

UNIVERSIDADE FEDERAL DE SANTA CATARINA
CENTRO TECNOLÓGICO
DEPARTAMENTO DE ENGENHARIA ELÉTRICA E ELETRÔNICA
CURSO DE GRADUAÇÃO EM ENGENHARIA ELÉTRICA

Marina Galarça Santa Catarina

**Identificação de modelos de telefones celulares a partir de impressões digitais
em gravações de áudio**

Florianópolis

2024

Marina Galarça Santa Catarina

**Identificação de modelos de telefones celulares a partir de impressões digitais
em gravações de áudio**

Trabalho de Conclusão do Curso de Graduação em Engenharia Elétrica do Centro Tecnológico da Universidade Federal de Santa Catarina como requisito para a obtenção do título de Bacharel em Engenharia Elétrica.

Orientador: Prof. Márcio Holsbach Costa, Dr.

Florianópolis

2024

Ficha de identificação da obra

Santa Catarina, Marina Galarça
Identificação de modelos de telefones celulares a
partir de impressões digitais em gravações de áudio /
Marina Galarça Santa Catarina ; orientador, Márcio
Holsbach Costa, 2024.
48 p.

Trabalho de Conclusão de Curso (graduação) -
Universidade Federal de Santa Catarina, Centro
Tecnológico, Graduação em Engenharia Elétrica,
Florianópolis, 2024.

Inclui referências.

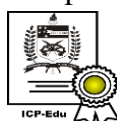
1. Engenharia Elétrica. 2. Perícia Forense. 3.
Identificação de celulares. 4. Aprendizado de máquina. I.
Costa, Márcio Holsbach. II. Universidade Federal de Santa
Catarina. Graduação em Engenharia Elétrica. III. Título.

Marina Galarça Santa Catarina

Identificação de modelos de telefones celulares a partir de impressões digitais em gravações de áudio

Este Trabalho Conclusão de Curso foi julgado adequado para obtenção do Título de “Bacharel em Engenharia Elétrica” e aceito, em sua forma final, pelo Curso de Graduação em Engenharia Elétrica.

Florianópolis, 14 de agosto de 2024.



Documento assinado digitalmente

Miguel Moreto

Data: 15/08/2024 11:42:43-0300

CPF: ***.850.100-**

Verifique as assinaturas em <https://v.ufsc.br>

Prof. Miguel Moreto, Dr.

Coordenador do Curso de Graduação em Engenharia Elétrica

Banca Examinadora:



Documento assinado digitalmente

Marcio Holsbach Costa

Data: 14/08/2024 10:58:46-0300

CPF: ***.170.680-**

Verifique as assinaturas em <https://v.ufsc.br>

Prof. Márcio Holsbach Costa, Dr.

Orientador

Universidade Federal de Santa Catarina



Documento assinado digitalmente

Eduardo Luiz Ortiz Batista

Data: 14/08/2024 11:20:44-0300

CPF: ***.521.889-**

Verifique as assinaturas em <https://v.ufsc.br>

Prof. Eduardo Luiz Ortiz Batista, Dr.

Universidade Federal de Santa Catarina



Documento assinado digitalmente

FERNANDO HENRIQUE LONZETTI

Data: 14/08/2024 12:09:56-0300

CPF: ***.839.359-**

Verifique as assinaturas em <https://v.ufsc.br>

Eng. Eletric. Fernando Henrique Lonzetti, M.Eng.

Polícia Científica de Santa Catarina

Dedico este trabalho a todos que tiveram a coragem de sonhar.
Aos que acreditam na educação e na ciência como direito
fundamental da vida.

AGRADECIMENTOS

Agradeço à minha família, por ser a rocha sólida em que me apoio nos momentos de dificuldade. Em especial, agradeço aos meus pais, Claudir e Márcia, por apoiarem os meus sonhos. Agradeço também aos meus irmãos e minha avó, por torcerem por mim. Agradeço ao meu tio, Sérgio, por dividir a profissão de engenheiro eletricista comigo e me aconselhar durante a graduação.

Aos meus amigos, por me arrancarem sorrisos e gargalhadas nos momentos mais difíceis. Aos colegas de curso, pelos longos anos de estudo, pela troca de conhecimentos e por todas as experiências vividas na graduação. E claro, aos amigos que a faculdade fez se tornarem família. Émillyn, Gabriel, Lucas, Maria Luiza, Marcello, Pedro, Rebecca e Vitória, agradeço por acreditarem em mim mais do que eu mesma já pude acreditar.

Ao professor Márcio Costa, que além de orientador desse trabalho, também foi um orientador para a vida. Aos demais professores do curso de graduação em Engenharia Elétrica, por toda sua competência e maestria nas aulas. Ao Departamento de Engenharia Elétrica e Eletrônica, pelo comprometimento com o ensino e a qualidade da formação de futuros engenheiros. À Universidade Federal de Santa Catarina, por ser uma instituição séria e acolhedora, comprometida com o ensino público e de qualidade.

Por fim, agradeço a todos que fizeram parte da minha trajetória ao longo da graduação e que contribuíram para minha formação pessoal, acadêmica e profissional. E agradeço a Deus, por guiar meu caminho.

RESUMO

A expansão do uso de semicondutores levou a uma era de digitalização da informação. Telefones fixos hoje são facilmente substituídos por *smartphones* que concentram uma grande quantidade de arquivos de mídia. Naturalmente, assim como em outros cenários do cotidiano, essa mudança trouxe novos desafios para os órgãos de justiça. Questões como a origem, veracidade e autoria de mídias como imagens, gravações de áudio e vídeos são algumas das preocupações presentes no campo da perícia forense. Este trabalho propõe o treinamento, avaliação e comparação de desempenho entre dois classificadores baseados em aprendizado de máquina para identificação de telefones celulares a partir de gravações de áudio.

Palavras-chave: Perícia Forense. Identificação de celulares. Aprendizado de máquina.

ABSTRACT

The expansion in the use of semiconductors has led to an era of digitalization of information. Landline telephones are now easily replaced by smartphones that hold a large number of media files. Naturally, as in other daily scenarios, this change has brought new challenges for law enforcement agencies. Issues such as the origin, veracity, and authorship of media such as images, audio recordings, and videos are some of the concerns present in the field of forensics. This work proposes training, assessing, and comparing the performance of two machine learning-based classifiers for identifying cell phones from audio recordings.

Keywords: Forensics. Cell phone identification. Machine learning.

LISTA DE FIGURAS

Figura 1 – Diagrama de blocos do processo de obtenção dos MFCCs.	23
Figura 2 – Hierarquia do aprendizado de máquina.....	24
Figura 3 – Ilustração da separação de dados por hiperplano.....	25
Figura 4 – Matriz de confusão de dimensão 2×2.....	29
Figura 5 – Curva ROC.....	31
Figura 6 – Correspondência entre modelo de telefone celular e índice.....	33
Figura 7 – Comparação entre forma de onda e envelope espectral em telefones Iphone5 e Desire C.....	36
Figura 8 – Envelope espectral de 4 modelos de telefone distintos da marca LG.....	37
Figura 9 – Matriz de confusão SVM.....	39
Figura 10 – Curva ROC e AUC do SVM.....	40
Figura 11 – Matriz de confusão LDA.....	41
Figura 12 – Curva ROC e AUC do LDA.....	42

LISTA DE TABELAS

Tabela 1 – Resultados do classificador SVM utilizando somente MFCCs.....	37
Tabela 2 – Resultados do classificador LDA utilizando somente MFCCs	38
Tabela 3 – Resultados do classificador LDA utilizando MFCCs e Δ -MFCCs	39

LISTA DE ABREVIATURAS E SIGLAS

GMM	Gaussian Mixture Model
GLDS	Generalized Linear Discriminant Sequence
GSV	Gaussian Super Vector
KNN	K-Nearest Neighbors
LDA	Linear Discriminant Analysis
LOO	Leave one out
MFCC	Mel Frequency Cepstral Coefficients
ROC	Receiver Operating Characteristic
SVM	Support Vector Machines

LISTA DE SÍMBOLOS

Δ	Delta (maiúsculo)
ξ	Xi (minúsculo)
δ	Delta (minúsculo)

SUMÁRIO

1	INTRODUÇÃO	15
1.2	OBJETIVOS	16
1.2.1	Objetivos específicos	16
2	FUNDAMENTAÇÃO TEÓRICA.....	17
2.1	PARÂMETROS DE IDENTIFICAÇÃO DE TELEFONES CELULARES	17
2.2	EXTRAÇÃO DE CARACTERÍSTICAS	20
2.3	APRENDIZADO DE MÁQUINA	22
2.3.1	Classificadores.....	23
2.3.1.1	<i>Máquinas de Vetor de Suporte (SVM)</i>	<i>24</i>
2.3.1.1.1	SVM Linear	24
2.3.1.1.2	SVM não linear.....	25
2.3.1.2	<i>Análise Discriminante Linear (LDA)</i>	<i>26</i>
2.4	MÉTODOS DE VALIDAÇÃO	27
2.4.1	<i>K-Fold</i>	27
2.4.2	<i>Leave one out</i>	27
2.5	MÉTODOS DE AVALIAÇÃO	27
2.5.1	Matriz de Confusão	28
2.5.2	Curva ROC.....	30
2.6	BANCO DE DADOS	31
3	METODOLOGIA.....	32
3.1	MATERIAIS.....	32
3.2	PRÉ-PROCESSAMENTO DE DADOS	32
3.3	EXTRAÇÃO DE CARACTERÍSTICAS	33
3.4	VALIDAÇÃO DO MODELO.....	34
3.5	TESTES ADICIONAIS.....	35

4	RESULTADOS E DISCUSSÃO.....	36
4.1	CLASSIFICAÇÃO COM VETOR DE CARACTERÍSTICAS DE MFCCs.....	37
4.2	CLASSIFICAÇÃO COM Δ -MFCCs ADICIONADOS AO VETOR.....	38
4.3	TESTES ADICIONAIS.....	42
4.4	DISCUSSÃO.....	43
5	CONCLUSÃO.....	45
	REFERÊNCIAS.....	46

1 INTRODUÇÃO

Em 1983 foi comercializado o primeiro modelo de telefone celular do mundo, o Dynatac 8000x, desenvolvido pela fabricante Motorola. Alguns anos depois os telefones celulares já se tornavam dispositivos indispensáveis no dia a dia da população. Com a popularização dos *smartphones* a partir dos anos 2000, muitas ferramentas como despertadores, agendas, câmeras e até mesmo computadores foram substituídas por aparelhos celulares.

