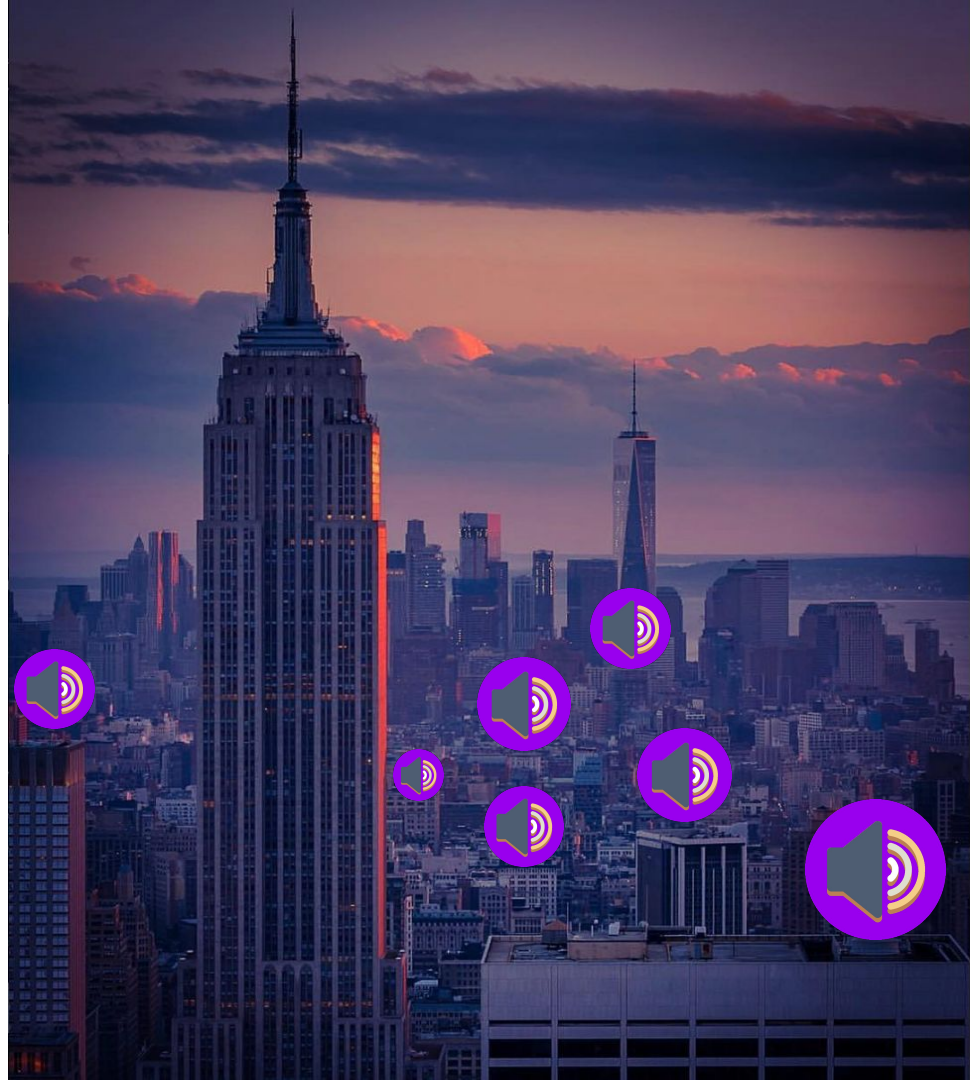


Processamento e Análise de Sinais Acústicos em Cenários Urbanos com ConvNets: Teoria e Prática

*Deborah Magalhães, Flávio Araújo, Jederson
Luz, Myllena Caetano, Fátima Medeiros*



Agenda

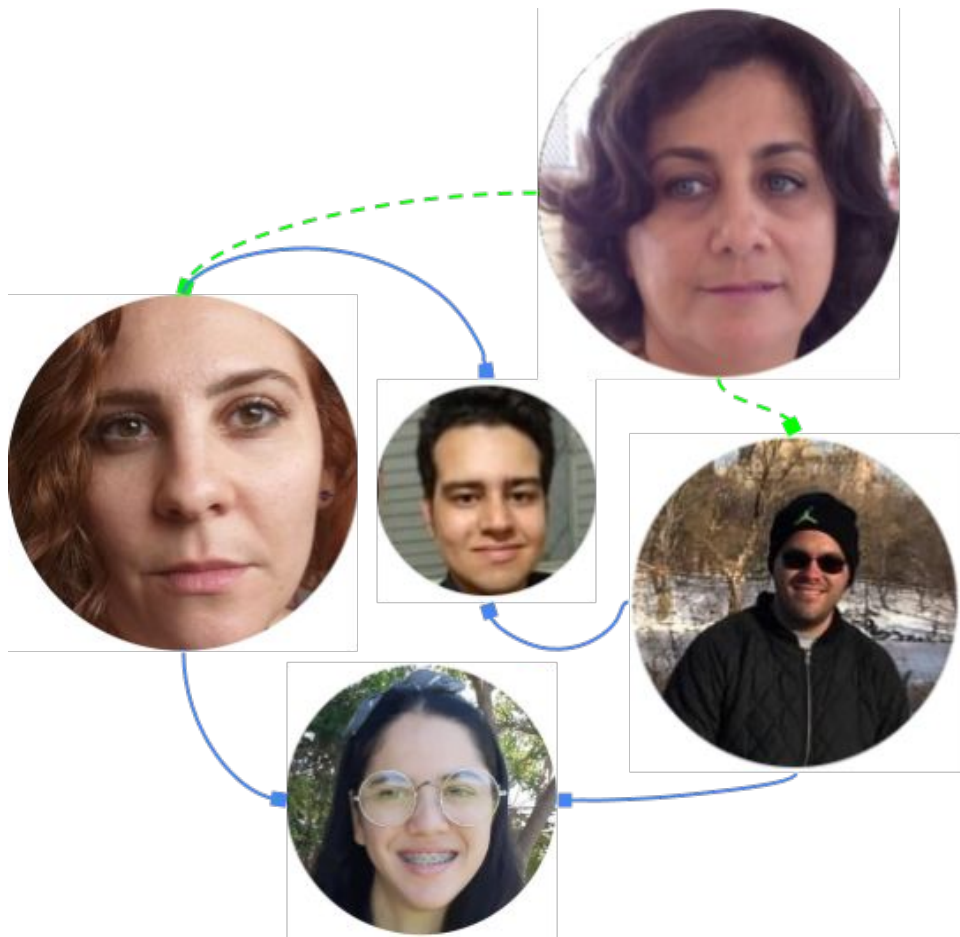
1. Introdução
2. Aquisição dos áudios
3. Pré-processamento
4. Extração de características
5. Classificação de áudios
6. Validação

Olá!

Esse minicurso foi desenvolvido em uma parceria entre o laboratório **Pavic** da Universidade Federal do Piauí e o laboratório **LabVis** da Universidade Federal do Ceará.

Onde nos encontrar:

- <http://www.gpi.ufc.br/>
- <https://github.com/deborahvm/AudioProcessing>



1.

Introdução





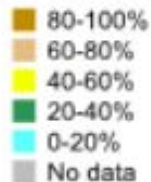
**A China apresentou um crescimento de
40% no número de habitantes urbanos
nos últimos 50 anos**



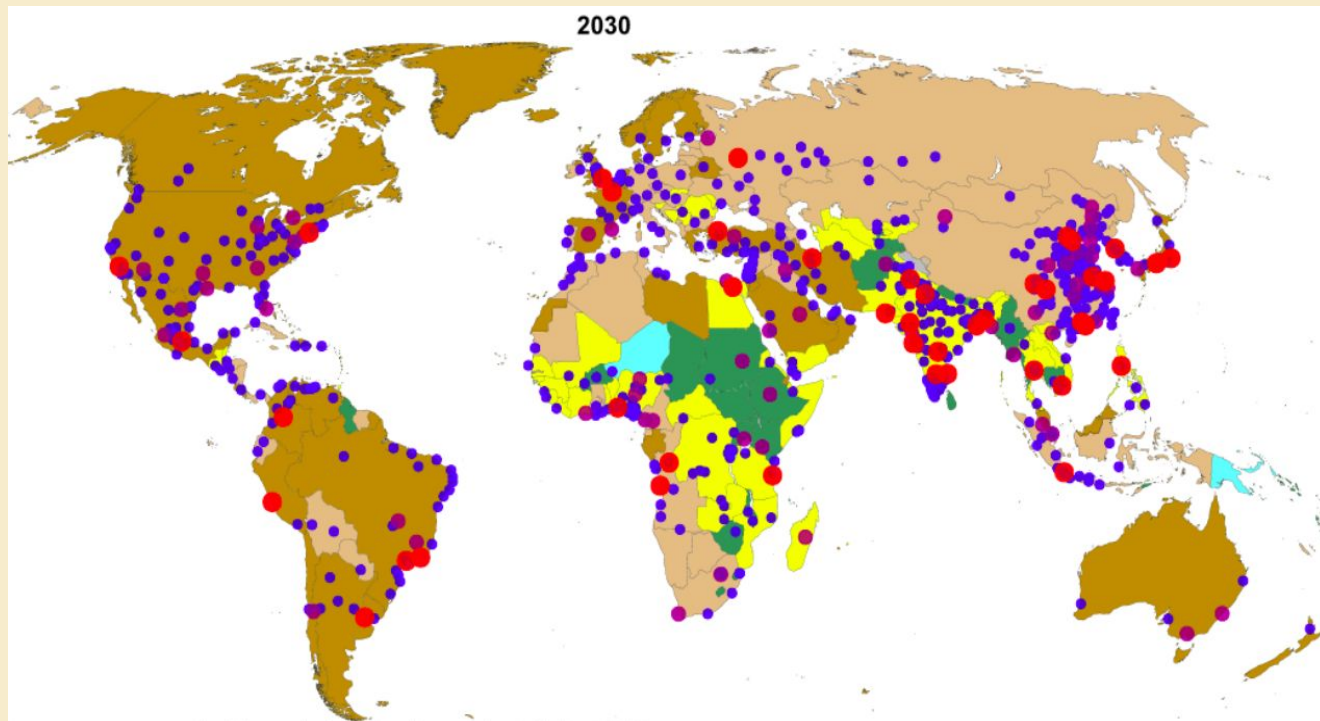
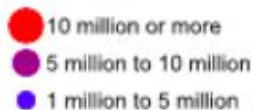
Em 2018, cerca de 55,3% da população mundial vivia em espaços urbanos, esse número chegou a 80% quando tratamos da Europa e América do Norte

Até 2030, as áreas urbanas devem abrigar 60% das pessoas em todo o mundo.

