

**CENTRO UNIVERSITÁRIO FEI
TATIANY MARCONDES HEIDERICH**

**RECONHECIMENTO AUTOMATIZADO DA DOR POR MOVIMENTOS FACIAIS
DE RECÉM-NASCIDOS INTERNADOS EM UNIDADE DE
TERAPIA INTENSIVA NEONATAL**

São Bernardo do Campo

2024

TATIANY MARCONDES HEIDERICH

**RECONHECIMENTO AUTOMATIZADO DA DOR POR MOVIMENTOS FACIAIS
DE RECÉM-NASCIDOS INTERNADOS EM UNIDADE DE
TERAPIA INTENSIVA NEONATAL**

Tese de Doutorado apresentada ao
Centro Universitário FEI para obtenção
do Título de Doutora em Engenharia
Elétrica. Orientada pelo Prof. Dr. Carlos
Eduardo Thomaz

São Bernardo do Campo

2024

Heiderich, Tatiany Marcondes

Reconhecimento automatizado da dor por movimentos faciais de recém-nascidos internados em Unidade de Terapia Intensiva Neonatal / Tatiany Marcondes Heiderich. São Bernardo do Campo, 2024.

187 f. : il.

Tese - Centro Universitário FEI.

Orientador: Prof. Dr. Carlos Eduardo Thomaz.

1. neonato. 2. segmentação facial. 3. dor. 4. expressão facial. 5. inteligência artificial. I. Thomaz, Carlos Eduardo , orient. II. Título.

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



APRESENTAÇÃO DE TESE
ATA DA BANCA EXAMINADORA
Programa de Pós-Graduação Stricto Sensu em Engenharia Elétrica

Doutorado

PGE-10

Aluno(a): Tatiany Marcondes Heiderich

Matrícula: 520.201-5

Título do Trabalho: RECONHECIMENTO AUTOMATIZADO DA DOR POR MOVIMENTOS FACIAIS DE RECÉM-NASCIDOS INTERNADOS EM UNIDADE DE TERAPIA INTENSIVA NEONATAL

Área de Concentração: Processamento de Sinais e Imagens

Orientador(a): Prof. Dr. Carlos Eduardo Thomaz

Data da realização da defesa: 27/11/2024

ORIGINAL ASSINADA

Avaliação da Banca Examinadora:

O exame iniciou com a apresentação oral da candidata que durou 30 minutos. Em seguida, a candidata foi arguida por cada um dos avaliadores individualmente em um tempo total de 2 horas e 40 minutos de arguição. Ao final, a banca avaliou a tese com mérito acadêmico e contribuições relevantes que justificam o título de doutorado. Os examinadores sugeriram alterações no texto final, reforçando a importância de descrever na discussão e na conclusão as principais contribuições da tese, além de detalhar melhor o perfil da amostra de voluntários selecionada. A candidata foi aprovada por unanimidade.

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

APROVADO

REPROVADO

MEMBROS DA BANCA EXAMINADORA

Prof. Dr. Carlos Eduardo Thomaz

Prof. Dr. Salvador Pinillos Gimenez

Prof. Dr. Marco Antonio Gutierrez

Profª Drª Josy Davidson Okida Vieira

Profª Drª Mandira Daripa Kawakami

Aprovação do Coordenador do Programa de Pós-graduação

Prof. Dr. Carlos Eduardo Thomaz

Dedico esta tese ao meu amado marido, Carlos G. Heiderich Junior, cujo apoio incondicional me deu forças em todos os momentos.

Às minhas filhas, Isabela e Helena, que iluminam meus dias com seus sorrisos e que, sem saber, me inspiraram a continuar a buscar formas de aliviar a dor de uma população tão especial.

Foi no olhar delas ao nascer que compreendi, de maneira profunda, o quanto é essencial para os pais saberem que seus filhos estão bem, sem sofrimento. Embora elas não tenham passado pela UTI, enxerguei no olhar delas a importância desse trabalho e o impacto que ele pode ter na vida de tantas famílias. Reafirmei minha convicção de que, além do toque terapêutico nas pequenas vidas da UTIN como fisioterapeuta neonatal, eu poderia contribuir de outra forma: através da pesquisa, do estudo e do conhecimento compartilhado.

Esta jornada é também para todos os profissionais, pais e mães que atravessam momentos de angústia, buscando conforto e esperança. Que esta dor, se inevitável, seja mais leve, cuidada com amor e atenção, e que os corações aflitos encontrem paz nas mãos dedicadas dos profissionais e, acima de tudo, no amparo de Deus, que conhece a dor mais profunda e nos deu a vida através do sacrifício de Seu filho, Jesus.

AGRADECIMENTOS

Agradeço primeiramente a Deus, fonte de toda sabedoria, pois me guiou, sustentou e permitiu enfrentar e superar os obstáculos de um curso de doutorado em engenharia, um campo tão desafiador e instigante para mim.

Aos meus pais, que agora estão em um lugar muito melhor, deixo minha eterna gratidão. Todo o aprendizado, o amor e os valores que me ensinaram continuam sendo o alicerce de quem sou.

Aos meus familiares: marido, filhas, irmãos, sobrinhos, tios e primos. Vocês são parte fundamental desta conquista! Valeu pelo apoio e carinho ao longo do caminho.

Ao meu orientador, Prof. Dr. Carlos Eduardo Thomaz, minha sincera gratidão. Obrigada por ter acreditado no meu trabalho e por todas as contribuições valiosas ao longo desta caminhada, por me ensinar que a pesquisa pode ser leve, mas cheia de profundidade e riqueza, individual, mas nunca sozinha. Sua orientação foi essencial para que este projeto se tornasse realidade e tantos outros iniciassem.

Deixo também um agradecimento especial aos meus queridos amigos Pedro Domingues e Lucas Buzuti. Pedro, que esteve ao meu lado em cada passo, ajudando com dedicação e generosidade, sem palavras para te agradecer. Lucas, pelo companheirismo nas longas jornadas na FEI, no LNCC, e na busca incansável por imagens faciais de neonatos (olha só que resultado deu!).

À Prof. Dra. Ruth, minha orientadora do coração, obrigada pelo apoio constante e pelas palavras encorajadoras que sempre me inspiraram. Às Dras. Marina e Rita, sou muito grata pelo suporte contínuo durante toda a pesquisa.

Agradeço imensamente aos amigos do grupo FEI e UNIFESP, que juntos formam a equipe mais dedicada e inspiradora de avaliação de dor neonatal que este país já viu. Nosso trabalho coletivo tem sido uma fonte constante de aprendizado e realização, inquietos vamos longe.

Aos pais e bebês que participaram deste projeto, meu agradecimento especial.

Aos profissionais avaliadores, que com tanto zelo e dedicação analisaram os bancos de dados deste estudo, agradeço por sua paciência e compromisso.

Agradeço ao Centro Universitário FEI e a todos os professores do meu curso pela excelência no ensino e pelo suporte inigualável.

Por fim, meu sincero agradecimento ao CNPQ, pelo apoio financeiro durante a pesquisa.

RESUMO

Recém-nascidos internados em Unidades de Terapia Intensiva Neonatal (UTIN) frequentemente passam por procedimentos que causam dor. A presença de dispositivos médicos fixados à face do neonato dificulta a avaliação adequada da dor. Este estudo teve como objetivo desenvolver um método automatizado para reconhecer a dor em neonatos, mesmo em condições de visibilidade facial limitada. Foi proposto e implementado um arcabouço computacional utilizando técnicas atuais de processamento de imagens e de Inteligência Artificial para segmentar regiões faciais específicas e para classificar movimentos faciais indicativos de dor, sendo estes: boca aberta, boca estirada, sulco nasolabial aprofundado, fronte saliente e fenda palpebral estreitada. Com o uso de quatro bases de dados distintas e de referência na literatura afim, o método foi treinado, validado e testado, permitindo a análise apenas das áreas visíveis da face. Uma abordagem inovadora foi aplicada, atribuindo pesos diferenciados aos movimentos faciais mais relevantes, com maior pontuação para os movimentos de maior poder discriminatório, como o sulco nasolabial aprofundado. O método proposto demonstrou ser eficiente e confiável, apresentando na validação um F1 score de 83%, valor-p de McNemar de 0,146, e Kappa de Cohen de 0,668, resultados compatíveis e, em alguns aspectos, superiores aos métodos tradicionais, especialmente em condições de visibilidade parcial da face do neonato. Esta pesquisa não pretende substituir os métodos tradicionais, mas sim otimizá-los, oferecendo uma solução prática e robusta para os desafios do diagnóstico de dor em neonatos na UTIN. A combinação de análise segmentada da face, avaliação apenas das regiões visíveis e pontuações diferenciadas para movimentos faciais mais discriminantes posiciona este arcabouço como uma contribuição significativa e viável para a evolução do manejo da dor neonatal, promovendo uma avaliação mais objetiva e precisa em cenários clínicos complexos.

Palavras-chave: neonato; segmentação facial; dor; expressão facial; inteligência artificial.

ABSTRACT

Newborns admitted to Neonatal Intensive Care Units (NICUs) often undergo procedures that cause pain. The presence of medical devices attached to the neonate's face poses challenges to the adequate assessment of pain. This study aimed to develop an automated method to recognize pain in neonates, even under conditions of limited facial visibility. A computational framework was proposed and implemented using state-of-the-art image processing and artificial intelligence techniques to segment specific facial regions and classify facial movements indicative of pain, namely: open mouth, stretched mouth, deepened nasolabial furrow, bulging forehead, and narrowed palpebral fissure. Using four distinct and literature-referenced databases, the method was trained, validated, and tested, allowing the analysis of only the visible areas of the face. An innovative approach was applied, assigning differentiated weights to the most relevant facial movements, with higher scores for movements with greater discriminatory power, such as the deepened nasolabial furrow. The proposed method proved to be efficient and reliable, demonstrating an F1 score of 83%, a McNemar p-value of 0.146, and a Cohen's Kappa of 0.668 during validation. These results are comparable to, and in some aspects surpass, traditional methods, particularly in conditions of partial facial visibility in neonates. This research does not aim to replace traditional methods but rather to optimize them, offering a practical and robust solution to the challenges of pain diagnosis in NICU neonates. The combination of segmented facial analysis, evaluation of only visible regions, and differentiated scores for more discriminative facial movements positions this framework as a significant and feasible contribution to the advancement of neonatal pain management, promoting a more objective and precise assessment in complex clinical scenarios.

Keywords: newborn; facial segmentation; pain; facial expression; artificial intelligence.

LISTA DE ILUSTRAÇÕES

Figura 1 - Fixação de sonda orogástrica/nasogástrica.....	29
Figura 2 - Fixação do CPAP.....	30
Figura 3 - Fixação da cânula traqueal.....	31
Figura 4 - Protetores oculares para fototerapia.....	32
Figura 5 - Exemplo do conjunto de dados iCOPE.....	34
Figura 6 - Exemplo do conjunto de dados FENP.....	35
Figura 7 - Exemplo do conjunto de dados USF-MNPAD-I.....	36
Figura 8 - Exemplo do conjunto de dados UNIFESP-I.....	38
Figura 9 - Posição das câmeras.....	39
Figura 10 - Exemplo de imagens faciais capturadas antes do procedimento doloroso.....	41
Figura 11 - Exemplo de imagens faciais capturadas após o procedimento doloroso.....	42
Figura 12 - Exemplo de imagens faciais do banco de dados UNIFESP-II.....	44
Figura 13 - Exemplo de imagens faciais sintéticas do banco de dados SynUnifesp.....	48
Figura 14 - Sistema de Codificação de Ações Faciais.....	50
Figura 15 - Pontos fiduciais.....	52
Figura 16 - Pontos fiduciais para detecção automática da dor neonatal.....	53
Figura 17 - Malhas triangulares.....	54
Figura 18 - Etapas para normalização espacial de imagens faciais.....	55
Figura 19 - Arquitetura da LeNet-5 - reconhecimento de dígitos.....	58
Figura 20 - Características faciais neonatais de dor.....	64
Figura 21 - Percepção visual dos médicos.....	69
Figura 22 - Percepção visual de profissionais de saúde.....	69
Figura 23 - Características mais discriminantes da face.....	73
Figura 24 - Parâmetros de nível de estrutura da expressão facial da dor infantil.....	74
Figura 25 - Atlas das imagens faciais das bases iCOPE e UNIFESP-I.....	75
Figura 26 - Sobreposição dos mapas em tons de vermelho sobre os atlas médios.....	76
Figura 27 - Telas de aplicação móvel.....	78
Figura 28 - Máscaras de segmentação geradas de cada modelo.....	80
Figura 29 - Fluxograma - Análise da face livre.....	84
Figura 30 - Fluxograma - Análise facial de neonatos internados em UTIN.....	85
Figura 31 - <i>MediaPipe</i> : face livre.....	88

Figura 32 – <i>MediaPipe</i> : face com dispositivos	88
Figura 33 - Pontos fiduciais: face livre.....	89
Figura 34 - Pontos fiduciais: face com dispositivos	90
Figura 35 - Exemplo das coordenadas para extração do recorte da boca	91
Figura 36 - Exemplo da extração das regiões faciais livres	92
Figura 37 - Exemplo da extração das regiões faciais com dispositivos.....	92
Figura 38 - Segmentação das regiões faciais livres	93
Figura 39 - Segmentação das regiões faciais de neonatos internados em UTIN.....	93
Figura 40 - <i>Resize</i> e <i>Padding</i> das regiões faciais livres	94
Figura 41 - <i>Resize</i> e <i>Padding</i> das regiões faciais de neonatos internados em UTIN	95
Figura 42 - Modelo para avaliação das regiões de faces livres.....	103
Figura 43 - Validação dos melhores modelos treinados.....	115
Figura 44 - Exemplo de imagem com diagnóstico de dor discordante no conjunto de dados UNIFESP-I.....	148
Figura 45 - Exemplo de imagem com diagnóstico de dor discordante no conjunto de dados UNIFESP-II.....	149
Figura 46 - Exemplo de imagem com diagnóstico de dor discordante no conjunto de dados SynUnifesp.	151
Figura 47 - Exemplo de imagem com diagnóstico de dor discordante no conjunto de dados iCOPE.....	152

LISTA DE TABELAS

Tabela 1 - Características das imagens dos neonatos - banco de dados UNIFESP-I	40
Tabela 2 - Características das imagens dos recém-nascidos do banco de dados UNIFESP-II.....	45
Tabela 3 - Escalas faciais para avaliação de dor e regiões faciais	67
Tabela 4 - Bases de dados.....	86
Tabela 5 - Matriz de confusão	98
Tabela 6 - Dados de treinamento e validação	104
Tabela 7 - Resultados das análises com destaque para o melhor F1 de cada movimento facial para cada conjunto de dados	107
Tabela 8 - Melhor F1 score dos modelos para cada movimento facial indicativo da presença de dor	111
Tabela 9 – Validação – Acurácia e Concordância entre os melhores modelos e o rótulo de movimento facial indicativo de dor do conjunto de dados UNIFESP-I.....	116
Tabela 10 – Validação - Acurácia e Concordância entre os melhores modelos e o rótulo de diagnóstico de dor do conjunto de dados UNIFESP-I	119
Tabela 11 – Validação - Acurácia e Concordância entre os diferentes conjuntos de score dos melhores modelos e o rótulo de diagnóstico de dor do conjunto de dados UNIFESP-I.....	122
Tabela 12 - Dados de teste	126
Tabela 13 – Teste - Acurácia e Concordância entre os melhores modelos e o rótulo de diagnóstico de dor do conjunto de dados iCOPE	127
Tabela 14 – Teste - Acurácia e Concordância entre os diferentes conjuntos de score dos melhores modelos e o rótulo de diagnóstico de dor do conjunto de dados iCOPE	130
Tabela 15 – Teste - Acurácia e Concordância entre os melhores modelos e o rótulo dos movimentos faciais indicativos de dor do conjunto de dados UNIFESP-II.....	134
Tabela 16 – Teste - Acurácia e Concordância entre os melhores modelos e o rótulo de diagnóstico de dor do conjunto de dados UNIFESP-II	136
Tabela 17 – Teste - Acurácia e Concordância entre os diferentes conjuntos de score dos melhores modelos e o rótulo de diagnóstico de dor do conjunto de dados UNIFESP-II.....	139

Tabela 18 – Novo Método de Classificação e Diagnóstico Automático da Dor	143
Tabela 19 – Dados de validação para um Novo Método de Classificação e Diagnóstico Automático da Dor.....	144
Tabela 20 – Dados de teste do Novo Método de Classificação e Diagnóstico Automático da Dor.....	145

LISTA DE ABREVIATURAS

APN-db	<i>Acute Pain in Neonates database</i>
APS	<i>Acute Pain Scale</i>
AUs	<i>Action Units</i>
BIIP	<i>InfanPain Assessment Scale for Preterm Infants</i>
BIPAP	<i>Bilevel Positive Airway Pressure</i>
CAPES	Coordenação de Aperfeiçoamento de Pessoal de Nível Superior
CNN	<i>Convolutional Neural Networks</i>
CPAP	<i>Continuous Positive Airway Pressure</i>
EVENDOL	<i>Evaluation Enfant Douleur</i>
FACS	<i>Facial Action Coding System</i>
FAPESP	Fundação de Amparo à Pesquisa do Estado de São Paulo
FEI	Fundação Educacional Inaciana "Pe. Sabóia de Medeiros"
FENP	<i>Facial Expression of Neonatal Pain</i>
IA	Inteligência Artificial
iCOPEvid	<i>Infant Classification Of Pain Expressions</i>
LIDS	<i>Liverpool Infant Distress Scale</i>
MCGRATH	<i>McGrath Facial Affective Scale</i>
MIPS	<i>Modified Infant Pain Scale</i>
ML	<i>Machine Learning</i>
MLP	<i>Multilayer Perceptron</i>
M-NFCS	<i>Modified - Neonatal Facial Coding System Modificado</i>
NBHR	<i>Newborn Baby Heart Rate Estimation Database</i>
N-CNN	<i>Neonatal - Convolutional Neural Networks</i>
NFCS	<i>Neonatal Facial Coding System</i>
NFCS-R	<i>"Neonatal Facial Coding System-Revised"</i>
NIPS	<i>Neonatal Infant Pain Scale</i>
NIRS	<i>Near-Infrared Spectroscopy</i>
NNICUPAT	<i>Nepean Neonatal Intensive Care Unit Pain Assessment Tool</i>
N-PASS	<i>Neonatal Pain, Agitation, and Sedation Scale</i>
ObsVAS	<i>Observer administered Visual Analogue Scale</i>

PASPI	<i>Pain Assessment Scale for Preterm Infants</i>
PIPP	<i>Premature Infant Pain Profile</i>
PIPP-R	<i>Premature Infant Pain Profile - Revisited</i>
P-MIPS	<i>Partial - Modified Infant Pain Scale</i>
ResNet	<i>Residual Network</i>
SAD-RN	<i>Software de Avaliação da Expressão Facial de Dor de Recém-Nascidos</i>
SVM	<i>Support Vector Machines</i>
TCLE	<i>Termo de Consentimento Livre e Esclarecido</i>
TPE	<i>Tree-structured Parzen Estimator</i>
UNIFESP	<i>Universidade Federal de São Paulo</i>
USF-MNPAD-I	<i>University of South Florida Multimodal Neonatal Pain Assessment Dataset</i>
UTI	<i>Unidades de Terapia Intensiva</i>
UTIN	<i>Unidades de Terapia Intensiva Neonatal</i>
ViT	<i>Vision Transformer</i>

SUMÁRIO

1	INTRODUÇÃO	17
1.1	JUSTIFICATIVA	24
1.2	OBJETIVO GERAL E ESPECÍFICOS	25
1.3	PUBLICAÇÕES	26
1.4	ESTRUTURA DO TRABALHO	26
2	CONCEITOS FUNDAMENTAIS.....	28
2.1	DISPOSITIVOS FACIAIS DE SUPORTE	28
2.1.1	Sonda orogástrica ou sonda nasogástrica	29
2.1.2	CPAP	30
2.1.3	Cânula traqueal.....	31
2.1.4	Protetor ocular para fototerapia.....	32
2.2	BANCOS DE DADOS DE IMAGENS FACIAIS DE RECÉM-NASCIDOS ..	33
2.2.1	Bancos de dados de imagens faciais UNIFESP-I	37
2.2.2	Bancos de dados de imagens faciais UNIFESP-II	43
2.2.3	Bancos de dados de imagens faciais sintéticas SynUnifesp	47
2.3	TECNOLOGIA PARA RECONHECIMENTO AUTOMÁTICO DA DOR	50
2.3.1	Reconhecimento Facial	50
2.3.2	Pontos Fiduciais.....	51
2.3.3	Malhas Triangulares.....	53
2.3.4	Técnicas de Pré-Processamento e Normalização de Imagem	54
2.3.5	Algoritmos de Aprendizado de Máquina	55
2.3.6	<i>Support Vector Machines</i>	57
2.3.7	<i>Convolutional Neural Networks</i>	57
2.3.8	<i>Neonatal - Convolutional Neural Networks</i>	59
2.3.9	<i>Transformers</i>	59
2.4	CONSIDERAÇÕES COMPLEMENTARES	60
3	TRABALHOS RELACIONADOS	61
3.1	AVALIAÇÃO DA DOR EM RECÉM-NASCIDOS	61
3.2	AVALIAÇÃO AUTOMÁTICA DA EXPRESSÃO FACIAL DE DOR EM RECÉM-NASCIDOS.....	70
3.3	CONSIDERAÇÕES COMPLEMENTARES	82
4	MATERIAIS E MÉTODOS	83

4.1	BASE DE DADOS - IMAGENS	86
4.2	MEDIAPIPE	88
4.3	EXTRAÇÃO DAS REGIÕES FACIAIS	90
4.4	MÁSCARA DE SEGMENTAÇÃO DAS REGIÕES FACIAIS.....	93
4.5	<i>RESIZE E PADDING</i>	93
4.6	TREINAMENTO COM <i>VISION TRANSFORMER</i>	95
4.7	VALIDAÇÃO E TESTE.....	96
4.8	ANÁLISE DE DADOS	96
4.8.1	Identificação do melhor modelo para avaliar cada uma das movimentações faciais indicativas da presença de dor	97
4.8.2	Acurácia e concordância entre os melhores modelos e a avaliação dos profissionais de saúde	98
5	EXPERIMENTOS E RESULTADOS	102
5.1	TREINAMENTO DAS IMAGENS COM A FACE LIVRE	102
5.2	RESULTADOS DO TREINAMENTO COM VIT	106
5.3	VALIDAÇÃO DOS MELHORES MODELOS TREINADOS	113
5.4	TESTE DOS MELHORES MODELOS VALIDADOS.....	125
5.4.1	Teste dos melhores modelos para as imagens com a face livre	126
5.4.2	Teste dos melhores modelos para as imagens de neonatos internados em UTIN.....	133
5.5	CLASSIFICAÇÃO E DIAGNÓSTICO DA DOR POR REGIÕES VISÍVEIS DA FACE.....	142
6	DISCUSSÃO	153
7	CONCLUSÃO	166
	REFERÊNCIAS.....	167
	ANEXO A - TRABALHOS PUBLICADOS COM A PARTICIPAÇÃO DA AUTORA DA TESE	179
	ANEXO B - SOLICITAÇÃO DE AUTORIZAÇÃO PARA USO DE IMAGENS – ICOPE	184
	ANEXO C - SOLICITAÇÃO DE AUTORIZAÇÃO PARA USO DE IMAGENS - UNIFESP-I.....	186
	ANEXO D - SOLICITAÇÃO DE AUTORIZAÇÃO PARA USO DE IMAGENS - UNIFESP-II.....	188

ANEXO E - SOLICITAÇÃO DE AUTORIZAÇÃO PARA USO DE IMAGENS - SYNUNIFESP	190
--	------------

1 INTRODUÇÃO

A previsão de desfechos clínicos em Unidades de Terapia Intensiva Neonatal (UTIN) frequentemente utiliza sistemas baseados na gravidade das condições clínicas do paciente para estimar a mortalidade⁽¹⁾. Recém-nascidos criticamente enfermos, internados em UTIN, são submetidos a uma série de procedimentos invasivos, que, embora indispensáveis para a manutenção da vida, estão fortemente associados a episódios de dor⁽²⁾.

Entre os principais procedimentos realizados em neonatos estão a aspiração oral ou nasal, aspiração traqueal, punções arteriais e venosas, passagem de sonda gástrica ou enteral, punção de calcâneo, inserção de cateter venoso central de inserção periférica, além de intubações e extubações traqueais⁽³⁾.

Um estudo realizado por Santiago *et al.* (2017)⁽⁴⁾ revelou que, devido às diversas morbidades apresentadas pelos neonatos, algumas intervenções são altamente prevalentes nas UTINs, como o uso de medicações (93,3%), aspiração de vias aéreas superiores (73,3%), oxigenoterapia (50%) e ventilação mecânica invasiva (50%), além das ventilações mecânicas não invasivas, como CPAP (*Continuous Positive Airway Pressure*) e BIPAP (*Bilevel Positive Airway Pressure*), destacando a complexidade do cuidado intensivo necessário para esses pacientes. Esses procedimentos, essenciais para o manejo das condições clínicas complexas desses pacientes, são fontes significativas de dor aguda e estresse fisiológico⁽⁴⁾.

A exposição repetida a esses estímulos dolorosos tem consequências que vão além da dor momentânea. Esses episódios dolorosos podem ter um impacto profundo no desenvolvimento cerebral do recém-nascido⁽⁵⁾. O sistema nervoso, ainda em maturação, é altamente vulnerável, e a exposição contínua à dor pode gerar alterações estruturais no cérebro, como danos neuronais e a redução das conexões axonais. Consequentemente, isso pode resultar em deficiências cognitivas e motoras, além de hipersensibilidade a estímulos dolorosos, que podem persistir por toda a vida^(5,6).

A curto prazo, a dor não tratada compromete a estabilidade respiratória, cardiovascular e metabólica, aumentando os riscos de morbimortalidade⁽⁷⁾. Além disso, a dor afeta o comportamento dos recém-nascidos, provocando irritabilidade, distúrbios no sono e recusa alimentar, interferindo também na relação mãe-filho⁽⁷⁾.

As consequências da dor persistente incluem a predisposição ao desenvolvimento de hipersensibilidade e condições dolorosas crônicas, além de déficits de atenção e concentração que podem se manifestar na vida escolar, mesmo na ausência de lesões neurológicas graves⁽⁸⁻¹⁰⁾.

A internação prolongada em Unidade de Terapia Intensiva (UTI), frequentemente associada a condições clínicas mais graves na admissão, pode ter um impacto negativo sobre o estado de saúde dos pacientes, aumentando o risco de infecções e complicações clínicas e, conseqüentemente, a mortalidade⁽¹⁾.

Um estudo realizado por Quaresma *et al.* (2018)⁽¹¹⁾, examinou os fatores determinantes da internação neonatal e revelou que as condições socioeconômicas maternas, como baixa escolaridade, etnia e falta de apoio familiar, estão fortemente associadas à necessidade de internação neonatal. Além disso, características do recém-nascido, como prematuridade e baixo peso ao nascer, também são fatores importantes para a hospitalização. Essas internações, muitas vezes, estão diretamente relacionadas à gravidade das condições clínicas no momento do nascimento, refletindo uma sobreposição entre os fatores que levam à internação e aqueles que influenciam a mortalidade neonatal⁽¹¹⁾.

Outro aspecto crítico no cenário da UTIN é a alta incidência de eventos adversos, definidos como complicações inesperadas que ocorrem durante o tratamento e que não estão diretamente relacionadas à evolução natural da doença⁽¹²⁾. Esses eventos têm sido objeto de intenso debate no âmbito internacional devido ao seu impacto substancial na morbimortalidade, prolongamento da hospitalização e aumento expressivo dos custos associados ao tratamento⁽¹³⁾.

Um estudo descritivo realizado na unidade neonatal de um hospital privado em Maringá⁽¹⁴⁾, entre junho de 2016 e março de 2017, caracterizou as internações na UTIN e documentou a ocorrência de eventos adversos registrados em prontuários médicos. Dentre esses, destacaram-se complicações associadas a procedimentos invasivos: 53,70% dos eventos estavam relacionados ao acesso venoso central, 14,33% ao cateter central de inserção periférica, 14,33% à intubação, 7,41% ao atraso na inserção do cateter vesical, e 5,27% ao cateter umbilical. Esses dados destacam a necessidade urgente de aprimorar as práticas clínicas em UTINs, promovendo intervenções que minimizem a ocorrência de eventos adversos e, conseqüentemente, reduzam o impacto negativo sobre os pacientes neonatais.

No estudo de Spironello e Cuman (2019)⁽¹⁴⁾ foi evidenciado que uma proporção considerável de eventos adversos estava associada a procedimentos dolorosos. Quando analisados mensalmente, os eventos adversos incluíam lesões de septo nasal devido ao uso de CPAP em 18,75% dos casos, lesões cutâneas causadas por fita adesiva em 14,06%, complicações relacionadas ao cateter nasoenteral em 13,28%, e lesões associadas ao cateter periférico central em 11,71% dos casos. Além disso, houve registros significativos de extubação acidental (10,93%), lesões cutâneas provocadas por oxímetro (4,68%), lesões de pele (3,9%), úlceras por pressão (2,34%) e complicações relacionadas ao cateter umbilical (7,81%).

No ano de 2021, um estudo conduzido por Cossul *et al.*, em uma UTIN do Distrito Federal, estimou que aproximadamente 70% dos pacientes sofreram algum tipo de evento adverso, sublinhando a magnitude do problema e a urgente necessidade de intervenções preventivas eficientes⁽¹⁵⁾. A estimativa de que um recém-nascido hospitalizado pode ser submetido a mais de 300 procedimentos dolorosos durante sua internação reforça ainda mais essa necessidade, destacando a importância de estratégias eficazes de manejo da dor para minimizar tais impactos e reduzir os índices de morbimortalidade neonatal^(7,16,17).

No contexto global, a Organização das Nações Unidas reconhece a importância da redução da mortalidade infantil como um indicador-chave para a saúde da população. Entre os “Objetivos de Desenvolvimento do Milênio”⁽¹⁸⁾, a redução da mortalidade infantil foi uma meta prioritária de 1990 a 2015. Atualmente, entre os “Objetivos de Desenvolvimento Sustentável”⁽¹⁹⁾, estabelecidos para serem alcançados até 2030, um dos principais objetivos relacionados à saúde visa a eliminação de mortes evitáveis de recém-nascidos e crianças menores de cinco anos, reafirmando o compromisso global com a promoção da saúde e bem-estar em todas as idades⁽²⁰⁾.

Nos últimos anos, observou-se uma redução significativa na taxa de mortalidade infantil em escala global⁽²¹⁾. De acordo com estimativas do "Grupo Interinstitucional das Nações Unidas para Estimativa da Mortalidade Infantil", o número de crianças que faleceram antes de completar cinco anos atingiu um mínimo histórico em 2022, com 4,9 milhões de óbitos. A taxa global de mortalidade infantil nessa faixa etária diminuiu 51% desde 2000. No Brasil, a redução foi ainda mais acentuada, alcançando uma queda de 60% no mesmo período⁽²²⁾.

No entanto, os óbitos neonatais apresentaram uma redução menos expressiva, permanecendo um grande desafio, especialmente para países em desenvolvimento,

como o Brasil⁽²¹⁾. Entre 2000 e 2018, houve uma redução na mortalidade neonatal evitável no Brasil e na maioria dos estados, embora as disparidades regionais sejam notáveis. Estados das regiões Norte e Nordeste continuam a apresentar taxas de mortalidade neonatal superiores à média nacional⁽²³⁾.

Em 2022, o "Grupo Interinstitucional das Nações Unidas" relatou que, globalmente, quase metade das mortes em crianças menores de cinco anos ocorreu no período neonatal, sendo a maioria causada por condições evitáveis ou tratáveis, como parto prematuro, complicações durante o nascimento, pneumonia, diarreia e malária⁽²²⁾. Em 2023, o Brasil atingiu a menor taxa de mortalidade infantil e fetal por causas evitáveis dos últimos 28 anos, conforme informações divulgadas pelo Painel de Monitoramento da Mortalidade Infantil e Fetal do Ministério da Saúde⁽²⁴⁾. Apesar do progresso, ainda persistem desafios, especialmente nas regiões mais vulneráveis do país, onde a qualidade do acesso aos cuidados de saúde ainda precisa ser aprimorada⁽²⁵⁾.

Fatores como o parto prematuro e complicações neonatais evitáveis ainda são os principais responsáveis por uma parte significativa dos óbitos infantis, exigindo esforços contínuos e investimentos em tecnologia e capacitação para melhorar a atenção neonatal e reduzir ainda mais esses índices⁽²⁴⁾. Por este motivo, é importante conhecer os fatores de óbito associados à cada faixa etária, bem como priorizar ações específicas para o controle da mortalidade⁽²⁶⁾. Nesse sentido, o Ministério da Saúde tem implementado diversas políticas para reorganizar a rede de cuidados e atender de maneira mais adequada às demandas por assistência neonatal⁽²⁷⁾.

Entre as iniciativas adotadas para aprimorar o cuidado neonatal no Brasil, destaca-se a Portaria nº 930, de 10 de maio de 2012, que estabelece diretrizes mínimas para o atendimento intensivo neonatal no Sistema Único de Saúde (SUS). O principal objetivo dessa portaria é assegurar uma atenção integral e humanizada aos recém-nascidos em estado grave, promovendo a redução da morbimortalidade neonatal⁽²⁸⁾. Além disso, o Ministério da Saúde implementou em 2017 a "Estratégia QualiNEO", em parceria com o Instituto Fernandes Figueira/Fiocruz, que teve como objetivo, ofertar apoio técnico de forma sistemática e integrada às maternidades prioritárias para qualificação das práticas de gestão e atenção ao recém-nascido⁽²⁹⁾.

Em 2023, foi instituída a Portaria GM/MS nº 889, que expandiu a rede de atendimento neonatal ao habilitar novos leitos em Unidades de Terapia Intensiva Neonatal (UTIN) e Unidades de Cuidado Intermediário Neonatal (UCIN)⁽³⁰⁾. Essa ação

fortaleceu ainda mais a infraestrutura das maternidades brasileiras, garantindo suporte especializado para recém-nascidos de risco, e ampliou a capacidade de atendimento em momentos críticos, assegurando que mais neonatos possam receber cuidados adequados em unidades especializadas.

Para que esses avanços tenham um impacto significativo na redução da mortalidade neonatal, é fundamental adotar intervenções eficazes que melhorem indicadores-chave. Isso inclui a qualificação da assistência pré-natal, o aprimoramento da nutrição materna e a garantia de cuidados imediatos ao recém-nascido. Essas medidas preventivas, quando associadas ao atendimento especializado nas unidades de cuidados intensivos, são essenciais para prevenir complicações neonatais e contribuir diretamente para a redução dos índices de mortalidade, especialmente em recém-nascidos prematuros e de alto risco.

Nesse contexto, um aspecto fundamental que ainda exige atenção especial é a avaliação e o manejo adequado da dor neonatal. A dor em recém-nascidos, particularmente em unidades de terapia intensiva, tem sido amplamente associada a desfechos negativos quando não é adequadamente tratada⁽³¹⁾. Considerando o impacto significativo da dor nos neonatos internados, os profissionais de saúde enfrentam o desafio de avaliá-la de maneira eficaz para garantir um tratamento adequado. Essa avaliação não apenas alivia o sofrimento imediato do neonato, mas também desempenha um papel crítico na redução da morbimortalidade neonatal⁽³¹⁾.

Nas UTINs, a avaliação da dor neonatal geralmente se baseia em sinais não verbais, como a expressão facial, postura corporal e alterações nos sinais vitais⁽³²⁾. Esses indicadores são essenciais para a identificação de dor aguda e crônica. A expressão facial, em particular, é amplamente utilizada por sua precisão e caráter não invasivo, sendo o método preferido quando o recém-nascido não utiliza dispositivos que obstruam a face⁽⁷⁾. No entanto, quando há o uso de dispositivos médicos fixados no rosto do neonato, como sondas ou máscaras de ventilação, a avaliação da dor torna-se mais difícil, exigindo que o profissional redirecione sua atenção para áreas descobertas da face e para sinais fisiológicos e comportamentais, como alterações nos sinais vitais e na movimentação corporal⁽³³⁾.

Essas dificuldades ressaltam a importância de uma observação atenta e detalhada por parte dos profissionais de saúde⁽³⁴⁾. No entanto, percebe-se uma lacuna significativa entre o conhecimento teórico sobre a avaliação da dor e sua aplicação prática nas UTINs. Um dos principais desafios é que a dor, normalmente expressa

verbalmente em pacientes adultos, se manifesta em neonatos por meio de sinais não verbais, o que requer uma interpretação acurada por parte da equipe de saúde⁽³³⁾.

Para reduzir essa lacuna, é fundamental que os profissionais de saúde recebam treinamento formal e contínuo sobre as particularidades da dor neonatal, em todos os níveis de sua formação. Além disso, a implementação de protocolos padronizados de avaliação e manejo da dor nas unidades neonatais é crucial para assegurar que a dor seja identificada e tratada de maneira eficaz, minimizando, assim, os efeitos adversos no desenvolvimento dos recém-nascidos⁽³⁵⁾.

A dificuldade em reconhecer e avaliar adequadamente a dor neonatal continua sendo um dos maiores obstáculos para a administração efetiva de analgesia em pacientes vulneráveis. A falta de familiaridade com as escalas de avaliação de dor neonatal e o manejo adequado de analgésicos, incluindo seus efeitos colaterais, agravam o subtratamento da dor. Esse cenário destaca a necessidade urgente de mais pesquisas e de uma ênfase maior no ensino médico voltado para a analgesia neonatal. Uma análise crítica das instituições formadoras de opinião e dos programas de ensino pode identificar as deficiências atuais e sugerir melhorias substanciais no cuidado neonatal, com foco no manejo da dor⁽³⁶⁾.

Para que as UTINs possam cumprir seu papel de oferecer cuidado integral e de alta qualidade a recém-nascidos em risco, é indispensável uma estrutura técnica adequada e profissionais qualificados. O avanço nas tecnologias médicas e a capacitação contínua da equipe de saúde são elementos fundamentais para a redução da morbimortalidade neonatal. Além de melhorar as taxas de sobrevivência, essas melhorias contribuem significativamente para a qualidade de vida dos neonatos, minimizando as sequelas associadas ao manejo inadequado da dor e às condições críticas tratadas nessas unidades⁽³⁷⁾.

Nos últimos anos, diversos métodos computacionais têm sido desenvolvidos para detectar automaticamente o fenômeno doloroso em neonatos, com o objetivo de auxiliar os profissionais de saúde na monitorização da dor e na identificação da necessidade de intervenções terapêuticas⁽³⁸⁻⁵⁰⁾. Esses métodos incluem o uso de técnicas de aprendizado de máquina, processamento de imagens e análise de expressões faciais para a avaliação da dor em pacientes neonatais. Estudos como os de Zamzmi *et al.* (2019)⁽⁴⁴⁾ têm explorado algoritmos que analisam características faciais para detectar sinais de dor em neonatos, destacando o potencial dessas abordagens para melhorar a precisão e a eficiência na avaliação da dor.

Heiderich *et al.* (2023)⁽⁵¹⁾, em um comentário sobre os desafios e as perspectivas para automatização da avaliação de dor em UTIN, descrevem que existem limitações significativas que ainda dificultam a implementação eficaz desses sistemas. Entre essas limitações estão a dificuldade em detectar pontos correspondentes à área inferior da face, como os movimentos do queixo e da língua, além dos movimentos rápidos da cabeça; a dificuldade em identificar rostos de neonatos em cenas complexas, em que a iluminação pode ser inadequada; e a incapacidade de detectar rostos que estão parcialmente cobertos por dispositivos médicos, como cânula traqueal. Outro desafio crítico é a escassez de bancos de dados contendo imagens faciais de neonatos para treinamento e teste de algoritmos, bem como o tamanho limitado dos conjuntos de dados disponíveis, o que restringe o desenvolvimento de modelos generalizáveis.

Em outras palavras, apesar dos avanços tecnológicos, ainda há uma falta significativa na disponibilidade de ferramentas adequadas para a avaliação e mensuração precisa da dor neonatal no ambiente hospitalar^(51,52). Atualmente, os métodos existentes dependem fortemente da observação clínica⁽⁵³⁾, o que, embora seja útil e confiável, pode ser limitado, especialmente em situações em que o recém-nascido utiliza dispositivos que obstruem a face ou quando sinais não verbais são sutis. Nesse sentido, o desenvolvimento de novas tecnologias automatizadas para a avaliação da dor neonatal torna-se uma necessidade urgente.

Para que essas novas tecnologias sejam eficazmente integradas à prática clínica, é essencial que sejam acompanhadas por pesquisas que validem sua utilização no cotidiano das UTINs. Essas pesquisas devem abordar não apenas a viabilidade técnica, mas também a implementação prática, garantindo que os profissionais de saúde estejam devidamente capacitados e conscientes da importância da avaliação contínua da dor em neonatos. A integração dessas tecnologias ao manejo clínico rotineiro tem o potencial de reestruturar o cuidado neonatal, proporcionando uma abordagem mais objetiva e precisa, minimizando os riscos de subtratamento da dor e, conseqüentemente, reduzindo o tempo de internação e melhorando os desfechos clínicos⁽⁵¹⁾.

Portanto, o futuro do cuidado neonatal está relacionado não apenas à criação de novas ferramentas tecnológicas, mas também à conscientização e capacitação dos profissionais de saúde para que possam utilizar essas inovações de maneira eficaz. Somente através da combinação de avanços tecnológicos e sensibilização da equipe

médica será possível oferecer uma assistência mais completa e humanizada, garantindo que a dor neonatal seja tratada de forma adequada e que os recém-nascidos tenham melhor oportunidade de um desenvolvimento saudável e livre de sofrimento desnecessário.

1.1 JUSTIFICATIVA

Recém-nascidos criticamente doentes frequentemente enfrentam longos períodos de hospitalização, permanecendo sob constante risco de eventos adversos e exposição contínua à dor^(1,2,5). Esses fatores contribuem significativamente para o aumento dos índices de morbidade e mortalidade neonatais, podendo resultar em repercussões clínicas importantes a curto, médio e longo prazo^(7-10,31). Dentre os diversos aspectos a serem considerados durante a internação, o manejo da dor neonatal é de extrema importância, devendo ser evitada, avaliada de maneira precisa e tratada adequadamente.

De acordo com diretrizes internacionais recentes⁽⁵⁴⁾, o gerenciamento eficaz da dor neonatal requer uma avaliação válida e contínua. Essas diretrizes recomendam que as unidades de saúde que cuidam de recém-nascidos adotem ferramentas de avaliação da dor que sejam capazes de abranger todas as idades e os diversos tipos de dor que podem ocorrer no ambiente neonatal. Além disso, ressaltam a necessidade de que a detecção e avaliação da dor sejam realizadas regularmente, e que futuras pesquisas se concentrem no desenvolvimento e validação de ferramentas de avaliação específicas para diferentes situações clínicas, garantindo que essas ferramentas sejam aplicadas de forma contínua.

Os trabalhos mais recentes relacionados à avaliação automática da dor no período neonatal estão avançando metodologicamente^(38,40,43,55,56), mas desafios significativos persistem, particularmente na identificação da dor em recém-nascidos criticamente doentes que utilizam dispositivos médicos fixados à face. A dificuldade de avaliar expressões faciais em neonatos cuja face está parcialmente coberta por dispositivos de suporte à vida, como sondas nasogástricas, cânulas traqueais e protetor ocular para fototerapia, limita a capacidade de detectar automaticamente e em tempo real sinais faciais indicativos de dor. Isso evidencia a necessidade urgente

de desenvolver novas técnicas para a detecção da mímica facial, especificamente adaptadas para neonatos internados em ambiente hospitalar⁽⁵¹⁾.

A avaliação da dor através da análise dos movimentos faciais já vem sendo testada na prática clínica em populações adultas, particularmente em pacientes não comunicantes, como idosos com demência ou em situações pós-operatórias⁽⁵⁷⁾. Estudos recentes demonstram resultados promissores, utilizando aprendizado profundo para avaliar a intensidade da dor pós-operatória com base no reconhecimento de expressões faciais em condições reais^(57,58). Esses avanços sugerem que a aplicação de Inteligência Artificial pode discriminar expressões faciais associadas a diferentes intensidades de dor e prever a intensidade da dor de forma precisa. Diante disso, a avaliação automatizada da dor em neonatos internados em UTINs, baseada na análise de movimentos faciais, surge como uma abordagem inovadora que pode contribuir significativamente para otimizar o manejo da dor em unidades neonatais.

1.2 OBJETIVO GERAL E ESPECÍFICOS

Esta tese tem como objetivo geral propor e implementar um arcabouço computacional de reconhecimento automatizado da dor, por regiões da face do recém-nascido internado em UTIN, que permanece com dispositivos fixados em sua face.

Especificamente, os objetivos são: (1) realizar a segmentação automática das regiões faciais do recém-nascido que permanecem, ou não, com a presença de dispositivos que ocluem parcialmente a face; (2) classificar as regiões faciais segmentadas como movimentos indicativos da presença ou ausência de dor; e (3) criar um método automático de classificação e diagnóstico da dor que utilize exclusivamente as regiões visíveis da face, garantindo que a avaliação seja eficaz, mesmo em condições de visibilidade parcial da face do neonato.

1.3 PUBLICAÇÕES

Partes dessa tese motivaram e influenciaram a produção científica de trabalhos em congressos nacionais, internacionais e periódicos. A lista completa destes trabalhos, da publicação mais recente para a mais antiga, encontra-se no ANEXO A.

1.4 ESTRUTURA DO TRABALHO

Este trabalho está estruturado nos seguintes capítulos:

No **Capítulo 2**, são apresentados os conceitos fundamentais relacionados aos dispositivos faciais de suporte à vida, bancos de dados disponíveis para estudo, e tecnologia usada no reconhecimento facial. Este capítulo aborda resumidamente conceitos que podem não ser familiarizados por leitores da área da saúde e da área da engenharia, e visa facilitar a compreensão dos capítulos seguintes.

No **Capítulo 3**, são discutidos os trabalhos relacionados, onde são revisadas as principais pesquisas disponíveis na literatura científica sobre avaliação manual e automática de dor neonatal, com o objetivo de delinear o estado da arte na área, destacando os avanços alcançados e as lacunas ainda presentes no conhecimento afim.

