

CENTRO UNIVERSITÁRIO DO PARÁ - CESUPA
ESCOLA DE NEGÓCIOS, TECNOLOGIA E INOVAÇÃO - ARGO
CURSO DE ENGENHARIA DA COMPUTAÇÃO

MATHEUS CARDOSO MODESTO
PEDRO GOMES DOS SANTOS JUNIOR
ROGER MORAES BRITO

**DESENVOLVIMENTO DE UMA INTELIGÊNCIA ARTIFICIAL PARA A
CLASSIFICAÇÃO DE SINAIS SONOROS DE TRÂNSITO**

BELÉM

2022

MATHEUS CARDOSO MODESTO
PEDRO GOMES DOS SANTOS JUNIOR
ROGER MORAES BRITO

**DESENVOLVIMENTO DE UMA INTELIGÊNCIA ARTIFICIAL PARA A
CLASSIFICAÇÃO DE SINAIS SONOROS DE TRÂNSITO**

Trabalho de conclusão de curso apresentado à Escola de Negócios, Tecnologia e Inovação do Centro Universitário do Estado do Pará como requisito para obtenção do título de Engenheiro da Computação na modalidade ARTIGO.

Orientador: Me. Pedro Giroto

BELÉM
2022

Dados Internacionais de Catalogação-na-publicação (CIP)
Biblioteca do CESUPA, Belém – PA

Modesto, Matheus Cardoso.

Desenvolvimento de uma inteligência artificial para a classificação de sinais sonoros de trânsito / Matheus Cardoso Modesto, Pedro Gomes dos Santos Junior, Roger Moraes Brito; orientador Pedro Giroto. – 2022.

Trabalho de Conclusão de Curso (Graduação) – Centro Universitário do Estado do Pará, Engenharia da Computação, Belém, 2022.

1. Inteligência artificial. 2. Tecnologia assistiva. 3. Rede neural artificial. 4. Deficientes auditivos. I. Santos Junior, Pedro Gomes dos Santos. II. Brito, Roger Moraes. III. Giroto, Pedro, orient. IV. Título.

CDD 23ª ed. 006.3

MATHEUS CARDOSO MODESTO
PEDRO GOMES DOS SANTOS JUNIOR
ROGER MORAES BRITO

**DESENVOLVIMENTO DE UMA INTELIGÊNCIA ARTIFICIAL PARA A
CLASSIFICAÇÃO DE SINAIS SONOROS DE TRÂNSITO**

Trabalho de conclusão de curso
apresentado à Escola de Negócios,
Tecnologia e Inovação do Centro
Universitário do Estado do Pará como
requisito para obtenção do título de
Engenheiro da Computação na
modalidade ARTIGO

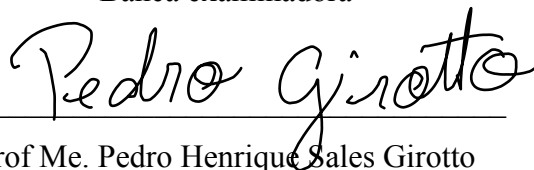
Data da aprovação: 8 / 06 / 22

Nota final aluno(a) I: 10

Nota final aluno(a) II: 10

Nota final aluno(a) III: 10

Banca examinadora



Prof Me. Pedro Henrique Sales Giroto

Orientador e Presidente da banca



Prof. Me. Polyana Santos Fonseca Nascimento

Examinadora interna

Prof. Me. Suzane Alfaia Dias

Examinadora interna

Eu, Matheus Cardoso Modesto, dedico o esforço que tive para finalizar este trabalho às pessoas que estiveram do meu lado me apoiando durante esta longa caminhada. Em especial a uma mulher lutadora, que não poupou esforços para que eu concluísse este trabalho e que sempre dedicou a sua vida pela felicidade de seus filhos e a pessoa que eu mais amo no mundo: minha mãe Josiany Cardoso Cardoso.

Eu, Pedro Gomes dos Santos Junior, dedico este trabalho de conclusão de curso a Deus, pelas graças concedidas ao longo da minha trajetória nesta graduação. A minha Mãe Lucia Helena Cardoso Marques e a minha irmã Luana Marques dos Santos, pelo apoio financeiro, emocional e por suas orações.

Eu, Roger Moraes Brito, dedico este trabalho ao meu falecido avô, Jacy. Uma das pessoas que mais me deram apoio e incentivo ao longo dessa jornada. Também dedico aos meus pais me propuseram esta oportunidade de ter chegado a este momento.

RESUMO

No Brasil, há um baixo número de pessoas que possuem deficiência auditiva e são condutoras de veículos no trânsito (IBGE, 2010), uma das causas disso são as dificuldades enfrentadas pelos mesmos no momento da emissão ou renovação de uma CNH (Carteira Nacional de Habilitação). A expansão do desenvolvimento da Tecnologia Assistiva tem possibilitado à comunidade de PcD, maior qualidade de vida e inclusão social. Este projeto tem como propósito desenvolver uma Inteligência Artificial, um modelo de RNA eficiente em classificar sons de apito, buzina e sirene, que em uma perspectiva futura, pode ser implementado em um sistema de alerta para dar assistência a condutores com deficiência auditiva. Além disso, o modelo de Rede Neural Artificial do tipo convolucional apresentado e a base de dados criados, dão margem ao uso desses componentes em outras aplicações.

Palavras-chave: Deficientes Auditivos; RNA; Tecnologia Assistiva.

ABSTRACT

In Brazil, there is a low number of people who are hearing impaired and are drivers of vehicles in traffic (IBGE, 2010), one of the causes of this are the difficulties faced by them when issuing or renewing a CNH (National Driver's License). The expansion of Assistive Technology development has enabled the PcD community to have a better quality of life and social inclusion. The purpose of this project is to develop an Artificial Intelligence, an Artificial Neural Network model which is efficient in classifying whistle, horn and siren sounds, which in a future perspective, can be implemented in an alert system to assist hearing impaired drivers. Furthermore, the presented convolutional-type Artificial Neural Network (CNN) model and the database created, give scope for the use of these components in other applications.

Keywords: Hearing Impaired; ANN; Assistive Technology.

LISTA DE FIGURAS

Figura 01 - Topologia inicial da RNA.....	24
Figura 02 - Fluxograma Realtime.....	26

LISTA DE TABELAS

Tabela 01 - Composição do Database.....	22
---	----

LISTA DE GRÁFICOS

Gráfico 01 - Gráficos da Acurácia e Erro.....	28
Gráfico 02 - Gráficos de previsão para sons (Teste V).....	28

LISTA DE ABREVIATURAS E SIGLAS

A	Apito
B	Buzina
CNH	Carteira Nacional de Habilitação
DTMF	Dual Tone Multi Frequency
IA	Inteligência Artificial
IDE	Integrated Development Environment
MFCC	Mel Frequency Cepstral Coefficients
MG	Minas Gerais
OMS	Organização Mundial da Saúde
PUC	Pontifícia Universidade Católica
RNA	Rede Neural Artificial
RNC	Rede Neural Convolucional
RNR	Redes Neurais Recorrentes
S	Sirene
TA	Tecnologia Assistiva
WAV	Waveform Audio File Format

SUMÁRIO

1. CONTEXTUALIZAÇÃO	13
1.1 REVISÃO BIBLIOGRÁFICA	13
1.1.1 Deficiência auditiva	13
1.1.2 A inteligência artificial	14
1.1.3 RNA e o reconhecimento de sons	16
1.1.4 A IA em benefício da tecnologia assistiva	17
1.2 APRESENTAÇÃO DO PROBLEMA	18
1.3 JUSTIFICATIVA	18
1.4 OBJETIVOS	19
1.4.1 Objetivo geral	19
1.4.2 Objetivos específicos	19
1.5. ESTRUTURA DO TRABALHO	19
2. DESENVOLVIMENTO DE UMA INTELIGÊNCIA ARTIFICIAL PARA A CLASSIFICAÇÃO DE SINAIS SONOROS DE TRÂNSITO	20
2.1 INTRODUÇÃO	21
2.2 MATERIAL E MÉTODOS	22
2.2.1 Base De Dados	22
2.2.2 Rede Neural Artificial	23
2.2.2.1 Argumentos nas camadas	25
2.2.3 Classificação em Tempo Real	25
2.4 RESULTADOS	26
Gráfico 02 - Gráficos de previsão para sons (Teste 5)	28
2.5 DISCUSSÃO	29
2.6 CONSIDERAÇÕES FINAIS	30
3. SÍNTESE	32
3.1 PRINCIPAIS CONCLUSÕES	32
3.2 DISCUSSÃO GERAL	32
3.3 PERSPECTIVA FUTURA	33
4 REFERÊNCIAS BIBLIOGRÁFICAS	34
APÊNDICE A - Gráficos da Acurácia e Erro Teste 1	37
APÊNDICE B - Gráficos de previsão para som (Teste 1)	38
APÊNDICE C - Gráficos da Acurácia e Erro Teste 2	39
APÊNDICE D - Gráficos de previsão para som (Teste 2)	40
APÊNDICE E - Gráficos da Acurácia e Erro Teste 3	41
APÊNDICE F - Gráficos de previsão para som (Teste 3)	42
APÊNDICE G - Gráficos da Acurácia e Erro Teste 4	43

APÊNDICE H - Gráficos de previsão para som (Teste 4)	44
APÊNDICE I - Gráficos da Acurácia e Erro Teste 6	45
APÊNDICE J - Gráficos de previsão para som (Teste 6)	46
APÊNDICE K - Gráficos da Acurácia e Erro teste 7	47
APÊNDICE L - Gráficos de previsão para som (Teste 7)	48
APÊNDICE M - Matriz de confusão Teste 1	49
APÊNDICE N - Matriz de confusão Teste 2	50
APÊNDICE O - Matriz de confusão Teste 3	51
APÊNDICE P - Matriz de confusão Teste 4	52
APÊNDICE Q - Matriz de confusão teste 5	53
APÊNDICE R - Matriz de confusão teste 6	54
APÊNDICE S - Matriz de confusão teste 7	55

1. CONTEXTUALIZAÇÃO

1.1 REVISÃO BIBLIOGRÁFICA

1.1.1 Deficiência auditiva

Deficiência auditiva ou perda auditiva é um distúrbio de comunicação devido uma baixa na capacidade de ouvir sons em um ou ambos ouvidos, é importante destacar que essa deficiência difere da surdez, segundo a Organização Mundial da Saúde (OMS, 2021), a surdez é descrita pela ausência ou perda total em um ou ambos ouvidos.

Sobre essas definições é importante afirmar que:

Muitas vezes os termos “surdo” e “deficiente auditivo” são tratados como sinônimos, quando na verdade existe uma gama ampla de pessoas que, apesar de terem em comum a perda parcial ou total da capacidade de ouvir, levam vidas completamente diferentes: há quem consiga escutar perfeitamente graças ao uso de aparelhos menores que um AirPods; quem se comunique apenas através de uma Língua de Sinais; quem já nasça com problemas auditivos e quem vá perdendo a capacidade de ouvir ao longo da vida (SCHMIDT. 2020).

Alguns deficientes auditivos têm dificuldade em ouvir somente sons agudos como a fala de pessoas, outros possuem uma perda auditiva muito maior e reconhecem apenas sons altos e muito graves como o ronco de motores e sons de turbinas de avião. Sobre essas diferenças a OMS classificou os graus de perda auditiva levando em consideração a média das frequências 500Hz, 1000Hz, 2000Hz e 4000z da via aérea. O quadro abaixo apresenta a distinção feita pela organização.

Quadro 1 – Classificação do grau da perda auditiva (OMS 2020)

Graus de perda auditiva	Média entre as frequências de 500 Hz, 1 kHz, 2 kHz e 4 kHz	Desempenho
Audição normal	< 20 dB	Nenhum problema em ouvir sons
Leve	20 < 35 dB	Pode apresentar dificuldade em ouvir o que é falado em locais ruidosos
Moderado	35 < 50 dB	Pode apresentar dificuldade em ouvir conversa particularmente em lugares ruidosos.
Moderadamente severo	50 < 65 dB	Dificuldade em participar de uma conversa especialmente em locais ruidosos. Mas pode ouvir se falarem com a voz mais alta sem dificuldade.
Severo	65 < 80 dB	Não ouve a maioria das conversas e pode ter dificuldade em ouvir sons elevados. Dificuldade extrema para ouvir em lugares ruidosos e fazer parte de uma conversa
Profundo	80 < 95 dB	Dificuldade extrema em ouvir voz em forte intensidade
Perda Auditiva completa / surdo	>95dB	Não consegue escutar nenhuma conversa e a maioria dos sons ambientais.

Fonte: Organização Mundial da Saúde (2021)

Como compensação do déficit de audição, essas pessoas acabam adquirindo melhores reflexos por meio de seus outros sentidos que são mais desenvolvidos, dos de quem possui a audição normal, mas o nível dessa de adaptação e habilidades compensatórias variam de acordo com o tipo da perda auditiva, grau e em qual estágio da vida ocorreu a perda (SOUZA et al., 2016). Por isso é muito comum que haja uma diferença em identificar objetos com mais rapidez que a maioria das pessoas, por exemplo.