Percentage urban



City population



**O aumento da
densidade
populacional
traz consigo
diversos
desafios**

Mobilidade

Aumento dos congestionamentos



Saúde

Aumento da poluição do ar e sonora,
aumento do esgoto e lixo sólido



Segurança

Aumento da criminalidade



Cidades Inteligentes

As cidades inteligentes oferecem melhores serviços e infraestrutura aos cidadãos



Urban IoT + Cidades Inteligentes ajudam a transpor os desafios gerados pela urbanização

Mobilidade

Gerenciamento do tráfego,
estacionamento e paradas de ônibus inteligentes



Saúde

Assistência de idosos, gerenciamento da poluição
do ar e sonora



Segurança

Vigilância e manutenção de espaços públicos
e ações antiterroristas



Monitoramento Acústico

O som é uma importante fonte de informação a respeito da vida urbana



Microfone como dispositivo de monitoramento

Custo



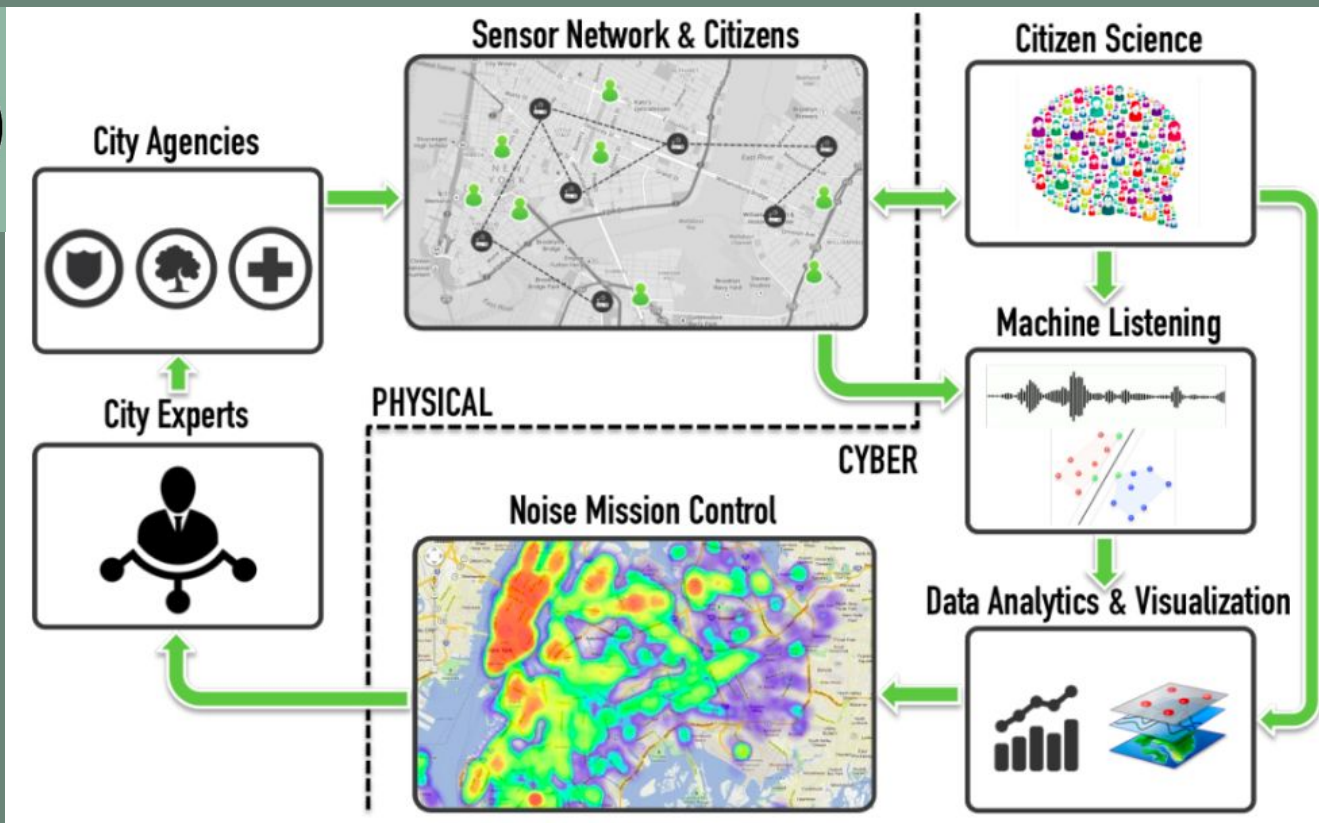
Robustez



Consumo de energia



PROJETO SONYC



Desafios do monitoramento acústico

Heterogeneidade



**Sobreposição
de eventos**



Interferência



Objetivo

Apresentar os passos para realizar a classificação automática de eventos sonoros urbanos :

- Pré-processamento dos áudios
- Aprendizagem de características



librosa

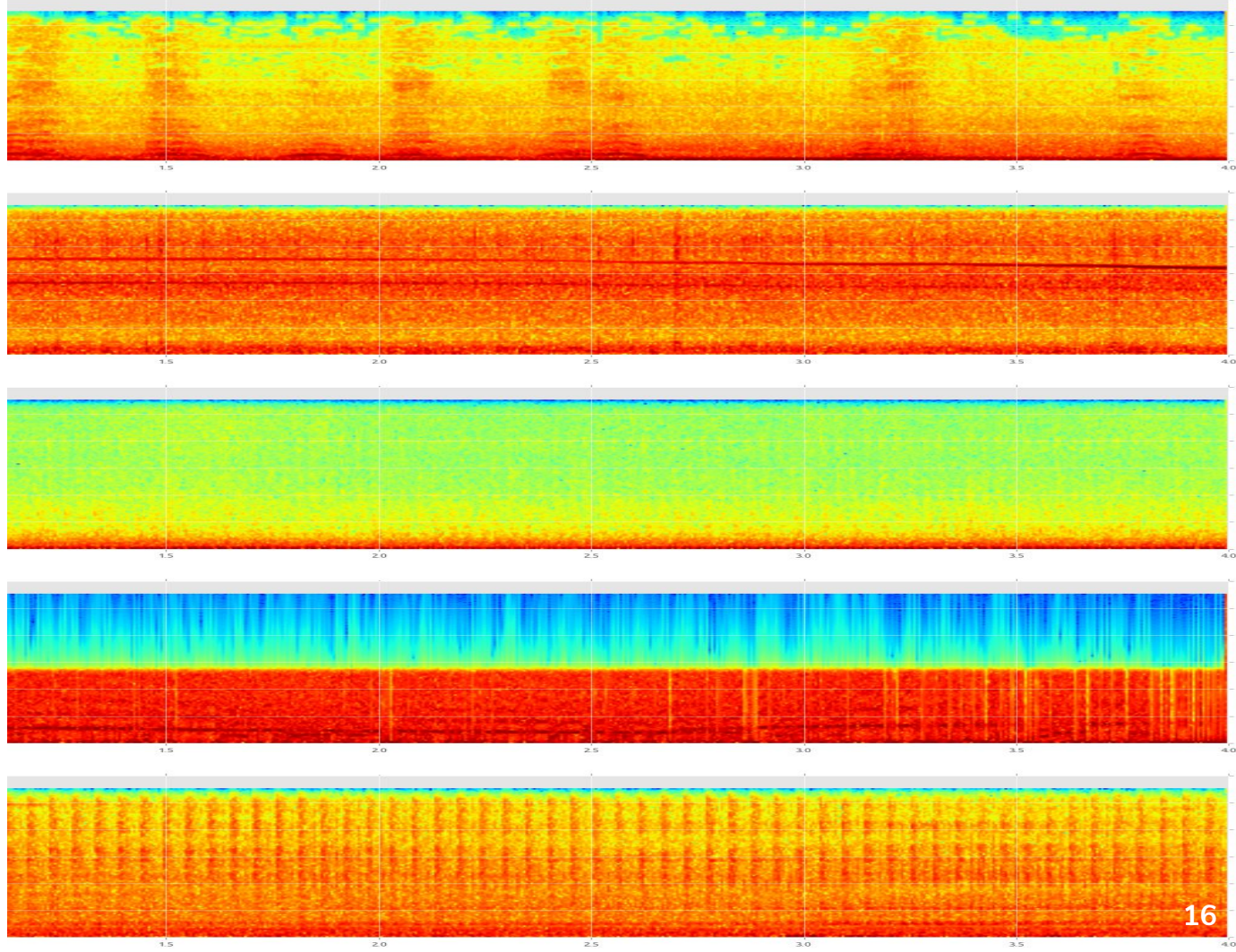


Keras






TensorFlow

2. Classificação de eventos sonoros urbanos



Classificação de eventos Sonoros

Extraír informação útil do sinal de áudio
para discriminar da melhor forma possível
diferentes classes sonoras

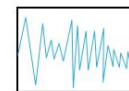
- Identificar o **evento sonoro**  em meio a uma **cena sonora**  

