

# Neonatal Pain Assessment From Facial Expression Using Deep Neural Networks

Lucas Fontes Buzuti\*, Tatiany M. Heideirich†, Marina C. M. Barros†, Ruth Guinsburg†, Carlos Eduardo Thomaz\*

\*Departamento de Engenharia Elétrica, Centro Universitário FEI, São Bernardo do Campo-SP, Brasil

†Departamento de Pediatria, Universidade Federal de São Paulo, São Paulo-SP, Brasil

**Abstract**—Currently, neonatal pain assessment varies among health professionals, leading to late intervention and flimsy treatment of pain in several occasions. Therefore, it is essential to understand the deficiencies of the current pattern of pain assessment tools in order to develop new ones, less subjective and susceptible to external variable influences. The aim of this paper is to investigate neonatal pain assessment using two models of Deep Learning: Neonatal Convolutional Neural Network trained end-to-end and ResNet trained using Transfer Learning. We used for training two distinct databases (COPE and Unifesp) and our results showed that the use of multi-racial databases might improve the evaluation of automatic models of neonatal pain assessment.

**Index Terms**—Automated pain recognition, deep learning, facial expression, neonatal pain assessment, pattern recognition

## I. INTRODUÇÃO

A definição de dor, pela Associação Internacional para o Estudo da Dor (IASP), é “uma experiência sensorial e emocional desagradável associada a danos teciduais reais, potenciais ou descrita em termos de tais danos”. A capacidade de comunicação verbal da dor, ou simplesmente o ato de apontar (escala visual analógica) 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 [1]. Tal crença foi refutada por diversos estudos científicos [2] [3].

Estudos relatam que experiências dolorosas repetidas vividas pelos neonatos estão associadas a alterações que podem prejudicar a curto e longo prazos suas vidas, sendo esses: alterações na sensibilidade e percepção da dor [4] [5] [6] [7], funcionamento do sistema de resposta ao estresse (altos níveis de cortisol) [8] [9] [10] [5], 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 [7] [11] [5], atraso no desenvolvimento corticoespinal [6] [5], alterações no número de conexões sinápticas e neuróglia (são células não neuronais do sistema nervoso central que proporcionam suporte e nutrição aos neurônios) e na alteração do grau de ramificação capilar que aumenta o suprimento de sangue e oxigênio [12] [13]. Tais alterações podem resultar em uma variedade de alterações comportamentais, de desenvolvimento e de aprendizagem [8] [14] [15].

Anand [16] relatou que os neonatos sentiam dor e, em geral, esta não era reconhecida e, por tanto, subtratada, por isso, recomendou o uso de analgésicos, que deveriam ser prescritos de acordo com os cuidados que cada neonato necessitasse. Trabalhos relataram que o uso excessivo de medicamentos analgésicos, tais como morfina e fentanil, poderiam causar efeitos colaterais. Zwicker et al. [17] relataram estes efeitos após observar o aumento de 10 vezes do uso de morfina (um agente comumente usado para o tratamento da dor neonatal), visto que está associado ao déficit do crescimento cerebelar no período neonatal, com um quadro de resultados piores para o desenvolvimento neurológico no período da primeira infância. Diversas revisões descrevem o fentanil como um analgésico extremamente potente e listam diversos efeitos colaterais tais como, neuroexcitação e depressão respiratória, para o uso de altas doses [15] [18].

Segundo Hummel et al. [19] e Simons et al. [20], em média, quatorze procedimentos dolorosos por dia são realizados em bebês na Unidade de Terapia Intensiva Neonatal (UTIN). Métodos de avaliação da dor comumente utilizados na pediatria, como a autoavaliação e o uso de escala visual analógica, com símbolos ou números para indicar diferentes níveis de dor, são considerados o padrão-ouro. Entretanto, estes métodos não são aplicáveis na neonatologia, visto que requerem uma capacidade de comunicação complexa, ainda não presente nos recém-nascidos. Os métodos atuais para avaliar a dor nessa população vulnerável dependem da atuação de profissionais bem qualificados, que observam as múltiplas repostas comportamentais e fisiológicas aliadas. No entanto, há uma dificuldade na utilização dessas escalas, relatada por Heideirich [21], pois, uma vez que os pacientes neonatais são pré-verbais e se encontram em diferentes fases do desenvolvimento cognitivo, ainda existem muitas dúvidas quanto à interpretação e à avaliação das respostas à dor neste paciente.

