



Universidade Federal de São Paulo – Campus São José dos Campos

# Predição de morte em pacientes de UTIs a partir de modelo de Machine Learning

Aprendizado de Máquina aplicado a Sinais e Imagens Médicas

Carlos Alberto Pessin de Souza



## CONTEXTO E RELEVÂNCIA

---

- Avaliação de risco e predição de prognóstico de pacientes de UTIs
- Decisão clínica por intervenção
- Priorização orientada de recursos

## TÉCNICAS VIGENTES

---

- Tradicionalmente baseadas em técnicas estatísticas
- APACHE (Acute Physiology and Chronic Health Evaluation)
  - Sistema de Score calculado a partir de 14 variáveis fisiológicas e históricas coletadas nas primeiras 24 horas após internação
  - Regressão logística
  - Cálculo relativamente simples
- Desempenho AUC–ROC (Badawi et al., 2018)
  - APACHE IV – 0,85
  - SOFA (Sequential Organ Failure Assessment) – 0,81
  - DRS (Discharge Readiness Score) – 0,88

## APACHE

## APACHE II SCORE

## AGE Points

≤ 44y	0
45-54y	2
55-64y	3
65-74y	5
≥75y	6

## CHRONIC HEALTH Points

Non-operative, or emergency post-op & any conditions below*	5
Elective operation & any conditions below*	2
*Cirrhosis w/ portal Hypertension or encephalopathy; class IV angina, chronic hypoxia, ↑CO <sub>2</sub> or polycythemia; chronic dialysis; immunocompromised	

## TOTAL APACHE SCORE = AP + CHP + APS

Sum Age Points (AP) + Chronic Health Points (CHP) + Acute Physiologic Score (APS) points. →

\*1 Sum all variables 1-12 for Acute Physiologic Score (APS) (use one variable each for 5 and 9).

Use the worst value from the preceding 24h.

APACHE II: a severity of disease classification system.  
Crit Care Med 1985;13:818-29.

## ACUTE PHYSIOLOGIC SCORE\*1 (APS)

Physiologic Variable		Points								
		4	3	2	1	0	1	2	3	4
1	Temp °F	≤85.9	86.0-89.5	89.6-93.1	93.2-96.7	96.8-101.2	101.3-102.1		102.2-105.7	≥105.8
	°C	≤29.9	30-31.9	32-33.9	34-35.9	36 - 38.4	38.5-38.9		39-40.9	≥41
2	HR, bpm	≤39	40-54	55-69		70-109		110-139	140-179	≥180
3	MAP, mmHg	≤49		50-69		70-109		110-129	130-159	≥160
4	RR, bpm	≤5		6-9	10-11	12-24	25-34		35-49	≥50
5	Oxygenation: Use A-a Gradient (5a) if FiO2 ≥0.5 or use PaO2 (5b) if FiO2 <0.5 (see page 17)									
5a	A-a Gradient					<200		200-349	350-499	≥500
5b	PaO2	≤54	55-60		61-70	>70				
6	Na <sup>+</sup> (S, mmo/L)	≤110	111-119	120-129		130-139	150-154	155-159	160-179	≥180
7	K <sup>+</sup> (S, mmo/L)	≤2.4		2.5-2.9	3.0-3.4	3.5-5.4	5.5-5.9		6.0-6.9	≥7.0
8	Cr (S, mg/dL)			<0.6		0.6-1.4		1.5-1.9	2.0-3.4	≥3.5
9	Arterial pH is preferred. Use venous HCO3 if no ABGs.									
9a	pH (arterial)	≤7.14	7.15-7.24	7.25-7.32		7.33-7.49	7.5-7.59		7.6-7.69	≥7.7
9b	HCO3 (venous)	≤14	15-17.9	18-21.9		22-31.9	32-40.9		41-51.9	≥52
10	WBC, cells/uL	≤1.0		1.0-2.9		3.0-14.9	15-19.9	20-39.9		≥40
11	Hct, %	≤20		20-29.9		30-45.9	46-49.9	50-59.9		≥60
12	GCS coma	Score = 15 – GCS Score (see below, Record e.g.: "GCS 9 = E2 V4 M3 at 17:35h".)								