O **Capítulo 4** trata dos materiais e métodos utilizados, detalhando as bases de dados, os algoritmos computacionais aplicados e os procedimentos adotados para coleta e análise dos dados. Esse capítulo é fundamental para assegurar a transparência metodológica e permitir a reprodutibilidade dos resultados.

No **Capítulo 5**, são apresentados os resultados da pesquisa, com descrições detalhadas das análises realizadas. Os resultados são exibidos com o suporte de tabelas que permitem ao leitor acompanhar as análises de forma visual e informativa.

O **Capítulo 6** aborda a discussão dos resultados, onde são comparados os achados desta pesquisa com os estudos revisados no Capítulo 3. Nesta seção, são discutidas as implicações dos resultados, as limitações da pesquisa, e o impacto que esses resultados atingidos podem ter no contexto da avaliação da dor em neonatos. Também são sugeridas possíveis direções para pesquisas futuras.

Por fim, o **Capítulo 7** apresenta as conclusões deste trabalho. Aqui são resumidas as principais contribuições da tese, destacando o impacto do estudo na

área de avaliação da dor neonatal e propondo melhorias e extensões que podem ser desenvolvidas em pesquisas futuras, com vistas a melhorar o atendimento a neonatos em ambientes clínicos.

2 CONCEITOS FUNDAMENTAIS

Este capítulo aborda os principais conceitos referentes aos dispositivos de suporte à vida que, frequentemente, são fixados à face de neonatos internados em UTIN. Além disso, explora os bancos de dados disponíveis para estudos na área e as tecnologias aplicadas ao reconhecimento facial, que têm sido fundamentais para a inovação no monitoramento e cuidado neonatal.

2.1 DISPOSITIVOS FACIAIS DE SUPORTE

Os recém-nascidos que necessitam de cuidados intensivos prolongados em UTIN frequentemente requerem o uso de dispositivos faciais de suporte para garantir funções vitais e terapêuticas essenciais. Esses equipamentos, embora fundamentais para a sobrevivência e o tratamento dos neonatos, podem cobrir parcial ou totalmente a face, interferindo na avaliação de sinais clínicos, como a expressão facial de dor^(7,59,60). A detecção precisa de pontos faciais é crucial para os métodos de análise de dor, tanto manuais quanto automatizados⁽⁵¹⁾. No entanto, a presença desses dispositivos dificulta a identificação desses sinais, tornando o processo de avaliação da dor mais complexo.

Entre os dispositivos faciais de suporte mais comuns estão a sonda orogástrica ou sonda nasogástrica; a cânula traqueal; o CPAP, que fornece uma pressão positiva contínua nas vias aéreas; os protetores oculares para fototerapia; entre outros. Além destes dispositivos, ainda existem materiais que são utilizados em situações específicas que também podem dificultar a visualização de certos pontos faciais, como o caso da touca, que muitas vezes é utilizada para aquecimento ou auxílio na fixação de alguns dispositivos de suporte, e impede a visualização da região superior da face.

Cada um desses dispositivos, além de seu papel terapêutico essencial, apresenta desafios distintos no que diz respeito à avaliação clínica e à aplicação de métodos computacionais para a análise da dor. A seguir, será detalhado o funcionamento e a importância de cada um desses dispositivos, bem como as implicações de seu uso na avaliação facial de dor neonatal.

2.1.1 Sonda orogástrica ou sonda nasogástrica

A sonda orogástrica ou nasogástrica é utilizada em recém-nascidos críticos para alimentação enteral, administração de medicamentos e drenagem de conteúdo gástrico. Essas sondas, feitas de materiais como silicone ou polivinil, são inseridas pela cavidade oral (sonda orogástrica) ou nasal (sonda nasogástrica) e seguem até o estômago. Em neonatos, a escolha entre inserção nasal ou oral depende da condição respiratória e anatômica do paciente⁽⁶¹⁾.

A fixação adequada dessas sondas é essencial para evitar deslocamentos acidentais, que podem resultar em complicações, como aspiração pulmonar. O método mais recomendado envolve o uso de fitas hipoalergênicas para fixar a sonda na bochecha ou na região nasal, minimizando o risco de deslocamento e garantindo a permeabilidade das vias aéreas, como mostrado na Figura 1. De acordo com diretrizes clínicas do Conselho Federal de Enfermagem (2019)⁽⁶²⁾, é importante verificar regularmente o posicionamento da sonda e sua integridade, especialmente durante procedimentos de reposicionamento ou manipulação do recém-nascido.

Figura 1 - Fixação de sonda orogástrica/nasogástrica



Fonte: Imagens disponíveis livremente na internet

Essas sondas orogástricas ou nasogástricas podem prejudicar a visualização de pontos faciais importantes, como boca, bochechas e sulco nasolabial, dificultando a análise precisa da expressão facial de dor por métodos computacionais ou clínicos.

2.1.2 CPAP

O CPAP (*Continuous Positive Airway Pressure*) nasal é amplamente utilizado em neonatos para manter a complacência pulmonar, especialmente em casos de prematuridade e deficiência de surfactante, ajudando a prevenir complicações respiratórias graves, como a displasia broncopulmonar. Ao aplicar uma pressão positiva contínua nas vias aéreas, o CPAP facilita a manutenção dos alvéolos abertos, promovendo trocas gasosas eficazes e diminuindo a necessidade de ventilação mecânica invasiva. O uso profilático imediato, logo após o nascimento, tem mostrado benefícios, especialmente em neonatos com risco aumentado de desenvolver síndrome do desconforto respiratório⁽⁶³⁾.

A aplicação do CPAP nasal em neonatologia ocorre por meio de prongas nasais conectadas a um sistema de ventilação não invasiva. Essas prongas, quando posicionadas adequadamente nas narinas, garantem a entrega de pressão positiva com menor risco de deslocamento. Além disso, o monitoramento contínuo da pele ao redor do nariz e a verificação regular do posicionamento das prongas são essenciais para prevenir traumas e garantir a eficácia do suporte ventilatório⁽⁶³⁾.

A fixação é realizada utilizando fitas adesivas hipoalergênicas ou dispositivos como toucas de fixação, que ajudam a manter a estabilidade do equipamento sem comprometer a integridade da pele delicada do recém-nascido. Exemplos de fixações do CPAP encontram-se na Figura 2.

Figura 2 - Fixação do CPAP



Fonte: A e B - Imagens disponíveis livremente na internet / C - Autor (com permissão)

A correta fixação do CPAP também pode afetar a visualização de áreas faciais importantes para a avaliação da dor, como o lábio superior, bochechas, sulcos nasolabiais e, em alguns casos, nariz, região entre os olhos e testa, o que pode comprometer métodos de avaliação de dor baseados na expressão facial.

2.1.3 Cânula traqueal

A cânula traqueal é um dispositivo médico fundamental para o manejo de vias aéreas em neonatos que necessitam de ventilação mecânica invasiva, particularmente quando falham as tentativas de ventilação não invasiva, como o uso de CPAP ou ventilação com pressão positiva intermitente. O procedimento de intubação traqueal envolve a inserção de uma cânula através da glote, utilizando vias orotraqueal ou nasotraqueal, para garantir a permeabilidade das vias aéreas e promover a ventilação artificial⁽⁶⁴⁾.

Em neonatos, a via orotraqueal é frequentemente a via preferencial, pois, além de ser tecnicamente mais rápida, apresenta uma maior taxa de sucesso na primeira tentativa. A inserção pela cavidade oral facilita o acesso direto às cordas vocais e minimiza a obstrução nasal, o que é especialmente importante em neonatos que apresentam dificuldade respiratória aguda ou que necessitam de ventilação prolongada. A cânula traqueal serve como um canal direto entre o ventilador mecânico e os pulmões, permitindo a administração de oxigênio e ventilação assistida de maneira eficaz e contínua, sendo essencial em quadros de insuficiência respiratória neonatal⁽⁶⁴⁾.

A fixação adequada da cânula traqueal é crucial para prevenir deslocamentos, evitar complicações como extubação acidental e garantir ventilação contínua e eficaz. Existem vários métodos de fixação, incluindo o uso de fitas hipoalergênicas dispostas em formato "H" ou "Y", dispositivos de fixação específicos como o NEO-fit ou o Neobar, que minimizam o risco de lesões na pele e oferecem estabilidade à cânula traqueal⁽⁶⁵⁾. Exemplos de fixações da cânula traqueal encontram-se na Figura 3.

Figura 3 - Fixação da cânula traqueal



Fonte: A e B - Imagens disponíveis livremente na internet / C - Autor (com permissão)

Em todos os casos, a verificação constante da posição da cânula é necessária. O treinamento e a padronização dos métodos de fixação da cânula traqueal são fundamentais para reduzir riscos de deslocamento da cânula e garantir a eficácia da ventilação e segurança do paciente. Entretanto, o uso desse dispositivo pode cobrir regiões faciais essenciais para a avaliação da dor por métodos visuais, como a boca, bochechas e sulcos nasolabiais, comprometendo a análise facial.

2.1.4 Protetor ocular para fototerapia

A fototerapia é um tratamento essencial para neonatos com diagnóstico de hiperbilirrubinemia (icterícia), uma condição comum nos primeiros dias de vida dos recém-nascidos. A fototerapia utiliza luz azul de alta intensidade para converter a bilirrubina em formas solúveis, facilitando sua excreção. No entanto, a exposição prolongada à luz intensa pode causar danos à retina, sendo fundamental o uso de protetores oculares adequados para prevenir complicações oculares, como degeneração retiniana⁽⁶⁶⁾.

A fixação dos protetores oculares deve ser firme, mas ao mesmo tempo confortável, para garantir que o dispositivo permaneça em posição, mesmo quando o bebê se movimenta. Além disso, o material de bloqueio da luz utilizado nos protetores oculares é crucial para evitar que a luz penetre e cause danos, sendo normalmente composto por tecidos opacos com camadas de espuma, que bloqueiam completamente a radiação luminosa prejudicial⁽⁶⁶⁾. Exemplo de protetores oculares encontram-se na Figura 4.

Figura 4 - Protetores oculares para fototerapia



Fonte: A e B - Imagens disponíveis livremente na internet / C - Autor (com permissão)

Devido à forma de fixação dos protetores oculares, é importante notar que algumas áreas faciais, como os olhos, sobrancelhas, a região entre os olhos e a testa, ficam cobertas durante o uso desses protetores. Isso pode comprometer a avaliação de dor com base na expressão facial, especialmente em métodos automatizados que dependem da visualização completa das características faciais.

2.2 BANCOS DE DADOS DE IMAGENS FACIAIS DE RECÉM-NASCIDOS

Nos últimos anos, automatizar o processo de diagnóstico da dor através da análise da expressão facial, que geralmente é feito pelo profissional de saúde à beira do leito, tem sido um tema de pesquisa bastante desafiador, e vem recebendo cada vez mais atenção dos pesquisadores da área de computação, visão computacional, ciências da vida e neurociência⁽⁶⁷⁾.

Esses estudos precisam de bancos de dados com uma vasta quantidade e qualidade de imagens faciais de recém-nascidos para serem assertivos na avaliação da dor, e esta é uma das dificuldades encontradas no meio científico. Fora a importância do número de imagens no banco, a diversidade racial é imprescindível, uma vez que muitos modelos precisam ser treinados com banco de imagens plurirracial⁽⁵⁰⁾.

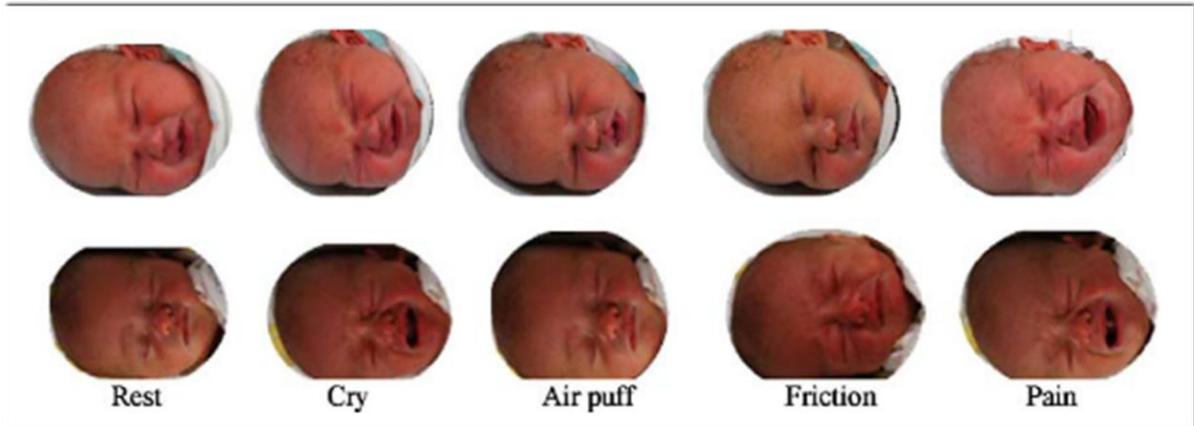
Percebe-se que existem poucas bases de dados voltadas para análise facial de recém-nascidos e crianças. Diferente dos programas de reconhecimento de expressão facial em adultos, que utilizam bancos de dados com grande quantidade de imagens, como o caso do banco de dados do UNBC-McMaster⁽⁶⁸⁾ que contém 48.398 imagens relacionadas à expressão de dor, os bancos de dados de dor em neonatos que estão disponíveis publicamente englobam uma quantidade menor de imagens.

Atualmente, existem alguns bancos de dados para análise da expressão facial de dor em recém-nascidos, e os que estão disponíveis publicamente são:

- O banco de dados iCOPE⁽⁶⁹⁾ (*Infant Classification Of Pain Expressions*) está relacionado à avaliação da expressão facial de dor em recém-nascidos (Figura 5) e contém 200 imagens coloridas de expressão facial de dor de 26 neonatos caucasianos, sendo 63 capturadas com o neonato em repouso, 18 capturadas durante

o choro, 23 capturadas durante estímulos de ar, 36 capturadas durante fricções e atritos, e 60 capturadas durante um procedimento doloroso. Pode-se perceber que, mesmo sendo um banco de dados público, fornece uma quantidade limitada de imagens quando comparado aos bancos de dados faciais de adultos.

Figura 5 - Exemplo do conjunto de dados iCOPE



Fonte: Brahnham *et al.*, 2006⁽⁶⁹⁾

- O banco de dados *Acute Pain in Neonates database* (APN-db)⁽⁷⁰⁾ é um banco de dados de expressões faciais de dor neonatal, composto por mais de 200 vídeos de 112 neonatos e crianças. As imagens foram capturadas durante um procedimento doloroso, sendo gravados 30 frames por segundo durante 2 minutos antes do procedimento e até 1 minuto depois, com o recém-nascido em seu leito ou incubadora. No entanto, o artigo que apresenta essa base de dados não mostra muitas imagens de exemplo, e também não apresenta os dados clínicos da amostra.

- O banco de dados *Facial Expression of Neonatal Pain* (FENP)⁽⁶⁷⁾ contém 11.000 imagens de expressões faciais associadas a 106 neonatos, e abrange quatro categorias de expressões, como: expressão de dor severa, expressão de dor leve, expressão de choro e expressão de calma, em que cada categoria contém 2750 imagens de expressão facial neonatal (Figura 6). No entanto, todas as imagens são de neonatos chineses, com características raciais bem marcantes, o que pode dificultar um pouco a correlação com outras bases de dados.

Figura 6 - Exemplo do conjunto de dados FENP



Fonte: Yan *et al.*, 2020⁽⁶⁷⁾

Há também o banco de dados disponível livremente pelo YouTube, que foi utilizado em 2014 para um estudo de revisão sistemática com objetivo de analisar o conteúdo de vídeos que mostravam crianças com menos de 12 meses de idade sendo imunizadas através de injeções intramusculares⁽⁷¹⁾. Este trabalho analisou 142 vídeos que tinham sido enviados por pais, enfermeiros, entre outros. No entanto, não foram vídeos controlados, que pudessem transmitir um grau de confiabilidade para análise de dor.

Um outro banco de dados de expressão facial de dor neonatal baseado em vídeo foi apresentado por Zamzmi *et, al* no ano de 2016⁽³⁹⁾. A amostra é composta por 18 recém-nascidos com idade gestacional média de 36 semanas, que foram filmados durante um procedimento doloroso (punção de calcanhar). Na época, os autores relataram estar trabalhando na coleta de vários dados comportamentais e fisiológicos, de aproximadamente 300 bebês, durante procedimentos dolorosos agudos e crônicos. Hoje, este grupo de estudos⁽⁷²⁾ já trabalha com um conjunto de dados multimodais de dor neonatal que contém respostas visuais, vocais e fisiológicas pós procedimentos dolorosos processuais e pós-operatórios clinicamente requeridos, de 58 neonatos (27 a 41 semanas de idade gestacional) internados em UTIN (Figura 7), denominado USF-MNPAD-I (*University of South Florida Multimodal Neonatal Pain Assessment Dataset*). Este banco de dados propôs avaliar a face do neonato em um ambiente clínico real e espera-se que em alguns casos os recém-nascidos estejam usando dispositivos necessários para a manutenção da vida que ficam fixados em sua face.

Figura 7 - Exemplo do conjunto de dados USF-MNPAD-I



Fonte: Salekin *et al.*, 2021⁽⁷²⁾

Todos estes bancos de dados de imagens faciais de recém-nascidos, alguns amplamente utilizados na pesquisa científica, apresentam limitações significativas, como o número reduzido de imagens disponíveis, a predominância de uma única etnia, e a falta de informações claras sobre aprovação em comitês de ética. Além disso, a maioria das bases de dados contém imagens de recém-nascidos a termo, o que restringe sua aplicabilidade a estudos envolvendo recém-nascidos enfermos, internados em UTIN. Alguns bancos, como o *Newborn Baby Heart Rate Estimation Database* (NBHR)⁽⁷³⁾, fornecem imagens faciais, mas seu foco principal está na monitorização de sinais fisiológicos, não na avaliação de dor.

É importante destacar que as bases de dados disponíveis não incluem, em sua maioria, imagens de neonatos com dispositivos médicos fixados na face, exceto pelo USF-MNPAD-I⁽⁷²⁾, que captura imagens em ambientes clínicos reais. Essa lacuna limita o desenvolvimento de métodos automatizados para avaliação de dor em neonatos, uma população que requer cuidados específicos e abordagens adaptadas.

Diante dessas limitações, dentre os bancos de dados possíveis e disponíveis publicamente, foi escolhido utilizar o banco de dados iCOPE para o desenvolvimento desta tese, pois os autores concederam autorização para o uso das imagens (ANEXO B). Além dele, foram utilizados outros três bancos de dados que não estão disponíveis publicamente: dois deles foram criados especificamente pela autora desta tese, denominados aqui de UNIFESP-I (com autorização formalizada no ANEXO C) e

UNIFESP-II (autorização no ANEXO D). O terceiro, denominado SynUnifesp, foi desenvolvido por um dos pesquisadores do Laboratório de Processamento de Imagens do Centro Universitário FEI, contendo imagens geradas artificialmente a partir do banco de dados UNIFESP-I (autorização de uso ANEXO E).

Esses bancos de dados, que não possuem acesso público, foram desenvolvidos com alto rigor metodológico e têm aprovação ética devidamente documentada. Embora o volume de imagens seja limitado, eles se destacam pela diversidade racial miscigenada e incluem tanto neonatos a termo quanto prematuros. Além disso, o banco SynUnifesp, que contém imagens geradas artificialmente, amplia as possibilidades de análise, proporcionando maior flexibilidade para o treinamento de modelos automatizados de avaliação de dor neonatal.

Nos próximos tópicos, cada um desses bancos de dados será descrito detalhadamente, explicando sua estrutura e relevância para o estudo em questão.

2.2.1 Bancos de dados de imagens faciais UNIFESP-I

Nos anos de 2009 a 2013 uma pesquisa com o objetivo de desenvolver um *software* para identificar a expressão facial de dor de recém-nascidos em tempo real foi realizada⁽³³⁾ e, através dela, o “Banco de Imagens Faciais de Recém-Nascidos da UNIFESP-I” foi criado.

Este Banco de Dados, denominado aqui de UNIFESP-I, contempla principalmente recém-nascidos a termo, que não fazem uso de dispositivos fixados à face. No entanto, é uma base que está sendo usada em diversas pesquisas pela autora e possibilitou com que parte da metodologia deste projeto fosse avançada. Este banco não está disponível publicamente, e só pode ser usado pela autora principal e pesquisadores colaboradores no projeto.

Este banco de dados (Figura 8) foi construído para um projeto do doutorado, realizado no Brasil, na Cidade de São Paulo, na Universidade Federal de São Paulo – UNIFESP⁽³³⁾.

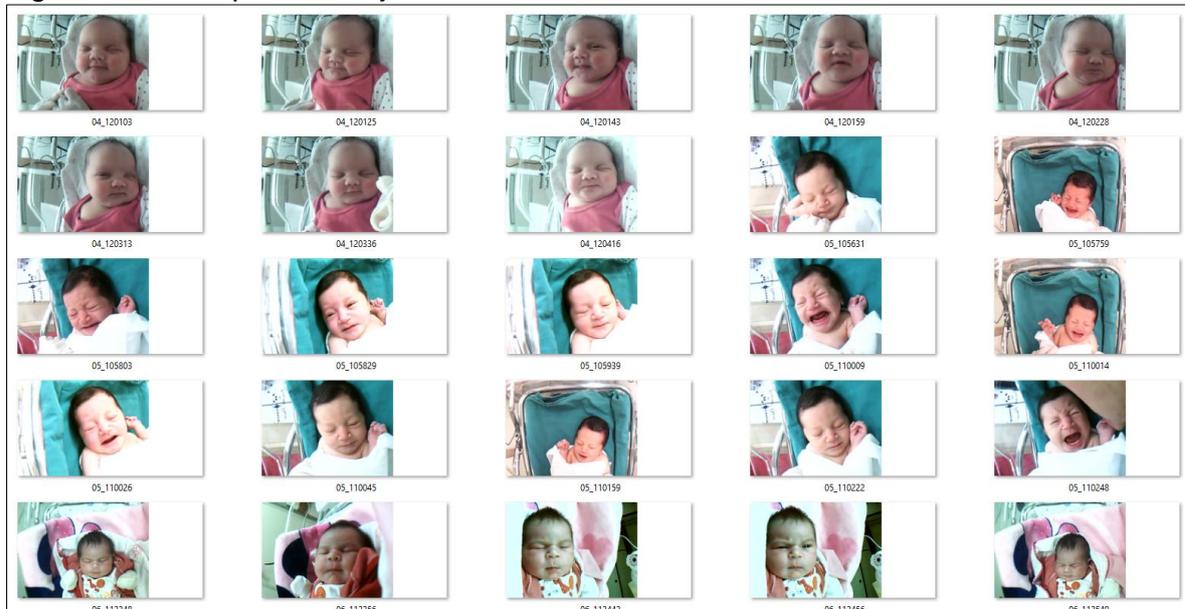
O estudo teve como objetivo desenvolver um *software* para identificar a expressão facial de dor do recém-nascido em tempo real, a ser utilizado nas unidades de terapia intensiva, para auxiliar os profissionais de saúde a monitorarem a presença

de dor e a necessidade de intervenção, denominado "*Software de Avaliação da Expressão Facial de Dor de Recém-Nascidos*" (SAD-RN).

O estudo teve início após aprovação do projeto pelo Comitê de Ética em Pesquisa da Universidade Federal de São Paulo - UNIFESP-EPM, no ano de 2009, sob o número 1299/09.

O projeto recebeu apoio financeiro da FAPESP (Fundação de Amparo à Pesquisa do Estado de São Paulo - Processo: 2012/50157-0 - Auxílio Pesquisa Regular), e a autora recebeu uma bolsa de estudos da CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior)⁽³³⁾.

Figura 8 - Exemplo do conjunto de dados UNIFESP-I.



Fonte: Heiderich, 2013⁽³³⁾

Para capturar as imagens faciais dos recém-nascidos, foram utilizadas 3 câmeras da marca *Foscam* (IP Wireless/Wied Camera), com as seguintes configurações: *high sensitive microphone*, *300k pixels*, *digital remote pan/tilt*, *color*, *wireless*, *night vision*, *motion detect*, *wi-fi*, *MJPEG video compression*. Estas câmeras foram posicionadas da seguinte forma: uma para capturar imagens frontais do recém-nascido; uma para capturar imagens do lado direito do neonato, caso ele virasse o rosto para o lado direito; uma para capturar imagens do lado esquerdo do neonato, caso ele virasse o rosto para o lado esquerdo (Figura 9).

Figura 9 - Posição das câmeras



Fonte: Heiderich, 2013⁽³³⁾

As imagens de cada recém-nascido foram adquiridas no próprio local onde o recém-nascido se encontrava alojado, sem que o mesmo fosse exposto à luz intensa, a fim de não causar desconforto. As imagens foram capturadas automaticamente, durante uma filmagem, por um *software* desenvolvido exclusivamente para o projeto⁽³³⁾.

Ao iniciar a filmagem foi solicitado ao profissional de saúde, que realizava o procedimento doloroso, que informasse ao pesquisador o momento exato da invasão da pele. Esta hora exata do momento da punção foi anotada manualmente (hora/ minuto/ segundo), e possibilitou identificar quais imagens foram adquiridas antes, durante e após o procedimento doloroso.

Foram coletadas imagens faciais de 30 recém-nascidos brasileiros, de raça miscigenada, com idade gestacional média de 37 ± 1 semana, com variação de 35-41 semanas, sendo sete pacientes prematuros tardios e 23 nascidos a termo. O peso ao nascer foi de 2962 ± 593 g, com variação de 2120 a 4100g, sendo sete recém-nascidos de baixo peso. Metade dos recém-nascidos avaliados era do sexo masculino. O parto foi vaginal em 11 e cesáreo em 19 pacientes. A mediana e a variação do boletim de Apgar no primeiro minuto de vida foi 9 (6-9) e, no quinto minuto, 9 (7-10). No momento da aquisição das imagens para o estudo, os neonatos tinham 2 ± 1 dia de vida (variação: 1-6 dias de vida).

Os procedimentos dolorosos realizados nesse grupo foram: duas injeções intramusculares na coxa para administração de vacina para hepatite B; três punções capilares no calcâneo para obtenção de glicemia; 12 punções venosas no calcâneo

para coleta de teste de triagem para erros inatos do metabolismo e 13 coletas de sangue por punção venosa no dorso da mão para coleta de exames solicitados pelos neonatologistas que cuidavam do paciente.

Dos 30 pacientes, seis estavam na UTIN e 24 no Alojamento Conjunto. Toda a aquisição de imagens só foi feita após consentimento informado da mãe do recém-nascido, com explicação detalhada do procedimento e a não aplicação de medidas de analgesia não farmacológica⁽³³⁾.

Foram capturadas e armazenadas 3.900 imagens faciais de 30 recém-nascidos, sendo que 935 (24%) foram coletadas antes da punção, 30 (0,8%) foram do momento exato do procedimento doloroso (invasão da pele) e 2935 (75,2%) foram coletadas após o procedimento doloroso, em tempos variados.

O tempo de captura das imagens antes do procedimento durou em média 07:15 min.; o tempo de captura das imagens após o procedimento durou em média 08:05 min.; e o tempo médio de procedimento foi 15:20 min. (Tabela 1).

Tabela 1 - Características das imagens dos neonatos - banco de dados UNIFESP-I (continua)

Nº	Data	Tempo de captura antes da punção	Tempo de captura depois da punção	Tempo total de captura das imagens	Nº de fotos entre o início do experimento e a punção	Nº de fotos entre a punção e o fim do experimento	Total de fotos
1	07/06/2013	00:14:00	00:07:30	00:21:30	40	91	131
2	10/06/2013	00:07:43	00:03:09	00:10:52	12	16	28
3	12/07/2013	00:12:00	00:02:43	00:14:43	43	10	53
4	12/07/2013	00:09:36	00:08:11	00:17:47	79	148	227
5	16/07/2013	00:03:46	00:08:19	00:12:05	16	126	142
6	18/07/2013	00:04:29	00:07:54	00:12:23	48	121	169
7	31/07/2013	00:06:34	00:09:55	00:16:29	19	116	135
8	30/07/2013	00:22:19	00:09:46	00:32:05	27	150	177
9	31/07/2013	00:08:13	00:09:44	00:17:57	25	104	129
10	31/07/2013	00:01:50	00:09:21	00:11:11	15	124	139
11	31/07/2013	00:06:43	00:05:05	00:11:48	74	22	96
12	01/08/2013	00:09:21	00:13:19	00:22:40	14	65	79
13	01/08/2013	00:02:00	00:09:11	00:11:11	14	135	149
14	02/08/2013	00:06:29	00:05:48	00:12:17	41	70	111
15	02/08/2013	00:02:51	00:09:20	00:12:11	32	198	230
16	02/08/2013	00:16:42	00:06:03	00:22:45	34	25	59
17	05/08/2013	00:10:08	00:08:50	00:18:58	12	110	122
18	05/08/2013	00:04:20	00:08:51	00:13:11	25	123	148

Tabela 1 - Características das imagens dos neonatos - banco de dados UNIFESP-I (conclusão)

19	05/08/2013	00:13:54	00:09:07	00:23:01	20	100	120
20	05/08/2013	00:05:10	00:06:40	00:11:50	59	108	167
21	05/08/2013	00:02:11	00:09:54	00:12:05	9	125	134
22	08/08/2013	00:02:25	00:09:12	00:11:37	19	108	127
23	08/08/2013	00:05:07	00:07:19	00:12:26	20	37	57
24	08/08/2013	00:02:46	00:08:54	00:11:40	23	76	99
25	09/08/2013	00:07:22	00:06:45	00:14:07	74	98	172
26	09/08/2013	00:06:25	00:07:11	00:13:36	46	91	137
27	09/08/2013	00:11:57	00:07:26	00:19:23	19	61	80
28	12/08/2013	00:04:41	00:09:09	00:13:50	9	104	113
29	13/08/2013	00:03:12	00:09:26	00:12:38	15	169	184
30	13/08/2013	00:03:27	00:08:21	00:11:48	52	134	186
Total		03:37:41	04:02:23	07:40:04	935	2965	3900
Média		00:07:15	00:08:05	00:15:20	31	99	130
Mínimo		00:01:50	00:02:43	00:10:52	9	10	28
Máximo		00:22:19	00:13:19	00:32:05	79	198	230

Fonte: Heiderich, 2013⁽³³⁾

As imagens capturadas antes do procedimento podem ser classificadas como imagens do recém-nascido sem dor, já que ele permanecia em repouso, mas não podemos descartar a possibilidade de algum desconforto alguns segundos antes da punção, devido à manipulação para o posicionamento adequado para o exame. Exemplos dessas imagens podem ser vistos na Figura 10.

Figura 10 - Exemplo de imagens faciais capturadas antes do procedimento doloroso



Fonte: Heiderich, 2013⁽³³⁾

As imagens capturadas durante o procedimento podem ser consideradas como imagens do recém-nascido com dor, já que o recém-nascido estava passando por um procedimento potencialmente doloroso. No entanto, as imagens capturadas após o procedimento doloroso, tanto podem ser consideradas como recém-nascido com dor, quanto como com leve desconforto, já que o tempo de recuperação e sensibilidade de cada recém-nascido pode variar (Figura 11).

Sendo assim, das 3.900 imagens capturadas, foram selecionadas 360 imagens para serem analisadas por seis especialistas na área de avaliação de dor neonatal. Os critérios de seleção dos especialistas na área de avaliação de dor neonatal foram: profissionais da área da saúde com especialização em neonatologia, que trabalhassem em centros formadores de recursos humanos para atuação em unidades de terapia intensiva neonatal, que concordassem em participar da pesquisa como avaliador assinando um termo de consentimento, e que declarassem sigilo e confidencialidade no resultado pontuado em cada imagem.

Figura 11 - Exemplo de imagens faciais capturadas após o procedimento doloroso



Fonte: Heiderich, 2013⁽³³⁾

Quanto à seleção das 360 imagens, os autores selecionaram 12 imagens de cada recém-nascido, sendo 122 imagens adquiridas antes do procedimento doloroso, durante o repouso; 30 imagens capturadas do momento exato do procedimento doloroso, quando da invasão da pele; e 208 imagens foram adquiridas no momento de recuperação, após o procedimento. As 360 imagens foram dispostas

aleatoriamente em folhas sulfite A4 e impressas a cores. Estas folhas foram entregues individualmente para cada avaliador e foi solicitado que cada um pontuasse cinco movimentos faciais indicativos da presença de dor (fronte saliente, fenda palpebral estreitada, sulco nasolabial aprofundado, boca aberta e estirada).

Quanto ao diagnóstico de dor, 198 imagens foram rotuladas como “sem dor” e 162 imagens rotuladas como “com dor”. Os cinco movimentos faciais foram avaliados individualmente. Se três ou mais avaliadores classificaram um movimento facial como indicativo de dor, esse movimento foi considerado "indicativo de dor". Caso menos de três profissionais concordassem que o movimento era indicativo de dor, ele foi classificado como "não indicativo de dor".

Portanto, o número de imagens em que os movimentos foram classificados como indicativos de dor foi:

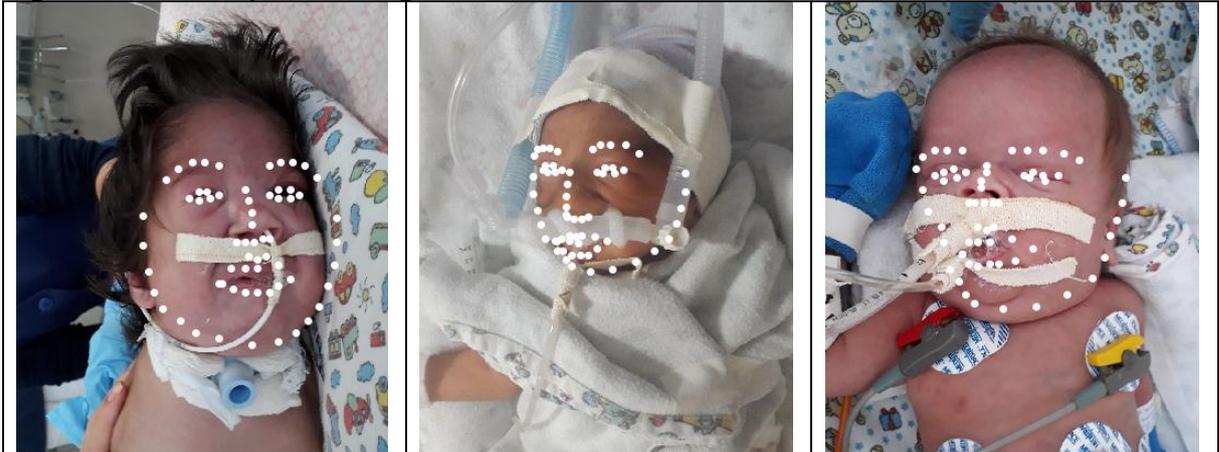
- Boca aberta: 212 imagens.
- Boca estirada: 59 imagens.
- Fronte saliente: 166 imagens.
- Fenda palpebral estreitada: 161 imagens.
- Sulco nasolabial aprofundado: 153 imagens.

2.2.2 Bancos de dados de imagens faciais UNIFESP-II

Nos anos de 2015 a 2020 a pesquisa de Heideirich⁽³³⁾ estendeu-se com o objetivo de validar o SAD-RN para ser utilizado com recém-nascidos prematuros internados na UTIN e, através dela, o segundo “Banco de Imagens Faciais de Recém-Nascidos da UNIFESP” foi criado, denominado UNIFESP-II (Figura 12).

Este banco de imagens é composto por 174 imagens faciais de dez neonatos, coletadas no período de março e abril do ano de 2019, na UTIN do Hospital São Paulo, hospital universitário da Escola Paulista de Medicina da Universidade Federal de São Paulo – UNIFESP-EPM. Este banco não está disponível publicamente, e só pode ser usado pela autora principal e pesquisadores colaboradores no projeto.

Figura 12 - Exemplo de imagens faciais do banco de dados UNIFESP-II



Fonte: Autor (com permissão)

O estudo teve início após aprovação do projeto pelo Comitê de Ética em Pesquisa da Universidade Federal de São Paulo - UNIFESP-EPM, no ano de 2015, sob o número 0566/2015.

Toda a aquisição de imagens só foi feita após consentimento informado da mãe do recém-nascido, com explicação detalhada do procedimento.

Para capturar as imagens faciais dos recém-nascidos, foi utilizada a câmera do celular marca *Samsung Galaxy S10 Lite*, e as imagens foram salvas como arquivo “.jpg”. Ao iniciar a filmagem foi solicitado ao profissional de saúde, que realizava o procedimento doloroso, que informasse ao pesquisador o momento exato da invasão da pele. Esta hora exata do momento foi anotada manualmente (hora/ minuto/ segundo), e possibilitou identificar quais imagens foram adquiridas antes, durante e após o procedimento doloroso.

Foram coletadas imagens faciais de dez recém-nascidos brasileiros, de raça miscigenada, sendo 8 prematuros com idade gestacional média de 32 semanas e 2 a termo de 37 e 40 semanas de idade gestacional. O peso médio ao nascer foi 1880g, com variação de 1040 a 3166g. Dos recém-nascidos avaliados, 50% eram do sexo masculino. O parto foi vaginal em 1 e cesáreo em 9 pacientes. A mediana e a variação do boletim de Apgar no primeiro minuto de vida foi 9 (5-9) e, no quinto minuto 9 (7-10). No momento da aquisição das imagens para o estudo, os neonatos tinham em média 20 dias de vida (variação: 1-80 dias de vida). Dois neonatos foram avaliados duas vezes, a primeira vez com um dia de vida e a segunda vez com doze dias de vida, pois em cada momento faziam uso de um dispositivo diferente fixado à face.

Os procedimentos dolorosos realizados nesse grupo foram: 9 punções capilares no calcâneo para obtenção de glicemia; 1 punção venosa no calcâneo para coleta de teste de triagem para erros inatos do metabolismo; 1 acesso venoso periférico; e 1 retirada de acesso venoso periférico; sendo todos os procedimentos solicitados pelos neonatologistas que cuidavam do paciente.

Foram capturadas e armazenadas 174 imagens faciais de dez recém-nascidos, sendo que 73 (42%) foram coletadas antes da punção, 12 (7%) foram do momento exato do procedimento doloroso (invasão da pele) e 89 (51%) foram coletadas após o procedimento doloroso, em tempos variados.

Na Tabela 2 podemos observar que o tempo de captura das imagens antes do procedimento durou em média 02:03 min.; o tempo de captura das imagens durante e após o procedimento durou em média 41 segs.; e o tempo médio de procedimento foi 02:44 min.

Tabela 2 - Características das imagens dos recém-nascidos do banco de dados UNIFESP-II

Nº	Data	Tempo de captura antes da punção	Tempo de captura durante e após a punção	Tempo total de captura das imagens	Nº de fotos entre o início do experimento e a punção	Nº de fotos entre a punção e o fim do experimento	Total de fotos
1	22/03/2019	00:01:18	00:00:10	00:01:28	8	7	15
2	27/03/2019	00:00:19	00:00:38	00:00:57	4	5	9
3	27/03/2019	00:02:35	00:02:28	00:05:03	5	4	9
4	29/03/2019	00:04:13	00:00:30	00:04:43	4	14	18
5	29/03/2019	00:01:58	00:00:44	00:02:42	4	13	17
6	29/03/2019	00:03:30	00:00:09	00:03:39	6	7	13
7	05/04/2019	00:00:59	00:00:17	00:01:16	9	8	17
7	17/04/2019	00:00:21	00:00:19	00:00:40	4	9	13
8	05/04/2019	00:01:26	00:00:21	00:01:47	11	9	20
8	17/04/2019	00:03:19	00:00:20	00:03:39	5	5	10
9	17/04/2019	00:03:19	00:01:49	00:05:08	5	12	17
10	17/04/2019	00:01:23	00:00:25	00:01:48	8	8	16
Total		00:24:17	00:08:10	00:32:50	73	101	174
Média		00:02:03	00:00:41	00:02:44	6	8	14
Mínimo		00:00:19	00:00:09	00:00:40	4	4	9
Máximo		00:04:13	00:02:28	00:05:08	11	14	20

Fonte: Autor

As imagens capturadas antes do procedimento doloroso podem ser classificadas como imagens do neonato sem dor, já que ele permanecia em repouso, mas, de forma análoga à base anterior, não podemos descartar a possibilidade de algum desconforto segundos antes da punção, devido à manipulação para o posicionamento adequado para o exame.

As imagens capturadas durante e após o procedimento doloroso podem ser consideradas como imagens do recém-nascido com dor, já que o recém-nascido estava passando, ou tinha acabado de passar, por um procedimento potencialmente doloroso. Entretanto, também não podemos descartar a possibilidade de algumas imagens capturadas após o procedimento doloroso serem de desconforto, já que o tempo de recuperação e sensibilidade de cada recém-nascido pode variar.

Sendo assim, 48 imagens do banco UNIFESP-II foram selecionadas para serem avaliadas por quatro especialistas na área de avaliação de dor neonatal. Os critérios de seleção dos especialistas foram: profissionais da área da saúde com especialização em neonatologia, que trabalhassem em centros formadores de recursos humanos para atuação em unidades de terapia intensiva neonatal, que concordassem em participar da pesquisa como avaliador assinando um termo de consentimento, e que declarassem sigilo e confidencialidade no resultado pontuado em cada imagem.

As 48 imagens foram dispostas aleatoriamente em um documento, em formato PDF, e foram entregues individualmente para cada avaliador. Foi solicitado para que cada profissional identificasse se cada um dos cinco movimentos faciais (fronte saliente, fenda palpebral estreitada, sulco nasolabial aprofundado, boca aberta e boca estirada) estava ocluído ou visível, já que o neonato fazia uso de dispositivos fixados em sua face, e quando a região estava visível, foi solicitado que o profissional pontuasse os movimentos faciais indicativos da ausência ou presença de dor.

Uma região facial foi considerada "visível" ou "não ocluída" quando dois ou mais avaliadores concordaram que ela estava claramente visível na imagem. Caso menos de dois avaliadores concordassem, a região foi classificada como "ocluída".

A quantidade de imagens em que cada movimento facial foi considerado visível foi a seguinte:

- Boca aberta: 39 imagens.
- Boca estirada: 38 imagens.
- Fronte saliente: 44 imagens.
- Fenda palpebral estreitada: 44 imagens.
- Sulco nasolabial aprofundado: 29 imagens.

De maneira semelhante, se dois ou mais avaliadores classificaram um movimento facial como indicativo de dor, ele foi considerado "indicativo de dor". Caso contrário, o movimento foi rotulado como "não indicativo de dor".

Assim, o número de imagens em que os movimentos faciais foram considerados indicativos de dor foi o seguinte:

- Boca aberta: 13/39 imagens.
- Boca estirada: 10/38 imagens.
- Fronte saliente: 18/44 imagens.
- Fenda palpebral estreitada: 17/44 imagens.
- Sulco nasolabial aprofundado: 11/29 imagens.

Para o diagnóstico de dor, a avaliação levou em conta a maioria dos movimentos indicativos de dor nas regiões faciais visíveis. Em casos de empate, como quando 4 ou 2 regiões estavam visíveis, optou-se por classificar a imagem como "com dor". No total, 32 imagens foram rotuladas como "sem dor" e 16 imagens foram rotuladas como "com dor".

2.2.3 Bancos de dados de imagens faciais sintéticas SynUnifesp

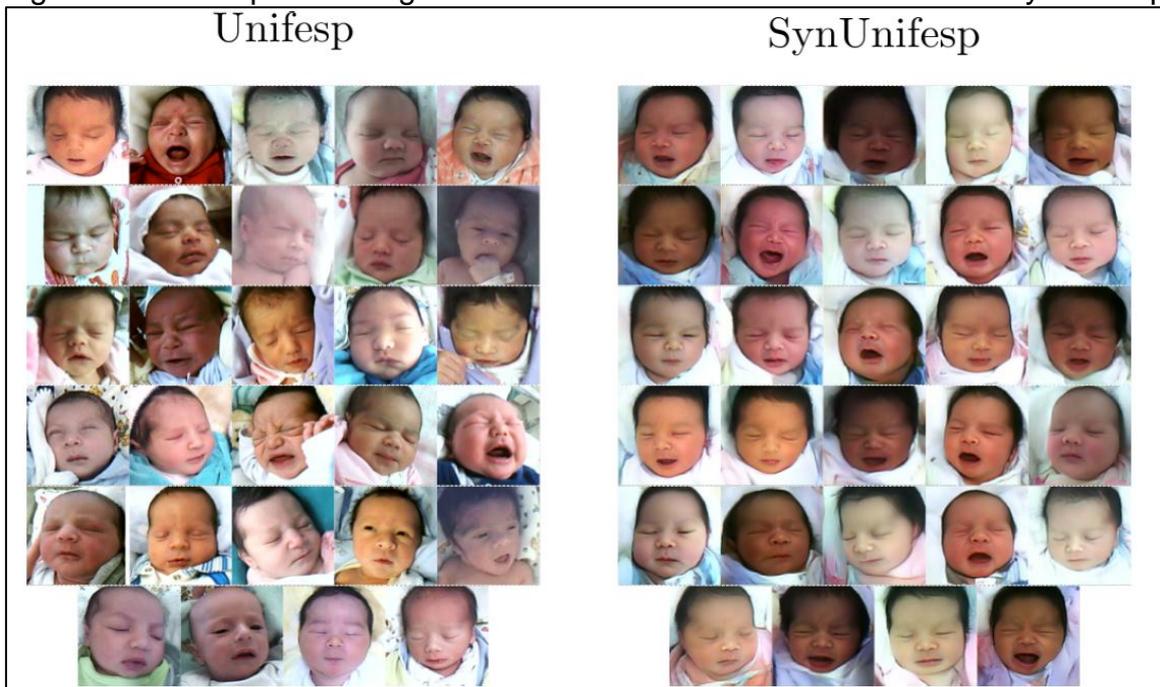
No ano de 2020, pesquisadores do Laboratório de Processamento de Imagens (FEI) observaram a necessidade de aumentar a base de dados de imagens faciais de neonatos. Um dos pesquisadores, Lucas Buziti, que trabalha com criação de imagens sintéticas, propôs e implementou um método computacional para fazer um banco de dados de imagens artificiais a partir das imagens do banco de dados UNIFESP-I⁽⁷⁴⁾.

Este banco de imagens sintéticas foi criado utilizando o método “*Variational AutoEncoder*” (VAE) hierárquico profundo chamado NVAE (*Nouveau VAE*), que se baseia em redes residuais regulares⁽⁷⁵⁾. Para treinar o NVAE, foram utilizadas 3.290 imagens das 3.900 imagens originais, sem rótulos, do conjunto de dados UNIFESP-I.

