

Hugo Gonçalves Lopes

**Forró para surdos: análise das faixas de
frequência do conjunto de entradas de uma
rede neural artificial**

Viçosa, MG

2022

Hugo Gonçalves Lopes

Forró para surdos: análise das faixas de frequência do conjunto de entradas de uma rede neural artificial

Monografia apresentada ao Departamento de Engenharia Elétrica do Centro de Ciências Exatas e Tecnológicas da Universidade Federal de Viçosa, para a obtenção dos créditos da disciplina ELT 402 – Projeto de Engenharia II – e cumprimento do requisito parcial para obtenção do grau de Bacharel em Engenharia Elétrica.

Orientador: Rodolpho Vilela Alves Neves

Viçosa, MG

2022

HUGO GONÇALVES LOPES

**FORRÓ PARA SURDOS: ANÁLISE DAS FAIXAS DE
FREQUÊNCIA DO CONJUNTO DE ENTRADAS DE UMA REDE
NEURAL ARTIFICIAL**

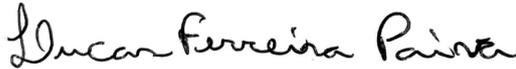
Monografia apresentada ao Departamento de Engenharia Elétrica do Centro de Ciências Exatas e Tecnológicas da Universidade Federal de Viçosa, para a obtenção dos créditos da disciplina ELT 402 – Projeto de Engenharia II e cumprimento do requisito parcial para obtenção do grau de Bacharel em Engenharia Elétrica.

Aprovada em 02 de março de 2022.

COMISSÃO EXAMINADORA



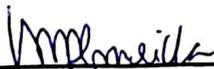
Prof. Dr. Rodolpho Vilela Alves Neves - Orientador
Universidade Federal de Viçosa



Lucas Ferreira Paiva - Coorientador
Universidade Federal de Viçosa

 Documento assinado digitalmente
DANIEL KHEDE DOURADO VILLA
Data: 02/03/2022 15:11:35-0300
Verifique em <https://verificador.iti.br>

Prof. Me. Daniel Khede Dourado Villa - Membro Avaliador
Universidade Federal de Viçosa



Prof. Me. Vinicius Martins Almeida - Membro Avaliador
Centro Universitário de Viçosa - Univiçosa

*Este trabalho é dedicado às crianças adultas que,
quando pequenas, sonharam em se tornar cientistas.*

Agradecimentos

Os agradecimentos principais são direcionados à Lucas Paiva, por fornecer o banco de dados utilizado neste projeto e ajudar em diversos momentos de dificuldades, ao professor Bonato, à todos os integrantes do Núcleo Interdisciplinar de Sinais (NIAS) e ao professor Rodolpho, orientador deste trabalho, por ajudarem na construção da apresentação e do trabalho.

Agradecimentos especiais são direcionados aos meus amigos e familiares pelo apoio emocional e financeiro ao longo de todo o curso de graduação, essenciais para que eu não desistisse apesar das dificuldades.

“Suba o primeiro degrau com fé. Não é necessário que você veja toda a escada. Apenas dê o primeiro passo.” (Martin Luther King)

Resumo

A dança é um dos grandes momentos de interação social na natureza, seja na busca de um parceiro ou em momentos de descontração. Para os seres humanos, a dança está altamente associada a musicalidade e ao ritmo das canções, o que implica que os participantes da dança precisam ouvir a música para interagir entre si. Entretanto, pessoas surdas ou com alguma deficiência auditiva são muitas vezes excluídas deste tipo de ambiente de interação social, dificultando ainda mais convivência delas na sociedade. Esta dificuldade de inserção social pode resultar em problemas ainda mais graves, como a depressão, chegando até a tendências suicidas. Logo, se faz necessário o uso de técnicas como uma RNA para converter sinais sonoros em outros que S/DA possam perceber. Esta técnica já existe na literatura, utilizando a transformada de Fourier em parte do espectro de frequência da música como entrada de uma RNA para determinar o tempo do passo base de músicas de forró, porém, não foi analisado como as variações na entrada da rede afetam o desempenho da mesma. Portanto, este trabalho propõe a variação do número de entradas e do espectro de frequência para determinar a melhor extração de características da música. Com isso, definiu-se uma extração de características de áudio que reduziu o erro em mais de 20% em comparação com o erro da técnica existente na literatura que pode ser utilizada para a aplicação em uma ferramenta capaz de inserir surdos e deficientes auditivos no ambiente de dança de forró, proporcionando uma melhor convivência destes indivíduos na sociedade local.

Palavras-chaves: extração de características; forró; surdos; rede neural.

Abstract

Dancing is one of the great moments of social interaction in nature, whether in search of a partner or in moments of relaxation. For humans, dance is highly associated with musicality and the rhythm of songs, which implies that dance participants need to hear the music in order to interact with each other. However, deaf or hearing impaired people are often excluded from this type of social interaction environment, making it even more difficult for them to live in society. This difficulty in social insertion can result in even more serious problems, such as depression, and even suicidal tendencies. Therefore, it is necessary to use techniques such as an ANN to convert sound signals into others that D/HH can perceive. This technique already exists in the literature, using the Fourier transform in part of the frequency spectrum of the music as input of an ANN to determine the base step tempo of forró music, however, it was not analyzed how the variations in the network input affect its performance. Therefore, this work proposes the variation of the number of inputs and the frequency spectrum to determine the best extraction of music characteristics. With this, an extraction of audio features was defined that reduced the error by more than 20% compared to the error of the technique existing in the literature that can be used for application in a tool capable of inserting the deaf and hearing impaired in the forró dance environment, providing a better coexistence of these individuals in the local society.

Key-words: feature extraction; forró; deaf; neural network.