## Score Mortality

0 - 4	4%
5 - 9	4%
10 - 14	15%
15 - 19	25%
20 - 24	40%
25 - 29	55%
30 - 34	75%
> 34	85%

## GLASGOW COMA SCALE (GCS) \*Teasdale G, Jennett B. Lancet 1974;2:81-84.

EYE Opening	Best VERBAL	Best MOTOR	Points	SCORE: Sum Points (eye+verbal+ motor categ).  Severe ≤ 8. Mod = 9-12. Minor ≥ 13.
		follows commands	6	
	oriented	localizes pain	5	
spontaneous	confused	withdraws to pain	4	
to command	inappropriate words	flexor response	3	
to painful stimuli	incomprehensible	extension (abnl)	2	
no response	no response	no response	1	

## APLICAÇÃO DE IA

---

- Informatização de dados médicos
- Aumento da capacidade de processamento
- Desenvolvimento de técnicas de IA
- Desafio WiDS (Women in Data Science) 2020 – Universidade de Stanford
  - Equipe vencedora
    - Classificador híbrido – AUC 0,915
    - APACHE IV – AUC 0,868



## OBJETIVO

---

- Desenvolvimento de um classificador baseado em técnicas de Machine Learning supervisionado para predição de morte em pacientes de UTI's a partir de múltiplas variáveis
- Consolidação da possibilidade de desenvolvimento de um modelo prático baseado em IA que possa ser mais eficiente do que as técnicas atualmente empregadas

## BANCO DE DADOS

---

- WiDS 2020
  - Global Open Source Severity of Illness Score (GOSSIS) – Massachusetts Institute of Technology (MIT)
- 91.713 pacientes
  - 83.781 sobreviveram (91,4%)
  - 7.907 morreram (8,6%)
- 147 UTIs
- 185 características para cada paciente

## BANCO DE DADOS

---

Tipo de dado	Exemplos	Quantidade
Demográficos	Idade, gênero, altura, status pré internação, entre outros	16
Exames laboratoriais	Concentração de hemoglobinas, sódio no sangue, entre outros	76
Vitais	Temperatura, pressão arterial, batimentos cardíacos, entre outros	52
Índices de risco	APACHE	41



## PRÉ-PROCESSAMENTO

---

### IMPUTAÇÃO DE DADOS

- Valores vazios no banco de dados
  - Não podem ser entrada no classificador
- 1. Remoção de características e padrões com mais de 20% de informações faltantes
  - Redução de erro de imputação
  - 46% das características removidas (100 restantes)
  - 8% dos pacientes removidos (84.716 restantes)

## PRÉ-PROCESSAMENTO

---

### IMPUTAÇÃO DE DADOS

#### 2. Imputação variáveis categóricas

- Vazio como nova categoria

#### 3. Imputação variáveis numéricas

- Valor gerado aleatoriamente a partir de distribuição normal de outros valores da característica
- Intervalo de CDF – 0,4 a 0,6

## PRÉ-PROCESSAMENTO

---

### REMOÇÃO DE OUTLIERS

- Redução de interferência de valores muito externos
  1. Intervalo – média  $\pm$  15 desvios-padrão
    - $<1\%$  de pacientes removidos

### SEPARAÇÃO TREINO E TESTE

- Para avaliação da capacidade de generalização do modelo a dados não vistos na fase de aprendizagem
- Dados de teste passam pelas mesmas operações definidas com os dados de treino
  1. Recorte aleatório mantendo proporção entre classes
    - 90% treino – 10% teste