Ainda que esses dispositivos contribuam positivamente no aspecto da praticidade de sua utilização, portabilidade e rapidez, muitas vezes podem ser utilizados para atividades ilegais ou criminosas. A transmissão de arquivos de mídia, como imagens e áudios se tornou uma das principais utilizações do uso de celulares no novo milênio. No entanto, a autenticidade e a origem desses arquivos de mídia passaram a ser algumas das preocupações no escopo da segurança da informação (TOCCHETTO; ESPINDULA, 2022).

Uma das razões pela qual a perícia de arquivos de mídia se tornou essencial é a facilidade em se fazer alterações nesses tipos de arquivos por meio de ferramentas simples de edição, disponíveis gratuitamente na rede mundial de computadores (web) (KOTROPOULOS, 2014). Por isso, garantir a verificação da origem e da autenticidade de arquivos multimídia é essencial para a validação desses arquivos como evidência em processos criminais (ZOU *et al.*, 2014).

A identificação de marca e modelo de telefones celulares contribui positivamente nesse processo de validação de evidências, uma vez que possibilita a determinação da origem dos arquivos apresentados. Em especial, a identificação do modelo do telefone celular pode trazer indicações sobre autoria de um arquivo de áudio no caso de múltiplas possibilidades de autores (PANDEY *et al.*, 2014).

O processo de identificação de dispositivos já é uma realidade desde os primórdios dos anos 2000 (BERDICH *et al.*, 2022). Isso é possível pela existência do que muitos autores denominam como “impressão digital”, ou seja, um padrão único presente no sinal, seja áudio, imagem ou vídeo, devido a características intrínsecas dos componentes do dispositivo. Em razão desse comportamento singular, por meio de métodos de extração de características, pode-se determinar padrões para a identificação de aparelhos de telefone celular.

Na literatura, diversos estudos sobre classificação de câmeras e microfones foram propostos nos últimos 20 anos. No entanto, pesquisas envolvendo a perícia de imagens estão

muito mais avançadas em relação a pesquisas com arquivos de áudio (KOTROPOULOS, 2014). Algumas áreas que merecem uma maior atenção são autenticação de áudio, autenticação do ambiente de gravação, autenticação do dispositivo fonte da gravação e autenticação da data da gravação (JIANG; LEUNG, 2016).

No ano de 2007, Kraetzer e Dittmann apresentaram um procedimento para a identificação de microfones baseado na extração de características por meio de coeficientes cepstrais de frequência mel (MFCCs) e uso de um classificador do tipo Naive Bayes, alcançando uma acurácia entre 60% e 75% (KOTROPOULOS, 2014). Apesar dos resultados tímidos, comprovou-se a possibilidade da identificação do modelo de microfone por meio de gravações de áudio (JIANG; LEUNG, 2019). De forma semelhante, este resultado pode ser estendido para a identificação de aparelhos celulares, uma vez que diferentes modelos utilizam diferentes tipos de microfones. Adicionalmente, verificou-se a possibilidade do emprego de classificadores simples baseados em técnicas de aprendizado de máquina. Atualmente, o aprendizado de máquina é uma ferramenta amplamente difundida nas mais diversas áreas como medicina, economia e processamento de sinais (BARUAH *et al.*, 2020).

Contribuindo para o crescimento da área de pesquisa de identificação de aparelhos celulares, este trabalho tem como objetivo a identificação de marca e modelo do aparelho celular produtor de uma gravação de áudio, utilizando métodos de aprendizado de máquina.

1.2 OBJETIVOS

O presente trabalho tem como objetivo a implementação, avaliação e comparação de métodos de identificação de marca e modelo de aparelhos celulares, por meio de gravações de áudio.

1.2.1 Objetivos específicos

- Estudar métodos de classificação segundo técnicas de aprendizado de máquina;
- Implementar métodos para classificação de marcas e modelos de telefones celulares a partir de gravações de áudio;
- Avaliar o desempenho dos métodos implementados.

2 FUNDAMENTAÇÃO TEÓRICA

Ao longo deste capítulo será descrita a fundamentação teórica necessária para a compreensão do tema.

2.1 PARÂMETROS DE IDENTIFICAÇÃO DE TELEFONES CELULARES

O estudo realizado neste trabalho se baseia na premissa de que diferentes telefones celulares possuem uma “assinatura espectral” específica. Tais assinaturas também são chamadas de impressão digital (JIN; WANG; YAN, 2018). Ou seja, possuem comportamento único em determinadas faixas de frequência do espectro. Considerando esse comportamento singular e utilizando métodos de extração de características, pode-se determinar padrões que permitam a identificação dos aparelhos celulares.

A literatura tem demonstrado a possibilidade de identificação do aparelho fonte de um arquivo de áudio, em decorrência das diferenças e não idealidades na implementação dos circuitos eletrônicos que o compõe (HANILÇI *et al.*, 2012). Mesmo considerando circuitos idênticos, a utilização de componentes de diferentes modelos e fabricantes impacta diretamente nas características da função de transferência do sistema (HANILÇI *et al.*, 2012).

Uma das principais diferenças entre marcas e modelos de aparelhos celulares reside nos tipos de microfones e sistemas de codificação-decodificação (CODEC) utilizados para sua confecção (LIN *et al.*, 2020). O CODEC é o componente eletrônico responsável pela amostragem e compressão do sinal de áudio. O processo de compressão segue um conjunto específico de regras que pode variar entre diferentes fabricantes (JIN *et al.*, 2018).

Embora possamos identificar impressões digitais tanto a partir de características do hardware quanto do software, as características dos microfones são mais difíceis de serem forjadas (BERDICH *et al.*, 2023). A identificação da fonte irá depender de um método que consiga, efetivamente, extrair as características desejadas uma vez que o sinal de saída pode conter perturbações provenientes de variabilidades linguísticas e ruídos (GARCIA-ROMERO, 2010). A extração de características pode ser realizada tanto no domínio do tempo quanto no domínio da frequência.

No ano de 2012, Hanilçi *et al.* propôs utilizar os princípios da identificação de microfones na identificação de telefones móveis, o que marcou um grande avanço para a pesquisa forense da década. No trabalho de HANILÇI *et al.* (2012), 14 telefones celulares

distintos foram classificados a partir de MFCCs utilizando uma Máquina de Vetores de Suporte (SVM) e a técnica de quantização vetorial (do inglês, *Vector Quantization*) obtendo respectivamente 96,42% e 92,56% de acurácia. Em 2013, Hanilçi e Ertas incluíram um processo de normalização da variância cepstral e da primeira derivada aos MFCCs, com isso aumentando as taxas de identificação para 98%. Em 2014, Hanilçi e Kinnunen propuseram utilizar os segmentos sem som das gravações para a obtenção das impressões digitais dos aparelhos, baseados na premissa de que a acurácia aumentaria devido à ausência da fala. A acurácia obtida foi de 98,39% para o classificador SVM utilizando MFCCs, o que significa um aumento de 1,97% comparado ao método proposto em 2012.

De forma independentemente, em 2014, Pandey *et al.* também propuseram o uso de segmentos sem fala na identificação de celulares, escolhendo como característica alvo a densidade espectral de potência (PSD, do inglês *Power Spectral Density*) do sinal. Utilizando o classificador SVM, os autores obtiveram uma acurácia média de 88%. Ainda em 2014, Aggarwal *et al.* propôs que ao eliminar a voz do narrador a função transferência do aparelho poderia ser estimada de maneira mais eficaz. A acurácia obtida foi de 90% para o classificador SVM.

Embora a escolha de uso de segmentos sem fala tenha apresentado um aumento nos percentuais de acurácia, o processo de remoção implica em maior tempo de processamento e necessidade de recursos computacionais adicionais (BARUAH, 2020). No caso de HANILÇI e KINNUNEN (2014) a seleção dos trechos de análise foi baseada na estimação e comparação dos valores de energia de cada segmento. No entanto, para uma detecção mais precisa essa etapa deveria ser realizada por um analista forense de forma manual (HANILÇI; KINNUNEN, 2014), o que pode ser apontado como uma desvantagem do método. Entretanto, segmentos de baixa energia podem se referir tanto a períodos de silêncio quanto a períodos de ruído, podendo impactar no treinamento do classificador (JIANG e LEUNG, 2019).

KOTROPOULOS (2014) propôs realizar o mapeamento de vetores de características em espaços de dimensões menores. Foram utilizados classificadores do tipo *Nearest Neighbors*, *Sparse Representation* e SVM que resultaram em acurácias superiores a 94%, chegando a atingir 100% em diversos casos.

ZOU *et al.* (2014) propuseram melhorias na extração de características combinando a energia do sinal com os MFCCs. A combinação resultou em um aumento da acurácia de classificação de cerca de 1%. O classificador Modelo de Mistura Gaussiano (GMM) obteve

uma acurácia máxima de 97,71%. Outro resultado importante obtido nesse trabalho foi que ao realizar uma tentativa de extração do ruído a acurácia diminuiu drasticamente.

Em JIANG e LEUNG (2016) foi proposto o uso de MFCCs e coeficientes cepstrais de frequência linear (LFCCs, do inglês *Linear Frequency Cepstral Coefficients*) no problema de identificação de celulares. O classificador utilizado foi um SVM ponderado. A acurácia obtida com o uso de MFCCs e LFCCs foi de 97,86% e 98,33%, respectivamente. Em trabalho posterior, no ano de 2019, os mesmos autores sugerem um método de projeção baseada em kernels com o intuito de obter melhores resultados de classificação, usando *Gaussian Super Vectors* (GSVs) para a obtenção das características. Ao comparar SVMs e Representação Esparsa, os autores constataram não haver diferenças significativas quanto ao método de classificação.

De acordo com JIN *et al.* (2018) embora os métodos de classificação que utilizam MFCCs apresentem elevada acurácia na identificação de marcas e modelos de *smartphones*, os resultados podem ser sensíveis a variações fonéticas e textuais no conteúdo dos áudios. Com o uso de classificadores de Limite de Variância e SVM, os autores obtiveram acurácias de 97,89% e 98,04%, respectivamente, na identificação de telefones celulares utilizando características extraídas do CODEC (*software*), superando os resultados obtidos por meio dos MFCCs. Contudo, essa metodologia é válida apenas para arquivos nos formatos MP3, AAC e M4A, que embora sejam os formatos de compressão predominantemente utilizados, não representam a totalidade.

Diferentemente dos demais autores que utilizavam características cepstrais, LUO *et al.* (2018) propuseram utilizar características baseadas na diferença de energia entre as bandas do sinal. O classificador escolhido também foi o SVM e os resultados de acurácia superaram 96%.

