

# CLASSIFICAÇÃO DE MODELOS CARDÍACOS TRIDIMENSIONAIS RELACIONADA À PRESENÇA OU À AUSÊNCIA DE CARDIOMIOPATIA

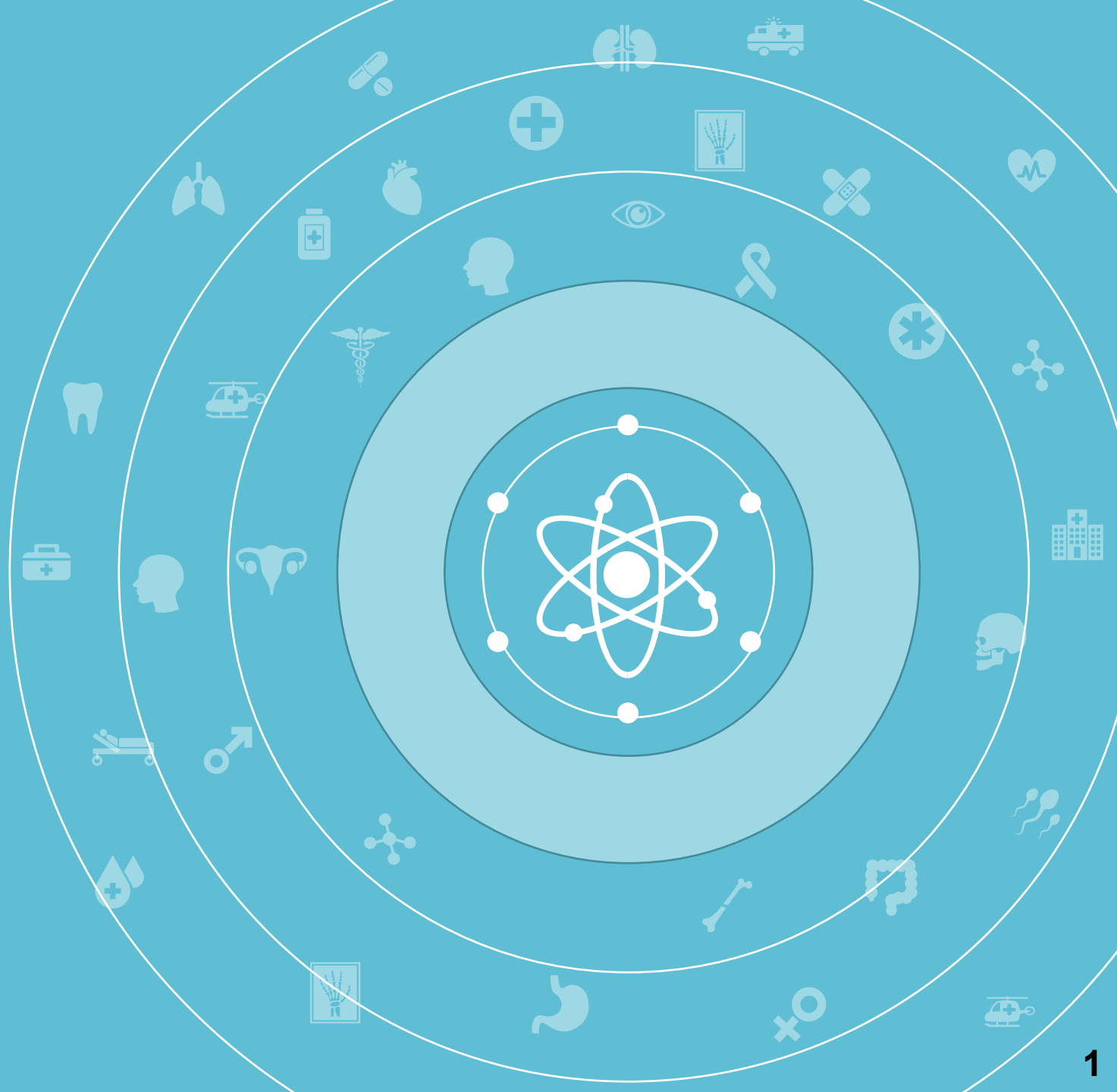
SIN5007 – Reconhecimento de Padrões  
1º sem. 2020

Profa. Dra. Arianne Machado Lima

Diego Pereira Dedize

Luiz Gustavo de Almeida Silva

Vagner Mendonça Gonçalves



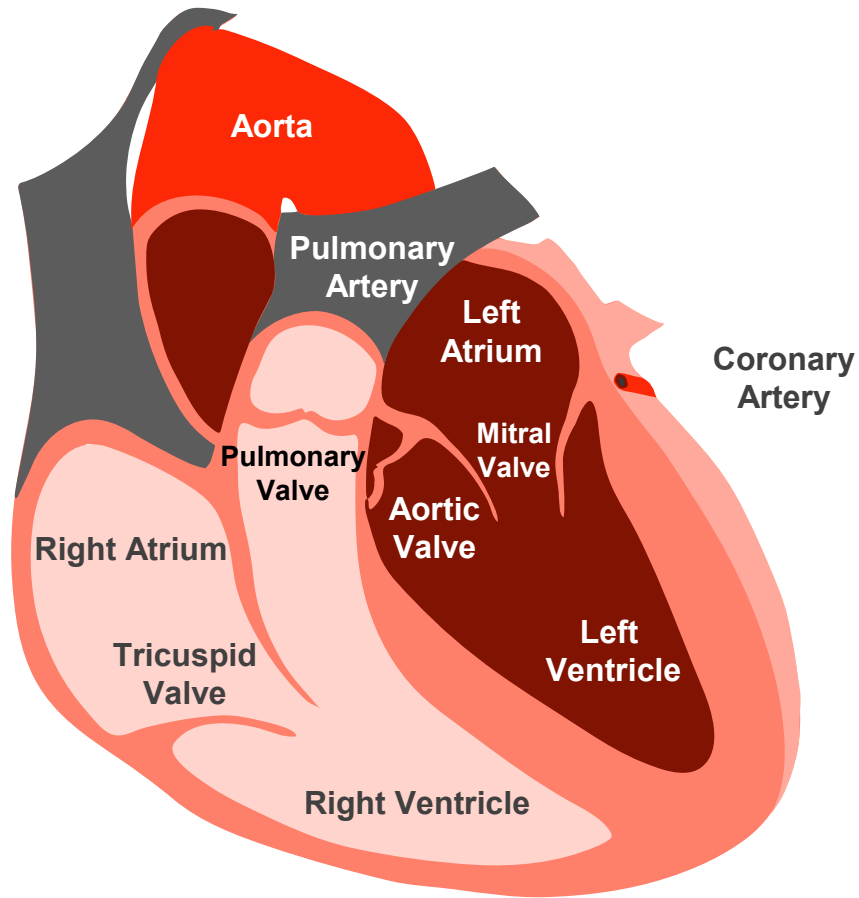
# AGENDA

- 01** DEFINIÇÃO DO PROBLEMA
- 02** CONJUNTO DE DADOS
- 03** PRÉ-PROCESSAMENTO
- 04** REDUÇÃO DA DIMENSIONALIDADE
- 05** MÉTODOS DE TREINAMENTO E VALIDAÇÃO
- 06** RESULTADOS E DISCUSSÕES
- 07** CONCLUSÕES
- 08** PRÓXIMOS PASSOS



# 01 DEFINIÇÃO DO PROBLEMA

# DEFINIÇÃO DO PROBLEMA



**Objetivo:** aplicar modelos de classificação e avaliar os resultados obtidos em um *dataset* composto por vetores de características extraídas de modelos 3D reconstruídos a partir de exames de Ressonância Magnética Cardíaca (RMC).

**Cardiomiopatia:** quadro clínico dado por uma anormalidade no miocárdio (estrutura muscular do coração) (KUMAR *et al.*, 2010 apud BERGAMASCO, 2018).

## MOTIVAÇÃO

Classificar modelos 3D provenientes de exames reais. Aplicar diferentes abordagens de classificação e contribuir com a linha de pesquisa apresentada por Bergamasco (2013, 2018).

## RMC

Cada amostra do *dataset* representa um modelo 3D do ventrículo esquerdo do coração do paciente, reconstruído a partir de fatias (*slices*) do exame de RMC, que pode sofrer ou não de algum tipo de Cardiomiopatia.

## FONTE DOS DADOS

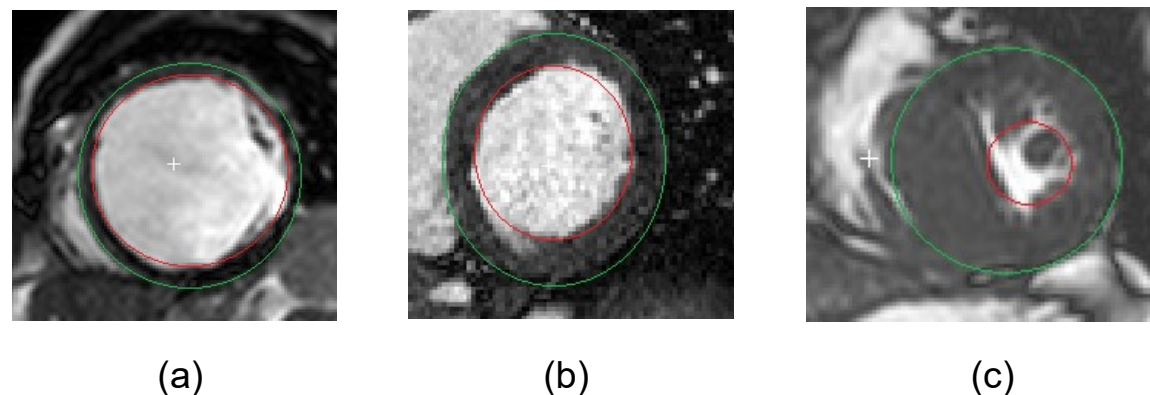
*Dataset* disponibilizado pelo Laboratório de Aplicações de Informática em Saúde (LApIS) - EACH/USP, coordenado pela Profa. Dra. Fátima L. S. Nunes.