Lista de figuras

Figura 1 – Relação de pensamentos suicidas entre pessoas surdas e ouvintes entre universitários.	11
Figura 2 – Notação de dança do passo base do forró, a cor roxa (mais escura) representa em qual pé o peso do corpo deve estar apoiado. Adaptado de Santos et al. (2018).	14
Figura 3 – Fluxograma do trabalho realizado em Paiva et al. (2020)	16
Figura 4 – Fluxograma do processo de obtenção e comparação das redes.	20
Figura 5 – EPAM e Coeficientes de Pearson para cada número de entradas.	22
Figura 6 – EPAM e Coeficientes de Pearson para cada faixa de frequência.	23
Figura 7 – Espectrograma de uma das músicas utilizadas.	24

Lista de abreviaturas e siglas

S/DA	Surdos e Deficientes Auditivos
EPAM	Erro Percentual Absoluto Médio
FFT	Fast Fourier Transform (Transformada Rápida de Fourier)
PMC	Perceptron MultiCamadas
RNA	Rede Neural Artificial
EQM	Erro Quadrático Médio

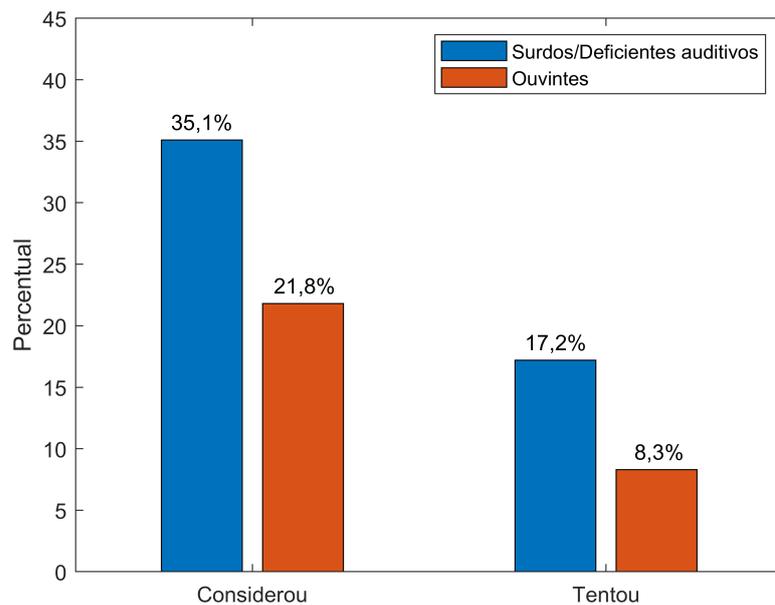
Sumário

1	INTRODUÇÃO	11
1.1	Objetivos	13
2	REFERENCIAL TEÓRICO	14
2.1	Forró	14
2.2	Transformada discreta de Fourier	15
2.3	Redes neurais artificiais (RNA)	15
2.4	Trabalhos existentes na literatura	16
3	MATERIAIS E MÉTODOS	18
3.1	Materiais	18
3.2	Geração de entradas	18
3.3	Rede PMC	18
4	RESULTADOS E DISCUSSÃO	21
4.1	Aumento de Entradas	21
4.2	Aumento do Espectro de Frequência	22
5	CONSIDERAÇÕES FINAIS	25
	REFERÊNCIAS	26

1 Introdução

O número de pessoas com deficiência auditiva no Brasil somam 9,7 milhões, segundo dados do Censo de 2010 do Instituto Brasileiro de Geografia e Estatística (IBGE), no entanto, no país, é comum que os surdos encontrem dificuldade para se comunicarem, seja com pessoas ouvintes ou surdas, gerando obstáculos que podem ter implicações no avanço social, emocional e cognitivo. À vista disso, a população surda pode ter sua saúde mental lesada (CHAVEIRO et al., 2014). Vários fatores podem dificultar a inclusão social e limitar a participação desses indivíduos em grupos sociais (SANTOS; VIEIRA; FARIA, 2013). De acordo com Fox, James e Barnett (2020), universitários surdos e deficientes auditivos têm uma taxa maior de pensamentos e tentativas de suicídio quando comparado com ouvintes. Os dados dessa pesquisa podem ser visualizados na Fig. 1.

Figura 1 – Relação de pensamentos suicidas entre pessoas surdas e ouvintes entre universitários.



Fonte: Adaptado de Fox, James e Barnett (2020).

É no sentido, de tornar realidade, a inclusão de surdos que foram criadas leis como a Lei Brasil (2002), que reconhece a Língua Brasileira de Sinais (Libras) como um meio legítimo de comunicação e expressão para os surdos, além do Decreto Brasil (2005) que regulamenta a Lei citada, e ainda trás em seu Capítulo II, Art. 3º a incorporação de Libras às disciplinas curriculares mesmo que a Libras se torne disciplina obrigatória nos cursos de formação de professores do ensino médio e superior e nos cursos de fonoaudiologia nas instituições de ensino públicas e privadas, na rede federal de ensino e nos sistemas de ensino

dos municípios, estados e distrito federal. No Capítulo 4, trata do uso e divulgação da Libras e do Português para a educação de surdos, e no Capítulo 5, trata da obrigatoriedade da formação de tradutores e intérpretes de Libras-Português. Ainda neste decreto, o capítulo VI garante o direito à educação das pessoas surdas ou com deficiência auditiva e o capítulo VII trata da garantia do direito à saúde das pessoas surdas ou com deficiência auditiva (BRASIL, 2005).

Neste contexto, este trabalho busca somar à inclusão do surdo na sociedade, baseado na comunidade universitária, dialogando com Lacerda (2006), que diz “além de práticas que se preocupam com a aprendizagem dos conteúdos curriculares, a inclusão do surdo perpassa pela sua inserção na cultura local”. E, dessa forma, entende-se que, a dança, tem forte presença em muitas culturas pelo mundo, e que dançar é uma das atividades aeróbicas grande influenciadora para um autoconhecimento, e é através dela que se busca uma facilidade para lidar com problemas do dia a dia, mudando a autoestima desde crianças até a terceira idade (FREITAS, 2019).

No Brasil, o forró, uma manifestação artístico-cultural, que vai além da dança e da música, acabou por conquistar todos os segmentos da sociedade com o seu ritmo contagiante e animado, principalmente a comunidade universitária. Existem vários programas que promovem aulas de dança e atividades culturais onde se dá a prática deste estilo de dança, a integração e a troca de conhecimentos entre os alunos. E mesmo sendo reduzido o número de surdos no ambiente universitário, e destes serem estigmatizados como seres não musicais (PAULA; PEDERIVA, 2017; HAGUIARA-CERVellini, 2003), alguns alunos surdos participam desses espaços. Na Universidade Federal de Viçosa (UFV), o projeto “Dança nas Moradias” é um exemplo onde, alunos surdos participam da aula de dança de forró e samba de gafieira, oferecidas por alunos do Curso de Dança da UFV para os moradores dos alojamentos.

