

A Fuzzy Approach to Drum Cymbals Classification

Tales H. A. Boratto , Alexandre A. Cury  and Leonardo Goliatt 

Abstract—The many factors that influence the sound of a cymbal, combined with the external aspects that modify its sound perception, make the study of the acoustics of these instruments more challenging. Within the context of machine learning, most researches involving cymbal classification from their sounds aim to identify those instruments according to their types. However, there is a lack of studies investigating the acoustic elements of cymbals using machine learning techniques as tools. Hence, this paper proposes to classify cymbals according to their constitutive materials, since the metallic alloy assumes a significant portion is responsible for their acoustics. In addition, there is an interest in evaluating a fuzzy logic approach as a classifier applied to three sets of attributes, formed from temporal features and Mel Frequency Cepstral Coefficients extracted from audio signals, comparing triangular and Gaussian membership functions. For this, 276 audios, referring to 4 drum cymbals, were collected from a standardized procedure for capturing the sounds that considered variations in microphones and environments. As a result, the implemented model achieved 94.72% of average accuracy with a standard deviation of 2.51%, considering the Gaussian membership function and Mel Frequency Cepstral Coefficients as the audio descriptor.

Index Terms—Bronze Alloys, Cymbals, Fuzzy, Music Information Retrieval.

I. INTRODUÇÃO

A extensa variedade de pratos de bateria e suas aplicações mostram o quanto estes instrumentos têm se tornando cada vez mais relevantes no contexto musical. Dentre os possíveis materiais com os quais podem ser fabricados, os bronzes, que são basicamente ligas metálicas de cobre (*Cu*) com adição de estanho (*Sn*), destacam-se por serem os mais utilizados. No entanto, a depender da concentração de estanho, diferentes ligas de bronzes podem ser obtidas, gerando propriedades mecânicas e acústicas distintas [1]. Sendo assim, grande parte das empresas adota uma nomenclatura de seus materiais baseada em suas composições químicas, conforme os exemplos mostrados pela Tabela I. A letra *B* refere-se a *Bronze* e a numeração subsequente indica a porcentagem de estanho em peso presente na liga.

Os pratos de bateria são classificados como instrumento idiofônicos, isto é, a vibração de seu próprio corpo é o que produz o seu som que, de acordo com Rossing [2], apresenta formas de ondas não repetitivas compostas por um grande número de frequências não relacionadas, sem um senso

Tales H. A. Boratto - Programa de Pós-Graduação em Modelagem Computacional (PGMC), Universidade Federal de Juiz de Fora (UFJF), Juiz de Fora, MG, Brasil. E-mail: tales.boratto@engenharia.ufjf.br

Alexandre A. Cury - Programa de Pós-Graduação em Engenharia Civil (PEC), Universidade Federal de Juiz de Fora (UFJF), Juiz de Fora, MG, Brasil. E-mail: alexandre.cury@ufjf.br

Leonardo Goliatt - Departamento de Mecânica Aplicada e Computacional (MAC), Universidade Federal de Juiz de Fora (UFJF), Juiz de Fora, MG, Brasil. E-mail: goliatt@gmail.com

TABELA I
EXEMPLOS DE NOMENCLATURA DE LIGAS DE BRONZE UTILIZADAS EM PRATOS DE BATERIA.

Liga de Bronze	Sn (% peso)	Cu (% peso)
B8	8	92
B10	10	90
B20	20	80

definido de tom. Ainda em [2], Rossing afirma que, quando um prato é tocado, as ondas se propagam da região de excitação a uma taxa que é inversamente proporcional às dimensões da flexão inicial da superfície feita durante o impacto. Na sequência, este pulso inicial, bem como suas reflexões nas bordas do instrumento, viajam até a região central, onde então são espalhados por todo o corpo do prato, fazendo-o vibrar [3].

Sabendo disso, pode-se dizer que todo fator que esteja relacionado a uma modificação estrutural destes instrumentos, ainda que mínima, contribuirá na formação de seu som final. Entretanto, alguns fatores apresentam uma maior relevância na construção acústica dos pratos, como é o caso do material constituinte, dos aspectos geométricos (formato, concavidade, tamanho diametral, tamanho de cúpula, espessura e gradiente de espessura), dos processos de fabricação submetidos e do acabamento superficial, como pode ser visto com maiores detalhes em [1]. Pinksterboer e Mattingly em [4] afirmam ainda que o estado de conservação, a idade e o uso do prato podem contribuir para uma modificação acústica destes instrumentos. Indo além, afirmam que a percepção do som de um prato é dependente do ambiente, da maneira do toque, do material utilizado para golpeá-lo, bem como do tempo de contato durante o impacto.

Por conta destas diversas possibilidades de resultados acústicos que podem ser obtidos, uma extensa variedade de pratos de bateria é encontrada no mercado. Entretanto, as modificações acústicas dificilmente são realizadas com o intuito de criar uma nova categoria de prato. Geralmente, estas alterações não descaracterizam o tipo de aplicação do prato. Os tipos mais comuns de categorias são:

- *Hi-hat*: Esta categoria é, na verdade, uma combinação de dois pratos côncavos que são montados juntos em uma estante apropriada e que regula a distância entre eles. Geralmente são utilizados na condução de uma música, apresentando diversas possibilidades de som devido à regulagem de distância entre os pratos.
- *Crash*: Refere-se aos pratos de geometria tradicional côncava. Normalmente são utilizados em momentos de destaque ou acentuações de uma música, apresentando

características acústicas específicas para esta tarefa.

- *Ride*: São pratos geometricamente próximos aos *Crashes* embora sejam, de maneira geral, mais pesados e com diâmetro maior. Geralmente são utilizados para a condução de uma música.
- *Splash*: De modo geral, também são próximos aos *Crashes*, com a diferença de serem menores, mais leves e utilizados em momentos de acentuações curtas e rápidas.
- *China*: Esta categoria de pratos possui uma geometria diferenciada em que a borda destes instrumentos segue a direção contrária à da superfície do prato. Por conta disso, os *chinas* apresentam características acústicas bastante peculiares e são geralmente utilizados em acentuações da uma música que necessitam de um efeito sonoro diferente.

Trabalhos que investigam a acústica dos pratos normalmente visam compreender em mais detalhes a maneira com que determinados aspectos influenciam no som destes instrumentos. Osamura *et al.* em [5] e Kuratani *et al.* em [6] dedicaram-se a estudar a influência da tensão residual nas características acústicas e vibratórias dos pratos de bateria. Já Ogawa *et al.* em [7] interessaram-se em estudar o efeito do tamanho da cúpula sobre as características acústicas destes instrumentos.

Há também aqueles que buscam compreender a relação com que o material constituinte de um instrumento musical apresenta em suas propriedades acústicas. Slamet *et al.* em [8] avaliaram o efeito da composição química de três diferentes ligas bronzes, cujas proporções de estanho variaram de 20% a 24% em peso, sobre as suas propriedades físicas, mecânicas e acústicas. Similarmente, Sugita *et al.* em [9] estavam interessados em investigar o efeito da composição química de bronzes de silício (Cu-Si) sobre as suas propriedades mecânicas e acústicas, com o objetivo de encontrar uma nova alternativa de material a ser utilizado em instrumentos musicais.