# 02 CONJUNTO DE DADOS

# CONJUNTO DE DADOS

- › Composto por amostras (vetores de características) que representam/descrevem modelos cardíacos tridimensionais (modelos 3D).
- › Reconstruções realizadas a partir de imagens RMC, fornecidas e classificadas por médicos especialistas do Departamento de Ressonância Magnética e Tomografia Computadorizada Cardiovascular do Instituto do Coração (InCor) da Universidade de São Paulo (USP) (BERGAMASCO, 2018).
- › A Figura 1 apresentada fatias de exames de RMC de cada um dos tipos de amostras.

Figura 1 - Diferentes tipos de Cardiomiopatias – fatia 7 retirado de exames de RMC: (a) caso com Cardiomiopatia Dilatada (CMD); (b) caso sem anomalia; (c) caso com Cardiomiopatia Hipertrófica (CMH). Nota-se pelas imagens, que para CMD a região interna do ventrículo esquerdo é maior e a parede ventricular mais fina. O oposto ocorre para a anomalia de CMH que apresenta uma região interna menor e uma parede ventricular mais espessa.



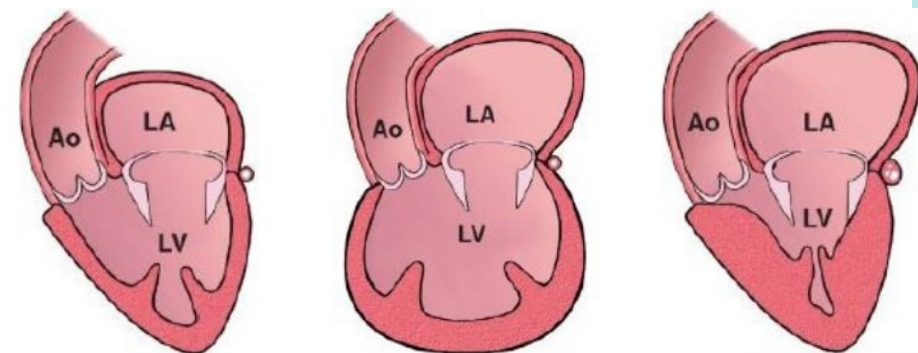
Fonte: Bergamasco (2018, p. 51)

# CONJUNTO DE DADOS

- › Classes originais (baseadas no estado de saúde do coração do paciente):
  - › **sem anomalia** (SAN);
  - › portador de **Cardiomiopatia Hipertrófica** (CMH); e,
  - › portador de **Cardiomiopatia Dilatada** (CMD).
- › Binarização
  - › **classificação par a par;**
  - › **foco principal: SAN X CMD.**
- › A Figura 2 apresenta ilustrações que representam secções de corações de cada um dos tipos de amostras descritos.
- › Quantidades de amostras (Total = 400):

Classes originais	Quantidade
SAN	101
CMH	183
CMD	116

Figura 2 - Ilustrações de secção do coração humano:  
(a) SAN; (b) CMD; (c) CMH.

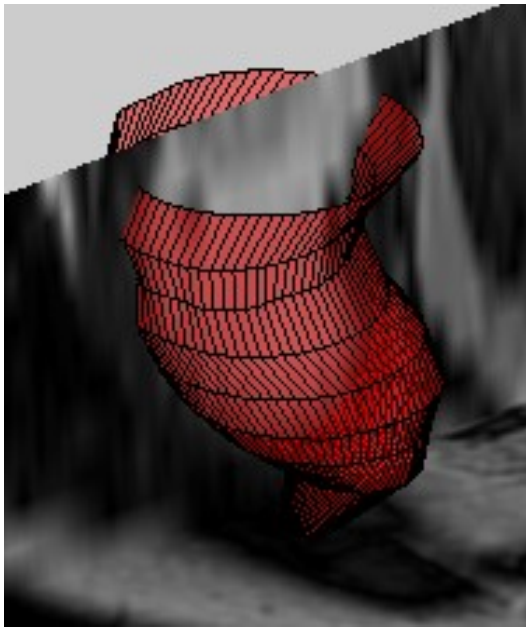


(a) (b) (c)

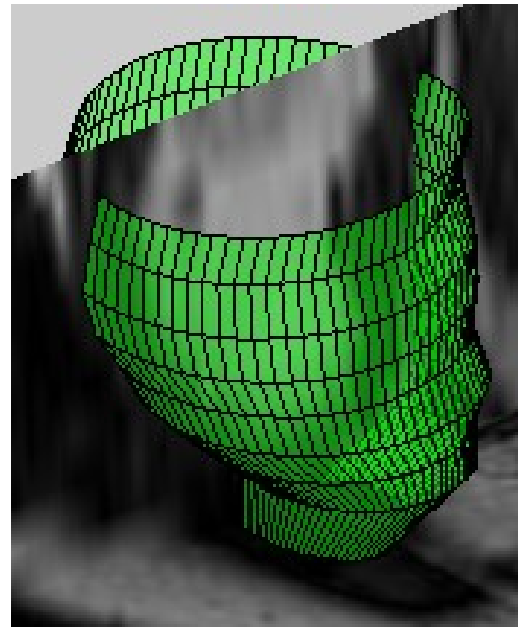
Fonte: adaptado de Bergamasco (2018, p. 49).

# CONJUNTO DE DADOS

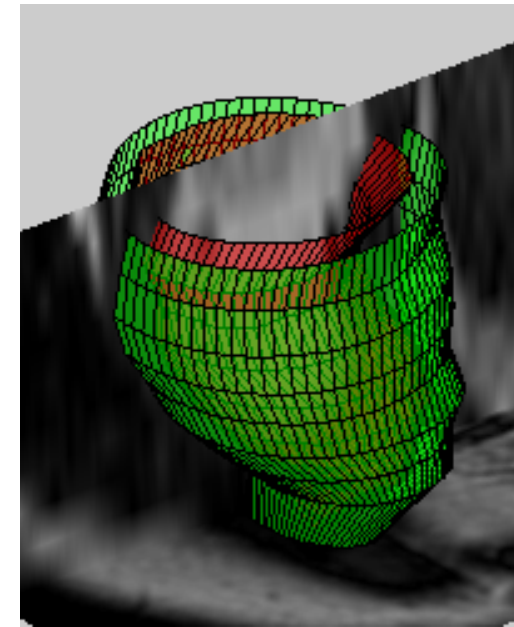
Figura 3 - Exemplo de reconstrução 3D das fatias: (a) reconstrução do endocárdio; (b) reconstrução do epicárdio; (c) reconstrução da parede ventricular.



(a)



(b)



(c)

Fonte: Bergamasco (2018, p. 101).



# CONJUNTO DE DADOS

CARACTERÍSTICA	TIPO	VALORES	OBSERVAÇÕES
Idade	<ul style="list-style-type: none"><li>Discreta</li><li>Racional</li></ul>	<ul style="list-style-type: none"><li>Unidade: anos.</li><li>Valor mínimo: 2.</li><li>Valor máximo: 87.</li><li>Qtde. de amostras com valor ausente: 5.</li></ul>	<ul style="list-style-type: none"><li>Intervalo de idades limites [0,100].</li></ul>
Sexo	<ul style="list-style-type: none"><li>Discreta</li><li>Nominal</li></ul>	<ul style="list-style-type: none"><li>Valores possíveis:<ul style="list-style-type: none"><li>Masculino (276 amostras);</li><li>Feminino (124 amostras).</li></ul></li><li>Qtde. de amostras com valor ausente: 0.</li></ul>	<ul style="list-style-type: none"><li>0 = Homem</li><li>1 = Mulher</li></ul>
Volume (ventrículo)	<ul style="list-style-type: none"><li>Discreta</li><li>Racional</li></ul>	<ul style="list-style-type: none"><li>Unidade: mililitros</li><li>Qtde. de amostras com valor ausente: 9</li></ul>	
Massa (ventrículo)	<ul style="list-style-type: none"><li>Discreta</li><li>Racional</li></ul>	<ul style="list-style-type: none"><li>Unidade: gramas</li><li>Qtde. de amostras com valor ausente: 9</li></ul>	
Extrator de Distâncias Local (EDL)	<ul style="list-style-type: none"><li>Contínua</li><li>Racional</li><li>Multivalorada</li></ul>	<ul style="list-style-type: none"><li>Unidade: voxels</li><li>Modelo 3D dividido em octantes</li><li>Quantidade de valores por amostra: 8</li><li>Características ordenadas</li></ul>	
Harmônicos Esféricos (SPHARMs)	<ul style="list-style-type: none"><li>Contínua</li><li>Racional</li><li>Multivalorada</li></ul>	<ul style="list-style-type: none"><li>Valores negativos, zero e positivos.</li><li>Quantidade de valores (coeficientes) por amostra: de 500 a 1000.</li></ul>	<ul style="list-style-type: none"><li>Cálculo de um coeficiente global (soma dos coeficientes).</li></ul>
Transformada de Hough	<ul style="list-style-type: none"><li>Contínua</li><li>Racional</li><li>Multivalorada</li></ul>	<ul style="list-style-type: none"><li>Quantidade de valores por amostra: 30.</li><li>Características ordenadas.</li></ul>	