Na literatura atual, há trabalhos mostrando que é possível potencializar o contato do surdo com a música através de estímulos visuais e táteis. Em Sharp et al. (2019) é feito um estudo se pessoas são capazes de determinar emoções através de estímulos táteis e/ou sonoros e é expandido em Sharp, Bacon e Champoux (2020) para surdos, Bossey (2020) realiza um estudo em festivais musicais do Reino Unido e sugere uma inclusão de surdos através de meios digitais, já em Mirzaei, Kán e Kaufmann (2020) é utilizado um fone de ouvido háptico para surdos serem capazes de localizar fontes sonoras em um ambiente virtual. Contudo, uma ferramenta de interpretação como uma Rede Neural Artificial (RNA) é imprescindível para que a música em tempo real possa ser transmitida por um meio que surdos e deficientes auditivos possam perceber, visto que é necessário a estimação do tempo através da compreensão de um sinal sonoro. Outra abordagem possível seria através da detecção de batidas da zabumba, entretanto a falha na detecção de uma batida resultaria no dobro do tempo estimado neste momento em uma aplicação em tempo real,

o que não ocorre na RNA pelas saídas estarem próximas às presentes no banco de dados.

A RNA é um algoritmo computacional treinado para classificar dados, aproximar funções e fazer predições (SILVA; SPATTI; FLAUZINO, 2016). A literatura sobre processamento de áudio e reconhecimento de padrões utilizando aprendizado de máquinas é extensa, podendo ser citados trabalhos como Salamon e Bello (2017) que utiliza Redes Neurais Convolucionais para a classificação do ambiente do áudio, já em Solanki e Pandey (2019), Yu et al. (2020) a rede neural é utilizada para identificação de instrumentos musicais e por fim, a identificação de emoções em músicas através de RNA é feita em Cai e Cai (2019). Estes trabalhos citados utilizam técnicas de redes neurais artificiais para a classificação de áudio. Porém, em Paiva et al. (2020) é apresentado um modelo de rede que é capaz de prever o tempo do compasso de uma música de forró, um grande avanço na criação de um aplicativo de sinalização do ritmo da música para a inclusão de surdos e deficientes auditivos no ambiente de forró, onde a música pode ser captada em tempo real por um microfone de um dispositivo móvel e então o compasso estimado através dessa rede neural.

Entretanto, em Paiva et al. (2020), para gerar as entradas da rede, as amostras foram submetidas à transformada de Fourier e, em seguida, foi feita a média das magnitudes das frequências a cada 10 Hz de 50 Hz à 300 Hz, resultando em 25 entradas para a rede PMC. Porém, essa redução de dados feita da saída da FFT para as entradas da rede também pode ter impacto no desempenho da mesma. Embora essa rede consiga estimar o compasso com erros próximos a 5%, também há o problema da mesma ter utilizado uma métrica para treino e outra para a avaliação de desempenho, o que não garante o menor valor para a métrica do teste. Em Paiva (2020) esta rede é aplicada em bancos de dados com diferentes naturezas de ruído, o que é necessário caso a mesma seja aplicada para a estimação em tempo real em um aplicativo móvel.

1.1 Objetivos

Este trabalho tem como objetivo geral analisar a geração das entradas a partir da transformada de Fourier para estimação do compasso de músicas de forró através de uma RNA. Para isso, tem-se como objetivos específicos:

- Replicar modelo de rede de Paiva (2020);
- Aumentar número de entradas para a faixa de frequência de 50 Hz à 300 Hz;
- Utilizar outras faixas de frequência para 50 entradas;
- Treinar redes para cada entrada de rede citada nos dois itens anteriores; e
- Comparar dados estatísticos para avaliar desempenho das diversas configurações de entrada.

2 Referencial Teórico

Este trabalho utilizou ferramentas de processamento de sinais e algoritmos que estão apresentados adiante neste capítulo. Também está presente a técnica para estimação do tempo do compasso de [Paiva et al. \(2020\)](#), onde não há uma variação das entradas para determinar a melhor extração de características.

2.1 Forró

Para [Junior e Volp \(2005\)](#), o gênero musical forró pode ser classificado em 3 categorias:

1. Forró Pé-de-serra

Originou-se no Nordeste brasileiro, por volta de 1940, com influência do ambiente rural do sertanejo.

2. Forró Universitário

Tem como origem o Forró Pé-de-serra, porém com influência de outros estilos musicais por jovens sulistas na década de 1990.

3. Forró Eletrônico

Também originado na década de 1990, surge a partir da utilização de instrumentos eletrônicos e com visual chamativo e linguagem estilizada.

O ritmo do forró é passado para o dançarino pelo ritmo da zabumba, onde quatro batidas definem o período de um compasso e dois compassos completam o período de um passo base. Os passos devem ser sincronizados de acordo com essas batidas de zabumba, como pode ser observado na Fig. 2 ([SANTOS et al., 2018](#); [SCHOENBERG, 1990](#)).

Figura 2 – Notação de dança do passo base do forró, a cor roxa (mais escura) representa em qual pé o peso do corpo deve estar apoiado. Adaptado de [Santos et al. \(2018\)](#).



Para este trabalho, a rede foi treinada para estimar o tempo de duração de um compasso (PAIVA et al., 2020), porém, segundo Schoenberg (1990) a duração de um passo completo consiste no tempo de dois compassos, logo a saída da rede deve ser multiplicada por 2 para sinalizar o passo completo.

2.2 Transformada discreta de Fourier

A transformada de Fourier permite que uma determinada onda possa ser representada por sua magnitude em função da frequência. No caso, será utilizada sua forma discreta descrita por

$$X_r = \sum_{n=0}^{N_0-1} x_n e^{-jr \frac{2\pi}{N_0} n}, \quad (2.1)$$

onde X_r e x_n representa o vetor de amostras na frequência e no tempo, respectivamente e N_0 o número de amostras, que ao invés de transformar um sinal contínuo no tempo, transforma um vetor de valores medidos no tempo (LATHI, 2007). Para a redução do volume de dados a ser tratado, a frequência de amostragem pode ser reduzida com o intuito de otimizar o tempo de processamento da transformada de Fourier (STEIJNS et al., 2020).