Metodologia

1

Aquisição de áudios

- Base pública
- UrbanSound8K



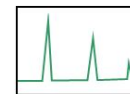
Tempo

2

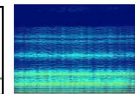
Pré-Processamento

- Uniformização dos dados
- Aumento dos dados
- Representação do sinal

- ✕ SoX
- ✕ MUDA
- ✕ LibROSA



Frequência

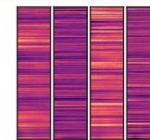


Espectrograma

3

Aprendizado de Características

- CNN
- ✕ Keras
- ✕ TensorFlow



Características acústicas

4

Classificação

- Divisão treino/teste: 80/20
- Random Forest

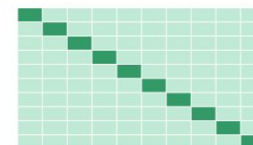


Classes

5

Validação

- Acurácia
- Matriz confusão



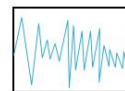
Matriz de Confusão

1. Aquisição dos dados

1

Aquisição de áudios

- Base pública
- UrbanSound8K



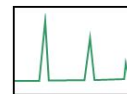
Tempo

2

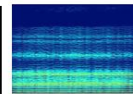
Pré-Processamento

- Uniformização dos dados
- Aumento dos dados
- Representação do sinal

- ✕ SoX
- ✕ MUDA
- ✕ LibROSA



Frequência

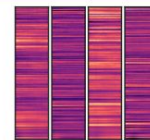


Espectrograma

3

Aprendizado de Características

- CNN
- ✕ Keras
- ✕ TensorFlow



Características acústicas

4

Classificação

- Divisão treino/teste: 80/20
- Random Forest

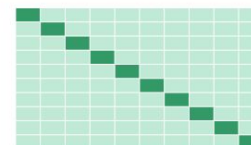


Classes

5

Validação

- Acurácia
- Matriz confusão



Matriz de Confusão

1. Aquisição dos dados

UrbanSound8K

- 8732 áudios com rótulos [0-4s], distribuídos em 10 pastas com classes misturadas
- Formato WAV (**diferentes taxas de amostragem e quantização**)
- 10 classes sonoras **desbalanceadas**

1. Aquisição dos dados



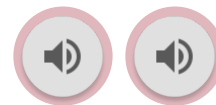
Ar condicionado

1000 amostras



Buzina de carro

429 amostras



Crianças brincando

1000 amostras



Latido de cachorro

1000 amostras



Motor de veículo

1000 amostras

1. Aquisição dos dados



Furadeira

1000 amostras



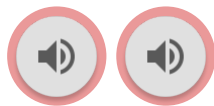
Tiro

374 amostras



Britadeira

1000 amostras



Sirene

929 amostras



Música de rua

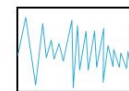
1000 amostras

2. Pré-processamento dos áudios

1

Aquisição de áudios

- Base pública
- UrbanSound8K



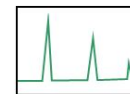
Tempo

2

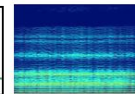
Pré-Processamento

- Uniformização dos dados
- Aumento dos dados
- Representação do sinal

- ✕ SoX
- ✕ MUDA
- ✕ LibROSA



Frequência

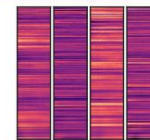


Espectrograma

3

Aprendizado de Características

- CNN
- ✕ Keras
- ✕ TensorFlow



Características acústicas

4

Classificação

- Divisão treino/teste: 80/20
- Random Forest

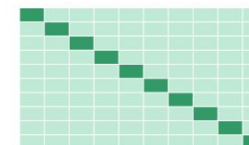


Classes

5

Validação

- Acurácia
- Matriz confusão



Matriz de Confusão

2. Pré-processamento dos áudios



Aumento
dos dados



Uniformização
dos dados



Representação
do sinal



SoX



MUDA



LibROSA



jams



2.1. Aumento dos dados

Aplicar alterações no conjunto de **treinamento** a fim de :

- Aumentar o número de amostras
- Balancear as classes

Tipos de alterações aplicadas

- Variações no tom: {1, 1.5, 2, 2.5, 3, 3.5}
- Ruído de fundo: trabalhadores na rua, tráfego de rua, e pessoas na rua



2.1. Aumento dos dados

1. Realizar a divisão treino e teste do conjunto de dados: 80/20 (pastas 1 e 2)
2. Gerar os arquivos de notação JAMS
3. Aplicar as alterações nos áudios



2.1. Aumento dos dados

Exemplos de saída dos áudios transformados:

- Variação do tom:



Motor de Veículo

Original / 3 semitons



Crianças brincando

Original / 3 semitons

- Ruído de fundo:



Motor de Veículo

Original / cidade



Crianças Brincando

Original / cidade



2.2. Uniformização dos dados

Etapa aplicada ao conjunto de treino e teste:

- Re-amostragem: 44 kHz
- Quantização: 16 bits

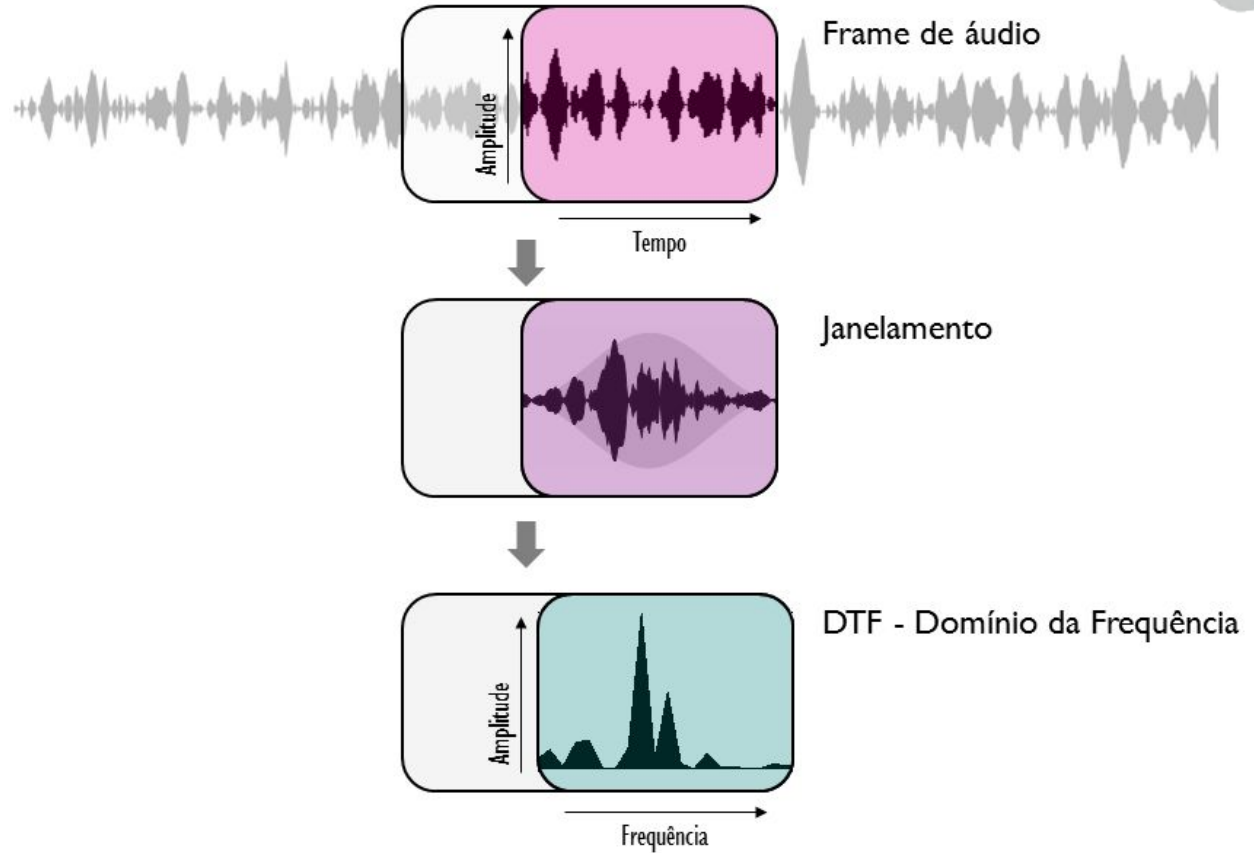
```
135544-6-0-0-ps-2.5-resample.wav:
```

```
File Size: 205k      Bit Rate: 706k  
Encoding: Signed PCM  
Channels: 1 @ 16-bit  
Samplerate: 44100Hz  
Replaygain: off  
Duration: 00:00:02.33
```

