

Avaliação de Dor em Expressão Facial Neonatal por meio de Redes Neurais Profundas

Lucas F. Buzuti¹, Carlos E. Thomaz¹

¹Departamento de Engenharia Elétrica – Laboratório de Processamento de Imagens
Centro Universitário FEI – São Bernardo do Campo – SP – Brasil

lucas.buzuti@outlook.com, cet@fei.edu.br

Abstract. Neonatal pain assessment might suffer variation among health professionals, leading to late intervention and flimsy treatment of pain in several occasions. In this context, the goal of this dissertation was to assess quantitatively and qualitatively models of Convolutional Neural Networks in the neonatal pain classification task using face images of two distinct databases (an international, named COPE, and other national, named UNIFESP). Our quantitative results showed the top performance of N-CNN on neonatal pain classification, with average accuracy of 87.2% and 78.7% for the databases COPE and UNIFESP, respectively. However, the quantitative analysis indicated that all neural models evaluated, including the N-CNN model, might learn artifacts from the images and not proper discriminating information, suggesting further research to apply such models in clinical practice.

Resumo. A avaliação da dor neonatal pode sofrer variações entre profissionais de saúde, resultando em intervenção tardia e tratamento inconsistente da dor. Neste contexto, o objetivo desta dissertação foi analisar, quantitativa e qualitativamente, modelos de Redes Neurais Convolucionais na tarefa de classificação automática da dor neonatal por meio de imagens de faces de dois bancos de dados distintos (um internacional, denominado COPE, e outro nacional, denominado UNIFESP). Os resultados quantitativos mostraram a superioridade da arquitetura N-CNN para avaliação automática da dor neonatal, com acurácias médias de 87.2% e 78.7% para os bancos de imagens COPE e UNIFESP, respectivamente. No entanto, a análise qualitativa evidenciou que todos os modelos neurais avaliados, incluindo a arquitetura Neonatal Convolutional Neural Network (N-CNN), podem aprender artefatos da imagem e não variações discriminantes das faces, mostrando a necessidade de mais estudos para aplicação de tais modelos na prática clínica em questão.

1. Introdução

A capacidade de comunicação verbal da dor não se aplica para o neonato. Por várias décadas, os pediatras acreditavam que os neonatos não sentiam ou não se lembravam da dor, uma vez que suas capacidades eram limitadas devido à ausência de substrato neurológico para percepção da mesma. Tal crença foi refutada por diversos estudos científicos principalmente a partir do final da década de 1980 [Anand and Carr 1989].

Estudos relatam que experiências dolorosas inesperadas e repetidas vividas pelos neonatos estão associadas a distúrbios que podem prejudicar a curto e longo

prazos suas vidas, sendo esses: alterações na sensibilidade e percepção da dor [Vinall et al. 2012][Brummelte et al. 2012], funcionamento do sistema de resposta ao estresse (altos níveis de cortisol) [Grunau et al. 2010][Grunau et al. 2004][Walker 2017] e crescimento pós-natal (menor ganho de peso corporal) [Vinall et al. 2012], entre outros. Fortes evidências em relação à exposição extensa à dor durante o período inicial da vida estão associadas a alterações estruturais e funcionais do cérebro. As alterações que ocorrem são: alterações na substância branca cerebral e na substância cinzenta subcortical [Brummelte et al. 2012][Vinall et al. 2012], atraso no desenvolvimento corticoespinal [Vinall et al. 2012], alterações no número de conexões sinápticas e alteração do grau de ramificação capilar que aumenta o suprimento de sangue e oxigênio [Bhutta and Anand 2002]. Tais alterações podem resultar em uma variedade de deficiências comportamentais, de desenvolvimento e de aprendizagem [Grunau et al. 2010].