Embora os classificadores clássicos, como o SVM, tenham apresentado até então acurácias elevadas na classificação, outros autores têm sugerido a utilização de Redes Neurais Convolucionais (CNN, do inglês *Convolutional Neural Network*) como alternativa na resolução do problema. VERMA e KHANNA (2019), QIN *et al.* (2018) e LIN (2020) abordaram o problema com o uso de CNNs. Li *et al.* (2017) optaram pelo uso de Supervetores Gaussianos Profundos e em um trabalho posterior no mesmo ano fizeram uso de uma Rede de Auto Codificação Profunda.

Em 2022, Berdich *et al.* realizaram uma comparação entre alguns classificadores clássicos e CNNs. Os resultados demonstram uma diferença irrisória entre o uso de um

classificador do tipo CNN e um baseado em Análise Discriminante Linear (LDA, do inglês *Linear Discriminant Analysis*). Esse resultado indica que embora as técnicas de aprendizado profundo sejam bem-vindas na exploração de alternativas para resolução do problema, elas não têm apresentado grandes vantagens em relação aos métodos clássicos. Uma possibilidade para justificar esse resultado seria o fato de que os bancos de dados disponíveis possuem uma quantidade de sinais bastante limitada. Além disso, as CNNs apresentam um custo computacional elevado enquanto o LDA é mais simples e requer menos memória (BERDICH *et al.*, 2022).

Mesmo com os avanços apresentados na descrição anterior, ainda são poucos os trabalhos publicados na literatura em referência à identificação de modelos e marcas de aparelhos celulares por meio de gravações de áudio. Os resultados obtidos em estudos já publicados anteriormente demonstram a possibilidade de se obter uma elevada acurácia na classificação dos aparelhos apoiado em técnicas tradicionais de classificação.

Em vista do apresentado, o problema analisado não é apenas uma linha de pesquisa científica, mas já se encontra em estágio de aplicação tecnológica. Dessa forma, desenvolvemos este trabalho de forma a adquirir conhecimento e experiência na implementação de sistemas de identificação de marcas e modelos de aparelhos celulares por meio de gravações de áudio. Para tanto, serão empregadas características baseadas em MFCCs e classificadores convencionais do tipo SVM e LDA.

2.2 EXTRAÇÃO DE CARACTERÍSTICAS

Um sinal de áudio gravado pode ser descrito como resultado da convolução entre o sinal de fala humana e as respostas ao impulso do ambiente de propagação e do dispositivo fonte da gravação (JIANG e LEUNG, 2016). Assumindo que as alterações devidas ao ambiente possam ser desconsideradas, existem diversos métodos para extração de características da resposta ao impulso do sistema de gravação a partir de sinais de áudio. Esses métodos se dividem em 5 grupos: características espectrais de curto-prazo, características da fonte de voz, características espectro-temporais, características prosódicas e características de alto nível (KINNUNEN e LI, 2010). Os métodos mais explorados em processamento de áudio são os que envolvem características espectrais do sinal (KOTROPOULUS, 2014), tanto características espectrais de curto-prazo como características espectro-temporais.

Ao analisar trabalhos previamente publicados na literatura da área pode-se notar a predominância de métodos de extração de características de curto-prazo. O método mais comum em processamento de fala são os MFCCs (HANILÇI *et al.*, 2012). Entretanto, revisando a bibliografia da área observamos que LFCCS, coeficientes cepstrais de previsão linear (LPCCs, do inglês *Linear Predict Cepstral Coefficients*), coeficientes de predição perceptual linear (PLPCs, do inglês *Perceptual Linear Predict Coefficients*) e coeficientes cepstrais normalizados de potência (PNCCs, do inglês *Power Normalized Cepstral Coefficients*) também sendo empregados.

No campo da perícia forense, em particular no processo de identificação de aparelhos celulares e no reconhecimento de voz, a utilização de MFCCs como método de extração de características é muito comum (HANILÇI *et al.*, 2012). Isso por que há uma similaridade entre o sistema humano de voz e as bandas de frequência Mel distribuídas no MFCC (HADOLTIKAR; RATNAPARKHE; KUMAR, 2019).

Proposta por Stevens, Volkman e Newmann em 1937, Mel é uma escala de medida que tem como objetivo quantificar o tom de uma voz. Para cada tom, um valor de frequência em Hertz é associado, conforme descreve a equação (1):

$$Mel(f) = 2595 \cdot \log \left(1 + \frac{f}{700} \right) \quad (1)$$

Que também pode ser escrita como:

$$Mel(f) = 1127 \cdot \ln \left(1 + \frac{f}{700} \right) \quad (2)$$

Os MFCCs, desenvolvidos por Davis e Mermestein em 1980 (DEUS, 2021), são os coeficientes que compõem a forma cepstral do gráfico de potência do sinal de áudio. O cepstro é uma forma gráfica que permite a identificação de picos de amplitude uniformemente espaçados ao longo do espectro de frequência ao aplicar uma transformação logarítmica no espectro de um sinal. O processo de obtenção destes coeficientes é realizado em 6 etapas, ilustradas no fluxograma apresentado na Figura 1. O sinal acústico original é primeiramente amostrado e então janelado em trechos menores, de 15 a 30ms. Utilizando a Transformada Rápida de Fourier é calculado o espectro de potência de cada segmento janelado, que é então ponderado por um banco de filtros psicoacústico (KINNUNEN e LI, 2010), ou seja, uma série

de filtros passa-banda (HANILÇI e ERTAS, 2013) na escala de frequência Mel, que tem por objetivo enfatizar determinadas frequências do espectro (BIMBOT *et al.*, 2004). Após este processo, é calculado o logaritmo dos sinais na saída do banco de filtros e aplicada a Transformada Discreta dos Cossenos, conforme a equação (3).

$$c_n = \sum_{m=1}^M \log Y(m) \cdot \cos\left[\frac{\pi n}{M} \left(m - \frac{1}{2}\right)\right] \quad (3)$$

Na sequência, os resultados são transformados em um vetor de características conhecido como vetor de MFCCs.

Figura 1 – Diagrama de blocos do processo de obtenção dos MFCCs.



Fonte: Autor.

O vetor de MFCCs é composto pela seleção dos menores coeficientes (c_n) obtidos pela Transformada Discreta dos Cossenos. Embora o tamanho do vetor possa ser definido arbitrariamente, geralmente um vetor com tamanho entre 15 e 20 coeficientes é utilizado para classificação.

2.3 APRENDIZADO DE MÁQUINA

O aprendizado de máquina, de maneira geral consiste em um conjunto de técnicas que mimetizam o processo de aprendizado humano, utilizando conceitos estatísticos. O processo todo se divide em quatro etapas: armazenamento, abstração, generalização e avaliação (LANTZ, 2015). A qualidade dos dados armazenados e a escolha do algoritmo de generalização são as partes mais importantes do processo e impactam diretamente na eficácia do modelo.

O aprendizado de máquina pode ser subdividido em duas categorias: aprendizado supervisionado e aprendizado não-supervisionado. No aprendizado supervisionado, utilizam-se rótulos para o treinamento de algoritmos que têm como objetivo prever valores de saída a

partir de um vetor de atributos, cujos rótulos do vetor de entrada são conhecidos (SOUZA, 2022). Por outro lado, no aprendizado não-supervisionado, os rótulos são desconhecidos e os dados de entrada são agrupados de acordo com métricas pré-definidas.

2.3.1 Classificadores

Os métodos de classificação fazem parte do grupo de técnicas de aprendizado supervisionado, conforme é ilustrado na Figura 2. Como no problema a ser discutido neste trabalho os rótulos dos modelos de *smartphones* são conhecidos *a priori*, utilizam-se métodos de classificação para esta tarefa. Métodos clássicos de classificação, como SVMs, LDA e KNNs são técnicas comumente aplicadas na identificação dos dispositivos fonte.

Figura 2 – Hierarquia do aprendizado de máquinas.



Fonte: (SANTANA, 2019).

A avaliação do desempenho de diferentes classificadores no reconhecimento de microfones foi abordada por JIANG e LEUNG (2019). Em seu estudo, os autores apontam que o classificador Naive Bayes, utilizado previamente no reconhecimento de microfones a partir de uma gravação de áudio, embora tenha apresentado uma pequena acurácia de classificação, evidenciou a possibilidade de identificação do tipo e marca de microfones a partir de uma gravação de áudio. No mesmo estudo, também é citada a predominância da escolha das SVMs como classificador em problemas de identificação de microfones, o que foi confirmado também por VERMA e KHANNA (2021) no problema de identificação de celulares.

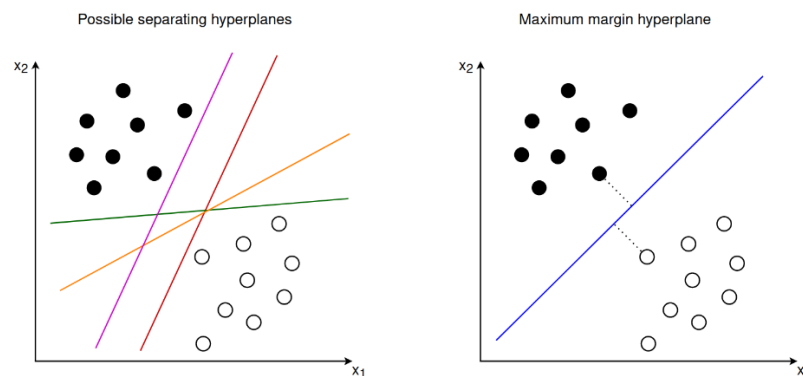
Em 2022, Berdich *et al.* mencionam SVMs, KNNs e CNNs como os classificadores mais empregados na tarefa de identificação de celulares. No entanto, Berdich *et al.* destacam

em seu trabalho a superioridade dos resultados obtidos com a escolha de LDA como classificador em comparação às SVMs.

2.3.1.1 Máquinas de Vetor de Suporte (SVM)

As SVMs são comumente empregadas em tarefas de classificação e regressão linear. A ideia central da utilização do classificador é delimitar um hiperplano que divide duas classes de dados distintas e linearmente separáveis, conforme ilustrado na Figura 3. Para isso, o método combina conceitos dos métodos de *Nearest Neighbors* e Regressão Linear (LANTZ,2015).

Figura 3 – Ilustração da separação de dados por hiperplano



Fonte: <https://www.inovex.de/de/blog/support-vector-machines-guide/>

Originalmente, as SVMs foram desenvolvidas para serem empregadas em problemas de classificação binária, no entanto, é possível realizar adaptações que permitem o uso do modelo em problemas de múltiplas classes. Uma solução para o caso de uma classificação de múltiplas classes é decompor o problema em subproblemas binários ou reformular o método para uma versão que atenda múltiplas classes, porém a segunda opção se torna menos interessante devido ao custo computacional elevado que apresenta (FACELI *et. al*, 2021).