Ao considerar o contexto de aprendizado de máquina [10], a utilização de modelos que auxiliem no estudo de pratos de bateria se restringem, em linhas gerais, a problemas de classificação conforme as suas categorias. Herrera *et al.* em [11] resolve um problema de classificação que envolve identificar diferentes partes de uma bateria acústica. Para isso, os autores optaram por avaliar o desempenho de 5 diferentes classificadores, além de particionar o problema em três etapas, aumentando a escala de complexidade:

- 1) Identificar se o conteúdo de áudio é referente a um tambor ou a um prato. Neste caso, foi possível atingir uma acurácia de 99,3% ao utilizar a Análise Discriminante Canônica (*Canonical Discriminant Analysis - CDA*) como seletor de atributos e classificador.
- 2) Classificar os áudios quanto a 5 classes, sendo 3 de tambores (Bumbo, Caixa e Tom) e 2 pratos (*Hi-hat* e Pratos em Geral). Nestas condições, os autores conseguiram alcançar uma acurácia de 97,4%, utilizando a técnica *ReliefF* para a seleção das características e o K^* [12] como classificador.
- 3) Identificar os áudios quanto a 5 sub-categorias dos tambores (Bumbo, Caixa, Tom Grave, Tom Médio, Tom Agudo) e 4 sub-categorias dos pratos (*Hi-hat* aberto, *Hi-*

hat fechado, *Ride* e *Crash*). Ao combinar o CDA para a seleção dos atributos e o K^* como classificador, um desempenho de classificação de 90,7% foi atingido.

Cavaco e Almeida em [13] fazem uso da Matriz de Fatoração Não-Negativa e do K-NN (com $K = 1$) para realizar a tarefa de classificar os pratos conforme as suas categorias, considerando 7 agrupamentos diferentes da base de dados:

- 1) *Splash* e *China*,
- 2) *Crashes* de 14" e 16",
- 3) *Splash* e *Crash* de 16",
- 4) *China* e *Crash* 16",
- 5) *hi-hat* fechado e *Ride*, ambos golpeados na borda,
- 6) *China* e *Crashes* de 14" e 16",
- 7) *Splash* e *Crashes* de 14" e 16".

Dentre os agrupamentos compostos por apenas 2 categorias de pratos, os autores conseguiram atingir um desempenho de classificação de 95%. Já para os agrupamentos que consideram 3 categorias, a taxa de acerto foi de 86%.

Souza *et al.* em [14] trazem uma abordagem mais aprofundada e propõem não só classificar os pratos em 5 categorias (*China*, *Crash*, *Hi-hat*, *Ride* e *Splash*), mas também buscam identificar a maneira como se deu o toque, fazendo com que o problema seja formulado para corresponder a um total de 12 classes. Cinco classificadores foram utilizados para solucionar este problema e em todos os casos os melhores desempenhos foram obtidos ao se utilizar a Máquina de Vetor Suporte. Ao considerar apenas a classificação quanto as 5 categorias de pratos, o modelo foi capaz de atingir um desempenho de 96,59%. No entanto, quando foram considerados os 12 rótulos, a acurácia alcançada pelo modelo foi de 91,54%.

Ainda assim, abordagens que focam na utilização de técnicas de aprendizado de máquina na avaliação dos aspectos acústicos de pratos de bateria ainda são pouco exploradas. Por conta disso, este trabalho tem como objetivo estudar o aspecto que proporciona a base do som dos pratos e, portanto, buscar classificá-los de acordo com os seus respectivos materiais. Indo além, optou-se por empregar uma abordagem da lógica *fuzzy* [15] como classificador, por ser um modelo que lida bem com incertezas e que, no melhor dos conhecimentos dos autores, ainda não foi testado para esta tarefa de classificação de pratos de bateria. Consequentemente, pretende-se averiguar a implementação de tal modelo de classificação com as funções de pertinência triangular e Gaussiana em diferentes conjuntos de atributos extraídos dos sinais.

Assim, 276 áudios distribuídos entre 3 classes de materiais foram captados em estúdio em um procedimento que considerou 2 ambientes, 2 microfones diferentes, 4 pratos de bateria e utilizou um pedal de bumbo para padronizar a força, o local, o modo e o tempo de contato no impacto. Coeficientes Mel-Cepstrais de Frequência (*Mel-Frequency Cepstral Coefficients - MFCC*) [16], [17] e Atributos Temporais foram extraídos dos sinais dando origem a três conjuntos de características que serão avaliados.

Apesar do enfoque musical, a classificação conforme os materiais a partir do som pode ser uma abordagem bastante útil quando observada da perspectiva de um fabricante para a certificação de seus produtos, por exemplo.

O presente artigo está organizado da seguinte forma: a presente seção tratou da introdução do problema de classificação de pratos, elencando alguns trabalhos relacionados na literatura. A seção II traz a formulação do problema, descrevendo os procedimentos de aquisição dos áudios, da extração de atributos, bem como a base de dados gerada. A seção III apresenta o modelo *fuzzy* utilizado para este problema. A seção IV aborda os resultados computacionais obtidos, as discussões pertinentes e elenca vantagens e limitações deste trabalho. Por fim, a seção V é reservada para as observações finais do trabalho.

II. FORMULAÇÃO DO PROBLEMA

A. Aquisição de Dados

Com o intuito de se obter uma captação dos sons da forma mais fiel à realidade, a determinação dos microfones ocorreu a partir da análise dos gráficos de resposta em frequência dos equipamentos disponíveis, optando-se por escolher aqueles com a curva mais plana possível próximo a 0 dB. Portanto a aquisição dos áudios foi conduzida em um estúdio de gravação equipado com um microfone condensador *AKG C414 XL II* ajustado no padrão cardioide e um microfone *Shure PG81*, também condensador e cardioide, cujas respostas em frequência são mostradas nas Fig. 1 e 2. Além disso, utilizou-se a interface de áudio *Behringer Xr18* (24 bits and 48 kHz) e o software *Reaper*, ambos configurados com ganhos nulos para minimizar a interferência nos sons captados.

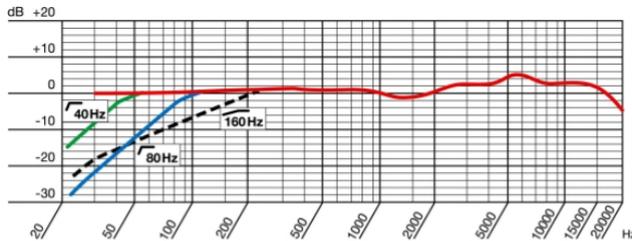


Fig. 1. Resposta em frequência do microfone AKG C414 XL II. Fonte: [18].

