

## Segmentação e pareamento automático de cromossomos para auxiliar em exames de cariótipo

Renan Gomes Miranda Costa

Prof. Dr. Luciano Antonio Digiampietri

Escola de Artes, Ciências e Humanidades / Universidade de São Paulo

[renancosta921@usp.br](mailto:renancosta921@usp.br) ; [digiampietri@usp.br](mailto:digiampietri@usp.br)

### Objetivos

Com o avanço da computação nas últimas décadas, diferentes métodos computacionais foram ou estão sendo desenvolvidos para auxiliar nas mais diferentes tarefas que, até então, eram desempenhadas exclusivamente por pessoas. Uma das áreas que se desenvolveu bastante foi a bioinformática que combina técnicas e métodos da computação, matemática e estatística para analisar diferentes dados biológicos (LESK, 2019). Dentre as diversas atividades relacionadas à bioinformática, o presente projeto está focado na Análise de Bioimagens (BLANKHEAD, 2022), isto é, a análise computacional de imagens biológicas, neste caso, imagens usadas em exames de cariótipo. O processo de cariotipação (do inglês, *karyotyping*) consiste em, a partir de uma imagem dos cromossomos, organizá-los e pareá-los para, em seguida, realizar a análise e possíveis diagnósticos iniciais (PONGALI e THAMMINENI, 2018). Este projeto visou a auxiliar nas primeiras tarefas desse processo, especificamente na segmentação automática dos cromossomos (DEVARAJ, 2022), na extração de características e no pareamento automático dos cromossomos a partir de imagens de exames de cariótipo.

### Métodos e Procedimentos

A segmentação de imagens foi realizada por meio da binarização da foto e aplicando um filtro *thresholding*. A partir disso, utilizou-se de erosão e dilatação para corrigir a foto. Após esse tratamento, foi executado um algoritmo, desenvolvido neste projeto, para segmentar a imagem nos diferentes cromossomos. Para a segunda etapa do projeto, utilizou-se um conjunto de dados já segmentado e anotado do *Laboratory of Biomedical Imaging* (<https://bioimlab.dei.unipd.it/Chromosome%20Data%20Set%204Class.htm>). A partir dele, extraiu-se características ligadas ao tamanho do cromossomo e características envolvendo os tons de cinza presentes na foto. Com estas informações, utilizou-se regressores e classificadores para tentar identificar o número do cromossomo e para tentar pareá-los.

### Resultados

Diferentes regressores foram utilizados para a identificação do cromossomo. O uso de regressores se justifica pelo fato dos cromossomos humanos serem numerados de acordo com seu tamanho (de forma decrescente), com exceção dos cromossomos sexuais. Por este motivo, para os regressores, foram excluídos dos conjuntos de dados os cromossomos sexuais. A Figura 1 ilustra a variação do campo Tamanho Relativo de cada cromossomo em relação ao maior cromossomo de cada pessoa do conjunto de dados.

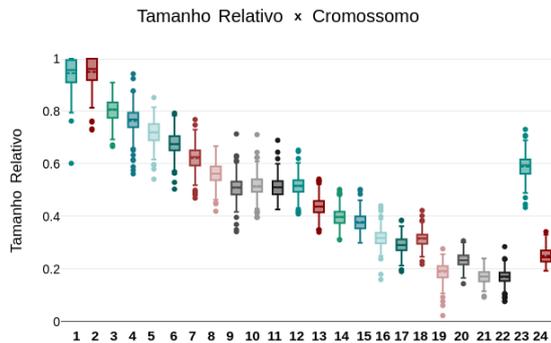


Figura 1: Box-plot do valor do campo Tamanho Relativo de cada cromossomo

A regressão linear obteve um erro absoluto médio de 1,8185. O tamanho relativo foi aquele que recebeu mais peso da regressão. O regressor que apresentou o melhor resultado foi o *Gradient Boosting*, com um erro absoluto médio de 1,625. Para a classificação, os melhores resultados foram obtidos com Florestas Aleatórias. Observou-se que, a maioria dos erros de classificação ocorreu por apenas uma unidade. Ou seja, o classificador acertou o número do cromossomo ou errou em apenas uma unidade, cerca de 70%.

Além dessa abordagem, utilizou-se, também, o pareamento automático no projeto. A identificação do cromossomo poderia ser usada para realizar o pareamento (parear dois cromossomos com a mesma classificação), porém também é possível extrair características da comparação entre pares de cromossomos e desenvolver um modelo de classificação que verifique se este par deve ou não ser pareado (isto é, se corresponde ao mesmo tipo de cromossomo). Os melhores resultados para a classificação de pares de cromossomos, considerando a validação cruzada em dez subconjuntos, foram obtidos usando Florestas Aleatórias, com acurácia igual a 71,5%.

## Conclusões

O presente projeto teve por objetivo auxiliar no problema de cariotipação, por meio da segmentação de imagens contendo

cromossomos, identificação dos cromossomos e pareamento dos mesmos.

Consideramos os resultados bastante satisfatórios, pois foi possível obter segmentações adequadas para imagens que não possuem cromossomos sobrepostos (sobreposições não foram tratadas no presente projeto). Quanto à identificação do número do cromossomo, apesar de o modelo desenvolvido não ter uma acurácia muito boa, observou-se que, na média, o sistema erra essa identificação por 1,625 pontos. Já na classificação para o pareamento propriamente dito, sem identificar previamente a numeração de cada cromossomo, foi obtida a acurácia de 71,5%. Conforme ilustrado na Figura 1, há sobreposições significativas entre os tamanhos dos cromossomos de forma que estão nas fotos, o que faz a tarefa de classificação focada nessa característica bastante desafiadora.

Como trabalhos futuros, sugere-se explorar características adicionais, como o tratamento de cromossomos que aparecem de lado ou sobrepostos nas imagens. Adicionalmente, notou-se a necessidade de um tratamento específico para os cromossomos sexuais (números 23 e 24 na Figura 1).

## Referências

- Bankhead, Pete. Introduction to Bioimage Analysis, 2022. <https://bioimagebook.github.io/index.html>
- Devaraj, Somasundaram; Madian, Nirmla; Suresh, S.. Mathematical approach for segmenting chromosome clusters in metaspread images. *Experimental Cell Research* 418.1, 113251, 2022.
- Lesk, Arthur. Introduction to bioinformatics. Oxford University Press, 2019.
- Raghavendra, Pongali; Pullaiah, Thammineni. Chapter 3 - Advancements in Genetic Applications for Cellular and Molecular Diagnostics in *Advances in Cell and Molecular Diagnostics*, Academic Press, 2018, Pages 57-84, ISBN 9780128136799, <https://doi.org/10.1016/B978-0-12-813679-9.0003-8>