



# Classificação de subtipos de câncer de mama com dados transcriptômicos estimados e limitados utilizando ensemble learning

Gabriel de Q. Sousa<sup>1</sup>; Beatriz A. Rodrigues<sup>1</sup>; Eduardo F. Nakamura<sup>1</sup>; Fabiola G. Nakamura<sup>1</sup>;  
1- Instituto de Computação/Universidade Federal do Amazonas (IComp/UFAM);

Dados Ausentes

Dados podem faltar por poeira ou arranhões nas lâminas

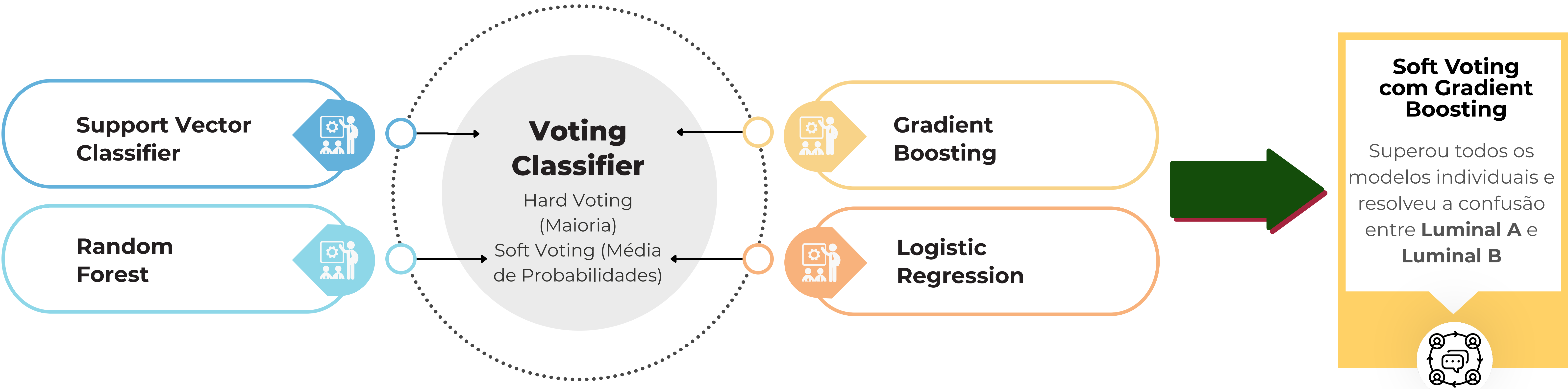
Limitação

A base de dados é derivada da CPTAC e filtrada com os 50 genes do PAM50. Total: 117 amostras

Dificuldade de Distinção

Confusão entre Luminal A e Luminal B

Classe	Quantidade de amostras
Luminal A	57
Basal	29
Luminal B	17
Her2	14



Métrica	Melhor classificador individual (RF)	Voting Soft + GB (Melhor resultado)
Acurácia	0.932	0.966
F1-Score	0.885	0.960

**Diversidade de modelos**  
Mitigou a dificuldade em distinguir Luminal A e Luminal B

**VotingClassifier**  
O uso do **Soft Voting** resultou em melhorias significativas de desempenho em dados limitados

**Trabalhos Futuros**

- Uso de modelos de deep learning no ensemble
- Utilização de transfer learning
- Testar diferentes datasets

Para mais detalhes, acesse o QR Code:



**Referências**

Rodrigues BA, et al. Um método de Estimação de Expressões Gênicas de Câncer de Mama... (2023)

Adedigba AP, et al. Deep learning-based mammogram classification using small dataset. (2019)

Albashish D, et al. Design an Ensemble Pretrained Deep Learning Model for Classification of Melanoma Skin Cancer Images. (2025)

Cao-Van K, et al. Soft-Voting Ensemble Model: An Efficient Learning Approach for Predictive Prostate Cancer Risk. (2024)

Azad M, et al. A novel ensemble learning method using majority based voting of multiple selective decision trees. (2025)

Batool A, Byun YC. Toward improving breast cancer classification using an adaptive voting ensemble learning algorithm. (2024)