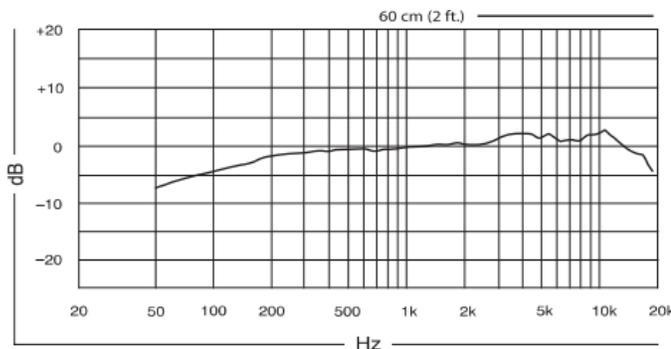


Fig. 2. Resposta em frequência do microfone Shure PG81. Fonte: [19].

Quatro pratos de mesma classificação (*Crashes* de 18 polegadas de diâmetro) distribuídos entre três classes de materiais (B8, B10 e B20) foram selecionados. Dois destes são similares, isto é, de mesmo modelo e fabricante, com a diferença de um estar saudável e o outro danificado com uma deformação

que se estende em 25,4 mm ao longo da borda, 10 mm no sentido do raio do prato e flexão máxima de cerca de 1 mm. A Tabela II apresenta um resumo dos pratos utilizados e as suas condições físicas.

TABELA II
PRATOS UTILIZADOS NA AQUISIÇÃO DE DADOS.

Pratos	Bronze / Classe	Condição Física
Krest Fusion	B8	Novo
Krest Fusion	B8	Usado, Danificado
N Gate K10 Expression	B10	Novo
Zildjian K Dark	B20	Usado, Sem dano

Como mostrado na Fig. 3, os pratos foram apoiados sobre uma estante apropriada, ajustada com uma pequena inclinação, de modo que o batedor do pedal pudesse atingir a borda do instrumento, e com uma pressão leve do feltro superior para, praticamente, apenas dar contato com o prato. Os microfones, por sua vez, foram posicionados a uma distância de aproximadamente 250 mm da borda na extensão do raio do instrumento. O pedal de bumbo foi fixado com uma certa inclinação em uma cadeira a fim de elevar a altura do equipamento. Apesar de seu acionamento ter sido feito de maneira manual, uma escala adesivada no pedal permitiu uma padronização do arco do batedor, o que possibilitou impactos de modos e intensidades aproximadas. Além disso, o uso de um pedal de bumbo permitiu que o tempo de contato do batedor com o prato durante o impacto não fosse o suficiente para abafar o som do instrumento.

O procedimento descrito para a aquisição dos sons foi realizado em dois ambientes distintos. Primeiramente, o local de captação foi isolado do restante da sala, utilizando-se colchões para minimizar a reflexão das ondas sonoras. Em seguida, o mesmo procedimento foi conduzido, desta vez no ambiente normal da sala, ou seja, sem os colchões.

Por fim, os sons captados foram separados individualmente em áudios de 21 segundos. No entanto, embora a captação tenha sido realizada em 48 kHz, o procedimento de renderização do *software* transformou a frequência de amostragem dos sinais para 44,1 kHz, através de uma operação de decimação (*Downsampling*).

B. Conjunto de Dados

Após a aquisição dos sinais, foi possível construir um conjunto de dados composto por 276 áudios de 21 segundos com frequências de amostragem de 44,1 kHz, considerando os quatro pratos estabelecidos. Subdividindo-se este conjunto de dados por classes, tem-se 134 áudios para o material B8, 76 para o B10, e 66 para o B20, conforme detalhado na Tabela III.

É importante destacar que a base de dados foi construída a partir dos sons brutos dos pratos e que, além das imperfeições do procedimento, foram considerando propositalmente áudios ruidosos com falas e uma renderização com tempo inferior aos 21 segundos previamente estabelecidos, corrigida por uma operação de *zero-padding*, a fim de estressar o modelo. Vale ressaltar, também, que este artigo visa a classificação por liga



Fig. 3. Configuração do procedimento para aquisição de dados. 1) Estante do Prato. 2) Prato. 3) Microfone Shure PG 81. 4) Microfone AKG C414 XL II. 5) Pedal de bumbo. 6) Batedor do Pedal. Fonte: Autor.

TABELA III
DESCRIÇÃO DO CONJUNTO DE DADOS AQUISITADOS.

Prato	Estado	Classe	Ambiente	Quantidade de Áudios
Krest Fusion	Sem Dano	B8	com colchão	30
			sem colchão	40
Krest Fusion	Danificado	B8	com colchão	32
			sem colchão	32
NGate K10 Expression	Sem Dano	B10	com colchão	30
			sem colchão	46
Zildjian K Dark	Sem Dano	B20	com colchão	30
			sem colchão	36

de bronze. Portanto, o prato danificado é colocado como sendo da mesma classe que seu semelhante não danificado, mesmo que a pequena deformação na borda tenha acarretado uma alteração acústica neste instrumento.

Para finalizar a preparação do conjunto de dados, 13 coeficientes mel cepstrais da frequência (vide Seção II-C) foram extraídos de cada sinal a partir de um algoritmo em *Python*

TABELA IV
CONJUNTOS DE ATRIBUTOS UTILIZADOS PARA A AVALIAÇÃO DE DESEMPENHO.

Conjunto	Atributos	Dimensão	Dimensão
		Inicial (Lin x Col)	Pós Filtro (Lin x Col)
C_1	Temporal	276 x 18	276 x 15
C_2	MFCC	276 x 13	276 x 13
C_3	Temporal + MFCC	276 x 31	276 x 28

implementado pelos autores, seguindo a lógica proposta por Malcom Slaney em [17]. Além disso, 18 características temporais também foram extraídas dos áudios, utilizando-se o pacote TSFEL [20] disponível para *Python*. Em seguida, um filtro para eliminar os atributos com médias absolutas inferiores a 1^{-10} foi aplicado com o objetivo de desconsiderar aqueles com valores praticamente residuais e, portanto, reduzir a quantidade de recursos computacionais a serem utilizados, além de evitar problemas de divisão por zero durante o cálculo das pertinências. Desta forma, três conjuntos de características foram construídos, como mostrado na Tabela IV.

C. Extração de Atributos

Os atributos utilizados neste problema foram extraídos dos domínios da que frequência e do tempo, conforme detalhado nos itens a seguir.

1) *Domínio Que Frequência*: De acordo com Oppenheim e Schaffer em [21], *Cepstrum* é o termo utilizado para se referir ao espectro do log do espectro de uma forma de onda no tempo. Já o *Mel Cepstrum de Frequência* é derivado do logaritmo do espectro obtido de um banco de filtros com frequências centrais e larguras de banda definidas por um intervalo de frequência constante.

