

UNIVERSIDADE FEEVALE

JÚLIO CÉSAR MULLER RODRIGUES

APRENDIZADO PROFUNDO E INTELIGÊNCIA ARTIFICIAL
APLICADOS NA ANÁLISE MÚSICAL

(Título Provisório)

Anteprojeto de Trabalho de Conclusão

Novo
Hamburgo 2022

JÚLIO CÉSAR MULLER RODRIGUES

APRENDIZADO PROFUNDO E INTELIGÊNCIA ARTIFICIAL
APLICADOS NA ANÁLISE MÚSICAL

(Título Provisório)

Anteprojeto de Trabalho de Conclusão de
Curso, apresentado como requisito
parcial à obtenção do grau de Bacharel
em Ciência da Computação pela
Universidade Feevale

Orientador: Dr. Gabriel da Silva Simões

Novo
Hamburgo 2022

RESUMO

Com o avanço tecnológico constante das últimas décadas o acesso da população a computadores e *smartphones* aumentou imensamente ao ponto em que a maioria dos moradores de uma casa possuem um *smartphone*. Essa crescente fomentou o desenvolvimento de diversos segmentos, um destes sendo o de *streaming* de música, onde plataformas como *Spotify*, *Deezer* e *Apple Music* dividem entre si uma parcela majoritária do mercado. O contato musical frequente proporcionado por estas plataformas aliados as milhares de postagens e anúncios em redes sociais onde há a presença de uma música de fundo ou onde a música é um dos elementos principais deste conteúdo acaba desenvolvendo em algumas pessoas o desejo de consumir música de maneira ativa a partir do estudo de um instrumento. Embora seja cativante o estudo de música possui uma série de complexidades e um das maiores é o estudo da teoria musical onde se é analisado a relação rítmica, melódica e harmônica das músicas, felizmente com o auxílio de algoritmos de inteligência artificial é possível criar ferramentas para auxiliar no estudo e aprendizados destes conceitos. Este trabalho tem como objetivo aplicar técnicas conhecidas de inteligência artificial, como *machine learning* para criar uma rede neural capaz de processar um arquivo de áudio, realizar a análise das notas musicais presentes no mesmo e realizar sugestões harmônicas e melódicas.

Palavras-chave: Inteligência artificial, *Deep Learning*, redes neurais, música;

MOTIVAÇÃO

Música já é quase uma parte integral da vida cotidiana, apresentando-se por diversas plataformas como rádios, serviços de *streaming*, acompanhando um vídeo postado em redes sociais, trilhas sonoras de filmes, séries e jogos, estando presentes em um som de fundo como em elevadores e anúncios de produtos. Somos expostos tão frequentemente e a tanto tempo que para muitas pessoas a apreciação musical, somente de maneira passiva (somente ouvir), deixa de ser suficiente fazendo com que passem a buscar ativamente um contato musical a partir do estudo de um instrumento.

Por mais popular e cativante que seja, o estudo de um instrumento musical não se trata de algo trivial, afinal existem diversas camadas de complexidade gerais atreladas diretamente à música. As complexidades gerais são partes fundamentais para desenvolver a capacidade de composição musical tais como ritmo, tempo, harmonia e melodia. Além disso, existem complexidades específicas relacionadas ao instrumento como curva de aprendizado e tipo de instrumentos (corda, sopro, percussão e elétricos; melódicos e harmônicos).

Dentre todas as complexidades citadas, a teoria musical, que é estudada principalmente na relação entre harmonia e melodia (apesar de também ser constituída pelo ritmo), é onde se encontra a maior dificuldade de muitos musicistas e estudantes. Felizmente, com a ajuda de algoritmos de inteligência artificial, é possível analisar e classificar diferentes tipos de sons e assim possivelmente ajudar na compreensão e na aplicação destes conceitos.

Este trabalho propõe-se em realizar um estudo com algoritmos de inteligência artificial e aprendizado profundo com o objetivo de criar um modelo capaz de processar um arquivo de áudio, classificar as notas musicais presentes no mesmo e gerar uma saída com sugestões harmônicas ou melódicas.

Inteligência artificial (IA) é um campo em rápido crescimento, com um amplo potencial de aplicabilidade em diversas áreas do conhecimento. IA é definida como um campo da ciência e da engenharia preocupado com a compreensão computacional do que é comumente chamado de comportamento inteligente com a criação de artefatos que exibem tal comportamento (IBM, 2020).

A IA é composta de vários algoritmos e técnicas inteligentes como aprendizado de máquina (machine learning - ML), robótica, *fuzzy logic* (FL), sistemas especialistas (SE), conhecimento base (*knowledge base* - KB), e a combinação de dois ou mais métodos (multimétodos) (IBM, 2020).

Machine learning é um subconjunto de inteligência artificial que usa técnicas (como o aprendizado profundo) que permitem que os computadores usem a experiência para aprimorar as tarefas. O processo de aprendizado consta das seguintes etapas: i) deve-se alimentar dados

em um algoritmo, ii) deve-se usar esses dados para treinar um modelo, iii) deve-se testar e implantar o modelo e por fim consumir o modelo implantado para realizar uma tarefa de previsão automatizada (MICROSOFT, 2020).