1.1.2 A inteligência artificial

A inteligência artificial (IA) é um ramo da computação que surgiu em meados dos anos 90, inicialmente para resolução de problemas e depois utilizada para fazer com que

computadores reproduzam comportamentos humanos. Os saltos tecnológicos ocorridos nas últimas décadas permitiram que a inteligência computacional atingisse patamares mais elevados. Sistemas contemporâneos são hábeis em mostrar uma capacidade genérica para aprender as leis físicas de ambientes nos quais forem inseridos, com o foco de concatenar informações colhidas de sensores múltiplos, visualizando e avaliando seus efeitos (KAELBLING, 1999), tornando possível o desenvolvimento e produção de dispositivos que realizam tarefas com algumas características humanas como identificar objetos, sons, imagens e executar demandas que antes eram de exclusividade humana.

O ser humano tem a capacidade de enxergar e identificar uma vasta gama de elementos que existem em um ambiente e, fornecer ao computador a percepção que leva um humano a ter essa habilidade, é um desafio (MORAIS, 2013). Afinal, segundo Rodrigues (2017), a capacidade de processamento do cérebro humano ainda está longe de ser simulada. É complexo dar a uma máquina os sentidos de forma perfeita de um ser humano, como a visão ou a audição. Todavia, mesmo com essas adversidades, é possível elaborar uma RNA com um desempenho que satisfaça esta proposta.

A inteligência artificial utiliza algoritmos de *machine learning* (aprendizado de máquina) - que treina a máquina para simular comportamentos humanos - e *deep learning* (aprendizado profundo) - que é uma subárea do *machine learning* que busca, através de algoritmos, trabalhar como os neurônios do cérebro humano para demonstrar um comportamento “inteligente”. As Redes Neurais Artificiais, por sua vez, são uma área do *machine learning*. O conceito sobre Redes Neurais Artificiais, refere-se às estruturas de modelos computacionais capazes de gerar alguns modelos matemáticos baseados no desenvolvimento neural de organismos inteligentes que, por meio de neurônios com pesos e possibilidades de programações, possuem a capacidade probabilística de captar conhecimento através de experiências, aprendizados e sensações (SILVEIRA, 2019).

Através do avanço das Inteligências Artificiais, na contemporaneidade, em alguns aspectos, já é possível simular o comportamento humano. Os dados repassados para uma rede, bem como a qualidade e quantidades deles, são fatores que contribuem para a efetividade dela. De acordo com Souza (2006), o treinamento de uma RNA é um dos fatores que contribui para o êxito da aplicação e os dados repassados a essa rede precisam ser inteligíveis para facilitar o aprendizado da rede a fim de não gerar uma confusão.

O funcionamento do *self learning* - autoaprendizagem -, característico da IA, leva a rede neural ao aperfeiçoamento e a torna capaz de corresponder às expectativas da sua implementação. O tempo de resposta e a precisão que uma IA pode ter torna favorável sua aplicação nas mais variadas áreas, como auxiliar um médico no diagnóstico de uma doença de um paciente e na interatividade entre ambos (LOBO, 2021).

Dentre os modelos de RNA, as redes do tipo convolucional - que fazem parte do *deep learning* - tem grande destaque, pois nos últimos anos trouxeram mudanças significativas para a comunidade computacional, (ZUO et al, 2015). E isso se dá porque a arquitetura de uma RNC é parecida com a do padrão de conectividade de neurônios no cérebro humano. Logo, com a implementação dela, é possível simular a capacidade cognitiva humana de classificar e reconhecer.

1.1.3 RNA e o reconhecimento de sons

As pesquisas na área da classificação de sons estão em ascensão e possuem bastante aplicabilidade. Embora, as elaborações no que se refere a esta temática ainda são um pouco escassas (SMALES, 2019). A classificação de sons inclui o aprendizado e catalogação dos mesmos para a previsão de suas categorias (DOSHI, 2021). As características dos áudios precisam ser extraídas para servirem como parâmetro de entrada, para a rede entender e identificar um determinado padrão.

Entre as possibilidades de representação de um som, o espectrograma se destaca, pois nele compreende-se informações de frequência e tempo (CHI; ZHANG; ZHANG, 2018). Mesmo assim, é trabalhoso conseguir construir um classificador com uma precisão que dê resultados confiáveis. Um elemento que pode influenciar bastante para reconhecimento adequado de sons é extrair a melhor representação paramétrica dos sinais acústicos (MUDA, 2010).

Rosso (2017) desenvolveu um aplicativo de celular utilizando o classificador K - Nearest Neighbor (KNN), que apesar de apresentar um bom desempenho, não foi o suficiente para garantir resoluções assertivas. Além disso, Rosso, identificou que uma alta qualidade do reconhecimento é proporcional à quantidade e diversidade de dados.

As Rede de Kohonen ou Mapas Auto-Organizáveis de Kohonen (SOM), fazem uso de um modelo de redes neurais competitivas. Para Lima e Minussi (2012), esse modelo

apresenta resultados agradáveis para o reconhecimento de sons, em especial os tons DTMF, e este modelo de RNA consegue garantir segurança e confiabilidade em seus resultados de classificação com os DTMF.

Os estudos e avanços na utilização de redes neurais convolucionais para a especificação de imagens apresentaram a possibilidade da sua aplicação em classificação de sons (SMALES, 2019). O emprego da combinação de modelos de RNA tem se tornado uma possibilidade para muitos pesquisadores da área, como a combinação de RNC e RNR. Carro et al., (2017) mostraram em seu estudo que a Rede Neural Recorrente combinada com uma Rede Neural Convolucional conseguiu melhores resultados de detecção.

1.1.4 A IA em benefício da tecnologia assistiva

Dentre os avanços tecnológicos, com a produção em larga escala de Inteligências Artificiais que fazem parte do dia a dia, existem também âmbitos específicos como o da TA, que é voltada para dar apoio às pessoas que possuem algum tipo de deficiência. TA é um termo ainda novo utilizado para identificar todo o conjunto de recursos e serviços que contribuem para proporcionar ou ampliar habilidades funcionais de pessoas com deficiência física ou cognitiva e consequentemente promover qualidade de vida e inclusão. (COOK & HUSSEY, 1995)

A possibilidade de reproduzir características humanas através da Inteligência Artificial fez com que pessoas com algum tipo de limitação física pudessem ter uma vida com maior inclusão social. Num sentido amplo percebemos que a evolução tecnológica caminha na direção de tornar a vida mais prática e sociável. Sem nos apercebermos, utilizamos constantemente ferramentas que foram especialmente desenvolvidas para favorecer e simplificar as atividades do cotidiano como os talheres, canetas, computadores, controle remoto, automóveis, telefones celulares, relógio etc. Isto é, uma longa lista de recursos assimilados à nossa rotina e que, num senso geral, “são instrumentos que facilitam nosso desempenho em funções pretendidas” (BERSCH & TONOLLI, 2006).

Várias são as pessoas que, possuem algum tipo de limitação física de diferentes categorias, já estão em processo de inclusão social através das novas tecnologias. Exemplificando todo o contexto abordado de pessoas com deficiência auditiva, já há tecnologias que auxiliam nas funções auditivas que incluem vários equipamentos

(infravermelho, FM), aparelhos para surdez, sistemas com alerta tátil-visual, celular com mensagens escritas e chamadas por vibração, softwares que favorecem a comunicação ao telefone celular transformando em voz o texto digitado no celular e em texto a mensagem falada, livros, textos e dicionários digitais em língua de sinais (HONORA, 2009).

Silveira (2019), em sua dissertação de mestrado, analisou a detecção de eventos sonoros para alertar surdos, testando alguns modelos de redes neurais artificiais e destacou a importância de uma tecnologia para alertar os surdos de situações de perigo. Também com a mesma intenção, mas trazendo para o cenário do trânsito, em 2017, cinco alunos da PUC - MG desenvolveram o Suricar, um aplicativo de celular para alertar condutores de veículos que possuem deficiência auditiva (ANDRADE, 2017).

1.2 APRESENTAÇÃO DO PROBLEMA

Souza et al. (2016), afirmam que, mesmo com direitos garantidos pelas leis de trânsito, os surdos convivem como seres “invisíveis” em um ambiente como o trânsito, onde a pessoa precisa de alguns sentidos como audição e visão. Motoristas com problemas de audição tem uma grande dificuldade em identificar e interpretar ruídos do trânsito, seja sons importantes como de buzinas, silvos de guardas de trânsito e sirenes. Isto dificulta a comunicação dos deficientes com outros condutores, pedestres e profissionais de trânsito (QUADROS, 2008).

Dados esses fatos, busca-se desenvolver uma Rede Neural Artificial que seja capaz de identificar padrões sonoros no trânsito.

1.3 JUSTIFICATIVA

A presença de problemas auditivos não significa que uma pessoa tenha que renunciar à sua vida social. Pelo contrário, frequentar eventos e participar regularmente de situações rotineiras é essencial para manter a qualidade de vida. Em vista disto, é necessário oferecer uma maior independência e inclusão aos deficientes auditivos no trânsito das ruas do país, de maneira que torne possível a participação e a interação deles na comunicação sonora que ocorre no trânsito.

Pessoas com deficiência auditiva até podem conseguir emitir a Carteira de Motorista e transitar pelas ruas, mas participar da comunicação sonora que ocorre neste âmbito exige um esforço notável e pode até levá-las a ocasionar ou sofrer um acidente de trânsito.

1.4 OBJETIVOS

1.4.1 Objetivo geral

Este projeto tem como finalidade principal desenvolver uma Inteligência Artificial através de uma rede neural profunda, que realize a detecção e classificação de sinais sonoros presente em um ambiente de trânsito de maneira eficiente, a fim contribuir para a construção de novas tecnologias assistivas para deficientes auditivos.

1.4.2 Objetivos específicos

Para que o objetivo geral seja alcançado, os seguintes objetivos específicos foram elencados:

- Estudar RNA para a detecção e classificação de eventos sonoros.
- Criar uma base de dados para o treinamento da rede.
- Desenvolver um modelo de RNA que processe e reconheça sinais sonoros de trânsito.
- Validar o modelo desenvolvido.

1.5. ESTRUTURA DO TRABALHO

O estudo a ser apresentado foi estruturado em quatro capítulos e apêndices. O primeiro capítulo contém a apresentação da contextualização, do problema/pesquisa, justificativa, objetivos e estrutura do trabalho. O segundo capítulo apresenta a introdução, a linha de pesquisa, metodologia e materiais utilizados, resultados obtidos através da RNA, discussões dos resultados e considerações finais. O terceiro capítulo reporta a síntese com conclusões gerais, discussão geral e as sugestões para trabalhos futuros. Por fim, apresentam-se as referências bibliográficas utilizadas.

2. DESENVOLVIMENTO DE UMA INTELIGÊNCIA ARTIFICIAL PARA A CLASSIFICAÇÃO DE SINAIS SONOROS DE TRÂNSITO

RESUMO

A capacidade humana de identificar sons é de suma importância nas atividades cotidianas, dentre elas a condução de veículos no trânsito, no entanto há uma parcela da sociedade que não recebe a atenção necessária para sua acessibilidade neste meio, os deficientes auditivos. O presente trabalho tem como objetivo estudar e desenvolver uma Inteligência Artificial que possa classificar sons do trânsito, tais como, apitos, sirenes, buzinas. Dentro do ecossistema de IA, foi escolhido o modelo de rede neural artificial do tipo convolucional, desenvolvido na linguagem Python. Para realizar seu treino, validação e teste, elaborou-se uma base de dados com 1.043 áudios. Mesmo com esta pequena quantidade de arquivos, obteve-se um modelo de Rede Neural Convolucional com 92% de acurácia, que, ao ser submetido a testes de classificação em tempo real, teve um desempenho admissível, mostrando que, por meio de uma futura implantação em veículos, pode ser usado para dar auxílio para pessoas com deficiência auditiva.

Palavras-chave: Deficientes auditivos; Rede Neural Convolucional; Sons do Trânsito;

ABSTRACT

The human ability to identify sounds is of utmost importance in everyday activities, among them the driving of vehicles in traffic, however there is a portion of society that does not receive the necessary attention for its accessibility in this environment, the hearing impaired. The present work aims to study and develop an Artificial Intelligence that can classify traffic sounds, such as whistles, sirens, and horns. Within the AI ecosystem, the convolutional artificial neural network model, developed in Python language, was chosen. To perform its training, validation and testing, a database with 1,043 audios was prepared. Even with this small amount of files, a Convolutional Neural Network model with 92% accuracy was obtained, which, when submitted to classification tests in real time, had an admissible performance, showing that, through a future implementation in vehicles, it can be used to provide aid for hearing impaired people.