2.3.1.1.1 SVM Linear

Em casos onde os conjuntos de dados são linearmente separáveis, define-se um hiperplano que separe os dados e então possibilite a classificação binária deles. O hiperplano é descrito pela equação (3):

$$h(x) = w \cdot x + b . \quad (3)$$

Entretanto, é possível que se obtenha diversos hiperplanos para o mesmo problema. Nesse caso, deve-se buscar o hiperplano que apresente a maior margem de segurança na divisão dos dados. Para encontrarmos o hiperplano ideal, minimizamos $\|w\|$ conforme a equação (4):

$$\text{Minimizar } J(w, b) = \frac{1}{2} \|w\|^2 \quad (4)$$

$$\text{Restrito por: } y_i(w \cdot x + b) - 1 \geq 0, \forall i = 1, \dots, n.$$

Essa restrição garante a inexistência de dados de treinamento entre as margens de separação do hiperplano. Por essa razão, é também conhecido como SVM de margens rígidas.

Embora na maioria das aplicações seja possível aproximar o problema por um modelo linear, nos casos reais existem pequenas divergências que implicam na não-linearidade do problema (FACELI *et al.*, 2021). Para contornar essa situação, são adicionadas variáveis de folga ξ para suavizar as margens, ou seja, permitir que existam dados de treinamento entre dois hiperplanos.

2.3.1.1.2 SVM não linear

Em conjuntos que não são linearmente separáveis ainda é possível realizar a classificação através de hiperplano linear, desde que seja feita uma transformação não linear no espaço de características do conjunto de dados para um espaço de características de dimensão maior e linearmente separável. Porém, como o espaço transformado pode ter dimensão infinita, as operações da função de mapeamento podem apresentar um custo computacional elevado, que não se justifica para a aplicação, já que o processo de separação depende somente do produto escalar entre os objetos no espaço de características, conforme discutido por FACELI *et al.* (2011). Assim, utilizam-se funções Kernel para definir o novo espaço de características.

2.3.1.2 Análise Discriminante Linear (LDA)

A LDA é uma técnica proposta por Ronald Aylmer Fischer em 1936, baseada na distribuição gaussiana (BERDICH *et al.*, 2013), que tem como objetivo a classificação de um conjunto de dados em múltiplas classes através da redução da dimensão do espaço descritivo (IBM, 2023). O método consiste na separação de classes através de hiperplanos, sendo similar ao SVM, e é bastante utilizado em problemas de classificação onde existem mais de duas classes.

A ideia central da LDA é juntar o Teorema de Bayes a outras regras probabilísticas com o intuito de prever se um dado de entrada pertence a um determinado grupo. O Teorema de Bayes e a Regra da Máxima Verossimilhança são combinados de forma que se obtém (LAMFO-UNB, 2020):

$$j = \arg \max \pi_i f_i(x) \quad (5)$$

onde j representa a classe, π representa a probabilidade de ocorrência da classe e f representa a função de verossimilhança de x dada a classe i .

Assumindo uma distribuição gaussiana, podemos classificar x na classe j conforme a equação (6):

$$\delta_i(x) = \log f_i(x) + \log \pi_i \quad (6)$$

onde δ representa a fronteira do hiperplano de separação.

Considerando que a covariância entre as classes seja idêntica, a equação (6) pode ser reescrita como:

$$\delta_i(x) = x^T \Sigma^{-1} \mu_i - \frac{1}{2} \cdot (\mu_i^T \Sigma^{-1} \mu_i) + \log \pi_i . \quad (7)$$

E assim obtemos a equação de separação das classes.

Além da aplicação em problemas de múltiplas classes, a LDA é uma excelente opção de classificador linear para problemas em que o tempo de processamento é crucial e deseja-se implementar um programa de baixo custo computacional (SANTANA, 2019).

2.4 MÉTODOS DE VALIDAÇÃO

O processo de aprendizado de métodos de classificação requer um conjunto de dados denominado de treinamento. Para verificar o desempenho do processo de aprendizado utiliza-se um conjunto de dados denominado de validação. O método de validação utilizado neste trabalho é conhecido como validação cruzada e suas principais variações são o *K-fold* e seu caso particular *Leave-one-out*.

2.4.1 K-Fold

Na técnica de *K-fold*, o conjunto de dados é dividido em K subconjuntos, selecionam-se $K-1$ amostras para realizar a etapa de treinamento e o subconjunto restante é utilizado para validação. Após a primeira validação, os subconjuntos são rotacionados e o processo se repete. Esse processo ocorre K vezes, até que todos os subconjuntos tenham sido utilizados tanto na etapa de treinamento quanto na de validação. O desempenho do classificador é avaliado através da média aritmética do desempenho obtido nos K subconjuntos de validação.

2.4.2 Leave one out

A técnica de *Leave-one-out* (LOO) é um caso particular de *K-fold*, onde o número de subconjuntos é equivalente ao número total de dados. Dessa forma, aplica-se o mesmo procedimento do *K-fold* e o desempenho é obtido pela média dos desempenhos individuais de cada validação.

2.5 MÉTODOS DE AVALIAÇÃO

O processo de aprendizado de máquina, conforme definido por Lantz (2015) é composto por quatro etapas, sendo a última delas a avaliação. Para avaliação de métodos de classificação binária, FACELI *et al.* (2011) sugere a utilização de curvas de Característica de Operação do Receptor (ROC, do inglês *Receiver Operating Characteristic*) e da matriz de confusão que são abordados a seguir.

2.5.1 Matriz de Confusão

Como parte das medidas de desempenho de classificadores binários, a matriz de confusão ilustra a concentração de resultados corretamente e incorretamente classificados. Entretanto, a matriz pode ser também adaptada para problemas superiores a duas classes.

No caso de um problema de classificação binária, a matriz de confusão gera uma matriz de dimensão 2×2 onde as linhas representam as classes e as colunas representam as classificações obtidas. A Figura 4 ilustra uma matriz de dimensão 2×2 , onde as siglas VP, FN, FP e VF correspondem a resultados verdadeiramente positivos, falsamente negativos, falsamente positivos e verdadeiramente falsos, respectivamente.

Figura 4: Matriz de Confusão de dimensão 2×2 .

	VERDADEIRO	FALSO
VERDADEIRO	VP	FN
FALSO	FP	VF

Fonte: autor.

A partir dos valores obtidos de VP, FN, FP e VF é possível determinar as seguintes medidas de desempenho:

- Taxa de erro na classe positiva: diz respeito à proporção de dados classificados erroneamente como positivos.

$$err_+(f) = \frac{FN}{VP + FN} \quad (8)$$

- Taxa de erro na classe negativa: diz respeito à proporção de dados classificados erroneamente como negativos.

$$err_-(f) = \frac{FP}{FP + VN} \quad (9)$$

- Taxa de erro total: representa o total de dados classificados incorretamente.

$$err(f) = \frac{FP + FN}{FP + FN + VP + VN} \quad (10)$$

- Acurácia: representa o total de dados classificados corretamente.

$$ac(f) = \frac{VP + VN}{FP + FN + VP + VN} \quad (11)$$

- Precisão: proporção de verdadeiros positivos entre todos classificados como positivos.

$$prec(f) = \frac{VP}{VP + FP} \quad (12)$$

- Sensibilidade: taxa de acerto na classe positiva.

$$sens(f) = \frac{VP}{VP + FN} \quad (13)$$

- Especificidade: taxa de acerto na classe negativa.

$$esp(f) = \frac{VN}{VN + FP} \quad (14)$$

- Medida F: a fim de evitar conclusões equivocadas avaliando isoladamente as medidas de precisão e sensibilidade, é possível analisá-las simultaneamente através da medida F.

$$F_k(f) = \frac{(k + 1)sens(f)prec(f)}{sens(f) + kprec(f)} \quad (15)$$

Com o uso da constante k é possível aumentar o peso da sensibilidade ou da precisão. Quando $k=1$, a sensibilidade e a precisão assumem o mesmo peso, e assim, obtém-se a medida F_1 .

$$F_1(f) = \frac{2sens(f)prec(f)}{sens(f) + kprec(f)} \quad (16)$$

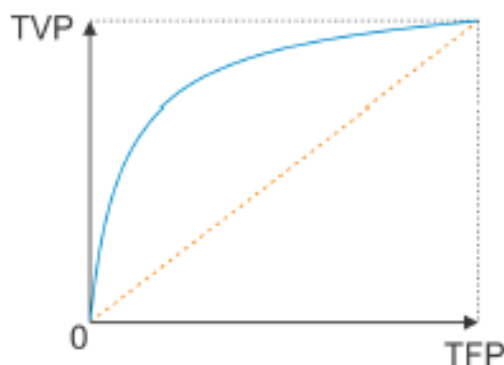
- Coeficiente de correlação de Matthews: indica a qualidade do classificador binário. A equação resulta em um valor entre -1 e 1 , onde 1 representa um classificador perfeito, 0 indica que o classificador é equivalente a um classificador aleatório e -1 indica um classificador ruim (GUBERT, 2019).

$$MCC(f) = \frac{VP \times VN - FP \times FN}{\sqrt{(VP + FP)(VP + FN)(VN + FP)(VN + FN)}} \quad (17)$$

2.5.2 Curva ROC

A curva ROC, obtida através de um gráfico bidimensional que correlaciona as medidas de precisão e sensibilidade, é mais uma forma de avaliação do desempenho de classificadores binários (FACELI *et al.*, 2011). Conforme ilustra a Figura 5, a abscissa representa a medida de precisão, também chamada de taxa de falsos positivos (TFP), enquanto a ordenada representa a sensibilidade, chamada de taxa de verdadeiros positivos (TVP). Idealmente, a melhor curva ROC é a que mais se aproxima do ponto $(1,0)$ no gráfico, pois indica o maior número de classificações corretas dos dados.

Figura 5 – Curva ROC



Fonte: (SOUZA, 2022).

Ao analisarmos múltiplas curvas ROC é possível utilizar uma outra medida de comparação, a medida da área sob a curva, do inglês *Area Under Curve* (AUC). Esse valor varia entre 0 e 1, sendo 1 o valor atribuído para um classificador perfeito e 0,5 para um classificador de desempenho semelhante ao classificador aleatório, quaisquer valores abaixo de 0,5 indicam um classificador de baixo desempenho.