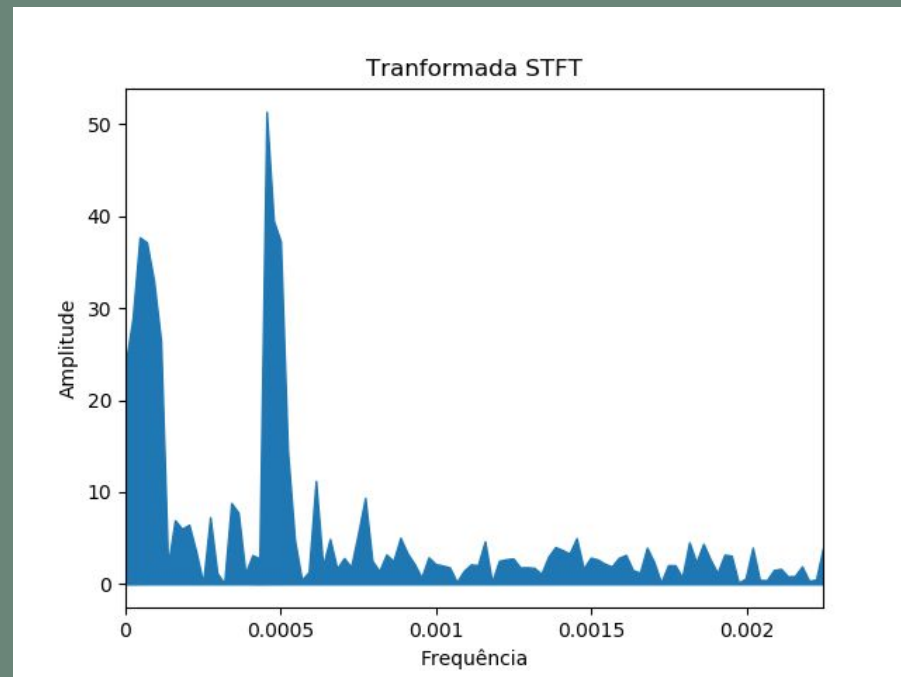
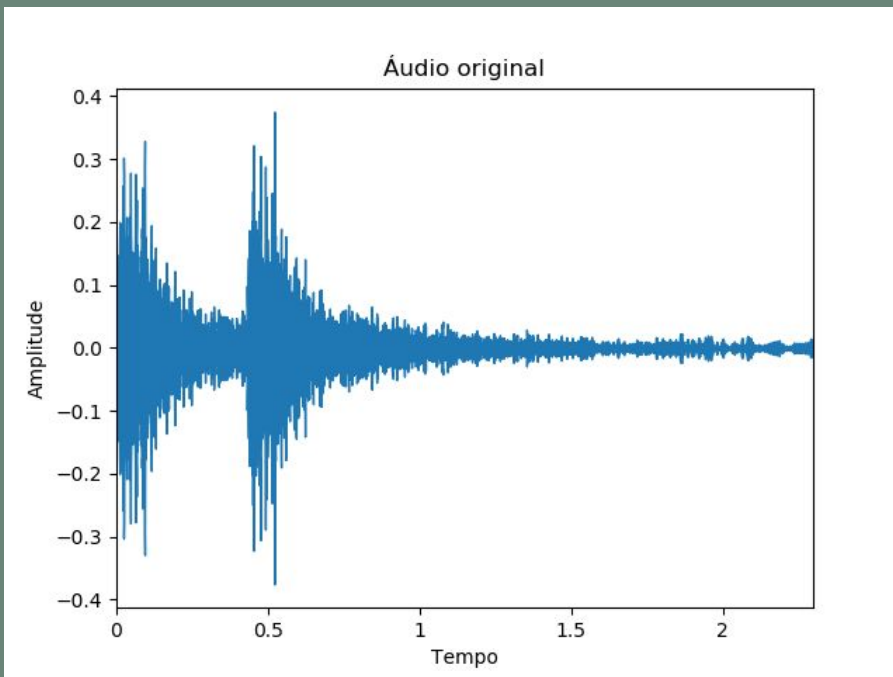
```
In:100% 00:00:02.33 [00:00:00.00] Out:103k [      |      ] Hd:0.0 Clip:0
```



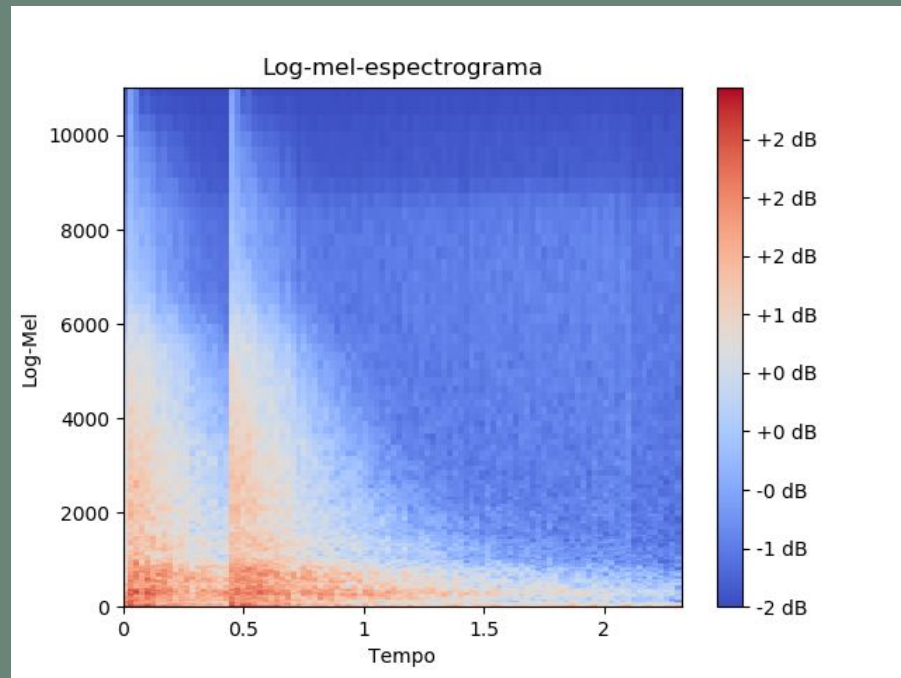
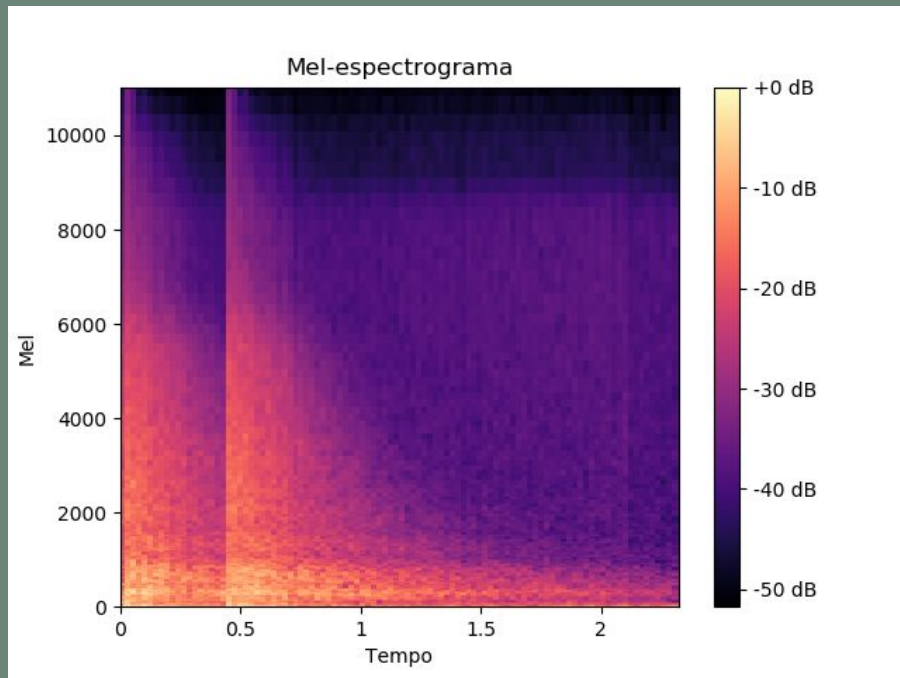
2.3. Representação do Sinal