O NVAE foi treinado para replicar as características clínicas presentes no conjunto de dados original, e obteve 50.000 imagens geradas, incluindo novos indivíduos com características clínicas semelhantes às do conjunto de dados (Figura 13).

Até o momento, este banco não está disponível publicamente, e só pode ser usado pelos pesquisadores colaboradores dos projetos entre a FEI e a UNIFESP. O projeto está em processo de análise para publicação em algumas revistas científicas, faz parte da tese de doutorado do aluno Lucas Buzuti, e foi concedido para uso neste projeto (ANEXO E).

Figura 13 - Exemplo de imagens faciais sintéticas do banco de dados SynUnifesp



Fonte: Buzuti, 2024⁽⁷⁴⁾ (com permissão)

Utilizando o NVAE treinado, foram selecionadas aleatoriamente 356 imagens sintéticas. Essas imagens foram entregues a três profissionais de saúde para serem rotuladas. Os critérios de seleção dos especialistas foram: profissionais da área da saúde com especialização em neonatologia, que trabalhassem em centros

formadores de recursos humanos para atuação em unidades de terapia intensiva neonatal, que concordassem em participar da pesquisa como avaliador assinando um termo de consentimento, e que declarassem sigilo e confidencialidade no resultado pontuado em cada imagem.

As 356 imagens foram dispostas aleatoriamente em um documento, em formato PDF, e foram entregues individualmente para cada avaliador. Foi solicitado para que cada profissional identificasse se cada um dos cinco movimentos faciais (fronte saliente, fenda palpebral estreitada, sulco nasolabial aprofundado, boca aberta e boca estirada) eram indicativos da ausência ou presença de dor. Além disso, os profissionais classificaram a dificuldade de avaliar cada imagem sintética em uma escala *Likert* de 1 a 7, em que 1 significava “extremamente difícil” e 7 indica “extremamente fácil”.

Quanto à dificuldade para avaliar o banco de dados artificial, os profissionais avaliadores não apresentaram dificuldade em avaliar e classificar as faces sintéticas dos neonatos.

Quanto ao diagnóstico de dor, 238 imagens sintéticas foram rotuladas como "sem dor" e 118 imagens rotuladas como "com dor", denominando este conjunto de 356 imagens sintéticas como SynUnifesp.

Os cinco movimentos faciais de frente saliente, fenda palpebral estreitada, sulco nasolabial aprofundado, boca aberta e boca estirada também foram avaliados individualmente. Se dois avaliadores classificaram um movimento facial como indicativo de dor, esse movimento foi considerado "indicativo de dor". Caso menos de dois profissionais concordassem que o movimento era indicativo de dor, ele foi classificado como "não indicativo de dor".

Portanto, o número de imagens em que os movimentos foram classificados como indicativos de dor foi:

- Boca aberta: 114 imagens.
- Boca estirada: 169 imagens.
- Frente saliente: 30 imagens.
- Fenda palpebral estreitada: 174 imagens.
- Sulco nasolabial aprofundado: 116 imagens.

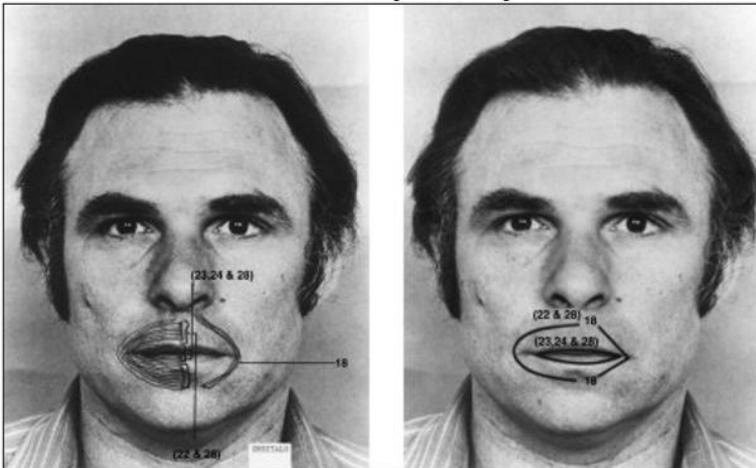
2.3 TECNOLOGIA PARA RECONHECIMENTO AUTOMÁTICO DA DOR

Nesta seção, serão apresentados os principais termos técnicos utilizados ao longo da descrição dos trabalhos relacionados à esta pesquisa, visando proporcionar uma compreensão mais clara e contextualizada dos conceitos fundamentais para o desenvolvimento e aplicação de tecnologias voltadas ao reconhecimento automático da dor neonatal.

2.3.1 Reconhecimento Facial

O reconhecimento facial é uma tecnologia biométrica de processamento de imagens e visão computacional que identifica ou verifica a identidade de um indivíduo com base nas características únicas de seu rosto^(76,77). Um dos métodos mais comuns é o uso de *Action Units* (AUs), formalizado por Ekman e Friesen (1978)⁽⁷⁸⁾, no sistema *Facial Action Coding System* (FACS), que descreve como os movimentos específicos dos músculos faciais estão diretamente associados a diferentes emoções, incluindo a dor (Figura 14).

Figura 14 - Sistema de Codificação de Ações Faciais



Fonte: Ekman e Friesen, 1978⁽⁷⁸⁾

As primeiras abordagens de reconhecimento facial surgiram antes dos anos 1990, focando em características simples da face em imagens homogêneas^(79,80). Com o avanço das tecnologias de codificação de vídeo e a crescente demanda por

precisão, surgiram novos métodos baseados em redes neurais, estatísticas e análise de movimento⁽⁷⁶⁾.

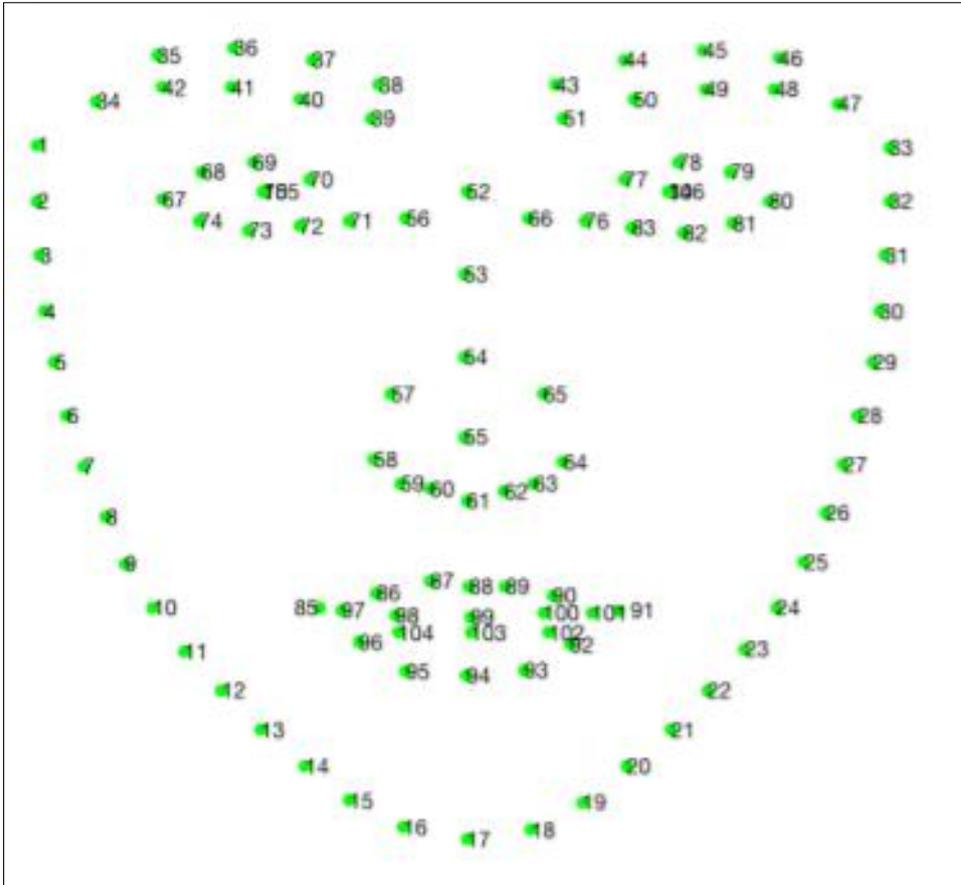
Atualmente, o reconhecimento facial utiliza algoritmos sofisticados que analisam diversos aspectos da face, como cor, textura e a posição de pontos-chave^(33,40,55). O objetivo é detectar a presença de uma face em uma imagem e compará-la a um banco de dados para identificar ou monitorar alterações. No contexto de recém-nascidos em UTINs, o reconhecimento facial pode permitir a detecção automática de alterações faciais que indicam dor, facilitando o acompanhamento contínuo do neonato pelos profissionais de saúde.

2.3.2 Pontos Fiduciais

Os pontos fiduciais (*Landmarks*) são pontos-chave que representam características específicas da face, usados para descrever sua estrutura básica. Esses pontos servem como âncoras para algoritmos de reconhecimento facial, reduzindo a quantidade de informação necessária para identificar ou monitorar expressões faciais. A quantidade de *landmarks* pode variar, desde um número pequeno, como cinco pontos (olhos, boca e nariz), até configurações mais complexas com 194 pontos ou mais⁽⁸¹⁾.

As bases de dados de *landmarks* faciais revistas por Sagonas *et al.* em 2016⁽⁸²⁾ apresentam de 5 a 194 pontos fiduciais; a base de dados de Liu *et al.* (2019)⁽⁸³⁾ possui anotações de 106 pontos fiduciais (Figura 15).

Figura 15 - Pontos fiduciais

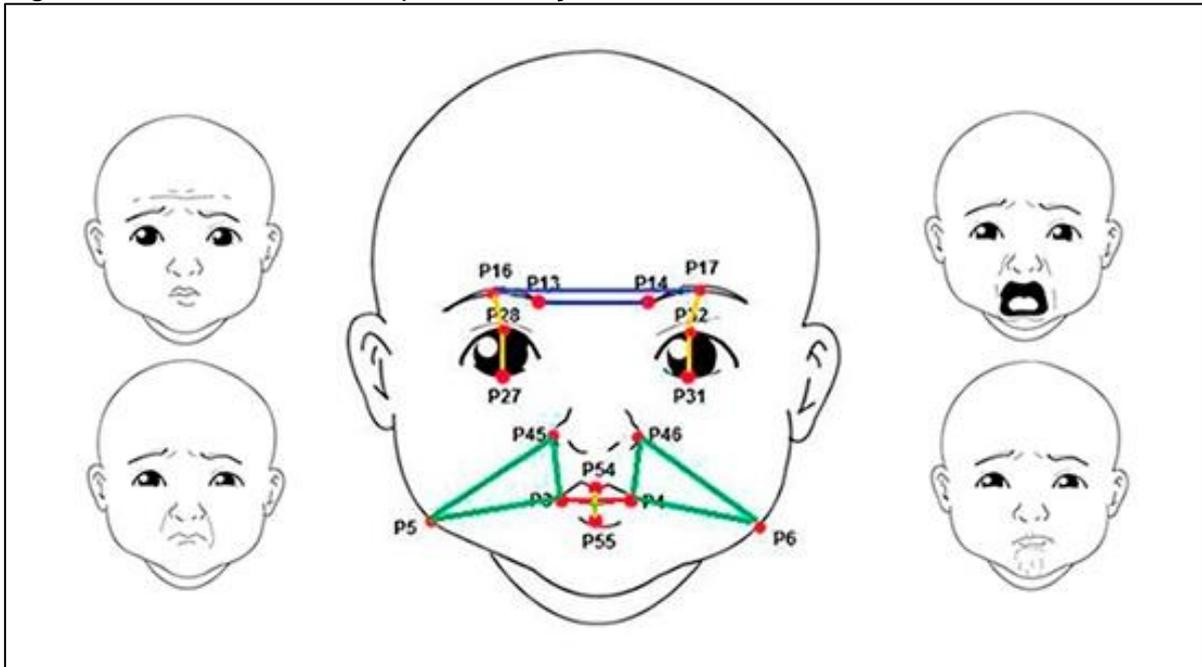


Fonte: Liu *et al.*, 2019⁽⁸³⁾

A ideia de usar *landmarks* remonta ao estudo de padrões faciais e à necessidade de reduzir a complexidade no reconhecimento facial. Ao longo dos anos, algoritmos de aprendizado de máquina foram refinados para identificar esses pontos com maior precisão. Como o caso do *MediaPipe*⁽⁸⁴⁾ da Google, que possui um modelo denominado "*Facemesh*"⁽⁸⁵⁾ que foi amplamente pré-treinado e proporciona uma densidade maior de pontos fiduciais comparado a outros métodos.

Na avaliação de dor os *landmarks* são essenciais para o monitoramento das expressões associadas à dor. Em recém-nascidos, as alterações em áreas como os olhos, boca e testa podem indicar desconforto, tornando os pontos fiduciais uma ferramenta crucial para a detecção automática da dor^(33,38,86) (Figura 16).

Figura 16 - Pontos fiduciais para detecção automática da dor neonatal



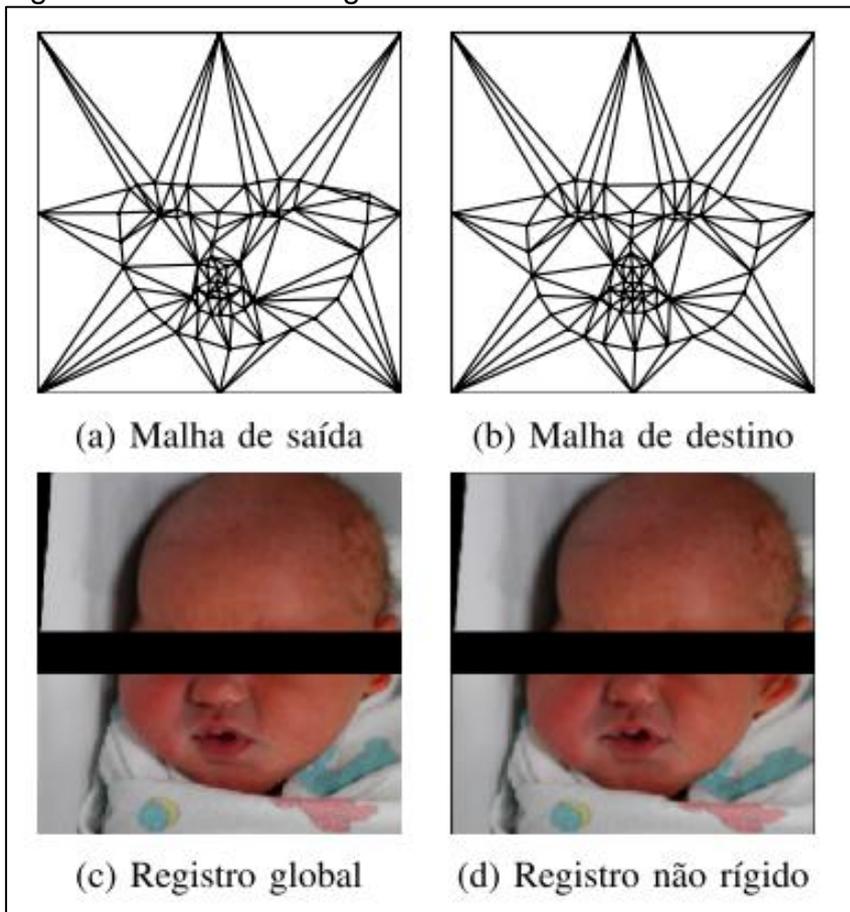
Fonte: Heiderich, 2013⁽³³⁾

2.3.3 Malhas Triangulares

As malhas triangulares são uma técnica usada para modelagem geométrica, em que a superfície da face é dividida em pequenos triângulos, a partir de pontos fiduciais médios, permitindo uma análise detalhada das características faciais (Figura 17).

Em um estudo de avaliação de dor neonatal⁽⁴³⁾, essa técnica foi utilizada para criar um atlas de faces neonatais que serviu como uma base normalizada para a análise de expressões faciais. Utilizou uma técnica de transformação afim para o alinhamento das imagens que preservou proporções e ângulos, permitindo translação, rotação e escalonamento da face para garantir que as imagens estivessem devidamente alinhadas antes das análises. Esta técnica permitiu que as expressões faciais de dor fossem comparadas de forma padronizada, aumentando a precisão das análises automatizadas e facilitando a detecção de padrões consistentes de dor em neonatos⁽⁴³⁾.

Figura 17 - Malhas triangulares



Fonte: Orona *et al.*, 2019⁽⁴³⁾

2.3.4 Técnicas de Pré-Processamento e Normalização de Imagem

As técnicas de pré-processamento e normalização de imagem são essenciais para preparar os dados antes de aplicá-los em modelos computacionais. A Figura 18 ilustra as etapas de processamento para normalização espacial⁽⁸⁷⁾. Estas técnicas incluem a remoção de ruído, segmentação de regiões de interesse (como a face) e ajustes de escala, ângulo e morfologia facial. Essas técnicas são fundamentais para garantir que as imagens sejam padronizadas, permitindo uma comparação eficaz entre diferentes imagens e a detecção precisa de padrões faciais associados à dor.

Em estudos de avaliação de dor neonatal, essas técnicas são aplicadas para eliminar informações desnecessárias e focar na segmentação precisa da face do neonato, melhorando a acurácia dos modelos na detecção de dor⁽⁴⁰⁾.

Figura 18 - Etapas para normalização espacial de imagens faciais



Fonte: Amaral *et al.*, 2009⁽⁸⁷⁾

2.3.5 Algoritmos de Aprendizado de Máquina

Os algoritmos de Aprendizado de Máquina (*Machine Learning* - ML) são uma categoria de técnicas computacionais que permitem que sistemas minerem os dados e façam previsões ou classificações com base em padrões detectados. O ML teve suas primeiras aplicações na década de 1950, mas evoluiu significativamente com o aumento da capacidade computacional e a disponibilidade de grandes quantidades de dados⁽⁸⁸⁾.

Existem basicamente três tipos de aprendizagem de máquina⁽⁸⁹⁾: 1 - Aprendizado Supervisionado, em que o algoritmo é treinado com um conjunto de dados rotulados, ou seja, os dados de entrada vêm com a resposta correta, e o objetivo é aprender uma função que mapeia entradas para saídas; 2 - Aprendizado Não Supervisionado, em que o algoritmo trabalha com dados não rotulados, e o objetivo é encontrar padrões ou estruturas ocultas nos dados; 3 - Aprendizado por Reforço, em que o algoritmo aprende através de tentativa e erro, recebendo recompensas ou punições com base nas ações que realiza.

Assim, o ML é um campo da inteligência artificial que se baseia no desenvolvimento de algoritmos capazes de aprender a partir de dados e tomar decisões ou realizar previsões automaticamente. Esses algoritmos são ajustados por meio de parâmetros, que são valores internos alterados matematicamente durante o processo de treinamento com o objetivo de otimizar o desempenho do modelo. Em contraste, os hiperparâmetros são definidos antes do treinamento e controlam o funcionamento do algoritmo, como a taxa de aprendizado (*learning rate*) e o tamanho do lote de dados (*batch size*)⁽⁸⁹⁾.

Uma ferramenta de otimização de hiperparâmetros que tem se destacado é a biblioteca Optuna, conhecida por sua eficácia e flexibilidade no ajuste fino de modelos de ML⁽⁹⁰⁾. A biblioteca permite a busca automática de hiperparâmetros enquanto maximiza uma métrica, utilizando um método de amostragem (*sampling*). O *Sampler* padrão, fornecido pela biblioteca, é o *Tree-structured Parzen Estimator* (TPE). O TPE otimiza diversos parâmetros, dentre eles destaca-se o “*Learning Rate*”, que controla a taxa com que o modelo ajusta os parâmetros durante o treinamento; o “*Weight Decay*”, que é uma técnica de regularização usada para evitar sobreajuste, penalizando pesos excessivos; e o “*Batch Size*”, que define a quantidade de amostras de dados usadas em cada iteração de treinamento. A escolha adequada desses hiperparâmetros é fundamental para garantir o desempenho do modelo na detecção de expressões faciais de dor em neonatos⁽⁹⁰⁾.

Em estudos de avaliação de dor neonatal, o ML é utilizado para treinar modelos a reconhecer expressões faciais de dor com base em bancos de dados de imagens. Esses algoritmos podem processar grandes volumes de dados de forma automatizada e continuar melhorando conforme novos dados são incorporados⁽⁸⁸⁾.

2.3.6 *Support Vector Machines*

O *Support Vector Machine* (SVM) é um dos métodos de aprendizado supervisionado mais amplamente utilizados para tarefas de classificação. Ele funciona criando um hiperplano que separa os dados em diferentes classes de maneira otimizada. O SVM é conhecido por sua capacidade de lidar com banco de dados pequenos, como os comumente usados na pesquisa neonatal^(41,42,91).

Desenvolvido por Vapnik na década de 1990, embora uma formulação linear antiga tenha sido publicada por Vapnik e Alex Chervonenkis em 1963⁽⁹²⁾, o SVM se tornou uma ferramenta popular em reconhecimento de padrões e análise de dados devido à sua eficácia em encontrar a fronteira de decisão discriminante.

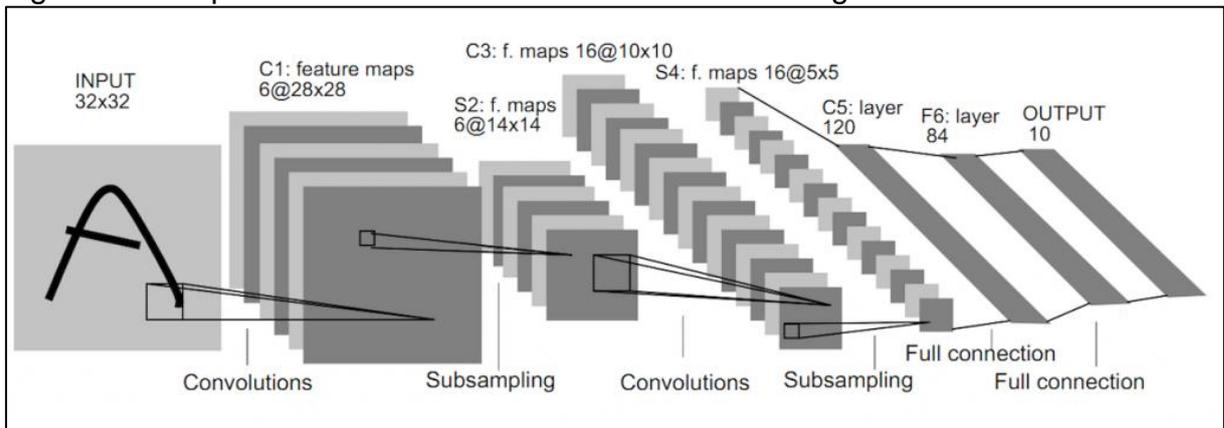
Na classificação de dor em neonatos, o SVM foi utilizado para distinguir entre expressões de dor e expressões neutras com alta precisão, contribuindo para a automatização da detecção de dor com base em imagens faciais^(41,42).

2.3.7 *Convolutional Neural Networks*

As Redes Neurais Convolucionais (*Convolutional Neural Networks* - CNNs) são um tipo especializado de rede neural profunda projetada para o processamento de dados visuais estruturados, como imagens. Introduzidas por Yann LeCun em 1998⁽⁹³⁾, as CNNs revolucionaram o campo de visão computacional e reconhecimento de padrões. LeCun aplicou a arquitetura denominada LeNet-5, para reconhecimento de dígitos manuscritos, demonstrando a eficácia dessas redes em tarefas de classificação de imagens.

A LeNet-5 é composta por sete camadas, excluindo a camada de entrada (Figura 19). Essas camadas incluem convoluções (que extraem características locais da imagem, como bordas e texturas), “*pooling*” (que reduzem a dimensionalidade dos mapas de características, mantendo as informações mais importantes) e camadas totalmente conectadas (que integram as características extraídas para realizar a classificação final)⁽⁹³⁾.

Figura 19 - Arquitetura da LeNet-5 - reconhecimento de dígitos



Fonte: Lecun *et al.*, 1998⁽⁹³⁾

O diferencial das CNNs reside em sua capacidade de aprender características visuais complexas automaticamente, através de camadas convolucionais, que detectam padrões locais nas imagens. Isso contrastava com os métodos tradicionais de aprendizado de máquina, que dependiam de características explicitamente extraídas pelos pesquisadores. Essas redes, otimizadas por meio do algoritmo de aprendizado denominado *backpropagation*⁽⁹³⁾, permitiram que características hierárquicas mais abstratas fossem extraídas das imagens conforme as camadas avançavam, o que é especialmente útil na identificação de microexpressões faciais associadas à dor.

No campo da saúde neonatal, as CNNs têm sido amplamente utilizadas para a detecção automática da dor em recém-nascidos, como no estudo de Brahnem *et al.* (2020)⁽⁴²⁾, em que as CNNs foram aplicadas para melhorar a detecção de dor em neonatos, classificando expressões faciais com alta precisão. Devido à sua arquitetura, as CNNs conseguem capturar variações sutis nas expressões faciais, tornando-as ferramentas úteis para essa tarefa.

Com a evolução das CNNs, especialmente com arquiteturas mais profundas, como a ResNet (*Residual Network*)⁽⁸⁹⁾, que utiliza blocos residuais para facilitar o treinamento, essas redes têm demonstrado resultados superiores em relação a outros métodos de aprendizado de máquina, especialmente em problemas de visão computacional complexos, como a classificação de dor neonatal^(50,94).

O poder das CNNs continua a expandir-se com o aumento da capacidade computacional e o uso de grandes volumes de dados, contribuindo significativamente para o desenvolvimento de sistemas automáticos de monitoramento clínico.

2.3.8 Neonatal - Convolutional Neural Networks

As Redes Neurais Convolucionais Neonatais (*Neonatal - Convolutional Neural Networks - N-CNN*) são variações de CNNs. Foi proposta por Zamzmi *et al.* (2019)⁽⁴⁴⁾ e é uma arquitetura leve, projetada especificamente para avaliar a dor em neonatos. Esta rede convolucional foi desenvolvida para trabalhar de forma eficiente com dados de recém-nascidos, extraindo características faciais e classificando expressões de dor. O modelo é composto por múltiplas camadas convolucionais que capturam informações de imagem, combinando-as em uma única representação robusta da face do neonato.

Nos estudos de Buzuti *et al.* (2020)⁽⁵⁰⁾ e Carlini *et al.* (2021)⁽⁹⁴⁾, as redes N-CNN foram avaliadas para o reconhecimento e classificação de padrões de dor neonatal. A alta acurácia alcançada nesses estudos demonstra a eficácia das CNNs em reconhecer padrões sutis de dor em imagens faciais, facilitando a automação do processo de monitoramento em UTINs.

2.3.9 Transformers

A arquitetura *Transformer*, introduzida por Vaswani *et al.* em 2017⁽⁹⁵⁾, é uma arquitetura de rede neural profunda inicialmente desenvolvida para processamento de linguagem natural, mas que rapidamente se expandiu para outras áreas, incluindo visão computacional e reconhecimento de padrões. O diferencial central do *Transformer* é o seu mecanismo de atenção, que permite ao modelo focar seletivamente em diferentes partes dos dados de entrada, seja uma sequência de palavras ou uma imagem, ao processar a informação. Isso contrasta com as CNNs, que utilizam convoluções fixas para analisar pequenas regiões da imagem de forma sequencial⁽⁹⁶⁾.

Com o desenvolvimento de variantes como o *Vision Transformer (ViT)*, a arquitetura *Transformer* passou a ser aplicada em tarefas de visão computacional. O ViT processa imagens dividindo-as em pequenos blocos chamados *patches*, que são tratados como *tokens* (símbolos), similar ao tratamento de palavras em modelos de linguagem. Essa abordagem permite ao *Transformer* captar relações espaciais entre

diferentes partes da imagem, proporcionando robustez em tarefas de classificação e reconhecimento de padrões visuais⁽⁹⁶⁾.

Uma modificação comum ao ViT para tarefas específicas de classificação é a substituição do *head* de classificação original por uma *Multilayer Perceptron* (MLP). O MLP é um tipo de rede neural composta por múltiplas camadas totalmente conectadas, em que cada neurônio de uma camada está ligado a todos os neurônios da camada seguinte. Essa estrutura permite que o MLP integre as informações extraídas pelo modelo *Transformer* e as transforme em saídas de classificação, como no caso da detecção de dor neonatal. Na prática, um MLP com até duas saídas pode ser utilizado para classificar a presença ou ausência de dor⁽⁹³⁾.

Na análise facial para a avaliação da dor neonatal, os *Transformers* oferecem uma vantagem importante ao processarem e integrarem diferentes partes da face de forma eficaz. Em estudos recentes, como de Sarmadi *et al.* (2024)⁽⁹⁷⁾ e de Domingues (2024)⁽⁹⁸⁾, a aplicação de *Transformers* na análise de imagens médicas, incluindo o reconhecimento de expressões faciais de dor, mostrou resultados superiores aos de arquiteturas baseadas em CNNs. Esses avanços sugerem que os *Transformers*, combinados com MLPs, podem capturar nuances faciais relevantes para a avaliação da dor em neonatos, mesmo em condições desafiadoras, como a presença de dispositivos médicos ou movimentos faciais mínimos.

2.4 CONSIDERAÇÕES COMPLEMENTARES

Os tópicos descritos aqui fecham o capítulo de conceitos fundamentais com uma visão dos principais métodos computacionais aplicados à avaliação automática da dor neonatal. A metodologia baseada em Inteligência Artificial, em particular, representa o avanço mais promissor para a automação e objetivação do diagnóstico de dor em ambientes clínicos complexos.

3 TRABALHOS RELACIONADOS

Este capítulo está dividido em duas seções, cada uma abordando aspectos distintos da avaliação da dor neonatal. Na primeira seção, são discutidos artigos relacionados à dor neonatal e às escalas de avaliação da dor, amplamente utilizadas à beira do leito pelos profissionais de saúde. Esses métodos clínicos manuais fornecem indicadores valiosos para a identificação e manejo da dor em recém-nascidos. Em seguida, a segunda seção apresenta pesquisas focadas na automação desse processo de avaliação, com ênfase em técnicas de análise facial automatizada de imagens. Para uma melhor compreensão do progresso tecnológico, os estudos estão organizados cronologicamente, demonstrando a evolução das metodologias automáticas de avaliação da dor desde seus primórdios até os achados mais recentes.

3.1 AVALIAÇÃO DA DOR EM RECÉM-NASCIDOS

Estudos realizados em fetos humanos e modelos animais demonstram que o sistema nervoso fetal é suficientemente desenvolvido para processar estímulos nociceptivos a partir da 24^a semana de gestação⁽⁹⁹⁾. Todos os componentes anatômicos, fisiológicos e bioquímicos necessários para a experiência dolorosa — transdução, transmissão, modulação e percepção da dor — estão presentes no recém-nascido, inclusive em neonatos pré-termo. Diversas condições clínicas podem precipitar ou agravar a dor, seja de natureza aguda ou crônica⁽¹⁰⁰⁾.

Em neonatos, especialmente os prematuros e aqueles em estado crítico, a dor é uma situação frequente, devido à necessidade de intervenções invasivas essenciais para a sobrevivência. Procedimentos como intubação traqueal, inserção de cateteres venosos periféricos e centrais, monitoramento contínuo e exames laboratoriais, realizados em pele altamente sensível, além de exposição prolongada a luzes e ruídos intensos, contribuem para o aumento da dor e estresse fisiológico⁽¹⁰¹⁾.

Um estudo recente de Harrison e Bueno (2023)⁽¹⁰²⁾ destacou que os procedimentos dolorosos em neonatos fornecem uma visão da reatividade infantil à dor, com a punção de calcâneo sendo identificada como o procedimento mais doloroso e comum em recém-nascidos hospitalizados. Esses achados sugerem que quantificar a intensidade da dor nesses bebês é um passo crucial para avaliar os

efeitos adversos de procedimentos repetidos e implementar estratégias preventivas que minimizem o sofrimento.

Diante desse cenário, a avaliação rotineira e precisa da dor deve ser realizada com a mesma frequência dos demais sinais vitais. No entanto, para que o manejo da dor seja efetivo, é fundamental compreender como o recém-nascido expressa o desconforto e a dor⁽³⁴⁾. O *National Pharmaceutical Council* e a *Joint Commission on Accreditation of Healthcare Organizations* enfatizam que a avaliação da dor deve contemplar três dimensões principais: os aspectos sensoriais, afetivos e cognitivos⁽¹⁰³⁾.

Nesta perspectiva, inúmeras escalas foram criadas e são utilizadas para avaliação de dor no recém-nascido, como as escalas multidimensionais, que utilizam parâmetros fisiológicos (medidas objetivas) e comportamentais (medidas subjetivas) a fim de se obter maior informação a respeito das respostas individuais à dor e de suas possíveis interações com o ambiente^(104,105).

A pesquisa de revisão sistemática, realizada por Giordano *et al.* (2019)⁽¹⁰⁶⁾, fornece uma visão ampla da validade e confiabilidade das escalas de dor e sedação existentes para diferentes populações-alvo e em diferentes contextos clínicos. No total, foram encontradas 65 escalas, sendo que cinquenta e sete escalas (88%) foram úteis para avaliar a dor, 13 escalas (20%) para avaliar a sedação e 4 escalas (6%) para avaliar ambas as condições. Quarenta e duas (65%) foram escalas comportamentais, e 23 (35%) foram escalas multidimensionais. Onze escalas (17%) foram validadas para neonatos em ventilação mecânica. Trinta e sete escalas (57%) foram validadas somente para recém-nascidos prematuros, 24 (37%) para recém-nascidos de termo e prematuros, 7 (11%) para somente recém-nascidos de termo, 7 (11%) para recém-nascidos prematuros, de termo e pequenos para idade gestacional, e 17 (26%) para bebês de termo e pequenos para idade gestacional.

Muitas pesquisas destacam o uso das principais escalas de dor neonatal usadas na beira do leito por profissionais de saúde⁽³⁴⁾. Maxwell *et al.* (2019)⁽¹⁰⁴⁾ e McPherson *et al.* (2021)⁽¹⁰⁵⁾ destacam que essas escalas são fundamentais para garantir um manejo adequado da dor, e citam cinco escalas que são amplamente utilizadas em ambientes de UTIN.

Uma dessas escalas é a *Premature Infant Pain Profile-Revised* (PIPP-R)⁽¹⁰⁷⁾, projetada para neonatos de 26 semanas até o termo. A PIPP-R considera parâmetros fisiológicos, como a frequência cardíaca e a saturação de oxigênio, além de

expressões faciais, como fenda palpebral estreitada e fronte saliente, que são indicativos de dor. Um dos diferenciais desta escala é a sua capacidade de ajustar a avaliação para neonatos prematuros, permitindo uma análise mais precisa da dor em pacientes extremamente vulneráveis. A PIPP-R é ideal para dor aguda, especialmente durante procedimentos dolorosos, como a punção venosa.

Outra escala usualmente empregada é a *Neonatal Infant Pain Scale (NIPS)*⁽¹⁰⁸⁾, usada para neonatos de 28 a 38 semanas de idade gestacional. A NIPS avalia componentes comportamentais como a expressão facial, o choro, o tônus dos braços e pernas, além do padrão respiratório e o estado de alerta do neonato. Esta escala é particularmente eficaz na avaliação da dor aguda, que ocorre durante intervenções clínicas.

A COMFORTneo⁽¹⁰⁸⁾, uma variante da escala COMFORT⁽¹⁰⁹⁾, é usada para avaliar tanto a dor aguda quanto crônica em neonatos entre 24 e 42 semanas. Essa escala é particularmente útil em situações de sedação prolongada, como em neonatos que estão em ventilação mecânica, e avalia parâmetros como a resposta respiratória, pressão arterial, tônus muscular e agitação.

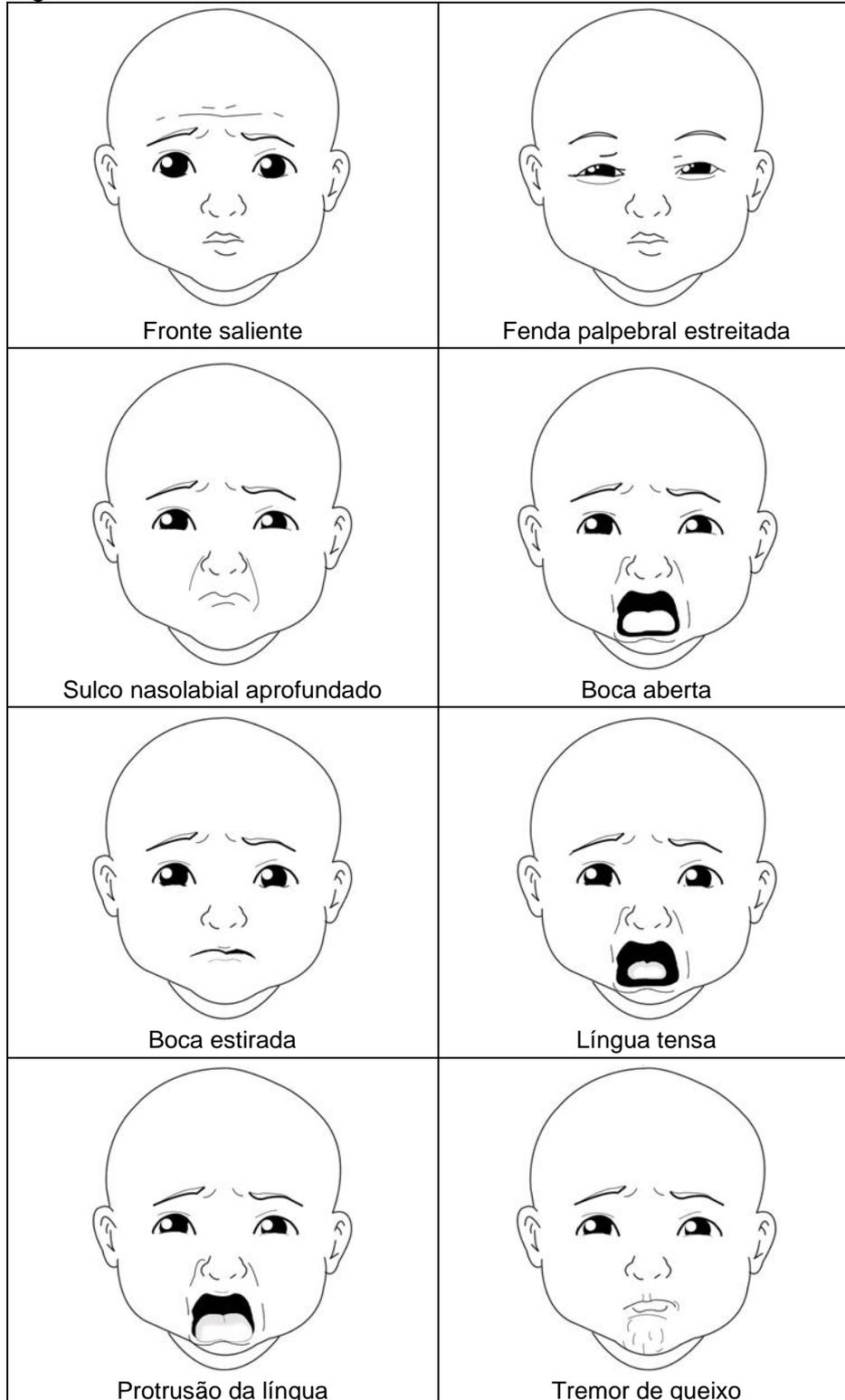
Além dessas, a *Neonatal Pain, Agitation, and Sedation Scale (N-PASS)*⁽¹¹⁰⁾ se destaca por sua versatilidade, sendo capaz de avaliar tanto a dor aguda quanto crônica, além da agitação e sedação em neonatos de até 100 dias. A N-PASS inclui tanto componentes fisiológicos (frequência cardíaca, pressão arterial, saturação de oxigênio) quanto comportamentais (expressão facial e irritabilidade), o que a torna uma ferramenta abrangente para o manejo da dor e sedação.

Outra escala relevante é a *Neonatal Facial Coding System (NFCS)*⁽¹¹¹⁾, que foca nas expressões faciais como principal indicador de dor^(7,59,60). As principais características faciais observadas incluem: Fronte saliente (*brow bulge*); fenda palpebral estreitada (*eye squeeze*); sulco nasolabial aprofundado (*nasolabial furrow*); boca aberta vertical ou horizontalmente (*vertical ou horizontal stretch*); estiramento dos lábios (*lip purse*); tremor do queixo (*chin quiver*) e língua tensionada (*taut tongue*), como observado na Figura 20. O número de características faciais observadas no NFCS pode variar (5, 7 ou 8 características), dependendo do contexto e das necessidades do estudo ou da prática clínica.

A observação da expressão facial constitui um método não invasivo de avaliação de dor em recém-nascidos prematuros e a termo. As características faciais indicativas de dor estão presentes em mais de 90% dos recém-nascidos submetidos

a estímulos dolorosos. O interessante é que as mesmas características não são demonstradas quando estes pacientes são submetidos a um estímulo desagradável, mas não doloroso (7,59,60).

Figura 20 - Características faciais neonatais de dor



Fonte: Heiderich, 2013⁽³³⁾

Um estudo realizado por Vaart *et al.*, no ano de 2019⁽¹¹²⁾, teve como objetivo investigar como as mudanças na expressão facial, o reflexo de retirada, a atividade cerebral, a frequência cardíaca e a saturação de oxigênio estavam relacionadas com os estímulos nocivos, e qual a precisão de cada modalidade de avaliação na discriminação entre estímulos nocivos e não nocivos. Investigaram se a combinação de respostas em cada modalidade melhorava a capacidade de discriminar os estímulos nocivos e não nocivos em 109 bebês que receberam punção no calcanhar clinicamente necessária e um estímulo não nocivo de controle. Como resposta, o modelo que combinou informações em todas as modalidades obteve boa capacidade discriminatória, que era maior do que o poder discriminatório dos modelos construídos a partir de modalidades individuais. Os autores concluíram a pesquisa relatando a importância de incluir múltiplas modalidades na avaliação da dor infantil.

O que chama a atenção neste estudo de Vaart *et al.* (2019)⁽¹¹²⁾ é que a avaliação da expressão facial foi feita pela análise individual de três movimentos faciais, considerados pela escala PIPP-R⁽¹⁰⁷⁾, a saber: fronte saliente, fenda palpebral estreitada e sulco nasolabial aprofundado. As respostas faciais dos bebês foram registradas com uma câmera de mão por 15 segundos antes e 30 segundos depois da punção no calcanhar. Dois pesquisadores treinados em análise da expressão facial avaliaram individualmente a duração do tempo de resposta de cada região facial nos 30 segundos após o estímulo. Os autores⁽¹¹²⁾ observaram que a duração do tempo de resposta da fronte, da fenda palpebral e do sulco nasolabial foi altamente correlacionada. As três expressões faciais individualmente podiam discriminar de forma semelhante entre as condições nocivas e as não nocivas, e combinando as três não produziam maior precisão, sensibilidade, especificidade ou acurácia. Isto demonstrou que a combinação destas três medidas de expressão facial não melhorou a discriminação entre o estímulo nocivo e o não nocivo em relação às medidas isoladas. Neste caso, como o sulco nasolabial teve a maior precisão e acurácia entre as medidas individuais, e como foi a característica com maior poder discriminatório, foi selecionado para uso no modelo multimodal do estudo.

No entanto, em 2021, outra pesquisa publicada pelo mesmo grupo de autores⁽¹¹³⁾, preferiu usar a fronte saliente como região facial de interesse para análise da expressão de dor. Este estudo tinha como objetivo determinar se recém-nascidos entre 28 e 40 semanas de idade gestacional apresentavam respostas fisiológicas, comportamentais, reflexivas e cerebrais discriminatórias a eventos nocivos e não

nocivos e, em caso positivo, investigar como essas respostas mudavam ao longo do desenvolvimento humano inicial. Como resultado, os autores relatam que a duração do tempo de resposta da frente saliente é maior em recém-nascidos acima de 36 semanas e menor em recém-nascidos abaixo de 33 semanas.

Esses dois últimos trabalhos citados utilizam apenas a escala multidimensional PIPP-R⁽¹⁰⁷⁾ para avaliação da dor neonatal. No entanto, é possível observar uma variedade de escalas disponíveis para a avaliação de dor em recém-nascidos apenas pela análise da expressão facial.

A avaliação comportamental da dor neonatal depende estritamente do olhar atento do profissional de saúde. Ao verificar a atenção visual de 143 adultos (84 profissionais de saúde e 59 não profissionais de saúde) ao avaliar a dor neonatal, Barros *et al.* (2021)⁽¹¹⁴⁾ identificaram que os pontos faciais mais observados foram a boca, os olhos e a testa, mas, dentre os 143 avaliadores, houveram 4 classes de adultos: aqueles com os atributos para identificar corretamente a dor e ausência de dor (n=80); aqueles com os atributos para identificar corretamente somente a dor (n=28), aqueles com os atributos para identificar corretamente somente a ausência de dor (n=34); e quem não identificou corretamente nenhuma das situações (n=1). Neste estudo, os autores puderam identificar que ser profissional de saúde e realizar fixação visual no sulco nasolabial ajudou a identificar de forma assertiva a presença/ausência de dor neonatal.

Em um estudo realizado por Silva *et al.* (2021)⁽¹¹⁵⁾, com objetivo de analisar a fixação do olhar de 38 pediatras durante o processo de decisão a respeito da presença/ausência de dor em 20 imagens de recém-nascidos, ao avaliar 4 regiões faciais (boca, olhos, testa e sulcos nasolabiais), os autores observaram que os pediatras fixam preferencialmente o olhar na boca. No entanto, percebeu-se que quanto maior a duração de fixação do olhar na boca e na testa há um aumento da percepção que a dor moderada/severa está presente.