O MFCC (*Mel Frequency Cepstral Coefficients* ou Coeficientes Mel-Cepstrais de Frequência, em português) são descritores de áudio que relacionam a frequência real e o tom percebido, mapeando as frequências para corresponder aproximadamente à capacidade de percepção auditiva humana. Tal mapeamento é linear até cerca de 1 kHz e logarítmico para frequências superiores [16], [22]. Por conta disso, os MFCC têm sido uma boa alternativa e amplamente utilizados em problemas de recuperação de informação musical, o que inclui o reconhecimento de instrumentos musicais, como é o caso dos trabalhos [14], [23]–[27].

O diagrama de blocos mostrado na Fig. 4 ilustra as etapas do procedimento para a obtenção dos MFCCs, cuja lógica foi utilizada para a implementação em *Python*. A etapa de janelamento consiste na aplicação da janela de *Hamming* [29], cuja formulação matemática é apresentada na Equação (1).

$$w(n) = 0,54 - 0,46 \cos\left(\frac{2\pi n}{N}\right), \quad 0 \leq n \leq N \quad (1)$$

onde $N + 1$ corresponde ao comprimento da janela em número de pontos [29]. Neste caso, optou-se por realizar um único janelamento do sinal, isto é, considerar $N + 1$ como sendo o tamanho do sinal de áudio.

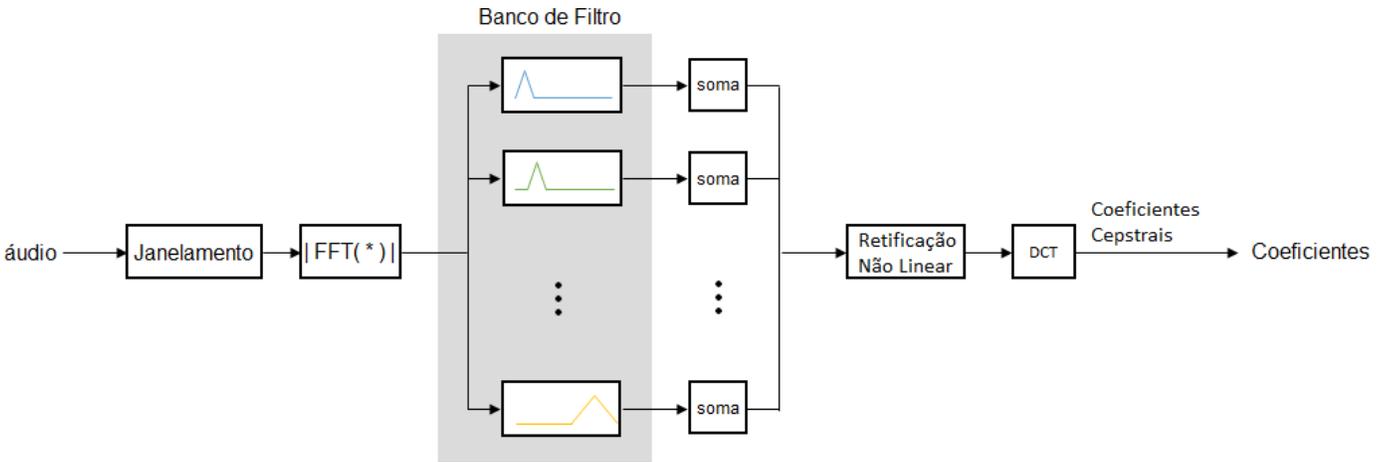


Fig. 4. Etapas para o cálculo dos MFCCs. Fonte: Adaptado de [28].

A Transformada Rápida de Fourier (*Fast Fourier Transform - FFT*) [29] é aplicada ao sinal obtido após a operação de janelamento e o valor absoluto deste resultado será utilizado como entrada na etapa subsequente.

O Banco de Filtro, por sua vez, foi construído de acordo com a descrição de Malcolm Slaney em [17]: “13 filtros linearmente espaçados (133,33 Hz entre frequências centrais), seguido por 27 filtros com espaçamento logarítmico (separados por um fator de 1,0711703 em frequência.)”

Com relação às demais etapas, a Retificação Não Linear consiste no cálculo do logaritmo na base 10 das saídas do banco de filtro e a Transformada Discreta de Cosseno (*Discrete Cosine Transform - DCT*) [29] é utilizada para a redução da dimensionalidade antes de retornar os coeficientes calculados.

2) *Domínio Temporal* : Para a extração das características temporais, utilizou-se o pacote TSFEL (*Time Series Feature Extraction Library*) [20] disponível para *Python*. Dessa forma, 18 atributos, descritos na Tabela V, foram extraídos dos sinais. Mais detalhes a respeito da implementação podem ser encontrados em [20].

III. MODELO EMPREGADO

O conceito da lógica difusa ou lógica *fuzzy* proposto por Zadeh em [31] manipula conceitos booleanos tradicionais, atribuindo graus de veracidade [32] às informações que variam entre o completamente falso e o completamente verdadeiro, passando a existir um grau de pertinência. Sendo assim, a estrutura da lógica *fuzzy* permite facilitar a tomada de decisões em um ambiente de imprecisão causado pela ausência de critérios bem definidos de pertencimento às classes [31].

Uma função de pertinência $f_A(x)$ para um conjunto difuso A no universo do discurso X é definida como $f_A(x) : X \rightarrow [0, 1]$, onde cada elemento de X é mapeado para um valor real entre 0 e 1. Assim, tomando um ponto qualquer $x \in X$, o valor obtido ao aplicar a função de pertinência neste ponto ($f_A(x)$) quantifica o grau de pertinência do elemento x ao conjunto difuso A , sendo, portanto, denominado valor ou grau de pertinência [31].

TABELA V

DESCRIÇÃO DOS ATRIBUTOS TEMPORAIS EXTRAÍDOS PELO TSFEL [20].

Atributos	Descrição
X_1	Energia absoluta do sinal.
X_2	Área sob a curva do sinal calculado com regra trapezoidal.
X_3	Autocorrelação do sinal.
X_4	Centroide ao longo do eixo do tempo
X_5	Entropia do sinal usando Entropia de Shannon [30].
X_6	Média do absoluto das diferenças do sinal.
X_7	Média das diferenças do sinal.
X_8	Mediana do absoluto das diferenças do sinal.
X_9	Mediana das diferenças do sinal.
X_{10}	Número de pontos de viragem negativos do sinal.
X_{11}	Número de picos em uma vizinhança definida do sinal.
X_{12}	Distância de pico a pico.
X_{13}	Número de pontos de viragem positivos do sinal.
X_{14}	Distância percorrida pelo sinal.
X_{15}	Inclinação do sinal.
X_{16}	Soma das diferenças absolutas do sinal.
X_{17}	Energia total do sinal.
X_{18}	Calcula a taxa com a qual o sinal cruza o zero.