Keywords: Hearing Impaired; Convolutional Neural Network; Traffic Sounds;

2.1 INTRODUÇÃO

Segundo o censo do IBGE de 2010, 5% dos brasileiros possuem um déficit na capacidade auditiva e menos de 1% dessas pessoas possuem carteira de motorista. A ausência de acessibilidade para deficientes auditivos no trânsito e na realização das avaliações na autoescola é um fator que causa transtornos para o indivíduo que deseja emitir ou renovar a CNH (BREAULT ET AL., 2020). Thorslund et al. (2013), dentre as conclusões de seu estudo sobre a influência da perda auditiva na segurança e mobilidade no trânsito, constatou que são baixas as chances de deficientes auditivos conseguirem licença para dirigir e que a perda auditiva afeta a atenção destes indivíduos ao dirigir um veículo.

A evolução da IA tem proporcionado à comunidade de deficientes uma maior qualidade de vida e inclusão social, devido ao seu potencial de conseguir reproduzir habilidades humanas. Na esfera da IA, quando a necessidade é fazer algum reconhecimento, as Redes Neurais Artificiais são bastante utilizadas como classificadores, aprendendo padrões e posteriormente certificando-os, como por exemplo, no reconhecimento de imagens feito pelo Google Lens.¹

Durante a pesquisa bibliográfica, encontrou-se poucos trabalhos sobre o uso de RNA para reconhecimentos de sons com a finalidade de auxiliar deficientes auditivos. Há alguns projetos como o Silence² e Suricar³ e outros estudos que visam reconhecer sons para alertar motoristas de determinados perigos. Desse modo, o estudo dessa área é importante quando se percebe os benefícios de sua aplicabilidade.

Neste trabalho será realizado um estudo e o desenvolvimento de uma RNA para tornar possível a elaboração de sistema que alerte deficientes auditivos (condutores de veículos motorizados) no trânsito, a fim de aumentar a acessibilidade desses indivíduos na sociedade. Pois, como foi identificado por Thorslund et al. (2013), há um interesse desse público em um sistema de alerta para motoristas independentemente da categoria auditiva.

¹ Ferramenta de pesquisa da Google, que utiliza a câmera do smartphone como entrada para realizar pesquisas e traduções.

² Dispositivo para identificar choro de bebe, para pais surdos.

³ Aplicativo de smartphone para alertar motoristas deficientes auditivos dos sons de trânsito.

2.2 MATERIAL E MÉTODOS

Este tópico apresenta seções de procedimentos práticos com a definição e construção da base de dados, a arquitetura inicial da RNA e o código para teste em tempo real.

Python foi a linguagem de programação escolhida para a implantação da rede e o ambiente para a construção do código foi o Google Colaboratory (Google Colab). Python é uma linguagem multiparadigma e sua sintaxe é muito mais simples que outras, como Java e C++, e permite construir soluções complexas com códigos pequenos (PYTHON SOFTWARE FOUNDATION, 2020). Além disso, é uma linguagem destacada pela sua alta capacidade no processamento de dados que podem ser usados no desenvolvimento de Machine learning com o uso das bibliotecas Tensor Flow.

O Google Colab, por sua vez, é um serviço em nuvem da Google que permite a edição de código fonte. Essa ferramenta possibilita o trabalho colaborativo, assim uma equipe pode trabalhar na criação de um código (COLABORATORY, 2022).

2.2.1 Base De Dados

Para a construção da base de dados (*database*) foram coletados áudios dos seguintes padrões: apito, buzina e sirene, dos repositórios na internet zapsplat.com e freesound.org. A fim de aumentar a variedade e quantidade dos dados, também foram gerados outros áudios com gravador de celular e o programa de computador Audacity. Ao todo o conjunto possui 1.043 áudios e esses arquivos possuem uma variação de 2 à 5 segundos de duração, são áudios monofásicos e em formato “.wav” que, conforme Fahy (2007), este formato ajuda a garantir uma boa fidelidade sonora. A Tabela 01 apresenta a distribuição e quantos arquivos foram adquiridos de cada categoria.

Tabela 01: Composição do Database

Conjunto	Quantidade
Apito	337
Buzina	345
Sirene	361

Fonte: Autoria própria (2022)

No google drive⁴, foi criada uma pasta nomeada como “database”, na qual, 3 subpastas foram criadas para comportar os padrões supracitados.

Criou-se também uma pasta (“database 2”) para armazenar áudios, fora da base oficial, que pertencem aos padrões reconhecidos pela RNA e áudios com sons fora desses padrões. Essa segunda base possui 45 áudios que serão usados para testar o modelo treinado e verificar se, de fato, a RNA foi capaz de aprender e generalizar⁵ mesmo que os áudios tenham outros ruídos no fundo.

2.2.2 Rede Neural Artificial

Entre os modelos/arquiteturas de RNA, foi escolhida a Rede Neural Convolutiva por ser uma arquitetura de “deep learning” (aprendizado profundo) e ser bastante indicado para a utilização em detecção e eventos sonoros (CAKIR E.; HEITTOLA, 2015). Sendo assim, adequada para os fins que este estudo almeja.

Durante as pesquisas, enquanto se estudava sobre a biblioteca Tensor Flow, que é usada para desenvolver/criar modelos de *machine learning*, foi encontrado um código de RNA para identificação de sons⁶, mas para reconhecer áudios com palavras chaves. O código foi de encontro com o que os autores pretendiam construir, pois é implantado na linguagem de programação Python. Então foi a partir desse código disponibilizado no site que iniciou o desenvolvimento e testes da RNA para reconhecer sinais sonoros de trânsito.

O código acessa a base de dados, mapeia os arquivos e com base nos títulos das pastas classifica cada áudio. Transforma os arquivos em espectrogramas⁷ e são essas novas representações dos dados que alimentarão a rede. Esses dados são passados para o modelo de RNA que retorna a matriz de confusão⁸ e a inferência através de representação gráfica.

Vale destacar que a arquitetura da RNC que estava inicialmente no código foi alterada, pois tinha uma configuração bem simples e não garantiria bons resultados para o problema deste trabalho. Assim, foi inserido ao código uma estrutura como uma pilha de

⁴ Google Drive é um serviço de armazenamento e sincronização de arquivos.

⁵ Refere-se a quão bem um modelo de aprendizado de máquina aprendeu.

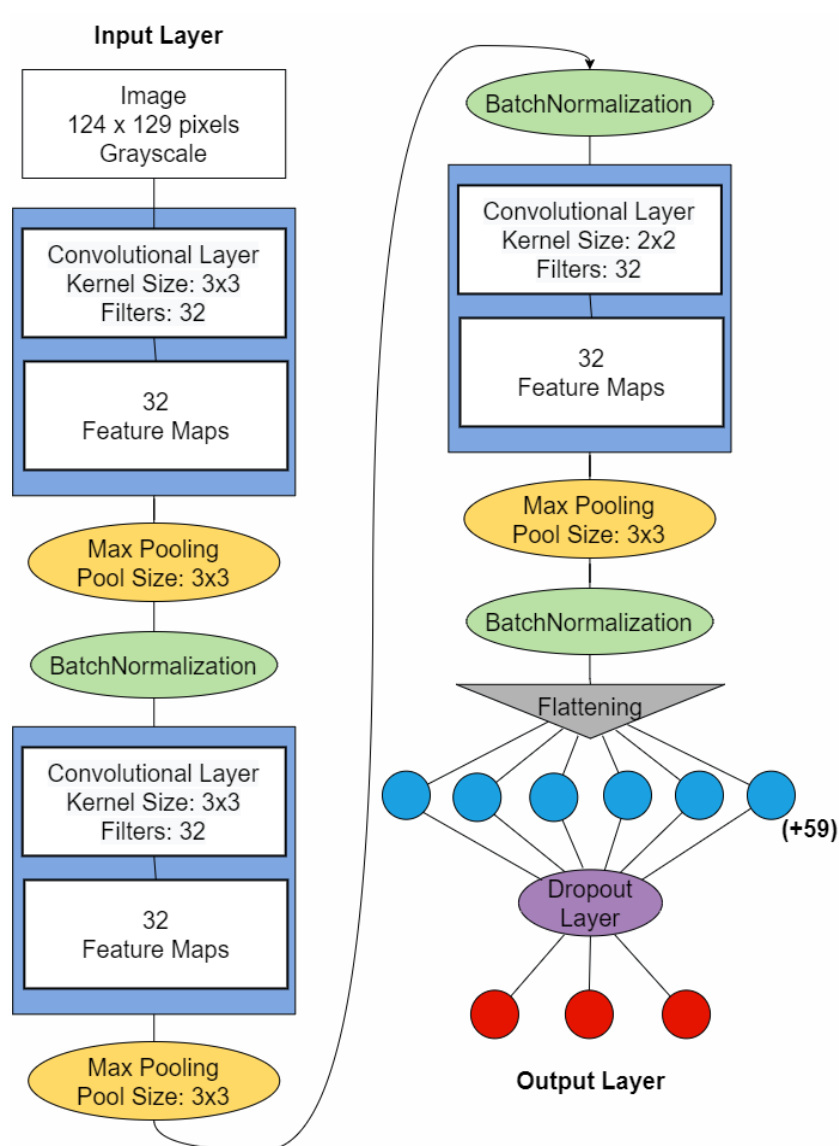
⁶ <https://www.tensorflow.org/tutorials/audio/simpleaudio?hl=%20pt-br>

⁷ Gráfico que mostra a intensidade por meio do escurecimento ou coloração do traçado, as faixas de frequência no eixo vertical e o tempo no eixo horizontal.

⁸ Matriz quadrada em que se compara os verdadeiros valores de uma classificação com os valores preditos através de algum modelo.

camadas, uma camada Conv2D⁹ de entrada, duas camadas Conv2D ocultas, essas três camadas contam com camadas MaxPooling2D¹⁰ e camadas de BatchNormalization¹¹. Após isso, tem um nivelamento da camada (Flatten), na sequência, uma camada oculta densa e uma camada de Dropout e, por fim, a camada densa de saída. A descrição da arquitetura em uma forma visual se encontra na figura 01.

Figura 1 - Topologia inicial da RNA



Fonte: Autoria própria (2022)

⁹Camadas convolucionais de duas dimensões.

¹⁰Camada de pool máximo para dados espaciais de duas dimensões.

¹¹Camada de normalização em lote, aplica uma transformação que mantém a saída média próxima a 0 e o desvio padrão da saída próximo a 1.

2.2.2.1 Argumentos nas camadas

Para escolher os argumentos que iriam compor as camadas da rede foi consultada a documentação da biblioteca TensorFlow para entender melhor os parâmetros. Então, optou-se que para o primeiro ensaio as camadas Conv2D possuísem como argumento 32 filtros, kernel de dimensão 3x3 (altura = 3 e a largura = 3) para as 1ª e 2ª Conv Layer e para a 3ª, kernel de dimensão 2x2. Além desses parâmetros, as camadas têm a função de ativação Unidade Linear Retificada (ReLU), com exceção da primeira que recebe a entrada dos dados.

Na MaxPooling2D a janela sobre a qual se deve obter o valor máximo tem tamanho 2x2. E o stride de 2x2, que especifica até onde a janela de pooling se moverá para cada etapa de pooling. Por último, o preenchimento (padding) que recebe atribuição “same” para resultar no preenchimento uniforme para a esquerda/direita ou para cima/para baixo da entrada, de modo que a saída tenha a mesma dimensão de altura/largura da entrada.

A primeira camada oculta densa possui 64 neurônios e função de ativação ReLu. Em seguida tem a camada de dropout a fim de evitar o overfitting no treinamento, realizando uma regularização, na qual 30% dos neurônios são desligados aleatoriamente juntamente com suas conexões.

A camada densa de saída tem como argumentos o valor inteiro 3 que é correspondente ao número de padrões e a função de ativação “softmax” que vai retornar probabilidade para cada classe.