No estudo de Tamanaka *et al.* (2022)⁽¹¹⁶⁾ foi relatada a existência de 52 escalas de avaliação da dor neonatal pela análise da expressão facial. A primeira parte da pesquisa foi uma revisão de literatura que identificou que muitas escalas analisavam o fenômeno careta como característico da presença de dor, mas 13 escalas em particular especificavam as regiões faciais avaliadas para o diagnóstico da dor, sendo essas: *Neonatal Facial Coding System* (NFCS); *NFCS Modificado* (M-NFCS); *McGrath Facial Affective Scale* (MCGRATH); *Nepean Neonatal Intensive Care Unit*

Pain Assessment Tool (NNICUPAT); Modified Infant Pain Scale (MIPS); Partial MIPS (P-MIPS); Premature Infant Pain Profile (PIPP); PIPP Revisited (PIPP-R); Evaluation Infant Douleur (EVENDOL); Acute Pain Scale (APS); Liverpool Infant Distress Scale (LIDS); Pain Assessment Scale for Preterm Infants (PASPI), e InfanPain Assessment Scale for Preterm Infants (BIIP). As regiões faciais avaliadas em cada escala encontram-se na Tabela 3, e foram consideradas 14 regiões faciais, pois o trabalho⁽¹¹⁶⁾ teve como objetivo, além de quantificar as regiões faciais avaliadas pela literatura, analisar quais áreas da face eram visualizadas por profissionais da área da saúde, utilizando um método de captação do olhar, de participantes médicos e não médicos, com um equipamento de rastreamento ocular.

Tabela 3 - Escalas faciais para avaliação de dor e regiões faciais

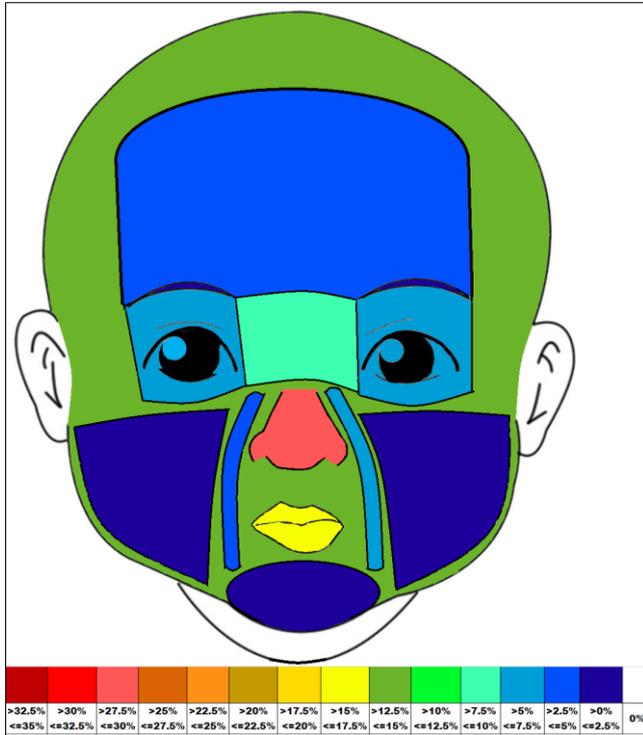
Regiões Faciais	Escalas faciais para avaliação de dor														
	NFCS	M-NFCS	MCG RATH	NNIC UPAT	MIPS	P-MIPS	PIPP	PIPP-R	EVENDOL	APN	LIDS	PASPI	BIIP	Total	Total (%)
Olho Esquerdo	1	1	1	1	1	1	1	1	1	1	1	1	1	13	100%
Olho Direito	1	1	1	1	1	1	1	1	1	1	1	1	1	13	100%
Entre Sobrancelha	1	1	1	1	1	1	1	1	1	0	1	1	1	12	92,3%
Testa	1	1	1	1	1	1	1	1	1	0	1	1	1	12	92,3%
Boca	1	1	1	1	1	1	0	0	1	0	1	1	1	10	76,9%
Sulco Nasolabial Esquerdo	1	1	1	1	1	1	1	1	0	1	0	1	1	11	84,6%
Sulco Nasolabial Direito	1	1	1	1	1	1	1	1	0	1	0	1	1	11	84,6%
Queixo	1	0	1	0	1	1	0	0	0	0	1	0	0	5	38,5%
Sobrancelha Esquerda	0	0	0	0	0	0	0	0	0	1	1	0	0	2	15,4%
Sobrancelha Direita	0	0	0	0	0	0	0	0	0	1	1	0	0	2	15,4%
Nariz	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0%
Bochecha Esquerda	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0%
Bochecha Direita	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0%
Outras Regiões da Face	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0%
Total	8	7	8	7	8	8	6	6	5	7	8	7	7	/	/

Fonte: Autor "adaptada de" Tamanaka *et al.*, 2022⁽¹¹⁶⁾

Um dado importante que deve ser destacado do estudo de Tamanaka *et al.* (2022)⁽¹¹⁶⁾ é a relação entre as regiões faciais abordadas pela literatura e as áreas faciais de interesse que os profissionais de saúde realmente olham quando avaliam um recém-nascido com expressão de dor. O estudo comparou quantitativamente a relevância entre as áreas faciais dos neonatos, encontradas na revisão literária atualizada sobre as escalas de dor, relacionadas com os achados da percepção visual de diferentes grupos de voluntários, incluindo médicos e outros profissionais de saúde, com o intuito de compreender melhor o uso de tais escalas na prática clínica.

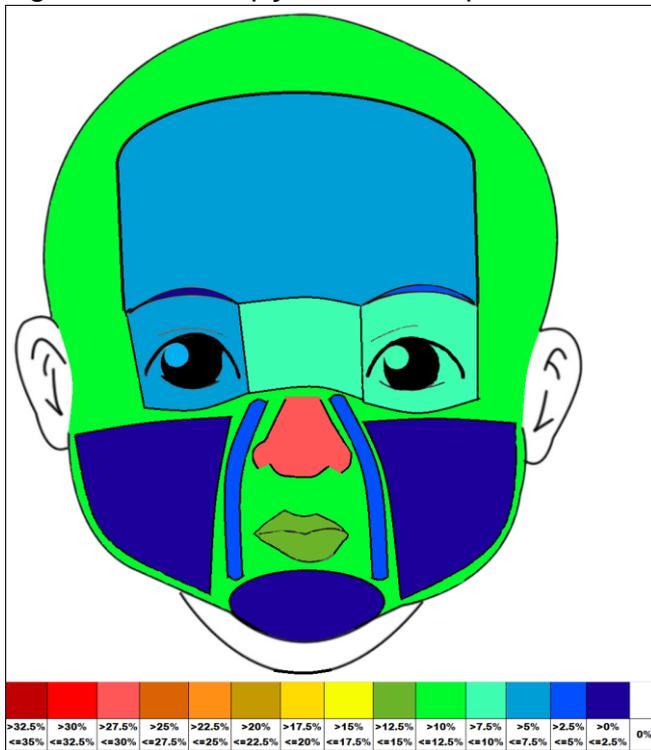
Para uma melhor visualização dos resultados de percepção visual⁽¹¹⁶⁾, uma imagem protótipo da face do neonato foi colorida de acordo com a duração média das fixações dos olhares para cada região da face do neonato, sendo representada por uma escala de cor, de tons frios para tons quentes, sendo o tom branco para as áreas que não foram observadas pelos participantes, até o tom vermelho escuro para as áreas mais visualizadas. Na Figura 21 e na Figura 22 é possível observar que os médicos e os outros profissionais de saúde fixaram seus olhares mais na região do nariz, da boca, das outras regiões da face, dos olhos, e da testa, pois estas são as áreas que apresentaram maior tempo de fixação do olhar durante a avaliação da dor respectivamente, o que difere das sugestões apontadas pela literatura, que enfatizam a necessidade do olhar para os olhos, as sobrancelhas, a testa, a boca, e os sulcos nasolabiais.

Figura 21 - Percepção visual dos médicos



Fonte: Tamanaka *et al.*, 2022⁽¹¹⁶⁾

Figura 22 - Percepção visual de profissionais de saúde



Fonte: Tamanaka *et al.*, 2022⁽¹¹⁶⁾

Com a subjetividade envolvida na avaliação do evento doloroso nessa faixa etária, caracterizada pela incapacidade do neonato em verbalizar, a falta de treinamento profissional específico dificulta o diagnóstico e o tratamento adequado da dor neonatal na prática clínica diária⁽¹⁰⁸⁾.

Apesar de muitos profissionais de saúde, especialmente da área de enfermagem, reconhecerem a presença de dor na população neonatal, sua avaliação ainda é frequentemente realizada de maneira observacional nos cenários clínicos. Esse processo exige capacitação contínua e treinamentos específicos para que as equipes de saúde adquiram o conhecimento e as habilidades necessárias para utilizar de forma eficaz os diversos instrumentos disponíveis. A variedade de escalas de dor e seus diferentes métodos de pontuação requerem uma compreensão técnica apurada para que a avaliação seja precisa e consistente, garantindo assim um manejo adequado da dor nos neonatos⁽¹¹⁷⁾.

Mesmo com a presença de profissionais de saúde atentos e treinados para a avaliação da dor neonatal, um desafio persistente é a frequência com que essa avaliação ocorre. Quando a dor é diagnosticada por meio da observação do comportamento, o neonato depende da disponibilidade do profissional para a análise visual constante. Já nas avaliações baseadas em sinais vitais, é necessário que o profissional esteja continuamente atento às variações para associá-las à dor, o que nem sempre é prático em um ambiente de UTIN. Uma solução promissora para esse problema seria a utilização de ferramentas computacionais capazes de identificar a dor de maneira automática e em tempo real. Esses programas de computador permitiriam uma avaliação sistemática e contínua, oferecendo um diagnóstico mais preciso e oportuno da dor em recém-nascidos criticamente enfermos.

3.2 AVALIAÇÃO AUTOMÁTICA DA EXPRESSÃO FACIAL DE DOR EM RECÉM-NASCIDOS

Diante da necessidade de uma avaliação tão específica e constante na vida do neonato internado em UTIN, foram desenvolvidos métodos computacionais que visam detectar o fenômeno doloroso de forma automática, a fim de auxiliar os profissionais de saúde a monitorarem a presença de dor e a necessidade de intervenção terapêutica. A automatização deste processo é feita pelo reconhecimento facial e

detecção de pontos fiduciais da face que possibilitam a visualização e acompanhamento da movimentação facial⁽¹¹⁸⁾.

Uma das primeiras tentativas de tornar a avaliação da expressão facial de dor do recém-nascido mais automatizada foi no trabalho realizado por Schiavenato *et al.* (2008)⁽¹¹⁹⁾, que avaliou a expressão facial de dor por meio do método de comparação de pontos fiduciais detectado manualmente. Para a comparação, foram utilizadas duas imagens, antes e depois de estímulo doloroso. A similaridade entre as expressões faciais foi analisada a partir de uma amostra de 57 neonatos e mostrou-se coerente quando comparada com os relatórios anteriores de expressão de dor detectados por escalas de dor.

O estudo de Schiavenato *et al.*⁽¹¹⁹⁾, por utilizar a detecção dos pontos fiduciais manualmente, tornava imprescindível que o profissional da saúde tivesse habilidade de apontar (“clique”) sobre os pontos específicos das regiões faciais, o que requirava, conseqüentemente, mais tempo para efetuar a análise. Ou seja, o sistema não permitia que a expressão de dor fosse detectada de forma imediata. Pelo contrário, era necessário que o profissional identificasse a face de dor para depois realizar a análise comparativa com outras faces, para assim obter a quantificação da distância entre os pontos. Por isso, o método empregado tinha grande interesse para a pesquisa clínica, mas não tanto para uso clínico.

Em 2013, Heiderich⁽³³⁾ desenvolveu um *software* para avaliar a dor neonatal, semelhante ao criado por Schiavenato *et al.* (2008)⁽¹¹⁹⁾. No entanto, a aquisição das imagens do recém-nascido e a comparação de pontos fiduciais da face eram realizadas de forma automática, com a utilização de um *SDK* (*Software Development Kit*) desenvolvido pela empresa “Luxand”, denominado “LuxandFaceSDK”, que permitiu a construção de aplicativos para reconhecimento da imagem facial, com funcionalidade baseada na identificação biométrica. Os neonatos eram filmados e as imagens faciais eram capturadas a cada 3 segundos. Cada imagem capturada passava por uma fase de detecção da face e dos pontos fiduciais, que eram sempre comparados no decorrer do tempo, imagem a imagem, e classificados em expressão facial neutra ou expressão de dor, dependendo da pontuação de movimentos faciais abordados pela escala de dor NFCS.

No estudo de Heiderich⁽³³⁾, os neonatos eram comparados entre si, pois era necessária uma foto dele em repouso para indicar a face neutra, e este seria o ponto de partida para a comparação com as imagens seguintes da filmagem, em que o

neonato passaria por um momento agudo de dor, a saber, a punção capilar. O programa foi bem-sucedido no que se refere à possibilidade de capturar as imagens de face, decodificar as imagens faciais de interesse para a avaliação da dor, e recodificar o resultado como dor presente ou ausente em intervalo de tempo quase instantâneo⁽³³⁾.

Uma questão levantada pelos autores⁽³³⁾ está relacionada ao uso de dispositivos fixados à face do recém-nascido, o que impediu a identificação de alguns pontos faciais e, assim, dificultou a identificação da face e dos pontos fiduciais. Sendo assim, os autores relataram a necessidade de aperfeiçoamento na detecção dos pontos fiduciais, pois a aplicação de um sistema informatizado para uso dentro de UTIN também deve ser viável para a monitoração da presença de dor à beira do leito de recém-nascidos criticamente doentes. Recomendaram que, para aplicação prática nestas unidades, seria preciso desenvolver a detecção de outros pontos fiduciais que facilitassem a identificação dos movimentos faciais característicos da presença de dor no neonato, assim como evitassem que movimentos não relacionados, como os de língua, bocejos e outros, fossem caracterizados pelo sistema como movimentos de dor.

Neste sentido, para o campo da pesquisa computadorizada de visão e emoção, a expressão facial humana é um indicativo promissor de comportamento para o reconhecimento da emoção e da dor, e tem sido estudada e utilizada para a avaliação automatizada da dor infantil há anos. Engenheiros e Cientistas da Computação podem programar algoritmos de Aprendizado de Máquina para extrair, rastrear e analisar expressões faciais humanas a partir de imagens gravadas e dados de vídeo para prever a dor e as emoções. Em comparação com os sistemas tradicionais de escala de dor baseados em observadores, o algoritmo de Aprendizado de Máquina tem uma oportunidade única de fornecer detecção e avaliação objetiva e contínua da dor para recém-nascidos e bebês⁽⁸⁸⁾.

Um dos primeiros estudos publicados para avaliação da dor infantil baseado em Aprendizado de Máquina, foi publicado por Zamzmi *et al.* em 2016⁽³⁹⁾, e teve como proposta uma abordagem multimodal automatizada que utilizava uma combinação de dados faciais, corporais e sinais vitais para avaliar a dor dos neonatos. Os autores⁽³⁹⁾ relatam que a combinação de diferentes indicadores para a avaliação da dor (ou seja, multimodal) alcançou a maior precisão geral, de 95%, e sugerem ser viável o

desenvolvimento de uma abordagem automatizada e multimodal para a avaliação da dor em neonatos.

Em 2018, Teruel⁽⁴⁰⁾ aperfeiçoou o trabalho de Heiderich (2015)⁽³⁸⁾, em que os neonatos eram comparados com eles mesmos, e desenvolveu um arcabouço computacional de detecção, interpretação e classificação de padrões em imagens frontais de faces para identificação automática de dor em recém-nascidos, a partir do banco de dados de imagens faciais de recém-nascido UNIFESP-I.

Na metodologia proposta pelo autor, Teruel (2018)⁽⁴⁰⁾, foram utilizadas diversas técnicas de pré-processamento e normalização de imagem, que foram essenciais para discriminar as imagens e permitir as classificações automáticas em espectros contínuos, possibilitando identificar o estado da dor do recém-nascido. Na Figura 23 podemos observar as características mais discriminantes da face, sendo que da esquerda para direita da imagem observa-se a evolução do estado facial sem dor para dor absoluta. A concordância deste modelo em relação à classificação feita por profissionais da área da saúde chegou a 72,77% de acurácia, e acredita-se que a discordância, em alguns casos, tenha origem em fatores de confusão ainda não estudados, como por exemplo, o estresse ou desconforto do recém-nascido.

Figura 23 - Características mais discriminantes da face



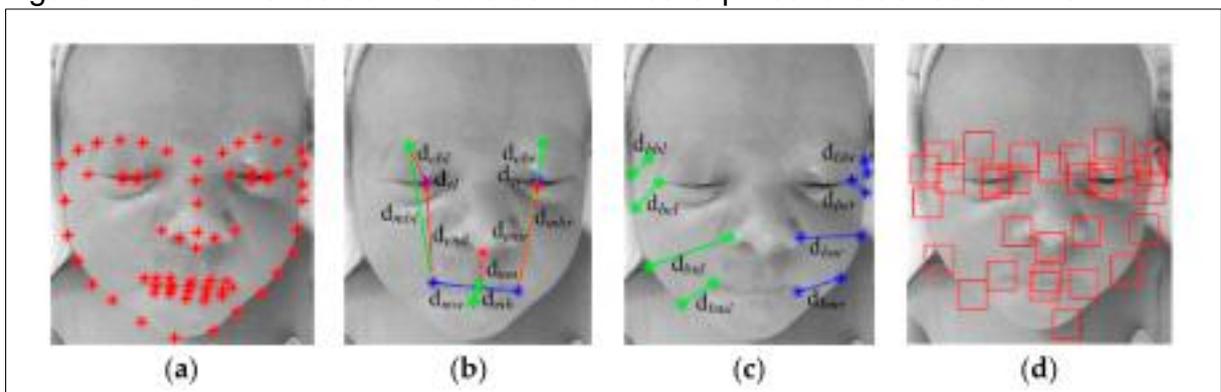
Fonte: Teruel, 2018⁽⁴⁰⁾

Em 2018, Zhi, Zamzmi, e Goldgolf⁽⁴¹⁾ criaram um método dinâmico de representação da atividade facial da dor, incluindo a configuração facial, os parâmetros de textura e a extração de descritores temporais com base em sequência (Figura 24). A representação da atividade facial da dor, incluindo características geométricas faciais e características de textura, foi estudada para o reconhecimento da expressão facial da dor. Para a dimensionalidade dos recursos faciais e reconhecimento da dor, foram utilizados os métodos como o classificador SVM.

Zhi, Zamzmi, e Goldgolf⁽⁴¹⁾ concluem que a avaliação da dor infantil com base no perfil também foi viável, pois seu desempenho foi quase tão bom quanto o de face

total. Observaram que a idade gestacional é o fator mais influente na avaliação da dor, sendo necessário construir modelos específicos para cada idade. Os autores⁽⁴¹⁾ relatam sobre a existência de poucos estudos realizados para analisar e avaliar a dor neonatal usando tecnologias de Visão Computacional e Aprendizado de Máquina, e referem que este fato pode se dar pelo número limitado de bancos de imagens neonatais e a crença de que os algoritmos projetados para avaliar dor em adultos teriam desempenhos semelhantes para os neonatos, o que não acontece na prática.

Figura 24 - Parâmetros de nível de estrutura da expressão facial da dor infantil

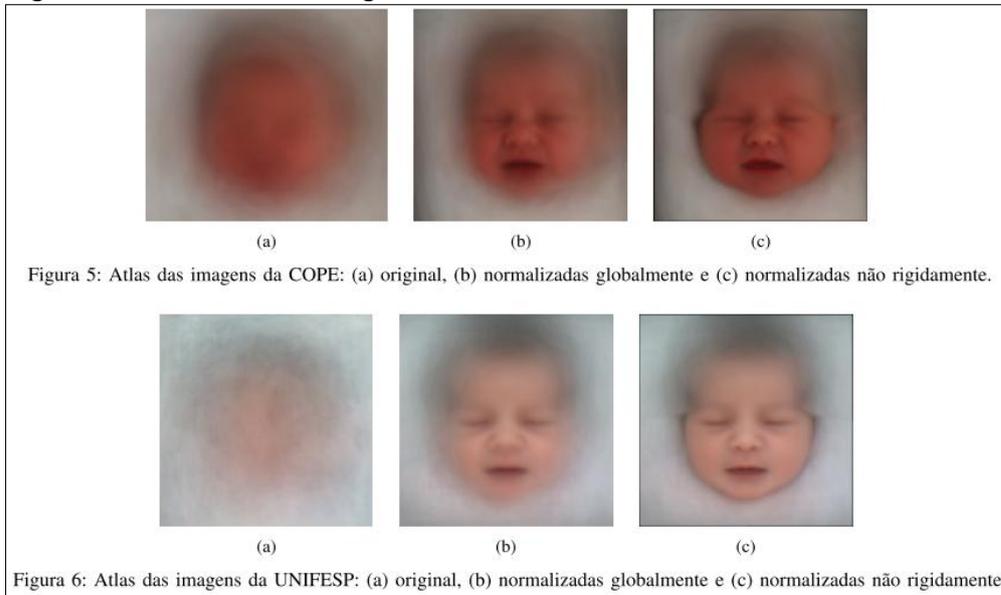


Fonte: Zhi, Zamzmi e Goldgolf, 2018⁽³⁹⁾

Legenda: (a) pontos faciais; (b) parâmetros de configuração facial; (c) parâmetros de postura da cabeça; (d) pontos de referência

Ainda no ano de 2019, Orona *et al.*⁽⁴³⁾ propuseram e implementaram uma estrutura computacional que utilizou malhas triangulares, com o objetivo de gerar um atlas espacialmente normalizado, de alta resolução, e potencialmente úteis para a avaliação automática da dor neonatal. A primeira etapa para construção do atlas foi o pré-processamento das imagens das bases iCOPE e UNIFESP-I, que consistia basicamente em remover as características que dificultariam a realização das outras etapas, como orientação das imagens, quantidade diferente de pixels e erros na forma como as imagens foram capturadas. Foram gerados os atlas de imagens de faces de recém-nascidos através do cálculo da média (Figura 25), com isso, os autores puderam alcançar uma evolução do método, dada a alta resolução dos detalhes faciais apresentados no atlas das imagens geradas, que proporcionou uma imagem colorida de alta definição, que pode ser utilizada como referência em trabalhos que desejem normalizar as bases apresentadas e que não possuem em mãos um atlas de alta resolução.

Figura 25 - Atlas das imagens faciais das bases iCOPE e UNIFESP-I



Fonte: Orona *et al.*, 2019⁽⁴³⁾

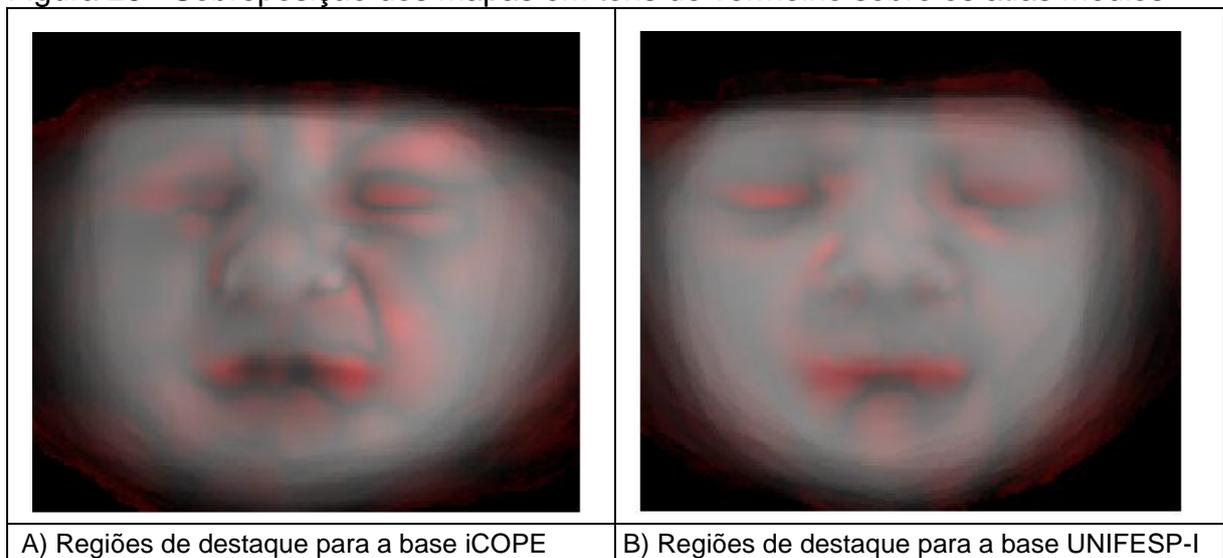
Brahnam *et al.* (2020)⁽⁴²⁾ teve como objetivo desenvolver um novo conjunto de dados de vídeo para detecção automática da dor neonatal chamado iCOPEvid, e apresentar um sistema para classificar os segmentos iCOPEvid em duas categorias de dor e não dor comparando-o com outros sistemas de avaliações humanas. Para o primeiro objetivo, os autores conseguiram reunir 234 vídeos de 49 neonatos que experimentaram um conjunto de estímulos nocivos (um período de descanso e um estímulo agudo à dor). E para o segundo objetivo, foi desenvolvido experimentalmente um sistema próprio comparando uma série de SVMs. Os resultados foram encorajadores e mostraram-se superiores às abordagens padrões, a alguns métodos de aprendizagem profunda, e a alguns avaliadores humanos. No entanto, os autores relatam que estudos adicionais usando CNN são necessários, pois são um subconjunto do Aprendizado de Máquina e utilizados com mais frequência para tarefas de classificação e visão computacional.

Sun *et al.* (2020)⁽⁴⁵⁾ criaram um sistema automatizado de detecção de desconforto com base no monitoramento por vídeo. O sistema primeiramente empregou o fluxo óptico para estimar as trajetórias de movimento do corpo do bebê através dos quadros de vídeo. A cadeia de processamento dependeu de três etapas. Primeiro, a extração de sinal 1D usando fluxo óptico, depois, a conversão do sinal 1D em representação de imagem característica, e por fim, a classificação de aprendizado

profundo das imagens características. Para cada segmento de vídeo, três representações de imagem que caracterizam trajetórias de movimento eram extraídas dos sinais de movimento no domínio do tempo e da frequência. Foi alcançada uma acurácia de 0,985, o que é promissor para uso na prática clínica. No entanto, os autores relatam a necessidade de estudo com mais dados de recém-nascidos para avaliar o sistema.

Silva (2020)⁽⁴⁶⁾ propôs a aplicação de análise estatística multivariada no contexto das imagens de recém-nascidos com e sem dor de forma a explorar, quantificar e determinar medidas comportamentais que auxiliassem na criação de modelos generalistas para a classificação de dor, tanto por sistemas automáticos como por profissionais da saúde. Com o objetivo de identificar as regiões da face mais destacadas no momento de uma experiência dolorosa, primeiramente foi aplicado um filtro estatístico em atlas de imagens como da Figura 26, sendo o resultado da transformação um mapa em que as regiões de maior destaque receberam tons de vermelho mais vivos, e as de menos destaque tons mais claros. Por meio deste processo, foram geradas as imagens das Figuras 26A e 26B para as bases iCOPE e UNIFESP-I, respectivamente. Analisando as figuras vemos que o entorno da boca, uma porção da região da bochecha, e as sobrancelhas são partes da face que recebem destaque em detrimento das outras em momentos que o neonato está experienciando dor.

Figura 26 - Sobreposição dos mapas em tons de vermelho sobre os atlas médios

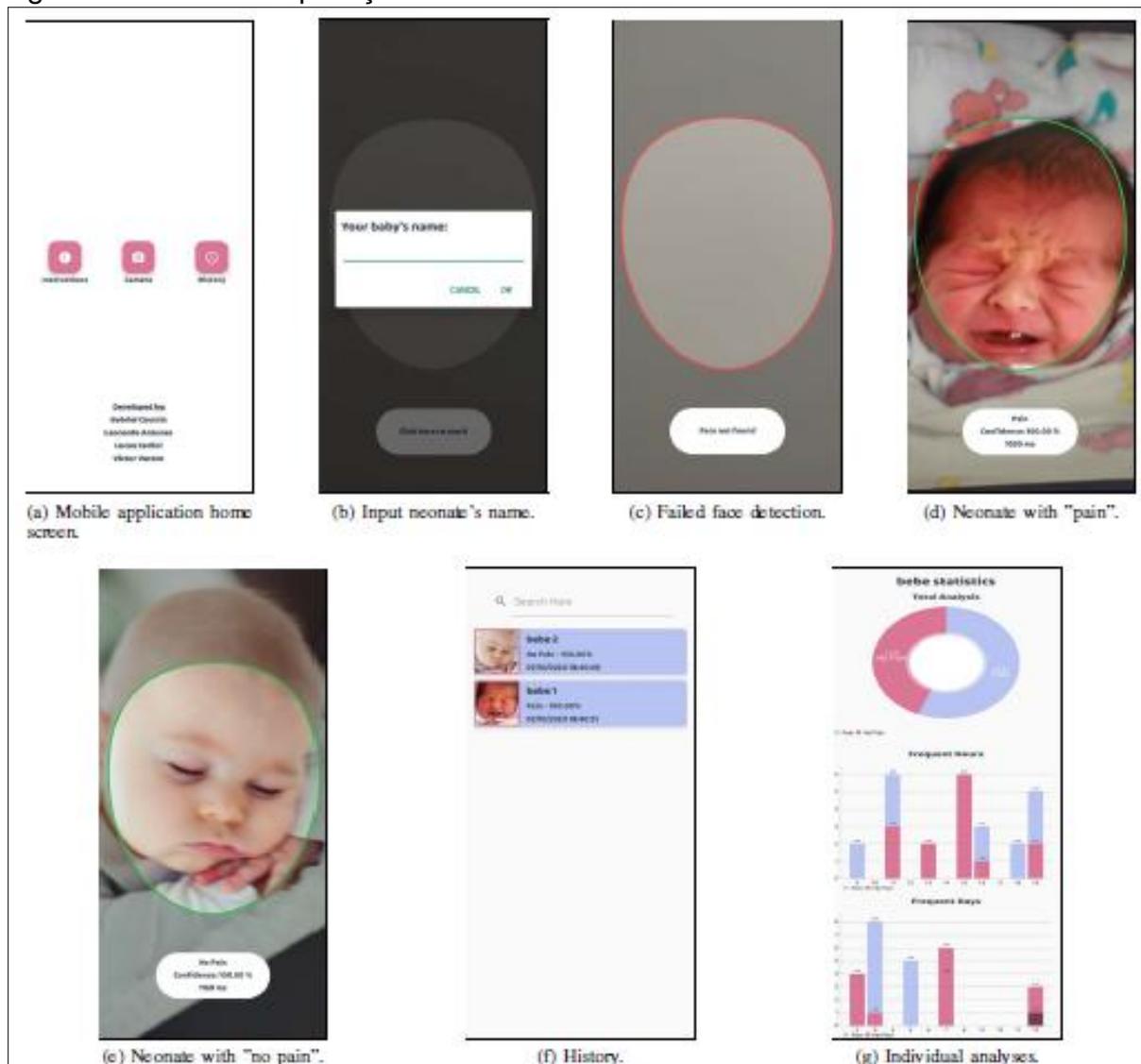


Fonte: Silva, 2020⁽⁴⁶⁾

Os autores⁽⁴⁶⁾ relatam que, desconsiderando as áreas externas à face, os pontos de destaque especificados servem como informação estatística discriminante para profissionais da saúde, pois poderiam dar maior atenção a estas regiões e provavelmente seriam mais assertivos quanto à avaliação de dor em momentos de urgência. Outra ênfase é que as imagens seriam úteis para a criação de algoritmos que fossem projetados de maneira a avaliar a dor por regiões de destaque, ou seja, todo o processamento e análise recairia sobre regiões mais discriminantes, ou então pesos maiores poderiam ser atribuídos a essas áreas, sendo possível criar uma nova escala de dor, ou enriquecer as existentes, utilizando estas regiões como pontos de destaque.

Os trabalhos de Buzuti *et al.* (2020)⁽⁵⁰⁾ e Carlini *et al.* (2021)⁽⁹⁴⁾ estudaram o uso de redes neurais profundas como abordagem para o reconhecimento e classificação de padrões de dor neonatal. O primeiro comparou o uso da arquitetura de redes neurais N-CNN de Zamzmi *et al.* (2019)⁽⁴⁴⁾ com outra CNN (ResNet50) a fim de encontrar o modelo que melhor se adaptasse à tarefa de reconhecimento de faces neonatais. O segundo abordou o uso das redes neurais para detecção de faces de recém-nascidos e classificação da dor no contexto de aplicativos móveis (Figura 27). Ambos os trabalhos demonstram a eficiência de redes convolucionais como estratégia para o reconhecimento de dor neonatal, apresentando resultados com acurácia média superior a 72.63% para a base de dados disposta pela UNIFESP-I.

Figura 27 - Telas de aplicação móvel



Fonte: Carlini *et al.*, 2021⁽⁹⁴⁾

Salekin *et al.*⁽⁴⁷⁾, em 2021, revisaram as práticas atuais e os desafios para a avaliação e o tratamento da dor nas UTINs através da Inteligência Artificial, e relataram a importância das combinações na análise de variáveis objetivas contínuas, ou seja, movimentos faciais e corporais, frequências de choro e dados fisiológicos (sinais vitais), para previsões de alta confiança sobre o tempo de início da dor, principalmente após a sedação pós-cirúrgica de neonatos.

Ainda no ano 2021, Hoti *et al.*⁽⁴⁸⁾ compararam o desempenho de um Sistema de Inteligência Artificial "*PainChek Infant*" (desenvolvido para a avaliação automática da dor em bebês de 0 a 12 meses, analisando as unidades de ação facial que indicam dor) com a escala "*Neonatal Facial Coding System-Revised*" (NFCS-R) e a escala

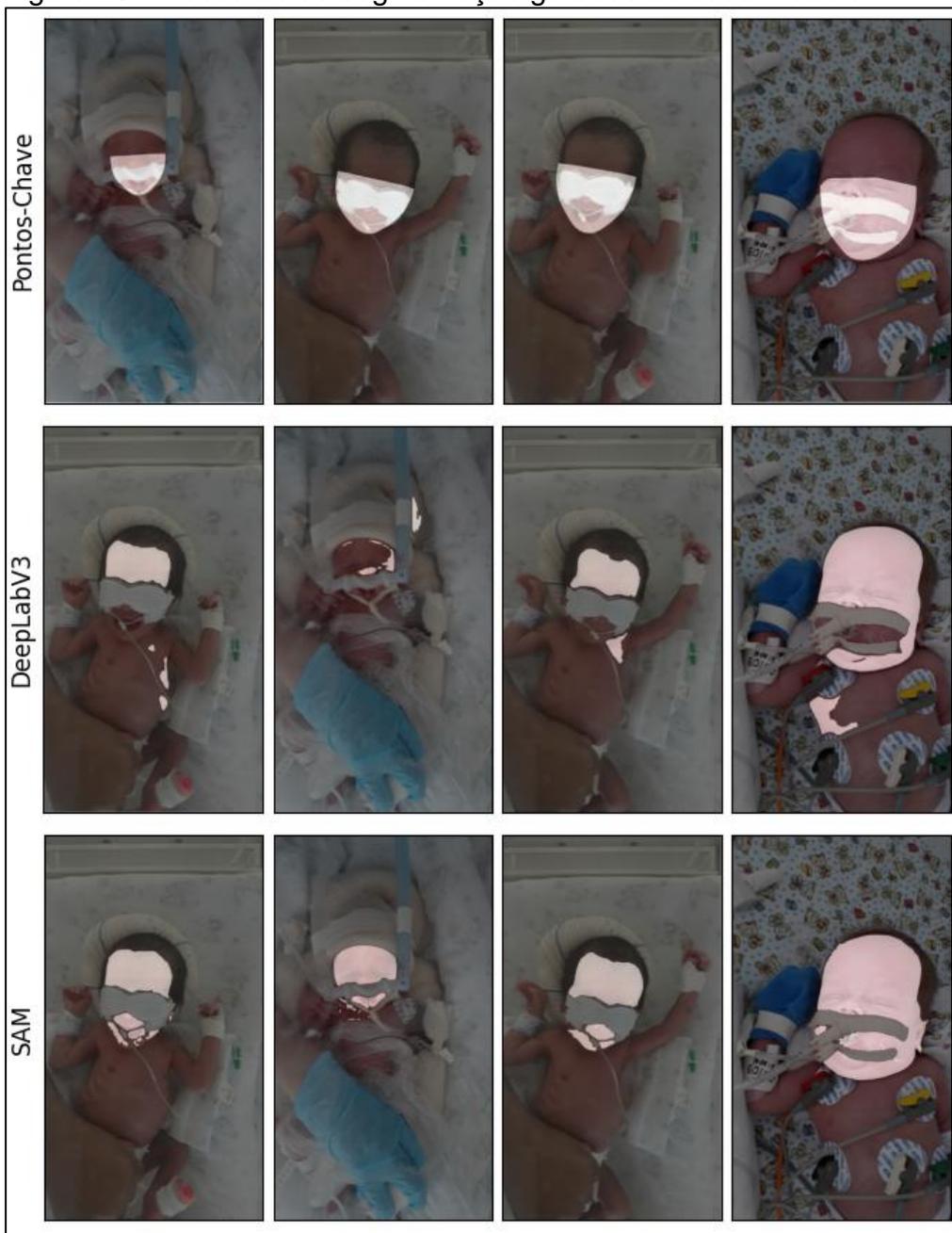
“*Observer administered Visual Analogue Scale*” (ObsVAS)⁽⁴⁸⁾. Neste estudo, a comparação entre os três métodos de avaliação de dor foi realizada com a análise de quatro segmentos de vídeos de 10 segundos (1 durante o repouso, 1 durante a preparação para a vacina, 1 durante a vacinação, 1 durante a fase de recuperação) de 30 bebês, por quatro avaliadores treinados em duas ocasiões separadas. Todas as ferramentas foram consideradas responsivas à dor induzida por procedimento, e as pontuações de dor da “*PainChek Infant*” mostraram boa correlação com as pontuações NFCS e ObsVAS, além de mostrarem boa confiabilidade entre os médicos com altos níveis de consistência interna. Os autores⁽⁴⁸⁾ relatam que o uso da “*PainChek Infant*” na análise automática da expressão facial pode oferecer um meio válido e confiável de monitorar a dor durante procedimentos dolorosos em bebês, mas sua utilidade na prática clínica requer mais pesquisas.

Em 2022, Zamzmi *et al.*⁽⁴⁹⁾ apresentaram um sistema abrangente de avaliação da dor que utiliza expressões faciais junto com o choro, movimento corporal e mudanças de sinais vitais. O sistema automático proposto gerou uma avaliação padronizada da dor comparável às obtidas pelas pontuações de dor descritas por enfermeiras à beira do leito. O sistema atingiu 95,56% de precisão em diferentes respostas à dor que foram registradas em um ambiente clínico desafiador. Os autores⁽⁴⁹⁾ relatam que a avaliação automática da dor neonatal é uma alternativa viável e mais eficiente que a avaliação manual. No entanto, discutem o impacto de diferentes fatores (por exemplo, idade gestacional) sobre a dor, e a necessidade de propor vários modelos específicos de grupos para avaliação da dor (por exemplo, modelos para neonato pré-termo e a termo), sendo necessária uma comparação do desempenho desses modelos com o desempenho dos modelos gerais atuais.

O estudo recente de Domingues (2024)⁽⁹⁸⁾ investigou métodos computacionais para segmentação de faces parcialmente ocluídas, com o objetivo de melhorar a análise de expressões faciais de dor em neonatos. Esse trabalho é extremamente relevante para a prática clínica, pois os equipamentos médicos utilizados em UTIN muitas vezes cobrem partes da face do recém-nascido, dificultando a detecção automática de dor a partir de imagens. Domingues⁽⁹⁸⁾ testou três modelos de segmentação: Segmentação por pontos-chave, DeepLabV3+, e o *Segment Anything Model* (SAM), em cenários com e sem oclusão facial (Figura 28). O SAM se destacou como o modelo de segmentação mais eficaz, superando os demais na capacidade de segmentar corretamente as regiões faciais em neonatos, mesmo com a presença de

oclusões. Após a segmentação, para classificar a expressão de dor em neonatos, foram utilizados quatro classificadores baseados em redes neurais profundas: VGG16, ResNet50, InceptionV3 e ViT. Apesar do sucesso na segmentação, os modelos de classificação não obtiveram resultados superiores a 66% de acurácia na detecção da expressão de dor em casos onde havia oclusão facial, sugerindo que a presença de barreiras físicas, como dispositivos médicos, ainda representa um grande desafio para a classificação precisa da dor em neonatos.

Figura 28 - Máscaras de segmentação geradas de cada modelo



Fonte: Domingues, 2024⁽⁹⁸⁾

Em 2024, Bergamasco *et al.*⁽¹²⁰⁾ propõem um sistema de avaliação automatizada da dor neonatal baseado na análise de expressões faciais. O trabalho desenvolve uma estrutura de *deep learning* que utiliza uma *Rede Neural Convolutiva* (CNN) pré-treinada, ajustada para dados de imagens capturadas durante o procedimento de punção no calcanhar em recém-nascidos. Os autores relatam que o modelo treinado alcançou uma precisão média de 87,4% e uma pontuação F1 de 75,4% para classificar imagens em categorias de dor ou não dor, e que os resultados demonstram a viabilidade da detecção automatizada de dor em recém-nascidos no cenário clínico real.

Outro estudo realizado em 2024 por Zhao *et al.*⁽¹²¹⁾ propõe a detecção automática de dor em neonatos combinando aprendizado profundo com avaliação manual para melhorar a precisão e confiança no diagnóstico. A rede neural proposta pelos pesquisadores foi pré-treinada com os modelos VGG16 e ResNet50, e adaptada para detecção de dor com base em dados limitados, utilizando técnicas como "*dropout*" e normalização das imagens. Quando a rede classificava um caso com alta confiança, a equipe era alertada. Se a confiança fosse baixa, uma avaliação manual era solicitada, proporcionando maior precisão e confiança no sistema. O modelo foi testado com dados limitados, incluindo os conjuntos de dados iCOPE e MNPAD, e demonstrou uma precisão superior em relação a algoritmos anteriores, alcançando 82,35% de acurácia quando a avaliação manual foi integrada.

Beam *et al.*, (2024)⁽¹²²⁾ discute sobre quatro categorias amplas de aplicativos de IA na UTIN que envolvem a interpretação de imagens, a modelagem de previsão de dados de registros eletrônicos de saúde, a integração de dados de monitoramento em tempo real e a documentação e faturamento. Os autores relatam que o progresso recente na IA foi significativamente impulsionado pelos avanços na aprendizagem profunda que se concentra em redes neurais de várias camadas, como as CNNs, e os *Transformers*. Este avanço é visto na aplicação de técnicas de aprendizagem profunda para diagnóstico de imagens médicas (cardiologia, dermatologia, oftalmologia, patologia, radiologia) capazes de ter desempenho igual ou superior ao de médicos; em prever resultados clínicos importantes a partir de dados de registros eletrônicos de saúde, dispositivos de monitoramento em tempo real e dispositivos vestíveis.

Embora o potencial da IA na UTIN seja vasto, Beam *et al.*, (2024)⁽¹²²⁾ relatam que os avanços tecnológicos por si só costumam ser insuficientes para a tradução

clínica e a adoção da IA em aplicativos de saúde. Existem limitações de aceitação do médico, de vieses inerentes aos dados médicos, preocupações éticas e regulatórias, desafios de implementação no mundo real e questões específicas dos modelos de linguagem de grande porte. Com isso, os autores afirmam que pesquisas futuras em neonatologia precisarão navegar cuidadosamente por essas questões, promovendo um diálogo contínuo entre desenvolvedores, clínicos, pacientes e órgãos reguladores, com testes iterativos e refinamento de modelos, a fim de melhorar a transparência, a interpretabilidade e a segurança dos sistemas de IA.

3.3 CONSIDERAÇÕES COMPLEMENTARES

Todos estes avanços e suas implicações clínicas nos levam a uma sensibilização para a melhor prática de automatização do processo de avaliação de dor neonatal. Até o momento, todas as pesquisas realizadas possibilitaram a detecção e diagnóstico da dor neste período, porém, todos os trabalhos até aqui desenvolvidos permitiram apenas a detecção e o reconhecimento da dor em imagens de recém-nascidos que permanecem com a face livre. Nenhum destes trabalhos apresentados anteriormente possibilitou a avaliação da dor em recém-nascidos enfermos, que permanecem internados na UTIN e fazem uso de dispositivos de suporte fixados à sua face.

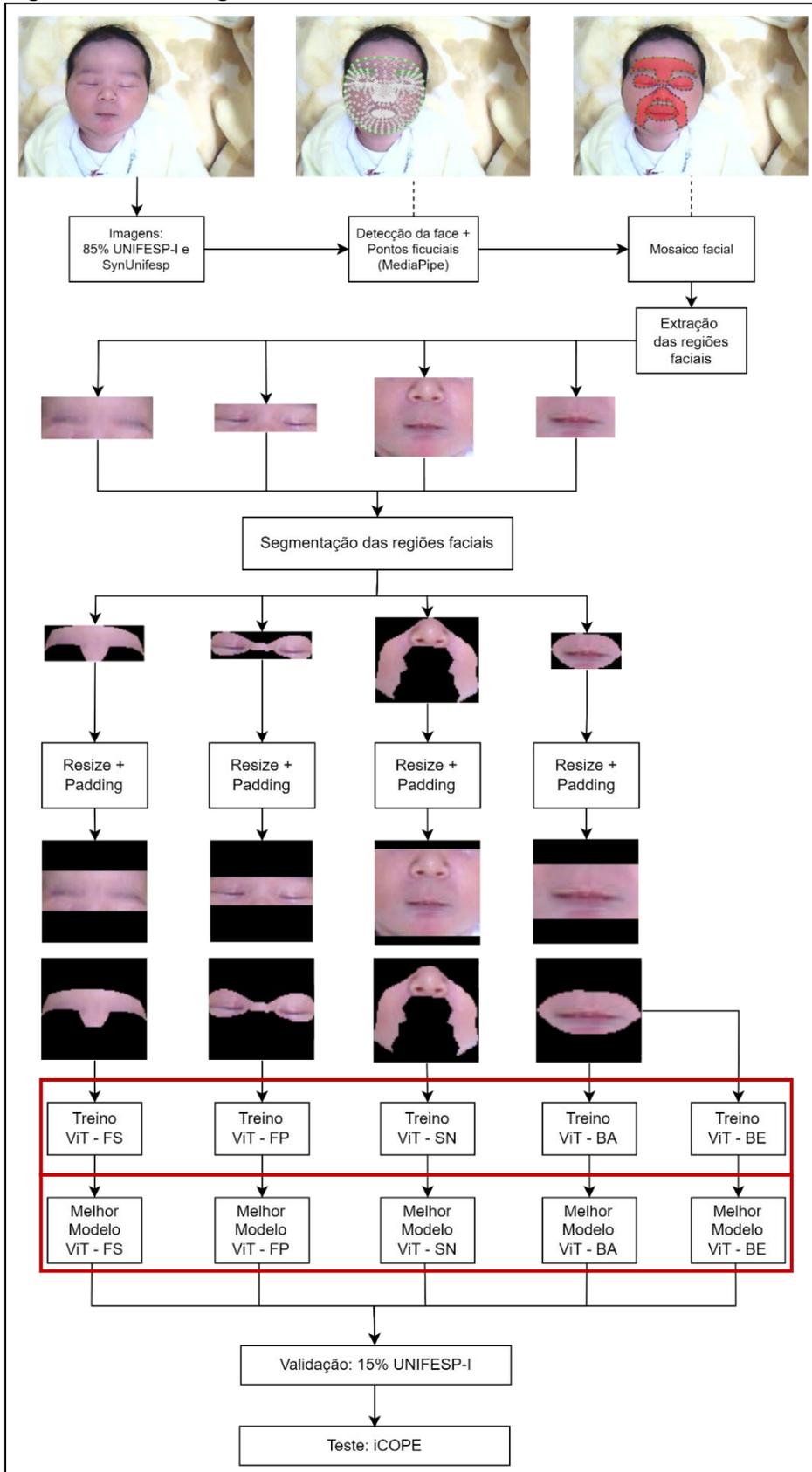
4 MATERIAIS E MÉTODOS

Neste estudo, optou-se por segmentar regiões específicas da face para a avaliação e classificação da dor neonatal. Esse método de segmentação surge como uma alternativa aos métodos tradicionais, que utilizam a face inteira do recém-nascido para identificar a dor. Na prática clínica, neonatos internados em UTIN frequentemente possuem partes da face cobertas por dispositivos médicos de suporte à vida, dificultando a detecção de pontos faciais essenciais para a avaliação completa da dor.