Poucos estudos como [22] e [21] foram realizados para analisar e avaliar a dor neonatal usando tecnologias de Visão Computacional e Aprendizado de Máquina. Por outro lado, uma variedade rica de métodos foi proposto para avaliar dor de adultos [23] [24] [25] [26] [27]. Os motivos destacados por Zamzmi [1] sobre a falta de estudos para o reconhecimento da dor neonatal, principalmente usando Aprendizado Profundo, se referem aos números limitados de bancos de imagens neonatais e à crença de que os algoritmos projetados para avaliar dor em adultos teriam desempenhos semelhantes para

os neonatos, o que não acontece na prática. Apenas dois trabalhos que utilizam a tecnologias de Aprendizado Profundo são [28] e [29], que aplicaram o Aprendizado por Transferência em algumas arquiteturas de rede neurais, tal como a ResNet [30], para classificar a dor neonatal.

Este artigo tem por objetivo investigar dois modelos de Aprendizado Profundo (DL), Neonatal Convolutional Neural Network (N-CNN) [29] e ResNet50 [29], em bases de imagens distintas, para avaliação facial automática da dor neonatal.

## II. METODOLOGIA

### A. RetinaFace

A localização automática da face é uma etapa pré-requisitada na análise de imagens faciais para muitas aplicações, como atributo facial [31] e reconhecimento de identidade facial [32]. Uma definição estreita de localização de face pode se referir à detecção de face tradicional [33], que visa estimar as caixas delimitadoras de face sem nenhuma escala e posição anterior. Este artigo fez uso do algoritmo RetinaFace proposto por Deng et al. [32], visto que utiliza de uma definição mais ampla de localização de face, assim encontrando faces em qualquer posição, incluindo: detecção de face, alinhamento de face, análise de face em pixel e regressão de correspondência densa em 3D. Esse tipo de localização densa da face fornece informações precisas da posição facial para todas as escalas diferentes.

No treinamento da RetinaFace [32] os autores utilizaram o erro de multi-tarefas (multi-task loss), no intuito de minimizar o erro da caixa de âncora  $i$ . Tal erro é definido sendo:

$$L = L_{cls}(p_i, p_i^*) + \lambda_1 p_i^* L_{box}(t_i, t_i^*) + \lambda_2 p_i^* L_{pts}(l_i, l_i^*) + \lambda_3 p_i^* L_{pixel}. \quad (1)$$

$L_{cls}(p_i, p_i^*)$  erro da classificação, em que  $p_i$  é a probabilidade prevista de  $i$  ser uma face,  $p_i^*$  será 1 para  $i$  positivo e 0 para  $i$  negativo, e  $L_{cls}$  é a função softmax para classes binárias (face/não face).  $L_{box}(t_i, t_i^*)$  erro de regressão da caixa de face, onde  $t_i = \{t_x, t_y, t_w, t_h\}_i$  e  $t_i^* = \{t_x^*, t_y^*, t_w^*, t_h^*\}_i$  representando as coordenadas da caixa prevista e a caixa do ground-truth associado ao  $i$  positivo.  $L_{pts}(l_i, l_i^*)$  erro de regressão dos pontos faciais, uma vez que  $l_i = \{l_{x1}, l_{y1}, \dots, l_{x5}, l_{y5}\}_i$  e  $l_i^* = \{l_{x1}^*, l_{y1}^*, \dots, l_{x5}^*, l_{y5}^*\}_i$  representam os cinco marcos faciais previstos e os cinco marcos faciais do ground-truth associado ao  $i$  positivo. O erro da regressão densa  $L_{pixel}$  é definida pela Equação (2). Os parâmetros de balanceamento dos erros  $\lambda_1, \lambda_2$  e  $\lambda_3$  são definidos em 0,25, 0,1 e 0,01, o que significa que há um aumento na importância de melhores locais de caixa e ponto de referência a partir dos sinais de supervisão [32].

$$L_{pixel} = \frac{1}{W * H} \sum_i^W \sum_j^H \|\mathcal{R}(\mathcal{D}_{P_{ST}}, P_{cam}, P_{ill})_{i,j} - I_{i,j}^*\|_1, \quad (2)$$

após a convolução gráfica, vide [32], os autores comparearam os parâmetros de forma e textura  $P_{ST} \in R^{128}$

para projetar uma malha colorida  $\mathcal{D}_{P_{ST}}$  em um plano da imagem 2D com parâmetros de câmera  $P_{cam} = [x_c, y_c, z_c, x'_c, y'_c, z'_c, f_c]$  (ou seja, localização da câmera, posição da câmera e distância focal) e parâmetros de iluminação  $P_{ill} = [x_l, y_l, z_l, r_l, g_l, b_l, r_a, g_a, b_a]$  (ou seja, localização da fonte de luz pontual, valores de cores e cores da iluminação ambiente). Com a face 2D renderizada  $\mathcal{R}(\mathcal{D}_{P_{ST}}, P_{cam}, P_{ill})$ , foi comparado a diferença de pixel da face 2D renderizada e a face original 2D, como mostra a Equação 2,  $W$  e  $H$  são a largura e altura de  $i$  do corte da face  $I_{i,j}^*$ , respectivamente.