# 03 PRÉ-PROCESSAMENTO

# PRÉ-PROCESSAMENTO

## BALANCEAMENTO DAS CLASSES

- › Classes originais:

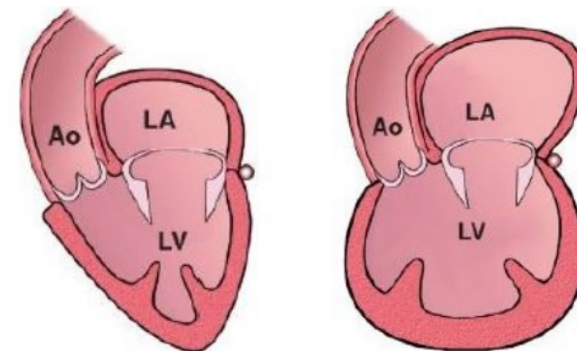
Classes originais	Quantidade de Amostras
<b>SAN</b>	<b>101</b>
CMH	183
<b>CMD</b>	<b>116</b>

- › Estratégia adotada (inicialmente):

### ***SAN versus CMD***

- › Sem necessidade, a priori, de aplicação de técnicas de balanceamento.

Figura 4 - Ilustrações de secção do coração humano:  
(a) SAN; (b) CMD.



(a)

(b)

Fonte: adaptado de Bergamasco (2018, p. 49).

# PRÉ-PROCESSAMENTO

## AMOSTRAS COM VALORES AUSENTES

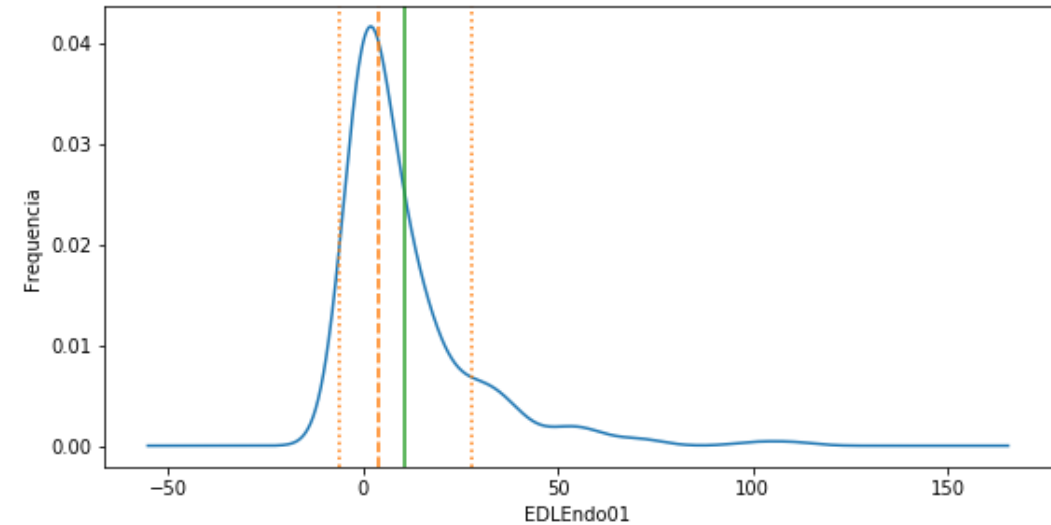
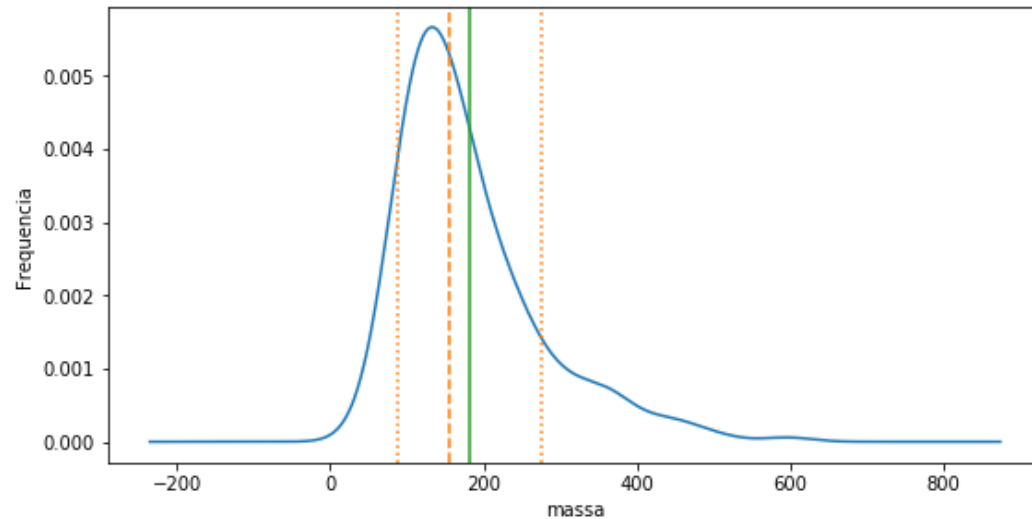
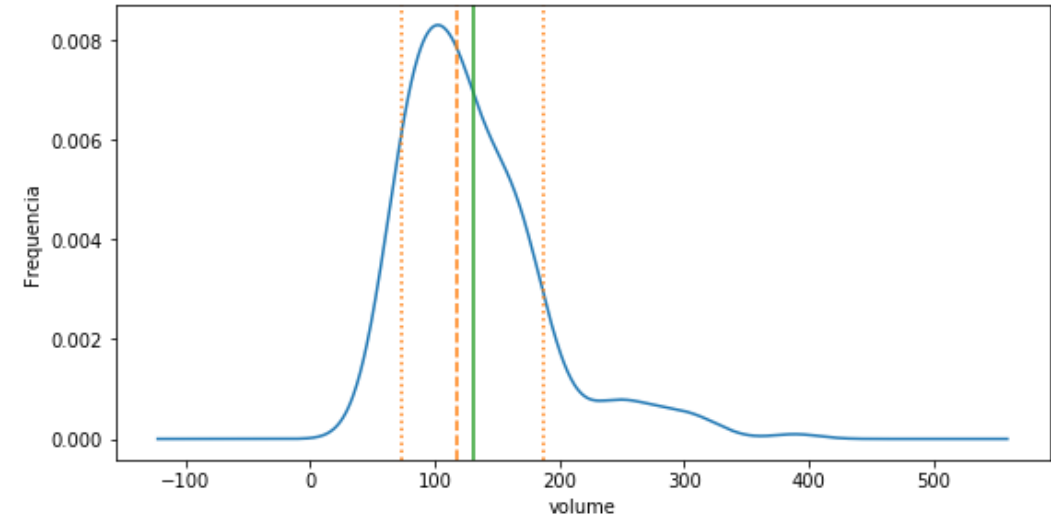
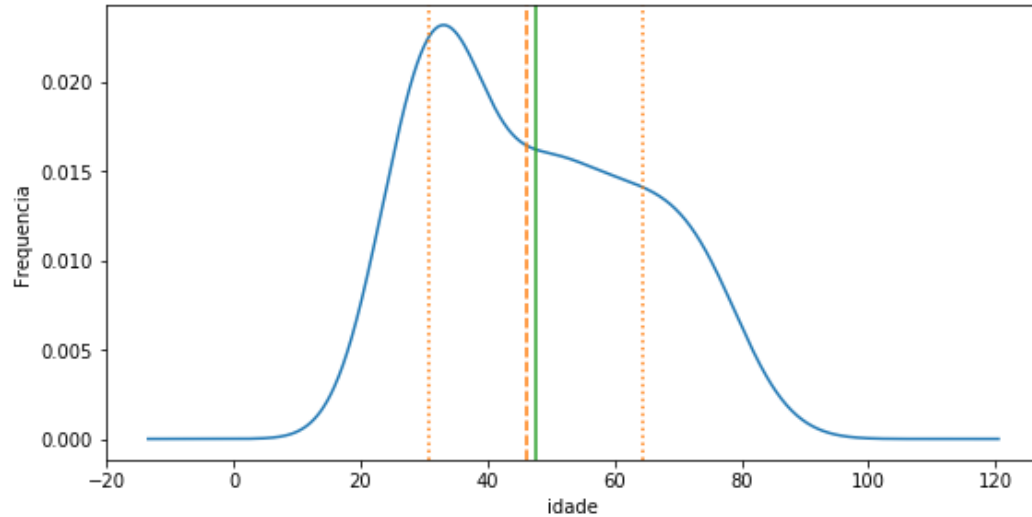
- › Idade (5), massa (4) e volume (4): poucas amostras apresentam valores ausentes.
  - optamos por preencher valores ausentes com a média encontrada na base.

## NORMALIZAÇÃO

- › Aplicação da classe **StandardScaler** da biblioteca **scikit-learn** (SCIKIT-LEARN, 2020d).
- › Normalização do valor  $x$  de uma característica da amostra  $\rightarrow z = \frac{x-\mu}{\sigma}$ , em que:
  - ›  $\mu$  é a média da característica para as amostras do *dataset*;
  - ›  $\sigma$  é o desvio-padrão da característica para as amostras do *dataset*.

# PRÉ-PROCESSAMENTO

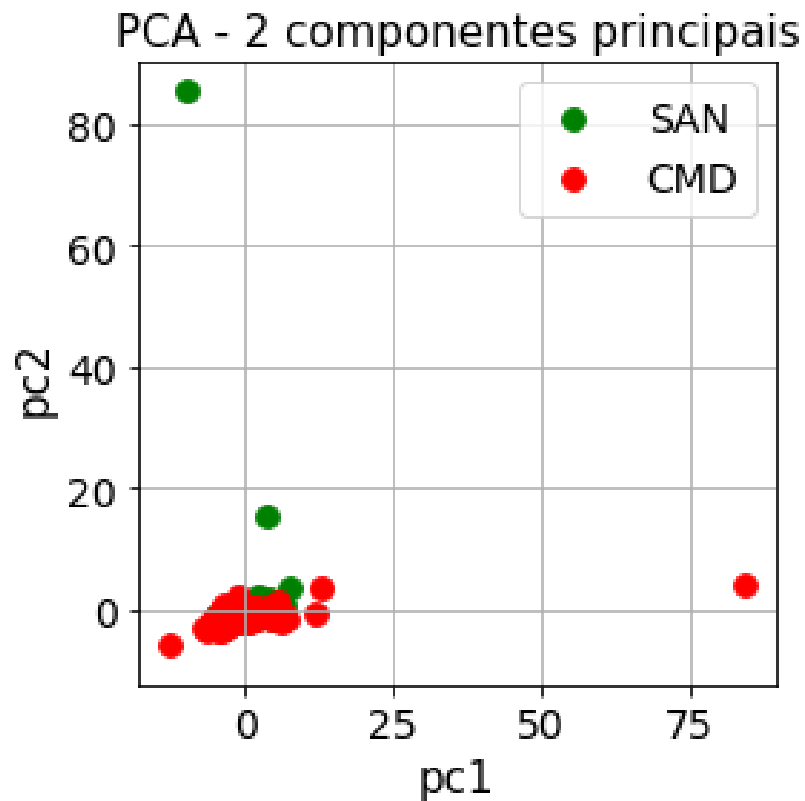
## DISTRIBUIÇÕES DAS CARACTERÍSTICAS - ANÁLISES PRELIMINARES



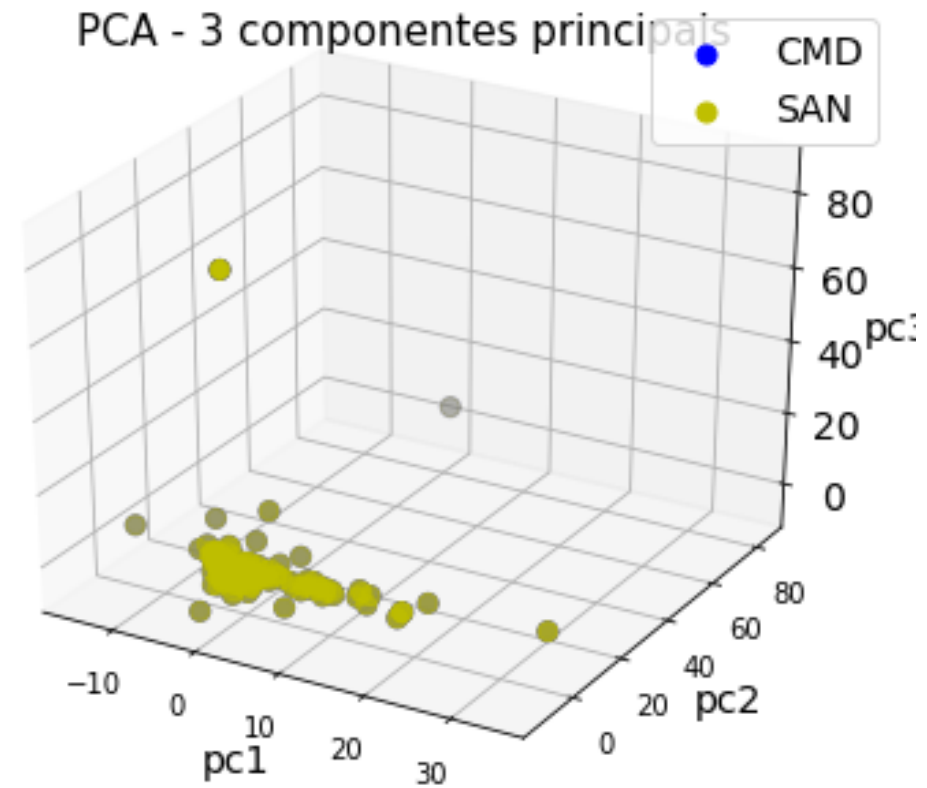
# 04 REDUÇÃO DA DIMENSIONALIDADE

# **ANÁLISE DE COMPONENTES PRINCIPAIS (PCA)**

# PCA



23,89%



32,46%

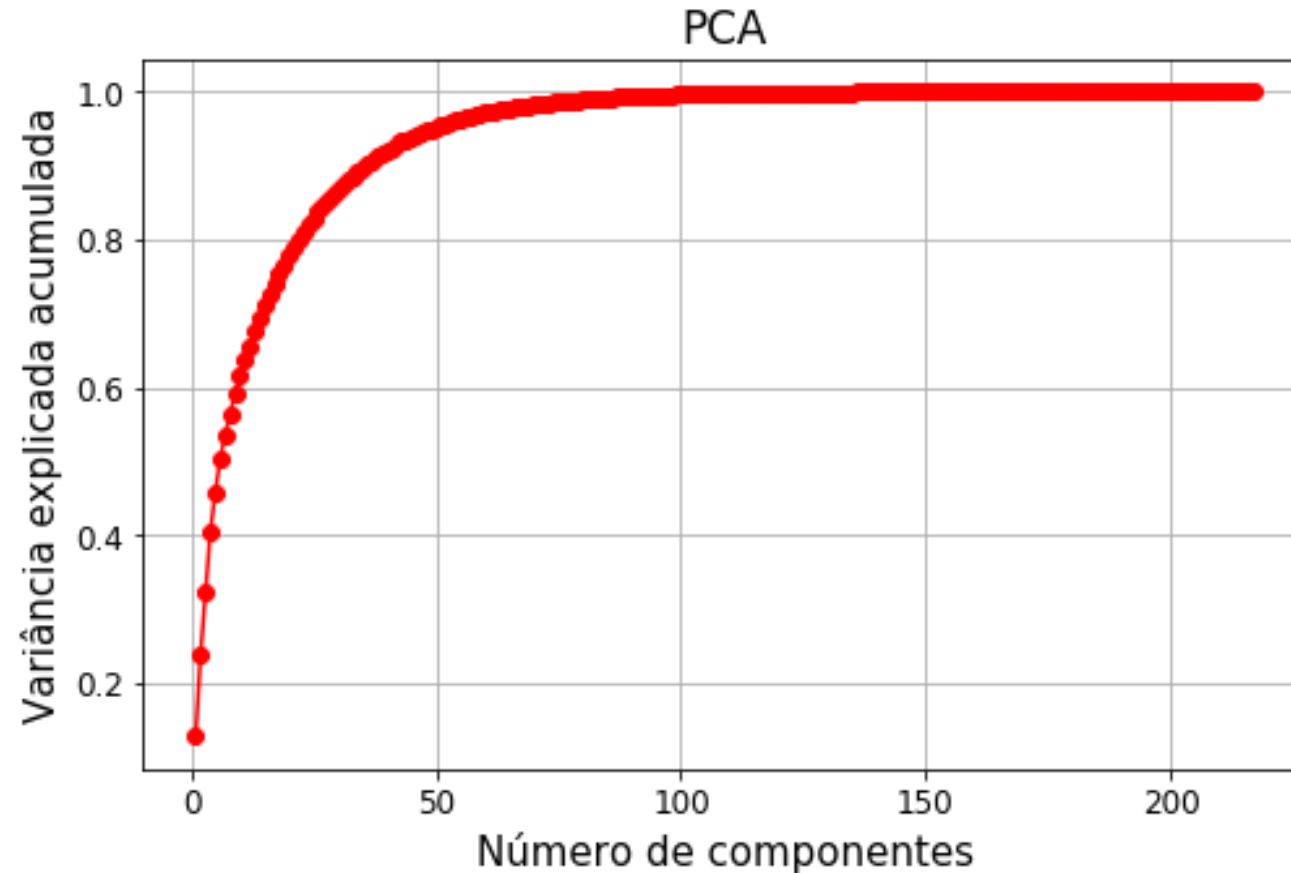
Variância Explicada Acumulada

- › Em tais visões, não é possível se separar claramente as classes.



# PCA

## VARIÂNCIA EXPLICADA ACUMULADA



- › Com 171 Componentes Principais (PC's), atinge-se 100% de Variância Explicada Acumulada.