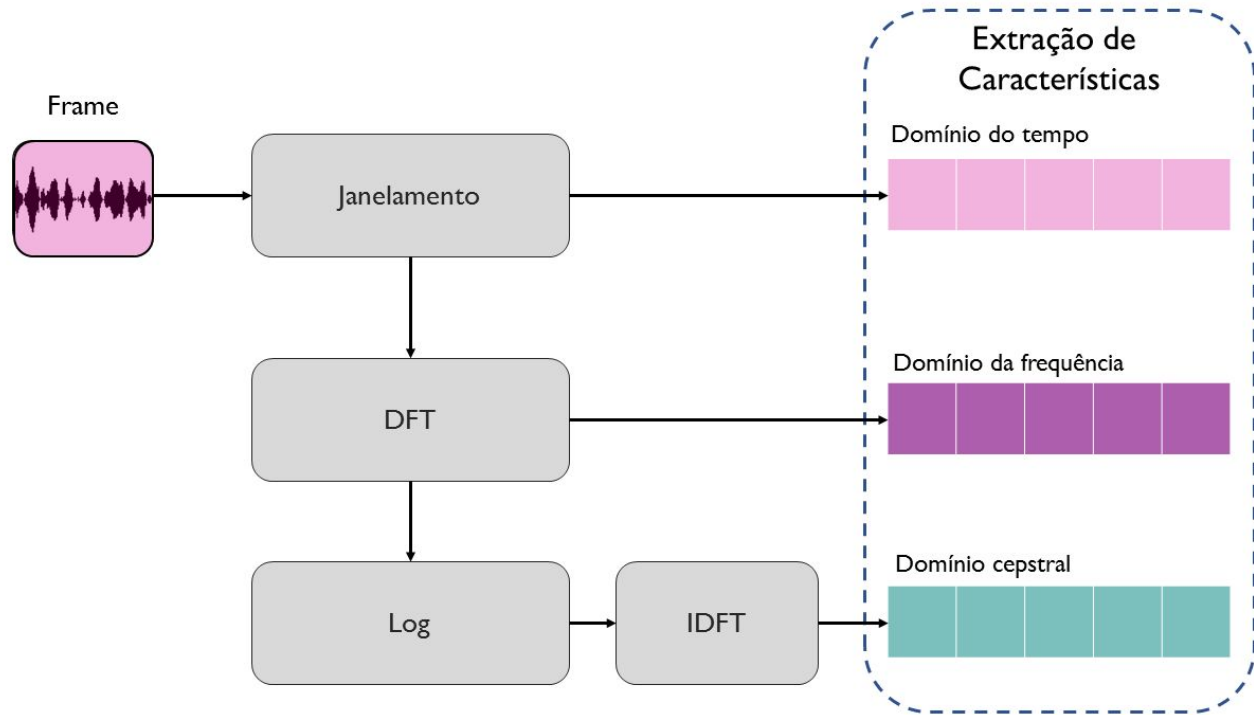
2.3. Representação do Sinal



2.3. Representação do Sinal



3. Extração de Características

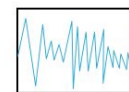


3. Aprendizado de Características

1

Aquisição de áudios

- Base pública
- UrbanSound8K



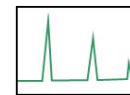
Tempo

2

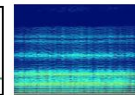
Pré-Processamento

- Uniformização dos dados
- Aumento dos dados
- Representação do sinal

- ✕ SoX
- ✕ MUDA
- ✕ LibROSA



Frequência

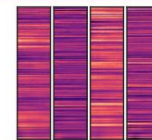


Espectrograma

3

Aprendizado de Características

- CNN
- ✕ Keras
- ✕ TensorFlow



Características acústicas

4

Classificação

- Divisão treino/teste: 80/20
- Random Forest

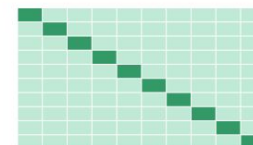


Classes

5

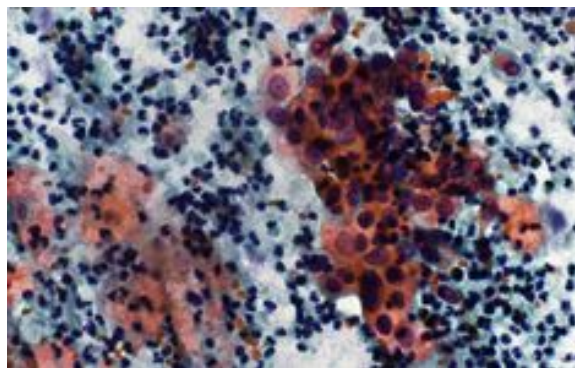
Validação

- Acurácia
- Matriz confusão



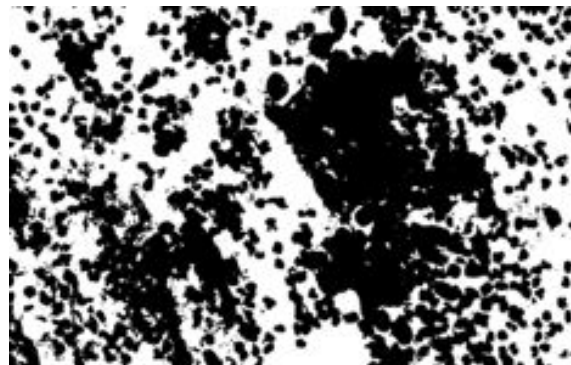
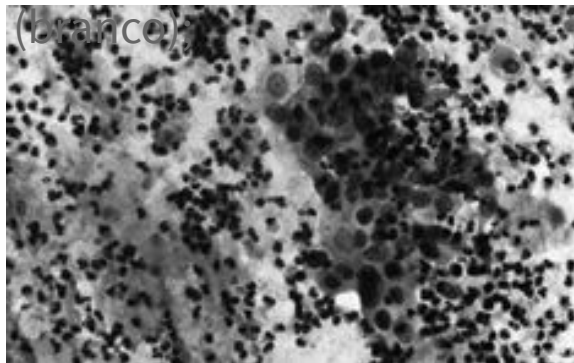
Matriz de Confusão

O que é uma imagem?



Colorida: União de 3 imagens nível de cinza.

Nível de cinza: composta por níveis de cinza que variam de 0 (preto) a 255 (branco).



Binária: composta por duas cores, o preto (0) e o branco (1).

Processo de aprendizagem

Dados os grupos abaixo



(a)



(b)

A qual grupo esse objeto pertence?



Processo de aprendizagem

Dados os grupos abaixo



(a)



(b)

A qual grupo esse objeto pertence?



Processo de aprendizagem

Dados os grupos abaixo



(a)



(b)

A qual grupo esse objeto pertence?



Processo de aprendizagem

Dados os grupos abaixo



(a)



(b)



(c)

A qual grupo esse objeto pertence?



Processo de aprendizagem

- Certamente, sua decisão foi tomada com base no grau de similaridade entre o objeto desconhecido e os grupos conhecidos:
 - Conhecidos como características, atributos ou features.
- A extração de features requer:
 - Conhecimento do domínio do problema;
 - Algumas vezes os problemas são bem específicos;
 - Em aplicações industriais isso representa 90% do tempo.

Métodos tradicionais X Deep learning

TRADITIONAL APPROACH

The traditional approach uses fixed feature extractors.



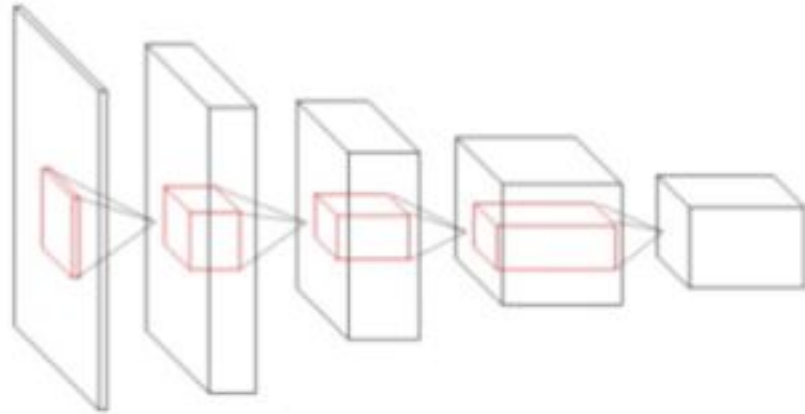
DEEP LEARNING APPROACH

Deep Learning approach uses trainable feature extractors.



O que é deep learning?

- Múltiplas definições, porém todas possuem em comum:
 - Múltiplas camadas de unidades de processamento;
 - As camadas formam uma hierarquia de features low-level para high-level.



O que é deep learning?

- Em 1998 Yann LeCun e seus colaboradores desenvolveram uma rede para reconhecimento de dígitos manuais:

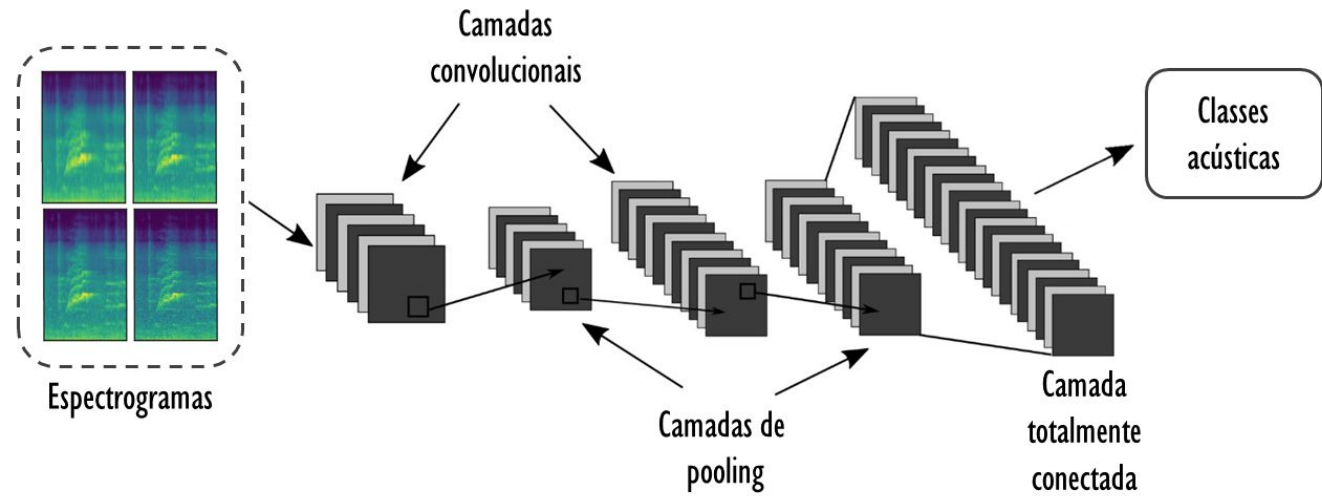
SVM: 94,35%;

SVM otimizada: 98,5%;

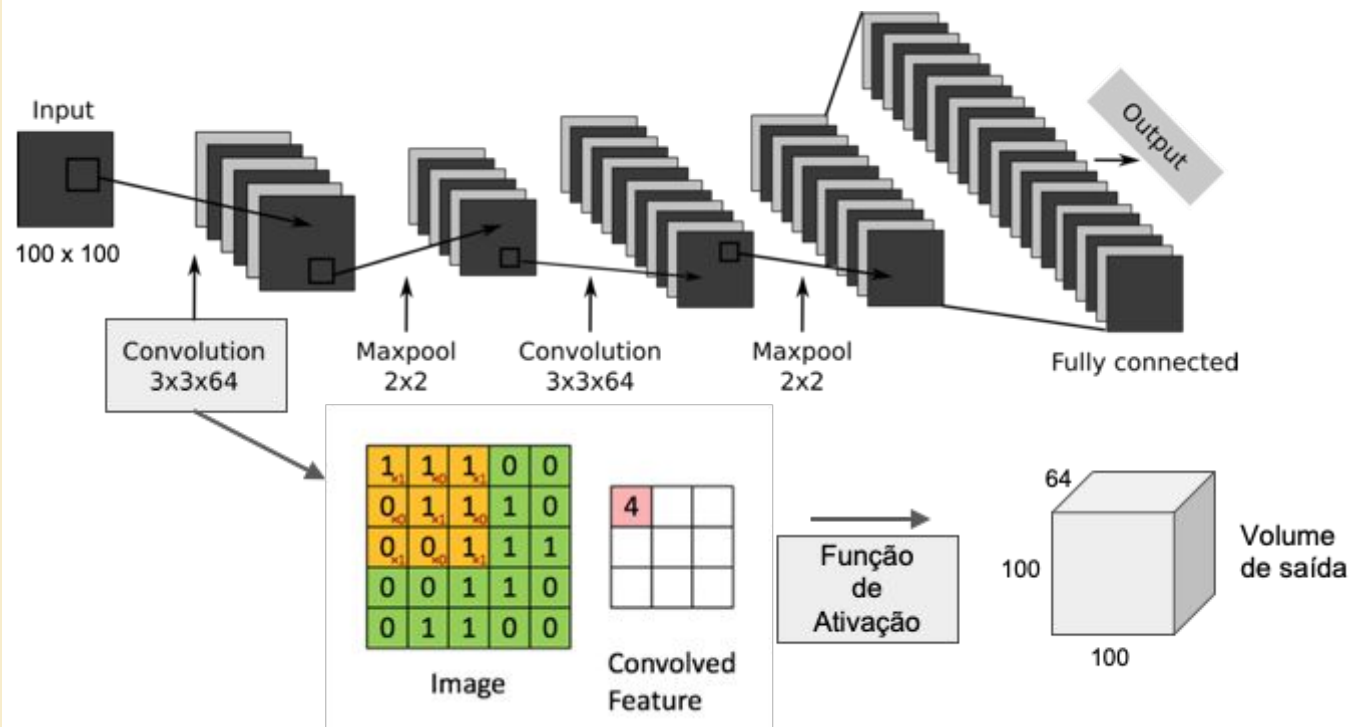
CNN: 99,79%.