Para este trabalho, foram utilizadas as funções de pertinência triangular e Gaussiana, conforme [33], representadas respectivamente pelas Equações 2 e 3.

$$f_A(x) = \begin{cases} 0 & x \leq a \\ \frac{x-a}{m-a} & a < x \leq m \\ \frac{b-x}{b-m} & m < x < b \\ 0 & x \geq b \end{cases} \quad (2)$$

$$f_A(x) = \exp\left\{-\frac{(x-m)^2}{2\sigma^2}\right\} \quad (3)$$

Os valores de a , m e b da Equação (2) referem-se, respectivamente, aos vértices esquerdo, superior e direito de um triângulo qualquer, conforme é ilustrado pela Fig. 5. Com relação à Equação (3), os termos m e σ são os parâmetros utilizados para descrever uma distribuição Gaussiana e representam o

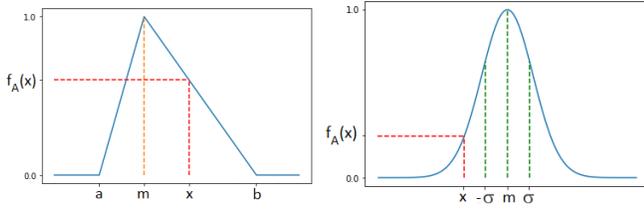


Fig. 5. Ilustração das funções de pertinências triangular (esquerda) e Gaussiana (direita), descritas respectivamente pelas Equações 2 e 3. Fonte: Elaborado pelo Autor.

valor médio e o grau de dispersão dos dados, respectivamente, conforme é ilustrado pela Fig. 5.

Para a construção dos conjuntos de pertinência para cada classe, inicialmente foi necessário dividir a base de dados em grupos de treinamento e teste, selecionando aleatoriamente 70% das amostras para treinar o modelo e 30% para avaliá-lo. Após o grupo de dados para treinamento ser definido, um particionamento segundo cada uma das classes foi realizado. Como resultado, foi possível obter um conjunto de treinamento para cada material. Com isso, tomando a matriz de atributos de um conjunto de treinamento (X) de uma determinada classe os conjuntos de pertinência triangular são construídos para cada atributo e definidos pelo menor valor (parâmetro a - Equação (4)), pela média (parâmetro m - Equação (5)) e pelo maior valor (parâmetro b - Equação (6)) presente na coluna referente à este atributo na matrix X . Já a construção dos conjuntos de pertinência Gaussiana é definida pela média (parâmetro m - Equação (5)) e desvio-padrão (parâmetro σ - Equação (7)) dos valores que um determinado atributo apresenta na matrix X . Sendo assim, para cada classe de material os parâmetros a , m , b e σ das Equações (2) e (3) são calculados para cada atributo (subíndice j) de acordo com as Equações (4), (5), (6) e (7):

$$a_j = \min(X_j) \quad (4)$$

$$m_j = \frac{1}{N} \sum_{i=1}^N x_{ij} \quad (5)$$

$$b_j = \max(X_j) \quad (6)$$

$$\sigma_j = \sqrt{\frac{\sum_{i=1}^N (x_{ij} - m_j)^2}{N}} \quad (7)$$

onde $i = 1, \dots, N$ representa o índice das amostras do conjunto de treinamento de cada material e j se refere aos atributos, cujo valor máximo é definido pelo número de colunas do conjunto de características em análise: C_1 , C_2 ou C_3 .

A identificação das classes dos materiais foi feita atribuindo-se os valores discretos aos rótulos: $B8 = 1$, $B10 = 2$ e $B20 = 3$. Dessa forma, foi possível tomar a acurácia como a principal métrica de desempenho devido à sua maior simplicidade de entendimento, uma vez que pode ser definida como a porcentagem de predições corretas das classes, sendo matematicamente descrita pela Equação (8).

$$Acuracia = \frac{1}{N} \sum_{k=1}^N I(f(x_k) = c_k) \quad (8)$$

onde $f(x_k)$ é a classe predita para a amostra de teste, c_k é a classe verdadeira desta amostra. Uma diferença nula entre a classe predita e a classe real significa acerto $I(verdadeiro) = 1$ e diferença não nula significa erro $I(falso) = 0$.

Vale ressaltar que a quantidade de classificações realizadas para cada observação em uma execução independente é exatamente o tamanho da dimensão do conjunto de características em questão, de modo que a decisão final fica determinada pela classe mais frequentemente atribuída para tal amostra. Além disso, ficou determinado que, em caso de empate, a classe de maior valor de rótulo dentre aquelas em situação de empate seria o rótulo final atribuído.

IV. RESULTADO EXPERIMENTAIS

Os experimentos computacionais foram conduzidos em *Python* e baseados nos pacotes *pandas* [34], *numpy* [35], *scikit-learn* framework [36], *seaborn* [37], *scipy* [38] e *TS-FEL* [20]. A base de dados gerada para este trabalho está disponível no repositório *Mendeley Data* [39] através do link <https://data.mendeley.com/datasets/9tytvdx24/1>. Além disso, todos os códigos são disponibilizados pelos autores mediante solicitação.

Os resultados da média e desvio-padrão das métricas de desempenho foram calculados a partir de 100 execuções independentes do algoritmo para ambas as funções de pertinência e para os três conjuntos de atributos previamente definidos. A Tabela VI mostra os resultados das acurácias médias e seus desvios-padrões obtidos. Para melhor visualização, um gráfico *box-plot* foi gerado para ilustrar o comportamento das acurácias, como mostrado na Fig. 6. As Fig. 7, 8 e 9 mostram as matrizes de confusão [40] médias para cada conjunto de características utilizado.

TABELA VI
ACURÁCIA MÉDIA COM DESVIO-PADRÃO PARA CADA CONJUNTO DE CARACTERÍSTICAS CONSIDERANDO FUNÇÕES DE PERTINÊNCIA TRIANGULAR E GAUSSIANA. O TEXTO EM NEGRITO IDENTIFICA O MELHOR RESULTADO MÉDIO OBTIDO.

Conjunto de Características	Função de Pertinência	Acurácia Média (%)	Desvio Padrão (%)
C_1	Triangular	75,90	4,25
	Gaussiana	77,11	4,23
C_2	Triangular	94,04	2,52
	Gaussiana	94,72	2,51
C_3	Triangular	82,90	3,85
	Gaussiana	84,92	3,99

Observando-se os resultados apresentados, é notória a superioridade dos MFCCs (conjunto C_2). As acurácias médias obtidas, 94,04% e 94,72%, respectivamente, para as funções de pertinência triangular e Gaussiana, mostram que este conjunto de características é bastante promissor para a classificação de pratos de bateria, pois conseguiu lidar bem com as variações de ambientes e de microfones impostas na aquisição dos áudios.