# PCA

## VARIÂNCIA EXPLICADA ACUMULADA PARA A DEFINIÇÃO DA QUANTIDADE DE PC'S

Qtde. de PC's	Variância Explicada Acumulada	Melhores Métricas – NB		
		Acurácia	Precisão	Revocação
30	86,73	60,89	74,06	39,77
35	89,70	60,43	71,17	40,68
40	92,12	59,05	68,45	38,94
45	93,89	60,43	72,95	40,61
50	95,24	60,89	72,60	42,35
<b>53</b>	<b>95,93</b>	<b>62,77</b>	<b>74,19</b>	<b>45,91</b>
55	96,33	61,84	71,98	45,91
60	97,13	61,36	71,00	45,91
<b>75</b>	<b>98,63</b>	<b>62,77</b>	<b>69,88</b>	<b>52,05</b>
171	100,00	59,55	62,83	49,47

# PCA

## VARIÂNCIA EXPLICADA ACUMULADA PARA A DEFINIÇÃO DA QUANTIDADE DE PC'S

Qtde. de PC's	Variância Explicada Acumulada	Melhores Métricas – SVM		
		Acurácia	Precisão	Revocação
30	86,73	79,35	83,61	76,97
35	89,70	83,40	87,46	82,80
40	92,12	83,46	91,64	76,82
45	93,89	83,03	89,75	77,73
50	95,24	84,42	85,90	85,45
53	95,93	85,74	90,58	82,88
55	96,33	83,94	82,38	90,45
60	97,13	84,76	84,11	88,71
<b>75</b>	<b>98,63</b>	<b>87,08</b>	<b>91,43</b>	<b>84,47</b>
171	100,00	73,20	74,83	76,59

# PCA

## VARIÂNCIA EXPLICADA ACUMULADA PARA A DEFINIÇÃO DA QUANTIDADE DE PC'S

Qtde. de PC's	Variância Explicada Acumulada	Melhores Métricas – MLP		
		Acurácia	Precisão	Revocação
30	86,73	82,90	84,19	86,21
35	89,70	84,78	91,21	80,15
<b>40</b>	<b>92,12</b>	<b>85,28</b>	<b>88,99</b>	<b>83,56</b>
45	93,89	85,24	87,45	85,23
50	95,24	84,78	90,37	81,06
53	95,93	83,42	84,46	85,23
55	96,33	82,94	87,40	80,00
60	97,13	81,99	86,71	79,17
75	98,63	81,00	85,57	77,42
171	100,00	58,07	61,55	59,39

# PCA

## PCA FEATURE IMPORTANCES

	feature	importancia
0	idade	12.85
1	sexo	11.03
2	volume	8.58
3	massa	8.11
4	EDLEndo01	5.24
5	EDLEndo02	4.54
6	EDLEndo03	3.26
7	EDLEndo04	2.86
8	EDLEndo05	2.72
9	EDLEndo06	2.42
10	EDLEndo07	2.09
11	EDLEndo08	1.97
12	EDLEpi01	1.86
13	EDLEpi02	1.78
14	EDLEpi03	1.73
15	EDLEpi04	1.52

	feature	importancia
16	EDLEpi05	1.47
17	EDLEpi06	1.30
18	EDLEpi07	1.27
19	EDLEpi08	1.17
20	EDLTotal01	1.11
21	EDLTotal02	1.09
22	EDLTotal03	1.04
23	EDLTotal04	1.00
24	EDLTotal05	0.95
25	EDLTotal06	0.83
26	EDLTotal07	0.79
27	EDLTotal08	0.74
28	HoughEndo	0.73
29	HoughEpi	0.69
30	HoughParede	0.65

	feature	importancia
31	OctEndoL0001	0.61
32	OctEndoL0002	0.60
33	OctEndoL0003	0.56
34	OctEndoL0004	0.55
35	OctEndoL0005	0.52
36	OctEndoL0006	0.50
37	OctEndoL0007	0.49
38	OctEndoL0008	0.48
39	OctEndoL0101	0.42
40	OctEndoL0102	0.40
41	OctEndoL0103	0.38
42	OctEndoL0104	0.36
43	OctEndoL0105	0.32
44	OctEndoL0106	0.31
45	OctEndoL0107	0.30

	feature	importancia
46	OctEndoL0108	0.29
47	OctEndoL0201	0.26
48	OctEndoL0202	0.25
49	OctEndoL0203	0.25
50	OctEndoL0204	0.24
51	OctEndoL0205	0.23
52	OctEndoL0206	0.22

# SELEÇÃO DE CARACTERÍSTICAS

# RANDOM FOREST FEATURE IMPORTANCE

- › **Tipo: embutido.**
- › Determina a importância de cada feature em relação à classe alvo.
- › Em seguida, selecionamos as características manualmente.

feature importância			feature importância			feature importância			feature importância		
0	idade	0.061532	15	OctEpiL1005	0.006870	30	OctParedeL0003	0.004602	45	TotalEpiSpharmL04	0.004055
1	massa	0.036820	16	OctEndoL0001	0.006640	31	OctEndoL0108	0.004584	46	OctEpiL0005	0.004048
2	OctParedeL0007	0.034217	17	OctEndoL0308	0.006627	32	OctEpiL0602	0.004548	47	OctEndoL1006	0.004037
3	HoughEndo	0.030190	18	OctEndoL0104	0.006123	33	OctParedeL0206	0.004543	48	TotalEndoSpharmL05	0.004018
4	volume	0.026472	19	OctEpiL0304	0.005803	34	OctParedeL0204	0.004515	49	OctParedeL0005	0.003990
5	OctParedeL0107	0.024630	20	OctParedeL0807	0.005659	35	EDLEpi07	0.004437	50	OctParedeL0907	0.003986
6	OctEpiL0107	0.023891	21	OctEndoL0005	0.005538	36	OctParedeL0103	0.004426	51	OctEpiL0704	0.003964
7	HoughEpi	0.016552	22	OctEndoL0103	0.005499	37	OctEpiL0104	0.004426	52	OctEpiL0402	0.003904
8	OctEndoL0107	0.016287	23	OctParedeL0503	0.005160	38	talParedeSpharmL09	0.004295			
9	HoughParede	0.016213	24	OctEndoL0207	0.004985	39	OctParedeL0902	0.004289			
10	OctEpiL0007	0.014885	25	OctEpiL0303	0.004944	40	OctEpiL0302	0.004250			
11	OctEndoL0007	0.011996	26	OctEndoL0302	0.004909	41	OctEpiL0106	0.004210			
12	EDLEndo07	0.010936	27	OctEpiL0406	0.004831	42	OctEpiL1004	0.004202			
13	OctEndoL0006	0.008245	28	OctParedeL0307	0.004736	43	OctParedeL1006	0.004132			
14	OctEpiL0006	0.007046	29	OctEndoL0003	0.004731	44	TotalEpiSpharmL00	0.004080			

# SELECTKBEST

- › **Tipo: filtro.**
- › Seleciona as melhores *features* com base em testes estatísticos univariados.
- › Mantém as *features* de pontuação mais alta.
- › Método para classificação utilizado: ***f\_classif* (F-Score).**

SELECTKBEST		SELECTKBEST		SELECTKBEST		SELECTKBEST	
0	idade	15	OctEndoL0107	30	OctEpiL0303	45	OctParedeL0505
1	sexo	16	OctEndoL0108	31	OctEpiL0405	46	OctParedeL0605
2	volume	17	OctEndoL0201	32	OctEpiL0605	47	OctParedeL0706
3	massa	18	OctEndoL0605	33	OctParedeL0003	48	TotalEndoSpharmL00
4	EDLEndo07	19	OctEpiL0003	34	OctParedeL0006	49	TotalEndoSpharmL01
5	EDLEpi07	20	OctEpiL0006	35	OctParedeL0007	50	TotalEpiSpharmL01
6	EDLTotal07	21	OctEpiL0007	36	OctParedeL0008	51	TotalParedeSpharmL00
7	HoughEndo	22	OctEpiL0008	37	OctParedeL0103	52	TotalParedeSpharmL01
8	HoughEpi	23	OctEpiL0103	38	OctParedeL0104		
9	HoughParede	24	OctEpiL0104	39	OctParedeL0105		
10	OctEndoL0003	25	OctEpiL0105	40	OctParedeL0107		
11	OctEndoL0006	26	OctEpiL0106	41	OctParedeL0108		
12	OctEndoL0007	27	OctEpiL0107	42	OctParedeL0201		
13	OctEndoL0103	28	OctEpiL0108	43	OctParedeL0306		
14	OctEndoL0104	29	OctEpiL0201	44	OctParedeL0405		



# RFE - RECURSIVE FEATURE ELIMINATION

- › **Tipo: camada.**
- › Selecciona subconjuntos recursivamente, considerando conjuntos cada vez menores de *features*.

RFE		RFE		RFE		RFE	
0	OctEndoL0001	15	OctEpiL0207	30	OctEpiL0907	45	OctParedelL1007
1	OctEndoL0007	16	OctEpiL0303	31	OctEpiL1007	46	TotalEndoSpharmL00
2	OctEndoL0103	17	OctEpiL0306	32	OctParedelL0007	47	TotalEndoSpharmL06
3	OctEndoL0106	18	OctEpiL0307	33	OctParedelL0008	48	TotalEndoSpharmL08
4	OctEndoL0107	19	OctEpiL0308	34	OctParedelL0105	49	TotalEpiSpharmL00
5	OctEndoL0108	20	OctEpiL0403	35	OctParedelL0107	50	TotalEpiSpharmL06
6	OctEndoL0202	21	OctEpiL0503	36	OctParedelL0108	51	TotalParedelSpharmL00
7	OctEndoL0206	22	OctEpiL0507	37	OctParedelL0201	52	TotalParedelSpharmL04
8	OctEndoL0303	23	OctEpiL0508	38	OctParedelL0202		
9	OctEndoL0308	24	OctEpiL0702	39	OctParedelL0203		
10	OctEndoL0407	25	OctEpiL0703	40	OctParedelL0207		
11	OctEndoL0808	26	OctEpiL0704	41	OctParedelL0208		
12	OctEndoL1003	27	OctEpiL0707	42	OctParedelL0308		
13	OctEpiL0007	28	OctEpiL0804	43	OctParedelL0604		
14	OctEpiL0107	29	OctEpiL0807	44	OctParedelL0608		

# SFS - SEQUENTIAL FORWARD SELECTION

- › **Tipo: heurístico.**
- › A cada iteração, novas características são adicionadas ao subconjunto corrente.