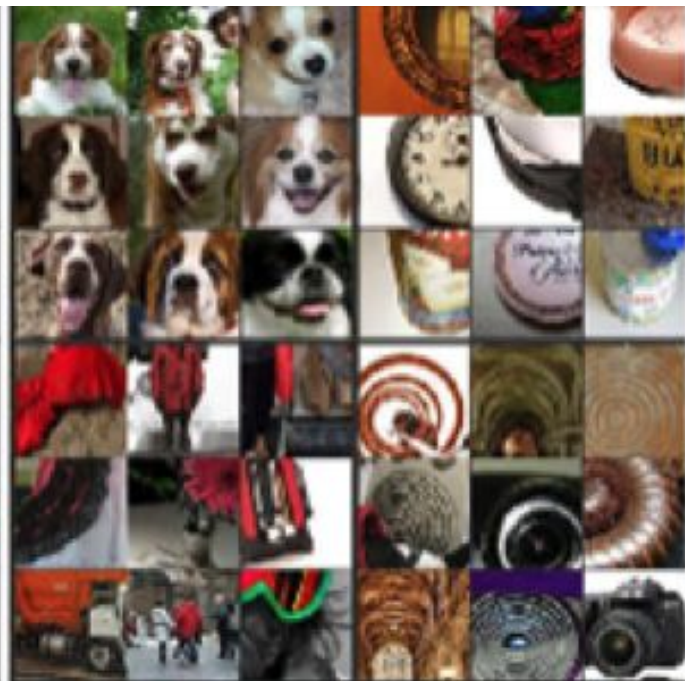
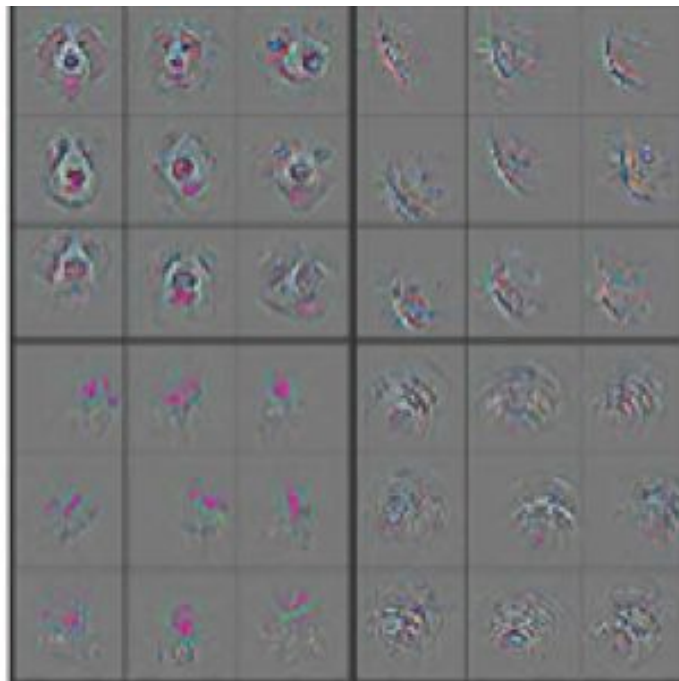
LeNet



