Classificação de texturas em imagens médicas através de modelos generativos e aprendizado autossupervisionado

Leonardo C. Gomide[1], **Alexei M. C. Machado** [1]

[1]Departamento de Ciência da Computação - PUC Minas - Belo Horizonte, Brasil

emails leocgomide@gmail.com, alexeimcmachado@gmail.com

Resumo. *Este trabalho propõe um Variational Autoencoder regularizado para análise de textura, composto por um encoder, um decoder e um módulo preditor, com uma função de perda tripla que regulariza simultaneamente a codificação da imagem, sua reconstrução e classificação. Experimentos na análise de densidade mamográfica alcançaram níveis de precisão superiores aos de outros estudos publicados, sugerindo que o modelo pode fornecer uma representação latente mais separável e contribuir para a melhoria da análise de textura.*

Abstract. *This work proposes a regularized Variational Autoencoder for texture analysis, composed of an encoder, a decoder and a predictor module, with a triple loss function that simultaneously regularizes the image encoding, its reconstruction, and classification. The method presents state-of-the-art classification accuracy for mammographies and generates more separable latent spaces that may contribute to texture analysis.*

1. Introdução

O uso de modelos de Deep Learning (DL) na Medicina possibilita auxílio a diagnósticos mais precisos, rápidos e acessíveis. No entanto, tão importante quanto a precisão de um método é sua capacidade de explicar e interpretar suas descobertas. Essa questão torna-se ainda mais desafiadora na análise de padrões de textura. Nesse caso, mapas de ativação ou atenção são ineficazes, pois a *imagem inteira* é relevante para a classificação, em vez de apenas algumas regiões de interesse.

Um exemplo desse cenário é a detecção precoce e a prevenção do câncer de mama por meio da análise mamográfica. A escala mais comumente utilizada para classificar a densidade mamária é o *Breast Imaging Reporting and Data System* (BI-RADS), que avalia a densidade com base na proporção de tecido adiposo e fibroglandular. A classificação da densidade mamária é importante, pois mamas de classes III e IV podem ocultar pequenas lesões, levando à detecção tardia da doença. No entanto, esses padrões apresentam grande variação e podem resultar em diagnósticos divergentes, mesmo entre especialistas.

Este trabalho tem como objetivo propor um novo *Autoencoder* Variacional (VAE) regularizado com um módulo preditor, capaz de obter uma representação latente mais adequada para a classificação de texturas. Além de permitir uma categorização mais precisa, o espaço latente pode gerar uma explicação qualitativa para a análise.

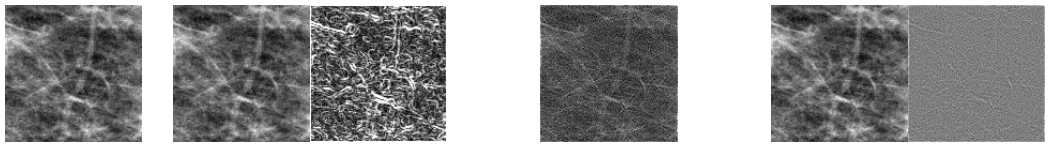
Uma solução para o problema da classificação do câncer de mama foi proposta por [Ha et al. 2018], utilizando uma rede residual e um *dataset* privado, alcançando acurácia de 0,72. Também usando imagens completas, [Zeiser et al. 2021] apresentaram um modelo de segmentação associado a mapas de calor interpretáveis. Poucos estudos abordam a análise de texturas BI-RADS usando recortes ao invés de imagens completas,

uma vez que isso oculta informações semânticas. [Lehman et al. 2019] apresentou um método para classificar imagens completas usando a ResNet-18 em 58.894 imagens de um banco de dados privado desbalanceado, no qual 86% das imagens pertenciam às classes com as maiores taxas de acerto, resultando em uma precisão ponderada de 67%. Já [Rungue 2019] aplicaram uma máquina de vetor de suporte (SVM) e diversos descritores de textura ao banco de dados público e balanceado IRMA de recortes mamográficos [Oliveira 2008], alcançando 69% de precisão.

No aspecto de interpretabilidade, [Liu et al. 2020] propuseram um método aplicado ao banco de dados MVTec-AD para obtenção de características visuais da representação latente de um VAE, no qual um mapa de atenção revelou o que cada variável latente representava na imagem de saída. A técnica foi posteriormente utilizada por [Song 2023] para demonstrar que era possível obter variáveis que representassem características reais da imagem.

