

Análise Experimental da Detecção de Células de Câncer de Mama

Danilo Calhes Paixão dos Santos RA: 11025215

Felipe Kawashita Kobayashi RA: 11090914

Novembro 2019

1 Resumo

O presente trabalho propõe a análise de imagens de biópsia com o uso de ferramentas de visão computacional a fim de detectar a presença ou ausência de células malignas do câncer de mama.

2 Motivação

A visão computacional é uma das áreas que envolve o aprendizado de máquina que tem crescido muito no últimos anos, no qual o objetivo é simular e replicar a visão humana através do uso de hardware e software. A construção de um modelo de visão computacional consiste em obter, analisar e processar os dados que serão utilizados para que seu modelo consiga aprender a informação que contém na imagem.

Na área médica, a visão computacional tem sido muito utilizada para o diagnóstico de pacientes através da análise de imagens médicas, como por exemplo o raio-x, pois de acordo com as características presentes naquela imagem é possível identificar uma possível anomalia, e a partir disso, realizar o diagnóstico.

O câncer de mama é o tipo da doença mais comum entre as mulheres no mundo e no Brasil, depois do de pele não melanoma, corres-

pondendo a cerca de 25% dos casos novos no mundo a cada ano. No Brasil, esse percentual é de 29%, no qual para o ano de 2018 foram estimados 59.700 novos casos[1].

Para realizar o diagnóstico, a confirmação da doença só pode ser feita através de biópsia, no qual consiste na retirada de uma pequena amostra das células, que são analisadas por um especialista através de um microscópio, com o intuito de identificar se são malignas ou benignas.

3 Objetivo

O objetivo desse projeto é realizar uma análise experimental de diferentes modelos que sejam capazes de identificar a presença ou a ausência de células malignas em imagens que contém o tecido mamário retirado através da biópsia, classificando-as como maligna ou benigna.

4 Metodologia

O *dataset* utilizado por esse projeto é o Breast Cancer Wisconsin (Diagnostic), que foi adquirido em [2], no qual contém dez características sobre cada núcleo de célula presente nas imagens obtidas através da biópsia.

Uma descrição detalhada sobre cada característica pode ser encontrada em [2].

As dez características são: (i) Raio; (ii) Textura; (iii) Perímetro; (iv) Área; (v) Suavidade; (vi) Compacidade; (vii) Concavidade; (viii) Pontos côncavos; (ix) Simetria; (x) Dimensão fractal.

Para cada uma dessas características, foram calculadas a média, erro padrão e o pior (média dos três maiores valores), ou seja, cada amostra do *dataset* possui no total 30 atributos. O *dataset* possui 569 amostras, onde 357 são amostras benignas e 212 malignas.

Os experimentos serão realizados utilizando diferentes classificadores, métodos de divisão dos dados e métodos de redução de dimensionalidade, que por fim, serão comparados e discutidos os resultados obtidos.

Referências

- [1] “Câncer de mama.”
<https://www.inca.gov.br/tipos-de-cancer/cancer-de-mama>. Accessed: 2019-11.
- [2] D. Dua and C. Graff, “UCI machine learning repository.”
<https://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Wisconsin+%28Diagnostic%29>, 2017.