**PONTIFÍCIA UNIVERSIDADE CATÓLICA DO RIO DE JANEIRO**

**Ferramenta de auxilio a analise oncológica de mamografias por machine learning**

**Guilherme Freitas de Araujo**

**Proposta Projeto Final I**

**CENTRO TÉCNICO CIENTÍFICO - CTC DEPARTAMENTO DE INFORMÁTICA**

Curso de Graduação em Engenharia da Computação

Rio de Janeiro, Junho de 2018



**Guilherme Freitas de Araujo**

**Ferramenta de auxilio a analise oncológica de mamografias por machine learning**

Proposta de Projeto de Conclusão de Curso, apresentado ao curso de **Engenharia de Computação** da PUC-Rio como requisito parcial para a obtenção do Bacharel em Engenheiria de Computação.

**Orientadora**:

Prof. Marley Maria B. R. Vellasco

**Coorientador**:

Prof. Italo de Oliveira Matias

Rio de Janeiro  
Junho de 2018

**Resumo**

Araujo, Guilherme; Vellasco, Marley; Matias, Italo. Ferramenta de auxilio a analise oncológica de mamografias por machine learning. Rio de Janeiro, 2018. 7p. Proposta de Projeto de Conclusão de Curso - Departamento de Informática. Pontifícia Universidade Católica do Rio de Janeiro.

Este trabalho tem como objetivo desenvolver uma ferramenta de auxilio ao diagnostico medicinal utilizando redes convolucionais para analise de tumores em imagens mamograficas.

**Palavras-chave**

Redes convolucionais; Inteligência artificial; Aprendizagem profunda;

**Abstract**

Araujo, Guilherme; Vellasco, Marley; Matias, Italo. Application to aid the oncological analysis of mammograms by machine learning. Rio de Janeiro, 2018. 7p. Capstone Project Report – Department of Informatics. Pontificial Catholic University of Rio de Janeiro.

This Project aims the development of a tool to assist medical diagnosis using concolutional networks for the analysis of tumors in mammographic images.

**Keywords**

Convolutional Networks; Artificial Intelligence; Deep Learning;

**Sumário**

[1 Introdução 6](#_Toc510541263)

[2 Situação Atual 7](#_Toc510541264)

[3 Propostas e Objetivos dotrabalho 8](#_Toc510541265)

[4 Plano de Ação 9](#_Toc510541266)

[5 Referências bibliográficas 10](#_Toc510541267)

# Introdução

Câncer de mama é o segundo tipo de câncer com mais ocorrência em mulheres nos Estados Unidos da America (EUA) [1] e estima-se que por volta de 268,000 novos casos serão diagnosticados no ano de 2018 nos EUA e que mais ou menos 41,000 pessoas morrerão devido a essa patologia nessa mesma época[1]. A analise clinica de câncer de mama normalmente envolve o uso de diagnostico por imagem e a mamografia é umas das técnicas utilizadas na detecção precoce do câncer de mama [1][2]. Mamografia é um exame de rastreio por imagem utilizando raios-X que permite visualmente analisar o tecido mamário [1]. Recomenda-se que mulheres acima dos 40 anos, ou consideradas dentro do grupo de alto risco de câncer mamário, façam exames de mamografia anualmente, pois a detecção precoce desse câncer permite um tratamento eficiente para cura do mesmo [1]. Diferentes características de tumores, tecidos e o erro humano podem causar diagnósticos errados dentro da área de oncologia [3].

O objetivo desse projeto é desenvolver uma ferramenta de auxilio ao profissional de medicina na analise de câncer de mama. Essa ferramenta utilizara redes convolucionais com aprendizagem profunda para classificar uma ou mais mamografias dentro da categoria de avaliação BI-RADS do American College of Radiology (ACR) [2]. Essa ferramenta facilitaria e aceleraria o processo de exame das imagens de raios-X resultantes da mamografia para detecção de categorização de tumores.

# 

# 2 Situação Atual

Estima-se que uma a cada cinco mamografias resultam em falsos negativos [1] e estima-se que 7% a 9% de mulheres que realizam mamografias anualmente receberam um resultado falso-positivo recomendando a etapa de biopsia [4].

# Propostas e Objetivos do trabalho

O projeto tem como objetivo desenvolver uma ferramenta que classifique diferentes imagens oriundas de mamografias no formato DICOM [5] de acordo com a classificação BI-RADS do ACR.