Porém, a Equação (2.1) retorna a magnitude na forma retangular, com parte real e imaginária, além de retornar valores proporcionais ao número de amostras do sinal e para o espectro positivo e negativo de frequência. Logo, a equação utilizada para obter a magnitude dos valores de frequência é representada por

$$X_r = \frac{2}{N_0} \left| \sum_{n=0}^{N_0-1} x_n e^{-jr \frac{2\pi}{N_0} n} \right|, \quad (2.2)$$

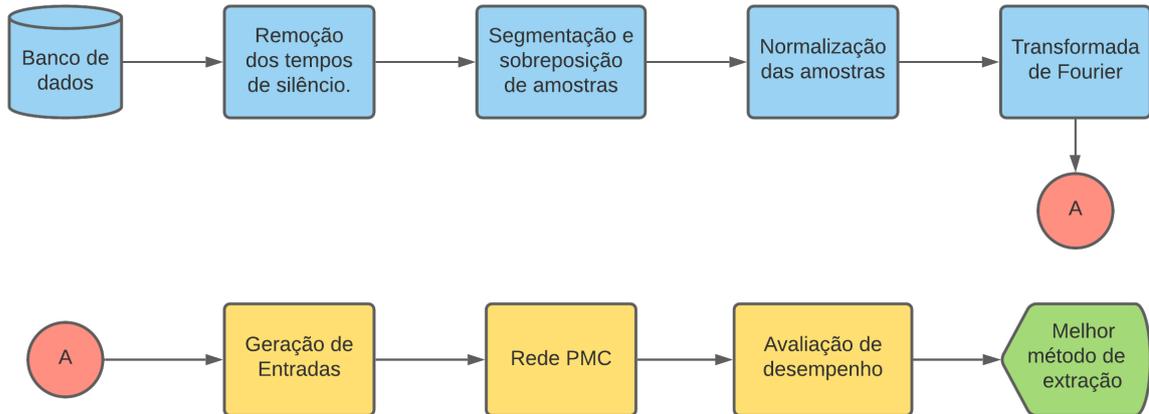
utilizando apenas metade dos valores da transformada, visto que os espectros positivos e negativos de frequência são espelhados.

2.3 Redes neurais artificiais (RNA)

Redes neurais artificiais são estruturas artificiais produzidas com o intuito de simular uma rede neural biológica. Através de algoritmos de treino é possível armazenar informações nessas redes para que a mesma realize determinada tarefa (HAYKIN, 2001).

O topologia de rede neural utilizada neste trabalho é a perceptron multicamadas, onde a rede possui várias camadas de neurônios, uma camada de entrada, uma ou mais camadas ocultas e enfim uma camada de saída (SILVA; SPATTI; FLAUZINO, 2016). Nessa topologia as informações são colocadas nos neurônios de entrada e transmitidas para a primeira camada oculta, onde serão multiplicadas pelos pesos obtidos no treinamento e passadas a seguir para próxima camada até a camada de saída através das funções de

Figura 3 – Fluxograma do trabalho realizado em Paiva et al. (2020)



ativação, que transmitem valores lógicos dependendo do valor de entrada (SILVA; SPATTI; FLAUZINO, 2016).

2.4 Trabalhos existentes na literatura

Em Paiva et al. (2020), é apresentado uma técnica para estimação de duração do compasso de músicas de forró, onde essa técnica é expandida para músicas com ruído real em Paiva (2020). Para garantir a generalização dos resultados, foram selecionadas músicas com tempos de compasso menores (mais rápidas) e maiores (mais lentas). Vale ressaltar que a seleção das músicas não realizou uma separação entre as categorias anteriormente citadas, por não ser necessário para o projeto. As músicas foram escolhidas por variedade rítmica e por popularidade no contexto universitário através da orientação de uma instrutora de dança.

A preparação das músicas para o banco de dados consiste em retirar o início e o final das músicas, segmentar e normalizar a amplitude do sinal em todo domínio do tempo. Cada música foi segmentada em trechos de três segundos com sobreposição de dois segundos entre os segmentos. Em seguida, a normalização foi feita dividindo-se cada valor pelo valor eficaz da amostra V_{rms} , definido por (2.3), para que eventual disparidade entre o volume das músicas não interfira nos resultados. A normalização deve ser feita por amostra em vista que no caso de previsão em tempo real, não haverá dados do futuro da música para a normalização, logo deve-se considerar essa informação inexistente na geração de amostras.

$$V_{rms} = \sqrt{\frac{1}{N} \sum_{i=1}^N V_i^2} \quad (2.3)$$

À cada segmento foi associado à duração do compasso da música a qual pertencia,

formando as entradas e saídas do banco de dados. Para isso, foi considerado que a duração do compasso das músicas escolhidas não varia ao longo da música. Essa consideração simplifica o método por permitir adotar a média dos tempos cronometrados para cada amostra de uma mesma música. Entretanto, a mesma consideração pode diminuir a capacidade de estimação da rede neural.

Com isso, gerou-se as entradas para uma RNA a partir da FFT de cada amostra, realizando a média das magnitudes para faixas de 10 Hz de 50 Hz à 300 Hz, que é a faixa onde o espectro de frequência da zabumba têm suas principais componentes. Essas entradas foram utilizadas para treinar uma rede PMC com uma camada oculta utilizando Levenberg–Marquardt como algoritmo de otimização, tendo como métrica de erro o EQM, sendo que posteriormente a avaliação do método é feito pelo Erro Percentual Absoluto Médio (EPAM), pois não necessariamente uma rede com o mínimo EQM terá também o mínimo EPAM. Os passos realizados podem ser visualizados na Fig. 3

3 Materiais e Métodos

A seguir estão presentes a descrição dos bancos de dados e métodos utilizados para a análise das variações de entradas de uma rede neural artificial.

3.1 Materiais

Os bancos de músicas utilizados para a realização deste projeto são os mesmos utilizados em [Paiva \(2020\)](#) com seus respectivos nomes e artistas ou bandas responsáveis pelas gravações das versões utilizadas, sendo eles:

- Banco de dados sem ruído contendo 82 músicas; e
- Banco de dados com ruído real contendo 37 músicas.

Após o banco já estar preparado com as amostras e saídas, o banco é dividido para a validação cruzada K-fold, sendo o valor de $K = 10$. Com isso, reduz-se o impacto de uma seleção de um banco de dados tendencioso para a rede.

3.2 Geração de entradas

Para avaliar o impacto que o número de entradas tem no desempenho da rede, diferentes números de entradas foram testados. Os valores escolhidos são 25, 50, 100 e 250, esses valores foram escolhidos com base nos valores resultantes da resolução em Hz das entradas, onde os valores de resolução para a faixa de 50 Hz à 300 Hz são 10 Hz, 5 Hz, 2,5 Hz e 1 Hz respectivamente.

Já para a avaliação da faixa de frequência, o número de entradas foi mantido fixo em 50 entradas pois ocorreu do modelo não convergir diversas vezes para os valores de 100 e 250 entradas, pois este aumento de entradas resultou em um excesso de variáveis para a rede ser treinada. As faixas escolhidas para teste foram 50 Hz à 300 Hz com base no espectro da zabumba ([PAIVA et al., 2020](#); [PAIVA, 2020](#)), 0 Hz à 300 Hz, 0 Hz à 800 Hz e 0 Hz à 1600 Hz, que nos gráficos de resultados estão representados pela variação de frequência de $\Delta_f = 250$, $\Delta_f = 300$, $\Delta_f = 800$, $\Delta_f = 1600$, respectivamente.