Este capítulo descreve os materiais e métodos utilizados para o desenvolvimento de um arcabouço automatizado de reconhecimento da expressão facial de dor de recém-nascidos internados em UTIN. O método foi baseado na análise de quatro regiões faciais: testa, fenda palpebral, sulco nasolabial e boca. Essas quatro regiões foram selecionadas por estarem diretamente associadas à Escala de Dor NFCS, amplamente utilizada no diagnóstico da dor neonatal em ambiente hospitalar⁽¹⁰⁴⁾. Além disso, essas regiões foram escolhidas por permitirem a identificação de sinais faciais de dor em imagens 2D, tais como: fronte saliente, fenda palpebral estreitada, sulco nasolabial aprofundado, boca aberta e boca estirada. Movimentos como o batimento de asa do nariz e tremores no queixo não foram considerados, pois não podem ser avaliados em uma análise estática, metodologia adotada nesta pesquisa.

Dessa forma, este capítulo detalha também a sequência de procedimentos para a automação do diagnóstico de dor. O processo é ilustrado no fluxograma da Figura 29, para neonatos com a face livre, e no fluxograma da Figura 30, para neonatos internados em UTIN que, muitas vezes, utilizam dispositivos de suporte fixados à face. Nestes fluxogramas estão apresentados os arcabouços propostos nesta pesquisa para a detecção automática da dor e, na cor vermelha, estão destacados os processos que sofreram alteração na sua configuração original, visando aprimorar o modelo à tarefa de classificação binária (movimento indicativo da presença de dor: sim ou não).

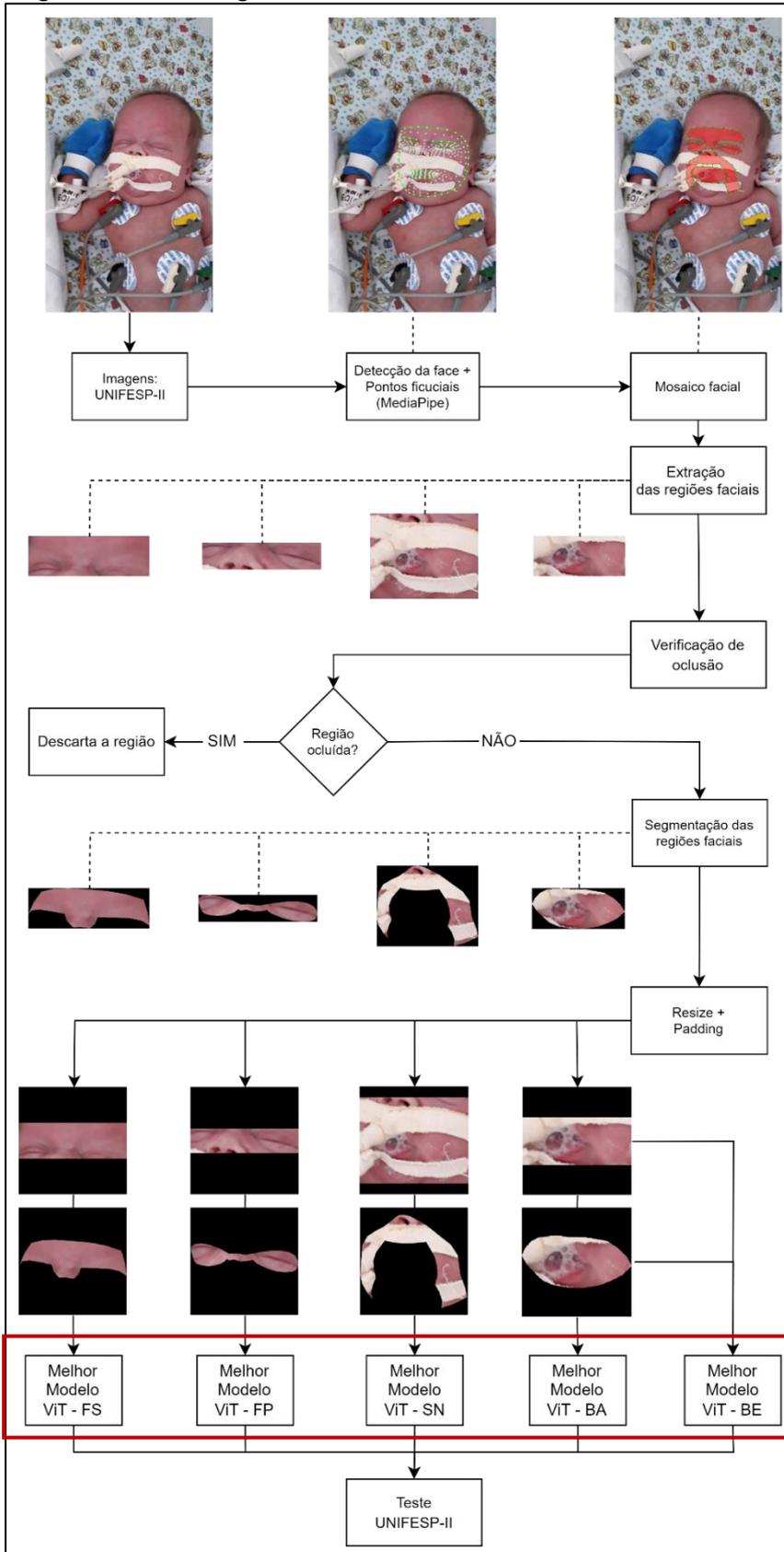
Figura 29 - Fluxograma - Análise da face livre



Fonte: Autor (com permissão)

Legenda: BA = Boca Aberta / BE = Boca Estirada / FP = Fenda Palpebral Estreitada / FS = Fronte Saliente / SN = Sulco Nasolabial Aprofundado / ViT = *Vision Transformer*

Figura 30 - Fluxograma - Análise facial de neonatos internados em UTIN



Fonte: Fonte: Autor (com permissão)

Legenda: BA = Boca Aberta / BE = Boca Estirada / FP = Fenda Palpebral Estreitada / FS = Fronte Saliente / SN = Sulco Nasolabial Aprofundado / ViT = *Vision Transformer*

4.1 BASE DE DADOS - IMAGENS

Quatro bases de dados foram utilizadas nesta pesquisa para o treinamento, validação e teste do arcabouço computacional proposto. As imagens faciais foram rotuladas de acordo com o diagnóstico de dor, classificadas como “face com dor” ou “face sem dor”. Em algumas bases, os dados também incluíam informações detalhadas sobre os movimentos faciais indicativos de dor, especificando se estavam “presentes” ou “ausentes”. Uma das bases, composta por imagens de neonatos com dispositivos médicos fixados à face, incluía ainda rótulos indicando se determinada região facial estava ocluída ou visível, permitindo análises específicas sobre os movimentos faciais em condições de visibilidade parcial.

A seguir, será apresentada uma descrição detalhada dessas quatro bases de dados. A Tabela 4 resume o número de imagens de cada base, juntamente com os rótulos de diagnóstico de dor e os movimentos faciais indicativos de dor.

Tabela 4 - Bases de dados

Base de Dados	Número de imagens	Faces com dor	Faces sem dor	Movimentos indicativos de dor				
				BA	BE	FS	FP	SN
UNIFESP-I	360	162	198	212	59	166	161	153
SynUnifesp	356	118	238	114	169	30	174	116
iCOPE	167	79	88	-	-	-	-	-
UNIFESP-II	48	16	32*	13	10	18	17	11

Fonte: Autor

Legenda: FS = Fronte Saliente / FP = Fenda palpebral estreitada / SN = Sulco nasolabial aprofundado / BA = Boca aberta / BE = Boca estirada / *Face sem dor ou não visível

UNIFESP-I

Este banco de dados contém imagens faciais de neonatos com o rosto completamente livre, sem obstrução das regiões importantes para a avaliação da dor. As imagens foram utilizadas como referência inicial para a detecção facial e subsequentes análises até a validação dos modelos (Figura 29).

Foram utilizadas as 360 imagens disponíveis no banco, que foram avaliadas e rotuladas por seis profissionais de saúde (Autorização – ANEXO C) quanto à presença ou não de cinco movimentos faciais indicativos de dor, e quanto ao diagnóstico de dor, quando 3 ou mais movimentos faciais estavam rotulados como indicativos de dor (Tabela 4).

SynUnifesp

Este banco de dados contém imagens sintéticas, geradas artificialmente, e foi utilizado com o objetivo de expandir o conjunto de dados (UNIFESP-I) e aprimorar o treinamento do modelo de detecção das regiões faciais (Figura 29).

Foram utilizadas as 356 imagens sintéticas disponíveis no banco, que foram avaliadas e rotuladas por três profissionais de saúde (Autorização – ANEXO E) quanto à presença ou não de cinco movimentos faciais indicativos de dor, e quanto ao diagnóstico de dor, quando 3 ou mais movimentos faciais estavam rotulados como indicativos de dor (Tabela 4).

iCOPE

Este banco de dados contém imagens faciais de neonatos com o rosto completamente livre, sem obstruções nas regiões importantes para a avaliação da dor. As imagens foram utilizadas apenas para a fase de teste, com objetivo de testar os melhores modelos de detecção dos movimentos faciais indicativos de dor em uma base independente (Figura 29).

Foram utilizadas 167 imagens disponíveis no banco iCOPE^(69,91,123,124) (Tabela 4), que foram fornecidas pelos desenvolvedores (Autorização – ANEXO B), que estavam apenas rotuladas como face “com dor” (79 imagens) e “face sem dor” (88 imagens). Neste caso, os autores não especificam como o diagnóstico de dor foi realizado, apenas fornecem as faces já rotuladas como “sem dor” e “com dor”.

UNIFESP-II

As imagens deste banco de dados são de neonatos cujas faces poderiam estar parcialmente cobertas por dispositivos de suporte. As imagens foram utilizadas para testar os melhores modelos de detecção dos movimentos faciais indicativos de dor (Figura 30).

Foram utilizadas as 174 imagens disponíveis no banco, que foram avaliadas e rotuladas por quatro profissionais de saúde (Autorização – ANEXO D) quanto à visibilidade de quatro regiões faciais, quanto a presença ou não de cinco movimentos faciais indicativos de dor, e quanto ao diagnóstico de dor, quando a maioria dos movimentos faciais visíveis estavam rotulados como indicativos de dor (Tabela 4).

4.2 MEDIAPIPE

O modelo "Facemesh"⁽⁸⁵⁾, do *MediaPipe*⁽⁸⁴⁾ da Google, foi a ferramenta utilizada para a detecção automática da face dos neonatos e identificação de 468 pontos fiduciais⁽¹²⁵⁾ no rosto dos recém-nascidos (Figuras 31 e 32).

Figura 31 – *MediaPipe*: face livre



Fonte: Autor (com permissão)

Figura 32 – *MediaPipe*: face com dispositivos



Fonte: Autor (com permissão)

A partir dos pontos fiduciais detectados pelo *MediaPipe*, foi gerado um mosaico facial, que serviu para a delimitação das quatro regiões de interesse na face dos recém-nascidos: testa, fenda palpebral, sulco nasolabial e boca (Figuras 33 e 34).

Os pontos fiduciais escolhidos para delimitar as quatro áreas de interesse são:

- **Boca:** 57, 43, 106, 182, 83, 18, 313, 406, 335, 273, 287, 410, 322, 391, 393, 164, 167, 165, 92, 186
- **Fenda Palpebral:** 226, 31, 228, 229, 230, 231, 232, 233, 244, 245, 122, 6, 351, 465, 464, 453, 452, 451, 450, 449, 448, 261, 446, 342, 445, 444, 443, 442, 441, 417, 168, 193, 221, 222, 223, 224, 225, 113
- **Sulco Nasolabial:** 2, 97, 98, 92, 61, 43, 204, 211, 170, 149, 150, 136, 135, 214, 207, 205, 36, 142, 126, 217, 174, 196, 197, 419, 399, 437, 355, 371, 266, 425, 427, 434, 364, 365, 379, 378, 395, 431, 424, 273, 291, 322, 327, 326
- **Fronte:** 168, 417, 285, 336, 296, 334, 293, 300, 298, 332, 297, 338, 10, 109, 67, 103, 68, 70, 63, 105, 66, 107, 55, 193

Utilizamos vários pontos fiduciais para cada região facial para obter maior controle dos contornos, sem a necessidade de aproximações, e consideramos uma pequena margem externa em cada região para não serem perdidos dados das bordas. Para todas as imagens validadas, treinadas e testadas os pontos fiduciais foram os mesmos.

Figura 33 - Pontos fiduciais: face livre



Fonte: Autor (com permissão)

Figura 34 - Pontos fiduciais: face com dispositivos

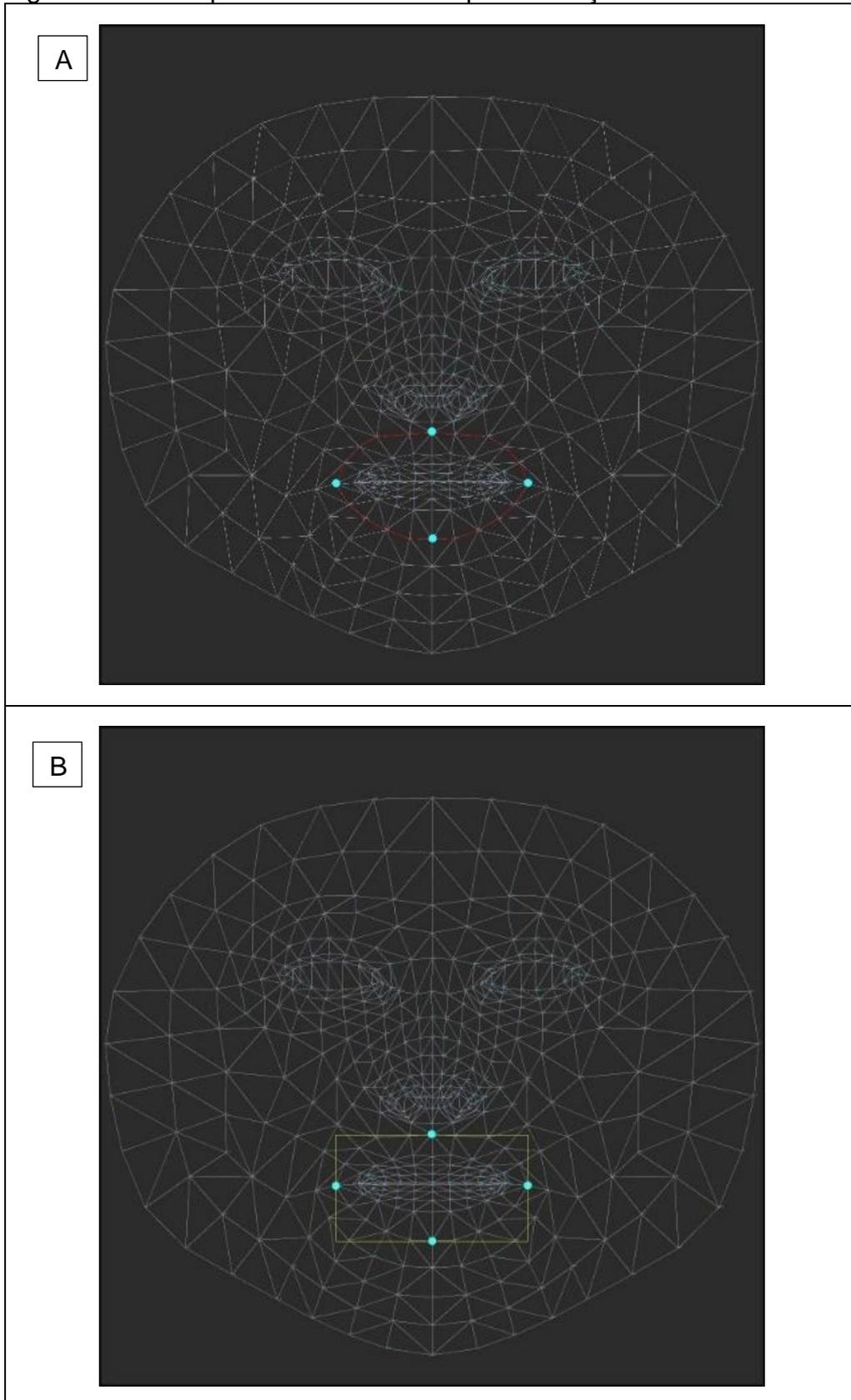


Fonte: Autor (com permissão)

4.3 EXTRAÇÃO DAS REGIÕES FACIAIS

Para trabalharmos com cada região facial de forma isolada, fizemos um processo de extração das regiões faciais determinando um retângulo de recorte. O retângulo foi definido pela menor e maior coordenada X e pela menor e maior coordenada Y. Com base nos pontos de coordenadas X e Y mais extremas, correspondentes aos pontos limites da região (Figura 35A – exemplo das coordenadas da boca), delimitamos a área para extração da região facial desejada (Figura 35B – exemplo da extração do recorte da boca).

Figura 35 - Exemplo das coordenadas para extração do recorte da boca

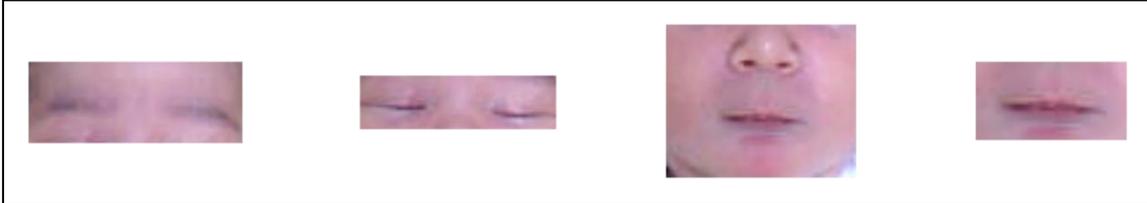


Fonte: Autor

Legenda: (A) Coordenadas da boca; (B) Extração do recorte da boca

Usamos esses pontos apenas para recortar as quatro regiões faciais escolhidas (boca, sulco nasolabial, fenda palpebral e frente) e separá-las do resto da imagem (Figuras 36 e 37).

Figura 36 - Exemplo da extração das regiões faciais livres



Fonte: Autor (com permissão)

Figura 37 - Exemplo da extração das regiões faciais com dispositivos



Fonte: Autor (com permissão)

No método utilizado para análise de regiões faciais de neonatos internado em UTIN, após essa etapa de extração das regiões faciais, o modelo passou por uma etapa de verificação de oclusão, conforme fluxograma da Figura 30. Se a região foi rotulada pelos avaliadores do banco de dados UNIFESP-II como uma região ocluída, não visível para avaliar os movimentos indicativos de dor, esta região foi descartada e não passou pela etapa de segmentação das regiões faciais livres e subsequentes análises.

Vale apontar que, em alguns casos, mesmo a região estando ocluída por algum dispositivo ou fixação de dispositivo, os avaliadores consideraram a mesma como região não ocluída, isto porque aquela oclusão provavelmente não o impediu de detectar o movimento facial indicativo de dor, como: frente saliente, fenda palpebral estreitada, sulco nasolabial aprofundado, boca aberta e boca estirada.

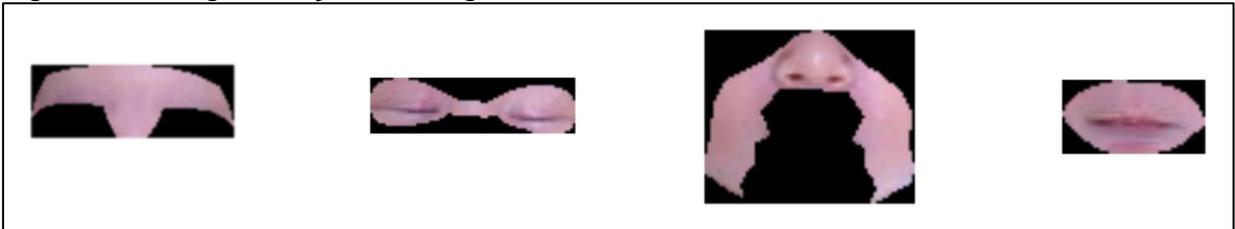
4.4 MÁSCARA DE SEGMENTAÇÃO DAS REGIÕES FACIAIS

A segmentação foi uma etapa crucial neste estudo, especialmente devido à necessidade de isolar as regiões específicas da face dos recém-nascidos (boca, sulco nasolabial, fenda palpebral e frente).

Foram utilizadas máscaras de segmentação para delimitar e separar as regiões de interesse da imagem, eliminando áreas irrelevantes e focando na análise dos movimentos faciais indicativos de dor.

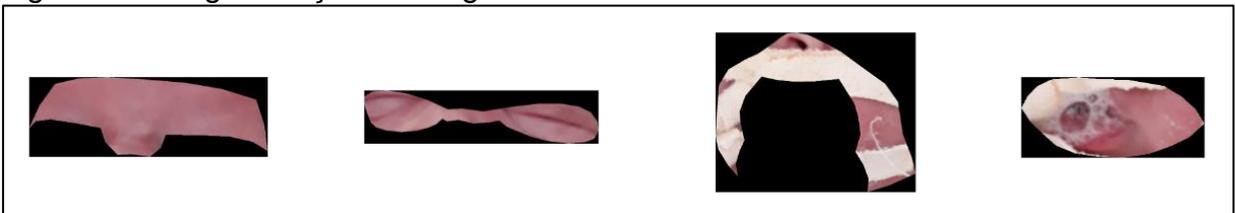
Essas máscaras foram feitas utilizando o mosaico facial. Primeiro, combinamos as coordenadas dos polígonos de cada região com a área extraída de cada região facial (Figura 35B – 35A). Assim, os valores de todos os pixels externos ao polígono foram substituídos por zero, e obtivemos a imagem segmentada por região facial (Figuras 38 e 39).

Figura 38 - Segmentação das regiões faciais livres



Fonte: Autor (com permissão)

Figura 39 - Segmentação das regiões faciais de neonatos internados em UTIN



Fonte: Autor (com permissão)

4.5 RESIZE E PADDING

Para garantir a compatibilidade das imagens com o modelo *Vision Transformer* (ViT) utilizado neste estudo, foi realizado um processo de redimensionamento e preenchimento das imagens:

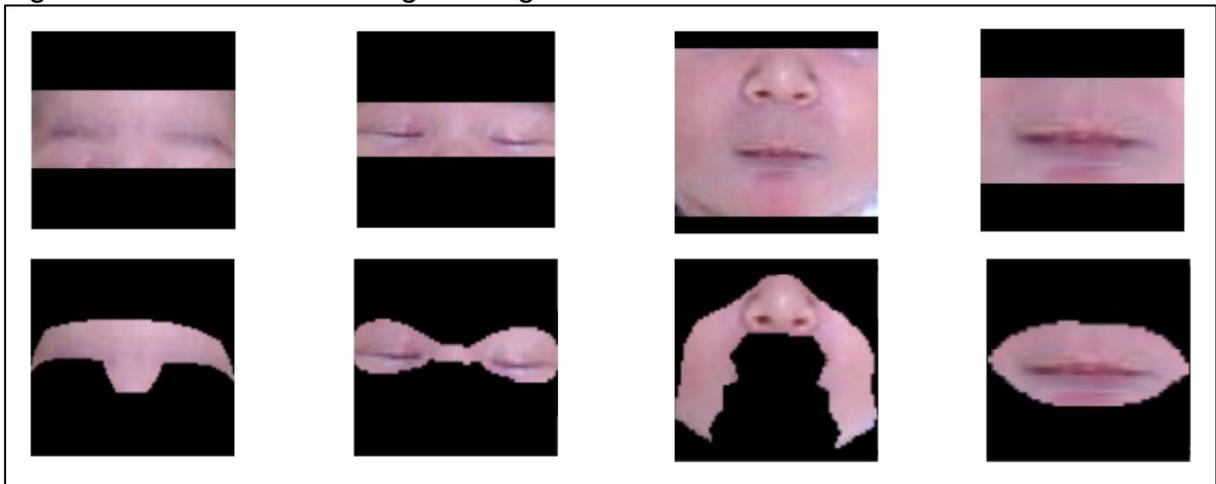
- **Resize** (Redimensionamento): Cada imagem recortada foi redimensionada de modo que sua maior dimensão (altura ou largura) fosse ajustada para 224 pixels. Esse redimensionamento foi feito sem distorção, mantendo a proporção original da imagem (*aspect ratio*).

- **Padding** (Preenchimento): Após o redimensionamento, a dimensão menor foi ajustada adicionando-se pixels pretos nas bordas, distribuídos igualmente nas partes superior e inferior ou laterais da imagem, até que ambas as dimensões alcançassem 224 pixels.

Esse procedimento garantiu que as imagens fossem processadas uniformemente pelo ViT, evitando distorções ou perda de informações visuais essenciais.

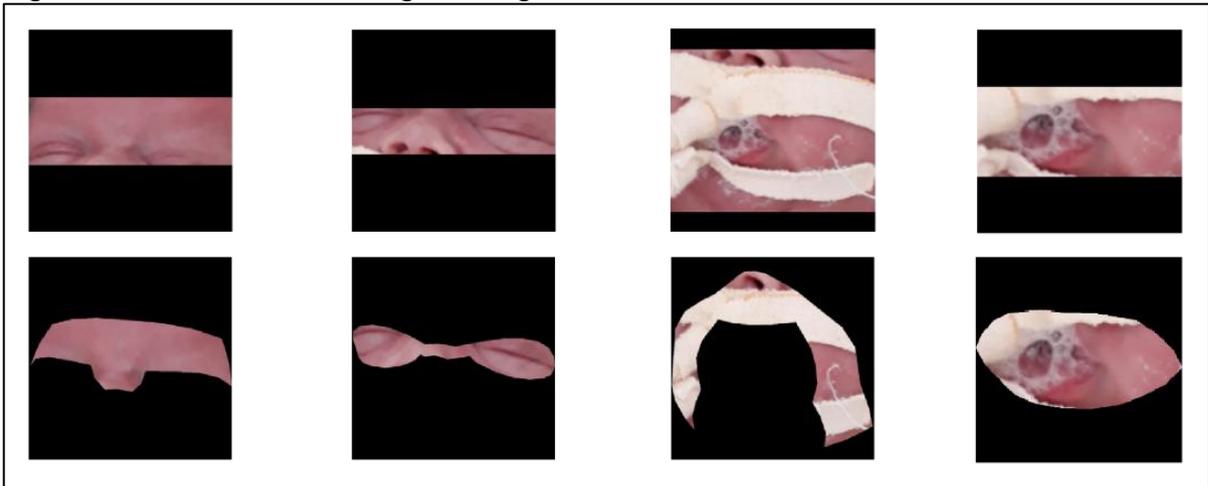
Fizemos esse procedimento tanto para as imagens das áreas extraídas de cada região facial quanto para as imagens da máscara de segmentação facial, como pode ser visualizado nas Figuras 40 e 41 respectivamente.

Figura 40 - *Resize* e *Padding* das regiões faciais livres



Fonte: Autor (com permissão)

Figura 41 - *Resize e Padding* das regiões faciais de neonatos internados em UTIN



Fonte: Autor (com permissão)

4.6 TREINAMENTO COM *VISION TRANSFORMER*

O *Vision Transformer* (ViT) foi utilizado para classificar presença ou ausência de cinco características faciais, sendo estas: boca aberta, boca estirada, sulco nasolabial aprofundado, frente saliente e fenda palpebral estreitada. Um modelo foi treinado para cada região, e vale ressaltar que a região da boca foi utilizada tanto para classificar o movimento de boca aberta quanto o de boca fechada.

A escolha do ViT se deu após o treinamento de alguns classificadores para cada região de interesse da face dos recém-nascidos “com dor” e “sem dor”. Esses classificadores estão descritos detalhadamente na pesquisa realizada por Domingues (2024)⁽⁹⁸⁾. Na pesquisa de Domingues (2024)⁽⁹⁸⁾, os modelos treinados para a tarefa de classificação da expressão de dor foram: VGG16, Resnet50, InceptionV3 e ViT. O estudo mostrou a superioridade da arquitetura ViT na classificação da dor, independente da estratégia de pré-processamento utilizada. Portanto, o ViT foi escolhido como modelo de treinamento para classificação dos movimentos faciais indicativos da ausência ou presença de dor.

Para otimização dos hiperparâmetros do modelo, usamos a biblioteca Optuna⁽⁹⁰⁾, com o *Sampler* padrão, fornecido pela biblioteca. Os hiperparâmetros escolhidos para serem otimizados foram: *learning rate*, *weight decay*, *batch size*, aplicação ou não das imagens com a máscara de segmentação, *loss* balanceada ou não, e congelamento ou não das camadas do *transformer* durante o treinamento de classificação. Além destes, a inclusão de imagens artificiais durante o treinamento

também foi avaliada, e a biblioteca buscou sempre maximizar o F1 score de validação (métrica escolhida para comparação entre os modelos de treinamento).

4.7 VALIDAÇÃO E TESTE

O conjunto de treinamento utilizado nesta pesquisa contemplou imagens das bases de dados UNIFESP-I (85%) e SynUnifesp (100%). Durante o processo de treinamento, o modelo analisou os dados de treinamento para aprender padrões e relações entre as variáveis. Nesse estágio, o algoritmo fez ajustes internos, alterando seus parâmetros para minimizar a função de erro ou perda, buscando melhorar sua capacidade de previsão.

No treinamento, o modelo usou essas informações aleatória e repetidamente (ao longo de várias iterações ou épocas) para aprender as características dos dados. Durante o processo de treinamento, o modelo foi periodicamente avaliado com o conjunto de validação para verificar se o desempenho melhorou. Métricas de F1 score foram calculadas para avaliar a capacidade do modelo.

O conjunto de validação contemplou 15% das imagens do banco de dados UNIFESP-I, priorizando as imagens dos neonatos que não estavam presentes no treinamento, além de monitorar e desempenhar papel crucial na otimização dos hiperparâmetros do modelo.

O conjunto de teste foi composto pela base de dados iCOPE, para imagens com a face livre, e UNIFESP-II, para imagens de neonatos internados em UTIN. Estes testes foram feitos com um subconjunto de dados independente, completamente novo para o modelo. O conjunto de teste só foi utilizado após o término do treinamento e da otimização dos hiperparâmetros, e serviu para avaliar o modelo de forma final, fornecendo uma estimativa de como ele se comporta ao lidar com dados desconhecidos.

4.8 ANÁLISE DE DADOS

Os cálculos e a análise estatística foram realizados utilizando o programa R 4.4.0 (R Core Team, 2024)⁽¹²⁶⁾ e foram divididos em dois momentos.

Primeiramente visamos identificar o melhor modelo para avaliar cada uma das movimentações faciais indicativas da presença de dor (boca aberta, boca estirada, fenda palpebral estreitada, fronte saliente e sulco nasolabial aprofundado). Em seguida, analisamos a acurácia e a concordância entre os melhores modelos gerados para cada uma das movimentações faciais e o rótulo das imagens com a avaliação realizada pelos profissionais de saúde para cada um dos bancos de dados.

4.8.1 Identificação do melhor modelo para avaliar cada uma das movimentações faciais indicativas da presença de dor

Com a finalidade de avaliar os melhores modelos de treinamento, realizamos uma análise dos resultados baseada no F1 score obtido para cada movimento facial indicativo de dor, comparando os desempenhos dos modelos treinados nos conjuntos de dados UNIFESP-I (somente imagens reais) e UNIFESP-I + SynUnifesp (imagens reais e artificiais).

Os resultados de todos os testes foram documentados e armazenados, permitindo uma análise comparativa detalhada do desempenho para diferentes configurações (uso de máscaras de segmentação, o tipo de *loss* e a estratégia de *fine-tuning*).

O método escolhido para análise dos dados foi a comparação entre os F1 scores porque essa métrica oferece um equilíbrio entre precisão (*precision*) e sensibilidade (*recall*). No contexto do reconhecimento de dor, é crucial garantir que o modelo não apenas minimize falsos positivos (alta precisão), mas também seja capaz de identificar corretamente a maior quantidade possível de casos reais de dor (alta sensibilidade). O F1 score é uma métrica harmônica que captura esse equilíbrio, proporcionando uma visão mais robusta do desempenho do modelo em situações em que um simples valor de acurácia poderia ser enganoso devido ao desbalanceamento das classes.

O F1 score é calculado pela Equação (1):

$$F1 = 2 \times \frac{\text{Precisão} \times \text{Sensibilidade}}{\text{Precisão} + \text{Sensibilidade}} \quad (1)$$

Em que:

- Precisão (*Precision*): Proporção de verdadeiros positivos (VP) entre os exemplos classificados como positivos (VP + falsos positivos=FP) [Equação (2)]:

$$\text{Precisão} = \frac{\text{VP}}{\text{VP} + \text{FP}} \quad (2)$$

- Sensibilidade (*Recall*): Proporção de VP entre todos os exemplos que são realmente positivos (VP + falsos negativos=FN) [Equação (3)]:

$$\text{Sensibilidade} = \frac{\text{VP}}{\text{VP} + \text{FN}} \quad (3)$$

4.8.2 Acurácia e concordância entre os melhores modelos e a avaliação dos profissionais de saúde

A análise de acurácia e de concordância, entre os melhores modelos gerados para cada uma das movimentações faciais e o rótulo das imagens com a avaliação realizada pelos profissionais de saúde para cada um dos bancos de dados, serviu como método de validação e de teste dos melhores modelos.

As análises foram feitas à partir de uma tabela de contingência, ou matriz de confusão (Tabela 5), calculando-se acurácia, sensibilidade, especificidade, taxa de falsos positivos, taxas de falsos negativos, F1 score, valor-p de McNemar, e Kappa de Cohen.

Tabela 5 - Matriz de confusão

	<i>Profissionais: Dor</i>	<i>Profissionais: Sem Dor</i>
Modelo: Dor	Verdadeiros Positivos (VP)	Falsos Positivos (FP)
Modelo: Sem Dor	Falsos Negativos (FN)	Verdadeiros Negativos (VN)

Fonte: Autor "Adaptado de" McNemar (1947)^[127]

Interpretação da Matriz de Confusão⁽¹²⁷⁾:

- VP (Verdadeiros Positivos): Imagens em que tanto o rótulo dos profissionais quanto o melhor modelo detectam dor.
- VN (Verdadeiros Negativos): Imagens em que tanto o rótulo dos profissionais quanto o melhor modelo indicam ausência de dor.
- FP (Falsos Positivos): Imagens em que o melhor modelo indica dor, mas o rótulo dos profissionais não.
- FN (Falsos Negativos): Imagens em que o melhor modelo não indica dor, mas o rótulo dos profissionais sim.

A acurácia foi calculada pela seguinte Equação (4)⁽¹²⁷⁾:

$$\text{Acurácia} = \frac{\text{VP} + \text{VN}}{\text{VP} + \text{VN} + \text{FP} + \text{FN}} \quad (4)$$

A especificidade foi calculada pela seguinte Equação (5)⁽¹²⁷⁾:

$$\text{Especificidade} = \frac{\text{VN}}{\text{VN} + \text{FP}} \quad (5)$$

A taxa de falsos positivos foi calculada pela seguinte Equação (6)⁽¹²⁷⁾:

$$\text{Falsos Positivos} = \frac{\text{FP}}{\text{FP} + \text{VN}} \quad (6)$$

A taxa de falsos negativos foi calculada pela seguinte Equação (7)⁽¹²⁷⁾:

$$\text{Falsos Negativos} = \frac{\text{FN}}{\text{FN} + \text{VP}} \quad (7)$$

A análise se deu tanto de forma regional (movimento facial indicativo de dor, rotulado pelos avaliadores de cada banco de dados) quanto de forma global (diagnóstico de dor, rotulado pelos avaliadores de cada banco de dados).

Para cada análise também aplicamos o Teste de McNemar⁽¹²⁷⁾, com a finalidade de identificar o nível de discordância entre os melhores modelos e os rótulos dos avaliadores. Optamos por este teste pois ele é usado para avaliar se há uma diferença significativa entre dois classificadores (neste caso, o modelo e o rótulo dos avaliadores) na classificação binária de uma única característica (movimentos faciais ou diagnóstico de dor).

O Teste de McNemar⁽¹²⁷⁾ também parte de uma tabela de contingência e pode fornecer uma análise descritiva das taxas de discordância a partir de uma análise dos resultados falsos positivos (modelo prediz dor, mas profissionais não) e falsos negativos (modelo não prediz dor, mas profissionais sim). O Teste de McNemar⁽¹²⁷⁾ é calculado pela seguinte Equação (8):

$$X^2_{McNemar} = \frac{(FP - FN)^2}{FP + FN} \quad (8)$$

O teste de McNemar tem as seguintes hipóteses:

- **Hipótese nula (H_0):** $FP = FN$, isto é, as probabilidades teóricas de Falsos Positivos e Falsos Negativos são iguais;
- **Hipótese alternativa (H_1):** $FP \neq FN$, isto é, as probabilidades teóricas de Falsos Positivos e Falsos Negativos são diferentes.

Ao focar nas taxas que indicam diferenças, o teste de McNemar calcula a discrepância entre as frequências observadas nos dados e as frequências esperadas sob a hipótese nula. Neste trabalho, a hipótese nula é de que não há diferença significativa entre as classificações dos rótulos dos avaliadores e as do modelo. Portanto, se o valor-p do Teste de McNemar for igual ou maior que 0,05, não rejeitamos a hipótese nula e concluímos que não há uma discordância significativa entre o modelo e o rótulo dos profissionais, e se o valor-p for menor que 0,05, rejeitamos a hipótese nula, sugerindo que o modelo e os profissionais discordam estatisticamente.

Outro teste utilizado foi o Kappa de Cohen⁽¹²⁸⁾, para avaliar, além da acurácia, o nível de concordância entre o melhor modelo e os profissionais, ajustando pela concordância que ocorreria ao acaso, sendo a Equação (9):

$$\kappa = \frac{P_o - P_e}{1 - P_e} \quad (9)$$

Em que:

- P_o é a proporção de concordância observada (acurácia) [Equação (10)].

$$P_o = \frac{VP + VN}{VP + VN + FP + FN} \quad (10)$$

- P_e é a proporção de concordância esperada ao acaso [Equação (11)].

$$P_e = \frac{(VP + FP)(VP + FN) + (VN + FN)(VN + FP)}{(VP + VN + FP + FN)^2} \quad (11)$$

A interpretação dos valores de Kappa de Cohen seguiu a seguinte escala⁽¹²⁸⁾:

- Kappa 1,0: Concordância perfeita.
- Kappa entre 0,81 – 1,0: Concordância quase perfeita.
- Kappa entre 0,61 – 0,80: Concordância substancial.
- Kappa entre 0,41 – 0,60: Concordância moderada.
- Kappa entre 0,21 – 0,40: Concordância leve.
- Kappa entre 0,00 – 0,20: Concordância pobre.
- Kappa < 0,00: Concordância abaixo do acaso (discrepância).

5 EXPERIMENTOS E RESULTADOS

Este capítulo fornece uma descrição técnica e detalhada do processo de implementação do *Vision Transformer* (ViT), apresentando as etapas de cada decisão tomada. As técnicas e configurações usadas foram selecionadas para maximizar a eficácia do modelo na tarefa de reconhecimento da dor em recém-nascidos.

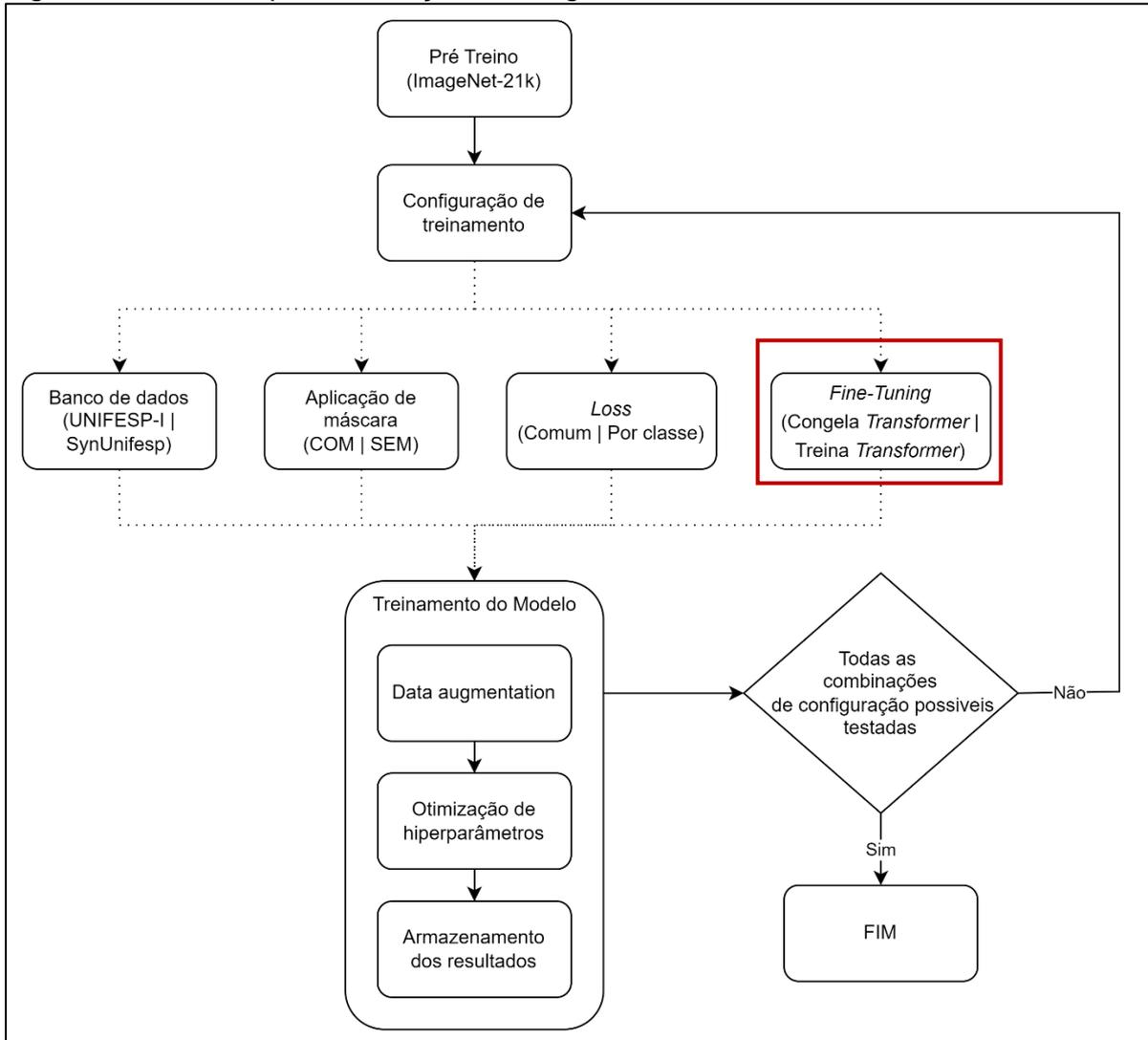
5.1 TREINAMENTO DAS IMAGENS COM A FACE LIVRE

Para a implementação do modelo, foram utilizadas as bibliotecas “*Hugging Face Transformers*” e “*PyTorch*”. A *Hugging Face* forneceu as ferramentas necessárias para o uso e modificação do ViT, enquanto o *PyTorch* foi utilizado para o treinamento e otimização do modelo, devido à sua flexibilidade e eficiência no manuseio de redes neurais profundas. As etapas do treinamento podem ser visualizadas no fluxograma da Figura 42.

O modelo foi inicialmente pré-treinado no conjunto de dados ImageNet-21k⁽¹²⁹⁾, contendo mais de 21.000 classes. Este pré-treinamento utilizou um modelo disponibilizado pela empresa Google (sem o *head* de classificação, apenas o *Transformer*) através da plataforma do *HuggingFace*, garantindo que o ViT possuísse um forte conhecimento base em reconhecimento de padrões visuais antes do *fine-tuning*⁽¹²⁹⁾.

Depois do pré-treinamento diferentes combinações foram feitas para verificar qual seria a melhor abordagem do modelo na avaliação das regiões faciais livres. O fluxograma da Figura 42 apresenta essas configurações de treinamento, e na cor vermelha destacamos a etapa em que modificamos o processo original, removendo o *head* de classificação e substituindo por uma rede MLP com duas saídas, sem nenhum treinamento prévio, com a finalidade de adaptar o modelo para uma tarefa de classificação binária.

Figura 42 - Modelo para avaliação das regiões de faces livres



Fonte: Autor

Para avaliar a eficácia de diferentes abordagens do modelo, foram testadas todas as combinações possíveis das seguintes alterações no treinamento:

Imagens Geradas Artificialmente:

- **Com versus Sem Adição:** Testou-se a inclusão de imagens geradas artificialmente no conjunto de treinamento (UNIFESP I e SynUnifesp), comparando os resultados com as imagens reais (UNIFESP-I).