Aprendizado profundo é um subconjunto de *machine learning* baseado em redes neurais artificiais. O processo de aprendizado é profundo porque a estrutura das redes neurais artificiais consiste em várias camadas: de entrada, saída e oculta. Cada camada contém unidades que transformam os dados de entrada em informações que a próxima camada pode usar para executar uma determinada tarefa preditiva. Graças a essa estrutura, um computador pode aprender por meio de seu próprio processamento de dados (PIMENTA, 2021).

É comum aos métodos de ML a capacidade de conduzir tarefas de previsão (quando a variável alvo é um valor) ou classificação (quando a variável alvo é uma classe). Vários métodos de ML, como subcategoria de IA, podem ser aplicados para cumprir esses propósitos usando suas próprias capacidades específicas. Em detalhes, eles incluem muitas técnicas, tais como redes neurais artificiais (*artificial neural network* - ANN), máquina de vetor de suporte (*support vector machine* - SVM), árvore de decisão (*decision tree* - DT) e Naive Bayes (NB) (IBM, 2020).

O processamento de áudio através da IA possui uma grande quantidade de aplicações em diversas áreas, podendo ser utilizada para detectar falhas em equipamentos e maquinário baseado nos sons emitidos pelos mesmos, sendo possível aplicar em sistemas de segurança para detectar invasões. Uma área bem desenvolvida no tratamento e processamento de áudio são as aplicações de conversão de discurso para texto (speech to text), utilizadas por diversos aplicativos, principalmente em smartphones, para redigir mensagens ou realizar pesquisas (DOSHI, 2021).

Até alguns anos atrás, os aplicativos de aprendizado de máquina de áudio costumavam depender de técnicas tradicionais de processamento de sinal digital para extrair recursos. Por exemplo, para entender a fala humana, os sinais de áudio podem ser analisados usando conceitos fonéticos para extrair elementos como fonemas. Tudo isso exigia muito conhecimento específico do domínio para resolver esses problemas e ajustar o sistema para um melhor desempenho (DOSHI, 2021).

No entanto, nos últimos anos, à medida que o *Deep Learning* se torna cada vez mais onipresente, ele também obteve um tremendo sucesso no manuseio de áudio. Com o aprendizado profundo, as técnicas tradicionais de processamento de áudio não são mais necessárias e podemos contar com a preparação de dados padrão sem exigir muita geração manual e personalizada de recursos. O mais interessante é que, com o aprendizado profundo, não lidamos com dados de áudio em sua forma bruta. Em vez disso, a abordagem comum usada

é converter os dados de áudio em imagens a partir de um espectrograma e, em seguida, usar uma arquitetura CNN padrão para processar essas imagens (DOSHI, 2021).

Conforme a temática apresentada, este trabalho tem como objetivo pesquisar por *datasets* de áudio que possuam notas musicais classificadas ou, na ausência dos mesmos, criar um com estas notas classificadas para utilizar como entrada para o treinamento de uma rede neural capaz de classificar as notas presentes em uma faixa de áudio.

OBJETIVOS

Objetivo geral

O objetivo geral deste trabalho é desenvolver uma rede neural capaz de processar um arquivo de áudio, classificar as notas musicais contidas no mesmo e gerar como saída uma sugestão harmônica ou melódica sobre o arquivo processado.

Objetivos específicos

- Encontrar/criar dataset com notas musicais para o treinamento da rede.
- Codificar a rede neural para o processamento do áudio.
- Realizar o fluxo de entrada para análise melódica.
 - Receber o arquivo de áudio.
 - Gerar o espectrograma do arquivo.
 - Classificar todas as notas encontradas no áudio.
 - Com o array das notas encontradas no áudio, sugerir possíveis harmonias.
- Realizar o fluxo de entrada para análise harmônica.
 - Receber o arquivo de áudio.
 - Gerar o espectrograma do arquivo.
 - Classificar todas as acordes encontradas no áudio.
 - Com o array dos acordes encontrados no áudio, sugerir possíveis escalas.

