

Universidade Federal de Minas Gerais
Departamento de Ciência da Computação

DCC057 Mineração de Dados
Trabalho Prático 3 — Relatório de Descrição dos Dados

Alexander Thomas Mol Holmquist
16 de março de 2021

1 Objetivo

Este relatório tem como objetivo clarificar aspectos sobre a descrição dos dados, principalmente em vista do plano inicial do projeto.

2 Percepções

Rapidamente ficou claro que os passos delineados pelo plano inicial do projeto são fruto de inexperiência com mineração de imagens. Por exemplo, é dito no plano inicial, como recomendação para o passo 1 da segunda fase, que deve-se gerar métricas como a média, variância, etc. Porém, isto não é interessante para imagens, porque o número de atributos é grande demais para permitir que uma análise individualizada dos dados seja feita..

Outro equívoco foi recomendar a produção de uma figura ilustrando o banco de dados de dimensionalidade reduzida, especificamente de duas dimensões. Novamente, se durante o desenvolvimento de tal plano, houvesse um entendimento básico da forma de um banco de dados de imagens, este passo não teria sido recomendado. A razão é a mesma do item anterior. Não é possível condensar uma imagem, que tem milhares e milhares de atributos, em poucos componentes principais.

3 Resultados

Apesar das falhas mencionadas, foi possível gerar duas saídas úteis para o projeto.

1. Proporção de imagens de câncer benigno e maligno

```
In [3]: showSeparation(classes)
Number of benign cases: 1440 (54.61%)
Number of malignant cases: 1197 (45.39%)
```

Figura 1: Proporção benigno/maligno

2. Exemplo de imagem para ambos os casos

```
In [4]: showOriginalImagesBenMal("train", saveDir="doc/relatorios/fase-2")
```

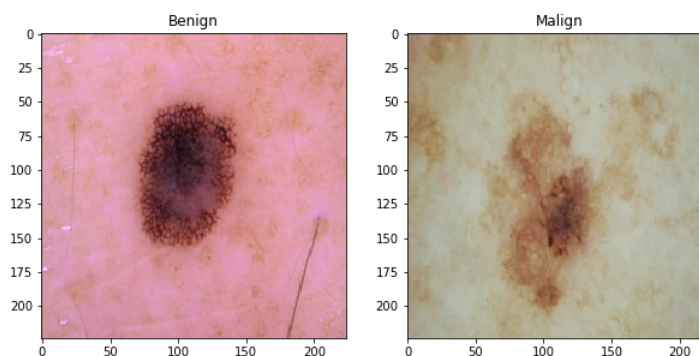


Figura 2: Exemplos de imagem para cada caso

Esta última imagem com os dois casos foi também salvo no diretório "doc/relatorios/fase-2", junto com o presente relatório.