## PRÉ-PROCESSAMENTO

---

### NORMALIZAÇÃO LINEAR

- Ordem de grandeza das variáveis pode interferir na avaliação do modelo
  1. Todas características dos dados de treino com média 0 e desvio padrão 1
    - Dados de teste passam por mesma operação

### REMOÇÃO DE OUTLIERS DE TREINO

- Redução de interferência de valores externos
  1. Intervalo – média  $\pm$  10 desvios-padrão
    - 1,5% de pacientes removidos

## PRÉ-PROCESSAMENTO

---

### DESBALANÇO ENTRE CLASSES

- 8,6% pacientes morreram vs. 91,4% pacientes sobreviveram
- Risco de desequilíbrio da superfície de decisão gerando classificador “viciado” em classificar pacientes como sobreviventes
  1. Subamostragem aleatória de 10% de pacientes da classe de pacientes sobreviventes
    - Redução brusca do número de dados
    - Pacientes que morreram passaram a representar pouco menos de 50%

## TRANSFORMAÇÃO DO ESPAÇO

---

### PCA

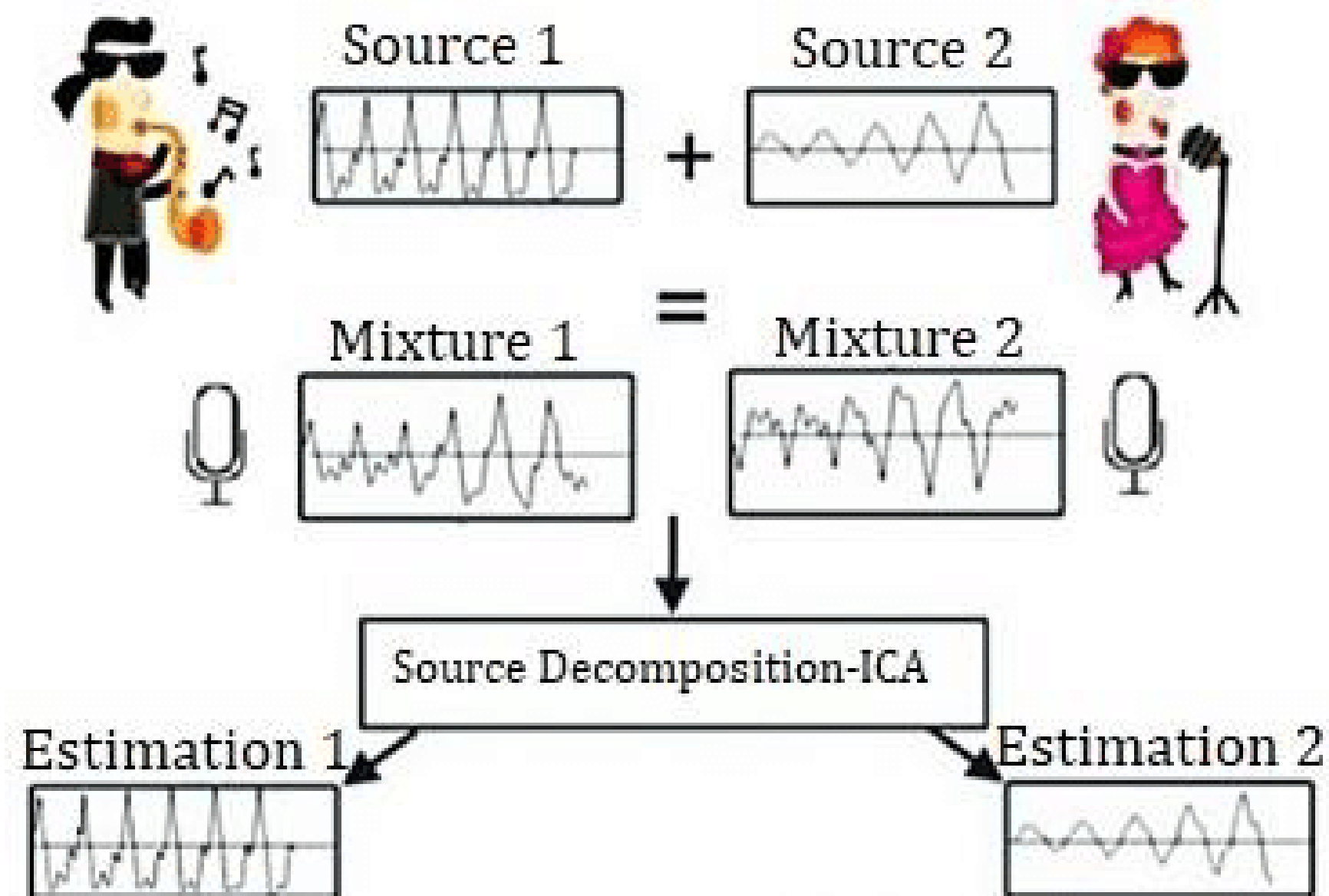
1. Transformação linear do espaço de características utilizando matriz de transformação que diagonaliza matriz de covariância
  - Novas componentes possuem mínima correlação linear
  - Nova matriz de covariância quantifica a variância que cada componente empresta ao conjunto geral
  - Componentes de menor variância podem ser interpretadas como ruído e removidas
2. Remoção de 45 componentes
  - <5% de variância removida