3.3 Rede PMC

A topologia de rede escolhida para este trabalho foi o perceptron multicamadas (PMC) com uma camada oculta, devido sua característica de aproximação universal de

funções (SILVA; SPATTI; FLAUZINO, 2016). Outra característica desta rede é necessitar de apenas uma camada oculta para mapear qualquer função contínua no espaço das funções reais, desde que, seja utilizada uma função de ativação contínua e limitada em sua imagem (SILVA; SPATTI; FLAUZINO, 2016).

O modelo utilizado para o treinamento da rede é semelhante à utilizada por Paiva (2020), que contém:

- Camada de entrada com normalização z-score, que mantém a média igual a 0 e desvio padrão igual a 1, sendo representado por

$$\mathbf{Z} = \frac{\mathbf{X} - \mu}{\sigma}, \quad (3.1)$$

onde \mathbf{X} é o batch utilizado de treinamento, μ e σ são a média e o desvio padrão de \mathbf{X} , respectivamente;

- Camada oculta, com 5 valores de neurônios testados segundo o Critério de Fletcher-Gloss descrito por

$$2\sqrt{n_0} + n_2 \leq n_1 \leq 2n_0 + 1, \quad (3.2)$$

onde n_0 , n_1 e n_2 representam, respectivamente, as camadas de entrada, oculta e saída da rede (SILVA, I. N.; SPATTI, D. H.; FLAUZINO, R. A., 2016), tendo com função de ativação a tangente hiperbólica;

- Uma camada de saída com ativação linear e normalização z-score na saída; e
- Treino com métrica de perda sendo o Erro Percentual Absoluto Médio (EPAM), que pode ser definido por

$$\text{EPAM}(\%) = \frac{1}{N} \sum_{i=1}^N \frac{|Y_i - Yd_i|}{Yd_i} \times 100\%, \quad (3.3)$$

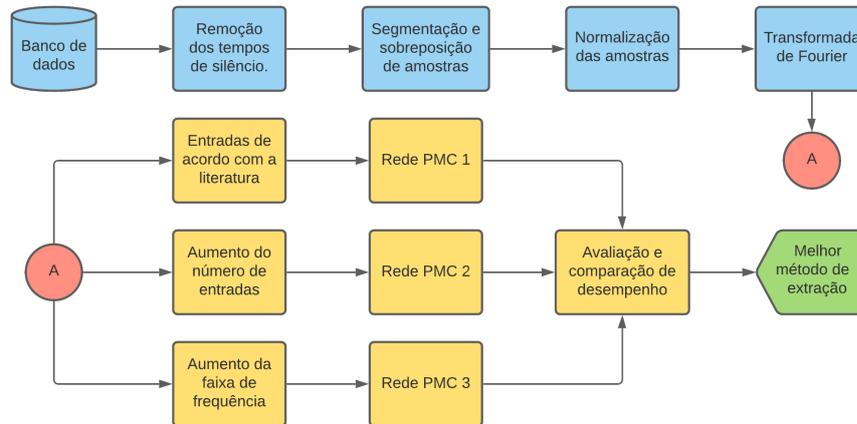
na qual N é o número de amostras durante o treinamento, Yd_i é a saída desejada e Y_i é a saída estimada da rede PMC, com algoritmo de otimização ADAM e batch size de 20, o que representa que os pesos da rede serão recalculados a cada 20 entradas;

Para evitar *overfitting*, o treinamento de cada rede foi finalizado sempre que EPAM dos dados de validação pararem de diminuir. Foi considerada a melhor rede aquela que obteve menor EPAM de todos os treinamentos realizados para cada método de extração de características.

Para avaliar a sensibilidade do modelo quanto sujeito a amostras de outra natureza, foi feito o cruzamento do banco de dados e então calculado o coeficiente de correlação de Pearson, que pode ser representado por

$$\rho = \frac{\text{cov}(\mathbf{X}, \mathbf{Y})}{\sqrt{\text{var}(\mathbf{X}) \cdot \text{var}(\mathbf{Y})}} \quad (3.4)$$

Figura 4 – Fluxograma do processo de obtenção e comparação das redes.



onde ρ representa o coeficiente de correlação, X os valores de saídas calculadas e Y os valores esperados. O valor de ρ pode variar de -1 à 1, sendo -1 representando uma correlação perfeita negativa, 0 sem correlação e 1 uma correlação perfeita positiva. Todo o processo descrito acima pode ser visualizado no fluxograma presente na Fig. 4.

4 Resultados e Discussão

A seguir, serão apresentados como o aumento do número de entradas e o aumento do espectro de frequência da rede neural afetaram o desempenho da mesma. Vale ressaltar que o aumento do número de entradas aumenta também a resolução das frequências de entrada, enquanto o aumento do espectro reduz a resolução e aumenta a variação de frequências de entrada.