SFS_15_CV10		SFS_15_CV10		SFS_15_CV10		SFS_15_CV10	
0	idade	15	EDLEpi04	30	HoughParede	45	OctEpiL0005
1	sexo	16	EDLEpi05	31	OctEndoL0001	46	OctEpiL0006
2	volume	17	EDLEpi06	32	OctEndoL0002	47	OctEpiL0106
3	massa	18	EDLEpi07	33	OctEndoL0003	48	OctEpiL0107
4	EDLEndo01	19	EDLEpi08	34	OctEndoL0004	49	OctParedeL0007
5	EDLEndo02	20	EDLTotal01	35	OctEndoL0005	50	OctParedeL0107
6	EDLEndo03	21	EDLTotal02	36	OctEndoL0006	51	OctParedeL0307
7	EDLEndo04	22	EDLTotal03	37	OctEndoL0007	52	TotalEndoSpharmL00
8	EDLEndo05	23	EDLTotal04	38	OctEndoL0008		
9	EDLEndo06	24	EDLTotal05	39	OctEndoL0102		
10	EDLEndo07	25	EDLTotal06	40	OctEndoL0103		
11	EDLEndo08	26	EDLTotal07	41	OctEpiL0001		
12	EDLEpi01	27	EDLTotal08	42	OctEpiL0002		
13	EDLEpi02	28	HoughEndo	43	OctEpiL0003		
14	EDLEpi03	29	HoughEpi	44	OctEpiL0004		

Simulações do SFS para 50 *features*:

k-Folds	Qtde. Vizinhos	Acurácia (kNN)
10	5	79,26
	15	80,67
	30	74,61
	90	73,24
	100	73,24
	180	74,71

# 05 MÉTODOS DE TREINAMENTO E VALIDAÇÃO

# NAIVE BAYES

## IMPLEMENTAÇÃO

- › `sklearn.naive_bayes.GaussianNB`.



Fonte: SCIKIT-LEARN (2020c, n.p.).

## CALIBRAÇÃO

- › Parâmetro **var\_smoothing** ( $\alpha$ ) → suavização das curvas de probabilidades para minimizar as chances de o classificador não conseguir fazer previsões para cenários não contemplados pelo conjunto de treinamento.
- › Variação de **0.0** (sem suavização) até **1.0** (inclusive), em intervalos de **0.1**.

# SVM

## IMPLEMENTAÇÃO

› sklearn.svm.SVC.

## CALIBRAÇÃO

- › **kernel** → [ 'rbf', 'linear', 'poly', 'sigmoid' ]
- › **C** → [ 0.0, 0.001, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0, 1.5, 10, 20 ]
- › **gamma** → [ 'auto', 'scale', 0.001, 0.0016, 0.0017, 0.0018, 0.0019, 0.002, 0.003, 0.005, 0.01, 0.02, 0.03, 0.04, 0.5, 1, 10 ]

sklearn.svm.SVC

```
class sklearn.svm.SVC(*, C=1.0, kernel='rbf', degree=3, gamma='scale', coef0=0.0,
shrinking=True, probability=False, tol=0.001, cache_size=200, class_weight=None, verbose=False,
max_iter=-1, decision_function_shape='ovr', break_ties=False, random_state=None) \[source\]
```

Fonte: SCIKIT-LEARN (2020f, n.p.).

auto = 1 / n

scale = 1 / (n \* var(X))

# MLP

## IMPLEMENTAÇÃO

› **sklearn.neural\_network.MLPClassifier.**

**sklearn.neural\_network.MLPClassifier**

```
class sklearn.neural_network.MLPClassifier(hidden_layer_sizes=(100,), activation='relu', *, solver='adam', alpha=0.0001, batch_size='auto', learning_rate='constant', learning_rate_init=0.001, power_t=0.5, max_iter=200, shuffle=True, random_state=None, tol=0.0001, verbose=False, warm_start=False, momentum=0.9, nesterovs_momentum=True, early_stopping=False, validation_fraction=0.1, beta_1=0.9, beta_2=0.999, epsilon=1e-08, n_iter_no_change=10, max_fun=15000)
```

[\[source\]](#)

Fonte: SCIKIT-LEARN (2020d, n.p.).

## CALIBRAÇÃO

- › **activation** → [ 'identity', 'logistic', 'tanh', 'relu' ]
- › **hidden\_layer\_sizes** → [ 8, 18, 36, 53, 328 ]
- › **learning\_rate\_init** → [ 0.01, 0.05, 0.1 ]
- › **max\_iter** → [ 30, 60, 90, 120, 150 ]

# RANDOM FOREST

## IMPLEMENTAÇÃO

› `sklearn.ensemble.RandomForestClassifier`.

## CALIBRAÇÃO

- › **n\_estimators** → [ 100, 250, 500, 750, 1000, 1250, 1500 ]
- › **max\_features** → [ 'sqrt', 'log2', 8, 18, 36, 53, 328 ]
- › **criterion** → [ 'gini', 'entropy' ]
- › **oob\_score** → [ False, True ]
- › **ccp\_alpha** → [ 0.0, 0.02, 0.04, 0.08, 0.16, 0.32, 0.64, 1.28 ]

### 3.2.4.3.1. `sklearn.ensemble.RandomForestClassifier`

```
class sklearn.ensemble.RandomForestClassifier(n_estimators=100, *, criterion='gini', max_depth=None, min_samples_split=2, min_samples_leaf=1, min_weight_fraction_leaf=0.0, max_features='auto', max_leaf_nodes=None, min_impurity_decrease=0.0, min_impurity_split=None, bootstrap=True, oob_score=False, n_jobs=None, random_state=None, verbose=0, warm_start=False, class_weight=None, ccp_alpha=0.0, max_samples=None) [source]
```

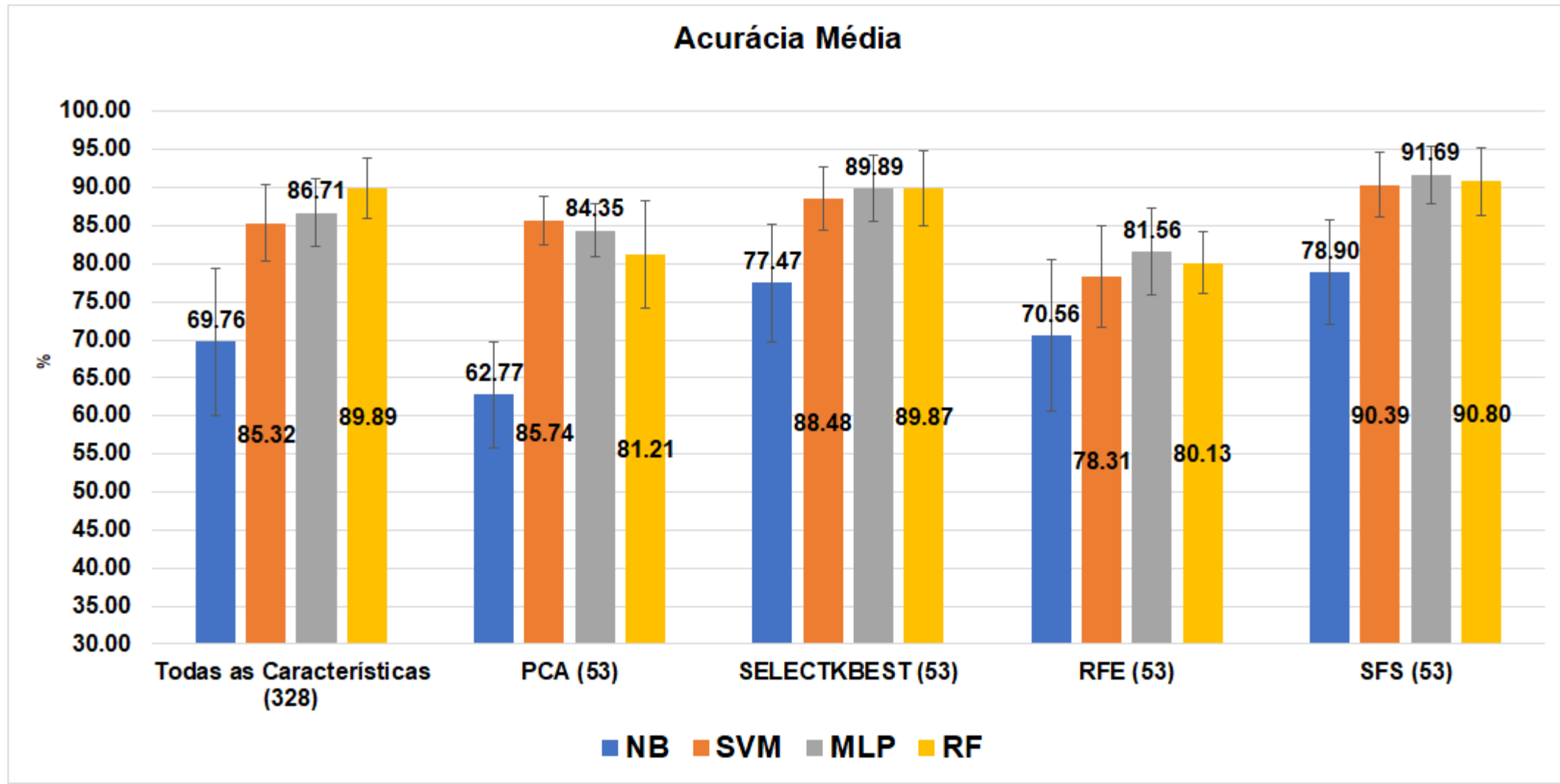
Fonte: SCIKIT-LEARN (2020a, n.p.).