Adicionalmente, ao tomar os resultados alcançados por C_1 e C_3 , é possível notar que o conjunto de atributos puramente temporal (C_1) foi aquele que atingiu os menores desempenhos médios, ao passo que o conjunto formado pela união das

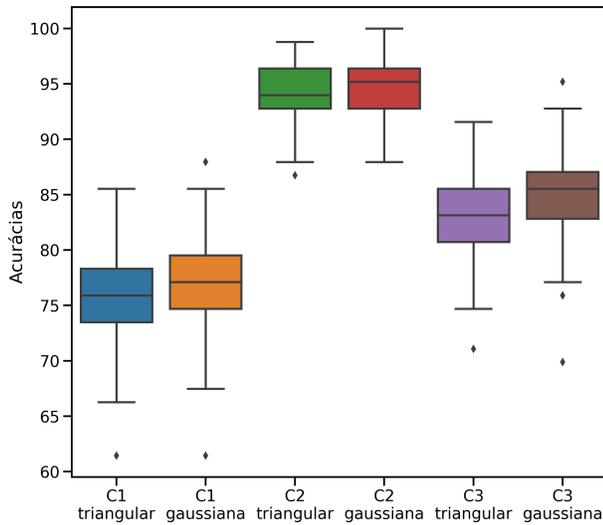


Fig. 6. Box-plot das acurácias para cada conjunto de características e função de pertinência.

características temporais com os MFCC, representado por C_3 , obteve desempenhos médios intermediários.

Dessa forma, pode-se entender que, dentre os conjuntos de atributos utilizados, o C_2 é o de maior relevância para o problema, uma vez que obteve, individualmente, os melhores desempenhos e, ao se associar ao conjunto C_1 , foi responsável por elevar os índices médios de desempenho deste último conjunto.

Com relação às funções de pertinência, a função Gaussiana apresentou, de maneira determinística, desempenhos de acurácia ligeiramente superiores (1,21% para C_1 , 0,68% para C_2 e 2,02% para C_3) aos obtidos pela função triangular. Aliado a isso, ainda que estatisticamente os resultados não aparentem mostrar uma diferença significativa entre as acurácias das funções de pertinência, o teste de hipótese ANOVA verificou que tal diferença é relevante para dois dos três casos, conforme mostrado pela Tabela VII. Portanto, pode-se afirmar que a função de pertinência Gaussiana é a mais indicada para o problema.

TABELA VII

RESULTADOS DOS TESTES ANOVA ENTRE AS FUNÇÕES DE PERTINÊNCIA TRIANGULAR E GAUSSIANA PARA CADA CONJUNTO DE CARACTERÍSTICAS. O VALOR P OBTIDO É SIGNIFICANTE PARA $p < 0,05$.

Conjunto	Valor F	Valor P
C_1	4,00162	0,04682
C_2	3,67891	0,05654
C_3	13,02149	0,00039

Ademais, a função de pertinência triangular está susceptível a um erro de classificação causado quando o valor de x na Equação (2) de uma amostra for inferior ao mínimo valor possível de a da primeira classe de pertinência ou superior ao valor máximo possível de b da terceira classe, mesmo que a classificação final seja determinada por todas as classificações individuais de cada atributo. Este é outro motivo para que a função de pertinência Gaussiana seja preferível para esta

aplicação, uma vez que leva em consideração apenas a média e o desvio-padrão de cada atributo das amostras de treinamento, não apresentando tal tipo de comportamento.

Em certa medida, era esperado que os maiores erros de classificação estivessem associados à liga de bronze B10, justamente por ser um material intermediário entre as ligas B8 e B20. Indo além, esperava-se que esta confusão fosse mais frequente com o bronze B8, por serem materiais mais similares entre si, compreendendo uma diferença de 2% de estanho em peso. No entanto, apesar de as matrizes de confusão das Fig. 7, 8 e 9 mostrarem que, de fato, os erros de classificação acontecem mais frequentemente para os bronzes B10, apenas os resultados para o conjunto de características C_2 apresentou a maior confusão entre B8 e B10.

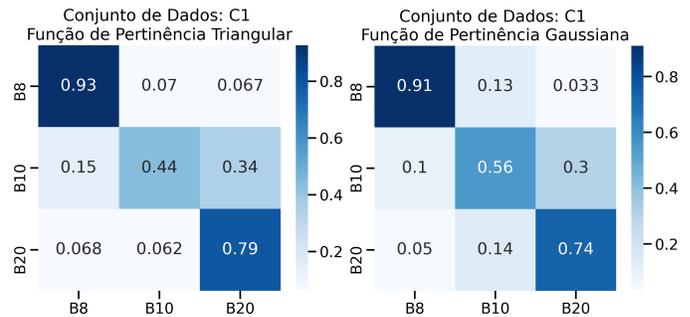


Fig. 7. Matriz de confusão média para 100 execuções independentes considerando o conjunto de atributos C_1 . Esquerda: Função de pertinência triangular. Direita: Função de pertinência Gaussiana.

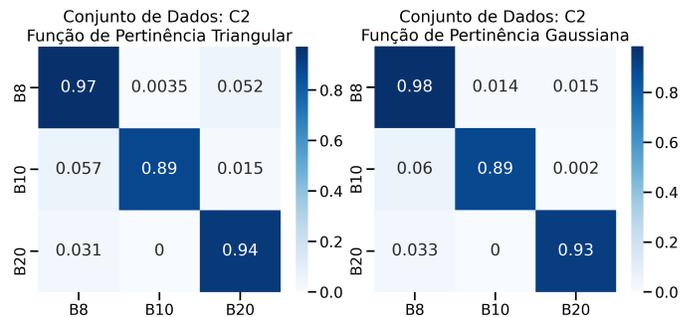


Fig. 8. Matriz de confusão média para 100 execuções independentes considerando o conjunto de atributos C_2 . Esquerda: Função de pertinência triangular. Direita: Função de pertinência Gaussiana.

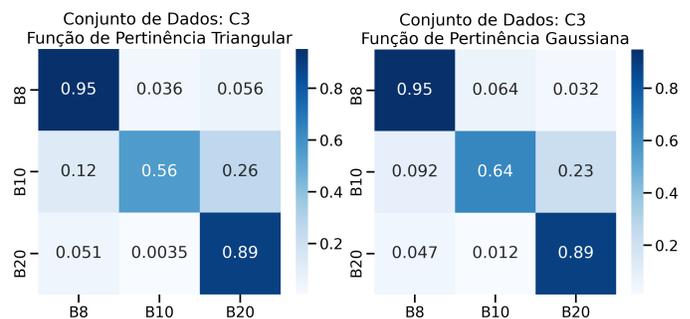


Fig. 9. Matriz de confusão média para 100 execuções independentes considerando o conjunto de atributos C_3 . Esquerda: Função de pertinência triangular. Direita: Função de pertinência Gaussiana.

A. Vantagens e Limitações

Apesar dos resultados promissores obtidos neste estudo e do procedimento de aquisição dos sons ter considerado diferentes perspectivas de captação, os autores enfatizam a necessidade de expandir o conjunto de dados em número de amostras e em variedade de pratos. Como resultado, é possível ter uma avaliação mais abrangente tanto do modelo, quanto da influência do material sobre as características acústicas destes instrumentos.

O procedimento experimental desenvolvido para padronizar a coleta dos áudios, embora ainda permita ser otimizado, provou ser eficiente, visto o elevado desempenho de classificação obtido. Por outro lado, a captação dos sons foi realizada em um ambiente controlado que pode não representar uma situação de um problema real, em que podem ter presença de ruídos indesejados.