Vale ressaltar que a validação foi feita usando sempre imagens do banco de dados UNIFESP-I, e as imagens sintéticas foram utilizadas apenas para treinamento. As imagens de validação e treinamento foram selecionadas aleatoriamente e fixadas,

assim pudemos comparar os resultados do treinamento dos modelos de forma coerente.

Aplicação de Máscaras:

- **Com Máscara versus Sem Máscara:** As imagens foram treinadas tanto com as máscaras de segmentação quanto sem a aplicação das máscaras de segmentação.

Função de Loss:

- **Loss Comum versus Loss Balanceada por Classe:** Foi utilizada a função de *loss* (Cross Entropy Loss) em dois cenários: um com *loss* comum e outro balanceado por classe, em que a classe com mais exemplos sofreu um ajuste menor por imagem.

Fizemos este ajuste pois queríamos balancear os dados de treinamento e de validação com relação às classes que representavam o movimento indicativo de dor (Tabela 6).

Tabela 6 - Dados de treinamento e validação

Base de Dados	Número de imagens	FACES com dor	FACES sem dor	Movimentos indicativos de dor				
				BA	BE	FP	FS	SN
UNIFESP-I	360	162	198	212 (59%)	59 (16%)	161 (45%)	166 (46%)	153 (42%)
UNIFESP-I e SynUnifesp	716	280	436	326 (45%)	228 (32%)	335 (47%)	196 (27%)	269 (37%)

Fonte: Autor

Legenda: BA = Boca Aberta / BE = Boca Estirada / FP = Fenda Palpebral Estreitada / FS = Fronte Saliente / SN = Sulco Nasolabial Aprofundado

Fine-Tuning do Modelo:

Foi utilizado um modelo da arquitetura ViT pré-treinado na base de dados Imagenet-21k⁽¹²⁹⁾. No entanto, o *head* de classificação original do modelo foi removido e substituído por uma rede MLP com duas saídas, sem nenhum treinamento prévio. A introdução desse MLP tem como objetivo adaptar o modelo à nova tarefa de classificação binária. Duas abordagens de treinamento foram testadas:

- **Congela *Transformer*:** Esta abordagem congela o *Transformer* e treina apenas o MLP. Nesta estratégia, todos os pesos do ViT, pré-treinados no ImageNet-21k, foram mantidos congelados, e apenas o MLP foi treinado. O objetivo dessa abordagem foi preservar as representações e padrões já aprendidos pelo ViT durante o pré-treinamento, minimizando o risco de *overfitting* para o conjunto de dados deste estudo.

- **Treina *Transformer*:** Este modelo treina o *Transformer* e o MLP simultaneamente. Nesta segunda abordagem, tanto o MLP quanto o ViT foram treinados juntos, permitindo o ajuste fino (*fine-tuning*) de toda a arquitetura. A condição testada aqui é que, apesar do número limitado de imagens no conjunto de dados utilizado, a diversidade das imagens poderia permitir que o ViT ajustasse suas representações para melhor capturar as diferenças entre as duas classes.

Para melhorar a robustez do modelo e prevenir *overfitting*, técnicas de aumento de dados foram aplicadas às imagens de treinamento. As transformações escolhidas foram:

Transformação Afim⁽¹³⁰⁾:

- Rotação: As imagens foram rotacionadas aleatoriamente entre -45 a +45 graus.
- Translações: As imagens foram deslocadas horizontalmente em $\pm 10\%$ e verticalmente em $\pm 30\%$.
- Escala: A escala das imagens foi ajustada aleatoriamente entre 70% a 90% do tamanho original.

Color Jitter⁽¹³⁰⁾:

- Brilho: Variabilidade no brilho com um fator de 0,4.
- Hue: Ajuste aleatório do matiz com um fator de 0,3.
- Saturação: Modificações na saturação com um fator de 0,6.

Blur Gaussiano⁽¹³⁰⁾:

- Kernel: Foi aplicada uma suavização com um filtro gaussiano de kernel 5x5.
- Sigma: O desvio padrão para o filtro foi escolhido aleatoriamente entre 0,1 e 2, seguindo uma distribuição uniforme.

Flip⁽¹³⁰⁾:

- Flip Vertical: Aplicado com uma probabilidade de 15%.
- Flip Horizontal: Aplicado com uma probabilidade de 15%.

Essas técnicas foram escolhidas para simular variabilidade nas condições de captura das imagens e aumentar a diversidade dos dados de treinamento.

Os hiperparâmetros otimizados foram⁽¹³⁰⁾:

- *Learning Rate*: Foi explorada uma faixa entre 10^6 e 10^3 .
- *Batch Size*: Os valores testados foram 4, 8, 16 e 32.
- *Weight Decay*: Foram testadas duas opções: 0 e 0,1.

A otimização foi direcionada para maximizar a métrica F1 de validação, que foi calculada utilizando 15% do banco de dados UNIFESP-I. Para cada configuração, foram realizadas 30 tentativas de otimização. Cada treinamento foi conduzido por 20 épocas, conforme testes preliminares que indicaram que esse número de épocas era suficiente para a convergência do modelo, e poderia ser parado de forma prematura pelo otimizador, caso não se demonstrasse promissor.

5.2 RESULTADOS DO TREINAMENTO COM ViT

Para avaliar o desempenho do modelo de reconhecimento automático de dor neonatal, treinado com diferentes combinações de dados e parâmetros, realizamos uma análise descritiva dos resultados com foco nos resultados de F1 para os conjuntos de dados UNIFESP-I (imagens reais) e UNIFESP-I + SynUnifesp (imagens reais combinadas com imagens geradas artificialmente), considerando o uso de máscaras de segmentação, o tipo de *loss* (comum *versus* balanceada) e a estratégia de *fine-tuning* aplicada (Tabela 7).

Tabela 7 - Resultados das análises com destaque para o melhor F1 de cada movimento facial para cada conjunto de dados

(continuação)

<i>Bancos de dados</i>	<i>Mov. facial</i>	<i>F1 score</i>	<i>Uso de máscara</i>	<i>Ajuste do modelo (loss)</i>	<i>Ajuste fino (fine-tuning)</i>
UNIFESP-I	BA	0.877	não	comum	treina transformer
UNIFESP-I	BA	0.757	não	comum	congela transformer
UNIFESP-I	BA	0.881	não	balanceado	treina transformer
UNIFESP-I	BA	0.510	não	balanceado	congela transformer
UNIFESP-I	BA	0.872	sim	comum	treina transformer
UNIFESP-I	BA	0.708	sim	comum	congela transformer
UNIFESP-I	BA	0.896	sim	balanceado	treina transformer
UNIFESP-I	BA	0.368	sim	balanceado	congela transformer
UNIFESP-I e SynUnifesp	BA	0.914	não	comum	treina transformer
UNIFESP-I e SynUnifesp	BA	0.680	não	comum	congela transformer
UNIFESP-I e SynUnifesp	BA	0.914	não	balanceado	treina transformer
UNIFESP-I e SynUnifesp	BA	0.652	não	balanceado	congela transformer
UNIFESP-I e SynUnifesp	BA	0.871	sim	comum	treina transformer
UNIFESP-I e SynUnifesp	BA	0.712	sim	comum	congela transformer
UNIFESP-I e SynUnifesp	BA	0.891	sim	balanceado	treina transformer
UNIFESP-I e SynUnifesp	BA	0.658	sim	balanceado	congela transformer
UNIFESP-I	BE	0.857	não	comum	treina transformer
UNIFESP-I	BE	0.653	não	comum	congela transformer
UNIFESP-I	BE	0.827	não	balanceado	treina transformer
UNIFESP-I	BE	0.333	não	balanceado	congela transformer
UNIFESP-I	BE	0.830	sim	comum	treina transformer
UNIFESP-I	BE	0.564	sim	comum	congela transformer
UNIFESP-I	BE	0.807	sim	balanceado	treina transformer
UNIFESP-I	BE	0.352	sim	balanceado	congela transformer
UNIFESP-I e SynUnifesp	BE	0.658	não	comum	treina transformer
UNIFESP-I e SynUnifesp	BE	0.461	não	comum	congela transformer
UNIFESP-I e SynUnifesp	BE	0.704	não	balanceado	treina transformer
UNIFESP-I e SynUnifesp	BE	0.262	não	balanceado	congela transformer
UNIFESP-I e SynUnifesp	BE	0.602	sim	comum	treina transformer
UNIFESP-I e SynUnifesp	BE	0.317	sim	comum	congela transformer
UNIFESP-I e SynUnifesp	BE	0.619	sim	balanceado	treina transformer
UNIFESP-I e SynUnifesp	BE	0.233	sim	balanceado	congela transformer
UNIFESP-I	FP	0.666	não	comum	treina transformer
UNIFESP-I	FP	0.125	não	comum	congela transformer
UNIFESP-I	FP	0.695	não	balanceado	treina transformer
UNIFESP-I	FP	0.064	não	balanceado	congela transformer
UNIFESP-I	FP	0.716	sim	comum	treina transformer
UNIFESP-I	FP	0.125	sim	comum	congela transformer
UNIFESP-I	FP	0.716	sim	balanceado	treina transformer
UNIFESP-I	FP	0.125	sim	balanceado	congela transformer

Tabela 7 - Resultados das análises com destaque para o melhor F1 de cada movimento facial para cada conjunto de dados

(conclusão)

UNIFESP-I e SynUnifesp	FP	0.673	não	comum	treina transformer
UNIFESP-I e SynUnifesp	FP	0.109	não	comum	congela transformer
UNIFESP-I e SynUnifesp	FP	0.659	não	balanceado	treina transformer
UNIFESP-I e SynUnifesp	FP	0.074	não	balanceado	congela transformer
UNIFESP-I e SynUnifesp	FP	0.644	sim	comum	treina transformer
UNIFESP-I e SynUnifesp	FP	0.142	sim	comum	congela transformer
UNIFESP-I e SynUnifesp	FP	0.606	sim	balanceado	treina transformer
UNIFESP-I e SynUnifesp	FP	0.135	sim	balanceado	congela transformer
UNIFESP-I	FS	0.769	não	comum	treina transformer
UNIFESP-I	FS	0.137	não	comum	congela transformer
UNIFESP-I	FS	0.723	não	balanceado	treina transformer
UNIFESP-I	FS	0.137	não	balanceado	congela transformer
UNIFESP-I	FS	0.730	sim	comum	treina transformer
UNIFESP-I	FS	0.638	sim	comum	congela transformer
UNIFESP-I	FS	0.769	sim	balanceado	treina transformer
UNIFESP-I	FS	0.608	sim	balanceado	congela transformer
UNIFESP-I e SynUnifesp	FS	0.854	não	comum	treina transformer
UNIFESP-I e SynUnifesp	FS	0.290	não	comum	congela transformer
UNIFESP-I e SynUnifesp	FS	0.862	não	balanceado	treina transformer
UNIFESP-I e SynUnifesp	FS	0	não	balanceado	congela transformer
UNIFESP-I e SynUnifesp	FS	0.784	sim	comum	treina transformer
UNIFESP-I e SynUnifesp	FS	0.686	sim	comum	congela transformer
UNIFESP-I e SynUnifesp	FS	0.770	sim	balanceado	treina transformer
UNIFESP-I e SynUnifesp	FS	0.072	sim	balanceado	congela transformer
UNIFESP-I	SN	0.945	não	comum	treina transformer
UNIFESP-I	SN	0.133	não	comum	congela transformer
UNIFESP-I	SN	0.867	não	balanceado	treina transformer
UNIFESP-I	SN	0.193	não	balanceado	congela transformer
UNIFESP-I	SN	0.862	sim	comum	treina transformer
UNIFESP-I	SN	0.590	sim	comum	congela transformer
UNIFESP-I	SN	0.857	sim	balanceado	treina transformer
UNIFESP-I	SN	0.363	sim	balanceado	congela transformer
UNIFESP-I e SynUnifesp	SN	0.872	não	comum	treina transformer
UNIFESP-I e SynUnifesp	SN	0.543	não	comum	congela transformer
UNIFESP-I e SynUnifesp	SN	0.857	não	balanceado	treina transformer
UNIFESP-I e SynUnifesp	SN	0.4	não	balanceado	congela transformer
UNIFESP-I e SynUnifesp	SN	0.86	sim	comum	treina transformer
UNIFESP-I e SynUnifesp	SN	0.463	sim	comum	congela transformer
UNIFESP-I e SynUnifesp	SN	0.826	sim	balanceado	treina transformer
UNIFESP-I e SynUnifesp	SN	0.238	sim	balanceado	congela transformer

Fonte: Autor

Legenda: Mov. = Movimento / BA = Boca Aberta / BE = Boca Estirada / FP = Fenda Palpebral Estreitada / FS = Fronte Saliente / SN = Sulco Nasolabial Aprofundado

Ao avaliar o desempenho do modelo para cada parâmetro analisado, pudemos observar os seguintes resultados:

- **Imagens Geradas Artificialmente:**

A inclusão das imagens geradas artificialmente aumentou o desempenho em alguns cenários, apresentando resultados mistos. Para boca aberta e frente saliente, por exemplo, o *F1 score* foi consistentemente superior com o conjunto de dados ampliado. No entanto, o impacto não foi tão claro para outros movimentos, como boca estirada, fenda palpebral estreitada e sulco nasolabial aprofundado, em que os *F1 scores* não tiveram grandes melhorias e a adição de imagens artificiais parece ter introduzido ruído, levando a uma queda no desempenho.

Isso sugere que as imagens geradas artificialmente podem ter ajudado o modelo a generalizar melhor para alguns tipos de movimentos faciais, mas não para todos.

- **Aplicação de Máscaras:**

Os movimentos que tiveram uma leve melhoria com uso de máscara de segmentação foram boca aberta e fenda palpebral estreitada, apenas no conjunto de dados UNIFESP-I. A maioria dos movimentos faciais, especialmente no conjunto de dados combinado, UNIFESP-I e SynUnifesp, teve uma redução de desempenho com o uso de máscara.

Esse fenômeno pode ser explicado pela possível identificação de padrões fora da região de interesse pelo modelo. Por exemplo, o movimento de sulco nasolabial aprofundado pode gerar uma movimentação na boca, e a máscara de segmentação impediria o modelo de capturar essa relação. Por outro lado, movimentos como boca aberta, que não geram movimentação significativa fora da região da boca, podem se beneficiar da máscara, que remove ruídos externos e foca na área correta de análise.

No geral, o uso de máscara não trouxe melhorias consistentes e, em alguns casos, reduziu o desempenho, especialmente no conjunto de dados ampliado. Nestes casos, quando o uso de máscara de segmentação não melhora o *F1 score* ou resulta em um desempenho semelhante ao treinamento sem máscara, seu uso pode ser questionado. Treinar com máscaras pode exigir mais tempo de processamento e recursos computacionais, sem necessariamente trazer benefícios práticos ao modelo.

- **Função de Loss**

O movimento com melhor desempenho com a função *loss* balanceada foi a boca aberta no conjunto de dados UNIFESP-I (F1 *score* de 0.896). Ao contrário, um desempenho consistentemente melhor com a *loss* comum foi o sulco nasolabial aprofundado no conjunto de dados UNIFESP-I, com um F1 *score* de 0.945 com *loss* comum contra 0.867 balanceada.

Para fenda palpebral estreitada e frente saliente, o F1 *score* foi idêntico para *loss* comum e balanceada nos dois conjuntos de dados. Isso sugere que o tipo de *loss* não teve impacto significativo nesses movimentos faciais.

Assim, para o objetivo de maximizar o F1 *score* em todos os movimentos faciais, a *loss* comum parece ser a escolha mais consistente. O uso da *loss* balanceada pode ser útil em situações de desbalanceamento severo entre classes, mas seu impacto não foi claramente positivo em todos os cenários investigados aqui.

- **Fine-Tuning do Modelo**

Em todos os movimentos faciais o treinamento completo do ViT (treinamento do *Transformer* e do MLP simultaneamente) resultou consistentemente em melhores F1 *scores* em comparação com o congelamento do *Transformer* e treinamento apenas do MLP.

Isso sugere que, mesmo com um número limitado de imagens, o ajuste fino do ViT é crucial para capturar as diferenças presentes nas classes de dor e não dor. O congelamento dos pesos parece limitar a capacidade do modelo de aprender com as especificidades do conjunto de dados, resultando em F1 *scores* significativamente mais baixos (muitas vezes abaixo de 0.5, como visto nos movimentos do sulco nasolabial aprofundado e fenda palpebral estreitada).

Para selecionar o melhor modelo para cada movimento facial indicativo da presença de dor, foi elaborada uma tabela com base nos melhores F1 *scores* observados para cada combinação de parâmetros (máscara, tipo de *loss*, *fine-tuning*) em cada movimento facial (Tabela 8).

Nos casos em que houve empate de F1 *score* dentro de um mesmo conjunto de dados, a escolha do modelo sem máscara foi preferida, quando possível, pois evita pré-processamento e poderia diminuir possíveis erros em imagens com oclusão,

quando os pontos não se alinham bem na região (um objetivo que foi analisado na segunda fase deste projeto, para as imagens do banco de dados UNIFESP-II).

Nos casos em que, mesmo com a escolha do modelo sem máscara, ainda houve empate de *F1 score* dentro de um mesmo conjunto de dados, a escolha do *loss*, comum ou balanceado, se deu de acordo com o balanceamento entre as classes. Em casos balanceados (exemplo: 45% *versus* 55%), o uso de *loss* comum é suficiente e não há necessidade de ajustes com uma *loss* balanceada.

Tabela 8 - Melhor *F1 score* dos modelos para cada movimento facial indicativo da presença de dor

<i>Movimento Facial</i>	<i>UNIFESP-I</i>	<i>UNIFESP-I + SynUnifesp</i>	<i>Melhor Conjunto</i>
Boca Aberta	0.896 (com máscara, loss balanceada)	0.914 (sem máscara, loss comum)	UNIFESP-I + SynUnifesp
Boca Estirada	0.857 (sem máscara, loss comum)	0.704 (sem máscara, loss balanceada)	UNIFESP-I
Fenda Palpebral Estreitada	0.716 (com máscara, loss comum)	0.673 (sem máscara, loss comum)	UNIFESP-I
Fronte Saliente	0.769 (sem máscara, loss comum)	0.862 (sem máscara, loss balanceada)	UNIFESP-I + SynUnifesp
Sulco Nasolabial Aprofundado	0.945 (sem máscara, loss comum)	0.872 (sem máscara, loss comum)	UNIFESP-I

Fonte: Autor

Mais detalhadamente:

- **Boca Aberta**

- UNIFESP-I: O melhor *F1 score* foi 0.896 (com máscara, *loss* balanceada, *fine-tuning* completo do *Transformer*). O uso de máscara trouxe uma pequena melhoria no desempenho em comparação com a ausência de máscara, e o uso da *loss* balanceada foi favorável. No entanto, vale ressaltar que a boca foi uma região em que as classes já estavam bem balanceadas nos conjuntos de dados de treinamento e validação.

- UNIFESP-I + SynUnifesp: O melhor *F1 score* foi 0.914 (sem máscara, *loss* comum, *fine-tuning* completo). A adição de imagens artificiais resultou em uma ligeira melhoria no desempenho. O uso de *loss* comum foi mais eficaz aqui, e o modelo desempenhou melhor sem máscara.

- **Boca Estirada**

- UNIFESP-I: O melhor F1 score foi 0.857 (sem máscara, *loss* comum, *fine-tuning* completo). O uso de máscara neste caso não foi benéfico, e a *loss* comum foi superior à balanceada.
- UNIFESP-I + SynUnifesp: O melhor F1 score foi 0.704 (sem máscara, *loss* balanceada, *fine-tuning* completo). A inclusão de imagens artificiais prejudicou o desempenho, levando a uma queda no F1 score.

- **Fenda Palpebral Estreitada**

- UNIFESP-I: O melhor F1 score foi 0.716 (com máscara, *loss* comum, *fine-tuning* completo). O uso de máscara ajudou a melhorar o desempenho, e a *loss* comum foi novamente a melhor.
- UNIFESP-I + SynUnifesp: O melhor F1 score foi 0.673 (sem máscara, *loss* comum, *fine-tuning* completo). O desempenho foi inferior com o conjunto de dados ampliado, mesmo sem a máscara, sugerindo que as imagens artificiais não ajudaram no reconhecimento automático deste movimento.

- **Fronte Saliente**

- UNIFESP-I: O melhor F1 score foi 0.769 (sem máscara, *loss* comum, *fine-tuning* completo). A máscara e a *loss* balanceada não proporcionaram diferenças no resultado, e de acordo com o critério de simplicidade, para reduzir a complexidade do treinamento e manter o desempenho igual, adotar os critérios sem uso de máscara de segmentação e *loss* comum pode ser uma boa opção para este caso.
- UNIFESP-I + SynUnifesp: O melhor F1 score foi 0.862 (sem máscara, *loss* balanceada, *fine-tuning* completo). O uso de imagens artificiais melhorou significativamente o desempenho, e a *loss* balanceada continuou a ser benéfica. A máscara não foi necessária para o melhor resultado.

- **Sulco Nasolabial Aprofundado**

- UNIFESP-I: O melhor F1 score foi 0.945 (sem máscara, *loss* comum, *fine-tuning* completo). O uso de máscara não foi necessário, e a *loss* comum proporcionou os melhores resultados.

- UNIFESP-I + SynUnifesp: O melhor F1 score foi 0.872 (sem máscara, *loss* comum, *fine-tuning* completo). A adição de imagens artificiais reduziu o desempenho para esse movimento, possivelmente introduzindo ruído ou desbalanceamento no conjunto de treino.

5.3 VALIDAÇÃO DOS MELHORES MODELOS TREINADOS

Após selecionar o melhor modelo para cada movimento facial (boca aberta, boca estirada, fenda palpebral estreitada, frente saliente e sulco nasolabial aprofundado), realizamos uma etapa de validação analisando a concordância entre o que o melhor modelo predisse como classificação e o que os profissionais de saúde avaliaram.

Para esta análise, utilizamos apenas os 15% das imagens presentes no conjunto de dados UNIFESP-I (54 imagens do total de 360 imagens do banco). Sendo que destas, 30 imagens estavam rotuladas como “face com dor” e 24 imagens estavam rotuladas como “face sem dor”. Quanto aos movimentos indicativos de dor, os mesmos estavam rotulados como:

- Boca aberta: 31 imagens (presente) / 23 imagens (ausente)
- Boca estirada: 28 imagens (presente) / 26 imagens (ausente)
- Fenda palpebral estreitada: 30 imagens (presente) / 24 imagens (ausente)
- Frente saliente: 27 imagens (presente) / 27 imagens (ausente)
- Sulco nasolabial aprofundado: 28 imagens (presente) / imagens 26 (ausente)

A comparação do melhor modelo com os profissionais de saúde foi realizada em duas situações distintas:

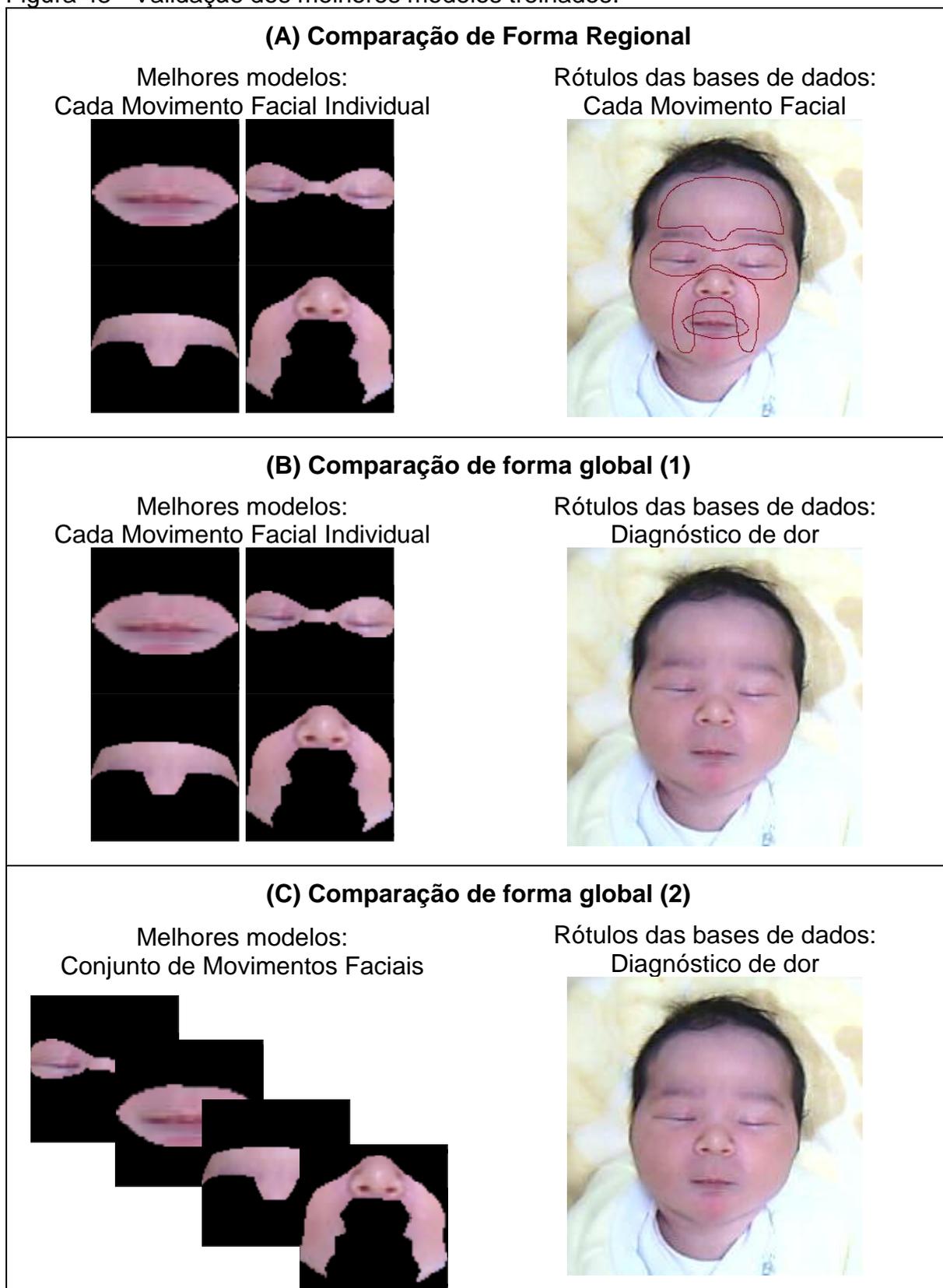
- **Forma regional:** Analisamos a concordância entre os profissionais de saúde e o melhor modelo, em relação à cada movimento facial individual. Por exemplo, observando o quanto o melhor modelo do movimento de boca aberta concordava com o rótulo de boca aberta indicado pelos profissionais de saúde, seguindo da mesma forma para os outros movimentos faciais (Figura 43A).

- **Forma global:** Analisamos a concordância entre os profissionais de saúde e o melhor modelo, em relação ao diagnóstico de dor. Para esta análise, observamos a concordância de duas formas distintas:

1- Melhores modelos de cada movimento facial *versus* o rótulo de diagnóstico de cada imagem apresentado no banco de dados UNIFESP-I (“com dor” e “sem dor”), como mostra a Figura 43B. Por exemplo, observando o quanto o melhor modelo do movimento de boca aberta concordava com o rótulo de imagem “com dor” pelos profissionais de saúde, seguindo da mesma forma para os outros movimentos faciais. Essa abordagem teve por objetivo identificar se apenas um movimento facial específico, identificado pelo melhor modelo, seria o suficiente para realizar o diagnóstico de dor.

2- Diferentes conjuntos de *score* dos melhores modelos de cada movimento facial *versus* o rótulo de diagnóstico de cada imagem apresentado no banco de dados UNIFESP-I (“com dor” e “sem dor”), como mostra a Figura 43C. Esta metodologia partiu do pressuposto de que as avaliações de dor dependem de um *score* para diagnóstico final de dor. Por exemplo, vários trabalhos em que a escala de dor NFCS foi utilizada usando a avaliação de 5 movimentos faciais, o diagnóstico de dor se dava quando 3 ou mais movimentos faciais eram detectados como indicativos de dor^(33,38,43,55,131–133). Por isso, nesta etapa da pesquisa, tentamos identificar qual conjunto de *scores* (1, 2, 3, 4 ou 5 movimentos faciais) seria mais assertivo para o diagnóstico de dor.

Figura 43 - Validação dos melhores modelos treinados.



Fonte: Autor (com permissão)

Comparação regional

Para realizar as análises de concordância entre os profissionais de saúde e o melhor modelo, em relação à cada movimento facial individual, foram calculadas as taxas de acurácia, sensibilidade, especificidade, falsos positivos, falsos negativos, F1 score, valor-p de McNemar, e Kappa de Cohen. Os resultados estão descritos na Tabela 9.

Tabela 9 – Validação – Acurácia e Concordância entre os melhores modelos e o rótulo de movimento facial indicativo de dor do conjunto de dados UNIFESP-I

<i>Métrica</i>	<i>BA</i>	<i>BE</i>	<i>FP</i>	<i>FS</i>	<i>SN</i>
Acurácia	88.89%	85.19%	79.63%	79.63%	90.74%
Sensibilidade	96.30%	88.46%	88.00%	80.77%	96.00%
Especificidade	81.48%	82.14%	72.41%	78.57%	86.21%
Taxa Falsos Positivos	18.52%	17.86%	27.59%	21.43%	13.79%
Taxa Falsos Negativos	3.70%	11.54%	12.00%	19.23%	4.00%
F1 score	89.66%	85.19%	80.00%	79.25%	90.57%
McNemar (valor-p)	0.221*	0.724*	0.228*	1.000*	0.371*
Kappa de Cohen	0.778	0.704	0.596	0.593	0.815

Fonte: Autor

Legenda: BA = Boca Aberta / BE = Boca Estirada / FP = Fenda Palpebral Estreitada / FS = Fronte Saliente / SN = Sulco Nasolabial Aprofundado / * p > 0,05 = não houve discordância

Podemos observar na Tabela 9 que para o movimento de boca aberta, a taxa de acurácia foi elevada, atingindo 88,89%, o que indica que a classificação do modelo teve um desempenho semelhante à classificação rotulada pelos avaliadores. A sensibilidade de 96,3% reflete a alta capacidade do modelo em detectar a presença de dor para este movimento, enquanto a especificidade de 81,48% demonstra uma boa habilidade em identificar corretamente a ausência de dor. As taxas de falsos positivos (18,52%) e falsos negativos (3,7%) são relativamente baixas, indicando que o modelo comete poucos erros tanto na identificação incorreta de dor quando não há, quanto na falta de identificação de dor quando deveria. O F1 score de 0,896 confirma que o modelo mantém um excelente equilíbrio entre precisão e sensibilidade. O valor do teste de McNemar ($p = 0,221$) sugere que não há discordância significativa entre as classificações do modelo e dos avaliadores ($p > 0,05$), portanto, a hipótese nula não pode ser rejeitada. O Kappa de Cohen foi de 0,778, indicando um nível de concordância substancial entre o modelo e os avaliadores.

Para o movimento de boca estirada, a taxa de acurácia foi de 85,19%, um pouco inferior à de boca aberta, mas ainda assim muito boa. A sensibilidade de 88,46% indica uma alta capacidade do modelo de identificar corretamente a dor, enquanto a especificidade de 82,14% mostra que o modelo tem uma boa capacidade de não confundir a ausência de dor como dor. A taxa de falsos positivos foi de 17,86%, enquanto a taxa de falsos negativos foi de 11,54%, indicando que o modelo erra mais em não detectar a dor do que em superestimá-la. O *F1 score* de 85,19% demonstra que o modelo é eficiente ao equilibrar a detecção de dor e os erros de decisão. O valor-p do teste de McNemar (0,724) indica que não há evidência de discordância significativa, logo, a hipótese nula não é rejeitada. O Kappa de Cohen foi de 0,704, também refletindo um nível de concordância substancial.

A taxa de acurácia para o movimento de fenda palpebral estreitada foi uma das mais baixas entre as regiões analisadas, atingindo 79,63%, com uma taxa relativamente elevada de falsos positivos (27,59%), indicando que o modelo tende a superestimar a presença de dor para este movimento. A sensibilidade de 88% e a especificidade de 72,41% são aceitáveis, mas revelam que o modelo pode cometer erros, especialmente em identificar corretamente a ausência de dor. O *F1 score* de 80% reflete um desempenho bom e equilibrado. O teste de McNemar ($p = 0,228$) não indica discordância significativa entre o modelo e os avaliadores, mantendo a hipótese nula. O Kappa de Cohen foi de 0,596, indicando um nível de concordância moderado.

Para o movimento de frente saliente, a taxa de acurácia foi também de 79,63%, com uma sensibilidade de 80,77%. No entanto, a taxa de falsos negativos (19,23%) revela que o modelo frequentemente não detecta a dor onde os avaliadores identificaram. A especificidade de 78,57% é boa, sugerindo que o modelo é razoavelmente eficaz em identificar corretamente quando não há dor. O *F1 score* de 79,25% reflete um desempenho também bom, apesar dos erros de classificação. O teste de McNemar ($p = 1,000$) indica que não há evidência de discordância significativa entre as classificações, e a hipótese nula não é rejeitada. O Kappa de Cohen foi de 0,593, também refletindo um nível de concordância moderado.

Por último, o movimento de sulco nasolabial aprofundado apresentou a maior taxa de acurácia, 90,74%, e uma sensibilidade elevada (96%), demonstrando que o modelo tem uma excelente capacidade de detectar a dor para este movimento. A especificidade de 86,21% indica que o modelo também é eficaz em identificar a ausência de dor. A taxa de falsos positivos foi de 13,79%, a mais baixa entre todas as

regiões, e a taxa de falsos negativos de 4%, também extremamente baixa, mostrando que o modelo erra muito pouco. O F1 score de 90,57% foi o mais alto e reflete um excelente equilíbrio entre sensibilidade e precisão, reforçando o bom desempenho do modelo para este movimento. O teste de McNemar ($p = 0,371$) sugere que não há discordância significativa entre as classificações do modelo e dos avaliadores, mantendo a hipótese nula. O Kappa de Cohen foi de 0,815, indicando um nível de concordância quase perfeito entre o modelo e os avaliadores.

No geral, o modelo apresentou boas taxas de acurácia com os avaliadores em todas as áreas analisadas, com destaque para boca aberta e sulco nasolabial aprofundado, que mostraram os melhores desempenhos tanto em sensibilidade quanto em F1 score e Kappa de Cohen. As áreas de fenda palpebral estreitada e fronte saliente apresentaram algumas dificuldades, especialmente com falsos positivos e falsos negativos, mas ainda assim mostraram concordância moderada a substancial, conforme indicado pelo Kappa de Cohen. Nenhum dos testes de McNemar revelou discordâncias significativas entre o modelo e os avaliadores, reforçando a não diferença entre suas classificações.

Comparação global: Melhores modelos de cada movimento facial *versus* o rótulo de diagnóstico apresentado no banco de dados UNIFESP-I

Para realizar as análises de concordância entre os melhores modelos de cada movimento facial e o rótulo de diagnóstico de cada imagem apresentado no banco de dados UNIFESP-I, foram também calculadas as taxas de acurácia, sensibilidade, especificidade, falsos positivos, falsos negativos, F1 score, valor-p de McNemar, e Kappa de Cohen. Os resultados estão descritos na Tabela 10.

Tabela 10 – Validação - Acurácia e Concordância entre os melhores modelos e o rótulo de diagnóstico de dor do conjunto de dados UNIFESP-I

<i>Métrica</i>	<i>BA</i>	<i>BE</i>	<i>FP</i>	<i>FS</i>	<i>SN</i>
Acurácia	75.93%	77.78%	75.93%	74.07%	87.04%
Sensibilidade	73.33%	73.33%	70.00%	70.00%	80.00%
Especificidade	79.17%	83.33%	83.33%	79.17%	95.83%
Taxa Falsos Positivos	20.83%	16.67%	16.67%	20.83%	4.17%
Taxa Falsos Negativos	26.67%	26.67%	30.00%	30.00%	20.00%
F1 score	77.19%	78.57%	76.36%	75.00%	87.27%
McNemar (valor-p)	0.579*	0.386*	0.267*	0.423*	0.131*
Kappa de Cohen	0.519	0.557	0.522	0.484	0.743

Fonte: Autor

Legenda: BA = Boca Aberta / BE = Boca Estirada / FP = Fenda Palpebral Estreitada / FS = Fronte Saliente / SN = Sulco Nasolabial Aprofundado / * p > 0,05 = não houve discordância

Podemos observar na Tabela 10 que o movimento de boca aberta apresentou uma acurácia de 75,93% em relação ao diagnóstico de dor, o que indica que, em aproximadamente 3/4 dos casos, o modelo de boca aberta concorda com o diagnóstico geral de dor. A sensibilidade de 73,33% mostra que o modelo é razoável em detectar dor quando a boca aberta é um sinal presente, mas não excelente. A especificidade de 79,17% indica que o modelo tem uma boa capacidade de identificar corretamente os casos sem dor, com uma taxa de falsos positivos de 20,83%, o que sugere que em cerca de 21% das vezes o modelo classifica erroneamente a dor quando não está presente. A taxa de falsos negativos de 26,67% reflete que o modelo deixa de identificar a dor em aproximadamente 27% dos casos em que ela deveria ser detectada. O F1 score de 0,772 mostra que há um equilíbrio razoável entre precisão e sensibilidade. O p-valor de McNemar (0,579) indica que não há uma discordância significativa entre o modelo e os rótulos de diagnóstico de dor, e o Kappa de Cohen (0,519) reflete um nível de concordância moderado.

Para o movimento de boca estirada observamos uma acurácia de 77,78%, um pouco superior à de boca aberta, indicando uma boa concordância com o diagnóstico de dor. A sensibilidade de 73,33% é igual à de boca aberta, mostrando que o modelo detecta dor com a mesma eficiência. No entanto, a especificidade de 83,33% é um pouco maior, o que significa que o modelo é ligeiramente melhor em não identificar falsamente dor onde não existe, com uma taxa de falsos positivos de 16,67%. A taxa de falsos negativos de 26,67% é semelhante à de boca aberta, o que indica que o modelo ainda não consegue detectar todos os casos de dor. O F1 score de 78,57%, um pouco superior, mostra que há um equilíbrio ligeiramente melhor entre a precisão

e a sensibilidade. O p-valor de McNemar (0,386) novamente não mostra discordância significativa, e o Kappa de Cohen de 0,557 indica um nível de concordância moderado, ligeiramente superior à de boca aberta.

O movimento de fenda palpebral estreitada teve uma acurácia de 75,93%, igual à de boca aberta, o que significa que o modelo concorda com o diagnóstico de dor em 3/4 dos casos. A sensibilidade de 70% mostra que o modelo tem uma capacidade razoável de identificar dor, mas é a menor entre os movimentos faciais analisados, o que pode indicar dificuldade em diagnosticar a dor. A especificidade de 83,33% é boa, com uma taxa de falsos positivos de 16,67%, sugerindo que o modelo é melhor em evitar diagnósticos falsos de dor. A taxa de falsos negativos de 30% é a mais alta entre todas as regiões, o que indica que o modelo frequentemente não detecta dor quando presente. O F1 score de 76,36% é o mais baixo, refletindo o impacto da sensibilidade reduzida. Mesmo assim, o p-valor de McNemar (0,267) não sugere discordâncias significativas entre o modelo e os diagnósticos de dor rotulados pelos profissionais de saúde, e o Kappa de Cohen (0,522) mostra um nível de concordância moderado, mas inferior a outras regiões.

Para o movimento de frente saliente, a acurácia foi de 74,07%, a mais baixa entre todas as regiões, o que sugere que o modelo para esta área tem um desempenho ligeiramente inferior. A sensibilidade de 70% também é baixa, indicando que o modelo frequentemente falha em detectar dor associada a frente saliente. A especificidade de 79,17% é razoável, com uma taxa de falsos positivos de 20,83%, sugerindo que o modelo ocasionalmente classifica a dor quando não presente. A taxa de falsos negativos de 30%, como para fenda palpebral, é alta, mostrando que o modelo frequentemente não detecta dor quando deveria. O F1 score de 75% reflete uma menor precisão e sensibilidade em relação às outras regiões. O p-valor de McNemar (0,423) não mostra discordância significativa, e o Kappa de Cohen (0,484) indica um nível de concordância moderado, mas o mais baixo entre todas as regiões analisadas.

Por último, o movimento de sulco nasolabial aprofundado destacou-se, com uma acurácia de 87,04%, a mais alta entre todas as regiões faciais. A sensibilidade de 80% mostra que o modelo é muito bom em detectar dor quando o sulco nasolabial aprofundado é um indicativo, e a especificidade de 95,83% é excelente, sugerindo que o modelo raramente classifica erroneamente a dor quando ela não está presente, com uma taxa de falsos positivos de apenas 4,17%. A taxa de falsos negativos de 20% é

a mais baixa entre as regiões faciais, indicando que o modelo erra menos ao não detectar a dor. O F1 score de 87,27% é o mais alto, refletindo um excelente equilíbrio entre precisão e sensibilidade. O p-valor de McNemar (0,131) novamente não mostra discordância significativa, e o Kappa de Cohen de 0,743 indica um nível de concordância quase perfeito com os rótulos de diagnóstico de dor, realizados pelos avaliadores.

No geral, o modelo apresentou bons resultados em todos os movimentos faciais, com destaque para o movimento de sulco nasolabial aprofundado, que teve o melhor desempenho global em todas as métricas, sendo a região mais confiável para o diagnóstico de dor. A boca estirada e boca aberta apresentaram resultados bons e relativamente similares, com boas concordâncias, apesar de uma menor sensibilidade. Os movimentos de fenda palpebral estreitada e fronte saliente foram as regiões com o desempenho mais fraco, especialmente devido à baixa sensibilidade e altas taxas de falsos negativos, o que compromete a capacidade de diagnóstico nessas áreas.

Comparação global: Diferentes conjuntos de score dos melhores modelos de cada movimento facial *versus* o rótulo de diagnóstico apresentado no banco de dados UNIFESP-I.

Para realizar as análises de concordância entre os diferentes conjuntos de score dos melhores modelos de cada movimento facial e o rótulo de diagnóstico de cada imagem apresentado no banco de dados UNIFESP-I, foram calculadas as taxas de acurácia, sensibilidade, especificidade, falsos positivos, falsos negativos, F1 score, valor-p de McNemar, e Kappa de Cohen. Os resultados estão descritos na Tabela 11.

Tabela 11 – Validação - Acurácia e Concordância entre os diferentes conjuntos de *score* dos melhores modelos e o rótulo de diagnóstico de dor do conjunto de dados UNIFESP-I

<i>Métrica</i>	<i>Score 1</i>	<i>Score 2</i>	<i>Score 3</i>	<i>Score 4</i>	<i>Score 5</i>
Acurácia	81.48%	87.04%	79.63%	77.78%	64.81%
Sensibilidade	96.67%	93.33%	80.00%	60.00%	36.67%
Especificidade	62.50%	79.17%	79.17%	100.00%	100.00%
Taxa Falsos Positivos	37.50%	20.83%	20.83%	0.00%	0.00%
Taxa Falsos Negativos	3.33%	6.67%	20.00%	40.00%	63.33%
F1 score	85.29%	88.89%	81.36%	75.00%	53.66%
McNemar (valor-p)	0.027	0.450*	1.000*	0.001	0.000
Kappa de Cohen	0.612	0.734	0.589	0.571	0.340

Fonte: Autor

Legenda: * $p > 0,05$ = não houve discordância

Quanto à concordância entre os diferentes conjuntos de *score* dos melhores modelos e o rótulo de diagnóstico apresentado no banco de dados UNIFESP-I, podemos observar na Tabela 11 que:

- Score 1 (um único movimento facial):

Para o Score 1, em que o diagnóstico de dor é baseado na detecção de apenas um movimento facial indicativo de dor, a acurácia foi de 81,48%, mostrando uma boa correspondência entre o modelo e os profissionais de saúde. A sensibilidade foi de 96,67%, a mais alta entre todos os scores, indicando que o modelo é extremamente eficaz em identificar corretamente a presença de dor quando apenas um movimento facial é considerado. Entretanto, a especificidade de 62,5% foi a mais baixa, o que significa que o modelo tende a superestimar a presença de dor em alguns casos, com uma taxa de falsos positivos de 37,5%. A taxa de falsos negativos foi de apenas 3,33%, sugerindo que o modelo raramente deixa de identificar a dor. O F1 score de 85,29% reflete um bom equilíbrio entre precisão e sensibilidade. No entanto, o p-valor de McNemar (0,027) é significativo, o que indica que há uma discordância estatisticamente significativa entre o modelo e os avaliadores. No entanto, o Kappa de Cohen de 0,612 indica um nível de concordância substancial.

- Score 2 (dois movimentos faciais):

O Score 2 teve a melhor acurácia, com 87,04%, indicando que o modelo que considera dois movimentos faciais tem o melhor desempenho em termos de

correspondência com o diagnóstico de dor geral. A sensibilidade de 93,33% também é muito alta, sugerindo que o modelo detecta corretamente a dor na maioria dos casos. A especificidade de 79,17% é boa, mostrando que o modelo é mais equilibrado em evitar falsos positivos do que o Score 1, com uma taxa de falsos positivos de 20,83%. A taxa de falsos negativos foi de 6,67%, o que é muito baixa, e o F1 score de 88,89% reflete um excelente equilíbrio entre precisão e sensibilidade. O p-valor de McNemar (0,450) indica que não há discordância significativa entre o modelo e os avaliadores. O Kappa de Cohen de 0,734 indica um nível de concordância substancial, o mais alto entre os scores.

- Score 3 (três movimentos faciais):

Para o Score 3, a acurácia caiu para 79,63%, indicando que o desempenho é ligeiramente inferior ao de Score 2. A sensibilidade foi de 80,0%, o que ainda é bom, mas mostra uma capacidade reduzida do modelo em detectar a dor corretamente. A especificidade de 79,17% é a mesma do Score 2, sugerindo que o modelo tem a mesma capacidade de evitar falsos positivos, com uma taxa de falsos positivos de 20,83%. A taxa de falsos negativos foi de 20,0%, mais alta do que nos scores anteriores, sugerindo que o modelo está deixando de detectar a dor em um número maior de casos. O F1 score de 81,36% ainda é bom, mas inferior ao de Score 2. O p-valor de McNemar (1.000) indica que não há discordância significativa entre o modelo e os avaliadores. O Kappa de Cohen de 0,589 reflete um nível de concordância moderado.