# 06 RESULTADOS E DISCUSSÕES



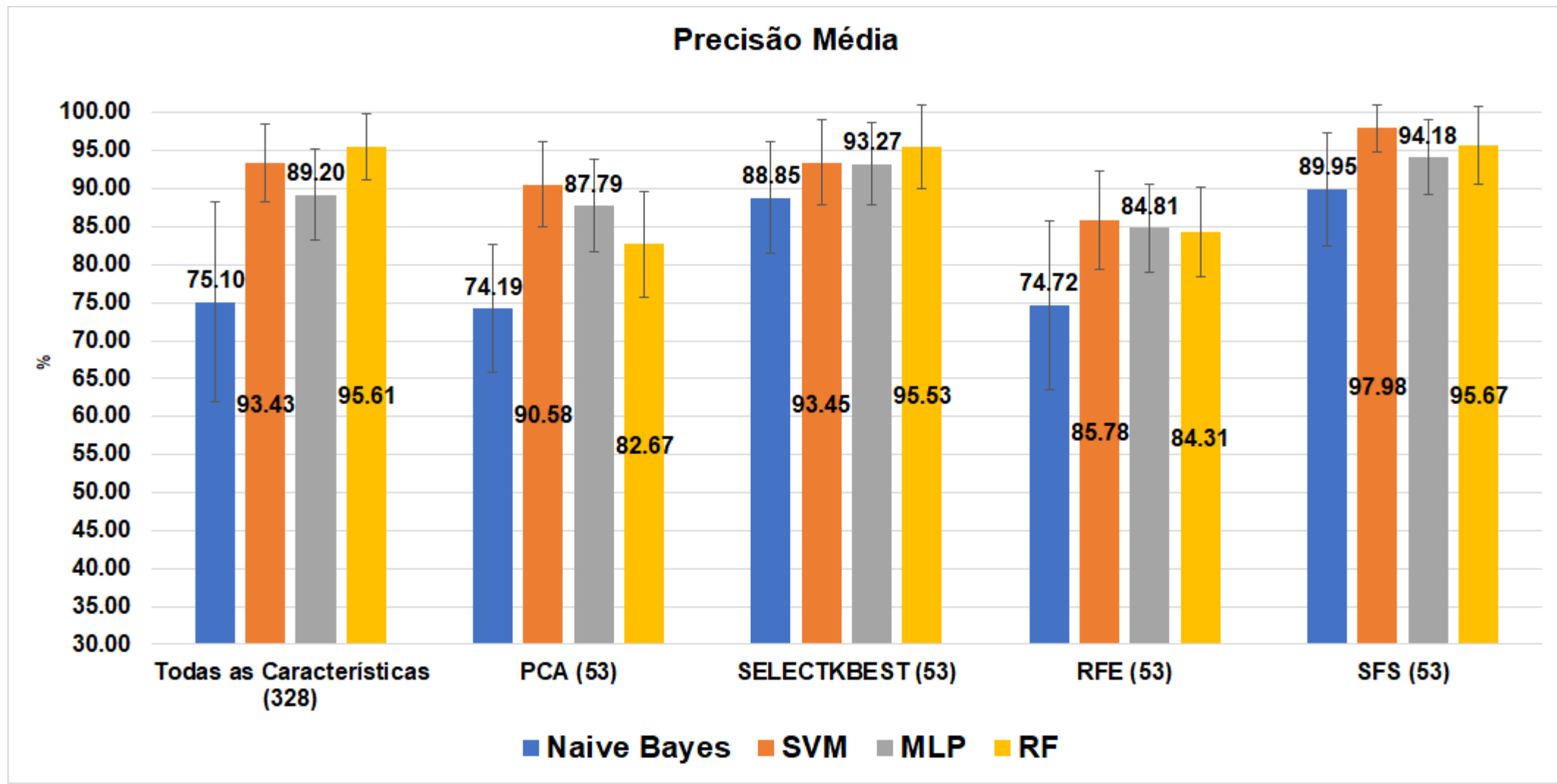
# RESULTADOS E DISCUSSÕES

Melhores resultados de Acurácia Média obtidos com os classificadores NB, SVM, MLP e RF.



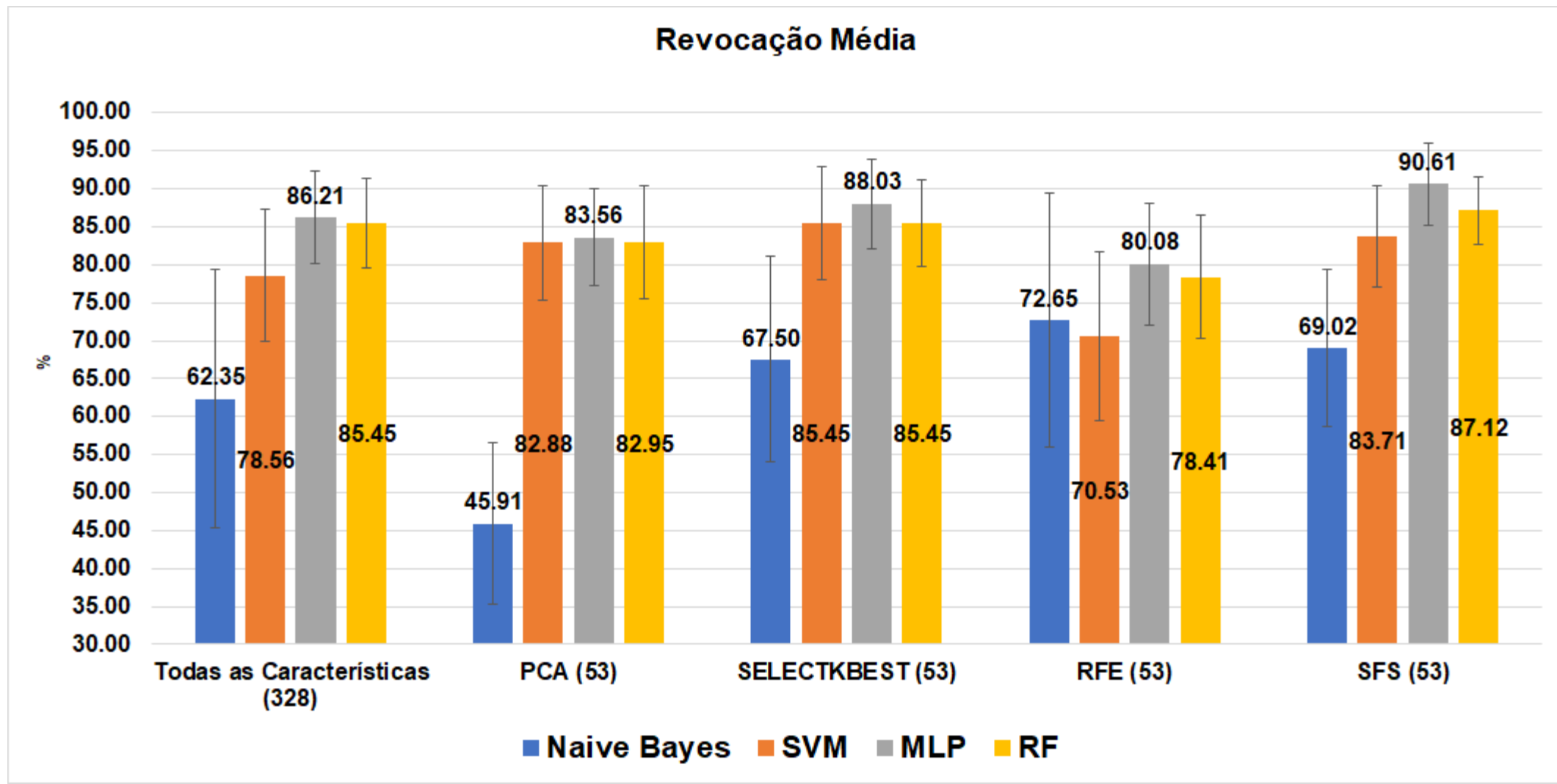
# RESULTADOS E DISCUSSÕES

Melhores resultados de Precisão Média obtidos com os classificadores NB, SVM, MLP e RF.



# RESULTADOS E DISCUSSÕES

Melhores resultados de Revocação Média obtidos com os classificadores NB, SVM, MLP e RF.