Os resultados obtidos são promissores e encorajam o aprofundamento da investigação deste problema, abrindo margem para avaliar outros modelos de classificação na busca por melhores desempenhos. Isso aumentaria a confiança para que empresas do ramo possam utilizar este método de análise para controlar a qualidade do material bruto utilizado (o que inclui detecção de fraude) ou até mesmo para dar uma garantia da liga de bronze utilizada na fabricação de seus pratos.

V. CONCLUSÕES

Este trabalho propôs a classificação de pratos de bateria conforme as suas respectivas ligas de bronze e investigou a utilização da lógica *fuzzy* como classificador, avaliando diferentes funções de pertinências em três conjuntos de atributos. Um procedimento desenvolvido para a aquisição dos áudios possibilitou a captação dos sons de quatro pratos de bateria considerando dois microfones e dois ambientes distintos. Atributos temporais e coeficientes mel cepstrais de frequência foram extraídos dos sinais e utilizados, individual e simultaneamente, como variáveis de entrada para treinar e avaliar o modelo de classificação, que considerou funções de pertinência triangular e Gaussiana.

Ainda que esta linha de pesquisa seja pouca investigada, o desempenho de classificação obtido neste estudo encoraja o aprofundamento e o desenvolvimento de novas pesquisas que busquem utilizar o aprendizado de máquina como ferramenta para examinar os principais elementos que influenciam na acústica dos pratos de bateria. Assim, pode-se dizer que as maiores contribuições deste artigo se resumem a:

- 1) Promover uma inicialização do uso de aprendizado de máquina para o estudo dos elementos acústicos em pratos de bateria
- 2) Propor um procedimento experimental para a aquisição dos áudios de pratos de bateria
- 3) Aplicar e explorar um modelo de classificação fuzzy neste tipo de problema

Os resultados experimentais mostraram uma superioridade do MFCC, visto que foi capaz de alcançar, individualmente, as melhores métricas de desempenho, além de ter sido o responsável por elevar os desempenhos do classificador utilizando os atributos puramente temporais, após associar-se a eles. Com

relação às funções de pertinência, a Gaussiana demonstrou-se ligeiramente superior, em termos de acurácia média. Sendo assim, o modelo implementado conseguiu atingir uma acurácia média de 94,72% com um desvio-padrão de 2,51% quando se considerou a função de pertinência Gaussiana e o MFCC para recuperar informações dos áudios.

AGRADECIMENTOS

Os autores agradecem o Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) (Processos 429639/2016-3 e 304329/2019-3), a Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG) (Processo TEC-PPM 00001-18), a Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) - Código de Financiamento 001 - pelo apoio financeiro para o desenvolvimento do presente trabalho, o estúdio ARGUS por disponibilizar o equipamento e a estrutura para a aquisição dos áudios e as colaborações do Grupo de Pesquisa em Ciência de Dados Aplicada à Engenharia (CIDENG-CNPq) e Grupo de Modelagem Computacional Aplicada (GMCA).

REFERENCES

- [1] T. H. A. Boratto, R. F. Marcomini, L. Goliatt, C. R. Pagotto, A. A. Cury, I. J. U. Pereira, and F. D. Nishida, "Análise dos efeitos de dois diferentes métodos de fabricação em pratos de bateria," in *XI Congresso Brasileiro de Engenharia de Fabricação (COBEF)*, 2021.
- [2] T. Rossing, *Springer handbook of acoustics*. Springer Handbooks, New York, NY: Springer New York, 2nd ed., 2014.
- [3] J. J. Harrison and A. J. Hill, "A scientific approach to microphone placement for cymbals in live sound," 2013.
- [4] H. Pinksterboer and R. Mattingly, *The Cymbal Book*. Musical Instruments Series, Hal Leonard Publishing Corporation, 1992.
- [5] K. Osamura, F. Kuratani, T. Koide, W. Ogawa, H. Taniguchi, Y. Monju, T. Mizuta, and T. Shobu, "The correlation between the percussive sound and the residual stress/strain distributions in a cymbal," *Journal of Materials Engineering and Performance*, vol. 25, no. 10, 2016.
- [6] F. Kuratani, T. Yoshida, T. Koide, T. Mizuta, and K. Osamura, "Understanding the effect of hammering process on the vibration characteristics of cymbals," *Journal of Physics: Conference Series*, vol. 744, p. 012110, sep 2016.
- [7] W. OGAWA, F. KURATANI, T. YOSHIDA, T. KOIDE, and T. MIZUTA, "Effect of bell size on sound characteristics of cymbals," *Transactions of the JSME (in Japanese)*, vol. 86, no. 881, pp. 19–00237–19–00237, 2020.
- [8] S. Slamet, D. Suyitno, I. Kusumaningtyas, and I. Miasa, "Effect of high-tin bronze composition on physical, mechanical, and acoustic properties of gamelan materials," 08 2021.
- [9] I. K. G. Sugita, R. Soekrisno, and D. Suyitno, "Mechanical and damping properties of silicon bronze alloys for music applications," *Int J Eng & Technol IJET-IJENS*, vol. 11, pp. 98–105, 01 2011.
- [10] T. M. Mitchell, *Machine Learning*. MGH, 1997.
- [11] P. Herrera, A. Yeterian, and F. Gouyon, "Automatic classification of drum sounds: A comparison of feature selection methods and classification techniques," in *Music and Artificial Intelligence* (C. Anagnostopoulou, M. Ferrand, and A. Smaill, eds.), (Berlin, Heidelberg), pp. 69–80, Springer Berlin Heidelberg, 2002.
- [12] J. G. Cleary and L. E. Trigg, "K*: An instance-based learner using an entropic distance measure," in *Machine Learning Proceedings 1995* (A. Prieditis and S. Russell, eds.), pp. 108–114, San Francisco (CA): Morgan Kaufmann, 1995.
- [13] S. Cavaco and H. Almeida, "Automatic cymbal classification using non-negative matrix factorization," in *2012 19th International Conference on Systems, Signals and Image Processing (IWSSIP)*, pp. 468–471, 2012.
- [14] V. M. A. Souza, G. E. A. P. A. Batista, and N. E. Souza-Filho, "Automatic classification of drum sounds with indefinite pitch," in *Proceedings of International Joint Conference on Neural Networks (IJCNN)*, 2015.
- [15] B. Y. George J. Klir, *Fuzzy Sets and Fuzzy Logic: Theory and Applications*. Prentice Hall, 1995.