## TRANSFORMAÇÃO DO ESPAÇO

### ICA

- *The Cocktail Party Problem*
- Assume que atuais componentes são uma combinação de fontes completamente independentes



## TRANSFORMAÇÃO DO ESPAÇO

---

### ICA

- Estratégia derivada do Teorema do Limite Central
  - Soma de sinais independentes gera uma gaussiana
  - Gaussianidade  $\leftrightarrow$  Independência
  - FastICA – busca iterativamente uma matriz de transformação linear que maximiza critérios de não-gaussianidade das fontes (negentropia)
- 1. Transformação linear do espaço de características
  - Construção de um espaço de mesmas dimensões cujas componentes possuem mínima correlação

## SELEÇÃO DE CARACTERÍSTICAS

---

### RANDOM FOREST

- Dentro das características restantes, é legítimo assumir que nem todas são relevantes para este processo de classificação
- Como quantificar?
  - Parâmetro feature importance – Random Forest
- 1. Remoção das 15 características menos relevantes
- Criação de um espaço reduzido
  - Diminuição da complexidade da tarefa de classificação
  - Favorecimento do foco do classificador nas características mais relevantes

## CLASSIFICADOR

---

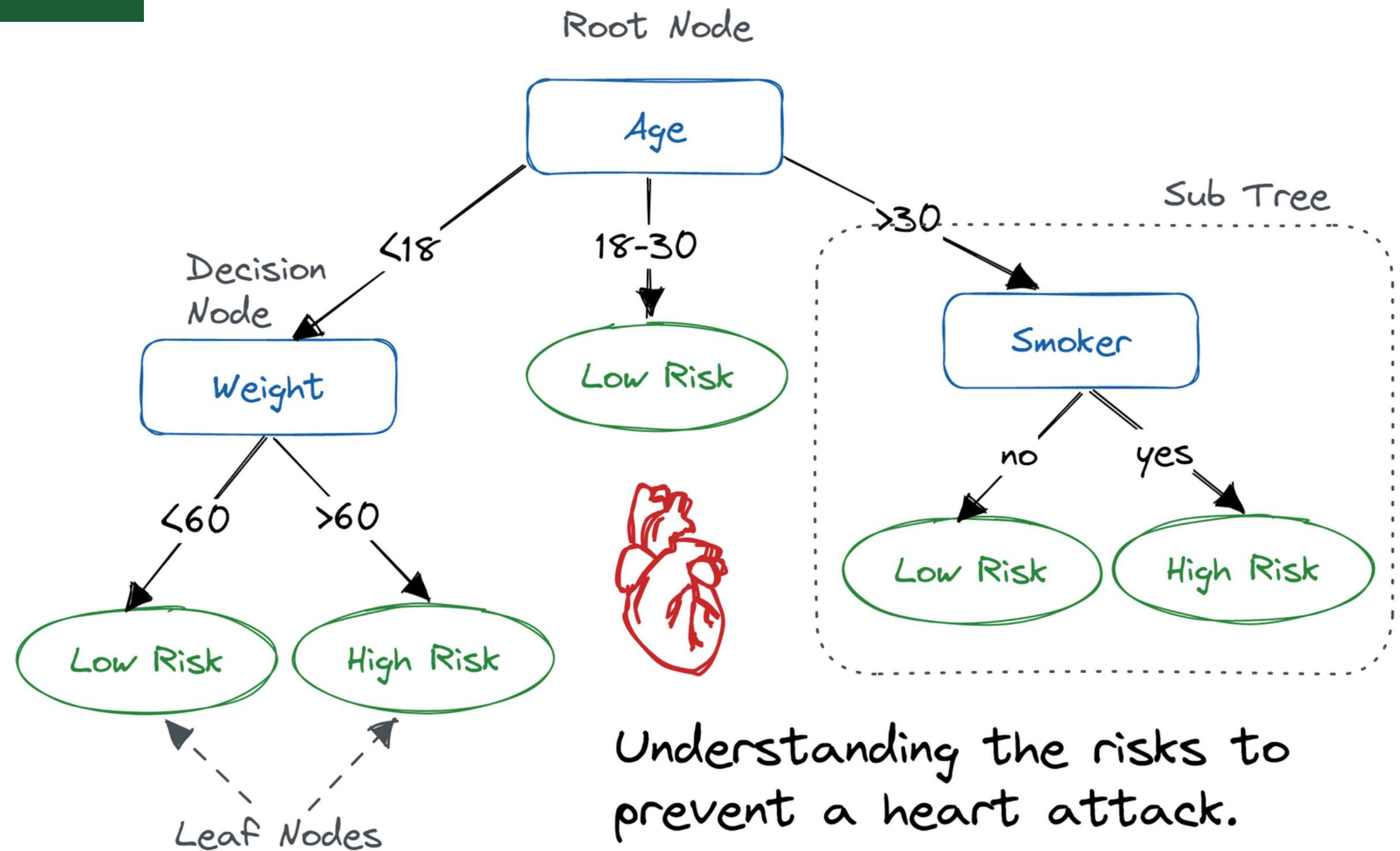
- Algoritmo responsável pela criação de uma regra/superfície de decisão que se adapte aos dados de treino disponíveis e que seja capaz de receber um novo padrão e classificá-lo

### RANDOM FOREST

- Criação de uma série de árvores de decisão, utilizando características e padrões diferentes, selecionados aleatoriamente
- A classificação é definida pela classe que mais vezes foi resultado da classificação em cada uma das árvores
- Maior capacidade de generalização da classificação
  - Menor chance de sobreajuste
- Modelo final construído com 1000 árvores