2.6 BANCO DE DADOS

Na realização deste trabalho foi utilizado o banco de dados MOBIPHONE, desenvolvido e apresentado no trabalho de KOTROPOULOS (2014). A escolha pela base MOBIPHONE teve como motivação o seu acesso público e gratuito, diferentemente das bases apresentadas em outros trabalhos.

Este banco de dados apresenta 24 gravações de áudio para cada um dos 21 modelos diferentes de telefones celulares pertencentes a 7 fabricantes distintos, totalizando 504 arquivos de áudio. Cada gravação tem duração média de 30s e contém 10 frases faladas por um narrador aleatório escolhido entre 12 homens e 12 mulheres selecionados a partir da TIMIT corpora (KOTROPOULOS, 2014). Todos os arquivos são armazenados em formato “.wav”.

A TIMIT é uma base de dados popular em processamento de fala, ela armazena gravações de frases feitas por 630 diferentes narradores, das quais 432 são de gênero masculino e 192 do gênero feminino, todos falantes de dialetos do inglês americano (HANILÇI *et al.*, 2012).

3 METODOLOGIA

A metodologia de pré-processamento, treinamento, validação e avaliação dos classificadores propostos será apresentada no presente capítulo. Neste trabalho será replicada a metodologia apresentada em HANILÇI *et al.* (2012).

3.1 MATERIAIS

Todos os testes realizados na execução deste trabalho foram feitos em um *notebook* do modelo Inspiron 15 3000 da fabricante Dell. Esse *notebook* possui um processador Intel Core I3 e memória de 4GB.

O software Anaconda Navigator foi empregado na execução do código que foi desenvolvido na linguagem Python. Para tanto, foram empregadas as bibliotecas: *librosa*, *scikit-learn*, *pandas*, *numpy* e *IPython*.

3.2 PRÉ-PROCESSAMENTO DE DADOS

O banco de dados mencionado no capítulo prévio, MOBIPHONE, foi empregado neste experimento por ser um banco de dados disponível publicamente de maneira gratuita. Conforme mencionado anteriormente, o banco conta com 24 gravações para cada um dos 21 modelos de celular que compõe o banco, totalizando 504 amostras de áudios a serem analisadas.

Na etapa de pré-processamento, antes que quaisquer cálculos fossem realizados foi feita a normalização dos sinais e uma associação entre os nomes dos modelos de smartphones e números inteiros para que pudessem ser utilizados como rótulos de entrada nos classificadores. A correspondência entre nome e índice pode ser vista na Figura 6.

Após a numeração, as amostras foram divididas de maneira aleatória entre treinamento e teste, com 70% das amostras compondo o grupo de treinamento e os 30% restantes o grupo de teste. Esse procedimento foi feito independentemente da marca e modelo do aparelho celular.

Figura 6 – Correspondência entre modelo de telefone celular e índice

	Índice
Apple iPhone 5	1
HTC desire c	2
HTC Sensation xe	3
LG GS290	4
LG L3	5
LG Optimus L5	6
LG Optimus L9	7
Nokia 5530	8
Nokia C5	9
Nokia N70	10
Samsung e1230	11
Samsung E2121B	12
Samsung E2600	13
Samsung Galaxy GT-I9100 s2	14
Samsung Galaxy Nexus S	15
Samsung GT-I8190 mini	16
Samsung GT-N7100	17
Samsung s5830i	18
Sony ericson c510i	19
Sony Ericson c902	20
Vodafone joy 845	21

Fonte: autor.

3.3 EXTRAÇÃO DE CARACTERÍSTICAS

Seguindo este processo, deu-se início à etapa de extração dos vetores de características obtidos pelo cálculo dos MFCCs. Os MFCCs são classificados como características espectrais de curto prazo (KINNUNEN e LI, 2010). Para realizar seu cálculo, cada arquivo de áudio foi dividido em frames 30 ms de duração com sobreposição de 15 ms a uma taxa de amostragem de 22.050Hz e foram computados os MFCCs, resultando em um vetor de parâmetros característicos para cada frame.

Durante a etapa de extração de características foram testados diversos tamanhos de vetores de MFCCs entre 5 e 25 coeficientes. Em sequência, foi realizado um aumento de dimensão do vetor de características de acordo com o procedimento descrito em HANILÇI *et al.* (2012). Para tanto, foi utilizado o kernel de Sequência Discriminante Linear Generalizada (GLDS, do inglês *Generalized Linear Discriminant Sequence*). O GLDS mapeia as características extraídas dos frames através de uma expansão polinomial, a ordem do polinômio pode ser definida de acordo com as necessidades. Para um vetor P de quatro parâmetros MFCC tem-se:

$$P = [x_1, x_2, x_3, x_4]^T. \quad (18)$$

A expansão polinomial de ordem 2 resulta em um vetor $P_{GLDS2}(P)$ de 15 parâmetros, cujos elementos são dados por:

$$P_{GLDS2}(P) = \begin{cases} x_1^2 & x_1x_2 & x_1x_3 & x_1x_4 \\ & x_2^2 & x_2x_3 & x_2x_4 \\ & & x_3^2 & x_3x_4 \\ & & & x_4^2 \\ x_1 & x_2 & x_3 & x_4 \\ & & & 1 \end{cases} \quad (19)$$

A forma de apresentação da equação (19) facilita o entendimento da regra de formação dos parâmetros, embora, em sua forma final, os elementos sejam organizados na forma de um vetor coluna.

Neste trabalho foram testados polinômios de grau 2 e 3, sendo o polinômio de segunda ordem o que demonstrou melhores resultados.

Por fim, foi calculada a média de cada um dos parâmetros ao longo dos frames, também conforme feito em HANILÇI *et al.* (2012).

Embora o vetor de MFCCs tenha propiciado resultados adequados para a classificação de telefones celulares, de acordo com a bibliografia desse trabalho, existem ainda alternativas para otimização dos MFCCs. Dessa forma, poderiam ser esperados resultados ainda mais promissores na classificação, caso utilizados.

Os trabalhos de HANILÇI e ERTAS (2013) e de HADOLTIKAR *et al.* (2019), sugerem adicionar a primeira e a segunda derivada dos MFCCs (Δ -MFCCs e $\Delta\Delta$ -MFCCs) no vetor de características como forma de aprimoramento. Esse procedimento também foi aplicado neste trabalho. Em sequência, a expansão de dimensionalidade e a média entre os frames foram realizadas da mesma maneira descrita anteriormente.

3.4 VALIDAÇÃO DO MODELO

Para a etapa de validação do modelo de treinamento do classificador foram escolhidos dois métodos de validação cruzada, *K-fold* e LOO. No processo de validação *K-fold* foram escolhidos 5 *folds* para validar o treinamento do classificador após a obtenção dos vetores característicos.

3.5 TESTES ADICIONAIS

Motivado pela intenção de mensurar na prática como os classificadores se comportam, foram conduzidos testes adicionais com dois telefones celulares de modelos mais atualizados. Foram selecionados um Galaxy A73 da marca Samsung e um Iphone 13 Pro da fabricante Apple para os testes.

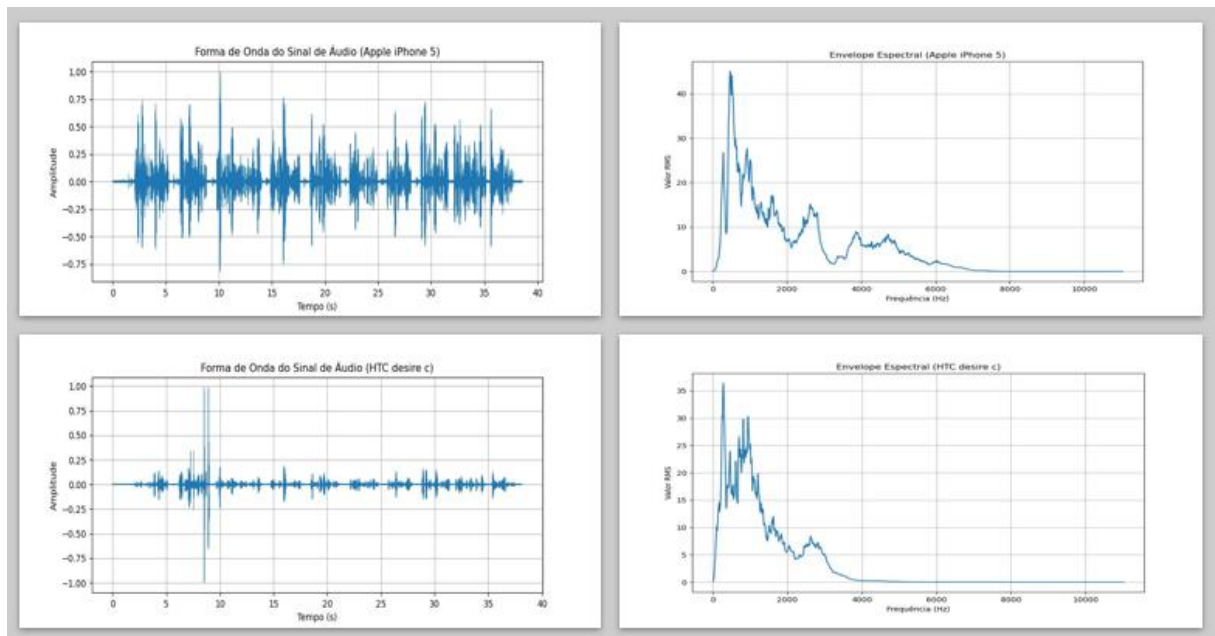
Foi utilizada somente a gravação de uma frase disponível. Para que o tamanho do áudio ficasse próximo de 30s, como o restante dos áudios da base, a gravação foi repetida o número suficiente de vezes para chegar a este tamanho.

Os testes foram conduzidos reproduzindo a gravação em um *notebook* Dell Inspiron 15 3000. Os celulares foram posicionados a uma distância de 15cm do *notebook* em um ambiente silencioso, porém não isolado acusticamente. Os arquivos foram convertidos em formato .wav para padronização.

4 RESULTADOS E DISCUSSÃO

Anterior a qualquer manipulação dos dados, os sinais de áudio foram observados e analisados. A Figura 7 ilustra os gráficos da forma de onda e do envelope espectral dos primeiros 2 modelos de telefone da base de dados, o modelo Iphone 5 da fabricante Apple e o modelo Desire C da fabricante HTC. Ambos os gráficos são referentes à gravação “speaker1.wav” da base MOBIPHONE. Portanto, o locutor e o texto são os mesmos para ambas as gravações.

