

Projeto de Sinais e Sistemas - Escolha de algoritmo classificador

Grupo 4

Integrantes:

Henrique César (Líder)
Esdras Gabriel
Gabriel Jatobá
João Victor Nascimento
Luiz Gustavo
Márcio André

Classificador escolhido:

Random Forest

Justificativa:

A equipe fez uma breve visualização do banco de dados recomendado pelo professor, o PTB-XL, e percebeu a sua robustez e, principalmente, a quantidade de informações contidas nele. Diante disso surgiu a primeira questão que deveríamos enfatizar para nossa escolha: um classificador que possa lidar com conjuntos de dados complexos e com múltiplos atributos. Uma vez estabelecida a dimensão que seriam seguidas, a questão seguinte passou a ser se seria usado um classificador a base de Machine Learning Tradicional ou Deep Learning, e, de forma completamente arbitrária, a equipe decidiu por seguir com algo tradicional.

Em sua publicação em 2001, Leo Breiman mostra como o Random Forest é pode ser um modelo de classificador com alta resistência a overfitting, mesmo quando há muitas variáveis correlacionadas — como é o caso de múltiplos extratores aplicados às 12 derivações do ECG do banco de dados que vamos usar. Um estudo feito por Saira Aziz em 2021 utilizou o classificador Random Forest com sucesso para classificar batimentos cardíacos e reforçou sua adequação a tarefas biomédicas supervisionadas. No mesmo ano, Sandra Śmigiel realizou um estudo com esse banco de dados e mostrou que há uma competitividade entre o uso de Deep Learning e modelos tradicionais baseados em árvores de decisão, Random Forest.

Dessa forma esse modelo foi escolhido pela equipe. Além do bom desempenho e da prática implementação dele, ainda há uma alta transparência, pois é possível quantificar a importância de cada variável no processo de decisão, ajudando na hora de entender os resultados que serão obtidos no projeto.

Referências:

- Aziz, S., Ahmed, S., & Alouini, M. S. (2021). *ECG-based machine-learning algorithms for heartbeat classification*. Scientific Reports. <https://www.nature.com/articles/s41598-021-97118-5>;

- Śmigiel, S., Pałczyński, K., & Ledziński, D. (2021). *ECG signal classification using deep learning techniques based on the PTB-XL dataset*. <https://www.mdpi.com/1099-4300/23/9/1121>;
- Wagner, P., et al. (2020). *PTB-XL, a large publicly available electrocardiography dataset*. Scientific Data. <https://www.nature.com/articles/s41597-020-0495-6>