2. Materiais e métodos

O conjunto de dados público utilizado nos experimentos foi fornecido pelo projeto IRMA [Oliveira 2008] e é composto por 5.024 segmentos de textura de 128×128 pixels extraídos de exames mamográficos, distribuídos uniformemente entre as quatro classes. Foram gerados quatro conjuntos distintos para o treinamento dos modelos: (a) *Simple*, contendo apenas os segmentos do conjunto original, sem qualquer pré-processamento; (b) *Sobel*, no qual o segmento é concatenado ao resultado da aplicação do filtro de Sobel; (c) *Added Laplacian*, composto pelas imagens originais somadas aos mapas de borda obtidos pelo filtro Laplaciano; e (d) *Laplacian*, no qual o segmento é concatenado à saída do filtro Laplaciano. Esses filtros foram utilizados como uma forma de aumentar a relevância das bordas das imagens nas amostras, com o objetivo de reduzir o efeito de suavização que os VAEs impõem à reconstrução. A Fig. 1 exemplifica os conjuntos de dados.



(a) Simple (b) Sobel (c) Added Laplacian (d) Laplacian

Figura 1. Exemplos de imagens dos *datasets* utilizados nos experimentos

A arquitetura do modelo proposto neste trabalho é composta por duas partes: um VAE e um módulo de predição, conforme detalhado na Fig. 2. O primeiro módulo utiliza um VAE treinado de forma autossupervisionada para obter uma representação latente das imagens. Essa representação latente também é passada para um módulo de predição, que deve retornar a classe da imagem. O modelo é treinado utilizando-se três funções de perda: a perda de reconstrução, calculada a partir do erro médio quadrático entre a imagem original e a imagem reconstruída; a perda por divergência de Kullback-Leibler (KL) entre a média e variância preditas [Kingma and Welling 2022]; e um termo de entropia cruzada entre a classe predita e a classe real (Fig. 2).

O efeito esperado da adição do módulo de predição e da perda de entropia cruzada a um modelo previamente não supervisionado é aumentar a regularização do espaço latente. O módulo de predição, ao atuar sobre as variáveis latentes, deve forçar a separação

entre as classes, priorizando características que um modelo de DL pode utilizar para prever a classe de uma imagem.

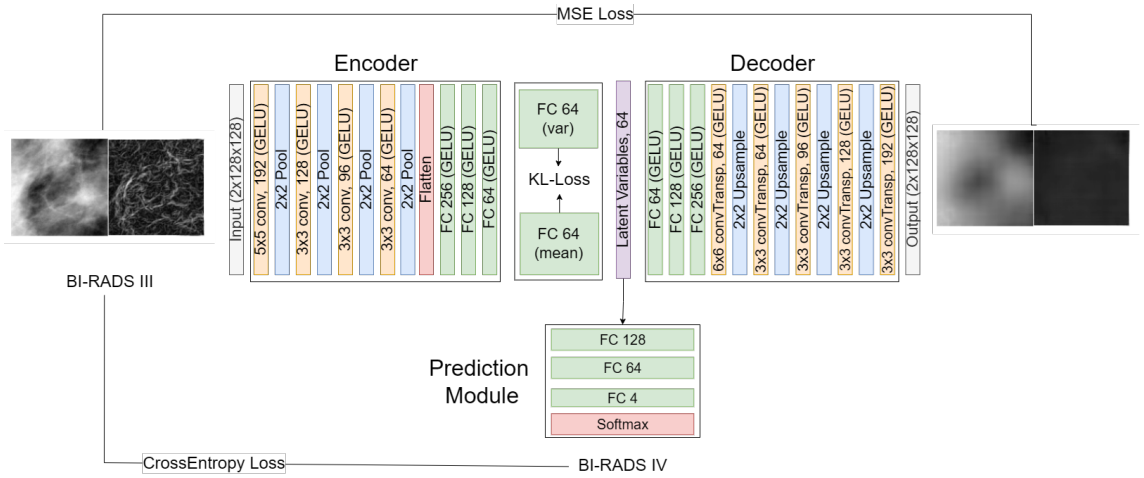


Figura 2. Arquitetura do modelo proposto. As camadas convolucionais utilizam *stride* 1 e *padding* válido.