- Score 4 (quatro movimentos faciais):

O Score 4 apresentou uma acurácia de 77,78%, o que indica uma diminuição na correspondência entre o modelo e os avaliadores. A sensibilidade de 60,0% é significativamente menor, mostrando que o modelo tem dificuldade em detectar a dor corretamente quando quatro movimentos faciais são considerados. No entanto, a especificidade foi de 100,0%, o que significa que o modelo não comete erros ao classificar casos sem dor, com uma taxa de falsos positivos de 0%. Por outro lado, a taxa de falsos negativos de 40,0% é bastante alta, sugerindo que o modelo frequentemente deixa de identificar a dor quando ela está presente. O F1 score de 75% reflete essa dificuldade. O p-valor de McNemar (0,001) é estatisticamente

significativo, sugerindo uma discordância considerável entre o modelo e os avaliadores. O Kappa de Cohen de 0,571 indica um nível de concordância moderado.

- Score 5 (cinco movimentos faciais):

Para o Score 5, a acurácia foi de apenas 64,81%, a mais baixa entre todos os scores. A sensibilidade de 36,67% foi a mais baixa, indicando que o modelo raramente detecta corretamente a dor quando todos os cinco movimentos faciais são considerados. A especificidade de 100,0% é excelente, sem falsos positivos, mas a taxa de falsos negativos de 63,33% é muito alta, mostrando que o modelo frequentemente deixa de detectar a dor. O F1 score de 53,66% reflete o desequilíbrio entre a sensibilidade baixa e a especificidade alta. O p-valor de McNemar (0,000) indica uma discordância estatisticamente significativa, e o Kappa de Cohen de 0,340 sugere um nível de concordância leve, o mais baixo entre os scores.

Podemos observar nesta análise que a sensibilidade diminui conforme o número de movimentos faciais (*Score*) aumenta, enquanto a especificidade tende a aumentar. E este é um comportamento relevante e tem implicações importantes tanto para a interpretação dos resultados quanto para o desenvolvimento de sistemas de diagnóstico baseados em múltiplos parâmetros.

A sensibilidade apresenta essa queda porque, ao exigir que mais movimentos faciais sejam detectados simultaneamente para confirmar a dor, o método se torna mais criterioso. Em outras palavras, torna-se menos provável que todos os movimentos faciais relevantes ocorram ao mesmo tempo, o que reduz a taxa de identificação dos verdadeiros positivos. Em contrapartida, a especificidade aumenta, pois o modelo se torna muito preciso ao identificar a ausência de dor à medida que mais movimentos são requeridos para confirmar a dor. Isso é esperado, pois, com uma maior exigência de movimentos faciais simultâneos, o modelo se torna mais conservador, errando menos ao diagnosticar dor quando ela não existe.

Diante disso, podemos sugerir que o uso de *scores* de um a três movimentos faciais oferece boas taxas de concordância, altas taxas de sensibilidade e F1 *score*, sendo suficiente para o diagnóstico de dor em muitos casos, especialmente no caso de dois movimentos (*Score 2*), que apresentou o melhor equilíbrio entre sensibilidade, especificidade e F1 *score*, além de uma concordância substancial. Aumentar o número de movimentos faciais para quatro ou mais não necessariamente melhora o

desempenho e pode até diminuir a sensibilidade e aumentar a taxa de falsos negativos.

Esses resultados sugerem que o diagnóstico de dor pode ser feito com menos movimentos faciais, desde que escolhidos com critério, sem necessidade de usar todos os cinco movimentos para obter uma concordância robusta.

Após esta análise regional e global da face, para validação dos melhores modelos, podemos inferir que é possível diagnosticar dor com maior acurácia ao focar nos movimentos de sulco nasolabial aprofundado, boca aberta, e boca estirada. Esses movimentos apresentaram os melhores desempenhos nas métricas analisadas, tanto isoladamente quanto em conjunto.

Além disso, quando apenas um desses movimentos é pontuado como indicativo de dor, há uma alta probabilidade de que o diagnóstico esteja correto, mesmo que esse seja o único movimento identificado. Portanto, esses movimentos faciais seriam fortes indicadores de dor e poderiam ser utilizados no diagnóstico automatizado da dor neonatal.

5.4 TESTE DOS MELHORES MODELOS VALIDADOS

Após a validação dos melhores modelo para avaliar cada movimento facial indicativo de dor (boca aberta, boca estirada, fenda palpebral estreitada, fronte saliente e sulco nasolabial aprofundado), realizamos testes de acurácia e concordância entre o que o melhor modelo predisse como classificação e os rótulos fornecidos pelos bancos de dados independentes iCOPE e UNIFESP-II.

As imagens do banco de dados iCOPE foram utilizadas para testar o diagnóstico de dor em faces livres. As imagens do banco de dados UNIFESP-II, foram utilizadas para testar o diagnóstico de dor em neonatos internados em UTIN, a fim de que o teste também fosse realizado em imagens que poderiam ter regiões faciais ocluídas por algum dispositivo de suporte (Tabela 12).

Vale ressaltar que, as regiões rotuladas pelos avaliadores como ocluídas foram descartadas da análise. No entanto, algumas regiões rotuladas como visíveis estavam parcialmente ocluídas. Isso aconteceu quando os neonatos faziam uso de dispositivo de suporte fixados à face, mas os avaliadores não consideraram como empecilho para análise do movimento facial. Nestes casos, quando o MediaPipe não identificou

a face do neonato de forma automática, a identificação e a segmentação das regiões faciais foram feitas de forma manual.

Tabela 12 - Dados de teste

<i>Base de Dados</i>	<i>Número de imagens</i>	<i>Faces com dor</i>	<i>Faces sem dor</i>	<i>BA</i>	<i>BE</i>	<i>FP</i>	<i>FS</i>	<i>SN</i>	
				<i>Regiões rotuladas como visíveis</i>					39 (81%)
<i>Movimentos indicativos de dor</i>					13 (27%)	10 (21%)	17 (35%)	18 (37%)	11 (23%)
iCOPE	167	79	88	-	-	-	-	-	

Fonte: Autor

Legenda: BA = Boca Aberta / BE = Boca Estirada / FP = Fenda Palpebral Estreitada / FS = Fronte Saliente / SN = Sulco Nasolabial Aprofundado / *Face sem dor ou não visível

Assim como na etapa de validação, os testes foram feitos pelos cálculos das taxas de acurácia, sensibilidade, especificidade, falsos positivos, falsos negativos, F1 score, valor-p de McNemar, e Kappa de Cohen.

As comparações dos melhores modelos com os rótulos fornecidos pelos bancos de dados iCOPE e UNIFESP-II também se deram de forma regional e de forma global, sendo que:

- **Forma regional:** Analisamos a concordância apenas para o conjunto de dados UNIFESP-II, pois o banco fornece os dados relacionados à cada movimento facial.
- **Forma global:** Analisamos a concordância tanto para o conjunto de dados UNIFESP-II quanto para iCOPE, pois os dois conjuntos forneciam o rótulo de diagnóstico de imagem facial “com dor” e “sem dor”.

5.4.1 Teste dos melhores modelos para as imagens com a face livre

As 167 imagens do banco de dados iCOPE foram utilizadas para testar o diagnóstico final de dor, porém, durante o processo de detecção da face, apenas 141

faces foram detectadas pelo Media Pipe. Destas, 61 imagens estavam rotuladas como “face com dor” e 80 imagens estavam rotuladas como “face sem dor”.

Vale ressaltar que não foi feita a comparação regional pois os desenvolvedores do banco iCOPE não disponibilizaram uma avaliação individual sobre cada movimento facial, apenas o diagnóstico final de face “com dor” ou “sem dor”.

Comparação global: Melhores modelos de cada movimento facial versus o rótulo de diagnóstico apresentado no banco de dados iCOPE

Para testar a concordância entre os melhores modelos de cada movimento facial e o rótulo de diagnóstico de cada imagem apresentado no banco de dados iCOPE, foram calculadas as taxas de acurácia, sensibilidade, especificidade, falsos positivos, falsos negativos, F1 score, valor-p de McNemar, e Kappa de Cohen. Os resultados estão descritos na Tabela 13.

Tabela 13 – Teste - Acurácia e Concordância entre os melhores modelos e o rótulo de diagnóstico de dor do conjunto de dados iCOPE

<i>Métrica</i>	<i>BA</i>	<i>BE</i>	<i>FP</i>	<i>FS</i>	<i>SN</i>
Acurácia	59.57%	80.85%	68.79%	51.06%	78.72%
Sensibilidade	90.16%	80.33%	77.05%	100.00%	96.72%
Especificidade	36.25%	81.25%	62.50%	13.75%	65.00%
Taxa Falsos Positivos	63.75%	18.75%	37.50%	86.25%	35.00%
Taxa Falsos Negativos	9.84%	19.67%	22.95%	0.00%	3.28%
F1 score	65.87%	78.40%	68.12%	63.87%	79.73%
McNemar (valor-p)	0.000	0.700*	0.024	0.000	0.000
Kappa de Cohen	0.243	0.612	0.384	0.121	0.587

Fonte: Autor

Legenda: BA = Boca Aberta / BE = Boca Estirada / FP = Fenda Palpebral Estreitada / FS = Fronte Saliente / SN = Sulco Nasolabial Aprofundado / * p > 0,05 = não houve discordância

Ao comparar os resultados dos melhores modelos para cada movimento facial validados no conjunto de dados UNIFESP-I com os resultados do rótulo do conjunto de dados de teste iCOPE, da Tabela 13, podemos observar variações significativas nas métricas de desempenho para os diferentes movimentos faciais.

Para o movimento de boca aberta, a acurácia no conjunto de dados UNIFESP-I foi maior (75,93%) em comparação com o teste iCOPE (59,57%). Em relação à

sensibilidade, o teste iCOPE apresentou um valor mais alto (90,16%) em relação ao UNIFESP-I (73,3%), mostrando que o modelo detecta dor de forma mais eficiente no teste. No entanto, a especificidade no iCOPE foi consideravelmente menor (36,25%) em comparação com o UNIFESP-I (79,2%), revelando uma maior dificuldade do modelo em identificar corretamente a ausência de dor no teste. O F1 score foi superior no UNIFESP-I (77,2%) em comparação ao iCOPE (65,87%), indicando um desempenho mais equilibrado no conjunto de validação. O Kappa de Cohen foi moderado no UNIFESP-I (0,519), enquanto no iCOPE (0,243) foi pobre, reforçando o menor nível de concordância no teste. O teste de McNemar foi significativo no iCOPE ($p = 0.000$), indicando discordância significativa entre o movimento de boca aberta e o rótulo de dor, enquanto no UNIFESP-I ($p = 0.579$) não houve discordância relevante.

A acurácia no iCOPE, de 80,85%, para o movimento de boca estirada, foi superior à do conjunto UNIFESP-I (77,78%). A sensibilidade também foi ligeiramente superior no iCOPE (80,33%) em comparação ao UNIFESP-I (73,3%), demonstrando que o modelo teve um desempenho mais eficiente no teste. A especificidade, por outro lado, foi semelhante em ambos os conjuntos, com 81,25% no iCOPE e 83,3% no UNIFESP-I, sugerindo que o modelo conseguiu identificar bem a ausência de dor nos dois casos. O F1 score foi praticamente o mesmo, com 78,6% no UNIFESP-I e 78,4% no iCOPE, mostrando um desempenho consistente. O Kappa de Cohen foi maior no iCOPE (0,612) do que no UNIFESP-I (0,557), indicando um nível de concordância substancial. O teste de McNemar não foi significativo em ambos os conjuntos (iCOPE $p = 0.700$ e UNIFESP-I $p = 0.386$), sugerindo que não houve discordância entre o modelo e os rótulos em ambos os casos.

Para o movimento de fenda palpebral estreitada, a acurácia no UNIFESP-I (75,93%) foi maior do que no iCOPE (68,79%). A sensibilidade foi um pouco superior no iCOPE (77,05%) em comparação ao UNIFESP-I (70,0%), mas a especificidade foi menor no iCOPE (62,5%) em comparação ao UNIFESP-I (83,3%). A taxa de falsos positivos foi maior no iCOPE, resultando em um F1 score inferior no iCOPE (68,12%) em relação ao UNIFESP-I (76,4%). O Kappa de Cohen foi moderado no UNIFESP-I (0,522) e menor no iCOPE (0,384), sugerindo um nível de concordância leve. O teste de McNemar foi significativo no iCOPE ($p = 0.024$), sugerindo discordância entre o modelo e os rótulos, enquanto no UNIFESP-I ($p = 0.267$) não houve discordância significativa.

A acurácia foi significativamente maior no UNIFESP-I (74,07%), para o movimento de frente saliente, do que no iCOPE (51,06%). A sensibilidade no iCOPE foi perfeita (100,0%) em comparação com o UNIFESP-I (70,0%), mas a especificidade no iCOPE foi extremamente baixa (13,75%) em comparação com o UNIFESP-I (79,2%), sugerindo uma alta taxa de falsos positivos no teste. Isso se reflete no F1 score, que foi muito menor no iCOPE (63,87%) em comparação ao UNIFESP-I (75%). O Kappa de Cohen foi substancialmente menor no iCOPE (0,121) em comparação ao UNIFESP-I (0,484), indicando um baixo nível de concordância. O teste de McNemar foi significativo no iCOPE ($p = 0.000$), indicando discordância significativa, enquanto no UNIFESP-I ($p = 0.423$) não houve discordância significativa.

Para o movimento de sulco nasolabial aprofundado, a acurácia foi ligeiramente menor no iCOPE (78,72%) em comparação ao UNIFESP-I (87,04%). A sensibilidade foi superior no iCOPE (96,7%) em comparação ao UNIFESP-I (80,0%), enquanto a especificidade foi menor no iCOPE (65,0%) em comparação ao UNIFESP-I (95,8%), indicando uma maior taxa de falsos positivos no teste. O F1 score foi ligeiramente menor no iCOPE (79,73%) em comparação ao UNIFESP-I (87,3%). O Kappa de Cohen foi substancial no UNIFESP-I (0,743) e moderado no iCOPE (0,587), sugerindo um nível de concordância menor. O teste de McNemar foi significativo no iCOPE ($p = 0.000$), sugerindo discordância significativa, enquanto no UNIFESP-I ($p = 0.131$) não houve discordância relevante.

No geral, a análise comparativa entre o conjunto de validação e o conjunto de teste revela que o modelo de boca estirada apresentou desempenho mais consistente entre os dois conjuntos de dados, UNIFESP-I e iCOPE respectivamente. Os movimentos de fenda palpebral estreitada, boca aberta e frente saliente tiveram um desempenho mais inconsistente no teste, com maior incidência de falsos positivos. O sulco nasolabial aprofundado permaneceu um dos melhores movimentos, embora com uma menor especificidade e aumento da taxa de falsos positivos no conjunto de teste.

Comparação global: Diferentes conjuntos de score dos melhores modelos de cada movimento facial versus o rótulo de diagnóstico apresentado no banco de dados iCOPE.

Para testar a concordância entre os diferentes conjuntos de *score* dos melhores modelos de cada movimento facial e o rótulo de diagnóstico de cada imagem apresentado no banco de dados iCOPE, foram calculadas analogamente as taxas de acurácia, sensibilidade, especificidade, falsos positivos, falsos negativos, F1 *score*, valor-p de McNemar, e Kappa de Cohen. Os resultados estão descritos na Tabela 14.

Tabela 14 – Teste - Acurácia e Concordância entre os diferentes conjuntos de *score* dos melhores modelos e o rótulo de diagnóstico de dor do conjunto de dados iCOPE

<i>Métrica</i>	<i>Score 1</i>	<i>Score 2</i>	<i>Score 3</i>	<i>Score 4</i>	<i>Score 5</i>
Acurácia	46.10%	58.87%	71.63%	82.98%	79.43%
Sensibilidade	100.00%	98.36%	98.36%	88.52%	59.02%
Especificidade	5.00%	28.75%	51.25%	78.75%	95.00%
Taxa Falsos Positivos	95.00%	71.25%	48.75%	21.25%	5.00%
Taxa Falsos Negativos	0.00%	1.64%	1.64%	11.48%	40.98%
F1 <i>score</i>	61.62%	67.42%	75.00%	81.82%	71.29%
McNemar (valor-p)	0.000	0.000	0.000	0.066*	0.000
Kappa de Cohen	0.044	0.244	0.462	0.660	0.563

Fonte: Autor

Legenda: * $p > 0,05$ = não houve discordância

Podemos observar na Tabela 14 que, ao comparar os diferentes conjuntos de *score* dos melhores modelos validados no banco de dados UNIFESP-I com o banco de dados de teste iCOPE, surgem diferenças importantes nas métricas de concordância.

- **Score 1 (um único movimento facial):**

No conjunto de dados iCOPE, a acurácia foi de 46,10%, indicando que o modelo teve um desempenho muito fraco ao diagnosticar dor com base em apenas um movimento facial. A sensibilidade foi perfeita (100,0%), ou seja, o modelo detectou corretamente todos os casos de dor, mas a especificidade foi extremamente baixa (5%), o que significa que o modelo cometeu muitos falsos positivos, classificando erroneamente a ausência de dor como dor. O F1 *score* foi de 61,62%, refletindo um desempenho desequilibrado, enquanto o Kappa de Cohen foi muito baixo (0,044), sugerindo uma quase inexistente concordância entre o modelo e os avaliadores no teste iCOPE. O valor-p de McNemar foi significativo ($p = 0.000$), confirmando uma discordância importante entre o modelo e os rótulos. Embora o modelo tenha

detectado todos os casos de dor, sua altíssima taxa de falsos positivos inviabiliza o uso de apenas um movimento facial para um diagnóstico confiável.

- Score 2 (dois movimentos faciais):

Para o *Score 2*, a acurácia foi um pouco melhor, de 58,87%, mas ainda abaixo de um nível aceitável para um diagnóstico confiável. A sensibilidade foi muito alta (98,36%), sugerindo que o modelo quase sempre detectou os casos de dor corretamente. No entanto, a especificidade foi de apenas 28,75%, indicando que o modelo continuou a classificar erroneamente muitos casos de ausência de dor como dor, com uma taxa de falsos positivos de 71,25%. O *F1 score* foi de 67,42%, refletindo uma leve melhora em relação ao *Score 1*. O Kappa de Cohen subiu para 0,244, mas ainda reflete um nível de concordância pobre. O teste de McNemar continuou significativo ($p = 0.000$), confirmando a discordância entre o modelo e os rótulos. Embora o desempenho tenha melhorado em comparação ao *Score 1*, o modelo ainda comete muitos erros ao identificar a ausência de dor.

- Score 3 (três movimentos faciais):

O *Score 3* apresentou uma acurácia significativamente melhor no iCOPE, de 71,63%, o que indica um desempenho mais confiável em comparação com os *scores* anteriores. A sensibilidade permaneceu alta (98,36%), enquanto a especificidade subiu para 51,25%, mostrando uma redução na taxa de falsos positivos (48,75%). O *F1 score* de 75% sugere um desempenho mais equilibrado, e o Kappa de Cohen subiu para 0,462, indicando um nível de concordância moderado. O teste de McNemar foi significativo ($p = 0.000$), o que ainda sugere discordância entre o modelo e os rótulos. O *Score 3* se mostrou mais confiável para o diagnóstico de dor, mas o modelo ainda cometeu erros ao classificar casos de ausência de dor.

- Score 4 (quatro movimentos faciais):

No iCOPE, o *Score 4* apresentou uma acurácia de 82,98%, a mais alta entre os *scores*, sugerindo que o modelo se aproximou de um nível mais aceitável de confiabilidade. A sensibilidade caiu para 88,52%, indicando que o modelo perdeu um pouco de capacidade de detectar todos os casos de dor. Por outro lado, a especificidade subiu significativamente para 78,75%, com uma taxa de falsos positivos de apenas 21,25%, sugerindo que o modelo ficou mais equilibrado ao diagnosticar a

ausência de dor. O F1 score foi de 81,82%, o mais alto entre os scores, e o Kappa de Cohen subiu para 0,660, indicando um nível de concordância substancial. O valor-p de McNemar ($p = 0.066$) não foi significativo, sugerindo que não houve discordância relevante entre o modelo e os avaliadores. O Score 4 parece ser a melhor opção para o diagnóstico de dor no iCOPE, com uma boa combinação de sensibilidade e especificidade.

- Score 5 (cinco movimentos faciais):

Para o Score 5, a acurácia foi de 79,43%, um pouco inferior ao Score 4, mas ainda consideravelmente boa. A sensibilidade caiu para 59,02%, o que indica que o modelo deixou de detectar muitos casos de dor. No entanto, a especificidade foi a mais alta de todas (95,0%), com uma taxa de falsos positivos muito baixa (5,0%), o que sugere que o modelo foi muito preciso em identificar corretamente a ausência de dor. O F1 score foi de 71,29%, refletindo um desempenho razoável, e o Kappa de Cohen foi de 0,563, indicando um nível de concordância moderado. O teste de McNemar foi significativo ($p = 0.000$), sugerindo discordância com os rótulos. Embora o Score 5 tenha a melhor especificidade, a queda acentuada na sensibilidade compromete sua confiabilidade para detectar dor em todos os casos.

Analisando os resultados do conjunto de dados iCOPE, o Score 4 se destaca como a melhor opção para o diagnóstico de dor. Ele apresentou a maior acurácia (82,98%), uma sensibilidade relativamente alta (88,52%), e uma especificidade considerável (78,75%), além de ter o maior F1 score (81,82%) e um nível de concordância substancial conforme o Kappa de Cohen (0,660). Além disso, o teste de McNemar não foi significativo, indicando que não houve discordância relevante entre o modelo e os avaliadores.

Em contrapartida, Scores 1 e 2 apresentaram muitos falsos positivos, com especificidades muito baixas, o que inviabiliza seu uso para diagnóstico confiável de dor. Embora o Score 5 tenha uma alta especificidade, a queda na sensibilidade o torna inadequado para detectar dor de forma eficiente.

Portanto, para o conjunto de dados iCOPE, o Score 4 é a escolha mais confiável, oferecendo um bom equilíbrio entre sensibilidade e especificidade, além de ter a menor discordância com os avaliadores.

Podemos observar que no conjunto de dados iCOPE a sensibilidade aumenta com o uso de mais scores. Isso significa que, à medida que mais movimentos são

considerados, o modelo é mais capaz em detectar a dor. O oposto ocorre com a especificidade, que tende a cair à medida que mais movimentos são utilizados. Isso sugere uma troca típica entre sensibilidade e especificidade, assim como no conjunto de validação com o banco de dados UNIFESP-I

A comparação entre os diferentes conjuntos de *score* validados no UNIFESP-I e testados no iCOPE mostra que o desempenho do modelo varia significativamente entre os dois conjuntos. O uso de dois ou mais movimentos faciais parece ser o mais eficiente para ambos os conjuntos, oferecendo o melhor equilíbrio entre sensibilidade e especificidade, além de boas taxas de concordância e F1 *score*. No entanto, o uso de todos os cinco movimentos faciais não necessariamente melhora o desempenho e, em alguns casos, pode aumentar a taxa de falsos negativos, especialmente no iCOPE.

Portanto, esta análise comparativa sugere que o diagnóstico de dor pode ser realizado com dois a quatro movimentos faciais, desde que escolhidos criteriosamente, sem necessidade de usar todos os cinco movimentos para obter um diagnóstico confiável.

5.4.2 Teste dos melhores modelos para as imagens de neonatos internados em UTIN

As 48 imagens do banco de dados UNIFESP-II foram utilizadas para testar os melhores modelos para os movimentos indicativos de dor e o diagnóstico final de dor (Tabela 12).

Durante o processo de detecção da face, apenas em 19 imagens a face foi detectada automaticamente pelo MediaPipe. Para as outras 29 imagens, a face e as regiões faciais indicativas de dor foram segmentadas de forma manual.

Comparação regional

Para testar a concordância entre os profissionais de saúde e o melhor modelo, em relação à cada movimento facial individual para o banco de dados UNIFESP-II, foram calculadas as taxas de acurácia, sensibilidade, especificidade, falsos positivos,

falsos negativos, F1 score, valor-p de McNemar, e Kappa de Cohen. Os resultados estão descritos na Tabela 15.

Tabela 15 – Teste - Acurácia e Concordância entre os melhores modelos e o rótulo dos movimentos faciais indicativos de dor do conjunto de dados UNIFESP-II

<i>Métrica</i>	<i>BA</i>	<i>BE</i>	<i>FP</i>	<i>FS</i>	<i>SN</i>
Acurácia	84.62%	81.58%	65.91%	72.73%	72.41%
Sensibilidade	70.59%	61.54%	56.25%	68.75%	66.67%
Especificidade	95.45%	92.00%	71.43%	75.00%	75.00%
Taxa Falsos Positivos	4.55%	8.00%	28.57%	25.00%	25.00%
Taxa Falsos Negativos	29.41%	38.46%	43.75%	31.25%	33.33%
F1 score	80.00%	69.57%	54.55%	64.71%	60.00%
McNemar (valor-p)	0.221*	0.450*	1.000*	0.773*	0.724*
Kappa de Cohen	0.679	0.567	0.273	0.426	0.393

Fonte: Autor

Legenda: BA = Boca Aberta / BE = Boca Estirada / FP = Fenda Palpebral Estreitada / FS = Fronte Saliente / SN = Sulco Nasolabial Aprofundado / * p > 0,05 = não houve discordância

Podemos observar na Tabela 15 que para o movimento de boca aberta, a taxa de acurácia foi de 84,62%, indicando que o modelo apresentou um desempenho próximo ao dos avaliadores. A sensibilidade de 70,59% mostra que o modelo detectou corretamente a maioria dos casos de dor, enquanto a especificidade de 95,45% reflete uma excelente capacidade de identificar corretamente a ausência de dor. As taxas de falsos positivos (4,55%) e falsos negativos (29,41%) indicam que o modelo classificou erroneamente alguns casos em que a dor estava presente. No entanto, o F1 score foi de 80%, o que sugere um bom equilíbrio entre precisão e sensibilidade. O valor-p de McNemar (0,221) indica que não há discordância significativa entre o modelo e os avaliadores. O Kappa de Cohen foi de 0,679, indicando um nível de concordância substancial entre o modelo e os profissionais de saúde.

Para o movimento de boca estirada, a taxa de acurácia foi de 81,58%, revelando que o modelo apresentou um bom desempenho ao detectar dor. A sensibilidade de 61,54% sugere que o modelo falhou em detectar uma parcela considerável dos casos de dor, enquanto a especificidade de 92,0% mostra que ele identificou bem a ausência de dor. A taxa de falsos positivos foi baixa (8,0%), mas a taxa de falsos negativos foi alta (38,46%), indicando que o modelo cometeu mais erros ao não identificar dor onde ela estava presente. O F1 score de 69,57% reflete um desempenho moderado, com um equilíbrio razoável entre precisão e sensibilidade. O

valor-p de McNemar (0,450) indica que não houve discordância significativa entre o modelo e os avaliadores. O Kappa de Cohen foi de 0,567, indicando um nível de concordância moderado.

No movimento de fenda palpebral estreitada, observamos que a taxa de acurácia foi de 65,91%, a mais baixa entre os movimentos analisados, indicando que o modelo teve mais dificuldade em detectar dor corretamente. A sensibilidade de 56,25% revela que o modelo falhou em identificar dor em quase metade dos casos, enquanto a especificidade de 71,43% mostra uma capacidade moderada de evitar falsos diagnósticos de dor. As taxas de falsos positivos (28,57%) e falsos negativos (43,75%) indicam que o modelo cometeu muitos erros tanto ao superestimar quanto ao subestimar a presença de dor. O F1 score de 54,55% foi o mais baixo, refletindo um desempenho desequilibrado. O valor-p de McNemar (1.000) mostra que não houve discordância significativa entre o modelo e os avaliadores. O Kappa de Cohen foi de 0,273, sugerindo um nível de concordância leve.

Para o movimento de frente saliente, a taxa de acurácia foi de 72,73%, sugerindo um desempenho razoável do modelo na detecção de dor. A sensibilidade de 68,75% indica que o modelo identificou a maioria dos casos de dor, enquanto a especificidade de 75,0% reflete uma capacidade moderada de evitar diagnósticos incorretos de dor. As taxas de falsos positivos (25,0%) e falsos negativos (31,25%) mostram que o modelo teve um número moderado de erros. O F1 score foi de 64,71%, mostrando um equilíbrio moderado entre precisão e sensibilidade. O valor-p de McNemar (0,773) indica que não houve discordância significativa entre o modelo e os avaliadores. O Kappa de Cohen foi de 0,426, indicando um nível de concordância moderado.

Para o movimento de sulco nasolabial aprofundado, a taxa de acurácia foi de 72,41%, sugerindo que o modelo teve um desempenho semelhante ao dos avaliadores. A sensibilidade foi de 66,67%, o que mostra que o modelo detectou corretamente a dor em dois terços dos casos, enquanto a especificidade de 75,0% indica que ele evitou bem os falsos positivos. As taxas de falsos positivos (25,0%) e falsos negativos (33,33%) mostram que o modelo cometeu um número moderado de erros. O F1 score foi de 60%, refletindo um desempenho razoável, com algum equilíbrio entre precisão e sensibilidade. O valor-p de McNemar (0,724) indica que não houve discordância significativa entre o modelo e os avaliadores. O Kappa de Cohen foi de 0,393, sugerindo um nível de concordância leve.

No geral, o movimento de boca aberta foi o que apresentou o melhor desempenho geral, com a maior concordância, alta especificidade, e o F1 score mais elevado. Isso indica que o modelo para boca aberta foi o mais eficaz para diagnosticar dor no banco de dados UNIFESP-II.

Comparação global: Melhores modelos de cada movimento facial versus o rótulo de diagnóstico apresentado no banco de dados UNIFESP-II

Para testar a concordância entre os melhores modelos de cada movimento facial e o rótulo de diagnóstico de cada imagem apresentado no banco de dados UNIFESP-II, foram calculadas as taxas de acurácia, sensibilidade, especificidade, falsos positivos, falsos negativos, F1 score, valor-p de McNemar, e Kappa de Cohen. Os resultados estão descritos na Tabela 16.

Tabela 16 – Teste - Acurácia e Concordância entre os melhores modelos e o rótulo de diagnóstico de dor do conjunto de dados UNIFESP-II

<i>Métrica</i>	<i>BA</i>	<i>BE</i>	<i>FP</i>	<i>FS</i>	<i>SN</i>
Acurácia	71.79%	76.32%	65.91%	68.18%	75.86%
Sensibilidade	68.75%	62.50%	52.94%	55.56%	60.00%
Especificidade	73.91%	86.36%	74.07%	76.92%	84.21%
Taxa Falsos Positivos	26.09%	13.64%	25.93%	23.08%	15.79%
Taxa Falsos Negativos	31.25%	37.50%	47.06%	44.44%	40.00%
F1 score	66.67%	68.97%	54.55%	58.82%	63.16%
McNemar (valor-p)	1.000*	0.505*	1.000*	0.789*	1.000*
Kappa de Cohen	0.423	0.501	0.273	0.330	0.453

Fonte: Autor

Legenda: BA = Boca Aberta / BE = Boca Estirada / FP = Fenda Palpebral Estreitada / FS = Fronte Saliente / SN = Sulco Nasolabial Aprofundado / * p > 0,05 = não houve discordância

Podemos observar na Tabela 16 que para o movimento de boca aberta, a acurácia no conjunto de dados UNIFESP-I foi maior (75,93%) em comparação com o UNIFESP-II (71,79%). Em termos de sensibilidade, o modelo no UNIFESP-I apresentou um valor mais alto (73,3%) em comparação ao UNIFESP-II (68,75%), o que indica que o modelo foi mais eficiente na detecção de dor no UNIFESP-I. A especificidade também foi superior no UNIFESP-I (79,2%) em relação ao UNIFESP-II (73,91%), revelando que o modelo foi mais eficaz em evitar diagnósticos incorretos de

dor no UNIFESP-I. O F1 score foi mais alto no UNIFESP-I (77,2%) em comparação com o UNIFESP-II (66,67%), sugerindo um desempenho mais equilibrado no conjunto de validação. O Kappa de Cohen foi moderado no UNIFESP-I (0,519), enquanto no UNIFESP-II foi menor (0,423), indicando um nível de concordância ligeiramente inferior no UNIFESP-II. O teste de McNemar não foi significativo em nenhum dos conjuntos ($p = 1.000$ no UNIFESP-II e $p = 0.579$ no UNIFESP-I), indicando que não houve discordância significativa em ambos os casos.

Para o movimento de boca estirada, a acurácia foi semelhante nos dois conjuntos, com 77,78% no UNIFESP-I e 76,32% no UNIFESP-II. No entanto, a sensibilidade foi maior no UNIFESP-I (73,3%) em comparação com o UNIFESP-II (62,5%), sugerindo que o modelo foi mais eficaz em detectar dor no UNIFESP-I. Por outro lado, a especificidade foi superior no UNIFESP-II (86,36%) em comparação ao UNIFESP-I (83,3%), mostrando que o modelo teve menos falsos positivos no UNIFESP-II. O F1 score foi inferior no UNIFESP-II (68,97%) em comparação ao UNIFESP-I (78,6%). O Kappa de Cohen foi moderado em ambos os conjuntos, com 0,557 no UNIFESP-I e 0,501 no UNIFESP-II, indicando uma leve queda no nível de concordância no UNIFESP-II. O teste de McNemar não foi significativo em nenhum dos conjuntos ($p = 0.505$ no UNIFESP-II e $p = 0.386$ no UNIFESP-I), sugerindo ausência de discordância significativa.

Para o movimento de fenda palpebral estreitada, a acurácia foi superior no UNIFESP-I (75,93%) em comparação ao UNIFESP-II (65,91%). A sensibilidade também foi maior no UNIFESP-I (70,0%) em relação ao UNIFESP-II (52,94%), indicando que o modelo teve mais dificuldade em identificar corretamente a dor no UNIFESP-II. A especificidade foi semelhante entre os dois conjuntos, com 74,07% no UNIFESP-II e 83,3% no UNIFESP-I. O F1 score foi consideravelmente maior no UNIFESP-I (76,4%) em comparação ao UNIFESP-II (54,55%), indicando um desempenho muito superior no UNIFESP-I. O Kappa de Cohen foi moderado no UNIFESP-I (0,522) e significativamente menor no UNIFESP-II (0,273), refletindo um nível de concordância leve no UNIFESP-II. O teste de McNemar não foi significativo em ambos os conjuntos ($p = 1.000$ no UNIFESP-II e $p = 0.267$ no UNIFESP-I), sugerindo que não houve discordância significativa.

Para o movimento de frente saliente, a acurácia foi maior no UNIFESP-I (74,07%) em comparação ao UNIFESP-II (68,18%). A sensibilidade foi também superior no UNIFESP-I (70,0%) em relação ao UNIFESP-II (55,56%), o que indica que

o modelo teve um desempenho inferior ao identificar dor no UNIFESP-II. A especificidade foi semelhante entre os dois conjuntos, com 76,92% no UNIFESP-II e 79,2% no UNIFESP-I. O F1 score foi superior no UNIFESP-I (75,0%) em comparação ao UNIFESP-II (58,82%), sugerindo um desempenho melhor no conjunto de validação. O Kappa de Cohen foi moderado no UNIFESP-I (0,484) e ligeiramente menor no UNIFESP-II (0,330), indicando um nível de concordância limitado no segundo conjunto. O teste de McNemar não foi significativo em nenhum dos conjuntos ($p = 0.789$ no UNIFESP-II e $p = 0.423$ no UNIFESP-I).

Para o movimento de sulco nasolabial aprofundado, a acurácia foi significativamente maior no UNIFESP-I (87,04%) em comparação ao UNIFESP-II (75,86%). A sensibilidade também foi superior no UNIFESP-I (80,0%) em relação ao UNIFESP-II (60,0%), indicando que o modelo foi mais eficaz em detectar dor no conjunto de validação. A especificidade foi também mais alta no UNIFESP-I (95,8%) em comparação com o UNIFESP-II (84,21%), mostrando uma menor taxa de falsos positivos no conjunto de validação. O F1 score foi superior no UNIFESP-I (87,3%) em relação ao UNIFESP-II (63,16%), sugerindo um desempenho muito mais robusto no UNIFESP-I. O Kappa de Cohen foi substancial no UNIFESP-I (0,743) e moderado no UNIFESP-II (0,453), refletindo um nível de concordância mais fraco no segundo conjunto. O teste de McNemar não foi significativo em nenhum dos conjuntos ($p = 1.000$ no UNIFESP-II e $p = 0.131$ no UNIFESP-I), indicando ausência de discordância significativa.

De forma geral, os modelos apresentaram melhor desempenho no conjunto de validação UNIFESP-I em relação ao UNIFESP-II para todos os movimentos faciais. A concordância, sensibilidade, F1 score e Kappa de Cohen foram consistentemente superiores no UNIFESP-I, sugerindo que o modelo teve maior precisão na detecção de dor e na identificação correta de ausência de dor nesse conjunto.

O movimento de sulco nasolabial aprofundado foi o que apresentou a maior discrepância entre os dois conjuntos, com um desempenho significativamente superior no UNIFESP-I. O fato de o sulco nasolabial ser a região menos visível devido à fixação dos dispositivos médicos pode ter influenciado o resultado do teste. Em contraste, o movimento de fenda palpebral estreitada teve o pior desempenho em ambos os conjuntos.

Comparação global: Diferentes conjuntos de *score* dos melhores modelos de cada movimento facial *versus* o rótulo de diagnóstico apresentado no banco de dados UNIFESP-II

Para testar a concordância entre os diferentes conjuntos de *score* dos melhores modelos de cada movimento facial e o rótulo de diagnóstico de cada imagem apresentado no banco de dados UNIFESP-II, foram calculadas as taxas de acurácia, sensibilidade, especificidade, falsos positivos, falsos negativos, F1 *score*, valor-p de McNemar, e Kappa de Cohen. Os resultados estão descritos na Tabela 17.

Tabela 17 – Teste - Acurácia e Concordância entre os diferentes conjuntos de *score* dos melhores modelos e o rótulo de diagnóstico de dor do conjunto de dados UNIFESP-II

<i>Métrica</i>	<i>Score 1</i>	<i>Score 2</i>	<i>Score 3</i>	<i>Score 4</i>	<i>Score 5</i>
Acurácia	54.17%	79.17%	72.92%	72.92%	66.67%
Sensibilidade	89.47%	68.42%	36.84%	31.58%	15.79%
Especificidade	31.03%	86.21%	96.55%	100.00%	100.00%
Taxa Falsos Positivos	68.97%	13.79%	3.45%	0.00%	0.00%
Taxa Falsos Negativos	10.53%	31.58%	63.16%	68.42%	84.21%
F1 <i>score</i>	60.71%	72.22%	51.85%	48.00%	27.27%
McNemar (valor-p)	0.000	0.752*	0.006	0.001	0.000
Kappa de Cohen	0.176	0.556	0.371	0.358	0.185

* $p > 0,05$ = não houve discordância

Ao comparar os diferentes conjuntos de *score* dos melhores modelos validados no banco de dados UNIFESP-I com o conjunto de dados de teste UNIFESP-II, da Tabela 17, observamos que:

- **Score 1 (um único movimento facial):**

Para o *Score 1*, a acurácia foi de 54,17%, o que indica um desempenho moderado, mas inferior ao esperado para um diagnóstico confiável. A sensibilidade foi alta (89,47%), sugerindo que o modelo foi capaz de identificar corretamente a maioria dos casos de dor. No entanto, a especificidade foi baixa (31,03%), o que significa que o modelo classificou muitos casos de ausência de dor como dor, resultando em uma taxa de falsos positivos muito alta (68,97%). O F1 *score* foi de 60,71, refletindo um equilíbrio mediano entre precisão e sensibilidade. O Kappa de Cohen foi baixo (0,176),

indicando um nível de concordância pobre entre o modelo e os avaliadores. O teste de McNemar foi significativo ($p = 0.000$), sugerindo uma discordância relevante entre o modelo e os avaliadores.

- Score 2 (dois movimentos faciais):

Para o *Score 2*, a acurácia foi de 79,17%, significativamente melhor em comparação com o *Score 1*. A sensibilidade foi de 68,42%, indicando que o modelo perdeu alguns casos de dor, mas ainda detectou a maioria. A especificidade foi alta (86,21%), mostrando que o modelo foi eficaz em evitar diagnósticos incorretos de dor. A taxa de falsos positivos foi baixa (13,79%) e o *F1 score* foi de 72,22%, sugerindo um desempenho equilibrado entre precisão e sensibilidade. O Kappa de Cohen foi moderado (0,556), indicando um bom nível de concordância entre o modelo e os avaliadores. O teste de McNemar não foi significativo ($p = 0.752$), sugerindo que não houve discordância relevante entre o modelo e os avaliadores.

- Score 3 (três movimentos faciais):

Para o *Score 3*, a acurácia foi de 72,92%, o que representa uma ligeira queda em relação ao *Score 2*. A sensibilidade foi baixa (36,84%), o que indica que o modelo falhou em detectar a maior parte dos casos de dor. Em contrapartida, a especificidade foi muito alta (96,55%), sugerindo que o modelo raramente classificou erroneamente a ausência de dor como dor. A taxa de falsos positivos foi muito baixa (3,45%), enquanto a taxa de falsos negativos foi elevada (63,16%). O *F1 score* foi de 51,85%, refletindo um desempenho desequilibrado, favorecendo a especificidade. O Kappa de Cohen foi moderado (0,371), sugerindo um nível de concordância limitado. O teste de McNemar foi significativo ($p = 0.006$), indicando discordância entre o modelo e os avaliadores.

- Score 4 (quatro movimentos faciais):

Para o *Score 4*, a acurácia foi de 72,92%, igual à do *Score 3*, mas com desempenho geral semelhante. A sensibilidade foi ainda mais baixa (31,58%), o que significa que o modelo deixou de identificar a maioria dos casos de dor. Em contrapartida, a especificidade foi perfeita (100,0%), sem falsos diagnósticos de dor onde não havia dor. A taxa de falsos positivos foi zero (0,0%), enquanto a taxa de falsos negativos foi muito alta (68,42%). O *F1 score* foi de 48,0%, refletindo esse

desempenho inferior, com maior foco na especificidade em detrimento da sensibilidade. O Kappa de Cohen foi de 0,358, indicando um nível de concordância limitado. O teste de McNemar foi significativo ($p = 0.001$), sugerindo uma discordância entre o modelo e os avaliadores.

- Score 5 (cinco movimentos faciais):

Para o *Score 5*, a acurácia foi a mais baixa (66,67%) entre os *Scores 2-5*. A sensibilidade foi a menor de todas (15,79%), indicando que o modelo falhou em detectar a maioria dos casos de dor. A especificidade foi perfeita (100,0%), o que significa que o modelo nunca classificou erroneamente a ausência de dor como dor. A taxa de falsos positivos foi zero (0,0%), mas a taxa de falsos negativos foi a mais alta (84,21%), indicando que a maior parte dos casos de dor foi perdida. O *F1 score* foi muito baixo (27,27%), sugerindo um desempenho muito desequilibrado. O Kappa de Cohen foi pobre (0,185), sugerindo um nível de concordância muito fraco entre o modelo e os avaliadores. O teste de McNemar foi significativo ($p = 0.000$), sugerindo uma discordância considerável entre o modelo e os avaliadores.

Entre os diferentes *scores*, o *Score 2* apresentou o melhor desempenho, com a maior acurácia (79,17%), alta especificidade (86,21%), e um Kappa de Cohen moderado (0,556), refletindo um bom nível de concordância entre o modelo e os avaliadores. Embora a sensibilidade tenha sido menor que no *Score 1* (68,4% contra 89,5%), o equilíbrio entre precisão e sensibilidade foi melhor no *Score 2*, como indicado pelo *F1 score* de 0,722. Os *Scores 3-5* apresentaram uma queda acentuada na sensibilidade, com *Score 5* sendo o menos eficiente, apesar da alta especificidade.

Vale levar em consideração que, no banco de dados UNIFESP-II, muitas regiões faciais não estavam visíveis (28 imagens com pelo menos uma região ocluída por dispositivos médicos). Nesses casos, foi raro que os cinco movimentos faciais estivessem disponíveis para a avaliação de dor.

A comparação entre os diferentes conjuntos de *score* validados no UNIFESP-I e testados no UNIFESP-II revela que o desempenho do modelo varia entre os dois conjuntos de dados. O uso de dois ou três movimentos faciais se mostrou o mais eficiente em ambos os conjuntos, oferecendo o melhor equilíbrio entre sensibilidade e especificidade, além de boas taxas de concordância e *F1 score*. O uso de todos os cinco movimentos faciais, no entanto, não necessariamente melhorou o desempenho,

especialmente no UNIFESP-II, em que se observou um aumento na taxa de falsos negativos.

Portanto, essa análise sugere que o diagnóstico de dor pode ser realizado com maior confiança ao utilizar dois ou três movimentos faciais, sem a necessidade de depender de todos os cinco para um diagnóstico preciso, sobretudo quando há oclusão parcial da face.

5.5 CLASSIFICAÇÃO E DIAGNÓSTICO DA DOR POR REGIÕES VISÍVEIS DA FACE

Com base nos dados obtidos ao longo da pesquisa, desenvolvemos um método automático de classificação e diagnóstico de dor para neonatos internados em UTIN, em que frequentemente há oclusão parcial da face devido a dispositivos de suporte à vida fixados. Esse método foi ajustado para operar de maneira eficiente, mesmo quando apenas parte do rosto está visível.

Propomos a implementação do diagnóstico de dor que leva em consideração apenas as áreas visíveis da face do neonato. Maximizando a precisão e eficiência do diagnóstico, e permitindo à equipe de saúde realizar uma avaliação confiável da dor, mesmo quando regiões faciais estejam parcialmente obstruídas. O método automático foi adaptado para lidar com essas condições de visibilidade limitada, garantindo maior segurança na detecção da dor sem depender da visualização completa de todas as áreas faciais.

Durante as análises, observamos que atribuir diferentes pesos aos movimentos faciais pode aumentar a precisão do diagnóstico. Por exemplo, o sulco nasolabial aprofundado se destacou como um dos movimentos mais importantes, com altas taxas de concordância, sensibilidade e F1 score consistente tanto nos conjuntos UNIFESP-I quanto UNIFESP-II (conjuntos que forneciam o rótulo relacionado aos movimentos faciais indicativos de dor). Movimentos como boca aberta e boca estirada também apresentaram desempenho robusto, enquanto fenda palpebral estreitada e fronte saliente tiveram um desempenho inferior.