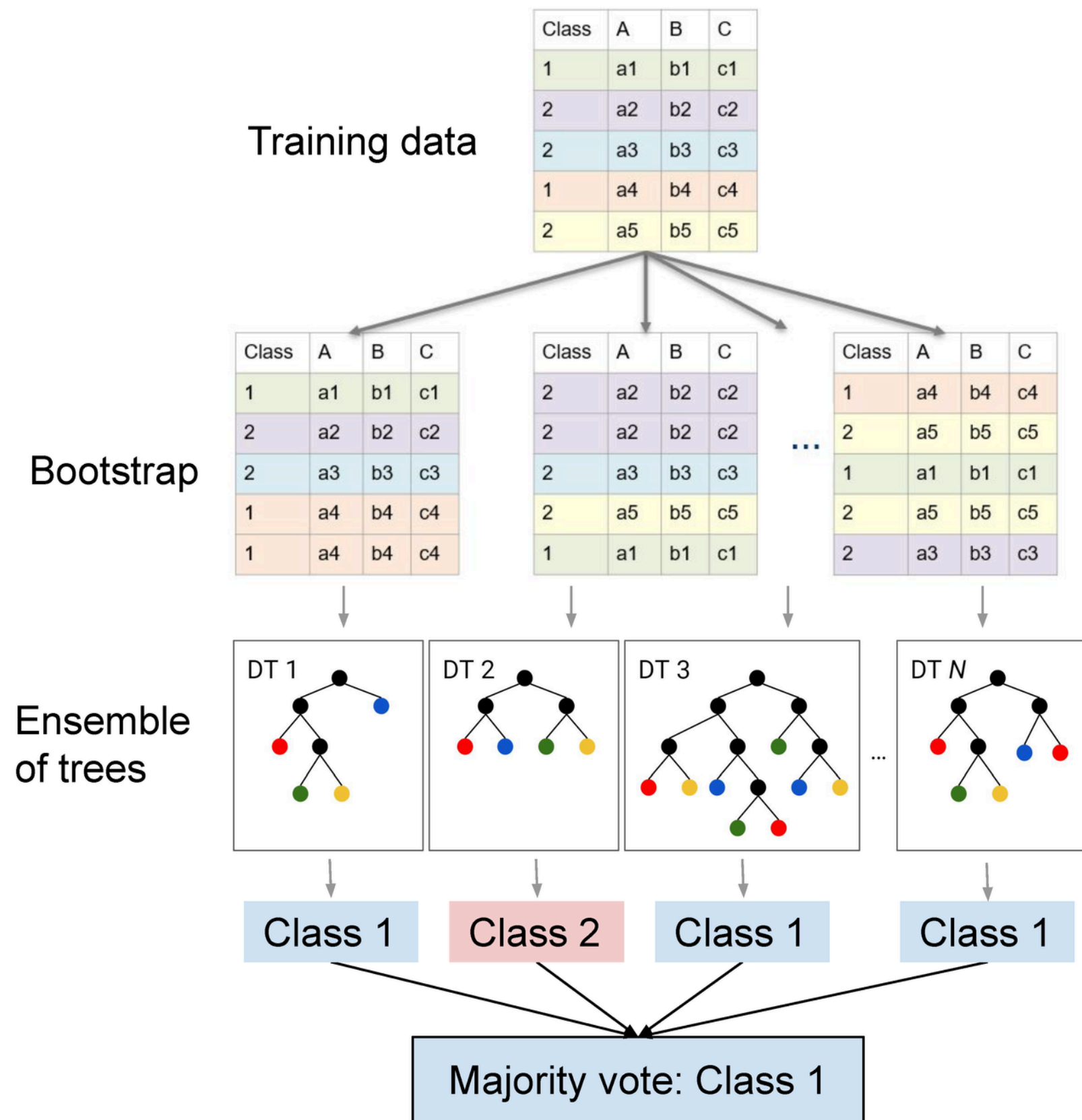
## CLASSIFICADOR

### RANDOM FOREST



## CLASSIFICADOR

### RANDOM FOREST





## CLASSIFICADOR

---

### SVM

- Construção de superfície de decisão que maximiza a distância entre padrões específicos (vetores de suporte do modelo) e este hiperplano
  - Costuma ser eficiente em problemas de alta dimensionalidade, apesar de sua relativa baixa complexidade
1. Foram testados diferentes kernels não-lineares
    - a. Rbf
    - b. Polinomial (de segundo a quinto grau)
    - c. Sigmóide