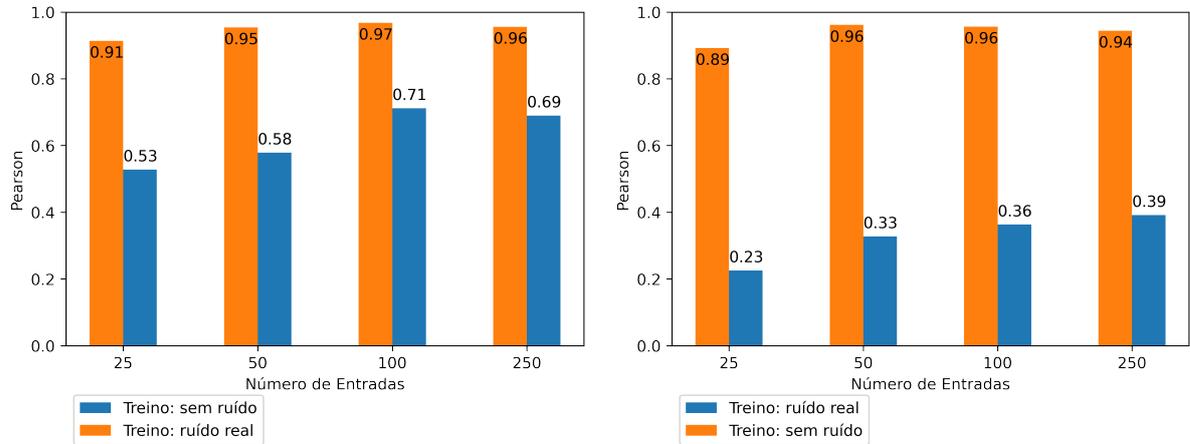
4.1 Aumento de Entradas

Na Fig. 6(c) pode-se conferir as menores médias e desvio padrão do EPAM dos K-folds para o banco de dados de ruído real e sem ruído para cada número de entradas. As linhas em preto representam o desvio padrão entre os K-folds e o valor representa a menor média entre os K-folds de acordo com a respectiva configuração de rede. Na mesma pode-se observar que o aumento de entradas para 50 reduziu o EPAM da rede de 5,29% para 3,98% e de 5,01% para 3,55% dos bancos de dados com ruído real e sem ruído, respectivamente. Entretanto, ao aumentar ainda mais as entradas da rede o EPAM permanece próximo ao de 50 entradas.

Para o banco de dados com ruído real, ao calcular a correlação de Pearson para avaliar se a saída da rede têm relação com os valores reais, obteve-se os valores encontrados na Fig. 6(a), onde nota-se que as amostras de teste para a rede com menor EPAM não foi capaz de obter uma correlação relevante para o cruzamento do banco de dados (representado em azul). Porém, obteve-se correlação expressiva para a rede treinada com o mesmo banco de dados que foi treinada (representado em laranja), indicando que todos os números de entradas testados nesse modelo não foram capazes de prever uma amostra fora do ambiente para o qual a rede foi treinada.

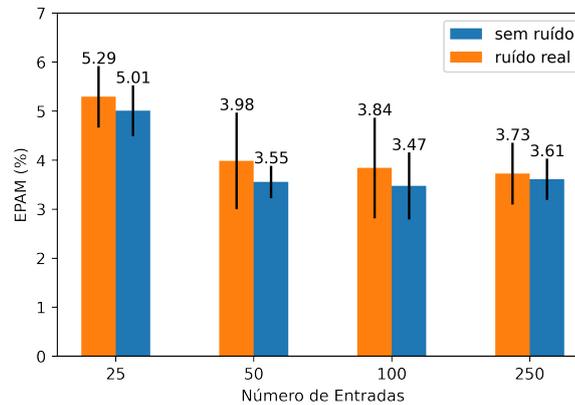
Já para o banco de dados sem ruído, obteve-se os valores encontrados na Fig. 6(b), na qual os resultados foram semelhantes às previsões com ruído real. Entretanto, a correlação do cruzamento do banco de dados se mostrou menor quando comparados ao cruzamento do banco de dados da Fig. 6(a). Isto se deve às músicas presentes no banco de dados com ruído real também estarem presentes no banco de dados sem ruído. Portanto, a rede treinada sem ruído previu no teste somente músicas com as quais teve contato no treino, porém, acrescidas de ruído, o que não acontece no cruzamento reverso.

Figura 5 – EPAM e Coeficientes de Pearson para cada número de entradas.



(a) Coeficientes de Pearson das amostras de teste com ruído real.

(b) Coeficientes de Pearson das amostras de teste sem ruído.



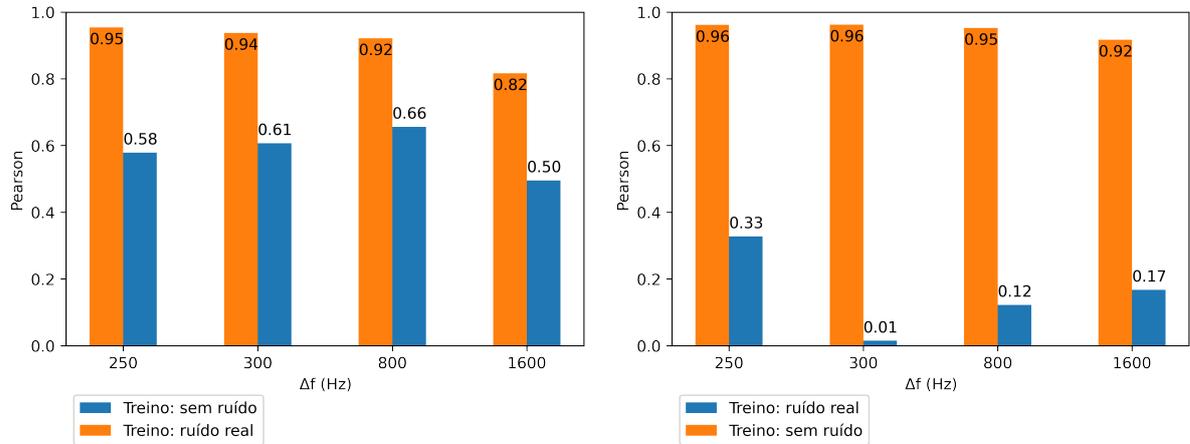
(c) EPAM das amostras de teste.

4.2 Aumento do Espectro de Frequência

Na Figura 7(c) encontra-se as menores médias e desvio padrão do EPAM dos K-folds para o banco de dados de ruído real e sem ruído para cada variação de frequência. Nota-se que o EPAM cresce para faixas de frequência maiores, indicando que as magnitudes das frequências acima de 300 Hz não contribuem positivamente para a eficácia da rede.

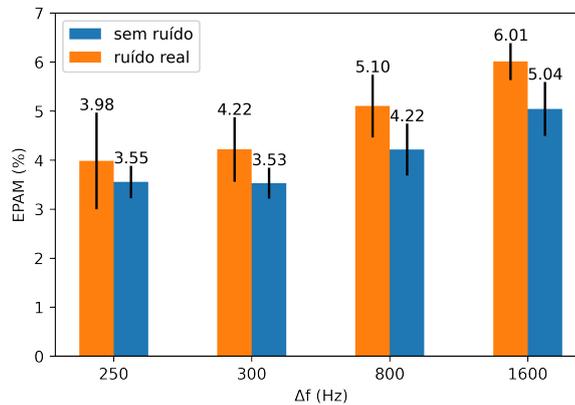
Nas Fig. 7(a) e Fig. 7(b), nota-se comportamento semelhante das correlações das Fig. 6(a) e Fig. 6(b) para o cruzamento do banco de dados, onde as previsões das amostras com ruído real são ligeiramente maiores, mas ainda insuficientes para afirmar que a rede é capaz de prever amostras de natureza de ruído diferente. Ao contrário quando a amostra é submetida à amostras no mesmo ambiente de ruído para a qual foi treinada, onde a correlação foi satisfatória pra todos os casos exceto para uma variação de frequência de 1600 Hz com ruído real. Nesse caso, também pode-se observar que o EPAM não foi extremamente alto para este caso, sendo então necessário a comparação de ambas as métricas para avaliação da eficácia da rede.