As imagens foram divididas aleatoriamente em três subconjuntos balanceados por classe, utilizados para treino, validação e teste do modelo, na proporção de 80%, 10% e 10%, respectivamente. Além disso, foi implementada uma validação cruzada de 5 *folds* para obter resultados mais robustos. O erro quadrático médio foi utilizado como métrica para avaliar a qualidade da reconstrução das imagens, enquanto a acurácia e o F1-score foram empregados para medir o desempenho do módulo de predição.

A capacidade de cada variável latente z gerada pelo modelo proposto em discriminar entre classes de texturas foi explorada por meio do cálculo do tamanho de efeito (*effect size*) absoluto médio entre cada par de classes. Embora o módulo de predição do modelo proposto seja crucial para a regularização, sua capacidade de atuar como um classificador é limitada, pois a função de perda deve simultaneamente favorecer a reconstrução da imagem inspecionada. Assim, para avaliar a capacidade das variáveis latentes geradas na discriminação entre classes de textura, foram utilizados cinco classificadores adicionais [Shehab 2022]: Naive Bayes (NB), regressão logística multinomial (MLR), K-nearest neighbors (KNN), máquinas de vetor de suporte com funções linear (LSVM) e de base radial (RSVM), além do XGBoost (XGB). Adicionalmente, foram investigadas três soluções de *ensemble* baseadas em votação: votação majoritária (HVE), votação ponderada por probabilidades (SVE) e máxima estimativa *a posteriori* (MAP).

3. Resultados experimentais

Para entender a contribuição de cada componente no desempenho do modelo, oito modelos foram treinados: quatro VAEs clássicos e quatro versões do modelo proposto, onde cada par de modelos teve uma representação de entrada diferente, conforme descrito na seção 2 e exemplificado na Fig. 1. A Tabela 1 mostra os resultados obtidos para os conjuntos de validação, após o treinamento usando cada modelo e representação de entrada. Na tarefa de reconstrução, a coluna MSE avalia a diferença entre as imagens reconstruídas e originais. Para a tarefa de predição, as colunas de acurácia e F1-Score mostram o desempenho do modelo treinado com a representação latente obtida. Embora o erro

de reconstrução do modelo proposto tenha sido maior que o do VAE, o método proposto obteve maior acurácia e F1-score em todos os cenários.

Tabela 1. Resultados experimentais

Dataset	Model	MSE	A	F1
Simple	VAE	1,22	57,0	0,57
	Proposto	1,30	65,2	0,65
Sobel	VAE	1,83	57,8	0,58
	Proposto	2,07	**65,8**	**0,66**
Laplacian	VAE	**0,93**	58,5	0,58
	Proposto	0,95	65,0	0,65
Added Lapl.	VAE	4,84	56,4	0,55
	Proposto	4,96	63,8	0,64

MSE: Erro de reconstrução; A: Acurácia; F1: F1-score.

Tabela 2. Matriz de confusão

R\P	I	II	III	IV
I	102	21	1	2
II	15	94	19	13
III	2	10	72	33
IV	1	10	39	69

Os valores preditos (P) são mostrados nas colunas para cada classe Real (R) nas linhas.

A superioridade das variáveis latentes geradas pelo método proposto quando comparado ao VAE padrão foi investigada, utilizando-se o novo espaço de variáveis como entrada para diferentes classificadores. A Tabela 3 mostra os resultados de um estudo comparativo que usou as variáveis latentes obtidas por várias combinações de modelos e dados de entrada, juntamente com diversos classificadores rasos e *ensembles*. A acurácia do MAP na classificação do conjunto de teste, neste caso, alcançou 70,2%. Além disso, a matriz de confusão relacionada a essa combinação de classificadores mostra que os erros de classificação têm maior probabilidade de ocorrer nas classes vizinhas (Tabela 2). Este resultado é 10 pontos percentuais superior ao obtido pelo VAE padrão.

Tabela 3. Resultados de classificação calculados para o VAE padrão (V) e o método proposto (P) com base nas entradas Simple, Sobel, Laplacian e Added Laplacian.