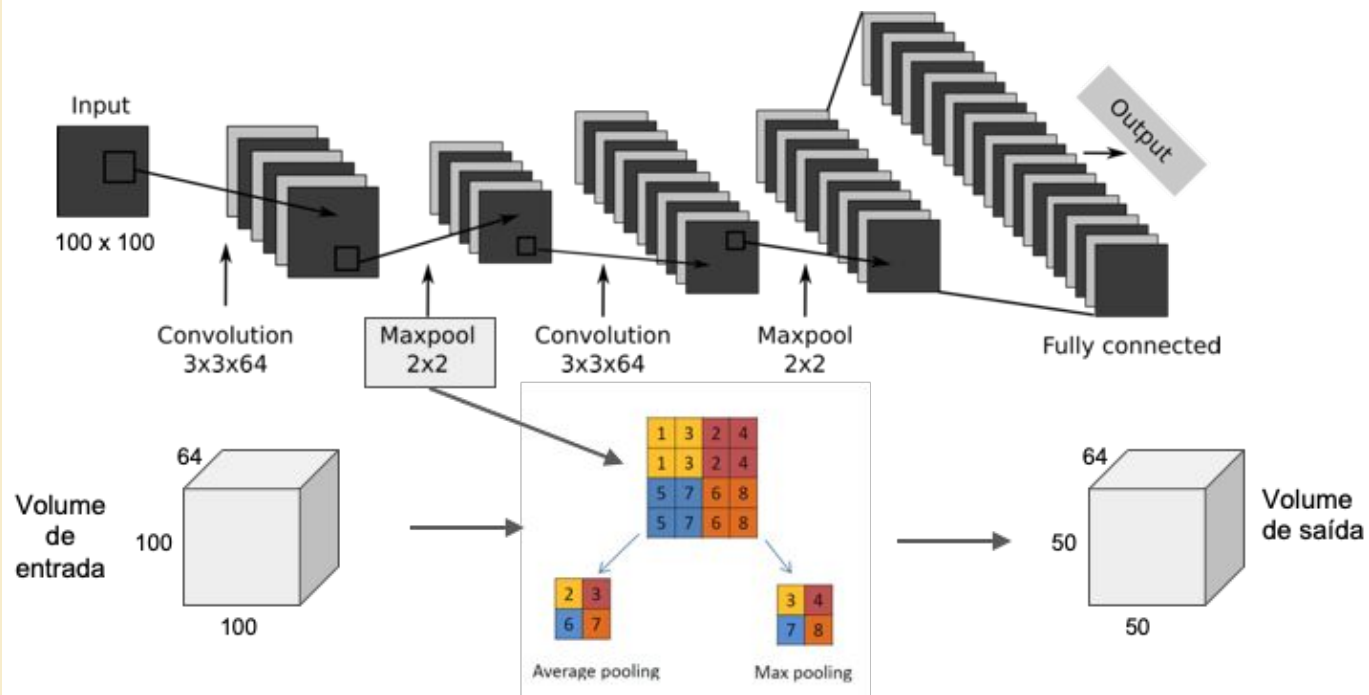
Camada convolucional



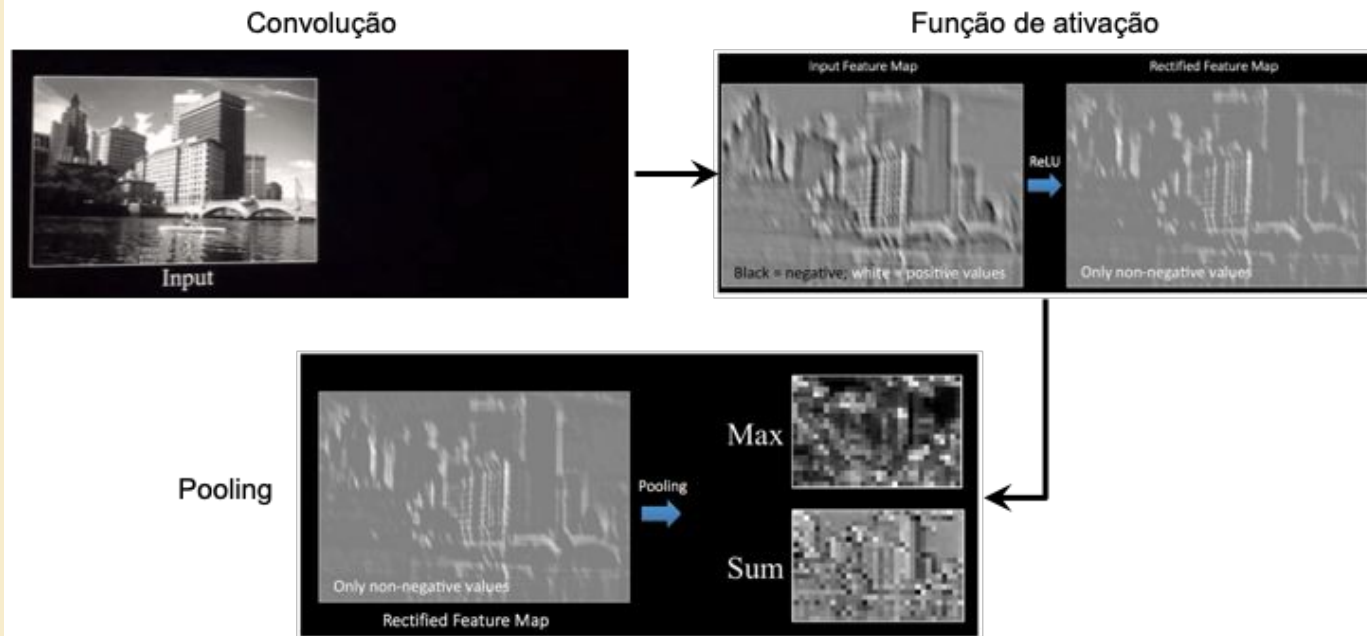
Mapa de ativação



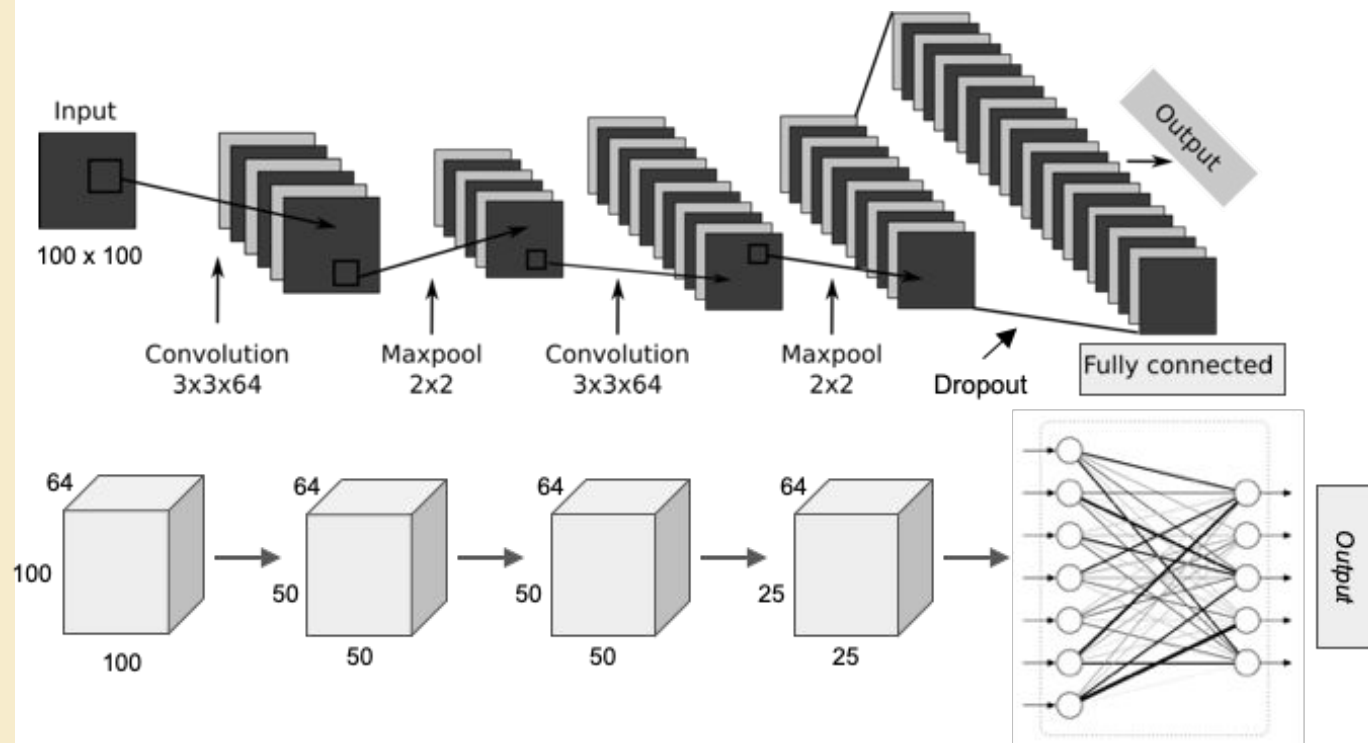
Camada de pooling



Camada de pooling

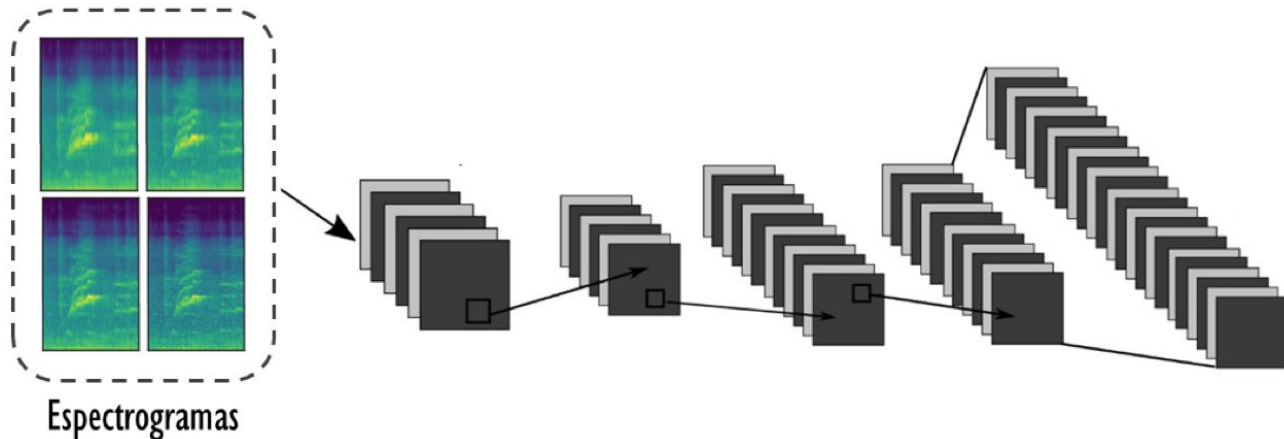


Camada totalmente conectada



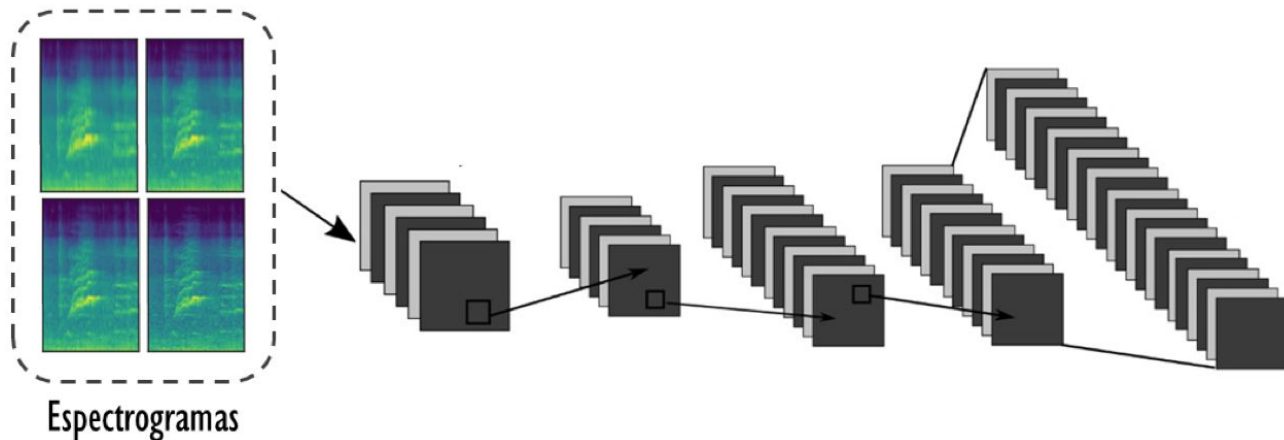
Treinando a CNN

- **Passo 1:** todos os filtros e pesos da rede são inicializados de forma aleatória;
- **Passo 2:** a rede recebe uma amostra de treino como entrada e realiza o processo de propagação, com isso são obtidos os valores de probabilidade da entrada pertencer a cada classe;



Treinando a CNN

- **Passo 3:** é calculado o erro total obtido na camada de saída;
- **Passo 4:** o algoritmo do backpropagation é utilizado para calcular os valores do gradiente do erro, em seguida os valores dos filtros e pesos são ajustados na proporção que eles contribuíram no erro total;



Treinando a CNN

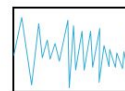
- **Passo 5:** os passos 2-4 são repetidos para todas as amostras do conjunto de treinamento;
- Devido ao ajuste realizado no passo 4, o erro obtido pela rede é menor a cada vez que uma mesma amostra passa pela rede. Essa redução no erro significa que a rede está aprendendo a classificar corretamente as amostras do treinamento;
- Caso o conjunto de treinamento seja abundante e variado o suficiente, a rede apresentará capacidade de generalização e conseguirá classificar corretamente novas amostras que não estavam presentes no processo de treinamento.

