

# Uma Revisão Sistemática das Técnicas de Justiça Algorítmica para Diagnóstico Radiológico: Avanços, Desafios e Perspectivas Futuras

Lucas Freire de Lima<sup>1</sup>, Luiz Fernando F. P. de Lima<sup>2</sup>, Maristela de Freitas Riquelme<sup>1</sup>,  
Danielle Rousy Dias Ricarte<sup>1</sup>

<sup>1</sup>Centro de Informática – Universidade Federal da Paraíba (UFPB) - João Pessoa, PB, Brasil

<sup>2</sup>Centro de Estudos e Sistemas Avançados do Recife (CESAR) - Recife, PE, Brasil

[lucas.freire@estudantes.ufpb.br](mailto:lucas.freire@estudantes.ufpb.br), [danielle@ci.ufpb.br](mailto:danielle@ci.ufpb.br),  
[lffpl@cesar.org.br](mailto:lffpl@cesar.org.br), [maristela.riquelme@academico.ufpb.br](mailto:maristela.riquelme@academico.ufpb.br)

**Abstract.** *Algorithmic fairness has gained prominence in the area of radiographic diagnostics, where artificial intelligence (AI) algorithms are applied to assist doctors in interpreting and diagnosing medical images. This systematic literature review addresses the current state of algorithmic fairness research in this context, investigating which techniques are on the rise associated with the use of AI algorithms for radiological diagnosis.*

**Resumo.** *A justiça algorítmica tem ganhado recente destaque na área de diagnóstico de radiografias, onde algoritmos de inteligência artificial (IA) são aplicados para auxiliar médicos na interpretação e diagnóstico de imagens médicas. Esta revisão sistemática da literatura aborda o estado atual da pesquisa em justiça algorítmica nesse contexto, investigando quais as técnicas em ascensão associadas ao uso de algoritmos de IA para diagnóstico radiológico.*

## 1. Introdução

Em um período de 10 anos, as publicações de IA em radiologia não passavam de um pouco mais de 100 por ano até 2016, para 700 a 800 artigos publicados entre 2016 e 2017 [Pesapane, Codari e Sardanelli 2018]. Porém, pela natureza sensível da área da medicina, desafios para a implementação de modelos de IA têm surgido ao longo do tempo, um desses desafios sem dúvida é a presença de viés no âmbito dos modelos de aprendizagem de máquina (AM), conscientemente ou não por seus projetistas [Macht 2022]. Erros médicos de diagnóstico por viés cognitivo são uma das causas principais de morbidade e mortalidade na área [Bubsy, Courtier e Glastonbury 2018], originando a necessidade de sistemas justos e que consigam mitigar tais comportamentos.

Dado tais preocupações, a justiça algorítmica ou equidade algorítmica (em inglês, *Algorithmic Fairness*), se é utilizada para encontrar, reconhecer e mitigar benefícios indevidos e discrepâncias entre atributos considerados sensíveis, resultando em um modelo que possa conseguir corrigir discriminações a sub-grupos dentro dos conjuntos de dados pelo seu sexo biológico, idade ou etnia, por exemplo [Dolata, Feuerriegel e Schwabe 2022]. Porém apesar de obtermos muitos artigos sobre Justiça Algorítmica em AM voltado ao direito ou educação, por exemplo, há uma baixa quantidade de artigos que analisam o tópico na área de classificação de imagens [Correa et al 2022].

Este estudo propõe investigar o conceito de justiça algorítmica, identificando técnicas para mitigar problemas sócio-algorítmicos na análise de imagens médicas que

possam nos mostrar a qualidade de artigos da área. Este artigo é parte integrante de um projeto de iniciação científica mais amplo, que tem como objetivo explorar e analisar a relevância dos dados de imagens médicas na criação de ferramentas de AM justas. Além desta introdução, este artigo expõe a metodologia da revisão sistemática da literatura, estratégias para a pesquisa, processo de seleção de artigos, resultados nos quais os artigos nos mostram em consonância com a metodologia apresentada, e as conclusões finais sobre todo o processo de realização desta revisão.