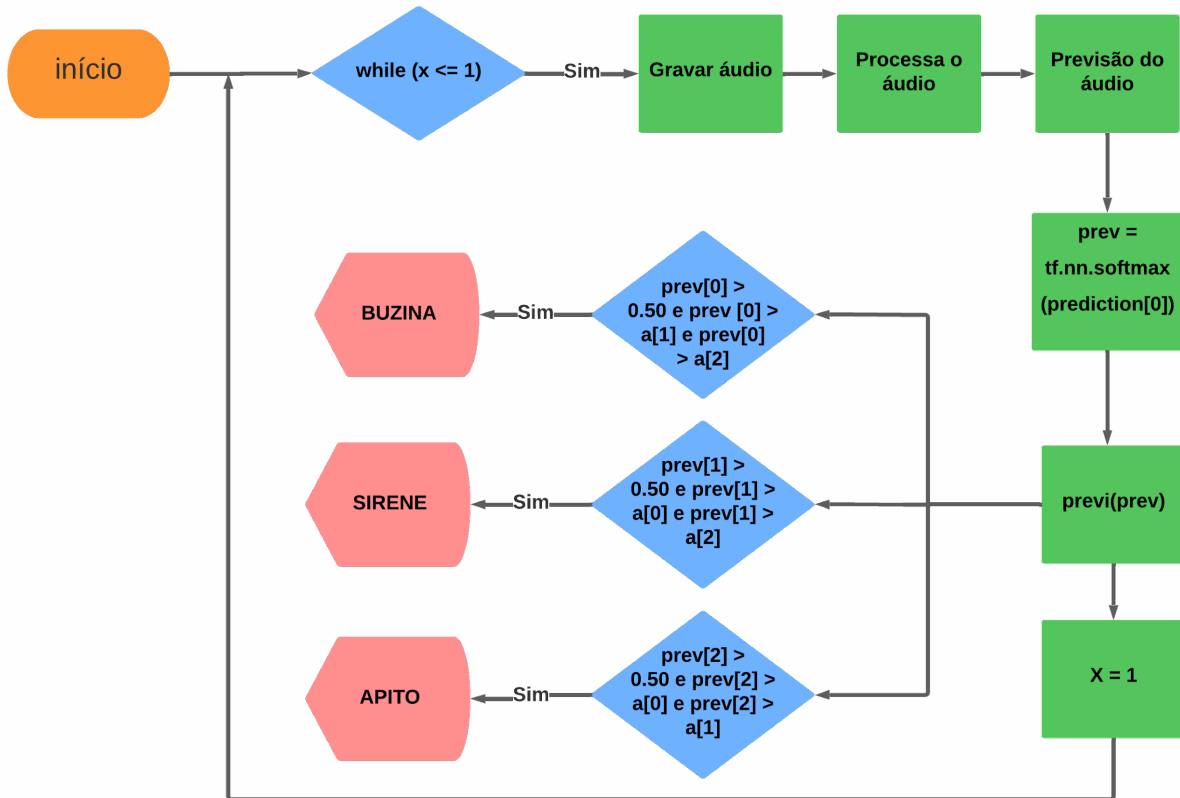
2.2.3 Classificação em Tempo Real

Para testar o comportamento do modelo da RNC em tempo real, foi criado um script em python, utilizando as bibliotecas PyAudio e WAV para captação e geração dos áudios, além da biblioteca TensorFlow. Para rodar o código foi utilizado a IDE PyCharm, pois o Colab apresenta conflitos ao reconhecer o microfone do computador para captar o som.

O script de captação consiste em um laço de repetição infinito do tipo “while”, simulando o reconhecimento em tempo real. No laço há duas funções: capturar o áudio e imprimir a previsão para informar a classificação.

A função de captação grava áudios de quatro segundos. Na função de imprimir a previsão, foi estabelecida uma estrutura condicional para considerar a previsão apenas quando for igual ou maior que 0.50 (50%), a fim de evitar enganos e classificar somente os padrões treinados. A Figura 02 ilustra o comportamento do código.

Figura 02 - Fluxograma Classificação em Tempo Real



Fonte: Autoria própria (2022)

2.4 RESULTADOS

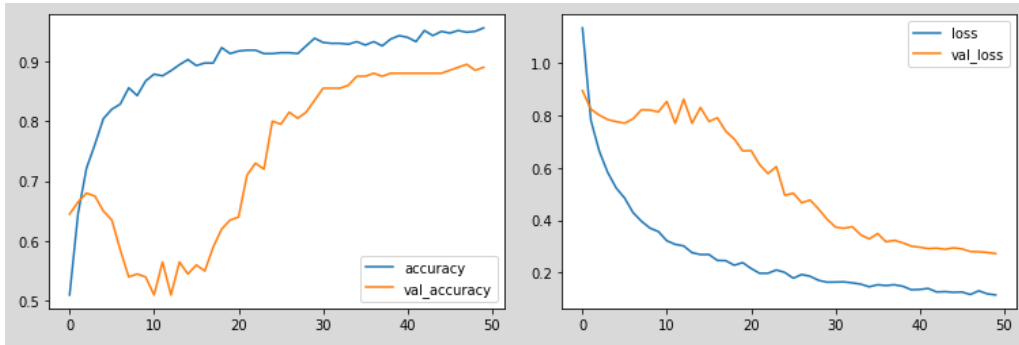
Para encontrar uma RNA que atenda ao problema, foram realizados 7 testes, com 50 épocas cada, modificando algumas características estruturais da RNC, como a distribuição de dados para treinamento, teste e validação e os argumentos das camadas. Os resultados desses testes podem ser observados no quadro 2.

Quadro 2 – Configuração das RNCs e seus resultados

Nº do Teste	Distribuição da Database	Alteração na Arquitetura da Rede	Acurácia	Teste de classificação			
				Áudios de Apito	Áudios de Buzina	Áudios de Sirene	Áudios diferente dos padrões
1	Treino: 620 Validação: 220 Teste: 203	Arquitetura Inicial	87%	Acertou com previsão acima de 0.50	Acertou com previsão abaixo de 0.50	Acertou com previsão acima de 0.50	Indicou um dos padrões com previsão acima de 0.50
2	Treino: 600 Validação: 243 Teste: 200	1ª camada convolucional 128 filtros e o kernel size de 4x4, 2ª camada 64 filtros	92%	Erro na classificação	Acertou com previsão acima de 0.50	Acertou com previsão acima de 0.50	Indicou um dos padrões com previsão acima de 0.50
3	Treino: 600 Validação: 300 Teste: 157	1ª e 2ª camada convolucional com 64 filtros e o kernel size de 3x3, 3ª camada 32 filtros e kernel size de 2x2	89%	Acertou com previsão entre 0.40 e 0.50	Erro na classificação	Acertou com previsão acima de 0.50	Indicou um dos padrões com previsão acima de 0.50
4	Treino: 700 Validação: 200 Teste: 143	Arquitetura Inicial	95%	Acertou com previsão acima de 0.50	Acertou com previsão acima de 0.50	Acertou com previsão acima de 0.50	Indicou um dos padrões com previsão acima de 0.50
5	Treino: 700 Validação: 200 Teste: 143	Na 3ª camada convolucional substituiu-se a camada de Batch Normalization por uma de dropout com 10% de neurônios para serem desligados aleatoriamente	92%	Acertou com previsão acima de 0.50	Acertou com previsão acima de 0.50	Acertou com previsão acima de 0.50	Indicou um dos padrões com previsão abaixo de 0.40
6	Treino: 700 Validação: 200 Teste: 143	Na 2ª e 3ª camada convolucional, foram trocadas as camadas de Batch Normalization por uma de dropout de 10% e de 20%, respectivamente	89%	Acertou com previsão acima de 0.50	Erro na classificação	Acertou com previsão acima de 0.50	Indicou um dos padrões com previsão abaixo de 0.50
7	Treino: 700 Validação: 200 Teste: 143	Inverteu-se as posições das Camada de MaxPolling e Batch Normalization nas camadas	92%	Acertou com previsão abaixo de 0.50	Acertou com previsão acima de 0.50	Acertou com previsão acima de 0.50	Indicou um dos padrões com previsão acima de 0.50

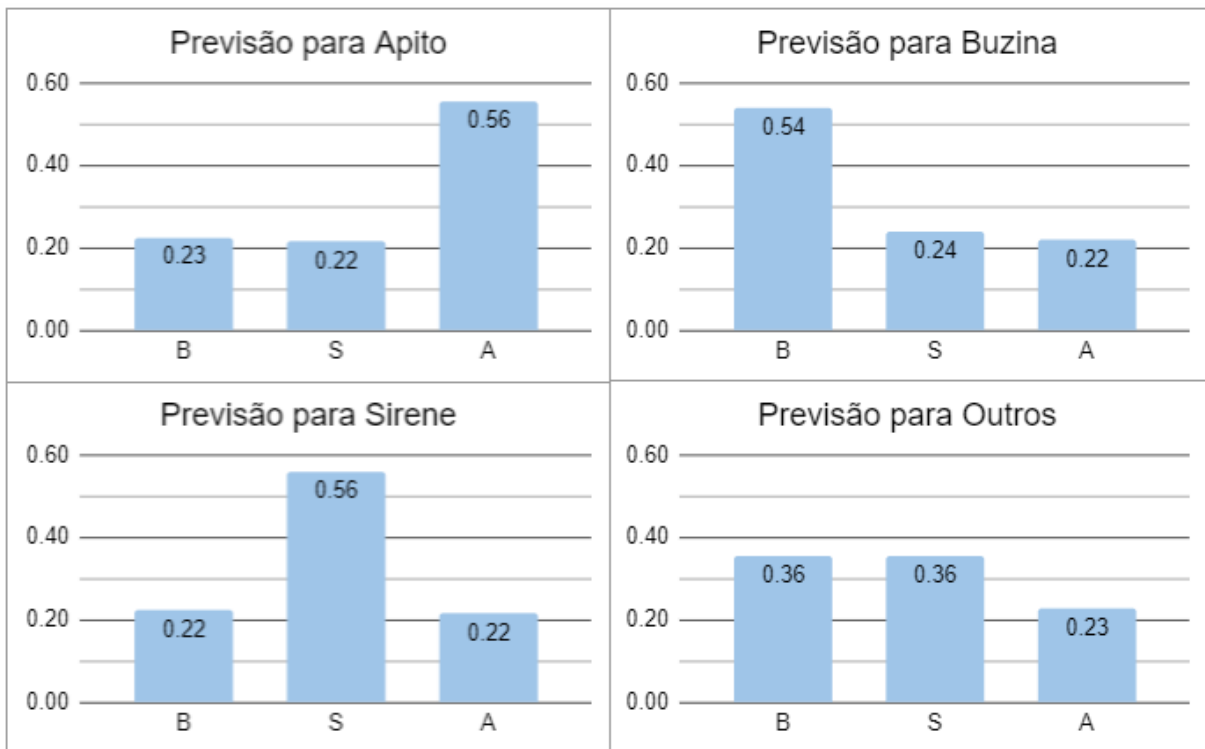
As matrizes de confusão, os gráficos de acurácia e erro e as previsões para áudios de ensaio de cada teste encontram-se mostrados no apêndice deste trabalho, com exceção do Teste 5, que está neste tópico (gráficos 01 e 02) por ter se destacado com melhor resultado.

Gráfico 01 - Gráficos da Acurácia e Erro (Teste 5)



Fonte: Autoria própria (2022)

Gráfico 02 - Gráficos de previsão para sons (Teste 5)



Fonte: Autoria própria (2022)

O modelo do Teste 5, foi utilizado para a classificação com o código de previsão em tempo real. Nesse teste ele apresentou uma performance mediana na classificação das categorias. Em alguns momentos o programa deixou de informar os padrões que estavam sendo emitidos no ambiente mas, por outro lado, na inexistência dos padrões de

classificação da rede não ocorreu nenhuma mensagem informando a identificação das classes.

2.5 DISCUSSÃO

Ao realizar as análises dos experimentos desenvolvidos é possível identificar que, apesar dos testes apresentarem resultados distintos, os gráficos de acurácia e erro possuíam um comportamento parecido, começando em sintonia, mas após algumas épocas os modelos evoluíram rapidamente nos dados de treino, enquanto nos dados de validação o desempenho reduzia, depois estabilizava e então evoluía de maneira gradual.

Mesmo com as camadas *BatchNormalization* e Dropout com seus respectivos argumentos, que colaboram para redução do Overfitting (TENSORFLOW, 2015) os gráficos dos testes apresentaram o distanciamento entre o desempenho dos dados de treinamento e de validação, ao longo dos ciclos. A tática aplicada no Teste 6 garantiu gráficos com performance um pouco melhores, mas em virtude de as informações serem perdidas com o processo de dropout nas camadas convolucionais, a sua acurácia ficou abaixo de 90% e apresentou erro na classificação de buzina. Ao aplicar a normalização depois que os dados saíssem das Conv2D (Teste 7), também não resolveu o problema da não generalização que se desenhava nos gráficos, ao contrário, só aumentou o overfitting.

O fato da base de dados ser constituída por um pouco mais de mil arquivos, sem dúvida, refletiu bastante nos resultados dos testes. Uma vez que não foi possível realizar uma distribuição de 80% para treinamento e o restante para os demais conjuntos, ou até mesmo dividi-los em 70% para dados de treino, 20% de teste e 10% para a validação, como geralmente é feito (SCACCIA, 2020). A aplicação de uma dessas distribuições, faria com que restassem poucos elementos para o conjunto de validação e teste. A divisão realizada no Teste 4 foi a que se apresentou como a melhor para a quantidade de dados da base.

Um fato curioso na obtenção desses resultados foi que o teste com maior acurácia (Teste 4), que teve como mudança apenas a redistribuição do database, não conseguiu responder da forma esperada quando foi inserido um áudio diferente dos padrões para ele classificar. O modelo que foi compatível com o esperado foi o do Teste 5, que além de apresentar uma acurácia acima dos 90%, os testes com os áudios que não fazem parte do

database foram bastante satisfatórios, o que demonstrou a capacidade de generalizar e classificar.

