

KNN ponderado e SVM na predição de Câncer de mama.

*

Arthur Felipe Reis Souza
Electrical Engineering Department,
Federal University of Minas Gerais,
Belo Horizonte, Brazil
arthurfreisouza@gmail.com

Antônio de Pádua Braga and Frederico Gualberto Ferreira Coelho
Electrical Engineering Department,
Federal University of Minas Gerais,
Belo Horizonte, Brazil
apbraga@cpdee.ufmg.br, fredgfc@ufmg.br

October 26, 2024

Abstract

1 Introdução

O câncer de mama é uma condição que se caracteriza pelo crescimento descontrolado de células malignas no tecido mamário, afetando tanto homens quanto mulheres, embora seja mais prevalente entre as mulheres. Os riscos para o desenvolvimento desse tumor aumentam com a idade, histórico familiar de casos anteriores e hábitos de vida, como obesidade, sedentarismo, e vícios em álcool e tabaco. Os principais sintomas incluem alterações no formato da mama, inchaço e irritação, dor nos seios e a presença de um nódulo na região mamária. Este relatório tem como objetivo, com base em um conjunto de dados de 568 pacientes, prever se o tumor é maligno ou benigno. Para essa predição, foram utilizados dois algoritmos: KNN ponderado e Máquina de Vetor de Suporte

*Insert applicable funding agency here. If none, delete this.

(SVM). Os hiperparâmetros de ambos os modelos foram ajustados para compreender seu desempenho, e a seleção dos parâmetros finais foi realizada por meio do algoritmo de GridSearch Cross-Validation.

As características que resultam em um rótulo são extraídas por meio de uma análise matemática das mamografias (raios X da região mamária), onde esses elementos representam distâncias nas imagens. Esses valores matemáticos são cruciais para a predição, pois alimentam o classificador, que determinará se o tumor é maligno ou benigno.

Figure 1: Foto resultante de uma mamografia.

2 Data Generation

3 KNN Function

4 New Data Generation

5 Results