## CLASSIFICADOR

---

### HÍBRIDO

- Consiste na combinação de diferentes classificadores
- A classificação final é dada pela classe que foi mais vezes resultado de classificação de um classificador individual
- Cada classificador pode se adaptar melhor a um conjunto de padrões
  - Combiná-los pode favorecer na eficiência e generalização do modelo
- Foram combinados os modelos que estavam tendo melhor desempenho nas validações iniciais
  - Random Forest, SVM rbf e SVM polinomial de grau 4

## VALIDAÇÃO CRUZADA

---

### K-FOLD

- Para avaliação dos modelos, os dados de treino foram redivididos em 10 partes
  - 9 dessas partes foram destinadas a treinamento e 1 para validação
  - Cada uma das partes foi selecionada aleatoriamente, mantendo-se a proporção entre as classes
- Toda a etapa de treinamento (a partir da normalização) foi repetida, 10 vezes, alternando a porção de dados utilizadas para validação

## VALIDAÇÃO CRUZADA

---

### MÉTRICAS

- Acurácia
- Sensibilidade
- Especificidade
- AUC-ROC
- A avaliação de cada modelo foi dada pela média das métricas entre as 10 repetições

## TESTE

---

- Modelo de melhor desempenho em validação foi treinado novamente com todo conjunto de treino (treino + validação)
- Testado nos dados de teste separados inicialmente
- Reavaliado, métrica final

## PRÉ-PROCESSAMENTO

---

- Banco de dados original
  - 91.713 pacientes e 180 características
  - 8,6% pacientes morreram vs. 91,4% pacientes sobreviveram
- Banco de dados final
  - 19.407 pacientes e 40 características
  - 49,4% pacientes morreram vs. 50,6% pacientes sobreviveram
- Redução do espaço em quase 16 milhões de variáveis (restando aproximadamente de 4,7% do total de parâmetros)



## PRÉ-PROCESSAMENTO

---

- Ajuste fino na remoção de variáveis
  - Redução da complexidade do modelo
    - Possibilidade de utilizar técnicas mais simples
    - Menor custo computacional
    - Aumento do “foco” em variáveis mais significativas
  - Redução na capacidade de aprendizagem
    - Exclusão de informação significativa para a classificação
- Alternativas para exclusão massiva
  - Técnica mais rebuscada de imputação de dados
  - Desbalanço de classes
    - Sobreamostragem da classe menos numerosa
    - Técnicas de regularização de classificador

# VALIDAÇÃO

## MÉTRICAS

Classificador	Random Forest		SVM – rbf		SVM – polinomial (4º grau)		Híbrido	
Métrica	Treino	Validação	Treino	Validação	Treino	Validação	Treino	Validação
Acurácia	88,77%	75.39%	75.82%	77.89%	99.99%	72.52%	81,25%	83.11%
Sensibilidade	89.42%	79.10%	71.92%	71.74%	99.98%	68.45%	70,48%	66.12%
Especificidade	88.15%	75.05%	79.49%	78.46%	100.00%	72.89%	91.39%	84.69%
AUC-ROC	0.888	0.771	0.757	0.751	1.000	0.707	0.809	0.754

# Resultados e Discussão

## TESTE

---

### RANDOM FOREST

Métrica	Dados de teste
Acurácia	75,57%
Sensibilidade	81,91%
Especificidade	74,98%
AUC-ROC	0,784

## AVALIAÇÃO DOS RESULTADOS

---

- O projeto de classificador atingiu desempenho regular na tarefa de classificação
  - 0,784 em AUC
  - DRS: 0,88
  - APACHE: 0,868
  - Modelo IA vencedor WiDs: 0,915
- A abrupta exclusão de 95% dos dados, bem como a simplicidade das técnicas utilizadas explicam tal desempenho inferior

## AVALIAÇÃO DOS RESULTADOS

---

- Considerando a aplicação do modelo, faz-se interessante a escolha de um modelo que privilegie a sensibilidade em relação à especificidade
  - Identificar os pacientes que precisam de intervenção
- Já na ótica da priorização de recursos, o bom desempenho em especificidade pode dar segurança ao bem identificar pacientes de menor risco

