



Evento	Salão UFRGS 2019: FEIRA DE INOVAÇÃO TECNOLÓGICA DA UFRGS - FINOVA
Ano	2019
Local	Campus do Vale - UFRGS
Título	Aprendizado de máquina aplicado à detecção e classificação de células tumorais para a construção de árvores de parentesco
Autores	ANGELO LUIZ ANGONEZI JULIANO LUIZ FACCIONI
Orientador	GUIDO LENZ

RESUMO

TÍTULO DO PROJETO: Aprendizado de máquina aplicado à detecção e classificação de células tumorais para a construção de árvores de parentesco.

Aluno: Angelo Luiz Angonezi

Orientador: Guido Lenz

Co-orientador: Juliano Luiz Faccioni

A heterogeneidade tumoral é um dos grandes fatores que limita o sucesso de terapias contra o câncer. Em um tumor, subpopulações de células podem resistir à terapia e causar recidiva após terapia, mesmo quando compõem uma pequena parcela da massa tumoral inicial. Podemos estudar a heterogeneidade tumoral em culturas celulares *in vitro* ao acompanharmos a dinâmica da formação de colônias a partir de células únicas. Observando aspectos como migração, divisão e expressão de marcadores de interesse, podemos gerar árvores de parentesco entre células e, a partir delas, extrair informações sobre a progressão da heterogeneidade em conjuntos de células tumorais, seja relativo a níveis de expressão de alguma proteína ou às mudanças nas taxas de divisão, por exemplo. A detecção e classificação de células tumorais em imagens de microscopia ainda se prova um desafio. Existem equipamentos capazes de capturar imagens sequenciais de culturas celulares por um longo período de tempo, o que favorece a construção de árvores de parentesco entre células. Nesse processo, é gerada uma enorme quantidade de dados, dificultando a contagem e classificação manual das imagens. Laboratórios que realizam contagem e rastreamento de forma manual estão sujeitos a um processo vagaroso e propenso a erros. Embora existam programas que busquem automatizar esse processo, eles apresentam limitações em casos mais complexos, como microscopia de luz visível ou células fortemente agrupadas. Assim, a automatização da análise de imagens de microscopia celular se faz necessária para possibilitar a identificação de células e análise de árvores de parentesco entre células tumorais. Algoritmos de aprendizado de máquina são uma excelente ferramenta para detectar objetos em imagens, visto que já existem programas que facilitam sua aplicação em distintos conjuntos de dados. O projeto em desenvolvimento tem por objetivo aplicar algoritmos de aprendizado de máquina para detectar e classificar células tumorais em imagens de microscopia, com o intuito de automatizar o processo de identificação de células em múltiplas imagens de maneira rápida e eficaz, e possibilitar a construção de árvores de parentesco em análises posteriores. O modelo de aprendizado de máquina foi treinado na linguagem de programação Python, utilizando o pacote `keras-retinanet` (<https://github.com/fizyr/keras-retinanet>). Python foi escolhido por ser uma das principais linguagens aplicadas no campo de aprendizado de máquina, e por familiaridade dos alunos com a linguagem. Para o treinamento e validação de modelos de aprendizado de máquina, é necessário ter um conjunto de dados conhecido e previamente identificado. Como dados de treinamento,

foram utilizadas marcações manuais em imagens da linhagem de glioblastoma A-172, obtidas no microscópio Lux2 (CytoSMART, Eindhoven, Países Baixos). Este equipamento é capaz de capturar imagens em intervalos regulares e operar dentro de uma incubadora e, portanto, manter estáveis as condições de cultivo celular (temperatura a 37°C e atmosfera com 5% CO₂). Utilizando a ferramenta LabelImg (<https://github.com/tzutalin/labelimg>), foram geradas 1335 marcações de células utilizando 30 imagens diferentes. Estes dados foram transformados e entregues para o pacote keras-retinanet, o qual foi treinado por 22 épocas (uma época corresponde à conclusão de uma etapa em que o modelo é treinado sobre todo conjunto de dados). Este treinamento gerou um programa capaz de identificar células e evitar outras informações na imagem. Os erros mais comuns foram a detecção de uma mesma célula múltiplas vezes e a não identificação de células, de modo que a contagem das mesmas se torne imprecisa. A fim de quantificar a taxa de falsos positivos e falsos negativos das detecções do modelo, bem como refinar sua capacidade de reconhecimento de células, um dos próximos objetivos é a marcação manual de um maior conjunto de dados, seguido de novas rodadas de treinos e validação. Desse modo, espera-se aprimorar o modelo, possibilitando além da detecção, a classificação de células de acordo com a sua morfologia, ou ainda aplicá-lo de maneira conjunta a outros programas como o *Family Tree Explorer* (FATE), para a construção de árvores de parentesco entre células.