### B. ResNet

A ResNet proposta por He et al. [30] revolucionou a corrida arquitetônica da CNN ao introduzir o conceito de aprendizado residual e o desenvolvendo de uma metodologia eficiente para o treinamento de redes profundas. Semelhante as Highway Networks [34] a ResNet também é colocada na categoria das CNNs com vários caminhos.

He et al. [30] adotaram na ResNet o aprendizado residual em todas as camadas empilhadas. A formulação matemática do bloco foi definida sendo:

$$y = \mathcal{F}(x, W_i) + x, \quad (3)$$

Tem-se  $x$  e  $y$  como os vetores de entrada e saída das camadas consideradas. A função  $\mathcal{F}(x, W_i)$  representa o mapeamento residual a ser aprendido. Supondo um bloco de construção residual com duas camadas, portanto, se tem:  $\mathcal{F} = W_2\sigma(W_1x)$ , na qual  $\sigma$  denota ReLU e os biases foram omitidos para simplificar as notações. A operação  $\mathcal{F} + x$  é realizada por um salto de conexão e uma adição elemento a elemento, tendo assim então  $y$ . He et al. também consideraram uma segunda não linearidade após a adição, ou seja,  $\sigma(y)$ .

### C. Neonatal Convolutional Neural Network

Devido aos números baixos de pesquisas relacionada à classificação da dor neonatal e principalmente pesquisas utilizando Aprendizado Profundo, Zamzmi et al. [29] propuseram uma topologia de CNN, denominada Neonatal Convolutional Neural Network (N-CNN), para extração de característica e classificação da dor neonatal. Segundo Zamzmi et al. [29], tal topologia foi a primeira modelada utilizando Aprendizado Profundo para classificar dor neonatal. Este artigo fez uso desse modelo conforme está descrito em [29], com uma vetorização entre a 9° camada e a 10° camada (tal vetorização não está descrita em [29]). Os parâmetros da N-CNN são sendo apresentados na Tabela I.

## III. MATERIAIS

Para a construção do arcabouço computacional para avaliar a dor neonatal nos modelos de Aprendizado Profundo, foi utilizado a linguagem de programação Python3 e o open source AI framework para aprendizado de máquina e computação numérica de alta performance da Google, denominado TensorFlow.

TABLE I  
PARÂMETROS DA N-CNN [29].

Branch	Layer	Type	Input	Filters	Filter Size	Activation	Regularization
Left	Layer 1	Max Pool 1	$120 \times 120 \times 3$	-	$10 \times 10$ , st. 10, pd. 0	-	-
Central	Layer 2	Conv 1	$120 \times 120 \times 3$	64	$5 \times 5$ , st. 1, pd. 0	Leaky ReLU (0.01)	-
	Layer 3	Max Pool 2	Layer 2	-	$3 \times 3$ , st. 3, pd. 0	-	-
Right	Layer 4	Conv 2	Layer 3	64	$2 \times 2$ , st. 1, pd. 0	Leaky ReLU (0.01)	-
	Layer 5	Max Pool 3	Layer 4	-	$3 \times 3$ , st. 3, pd. 0	-	Dropout (0.1)
	Layer 6	Conv 3	$120 \times 120 \times 3$	64	$5 \times 5$ , st. 1, pd. 0	Leaky ReLU (0.01)	-
	Layer 7	Max Pool 4	Layer 6	-	$10 \times 10$ , st. 10, pd. 0	-	Dropout (0.1)
Merge Layer (Left — Central — Right)							
	Layer 8	Conv 4	Merge Layer	64	$2 \times 2$ , st. 1, pd. 0	ReLU	-
	Layer 9	Max Pool 5	Layer 8	-	$2 \times 2$ , st. 2, pd. 0	-	-
Vectorization(Layer 9)							
	Layer 10	FC1	Layer 9	-	-	ReLU	L2 Regularizer (0.01) Dropout (0.1)
	Layer 11	FC2	Layer 10	-	-	Sigmoid	-

Bancos de imagens para proporcionar estudos utilizando Visão Computacional, Aprendizado de Máquina e principalmente Aprendizado Profundo na análise e avaliação da dor neonatal ainda são poucos. Este artigo fez uso de dois bancos neonatais: COPE [35], um dos primeiros bancos de imagens na finalidade de analisar e avaliar a dor neonatal, e o banco de imagens neonatal da Unifesp [36].

#### A. COPE

Para o desenvolvimento do COPE, Brahn et al. [35] aplicaram estímulos que provocam expressões faciais, com o objetivo de serem utilizados nas avaliações e análises da dor em neonatos. Os estímulos aplicados foram: punção do calcanhar (teste do pezinho), fricção na superfície lateral externa do calcanhar, transporte de um berço para outro e estímulo aéreo. O estímulo aéreo no nariz teve a intenção de provocar aperto nos olhos, simulando mudanças de iluminação [35]. O banco de imagens COPE contém 288 imagens coloridas com  $3008 \times 2000$  pixels de 26 recém-nascidos caucasianos, 13 meninos e 13 meninas.

#### B. Unifesp

Heiderich et al. [36] construíram um banco de imagens neonatais (banco de imagens Unifesp) para a elaboração de um software capaz de identificar automaticamente a expressão de dor do recém-nascido. O *back-end* do software utilizou a distância entre pontos específicos identificados no rosto do recém-nascido e a escala unidimensional NFCS [37], assim classificando a existência ou não da dor. O banco de imagens foi construído a partir de fotos capturadas antes, durante e depois de procedimentos dolorosos aplicados a essa população, como: punção venosa, capilar ou injeção intramuscular (procedimentos comuns e necessários). O banco de imagens da Unifesp contém 360 imagens coloridas com  $450 \times 233$  pixels de 30 recém-nascidos entre 34 e 41 semanas e idade gestacional e entre 24 e 168 horas de vida (prematuros tardios ou a termo).

### IV. IMPLEMENTAÇÃO

Para detectar somente a face do neonato, utilizou-se do detector facial RetinaFace já treinado em cada imagem do banco.

O detector retornou as coordenadas de cada face contida na imagem<sup>1</sup> e 5 pontos faciais (landmarks). O detector também enviou as coordenadas desses pontos e uma mensagem de falha para indicar se a detecção não ocorreu<sup>2</sup>. Para cada imagem foram utilizadas as coordenadas detectadas para registrar e cortar a região exata da face do neonato. Em seguida, cada imagem sofreu um específico redimensionamento conforme a dimensão de entrada das topologias. Para os experimentos com N-CNN cada imagem foi redimensionada para  $120 \times 120$  e com ResNet50 o redimensionamento foi de  $224 \times 224$ . O método utilizado para redimensionar as imagens foi o bicubico.

Em todo o arcabouço foi utilizado o aumento dos dados (Data Augmentation). Portanto, este artigo fez o uso do mesmo aumento de dados utilizado no trabalho [29], no conjunto de treinamento e validação. Para a construção dos conjuntos de treinamento, validação e teste, após o conjunto original sofrer o corte da região exata da face do bebê, por meio do detector facial RetinaFace, foi dividido randomicamente em 50% gerando o conjunto  $I_{test}$  e  $I^*$ . No conjunto de dados  $I^*$  aplicou-se o aumento de dados, sendo:

- Cada imagem foi randomicamente rotacionada até  $30^\circ$  ( $1^\circ$  a  $30^\circ$ ), para gerar um total de 12 imagens para cada imagem
- Cada imagem rotacionada foi invertida horizontalmente e verticalmente, assim gerando um total de 24 imagens para cada imagem rotacionada.

Esse procedimento gerou um total de  $36|I^*|$  e o novo número de amostra do conjunto ficou  $|I^*| = 36|I^*| + |I^*|$ . Os conjuntos de treinamento e validação foram obtidos a partir da divisão randomicamente do  $I^*$ , ficando 80% o conjunto de treinamento  $I_{train}$  e 20% o conjunto de validação  $I_{val}$ .

Os modelos avaliados por este artigo foram três: N-CNN [29], ResNet50 aplicando TL proposta por Zamzmi em [29] e ResNet50 aplicando TL proposta por este artigo. A ResNet50 proposta em [29], está removendo a última camada que define as classes do modelo e adicionando 1 único neurônio com

<sup>1</sup>As imagens no banco só contêm apenas uma face, sendo essa face de um bebê.

<sup>2</sup>Não houve falha nos bancos de imagens.

função sigmoid, em que 1 indica o estado dor e 0 o estado sem dor. O modelo adotado em [29] está sendo mostrada na Tabela II. Este artigo propôs também uma ResNet50, denominada ResNet50(ours), com a aplicação do TL, entretanto, ao invés de utilizar 1 único neurônio utilizou-se 2 neurônios com a função softmax, em que [0 1] define o estado dor e [1 0] o estado sem dor. A Tabela III mostra o modelo proposto.

TABLE II  
ARQUITETURA RESNET50 PROPOSTA EM [29].

Global Average Pooling	Base model output
Dropout	0.5
Full 1	1, sigmoid
Total parameters	23.688.065

TABLE III  
ARQUITETURA RESNET50(OURS) PROPOSTO POR ESTE ARTIGO.

Global Average Pooling	Base model output
Dropout	0.5
Full 1	2, softmax
Total parameters	23.788.418

Os modelos foram treinados usando o tamanho de lote (16) e taxa de aprendizado (0.0001) com o algoritmo de descida de gradiente RMSprop. [38].

## V. EXPERIMENTOS E RESULTADOS

Os bancos de imagens que foram utilizados neste artigo passaram pela etapa de pré-processamento, mas antes de tal etapa, as imagens foram selecionadas para a construção dos conjuntos de dados contendo os estados “Dor” e “Sem Dor”. O banco de imagens da Unifesp já contém os dois estados que este artigo fez uso, mas houve a necessidade de não utilizar 4 imagens, uma vez que estavam sem seus devidos rótulos. Portanto, utilizou-se 356 imagens do banco de imagens da Unifesp. Em relação ao segundo banco utilizado, banco de imagens COPE, visto que é um banco que além dos estados necessários para o experimento há mais estados (choro, estímulo de ar e fricção), contabilizando um total de 288 imagens. Sendo assim, foram selecionadas as imagens com os rótulos *rest* e *pain* sendo-os sem dor e dor, respectivamente. Após tal seleção, contabilizou-se 153 imagens do banco COPE. Após a seleção das imagens, os bancos passaram pela etapa de pré-processamento. Logo, os conjuntos de imagens ficaram:

- Banco de imagens Unifesp:  $I_{train} = 5269$  imagens,  $I_{val} = 1317$  imagens,  $I_{test} = 178$  imagens
- Banco de imagens COPE:  $I_{train} = 2279$  imagens,  $I_{val} = 570$  imagens,  $I_{test} = 76$  imagens

Nas Tabelas IV e V são mostradas as eficiências de cada modelo treinado, validado e testado com o banco da Unifesp e COPE, respectivamente, e também mostra uma segunda acurácia quando os modelos já treinados são testados com outro banco (os modelos que foram treinados com o banco de imagens da Unifesp após o treinamento foram submetidos ao

banco de imagens da COPE, assim computando essa segunda acurácia, e mutuamente).

TABLE IV  
RESULTADO DA AVALIAÇÃO DA DOR NEONATAL EM EXPRESSÃO FACIAL DO MODELO TREINADO, VALIDADO E TESTADO COM O BANCO DE IMAGENS UNIFESP (ACURÁCIA DE TESTE 2° COLUNA) E NA 3° COLUNA UMA SEGUNDA ACURÁCIA, NO QUAL O MODELO APÓS O TREINAMENTO FOI SUBMETIDO AO BANCO DE IMAGENS COPE.

Model (Unifesp)	Accuracy (Unifesp)	Accuracy (COPE)
N-CNN [29]	80.1%	70.3 %
ResNet50 [29]	78.6%	59.4%
ResNet50 (ours)	78.4%	57.8%

TABLE V  
RESULTADO DA AVALIAÇÃO DA DOR NEONATAL EM EXPRESSÃO FACIAL DO MODELO TREINADO, VALIDADO E TESTADO COM O BANCO DE IMAGENS COPE (ACURÁCIA DE TESTE 2° COLUNA) E NA 3° COLUNA UMA SEGUNDA ACURÁCIA, NO QUAL O MODELO APÓS O TREINAMENTO FOI SUBMETIDO AO BANCO DE IMAGENS UNIFESP.

Model (COPE)	Accuracy (COPE)	Accuracy (Unifesp)
N-CNN [29]	81.2%	43.2%
ResNet50 [29]	71.9%	53.4%
ResNet50 (ours)	87.5%	53.4%

Os resultados trazidos das Tabelas IV e V demonstram em uma primeira análise que o modelo ResNet50 proposto por este artigo é superior ao modelo ResNet50 proposto em [29], uma vez que o modelo treinado nos dois bancos de imagens obteve uma acurácia relativamente boa (acima de 70%). Tal afirmação pode ser validada analisando a acurácia no modelo junto com uma segunda acurácia, a qual foi obtida introduzindo outro banco de imagens. Sendo assim na Tabela IV a 3° e 4° linha indicam uma acurácia homogênea tanto do modelo (2° coluna) quanto da segunda acurácia (3° coluna), mas impedindo extrair uma conclusão de qual modelo é superior. Na Tabela V a 3° e 4° linha indicam que o modelo proposto em [29] foi inferior ao modelo proposto por este artigo, porém, quando os modelos foram testados com outro banco de imagens obtiveram uma mesma acurácia (3° coluna). Portanto, a alteração do modelo ResNet50 proposto por este artigo pode ser considerada melhor ao modelo ResNet50 proposto em [29], para o banco de imagens COPE, visto que obteve uma acurácia de 87.5% contra os 71.9%, mas ressaltando que tal afirmação está sendo realizada com ressalva, devido ao número de imagens dos dois bancos utilizados e principalmente às características dos bancos, uma vez que o banco da COPE consiste apenas de neonatos caucasianos e o banco da Unifesp consiste em um conjunto maior de raças, sendo um banco plurirracial.

O argumento em relação ao número de imagens presentes nos bancos e principalmente nas características dos bancos, torna-se mais visível quando analisa o modelo N-CNN treinado em ambos os bancos de imagens e testado de forma a verificar seu desempenho em outro banco. Os valores da 2° linha nas Tabelas IV e V, mostram que a N-CNN treinada com a COPE foi aproximadamente 1% melhor que a N-CNN treinada com o banco da Unifesp, entretanto, ao analisar a

N-CNN treinada com o banco da Unifesp mostrou-se mais flexível ao ser submetida a um teste com outro banco de imagens, obtendo uma acurácia de 70.3% e a N-CNN treinada com a COPE obteve 43.2% de acurácia. Logo, é visível que o número de imagens para treinamento é relevante no desempenho de um modelo de DL e mesmo impondo um aumento de dados não significa que o problema do número de imagens de treinamento foi resolvido, pois o aumento de dados não cria novos dados, apenas tenta resolver o problema de Posição, Orientação e Escala (POS). Além da questão do número de imagens, o que se torna mais relevante nesse artigo não é o fato que a N-CNN proposta em [29] foi melhor que a ResNet, um modelo que é considerado o estado-da-arte, uma vez que a N-CNN foi modelada para a tarefa de classificação da dor neonatal e a ResNet se mostrou com bom desempenho nas tarefas de reconhecimento e localização de imagens para muitas tarefas de reconhecimento visual. A grande relevância está na constatação de que, fora a importância do número de imagens no banco, a diversidade racial das imagens é imprescindível, uma vez que os modelos que foram treinados com o banco de imagens plurirracial (Banco da Unifesp) obtiveram uma flexibilidade melhor quando foram testados com outro banco de imagens, assim sendo superior aos modelos treinados com apenas neonatos caucasianos (Banco da COPE).

## VI. CONCLUSÃO E TRABALHOS FUTUROS

Este artigo avaliou a dor neonatal por meio da expressão facial explorando e propondo modelo de Aprendizado Profundo. Primeiro, explorou-se o modelo Neonatal Convolutional Neural Network (N-CNN), proposta por Zamzmi em 2019 [29] e afirmada pela mesma sendo a primeira topologia para realizar tal avaliação; segundo, o modelo ResNet50 modificada por Zamzmi em 2019 [29] através do conceito de Aprendizado por Transferência (TL). Modificações da ResNet50 desenvolvida por Zamzmi et al. [29] foram propostas, assim criando uma outra ResNet50 a partir da primeira.

Os experimentos com dois bancos de imagens distintos, Unifesp e COPE, demonstraram que a alteração do modelo Resnet50 proposto por este artigo gerou resultados melhores de classificação, para o banco de imagens COPE. Verificou-se também que o número de imagens no banco influenciam na acurácia dos modelos, mesmo tentando contornar o problema com técnica de aumento de dados, uma vez que tal técnica não cria novos dados, mas, tenta resolver o problema de POS. Além da questão do número de imagens, os resultados mostraram que treinar modelos com bancos de imagens plurirraciais pode aumentar a robustez desses modelos, pois bancos plurirraciais possibilitam um espectro maior de semânticas do estado “Dor” e “Sem Dor”.

Vislumbra-se, como trabalhos futuros, estender essas análises baseadas em modelos de Aprendizado Profundo utilizando validações cruzadas, matrizes de confusão e mapas de ativação dos grupos de padrões de interesse.

## AGRADECIMENTOS

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Código de Financiamento 001 e do Centro Universitário FEI.

## REFERÊNCIAS

- [1] G. Zamzmi, “Automatic multimodal assessment of neonatal pain,” Ph.D. dissertation, University of South Florida, 2018a.
- [2] K. J. Anand and D. B. Carr, “The neuroanatomy, neurophysiology, and neurochemistry of pain, stress, and analgesia in newborns and children,” *Pediatric Clinics of North America*, vol. 36, no. 4, pp. 795–822, 1989.
- [3] B. Golianu, E. J. Krane, K. S. Galloway, and M. Yaster, “Pediatric acute pain management,” *Pediatric Clinics of North America*, vol. 47, no. 3, pp. 559–587, 2000.
- [4] O. F. COMMITTEE *et al.*, “Prevention and management of procedural pain in the neonate: An update,” *Pediatrics*, vol. 137, no. 2, p. e20154271, 2016.
- [5] J. Vinall, S. P. Miller, V. Chau, S. Brummelte, A. R. Synnes, and R. E. Grunau, “Neonatal pain in relation to postnatal growth in infants born very preterm,” *Pain*, vol. 153, no. 7, pp. 1374–1381, 2012.
- [6] M. DiLorenzo, R. Pillai Riddell, and L. Holsti, “Beyond acute pain: understanding chronic pain in infancy,” *Children*, vol. 3, no. 4, p. 26, 2016.
- [7] S. Brummelte, R. E. Grunau, V. Chau, K. J. Poskitt, R. Brant, J. Vinall, A. Gover, A. R. Synnes, and S. P. Miller, “Procedural pain and brain development in premature newborns,” *Annals of neurology*, vol. 71, no. 3, pp. 385–396, 2012.
- [8] R. E. Grunau, M. T. Tu, M. F. Whitfield, T. F. Oberlander, J. Weinberg, W. Yu, P. Thiessen, G. Gosse, and D. Scheifele, “Cortisol, behavior, and heart rate reactivity to immunization pain at 4 months corrected age in infants born very preterm,” *The Clinical journal of pain*, vol. 26, no. 8, p. 698, 2010.
- [9] R. E. Grunau, J. Weinberg, and M. F. Whitfield, “Neonatal procedural pain and preterm infant cortisol response to novelty at 8 months,” *Pediatrics*, vol. 114, no. 1, pp. e77–e84, 2004.
- [10] S. M. Walker, “Translational studies identify long-term impact of prior neonatal pain experience,” *Pain*, vol. 158, pp. S29–S42, 2017.
- [11] A. Marchant, “‘neonates do not feel pain’: a critical review of the evidence,” *Bioscience Horizons: The International Journal of Student Research*, vol. 7, 2014.
- [12] A. T. Bhutta and K. Anand, “Vulnerability of the developing brain: neuronal mechanisms,” *Clinics in perinatology*, vol. 29, no. 3, pp. 357–372, 2002.
- [13] K. Anand and F. M. Scalzo, “Can adverse neonatal experiences alter brain development and subsequent behavior?” *Neonatology*, vol. 77, no. 2, pp. 69–82, 2000.
- [14] R. Grunau, “Self-regulation and behavior in preterm children: effects of early pain,” *Progress in pain research and management*, vol. 26, pp. 23–56, 2003.
- [15] B. Stevens, C. Johnston, P. Petryshen, and A. Taddio, “Premature infant pain profile: development and initial validation,” *The Clinical journal of pain*, vol. 12, no. 1, pp. 13–22, 1996.
- [16] K. Anand, “International evidence-based group for neonatal pain consensus statement for the prevention and management of pain in the newborn,” *Arch Pediatr Adolesc Med*, vol. 155, no. 2, pp. 173–180, 2001.
- [17] J. G. Zwicker, S. P. Miller, R. E. Grunau, V. Chau, R. Brant, C. Studholme, M. Liu, A. Synnes, K. J. Poskitt, M. L. Stiver *et al.*, “Smaller cerebellar growth and poorer neurodevelopmental outcomes in very preterm infants exposed to neonatal morphine,” *The Journal of pediatrics*, vol. 172, pp. 81–87, 2016.
- [18] R. Guinsburg, “Avaliação e tratamento da dor no recém-nascido,” *J Pediatr (Rio J)*, vol. 75, no. 3, pp. 149–60, 1999.
- [19] P. Hummel and M. van Dijk, “Pain assessment: current status and challenges,” in *Seminars in Fetal and Neonatal medicine*, vol. 11, no. 4. Elsevier, 2006, pp. 237–245.
- [20] S. H. Simons, M. van Dijk, K. S. Anand, D. Roofthoof, R. A. van Lingen, and D. Tibboel, “Do we still hurt newborn babies?: A prospective study of procedural pain and analgesia in neonates,” *Archives of pediatrics & adolescent medicine*, vol. 157, no. 11, pp. 1058–1064, 2003.

- [21] T. M. Heiderich, “Desenvolvimento de software para identificar a expressão facial de dor do recém-nascido.” 2013.
- [22] M. Schiavenato, J. F. Byers, P. Scovanner, J. M. McMahon, Y. Xia, N. Lu, and H. He, “Neonatal pain facial expression: Evaluating the primal face of pain,” *Pain*, vol. 138, no. 2, pp. 460–471, 2008.
- [23] M. Velana, S. Gruss, G. Layher, P. Thiam, Y. Zhang, D. Schork, V. Kessler, S. Meudt, H. Neumann, J. Kim *et al.*, “The senseemotion database: A multimodal database for the development and systematic validation of an automatic pain-and emotion-recognition system,” in *IAPR Workshop on Multimodal Pattern Recognition of Social Signals in Human-Computer Interaction*. Springer, 2016, pp. 127–139.
- [24] D. L. Martinez, O. Rudovic, D. Doughty, J. A. Subramony, and R. Picard, “Automatic detection of nociceptive stimuli and pain intensity from facial expressions,” *The Journal of Pain*, vol. 18, no. 4, p. S59, 2017.
- [25] P. Rodriguez, G. Cucurull, J. González, J. M. Gonfaus, K. Nasrollahi, T. B. Moeslund, and F. X. Roca, “Deep pain: Exploiting long short-term memory networks for facial expression classification,” *IEEE transactions on cybernetics*, 2017.
- [26] S. Walter, S. Gruss, H. Ehleiter, J. Tan, H. C. Traue, P. Werner, A. Al-Hamadi, S. Crawcour, A. O. Andrade, and G. M. da Silva, “The biovid heat pain database data for the advancement and systematic validation of an automated pain recognition system,” in *2013 IEEE international conference on cybernetics (CYBCO)*. IEEE, 2013, pp. 128–131.
- [27] P. Werner, A. Al-Hamadi, and R. Niese, “Comparative learning applied to intensity rating of facial expressions of pain,” *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 28, no. 05, p. 1451008, 2014.
- [28] G. Zamzmi, D. Goldgof, R. Kasturi, and Y. Sun, “Neonatal pain expression recognition using transfer learning,” *arXiv preprint arXiv:1807.01631*, 2018b.
- [29] G. Zamzmi, R. Paul, D. Goldgof, R. Kasturi, and Y. Sun, “Pain assessment from facial expression: Neonatal convolutional neural network (n-cnn),” in *2019 International Joint Conference on Neural Networks (IJCNN)*. IEEE, 2019, pp. 1–7.
- [30] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.
- [31] F. Zhang, T. Zhang, Q. Mao, and C. Xu, “Joint pose and expression modeling for facial expression recognition,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 3359–3368.
- [32] J. Deng, J. Guo, Y. Zhou, J. Yu, I. Kotsia, and S. Zafeiriou, “Retinaface: Single-stage dense face localisation in the wild,” *arXiv preprint arXiv:1905.00641*, 2019.
- [33] P. Viola and M. J. Jones, “Robust real-time face detection,” *International journal of computer vision*, vol. 57, no. 2, pp. 137–154, 2004.
- [34] R. K. Srivastava, K. Greff, and J. Schmidhuber, “Highway networks,” *arXiv preprint arXiv:1505.00387*, 2015.
- [35] S. Brahnam, C.-F. Chuang, F. Y. Shih, and M. R. Slack, “Machine recognition and representation of neonatal facial displays of acute pain,” *Artificial intelligence in medicine*, vol. 36, no. 3, pp. 211–222, 2006.
- [36] T. M. Heiderich, A. T. F. S. Leslie, and R. Guinsburg, “Neonatal procedural pain can be assessed by computer software that has good sensitivity and specificity to detect facial movements,” *Acta Paediatrica*, vol. 104, no. 2, pp. e63–e69, 2015.
- [37] R. V. Grunau and K. D. Craig, “Pain expression in neonates: facial action and cry,” *Pain*, vol. 28, no. 3, pp. 395–410, 1987.
- [38] T. Tieleman and G. Hinton, “Lecture 6.5-rmsprop: Divide the gradient by a running average of its recent magnitude,” *COURSERA: Neural networks for machine learning*, vol. 4, no. 2, pp. 26–31, 2012.