XXXIII Congresso de Iniciação Científica da Unicamp

FaceShield: Defacing de Imagens Médicas em Tomografia Computadorizada

Palavras Chave: Tomografia computadorizada, Anonimização Facial, Segurança de Informação

Hospital Israelita Albert Einstein (HIAE)
Faculdade de Engenharia Elétrica e de Computação (FEEC)

Autores
João Vitor Souza de Alcantara [IMECC]
Profa. Leticia Rittner (Orientadora)[FEEC]
Bruna Garbes Goncalves Pinto [HIAE]

---

1 Introdução

O avanço da medicina diagnóstica e da pesquisa em saúde está intrinsecamente ligado à disponibilidade de grandes volumes de dados de alta qualidade. Imagens médicas, como as de Tomografia Computadorizada (TC), são fundamentais para o treinamento de algoritmos de inteligência artificial, estudos epidemiológicos e validação de novas técnicas clínicas. No entanto, a utilização desses dados impõe um desafio crítico e inegociável: a proteção da privacidade do paciente. Regulamentações rigorosas, como a Lei Geral de Proteção de Dados (LGPD) no Brasil e a Health Insurance Portability and Accountability Act (HIPAA) nos Estados , exigem a completa anonimização dos dados antes de seu compartilhamento [1].

Em exames de neuroimagem, como as TCs de crânio, a reconstrução 3D do volume de dados frequentemente expõe a face do paciente, constituindo uma violação direta dessas normas de privacidade. A remoção dessas características faciais identificáveis, processo conhecido como *defacing*, é, portanto, uma etapa indispensável para o uso secundário desses exames em pesquisa [2]. Contudo, o desafio reside em realizar essa remoção de forma precisa e segura, garantindo que nenhuma informação clinicamente relevante, como estruturas cerebrais, vasos ou tecidos moles adjacentes, seja comprometida.

Para endereçar esse desafio, este trabalho apresenta o FaceShield: um projeto que visa o desenvolvimento de um pipeline automatizado para o defacing de imagens de TC. O objetivo central é desenvolver um sistema que remova as características faciais de forma robusta, preservando ao máximo as áreas de interesse para a análise clínica e a pesquisa. Os algoritmos iniciais que deram inicio ao projeto são inspirados no trabalho de Felipe Kitamura (disponível em `https://github.com/kitamura-felipe`)

Nossa abordagem se baseia na utilização de um pipeline que integra técnicas clássicas de processamento de imagem e o uso de algoritmos de Deep Learning. A etapa inicial de segmentação facial é realizada através de operações de morfologia matemática, um método que remove as camadas mais externas da cabeça (pele e gordura) aplicando sequencialmente etapas de binarização, erosão e dilatação. Embora eficaz para a descaracterização, esta técnica apresenta o risco de remover informações anatômicas importantes devido à sua natureza subtrativa, e portanto precisa ter uma escolha de parâmetros cuidadosa.

2 Metodologia

2.1 Base de Dados

Os dados disponibilizados para esse projeto são principalmente provenientes do banco de imagens do Hospital Israelita Albert Einstein (HIAE). O conjunto de dados principal consiste em aproximadamente 1400 exames de TC de crânio em formato volumétrico. A amostra é composta por 58% (812) de sujeitos do sexo feminino e 42% (588) do sexo masculino. A faixa etária dos pacientes varia de 18 a 85 anos, com uma idade média de 45 anos e desvio padrão de ±18,44 anos. Estes dados foram previamente preparados e disponibilizados pela equipe Neuro do HIAE para uso em projetos de pesquisa.

2.2 Pipeline de Defacing Proposto

Este estudo propõe um pipeline de pré-processamento e anonimização de exames de TC do crânio, com o objetivo de proteger a identidade dos pacientes sem comprometer a integridade clínica das imagens. O fluxo metodológico está estruturado nas etapas descritas a seguir.

2.2.1 Aquisição e Pré-processamento Inicial

Os exames utilizados foram obtidos em formato DICOM e convertidos para o padrão NIfTI (.nii.gz), visando facilitar o manuseio e integração com bibliotecas de análise médica modernas. Durante essa conversão, foi aplicada uma padronização para a orientação RAS (Right-Anterior-Superior), e os volumes foram reamostrados para uma resolução isotrópica de 1 mm^3, de forma a garantir consistência espacial entre os exames e compatibilidade com as ferramentas subsequentes do pipeline. Essa etapa foi realizada utilizando um pipeline composto pelo *SimpleITK* e *Monai*.