## **2. Metodologia**

Esta seção apresenta como o processo de revisão sistemática da literatura [Kitchenham 2004] foi realizado para a identificação do estado da arte em técnicas de justiça algorítmica em modelos de aprendizagem de máquina, assim como *insights* sobre quais técnicas serão mais apropriadas para a construção e condução do modelo. Ela apresenta também os devidos objetivos a serem atingidos, engenhos de busca selecionados, strings, critérios da revisão e resultados colhidos a partir da coleta de artigos da revisão sistemática, sendo, o autor principal deste artigo, responsável pela pesquisa e os autores secundários como revisores do processo.

### **2.1 Questões de Pesquisa**

A Revisão Sistemática da Literatura (RSL) em questão tem como objetivo apontar as tecnologias e aplicações de justiça algorítmica que possam contribuir com o desenvolvimento e aprimoramento de modelos de AM justos para diagnósticos médicos. A revisão sistemática se apoia nas seguintes questões norteadoras: (QP01) Quais são os conjuntos de dados (do inglês *datasets*) utilizados na realização da pesquisa? (QP02) Quais são os conceitos de justiça algorítmica utilizados pela pesquisa deste artigo? (QP03) Quais as métricas de avaliação utilizadas pela pesquisa? (QP04) Quais as técnicas de AM para obtenção de modelos mais justos utilizadas pela pesquisa? E (QP05) Quais as principais dificuldades para a obtenção de modelos de AM mais justos encontrados pela pesquisa? As questões buscaram extrair informações-chave contidas nos artigos de pesquisa encontrados, assim visando atingir o objetivo geral da RSL.

### **2.2 Estratégias de Busca**

Foram adotados na RSL como engenhos de busca: a ACM Digital Library, o IEEE Xplore, Scopus, Pubmed, Springer e ScienceDirect. A *string* utilizada no processo de busca foi: ("fairness" and "image classification" and "x-ray"). Foram escolhidos artigos em um período de 5 anos (2018 a 2023) e que em seu conteúdo estejam contidas técnicas de justiça algorítmica, discussões sobre viés em AM, e/ou que aborde o assunto de AM no diagnóstico de radiografias. Foram considerados apenas artigos em inglês. Todos os aspectos descritos representaram os critérios de inclusão de artigos para a RSL. Como critérios de exclusão foi estabelecido os pontos: artigos que não estão dentro do período de tempo selecionado, artigos que não discorrem sobre viés em AM

ou justiça algorítmica, *short papers* e artigos que não estejam na área de diagnóstico de imagens.

### 2.3 Seleção

O processo de seleção de artigos foi dividido em três partes: pesquisa nos engenhos de busca com a *string* estabelecida, a filtragem dos artigos por título em conformidade com os critérios adotados, e uma filtragem final através da leitura dos *abstracts* dos artigos restantes.

**Tabela 1. Etapas de seleção de artigos do processo de revisão sistemática.**

Bases	Pesquisa da string de busca nas bases	Filtragem de artigos por título	Filtragem de artigos por leitura de <i>abstracts</i>
ACM	12	2	0
IEEE	2	0	0
ScienceDirect	338	3	0
Springer	46	2	2
Pubmed	860	5	2
<b>Total</b>	1258	12	4

### 2.4 Resultados da pesquisa

Com a realização do processo de pesquisa, os seguintes artigos foram selecionados: (A1) *Addressing fairness in artificial intelligence for medical imaging* [Ricci Lara *et al.* 2022], (A2) *Algorithmic fairness in artificial intelligence for medicine and healthcare* [Chen *et al.* 2023], (A3), *SecureFed: federated learning empowered medical imaging technique to analyze lung abnormalities in chest X-rays* [Makkar e Santosh 2023] e (A4) *Improving model fairness in image-based computer-aided diagnosis* [Lin *et al.* 2023]. Os artigos A1 e A2 foram extraídos do engenho de busca Pubmed e A3 e A4 extraídos do Springer. Pode-se ver uma discrepância entre quantidade de publicações obtidas pela pesquisa da string nos engenhos de busca e na filtragem por título devido a não abordagem de conceitos de justiça algorítmica dos mesmos, focando apenas em métricas de performance, caindo em um de nossos critérios de exclusão.

## 3. Resultados

Nesta seção apresentamos os resultados revelados pela RSL, discriminando algumas técnicas, insights e desafios cruciais no desenvolvimento de ferramentas de aprendizado de máquina para análise de imagens médicas, fornecendo uma base para avançar na compreensão e na aplicação eficaz de algoritmos em diagnósticos clínicos, contribuindo assim para a melhoria dos cuidados de saúde.