Em relação ao teste com o código para classificação em tempo real, o melhor modelo adquirido não repetiu o comportamento dos testes feitos com os áudios da segunda base de dados. Durante alguns momentos de testes, com os sons das categorias, a taxa de predição para a classe correta não atingiu a medida estabelecida (0.50). Levando em consideração este fato, poderia ser diminuído o valor para 0.40 ou 0.45, mas isso faria com que programa, em um dado instante, classifica-se de forma precipitada, pois os sons possuem algumas familiaridades e, dependendo do momento em que é captado o áudio, o modelo poderia prever uma informação incorreta. Então, é importante que a previsão só seja considerada quando for maior ou igual a 50%.

O instante em que os sons ocorrem no intervalo de tempo em que é gravado o áudio, para ser classificado, influenciou nas taxas de previsão abaixo da métrica estabelecida. Por exemplo, quando o sinal sonoro era captado nos instantes finais da gravação o áudio não era classificado. Então o modo como os áudios foram captados teve relação com a baixa taxa de classificação nesse teste.

Além disso é válido destacar que como a captação dos sons foi feita por um microfone de notebook que possui baixo desempenho e que há uma relação direta entre a qualidade sonora captada e o rendimento do modelo nos testes com o código de verificação em tempo real, um microfone dedicado garantiria melhores resultados.

2.6 CONSIDERAÇÕES FINAIS

Mesmo com uma base de dados pequena, o uso da técnica de dropout no final da terceira camada convolucional da rede (Teste 5) garantiu o desenvolvimento de um modelo modesto que classificou bem os dados estáticos e dados captados em tempo real, tornando possível a utilização da RNA do tipo convolucional como classificador de sinais sonoros do trânsito. Além disso, os resultados obtidos neste trabalho possibilitam a construção de um sistema de alerta para motoristas com deficiência auditiva e viabilizam a aplicação do modelo desenvolvido na identificação de outros sons. Os materiais gerados

neste trabalho estão disponíveis em código aberto no GitHub¹² para novos estudos e implementações.

¹² <https://github.com/PedroGomesjr/RNA-sinais-sonoros-de-transito>

3. SÍNTESE

3.1 PRINCIPAIS CONCLUSÕES

A Rede Neural Convolucional demonstrou ser uma poderosa ferramenta para auxiliar e suprir necessidades físicas, como na área de identificação de sons para deficientes auditivos, que pode proporcionar, a essa parcela da população, mais segurança ao dirigir. A metodologia utilizada pode ser aplicada para a classificação de diferentes padrões de sons de alerta. Além disso, observou-se que com poucos dados já se foi capaz de obter ótimos resultados.

Desempenhar esforços para a elaboração de tecnologias para deficientes, como os auditivos, é de suma importância para ter uma sociedade inclusiva e ambientes interativos entre todos. Este trabalho acadêmico trouxe observações relevantes acerca do tema estudado que poderão ser utilizadas em outros estudos no meio acadêmico, pois pessoas interessadas no assunto poderão usar da metodologia aplicada neste projeto para conseguir ir além na resolução do problema, ou usá-la para uma possível solução em outros campos de atuação, afinal, esta obra se limitou exclusivamente ao desenvolvimento de uma RNA.

3.2 DISCUSSÃO GERAL

A quantidade de arquivos nos conjuntos de dados da RNA não foi empecilho tão grande visto que se obteve um modelo com bom desempenho nas classificações dos padrões. Afinal, a base garantiu a variedade de dados no treinamento, validação e testes (SAVIETTO, 2021). É fato que um database com o dobro ou triplo de dados poderia garantir uma RNA com performance maior do que a obtida. Porém os padrões que este trabalho se propõe a classificar possuem pouca variação.

Durante o processo de testes poderia ser considerado a aplicação do Image Data Generator¹³ para aumentar a quantidade de dados (TENSORFLOW, 2015), mas como a RNC em questão trabalha com imagens de espectrograma, esse tipo de aplicação poderia

¹³ É uma técnica para criar artificialmente novos dados de treinamento a partir de dados de treinamento existentes. Isso é feito pegando o conjunto de dados e aplicando uma série de transformações aleatórias a cada imagem como rotação aleatória, redimensionamento, cisalhamento, etc.

afetar o desempenho da mesma, pois é melhor indicado quando a base de dados é composta por imagens de objetos (BHANDARI, 2020).

3.3 PERSPECTIVA FUTURA

A conclusão deste artigo mostrou a viabilidade de uma futura aplicação da RNC em um sistema embarcado, a fim de dar assistência aos motoristas deficientes auditivos. Com esta ferramenta o projeto poderá ser levado a outro patamar, sendo capaz, por exemplo, de comunicar ao deficiente auditivo de qual direção está sendo emitido o sinal por meio de sensores instalados ao redor do veículo. Para a construção deste sistema é proposto analisar, juntamente com a comunidade de deficientes auditivos, qual seria a melhor forma de informar os sons identificados, compreender quais cores podem representar cada padrão e estudar a melhor maneira de transmitir sua visualização.

O código da classificação em tempo real deverá ser otimizado, pois o utilizado no trabalho foi apenas para testes iniciais. Quanto ao modelo utilizado para esse sistema, pode-se propor um estudo acentuado em outros tipos de arquiteturas de RNA, como a mescla da rede neural recorrente com a convolucional, e em outras formas de extração de dados dos áudios, como MFCC, a fim de melhorar a taxa de acertos da inteligência desenvolvida.

4 REFERÊNCIAS BIBLIOGRÁFICAS

ANDRADE, Laurie. SuriCar: aplicativo para auxiliar deficientes auditivos in: AUTOPAPO. São Paulo, 16 dez. 2017. Disponível em: <<https://autopapo.uol.com.br/noticia/suricar-aplicativo-deficientes-auditivos/>> Acesso em: 20 ago. 2021.

BREAULT, Camille; DÉRY, Julien; BEAUDRY, Maryse; CHÉNARD, Sylvie; GÉLINAS, Isabelle; MORALES, Ernesto e LAMONTAGNE, Marie-Eve Lamontagne. **From knowledge to action: measuring the gaps between the evidence and adapted driver education services for young adults with disabilities.** Transportation Research Interdisciplinary Perspectives, Québec, v. 4, p. 1-12. 2020. Disponível em: <https://www.sciencedirect.com/science/article/pii/S2590198219300648>. Acesso em 13 out. 2021.

BHANDARI, Aniruddha. Image Augmentation on the fly using Keras ImageDataGenerator. Analytics Vidhya, 2020. Disponível em: <<https://www.analyticsvidhya.com/blog/2020/08/image-augmentation-on-the-fly-using-keras-imagedatagenerator/>>. Acesso em: 01, mai de 2022.

CAKIR E.; HEITTOLA, T. H. H. V. T. **Polyphonic sound event detection using multi label deep neural networks.** IEEE International Joint Conference on Neural Networks (IJCNN), [S.l.], 2015.

COOK, A.M. & HUSSEY, S. M. (1995) Assistive Technologies: Principles and Practices. St. Louis, Missouri. Mosby - Year Book, Inc.

FAHY, Frank J. **Foundations of Engineering Acoustics.** Oxford: Academic Press, 2000.

HONORA, Márcia; FRIZANCO, Mary. **Livro ilustrado de Língua Brasileira de Sinais: desvendando a comunicação usada pelas pessoas com surdez.** São Paulo: Ciranda Cultural, 2009.

KAELBLING, Leslie P. **Robotics and learning. Computational intelligence.** The MIT encyclopedia of the cognitive sciences. Cambridge: The MIT Press.1999, p.723 - 724.

LEFÉVRE, F, Lefèvre AMC. **Curso teórico prático de introdução ao de introdução ao Discurso do Sujeito Coletivo e ao software Qualiquantisoft.** Instituto de Pesquisa do Discurso do Sujeito Coletivo (IpDsc), 2012.

LIMA, Fernando P. and Carlos R. Minussi. **“RECONHECIMENTO DE TONELADAS DTMF UTILIZANDO UMA REDE NEURAL DE KOHONEN.”** (2012).

LOBO, Luiz Carlos. **Inteligência Artificial e Medicina. Revista Brasileira de Educação Médica [online].** 2017, v. 41, n. 2 [Acessado 20 Setembro 2021] , pp. 185-193. Disponível em: <<https://doi.org/10.1590/1981-52712015v41n2esp>>. ISSN 1981-5271. <https://doi.org/10.1590/1981-52712015v41n2esp>.

LOPES, M. C. **Surdez educação**. Belo Horizonte: Autêntica, 2007.

MORAIS, Járson Néri Silva de. **Aprendizado computacional dinâmico baseado em visão de máquina**. Belém, 2013. 46f. Trabalho de Conclusão de Curso (Bacharelado em Ciência da Computação) - Centro Universitário do Estado do Pará, 2013.

MUDA, L. B. M. E. I. **Voice Recognition Algorithms using Mel Frequency Cepstral Coefficient (MFCC) and Dynamic Time Warping (DTW) Techniques**. Journal Of Computing, [S.l.], v.2, p.138–143, 2010.

OMS - Organização Mundial da Saúde. **World Report on Hearing**. Geneva: OMS, 2021. Disponível em: <https://www.who.int/teams/noncommunicable-diseases/sensory-functions-disability-and-rehabilitation/highlighting-priorities-for-ear-and-hearing-care>. Acesso em: 8 fev. 2022

PYTHON SOFTWARE FOUNDATION. **Python, 2020**. Página de documentação. Disponível em: <<https://docs.python.org/pt-br/3/tutorial/index.html>. Acesso em: 06 de abr. de 2022

POR QUE A LINGUAGEM PYTHON É TÃO POPULAR EM MACHINE LEARNING E INTELIGÊNCIA ARTIFICIAL?. Data Science Academy, 2020. Disponível em: <<https://blog.dsacademy.com.br/por-que-a-linguagem-python-e-tao-popular-em-machine-learning-e-inteligencia-artificial/>>. Acesso em: 26, mar de 2022.

QUADROS, Rinoce Müller de. **Educação de surdos: a aquisição da linguagem**. Porto Alegre: Artmed, 2008.

QUEIROZ, Malthus Oliveira de. **Dicionário Sucesso da Língua Portuguesa**. 1º ed. Conforme o Novo Acordo Ortográfico. Recife, PE: Distribuidora de Edições Pedagógicas, 2015.

RODRIGUES, Leonardo C. **Fundamentos, Tecnologias E Aplicações De Veículos Autônomos**. Trabalho de Conclusão de Curso (Bacharel em Engenharia Eletrônica) - Universidade Tecnológica Federal do Paraná, 2017.

ROSSO, Pedro Augusto Di Francia. **Desenvolvimento e Implementação de Aplicativo de Segurança para Sistemas de Áudio em Smartphones**. Trabalho de Conclusão de Curso (graduação) - Universidade Federal de Santa Catarina, Campus Araranguá, Graduação em Engenharia de Computação, Araranguá, 2017.

SAVIETTO, João Vitor. Machine Learning: Métricas, Validação Cruzada, Bias e Variância. Medium, 2021. Disponível em: <<https://medium.com/@jvsavietto6/machine-learning-m%C3%A9tricas-valida%C3%A7%C3%A3o-cruzada-bias-e-vari%C3%A2ncia-380513d97c95>>. Acesso em: 01, mai de 2022.

SCACCIA, Kevin. Validação Cruzada Aninhada com Scikit-learn.Data ML, 2020. Disponível em: <https://dataml.com.br/validacao-cruzada-aninhada-com-scikit-learn/>>. Acesso em: 30, abr. de 2022.

SCHMIDT, Renata. "Surdo e Deficiente Auditivo é a mesma coisa? Entenda a diferença!". Guia de rodas, São Paulo, 07 de Setembro. 2020. Disponível em: <<https://guiaderodas.com/surdo-e-deficiente-auditivo-e-a-mesma-coisa-entenda-a-diferencia/>>. Acesso em: 29 de Maio de 2022.

SILVEIRA, Douglas Severo . **Estudo para detecção de eventos sonoros como comunicação**. 2019. 93 f. Dissertação (mestrado) – Universidade Federal do Rio Grande –FURG, Programa de Pós-Graduação em Computação, Rio Grande/RS, 2019.

Souza, Vânia Mendes de et al. A inclusão de surdos no trânsito. Revista CEFAC [online]. 2016, v. 18, n. 3 [Acessado 30 Maio 2022], pp. 677-687. Disponível em: <<https://doi.org/10.1590/1982-0216201618317615>>. Acesso em: 29 de Maio de 2022.

SOUZA, Carlos Francisco Soares de. **Utilização de redes neurais artificiais no diagnóstico do exame fonético-fonológico**. Belém, 2006. 80f. Trabalho de Conclusão de Curso (Bacharelado em Ciência da Computação) - Centro Universitário do Pará, 2006.

