

João Victor Sousa

## **Título do seu TCC**

Ilhéus - Bahia

2025

João Victor Sousa

## **Título do seu TCC**

Trabalho apresentado ao Curso de Y da Universidade X como requisito parcial para obtenção do título de Bacharel.

Universidade Estadual de Santa Cruz  
Curso de Graduação em Ciência da Computação

Orientador: Otacílio José Pereira

Ilhéus - Bahia

2025

# Resumo

Escreva aqui o resumo do seu trabalho. Inclua objetivos, metodologia, resultados e conclusões.

**Palavras-chave:** palavra1. palavra2. palavra3.

# Sumário

Sumário	3	
1	INTRODUÇÃO	5
2	METODOLOGIA	6
2.1	Particionamento dos dados e classes	6
2.2	Métricas	7
2.3	Modelos utilizados	8
2.4	Organização da entrada	9
3	RESULTADOS	10
4	CONCLUSÃO	11
	REFERÊNCIAS	12

# Lista de Siglas

AUC	Area Under Curve
RNN	Redes Neurais Recorrentes
CNN	Redes Neurais Convolucionais
LSTM	Long Short-Term Memory
AAMI	Association for the Advancement of Medical Instrumentation
GRU	Gated Unit Recurrent
ECG	Eletrocardiograma
TP	Verdadeiro Positivo
FP	Falso Positivo
TN	Verdadeiro Negativo
FN	Falso Negativo

# 1 Introdução

Texto da introdução.

## 2 Metodologia

Primeiramente, foi necessário definir qual banco de dados seria utilizado para o treinamento e a validação. Optou-se pelo *MIT-BIH Arrhythmia Database* (Moody e Mark 2005), recomendado pela AAMI. O banco é composto por 58 registros de eletrocardiograma (ECG), cada um com 30 minutos de duração. Os 23 primeiros registros foram selecionados aleatoriamente a partir de um conjunto de 4000 gravações de 24 horas realizadas em pacientes ambulatoriais do Beth Israel Deaconess Medical Center. Os 25 registros restantes foram escolhidos de modo a incluir arritmias raras, mas clinicamente significativas.

Uma característica importante desse banco é a anotação de cada batimento cardíaco em torno do complexo R, realizada por três cardiologistas independentes.

### 2.1 Particionamento dos dados e classes

Os dados foram particionados seguindo a estratégia inter-paciente proposta por Chazel et al. (apud Silva et al. 2025), na qual batimentos de um mesmo paciente não podem aparecer simultaneamente nos conjuntos de treinamento e validação. O objetivo é garantir a capacidade de generalização do modelo para diferentes pacientes. Além disso, conforme recomendado pela AAMI, registros de pacientes com marcapasso foram excluídos.

Os registros 101, 106, 108, 109, 112, 114, 115, 116, 118, 119, 122, 124, 201, 203, 205, 207, 208, 209, 215, 220, 223 e 230 foram utilizados para treinamento (conjunto DS1). Os demais (100, 103, 105, 111, 113, 117, 121, 123, 200, 202, 210, 212, 213, 214, 219, 221, 222, 228, 231, 232, 233 e 234) formaram o conjunto de teste (DS2).

De acordo com a AAMI (apud Silva et al. 2025), são definidas cinco classes de arritmia: N, SVEB, VEB, F e Q, correspondentes a batimento normal, batimento supraventricular ectópico, batimento ventricular ectópico, fusão de batimento ventricular e normal e batimento não classificado, respectivamente. A Tabela 1 apresenta a distribuição dessas classes no conjunto de dados.

Tabela 1 – Particionamento inter-paciente proposto por Chazel et al.

Conjunto	N	SVEB	VEB	F	Q	Total
DS1	45 866	944	3 788	415	8	51 021
DS2	44 259	1 837	3 221	388	7	49 712
Total	90 125	2 781	7 009	803	15	100 733

Fonte: Adaptado de Silva et al. (2025).

O conjunto DS1 foi então subdividido em treinamento e validação por meio de

validação cruzada, inicialmente com duas partições (dois *folds*) e, posteriormente, com cinco partições (cinco *folds*) nos modelos finais.

Considerando o desbalanceamento dos dados e visando maior simplicidade, adotou-se a classificação binária, na qual a classe positiva corresponde a arritmia e a classe negativa a batimentos normais.

## 2.2 Métricas

As métricas utilizadas para avaliar o desempenho dos modelos foram: sensibilidade, precisão, acurácia, *F1-score* e AUC (*Area Under the Curve*).

A sensibilidade representa a capacidade do modelo em identificar corretamente as classes positivas, isto é, os batimentos arrítmicos. Sua equação é dada por:

$$\text{Sensibilidade} = \frac{TP}{TP + FN} \quad (2.1)$$

em que  $TP$  são os verdadeiros positivos e  $FN$  os falsos negativos.

A precisão, por sua vez, indica a proporção de batimentos classificados como arrítmicos que realmente pertencem a essa classe:

$$\text{Precisão} = \frac{TP}{TP + FP} \quad (2.2)$$

onde  $FP$  representa os falsos positivos. Precisão e sensibilidade estão relacionadas por um *trade-off*. No contexto médico, prioriza-se elevada sensibilidade, ainda que à custa de menor precisão, uma vez que falsos negativos são mais prejudiciais que falsos positivos.