Figura 6 – EPAM e Coeficientes de Pearson para cada faixa de frequência.



(a) Coeficientes de Pearson das amostras de teste com ruído real.

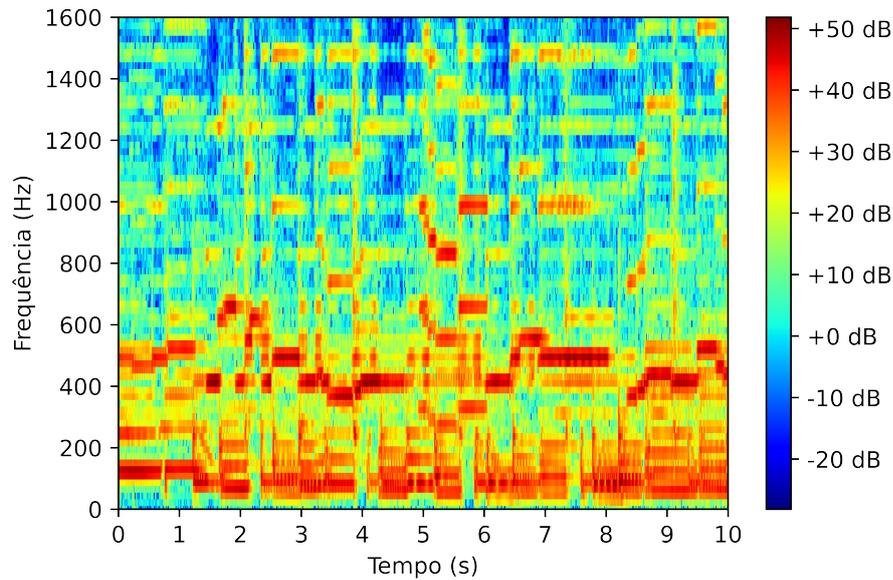
(b) Coeficientes de Pearson das amostras de teste sem ruído.



(c) EPAM das amostras de teste.

A provável razão do aumento do EPAM para os valores mais altos de Δf pode ser observada na Figura 7, onde nota-se maior variação das magnitudes das frequências de 600 Hz à 1600 Hz ao longo do tempo, o que resulta em diferentes valores de entrada para uma mesma saída da RNA, dificultando o treinamento da rede por ter maiores variações de entrada para uma mesma saída.

Figura 7 – Espectrograma de uma das músicas utilizadas.



Em resumo, o aumento do número de entradas da rede para a mesma faixa de frequência foi capaz de reduzir o erro em 24,76% e 29,14% do EPAM para músicas com ruído real e sem ruído, respectivamente. Com isso, aumentar o número de entradas para 50, mantendo a faixa de frequência de 50 Hz à 300 Hz foi o melhor método de extração de características na questão de erro. Vale ressaltar que o aumento de entradas para 100 e 250 resultou em erros menores, entretanto, durante o treinamento havia casos da rede não convergir devido ao alto número de parâmetros a serem treinados em proporção ao banco de dados, portanto, ambos os valores não foram considerados como melhores. Porém, vale ressaltar que nenhum método foi capaz de prever corretamente a duração do compasso quando submetido à uma amostra de natureza de ruído diferente do treino. Tendo a rede treinada por músicas sem ruído uma diferença positiva, porém insuficiente, ao prever as mesmas músicas com ruído.

5 Considerações Finais

Este trabalho analisou duas alterações nas extrações de características de músicas de forró existente na literatura. Sendo apenas o aumento do número de entradas da rede para 50 mantendo a faixa de frequência de 50 Hz à 300 Hz capaz de reduzir o EPAM em mais de 20% para ambos os bancos de dados. Porém, mesmo essa configuração de entrada não fez com que a rede fosse capaz de prever amostras fora do ambiente de ruído para a qual foi treinada. Com isso, se faz necessário um estudo do quanto a mistura dos bancos de dados irá beneficiar a previsão de diferentes ambientes, levando em conta diferentes proporções de músicas em diferentes ambientes de ruído, que será o próximo objetivo de estudo em trabalhos futuros. Em [Paiva \(2020\)](#) a mistura total dos dois bancos de dados reduziu o valor do EPAM, contudo, havia amostras de teste utilizadas no treino quando esta avaliação foi realizada.

Dessa forma, espera-se que o avanço para a criação de um aplicativo que realize essa sinalização se aproxime de se tornar concreta para que ajude no aprendizado da dança de surdos, para que os mesmos possam ser incluídos no ambiente de dança de forró, melhorem sua qualidade de vida através de mais interações sociais. Tendo como objetivo final um aplicativo que possa captar o áudio através do microfone de dispositivos móveis e transmitir o ritmo da música através de estímulos táteis ou visuais para o usuário.