Figura 7 – Comparação entre forma de onda e envelope espectral em telefones Iphone5 e Desire C

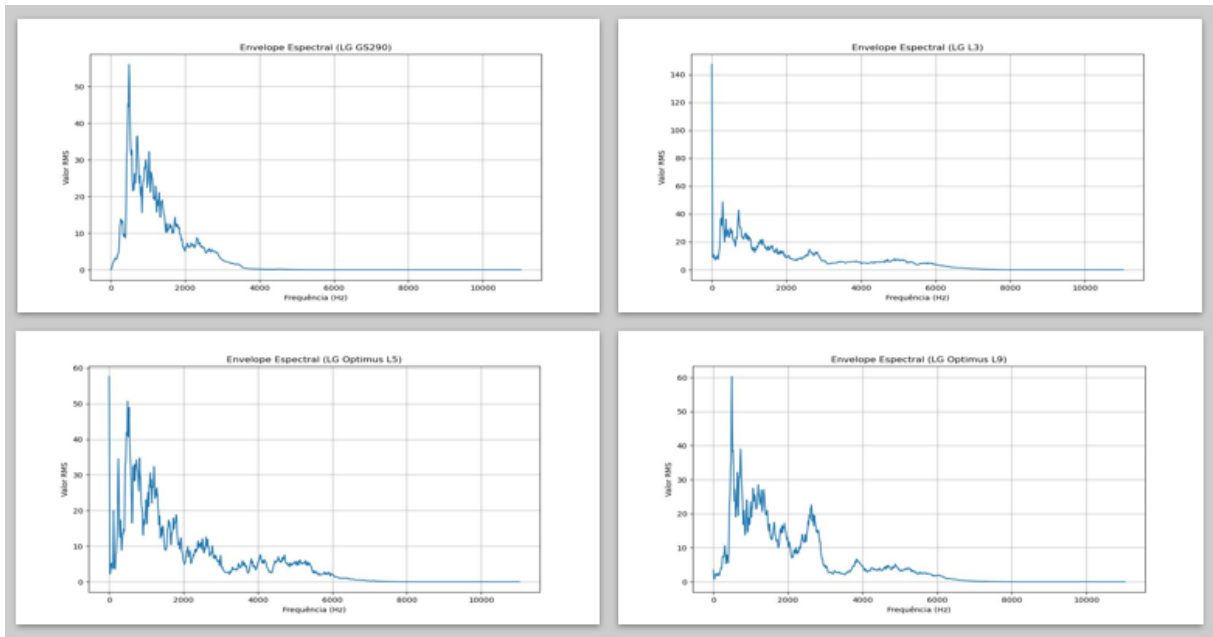


Fonte: autor.

Observando a forma de onda do sinal de ambos os telefones, podem ser percebidas diferenças. Entretanto, os gráficos do envelope espectral representam com mais clareza as diferenças entre as marcas.

Ainda assim, podemos questionar se esta diferença também acontece entre telefones da mesma marca. Para exemplificar, a Figura 8 compara os envelopes espectrais dos 4 modelos distintos de celulares da fabricante LG presentes na base MOBIPHONE, onde é percebido que mesmo entre modelos do mesma fabricante há diferenças expressivas. Nessa figura o sinal utilizado também foi o “speaker1.wav”.

Figura 8 – Envelope espectral de 4 modelos de telefone distintos da marca LG



Fonte: autor.

4.1 CLASSIFICAÇÃO COM VETOR DE CARACTERÍSTICAS DE MFCCS

Em uma primeira abordagem, utilizamos apenas os MFCCs como método de extração de características dos sinais. Os resultados são resumidos nas Tabelas 1 e 2.

Embora a precisão e a medida F1 do classificador SVM sejam um pouco maiores para 25 coeficientes, o vetor de 20 coeficientes MFCC apresenta um compromisso interessante entre (maior) acurácia e (menor) custo computacional para ambos os classificadores. Por esse motivo, 20 foi selecionado como o tamanho utilizado para o vetor de características.

Tabela 1 – Resultados do classificador SVM utilizando somente MFCCs

SVM				
Nº mfccs	Acurácia	Precisão	Sensibilidade	F1
5	0,8158	0,8334	0,8158	0,8126
10	0,9211	0,9436	0,9211	0,9208
15	0,9408	0,9534	0,9408	0,9375
20	0,9539	0,9621	0,9539	0,9497
25	0,9539	0,9631	0,9539	0,9501

Fonte: autor.

Tabela 2 – Resultados do classificador LDA utilizando somente MFCCs

LDA				
Nº mfccs	Acurácia	Precisão	Sensibilidade	F1
5	0,8618	0,8921	0,8618	0,8603
10	0,9803	0,9842	0,9803	0,9809
15	0,9803	0,9842	0,9803	0,9809
20	0,9803	0,9865	0,9803	0,9816
25	0,8816	0,9001	0,8816	0,8798

Fonte: autor.

Estabelecidos os parâmetros de extração de características, a etapa seguinte é a validação. A validação LOO apresentou resultados notavelmente mais promissores que o *K-fold*, sugerindo que o método de validação escolhido pode apresentar resultados bastante díspares em bases de dados pequenas.

4.2 CLASSIFICAÇÃO COM Δ -MFCCS ADICIONADOS AO VETOR

Posteriormente aos resultados obtidos apenas com os MFCCs e inspirado nos trabalhos de HANILÇI e ERTAS (2013) e de HADOLTIKAR *et al.* (2019) foi avaliada a pertinência de adicionar a primeira ou a segunda derivada dos MFCCs (Δ -MFCCs e $\Delta\Delta$ -MFCCs) ao vetor de características.

Os métodos de otimização propostos tanto em HANILÇI e ERTAS (2013) quanto em HADOLTIKAR *et al.* (2019) foram testados e os resultados obtidos demonstraram que no conjunto alvo, apenas o uso da primeira derivada dos coeficientes apresentou melhorias nos resultados tanto da medida F1 quanto da acurácia dos classificadores. Ao compararmos com a segunda derivada os resultados permaneceram os mesmos.

Na Tabela 3 é possível visualizar uma comparação entre os resultados obtidos utilizando os classificadores SVM e LDA, ambos com os mesmos parâmetros, uma combinação de 20 MFCCs e 20 Δ -MFCCs. Observando a medida F1 e a acurácia, notamos que o classificador LDA possui melhor desempenho em comparação ao SVM, conforme sugerido por BERDICH *et al.* (2022), embora ambos apresentem resultados satisfatórios para o problema proposto.

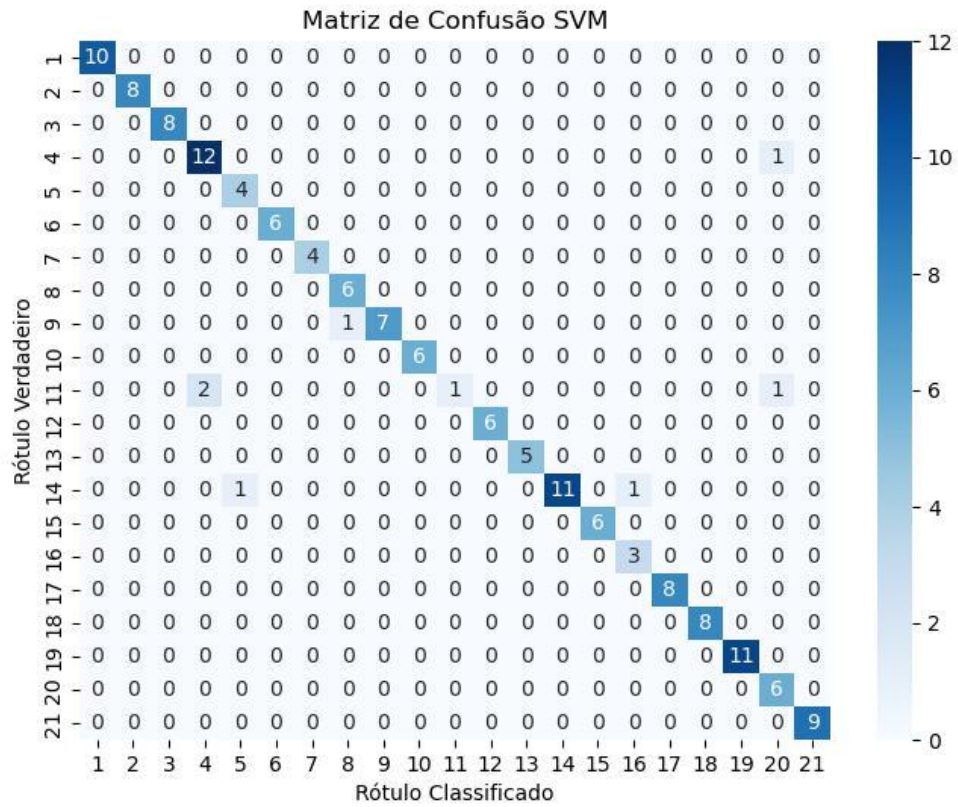
Tabela 3 – Resultados do classificador LDA utilizando MFCCs e Δ -MFCCs

Classificador	MFCCs + Delta MFCCs			
	Acurácia	Precisão	Sensibilidade	F1
SVM	0,9539	0,9621	0,9539	0,9497
LDA	0,9934	0,9951	0,9934	0,9938

Fonte: autor.