4. Classificação

1

Aquisição de áudios

- Base pública
- UrbanSound8K



Tempo

2

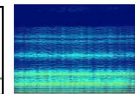
Pré-Processamento

- Uniformização dos dados
- Aumento dos dados
- Representação do sinal

- ✕ SoX
- ✕ MUDA
- ✕ LibROSA



Frequência

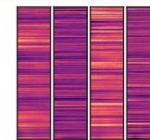


Espectrograma

3

Aprendizado de Características

- CNN
- ✕ Keras
- ✕ TensorFlow



Características acústicas

4

Classificação

- Divisão treino/teste: 80/20
- Random Forest

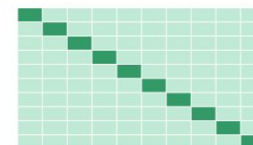


Classes

5

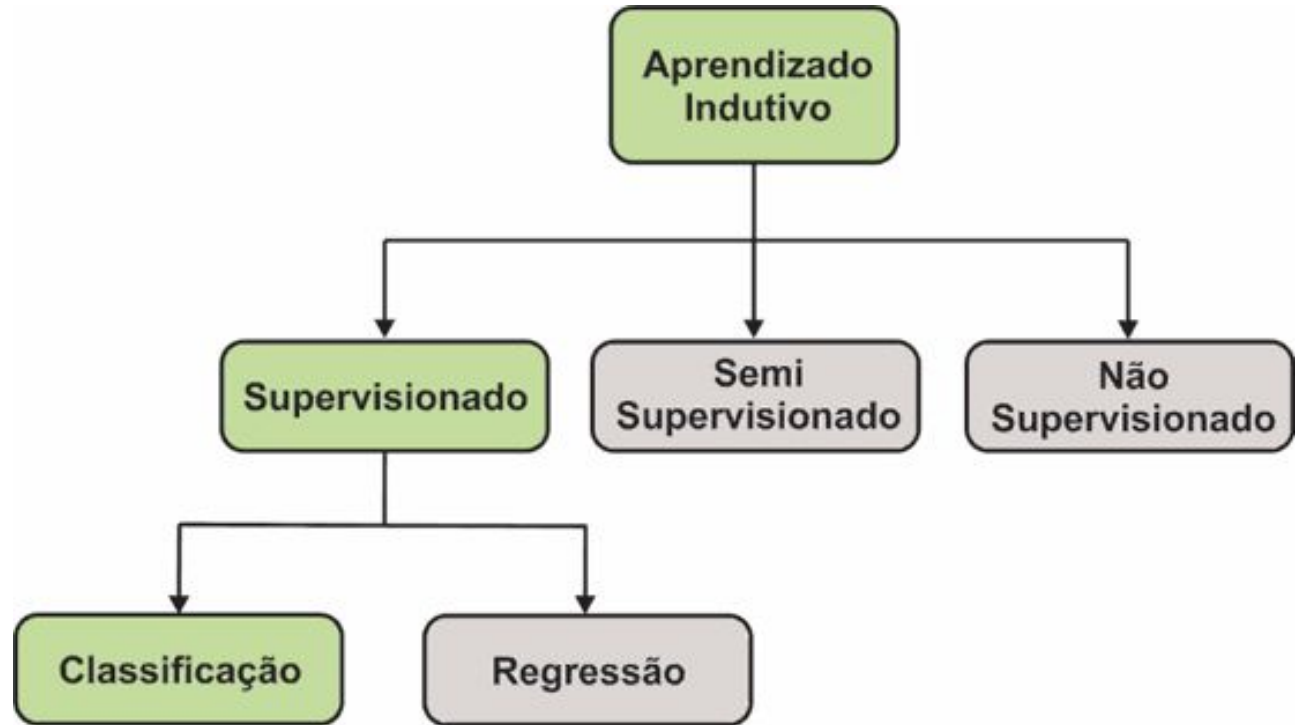
Validação

- Acurácia
- Matriz confusão

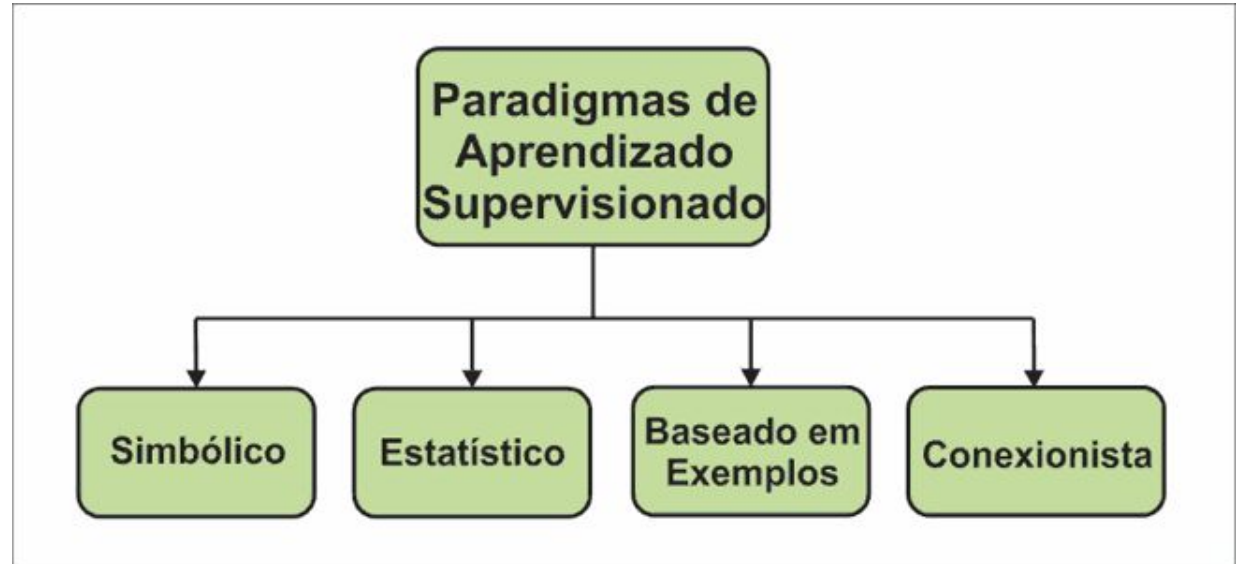


Matriz de Confusão

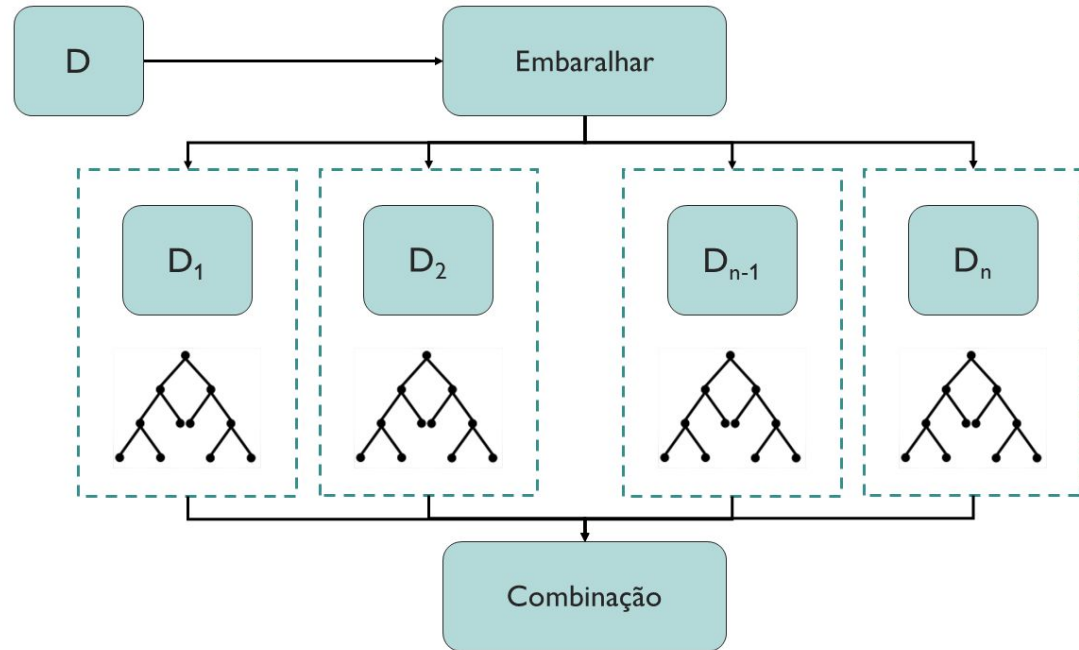
Aprendizado de máquina



Paradigmas do aprendizado de máquina



Random forest

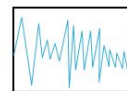


5. Validação

1

Aquisição de áudios

- Base pública
- UrbanSound8K



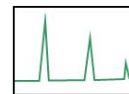
Tempo

2

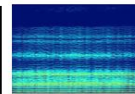
Pré-Processamento

- Uniformização dos dados
- Aumento dos dados
- Representação do sinal

- ✕ SoX
- ✕ MUDA
- ✕ LibROSA



Frequência

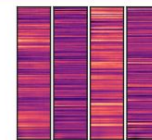


Espectrograma

3

Aprendizado de Características

- CNN
- ✕ Keras
- ✕ TensorFlow



Características acústicas

4

Classificação

- Divisão treino/teste: 80/20
- Random Forest

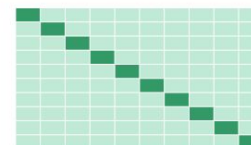


Classes

5

Validação

- Acurácia
- Matriz confusão



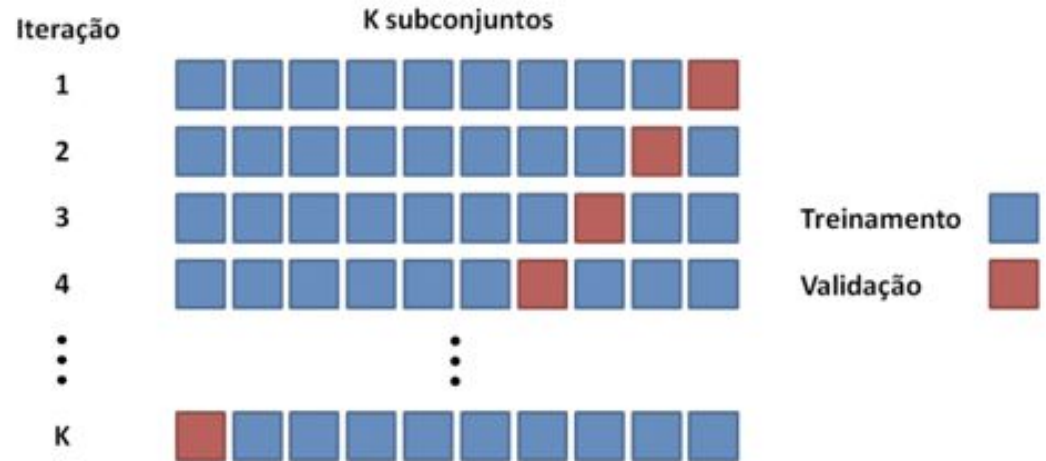
Matriz de Confusão

Divisão treino/teste



Cross-validation

Cross-Validation



Matriz de confusão

Ar condicionado	Buzina de carro	Crianças brincando	Latido de cachorro	Furadeira	Motor de veículo	Tiro	Britadeira	Sirene	Música de rua
79	0	3	1	7	14	0	0	0	6
0	31	0	0	3	0	0	0	2	2
29	0	178	13	9	9	0	1	15	11
3	0	6	116	4	7	0	0	0 8	0
3	1	0	0	103	0	1	3	0	5
56	0	0	2	0	134	0	105	3	0
1	0	1	0	1	26	23	0	1	0
19	0	0	0	18	0	0	82	0	4
0	0	3	4	4	4	0	21	147	7
9	0	9	1	8	1	0	2	1	165



Muito Obrigado!

Se você tiver qualquer dúvida ou sugestão a respeito desse minicurso, por favor fale conosco:

- deborah.vm@ufpi.edu.br
- flavio86@ufpi.edu.br