Além disso, os exames foram submetidos inicialmente a um algoritmo de extração de métricas clínicas (com foco em volumetria intracraniana) que servirá posteriormente como referência para avaliação do impacto do processo de anonimização. O algoritmo em questao se trata de uma implementaçao para calcular o Volume intracraniano total (tecido cerebral, líquor e espaços vazios).

2.2.2 Anonimização via *Defacing*

Para assegurar a privacidade em exames tridimensionais de TC, o um pipeline automatizado de anonimização que atua diretamente na imagem volumétrica adquirida. A estratégia permite a supressão ou substituição da região facial, com flexibilidade no método de preenchimento.

O algoritmo recebe como entrada um volume tridimensional, podendo ser um diretório contendo múltiplos arquivos DICOM ou um único arquivo no formato NIfTI. As imagens são então convertidas para unidades Hounsfield (HU), utilizadas na quantificação da densidade dos tecidos anatômicos em exames de tomografia. Em seguida o algoritmo consiste na identificação da região a ser ocultada. Para tal, o volume é binarizado, segmentando voxels com densidade inferior a um limiar pré-definido (geralmente -800 HU), correspondentes ao ar. Em seguida, é extraída a maior componente conexa dessa máscara binária, presumivelmente representando a cavidade aérea ao redor da cabeça do paciente.

Posteriormente, aplica-se uma dilatação morfológica com elemento estruturante elíptico tridimensional, expandindo a região segmentada. A diferença entre o volume dilatado e a máscara original define a região facial a ser anonimizada.

2.2.3 Estratégias de Substituição (*Replacers*)

A substituição dos valores de voxel na região segmentada pode ocorrer por diferentes métodos, definidos via o parâmetro *replacer*. As principais estratégias implementadas são:

- *Face*: valores são amostrados aleatoriamente a partir de regiões do próprio volume identificadas como tecidos moles (e.g., pele, gordura subcutânea), de forma a preservar características de textura, mas eliminar qualquer informação anatômica facial identificável.
- *Air*: toda a região dilatada é preenchida com o valor correspondente ao ar (-1000 HU), resultando na completa exclusão da estrutura facial em visualizações tridimensionais.
- *Blur*: a região facial é suavizada por meio de um filtro gaussiano, com redefinição das áreas de ar para -1000 HU. Essa técnica preserva continuidade espacial, embora desfocalize detalhes anatômicos. No entanto, essa implementação ainda não apresentou resultados promissores nas etapas de avaliação da integridade.

Adicionalmente, o algoritmo permite a definição de valores constantes personalizados para preenchimento (e.g., -500 HU). Além de uma estrutura que facilita a inclusão de novos métodos para implementação.

2.2.4 Verificação Pós-Anonimização com Detecção Facial

Para avaliar a eficácia da anonimização na remoção de características faciais, é feita uma reconstrução 2D da região craniofacial, por projeção de intensidade máxima (MIP) ou corte axial central. Essa imagem é analisada por um sistema de detecção facial baseado no framework *InsightFace* [3] com a arquitetura RetinaFace. A detecção de qualquer estrutura facial indica que um rosto ainda pode ser reconhecido, enquanto a ausência de detecção sugere anonimização eficaz, sem feições reconhecíveis por visão computacional.

Além disso, realiza-se uma validação por reconhecimento facial profundo. As imagens pré e pós-anonimização são convertidas em embeddings de 512 dimensões pelo modelo *ArcFace*[4], e a similaridade é medida pela distância cosseno. Um limiar de 0,4, definido empiricamente, determina se há correspondência biométrica. Distâncias abaixo desse valor indicam possível falha na anonimização, enquanto valores maiores confirmam que o exame anonimizado não é reconhecível mesmo por sistemas avançados.

2.2.5 Avaliação da Integridade Clínica Pós-Anonimização

A preservação da utilidade clínica das imagens é um requisito fundamental em aplicações médicas. Portanto, após a aplicação do defacing, os exames passam por uma etapa de validação quantitativa que visa garantir a integridade das principais medidas volumétricas.