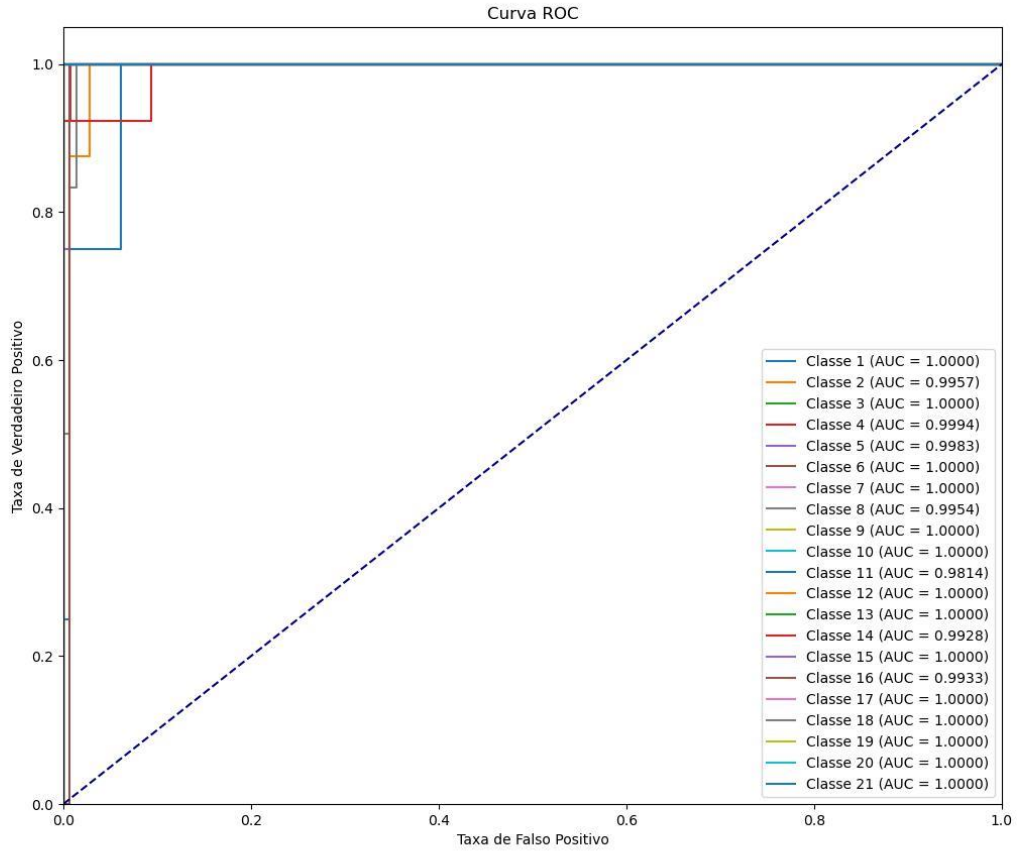
As Figuras 9 e 10 ilustram as medidas de avaliação propostas neste trabalho para o classificador SVM. Já as Figuras 11 e 12 ilustram as medidas para o classificador LDA.

Figura 9 – Matriz de confusão SVM



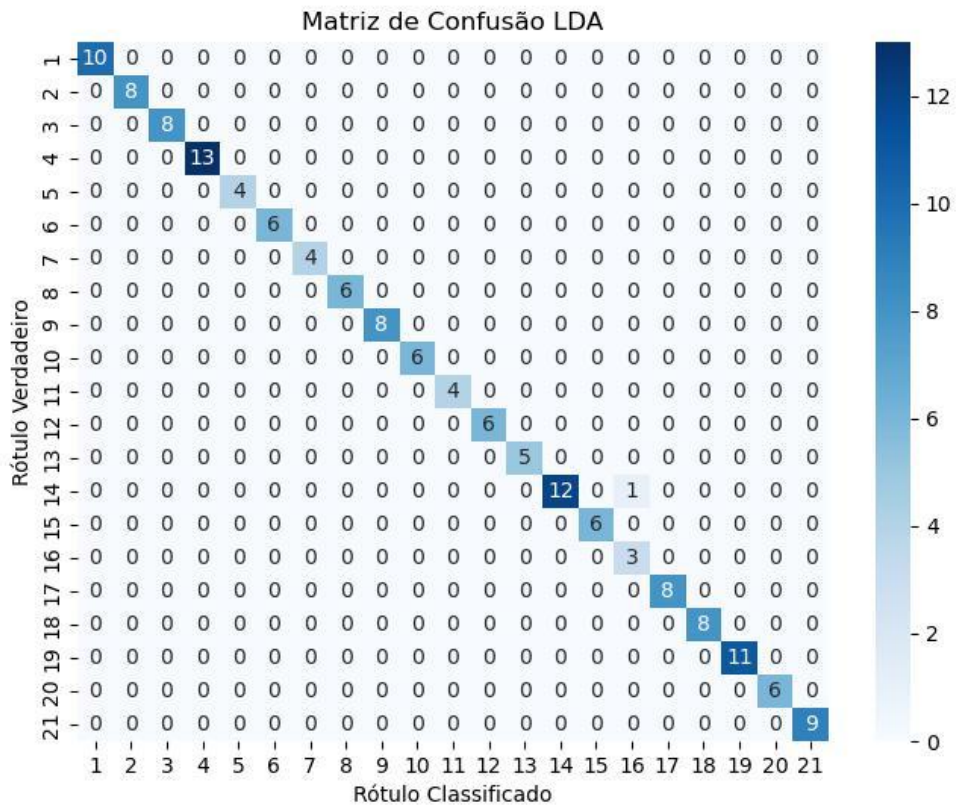
Fonte: autor.

Figura 10 – Curva ROC e AUC do SVM



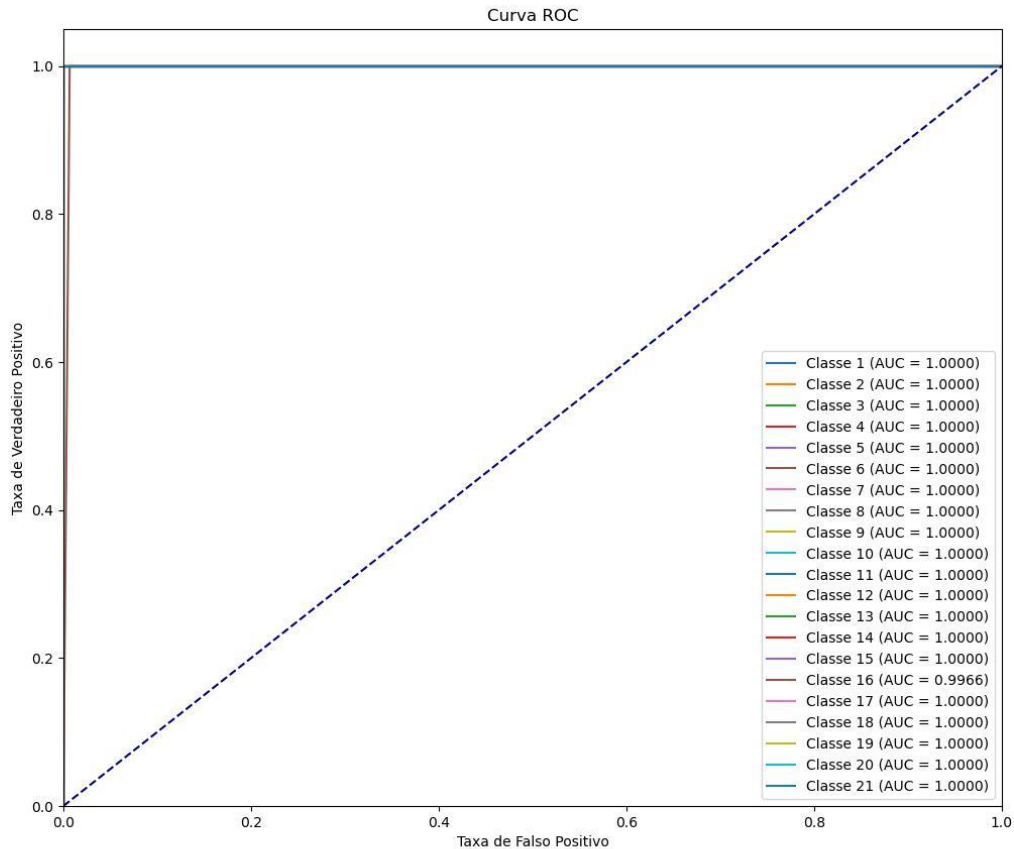
Fonte: autor.

Figura 11 – Matriz de confusão LDA



Fonte: autor.

Figura 12 – Curva ROC e AUC do LDA



Fonte: autor.

4.3 TESTES ADICIONAIS

Por fim, foram realizados testes com os telefones Samsung Galaxy A73 e Iphone 13 Pro. O intuito não foi verificar a eficácia dos classificadores, visto que esses modelos são muito mais atuais, sendo apenas uma investigação adicional.

Selecionando como características apenas os MFCCs, o Samsung Galaxy A73 foi classificado como LG L3 (índice 5) pelo classificador SVM e como Samsung s5830i (índice 18) pelo classificador LDA. O Iphone 13 Pro foi classificado como Samsung s5830i (índice 18) por ambos os classificadores.

Utilizando MFCCs e Δ -MFCCs obtivemos LG L3 (índice 5) pelo classificador SVM e LG Optimus L5 (índice 6) pelo classificador LDA para o Samsung Galaxy A73, enquanto o

Iphone 13 Pro foi classificado como Samsung s5830i (índice 18) pelo classificador SVM e LG Optimus L5 (índice 6) pelo classificador LDA.

4.4 DISCUSSÃO

Avaliando os resultados apresentados nas Figuras 9 a 12 confirmamos a hipótese proposta pela literatura sobre a possibilidade da identificação de aparelhos de telefone celular através de métodos de classificação baseados em aprendizado de máquina.

Embora a base de dados empregada nesta tarefa seja pequena para o problema, ela é suficiente para que se façam algumas generalizações. Os MFCCs realmente se apresentam como um método adequado de extração de características. O classificador SVM obteve uma acurácia de 95,39%, sendo é um resultado próximo ao obtido por HANILÇI *et al.* (2012), de 96,42%, que utiliza um procedimento similar.

Entretanto ainda há espaço para a melhoria dos vetores de características já que em KOTROPOULOS (2014) a acurácia de um classificador SVM utilizando a mesma base de dados chegou a valores de 100% ao escolher diferentes métodos de extração de características.

A base de dados mencionada conta com modelos de telefones celulares que não são atuais. Nesse sentido, a ampliação para uma base que inclua modelos mais modernos de aparelhos seria importante para aplicações reais. No entanto, conjuntos de dados muito grandes podem apresentar um tempo excessivo para o aprendizado do método SVM, a depender dos recursos computacionais disponíveis, sendo necessário um estudo mais aprofundado sobre o uso deste classificador. Por outro lado, o LDA não apresenta as mesmas fragilidades, podendo ser empregado em bases maiores.

Quanto aos métodos de validação cruzada, na hipótese de uma base de dados maior o método LOO pode se tornar mais custoso pelo volume de dados a ser processado embora tenha a vantagem de apresentar uma validação mais precisa, uma vez que testa todos os dados do conjunto. Em contrapartida, o método *K-fold* pode ser ajustado para diferentes números de *folds* estabelecendo um compromisso entre acurácia e custo computacional.

Outro ponto importante a ser destacado é o fato de que todas as gravações de áudio foram feitas em ambientes controlados com minimização de ruídos, o que não representa situações reais de arquivos de áudio encaminhados como evidências periciais. Acredita-se

que, em geral, os arquivos submetidos serão gravados em ambientes ruidosos e este cenário deve ser considerado em trabalhos futuros para que se tenha um resultado mais realista.