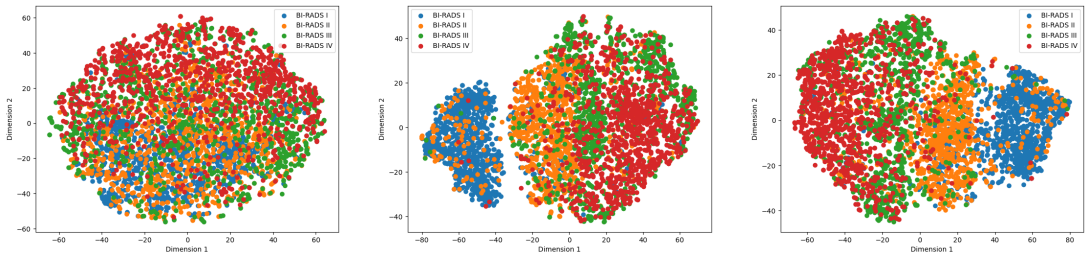
	Simple		Sobel		Laplacian		Added	
	V	P	V	P	V	P	V	P
NB	58,6	65,4	56,4	67,4	59,4	63,4	50,5	63,0
MLR	58,1	64,4	57,1	66,6	56,9	62,9	54,4	62,0
KNN	54,7	65,4	54,7	65,0	57,3	63,8	43,5	60,6
RSVM	58,9	66,2	58,4	68,8	60,2	63,4	53,5	62,4
LSVM	56,9	65,8	56,1	67,6	57,5	62,0	53,5	63,0
XGB	60,2	64,0	58,1	69,2	59,2	64,2	51,7	63,6
HVE	59,8	65,8	58,8	69,0	60,4	65,0	55,5	64,0
SVE	60,2	65,6	57,4	68,8	62,8	63,6	52,7	63,4
MAP	60,2	66,0	61,8	**70,2**	60,0	64,2	54,7	65,0

NB:*Naive Bayes*; MLR:*Multinomial logistic regression*; KNN:*K-nearest neighbors*; LSVM:*Support vector machine with linear kernel*; RSVM:*Support vector machines with radial basis kernel*; XGB:XGBoost; HVE:*Hard voting ensemble*; SVE:*Soft voting ensemble*; MAP:*Maximum a posteriori estimator*.

Uma análise quantitativa da tabela também confirma que o método proposto é consistentemente superior ao VAE padrão. Isso mostra que a concatenação de mapas de bordas, como Sobel ou Laplaciano, à entrada, aliada à função de perda tripla do modelo, é capaz de atenuar o comportamento de filtro passa-baixa da suavização gaussiana. De fato, a tendência dos modelos de DL em filtrar informações de alta frequência já foi discutida por [Rahaman 2019], que nomearam o fenômeno como *o viés espectral*. Isso reforça o fato de que o arranjo e a qualidade das bordas são características importantes para prever

padrões de textura como os mostrados em mamografias.

Uma forma simples de investigar se uma variável é capaz de discriminar entre classes é calcular o tamanho de efeito correspondente. Em cenários multiclasses, uma média dos tamanhos de efeito entre pares pode revelar o quão separadas estão as médias das classes no espaço latente. A Tabela 4 mostra as médias e desvios-padrões das variáveis latentes mais discriminantes de cada modelo, para cada um dos quatro grupos BI-RADS. Apenas as variáveis que apresentam tamanhos de efeito médios superiores a 1 são exibidas. Pode-se observar que o número de variáveis discriminantes aumenta à medida que passamos do VAE padrão para o método proposto. A separabilidade das classes também pode ser avaliada por meio de ferramentas de visualização. A Fig. 3 mostra uma representação 2D do espaço latente de cada um dos modelos, obtida a partir da redução de dimensionalidade usando o t-SNE [van der Maaten and Hinton 2008]. Pode-se observar que o modelo proposto gera um mapa onde as classes estão mais separadas. Isso corrobora a hipótese de que o módulo de predição proposto está realizando uma regularização nas variáveis latentes para construir um espaço onde as classes são melhor distinguidas.



(a) VAE (b) Proposto (c) Proposto+Sobel

Figura 3. Representações latentes dos modelos

Tabela 4. Média (desvio-padrão) das variáveis latentes (LV) mais discriminantes para cada classe BI-RADS (BR), e tamanho de efeito médio (ES).