O propósito deste trabalho foi analisar quantitativa e qualitativamente modelos de Redes Neurais Convolucionais na tarefa de classificação automática da dor neonatal baseados em imagens de faces de dois bancos de dados distintos (um internacional, denominado COPE [Brahnam et al. 2006], e outro nacional, denominado UNIFESP [Heiderich et al. 2015]). A contribuição foi implementar, avaliar e comparar modelos existentes de redes neurais (Neonatal Convolutional Neural Network (N-CNN) [Zamzmi et al. 2019] e ResNet50 [He et al. 2016]) desenvolver um arcabouço computacional de avaliação da dor neonatal menos subjetiva e que não sofra variações entre profissionais de saúde.

2. Metodologia

Esta seção descreve a metodologia proposta e implementada neste trabalho, composta das seguintes etapas: detecção facial, aumento de dados, reconhecimento da dor usando topologias de Deep Learning (DL), tais como Neonatal Convolutional Neural Network (N-CNN) e ResNet50, e mapeamento de ativação de classe ponderada por gradiente como métrica qualitativa. A Figura 1 descreve o arcabouço proposto nas etapas de treinamento e avaliação.

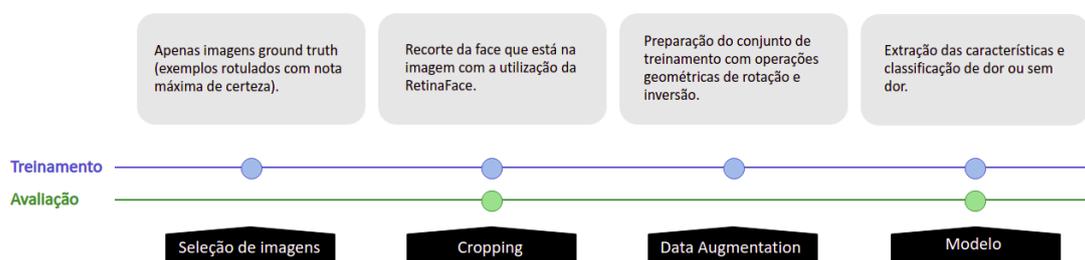


Figura 1. Etapas de treinamento e avaliação (arcabouço computacional).

2.1. Seleção de imagens

Foram escolhidas apenas os *ground truths* (imagens rotuladas como de dor e não-dor) para o conjunto de dados COPE, foram utilizadas 153 imagens coloridas com de 26 recém-nascidos caucasianos. Para a UNIFESP foram utilizadas 356 imagens coloridas de 30 recém-nascidos.

2.2. Cropping

Este trabalho fez uso do detector facial RetinaFace [Deng et al. 2019], já treinado. Em cada imagem do conjunto de dados o algoritmo retornou as coordenadas de cada face e 5 pontos faciais (landmarks). Para cada imagem foram utilizadas as coordenadas correspondentes detectadas para registrar e cortar a região exata da face do bebê. Em seguida, cada imagem sofreu um redimensionamento específico conforme a dimensão de entrada das topologias. Para os experimentos com N-CNN cada imagem foi redimensionada para 120×120 e com ResNet50 o redimensionamento foi de 224×224 . O método utilizado para redimensionar as imagens foi o bi-cubico [Zamzmi et al. 2019].

2.3. Data Augmentation

Para a construção dos conjuntos de treinamento, validação e teste, após o conjunto original sofrer o corte da região exata da face do bebê, foi dividido randomicamente em 50% gerando o conjunto I_{test} e I^* . No conjunto de dados I^* aplicou-se operações geométricas de rotação de inversão. Esse procedimento gerou um total de $36|I^*|$ e o novo número de amostras do conjunto ficou $|I^*| = 36|I^*| + |I^*|$. O conjunto de treinamento e validação foram obtidos a partir da divisão randomicamente do I^* , ficando 70% o conjunto de treinamento I_{train} e 30% o conjunto de validação I_{val} .