O *F1-score* é a média harmônica entre precisão e sensibilidade, buscando um equilíbrio entre ambas:

$$F1\text{-score} = \frac{2 \cdot \text{Precisão} \cdot \text{Sensibilidade}}{\text{Precisão} + \text{Sensibilidade}} \quad (2.3)$$

A acurácia corresponde ao acerto global do modelo, considerando tanto as classes positivas quanto as negativas:

$$\text{Acurácia} = \frac{TP + TN}{TP + TN + FP + FN} \quad (2.4)$$

A AUC mede a capacidade do modelo em separar as classes positivas das negativas, variando entre 0 e 1. Valores próximos de 1 indicam separação perfeita, enquanto 0,5 corresponde a um modelo com desempenho equivalente ao acaso.

Essa métrica é calculada a partir da área sob a curva ROC, ilustrada na Figura 1.



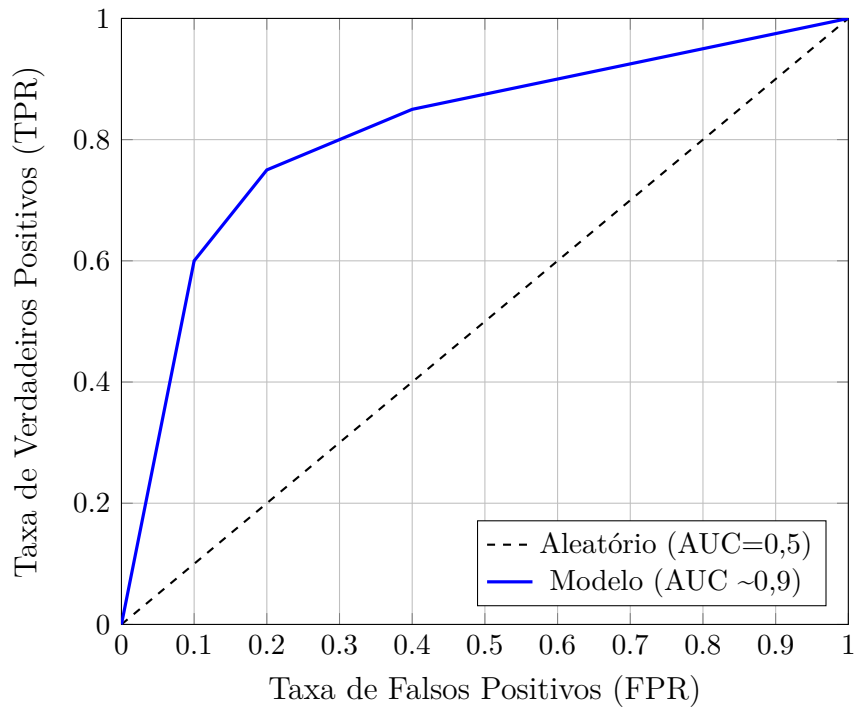


Figura 1 – Exemplo de curva ROC: comparação entre modelo aleatório e modelo com bom desempenho.

A matriz de confusão, por fim, fornece uma representação tabular dos acertos e erros do modelo, como exemplificado na Tabela 2.

Tabela 2 – Exemplo de matriz de confusão binária

Classe Verdadeira	Classe Predita	
	Positiva	Negativa
Positiva	TP	FN
Negativa	FP	TN

Fonte: Elaborado pelo autor.

Essas métricas em conjunto permitem avaliar não apenas a proporção global de acertos, mas também a capacidade do modelo em detectar corretamente arritmias, aspecto essencial em aplicações médicas.

## 2.3 Modelos utilizados

Como o diagnóstico de arritmias depende tanto do aspecto morfológico dos ECGs quanto do aspecto sequencial, isto é, o ritmo, foram escolhidos modelos RNNs que possuem bom desempenho em dados sequenciais como séries temporais e linguagem natural (James et al. 2023). Já as redes CNN possuem grande capacidade de extração de características e podem aprender os aspectos morfológicos dos ECGs.

## 2.4 Organização da entrada

Foi escolhido de cada ECG o lead MLII que facilita a visualização do complexo QRS. Quando disponível. O ECG foi limpo utilizando a biblioteca neurokit2.

Cada batimento foi segmentado utilizando a biblioteca neurokit2 (Makowski et al. 2021). Com os batimentos segmentados foi montado uma matriz tridimensional no formato: (*batch size*, tamanho da sequência, características). As características inicialmente correspondem as amostras do ECGs e demais características como intervalo RR, média do intervalo RR para o paciente e a média do 5 intervalos RR anteriores e posteriores. Em cada sequência de batimentos, só há batimentos exclusivos de um único paciente.

## 3 Resultados

Texto dos resultados.

## 4 Conclusão

Texto da conclusão.

# Referências

James et al. 2023 JAMES, G. et al. *An Introduction to Statistical Learning: With Applications in Python*. Cham: Springer, 2023. E-book. Disponível em: <<https://www.statlearning.com/>>. Acesso em: 28 ago. 2025.

Makowski et al. 2021 MAKOWSKI, D. et al. NeuroKit2: A python toolbox for neurophysiological signal processing. *Behavior Research Methods*, Springer Science and Business Media LLC, v. 53, n. 4, p. 1689–1696, feb 2021. Disponível em: <<https://doi.org/10.3758%2Fs13428-020-01516-y>>.

Moody e Mark 2005 MOODY, G. B.; MARK, R. G. *MIT-BIH Arrhythmia Database*. 2005. Disponível em: <<https://www.physionet.org/physiobank/database/mitdb/>>. Acesso em: 28 set. 2025.

Silva et al. 2025 SILVA, G. A. L. et al. *A Systematic Review of ECG Arrhythmia Classification: Adherence to Standards, Fair Evaluation, and Embedded Feasibility*. 2025. ArXiv:2503.07276v1. Acesso em: 28 set. 2025.