### **3.1 (QP01) Quais são os conjuntos de dados utilizados na realização da pesquisa?**

Em relação aos conjuntos de dados, pode-se observar que o MIMIC Chest X-RAY é o mais comum em trabalhos nesta linha de pesquisa, sendo presente nos artigos A1, A2 e A4. O NIH Chest X-Ray está presente nos artigos A1 e A2 assim como o CheXpert, enquanto há uma aparição única de datasets em artigos como o Emory University Hospital Chest X-Ray no artigo A1 e MIDRC(COVID-19) em A4. O conjunto de dados de A3 é uma junção de três conjuntos de dados radiológicos abertos onde há presença de COVID-19 mais um conjunto de dados onde há presença de radiografias torácicas de pacientes saudáveis e pacientes com quadros de pneumonia. Nota-se que todos os conjuntos utilizados são open-source.

### **3.2 (QP02) Quais são os conceitos de justiça algorítmica utilizados pelo artigo?**

O artigo A1 denota que modelos de diagnóstico de imagens radiológicas são preferíveis a seguir o critério de Igualdade de Oportunidades, um critério de obtenção de algoritmos justos na área de AM no qual a taxa de verdadeiros positivos (TVPs) do modelo devem ser iguais para todos os subgrupos. Esta recomendação é seguida pelos artigos A2 a A4, dado ao fato de que a comparação de TVPs entre subgrupos são independentes de cálculos probabilísticos pré-teste.

### **3.3 (QP03) Quais as métricas de avaliação utilizadas pelo artigo?**

Há uma variedade de métricas utilizadas para a avaliação dos modelos de AM: A1 cita como exemplos a AUC (sigla para *area under ROC curve*) e acurácia. No artigo A2, são utilizados as métricas de acurácia, taxa de verdadeiros positivos (TVP), e taxa de falsos negativos (TFN). Em A3, são usados acurácia, a probabilidade de predição imediata para avaliar o desempenho de Modelos Ocultos de Markov (sigla para *Hidden Markov Models*, em inglês). Já A4, usa como métricas a AUC, a curva ROC (sigla para *Receiver Operator Characteristic*), e uma métrica baseada no conceito de Justiça Par-a-Par (*Pairwise Fairness*) chamado PFD (*Pairwise Fairness Difference*), que calcula a diferença das taxas de predição entre subgrupos.

### **3.4 (QP04) Quais as técnicas de AM para obtenção de modelos mais justos utilizadas pela pesquisa?**

O artigo A1 propõe técnicas utilizadas antes, durante e depois do treinamento do modelo: técnicas de reamostragem de dados para obter mais representatividade no conjunto de dados, técnicas de Aumento de Dados (*data augmentation*) e aprendizado

adversarial (*Adversarial Learning*), e métodos de pós-processamento e correção do modelo, respectivamente. A2 nos mostra técnicas que podem melhorar qualitativamente a justiça de um modelo com o uso de Aprendizado Federado, Aprendizado Representacional (*Representational Learning*) e Desentrelaçamento Representacional (*Representational Disentanglement*). A3 desenvolve um modelo de Aprendizado Federado utilizando Modelos Ocultos de Markov (HMMs, em inglês) para a obtenção de um sistema descentralizado na qual a possibilidade de vazamento de dados através do modelo resultando em vieses seja minimizado, o mesmo citado como um dos métodos apresentados pelo A2. A4 adota o Erro Marginal de Classificação (Marginal Ranking Loss) utilizando um ranqueamento para obter aumento de imparcialidade com uma mínima perda de desempenho, adequadamente quantificado através de técnicas de justiça par-a-par como o PFD.

### **3.5 (QP05) Quais as principais dificuldades para a obtenção de modelos de AM justos encontrados pela pesquisa?**

Todas as pesquisas encontradas relataram a dificuldade de balanceamento dos conjuntos de dados para obter modelos de AM justos. Tanto devido a não inclusão e discrepâncias entre a nacionalidade e ascendência de indivíduos, como a retratação histórica de má cobertura de atendimento médico de algum subgrupo representado nos dados. A1 também ressalta a falta de representatividade entre conjuntos e A2 cita presença de *Dataset Shift*, um desvio de representatividade de um conjunto da realidade ao longo do tempo devido a mudanças históricas e genéticas da população, assim como diferenças entre métodos de aquisição de imagens para formar esse conjunto.