SMALES, Mike. Sound classification using deep learning. in: MEDIUM. São Paulo, 19 dez. 2019. Disponível em: <<https://mikesmales.medium.com/sound-classification-using-deep-learning-8bc2aa1990b>> Acesso em: 20 jan. 2022..

TENSORFLOW. **Tensor Flow: reconhecimento de áudio simples: reconhecendo palavras-chave**, c2018. Tutoriais. Disponível em: <<https://tensorflow.org/>>. Acesso em: 05 de dez. de 2021.

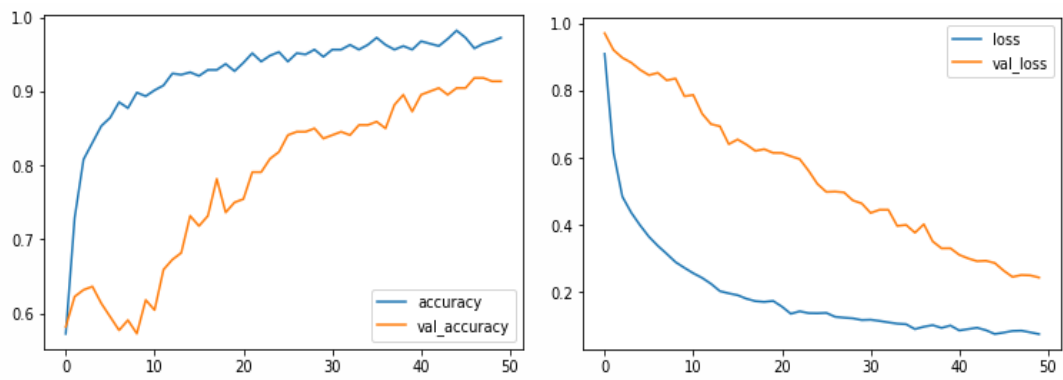
THORSLUND, B., Peters, B., Lyxell, B. *et ai*. **A influência da perda auditiva na segurança e mobilidade do transporte**. *EUR. Transp. Res. Rev.* 5, 117–127 (2013)

YANG, J e COUGHLIN, J. F. **In-vehicle technology for self - driving cars: Advantages and Cambridge**. v.15, n.2, p.333-340, mar. 2014.

Z. ZUO *et al.*, "**Convolutional recurrent neural networks: Learning spatial dependencies for image representation**" *2015 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2015, pp. 18-26, doi: 10.1109/CVPRW.2015.7301268.

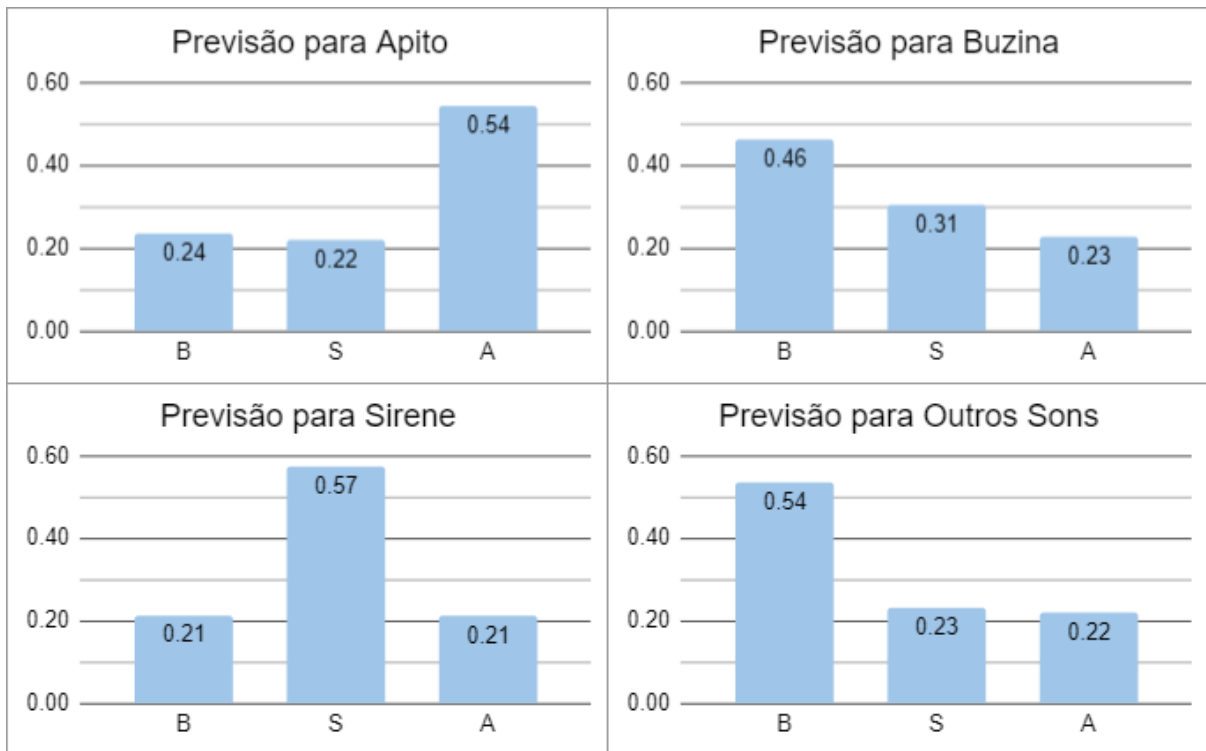
ZHANG, Chi. ZHANG, Yue. CHEN, Chen. SongNet: Real-time Music Classification. Stanford University, 2020. Disponível em: <<https://cs229.stanford.edu/proj2018/report/53.pdf>>. Acesso em: 20 jan. 2022.

APÊNDICE A - Gráficos da Acurácia e Erro Teste 1



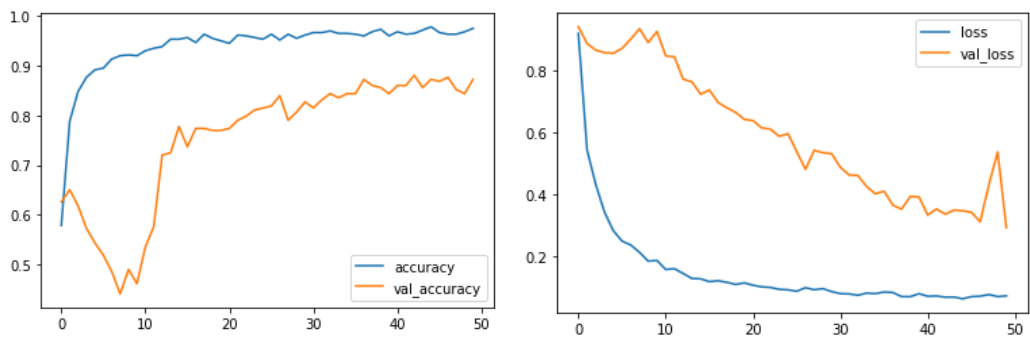
Fonte: Autoria própria (2022)

APÊNDICE B - Gráficos de previsão para som (Teste 1)



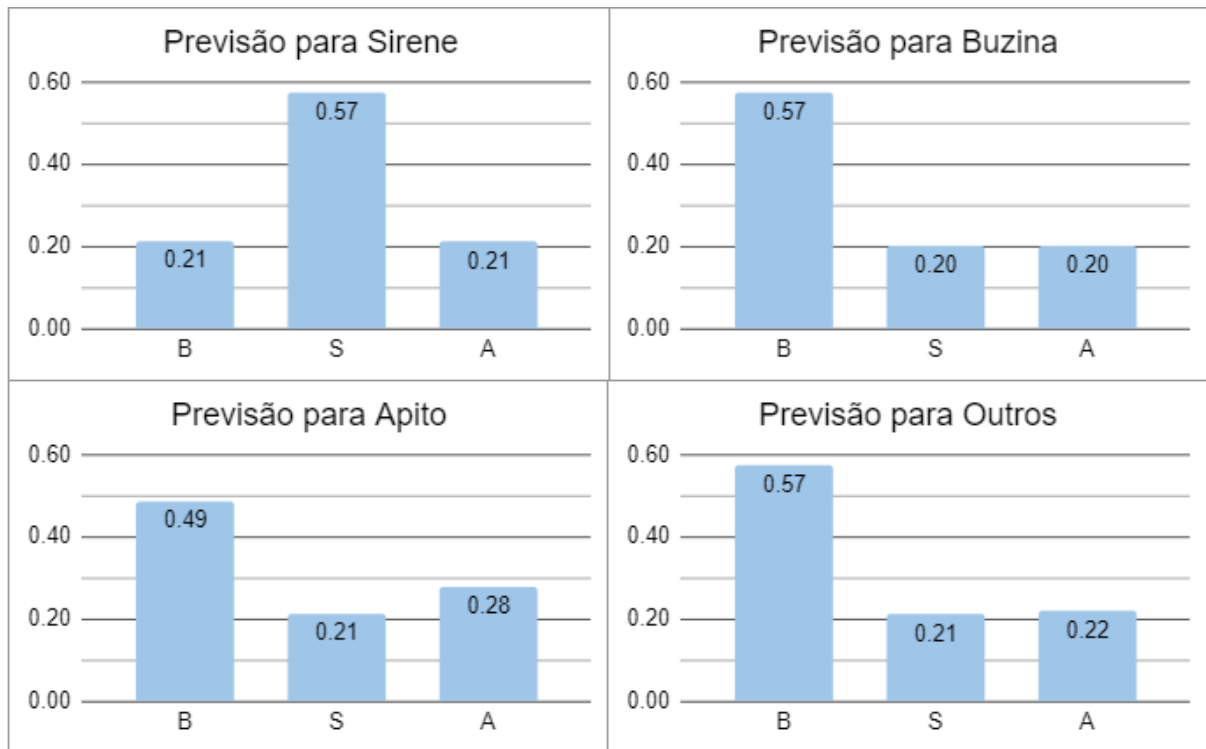
Fonte: Autoria própria (2022)

APÊNDICE C - Gráficos da Acurácia e Erro Teste 2



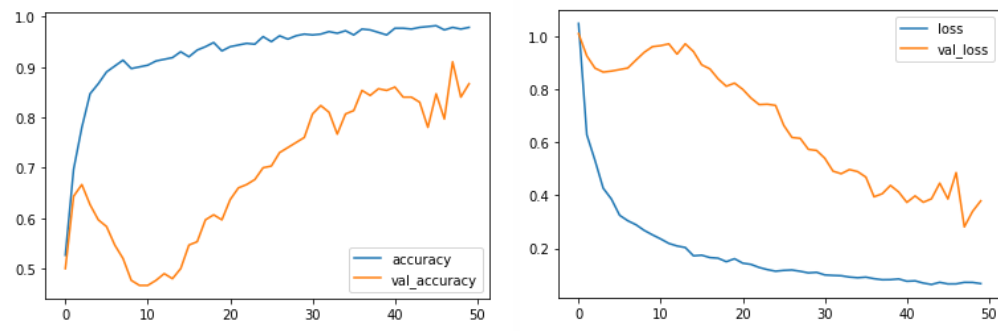
Fonte: Autoria própria (2022)

APÊNDICE D - Gráficos de previsão para som (Teste 2)



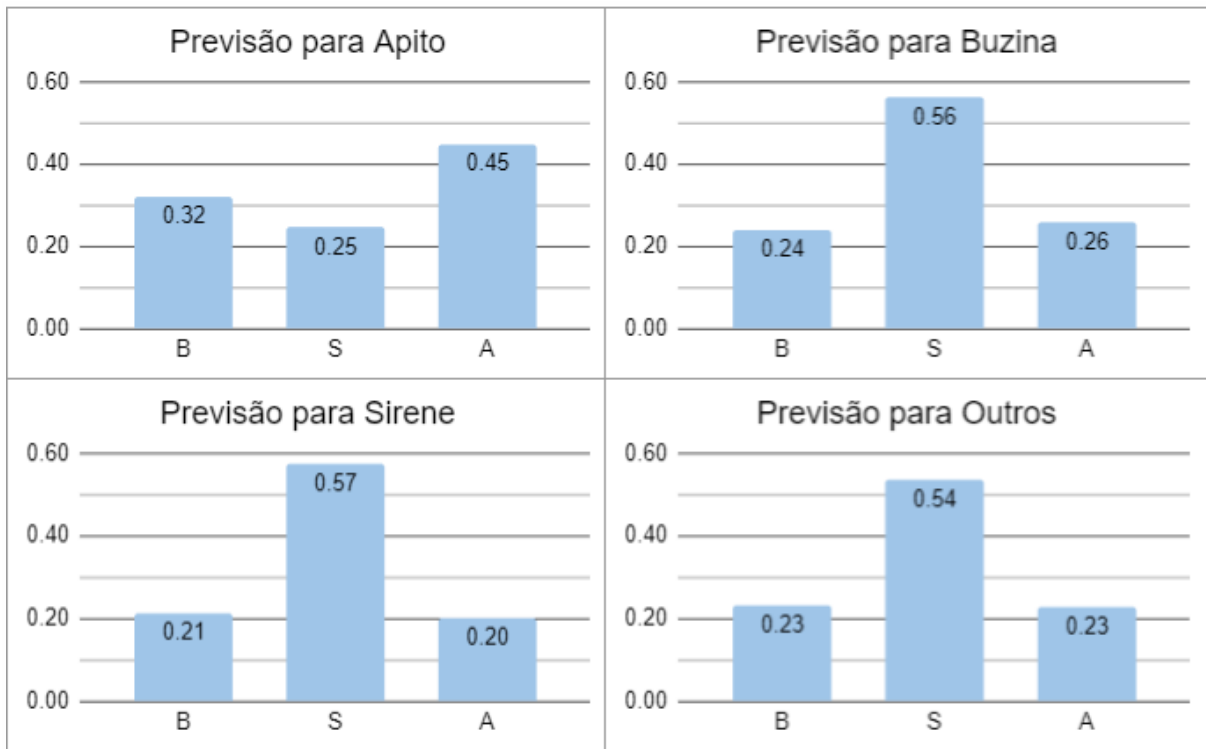
Fonte: Autoria própria (2022)