2.4. Modelo

Os modelos abordados neste trabalho, N-CNN [Zamzmi et al. 2019] e ResNet50, são modelos de Deep Learning utilizados em estudos recentes na classificação da dor neonatal. A ResNet50 foi o primeiro modelo a introduzir o aprendizado residual em todas as camadas empilhadas. Tal aprendizado é constituído por um bloco matemático, que foi definido sendo: $y = \mathcal{F}(x, W_i) + x$, representando o mapeamento residual a ser aprendido pela próxima camada. Uma topologia de redes neurais convolucional, N-CNN, foi proposta para extração de característica e classificação da dor neonatal, sendo o primeiro modelo de Deep Learning para essa tarefa. A N-CNN têm 4 camadas convolucionais seguidas de max pooling e 2 camadas totalmente conectadas para fazer a classificação.

2.5. Métricas de Avaliação

As métricas de avaliações topológicas abordadas neste trabalho foram as comumente conhecidas: função de erro, acurácia e matriz de confusão. Todas essas métricas foram avaliadas com k -fold igual a 3. Além disso, abordou-se também a técnica de visualização denominada Grad-CAM [Selvaraju et al. 2017].

3. Resultados

A Tabela 1 mostra a acurácia de treinamento, validação e teste de cada modelo treinado nos bancos de imagens UNIFESP e COPE. Verifica-se que em ambos os bancos de imagens o modelo N-CNN teve o melhor desempenho estatisticamente, 87.2% (COPE) e 78.7% (UNIFESP).

Com a utilização do Grad-CAM foi possível visualizar nas imagens atribuídas ao modelo os locais discriminantes dos estados dor e não-dor, como mostram os resultados da Figura 2.

Modelo	Treinamento		Validação		Teste	
	COPE	UNIFESP	COPE	UNIFESP	COPE	UNIFESP
N-CNN [Zamzmi et al. 2019]	100% ± 0.0	100% ± 0.0	99.6% ± 0.2%	99.5% ± 0.3%	87.2% ± 3.6%	78.7% ± 1.7%
ResNet50 [Zamzmi et al. 2019]	100% ± 0.0%	100% ± 0.0%	99.3% ± 0.2%	99.4% ± 0.2%	82.9% ± 4.7%	76.0% ± 0.7%
ResNet50 (proposta)	100% ± 0.0%	100% ± 0.0%	99.5% ± 0.2%	99.2% ± 0.0%	83.8% ± 4.1%	74.7% ± 1.7%

Tabela 1. Resultado estatístico da acurácia média dos modelos treinados, validados e testados com os bancos COPE e UNIFESP.

Computando o Grad-CAM nas imagens neonatais para o modelo melhor avaliado estatisticamente, N-CNN, em ambos os bancos de imagens, Figura 2, observa-se que o modelo aprendeu locais discriminantes do estado dor e não-dor. Os locais mais discriminantes foram as regiões da boca, sulco nasolabial e sobrancelhas, para o banco de imagens UNIFESP. Regiões essas comumente analisadas pelos profissionais de saúde para identificar a dor nos recém-nascidos. Para o banco de imagens COPE, o modelo também extraiu regiões discriminantes. No entanto, as regiões discriminantes de maior ponderação estão localizadas mais ao redor da face do neonato, como artefatos.

Banco	Rótulo	Predição	Imagem Original	Grad-CAM	Grad-CAM + Imagem	Segmentação
UNIFESP	Dor	Dor				
UNIFESP	Não-dor	Não-dor				
COPE	Dor	Dor				
COPE	Não-dor	Não-dor				

Figura 2. Resultados do Grad-CAM do modelo N-CNN nos bancos UNIFESP e COPE.

4. Conclusão

Entre os três modelos abordados neste trabalho, o modelo N-CNN [Zamzmi et al. 2019] foi o melhor em ambos os bancos de imagens e mostrou bom desempenho na análise quantitativa e qualitativa. Interessante observar que esse modelo extraiu regiões discriminantes, de dor ou não-dor, que são avaliadas pelos profissionais de saúde. Entende-se

que o modelo N-CNN é a referência atual do estado-da-arte para classificação automática da dor neonatal, obtendo como acurácias médias 87.2% e 78.7% para as bases COPE e UNIFESP, respectivamente. No entanto, a análise qualitativa evidenciou que o modelo N-CNN pode aprender artefatos da imagem e não variações discriminantes, mostrando a necessidade de mais estudos para aplicação prática na área clínica em questão.