As métricas calculadas anteriormente ao processo de anonimização são recalculadas nas versões anonimizadas e comparadas. Dentre os parâmetros avaliados, destaca-se o Volume Intracraniano Total (ICV). Variações relativas superiores a 3% (limiar alterável, mas escolhido arbitrariamente para os experimentos) em qualquer uma das métricas-chave são consideradas indicativas de distorções indesejadas decorrentes do processo de anonimização, resultando na exclusão do exame da base de dados final.

3 Resultados preliminares

Os testes iniciais realizados com diferentes abordagens de preenchimento para *defacing* demonstram variações significativas quanto à preservação anatômica e à efetividade na ocultação de características faciais. A Figura 1 apresenta comparações visuais entre o exame original e três técnicas distintas: *Face*, *Blur* e *Air* com Substituição de Voxel.

A técnica de preenchimento *Face*, que substitui a região facial por uma máscara segmentada, remove com sucesso a identidade visual, mas pode impactar informações adjacentes, especialmente nas proximidades do nariz e mandíbula, o que pode comprometer pesquisas clínicas, a depender do contexto da

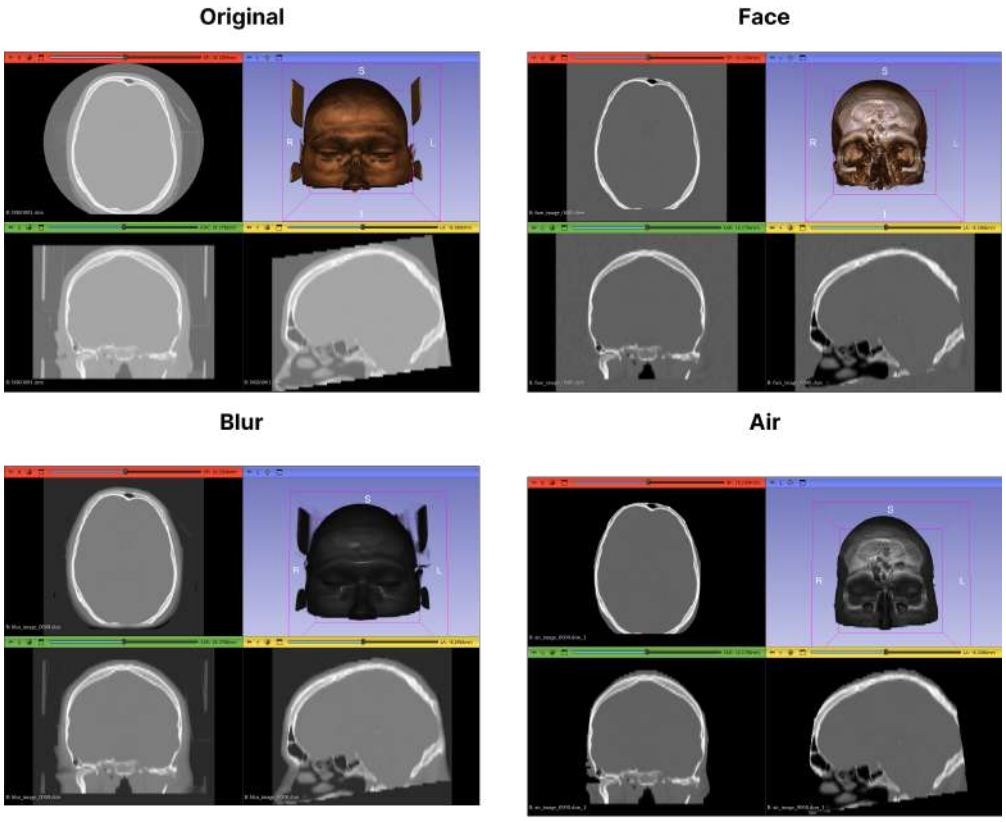

Figura 1: Comparação visual entre diferentes métodos de desidentificação facial em exames de tomografia computadorizada. O exame exibido na figura é um dado público disponibilizado pelo Slicer 3D. Dado original sem defacing (sup. esq.); defacing usando a técnica *face* (sup. dir.); defacing usando a técnica *blur* (inf. esq.); e defacing usando a técnica *air* (inf. dir.)

pesquisa. A técnica *Blur*, baseado em desfoque seletivo na região facial, oferece um bom equilíbrio entre privacidade e preservação estrutural. Com ajustes em parâmetros como o filtro (*Kernel*), a malha facial torna-se irreconhecível. No entanto, a intensidade do desfoque deve ser cuidadosamente calibrada para não comprometer a qualidade diagnóstica. Esse método segue em busca para tentar minimizar o impacto na qualidade. Por fim, a técnica *Air*, que substitui a região da face por valores de densidade equivalentes ao ar ($HU \simeq -1000$), remove quase completamente os traços faciais, mas pode introduzir descontinuidades abruptas no volume, perceptíveis em cortes sagitais e coronais. Esses artefatos podem dificultar o uso do volume em modelos de reconstrução tridimensional ou em algoritmos sensíveis a bordas.