### **3.6 Discussão**

Apesar de que, dada a quantidade reduzida, porém esperada, de artigos obtidos sobre justiça algorítmica, atesta-se que tais artigos conseguem obter uma excelente noção de vieses na área de diagnóstico de radiografias, assim como métodos para mitigá-los, demonstrando um crescimento da importância da justiça algorítmica por parte de projetistas de modelos de AM para minimizar aspectos de discriminação no âmbito da saúde, concedendo ao tópico sua devida importância e desenvolvimento acadêmico.

Como cada contexto de atuação de modelos de AM requer uma abordagem singular de justiça algorítmica, não há um esforço unificado da comunidade acadêmica para um desenvolvimento profundo do tema, afetando também a área da saúde, contendo poucos artigos sobre o tema, porém, assim como avanços históricos nos modelos de AM como Deep Learning demandaram esforços para comprovação de confiabilidade e imparcialidade de tais modelos, novos avanços tecnológicos do campo poderão dar margem a novas técnicas e conceitos de justiça algorítmica, tanto na área da saúde como em outros campos aplicáveis.

## **4. Conclusão**

Com base no processo de revisão sistemática da literatura, podemos identificar tecnologias na área de justiça algorítmica aplicadas ao diagnóstico de imagens radiológicas, que além de ajudarem a oferecer um tratamento de saúde justo a todos que precisam, podem servir de panorama atual para o estado acadêmico da justiça algorítmica.

Este artigo, sendo parte de um projeto de pesquisa maior em justiça algorítmica em diagnóstico de radiografias, terá como etapa seguinte a construção de semi-experimentos adotando algumas das técnicas identificadas nesta RSL a fim de avaliar as melhores práticas na mitigação de problemas sócio-algorítmicos.

## 5. Referências

- Chen, R.J. *et al.* (2023). Algorithmic fairness in artificial intelligence for medicine and healthcare. In *Nature Biomedical Engineering*, 7(6), 719–742. <https://doi.org/10.1038/s41551-023-01056-8>
- Correa, R., Shaan, M., Trivedi, H. (2022) A systematic review of ‘fair’ AI model development for image classification and prediction. In *J. Med. Biol. Eng.* **42**, 816–827. <https://doi.org/10.1007/s40846-022-00754-z>
- Dolata, M., Feuerriegel, S., e Schwabe, G. (2022). A sociotechnical view of algorithmic fairness. In *Information Systems Journal*, 32(4), 754–818. <https://doi.org/10.1111/isj.12370>
- Kitchenham, B. (2004). Procedures for performing systematic reviews. In, Keele Univ.. 33.
- Lin, M. *et al.* (2023). Improving model fairness in image-based computer-aided diagnosis. In *Nat Commun* **14**, 6261. <https://doi.org/10.1038/s41467-023-41974-4>
- Macht, B. (2022). Considering the potential impact of data bias on AI/ML and the Medical Device Ecosystem. In *Biomedical Instrumentation & Technology*, p. 127-129.
- Makkar, A., Santosh, K. (2023). SecureFed: federated learning empowered medical imaging technique to analyze lung abnormalities in chest X-rays. In *Int. J. Mach. Learn. & Cyber.* **14**, 2659–2670. <https://doi.org/10.1007/s13042-023-01789-7>
- Pesapane, F., Codari, M., & Sardanelli, F. (2018). Artificial intelligence in medical imaging: threat or opportunity? Radiologists again at the forefront of innovation in medicine. In *European radiology experimental*, 2(1), 35. <https://doi.org/10.1186/s41747-018-0061-6>
- Ricci Lara, M.A., Echeveste, R. e Ferrante, E. (2022). Addressing fairness in artificial intelligence for medical imaging. In *Nat Commun* **13**, 4581. <https://doi.org/10.1038/s41467-022-32186-3>
- Busby, L. P., Courtier, J. L., e Glastonbury, C. M. (2018). Bias in Radiology: The How and Why of Misses and Misinterpretations. *Radiographics : a review publication of the Radiological Society of North America, Inc*, 38(1), 236–247. <https://doi.org/10.1148/rg.2018170107>