Algumas limitações que impactaram os desempenhos dos modelos neurais avaliados foram observadas neste trabalho. A limitação mais relevante, se refere ao número disponível de imagens dos bancos, COPE e UNIFESP. Apesar do aumento de dados proposto e implementado, que adicionou mais imagens aos bancos originais, esse aumento não permitiu a criação de novas imagens com novos atributos semânticos, pois se deu apenas por transformações geométricas afins das imagens já existentes.

Vislumbra-se, como trabalhos futuros: estender essas análises baseadas em modelos de Aprendizado Profundo para outros modelos; modelar uma nova arquitetura para a classificação da dor neonatal e compreender como as extrações de regiões discriminantes são avaliadas por esses modelos utilizando outros métodos de visualização; e gerar imagens sintéticas a partir de bancos plurirraciais como a base da UNIFESP.

Agradecimento

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001, da Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) processo nº 2018/13076-9 e do Centro Universitário FEI.

Referências

- Anand, K. J. and Carr, D. B. (1989). The neuroanatomy, neurophysiology, and neurochemistry of pain, stress, and analgesia in newborns and children. *Pediatric Clinics of North America*, 36(4):795–822.
- Bhutta, A. T. and Anand, K. (2002). Vulnerability of the developing brain: neuronal mechanisms. *Clinics in perinatology*, 29(3):357–372.
- Brahnam, S., Chuang, C.-F., Shih, F. Y., and Slack, M. R. (2006). Machine recognition and representation of neonatal facial displays of acute pain. *Artificial intelligence in medicine*, 36(3):211–222.
- Brummelte, S., Grunau, R. E., Chau, V., Poskitt, K. J., Brant, R., Vinall, J., Gover, A., Synnes, A. R., and Miller, S. P. (2012). Procedural pain and brain development in premature newborns. *Annals of neurology*, 71(3):385–396.
- Deng, J., Guo, J., Zhou, Y., Yu, J., Kotsia, I., and Zafeiriou, S. (2019). Retinaface: Single-stage dense face localisation in the wild. *arXiv preprint arXiv:1905.00641*.
- Grunau, R. E., Tu, M. T., Whitfield, M. F., Oberlander, T. F., Weinberg, J., Yu, W., Thiesen, P., Gosse, G., and Scheifele, D. (2010). Cortisol, behavior, and heart rate reactivity to immunization pain at 4 months corrected age in infants born very preterm. *The Clinical journal of pain*, 26(8):698.
- Grunau, R. E., Weinberg, J., and Whitfield, M. F. (2004). Neonatal procedural pain and preterm infant cortisol response to novelty at 8 months. *Pediatrics*, 114(1):e77–e84.

- He, K., Zhang, X., Ren, S., and Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778.
- Heiderich, T. M., Leslie, A. T. F. S., and Guinsburg, R. (2015). Neonatal procedural pain can be assessed by computer software that has good sensitivity and specificity to detect facial movements. *Acta Paediatrica*, 104(2):e63–e69.
- Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., and Batra, D. (2017). Grad-cam: Visual explanations from deep networks via gradient-based localization. In *Proceedings of the IEEE international conference on computer vision*, pages 618–626.
- Vinall, J., Miller, S. P., Chau, V., Brummelte, S., Synnes, A. R., and Grunau, R. E. (2012). Neonatal pain in relation to postnatal growth in infants born very preterm. *Pain*, 153(7):1374–1381.
- Walker, S. M. (2017). Translational studies identify long-term impact of prior neonatal pain experience. *Pain*, 158:S29–S42.
- Zamzmi, G., Paul, R., Goldgof, D., Kasturi, R., and Sun, Y. (2019). Pain assessment from facial expression: Neonatal convolutional neural network (n-cnn). In *2019 International Joint Conference on Neural Networks (IJCNN)*, pages 1–7. IEEE.