A Tabela 1 apresenta a comparação dos volumes intracranianos (ICV) estimados, em milímetros cúbicos, e o número de voxels identificados como pertencentes ao volume cerebral, considerando diferentes estratégias de pré-processamento aplicadas a um mesmo volume de imagem de TC. O volume por voxel permanece constante entre as versões ($0,4667mm$), permitindo comparações diretas.

Observa-se que as técnicas de *air* e *blur* reduziram o volume intracraniano em relação à imagem original, com uma variação aproximada de 8% a 9%. Essa diminuição pode ser atribuída à exclusão de regiões periféricas ou artefatos de densidade similar ao tecido cerebral que não foram mais identificados após os respectivos processamentos. O efeito do desfoque, por exemplo, tende a suavizar contornos e bordas, o que pode comprometer a detecção de regiões finas ou com contraste marginal em relação ao fundo.

Tabela 1: Comparação do Volume Intracraniano (ICV) e número de voxels após diferentes processamentos

Versão	**ICV (mm^3)**	**Nº de Voxels**	**Volume por Voxel (mm^3)**
Original	1.962.265,06	4.204.111	0,4667
Air	1.813.532,66	3.885.455	0,4667
Face	3.593.833,25	7.699.711	0,4667
Blur	1.802.664,87	3.862.171	0,4667

A técnica de preenchimento *face*, nesta versão preliminar, apresentou comportamento inesperado, indicando necessidade de ajustes ou revisão do pipeline de substituição de voxels. Tal variação extrema sugere que a técnica de anonimização utilizada inseriu artefatos ou alterou significativamente a morfologia do volume analisado. É possível que regiões anteriormente não classificadas como intracranianas tenham sido erroneamente incluídas devido a modificações no gradiente de intensidade da imagem ou à introdução de estruturas com atenuações semelhantes ao tecido cerebral ou ósseo.

Esse resultado evidencia a necessidade de criteriosa avaliação das etapas de pré-processamento em estudos quantitativos de neuroimagem. Processamentos inadequados, especialmente os que envolvem *defacing*, podem comprometer seriamente a acurácia de medidas volumétricas, afetando a validade de análises morfométricas ou estudos longitudinais.

4 Considerações finais

O pipeline FaceShield contribui para a anonimização em larga escala de imagens médicas, buscando equilibrar segurança da informação e preservação clínica. Os testes reforçam os desafios de garantir anonimato sem comprometer a integridade anatômica, evidenciando a importância de validações rigorosas.

Técnicas de preenchimento para *defacing*, como *face*, que substituem a região facial por tecidos simulados, ocultam melhor feições visuais e biométricas, mas afetam significativamente métricas intracranianas. Já as técnicas de preenchimento *air* e *blur* preservam melhor a estrutura anatômica, ainda que com menor eficácia visual, demandando ajustes finos.

Como próximos passos, propõe-se investigar o impacto do *defacing* em algoritmos de inteligência artificial e explorar abordagens adaptativas com aprendizado profundo, visando aprimorar a anonimização com mínima perda de informações.

Referências

[1] Yanming Zhu, Xuefei Yin, Alan Wee-Chung Liew, and Hui Tian. Privacy-preserving in medical image analysis: A review of methods and applications, 2024. URL https://arxiv.org/abs/2412.03924.

[2] Scott A. Collins, Jing Wu, and Harrison X. Bai. Facial de-identification of head CT scans. 296(1): 22–22. ISSN 0033-8419, 1527-1315. doi: 10.1148/radiol.2020192617. URL http://pubs.rsna.org/doi/10.1148/radiol.2020192617.

[3] Jia Guo. Insightface: 2d and 3d face analysis project. https://github.com/deepinsight/insightface, 2021. Accessed: 2025-08-04.

[4] Jiankang Deng, Jia Guo, Niannan Xue, and Stefanos Zafeiriou. Arcface: Additive angular margin loss for deep face recognition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4690–4699, 2019.