6	0	3	8	34	0	0	0	m = rock
13	1	1	1	1	5	3	0	0	16	0	7	0	57	1	3	n = rockbr	10	0	0	1	1	10	1	0	0	18	0	0	0	65	2	1	n = rockbr
2	0	0	1	0	8	0	3	0	9	0	0	0	6	43	22	o = samba	1	0	0	0	0	6	0	0	0	12	0	0	0	6	52	17	o = samba
5	0	2	0	0	10	1	0	0	3	0	0	0	0	9	64	p = sertanejo	7	0	1	0	0	8	0	0	0	6	0	0	0	2	5	65	p = sertanejo

Figura 3. Matriz de confusão para amostras de 30s e 60s, respectivamente.

a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	<-- classified as
42	0	0	0	0	18	0	0	0	3	0	0	0	11	3	7	a = axe
0	61	3	6	5	0	0	5	9	0	0	2	9	0	0	0	b = blues
0	0	92	2	0	0	0	3	0	0	0	2	1	0	0	0	c = classical
0	2	1	57	4	0	3	14	0	0	6	6	7	0	0	0	d = country
0	4	1	1	57	0	11	0	2	0	13	5	6	0	0	0	e = disco
7	0	0	0	0	80	0	0	0	5	0	0	0	7	6	5	f = forro
0	0	0	4	9	0	55	0	9	0	8	9	6	0	0	0	g = hiphop
0	4	8	1	3	0	2	68	1	0	5	5	3	0	0	0	h = jazz
0	3	0	0	3	0	5	0	84	0	0	2	3	0	0	0	i = metal
3	0	0	0	0	4	0	0	0	87	0	0	0	14	9	5	j = mpb
0	0	0	6	7	0	5	2	0	0	76	3	1	0	0	0	k = pop
0	1	2	11	7	0	15	3	1	0	5	52	3	0	0	0	l = reggae
0	9	0	11	21	0	2	6	8	0	6	7	30	0	0	0	m = rock
6	0	0	0	0	7	0	0	0	18	0	0	0	74	3	1	n = rockbr
0	0	0	0	0	6	0	0	0	12	0	0	0	4	60	12	o = samba
2	0	1	0	0	11	1	0	0	5	0	0	0	4	5	65	p = sertanejo

Figura 4. Matriz de confusão para amostras de 40s.

Nas Figuras 3 e 4 há as matrizes de confusão obtidas pela aplicação do algoritmo Floresta Aleatória utilizando CV para 30, 60 e 40 segundos de amostras brasileiras, indicando que durante a classificação há confusão principalmente do gêneros *Rock* e *Axé*, sendo que *Axé* é bastante confundido com *Forró* e *Rock Nacional* (identificado nas matrizes como "rockbr"), enquanto que *Rock* é confundido com quase todos os demais gêneros da base, principalmente *Disco* e *Blues*. Parte destes resultados podem ser justificados através da literatura existente, que indica a presença de falhas contidas na base *GTZAN* e seus efeitos na avaliação de tarefas como *MGR*, mostrando que há diversos artistas repetidos e também músicas classificadas de forma errônea, por exemplo [Sturm 2013]. Outra justificativa é a escolha das características musicais utilizadas como atributos, que podem ser melhoradas e determinadas de forma que representem mais precisamente as definições de cada um dos gêneros, principalmente os brasileiros, ainda pouco explorados.

6. Considerações Finais

Neste trabalho, a problemática de reconhecimento automático de gêneros foi abordada com o uso de técnicas de aprendizagem de máquina supervisionada e aplicada a seis dos principais gêneros musicais brasileiros: *Axé*, *Forró*, *MPB*, *Rock*, *Samba* e *Sertanejo*. Por meio de estudos iniciais e testes de ambas organizações de faixas musicais propostas: *GTZAN* com gêneros brasileiros e somente gêneros brasileiros, foram obtidas precisões gerais de 0,658 e 0,642 e revocações gerais de 0,645 e 0,656 para cada uma das amostras, respectivamente.

Com isso, é possível inferir que a adição de novos gêneros a uma base popularmente conhecida não altera negativamente os resultados obtidos através da aplicação de algoritmos de *ML* clássicos, contribuindo apenas para uma diversificação e abrangência

maior de gêneros. Há muitos materiais e ferramentas disponíveis para a comunidade científica quando trata-se de dados para pesquisas em *MIR*, contudo há uma grande escassez de representatividade de gêneros mais regionais e diversificados. Como trabalhos futuros, pretende-se aumentar a quantidade de gêneros musicais brasileiros entre as amostras e melhorar as precisões e revocações obtidas através de outras combinações entre atributos e técnicas de aprendizagem de máquina, incluindo as não supervisionadas.

Referências

- Aucouturier, J.-J. and Pachet, F. (2003). Representing musical genre: A state of the art. *Journal of New Music Research*, 32(1):83–93.
- Biau, G. (2012). Analysis of a random forests model. *Journal of Machine Learning Research*, 13(Apr):1063–1095.
- BRÉSCIA, V. P. (2009). A música como recurso terapêutico. In *Encontro Paranaense Congresso Brasileiro de Psicoterapias Corporais, XIV, IX*.
- Haggblade, M., Hong, Y., and Kao, K. (2011). Music genre classification. *Department of Computer Science, Stanford University*.
- Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., and Witten, I. H. (2009). The weka data mining software: An update. *SIGKDD Explor. Newsl.*, 11(1):10–18.
- Jeong, I.-Y. and Lee, K. (2016). Learning temporal features using a deep neural network and its application to music genre classification. In *Ismir*, pages 434–440.
- Lima, M., Valle, V., Costa, E., Lira, F., and Gadelha, B. (2019). Software engineering repositories: Expanding the promise database. In *Proceedings of the XXXIII Brazilian Symposium on Software Engineering*, pages 427–436.
- McFee, B., Raffel, C., Liang, D., Ellis, D. P., McVicar, M., Battenberg, E., and Nieto, O. (2015). librosa: Audio and music signal analysis in python. In *Proceedings of the 14th python in science conference*, volume 8.
- Moreira, S. V., França, C. C., Moreira, M. A., and Lana-Peixoto, M. A. (2009). Identidade musical em pacientes com esclerose múltipla. *Arquivos de Neuro-Psiquiatria*, 67(1):46–49.
- Naik, A. and Samant, L. (2016). Correlation review of classification algorithm using data mining tool: Weka, rapidminer, tanagra, orange and knime. *Procedia Computer Science*, 85:662–668.
- Rajvanshi, N. and Chowdhary, K. (2017). Comparison of svm and naïve bayes text classification algorithms using weka. *International Journal of Engineering Research and*, 6:09.
- Ramírez, J. and Flores, M. J. (2019). Machine learning for music genre: multifaceted review and experimentation with audioset. *Journal of Intelligent Information Systems*, pages 1–31.
- Silla, C. N., Koerich, A. L., and Kaestner, C. A. (2008). A machine learning approach to automatic music genre classification. *Journal of the Brazilian Computer Society*, 14(3):7–18.

- Silla Jr, C. N., Koerich, A. L., and Kaestner, C. A. (2008). The latin music database. In *ISMIR*, pages 451–456.
- Sturm, B. L. (2012). A survey of evaluation in music genre recognition. In *International Workshop on Adaptive Multimedia Retrieval*, pages 29–66. Springer.
- Sturm, B. L. (2013). The gtzan dataset: Its contents, its faults, their effects on evaluation, and its future use. *arXiv preprint arXiv:1306.1461*.