APÊNDICE E - Gráficos da Acurácia e Erro Teste 3



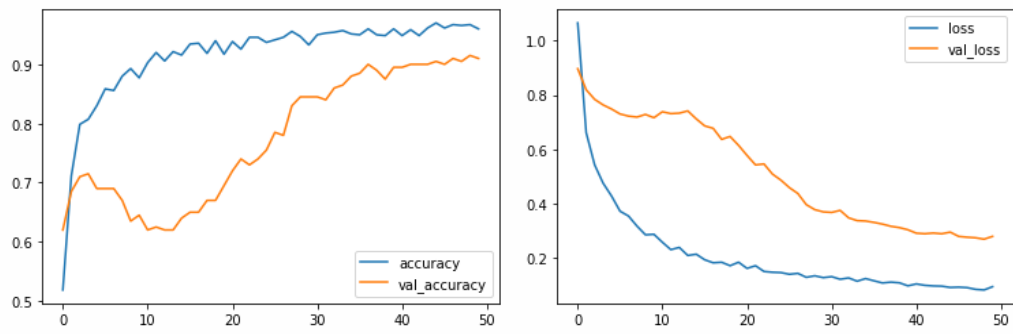
Fonte: Autoria própria (2022)

APÊNDICE F - Gráficos de previsão para som (Teste 3)



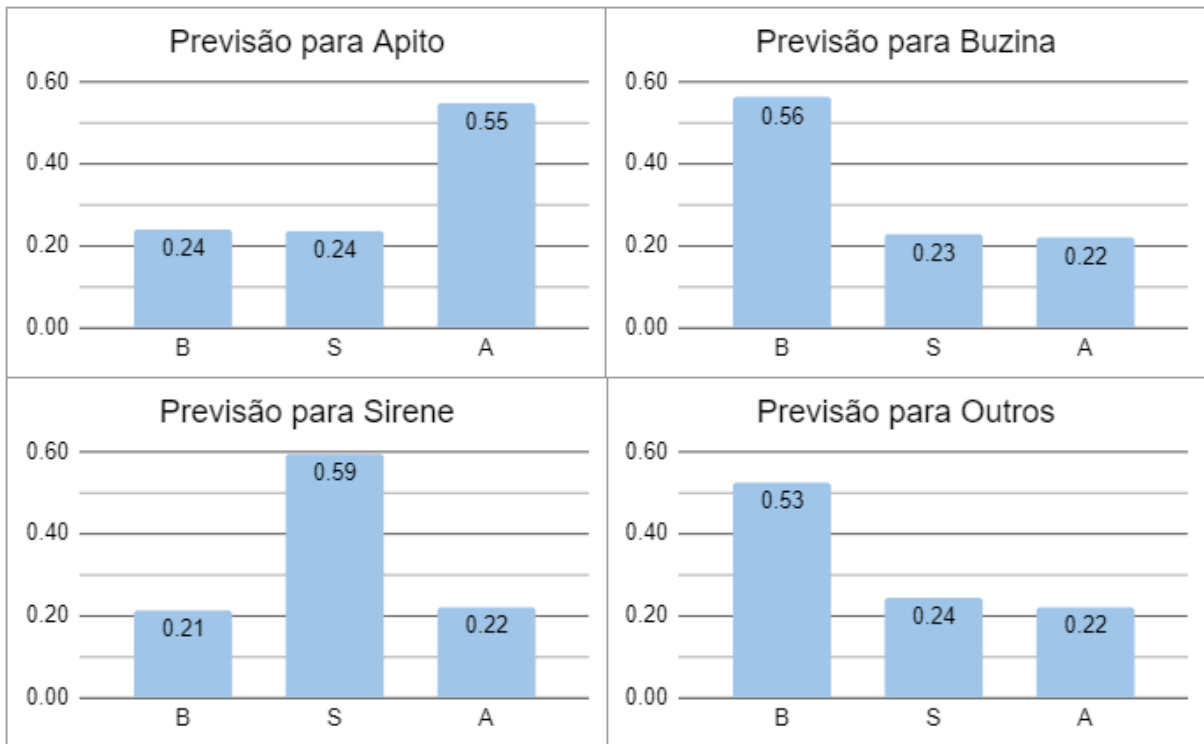
Fonte: Autoria própria (2022)

APÊNDICE G - Gráficos da Acurácia e Erro Teste 4



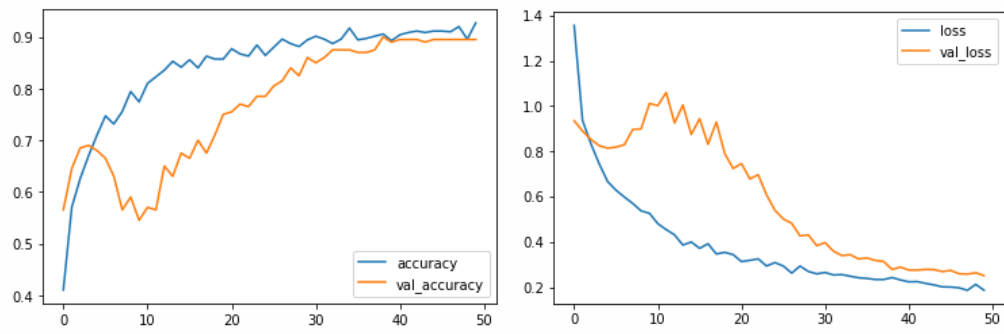
Fonte: Autoria própria (2022)

APÊNDICE H - Gráficos de previsão para som (Teste 4)



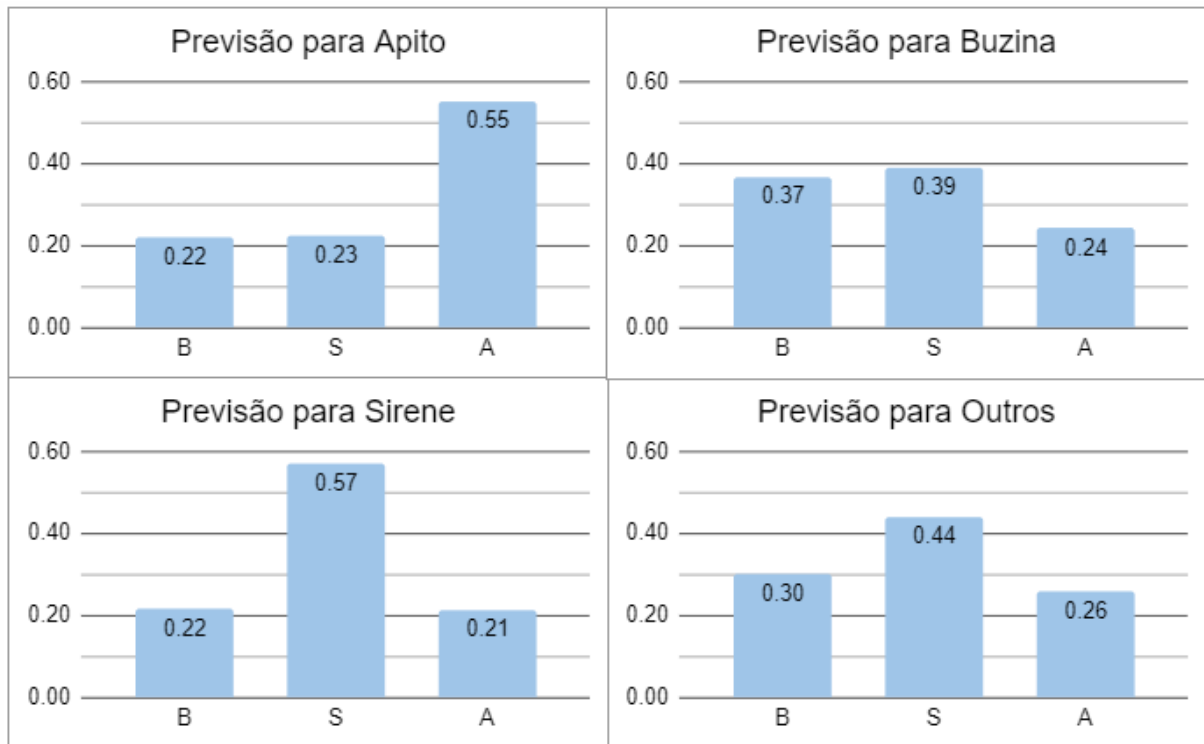
Fonte: Autoria própria (2022)

APÊNDICE I - Gráficos da Acurácia e Erro Teste 6



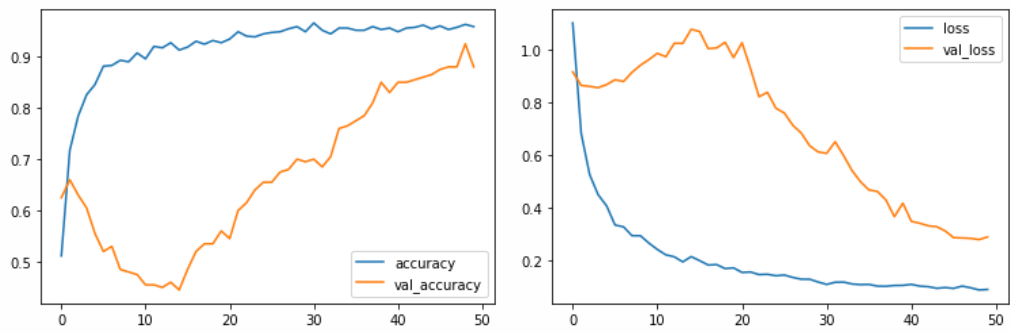
Fonte: Autoria própria (2022)

APÊNDICE J - Gráficos de previsão para som (Teste 6)



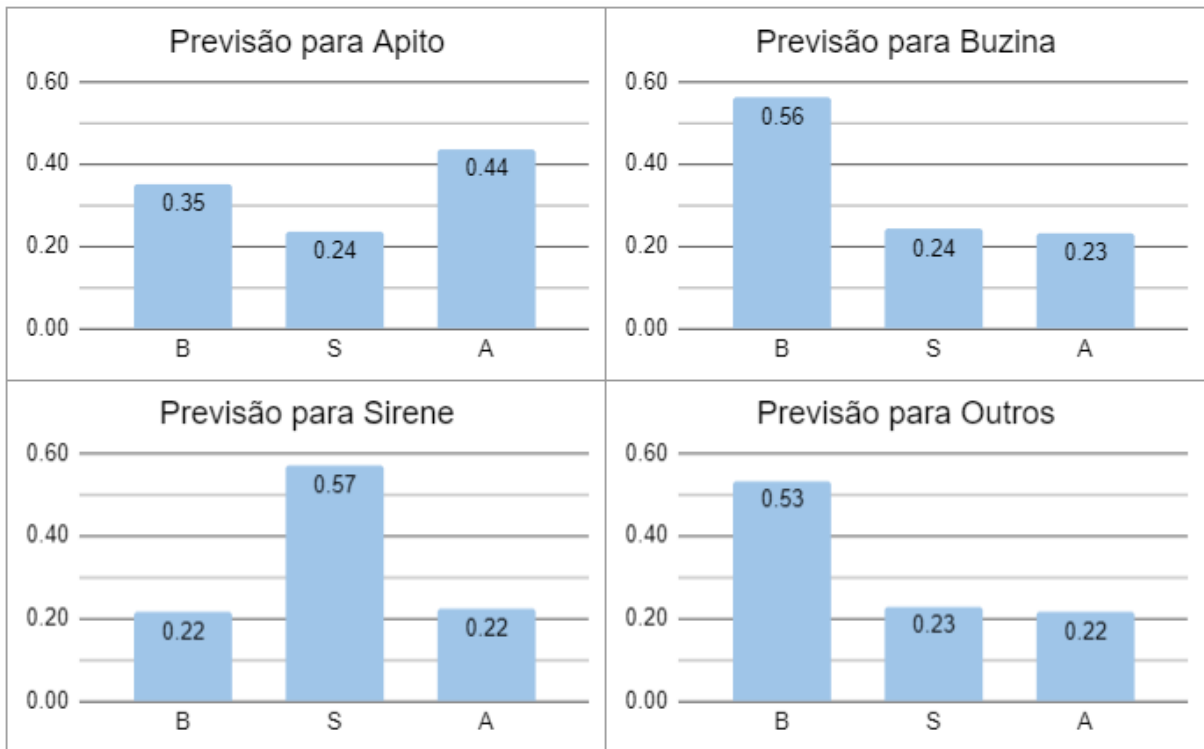
Fonte: Autoria própria (2022)