Referências

- BOSSEY, A. Accessibility all areas? UK live music industry perceptions of current practice and Information and Communication Technology improvements to accessibility for music festival attendees who are deaf or disabled. *International Journal of Event and Festival Management*, v. 11, n. 1, p. 6–25, feb 2020. ISSN 1758-2954. Disponível em: <<https://www.emerald.com/insight/content/doi/10.1108/IJEFM-03-2019-0022/full/html>>. Citado na página 12.
- BRASIL. Lei nº. 10.436, de 24 de abril de 2002. *Dispõe sobre a Língua Brasileira de Sinais - Libras e dá outras providências.*, Brasília, DF, Abril 2002. Citado na página 11.
- BRASIL. Decreto nº. 5.626, de 22 de dez. de 2005. *Regulamenta a Lei nº 10.436, de 24 de abril de 2002, que dispõe sobre a Língua Brasileira de Sinais - Libras, e o art. 18 da Lei nº 10.098, de 19 de dezembro de 2000.*, Brasília, DF, Dezembro 2005. Citado 2 vezes nas páginas 11 e 12.
- CAI, L.; CAI, Q. Music creation and emotional recognition using neural network analysis. *Journal of Ambient Intelligence and Humanized Computing*, Springer Berlin Heidelberg, n. 0123456789, 2019. ISSN 18685145. Disponível em: <<https://doi.org/10.1007/s12652-019-01614-6>>. Citado na página 13.
- CHAVEIRO, N. et al. Qualidade de vida dos surdos que se comunicam pela língua de sinais: revisão integrativa. *Interface - Comunicação, Saúde, Educação*, v. 18, p. 101–114, 01 2014. Citado na página 11.
- FOX, M. L.; JAMES, T. G.; BARNETT, S. L. Suicidal behaviors and help-seeking attitudes among deaf and hard-of-hearing college students. *Suicide and Life-Threatening Behavior*, v. 50, n. 2, p. 387–396, 2020. Disponível em: <<https://onlinelibrary.wiley.com/doi/abs/10.1111/sltb.12595>>. Citado na página 11.
- FREITAS, F. B. *Benefícios psicológicos da prática da dança em pessoas com diagnóstico de ansiedade e depressão: Revisão bibliográfica*. 2019. Monografia (Bacharel em Psicologia), UFM (Universidade Federal do Maranhão), São Luís, Maranhão, Brasil. Citado na página 12.
- HAGUIARA-CERVellini, N. *A musicalidade do surdo: representação e estigma*. [S.l.]: Plexus Editora, 2003. Citado na página 12.
- HAYKIN, S. Redes neurais: Princípios e práticas; trad. *Paulo Martins Engel*, v. 2, p. 27–32, 2001. Citado na página 15.
- JUNIOR, A. C. de Q.; VOLP, C. M. Forró universitário: a tradução do forró nordestino no sudeste brasileiro. *Motriz. Journal of Physical Education. UNESP*, p. 117–120, 2005. Citado na página 14.
- LACERDA, C. B. F. de. A inclusão escolar de alunos surdos: o que dizem alunos, professores e intérpretes sobre esta experiência. *Cadernos CEDES*, v. 26, n. 69, p. 163–184, 2006. Citado na página 12.

- LATHI, B. P. *Sinais e Sistemas Lineares - 2ª Edição*. [S.l.]: Bookman, 2007. Citado na página 15.
- MIRZAEI, M.; KÁN, P.; KAUFMANN, H. EarVR: Using Ear Haptics in Virtual Reality for Deaf and Hard-of-Hearing People. *IEEE Transactions on Visualization and Computer Graphics*, v. 26, n. 5, p. 2084–2093, 2020. ISSN 19410506. Citado na página 12.
- PAIVA, L. F. *Estimação do compasso musical do forró utilizando rede perceptron multicamadas*. 2020. Monografia (Bacharel em Engenharia Elétrica), UFV (Universidade Federal de Viçosa), Viçosa, Minas Gerais, Brasil. Citado 5 vezes nas páginas 13, 16, 18, 19 e 25.
- PAIVA, L. F. et al. Estimação do compasso musical do forró utilizando rede perceptron multicamadas. In: *Anais do XXIII Congresso Brasileiro de Automática*. Porto Alegre: [s.n.], 2020. Citado 6 vezes nas páginas 8, 13, 14, 15, 16 e 18.
- PAULA, T. R. M. de; PEDERIVA, P. L. M. Musical Experience in Deaf Culture. *International Journal of Technology and Inclusive Education*, 2017. Disponível em: <<https://www.semanticscholar.org/paper/Musical-Experience-in-Deaf-Culture-Paula-Pederiva/1e450c197d715f94ee8b062a5eb85f26c33adf5e>>. Citado na página 12.
- SALAMON, J.; BELLO, J. P. Deep Convolutional Neural Networks and Data Augmentation for Environmental Sound Classification. *IEEE Signal Processing Letters*, v. 24, n. 3, p. 279–283, 2017. ISSN 10709908. Citado na página 13.
- SANTOS, A. et al. You are off the beat!: Is accelerometer data enough for measuring dance rhythm? In: . [S.l.: s.n.], 2018. p. 1–8. Citado 2 vezes nas páginas 8 e 14.
- SANTOS, T. M.; VIEIRA, L. C.; FARIA, C. A. Deficiência auditiva e mercado de trabalho: uma visão de empregadores da cidade de Uberlândia-MG. *Psicologia: teoria e prática*, scielopepsic, v. 15, p. 92 – 103, 08 2013. ISSN 1516-3687. Disponível em: <http://pepsic.bvsalud.org/scielo.php?script=sci_arttext&pid=S1516-36872013000200007&nrm=iso>. Citado na página 11.
- SCHOENBERG, A. *Fundamentos da composição musical*. [S.l.]: Edusp, 1990. Citado 2 vezes nas páginas 14 e 15.
- SHARP, A.; BACON, B. A.; CHAMPOUX, F. Enhanced tactile identification of musical emotion in the deaf. *Experimental Brain Research*, Springer Berlin Heidelberg, v. 238, n. 5, p. 1229–1236, 2020. ISSN 14321106. Disponível em: <<https://doi.org/10.1007/s00221-020-05789-9>>. Citado na página 12.
- SHARP, A. et al. Musicians show better auditory and tactile identification of emotions in music. *Frontiers in Psychology*, v. 10, n. AUG, p. 1–7, 2019. ISSN 16641078. Citado na página 12.
- SILVA, I. N.; SPATTI, D. H.; FLAUZINO, R. A. *Redes Neurais Artificiais para Engenharia e Ciências Aplicadas: Fundamentos Teóricos e Aspectos Práticos*. 2ª edição. ed. São Paulo: Artliber, 2016. Citado 4 vezes nas páginas 13, 15, 16 e 19.
- SILVA, I. N.; SPATTI, D. H.; FLAUZINO, R. A. *Redes Neurais Artificiais para Engenharia e Ciências Aplicadas: Fundamentos Teóricos e Aspectos Práticos*. 2ª edição. [S.l.]: Artliber, 2016. Citado na página 19.

SOLANKI, A.; PANDEY, S. Music instrument recognition using deep convolutional neural networks. *International Journal of Information Technology*, Springer Singapore, 2019. ISSN 2511-2104. Disponível em: <<https://doi.org/10.1007/s41870-019-00285-y>>. Citado na página 13.

STEIJNS, F. et al. Ambulatory electrocardiographic monitoring and ectopic beat detection in conscious mice. *Sensors*, v. 3867, n. 14, 2020. Citado na página 15.

YU, D. et al. Predominant Instrument Recognition Based on Deep Neural Network with Auxiliary Classification. *IEEE/ACM Transactions on Audio Speech and Language Processing*, IEEE, v. 28, p. 852–861, 2020. ISSN 23299304. Disponível em: <<https://ieeexplore.ieee.org/document/8979336/>>. Citado na página 13.