# RESULTADOS E DISCUSSÕES

## ANÁLISE DOS CONJUNTOS DE *FEATURES*

		Conjunto de Features				
		Completo	PCA Imp.	SFS	RFE	KBEST
Grupos de Features	Idade	1	1	1	0	1
	Sexo	1	1	1	0	1
	Volume	1	1	1	0	1
	Massa	1	1	1	0	1
	EDL	24	24	24	0	3
	HoughEndo	1	1	1	0	1
	HoughEpi	1	1	1	0	1
	HoughParede	1	1	1	0	1
	SpharmsOcts	264	22	10	8	8
	SpharmsTotais	33	0	0	0	0
Qtde. Features		328	53	41	8	18

# RESULTADOS E DISCUSSÕES

## ANÁLISE DOS CONJUNTOS DE *FEATURES*

Acurácia de classificação  
para o conjunto PCA  
Importances

Naive Bayes: 77,49%

SVM: 89,42%

MLP: 89,46%

Random Forest: 90,35%

		Conjunto de Features				
		PCA Imp.	SFS	RFE	KBEST	
Grupos de Features	Idade	1	1	1	0	1
	Sexo	1	1	1	0	1
	Volume	1	1	1	0	1
	Massa	1	1	1	0	1
	EDL	24	24	24	0	3
	HoughEndo	1	1	1	0	1
	HoughEpi	1	1	1	0	1
	HoughParede	1	1	1	0	1
	SpharmsOcts	264	22	10	8	8
	SpharmsTotais	33	0	0	0	0
Qtde. Features		328	53	41	8	18

# RESULTADOS E DISCUSSÕES

## ANÁLISE DOS CONJUNTOS DE *FEATURES*

		Conjunto de Features				
		Completo	RF Imp.	SFS	RFE	KBEST
Grupos de Features	Idade	1	1	1	0	1
	Sexo	1	0	0	0	0
	Volume	1	1	1	0	1
	Massa	1	1	1	0	1
	EDL	24	2	2	0	2
	HoughEndo	1	1	1	0	1
	HoughEpi	1	1	1	0	1
	HoughParede	1	1	1	0	1
	SpharmsOcts	264	41	13	12	17
	SpharmsTotais	33	4	0	1	0
Qtde. Features		328	53	21	13	25

# RESULTADOS E DISCUSSÕES

## ANÁLISE DOS CONJUNTOS DE *FEATURES*

Acurácia de classificação  
para o conjunto RF  
Importances

Naive Bayes: 77,94%

SVM: 88,51%

MLP: 87,10%

Random Forest: 92,19%

		Conjunto de Features			
		RF Imp.	SFS	RFE	KBEST
Grupos de Features	Idade	1	1	0	1
	Sexo	1	0	0	0
	Volume	1	1	0	1
	Massa	1	1	0	1
	EDL	24	2	0	2
	HoughEndo	1	1	0	1
	HoughEpi	1	1	0	1
	HoughParede	1	1	0	1
	SpharmsOcts	264	41	12	17
	SpharmsTotais	33	4	1	0
Qtde. Features		328	53	13	25

# RESULTADOS E DISCUSSÕES

## COMBINAÇÃO ENTRE *RANKINGS*

**Acurácia Média de classificação para a combinação entre as *features* dos *rankigs* PCA e RF importances:**

**Naive Bayes: 76,13%**

**SVM: 88,98%**

**MLP: 89,03%**

**Random Forest: 91,28%**



# 07 CONCLUSÕES

# CONCLUSÕES

- › Os resultados do trabalho demonstram que, para o problema apresentado, é possível se obter resultados significativos com as técnicas de Aprendizado Supervisionado experimentadas. Chegou-se a superar 90% de acurácia, com parametrizações que evidenciam a não ocorrência de *overfitting*.
- › O conjunto de *features* obtido por meio do selecionador SFS foi o que apresentou melhores resultados de classificação. As evidências são de que sua natureza heurística levou à seleção das *features* que mais contribuem para a distinção entre as classes, conforme o evidenciado por meio da análise comparativa com os *rankings* PCA importances e RF importances.
- › Por outro lado, os resultados menos expressivos se mostraram para o conjunto de *features* obtido por meio do selecionador RFE. Analisando-se tal conjunto, tomando-se também como referência os referidos *rankings*, observou-se que o RFE selecionou as características menos importantes para a distinção entre as classes, o que pode se justificar pela sua natureza gulosa com o descarte de *features* importantes para o problema, sem possibilidade de reconsideração.

# CONCLUSÕES

- › A hipótese para os resultados menos expressivos observados para os conjuntos formados por todas as características e pelas componentes principais (comparando-se com o SFS e o SELECTKBEST) se baseia na existência de *features* que pouco contribuem para a distinção entre as classes, mais atrapalhando do que ajudando na construção do modelo.
- › O classificador Naive Bayes, na maior parte das métricas, demonstrou resultados significativamente inferiores aos demais. A hipótese para essa observação se baseia no fato de que o Naive Bayes considera as *features* de forma independente para a estimação das probabilidades e resolução do problema. Além disso, a quantidade de amostras pode não ter sido suficiente para cobrir todas as possibilidades/combinções de classificação possíveis.

# CONCLUSÕES

- › De forma geral, os classificadores SVM (Ac. 90,39% - SFS), Random Forest (Ac. 90,80% - SFS) e MLP (Ac. 91,69% - SFS) viabilizaram resultados significativos (bons!) e próximos entre si.
- › Considerando que a maior parte das *features* são não categóricas, bem como os resultados apresentados, entendemos que o MLP se configura como o classificador mais apropriado para o problema de classificação abordado, devido à sua forma mais flexível de tratar as entradas e de construir o modelo de classificação.
- › É importante ressaltar que ainda temos muitas possibilidades de configuração de MLP a explorar, o que ainda pode melhorar os resultados aqui apresentados.

# 08 PRÓXIMOS PASSOS

# PRÓXIMOS PASSOS

- › Complementar a base de dados com mais amostras, inclusive para gerar um conjunto de teste a parte dos *k-folds*.
- › Explorar mais os classificadores em termos de calibração, em especial o MLP, acrescentando camadas e, talvez, evoluindo para alguma arquitetura de *Deep Learning* visando a melhorar o modelo.
- › Discutir com especialistas da área médica e identificar possíveis novas características para complementar a base e auxiliar na resolução do problema.
- › Explorar a classificação multiclasse para incluir a classe CMH (Cardiomiopatia Hipertrófica) nas análises.
- › Explorar comitês de classificadores para avaliar o potencial dessa técnica no contexto do problema.

# REFERÊNCIAS

- › BERGAMASCO, Leila Cristina Carneiro. **Recuperação de imagens cardíacas tridimensionais por conteúdo**. 2013. 134 f. Dissertação (Mestrado em Ciências) - Programa de Pós-graduação em Sistemas de Informação, Escola de Artes, Ciências e Humanidades, Universidade de São Paulo, São Paulo, 2013.
- › BERGAMASCO, Leila Cristina Carneiro. **Recuperação de objetos médicos 3D utilizando harmônicos esféricos e redes de fluxo**. 2018. 181 f. Tese (Doutorado em Ciências) - Escola Politécnica, Departamento de Engenharia da Computação e Sistemas Digitais, Universidade de São Paulo, São Paulo, 2018.
- › KUMAR, V.; ABBAS, A. K.; FAUSTO, N.; ASTER, J. C.. **Robbins & Cotran – Patologia: Bases Patológicas das Doenças**. 8 ed. Rio de Janeiro: Elsevier, 2010.
- › SCIKIT-LEARN. **sklearn.ensemble.RandomForestClassifier** — scikit-learn 0.23.1 documentation. Disponível em: [https://scikit-learn.org/stable/modules/generated/sklearn.naive\\_bayes.GaussianNB.html](https://scikit-learn.org/stable/modules/generated/sklearn.naive_bayes.GaussianNB.html). Acesso em: 06 jul. 2020a.
- › SCIKIT-LEARN. **sklearn.model\_selection.GridSearchCV** — scikit-learn 0.23.1 documentation. Disponível em: [https://scikit-learn.org/stable/modules/generated/sklearn.model\\_selection.GridSearchCV.html](https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.GridSearchCV.html). Acesso em: 06 jul. 2020b.
- › SCIKIT-LEARN. **sklearn.naive\_bayes.GaussianNB** — scikit-learn 0.23.0 documentation. Disponível em: <https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestClassifier.html>. Acesso em: 08 mai. 2020c.
- › SCIKIT-LEARN. **sklearn.neural\_network.MLPClassifier** — scikit-learn 0.23.1 documentation. Disponível em: [https://scikit-learn.org/stable/modules/generated/sklearn.neural\\_network.MLPClassifier.html](https://scikit-learn.org/stable/modules/generated/sklearn.neural_network.MLPClassifier.html). Acesso em: 22 jun. 2020d.
- › SCIKIT-LEARN. **sklearn.preprocessing.StandardScaler** — scikit-learn 0.23.1 documentation. Disponível em: <https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.StandardScaler.html>. Acesso em: 14 jul. 2020e.
- › SCIKIT-LEARN. **sklearn.svm.SVC** — scikit-learn 0.23.1 documentation. Disponível em: <https://scikit-learn.org/stable/modules/generated/sklearn.svm.SVC.html#sklearn.svm.SVC>. Acesso em: 01 jun. 2020f.

**MUITO OBRIGADO!**