O resultado esperado no teste adicional era de que os novos telefones celulares fossem classificados dentre os modelos da mesma marca, porém esse resultado não foi verdadeiro em todos os casos. Algumas hipóteses das razões pelas quais isso aconteceu são o baixo número de celulares da marca Apple na base MOBIPHONE e os novos modelos de CODECs utilizados nesses telefones. Além disso, na Figura 9 podemos observar que em 3 ocasiões o SVM classificou celulares da marca Samsung como sendo pertencentes à marca LG. O mesmo resultado foi observado também no teste adicional, sugerindo que haja alguma similaridade entre os CODECs de ambas as marcas.

5 CONCLUSÃO

Embora comprovada a possibilidade da identificação de telefones celulares como fontes de arquivos de áudio através de classificadores, deve-se olhar com atenção para as nuances presentes em aplicações reais do problema. Conforme mencionado no capítulo 5, fatores como ruído e atualidade dos modelos de aparelho não foram considerados na hipótese, mas podem trazer mudanças significativas nos resultados de classificação.

A ampliação da base de dados também força a necessidade de se repensar nos métodos empregados nas etapas de validação e classificação, buscando menores custos computacionais nos métodos de treinamento.

Contudo, salvo as observações acima, os resultados de 95,39% e 99,34% de acurácia para os classificadores SVM e LDA, respectivamente, são extremamente satisfatórios para uma primeira abordagem do problema.

Em uma sociedade onde os aparelhos de telefone celular se tornaram itens indispensáveis no cotidiano e são manipulados por usuários de praticamente todas as faixas etárias indiscriminadamente, é desejado que se desenvolvam métodos práticos e eficazes para garantir a segurança da população quando se trata de informação multimídia. O presente trabalho abre portas para o aprimoramento de técnicas que contribuam para a resolução dessa problemática.

REFERÊNCIAS

- AGGARWAL, Rachit; SINGH, Shivam; ROUL, Amulva K.; KHANN; Nitin. **Cellphone identification using noise estimates from recorded audio**. 2014 International Conference on Communication and Signal Processing. IEEE, 2014. p. 1218-1222.
- BARUAH, Ujwala; LASKAR, Rabul Hussain; PURKAYASHTHA, Biswajit. **Speaker verification systems: A comprehensive review**. Smart Computing Paradigms: New Progresses and Challenges: Proceedings of ICACNI 2018, v. 1, p. 195-207, 2020.
- BERDICH, Adriana; GROZA, Bogdan; LEVY, Efrat; SHABTAI, Asaf; ELOVICI, Yuval; MAYRHOFER, René. **Fingerprinting smartphones based on microphone characteristics from environment affected recordings**. IEEE Access, v. 10, p. 122399-122413, 2022.
- BERDICH, Adriana; GROZA, Bogdan; MAYRHOFER, René. **A survey on fingerprinting technologies for smartphones based on embedded transducers**. IEEE Internet of Things Journal, v. 10, n. 16, p. 14646-14670, 2023.
- BIMBOT, Frédéric; BONASTRE, Jean-François; FREDOUILLE, Corinne; GRAVIER, Guillaume; MAGRIN-CHAGNOLLEAU, Ivan; MEIGNIER, Sylvain; MERLIN, Teva; ORTEGA-GARCÍA, Javier; PETROVSKA-DELACRÉTAZ, Dijana; REYNOLDS, Douglas. **A tutorial on text-independent speaker verification**. EURASIP Journal on Advances in Signal Processing, v. 2004, p. 1-22, 2004.
- DEUS, Maria Regina Santos de. **Análise de áudio de voz para identificação do emissor utilizando técnicas de processamento de sinais e redes neurais artificiais**. Monografia – Graduação em Ciências da Computação, Escola Politécnica, Pontifícia Universidade Católica de Goiás. Goiânia, 2021.
- DROPBOX. **Mobiphone**. Disponível em: <https://www.dropbox.com/sh/9n7fy7moi825bgk/AABsy5c4J0hnrEHqbBvqhkua>. Acesso em: 17 abr. 2023.
- FACELI, Katti; LORENA, Ana Carolina; GAMA, João; CARVALHO, A. C. P. L. F. de; ALMEIDA, T. A. de. **Inteligência artificial: Uma abordagem de aprendizado de máquina**. [S.l.]: LTC, 2011.
- GARCIA-ROMERO, Daniel; ESPY-WILSON, Carol Y. **Automatic acquisition device identification from speech recordings**. 2010 IEEE International Conference on Acoustics, Speech and Signal Processing. IEEE, 2010. p. 1806-1809.
- GUBERT, Paulo Henrique. **Contribuição à matriz de informações em sistemas de imagética motora baseadas no método de CSP**. 2019. Mestrado em Engenharia Elétrica – Universidade Federal de Santa Catarina.
- HADOLTIKAR, Vishal A.; RATNAPARKHE, Varsha R.; KUMAR, Rajesh. **Optimization of MFCC parameters for mobile phone recognition from audio recordings**. 2019 3rd International Conference on Electronics, Communication and Aerospace Technology (ICECA). IEEE, 2019. p. 777-780.

HANILÇI, Cemal; ERTAS, Figen; ERTAS, Tuncay; ESKIDERE, Ömer. **Recognition of brand and models of cell-phones from recorded speech signals**. IEEE Transactions on Information Forensics and Security, v. 7, n. 2, p. 625-634, 2012.

HANILÇI, Cemal; ERTAS, Figen. **Optimizing acoustic features for source cell-phone recognition using speech signals**. Proceedings of the First ACM Workshop on Information Hiding and Multimedia Security. 2013. p. 141-148.

HANILÇI, Cemal; KINNUNEN, Tomi. **Source cell-phone recognition from recorded speech using non-speech segments**. Digital Signal Processing, v. 35, p. 75-85, 2014.

IBM. **What is LDA?** Disponível em: <https://www.ibm.com/topics/linear-discriminant-analysis>. Publicado em: 23 nov. 2023. Acesso em: 12 mai. 2024.

JIANG, Yuechi; LEUNG, Frank H. F. **Mobile phone identification from speech recordings using weighted support vector machine**. IECON 2016-42nd Annual Conference of the IEEE Industrial Electronics Society. IEEE, 2016. p. 963-968.

JIANG, Yuechi; LEUNG, Frank H. F. **Source microphone recognition aided by a Kernel-based projection method**. IEEE Transactions on Information Forensics and Security, v. 14, n. 11, p. 2875-2886, 2019.

JIN, Chao; WANG, Rangding; YAN, Diqun. **Source smartphone identification by exploiting encoding characteristics of recorded speech**. Digital Investigation, v. 29, p. 129-146, 2018.

KINNUNEN, Tomi; LI, Haizhou. **An overview of text-independent speaker recognition: From features to supervectors**. Speech communication, v. 52, n. 1, p. 12-40, 2010.

KOTROPOULOS, Constantine L. **Source phone identification using sketches of features**. IET Biometrics, v. 3, n. 2, p. 75-83, 2014.

LAMFO-UNB. **Análise Discriminante Linear**. Disponível em: <https://lamfo.unb.br/wp-content/uploads/2020/12/An%C3%A1lise-Discriminante-Linear.pdf>. Publicado em: 10 out. 2020. Acesso em: 12 mai. 2024.

LANTZ, Brett. **Machine Learning with R**. 2.ed. Packt Publishing, 2015.

LI, Yanxiong; ZHANG, Xue; LI, Xianku; FENG, Xiaohui; YANG, Jichen; CHEN, Aiwu; HE, Qianhua. **Mobile phone clustering from acquired speech recordings using deep Gaussian supervector and spectral clustering**. 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2017. p. 2137-2141.

LI, Yanxiong; ZHANG, Xue; LI, Xianku; ZHANG, Yuhan; YANG, Jichen; HE, Qianhua. **Mobile phone clustering from speech recordings using deep representation and spectral clustering**. IEEE Transactions on Information Forensics and Security, v. 13, n. 4, p. 965-977, 2017.

LIN, Xiaodan; ZHU, Jianqing; CHEN, Donghua. **Subband aware CNN for cell-phone recognition**. IEEE Signal Processing Letters, v. 27, p. 605-609, 2020.

LUO, Da; KORUS, Paweł; HUANG, Jiwu. **Band energy difference for source attribution in audio forensics**. IEEE Transactions on Information Forensics and Security, v. 13, n. 9, p. 2179-2189, 2018.

PANDEY, Vandana; VERMA, Vicky Kumar; KHANNA, Nitin. **Cell-phone identification from audio recordings using PSD of speech-free regions**. 2014 IEEE Students' Conference on Electrical, Electronics and Computer Science. IEEE, 2014. p. 1-6.

QIN, Tianyun; WANG, Rangding; YAN, Diqun; LIN, Lang. **Source cell-phone identification in the presence of additive noise from CQT domain**. Information, v. 9, n. 8, p. 205, 2018.

SANTANA, Thiago O. de. **Comparação entre técnicas de aprendizado de máquina no processo de identificação biométrica através de imagens da orelha**. 2019. Monografia. Graduação em Ciência da Computação – Universidade Federal de Ouro Preto.

STEVENS, Stanley Smith; VOLKMANN, John; NEWMAN, Edwin Broomell. **A scale for the measurement of the psychological magnitude pitch**. The Journal of the Acoustical Society of America, v. 8, n. 3, p. 185-190, 1937.

SOUZA, Lucas F. **Predição da fertilidade de reprodutores suínos através de parâmetros de cinética espermática e aprendizado de máquina**. 2022. Monografia. Graduação em Engenharia Elétrica – Universidade Federal de Santa Catarina.

TOCCHETTO, Domingos; ESPINDULA, Alberi. **Criminalística: Procedimentos e Metodologias**, 5 ed, Editora Millennium, 2022.

VERMA, Vinay; KHATURIA, Preet; KHANNA, Nitin. **Cell-phone identification from recompressed audio recordings**. 2018 Twenty Fourth National Conference on Communications (NCC). IEEE, 2018. p. 1-6.

VERMA, Vinay; KHANNA, Nitin. **CNN-based system for speaker independent cell-phone identification from recorded audio**. CVPR Workshops. 2019. p. 53-61.

VERMA, Vinay; KHANNA, Nitin. **Speaker-Independent source cell-phone identification for re-compressed and noisy audio recordings**. Multimedia Tools and Applications, v. 80, n. 15, p. 23581-23603, 2021.

ZOU, Ling; YANG, Jichen; HUANG, Tangsen. **Automatic cell phone recognition from speech recordings**. 2014 IEEE China Summit & International Conference on Signal and Information Processing (ChinaSIP), 2014, pp. 621-625.

ZOU, Ling; HE, Qianhua; WU, Junfeng. **Source cell phone verification from speech recordings using sparse representation**. Digital Signal Processing, v. 62, p. 125-136, 2017.