**Nome: Marcus Vinicius Oliveira Nunes**

**Assunto: Relatório do projeto final**

**RA:2554100**

**1 - Objetivo do Trabalho**

Desenvolver um modelo de aprendizado de máquina capaz de distinguir as tomografias torácicas de cânceres pulmonares e uma tomografia torácica normal.

**2 - Base de dados**

A base de dados utilizada para o desenvolvimento do trabalho foi a “Chest CT-Scan images Dataset” que é basicamente uma base de dados com imagens de tomografia computadorizada do tórax. As imagens são divididas em 3 classificações de câncer no pulmão e outra classificação de sem câncer. Os tipos de classes são:

* **Adenocarcinoma:** 338 imagens
* **Carcinoma de células grandes:** 187 imagens
* **Carcinoma de células escamosas:** 224 imagens
* **Normal:** 251 imagens

Do total de 1000 imagens foram divididas em 70% para treinamento, 20% para teste e 10% para validação.

**2.1 - Data augmentation ou aumento de dados**

Afim de aumentar a capacitação do modelo no reconhecimento das imagens foi aplicado o “Data Augmentation” que é o aumento dos dados durante o treinamento aplicando mudanças nos dados que já existem na base de dados. Um exemplo de mudança é virar a imagem de cabeça para baixo, mudar a imagem de lado, aplicar zoom, mudar o eixo etc. Algumas das mudanças aplicadas neste trabalho estão abaixo:

