CENTRO UNIVERSITÁRIO FEI

LUCAS FONTES BUZUTI

AVALIAÇÃO DE DOR EM EXPRESSÃO FACIAL NEONATAL POR MEIO DE REDES NEURAIS PROFUNDAS

São Bernardo do Campo

2020

LUCAS FONTES BUZUTI

AVALIAÇÃO DE DOR EM EXPRESSÃO FACIAL NEONATAL POR MEIO DE REDES NEURAIS PROFUNDAS

Dissertação de Mestrado, apresentado ao Centro Universitário FEI para obtenção do título de Mestre em Engenharia Elétrica. Orientado pelo Prof. Dr. Carlos Eduardo Thomaz.

São Bernardo do Campo

2020

Fontes Buzuti, Lucas.
Avaliação de dor em expressão facial neonatal por meio de Redes Neurais Profundas / Lucas Fontes Buzuti. São Bernardo do Campo, 2020.
103 p. : il.

Dissertação - Centro Universitário FEI.
Orientador: Prof. Dr. Carlos Eduardo Thomaz.

1. Reconhecimento Automático da Dor. 2. Expressão Facial. 3. Avaliação da Dor Neonatal. 4. Aprendizado Profundo. 5. Rede Neural Convolucional. I. Eduardo Thomaz, Carlos, orient. II. Título.

Elaborada pelo sistema de geração automática de ficha catalográfica da FEI com os dados fornecidos pelo(a) autor(a).

centro universitário FEI

APRESENTAÇÃO DE DISSERTAÇÃO
ATA DA BANCA EXAMINADORA

Programa de Pós-Graduação Stricto Sensu em Engenharia Elétrica

Mestrado

PGE-10

Aluno: Lucas Fontes Buzuti **Matrícula:** 119124-6

Título do Trabalho: Avaliação de dor em expressão facial neonatal por meio de redes neurais profundas.

Área de Concentração: Processamento de Sinais e Imagens

Orientador: Prof. Dr. Carlos Eduardo Thomaz

Data da realização da defesa: 07/08/2020

ORIGINAL ASSINADA

Avaliação da Banca Examinadora:

São Bernardo do Campo, / / .

MEMBROS DA BANCA EXAMINADORA	
Prof. Dr. Carlos Eduardo Thomaz	Ass.: _____
Prof. Dr. Paulo Sergio Silva Rodrigues	Ass.: _____
Profª Drª Daniela Testoni	Ass.: _____

A Banca Julgadora acima-assinada atribuiu ao aluno o seguinte resultado:

APROVADO ☒ REPROVADO ☐

VERSÃO FINAL DA DISSERTAÇÃO

APROVO A VERSÃO FINAL DA DISSERTAÇÃO EM QUE FORAM INCLUÍDAS AS RECOMENDAÇÕES DA BANCA EXAMINADORA

_____

Aprovação do Coordenador do Programa de Pós-graduação

_____

Prof. Dr. Carlos Eduardo Thomaz

Dedico esta dissertação à minha mãe Marinalva Fontes grande colaboradora e incentivadora. Nenhuma palavra pode descrever o quanto sou grato por sua presença na minha vida. Obrigado por seu amor e apoio incondicional.

AGRADECIMENTOS

Nesses anos de mestrado, de muito estudo, esforço e empenho, gostaria de agradecer a algumas pessoas que me acompanharam e foram fundamentais para a realização de mais este sonho. Por isso, expresso aqui, através de palavras sinceras, um pouquinho da importância que elas tiveram, e ainda têm, nesta conquista e a minha sincera gratidão a todas elas. Primeiramente, agradeço à minha mãe Marinalva pela compreensão, ao ser privada em muitos momentos da minha companhia e atenção, e pelo profundo apoio, me estimulando nos momentos mais difíceis. Obrigado por desejar sempre o melhor para mim, pelo esforço que fez para que eu pudesse superar cada obstáculo em meu caminho e chegar aqui e, principalmente, pelo amor imenso que você tem por mim. À você, minha mãe, sou eternamente grato por tudo que sou, por tudo que consegui conquistar e pela felicidade que tenho.

Meu eterno agradecimento ao meu orientador, Dr. Carlos Eduardo Thomaz, um querido e grande amigo, pela pessoa e profissional que é. Obrigado por sua dedicação, por sempre ter acreditado e depositado sua confiança em mim ao longo de todos esses anos de trabalho que se iniciaram ainda na graduação. Sem sua orientação, apoio, confiança e amizade, não somente neste trabalho, mas em todo o caminho percorrido até aqui, nada disso seria possível.

Finalizo agradecendo à CAPES e ao Centro Universitário FEI pela bolsa de estudos do mestrado.

“Out of clutter, find simplicity. From discord, find harmony. In the middle of difficulty lies opportunity.”

Albert Einstein

“O que as vitórias têm de mau é que não são definitivas. O que as derrotas têm de bom é que também não são definitivas.”

José Saramago

RESUMO

A avaliação da dor neonatal pode sofrer variações entre profissionais de saúde, resultando em intervenção tardia e tratamento inconsistente da dor. Portanto, faz-se fundamental desenvolver ferramentas computacionais de avaliação da dor menos subjetivas e que não sofram influências de variáveis externas. Modelos de Aprendizado Profundo, especialmente baseados em Redes Neurais Convolucionais, ganharam popularidade nas últimas décadas devido à ampla gama de aplicações bem-sucedidas em análise de imagens, reconhecimento de objetos e reconhecimento de emoções humanas. Neste contexto, o objetivo geral desta dissertação foi analisar, quantitativa e qualitativamente, modelos de Redes Neurais Convolucionais na tarefa de classificação automática da dor neonatal por meio de um arcabouço computacional baseado em imagens de faces de dois bancos de dados distintos (um internacional, denominado COPE, e outro nacional, denominado UNIFESP). Como objetivos específicos foram implementados, avaliados e comparados três modelos existentes de redes neurais usados na literatura afim: Neonatal Convolutional Neural Network (N-CNN) e dois tipos da arquitetura ResNet50. 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 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.

Palavras-chave: Reconhecimento automático da dor, expressão facial, avaliação da dor neonatal, aprendizado profundo, classificação, rede neural convolucional

ABSTRACT

Neonatal pain assessment might suffer variation among health professionals, leading to late intervention and flimsy treatment of pain in several occasions. Therefore, it is essential to develop computational tools of pain assessment, less subjective and susceptible to external variable influences. Deep learning models, especially Convolutional Neural Networks, have gained ground in the last decade, due to many successful applications in image analysis, object recognitions and human emotion recognitions. In this context, the general aim this dissertation was analyse quantitatively and qualitatively models of Convolutional Neural Networks in the task neonatal pain classification through a computacional framework based in face images of two distinct databases (an international, named COPE, and other national, named UNIFESP). How specific aims were implemented, evaluated and compared the performance of three existent models used in literature: Neonatal Convolutional Neural Network (N-CNN) and two type of ResNet50 models. The quantitative results showed the excellence of N-CNN to neonatal pain assessment automatic, with average accuracy of 87.2% and 78.7% for the databases COPE and UNIFESP, respectively. However, the quantitative analysis showed that all neural models evaluated, including N-CNN models, can learn artifacts from the imagens and not variation discriminating in faces, thus showed the necessity more studies to apply this models in clinical practice.

Keywords: Automated pain recognition, facial expression, neonatal pain assessment, deep learing, classification, convolutional neural network

LISTA DE ILUSTRAÇÕES

LISTA DE TABELAS

LISTA DE ABREVIATURAS

AAM	Modelo de Aparência Ativa (Active Appearance Model)
ANOVA	Análise Estatística (Statistical Analysis)
AUC	Área Sob a Curva ROC (Area under the ROC Curve)
CAPP	Canonical Normalized Appearance
CapsNet	Rede Neural de Cápsula (Capsule Neural Network)
CNN	Rede Neural Convolucional (Convolutional Neural Network)
COPE	Classification of Pain Expressions
CRIES	Choro, nível de O2, aumento da VS, insônia (Crying, requires O2 increased VS sleepless)
CV	Visão Computacional (Computer Vision)
DL	Aprendizado Profundo (Deep Learning)
ELBP	Padrão Binário Alongado (Elongated Binary Pattern)
ELTP	Padrão Ternário Alongado (Elongated Ternary Pattern)
FACS	Sistema de Codificação de Ação Facial (Facial Action Coding System)
FSVM	Máquina de Vetor de Suporte Fuzzy (Fuzzy Support Vector Machine)
GSR	Resposta Galvânica da Pele (Galvanic Skin Response)
IASP	Associação Internacional para o Estudo da Dor (Internacional Association for the Study of Pain)
KNN	K Vizinhos Mais Próximos (K-Nearest Neighbors)
LBP	Padrão Binário Local (Local Binary Pattern)
LDA	Análise Discriminante Linear (Linear Discriminant Analysis)
LPC	Coeficiente Preditivo Linear (Linear Predictive Coefficient)
LPCC	Coeficientes Cepstrais de Predição Linear (Linear Prediction Cepstral Coefficients)
LTP	Padrão Ternário Local (Local Ternary Pattern)
MFCC	Coeficientes de Frequências Mel-Cepstrais (Mel-Frequency Cepstral Coefficients)
ML	Aprendizado de Máquina (Machine Learning)
NFCS	Sistema de Codificação Facial Neonatal (Neonatal Facial Coding System)
NIPS	Escala Neonatal de Dor Infantil (Neonatal Infant Pain Scale)

NIRS	Espectroscopia no Infravermelho Próximo (Near-Infrared Spectroscopy)
NNSOA	Algoritmo de Otimização Simultânea de Rede Neural (Neural Network Simultaneous Optimization Algorithm)
N-PASS	Escala Neonatal de Dor, Agitação e Sedação (Neonatal Pain, Agitation, and Sedation Scale)
PCA	Análise de Componentes Principais (Principal Component Analysis)
POS	Posição, Orientação e Escala (Position, Orientation and Scale)
ROC	Receiver Operating Characteristic
RVM	Máquina de Vetor de Relevância (Relevance Vector Machine)
SAPP	Similarity Normalized Appearance
SFFS	Busca Sequencial Flutuante para Frente (Sequential Floating Forward Selection)
SPTS	Similarity Normalized Shape
STD	Desvio Padrão (Standard Deviation)
STE	Energia de Curta Duração (Short-time Energy)
SVM	Máquina de Vetores de Suporte (Support Vector Machine)
UNIFESP	Universidade Federal de São Paulo
UTIN	Unidade de Terapia Intensiva Neonatal

SUMÁRIO

1 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 (ZAMZMI, 2018a). Tal crença foi refutada por diversos estudos científicos (ANAND; CARR, 1989; GOLIANU et al., 2000).

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 (COMMITTEE et al., 2016; VINALL et al., 2012; DILORENZO et al., 2016; BRUMMELTE et al., 2012), funcionamento do sistema de resposta ao estresse (altos níveis de cortisol) (GRUNAU et al., 2010, 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; MARCHANT, 2014; VINALL et al., 2012), atraso no desenvolvimento corticoespinhal (DILORENZO et al., 2016; VINALL et al., 2012), 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 (BHUTTA; ANAND, 2002; ANAND; SCALZO, 2000). Tais alterações podem resultar em uma variedade de deficiências comportamentais, de desenvolvimento e de aprendizagem, segundo Grunau et al. (2010), Grunau (2003) e Stevens et al. (1996).

Anand (2001) 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. (2016) relatou estes efeitos após observar, em seu estudo, que o aumento de 10 vezes no uso de morfina (um agente comumente usado para o tratamento da dor neonatal) estava associado ao comportamento 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 (STEVENS et al., 1996; GUINSBURG, 1999).

De acordo com Zamzmi (2018a), estudos afirmam que há falha em reconhecer e tratar a dor neonatal e também na administração de determinados medicamentos analgésicos na ausência de dor (pouco tratamento versus excesso de tratamento). Apesar do neonato ser incapaz de manifestar verbalmente suas dores, seu corpo responde a estímulos dolorosos, definidos de três maneiras divergentes: comportamentais (tais como: expressão facial e choro), fisiológicas (tais como: frequência cardíaca e pressão sanguínea) e metabólicas (ZAMZMI, 2018a; ZAMZMI et al., 2017). Em seu estudo, Anand et al. (2007) afirmaram que as respostas aos estímulos diferem entre os tipos de dores, sendo: dor aguda e crônica (ANAND et al., 2007 apud ZAMZMI et al., 2017). Zamzmi (2018a) afirma que os neonatos têm uma resposta comportamental geralmente mais intensa ao estímulo doloroso crônicos em comparação à resposta à dor prolongada aguda, uma vez que pode ser atribuído às baixas reservas físicas para sustentar uma resposta ao nível de sedação ou analgesia. Também em outro trabalho, Zamzmi et al. (2017) relatam que os bebês podem experimentar diferentes tipos de dores simultaneamente.

Segundo Hummel e Dijk (2006) e Simons et al. (2003), 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 comunicação contactual ainda não presente no recém-nascido. Os métodos atuais para avaliar a dor nessa população vulnerável dependem da observação de profissionais bem qualificados, que observam as múltiplas repostas comportamentais e fisiológicas aliadas. No entanto, há uma dificuldade na utilização dessas informações, relatada por Heiderich (2013), 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 (SCHIAVENATO et al., 2008; HEIDERICH, 2013; ZAMZMI et al., 2019) 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 proposta para avaliar dor de adultos (VELANA et al., 2016; MARTINEZ et al., 2017; RODRI-

GUEZ et al., 2017; WALTER et al., 2013; WERNER et al., 2014). Os motivos destacados por Zamzmi (2018a) 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. Dois dos poucos trabalhos que utilizam a tecnologias de Aprendizado Profundo são (ZAMZMI et al., 2018b, 2019), em que aplicam o Aprendizado por Transferência em algumas arquiteturas de rede neurais, tal como a Residual Network (ResNet) (HE et al., 2016), para classificar a dor neonatal.

1.1 OBJETIVO

O objetivo geral desta dissertação é analisar quantitativa e qualitativamente modelos de Redes Neurais Convolucionais na tarefa de classificação automática da dor neonatal por meio de um arcabouço computacional baseado em imagens de faces de dois bancos de dados distintos (um internacional, denominado Classification of Pain Expressions (COPE), e outro nacional, denominado UNIFESP). Como objetivos específicos são implementados, avaliados e comparados três modelos existentes de redes neurais: Neonatal Convolutional Neural Network (N-CNN), proposta por Zamzmi et al. (2019), ResNet50 proposta por Zamzmi et al. (2019) e uma alteração nessa ResNet50 proposta por esta dissertação.

1.2 ESTRUTURA DESTA DISSERTAÇÃO

Esta dissertação é estruturada por 5 capítulos. No capítulo 2 seguinte, está descrevendo o estado da arte em avaliação da dor neonatal, os trabalhos relacionados ao assunto estudado nesta dissertação e os métodos computacionais de relevância até então para avaliação automática da dor neonatal. No capítulo 3 são descritos os materiais, os detalhes da metodologia e as métricas de avaliações topológicas propostas nesta dissertação. No capítulo 4 são descritos os experimentos e resultados dos modelos avaliados. Por fim, o capítulo 5 descreve a conclusão dos experimentos obtidos por esta dissertação.

2 TRABALHOS RELACIONADOS

Este capítulo descreve o estado da arte em avaliação da dor neonatal, os trabalhos relacionados ao assunto estudado nesta dissertação e os métodos computacionais de relevância até então para avaliação automática da dor neonatal.

2.1 AVALIAÇÃO DA DOR NEONATAL

A avaliação realizada por profissionais qualificados é de suma importância no reconhecimento do sofrimento por parte do recém-nascido, visto que o recém-nascido não consegue expressar a dor (ANAND et al., 2007). A avaliação precisa da dor proporciona aos profissionais entender a situação em que essa população vulnerável se encontra e prosseguir com tratamento imediato, pois sérios problemas a curto e longo prazos podem ocorrer com os neonatos expostos à dor.

Anand e Craig (1996) demostraram que os sistemas anatômicos sensoriais e neuroquímicos estão suficientemente desenvolvidos no nascimento para permitirem a percepção da dor. É essencial que haja condições adequadas para avaliação da dor em recém-nascidos, uma vez que fortes evidências em relação à exposição extensa à dor durante o período inicial da vida estão associadas às alterações estruturais e funcionais do cérebro (BRUMMELTE et al., 2012; MARCHANT, 2014; VINALL et al., 2012; DILORENZO et al., 2016; VINALL et al., 2012; BHUTTA; ANAND, 2002; ANAND; SCALZO, 2000). As alterações podem resultar em uma variedade de deficiências comportamentais, de desenvolvimento e de aprendizagem segundo Grunau et al. (2010), Grunau (2003) e Stevens et al. (1996).

Segundo Guinsburg e Cuenca (2010), é necessário dispor de instrumentos que traduzem a linguagem da dor para atuar de forma terapêutica em situações possivelmente dolorosas. Mediante ao fato de que é necessário a tradução da dor, foram desenvolvidas escalas unidimensionais para avaliação da resposta comportamental à dor e ferramentas multidimensionais combinando parâmetros objetivos e subjetivos relacionados à dor em neonatos (GUINSBURG; CUENCA, 2010; SCOPEL et al., 2007). Melo et al. (2014) afirmam que os instrumentos unidimensionais são designados para quantificar a presença ou ausência da dor e obtêm informações rápidas, não invasivas, sobre a validação da dor e a analgesia, e os instrumentos multidimensionais são empregados para avaliar componentes sensoriais afetivos e avaliativos que estão refletidos na linguagem usada para descrever a experiência dolorosa. De acordo com Pasero e

McCaffery (1999), escalas de dores foram desenvolvidas para obtenção de informações a respeito das características da dor, da maneira como o paciente a expressa e os efeitos dessa sobre seu cotidiano.

Anand et al. (2007) afirmam que as respostas aos estímulos diferem entre os tipos de dores (ANAND et al., 2007 apud ZAMZMI et al., 2017). Os tipos de dores são definidos como: a dor aguda é geralmente associada por um curto estímulo doloroso e finaliza assim que a causa da dor é removida, em outras palavras, é desencadeada por um estímulo claro, tal como um procedimento cirúrgico, e tem um ponto de início claramente definido e um ponto final esperado, a intensidade desse tipo de dor diminui em função do tempo a partir da ocorrência do estímulo; a dor crônica neonatal é uma dor persistente e contínua, dura além do tempo normal de cicatrização de três meses, não tem um ponto final esperado, e pode evoluir de modo persistente, em decorrência de uma doença de base, ou de modo recorrente, que se caracteriza por surtos com duração, intensidade e frequência diversificados, separados por períodos assintomáticos (OKADA et al., 2001). Devido às baixas reservas físicas para sustentar uma resposta ao nível de sedação ou analgesia, Zamzmi (2018a) afirma que os recém-nascidos têm uma resposta comportamental geralmente mais intensa ao estímulo doloroso crônico e relata que são capazes de experimentar diferentes tipos de dores simultaneamente.

A avaliação em recém-nascidos representa um grande desafio, mesmo com a utilização das escalas de dores. Arias e Guinsburg (2012) afirmam que a falta de uma escala padrão-ouro capaz de medir a intensidade da dor é uma das principais limitações para alcançar o alívio da dor neonatal, visto que o padrão-ouro atual para avaliação clínica da dor refere-se somente a adultos. Melo et al. (2014) ressaltam que, para alguns autores, as escalas multidimensionais são mais adequadas na avaliação da dor neonatal, por terem respostas comportamentais associadas às respostas fisiológicas, tornando as abordagens mais completas. Em contraste, Guinsburg e Cuenca (2010) ressaltam que "..., escalas que levam em consideração parâmetros fisiológicos poderiam ser menos sensíveis à expressão da dor no período neonatal. Em essência, as escalas unidimensionais parecem ser ferramentas mais sensíveis para identificar os indivíduos com dor, quando comparadas às escalas multidimensionais.". De fato, Arias e Guinsburg (2012) constataram que as escalas unidimensionais foram mais sensíveis à detecção da dor, durante a primeira hora de vida com uma injeção de vitamina K, do que as escalas multidimensionais.

No contexto da avaliação da dor neonatal, recomenda-se que pelo menos um dos instrumentos empregados nas múltiplas escalas pelos diferentes profissionais de saúde seja uma escala unidimensional comportamental (GUINSBURG; CUENCA, 2010). Na Tabela 1, pode-

se encontrar as escalas de dor neonatal mais utilizadas no campo da neonatologia. Mais detalhes sobre as escalas de dor podem ser encontrados em (DA SILVA; DA SILVA, 2010).

Tabela 1 – Exemplos de escalas de dor neonatal comuns.

Pain scale	**Pain type**	**Age range**	**Behavioral measures**	**Physiological measures**	**Psychometric properties**
NIPS (LAWRENCE et al., 1993)	Procedural	28–38 gestations weeks	Facial expression, crying, arms/legs, movement, and arousal state	Breathing patterns	Inter-rater reliability: (r = 0.92–0.97) Internal consistency: (Cronbach's α = 0.87–0.95) Content validity Concurrent validity: (r = 0.53–0.83)
NFCS (GRUNAU; CRAIG, 1987)	Procedural	Preterm $\geq$ 25 gestations weeks to term infants	Brow bulge, eye squeeze nasolabial furrow, open lips, horizontal mouth, vertical mouth, lips pursed, taut tongue, chin quiver, and tongue protrusion	N/A	Inter-rater and intra rater reliability > 0.85 Internal consistency: (Cronbach's α = 0.87–0.95) Content and face validity Construct validity
N-PASS (HUMMEL et al., 2008)	Postoperative	23-40 gestations weeks	Facial expression, behavior movements, crying/irritability, and extremities tone	Heart rate, respiratory rate, blood pressure, and oxygen saturation	Inter-rater reliability: (r = 0.85–0.95) Intrarater reliability: (r = 0.87) Internal consistency: (Cronbach's α = 0.84–0.89) Construct validity: ($P < 0.0001$)
CRIES (KRECHEL; BILDNER, 1995)	Postoperative	32-60 gestations weeks	Facial expression, crying, and, sleeping state	Requires increased oxygen and VS	Inter-rater reliability: (r = 0.98) Construct and content validity

Fonte: Zamzmi et al., 2017

Segundo Heiderich (2013) e Heiderich et al. (2015), há três limitações agregadas às escalas de dor neonatal:

a) A avaliação da dor pelos profissionais são realizadas em diferentes intervalos de tempo e ocorre a incapacidade de fornecer uma avaliação contínua da dor. O monitoramento contínuo é importante, uma vez que os bebês podem sofrer com a ocorrência da dor quando são deixados sem vigilância. Isto acaba sendo verídico para a dor prolongada aguda, que requer cuidados intensivos contínuos da detecção da dor para ocorrer a intervenção.

b) Embora as escalas sejam baseadas nas observações diretas e na interpretação do profissional em múltiplas respostas (comportamentos, fisiológico e metabólico), a avaliação pode ser tendenciosa e afetada por diversos fatores próprios e particulares de uma pessoa ou de um grupo, como o viés cognitivo, a identidade, a cultura e o gênero (MILLER; NEWTON, 2006; SAMOLSKY DEKEL et al., 2016; PILLAI RIDDELL; CRAIG, 2007; PILLAI RIDDELL et al., 2004). Uma avaliação errada pode levar a uma ocorrência de tratamento inconsistente da dor.

c) Há um compromisso substancial de tempo e necessita de um grande número de cuidadores bem treinados para utilização adequada da escala de dor. O custo substancial torna-se inviável em países subdesenvolvidos, onde que os profissionais e recursos médicos são escassos.

2.2 AVALIAÇÃO AUTOMÁTICA DA DOR NEONATAL

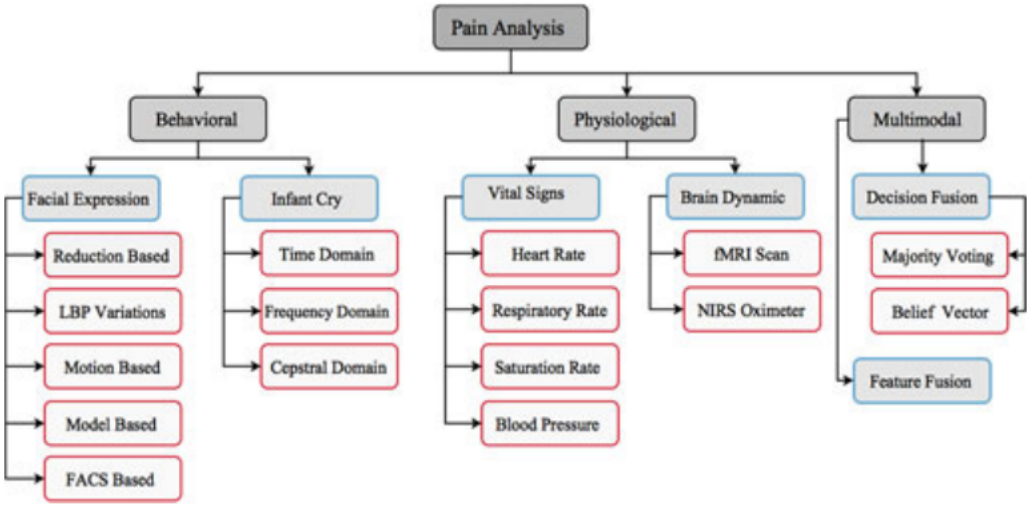
Nos últimos anos, tem ocorrido um crescimento no uso de métodos de Aprendizado de Máquina (ML) para entender as respostas comportamentais à dor utilizando bases de expressões faciais (WERNER et al., 2014; VELANA et al., 2016; SIKKA et al., 2015; PAL et al., 2006), sons de choros (PAL et al., 2006; VARALLYAY et al., 2004; PAI, 2016) e de movimentos do corpo (ZAMZMI et al., 2016). Entretanto, esse crescimento é muito menor em comparação com métodos de detecção da dor em adultos. De fato é um erro pensar que métodos desenvolvidos para adultos possam ser usados em recém-nascidos. Primeiro, a dinâmica da dor e a morfologia facial variam entre bebês e adultos. Por exemplo, expressões faciais de bebês com movimentos adicionais não são decodificados pelo Sistema de Codificação de Ação Facial (FACS) (EKMAN, 1997). Há uma extensão do FACS que é conhecida como Sistema de Codificação de Ação Facial do Bebê (Baby-FACS) (OSTER, 2006 apud ZAMZMI, 2018a). Também há diferença de som e movimento dos neonatos durante a dor. Adicionalmente, o estágio de pré-processamento dos dados, rastreamento de face e corpo tornam-os mais desafiadores, uma vez que a população neonatal é frágil e considerada não cooperativa. Embora tenha ocorrido um crescimento no uso dessas bases, Zamzmi et al. (2017) expressaram que os sistemas automatizados podem ser usados para detectar emoções de respostas fisiológicas, tais como dilatação da pupila (PARTALA et al., 2000; PARTALA; SURAKKA, 2003; AL-OMAR et al., 2013; LANATÀ et al., 2011), Resposta Galvânica da Pele[1] (GSR) (GRUSS et al., 2015; KÄCHELE et al., 2015), alterações na frequência cardíaca (GRUSS et al., 2015; FAYE et al., 2010) e alterações hemodinâmicas cerebrais (BROWN et al., 2011; RANGER; GÉLINAS, 2014).

Neste contexto, foram introduzidos métodos para extração de características relevantes da dor em dados comportamentais ou fisiológicos de neonatos. Esses métodos podem ser categorizados em: Representações das Características à Resposta Comportamental; Representações das Características à Resposta Fisiológica; Junção das Respostas à Dor. Também cada catego-

[1] A alteração na resistência elétrica da pele associada com a descarga nervosa simpática, traduzido (Miller-Keane M. Miller-Keane Encyclopedia and Dictionary of Medicine, Nursing, and Allied Health. 7th ed. Philadelphia, PA: Saunders, An Imprint of Elsevier, Inc.; 2003).

ria pode ser dividida em subcategorias. Na Figura 1, encontra-se um diagrama em árvore com essas categorias e subcategorias principais.

Figura 1 – Diagrama em árvore dos métodos de análise automática de dor neonatal.



Fonte: Zamzmi et al., 2017

A análise de dor baseada em respostas comportamentais pode ser definida como uma tarefa de extração automática de características relevantes da dor, na qual a partir de indicadores comportamentais da dor, tais como expressão facial e som de choro, pode ser realizada uma análise até a solução do problema. Nas Tabelas 2 e 3 podem ser encontrados estudos dos indicadores comportamentais de expressão facial e som de choro, respectivamente.

A análise da dor baseada em repostas fisiológicas pode ser definida como o processo de extrair características relevantes da dor das respostas fisiológicas do corpo do bebê, tais como alterações dos sinais vitais (por exemplo, aumento da frequência cardíaca) e atividade hemodinâmica cerebral, como ilustrado na Figura 1. Na Tabela 4, está resumido os métodos existentes que fazem o uso de medidas fisiológicas para avaliação da dor.

Há um interesse crescente em compreender as respostas comportamentais e fisiológicas. Modelos de Inteligência Artificial (ZAMZMI et al., 2018b, 2019; SALEKIN et al., 2019), Visão Computacional e Estatística (TERUEL et al., 2019; ORONA et al., 2019) estão sendo utilizados para análise automática da dor neonatal. Tais modelos estão surgindo da crença/carência de monitoramento contínuo e consistente da dor. Esses trabalhos, até então, utilizam de um único indicador ou modalidade para a detecção da dor neonatal. Visto que a dor é manifestada por várias modalidades, as escalas de dor neonatal existentes são multimodais, atribuindo indicadores comportamentais e fisiológicos para tal avaliação. Segundo Zamzmi et al. (2017),

Tabela 2 – Resumo dos métodos de ML para analisar a expressão da dor.

Ref. and Year	Database	Category	Extraction method	Classification	Results
(BRAHNAM et al., 2006)	COPE database Subjects: 26, half girls Race: Caucasian Age range: 18 hours to 3 days Stimuli: Pain stimulus and 3 other stimuli: air, friction, and rest/cry Data: 204 static images	Feature reduction based	Column stacking image's intensities and dimensionality reduction PCA	PCA/LDA with L1 and SVM: Pain/no-pain, pain/rest, pain/cry pain/air-puff, and pain/friction Testing protocol: Tenfold cross-validation	SVM avg. accuracy: Pain/no-pain (88%) Pain/rest (95%) Pain/cry (80%) Pain/air-puff (83%) Pain/friction (93%)
(BRAHNAM et al., 2007)	COPE database	Feature reduction based	Column stacking image's intensities and dimensionality reduction	NNSOA, PCA/LDA and SVM: Pain (60 images) versus no-pain (144 images) Testing protocol: Leave-one-subject-out cross-validation	Average accuracy: NNSOA (90.20%) SVM (82.35%) PCA with L1 (80.35%) LDA with L1 (76.96%)
(GHOLAMI et al., 2010)	COPE database	Feature reduction based	Column stacking image's intensities	RVM Pain/no-pain Pain intensity estimation Testing protocol: Leave-one-image-out cross-validation	Weighted Kappa Coeff.: 0.47 (expert/RVM) 0.46 (nonexpert/RVM)
(SILVA, 2020)	UNIFESP database COPE database	Feature reduction based	Dimensionality reduction PCA	PCA MLDA	Accuracy with $k-fold=5$: UNIFESP 0.8727 $\pm$ 0.0874, COPE 0.6889 $\pm$ 0.0712 Accuracy with $Leave-One-Out$: UNIFESP 0.9573, COPE 0.7660
(TERUEL et al., 2019)	UNIFESP database COPE database	Feature reduction based	Dimensionality reduction PCA	PCA MLDA	MLDA 72.77%
(NANNI et al., 2010)	COPE database	LBP variation based	LBP, LTP, ELTP and ELBP descriptors	SFFS feature selection and SVMs Testing protocol: Leave-one-out cross-validation (SFFS) Train/Test split	Highest (0.93) area under the curve of ROC (ELTP)
(ZAMZMI et al., 2015)	Subjects: 10, half girls Race: Caucasian Age range: 32–41 gestations weeks Stimuli: Pain stimulus (i.e., heel lancing) and normal state Data: Ten videos and NIPS scores	Motion-based (optical flow)	Strain magnitude estimated from flow vectors	KNN and SVM: Pain versus no-pain expression Testing protocol: Tenfold cross-validation	Highest overall accuracy: KNN(96%)
(FOTIADOU et al., 2014)	Subjects: 10 Race: N/A Age range: N/A Stimuli: Heel puncture, diaper change, hunger, and resting Data: 15 videos, ranges from few seconds to minutes	Model-based AAM	AAM-based features SPTS, SAPP, and CAPP	SVM: Discomfort versus comfort Testing protocol: Leave-one-subject-out cross-validation	AUC of ROC: 0.98
(SIKKA et al., 2015)	Subjects: 50 children, 35% boys Race: 35 Hispanic, 9 non-Hispanic white, 5 Asian, and 1 Native American Age range: 5 to 18 years Stimuli: Appendectomy (ongoing) and pressing surgical site (transient) Data: Videos, self-report, and by proxy rating by a nurse and parent	FACS-based (opticalflow)	Strain magnitude estimated from flow vectors	KNN and SVM: Pain expression versus no-pain expression Testing protocol: Tenfold cross-validation	Highest overall accuracy: KNN(96%)

Fonte: Adaptado de Zamzmi et al., 2017

a multimodalidade permite uma avaliação confiável da dor, em casos de falta dos dados. De fato, a multimodalidade têm as suas vantagens, pois tenta retratar as escalas de dor neonatal, entretanto, a multimodalidade é uma união de respostas comportamentais e fisiológicas, sendo assim não basta apenas concatenar métodos de respostas comportamentais a métodos de respostas fisiológicas. O que de fato necessita-se é de uma compreensão metodológica na saída e internamente dos modelos, e abrir o campo de visão para novos modelos que estão surgindo e a surgir, para assim começar a pensar e ter sucesso nos modelos multimodais.

Tabela 3 – Resumo dos métodos de ML para analisar o choro da dor.

Ref. and Year	Database	Category	Extraction method	Classification	Results
(VEMPADA et al., 2012)	Subjects: 120 infants Race: N/A Age range: 12–40 weeks Stimuli: N/A Data: 120 samples; 30 Pain, 60 hunger, and 30 wet-diaper	Time-domain analysis	Short-time energy (STE) and pause duration	SVM Testing protocol: Splitting samples into train and test	SVM avg. accuracy: Pain/no-pain (88%) Pain/rest (95%) Pain/cry (80%) Pain/air-puff (83%) Pain/friction (93%)
(PAL et al., 2006)	N/A	Frequency-domain analysis	F0 fundamental frequency and 3 first formants	k-means: Pain, hunger, fear, sadness, and anger	Classification accuracy: 91%
(FULLER; HORII, 1988)	Subjects: 41 infants Race: Caucasian Age range: 2–6 months Stimuli: Immunization for pain, feeding time for hunger, naptime for fussy, and fondling for cooing Data: 109 samples; 16 hunger, 23 cooing, 42 pain, and 28 fussy	Frequency-domain analysis	Mean value of spectral energy	Statistical analysis (ANOVA)	Unique spectral characteristics of pain-induced cry
(PAI, 2016)	Subjects: 27 infants Race: Caucasian, Hispanic, African American, and Asian Avg. age: 36 gestation weeks Stimuli: Immunization and and heel lancing Data: 34 samples; NIPS score	Frequency-domain analysis	LPC and statistics (e.g., mean and STD)	KNN: Whimper cry Vigorous cry Testing protocol: Tenflod cross-validation	Average accuracy: 76.47%
(PETRONI et al., 1995)	Subjects: 16 Race: N/A Age range: 2–6 months Stimuli: Immunization (pain), jack-in-the-box (fear), and head restraint (anger). Data: 230 cry samples	Cepstral-domain analysis	Ten MFCC coeff.	Neural network: Pain cry versus no-pain cry (fear and anger) Testing protocol: Tenfold cross-validation	Classification accuracy: Pain (92.0%) No-pain (75.7%)
(MONTIEL; GARCIA, 2006)	Subjects, race, age, and stimuli: N/A Data: 1627 samples; 209 pain, 759 hunger, and 659 others (Data collected and labeled by doctors)	Cepstral-domain analysis	16 MFCC coeff. Dimensionality reduction (PCA)	FSVM: Pain cry Hunger cry No-pain-no-hunger cry Testing protocol: Tenfold cross-validation	Average accuracy: 97.83%
(ABDULAZIZ; AHMAD, 2010)	Subjects and race: N/A Age range: Newborns to 1 year Stimuli: Immunization (pain) and spontaneous emotions Data: 180 sample; 150 pain and 30 no-pain	Cepstral-domain analysis	12 MFCC coeff. 16 LPCC coeff.	Neural network: Pain/no-pain Testing protocol: Splitting samples to train and test	Classification accuracy: MFCC: 76.2% LPCC: 68.5%
(VEMPADA et al., 2012)	Database in first row	Cepstral-domain analysis	13 MFCC, Δ MFCC, and Δ Δ MFCC	SVM: Pain, hunger, and wet-diaper Testing protocol: Splitting samples to train and test	Accuracy per class: Pain (30.56%) Hunger (66.67%) Wet-diaper (86.11%)

Fonte: Zamzmi et al., 2017

Tabela 4 – Resumo das publicações para análise da dor usando medidas fisiológicas.

Ref. and Year	Measures	Database	Extracted data	Analysis method	Results
(LINDH et al., 1999)	Vital signs	Subjects: 25 infants Age range: 72–96 h Stimuli: Baseline, sham heel prick, sharp heel prick, and heel squeezing	Heart Rate HR_{mean} , the power in low-frequency (P_{LF}), and high-frequency (P_{HF}) and total heart rate variability (P_{tot})	Multivariate statistics	Increase in HR_{mean} , P_{tot}, and P_{LF}, between baseline and sharp prick
(FAYE et al., 2010)	Vital signs	Subjects: 28 infants Age: > 34 gestational weeks Stimuli: Baseline and a major surgery (postoperative)	Heart rate variability index (HRVI)	Linear regression analysis	Sensitivity (90%) Specificity (75%) Area under ROC (0.81)
(BARTOCCI et al., 2006)	Cerebral hemo-dynamics (NIRS)	Subjects: 40 infants, half male Age: $\geq$ 26 gestational weeks Stimuli: Baseline, tactile, and venipuncture pain stimulus	Difference of concentration of oxygenated [HbO_2] and de-oxygenated [HbH] and total ($HbH + HbO_2$) hemo-globin from baseline	Student t-test ANOVA Newman–Keuls posthoc test	[HbO_2] increases in both hemispheres; more pronounced increase in male
(SLATER et al., 2006)	Cerebral hemo-dynamics (NIRS)	Subjects: 18 infants Age: 25–45 postmentsural weeks Stimuli: Baseline and heel lancing	Vital signs data and mean of [HbO_2], [HbH], and $HB_{total} = HbH + HbO_2$	Statistical t-test	Significant increase in [HB_{total}]; more pronounced increase in awake infants
(RANGER et al., 2013)	Cerebral hemo-dynamics (NIRS)	Subjects: 40 infants Age: < 12 months Stimuli: Baseline (T_0), tactile (T_1), and painful (T_2) stimuli	$[HbH]_{mean}$, $[HbO2]_{mean}$, and $[HR]_{mean}$	Univariate linear regression	ΔHbH differed significantly between T_0 and T_2

Fonte: Zamzmi et al., 2017

2.3 ANÁLISE COMPORTAMENTAL DE DOR NEONATAL

A representação de características da resposta comportamental tem como definição a tarefa de extração de características relevantes da dor de uma resposta comportamental, como expressão facial e som de choro. Devido a subjetividade da dor, há um interesse relevante em métodos que são capazes de extrair e aprender características em faces e sinais sonoros. O choro infantil é um sinal que pode indicar desconforto, fome ou dor, transmitindo informações que ajudam os profissionais de saúde a avaliar o estado emocional do bebê e a reagir adequadamente. Na Tabela 3, pode-se encontrar estudos relacionados a resposta comportamental para análise do som do choro (ZAMZMI et al., 2017).

2.3.1 Expressão Facial

Nesta seção, serão discutidos os métodos existentes que foram utilizados para extrair características na expressão facial de recém-nascidos e classificá-los com dor ou sem dor.

A expressão facial é um dos indicadores mais comuns na percepção de emoções, sendo assim especialmente para dor. Já que expressar emoções ocorre nos movimentos musculares faciais, Zamzmi et al. (2017) definem a expressão facial da dor sendo "os movimentos e distorções nos músculos faciais associados a um estímulo doloroso, traduzido". As distorções dos movimentos faciais associados à dor em bebês estão no aprofundamento do sulco nasolabial, abaixamento da sobrancelha, olhos estreitados, alongamento vertical e horizontal da boca, franzir os lábios, abertura labial, protrusão da língua, língua tensa e tremor no queixo (GRUNAU; CRAIG, 1987).

O reconhecimento da expressão facial da dor tem os mesmos estágios de um reconhecimento qualquer, na qual se faz o uso de expressões faciais. Esses estágios consistem basicamente em: *Primeiro*, detecção e normalização (se houver necessidade) da face; *Segundo*, extração de características; e *Terceiro*, reconhecimento/classificação da expressão facial. O segundo e terceiro estágios são considerados como um só ao utilizar métodos de DL, uma vez que são capazes de extrair características a um nível superior em comparação com os métodos tradicionais (TAIGMAN et al., 2014; BALABAN, 2015). Em relação à detecção facial, não há duvida que já é uma área madura e muito bem estruturada. Em trabalho recente, Deng et al. (2019) propuseram um detector robusto de face de estágio único, chamado RetinaFace, e

foi capaz de detectar aproximadamente 900 faces de um total de 1151 pessoas. Na Tabela 2 destacam-se os principais trabalhos para o reconhecimento da expressão facial da dor.

As subseções seguintes descrevem os métodos de extração de características que foram propostos para o reconhecimento da dor em expressão facial, tomando-se como referência o trabalho de revisão realizado por Zamzmi et al. (2017) e acrescentando o tópico de Aprendizado Profundo.

2.3.1.1 Métodos Baseados em Redução de Dimensionalidade

Zamzmi et al. (2017) citam dois métodos para redução de características que foram utilizados na classificação da dor neonatal, Análise de Componentes Principais (PCA) e Busca Sequencial Flutuante para Frente (SFFS). Cada imagem é convertida em um vetor de dimensão $N_x \times N_y \times 1$, em que N_x e N_y representam a largura e altura da imagem, e o 1 representa o canal (neste caso escala de cinza, com tonalidade de pixels variando de 0 a 255). Sendo assim, PCA e SFFS podem ser aplicados para reduzir tal vetor.

PCA é um procedimento matemático/estatístico que utiliza uma transformação ortogonal para converter um conjunto de observações de variáveis possivelmente correlacionadas em um conjunto de valores de variáveis linearmente não correlacionadas, chamadas de componentes principais, assim reduzindo a dimensionalidade de um determinado espaço de características. Esses componentes representam as dimensões ao longo das quais os dados estão mais espalhados. Em outras palavras, o PCA é a combinação linear das variáveis iniciais e alocadas em ordem decrescente por suas variâncias. Assim, o procedimento matemático é dado por:

$$\begin{aligned} Y^\top &= X^\top W \\ &= V\Sigma^\top W^\top W \\ &= V\Sigma^\top, \end{aligned} \tag{1}$$

no qual, $X^\top$ é a matriz transposta de dados com média empírica nula, cada uma das n linhas representa um dado diferente e cada uma das m colunas são as respectivas variáveis dos dados. Sendo a Decomposição em Valores Singulares de $X = W\Sigma V^\top$, em que W é a matriz $m \times m$ de autovetores da matriz de covariância de $XX^\top$, e a matriz $m \times n$ Σ é uma matriz diagonal com números reais não negativos na diagonal, e por fim a matriz $n \times n$ V é a matriz de autovalores de $XX^\top$. Sendo W uma matriz ortogonal e cada linha de $Y^\top$ sendo a rotação da linha

correspondente de $X^\top$, logo a primeira coluna de $Y^\top$ é construída com a componente de maior variância, a segunda coluna é composta pela componente de segunda maior variância, e assim por diante.

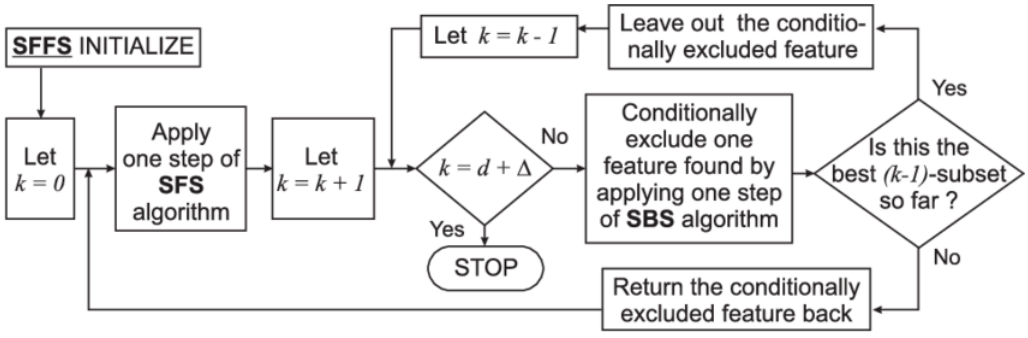
Pode-se representar a Equação (1) em uma dimensionalidade reduzida na qual quer-se projetar X, o espaço reduzido será definido pelos primeiros L vetores singulares W_L:

$$Y = W_L^\top X = \Sigma_L V^\top, \tag{2}$$

onde $\Sigma_L = I_{L\times m}\Sigma$, e $I_{L\times m}$ é a matriz identidade $L \times m$.

Diferentemente do PCA, o método SFFS é um algoritmo e não uma formulação matemática (SOMOL et al., 2010). O método SFFS se enquadra nos métodos de seleção de característica sequencial, visto que são uma família de algoritmos de pesquisa gulosos e são usados para reduzir um espaço inicial de característica d-dimensional para um subespaço k-dimensional, em que modulando k e d adiciona-se ou remove-se sequencialmente uma única característica até que não haja alguma melhoria do desempenho do algoritmo. O SFFS é conhecido como uma extensão da Busca Sequencial Direta. Zamzmi et al. (2017) afirmaram que no SFFS, de acordo com uma função de critério, há a eliminação da pior característica do subconjunto, sendo assim, aumentando e diminuindo dinamicamente o número de características até alcançar o melhor subconjunto. A Figura 2 demonstra o diagrama de blocos do algoritmo SFFS.

Figura 2 – Algoritmo Busca Sequencial Flutuante para Frente (SFFS).



Fonte: Somol et al., 2010

Brahnam et al. (2006) desenvolveram um dos principais trabalhos relacionados para avaliação da dor em recém-nascidos baseado na análise de expressão facial. Nesse estudo, utilizou-se imagens do banco de dados COPE obtendo um taxa de reconhecimento da dor versus não dor de 88% com o classificador Máquina de Vetores de Suporte (SVM), demostrando superi-

oridade a dois métodos comumente usados na época para se classificar faces, PCA e Análise Discriminante Linear (LDA), cada um usando a métrica de distância $L1$.

Para a extração de característica, Brahnam et al. (2006) processaram as imagens para extrair somente a face para então converter-las em escala de cinza e redimensiona-las para 100×120 pixels. Por fim, foi transformada cada imagem em um vetor de dimensão 12000. Esse procedimento foi realizado para as 204 imagens, portanto, uma matriz de dimensão 204×12000 contendo todas as imagens do conjunto de dados foi construída, assim, aplicou-se o PCA nessa matriz para redução de dimensionalidade. Na classificação, classificadores baseados em distância tais como, LDA e SVM foram utilizados para classificar as imagens dos bebês nos seguintes pares: dor/não-dor, dor/repouso, dor/choro, dor/sopro de ar e dor/atrito. O LDA pode ser definido como um método estatístico de aprendizado supervisionado, visto que transforma os dados em um subespaço que maximiza a separação das classes. O SVM também é um algoritmo estatístico de aprendizado supervisionado, em que constrói um hiperplano de separação ideal que melhor segrega novos dados. Os resultados mostrados pelo Brahnam et al. (2006) alcançaram a melhor taxa de reconhecimento e superaram outros classificadores, portanto, classificando dor versus ausência de dor em 88%, dor versus repouso em 94,62% , dor versus choro em 80%, dor versus sopro no ar em 83,33% e dor versus atrito em 93%. Para atingir esses resultados a configuração do SVM foi: um núcleo polinomial de grau 3 avaliado por meio da validação cruzada dividindo-se o conjunto original em dez subconjuntos.

Estendendo o trabalho discutido acima, Brahnam et al. (2007) incluíram em sua análise um Algoritmo de Otimização Simultânea de Rede Neural (NNSOA) para classificação juntamente com SVM, PCA com distância L1 e LDA com distância L1. Além disso, utilizou-se da validação cruzada *leave-one-subject-out*. Os resultados mostraram que o NNSOA tem a maior taxa média de classificação 90,20% ao classificar imagens de bebês como dor ou sem dor, enquanto que, SVM, PCA e LDA alcançaram taxas médias de classificação de 82,35%, 80,39% e 76,96%, respectivamente. Devido a superioridade da NNSOA, em estudos subsequentes Brahnam et al. (2008) utilizaram o algoritmo SFFS e seleção de características, também conhecida como seleção de variáveis (VAR) para extração de características nas imagens da COPE com o classificador NNSOA, além do PCA também como extrator de características e do SVM linear. Os resultados mostraram que o NNSOA+VAR alcançou uma maior taxa de classificação média de 95.38% ao classificar imagens de bebês com dor ou sem dor. Na Tabela 5, pode-se visualizar os resultados obtidos por Brahnam et al. (2008) em seus estudos.

Tabela 5 – Resultados da classificação da dor neonatal aplicando o NNSOA e o Linear SVM a um intervalo de confiança de 95%.

Method	**95% Confidence Interval**	**Standard Deviation**
Linear SVM+VAR	87.53% $\pm$ 6.47%	16.01%
Linear SVM+SFFS	89.17% $\pm$ 5.69%	14.09%
NNSOA+VAR	95.38% $\pm$ 2.81%	6.97%
NNSOA+SFFS	93.18% $\pm$ 4.40%	10.90%
Linear SVM+PCA	82.35% $\pm$ 6.20%	15.34%
NNSOA+PCA	90.20% $\pm$ 4.16%	13.30%

Fonte: Brahnam et al., 2008

Em seu trabalho, Teruel et al. (2019) propuseram uma sequência de procedimentos computacionais para detecção, interpretação e classificação de padrões em imagens bidimensionais frontais de faces para o reconhecimento automático da dor neonatal. Nos procedimentos, Teruel et al. (2019) utilizaram o banco de imagens de recém-nascidos da Universidade Federal de São Paulo (HEIDERICH et al., 2015). Em sua sequência de procedimentos computacionais, os autores fizeram o uso das seguintes etapas: pré-processamento; marcação dos pontos fiduciais; normalização espacial das imagens; construção de um atlas de referência inicial e atlas de recém-nascidos; e análise estatística multivariada (PCA e LDA). O resultado da classificação obtido por Teruel et al. (2019) foram de 72.77% de acurácia.

Estendendo o trabalho (TERUEL et al., 2019), Silva (2020) propôs um arcabouço computacional de interpretação e reconhecimento de padrões em imagens de faces para avaliação automática de dor em recém-nascidos. Silva (2020) concentrou seu trabalho na investigação, implementação e integração de técnicas de detecção, segmentação, normalização espacial e classificação de imagens de faces baseadas em informações extraídas por mineração estatística de dados (PCA e Maximum uncertainty LDA-base (THOMAZ et al., 2007)). O arcabouço computacional proposto por Silva (2020) utilizou duas bases de imagens distintas, COPE (BRAHNAM et al., 2006) e uma base desenvolvida pela UNIFESP (HEIDERICH et al., 2015). Os resultados experimentais obtido pelo autor, mostrou que é possível classificar estatisticamente dor e não-dor através de imagens de faces, mas também evidenciar regiões faciais discriminantes para o fenômeno dor, auxiliando na construção de escalas de dor neonatal mais gerais e assertivas. Entretanto, Silva (2020) destaca que há limitações em seu trabalho uma vez que utilizou uma seleção específica anterior à normalização e classificação, escolhendo então apenas as imagens com ângulos próximos ao de perfil. Silva (2020) também afirma que, para alcançar os melhores resultados foram removidos componentes do PCA.

Mesmo não utilizando método de redução de dimensionalidade, vale citar o trabalho realizado por Gholami et al. (2010) nessa seção, uma vez que apresentou um algoritmo de máquina de kernel esparso, conhecido como máquina de vetor de relevância (RVM), para estimar o nível de intensidade da dor na expressão detectada, em vez de detectar a presença ou ausência da expressão da dor. O RVM nada mais é que um tratamento bayesiano de um modelo linear generalizado de forma funcional idêntica ao SVM. Aplicando o algoritmo RVM com kernel linear, resultou-se em uma acurácia de classificação de 91%. Para medir a avaliação da intensidade da dor foi utilizado um coeficiente κ, em que confronta a intensidade de dor indicada pelos examinadores (profissionais de saúde e não-profissionais) e o algoritmo RVM, para então, comparar a nota de 0 a 100 dada pelos examinadores com a incerteza para a classe "dor" (probabilidade a posteriori) dada pelo algoritmo RVM.

Uma questão peculiar que aborda os trabalhos citados até então é o pré-processamento das imagens de faces, que não se encontram em posições frontais. Os sistemas de análises automáticas de expressões faciais que fazem do uso explícito de extração de características, focam em imagens de visão frontal da face. De acordo com observações clínicas, movimentos com a cabeça ocorrem frequentemente durante experiências dolorosas e essa movimentação resulta em múltiplas visões da face, podendo acarretar em falhas no reconhecimento da dor, visto que o detector poderá entregar faces frontais ou de perfil. Reconhecimento de expressões em imagens faciais de perfil para os métodos descritos até aqui é uma questão desafiadora, pois são métodos que se faziam em invariância translacional e rotacional. Uma alternativa ao PCA é o Autoencoder, um redutor de dimensionalidade não linear (BUZUTI; THOMAZ, 2019).

2.3.1.2 Métodos Baseados em Variação de Padrões Binários Locais

O algoritmo de Padrão Binário Local (LBP) e suas variantes são um dos descritores de textura mais populares em Visão Computacional. Sua popularidade pode ser atribuída a sua simplicidade, baixa complexidade computacional e robustez às variações de iluminação, e alinhamento (OJALA et al., 2002).

Em sua forma mais simples, o LBP descreve a textura da imagem comparando o valor em escala de cinza de um pixel central X com os valores de seus P vizinhos dentro de um círculo predefinido de raio R e considerando a saída da comparação como um valor binário, assim criando um vetor de característica (ZAMZMI et al., 2017). Em outras palavras, cria-se o vetor de característica dividindo a janela examinada (imagem) em células, por exemplo 16×16

pixels para cada célula; cada pixel em uma célula será comparada com cada um dos seus 8 pixels vizinhos seguindo ao longo de um círculo, ou seja, podendo ser no sentido horário ou anti-horário; se o valor do pixel central for maior que o valor do vizinho, escreverá 0, caso contrário 1. Isso fornecerá um número binário de 8 dígitos, em que é convertido em decimal; sobre a célula é calculado seu histograma, ou seja, calcula-se a frequência de cada combinação de quais pixels são menores e quais são maiores do que o centro. Portanto, este histograma pode ser visto como um vetor de característica 256-dimensional; concatenando os histogramas de todas as células, um vetor de característica na janela examinada será fornecida. Na posse do vetor de característica pode-se então processá-lo em algum algoritmo de ML para classificação, tais como SVM e Redes Neurais.

O LBP original não é invariável para rotação. Portanto, uma extensão foi proposta por Jin et al. (2004), conhecida como LBP aprimorado, para tornar sua rotação invariável executando operações de deslocamento bit a bit nos padrões binários e escolhendo o menor valor como saída (ZAMZMI et al., 2017). O LBP aprimorado é menos sensível ao ruído e um descritor que reduz o ruído da imagem comparando a intensidade dos pixels vizinhos com o valor médio local em vez do pixel central.

Ao invés de utilizar padrões binários, Tan e Triggs (2010) propuseram a utilização de padrões ternários, em que a diferença entre o pixel central X e seu vizinho P será representado por uma função de 3 valores, ou seja, valores ternários de 0, 1 e -1 em vez de binários como mostra a Equação (3). Esse outro método é uma extensão do LBP e é denominado de Padrão Ternário Local (LTP), tal que:

$$f(P,X,T) = \begin{cases} 1, & P \geq x + t \\ -1, & P \leq x - t \\ 0, & |X - P| < t \end{cases}, \tag{3}$$

em que T é um *threshold* definido pelo usuário. Trabalhos subsequentes propuseram variantes dos métodos LBP e LTP que fazem uso de células elípticas para analisar os pixels da vizinhança ao invés de uma célula circular, pois segundo Liao e Chung (2007) uma célula elíptica permite capturar a estrutura anisotrópica[2] de imagens faciais com mais eficiência. Essa afirmação da estrutura anisotrópica em imagens faciais é um ponto em que Hinton et al. (2012) enfatizam dizendo que "A visão computacional é a computação gráfica inversa; portanto, os níveis mais

[2]Na computação gráfica, a filtragem anisotrópica é um método para melhorar a qualidade das texturas das imagens que estão em ângulos de visão oblíquos em relação à câmera, visto que a projeção da textura não parece ortogonal.

altos de um sistema de visão devem se parecer com as representações usadas nos gráficos, traduzido", e complementam dizendo "Os programas gráficos usam modelos hierárquicos, no qual a estrutura espacial é modelada por matrizes que representam a transformação de um quadro de coordenadas incorporado no todo para um quadro de coordenadas incorporado em cada parte, traduzido". As variantes propostas são denominadas de Padrão Binário Alongado (ELBP) (LIAO; CHUNG, 2007) e Padrão Ternário Alongado (ELTP) (NANNI et al., 2010).

Em seu trabalho, Nanni et al. (2010) apresentaram métodos baseados em textura para detecção de dor em expressões faciais Os métodos utilizados foram LBP, LTP, ELBP e ELTP. Neste estudo foi utilizado a base de dados COPE. Mas antes da execução do descritor uma etapa de pré-processamento foi executada nas imagens, em que foram redimensionadas, alinhadas, recortadas para a obtenção da região facial e divididas em células de dimensões 25×25, para então, os métodos baseados em texturas serem aplicados a essas células e extraírem características relevantes (dor ou sem dor). Na determinação das células mais discriminantes o algoritmo SFFS foi aplicado no conjunto de treinamento e avaliado por meio da validação cruzada *leave-one-out*. Na etapa de classificação das imagens dos neonatos, foi usado um classificador SVM base radial e avaliado em um conjunto de teste. Os resultados mostraram que o descritor de textura ELTP alcançou a área mais alta (aprox. 0,93) da Área Sob a Curva ROC (AUC) em comparação com os outros descritores de textura. Também mostraram que a expressão da dor afeta sub-regiões da face e, assim, dividir toda a imagem em células pode melhorar o desempenho.

Lu et al. (2016) propuseram também um *framework* para detecção de dor neonatal em imagens de expressões faciais, através de variações de Padrões Binários Locais. O *framework* proposto inclui os passos de pré-processamento das imagens, extração de características, redução de dimensionalidade e classificação. Para o pré-processamento, foi realizada a normalização em escala de cinza, a qual converte imagens coloridas em imagens em tons de cinza e realça o contraste e o brilho de regiões individuas. Para melhoramento do contraste geral da imagem, foi adotado a equalização por histograma. No estágio de extração de características, descritores como Código de Gradiente Local (LGC), Padrão Direcional Local (LDP), Padrão de Textura Direcional Local (LDTP), além das variantes do LBP com diferentes tamanhos foram utilizados para extrair características de expressões faciais das imagens. Para a redução de dimensionalidade foi utilizado o PCA. O classificador baseado em representação esparsa foi utilizado para classificar quatro classes de expressões faciais: calmo, chorando, dor moderada e

dor severa. Os resultados mostraram que a melhor acurácia média, 85,5%, foi alcançada através da utilização do descritor Padrão Binário Local Ponderado (WLBP).

Uma limitação destes trabalhos citados é o uso de descritores de textura estáticos em arcabouços de classificação de expressões faciais da dor neonatal. Estes lidam apenas com informações espaciais e ignoram o padrão dinâmico das expressões faciais. Segundo Zamzmi et al. (2017), para mensurar as informações espaço-temporais, descritores dinâmicos de textura podem ser explorados. Entretanto, Silva (2020) afirma que "na verdade, existe uma limitação em utilizar imagens 2D estáticas para reconhecimento de expressões de dor. Imagens estáticas ignoram informações temporais e a dinâmica das expressões. Isso afeta a capacidade de compreender a expressão facial e sua evolução ao longo do tempo". O ponto não está na limitação de imagens 2D, mas sim na limitação dos modelos não trabalharem com hierarquias de partes. Hinton et al. (2012) definiram hierarquia de partes como: uma entidade visual de nível superior presente se várias entidades visuais de nível inferior puderem concordar com suas previsões para sua pose (relação com a câmera), assim, modelos hierárquicos por partes podem aprender a pose de uma entidade ao longo do tempo. E outra limitação destes trabalhos são as oclusões tais como auto-oclusão, máscara de oxigênio, ventilação invasiva e não invasiva, e outros, que são comuns em ambientes clínicos e seguem como artefatos restritivos para análises.

2.3.1.3 Métodos Baseados em Movimento

Métodos baseados em movimento podem ser definidos como métodos que estimam os vetores de movimento para um pixel (direto) ou características (indiretas) entre os consecutivos quadros de vídeos. O fluxo óptico é um método conhecido de estimativa de movimento, no qual estima diretamente a velocidade do pixel nos consecutivos quadros. Depende do princípio de conservação do brilho e fornece uma correspondência densa de pixel a pixel. Para mais detalhes sobre o algoritmo de fluxo óptico e sua implementação, sugere-se (CORREIA; CAMPILHO, 2002).

Zamzmi et al. (2015) introduziram um método baseado em movimento para detectar a expressão de dor dos recém-nascidos em vídeos. O conjunto de dados utilizado neste trabalho foi coletado de dez recém-nascidos, com idades entre 32 e 41 semanas de vida, internados na UTIN do Hospital Geral de Tampa na Florida, Estados Unidos. Os vídeos foram gravados com bebês que eram submetidos a um procedimento doloroso agudo de rotina (por exemplo, punção no calcanhar). Especificamente, os bebês foram registrados antes dos procedimentos dolorosos

em um estado normal para obter a linha de base, um estado sem dor. Em seguida, eles foram registrados durante os procedimentos dolorosos desde o início até o final do procedimento. Para obter os devidos rótulos, utilizou-se de profissionais de saúde que avaliaram a dor dos bebês e forneceram notas usando escala de dor neonatal NIPS, mostrada na Tabela 1.

No pré-processamento, Zamzmi et al. (2015) detectaram em cada quadro o rosto dos bebês, para então, extrair os 68 pontos faciais. Esses pontos foram usados para alinhar a face, cortá-la e dividi-la em quatro regiões. Para extrair características relevantes da dor os vetores de fluxo óptico foram calculados entre quadros consecutivos e usados para estimar as magnitudes das deformações ópticas. Em seguida, um detector de pico foi aplicado nas curvas de deformação para encontrar as magnitudes máximas de deformação que correspondem as expressões faciais. Para a classificação, as características de deformações extraídas foram usadas para treinar diferentes classificadores de ML: SVM e KNN[3]. Para avaliar o modelo treinado e estimar o desempenho da generalização, foi aplicado um protocolo de avaliação de validação cruzada com *k-fold* ($k = 10$). O KNN alcançou a maior precisão geral (96%) para classificar as expressões faciais dos bebês como dor ou sem dor.

Apesar da popularidade e eficiência do fluxo óptico na estimativa de movimento, a violação da restrição de suavidade e as auto-oclusões podem causar falha no fluxo óptico e fornecer cálculos imprecisos de fluxo. Outro fator que afeta os resultados do fluxo óptico são as descontinuidades de movimento e as variações de iluminação.

2.3.1.4 Métodos Baseados em Modelo

Algoritmos baseados em modelo têm o conceito da procura de parâmetros ideais de um modelo de objeto, no qual o melhor irá corresponder ao modelo e à imagem de entrada. O Modelo de Aparência Ativa (AAM) é um algoritmo baseado em modelos conhecidos que usam aparência, ou seja, combinação de textura e forma, para combinar um modelo de imagem com uma nova imagem. É um algoritmo que pode ser utilizado em várias aplicações, como reconhecimento de rosto (EDWARDS et al., 1998), reconhecimento de expressão facial (CHEON; KIM, 2009) e análise de imagens médicas (BEICHEL et al., 2005). Para ajustar a uma imagem, o erro entre o modelo representativo e a imagem de entrada deve ser minimizado, assim sendo um problema de otimização não linear.

[3]Um algoritmo simples, não paramétrico, que armazena todas as instâncias antecipadamente e classifica uma nova instância com base em uma medida de similaridade, geralmente uma função de distância.

Em seu trabalho, Fotiadou et al. (2014) discutiram o uso do AAM na detecção da expressão da dor dos bebês durante um procedimento doloroso agudo. O método apresentado por esses autores adota o método proposto por Lucey et al. (2010), em que analisa a expressão da dor em adultos. Para processar o método, os autores utilizaram um banco de dados contendo expressão facial de dez bebês internados na UTIN de um hospital localizado em Veldhoven, na Holanda. Os bebês foram registrados em quatro estados: punção do calcanhar, troca de fraldas, fome e repouso/sono. Todos os vídeos foram gravados em condição de luz ambiente.

Fotiadou et al. (2014) aplicaram o rastreador AAM em cada vídeo para obter pontos de referência faciais através dos quadros. Em seguida, três características foram extraídas da face rastreada com base nos parâmetros do AAM: a forma normalizada de similaridade (um vetor de característica que contém as coordenadas dos pontos de referência após remover todas as variações geométricas rígidas), a aparência normalizada de similaridade (um vetor que representa a aparência após a remoção de variações e escalas geométricas rígidas) e a aparência normalizada canônica (representação da aparência após remover toda a variação de forma não rígida). Um total de 15 vídeos de 8 crianças foi usado para construir o sistema automatizado de detecção de expressão da dor. Os autores excluíram das análises posteriores os vídeos de dois bebês, uma vez que incluíam oclusão severa causada por grande rotação da face ou movimento das mãos. O sistema proposto classificou a expressão facial dos bebês como desconforto ou conforto usando as características extraídas e um classificador SVM. Para avaliar o desempenho do classificador, realizou-se a validação cruzada *leave-one-subject-out*. O resultado de 0,98 da AUC mostrou que o sistema proposto pode detectar desconforto automaticamente.

Segundo Zamzmi et al. (2017), no trabalho proposto por Fotiadou et al. (2014) há três limitações. Primeiro, os estados emocionais de cada classe não foram claramente especificados. Não foi definido claramente se a classe de desconforto contém apenas a punção no calcanhar ou se contém a punção no calcanhar, além de troca de fraldas e fome, pois Zamzmi et al. (2017) acreditam que os dois últimos estados são diferentes da dor e devem ser tratados separadamente. Segundo, todas as experiências foram realizadas usando um AAM específico para cada individuo, construído especificamente para cada bebê, isso pode levar a problemas de escalabilidade na prática. Terceiro, o método proposto exige uma investigação mais aprofundada em um conjunto de dados maior, uma vez que foi avaliado em um pequeno conjunto de dados com 8 sujeitos. Além das três limitações em que Zamzmi et al. (2017) apontaram, há mais uma limitação que fez Fotiadou et al. (2014) excluírem os vídeos de dois bebês, em que havia oclusão severa causada por grande rotação da face ou movimento das mãos. Sendo que, o sistema

automático proposto por eles não é robusto à equivariância da rotação, impedindo que o sistema se adapte a grandes rotações.

2.3.1.5 Métodos Baseados em FACS

O Sistema de Codificação de Ação Facial (FACS) é um sistema abrangente que usa um conjunto de códigos numéricos para descrever os movimentos dos músculos faciais para todas as expressões faciais observáveis. Os códigos numéricos do FACS são conhecidos como unidades de ação (AUs). O sistema de Sistema de Codificação Facial Neonatal (NFCS) (GRUNAU; CRAIG, 1987) é uma extensão do FACS projetada especificamente para observar os movimentos faciais relevantes da dor dos bebês.

Segundo Zamzmi et al. (2017), a grande maioria dos métodos no campo do reconhecimento automático de expressões faciais usa o FACS para detectar as expressões faciais, e complementa dizendo que não há nenhum conhecimento da existência de método baseado em FACS projetado especificamente para detectar a expressão facial de dores neonatal, entretanto, há estudo que utiliza a extensão do FACS, o Sistema de Codificação Facial Neonatal (NFCS), para classificar a dor neonatal (HEIDERICH et al., 2015). Em seu trabalho, Heiderich et al. (2015) desenvolveram um software para analisar imagens faciais de recém-nascidos em tempo real para monitorar os seus estados (repouso ou dor). Para validar o desempenho do software, foram obtidas imagens faciais durante o monitoramento de 30 neonatos submetidos a procedimentos dolorosos relacionados ao manejo diário. Os autores também utilizaram profissionais de saúde com experiência no reconhecimento da dor neonatal para avaliar o desempenho do software, em um total de seis profissionais. O trabalho obteve um concordância entre os examinadores e a avaliação do software de 97,5%.

Em uma população distinta de neonatos, Sikka et al. (2015) apresentaram um método baseado em FACS para descrever as expressões faciais de dor de crianças. O método proposto foi aplicado a sequências de vídeo de 50 crianças gravadas durante condições de dor contínua e transitória[4], tal que um total de 35 crianças eram hispânicas, 9 não-hispânicas brancas, 5 asiáticas e 1 nativa americana. A idade dos indivíduos variavam de 5 a 18 anos, sendo que 35% dos indivíduos eram meninos. Os dados foram coletados em três visitas. Primeira, nas 24 horas seguintes à cirurgia de apendicectomia. Segundo, um dia após a primeira visita. Terceiro em uma visita de acompanhamento. Em cada visita, as expressões faciais dos indivíduos foram

[4]A dor transitória foi desencadeada pressionando manualmente o local cirúrgico por 2 a 10 segundos.

gravadas usando uma câmera de vídeo Canon VIXIA-HF-G10 colocada na posição vertical. Juntamente com a gravação, foram coletadas classificações autorreferidas pelos indivíduos mediante a autorização por procuração de ambos os pais e uma enfermeira para obter os rótulos *ground truth*.

Sikka et al. (2015) utilizaram ferramentas computacionais de reconhecimento de expressão para extrair características úteis dos vídeos gravados, com essas ferramentas foram capazes de detectar várias AUs. Um método de seleção de característica foi então aplicado para selecionar 14 AUs e diferentes estatísticas foram calculadas para cada uma dessas AUs, para então formar os vetores de características. Com as características extraídas, foi construído um modelo de regressão logística com avaliação de validação cruzada $k = 10$. A classificação binária da expressão facial como dor ou não-dor alcançou uma precisão de 0,84 e 0,94 AUC para dor contínua e transitória. Zamzmi et al. (2017) fizeram um contraponto na limitação do trabalho de Sikka et al. (2015) em que há condições restritas de luz e movimento, uma vez que o algoritmo apresentado requer iluminação e movimentos moderados o que pode ser difícil de ser realizado em contextos clínicos especialmente no caso de bebês na UTIN, já que são considerados uma população não cooperativa.

Outra deficiência dos métodos baseados em FACS está no extenso tempo necessário para rotular as AUs em cada quadro do vídeo, para obter o *ground truth*. Littlewort et al. (2009) relataram que um especialista precisa de cerca de três horas para rotular um minuto de uma sequência de vídeo, sendo que uma maneira de reduzir o custo da rotulagem é detectar automaticamente AUs em cada quadro e usá-las como rótulos.

2.3.1.6 Métodos de Aprendizado Profundo

Os métodos apresentados nas seções anteriores fazem o uso de extrações de características tradicionais para classificação de imagens. Extrações de características profundas provenientes de CNNs mostraram bom desempenho em várias tarefas de classificação desse tipo de dados. A principal diferença entre características tradicionais e profundas é que as características extraídas pela CNN são aprendidas em vários níveis de abstrações diretamente dos dados, em contraste com as características tradicionais projetadas para extrair um determinado conjunto de características escolhidas.

2.3.1.6.1 Transferência de Aprendizado

Zamzmi et al. (2018b) contribuíram com um novo pipeline para o reconhecimento da expressão da dor neonatal utilizando DL, especialmente Aprendizado por Transferência (TL). Em um dos seus tralhados, Zamzmi et al. (2018b), propuseram o uso de quatro arquiteturas de CNN pré-treinadas, sendo: VGG-F, VGG-M, VGG-S e VGG-Face. Essas redes são ilustradas nas Tabelas 6 a 9, respectivamente, e mostraram que tais CNNs pré-treinadas podem ser usadas para extrair informações úteis de características para classificação da expressão da dor. As arquiteturas VGG-F, M e S foram originalmente treinadas no conjunto de dados ImageNet, contendo aproximadamente 1,2 milhões de imagens e 1000 classes. Já o VGG-Face foi treinado em um grande conjunto de dados faciais com aproximadamente 2,6 milhões de imagens de rosto de 2622 indivíduos.

Tabela 6 – Arquitetura VGG-F (CHATFIELD et al., 2014); $k \times n \times n$ indica o número de filtros e seu tamanho; st. e pad indica o passo da convolução e o *padding*. Cada camada exceto a Full 8 é seguida pela ReLU.

Conv 1	$64 \times 11 \times 11$, st. 4, pad 0
Conv 2	$256 \times 5 \times 5$, st. 1, pad 2
Conv 3	$256 \times 5 \times 5$, st. 1, pad 1
Conv 4	$256 \times 5 \times 5$, st. 1, pad 1
Conv 5	$256 \times 5 \times 5$, st. 1, pad 1
Full 6	4096 dropout
Full 7	4096 dropout
Full 8	1000 softmax

Fonte: Zamzmi et al., 2018b

Tabela 7 – Arquitetura VGG-M (CHATFIELD et al., 2014); $k \times n \times n$ indica o número de filtros e seu tamanho; st. e pad indica o passo da convolução e o *padding*. Cada camada exceto a Full 8 é seguida pela ReLU.

Conv 1	$96 \times 7 \times 7$, st. 2, pad 0
Conv 2	$256 \times 5 \times 5$, st. 2, pad 1
Conv 3	$512 \times 5 \times 5$, st. 1, pad 1
Conv 4	$512 \times 5 \times 5$, st. 1, pad 1
Conv 5	$512 \times 5 \times 5$, st. 1, pad 1
Full 6	4096 dropout
Full 7	4096 dropout
Full 8	1000 softmax

Fonte: Zamzmi et al., 2018b

Tabela 8 – Arquitetura VGG-S (CHATFIELD et al., 2014); $k \times n \times n$ indica o número de filtros e seu tamanho; st. e pad indica o passo da convolução e o *padding*. Cada camada exceto a Full 8 é seguida pela ReLU.

Conv 1	$96 \times 7 \times 7$, st. 2, pad 0
Conv 2	$256 \times 5 \times 5$, st. 1, pad 1
Conv 3	$512 \times 5 \times 5$, st. 1, pad 1
Conv 4	$512 \times 5 \times 5$, st. 1, pad 1
Conv 5	$512 \times 5 \times 5$, st. 1, pad 1
Full 6	4096 dropout
Full 7	4096 dropout
Full 8	1000 softmax

Fonte: Zamzmi et al., 2018b

Tabela 9 – Arquitetura VGG-Face (DING et al., 2017), st. e pad indicam o passo da convolução e o *padding*. Cada camada (por exemplo, Conv 1-1) seguida por ReLU e cada bloco (por exemplo, Conv 1-1 e Conv 1-2) seguidos de pool.

Conv 1-1	$64 \times 3 \times 3$, st. 1, pad 1
Conv 1-2	$64 \times 3 \times 3$, st. 1, pad 1
Conv 2-1	$128 \times 3 \times 3$, st. 1, pad 1
Conv 2-2	$128 \times 3 \times 3$, st. 1, pad 1
Conv 3-1	$256 \times 3 \times 3$, st. 1, pad 1
Conv 3-2	$256 \times 3 \times 3$, st. 1, pad 1
Conv 3-3	$256 \times 3 \times 3$, st. 1, pad 1
Conv 4-1	$512 \times 3 \times 3$, st. 1, pad 1
Conv 4-2	$512 \times 3 \times 3$, st. 1, pad 1
Conv 4-3	$512 \times 3 \times 3$, st. 1, pad 1
Conv 5-1	$512 \times 3 \times 3$, st. 1, pad 1
Conv 5-2	$512 \times 3 \times 3$, st. 1, pad 1
Conv 5-3	$512 \times 3 \times 3$, st. 1, pad 1
Full 6	4096 dropout
Full 7	4096 dropout
Full 8	2622

Fonte: Zamzmi et al., 2018b

Na fase de pré-processamento das imagens, Zamzmi et al. (2018b) aplicaram o ZFace (JENI et al., 2015), um detector facial, em cada vídeo para detectar o rosto e obter 49 pontos de referências faciais. Para cada quadro de vídeo foi usado os pontos de detecção para registrar e cortar a região exata da face dos bebês. Aplicaram esse detector em 200 vídeos para assim detectar o rosto e os pontos de referência. Em seguida, selecionaram os quadros-chave de cada vídeo, removendo os semelhantes. Os quadros selecionados de todos os vídeos (ou seja, 3026 quadros) foram redimensionados para 224×224, conforme os requisitos de tamanho da imagem

impostas pelas CNNs, $244 \times 224 \times 3$ imagens RGB. Todo esse pipeline foi construído com uma base de dados desenvolvida pela equipe de neonatologia do Hospital Geral de Tampa.

Zamzmi et al. (2018b) em seus experimentos abordaram a extração de características profundas nas camada superior e inferior. A abstração de características nas camadas superiores obtiveram um resultado de 0.841 na AUC (90.34% de acurácia) superior a abstração nas camadas inferiores que ficou em 0.797 na AUC (88.23% de acurácia), ressaltando a obtenção desses resultados pela VGG-Face. Os autores também afirmaram que a VGG-Face foi superior que as VGG-F, M e S tanto nas extrações de características na camada superior e inferior, devido ao conjunto de dados em que as topologias foram treinadas, pois a VGG-Face fez uso de banco de dados com imagens faciais em seu treinamento e as demais VGGs utilizaram o ImageNet.

Em trabalhos subsequentes, Zamzmi et al. (2019) utilizaram o TL na arquitetura ResNet50 (HE et al., 2016) para também reconhecer a expressão da dor neonatal. A Tabela 10 mostra as novas camadas na ResNet que suportaram o TL. O pré-processamento das imagens foi o mesmos realizado em (ZAMZMI et al., 2018b), e tendo o acréscimo da técnica *Data Augmentation* no conjunto de dados de treinamento. O pipeline do trabalho foi construído através do banco de dados NPAD (ZAMZMI et al., 2018b), utilizado para o treinamento, validação e teste, e COPE (BRAHNAM et al., 2006), somente utilizado para teste.

Tabela 10 – Arquitetura ResNet50 modificada para o Aprendizado por Transferência.

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

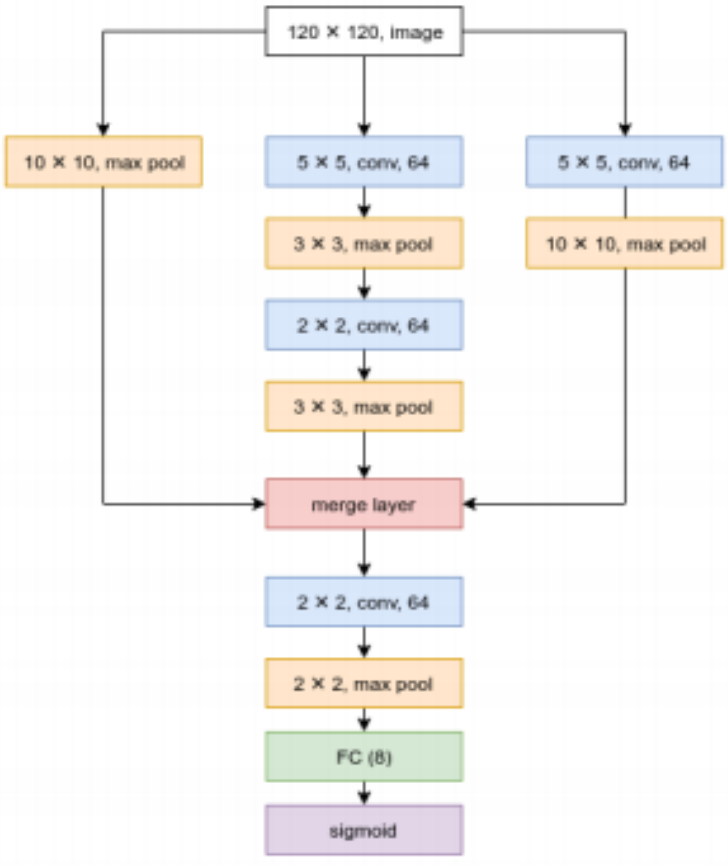
Fonte: Zamzmi et al., 2019

Os resultados demostraram 87.1% de acurácia no banco de dados NPAD e 82.9% no COPE. É importante ressaltar que, o trabalho (ZAMZMI et al., 2019) não fez uso somente da ResNet50, mas de uma proposta que foi introduzir uma nova topologia de CNN na questão do reconhecimento da expressão da dor neonatal, denominada Neonatal Convolutional Neural Network (N-CNN).

2.3.1.6.2 Neonatal Convolutional Neural Network

Zamzmi et al. (2019) apresentaram uma nova topologia de Rede Neural Convolucional, denominada Neonatal Convolutional Neural Network (N-CNN). Esta rede foi projetada e treinada do zero (end-to-end) para reconhecer a dor neonatal. Os autores afirmam não ter conhecimento de algum trabalho que propôs uma topologia própria para o reconhecimento da dor neonatal, sendo assim, complementam dizendo ser a primeira CNN projetada e construída para reconhecer a emoção da dor neonatal. A N-CNN proposta mostrou ser uma solução aos métodos implementados até então, ou seja, métodos que fazem uso de extração de características tradicionais, uma vez que tais métodos necessitam que os dados estejam normalizados e alinhados. O trabalho foi desenvolvido por meio do banco de dados NPAD (ZAMZMI et al., 2018b), que contém 200 vídeos para um total de 31 indivíduos, e COPE (BRAHNAM et al., 2006) que contém 288 fotos coloridas de 26 recém-nascidos. O NPAD foi utilizado para treinamento, validação e teste, enquanto que a base COPE foi utilizada somente para teste.

Figura 3 – Topologia da Neonatal Convolutional Neural Network (N-CNN).



Fonte: Zamzmi et al., 2019

O sistema proposto é constituído por dois estágios: detecção e pré-processamento da face, e reconhecimento da dor usando a topologia N-CNN. Na primeira etapa do sistema, os autores aplicaram a ZFace (JENI et al., 2015), que é um detector facial, em cada quadro de vídeo para abstrair o rosto e obter 49 pontos de referências faciais. Detectado a face em cada quadro, utilizaram os pontos de referências para registrar e cortar a região exata da face da criança e quadros muito semelhantes foram removidos, resultando em um número total de 3026 quadros RGB. Os 3026 quadros foram redimensionados para 120×120, utilizando o método de interpolação bi-cúbica. Na segunda etapa, visto que ocorre a fase de treinamento, os autores utilizaram-se do método *Data Augmentation*, sendo assim, gerando um aumento de 36 vezes o conjunto de treinamento. A topologia da N-CNN é apresentada na Figura 3. Nesse mesmo estudo os autores também utilizaram a ResNet50 (HE et al., 2016) com TL, como apresentado na Tabela 10, e analisaram a N-CNN contra a ResNet50. A ResNet50 foi treinada, validada e testada nas mesmas condições que a N-CNN.

Zamzmi et al. (2019) demostraram em seus experimentos que a topologia proposta por eles para o reconhecimento da expressão facial da dor neonatal foi melhor que o estado da arte ResNet50 (uma variação da ResNet). Seus resultados demostraram 82.9% de acurácia da ResNet50 contra 84.5% da N-CNN para base COPE. Em sua própria base NPAD, a ResNet50 obteve 87.1% de acurácia contra 91.0% da N-CNN.

3 MATERIAIS E MÉTODOS

Este capítulo descreve os materiais, os detalhes da metodologia e as métricas de avaliações topológicas propostas nesta dissertação. A metodologia é composta por: detecção facial (implementação não autoral), extração de características (implementação autoral) e classificação (implementação autoral).

Para a detecção facial das imagens dos recém-nascidos utilizou-se a *RetinaFace* (DENG et al., 2019) já treinada, código disponível publicamente, podendo ser considerado o algoritmo do estado da arte em relação à detecção facial. Na fase de extrações de características, esta dissertação fez o uso de camadas convolucionais junto com função *pooling* e funções de ativação (tais como, ReLU e Leaky ReLU) (KHAN et al., 2019). Na classificação, sequências de camadas totalmente conectadas com *Dropout* e funções de ativação (tais como, ReLU, Sigmoid e Softmax) foram utilizadas para classificar o conjunto de dados em dois estados “Dor” ou “Sem Dor”.

As métricas de avaliações topológicas foram: função de erro; acurácia; matriz de confusão. Todas essas métricas foram avaliadas com *k-fold* igual a 3. Além disso, utilizou-se o método de visualização nas CNN denominado Mapeamento de Ativação de Classe Ponderada por Gradiente (Grad-CAM) (SELVARAJU et al., 2017).

3.1 HARDWARE E SOFTWARE

Todo o arcabouço desta dissertação foi realizado em um servidor próprio da FEI. As especificações do servidor são:

a) Sistema Operacional Linux, arquitetura x86_64
b) CPU, Intel(R) Xeon(R) Gold 5118 2.30GHz
c) GPU, 2x Nvidia V100 16Gb
d) RAM, 188Gb

Para a construção do arcabouço também foi utilizada a linguagem de programação Python3 e toda a metodologia foi modelada no open source AI framework para ML e computação numérica de alta performance da Google. Esse framework é denominado TensorFlow.

3.2 BANCOS DE DADOS

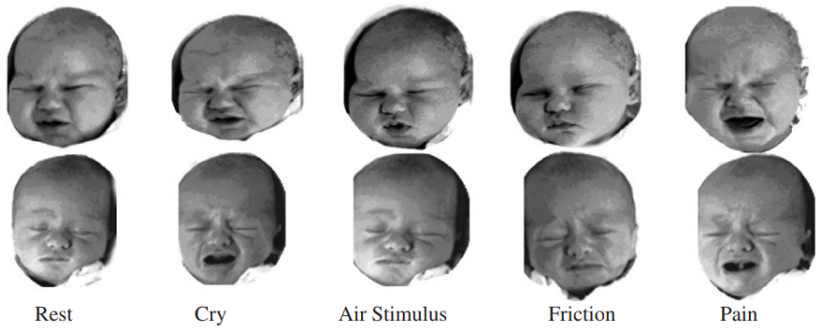
Bancos de imagens para proporcionar estudos utilizando as tecnologias de CV, ML e principalmente DL para analisar e avaliar a dor neonatal ainda são poucos. Esta seção descreve dois bancos neonatais investigados aqui: Classification of Pain Expressions (COPE) (BRAHNAM et al., 2006), um dos primeiros bancos de imagens criados para analisar e avaliar a dor neonatal, e o banco de imagens neonatal da UNIFESP (HEIDERICH et al., 2015).

3.2.1 COPE

Um dos primeiros estudos no reconhecimento automático da dor foi proposto por BRAHNAM et al. em 2006. Para o desenvolvimento do COPE, Brahnam et al. (2006) escolheram estímulos que provocam expressões faciais, com o objetivo de serem utilizados nas avaliações e análises da dor em neonatos.

Os estímulos introduzidos pelos autores são: (1) punção do calcanhar (teste do pezinho), (2) fricção na superfície lateral externa do calcanhar, (3) transporte de um berço para outro e (4) estímulo aéreo. A introdução do estímulo aéreo no nariz teve a intenção de provocar aperto nos olhos, simulando mudanças de iluminação (BRAHNAM et al., 2006).

Figura 4 – Exemplos das cinco expressões faciais da base de dados COPE.



Fonte: Brahnam et al., 2006

O trabalho para o desenvolvimento do COPE cumpriu os protocolos e diretrizes éticas para pesquisas envolvendo seres humanos no St. John's Health System, Inc. Os pais foram recrutados na unidade neonatal do hospital de St. John. Foram tiradas 288 fotos coloridas com 3008×2000 pixels de 26 recém-nascidos caucasianos, 13 meninos e 13 meninas, com idades entre 18h e 3 dias. Uma câmera digital Nikon D100 foi utilizada para tirar as fotografias sob

condições de iluminação ambiente em uma sala separada de outros recém-nascidos. A Figura 4 ilustra cinco exemplos de expressão facial contidas no COPE.

3.2.2 UNIFESP

No trabalho desenvolvido por Heiderich et al. (2015), foi construído um banco de imagens neonatal para a elaboração de um software capaz de identificar automaticamente expressões de dor neonatal. O *back-end* do software utilizou a distância entre pontos específicos identificados no rosto do recém-nascido e a escala unidimensional NFCS (GRUNAU; CRAIG, 1987), 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 aplicados a essa população. Tais procedimentos considerados foram: punção venosa, capilar ou injeção intramuscular (procedimentos comuns e necessários), também foi considerado a fragilidade nesses indivíduos pois são procedimentos invasivos, dolorosos e/ou estressantes. A captura dessas imagens foram feitas a cada 3 segundos (HEIDERICH et al., 2015).

A partir do conjunto de imagens de cada recém-nascido, foram selecionadas 360 imagens coloridas com 450×233 pixels de 30 recém-nascidos entre 34 e 41 semanas de idade gestacional e entre 24 e 168 horas de vida (prematuros tardios ou a termo), alimentados e saudáveis (sem más-formações congênitas, anomalias faciais, necessidade de suporte ventilatório, sonda gástrica ou injeções intramusculares e/ou subcutâneas) (HEIDERICH et al., 2015). Esses recém-nascidos estavam internados no Hospital São Paulo (Hospital Universitário da Escola Paulista de Medicina da Universidade Federal de São Paulo). Todos os recém-nascidos selecionados participaram da pesquisa com o consentimento dos familiares ou responsáveis (HEIDERICH et al., 2015) com a permissão de uso das imagens pelo comitê de ética, sob o número de rastreamento *CCAE: 66743417.0.0000.5505* e o número de parecer *2.035.113*. Não houve restrição de seleção dos recém-nascidos quanto a gênero, raça ou cor. Após a seleção das imagens, um relatório e com disposição aleatória dessas imagens foram então submetidas à avaliação de profissionais da área de saúde com especialização em neonatologia.

A avaliação feita pelos profissionais foi baseada na escala NFCS, identificando em cada imagem: fronte saliente; fenda palpebral; sulco nasolabial; boca aberta e boca estirada. A cada três ou mais características positivas a imagem do recém-nascido era então classificada com dor, caso contrário sem dor. Essa avaliação agregou valor de informação nas imagens, classificando

164 imagens com dor e 196 sem dor, sendo essa classificação humana tomada como referência e usada na construção do algoritmo sequencial de procedimentos computacionais para determinação da existência de dor, de maneira automática. No relatório, as imagens classificadas com desconforto foram consideradas no estado de dor, devido a características de comparação com a escala unidimensional NFCS. A Figura 5 ilustra algumas imagens da base de dados UNIFESP.

Figura 5 – Exemplos da base de dados UNIFESP.

Fonte: Silva, 2020

3.3 METODOLOGIA

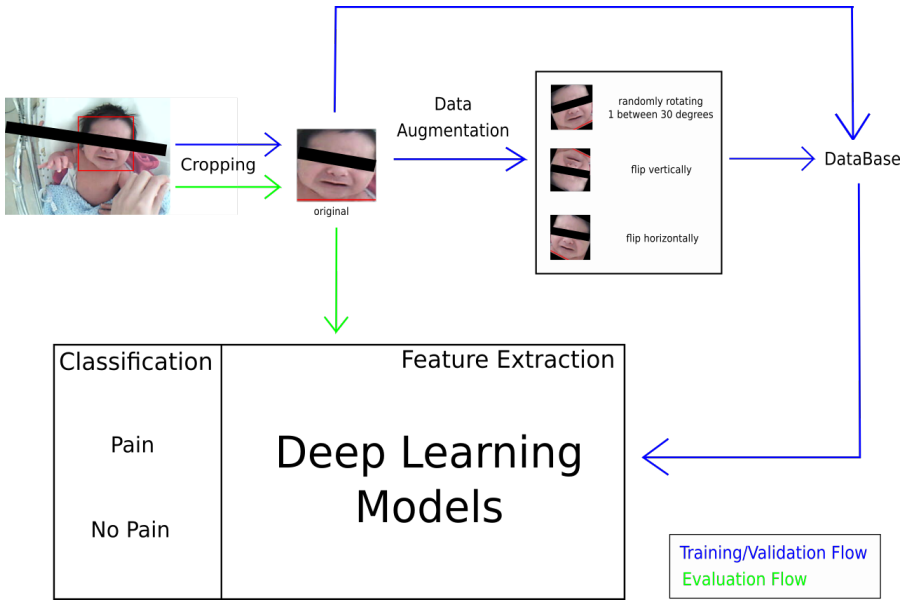
Esta seção descreve o arcabouço desta dissertação que consiste em: detecção facial, aumento de dados e reconhecimento da dor usando topologias de DL, tais como N-CNN (ZAMZMI et al., 2019) e ResNet50 (HE et al., 2016). A Figura 6 ilustra o arcabouço desta dissertação nas etapas de treinamento e avaliação.

3.3.1 Detecção Facial e Aumento de Dados

3.3.1.1 RetinaFace

A localização automática da face é uma etapa de pré-processamento da análise de imagens faciais para muitas aplicações, tais como atributo facial (ZHANG et al., 2018) e reconhecimento de identidade facial (DENG et al., 2019). Uma definição estreita de localização de face pode se referir à detecção de face tradicional (VIOLA; JONES, 2004), que visa estimar as

Figura 6 – Arcabouço computacional.



Fonte: Autor

caixas delimitadoras de face sem nenhuma escala e posição anterior. Esta dissertação fez uso do algoritmo RetinaFace (DENG et al., 2019), 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.

Com a RetinaFace (DENG et al., 2019) os autores aprimoraram a estrutura de detecção facial de estágio único e propuseram um método de localização de face densa sendo o estado da arte, explorando as perdas de várias tarefas provenientes de sinais fortemente supervisionados e auto-supervisionados (DENG et al., 2019). A Figura 7 mostra a estrutura na RetinaFace proposta.

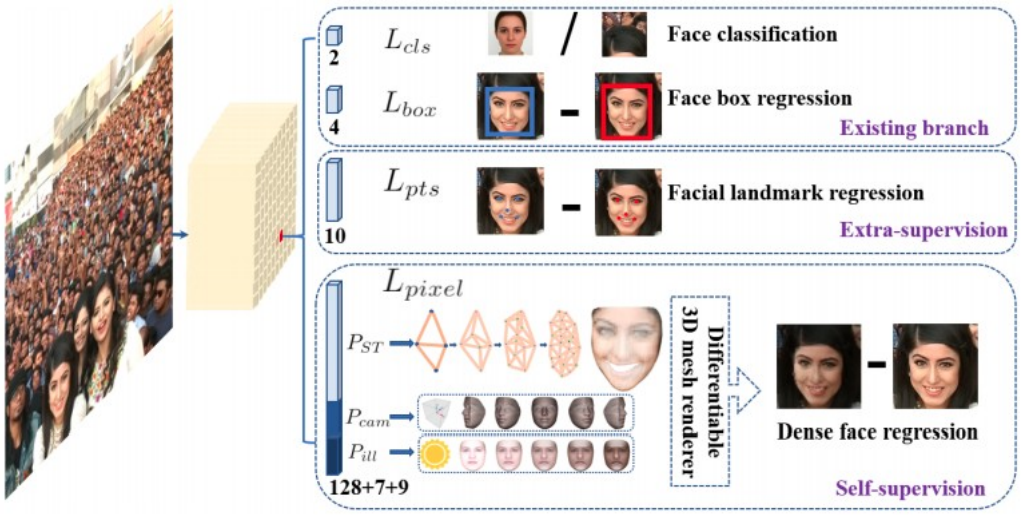
No treinamento da RetinaFace os autores utilizaram o erro de multi-tarefas (multi-task loss), no intuito de minimizar o erro da caixa de âncora [1] i. Tal erro é definido sendo:

$$L = L_{cls}(p_i,p_i^*) + \lambda_1 p_i^* L_{box}(t_i,t_i^*) + \lambda_2 p_i^* L_{pts}(l_i,l_i^*) + \lambda_3 p_i^* L_{pixel}. \quad (4)$$

O termo $L_{cls}(p_i,p_i^*)$ é o 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

[1] As caixas de âncoras (anchor boxes) são informações da face detectada, sendo a probabilidade de ser uma face e os pontos de coordenadas para a localização da mesma.

Figura 7 – O RetinaFace emprega o aprendizado multitarefa extra-supervisionado e auto-supervisionado em paralelo com os ramos de classificação e regressão existente da caixa. Cada âncora positiva produz: uma pontuação de face; uma caixa facial; cinco pontos de referências faciais; densos vértices de face 3D projetados no plano da imagem.



Fonte: Deng et al., 2019

binárias (face/não face). $L_{box}(t_i,t_i^*)$ erro de regressão da caixo 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 $l_i = \{l_{x1},l_{y1},\ldots,l_{x5},l_{y5}\}_i$ e $l_i^* = \{l_{x1}^*,l_{y1}^*,\ldots,l_{x5}^*,l_{y5}^*\}_i$ 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 (5). Os parâmetros de balanceamento dos erros λ_1, λ_2 e λ_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 (DENG et al., 2019).

$$L_{pixel} = \frac{1}{W * H} \sum_{i}^{W} \sum_{j}^{H} \parallel \mathcal{R}(\mathcal{D}_{P_{ST}},P_{cam},P_{ill})_{i,j} - I_{i,j}^* \parallel_1, \tag{5}$$

após a convolução gráfica, vide (DENG et al., 2019), os autores computaram os parâmetros de forma e textura $P_{ST} \in \mathbb{R}^{128}$ 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 5, W e H são a largura e altura de i do corte da face $I^*_{i,j}$, respectivamente.

Portanto esta dissertação fez uso do detector facial RetinaFace[2] disponível publicamente, já treinado. Em cada imagem do banco de dados o algoritmo retornou as coordenadas de cada face contida na imagem[3] e 5 pontos faciais (landmarks). O detector envia as coordenadas desses pontos e uma mensagem de falha para indicar as detecções que falharam[4]. Para cada imagem foram utilizadas as coordenadas 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 conforme utilizado em (ZAMZMI et al., 2019).

3.3.1.2 Data Augmentation

Ter um grande conjunto de dados é importante para o desempenho de um modelo de DL. Uma alternativa de melhorar o desempenho do modelo é aumentando os dados (Data Augmentation), uma técnica trazida para área por um grupo da Universidade de Toronto (KRIZHEVSKY et al., 2012). As estruturas de DL geralmente têm utilizado aumento de dados, que pode impactar computacionalmente. Entretanto, esta dissertação contorna tal problema pois os modelos foram computados em um servidor dedicado de alta performance e também houve uma obrigatoriedade do uso de tal técnica, devido ao número de imagens na base de dados não ser suficiente.

São muitas maneiras possíveis para aumentar a base de imagens, mas os métodos mais comuns são aplicar combinações de operações sobre a imagem original, tais como: translação; rotação; modificação da perspectiva; achatamento e alongamento; distorção de lentes. Esta dissertação fez o uso do mesmo aumento de dados do trabalho (ZAMZMI et al., 2019), 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, foi dividido randomicamente em 50% gerando o conjunto I_{test} e I^*. No conjunto de dados I^* aplicou-se o aumento de dados, sendo:

[2]https://github.com/deepinsight/insightface/tree/master/RetinaFace

[3]As imagens no banco só contem apenas uma face, sendo essa face de um bebê.

[4]Não ocorreu falha nos bancos de imagens.

a) Cada imagem foi randomicamente rotacionada até $30°$ ($1°$ a $30°$), para gerar um total de 12 imagens para cada imagem

b) 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 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}.

Segundo Goodfellow et al. (2016), aumentar o tamanho do conjunto de treinamento, adicionando cópias extras dos exemplos de treinamento que foram modificados com transformações que não alteram a classe, melhora a generalização de um classificador. O reconhecimento de objetos é uma tarefa de classificação especialmente adaptável a essa forma de aumento do conjunto de dados, porque a classe é invariável a tantas transformações e a entrada pode ser facilmente transformada com muitas operações geométricas, assim, todo o aumento feito e imposto ao modelo tem intenção de reduzir o erro de generalização (mas não o erro de treinamento). Portanto o Data Augmentation é uma técnica de regularização (GOODFELLOW et al., 2016). Logo, foi aplicado o aumento de dados no conjunto de validação para comprovar a robustez do modelo a essas operações geométricas.

Tendo como objetivo do treinamento da rede encontrar o melhor desempenho em novos dados, Bishop et al. (1995) afirmam que a abordagem mais simples é avaliar a função de erro usando dados independentes dos utilizados para o treinamento. Portanto, a rede foi treinada pela minimização de uma função de erro apropriada definida em relação a um conjunto de dados de treinamento. Bishop et al. (1995) complementam dizendo que: O desempenho da rede deve ser então comparado através da avaliação da função de erro usando um conjunto de validação independente, e a rede com o menor erro e/ou maior acurácia em relação ao conjunto de validação deve ser selecionada. Porém, como esse procedimento afirmado por Bishop et al. (1995) pode, por si só, levar a algum *overfitting* no conjunto de validação, o desempenho da rede selecionada é então confirmado medindo seu desempenho em um terceiro conjunto de dados independente chamado de conjunto de teste.

3.3.2 Reconhecimento da dor

3.3.2.1 ResNet

A capacidade de aprendizado das CNNs se deve em grande parte ao uso de vários estágios de extração de característica, que podem aprender automaticamente as representações dos dados. De fato, várias ideias interessantes para trazer avanços para CNNs foram exploradas, como o uso de diferentes funções de ativação e perda, otimização de parâmetros, regularização e inovações arquitetônicas. Mediante a esses avanços exploratórios, esta dissertação faz uso da ResNet (HE et al., 2016), uma topologia inovadora introduzida em 2015 e vencedora do 2015-ILSVRC (Large Scale Visual Recognition Challenge).

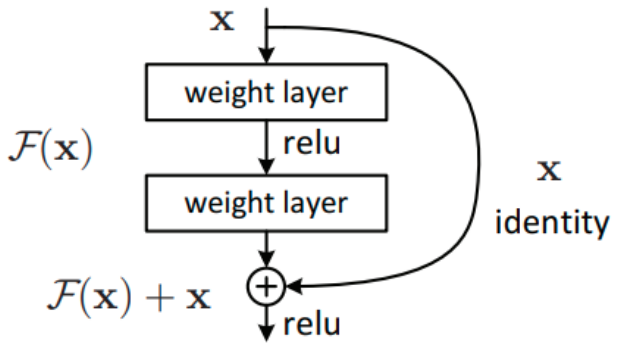
A ResNet proposta por He et al. (2016) revolucionou a corrida arquitetônica da CNN ao introduzir o conceito de aprendizado residual e o desenvolvendo por meio de uma metodologia eficiente para o treinamento de redes profundas. Semelhante às Highway Networks (SRIVASTAVA et al., 2015) a ResNet também é colocada na categoria das CNNs com vários caminhos. A ResNet proposta por He et al. (2016) com 152 camadas de profundidade foi a topologia que venceu a competição 2015-ILSVRC. A ResNet, em que é 20 e 8 vezes mais profunda do que a AlexNet (KRIZHEVSKY et al., 2012) e a VGG (SIMONYAN; ZISSERMAN, 2014), mostrou menos complexidade computacional do que as redes propostas anteriormente, comprovando aquilo que Bengio et al. (2013) havia mostrado empiricamente em 2013, ou seja, redes profundas são computacionalmente mais eficientes para tarefas complexas. He et al. (2016) também demostraram empiricamente que as ResNets com 50/101/152 camadas apresentam menos erro na tarefa de classificação de imagens do que uma rede comum de 34 camadas. Além disso, a ResNet obteve uma melhoria no erro de 28% no famoso conjunto de dados de referência de reconhecimento de imagem denominado COCO (LIN et al., 2014). O bom desempenho da ResNet nas tarefas de reconhecimento e localização de imagens mostrou que a profundidade representacional é de importância para muitas tarefas de reconhecimento visual.

Os autores adotaram na ResNet o aprendizado residual em todas as camadas empilhadas. Tal aprendizado é mostrado na Figura 8. A formulação matemática do bloco foi definida sendo:

$$y = \mathcal{F}(x,W_i) + x, \tag{6}$$

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. Como na Figura 8 que possui duas camadas,

Figura 8 – Aprendizagem residual. Bloco de construção residual.



Fonte: He et al., 2016

$\mathcal{F} = W_2\sigma(W_1x)$, na qual σ 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. Os autores também consideraram uma segunda não linearidade após a adição, ou seja, $\sigma(y)$.

Embora a ResNet152 tenha sido a vencedora do 2015-ILSVRC, esta dissertação utilizou-se da ResNet50 para comparar com o trabalho (ZAMZMI et al., 2019). Zamzmi et al. (2019) em seu trabalho aplicaram o TL na ResNet50, assim removendo a última camada que define as classes da topologia e adicionando 1 único neurônio com função sigmoid, em que 1 indica o estado dor e 0 o estado sem dor. A topologia adotada por Zamzmi et al. (2019) foi mostrada anteriormente na Tabela 10. Esta dissertação, implementou também a ResNet50, denominada ResNet50(ours), com a aplicação do TL, entretanto, ao invés de utilizar 1 único neurônio utilizou-se 2 neurônios com a função sotfmax, em que [0 1] define o estado dor e [1 0] o estado sem dor. A Tabela 11 mostra a topologia implementada por esta dissertação.

Tabela 11 – Arquitetura ResNet50(ours) modificada para o Aprendizado por Transferência.

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

Fonte: Autor

O problema de utilizar 1 único neurônio é a distribuição dos pesos na camada de classificação para descriminar dois estados, uma vez que esses pesos terão uma combinação linear para dois estados. Utilizando 2 neurônios, no qual cada neurônio representa um estado, os pesos terão uma combinação linear para cada estado, em outras palavras, cada estado terá uma equação própria para classificação, sendo representada como:

$$y^c = \sum_k w_k^c \underbrace{\frac{1}{Z}\sum_i \sum_j}_{\text{global average pooling}} A_{ij}^k, \tag{7}$$

onde c é a classe, $y^c \in Y_{c\times 1}$ são os *logits scores* (antes na função de ativação) e A^k são os mapas de ativação de características da última camada convolucional, e k representa o número dos mapas ($\therefore A^k \in \mathbb{R}^{i\times j}$, i e j representam as dimensões largura e altura de cada mapa). A equação de classificação com 1 único neurônio é:

$$y = \sum_k w_k \underbrace{\frac{1}{Z}\sum_i \sum_j}_{\text{global average pooling}} A_{ij}^k, \tag{8}$$

y é o *logit score* (antes na função de ativação). Um outro problema de utilizar 1 único neurônio com a função sigmoid será a saturação da função, pois a rede em seu aprendizado forçará essa ativação trabalhar nos extremos e, após o treinamento, um threshold deverá ser definido, para delimitar o que serão os estados (exemplo, $threshold \geq 0.5$ será o estado *Dor* e $threshold < 0.5$ será o estado *Sem Dor*). Ao utilizar 2 neurônios com softmax não precisará de um threshold, uma vez que os 2 neurônios representarão uma matriz $Y_{2\times 1}$, que tem seu $|Y| = 2$, logo a posição y_{11} representa a probabilidade do estado *Sem Dor* e a posição y_{21} representa a probabilidade do estado *Dor*.

As Equações (9) e (10) representam as equações de erro dos modelos ResNet50 proposta por Zamzmi et al. (2019) e da ResNet50(ours) proposta por esta dissertação, respectivamente.

$$\mathcal{L}(S,t) = -\frac{1}{m}\sum_i^m t_i ln(S_i) + (1-t_i) ln(1-S_i), \tag{9}$$

$$\mathcal{L}(S,t) = -\frac{1}{m}\sum_i^n \sum_j^m t_{ij} \log(S_{ij}), \tag{10}$$

no qual para a Equação (9) S representa a função sigmoid ($\therefore S = S(y)$) e para a Equação (10) S representa a função softmax ($\therefore S = S(Y_{c\times 1})$) e t são os *ground-truths*, e m é o número de imagens no lote. Ambas as ResNet50 foram treinadas usando o mesmo tamanho de lote (16) e taxa de aprendizado (0.0001) com o algoritmo de descida de gradiente RMSprop (TIELEMAN; HINTON, 2012).

3.3.2.2 N-CNN

Zamzmi et al. (2019) propuseram uma topologia de CNN, denominada N-CNN, para extração de característica e classificação da dor neonatal. Segundo Zamzmi et al. (2019), tal topologia foi a primeira modelada utilizando DL para classificar dor neonatal. Esta dissertação faz uso desse modelo conforme está descrito no artigo (ZAMZMI et al., 2019) e demonstrado na Tabela 12.

Tabela 12 – Parâmetros da N-CNN.

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

Fonte: Adaptado de Zamzmi et al., 2019

No final da extração de característica do modelo os mapas de ativação de características são vetorizados e seu resultado é a entrada para a primeira camada totalmente conectada. Ao todo são anexadas duas camadas totalmente conectadas para que ocorra a classificação, expressa pela Equação 11.

$$y = \sum_{n} w_n ReLU \left(\sum_{m} w_m vec(A)_m \right)_n , \tag{11}$$

onde y é o *logit score* (antes na função de ativação), $A^k \in A$ são os mapas de ativação de características da última camada convolucional sendo k o número do mapas ($\therefore A \in \mathbb{R}^{i \times j \times k}$, i e j representam as dimensão largura e altura dos mapas), então, $vec(A) \in \mathbb{R}^{ijk}$ e, $m = ijk$ e n representa o número de neurônios na camada intermediária. A primeira camada totalmente conectada possui 8 unidades e é seguida pela regularização L2 e Dropout para evitar o over-fitting. A segunda camada totalmente conectada executa a classificação binária usando uma função sigmoide, que gera um valor de 0 a 1, visto que 0 representa o estado *Sem Dor* e 1 o estado *Dor*. Após o treinamento um threshold deverá ser definido, para delimitar o que será os

estados (exemplo, $threshold \geq 0.5$ será o estado *Dor* e $threshold < 0.5$ será o estado *Sem Dor*). O modelo foi treinado sobre a função de erro

$$\mathcal{L}(S,t) = -\frac{1}{m}\sum_{i}^{m} t_i ln(S_i) + (1-t_i) ln(1-S_i), \tag{12}$$

no qual S representa a função sigmoid ($\because S = S(y)$) e t são os *ground-truths*, e m é o número de imagens no lote. O modelo foi treinado usando o tamanho de lote (16) e taxa de aprendizado (0.0001) com o algoritmo de descida de gradiente RMSprop (TIELEMAN; HINTON, 2012).

3.4 MÉTRICAS DE AVALIAÇÃO

As métricas de avaliações topológicas utilizadas por esta dissertação foram as comumente conhecidas: função de erro, em que se monitora o aprendizado do modelo e verifica se o mesmo está sofrendo *overfitting*; acurácia, monitoramento do desempenho do modelo ao longo do aprendizado e matriz de confusão que verifica o desempenho da classificação do modelo em cada classe. Todas essas métricas foram avaliadas com *k-fold* igual a 3. Além disso, utilizou-se também o método de visualização denominado Mapeamento de Ativação de Classe Ponderada por Gradiente (Grad-CAM) (SELVARAJU et al., 2017).

3.4.1 Mapeamento de Ativação de Classe Ponderada por Gradiente

Trabalhos anteriores afirmaram que representações mais profundas em uma CNN capturam construções visuais de nível superior (BENGIO et al., 2013; MAHENDRAN; VEDALDI, 2016). Além disso, as camadas convolucionais naturalmente retêm informações espaciais perdidas nas camadas totalmente conectadas, de modo que pode-se esperar que as últimas camadas convolucionais tenham o melhor compromisso entre a semântica de alto nível e as informações espaciais detalhadas. Portanto, os neurônios nessas camadas procuram informações semânticas específicas da classe na imagem, tais como partes do objeto.

Na literatura de métodos de visualização para Rede Neural Convolucional podem ser encontrados diversos trabalhos que propuseram métodos para interpretar internamente esses modelos, uma vez que são considerados caixa preta. Entre tantos trabalhos, o que se destaca é o método de *Deconvolutional Neural Network* (DeconvNet) proposto por Zeiler e Fergus (2014). Entretanto, esta dissertação fez uso do método de Mapeamento de Ativação de Classe

Ponderada por Gradiente, onde tal método visa na localização dos estados que foram treinados, já o método DeconvNet visa a visualização nos filtros convolucionais.

Para avaliar as informações semânticas de cada classe capturada pela camada convolucional, aplicou-se do Mapeamento de Ativação de Classe Ponderada por Gradiente (Grad-CAM) (SELVARAJU et al., 2017). O Grad-CAM usa as informações de gradiente que fluem para a última camada convolucional da CNN para atribuir valores de importância a cada neurônio para uma decisão específica de interesse. Selvaraju et al. (2017) propuseram o Grad-CAM para modelos que fazem uso do *global-average-pooling* na transição da última camada convolucional para a camada totalmente conectada. Esta dissertação também utilizou modelos que fazem uso da vetorização, no qual vetoriza os mapas de características da última camada convolucional para a camada totalmente conectada. Portanto, para utilizar o Grad-CAM nesses modelos, precisou fazer uma adaptação na computação dos pesos de importância dos neurônios α_k^c ou α_k.

Conforme mostrado na Figura 9 e descrito em (SELVARAJU et al., 2017), para obter o mapa de localização discriminativa de classe Grad-CAM $L^c_{Grad-CAM} \in \mathbb{R}^{u \times v}$ ou $L_{Grad-CAM} \in \mathbb{R}^{u \times v}$, no qual u é a largura e v a altura para qualquer classe c, primeiro computa-se o gradiente do *logit score* para qualquer classe[5] c, y^c ou y (antes na função de ativação) em relação às ativações dos mapas de características A^k de uma camada convolucional, no caso desta dissertação a última camada convolucional, portanto, $\frac{\partial y^c}{\partial A^k}$ e $\frac{\partial y}{\partial A^k}$. Esses gradientes computados sofrem a operação *global-average-pooling* sobre as dimensões de largura e altura dos mapas de características (indexadas por i e j) para obter os pesos de importância do neurônio classe α_k^c ou α_k, se têm:

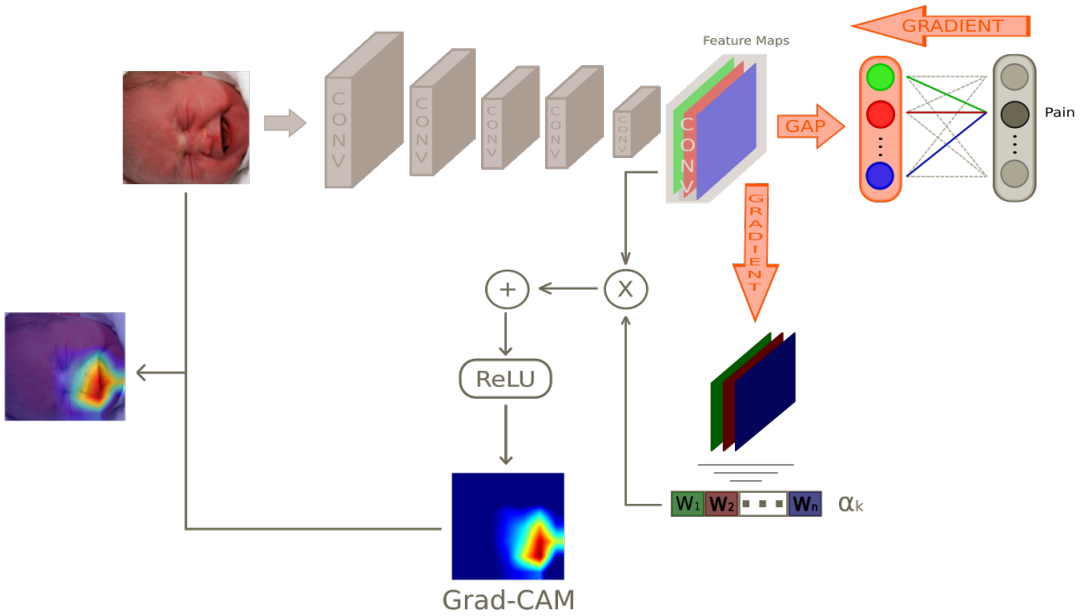
$$\alpha_k^c = \overbrace{\frac{1}{Z}\sum_i\sum_j}^{\text{global average pooling}} \underbrace{\frac{\partial y^c}{\partial A_{ij}^k}}_{\text{gradients via backprop}}, \tag{13}$$

$$\alpha_k = \overbrace{\frac{1}{Z}\sum_i\sum_j}^{\text{global average pooling}} \underbrace{\frac{\partial y}{\partial A_{ij}^k}}_{\text{gradients via backprop}}. \tag{14}$$

O cálculo de α_k^c ou α_k durante a retropropagação de gradientes em relação as ativações, o cálculo exato equivale a produtos matriciais sucessivos das matrizes de peso e o gradiente em

[5] A definição de classe depende se o modelo tem 1 ou 2 neurônios na camada de classificação, logo, o modelo que tem 2 neurônios, cada um representa uma classe. Mas para o modelo que tem 1 único neurônio, tal neurônio representa todas as classes. Portanto, ao dizer classe também está dizendo neurônio.

Figura 9 – Visão geral do Mapeamento de Ativação de Classe Ponderada por Gradiente (Grad-CAM) com *global-average-pooling*. Dada uma imagem e a classe de interesse (exemplo, Dor) como entrada, computa-se a imagem através da CNN até a tarefa específica para obter o *logit score* da classe desejada. Os gradientes são zerados para todas classes, exceto a classe desejada (exemplo, Dor) definida como 1. Então, esse sinal é retropropagado para a camada convolucional de interesse (mapas de características), em que combina a localização dos mapas de características passando pela função ReLU e gerando o Grad-CAM (mapa de calor azul) que representa o local onde o modelo deve procurar a classe de interesse para tomar uma decisão específica.



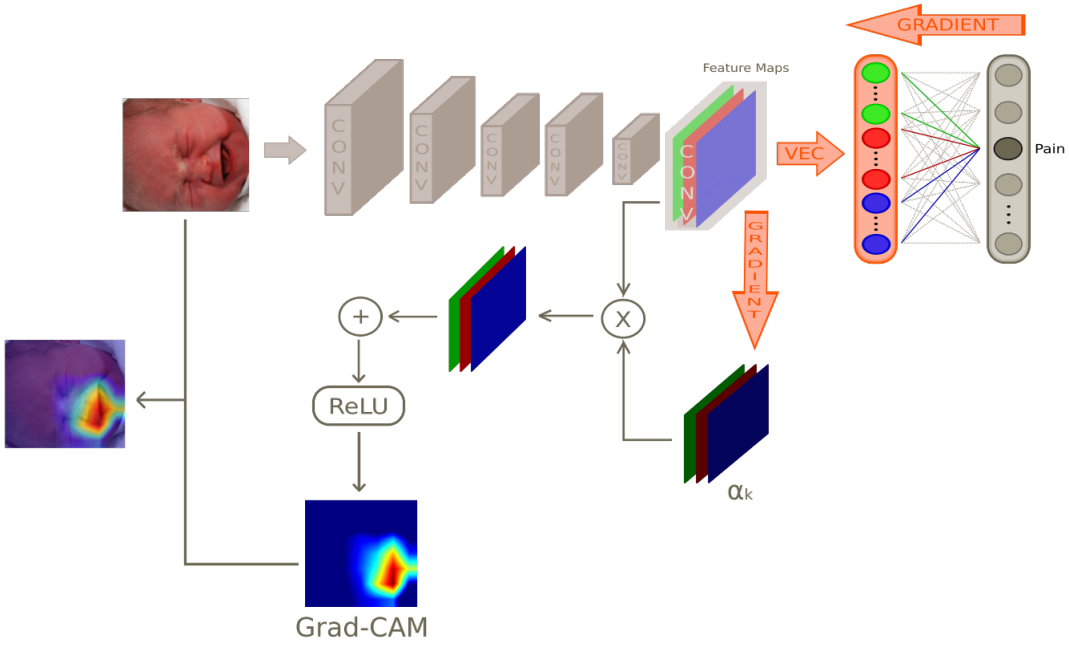
Fonte: Autor

relação às funções de ativação até a camada de convolução final para a qual os gradientes estão sendo propagados. Portanto, esse peso α_k^c ou α_k representa uma linearização parcial da rede profunda em direção a A e captura a importância do mapa de características k para uma classe de destino c.

Realiza-se uma combinação ponderada de mapas de ativação usada como atributo para uma função ReLU, assim obtém-se:

$$L^c_{Grad-CAM} = ReLU \underbrace{\left(\sum_k \alpha_k^c A^k \right)}_{\text{linear combination}}, \tag{15}$$

Figura 10 – Visão geral do Mapeamento de Ativação de Classe Ponderada por Gradiente (Grad-CAM) com vetorização. Dada uma imagem e a classe de interesse (exemplo, Dor) como entrada, computa-se a imagem através da CNN até a tarefa específica para obter o *logit score* da classe desejada. Os gradientes são zerados para todas classes, exceto a classe desejada (exemplo, Dor) definida como 1. Então, esse sinal é retropropagado para a camada convolucional de interesse (mapas de características), em que combina a localização dos do mapas de características passando pela função ReLU e gerando o Grad-CAM (mapa de calor azul) que representa o local onde o modelo deve procurar a classe de interesse para tomar uma decisão específica.



Fonte: Autor

$$L_{Grad-CAM} = ReLU \underbrace{\left(\sum_k \alpha_k A^k\right)}_{\text{linear combination}}, \tag{16}$$

resultando em um mapa de calor aproximado do mesmo tamanho dos mapas de características convolucionais. Selvaraju et al. (2017) destacaram que os mapas Grad-CAM se tornam progressivamente piores à medida que se avança para as camadas convolucionais anteriores, pois possuem campos receptivos menores e focam apenas em características locais menos semânticas. Aplica-se a função ReLU na combinação linear dos mapas, uma vez que o interesse está apenas nas características que influenciam positivamente a classe desejada, ou seja, pixels

cuja intensidade deve ser aumentada para aumentar y^c ou y. Segundo Selvaraju et al. (2017), é provável que os pixels negativos pertençam a outras categorias na imagem e sem a ReLU os mapas de localização às vezes destacam mais do que apenas a classe desejada e apresentam desempenho pior na localização.

Como mencionado anteriormente, esta dissertação também está utilizando modelos que usam a vetorização entre a última camada convolucional e a camada totalmente conectada, portanto, o cálculo de $L_{Grad-CAM}$ e α_k sofreram modificações. As modificações são ilustradas na Figura 10 e descritas pelas Equações (17) e (18).

Devido a vetorização não se aplica a equação do *global-average-pooling* no α_k, uma vez que a vetorização utiliza-se de todos os pixels para a camada de classificação. Logo, o cálculo de $L_{Grad-CAM}$ será a multiplicação de cada pixel de cada mapa de característica por cada gradiente calculado até a última camada convolucional, logo, α_k e $L_{Grad-CAM}$ são definidos como:

$$\alpha_k = \underbrace{\frac{\partial y}{\partial A^k}}_{\text{gradients via backprop}}, \tag{17}$$

$$L_{Grad-CAM} = ReLU \underbrace{\left(\sum_k \alpha_k A^k \right)}_{\text{linear combination}}, \tag{18}$$

sendo, $\alpha_k, A^k \in \mathbb{R}^{i \times j}$ e $L_{Grad-CAM} \in \mathbb{R}^{u \times v}$.

4 EXPERIMENTOS E RESULTADOS

4.1 ANÁLISE QUANTITATIVA

Os bancos de imagens COPE e UNIFESP passaram pela etapa de pré-processamento, mas antes de tal etapa, as imagens desses bancos foram selecionadas para construir um conjunto de dados de cada banco contendo apenas imagens dos neonatos nos estados “Dor” e “Sem Dor”. No banco de imagens da UNIFESP os dois estados que esta dissertação faz uso já estão presentes, ou seja, apenas imagens de neonatos nos estados “Dor” e “Sem Dor”. Mas houve a necessidade de retirar 4 imagens, uma vez que estavam sem seus devidos rótulos. Portanto, utilizou-se 356 imagens de 360 do banco de imagens da UNIFESP. Em relação ao banco de imagens COPE, há mais estados (choro, estímulo de ar e fricção), contabilizando um total de 288 imagens. Sendo assim, foram selecionados as imagens com os rótulos *rest* e *pain*, sendo esses sem dor e com dor, respectivamente. Após tal seleção, contabilizou-se 153 imagens do banco COPE.

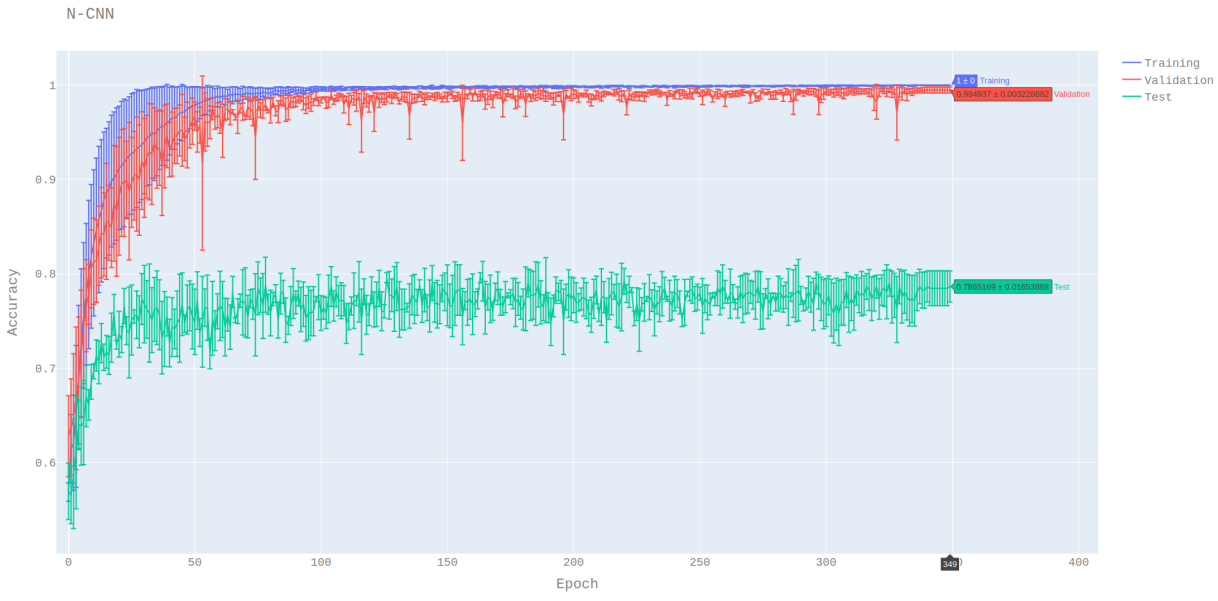
A RetinaFace foi aplicada em ambos os bancos de imagens após a seleção e obteve as coordenadas de onde a face do bebê está localizada, para então, selecionar somente a face, uma vez que se pretende que o modelo aprenda a distinguir os atributos que estão contidos na face. Realizado tal processamento de seleção, as imagens foram impostas ao aumento de dados, visando aumentar apenas os conjunto de treinamento e validação. Dividiu-se então o banco de imagens em dois conjuntos I_{test} e I^*, 50% das imagens para cada conjunto, e no I^* foi onde que ocorreu o aumento, sendo que o $|I^*| = 36|I^*| + |I^*|$ após tal aumento. O I^* foi dividido em 70% para o conjunto de treinamento I_{train} e 30% para o conjunto de validação I_{val}. Logo, os conjuntos de dados ficaram:

a) UNIFESP: $I_{train} = 4610$ imagens, $I_{val} = 1976$ imagens, $I_{test} = 178$ imagens
b) COPE: $I_{train} = 1968$ imagens, $I_{val} = 844$ imagens, $I_{test} = 76$ imagens

A partir dos conjuntos definidos introduziu-se I_{train}, I_{val} e I_{test} em cada modelo, mas somente os parâmetros dos modelos que utilizaram I_{train} para treinamento foram atualizados. O I_{val} foi utilizado para validar os parâmetros dos modelos no treinamento, no intuito de verificar se o modelo sofreu overfitting durante o treinamento. A métrica utilizada para tal verificação foi a função de erro dos devidos modelos, mas ao computar tais erros tanto de I_{train} e I_{val} a atualização dos parâmetros ocorreu apenas com os erros referente à I_{train}. Para verificar a

eficiência de cada modelo utilizou-se I_{test}, conjunto de dados distinto, de tal forma que foi computada a acurácia de cada modelo com a introdução de I_{val} e I_{test} nos mesmos períodos de épocas de treinamento. A acurácia de I_{test} é que irá validar a eficiência do modelo para dados distintos, em outras palavras, a métrica de acurácia irá indicar o quanto eficiente o modelo está sendo para dados que não foram usados no treinamento, medindo assim sua capacidade de generalização. Não houve um critério para determinar a quantidade de épocas que cada modelo deveria ser treinado, o que ocorreu foi deixar cada modelo encontrar o seu plato de erro e de acurácia, mas monitorando a métrica de erro. Para analisar o comportamento de cada modelo nos seus respectivos treinamentos utilizou-se da validação cruzada com *k-fold=3*. Portanto, as métricas de avaliações, função de erro e acurácia de cada modelo estão sendo analisadas pela média e desvio padrão. Nas Figuras 12, 14, 16, 18, 20 e 22 estão mostrando a curva média do treinamento de cada modelos nos bancos de imagens UNIFESP e COPE, a métrica dos respectivos gráficos é a função de erro médio.

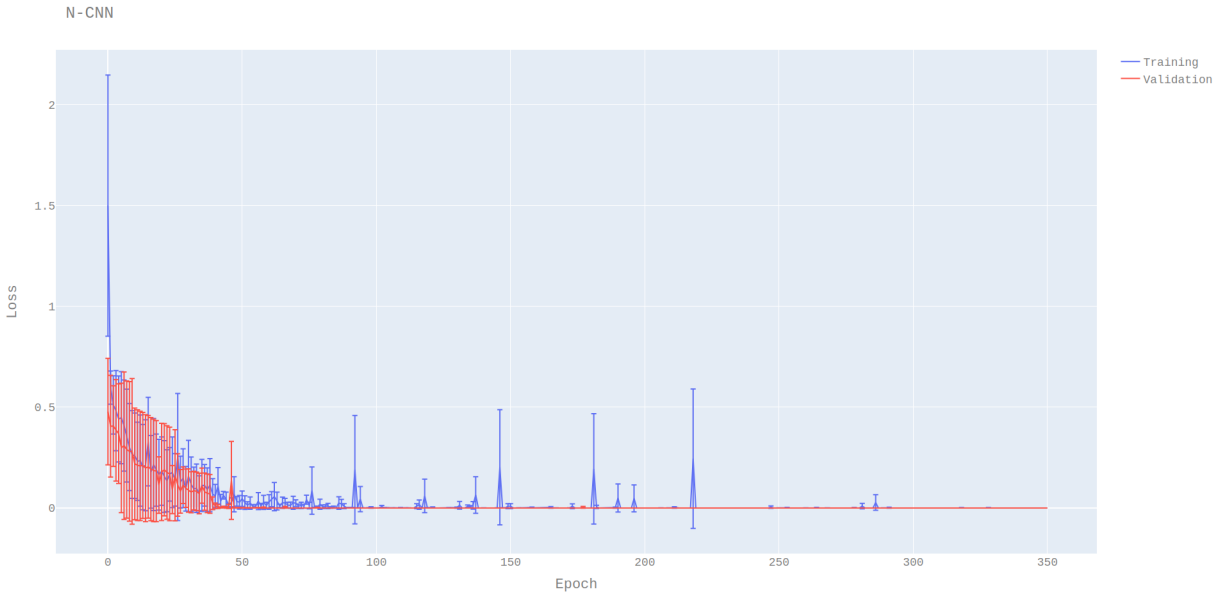
Figura 11 – Gráficos da métrica de acurácia média do treinamento, validação e teste do modelo N-CNN proposto por Zamzmi et al. (2019) treinado com o banco de imagens da UNIFESP. Tal gráfico mostra a convergência do modelo para um determinado plato ao longo do tempo na época 349 com 78.7% de acurácia média.



Fonte: Autor

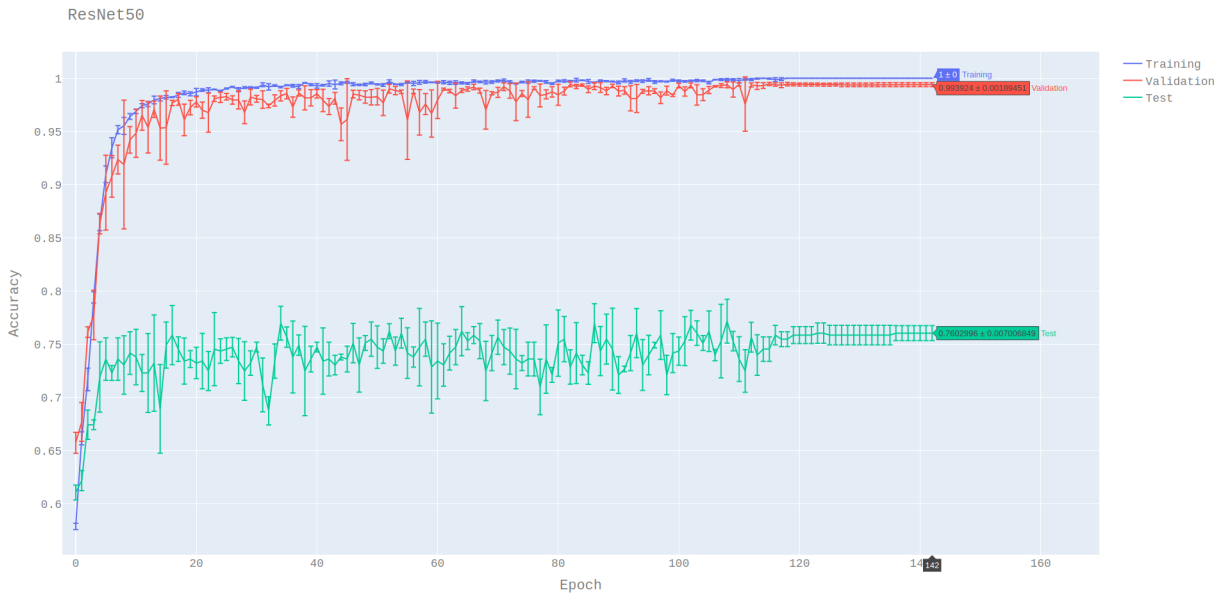
Um fato peculiar ocorrido no treinamento dos modelos foi o número de épocas necessárias para as mesmas entrarem em um plato, onde não haverá mais atualizações dos parâmetros. Os modelos treinados com o banco de imagens COPE pararam de encontrar a solução ótima em

Figura 12 – Gráficos da curva média de treinamento do modelo N-CNN proposto por Zamzmi et al. (2019) treinado com o banco de imagens da UNIFESP. O gráfico mostra que o modelo não sofreu overfitting e convergiu para um determinado erro ao longo do tempo.



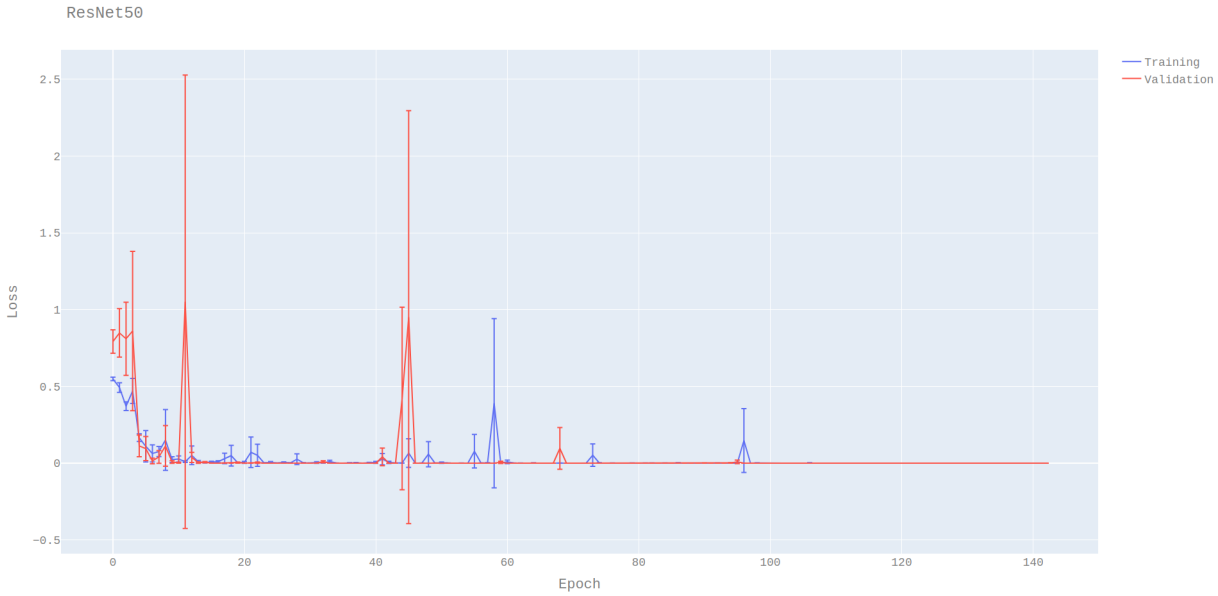
Fonte: Autor

Figura 13 – Gráficos da métrica de acurácia média do treinamento, validação e teste do modelo ResNet50 proposto por Zamzmi et al. (2019) treinado com o banco de imagens da UNIFESP. Tal gráfico mostra a convergência do modelo para um determinado plato ao longo do tempo na época 142 com 76.0% de acurácia média.



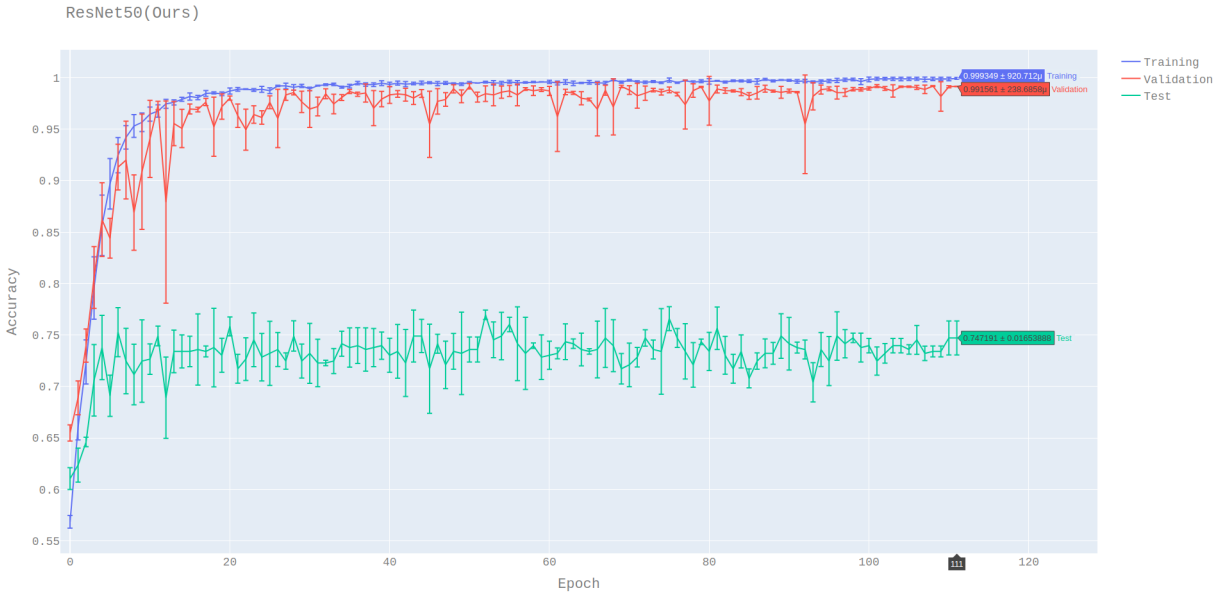
Fonte: Autor

Figura 14 – Gráficos da curva média de treinamento do modelo ResNet50 proposto por Zamzmi et al. (2019) treinado com o banco de imagens da UNIFESP. O gráfico mostra que o modelo não sofreu overfitting e convergiu para um determinado erro ao longo do tempo.



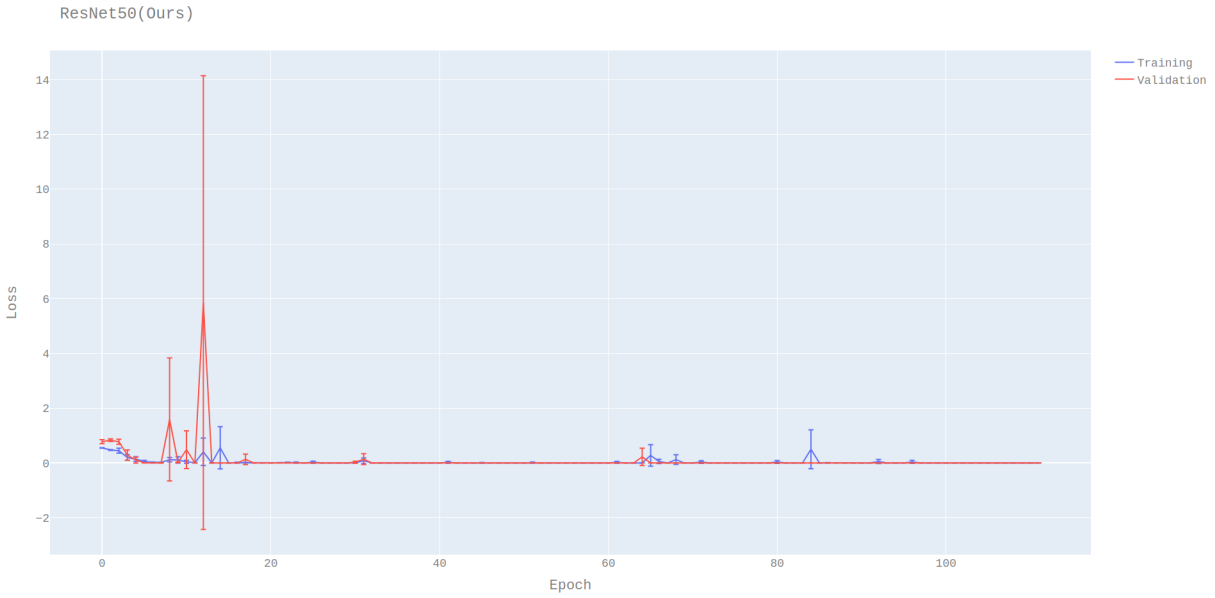
Fonte: Autor

Figura 15 – Gráficos da métrica de acurácia média do treinamento, validação e teste do modelo ResNet50 proposto por esta dissertação treinado com o banco de imagens da UNIFESP. Tal gráfico mostra a convergência do modelo para um determinado plato ao longo do tempo na época 111 com 74.7% de acurácia média.



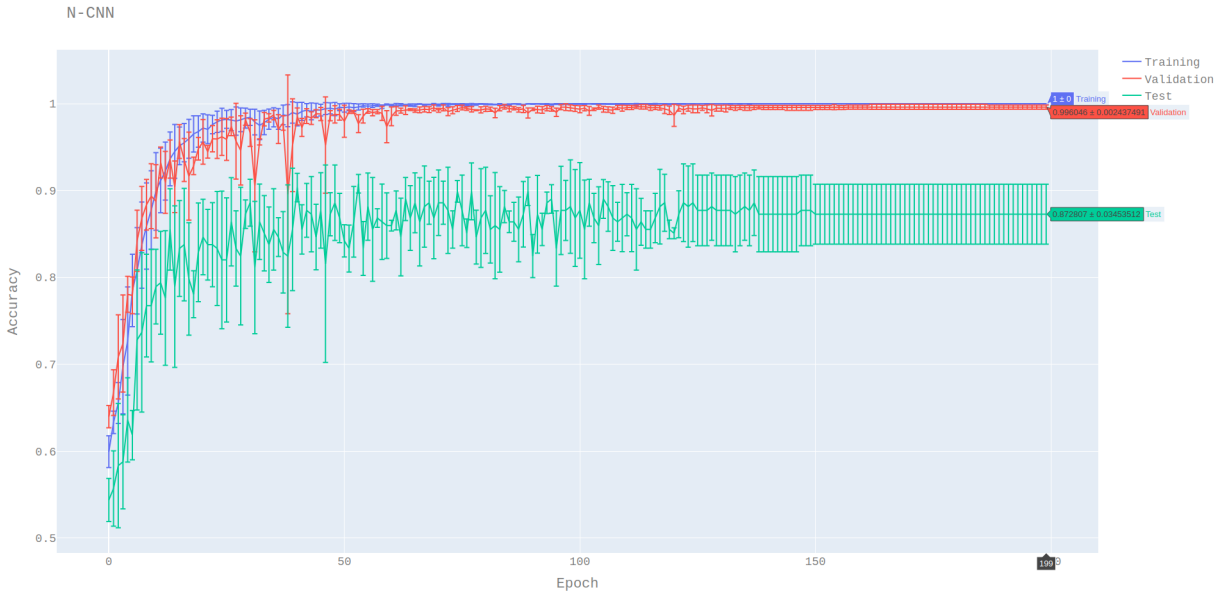
Fonte: Autor

Figura 16 – Gráficos da curva média de treinamento do modelo ResNet50 proposto por esta dissertação treinado com o banco de imagens da UNIFESP. O gráfico mostra que o modelo não sofreu overfitting e convergiu para um determinado erro ao longo do tempo.



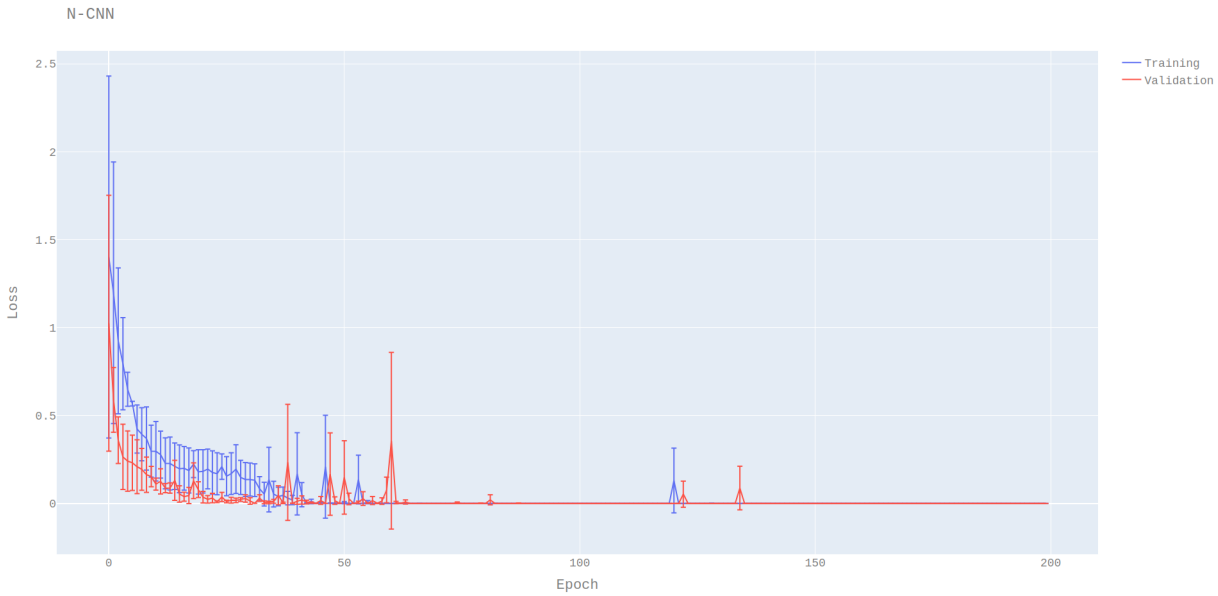
Fonte: Autor

Figura 17 – Gráficos da métrica de acurácia média do treinamento, validação e teste do modelo N-CNN proposto por Zamzmi et al. (2019) treinado com o banco de imagens COPE. Tal gráfico mostra a convergência do modelo para um determinado plato ao longo do tempo na época 199 com 87.3% de acurácia média.



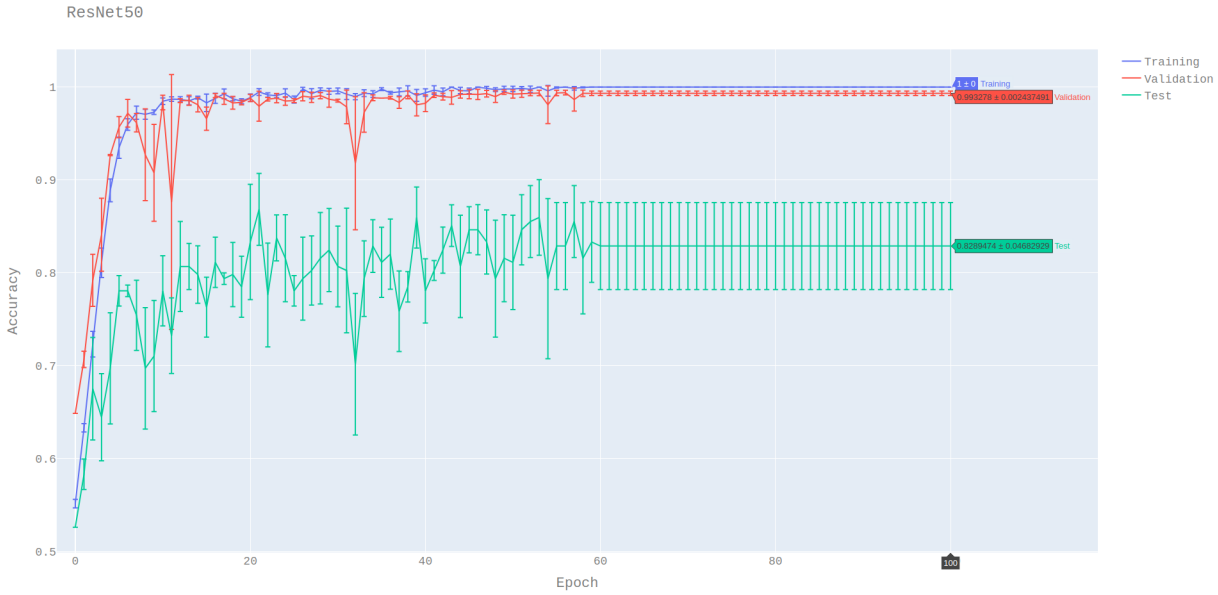
Fonte: Autor

Figura 18 – Gráficos da curva média de treinamento do modelo N-CNN proposto por Zamzmi et al. (2019) treinado com o banco de imagens da COPE. O gráfico mostra que o modelo não sofreu overfitting e convergiu para um determinado erro ao longo do tempo.



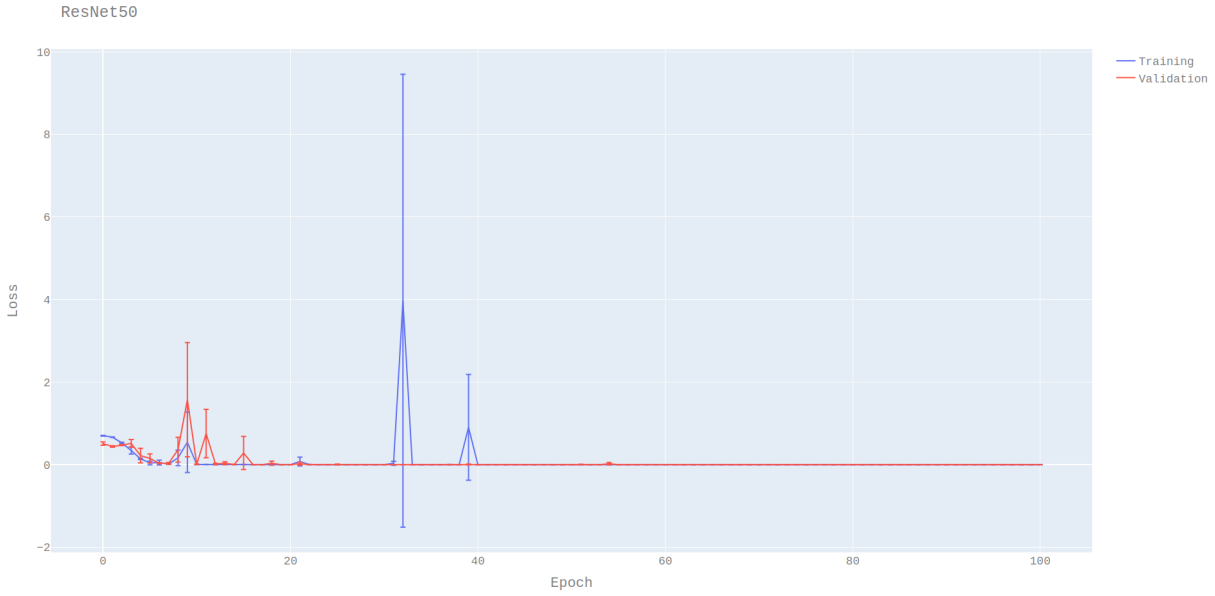
Fonte: Autor

Figura 19 – Gráficos da métrica de acurácia média do treinamento, validação e teste do modelo ResNet50 proposto por Zamzmi et al. (2019) treinado com o banco de imagens da COPE. Tal gráfico mostra a convergência do modelo para um determinado plato ao longo do tempo na época 100 com 83.0% de acurácia média.



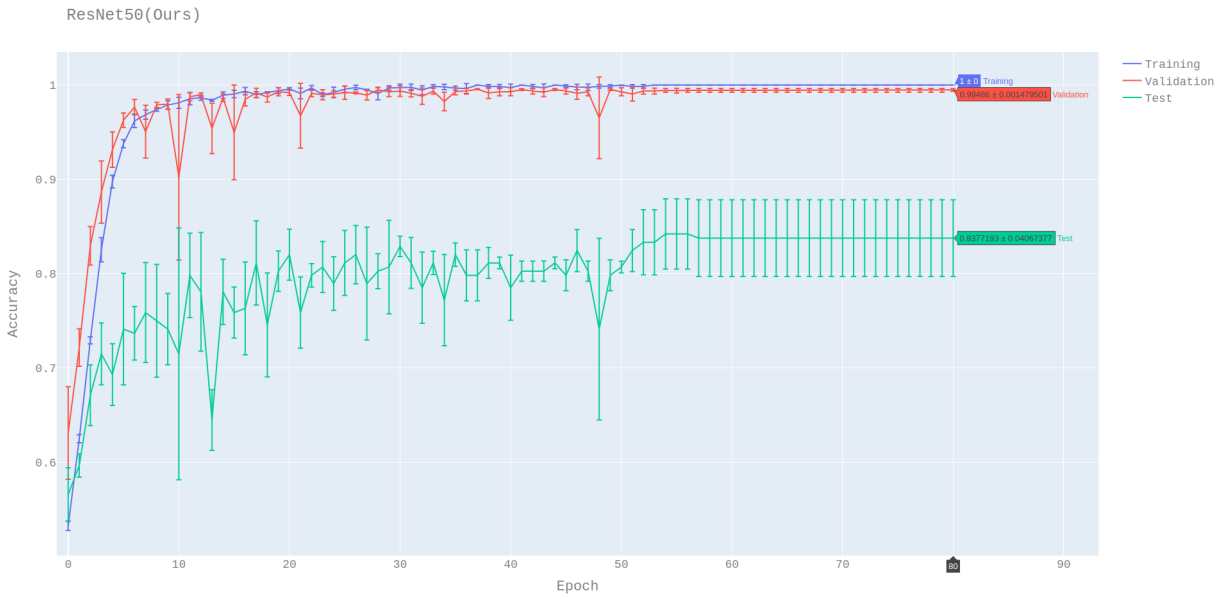
Fonte: Autor

Figura 20 – Gráficos da curva média de treinamento do modelo ResNet50 proposto por Zamzmi et al. (2019) treinado com o banco de imagens da COPE. O gráfico mostra que o modelo não sofreu overfitting e convergiu para um determinado erro ao longo do tempo.



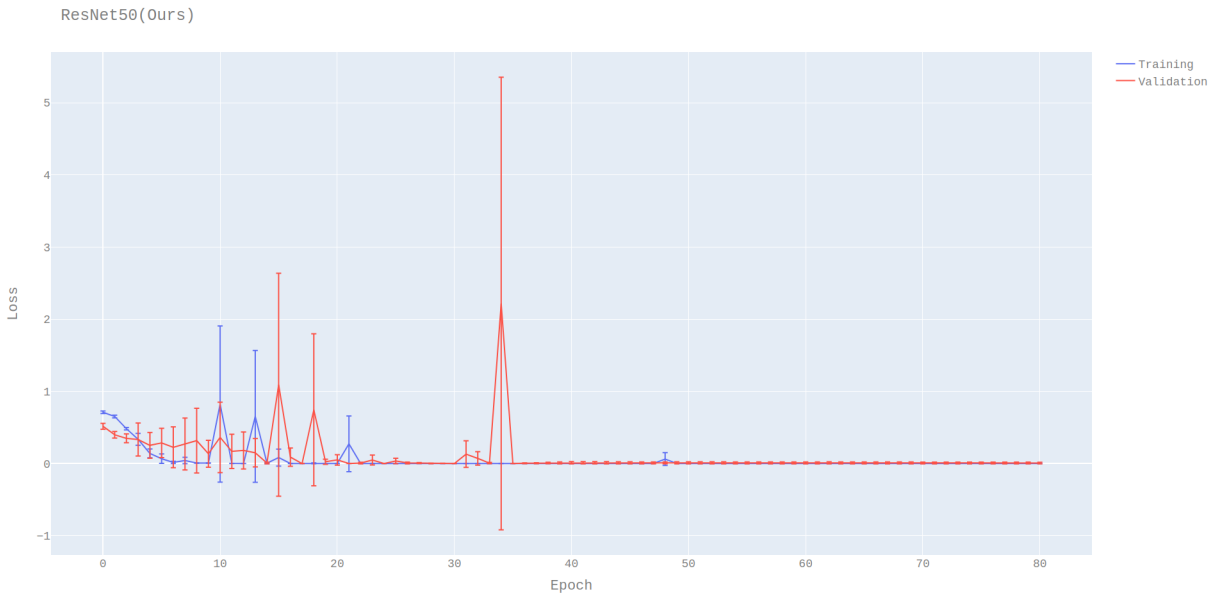
Fonte: Autor

Figura 21 – Gráficos da métrica de acurácia média do treinamento, validação e teste do modelo ResNet50 proposto por esta dissertação treinado com o banco de imagens da COPE. Tal gráfico mostra a convergência do modelo para um determinado plato ao longo do tempo na época 80 com 84.0% de acurácia média.



Fonte: Autor

Figura 22 – Gráficos da curva média de treinamento do modelo ResNet50 proposto por esta dissertação treinado com o banco de imagens da COPE. O gráfico mostra que o modelo não sofreu overfitting e convergiu para um determinado erro ao longo do tempo.



Fonte: Autor

menos épocas. Em contraste, os modelos treinados com o banco de imagens UNIFESP necessitaram de mais épocas para pararem de encontrar a solução ótima. Nas Figuras 11, 13, 15, 17, 19 e 21 pode-se visualizar a curva média de acurácia de cada modelo nos bancos de imagens UNIFESP e COPE, e a época em que cada modelo entraram em seus respectivos plato. Tal fato está relacionado ao número de imagens nos bancos, visto que do banco de imagens COPE está sendo utilizado 153 imagens e do banco de imagens UNIFESP está sendo utilizado 356 imagens. Embora a técnica de aumento de dados esteja sendo aplicada, tal técnica não consegue gerar imagens que tenham as mesmas relações das imagens originais. Em outras palavras, não consegue criar novas imagens nunca vista antes com a mesma similaridade e semântica das imagens originais, em comparação com estudos que fazem uso de DL para geração de dados, como as *Generative Adversarial Networks* (GAN's) (KARRAS et al., 2019; SONG; ERMON, 2019).

Os resultados das Tabelas 13 e 14 demostram as acurácias de treinamento, validação e teste de cada modelo treinado com os bancos de imagens UNIFESP e COPE, respectivamente. Nos dois bancos de imagens, os modelos tiveram uma acurácia média e um desvio padrão satisfatórios para o conjunto de validação, acima de 95.0% e abaixo de 2.0%, assim validando

os parâmetros de cada modelo e a sua robustez para as operações geométricas do aumento de dados.

Tabela 13 – Resultado da acurácia média da avaliação da dor neonatal em expressão facial dos modelos treinados, validados e testados com o banco de imagens UNIFESP.

	Accuracy		
Model	**Training**	**Validation**	**Test**
N-CNN (ZAMZMI et al., 2019)	1.00 ± 0.0000	0.995 ± 0.0032	0.787 ± 0.0165
ResNet50 (ZAMZMI et al., 2019)	1.00 ± 0.0000	0.994 ± 0.0019	0.760 ± 0.0070
ResNet50 (ours)	1.00 ± 0.0000	0.992 ± 0.0000	0.747 ± 0.0165

Fonte: Autor

Tabela 14 – Resultado da acurácia média da avaliação da dor neonatal em expressão facial dos modelos treinados, validados e testados com o banco de imagens COPE.

	Accuracy		
Model	**Training**	**Validation**	**Test**
N-CNN (ZAMZMI et al., 2019)	1.00 ± 0.0000	0.996 ± 0.0024	0.872 ± 0.0345
ResNet50 (ZAMZMI et al., 2019)	1.00 ± 0.0000	0.993 ± 0.0024	0.829 ± 0.0468
ResNet50 (ours)	1.00 ± 0.0000	0.995 ± 0.0015	0.838 ± 0.0407

Fonte: Autor

A última coluna das Tabelas 13 e 14 mostra a acurácia média do conjunto de teste, um conjunto totalmente distinto dos conjuntos de treinamento e validação. Analisando a acurácia média de teste dos modelos ResNet50 proposta por Zamzmi et al. (2019) e ResNet50 proposta por esta dissertação, ResNet50(ours), analisa-se que a ResNet50(ours) foi ligeiramente melhor (em 1.0%) ao comparar com a ResNet50 proposta por Zamzmi et al. (2019), para o banco de imagens COPE. Entretanto, o desvio padrão de ambos os modelos ficaram próximos dizendo estatisticamente que os dois modelos são essencialmente iguais. Analisando esses mesmos modelos para o banco de imagens UNIFESP, percebe-se que a ResNet50 proposta por Zamzmi et al. (2019) foi melhor em 1.0% ao comparar com a ResNet50(ours). Mas ambos os modelos obtiveram desvios próximos. Logo, fica incerto definir qual modelo foi melhor: com 1 neurônio ou com 2 neurônio na camada de classificação, visto que ambos os modelos estatisticamente se mostraram equivalentes. Entretanto, quando analisa-se os gráficos dos modelos ResNet50 proposto por Zamzmi et al. (2019) e ResNet50 proposto por esta dissertação (Figuras 13, 15, 19 e 21), percebe-se claramente que o modelo ResNet50(ours) precisou de menos épocas para convergir, tendo sido mais eficiente computacionalmente. Isso ocorre devido ao número de

neurônios, visto que o modelo ResNet50(ours) está atribuindo parâmetros distintos para cada classe de treinamento.

Analisando a acurácia dos modelos ResNet50 e N-CNN para definir qual o melhor modelo para ser usado na classificação da dor neonatal, entende-se que em ambos os bancos de imagens o modelo N-CNN teve o melhor desempenho estatisticamente. O desempenho da N-CNN está na questão topológica, pois tanto a N-CNN como a ResNet50 fazem o uso de blocos residuais, mas a ResNet50 possui muitas camadas, portanto, para modelar um modelo de DL que possa ter um bom desempenho na classificação da dor neonatal entende-se que o modelo precisa de blocos residuais e principalmente não ser muito profundo, como a ResNet50.

Tabela 15 – Matriz de confusão média dos modelos N-CNN e ResNet50 propostos por Zamzmi et al. (2019), e ResNet50 proposto por esta dissertação treinados com o banco de imagens UNIFESP.

	Rest		**Pain**	
Model	**TP**	**FP**	**TN**	**FN**
N-CNN (ZAMZMI et al., 2019)	0.85 ± 0.0082	0.15 ± 0.0082	0.71 ± 0.0455	0.29 ± 0.0455
ResNet50 (ZAMZMI et al., 2019)	0.85 ± 0.0287	0.15 ± 0.0287	0.65 ± 0.0262	0.35 ± 0.0262
ResNet50 (ours)	0.84 ± 0.0573	0.16 ± 0.0573	0.65 ± 0.0330	0.35 ± 0.0330

Fonte: Autor

Tabela 16 – Matriz de confusão média dos modelos N-CNN e ResNet50 propostos por Zamzmi et al. (2019), e ResNet50 proposto por esta dissertação treinados com o banco de imagens COPE.

	Rest		**Pain**	
Model	**TP**	**FP**	**TN**	**FN**
N-CNN (ZAMZMI et al., 2019)	0.97 ± 0.0236	0.03 ± 0.0236	0.77 ± 0.0464	0.23 ± 0.0464
ResNet50 (ZAMZMI et al., 2019)	0.95 ± 0.0163	0.05 ± 0.0163	0.70 ± 0.0801	0.30 ± 0.0801
ResNet50 (ours)	0.96 ± 0.0189	0.04 ± 0.0189	0.70 ± 0.0822	0.30 ± 0.0822

Fonte: Autor

As Tabelas 15 e 16 mostram as matrizes de confusão média de cada modelo nos respectivos bancos de imagens, UNIFESP e COPE. Analisando as duas últimas linhas nessas Tabelas comprova-se as afirmações ditas nos parágrafos anteriores, uma vez que os resultados em ambos os bancos com os modelos ResNet50 proposta por Zamzmi et al. (2019) e ResNet50 proposta por esta dissertação são muito próximos. Como dito antes o modelo N-CNN foi melhor em ambos os bancos de imagens. Um fato interessante dos resultados mostrados nas Tabelas 15 e 16 foi que todos os modelos em ambos os bancos tiveram dificuldade de classificar o estado ”Dor“, ficando entre 65.0% e 77.0% de acerto. Já no estado ”Sem Dor“ os modelos ficaram entre 84.0% e 97.0% de acurácia.

4.2 ANÁLISE QUALITATIVA

Analisar apenas a estatística do modelo não garante a eficiência do mesmo, em outras palavras, não se tem a garantia se o modelo aprendeu o que foi imposto no treinamento, logo, o modelo pode ter aprendido artefatos arredores do objeto de interesse. Por esse fato, é relevante aplicar métodos de visualização após o treinamento do modelo. O método que esta dissertação fez uso foi o Mapeamento de Ativação de Classe Ponderada por Gradiente (Grad-CAM). (SELVARAJU et al., 2017).

Devido a utilização da validação cruzada, cada modelo foi computado três vezes em cada banco de imagens (*k-fold=3*). Portanto, para o cálculo do Grad-CAM estabeleceu-se o critério de usar o melhor resultado do *k-fold* de cada modelo. Nas Tabelas 17 e 18 são mostrados os resultados do melhor *k-fold* de cada modelo para os bancos de imagens UNIFESP e COPE, respectivamente. A partir desses modelos treinados foi calculado o Grad-CAM.

Tabela 17 – Resultado da acurácia do melhor *k-fold* de cada modelo treinado, validado e testado com o banco de imagens UNIFESP.

		Accuracy		
Model	**The best k-fold**	**Training**	**Validation**	**Test**
N-CNN (ZAMZMI et al., 2019)	k_2	1.00	0.996	0.810
ResNet50 (ZAMZMI et al., 2019)	k_2	1.00	0.993	0.770
ResNet50 (ours)	k_1	1.00	0.991	0.770

Fonte: Autor

Tabela 18 – Resultado da acurácia do melhor *k-fold* de cada modelo treinado, validado e testado com o banco de imagens COPE.

		Accuracy		
Model	**The best k-fold**	**Training**	**Validation**	**Test**
N-CNN (ZAMZMI et al., 2019)	k_3	1.00	0.995	0.921
ResNet50 (ZAMZMI et al., 2019)	k_3	1.00	0.995	0.895
ResNet50 (ours)	k_2	1.00	0.992	0.895

Fonte: Autor

Com a utilização do Grad-CAM, busca-se compreender no nível de semântica como os modelos convolucionais discriminam as classes que foram treinadas, logo, as classes que esta dissertação buscou compreender foram “Dor” e “Sem Dor” na área da neonatologia.

Para estabelecer uma análise igualitária dos modelos nos seus respectivos banco de imagens, selecionou-se as mesmas imagens para serem analisadas. Entretanto, na seleção das ima-

Tabela 19 – Matriz de confusão do melhor *k-fold* dos modelos N-CNN e ResNet50 propostos por Zamzmi et al. (2019), e ResNet50 proposto por esta dissertação treinados com o banco de imagens UNIFESP.

		Rest		Pain	
Model	**The best k-fold**	**TP**	**FP**	**TN**	**FN**
N-CNN (ZAMZMI et al., 2019)	k_2	0.84	0.16	0.77	0.23
ResNet50 (ZAMZMI et al., 2019)	k_2	0.89	0.11	0.63	0.37
ResNet50 (ours)	k_1	0.91	0.09	0.61	0.39

Fonte: Autor

Tabela 20 – Matriz de confusão do melhor *k-fold* dos modelos N-CNN e ResNet50 propostos por Zamzmi et al. (2019), e ResNet50 proposto por esta dissertação treinados com o banco de imagens COPE.

		Rest		Pain	
Model	**The best k-fold**	**TP**	**FP**	**TN**	**FN**
N-CNN (ZAMZMI et al., 2019)	k_3	1.00	0.00	0.83	0.17
ResNet50 (ZAMZMI et al., 2019)	k_3	0.97	0.03	0.81	0.19
ResNet50 (ours)	k_2	0.97	0.03	0.81	0.19

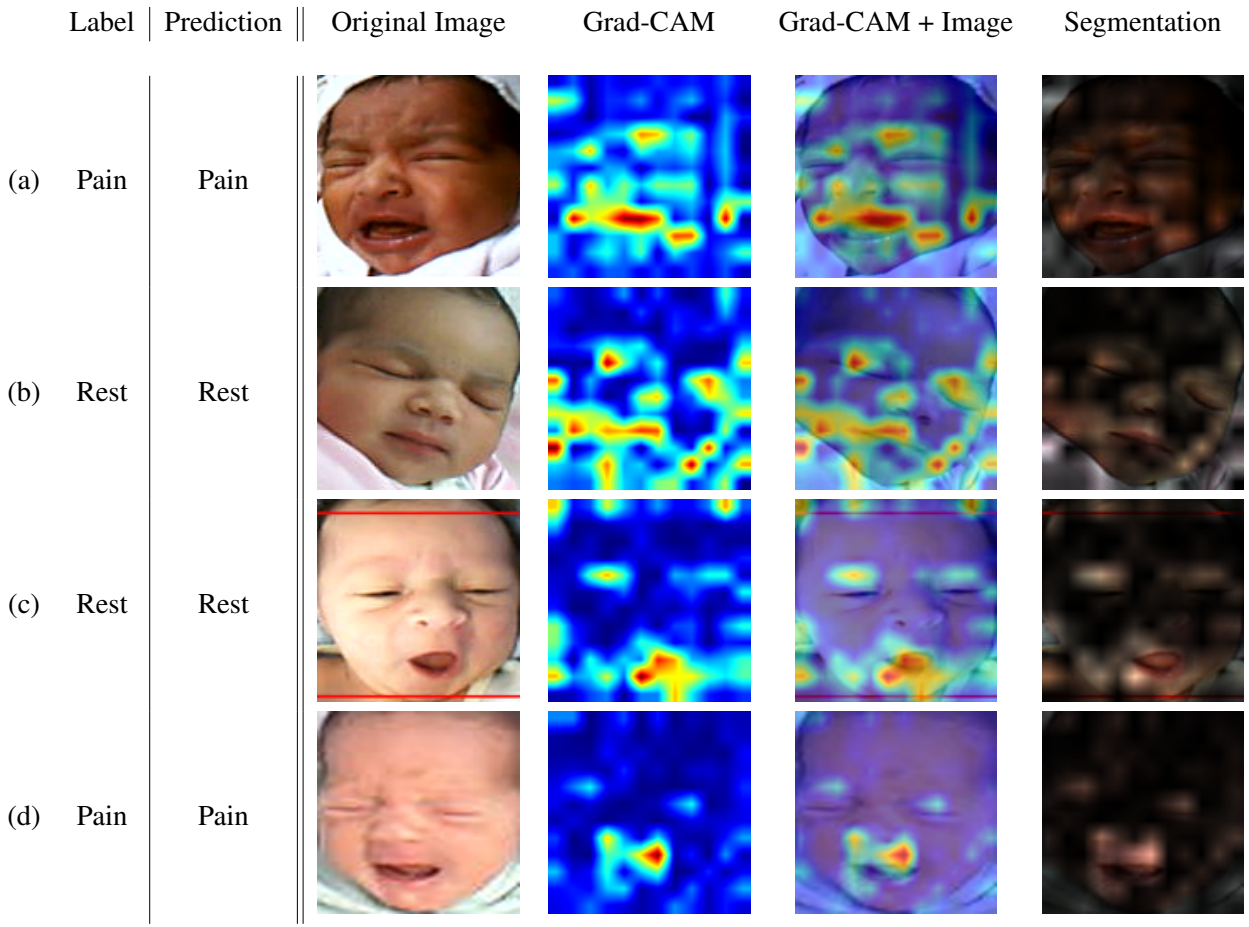
Fonte: Autor

gens dos falsos positivos e falsos negativos foi mais desafiador devido ao desempenho de cada modelo, mostrado nas Tabelas 19 e 20. Portanto, para alguns modelos não foi possível selecionar tais imagens falsos positivos ou falsos negativos, um nítido exemplo é o modelo N-CNN treinado com o banco de imagens COPE mostrado na 2° linha da Tabela 20, sendo que classificou 100% das imagens com estado "Sem Dor" corretamente, logo, impossibilitando a seleção de falsos positivos do estado "Sem Dor" para análise.

Com a utilização do Grad-CAM foi possível visualizar nas imagens atribuídas aos modelos os locais discriminantes dos estados "Dor" e "Sem Dor", como mostram as Figuras 23 à 34. Todas essas visualizações foram realizadas com imagens consideradas verdadeiros positivos, verdadeiros negativos, falsos positivos e falsos negativos, para revelar os locais discriminantes quando o modelo acerta a classificação e quando o modelo erra a classificação.

Computando o Grad-CAM de 4 imagens neonatais, sendo 2 imagens sem dor (verdadeiros positivos) e 2 imagens com dor (verdadeiros negativos), no modelo N-CNN treinado com o banco de imagens UNIFESP, mostrado na Figura 23, observa-se que em todas as 4 imagens o modelo aprendeu locais discriminantes do estado "Dor" e "Sem Dor". Os locais mais discriminantes foram as regiões da boca, sulco nasolabial e sobrancelhas, regiões comumente analisadas pelos profissionais de saúde para identificar a dor nos recém-nascidos. Porém não

Figura 23 – Resultado do Mapeamento de Ativação de Classe Ponderada por Gradiente (Grad-CAM) do modelo N-CNN treinado com o banco de imagens UNIFESP. As Figuras Grad-CAM (a-d) estão representando os locais de detecção dos estados “Dor” ou “Sem Dor” da Figura *Original Image* correspondente. As Figuras *Segmentation* mostram os locais descriminantes do estado correspondente à Figura *Original Image*.



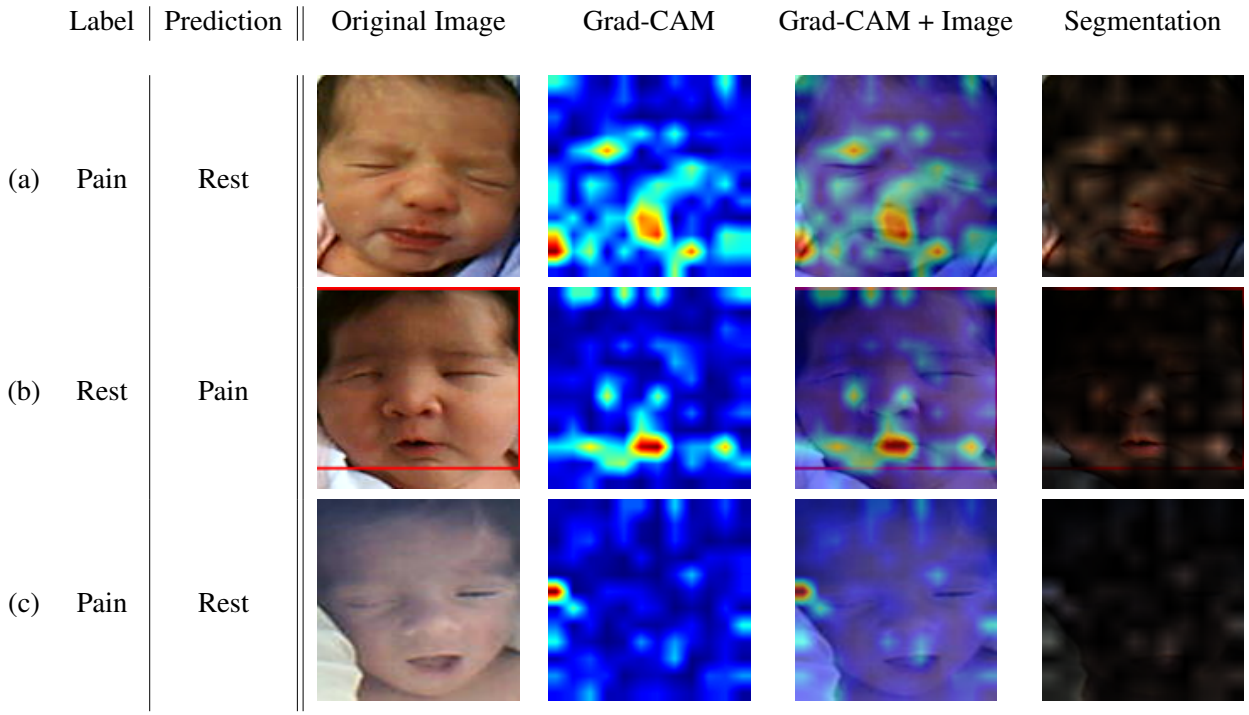
Fonte: Autor

houve discriminações nas regiões do nariz e olhos nas Figuras 23c e 23d, uma vez que são regiões de avaliações para identificar a dor nos recém-nascidos pelos profissionais de saúde. Tais não extrações dos locais discriminantes comumente analisados pelos profissionais de saúde não significa que o modelo é desprovido de eficiência, o que realmente significa é que o modelo foi capaz de classificar o estado do neonato sem precisar localizar essas regiões discriminantes. Além desses locais mais discriminantes, o modelo também localizou regiões não muito perceptivas mas discriminantes, tais regiões estão sendo mostradas na Figura 23 coluna *Segmentation*.

A Figura 24 apresenta os resultados do Grad-CAM de 3 imagens neonatais, imagens que foram classificadas como falso positivo e falso negativo pelo modelo N-CNN treinado com o banco da UNIFESP. Ao analisar os resultados do Grad-CAM do modelo N-CNN percebe-se que o modelo teve regiões discriminantes extraídas na face do recém-nascido, mas os pontos discriminantes, em sua maioria, são artefatos não significativos para predizer se o recém-nascido está com dor ou sem dor, como mostra a Figura 24c. Também na Figura 24a foi extraído região discriminante fora da área de interesse (face do neonato). Apesar de haver região discriminante da área de interesse o modelo classificou incorretamente, presumindo que a região discriminante fora da área de interesse sobressaiu na decisão do modelo. Os pontos discriminantes na Figura 24a têm mais percepções na região da boca e parte das bochechas, entretanto, tais regiões foram insuficientes para o modelo classificar corretamente a imagem.

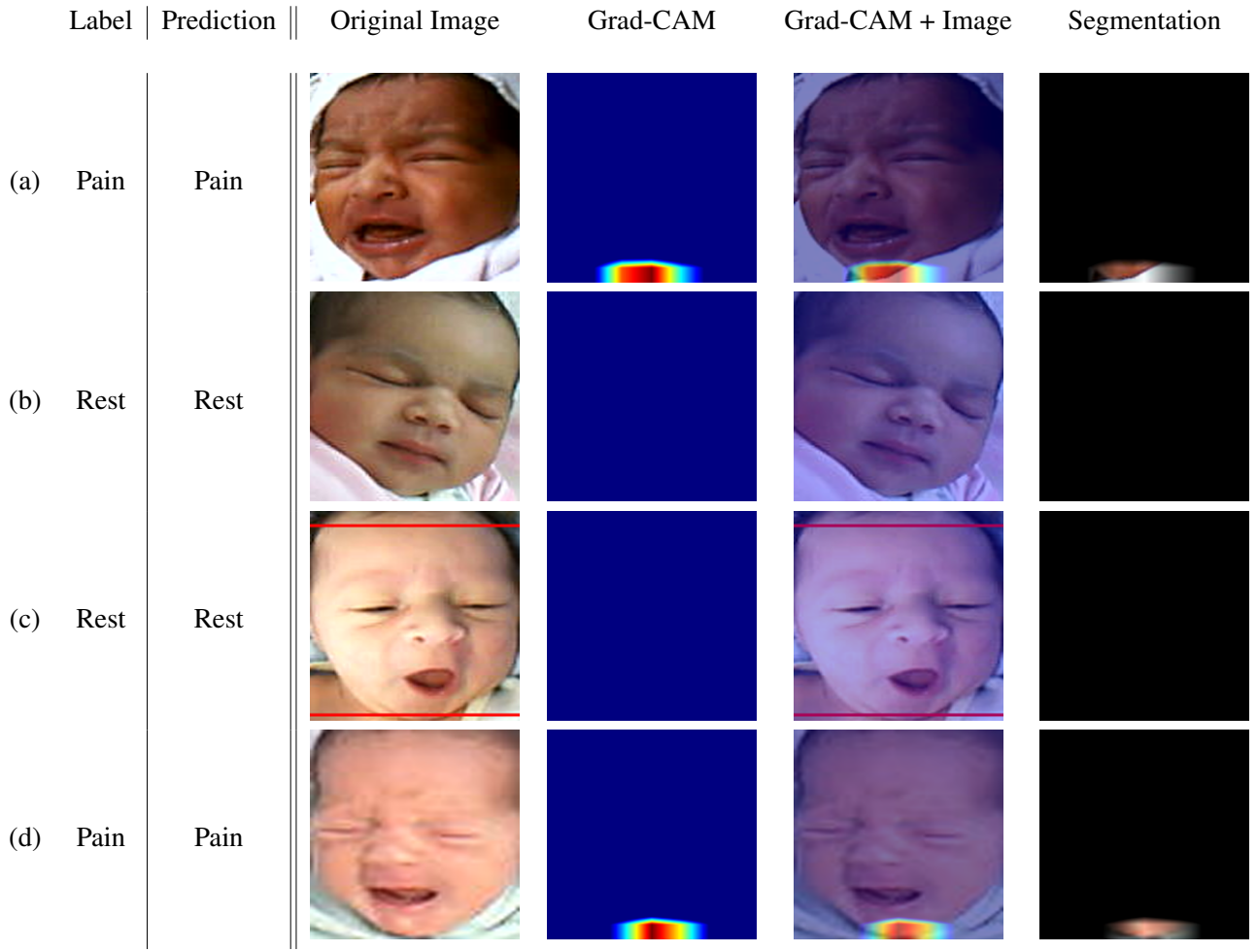
Os cálculos dos Grad-CAMs dos modelos ResNet50 treinados com o banco de imagens UNIFESP foram calculados utilizando as mesmas imagens do cálculo do Grad-CAM para o modelo N-CNN, como mostram as Figuras 25 à 28 *Original Image*. As imagens classificadas como verdadeiros positivos e verdadeiros negativos em ambos os modelos ResNet50 obtiveram regiões discriminantes no mesmo ponto da face neonatal, sendo essa região o queixo, mostrado nas Figuras 25a e 25d, e 26a e 26d. Entretanto, há diferença entre a região discriminada pela ResNet50 proposta por Zamzmi et al. (2019) e a ResNet50 proposta por esta dissertação. No modelo ResNet50 proposto por Zamzmi et al. (2019) a área discriminante é maior, fazendo o modelo discriminar além do queixo parte da roupa no recém-nascido. Já o modelo ResNet50 proposto por esta dissertação obteve uma área menor, discriminando apenas o queixo. Logo, tal área maior é devido a utilização de um único neurônio na camada de classificação, assim fazendo o modelo discriminar a dor do recém-nascido pelo queixo e a roupa, Figuras 25a e 25d. Então, utilizar a quantidade de neurônios correspondente ao número de classes é melhor, pois o modelo terá parâmetros mais específicos para cada classe.

Figura 24 – Resultado do Mapeamento de Ativação de Classe Ponderada por Gradiente (Grad-CAM) do modelo N-CNN treinado com o banco de imagens UNIFESP. Entretanto, esses resultados são dos falsos positivos e falsos negativos que o modelo classificou (classificação incorreta). As Figuras Grad-CAM (a-c) estão representando os locais de detecção dos estados "Dor" ou "Sem Dor" da Figura *Original Image* correspondente. As Figuras *Segmentation* mostram os locais descriminantes do estado correspondente à Figura *Original Image*.



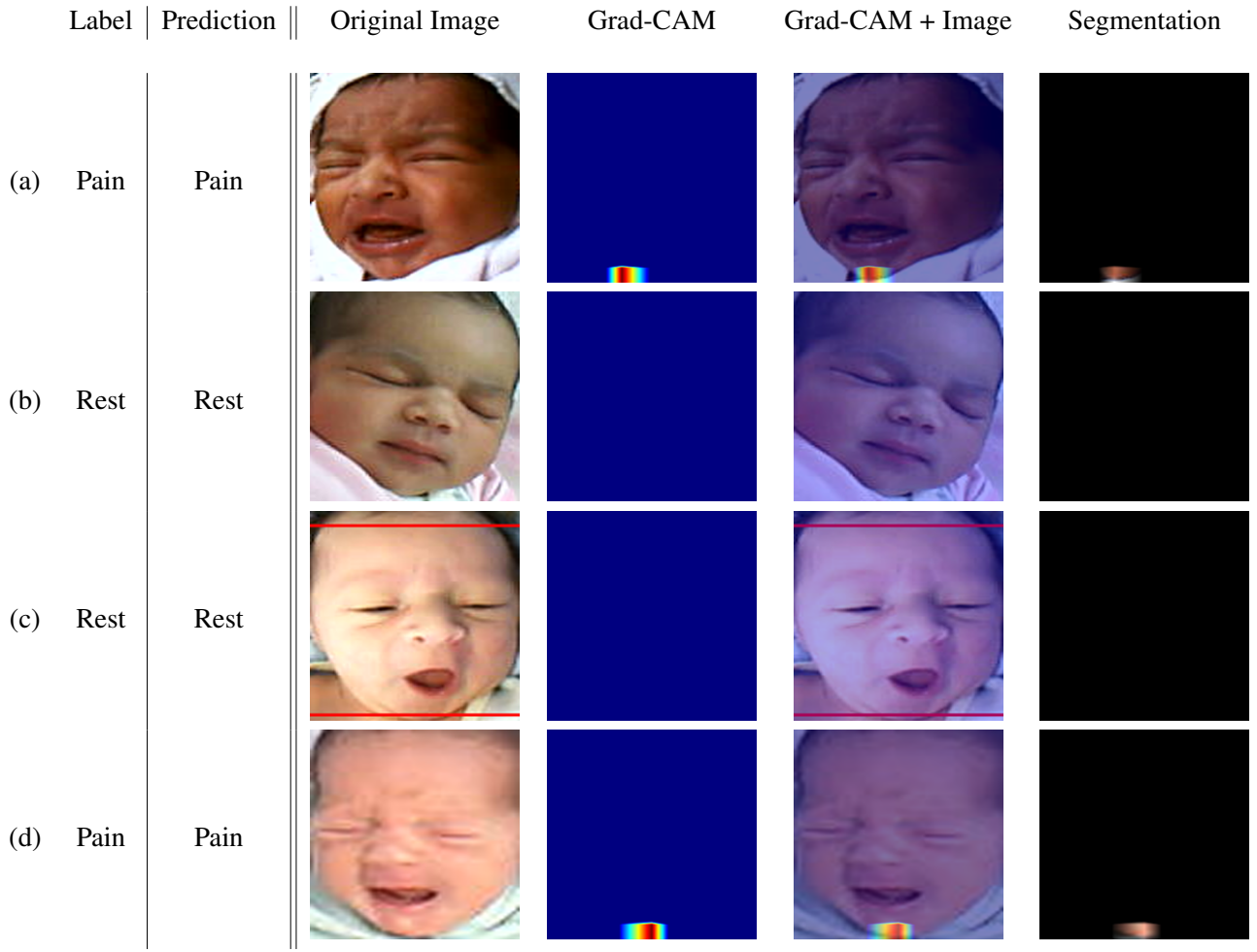
Fonte: Autor

Figura 25 – Resultado do Mapeamento de Ativação de Classe Ponderada por Gradiente (Grad-CAM) do modelo ResNet50 proposto por Zamzmi et al. (2019) treinado com o banco de imagens UNIFESP. As Figuras Grad-CAM (a-d) estão representando os locais de detecção dos estados “Dor” ou “Sem Dor” da Figura *Original Image* correspondente. As Figuras *Segmentation* mostram os locais descriminantes do estado correspondente à Figura *Original Image*.



Fonte: Autor

Figura 26 – Resultado do Mapeamento de Ativação de Classe Ponderada por Gradiente (Grad-CAM) do modelo ResNet50 proposto por esta dissertação treinado com o banco de imagens UNIFESP. As Figuras Grad-CAM (a-d) estão representando os locais de detecção dos estados “Dor” ou “Sem Dor” da Figura *Original Image* correspondente. As Figuras *Segmentation* mostram os locais descriminantes do estado correspondente à Figura *Original Image*.



Fonte: Autor

As Figuras 25b e 25c, e 26b e 26c, estão mostrando que ambos modelos ResNet50 não extraíram regiões discriminantes nessas faces de recém-nascidos. Tal resultado pode estar relacionado com a topologia, visto que a ResNet50 possui muita camadas e foi modelada para multi-classes. Logo, há perdas de regiões discriminantes dos estados “Dor” e “Sem Dor” nas camadas convolucionais superiores.

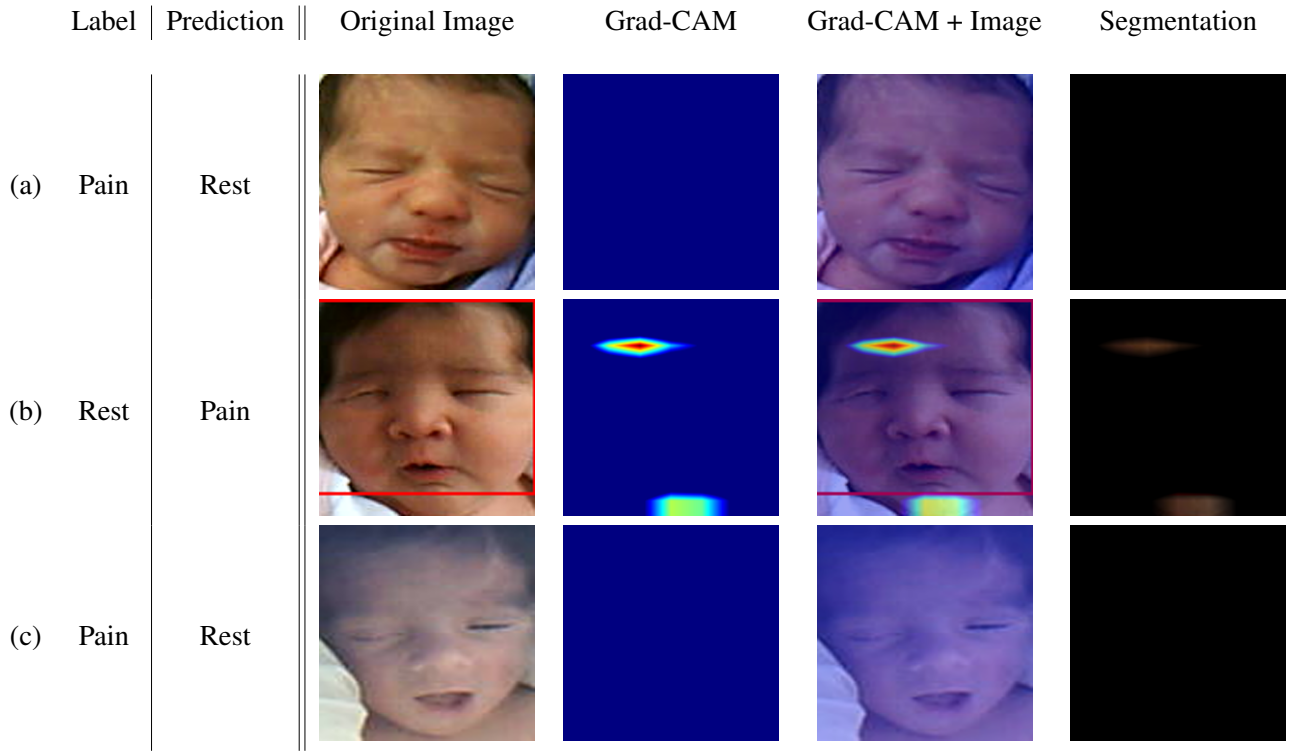
Os resultados dos cálculos dos Grad-CAMs das imagens falsos positivos e falsos negativos dos modelos ResNet50 treinados com o banco da UNIFESP, ilustrados nas Figuras 27 e 28, mostram que a ResNet50 proposta por esta dissertação não extraiu nas 3 imagens neonatais regiões que discriminam os estados “Dor” ‘ou “Sem Dor”. Das 3 faces neonatais analisadas na ResNet50 proposta por Zamzmi et al. (2019), constatou que o modelo extraiu regiões discriminantes em 1 imagem como mostra a Figura 27b, a região de maior ponderação foi a testa.

As Figuras 29 à 34 apresentam os resultados do cálculo do Grad-CAM dos modelos ResNet50 e N-CNN treinados com o banco de imagens COPE. Em todos os cálculos do Grad-CAM para imagens classificadas como verdadeiros positivos e verdadeiros negativos foram utilizadas as mesmas imagens da face do recém-nascido, totalizando 4 imagens de recém-nascido, 2 imagens com o estado “Dor” e 2 imagens com estado “Sem Dor”, como mostram as Figuras 29, 31 e 32. Para as imagens classificadas como falsos positivos e falsos negativos, Figuras 30, 33 e 34, não foi possível utilizar as mesmas imagens na análise devido ao desempenho de cada modelo, como é mostrado a Tabela 20.

Analisando o Grad-CAM das 4 imagens neonatais classificadas como verdadeiros positivos e verdadeiros negativos no modelo N-CNN treinado com o banco de imagens COPE, observa-se que o modelo extraiu regiões discriminantes nas 4 imagens, Figura 29. No entanto, as regiões discriminantes de maior ponderação estão localizadas mais ao redor da face nos neonatos. Na Figura 29a é perceptível que o modelo está classificando essa imagem pelo artefato que o neonato está envolto. Embora o modelo extraiu regiões discriminantes de maior ponderação ao redor da face nos neonatos nas Figuras 29b, 29c e 29d, também extraiu regiões discriminantes com maior ponderação na face, tais como: boca e bochechas, o que fez o modelo tomar uma decisão que levou a classificação correta. Entretanto, o modelo também está tomando uma decisão se é o estado “Dor” ou “Sem Dor” com artefatos que não correspondem a esses estados.

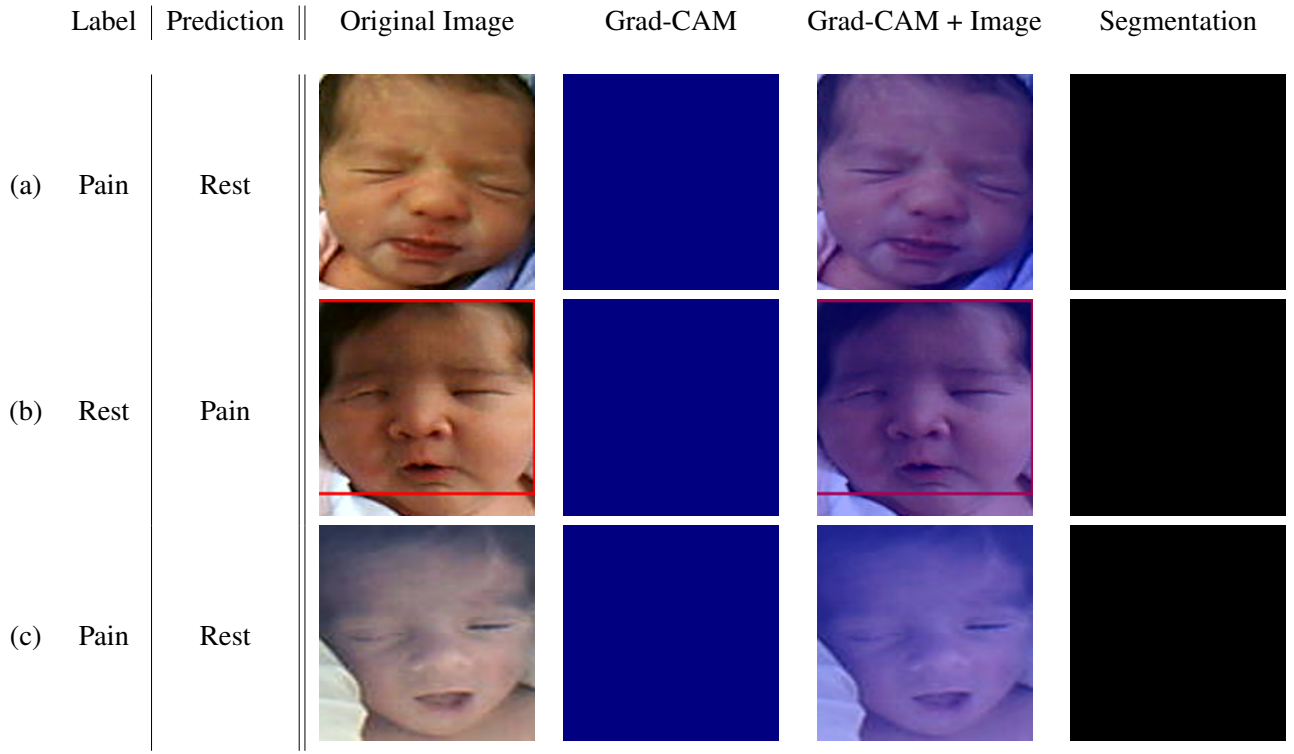
O modelo N-CNN treinado com o banco de imagens COPE não obteve falsos positivos, mas isso não significa que o modelo é bom ou o melhor, uma vez que levou a classificações de artefatos descrito no paragrafo anterior. Analisando os falsos negativos do modelo, observa-

Figura 27 – Resultado do Mapeamento de Ativação de Classe Ponderada por Gradiente (Grad-CAM) do modelo ResNet50 proposto por Zamzmi et al. (2019) treinado com o banco de imagens UNIFESP. Entretanto, esses resultados são dos falsos positivos e falsos negativos que o modelo classificou (classificação incorreta). As Figuras Grad-CAM (a-c) estão representando os locais de detecção dos estados "Dor" ou "Sem Dor" da Figura *Original Image* correspondente. As Figuras *Segmentation* mostram os locais descriminantes do estado correspondente à Figura *Original Image*.



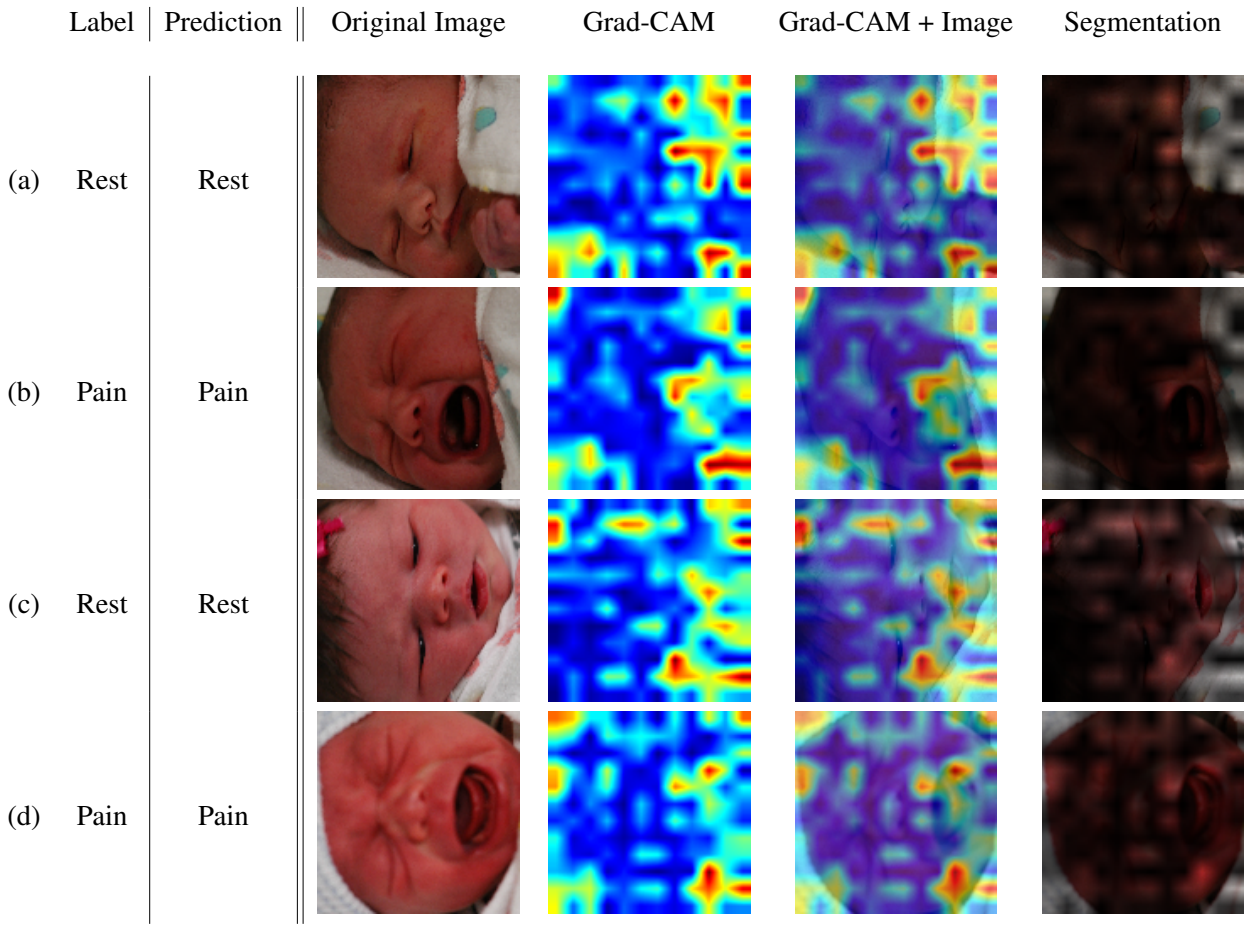
Fonte: Autor

Figura 28 – Resultado do Mapeamento de Ativação de Classe Ponderada por Gradiente (Grad-CAM) do modelo ResNet50 proposto por esta dissertação treinado com o banco de imagens UNIFESP. Entretanto, esses resultados são dos falsos positivos e falsos negativos que o modelo classificou (classificação incorreta). As Figuras Grad-CAM (a-c) estão representando os locais de detecção dos estados "Dor" ou "Sem Dor" da Figura *Original Image* correspondente. As Figuras *Segmentation* mostram os locais descriminantes do estado correspondente à Figura *Original Image*.



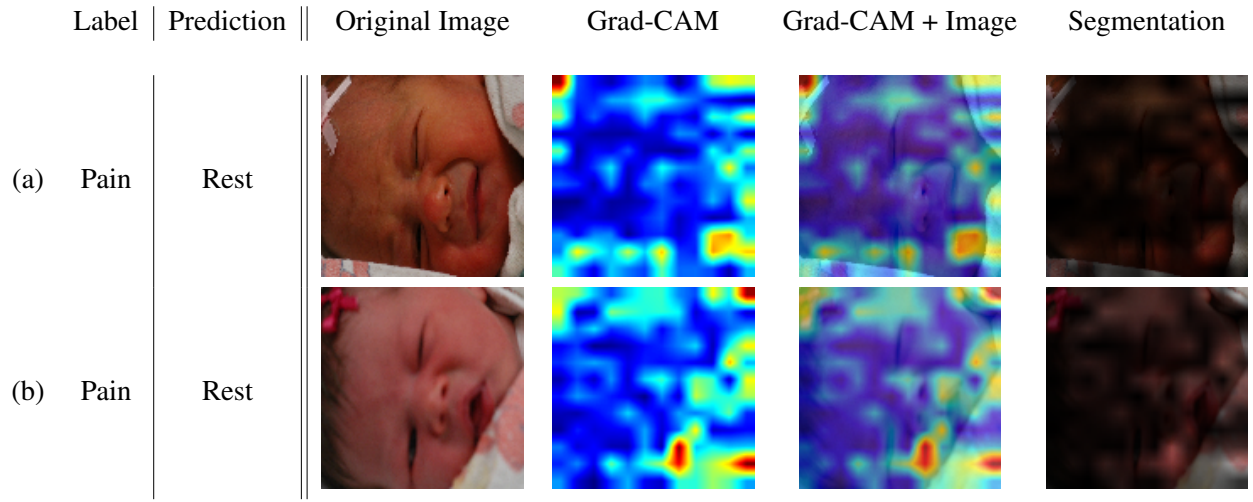
Fonte: Autor

Figura 29 – Resultado do Mapeamento de Ativação de Classe Ponderada por Gradiente (Grad-CAM) do modelo N-CNN treinado com o banco de imagens COPE. As Figuras Grad-CAM (a-d) estão representando os locais de detecção dos estados “Dor” ou “Sem Dor” da Figura *Original Image* correspondente. As Figuras *Segmentation* mostram os locais descriminantes do estado correspondente à Figura *Original Image*.



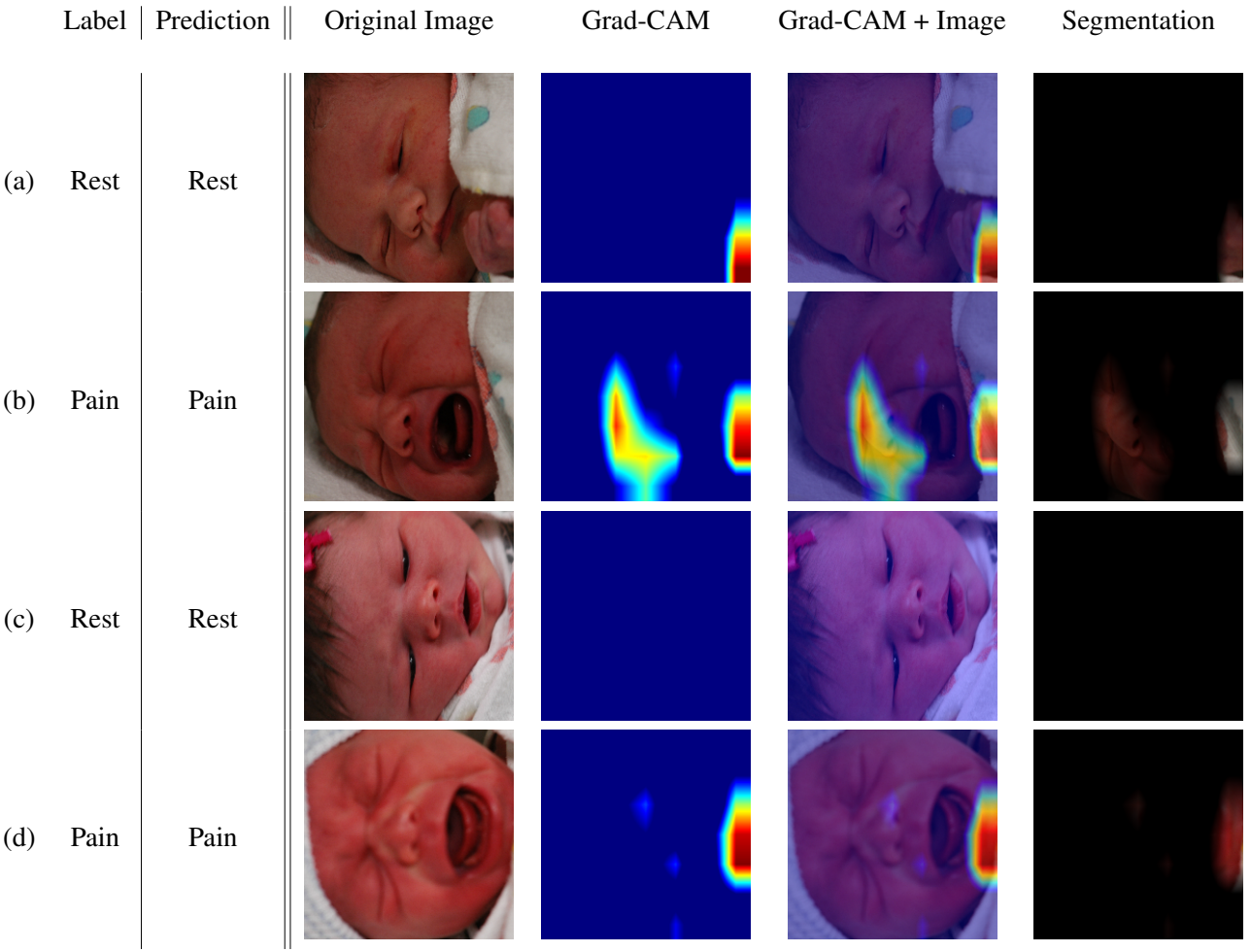
Fonte: Autor

Figura 30 – Resultado do Mapeamento de Ativação de Classe Ponderada por Gradiente (Grad-CAM) do modelo N-CNN treinado com o banco de imagens COPE. Entretanto, esses resultados são dos falsos positivos e falsos negativos que o modelo classificou (classificação incorreta). As Figuras Grad-CAM (a-b) estão representando os locais de detecção dos estados "Dor" ou "Sem Dor" da Figura *Original Image* correspondente. As Figuras *Segmentation* mostram os locais descriminantes do estado correspondente à Figura *Original Image*.



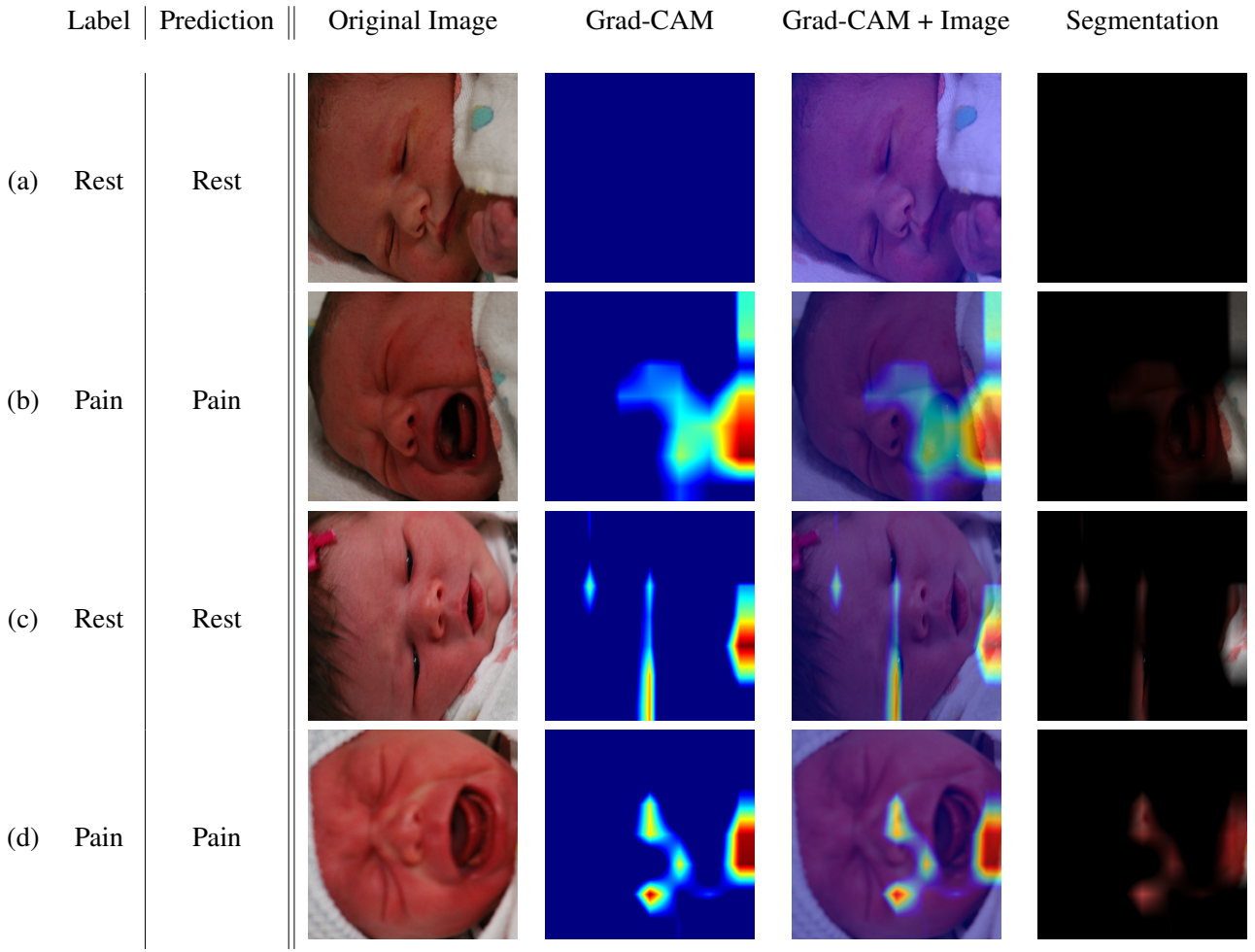
Fonte: Autor

Figura 31 – Resultado do Mapeamento de Ativação de Classe Ponderada por Gradiente (Grad-CAM) do modelo ResNet50 proposto por Zamzmi et al. (2019) treinado com o banco de imagens COPE. As Figuras Grad-CAM (a-d) estão representando os locais de detecção dos estados “Dor” ou “Sem Dor” da Figura *Original Image* correspondente. As Figuras *Segmentation* mostram os locais descriminantes do estado correspondente à Figura *Original Image*.



Fonte: Autor

Figura 32 – Resultado do Mapeamento de Ativação de Classe Ponderada por Gradiente (Grad-CAM) do modelo ResNet50 proposto por esta dissertação treinado com o banco de imagens COPE. As Figuras Grad-CAM (a-d) estão representando os locais de detecção dos estados “Dor” ou “Sem Dor” da Figura *Original Image* correspondente. As Figuras *Segmentation* mostram os locais descriminantes do estado correspondente à Figura *Original Image*.



Fonte: Autor

se extrações de regiões discriminantes com maior ponderação ao redor da face dos neonatos mostrados na Figura 30.

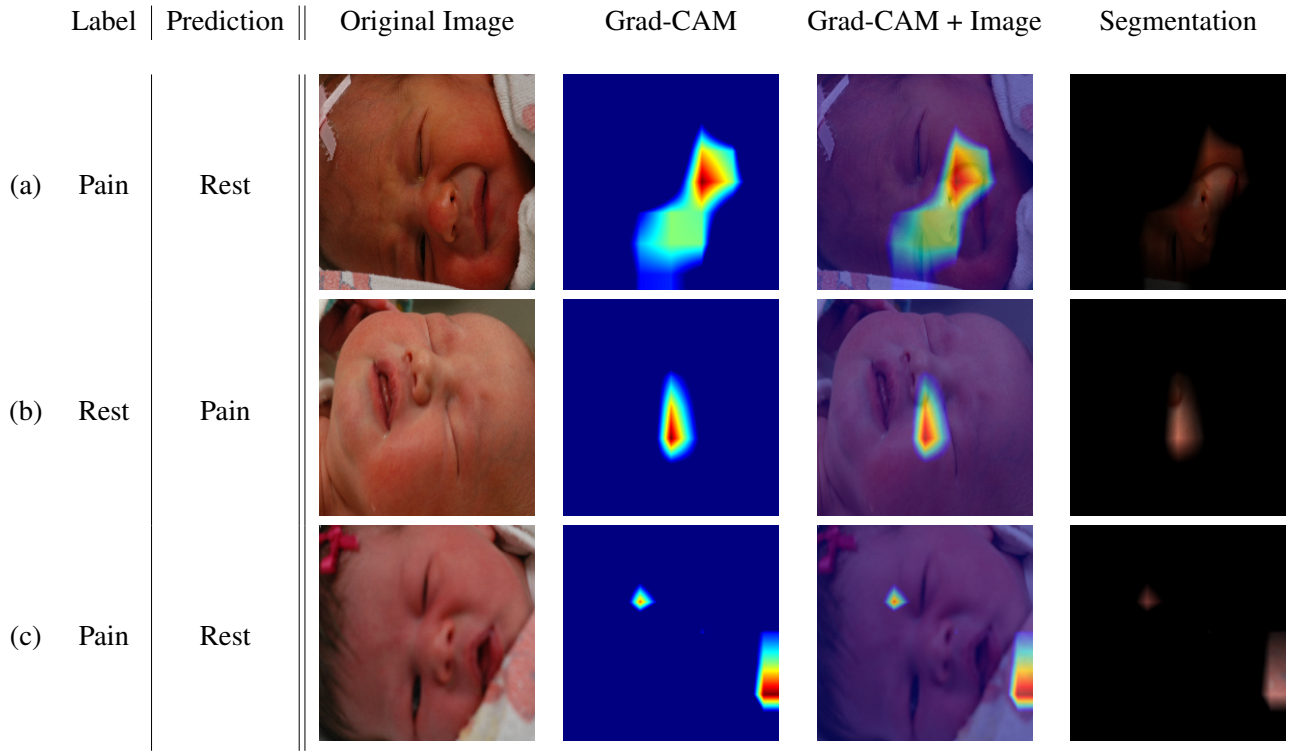
Ambos os modelos ResNet50 extraíram regiões discriminantes nas 4 imagens dos recém-nascidos classificadas como verdadeiros positivos e verdadeiros negativos, como mostram as Figuras 31 e 32. Entretanto, as Figuras 31a, 31b, 31c, 32a, 32b e 32c estão mostrando artefatos nas classificações, visto que as regiões discriminantes extraídas em ambos os modelos estão em locais que não identificam o estado "Dor" ou "Sem Dor", ou nem estão discriminando algo. Logo, pode-se considerar essas imagens classificadas com artefatos que não correspondem o estado "Dor" ou "Sem Dor'. Analisando as Figura 31d e 32d, observa-se que ambos os modelos extraíram regiões discriminantes em locais de interesse (face neonatal), entretanto, a extração de regiões discriminantes da ResNet50 proposta por esta dissertação foi mais ponderada, sendo a região do queixo, nariz e bochecha. A ResNet50 proposta por Zamzmi et al. (2019) discriminou a região do queixo. Dessa forma, tem-se a mesma evidência descrita na análise dos modelos ResNet50 treinados com o banco de imagens UNIFESP, sendo melhor utilizar a quantidade de neurônios correspondente ao número de classes, visto que o modelo terá parâmetros mais específicos para cada classe.

O cálculo do Grad-CAM das imagens classificadas como falsos positivos e falsos negativos do modelo ResNet50 proposto por Zamzmi et al. (2019) treinado com o banco de imagens COPE demostra que tal modelo extraiu regiões discriminantes na face do recém-nascido como mostra as Figuras 33a e 33b, entretanto, essas regiões discriminantes são insuficientes ou extrai parte de uma região de interesse levando o modelo a tomar uma decisão errada. Analisando a Figura 33a *Segmentation* observa-se parte da expressão facial discriminada que faz o modelo predizer que é o estado "Sem Dor". Na Figura 33c o modelo toma uma decisão através de uma região discriminante fora da região de interesse (face neonatal).

Analisando o Grad-CAM das imagens classificadas como falsos positivos e falsos negativos do modelo ResNet50 proposto por esta dissertação treinado com o banco da COPE, observa-se que o modelo não extraiu regiões discriminantes na face do recém-nascido, mas sim ao redor como cabeça e artefato envolto do recém-nascido como mostram as Figuras 34b e 34c. A Figura 34a demostra que o modelo não extraiu regiões discriminantes na imagem.

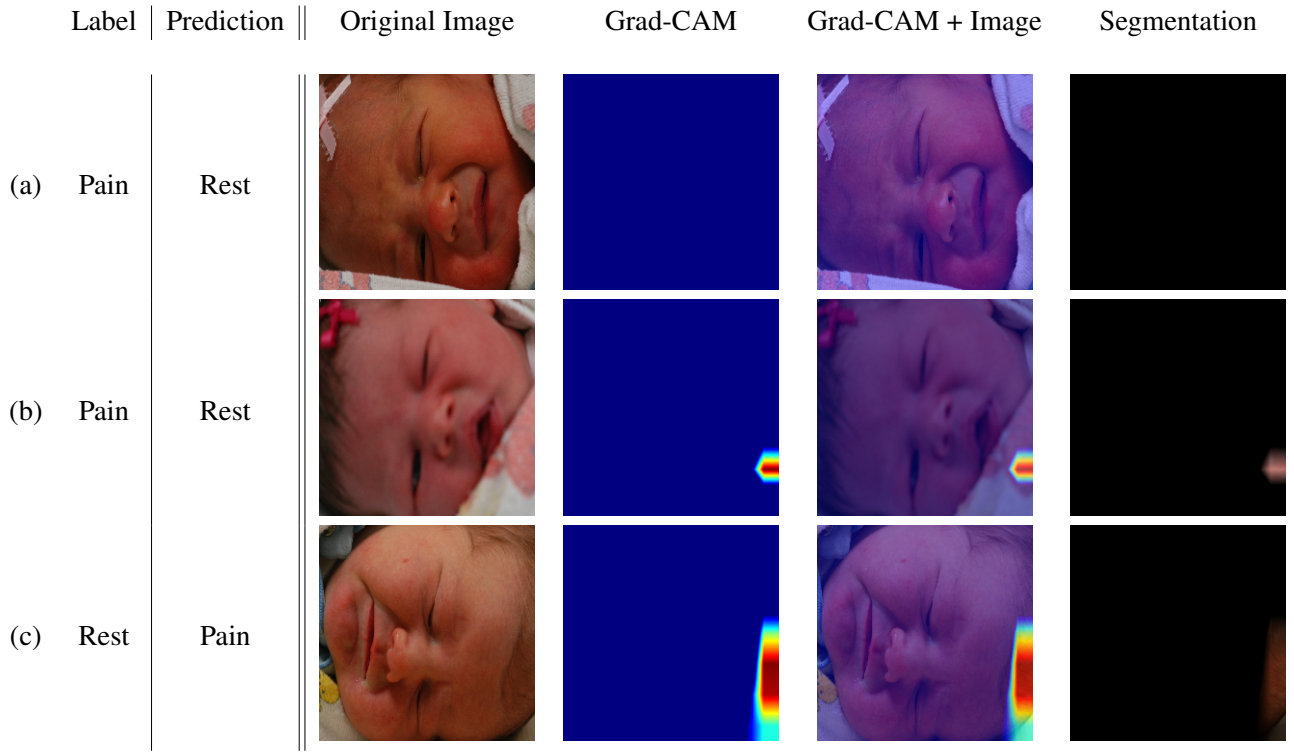
Analisando os resultados dos Grad-CAMs dos modelos em ambos os banco de imagens, conclui-se que o modelo N-CNN treinado com o banco de imagens UNIFESP foi o mais eficiente na extração de regiões discriminantes, extraindo regiões avaliadas pelos profissionais de saúde para o diagnostico clínico da dor neonatal.

Figura 33 – Resultado do Mapeamento de Ativação de Classe Ponderada por Gradiente (Grad-CAM) do modelo ResNet50 proposto por Zamzmi et al. (2019) treinado com o banco de imagens COPE. Entretanto, esses resultados são dos falsos positivos e falsos negativos que o modelo classificou (classificação incorreta). As Figuras Grad-CAM (a-c) estão representando os locais de detecção dos estados "Dor" ou "Sem Dor" da Figura *Original Image* correspondente. As Figuras *Segmentation* mostram os locais descriminantes do estado correspondentes à Figura *Original Image*.



Fonte: Autor

Figura 34 – Resultado do Mapeamento de Ativação de Classe Ponderada por Gradiente (Grad-CAM) do modelo ResNet50 proposto por esta dissertação treinado com o banco de imagens COPE. Entretanto, esses resultados são dos falsos positivos e falsos negativos que o modelo classificou (classificação incorreta). As Figuras Grad-CAM (a-c) estão representando os locais de detecção dos estados "Dor" ou "Sem Dor" da Figura *Original Image* correspondente. As Figuras *Segmentation* mostram os locais descriminantes do estado correspondente à Figura *Original Image*.



Fonte: Autor

5 CONCLUSÃO

Esta dissertação investigou métodos recentes de redes neurais para reconhecimento de padrões da dor neonatal por meio da expressão facial.

Os experimentos com duas bases distintas (UNIFESP e COPE) demostraram que os dois modelos ResNet50, proposto por Zamzmi et al. (2019) e por esta dissertação, são quantitativamente iguais em acurácia, indicando que não houve diferença entre utilizar um ou dois neurônios na camada de classificação. Entretanto, ao analisar os gráficos de convergência de aprendizado de ambos os modelos, observou-se claramente que o modelo ResNet50 proposto por esta dissertação precisou de menos épocas para convergir. Isso ocorreu devido ao número de neurônios, visto que o modelo ResNet50 proposto aqui permite atribuição de parâmetros distintos para cada classe de treinamento. Na analise qualitativa, o modelo ResNet50 proposto mostrou também ser ligeiramente mais discriminante nas características de dor ou sem dor do objeto de interesse (face do neonato) quando comparado com o modelo ResNet50 proposto por Zamzmi et al. (2019).

Entre os três modelos avaliados por esta dissertação, o modelo N-CNN proposto por Zamzmi et al. (2019) foi o melhor em ambos os bancos de imagens e mostrou bom desempenho na análise quantitativa de acurácia e na análise qualitativa. Interessante observar que esse modelo extraiu regiões discriminantes, dor ou sem 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 acurarias médias 87.2% e 78.7% para as bases COPE e UNIFESP, respectivamente. No entanto, a análise qualitativa evidenciou que todos os três modelos neurais avaliados, incluindo a arquitetura 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.

Algumas limitações que impactaram os desempenhos dos modelos neurais avaliados foram observadas nesta dissertação. Uma limitação 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 das imagens já existentes. Outra limitação enfrentada por esta dissertação foi o tempo de processamento para treinamento dos modelos neurais. Embora todos os modelos tenham sido

computados em um servidor dedicado, cada época de processamento teve duração média de aproximadamente 50 minutos, especialmente nas arquiteturas ResNet50.

Vislumbra-se, como trabalhos futuros: estender essas análises baseadas em modelos de Aprendizado Profundo para outros modelos, tal como a Rede Neural de Cápsula (CapsNet) (HINTON et al., 2018); 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 (QIN et al., 2018); e gerar imagens sintéticas a partir de bancos plurirraciais como a base da UNIFESP.

REFERÊNCIAS

ABDULAZIZ, Yousra; AHMAD, Sharrifah Mumtazah Syed. Infant cry recognition system: A comparison of system performance based on mel frequency and linear prediction cepstral coefficients. In: IEEE. 2010 International Conference on Information Retrieval & Knowledge Management (CAMP). [S.l.: s.n.], 2010. P. 260–263.

ANAND, K.J.; STEVENS, B.J.; MCGRATH, P.J. Pain in neonates and infants. **Elsevier Health Sciences**, 2007.

ANAND, Kanwaljeet JS; CARR, David B. The neuroanatomy, neurophysiology, and neurochemistry of pain, stress, and analgesia in newborns and children. **Pediatric Clinics of North America**, Elsevier, v. 36, n. 4, p. 795–822, 1989.

ANAND, Kanwaljeet JS; CRAIG, Kenneth D. New perspectives on the definition of pain. **Pain-Journal of the International Association for the Study of Pain**, [Amsterdam]: Elsevier/North-Holland, 1975-, v. 67, n. 1, p. 3–6, 1996.

ANAND, KJ. International Evidence-Based Group for Neonatal Pain Consensus statement for the prevention and management of pain in the newborn. **Arch Pediatr Adolesc Med**, v. 155, n. 2, p. 173–180, 2001.

ANAND, KJS; SCALZO, Frank M. Can adverse neonatal experiences alter brain development and subsequent behavior? **Neonatology**, Karger Publishers, v. 77, n. 2, p. 69–82, 2000.

ARIAS, Maria Carmenza Cuenca; GUINSBURG, Ruth. Differences between uni-and multidimensional scales for assessing pain in term newborn infants at the bedside. **Clinics**, SciELO Brasil, v. 67, n. 10, p. 1165–1170, 2012.

BALABAN, Stephen. Deep learning and face recognition: The state of the art. In: INTERNATIONAL SOCIETY FOR OPTICS e PHOTONICS. BIOMETRIC and Surveillance Technology for Human and Activity Identification XII. [S.l.]: SPIE, 2015. v. 9457, p. 68–75.

BARTOCCI, Marco et al. Pain activates cortical areas in the preterm newborn brain. **Pain**, Elsevier, v. 122, n. 1-2, p. 109–117, 2006.

BEICHEL, Reinhard et al. Robust active appearance models and their application to medical image analysis. **IEEE transactions on medical imaging**, IEEE, v. 24, n. 9, p. 1151–1169, 2005.

BENGIO, Yoshua; COURVILLE, Aaron; VINCENT, Pascal. Representation learning: A review and new perspectives. **IEEE transactions on pattern analysis and machine intelligence**, IEEE, v. 35, n. 8, p. 1798–1828, 2013.

BHUTTA, Adnan T; ANAND, KJS. Vulnerability of the developing brain: neuronal mechanisms. **Clinics in perinatology**, Elsevier, v. 29, n. 3, p. 357–372, 2002.

BISHOP, Christopher M et al. **Neural networks for pattern recognition**. [S.l.]: Oxford university press, 1995.

BRAHNAM, Sheryl; NANNI, Loris; SEXTON, Randall S. Neonatal Facial Pain Detection Using NNSOA and LSVM. In: IPCV. [S.l.: s.n.], 2008. P. 352–357.

BRAHNAM, Sheryl et al. Machine assessment of neonatal facial expressions of acute pain. **Decision Support Systems**, Elsevier, v. 43, n. 4, p. 1242–1254, 2007.

BRAHNAM, Sheryl et al. Machine recognition and representation of neonatal facial displays of acute pain. **Artificial intelligence in medicine**, Elsevier, v. 36, n. 3, p. 211–222, 2006.

BROWN, Justin E et al. Towards a physiology-based measure of pain: patterns of human brain activity distinguish painful from non-painful thermal stimulation. **PloS one**, Public Library of Science, v. 6, n. 9, 2011.

BRUMMELTE, Susanne et al. Procedural pain and brain development in premature newborns. **Annals of neurology**, Wiley Online Library, v. 71, n. 3, p. 385–396, 2012.

BUZUTI, Lucas; THOMAZ, Carlos. Understanding fully-connected and convolutional layers in unsupervised learning using face images. In: ANAIS do XV Workshop de Visão Computacional. São Bernardo do Campo: SBC, 2019. P. 13–18. Disponível em: <https://sol.sbc.org.br/index.php/wvc/article/view/7621>.

CHATFIELD, Ken et al. Return of the devil in the details: Delving deep into convolutional nets. **arXiv preprint arXiv:1405.3531**, 2014.

CHEON, Yeongjae; KIM, Daijin. Natural facial expression recognition using differential-AAM and manifold learning. **Pattern Recognition**, Elsevier, v. 42, n. 7, p. 1340–1350, 2009.

COMMITTEE, ON FETUS et al. Prevention and Management of Procedural Pain in the Neonate: An Update. **Pediatrics**, v. 137, n. 2, e20154271, 2016.

CORREIA, Miguel V; CAMPILHO, Aurélio C. Real-time implementation of an optical flow algorithm. In: IEEE. OBJECT recognition supported by user interaction for service robots. [S.l.: s.n.], 2002. v. 4, p. 247–250.

DA SILVA, Tiago Pereira; DA SILVA, Lincoln Justo. Pain scales used in the newborn infant: a systematic review. **Acta medica portuguesa**, v. 23, n. 3, p. 437–54, 2010.

DENG, Jiankang et al. Retinaface: Single-stage dense face localisation in the wild. **arXiv preprint arXiv:1905.00641**, 2019.

DILORENZO, Miranda; PILLAI RIDDELL, Rebecca; HOLSTI, Liisa. Beyond acute pain: understanding chronic pain in infancy. **Children**, Multidisciplinary Digital Publishing Institute, v. 3, n. 4, p. 26, 2016.

DING, Hui; ZHOU, Shaohua Kevin; CHELLAPPA, Rama. Facenet2expnet: Regularizing a deep face recognition net for expression recognition. In: IEEE. 2017 12th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017). [S.l.: s.n.], 2017. P. 118–126.

EDWARDS, Gareth J; COOTES, Timothy F; TAYLOR, Christopher J. Face recognition using active appearance models. In: SPRINGER. EUROPEAN conference on computer vision. [S.l.: s.n.], 1998. P. 581–595.

EKMAN, Rosenberg. **What the face reveals: Basic and applied studies of spontaneous expression using the Facial Action Coding System (FACS)**. [S.l.]: Oxford University Press, USA, 1997.

FAYE, Papa M et al. Newborn infant pain assessment using heart rate variability analysis. **The Clinical journal of pain**, LWW, v. 26, n. 9, p. 777–782, 2010.

FOTIADOU, Eleni et al. Video-based facial discomfort analysis for infants. In: INTERNATIONAL SOCIETY FOR OPTICS e PHOTONICS. VISUAL Information Processing and Communication V. [S.l.: s.n.], 2014. v. 9029, 90290f.

FULLER, Barbara F; HORII, Yoshiyuki. Spectral energy distribution in four types of infant vocalizations. **Journal of Communication Disorders**, Elsevier, v. 21, n. 3, p. 251–261, 1988.

GHOLAMI, Behnood; HADDAD, Wassim M; TANNENBAUM, Allen R. Relevance vector machine learning for neonate pain intensity assessment using digital imaging. **IEEE Transactions on biomedical engineering**, IEEE, v. 57, n. 6, p. 1457–1466, 2010.

GOLIANU, Brenda et al. Pediatric acute pain management. **Pediatric Clinics of North America**, Elsevier, v. 47, n. 3, p. 559–587, 2000.

GOODFELLOW, Ian; BENGIO, Yoshua; COURVILLE, Aaron. **Deep learning**. [S.l.]: MIT press, 2016.

GRUNAU, RE. Self-regulation and behavior in preterm children: effects of early pain. **Progress in pain research and management**, Seattle; IASP Press; 1999, v. 26, p. 23–56, 2003.

GRUNAU, Ruth E; WEINBERG, Joanne; WHITFIELD, Michael F. Neonatal procedural pain and preterm infant cortisol response to novelty at 8 months. **Pediatrics**, Am Acad Pediatrics, v. 114, n. 1, e77–e84, 2004.

GRUNAU, Ruth E et al. Cortisol, behavior, and heart rate reactivity to immunization pain at 4 months corrected age in infants born very preterm. **The Clinical journal of pain**, NIH Public Access, v. 26, n. 8, p. 698, 2010.

GRUNAU, Ruth VE; CRAIG, Kenneth D. Pain expression in neonates: facial action and cry. **Pain**, Elsevier, v. 28, n. 3, p. 395–410, 1987.

GRUSS, Sascha et al. Pain intensity recognition rates via biopotential feature patterns with support vector machines. **PloS one**, Public Library of Science, v. 10, n. 10, 2015.

GUINSBURG, Ruth. Avaliação e tratamento da dor no recém-nascido. **J Pediatr (Rio J)**, v. 75, n. 3, p. 149–60, 1999.

GUINSBURG, Ruth; CUENCA, Maria Carmenza. A linguagem da dor no recém-nascido. **São Paulo: Sociedade Brasileira de Pediatria.[Internet]**, 2010.

HE, Kaiming et al. Deep residual learning for image recognition. In: PROCEEDINGS of the IEEE conference on computer vision and pattern recognition. [S.l.: s.n.], 2016. P. 770–778.

HEIDERICH, Tatiany Marcondes. Desenvolvimento de software para identificar a expressão facial de dor do recém-nascido. Universidade Federal de São Paulo (UNIFESP), 2013.

HEIDERICH, Tatiany Marcondes; LESLIE, Ana Teresa Figueiredo Stochero; GUINSBURG, Ruth. Neonatal procedural pain can be assessed by computer software that has good sensitivity and specificity to detect facial movements. **Acta Paediatrica**, Wiley Online Library, v. 104, n. 2, e63–e69, 2015.

HINTON, Geoffrey E; SABOUR, Sara; FROSST, Nicholas. Matrix capsules with EM routing. In: INTERNATIONAL Conference on Learning Representations. [S.l.: s.n.], 2018. Disponível em: <https://openreview.net/forum?id=HJWLfGWRb>.

HINTON, Geoffrey E et al. Does the brain do inverse graphics? In: BRAIN and Cognitive Sciences Fall Colloquium. [S.l.: s.n.], 2012. v. 2.

HUMMEL, P et al. Clinical reliability and validity of the N-PASS: neonatal pain, agitation and sedation scale with prolonged pain. **Journal of perinatology**, Nature Publishing Group, v. 28, n. 1, p. 55–60, 2008.

HUMMEL, Pat; DIJK, Monique van. Pain assessment: current status and challenges. In: ELSEVIER, 4. SEMINARS in Fetal and Neonatal medicine. [S.l.: s.n.], 2006. v. 11, p. 237–245.

JENI, László A; COHN, Jeffrey F; KANADE, Takeo. Dense 3D face alignment from 2D videos in real-time. In: IEEE. 2015 11th IEEE international conference and workshops on automatic face and gesture recognition (FG). [S.l.: s.n.], 2015. v. 1, p. 1–8.

JIN, Hongliang et al. Face detection using improved LBP under Bayesian framework. In: IEEE. THIRD International Conference on Image and Graphics (ICIG'04). [S.l.: s.n.], 2004. P. 306–309.

KÄCHELE, Markus et al. Multimodal data fusion for person-independent, continuous estimation of pain intensity. In: SPRINGER. INTERNATIONAL Conference on Engineering Applications of Neural Networks. [S.l.: s.n.], 2015. P. 275–285.

KARRAS, Tero; LAINE, Samuli; AILA, Timo. A style-based generator architecture for generative adversarial networks. In: PROCEEDINGS of the IEEE Conference on Computer Vision and Pattern Recognition. [S.l.: s.n.], 2019. P. 4401–4410.

KHAN, Asifullah et al. A survey of the recent architectures of deep convolutional neural networks. **arXiv preprint arXiv:1901.06032**, 2019.

KRECHEL, Susan W; BILDNER, JUDY. CRIES: a new neonatal postoperative pain measurement score. Initial testing of validity and reliability. **Pediatric Anesthesia**, Wiley Online Library, v. 5, n. 1, p. 53–61, 1995.

KRIZHEVSKY, Alex; SUTSKEVER, Ilya; HINTON, Geoffrey E. Imagenet classification with deep convolutional neural networks. In: ADVANCES in neural information processing systems. [S.l.: s.n.], 2012. P. 1097–1105.

LANATÀ, Antonio et al. Eye tracking and pupil size variation as response to affective stimuli: a preliminary study. In: IEEE. 2011 5th International Conference on Pervasive Computing Technologies for Healthcare (PervasiveHealth) and Workshops. [S.l.: s.n.], 2011. P. 78–84.

LAWRENCE, Jocelyn et al. The development of a tool to assess neonatal pain. **Neonatal network: NN**, v. 12, n. 6, p. 59–66, 1993.

LIAO, Shu; CHUNG, Albert CS. Face recognition by using elongated local binary patterns with average maximum distance gradient magnitude. In: SPRINGER. ASIAN conference on computer vision. [S.l.: s.n.], 2007. P. 672–679.

LIN, Tsung-Yi et al. Microsoft coco: Common objects in context. In: SPRINGER. EUROPEAN conference on computer vision. [S.l.: s.n.], 2014. P. 740–755.

LINDH, Viveca; WIKLUND, Urban; HÅKANSSON, Stellan. Heel lancing in term new-born infants: an evaluation of pain by frequency domain analysis of heart rate variability. **Pain**, Elsevier, v. 80, n. 1-2, p. 143–148, 1999.

LITTLEWORT, Gwen C; BARTLETT, Marian Stewart; LEE, Kang. Automatic coding of facial expressions displayed during posed and genuine pain. **Image and Vision Computing**, Elsevier, v. 27, n. 12, p. 1797–1803, 2009.

LU, Guanming et al. Sparse representation based facial expression classification for pain assessment in neonates. In: IEEE. 2016 12th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD). [S.l.: s.n.], 2016. P. 1615–1619.

LUCEY, Patrick et al. Automatically detecting pain in video through facial action units. **IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)**, IEEE, v. 41, n. 3, p. 664–674, 2010.

MAHENDRAN, Aravindh; VEDALDI, Andrea. Visualizing deep convolutional neural networks using natural pre-images. **International Journal of Computer Vision**, Springer, v. 120, n. 3, p. 233–255, 2016.

MARCHANT, Amy. ‘Neonates do not feel pain’: a critical review of the evidence. **Bioscience Horizons: The International Journal of Student Research**, Narnia, v. 7, 2014.

MARTINEZ, D Lopez et al. Automatic detection of nociceptive stimuli and pain intensity from facial expressions. **The Journal of Pain**, Elsevier, v. 18, n. 4, s59, 2017.

MELO, Gleicia Martins de et al. Escalas de avaliação de dor em recém-nascidos: revisão integrativa. **Revista Paulista de Pediatria**, SciELO Brasil, v. 32, n. 4, p. 395–402, 2014.

MILLER, Carly; NEWTON, Sarah E. Pain perception and expression: the influence of gender, personal self-efficacy, and lifespan socialization. **Pain Management Nursing**, Elsevier, v. 7, n. 4, p. 148–152, 2006.

MONTIEL, Sandra E. B.; GARCIA, Carlos A. R. Fuzzy support vector machines for automatic infant cry recognition. In: INTELLIGENT Computing in Signal Processing and Pattern Recognition. [S.l.]: Springer, 2006. P. 876–881.

NANNI, Loris; BRAHNAM, Sheryl; LUMINI, Alessandra. A local approach based on a local binary patterns variant texture descriptor for classifying pain states. **Expert Systems with Applications**, Elsevier, v. 37, n. 12, p. 7888–7894, 2010.

OJALA, Timo; PIETIKAINEN, Matti; MAENPAA, Topi. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. **IEEE Transactions on pattern analysis and machine intelligence**, IEEE, v. 24, n. 7, p. 971–987, 2002.

OKADA, Massako et al. Dor em pediatria. **Revista de Medicina**, v. 80, p. 135–156, 2001.

AL-OMAR, Dania; AL-WABIL, Areej; FAWZI, Manar. Using pupil size variation during visual emotional stimulation in measuring affective states of non communicative individuals. In: SPRINGER. INTERNATIONAL Conference on Universal Access in Human-Computer Interaction. [S.l.: s.n.], 2013. P. 253–258.

ORONA, Pedro et al. Atlas of neonatal face images using triangular Meshes. In: ANAIS do XV Workshop de Visão Computacional. São Bernardo do Campo: SBC, 2019. P. 19–24. Disponível em: <https://sol.sbc.org.br/index.php/wvc/article/view/7622>.

OSTER, Harriet. Baby FACS: Facial Action Coding System for infants and young children. **Unpublished monograph and coding manual. New York University**, 2006.

PAI, Chih-Yun. **Automatic pain assessment from infants’ crying sounds**. 2016. Diss. (Mestrado) – University of South Florida.

PAL, Pritam; IYER, Ananth N; YANTORNO, Robert E. Emotion detection from infant facial expressions and cries. In: IEEE. 2006 IEEE International Conference on Acoustics Speech and Signal Processing Proceedings. [S.l.: s.n.], 2006. v. 2, p. ii–ii.

PARTALA, Timo; JOKINIEMI, Maria; SURAKKA, Veikko. Pupillary responses to emotionally provocative stimuli. In: PROCEEDINGS of the 2000 symposium on Eye tracking research & applications. [S.l.: s.n.], 2000. P. 123–129.

PARTALA, Timo; SURAKKA, Veikko. Pupil size variation as an indication of affective processing. **International journal of human-computer studies**, Elsevier, v. 59, n. 1-2, p. 185–198, 2003.

PASERO, Chris; MCCAFFERY, Margo. **Pain: clinical manual**. [S.l.]: Mosby St. Louis, 1999.

PETRONI, Marco et al. Identification of pain from infant cry vocalizations using artificial neural networks (ANNs). In: INTERNATIONAL SOCIETY FOR OPTICS e PHOTONICS. APPLICATIONS and Science of Artificial Neural Networks. [S.l.: s.n.], 1995. v. 2492, p. 729–738.

PILLAI RIDDELL, Rebecca R; BADALI, Melanie A; CRAIG, Kenneth D. Parental judgments of infant pain: Importance of perceived cognitive abilities, behavioural cues and contextual cues. **Pain Research and Management**, Hindawi, v. 9, n. 2, p. 73–80, 2004.

PILLAI RIDDELL, Rebecca R; CRAIG, Kenneth D. Judgments of infant pain: The impact of caregiver identity and infant age. **Journal of pediatric psychology**, Oxford University Press, v. 32, n. 5, p. 501–511, 2007.

QIN, Zhuwei et al. How convolutional neural network see the world-A survey of convolutional neural network visualization methods. **Mathematical Foundations of Computing**, v. 1, p. 149, 2018.

RANGER, Manon; GÉLINAS, Céline. Innovating in pain assessment of the critically ill: exploring cerebral near-infrared spectroscopy as a bedside approach. **Pain Management Nursing**, Elsevier, v. 15, n. 2, p. 519–529, 2014.

RANGER, Manon et al. A multidimensional approach to pain assessment in critically ill infants during a painful procedure. **The Clinical journal of pain**, NIH Public Access, v. 29, n. 7, p. 613, 2013.

RODRIGUEZ, Pau et al. Deep pain: Exploiting long short-term memory networks for facial expression classification. **IEEE transactions on cybernetics**, IEEE, 2017.

SALEKIN, Md Sirajus et al. Multi-Channel Neural Network for Assessing Neonatal Pain from Videos. In: IEEE. 2019 IEEE International Conference on Systems, Man and Cybernetics (SMC). [S.l.: s.n.], 2019. P. 1551–1556.

SAMOLSKY DEKEL, Boaz Gedaliahu et al. Medical evidence influence on inpatients and nurses pain ratings agreement. **Pain Research and Management**, Hindawi, v. 2016, 2016.

SCHIAVENATO, Martin et al. Neonatal pain facial expression: Evaluating the primal face of pain. **Pain**, Elsevier, v. 138, n. 2, p. 460–471, 2008.

SCOPEL, Evânea; ALENCAR, Márcia; CRUZ, Roberto Moraes. Medidas de avaliação da dor. **Lecturas: Educación física y deportes**, Tulio Guterman, n. 105, p. 34, 2007.

SELVARAJU, Ramprasaath R et al. Grad-cam: Visual explanations from deep networks via gradient-based localization. In: PROCEEDINGS of the IEEE international conference on computer vision. [S.l.: s.n.], 2017. P. 618–626.

SIKKA, Karan et al. Automated assessment of children's postoperative pain using computer vision. **Pediatrics**, Am Acad Pediatrics, v. 136, n. 1, e124–e131, 2015.

SILVA, Pedro Augusto S. O. **Interpretação e reconhecimento de padrões para avaliação de dor em imagens faciais de recém-nascidas**. 2020. Diss. (Mestrado) – Centro Universitário FEI.

SIMONS, Sinno HP et al. Do we still hurt newborn babies?: A prospective study of procedural pain and analgesia in neonates. **Archives of pediatrics & adolescent medicine**, American Medical Association, v. 157, n. 11, p. 1058–1064, 2003.

SIMONYAN, Karen; ZISSERMAN, Andrew. Very deep convolutional networks for large-scale image recognition. **arXiv preprint arXiv:1409.1556**, 2014.

SLATER, Rebeccah et al. Cortical pain responses in human infants. **Journal of Neuroscience**, Soc Neuroscience, v. 26, n. 14, p. 3662–3666, 2006.

SOMOL, Petr; NOVOVICOVÁ, Jana; PUDIL, Pavel. Efficient feature subset selection and subset size optimization. **Pattern recognition recent advances**, IntechOpen, p. 1–24, 2010.

SONG, Yang; ERMON, Stefano. Generative modeling by estimating gradients of the data distribution. In: ADVANCES in Neural Information Processing Systems. [S.l.: s.n.], 2019. P. 11895–11907.

SRIVASTAVA, Rupesh Kumar; GREFF, Klaus; SCHMIDHUBER, Jürgen. Highway networks. **arXiv preprint arXiv:1505.00387**, 2015.

STEVENS, Bonnie et al. Premature infant pain profile: development and initial validation. **The Clinical journal of pain**, LWW, v. 12, n. 1, p. 13–22, 1996.

TAIGMAN, Yaniv et al. Deepface: Closing the gap to human-level performance in face verification. In: PROCEEDINGS of the IEEE conference on computer vision and pattern recognition. [S.l.: s.n.], 2014. P. 1701–1708.

TAN, Xiaoyang; TRIGGS, Bill. Enhanced local texture feature sets for face recognition under difficult lighting conditions. **IEEE transactions on image processing**, IEEE, v. 19, n. 6, p. 1635–1650, 2010.

TERUEL, Gilberto et al. Análise e Reconhecimento de Dor em Imagens 2D Frontais de Recém-Nascidos a Termo e Saudáveis. In: ANAIS Estendidos do XIX Simpósio Brasileiro de Computação Aplicada à Saúde. Niterói: SBC, 2019. P. 97–102. Disponível em: <https://sol.sbc.org.br/index.php/sbcas_estendido/article/view/6291>.

THOMAZ, Carlos E et al. A multivariate statistical analysis of the developing human brain in preterm infants. **Image and Vision Computing**, Elsevier, v. 25, n. 6, p. 981–994, 2007.

TIELEMAN, Tijmen; HINTON, Geoffrey. Lecture 6.5-rmsprop: Divide the gradient by a running average of its recent magnitude. **COURSERA: Neural networks for machine learning**, v. 4, n. 2, p. 26–31, 2012.

VARALLYAY, G Jr et al. Acoustic analysis of the infant cry: classical and new methods. In: IEEE. THE 26th Annual International Conference of the IEEE Engineering in Medicine and Biology Society. [S.l.: s.n.], 2004. v. 1, p. 313–316.

VELANA, Maria et al. The senseemotion database: A multimodal database for the development and systematic validation of an automatic pain-and emotion-recognition system. In: SPRINGER. IAPR Workshop on Multimodal Pattern Recognition of Social Signals in Human-Computer Interaction. [S.l.: s.n.], 2016. P. 127–139.

VEMPADA, Ramu Reddy; KUMAR, B Siva Ayyappa; RAO, K Sreenivasa. Characterization of infant cries using spectral and prosodic features. In: IEEE. 2012 National Conference on Communications (NCC). [S.l.: s.n.], 2012. P. 1–5.

VINALL, Jillian et al. Neonatal pain in relation to postnatal growth in infants born very preterm. **Pain**, Elsevier, v. 153, n. 7, p. 1374–1381, 2012.

VIOLA, Paul; JONES, Michael J. Robust real-time face detection. **International journal of computer vision**, Springer, v. 57, n. 2, p. 137–154, 2004.

WALKER, Suellen M. Translational studies identify long-term impact of prior neonatal pain experience. **Pain**, LWW, v. 158, s29–s42, 2017.

WALTER, Steffen et al. The biovid heat pain database data for the advancement and systematic validation of an automated pain recognition system. In: IEEE. 2013 IEEE international conference on cybernetics (CYBCO). [S.l.: s.n.], 2013. P. 128–131.

WERNER, Philipp; AL-HAMADI, Ayoub; NIESE, Robert. Comparative learning applied to intensity rating of facial expressions of pain. **International Journal of Pattern Recognition and Artificial Intelligence**, World Scientific, v. 28, n. 05, p. 1451008, 2014.

ZAMZMI, Ghada. **Automatic Multimodal Assessment of Neonatal Pain**. 2018a. Ph.D. dissertation – University of South Florida.

ZAMZMI, Ghada et al. A review of automated pain assessment in infants: Features, classification tasks, and databases. **IEEE reviews in biomedical engineering**, IEEE, v. 11, p. 77–96, 2017.

ZAMZMI, Ghada et al. An approach for automated multimodal analysis of infants' pain. In: IEEE. 2016 23rd International Conference on Pattern Recognition (ICPR). [S.l.: s.n.], 2016. P. 4148–4153.

ZAMZMI, Ghada et al. Neonatal pain expression recognition using transfer learning. **arXiv preprint arXiv:1807.01631**, 2018b.

ZAMZMI, Ghada et al. Pain assessment from facial expression: Neonatal convolutional neural network (N-CNN). In: IEEE. 2019 International Joint Conference on Neural Networks (IJCNN). [S.l.: s.n.], 2019. P. 1–7.

ZAMZMI, Ghada et al. Pain assessment in infants: Towards spotting pain expression based on infants' facial strain. In: IEEE. 2015 11th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG). [S.l.: s.n.], 2015. v. 5, p. 1–5.

ZEILER, Matthew D; FERGUS, Rob. Visualizing and understanding convolutional networks. In: SPRINGER. EUROPEAN conference on computer vision. [S.l.: s.n.], 2014. P. 818–833.

ZHANG, Feifei et al. Joint pose and expression modeling for facial expression recognition. In: PROCEEDINGS of the IEEE Conference on Computer Vision and Pattern Recognition. [S.l.: s.n.], 2018. P. 3359–3368.

ZWICKER, Jill G et al. Smaller cerebellar growth and poorer neurodevelopmental outcomes in very preterm infants exposed to neonatal morphine. **The Journal of pediatrics**, Elsevier, v. 172, p. 81–87, 2016.