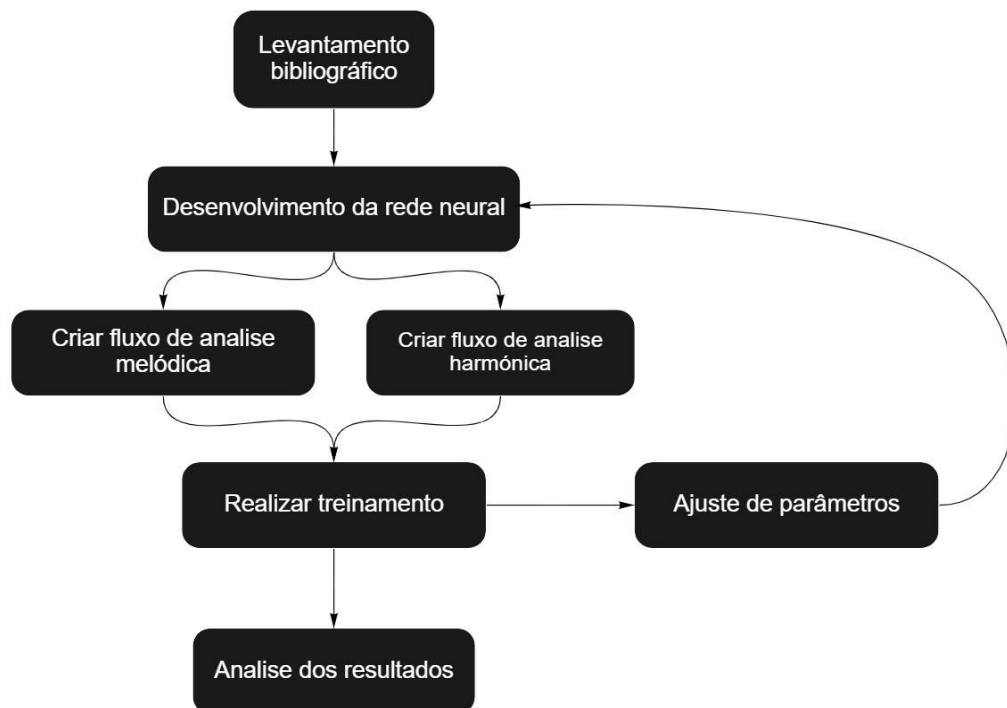
METODOLOGIA

O objetivo deste trabalho é desenvolver uma rede neural capaz de processar um arquivo de áudio e classificar as notas musicais contidas no mesmo, então pode-se dizer que este trabalho constitui-se de uma pesquisa aplicada sobre o uso de *machine learning* para classificar musicalmente arquivos de áudio.

A pesquisa será feita utilizando procedimentos de uma revisão sistemática em conjunto com experimentação em ambiente local, visando a criação de rede neural para a classificação dos arquivos.

O processo será dividido em três macro etapas: levantamento bibliográfico, desenvolvimento da rede neural com python, por fim, será feita a análise dos resultados obtidos pelos diferentes treinamentos.

O fluxograma resumido do processo de pesquisa está presente na figura 1.



Fonte: Autor

A etapa do levantamento bibliográfico contempla a busca por bibliografia relacionada ao processamento de arquivos de áudio através de redes neurais, classificação destes arquivos e reconhecimento de notas musicais em arquivos de áudio. O objetivo desta etapa é levantar conhecimento para servir de apoio durante a etapa de desenvolvimento da rede neural.

A segunda etapa consiste no desenvolvimento da rede neural, utilizando como base os conhecimentos obtidos durante o levantamento bibliográfico, nesta etapa será realizado a codificação da rede neural, juntamente do treinamento e ajustes dos parâmetros da rede em busca da melhor acurácia possível.

A terceira etapa é onde serão analisados os resultados obtidos pela rede neural, sua precisão e viabilidade, além de estudar pontos de melhorias e extensões para aplicações futuras.

CRONOGRAMA

Trabalho de Conclusão I

Etapa	Meses			
	Ago	Set	Out	Nov
Redação do anteprojeto	■	■		
Revisão do anteprojeto	■	■		
Entrega do anteprojeto		■		
Levantamento bibliográfico		■		
Pesquisa/criação de datasets compatíveis com o projeto		■	■	
Desenvolvimento inicial da rede neural		■	■	■
Redação TC I		■	■	■
Revisão TC I		■	■	■
Entrega do TC I				■

Trabalho de Conclusão II

Etapa	Meses			
	Mar	Abr	Mai	Jun
Finalização do desenvolvimento da rede neural	■	■		
Testes com a rede neural	■	■		
Análise dos resultados	■	■		
Redação TC II	■	■	■	■
Revisão TC II	■	■	■	■
Entrega do TCC II				■
Apresentação do trabalho à banca				■

BIBLIOGRAFIA

IBM. O que é inteligência artificial?. Disponível em:

<https://www.ibm.com/br-pt/cloud/learn/what-is-artificial-intelligence> . Acesso em: 20 de Agosto de 2022

PIMENTA, Igor. Guia Completo Inteligência Artificial: o que é, conceito e métodos de IA. disponível em <https://www.take.net/blog/tecnologia/inteligencia-artificial/> , acessado 31 de Agosto de 2022.

MICROSOFT, Aprendizado profundo x Aprendizado de máquina em Azure Machine Learning versus Machine Learning, disponível em:

<https://docs.microsoft.com/pt-br/azure/machine-learning/concept-deep-learning-vs-machine-learning> , acessado 01 de setembro de 2022.

DOSHI, Ketan. Audio Deep Learning Made Simple (Part 1): State-of-the-Art Techniques, Ketan Doshi. Disponível em:

<https://towardsdatascience.com/audio-deep-learning-made-simple-part-1-state-of-the-art-techniques-da1d3dff2504>. Acessado em 20 de Agosto de 2022