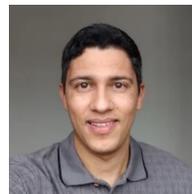
- [16] B. Logan, "Mel frequency cepstral coefficients for music modeling," *Proc. 1st Int. Symposium Music Information Retrieval*, 11 2000.
- [17] M. Slaney, "Auditory Toolbox Version 2," Technical Report #1998-010, Interval Research Corporation, 1998.
- [18] AKG, "C414 XLII - Reference multipattern condenser microphone." Last accessed 28 March 2021.
- [19] Shure, "PGA81 - Microfone condensador cardioide para instrumento." Last accessed 02 April 2021.
- [20] M. Barandas, D. Folgado, L. Fernandes, S. Santos, M. Abreu, P. Bota, H. Liu, T. Schultz, and H. Gamboa, "Tsfel: Time series feature extraction library," *SoftwareX*, vol. 11, p. 100456, 2020.
- [21] A. V. Oppenheim and R. W. Schaffer, "From frequency to quefrequency: a history of the cepstrum," *IEEE Signal Processing Magazine*, vol. 21, no. 5, pp. 95–106, 2004.
- [22] L. Muda, M. Begam, and I. Elamvazuthi, "Voice recognition algorithms using mel frequency cepstral coefficient (MFCC) and dynamic time warping (DTW) techniques," *CoRR*, vol. abs/1003.4083, 2010.
- [23] D. Bhalke, C. R. Rao, and D. Bormane, "Hybridisation of mel frequency cepstral coefficient and higher order spectral features for musical instruments classification," *Archives of Acoustics*, vol. 41, no. 3, 2016.
- [24] R. Loughran, J. Walker, M. O'Farrell, and M. O'Neill, "Comparison of features in musical instrument identification using artificial neural networks," in *5th International Symposium on Computer Music Modeling and Retrieval*, pp. 19–33, Springer, 2008.
- [25] R. Loughran, J. Walker, M. O'Neill, and M. O'Farrell, "The use of mel-frequency cepstral coefficients in musical instrument identification," Feb 2018.
- [26] M. Malík and R. Orješek, "The comparison of selected audio features and classification techniques in the task of the musical instrument," 05 2016.
- [27] M. S. Nagawade and V. R. Ratnaparkhe, "Musical instrument identification using mfcc," in *2017 2nd IEEE International Conference on Recent Trends in Electronics, Information Communication Technology (RTEICT)*, pp. 2198–2202, 2017.
- [28] MATLAB, "Extract cepstral features from audio segment." Last accessed 28 April 2021.
- [29] Alan V. Oppenheim, Ronald W. Schaffer, and John R. Buck, *Discrete-Time Signal Processing*. Prentice-Hall Signal Processing Series, Prentice Hall, 2nd ed ed., 1999.
- [30] J. P. Crutchfield and D. P. Feldman, "Regularities unseen, randomness observed: Levels of entropy convergence," *Chaos: An Interdisciplinary Journal of Nonlinear Science*, vol. 13, no. 1, pp. 25–54, 2003.
- [31] L. Zadeh, "Fuzzy sets," *Information and Control*, vol. 8, no. 3, pp. 338–353, 1965.
- [32] V. Cherkassky, "Fuzzy inference systems: A critical review," in *Computational Intelligence: Soft Computing and Fuzzy-Neuro Integration with Applications* (O. Kaynak, L. A. Zadeh, B. Türkşen, and I. J. Rudas, eds.), (Berlin, Heidelberg), pp. 177–197, Springer Berlin Heidelberg, 1998.
- [33] E. M. Jyh-Shing Roger Jang, Chuen-Tsai Sun, *Neuro-Fuzzy and Soft Computing: A Computational Approach to Learning and Machine Intelligence*. Prentice Hall, 1 ed., 1997.
- [34] Wes McKinney, "Data Structures for Statistical Computing in Python," in *Proceedings of the 9th Python in Science Conference* (Stéfan van der Walt and Jarrod Millman, eds.), pp. 56 – 61, 2010.
- [35] C. R. Harris, K. J. Millman, S. J. van der Walt, R. Gommers, P. Virtanen, D. Cournapeau, E. Wieser, J. Taylor, S. Berg, N. J. Smith, R. Kern, M. Picus, S. Hoyer, M. H. van Kerkwijk, M. Brett, A. Haldane, J. F. del Río, M. Wiebe, P. Peterson, P. Gérard-Marchant, K. Sheppard, T. Reddy, W. Weckesser, H. Abbasi, C. Gohlke, and T. E. Oliphant, "Array programming with NumPy," *Nature*, vol. 585, pp. 357–362, Sept. 2020.
- [36] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, et al., "Scikit-learn: Machine learning in python," *The Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.
- [37] M. L. Waskom, "seaborn: statistical data visualization," *Journal of Open Source Software*, vol. 6, no. 60, p. 3021, 2021.
- [38] P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, S. J. van der Walt, M. Brett, J. Wilson, K. J. Millman, N. Mayorov, A. R. J. Nelson, E. Jones, R. Kern, E. Larson, C. J. Carey, I. Polat, Y. Feng, E. W. Moore, J. VanderPlas, D. Laxalde, J. Perktold, R. Cimrman, I. Henriksen, E. A. Quintero, C. R. Harris, A. M. Archibald, A. H. Ribeiro, F. Pedregosa, P. van Mulbregt, and SciPy 1.0 Contributors, "SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python," *Nature Methods*, vol. 17, pp. 261–272, 2020.
- [39] T. H. Boratto, A. Cury, and L. Goliatt, "Crash cymbal sounds." <https://data.mendeley.com/datasets/9tytvdx24/1>, 2022.
- [40] A. Géron, *Hands-On Machine Learning with Scikit-Learn and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems*. O'Reilly Media, 1 ed., 2017.



Tales H. A. Boratto Mestrando em Modelagem Computacional pela Universidade Federal de Juiz de Fora. Possui graduação em Ciências Exatas (2019) e em Engenharia Mecânica (2020) pela Universidade Federal de Juiz de Fora. Suas pesquisas envolvem as áreas de Ciência dos Materiais, com ênfase em Ligas de Bronze, Processos de Fabricação, Recuperação de Informação Musical (MIR) e Aprendizado de Máquina.



Alexandre A. Cury Graduado em Engenharia Civil pela Universidade Federal de Juiz de Fora (2006), mestre em Modelagem Computacional pela Universidade Federal de Juiz de Fora (2008) e doutor em Engenharia Civil pela Universidade Paris-Est (École Nationale des Ponts et Chaussées) em 2010. Atua, principalmente, nos seguintes temas: monitoramento de integridade estrutural, análise de vibrações, detecção de danos, identificação modal, análise estatística avançada e confiabilidade estrutural.



Leonardo Goliatt Possui graduação em Engenharia Civil pela Universidade Federal de Juiz de Fora (2003) e Doutorado em Modelagem Computacional pelo Laboratório Nacional de Computação Científica (2009). Professor associado do Departamento de Mecânica Aplicada e Computacional e membro permanente do Programa de Pós-Graduação em Modelagem Computacional da Universidade Federal de Juiz de Fora. Atua na área de Computação Evolucionária, Inteligência Computacional e Ciência de Dados, atuando principalmente nos seguintes temas: aprendizado de máquina, meta heurísticas, metamodelos, otimização estrutural e simulação.