Será utilizada uma rede convolucional, ideal para classificação de imagens, pois dispensa boa parte do pré-processamento da base de dados. Algumas arquiteturas diferentes serão estudadas para se decidir qual deve ser abordada no projeto, utilizando benchmarks de classificação de imagens como CIPHAR-10 [6] e ILSVRC [7]. Diferentes configurações dos parâmetros da rede serão testadas visando aperfeiçoar a classificação resultante da rede.

Deseja-se ter ao final do projeto uma aplicação que possa servir para uso clinico real auxiliando profissionais de medicina no dia-a-dia, e ter uma rede robusta que sirva como fonte de estudo para futuro desenvolvimento relacionando oncologia e processamento de imagens com inteligência artificial.

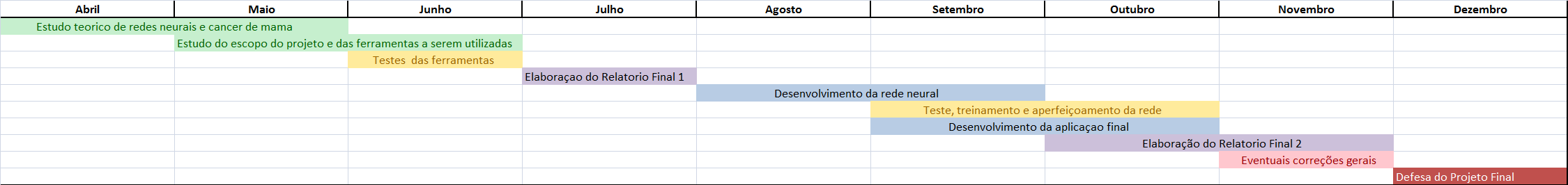
# Plano de Ação

A primeira fase do projeto foi de estudo teórico da área de câncer de mama e de redes convolucionais para machine learning com imagens. A segunda fase envolveu a decisão de que linguagens, ferramentas e bibliotecas seriam utilizadas no desenvolvimento do projeto.

Python foi linguagem escolhida para o desenvolvimento da rede neural utilizada, pelo fato de ser uma linguagem que possui uma grande quantidade de ferramentas relacionadas a processamento de imagens e redes neurais. A principal biblioteca a ser utilizada será o TensorFlow [8], um framework para machine learning open source criada pela Google.

O segundo semestre envolve todo o desenvolvimento, treinamento, teste e aperfeiçoamento da rede utilizando a base de dados. Além do desenvolvimento do aplicativo que utilizara a rede para processar as imagens carregadas no programa. E por fim redigir o relatório final e apresentar o projeto para banca.

**Cronograma do projeto**



# Referências bibliográficas

1. <https://www.cancer.org/cancer/breast-cancer/>
2. <https://cancerstatisticscenter.cancer.org/?_ga=2.251561797.238454592.1529880382-792691082.1529880382#!/>
3. <https://www.breastcancer.org/symptoms>
4. <http://acpjc.acponline.org/Content/119/2/issue/ACPJC-1993-119-2-038.htm>
5. <https://www.ncbi.nlm.nih.gov/pubmed/22824119>
6. <https://www.acr.org/Clinical-Resources/Reporting-and-Data-Systems/Bi-Rads>
7. <https://onlinelibrary.wiley.com/doi/abs/10.1002/cncr.30245?systemMessage=Wiley+Online+Library+will+be+unavailable+on+Saturday+3rd+September+2016+at+08.30+BST%2F+03%3A30+EDT%2F+15%3A30+SGT+for+5+hours+and+Sunday+4th+September+at+10%3A00+BST%2F+05%3A00+EST%2F+17%3A00+SGT+for+1+hour++for+essential+maintenance.+Apologies+for+the+inconvenience>
8. <https://www.forbes.com/sites/janetwburns/2016/08/29/artificial-intelligence-can-help-doctors-assess-breast-cancer-risk-thirty-times-faster/#7328a2165908>
9. <https://www.ncbi.nlm.nih.gov/pubmed/22007042>
10. <https://www.mitpressjournals.org/doi/pdf/10.1162/neco_a_00990>
11. <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0100-39842004000400005&lng=pt&tlng=pt>
12. https://www.sciencedirect.com/science/article/pii/S1532046417300813