APÊNDICE K - Gráficos da Acurácia e Erro teste 7



Fonte: Autoria própria (2022)

APÊNDICE L - Gráficos de previsão para som (Teste 7)



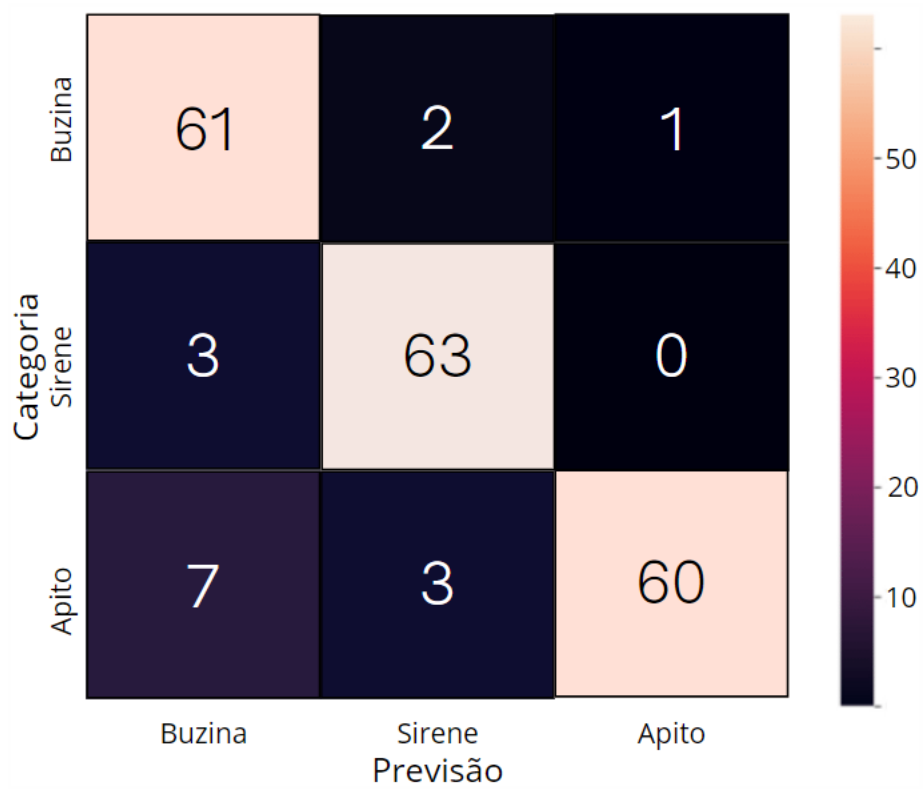
Fonte: Autoria própria (2022)

APÊNDICE M - Matriz de confusão Teste 1

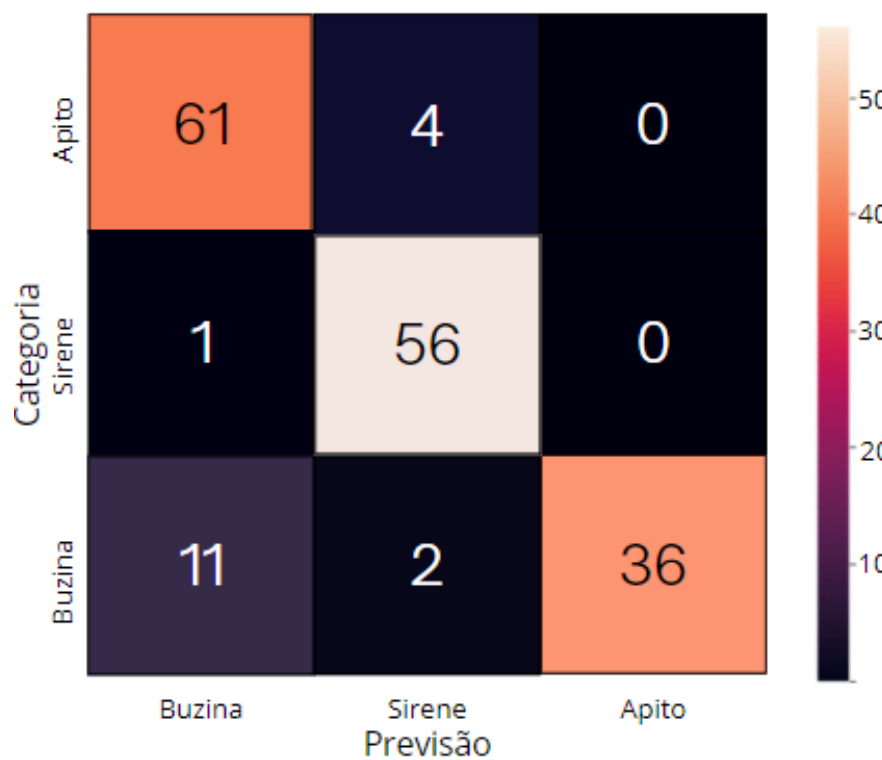


Fonte: Autoria própria (2022)

APÊNDICE N - Matriz de confusão Teste 2

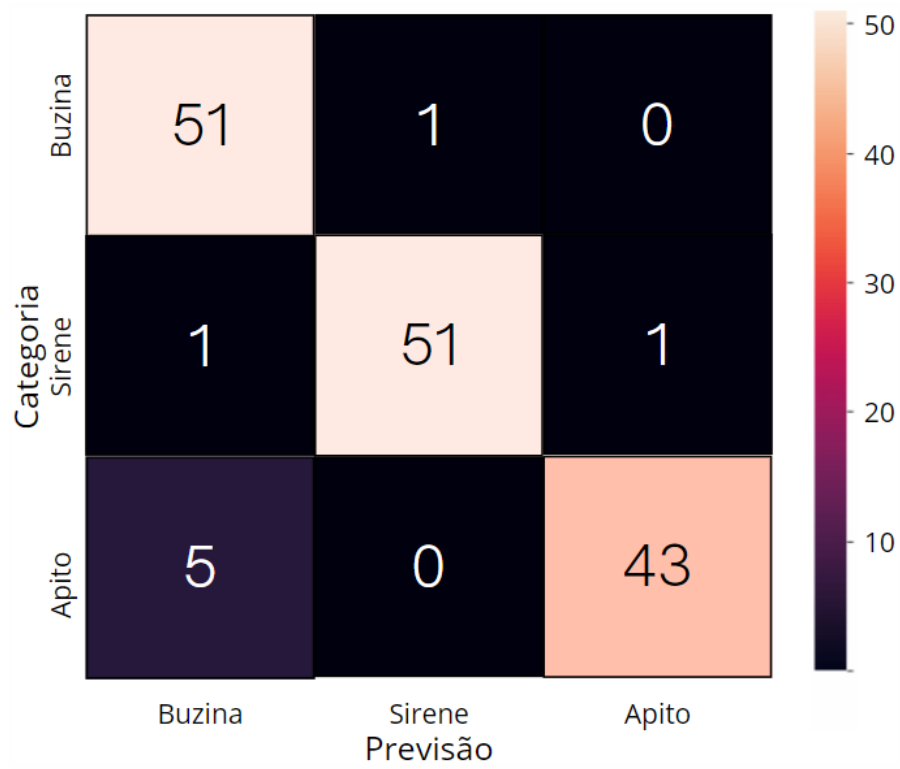


APÊNDICE O - Matriz de confusão Teste 3



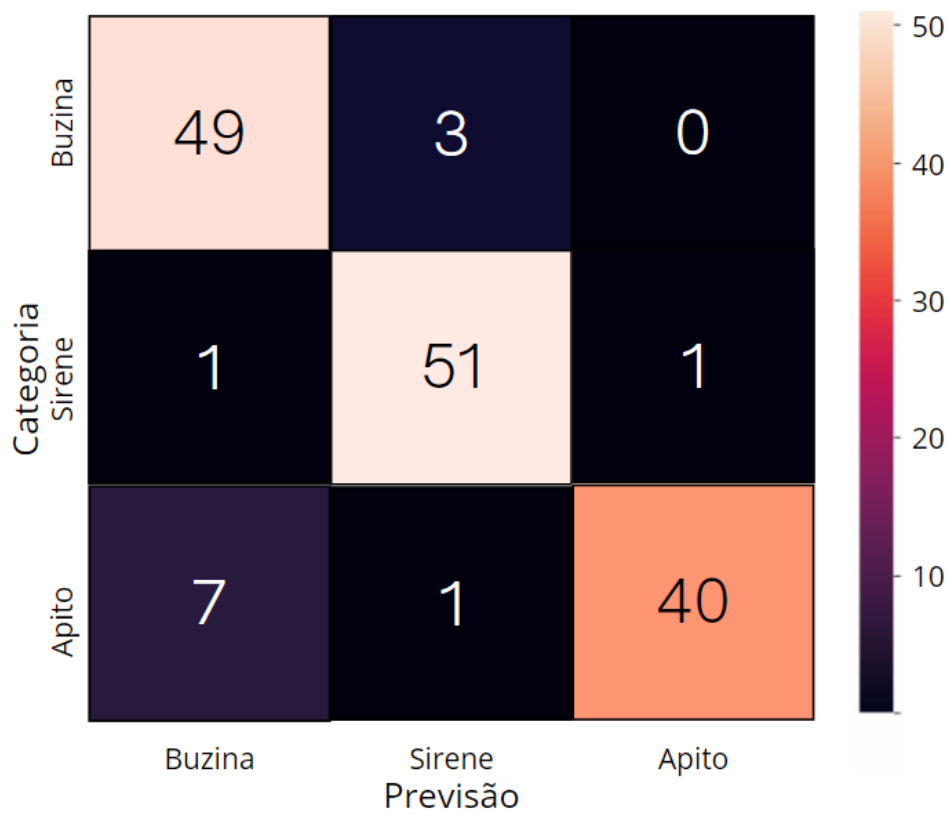
Fonte: Autoria própria (2022)

APÊNDICE P - Matriz de confusão Teste 4



Fonte: Autoria própria (2022)

APÊNDICE Q - Matriz de confusão teste 5



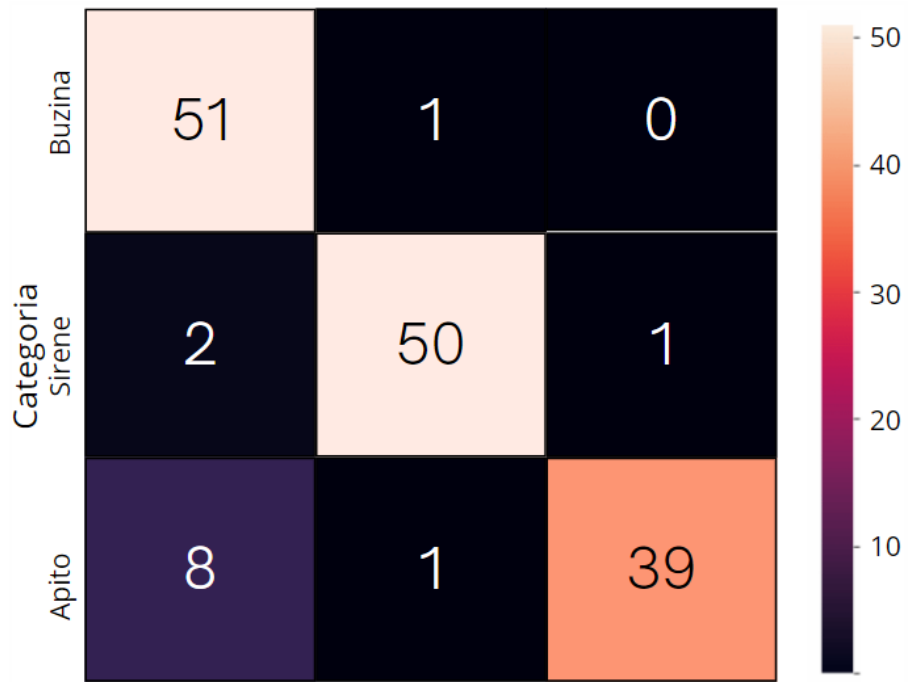
Fonte: Autoria própria (2022)

APÊNDICE R - Matriz de confusão teste 6



Fonte: Autoria própria (2022)

APÊNDICE S - Matriz de confusão teste 7



Fonte: Autoria própria (2022)