Met	LV	BR-I	BR-II	BR-III	BR-IV	ES
V	56	0,01(1,07)	0,37(1,11)	1,28(1,2)	2,04(1,22)	1,01
VS	8	1,36(1,16)	1,06(1,14)	0,25(1,21)	-0,69(1,15)	1,00
	50	-0,64(0,85)	0,04(1,05)	0,69(1,07)	1,21(0,98)	1,06
P	23	-0,97(0,71)	-0,65(0,84)	0,36(1,08)	0,72(1,06)	1,09
	30	-0,98(0,74)	-0,55(1,02)	0,89(1,23)	0,85(1,1)	1,15
	44	-0,6(1,03)	-0,31(1,16)	0,63(1,38)	1,83(1,55)	1,03
PS	34	1,28(0,72)	0,25(0,81)	-0,42(0,8)	-0,36(0,88)	1,21
	43	-1,19(0,78)	-0,52(0,82)	0,29(0,87)	0,03(0,81)	1,02
	48	0,87(1,14)	0,82(1,12)	-0,51(1,21)	-1,08(1,12)	1,05
	55	1,42(0,94)	0,28(0,84)	-0,46(1,00)	0,13(0,8)	1,05

V: VAE; VS: VAE+Sobel. P: Proposto; PS: Proposto+Sobel;

Finalmente, vale ressaltar que a acurácia do modelo proposto (70,2%) é competitiva com o estado da arte para a análise do conjunto de dados IRMA, que variou de 67% a 69% em estudos anteriores, principalmente considerando que os experimentos de [Rungue 2019] não utilizaram conjuntos de validação e teste separados para ajuste do modelo. No entanto, a possibilidade de se obterem resultados que são mais facilmente

interpretados pode ser a contribuição mais importante do modelo proposto neste trabalho, pois a interpretabilidade se torna um fator crucial para a aceitação de abordagens de DL em aplicações na saúde.

4. Conclusão

Este estudo apresentou um autoencoder variacional com um módulo de previsão adaptado para a classificação de texturas. Os resultados obtidos revelaram o potencial do método proposto para representar recortes de textura em um espaço de dimensões reduzidas, onde as variáveis latentes são simultaneamente discriminantes e interpretáveis. Aplicado ao problema de análise da densidade mamográfica, o método foi capaz de produzir um espaço latente compacto, onde as amostras de textura foram naturalmente agrupadas, alcançando uma acurácia de classificação competitiva com outros estudos na literatura. A continuação deste trabalho deve considerar a geração de novas imagens plausíveis à medida que amostras sejam extraídas de suas distribuições e mapeadas de volta para o domínio da imagem. Outra tarefa a ser investigada é a validação do modelo em um número maior de bases de dados para se determinar se os resultados obtidos com mamografias são consistentes e generalizáveis para outros domínios e classes de problemas.

Agradecimentos — Este artigo é resultado de projetos parcialmente financiados pelo Fundo de Incentivo à Pesquisa FIP-PUCMinas 2025/32467, e pela FAPEMIG APQ-02753-24 e APQ-06556-24.

Referências

Ha, R. et al. (2018) Convolutional neural network based breast cancer risk stratification using a mammographic dataset. *Academic Radiology*.

Kingma, D. P. and Welling, M. (2022). Auto-encoding variational bayes. *arXiv*.

Lehman, C. et al. (2019) Mammographic breast density assessment using deep learning: Clinical implementation. *Radiology*.

Liu, W. et al. (2020) Towards visually explaining variational autoencoders. In *IEEE Computer Vision and Pattern Recognition*.

Oliveira, J. . (2008). Toward a standard reference database for computer-aided mammography. In *Medical Imaging: Computer-Aided Diagnosis*.

Rahaman, N. (2019). On the spectral bias of neural networks. In *International Conference on Machine Learning*.

Rungue, A. H. A. (2019). Analysis of SVM parametrization in the classification of mammographic texture images. In *Brazilian Congress on Automation*.

Shehab, M. (2022). Machine learning in medical applications: A review of state-of-the-art methods. *Computers in Biology and Medicine*.

Song, Y. (2023). Latent traversals in generative models as potential flows. *arXiv*.

van der Maaten, L. and Hinton, G. (2008). Visualizing high-dimensional data using t-sne. *Journal of Machine Learning Research*.

Zeiser, F. et al. (2021) Deepbatch: A hybrid deep learning model for interpretable diagnosis of breast cancer in whole-slide images. *Expert Systems with Applications*.