# Conclusão

- A análise de risco de morte em pacientes internados em UTI's é extremamente relevante tanto para o cuidado clínico como pode ser essencial na administração de recursos
- A informatização dos dados médicos, junto com o desenvolvimento da capacidade de processamento e das técnicas de Inteligência Artificial, traz à tona o potencial de desenvolvimento de técnicas que possam ser mais eficientes e práticas para tal função
- O trabalho foi bem sucedido em desenvolver um modelo baseado em Machine Learning de classificação e, apesar de ter atingido resultados inferiores a outras técnicas, demonstra o potencial deste tipo de ferramenta



# Referências Bibliográficas

---

Badawi, O., Liu, X., Hassan, E., Amelung, P. J., & Swami, S (2018). Evaluation of ICU Risk Models Adapted for Use as Continuous Markers of Severity of Illness Throughout the ICU Stay\*. Critical Care Medicine 46(3), P361–367. <https://doi.org/10.1097/CCM.0000000000002904>

Breiman, L. (2001). Random Forests. Machine Learning 45, P5–32. <https://doi.org/10.1023/A:1010933404324>

Carvalho, G. A., Rezende, A. A. B., Reis, G. R., & Gardenghi, G. (2020). Uso do escore APACHE IV como preditor de mortalidade e tempo de permanência em uma unidade de terapia intensiva. Revista Pesquisa em Fisioterapia 10(1), P9-15. <https://doi.org/10.17267/2238-2704rpf.v10i1.2606>

Clifford, G. D. (2005). Singular Value Decomposition & Independent Component Analysis for Blind Source Separation – Course materials for HST582J/6.555J/16.456J, Biomedical Signal and Image Processing. Massachusetts Institute of Technology OpenCourseWare <http://ocw.mit.edu>

Cohen, S., Dagan, N., Cohen-Inger, N., Ofer, D., & Rokach, L. (2021). ICU Survival Prediction Incorporating Test-Time Augmentation to Improve the Accuracy of Ensemble-Based Models. IEEE Access 9, P91584–91592. <https://doi.org/10.1109/ACCESS.2021.3091622>

Emanuel, E. J., Persad, G., Upshur, R., Thome, B., Parker, M., Glickman, A., Zhang, C., Boyle, C., Smith, M., & Phillips, J. P. (2020). Fair Allocation of Scarce Medical Resources in the Time of Covid-19. New England Journal of Medicine 382(21), P2049–2055. <https://doi.org/10.1056/NEJMs2005114>

Freitas, E. R. F. S. (2010). Perfil e gravidade dos pacientes das unidades de terapia intensiva: aplicação prospectiva do escore APACHE II. Rev. Latino-Am. Enfermagem 18(3), P317–323. <https://doi.org/10.1590/S0104-11692010000300004>

Jentzer, J. C., Diepen, S. D., Murphree, D. H., Ismail, A. S., Keegan, M. T., Morrow, D. A., Barsness, G. W., & Anavekar, N. S. (2020). Admission diagnosis and mortality risk prediction in a contemporary cardiac intensive care unit population. American Heart Journal 224, P57–64. <https://doi.org/10.1016/j.ahj.2020.02.018>

Pisner, D. A., & Schnyer, D. M. (2020). Chapter 6 – Support vector machine. Machine Learning – Methods and Applications to Brain Disorders. P101–121. <https://doi.org/10.1016/B978-0-12-815739-8.00006-7>.

Zimmerman, J. E., Kramer, A. A., McNair, D. S., & Malila, F. M. (2006). Acute Physiology and Chronic Health Evaluation (APACHE) IV: hospital mortality assessment for today's critically ill patients. Crit Care Med. 34(5), P1297–1310. <https://doi.org/10.1097/01.ccm.0000215112.84523.f0>