Assim, sugerimos um novo método de classificação e diagnóstico de dor, considerando as regiões visíveis e atribuindo pontuações diferentes para os movimentos mais relevantes (Tabela 18).

Neste método, cada região visível da face é avaliada para determinar se o movimento facial indicativo de dor está presente ou ausente. Os movimentos de frente saliente e fenda palpebral estreitada recebem Score 1, boca aberta e boca estirada recebem Score 2, e o sulco nasolabial aprofundado recebe Score 3. A pontuação final deve ser interpretada da seguinte forma: com cinco movimentos visíveis, um Score ≥ 4 indica dor; com quatro movimentos, um Score ≥ 3 indica dor; com três movimentos, um Score ≥ 2 indica dor; e com um ou dois movimentos, um Score ≥ 1 indica dor.

Tabela 18 – Novo Método de Classificação e Diagnóstico Automático da Dor

<i>Região</i>	<i>Visível</i>	<i>Movimento</i>	<i>Ausente</i>	<i>Presente</i>
Testa	()	Fronte Saliente	() Score 0	() Score 1
Olhos	()	Fenda Palpebral Estreitada	() Score 0	() Score 1
Boca	()	Boca Aberta	() Score 0	() Score 2
	()	Boca Estirada	() Score 0	() Score 2
Sulco Nasolabial	()	Sulco Nasolabial Aprofundado	() Score 0	() Score 3

Fonte: Autor

Legenda: *Score: 5 movimentos visíveis = Score ≥ 4 indica dor; 4 movimentos visíveis = Score ≥ 3 indica dor; 3 movimentos visíveis = Score ≥ 2 indica dor; 1 ou 2 movimentos visíveis = Score ≥ 1 indica dor.*

O novo método atribui um peso maior (3) ao movimento de sulco nasolabial aprofundado devido à sua alta confiabilidade no diagnóstico da dor. Já os movimentos de boca aberta e boca estirada recebem peso 2, pois, embora sejam indicadores importantes de dor, não são tão consistentes quanto o sulco nasolabial aprofundado. Os movimentos de fenda palpebral estreitada e frente saliente recebem peso 1, devido ao seu desempenho inferior, mas ainda são relevantes para o diagnóstico quando visíveis.

Ao definir um diagnóstico de dor com scores diferenciados, dependendo da quantidade de movimentos faciais visíveis, foi possível desenvolver uma classificação sensível o suficiente para detectar a dor quando as regiões mais relevantes estão disponíveis, além de manter uma especificidade robusta para evitar diagnósticos equivocados baseados em movimentos menos confiáveis.

Para validar este novo método, combinamos dados dos conjuntos UNIFESP-I (15% das imagens que foram usadas na fase de teste dos modelos) e UNIFESP-II, que forneciam rótulos relacionados aos movimentos faciais reais indicativos de dor. Aplicamos os novos pesos a cada movimento facial visível, e a classificação das

imagens foi realizada com base nesse novo método. Os resultados comparativos estão apresentados na Tabela 19.

Tabela 19 – Dados de validação para um Novo Método de Classificação e Diagnóstico Automático da Dor

<i>Métrica</i>	<i>UNIFESP-I</i>	<i>UNIFESP-II</i>	<i>Novo Método</i>
Acurácia	79.63%	79.17%	83.33%
Sensibilidade	80.00%	68.42%	85.71%
Especificidade	79.17%	86.21%	81.13%
Taxa Falsos Positivos	20.83%	13.79%	18.87%
Taxa Falsos Negativos	20.00%	31.58%	14.29%
F1 score	81.36%	72.22%	83.17%
McNemar (valor-p)	1.000*	0.752*	0.628*
Kappa de Cohen	0.589	0.556	0.667

Fonte: Autor

Legenda: * $p > 0,05$ = não houve discordância

Durante o processo de validação, os resultados demonstraram que o modelo utilizando o novo método de classificação e diagnóstico automático da dor, para as duas bases de dados em conjunto, teve um desempenho superior em comparação ao modelo que utilizou o método tradicional, considerando as imagens de teste das bases de dados UNIFESP-I e UNIFESP-II.

Vale lembrar que, no modelo tradicional, a presença de dor no banco de dados UNIFESP-I era identificada quando 3 ou mais movimentos faciais indicativos de dor estavam presentes. Já para o banco UNIFESP-II, que envolvia neonatos com parte da face ocluída por dispositivos médicos, a avaliação de dor considerava a maioria dos movimentos visíveis, uma vez que algumas regiões faciais poderiam estar obstruídas.

Com este novo método, a acurácia subiu para 83,33%, e a sensibilidade foi particularmente elevada (85,71%), trazendo uma melhoria significativa na detecção de dor em comparação com métodos anteriores, especialmente ao UNIFESP-II (68,42%). A taxa de falsos negativos foi significativamente reduzida no novo método (14,29%), o que contribuiu para o aumento da sensibilidade.

Em termos de especificidade, o novo método (81,13%) apresentou um desempenho um pouco inferior ao UNIFESP-II (86,21), mas superior ao UNIFESP-I (79,17%), demonstrando que foi eficaz em evitar diagnósticos falsos de dor. A

pontuação F1 foi a mais alta no novo método (83,17%), refletindo um bom equilíbrio entre precisão e sensibilidade .

O Kappa de Cohen também foi superior no novo método (0,667), o que sugere uma avaliação mais robusta da dor. O teste de McNemar ($p = 0,628$) não foi significativo, indicando que não houve discordâncias relevantes no novo método.

O novo método de classificação e diagnóstico do dor baseado em regiões visíveis da face, com diferentes pesos para cada movimento facial, mostrou ser eficiente e confiável. Esse método permite um diagnóstico preciso, mesmo em cenários de visibilidade parcial da face, e oferece ser uma ferramenta útil para automatizar o diagnóstico clínico do dor em neonatos na UTIN.

Para testar o novo método, fizemos uma análise individual com todos os bancos de dados, sendo testadas: 360 imagens do banco de dados UNIFESP-I; 48 imagens do banco de dados UNIFESP-II; 356 imagens do banco de dados SynUnifesp e 167 do banco de dados iCOPE. A classificação das imagens foi realizada da forma tradicional e com base no novo método, e os resultados comparativos com os rótulos de diagnóstico de cada banco estão apresentados na Tabela 20.

Tabela 20 – Dados de teste do Novo Método de Classificação e Diagnóstico Automático da Dor

Métrica	UNIFESP-I	UNIFESP-I	UNIFESP-II	UNIFESP-II	SynUnifesp	SynUnifesp	iCOPE	iCOPE
	<i>Tradicional</i>	<i>Novo Método</i>						
Acurácia	90.28%	90.28%	79.17%	79.17%	87.04%	85.35%	71.63%	70.21%
Sensibilidade	88.24%	92.81%	68.42%	73.68%	95.45%	97.27%	98.36%	98.36%
Especificidade	91.79%	88.41%	86.21%	82.76%	83.27%	80.00%	51.25%	48.75%
Taxa Falsos Positivos	8.21%	11.59%	13.79%	17.24%	16.73%	20.00%	48.75%	51.25%
Taxa Falsos Negativos	11.76%	7.19%	31.58%	26.32%	4.55%	2.73%	1.64%	1.64%
F1 score	88.52%	89.03%	72.22%	73.68%	82.03%	80.45%	75.00%	74.07%
McNemar (valor-p)	1.000*	0.043	0.752*	1.000*	0.000	0.000	0.000	0.000
Kappa de Cohen	0.801	0.803	0.556	0.564	0.722	0.693	0.462	0.437

Fonte: Autor

Legenda: * $p > 0,05$ = não houve discordância

Os resultados apresentados na Tabela 20 comparam o desempenho do método de classificação e diagnóstico da dor entre os melhores modelos, usando o método tradicional e o novo método proposto, utilizando diferentes bancos de dados.

No banco de dados UNIFESP-I, tanto o método tradicional quanto o novo apresentaram a mesma taxa de acurácia (90,28%), sugerindo que o novo método não invalida o tradicional. No entanto, o novo método apresentou uma sensibilidade superior (92,81% versus 88,24%), demonstrando uma maior capacidade de detectar a presença de dor. A especificidade foi menor no novo método (88,41 contra 91,79%), indicando uma leve tendência do novo método gerar mais falsos positivos. Isso é reforçado pela taxa de falsos positivos que aumentou de 8,21% para 11,59%. A taxa de falsos negativos, por outro lado, foi reduzida no novo método, de 11,76% para 7,19%, o que reflete a melhora na detecção de dor.

Para o banco de dados UNIFESP-II, composto por imagens de neonatos com dispositivos médicos fixados à face, a acurácia permaneceu a mesma (79,17%) para ambos os métodos, sem diferenças aparentes. Isso é esperado, já que as imagens do UNIFESP-II foram rotuladas levando em consideração os dispositivos. No entanto, o novo método mostrou uma melhora na sensibilidade (73,68% contra 68,42%) e queda na especificidade (82,76% versus 86,21%), o que sugere que, embora o novo método tenha melhorado a capacidade de detectar dor, houve um leve aumento nas classificações de falsos positivos. A taxa de falsos negativos também foi reduzida no novo método (26,32% contra 31,58%), reforçando a eficiência do novo método em condições em que a face está parcialmente ocluída.

No banco de dados SynUnifesp, o método tradicional apresentou uma leve vantagem em termos de acurácia (87,04% versus 85,35%). No entanto, o novo método teve uma sensibilidade superior (97,27% contra 95,45%), o que demonstra uma melhoria na detecção da dor. A especificidade do novo método foi menor (80,0% contra 83,27%), refletindo um aumento na taxa de falsos positivos (20,0% no novo método, contra 16,73% no tradicional). Por outro lado, a taxa de falsos negativos foi significativamente reduzida, de 4,55% para 2,73%, indicando que o novo método foi mais eficaz em identificar corretamente a dor.

No banco de dados iCOPE o novo método apresentou resultados bastante semelhantes ao método tradicional, com uma ligeira queda na acurácia (70,21% contra 71,63%) e na especificidade (48,75% versus 51,25%). A sensibilidade se manteve igual (98,36%), sugerindo que ambos os métodos são igualmente eficazes na detecção de dor. O novo método, no entanto, gerou uma taxa de falsos positivos um pouco menor (48,75% versus 51,25%), mas manteve a mesma taxa de falsos negativos (1,64%).

Esses resultados indicam que o novo método de classificação e diagnóstico de dor não substitui os métodos tradicionais, mas otimiza a avaliação, especialmente em neonatos com dispositivos médicos na face (UNIFESP-II). O novo método é eficaz em reduzir a taxa de falsos negativos e aumentar a sensibilidade, o que é crucial para a detecção precisa da dor. A queda na especificidade, embora presente em alguns casos, não compromete significativamente o desempenho geral. Em bancos de dados com neonatos sem dispositivos o novo método também mostrou resultados competitivos, com desempenho semelhante ao método tradicional.

Portanto, o novo método oferece uma melhoria na detecção de dor, especialmente em cenários em que a face está parcialmente ocluída, sem comprometer a precisão em comparação com os métodos tradicionais.

Ao analisar as imagens que apresentaram discordância entre o método tradicional e o novo, percebemos que elas representam casos em que o diagnóstico pode ser incerto. Após a fase de validação e teste, observamos que o novo método tende a priorizar a sensibilidade, sugerindo um possível diagnóstico de dor em situações de dúvida. Essa discordância, que leva à sugestão de dor, pode ser útil para alertar o profissional de saúde a reavaliar o paciente com mais atenção. Essa revisão cuidadosa é especialmente importante quando partes da face estão ocluídas por dispositivos médicos, dificultando a observação direta dos sinais de dor.

No caso de um exemplo do conjunto de dados UNIFESP-I, a Figura 44 ilustra uma face que foi classificada tanto pelos avaliadores quanto pelo modelo, usando o método tradicional, como "sem dor". No entanto, o novo método rotulou essa face como "com dor". Essa diferença ocorreu porque o novo método identificou o sulco nasolabial aprofundado e a fenda palpebral estreitada como movimentos indicativos de dor. No novo método, quando todas as cinco regiões faciais estão visíveis e a face está livre de obstruções, um Score de 4 — sendo 3 pontos do sulco nasolabial aprofundado e 1 ponto da fenda palpebral estreitada — é suficiente para classificar o neonato com o diagnóstico de dor. Nesse caso, o novo método recomenda que o avaliador revise o diagnóstico com maior atenção, já que indicadores faciais relevantes foram identificados.

Figura 44 - Exemplo de imagem com diagnóstico de dor discordante no conjunto de dados UNIFESP-I.



Fonte: Autor (com permissão)

Exemplificando com uma imagem do conjunto de dados UNIFESP-II, a Figura 45 ilustra uma situação de discordância entre o método tradicional e o novo. Nesta imagem, as regiões da testa e do sulco nasolabial estavam ocluídas, e a imagem foi rotulada pelos avaliadores como "sem dor". O método tradicional seguiu a mesma linha de raciocínio, diagnosticando "sem dor", pois o único movimento visível, a boca aberta, não foi suficiente para classificar a imagem como "com dor". No método tradicional, apenas um movimento indicativo de dor, como o de boca aberta, não é suficiente para diagnosticar dor. Contudo, o novo método adotou uma abordagem

diferente. Ele atribuiu um peso maior ao movimento de boca aberta (Score 2) e, considerando que duas regiões faciais (testa e sulco nasolabial) estavam ocluídas, o método classificou a imagem como "com dor". Segundo o novo método, com três movimentos faciais visíveis e uma pontuação de Score 2, a imagem é rotulada como "com dor", o que diferenciou o diagnóstico do método tradicional.

Esse exemplo ilustra como o novo método pode ser mais sensível na detecção de dor, atribuindo maior peso aos movimentos visíveis e otimizando o diagnóstico em cenários de oclusão facial parcial. Isso reforça a importância de uma reavaliação clínica para garantir a precisão no diagnóstico de dor neonatal.

Figura 45 - Exemplo de imagem com diagnóstico de dor discordante no conjunto de dados UNIFESP-II.



Fonte: Autor (com permissão)

Observando uma imagem do conjunto de dados SynUnifesp, a Figura 46 ilustra uma outra situação de discordância entre o método tradicional e o novo método. Nesta imagem, o modelo usando o método tradicional discordou dos avaliadores, enquanto o novo método concordou com o diagnóstico de dor. Os avaliadores rotularam 3 movimentos como indicativos de dor: sulco nasolabial aprofundado, boca aberta e frente saliente. No método tradicional, o modelo identificou apenas o sulco nasolabial aprofundado e a boca aberta como indicativos de dor, mas como apenas 2 movimentos não são suficientes para o diagnóstico de dor no método tradicional, pelo menos 3 movimentos deveriam ter sido identificados. No entanto, no novo método, esses dois movimentos recebem uma pontuação mais alta: Score 3 para o sulco nasolabial aprofundado e Score 2 para a boca aberta. Com a face completamente visível, o novo método determina que um Score acima de 4 indica dor. Dessa forma, o novo método, ao atribuir maior peso a essas regiões, classifica corretamente a imagem como "com dor", concordando com os avaliadores.

Figura 46 - Exemplo de imagem com diagnóstico de dor discordante no conjunto de dados SynUnifesp.



Fonte: Base SynUnifesp (com permissão)

No conjunto de dados iCOPE, a única imagem que apresentou uma diferença entre o rótulo do método tradicional e o do novo método proposto foi a da Figura 47. Tanto os avaliadores quanto o método tradicional classificaram essa imagem como "sem dor", enquanto o novo método a rotulou como "com dor". Essa discrepância pode ter ocorrido porque, no novo método, a presença de boca aberta recebe um Score maior, o que influenciou o diagnóstico. Nesses casos, o novo método sugere ao avaliador que reavalie o neonato para confirmar ou descartar a presença de dor, promovendo uma ação mais cuidadosa e criteriosa.

Figura 47 - Exemplo de imagem com diagnóstico de dor discordante no conjunto de dados iCOPE



Fonte: iCOPE^(69,91,123,124) (com permissão)

Ao analisar qualitativamente as imagens que apresentaram discordância entre o método tradicional e o novo método, observamos que essas situações muitas vezes envolvem incertezas no diagnóstico. O novo método, priorizando a sensibilidade, é capaz de identificar sinais de dor que o método tradicional pode deixar passar, alertando o profissional de saúde a revisar o estado do paciente com maior atenção. Isso é particularmente útil em cenários clínicos em que partes do rosto do neonato estão obstruídas, dificultando a visualização direta de sinais de dor.

Exemplos práticos, como os apresentados nas Figuras 44 a 47, ilustram essa capacidade aprimorada do novo método de detectar dor em neonatos, atribuindo maior peso a movimentos faciais visíveis, como o sulco nasolabial aprofundado e a boca aberta. Esses casos reforçam a necessidade de uma reavaliação cuidadosa, com o novo método atuando como uma ferramenta de reforço para melhorar o diagnóstico de dor neonatal.

6 DISCUSSÃO

Na prática clínica, a avaliação da dor neonatal depende da observação detalhada por parte dos profissionais de saúde, especialmente quando a face do recém-nascido está completamente livre. No entanto, quando dispositivos de suporte à vida estão fixados à face do neonato, a avaliação da dor se torna mais desafiadora, já que parte do rosto fica obstruída.

Profissionais de saúde costumam redirecionar o olhar para áreas não cobertas ou baseiam-se em sinais complementares, como sinais vitais e movimento corporal⁽³⁴⁾. Isso torna evidente a necessidade de um método automatizado capaz de auxiliar na identificação da dor, mesmo em cenários em que o rosto está parcialmente coberto.

Interpretação dos Resultados

Nossos experimentos com recém-nascidos, utilizando o *Vision Transformer* (ViT), mostraram que o arcabouço computacional proposto é eficaz para diagnosticar dor. A implementação do ViT para o reconhecimento de dor neonatal demonstrou que o ajuste fino completo do modelo foi essencial para garantir a detecção precisa dos movimentos faciais.

A criação de bancos de dados sintéticos robustos pode acelerar o desenvolvimento de modelos mais precisos, conforme sugerido por Heiderich *et al.* (2023)⁽⁵¹⁾. No entanto, como nossos resultados indicaram, ao incluir imagens sintéticas no modelo de treinamento enfrentamos algumas dificuldades, já que, enquanto movimentos como fronte saliente se beneficiaram da adição de imagens artificiais, outros, como fenda palpebral estreitada, tiveram seu desempenho prejudicado. Esses resultados mostram que o uso de dados sintéticos precisa ser aplicado com cautela para não introduzir ruído na análise.

Neste estudo, movimentos como sulco nasolabial aprofundado e boca aberta, que consistentemente apresentaram os melhores resultados em termos de F1 score e concordância com os profissionais de saúde, apresentaram F1 scores elevados durante o treinamento, o que sugere que o modelo capturou bem as nuances desses movimentos para identificar a dor. Isso indica que essas áreas faciais têm um alto

valor discriminatório, alinhando-se à literatura existente, que também reconhece essas regiões como essenciais para o diagnóstico preciso da dor neonatal^(46,112,134).

Estudos anteriores destacam a importância de regiões faciais específicas, como fronte, fenda palpebral e sulco nasolabial, para a avaliação da dor neonatal. Vaart *et al.* (2019)⁽¹¹²⁾ observaram que essas regiões faciais, após estímulos nocivos, são altamente correlacionadas com a presença de dor, especialmente o movimento de sulco nasolabial aprofundado.

Principais Achados

Durante a fase de validação dos modelos, o sulco nasolabial destacou-se como o movimento com a maior concordância em relação aos rótulos de dor fornecidos pelos avaliadores, apresentando excelente sensibilidade e especificidade, além de baixas taxas de erro. Essa região mostrou-se a mais confiável para a detecção de dor. O modelo para boca aberta também apresentou um desempenho consistente, com alta sensibilidade e especificidade, demonstrando eficácia moderada na previsão do diagnóstico de dor. O modelo para boca estirada foi robusto, com uma leve diminuição na sensibilidade em comparação com boca aberta, mas compensando com maior especificidade e um F1 score equilibrado, tornando-o mais eficaz para o diagnóstico global. Já o desempenho da fenda palpebral estreitada foi inferior em termos de precisão, principalmente devido à maior ocorrência de falsos positivos, indicando maior dificuldade na detecção de dor nessa região. Da mesma forma, a fronte saliente apresentou desafios, com menor sensibilidade e uma taxa de falsos negativos mais elevada, o que dificultou a detecção adequada da dor.

Na avaliação dos diferentes Scores para o diagnóstico da dor neonatal, o Score 1 mostrou alta sensibilidade, porém com uma especificidade menor, sugerindo que o modelo tende a superestimar a dor quando apenas um movimento facial é considerado. O Score 2, por outro lado, foi o mais equilibrado, com altos valores de sensibilidade e especificidade, além de um excelente equilíbrio entre precisão e sensibilidade, sendo a opção mais eficaz para o diagnóstico global de dor. O Score 3 mostrou resultados aceitáveis, enquanto os Scores 4 e 5 apresentaram quedas acentuadas na sensibilidade e aumentos na taxa de falsos negativos, comprometendo

o desempenho e indicando que essas abordagens não são ideais para o diagnóstico de dor.

Observamos ainda que, conforme o número de movimentos faciais exigidos para o diagnóstico aumentava, a sensibilidade dos modelos diminuía enquanto a especificidade aumentava. Esse comportamento reflete a clássica troca (*trade-off*) entre sensibilidade e especificidade observada em muitas avaliações de classificação e diagnóstico, conhecida como curva ROC (*Receiver Operating Characteristic*)⁽¹³⁵⁾. À medida que se aumenta o critério para detecção positiva, o modelo torna-se mais específico, porém com a perda de verdadeiros positivos, o que reduz a sensibilidade.

Clinicamente, essa dinâmica exige um equilíbrio delicado. Aumentar o número de movimentos faciais necessários para o diagnóstico de dor pode evitar falsos diagnósticos e intervenções desnecessárias, mas pode também comprometer a identificação de recém-nascidos em dor, o que pode ter sérias implicações para o manejo clínico. O *Score 2*, com sua alta sensibilidade (0,933) e especificidade satisfatória (0,792), parece atingir um bom ponto de equilíbrio entre esses dois fatores.

Com base nas análises regionais e globais, podemos concluir que os movimentos de sulco nasolabial aprofundado, boca aberta, e boca estirada são indicadores particularmente fortes de dor. Mesmo quando apenas um desses movimentos é pontuado, há uma alta probabilidade de que o diagnóstico de dor esteja correto, tornando-os componentes críticos no diagnóstico automatizado da dor neonatal.

Com relação aos testes feitos usando as bases de dados iCOPE e UNIFESP-II, observamos que, para a base de dados iCOPE, que contemplava imagens de neonatos com a face livre, a análise comparativa com o conjunto de validação UNIFESP-I revelou uma variação significativa no desempenho dos modelos de reconhecimento facial de dor neonatal. Esses resultados destacam a importância de testar modelos de aprendizado em diferentes conjuntos de dados para avaliar sua capacidade de generalização.

Assim, embora alguns movimentos faciais, como boca estirada e sulco nasolabial aprofundado, tenham mostrado um desempenho razoavelmente consistente entre os dois conjuntos de dados, o desempenho geral caiu significativamente no iCOPE, especialmente em termos de especificidade e concordância. Isso indica que o modelo tende a superestimar a presença de dor em situações em que ela não está presente.

É importante ressaltar que, nesta pesquisa, ao rotular as imagens do banco de dados UNIFESP-I como "com dor" ou "sem dor", ou ao identificar movimentos indicativos de dor, frequentemente nos deparamos com empates na classificação dos avaliadores. Nessas situações, sempre optamos por classificar como "indicativo de dor", o que provavelmente influenciou o modelo a superestimar a presença de dor em alguns casos em que ela não estava presente. Essa abordagem pode ter aumentado a sensibilidade do modelo para detectar automaticamente a dor, o que, por um lado, ajudou a identificar mais casos de dor, mas também elevou o risco de falsos diagnósticos.

Escolhemos essa estratégia com o objetivo de promover uma ação clínica mais cuidadosa, especialmente em situações em que os movimentos faciais podem ser sutis e gerar dúvidas sobre a presença ou ausência de dor. Nestes casos, sugerimos que o avaliador, na UTIN, revise seu olhar para o paciente e reconsidere a avaliação, para que não haja dúvidas quanto ao diagnóstico automático de dor neonatal.

Outras questões podem ter influenciado significativamente o desempenho do modelo, como o fato de o conjunto de dados iCOPE ser de uma amostra diferente, com qualidade de imagem distinta, com diferenças na iluminação e ângulos das imagens. A raça dos neonatos também pode ter sofrido influência sobre o resultado, já que os recém-nascidos do banco iCOPE são todos caucasianos e do banco UNIFESP-I são de raça miscigenada. O rótulo de dor nos bancos de dados foi atribuído por diferentes profissionais, o que gera uma variabilidade interobservador, e a subjetividade na avaliação dos sinais faciais de dor também pode impactar a acurácia do modelo. O número de imagens nos dois conjuntos de dados também pode influenciar os resultados, já que o conjunto iCOPE tinha um número maior de imagens em relação ao UNIFESP-I.

Mesmo com uma abordagem para mitigar esses problemas, incorporando técnicas de aumento de dados para adaptar o modelo às novas condições, seguimos com os mesmo desafios comuns em aprendizado de máquina, em que os modelos podem aprender bem um conjunto de dados, mas falhar quando testados em contextos novos e diferentes.

Com relação aos testes feitos usando a base de dado UNIFESP-II, em que algumas regiões faciais poderiam estar ocluídas por dispositivos de suporte à vida fixados à face do neonato, a análise por regiões da face indicou que o movimento de boca aberta apresentou a melhor acurácia e um Kappa de Cohen que indicava uma

concordância substancial. A boca estirada teve uma acurácia um pouco menor, um Kappa indicando uma concordância moderada e um F1 score razoável, com o desempenho inferior ao de boca aberta. A fenda palpebral estreitada teve o desempenho mais fraco, com Kappa indicando uma concordância leve, e o F1 score foi o mais baixo entre os movimentos. A frente saliente apresentou um Kappa indicando uma concordância moderada, com F1 score sugerindo um desempenho mediano. O sulco nasolabial aprofundado teve um Kappa indicando uma concordância leve, e um F1 score sugerindo um desempenho razoável, mas não tão robusto quanto outros movimentos.

Um ponto importante a ser considerado é o fato de o sulco nasolabial ser a região menos visível, no conjunto de dados UNIFESP-II, devido à fixação dos dispositivos médicos, o que pode ter influenciado os resultados do teste. Vale ressaltar que este foi o movimento que trouxe melhores resultados comparativos na fase de validação dos melhores modelos, e é um movimento destacado como importante para o diagnóstico de dor também em outros trabalhos^(55,56,114–116,133,134,136–138).

Quanto à proposta de um novo método de classificação e diagnóstico da dor, os resultados apontam que esta pode ser uma metodologia inovadora, com o potencial de modernizar o processo de avaliação da dor em neonatos. Heiderich *et al.* (2023)⁽⁵¹⁾ também sugerem o uso de imagens segmentadas como uma solução viável para melhorar o diagnóstico da dor, o que é consistente com nossa escolha metodológica, de trabalhar com máscaras de segmentação das regiões faciais mais discriminantes, e pontuar conforme a visibilidade de cada região.

O novo método foi desenvolvido com base nos resultados desta pesquisa e também se apoia em estudos prévios, como o de Silva (2020)⁽⁴⁶⁾, que destacou a importância das regiões do sulco nasolabial e da boca como altamente discriminatórias para o diagnóstico da dor, reforçando a importância de atribuir maior peso a essas regiões em ferramentas automatizadas de diagnóstico. Esse sistema permite que a equipe de saúde tome decisões rápidas e confiáveis, mesmo quando partes do rosto do paciente estão parcialmente ocluídas por dispositivos médicos.

Trabalhos Relacionados

Os resultados obtidos nesta pesquisa se relacionam diretamente com métodos manuais de avaliação comportamental para dor aguda amplamente utilizados na prática clínica neonatal. Escalas como a PIPP-R, NIPS, COMFORTneo, N-PASS, e a NFCS, mencionadas por estudos como os de Maxwell *et al.* (2019)⁽¹⁰⁴⁾ e McPherson *et al.* (2021)⁽¹⁰⁵⁾, são ferramentas fundamentais para o manejo da dor em neonatos. Essas escalas utilizam expressões faciais como um de seus principais indicadores, destacando características como sulco nasolabial aprofundado, boca aberta, fenda palpebral estreitada, e frente saliente, que também foram os movimentos mais relevantes para o modelo automatizado proposto nesta pesquisa.

Enquanto os métodos manuais são baseados na observação direta e subjetiva da equipe de saúde, o sistema automatizado desenvolvido neste estudo fornece uma abordagem objetiva e padronizada, mantendo a capacidade de identificar movimentos faciais importantes mesmo em condições de visibilidade parcial da face. Isso representa uma evolução significativa, uma vez que estudos como o de Giordano *et al.* (2019)⁽¹⁰⁶⁾ ressaltam a necessidade de ferramentas confiáveis e adaptáveis para populações neonatais, especialmente em cenários clínicos complexos, como aqueles envolvendo ventilação mecânica. Assim, o sistema proposto oferece uma alternativa inovadora e complementar às escalas manuais, alinhando-se aos objetivos de modernização e precisão no manejo da dor neonatal.

Os resultados desta pesquisa também dialogam diretamente com as descobertas de Vaart *et al.* (2019)⁽¹¹²⁾, reforçando a importância do sulco nasolabial aprofundado como um dos movimentos faciais mais relevantes para a discriminação de dor em neonatos. No estudo de Vaart *et al.* (2019)⁽¹¹²⁾, o sulco nasolabial foi identificado como a característica facial com maior precisão e acurácia entre os movimentos analisados, demonstrando ser a medida individual com maior poder discriminatório entre estímulos nocivos e não nocivos. Essa relevância também justifica a decisão de atribuímos um peso maior ao movimento do sulco nasolabial aprofundado no sistema automatizado de diagnóstico proposto, pois seu desempenho consistente o torna um indicador confiável na detecção de dor.

Outro ponto a ser considerado é que, assim como Vaart *et al.* (2019)⁽¹¹²⁾ destacaram que a combinação de modalidades nem sempre melhora a acurácia quando uma medida individual já possui alto poder discriminatório, nossos resultados

sugerem que o foco em regiões faciais específicas, como o sulco nasolabial, pode otimizar o diagnóstico de dor. Esse enfoque também reforça a necessidade de sistemas automatizados priorizarem características faciais altamente discriminatórias, como o sulco nasolabial, para aumentar a eficiência e a precisão na avaliação da dor neonatal, mesmo em condições de visibilidade parcial da face.

Os resultados de Barros *et al.* (2021)⁽¹¹⁴⁾ também destacam a importância do sulco nasolabial como uma região facial crítica para a avaliação da dor neonatal, corroborando a abordagem adotada neste trabalho. No estudo, a atenção visual ao sulco nasolabial foi identificada como um fator determinante para a avaliação precisa da dor, especialmente entre profissionais de saúde. Esse achado complementa nossa proposta de um sistema automatizado que integra a análise do sulco nasolabial como uma das principais áreas de interesse. A inclusão dessa região como um ponto central no sistema automatizado proposto não só valida sua importância clínica, mas também oferece uma alternativa moderna e objetiva para complementar a avaliação manual realizada por profissionais de saúde.

Os resultados apresentados por Silva *et al.* (2020)⁽⁴⁶⁾ também corroboram diretamente a decisão tomada neste trabalho de atribuir pesos diferentes às regiões faciais no sistema automatizado de diagnóstico de dor. A análise estatística multivariada conduzida por Silva *et al.* (2020)⁽⁴⁶⁾ revelou que o entorno da boca, incluindo o sulco nasolabial, as sobrancelhas e porções das bochechas, são as áreas faciais mais destacadas durante uma experiência dolorosa em neonatos. Essas regiões foram identificadas como os principais pontos de discriminação entre estados de dor e ausência de dor, evidenciando sua relevância para sistemas de avaliação automática e para o julgamento clínico de profissionais de saúde.

Com base nessas evidências, o sistema aqui proposto atribui peso maior ao movimento do sulco nasolabial aprofundado, que é uma das regiões mais frequentemente destacadas nos mapas estatísticos gerados por Silva *et al.* (2020)⁽⁴⁶⁾. Movimentos da boca aberta e boca estirada, que também pertencem à área ao redor da boca, recebem peso intermediário, enquanto as sobrancelhas (representadas no sistema pelos movimentos de frente saliente) e outras áreas faciais com menor relevância estatística recebem peso menor. Essa abordagem reflete diretamente a sugestão de Silva, de que áreas mais discriminantes devem ter maior relevância em algoritmos e escalas de dor.

O estudo de Heiderich *et al.* (2021)⁽⁵¹⁾ destacou uma das principais dificuldades em sistemas automatizados de avaliação facial: a presença de dispositivos médicos fixados à face do recém-nascido. Os resultados apresentados neste trabalho abordam diretamente essa dificuldade, propondo um sistema automatizado que avalia apenas as regiões visíveis da face, adaptando-se à realidade dos neonatos internados em UTIN. Diferentemente do método utilizado por Heiderich *et al.* (2021)⁽⁵¹⁾, que dependia de uma imagem de face neutra como ponto de comparação, o sistema aqui proposto não exige referência prévia e é capaz de atribuir pesos distintos aos movimentos faciais mais relevantes. Além disso, nossa proposta alinha-se às recomendações de Heiderich, ao sugerir que apenas movimentos faciais visíveis sejam avaliados. Essa estratégia evita a inclusão de dados irrelevantes ou confusos, como movimentos não relacionados à dor. Assim, o sistema desenvolvido não apenas reconhece as limitações apontadas por Heiderich, mas também apresenta uma solução prática ao considerar a realidade clínica de neonatos em UTIN. Avaliar apenas as regiões visíveis, atribuindo maior peso a movimentos confiáveis, garante que o sistema seja eficiente, reduzindo erros associados à oclusão facial e promovendo uma aplicação mais robusta em contextos hospitalares.

Até o momento, não foram identificados trabalhos na literatura que abordem a avaliação automática de partes específicas da face do recém-nascido para o diagnóstico de dor, considerando apenas as regiões visíveis ou atribuindo diferentes pontuações aos movimentos faciais com maior poder discriminatório. Os estudos existentes geralmente avaliam a face como um todo, sem distinção entre as áreas mais relevantes ou as condições de visibilidade parcial, como ocorre frequentemente em neonatos internados em UTIN. Além disso, não há registros de propostas que utilizem pesos diferenciados para priorizar os movimentos faciais mais confiáveis, como o sulco nasolabial aprofundado, ou que adaptem a avaliação à realidade clínica da presença de dispositivos médicos que ocluem parcialmente a face. Essa lacuna evidencia a originalidade e a inovação do sistema proposto neste trabalho, que combina essas abordagens para oferecer uma solução mais precisa, robusta e adaptada à prática clínica.

Limitações

As bases de dados utilizadas apresentaram um número limitado de imagens faciais de neonatos, e os conjuntos de treinamento e validação, embora priorizassem a seleção de neonatos diferentes em cada conjunto, não sofreram mais de uma atualização cruzada, podendo impactar a generalização dos resultados. Estudos com mais conjuntos de treinamento e validação, dados maiores e mais diversificados são necessários para validar a robustez do modelo em diferentes cenários de otimização.

Outra limitação importante foi a segmentação facial em imagens com oclusão parcial, causada por dispositivos de suporte à vida fixados à face do neonato. Embora nosso arcabouço tenha conseguido segmentar muitas dessas imagens, algumas delas, parcialmente obstruídas, precisaram ser segmentadas manualmente, tornando o processo semiautomático. Isso ressalta a necessidade de, em estudos futuros, serem desenvolvidos métodos totalmente automatizados, capazes de detectar todas as regiões faciais, inclusive a presença de dispositivos.

Outra limitação foi a natureza estática das imagens analisadas. Movimentos faciais dinâmicos, como o tremor do queixo ou o batimento das asas do nariz, não puderam ser capturados, pois não são visíveis em imagens 2D estáticas. Estudos futuros devem explorar a análise de expressões faciais em vídeos, o que permitiria capturar movimentos mais sutis e melhorar a detecção de dor em tempo real.

Este estudo concentrou-se exclusivamente na análise de imagens faciais. Pesquisas futuras podem integrar outros sinais fisiológicos importantes, como frequência cardíaca e saturação de oxigênio, com o objetivo de fornecer um diagnóstico mais completo da dor. Isso seria particularmente útil em situações em que o rosto dos neonatos está significativamente obstruído, complementando a avaliação facial com dados fisiológicos para melhorar a acurácia na detecção da dor.

Adicionalmente, a validação do modelo foi limitada a um ambiente experimental, e sua aplicação prática em ambientes clínicos reais ainda precisa ser testada para verificar sua eficácia e robustez na detecção de dor em condições mais complexas e dinâmicas. Isso reforça a importância de continuar explorando e ajustando o arcabouço em situações do mundo real, garantindo que ele atenda às necessidades clínicas com precisão e confiabilidade.

Pontos Fortes

O estudo apresentou várias inovações importantes, como o uso do *Vision Transformer* (ViT), que se mostrou altamente eficaz na detecção de dor em neonatos, atingindo F1 scores elevados para movimentos faciais críticos como sulco nasolabial e boca aberta.

O uso de máscaras de segmentação facial, aliado à identificação automática dos pontos fiduciais com o MediaPipe, foi outro ponto forte, permitindo a delimitação precisa das áreas faciais de interesse e reduzindo os erros de detecção, e um refinamento na detecção de dor.

Esses avanços técnicos tornam o arcabouço proposto uma alternativa promissora aos procedimentos tradicionais de avaliação da dor, como as escalas multidimensionais manuais, já que, além de ser objetivo e replicável, o modelo apresentou concordância substancial com os profissionais de saúde, evidenciada pelos altos valores de Kappa de Cohen, evidenciando a confiabilidade do modelo na prática clínica.

A abordagem automatizada, com a capacidade de replicar as observações manuais dos profissionais, pode representar um grande avanço no manejo da dor neonatal em ambientes hospitalares.

Implicações Práticas

Os achados sugerem que o desenvolvimento de uma ferramenta automatizada para a avaliação de dor neonatal é uma solução promissora, especialmente em situações em que a face do neonato está parcialmente obstruída.

Uma das principais contribuições deste estudo foi o desenvolvimento de um método de classificação e diagnóstico de dor por regiões visíveis da face. Esse método atribui diferentes pesos aos movimentos faciais, com base em sua relevância para o diagnóstico de dor, permitindo uma avaliação mais precisa e sensível, mesmo em situações de oclusão parcial da face. Com essa abordagem, o diagnóstico é determinado pela presença de movimentos faciais mais confiáveis, como o sulco nasolabial e a boca, enquanto movimentos com desempenho inferior, como a fenda palpebral estreitada e a fronte saliente, têm menor peso no diagnóstico final.

Teoricamente, este estudo contribui para o campo da avaliação automatizada da dor ao destacar a importância de uma abordagem segmentada na análise facial. Enquanto muitos métodos tradicionais focam na face como um todo, nossos achados indicam que uma abordagem baseada em regiões faciais específicas pode ser mais factível. Isso abre novas perspectivas para o desenvolvimento de ferramentas mais precisas e personalizadas de reconhecimento de dor.

Na prática clínica, reforçamos que os profissionais de saúde considerem que algumas regiões faciais são mais discriminatórias para o diagnóstico de dor. Assim, seria benéfico ajustar a posição dos dispositivos médicos de forma a manter visíveis áreas importantes, como o sulco nasolabial e a boca. O uso de esparadrapos com cores menos contrastantes com a pele ou materiais mais finos também poderiam melhorar a visibilidade dessas regiões, facilitando a avaliação dos sinais faciais de dor. Além disso, o reposicionamento periódico dos dispositivos, sempre que clinicamente possível, ajudaria a garantir que as áreas cruciais para a avaliação facial permaneçam acessíveis.

Trabalhos Futuros

Para pesquisas futuras, seria importante desenvolver métodos automatizados para detectar e lidar com dispositivos médicos que obstruem a face, bem como aprimorar a análise de imagens dinâmicas para captar melhor os movimentos faciais em tempo real. A criação de um banco de dados sintético mais abrangente, incluindo variações de iluminação e movimento, também seria um avanço significativo para validar o modelo em ambientes clínicos reais.

Outra abordagem importante seria a integração do método de avaliação automática da dor desenvolvido nesta pesquisa com a análise contínua de outros parâmetros fisiológicos, como frequência cardíaca, frequência respiratória e saturação de oxigênio, permitindo um monitoramento abrangente e em tempo real à beira do leito. Essa abordagem multimodal pode ampliar a precisão do diagnóstico de dor, considerando tanto indicadores comportamentais quanto fisiológicos, e auxiliar no manejo clínico de neonatos internados em UTIN. Além disso, a incorporação desses dados em um sistema automatizado pode proporcionar um alerta precoce para

mudanças no estado do paciente, otimizando as intervenções médicas e melhorando a qualidade do cuidado neonatal.

Outro aspecto relevante seria a realização de estudos comparativos que investiguem o foco dos profissionais de saúde em regiões faciais específicas, como sugerido por Tamanaka *et al.* (2022)⁽¹¹⁶⁾. Isso ajudaria a avaliar se a atenção a áreas como o nariz e os olhos influencia o diagnóstico de dor e se o treinamento direcionado para regiões mais discriminantes, como o sulco nasolabial e a boca, poderia melhorar a precisão diagnóstica.

Contribuições para o Estado da Arte, Sociais e Éticas

As contribuições deste trabalho avançam o estado da arte no reconhecimento automatizado da dor neonatal, introduzindo inovações significativas em várias frentes. Utilizamos a inteligência artificial para desenvolver uma abordagem baseada no Vision Transformer, com uma modificação arquitetural importante: a substituição do head de classificação original por uma rede MLP com duas saídas, configurada para realizar a classificação binária de dor sem a necessidade de treinamento prévio. Essa modificação permitiu adaptar o modelo à nova tarefa de forma eficiente e direcionada. Além disso, propomos a segmentação da face em regiões específicas, diferenciando nossa abordagem das tradicionais que avaliam a face como um todo, priorizando a análise das regiões visíveis, uma solução prática para cenários onde dispositivos médicos obstruem parte da face. Por fim, introduzimos um novo sistema de pontuação que atribui pesos diferenciados aos movimentos faciais mais discriminantes, como o sulco nasolabial aprofundado, otimizando a precisão e sensibilidade do diagnóstico. Essas contribuições não apenas ampliam as possibilidades de aplicação da IA em contextos clínicos, mas também apresentam um avanço significativo na avaliação objetiva e automatizada da dor neonatal.

Este estudo marca um avanço significativo na aplicação de técnicas automatizadas para a avaliação da dor neonatal. A segmentação facial e o uso do ViT para identificar movimentos faciais indicativos de dor demonstraram ser estratégias robustas, mesmo em situações desafiadoras, como a oclusão parcial da face. Apesar de limitações inerentes, como a oclusão parcial da face e a natureza estática das imagens, as contribuições teóricas e práticas deste trabalho oferecem um caminho

promissor para o desenvolvimento de ferramentas clínicas automatizadas, capazes de transformar o manejo da dor neonatal em um processo mais ágil, acessível e preciso.

No contexto da saúde neonatal, a automação na detecção da dor pode trazer benefícios significativos, como diagnósticos mais rápidos e eficazes, aliviando o sofrimento dos neonatos e melhorando a qualidade do cuidado. Entretanto, desafios importantes precisam ser considerados, como as questões éticas relacionadas à privacidade das imagens e à interpretação automatizada de sinais clínicos. Além disso, a aceitação clínica de sistemas baseados em IA requer treinamentos adequados para profissionais de saúde, garantindo que saibam complementar e interpretar os resultados fornecidos por essas ferramentas.

7 CONCLUSÃO

Esta tese apresentou um arcabouço computacional inovador para o reconhecimento automatizado de dor em recém-nascidos internados em Unidade de Terapia Intensiva Neonatal (UTIN), focando na análise de regiões faciais mesmo quando parcialmente ocluídas por dispositivos médicos.

Os resultados indicam que regiões faciais específicas, como o sulco nasolabial aprofundado e a boca aberta, desempenham papéis críticos no diagnóstico de dor neonatal, apresentando as melhores taxas de concordância com os profissionais de saúde. Por outro lado, movimentos como a fenda palpebral estreitada e a fronte saliente apresentaram um desempenho inferior, com maior taxa de falsos positivos e menor concordância, sugerindo que esses movimentos são mais desafiadores para a análise automatizada.

O objetivo principal foi alcançado com o desenvolvimento de uma metodologia capaz de segmentar e analisar automaticamente partes específicas da face do recém-nascido, considerando apenas as regiões visíveis e atribuindo diferentes pontuações aos movimentos faciais com maior poder discriminatório para o diagnóstico de dor.

O sucesso do modelo em diferentes bases de dados reforça a viabilidade de sua aplicação clínica, sendo capaz de fornecer diagnósticos precisos, mesmo em condições desafiadoras, como a obstrução parcial da face do neonato. O arcabouço também abre novas possibilidades para aprimorar o manejo da dor neonatal em UTINs, contribuindo para diagnósticos mais rápidos e acurados, além de oferecer uma ferramenta confiável para profissionais de saúde.

Embora o modelo tenha se mostrado robusto, o estudo destacou limitações, como a necessidade de aprimorar a segmentação automática em casos de oclusão facial e de desenvolver métodos que capturem movimentos faciais dinâmicos em vídeos. Estudos futuros podem integrar outros sinais fisiológicos, como frequência cardíaca e saturação de oxigênio, para um diagnóstico ainda mais preciso, especialmente em situações em que a face está significativamente obstruída.

Esta pesquisa representa um avanço significativo no campo da avaliação automatizada de dor neonatal, propondo um arcabouço inovador e um novo método automático de classificação e diagnóstico da dor, que combina acurácia, precisão e aplicabilidade clínica, com o potencial de transformar o diagnóstico da dor em UTINs.

REFERÊNCIAS

- 1 ABELHA, F. J. et al. Mortalidade e o Tempo de Internação em uma Unidade de Terapia Intensiva Cirúrgica. **Rev Bras Anesthesiol**, v. 56, n. 1, p. 34–45, 2006.
- 2 MACEDO, J. S.; MÜLLER, A. B. Dor e medidas não-farmacológicas em prematuros hospitalizados. **Revista Saúde - UNG-Ser**, v. 15, n. 1/2, p. 23, 2021.
- 3 BONUTTI, D. P. et al. Dimensioning of painful procedures and interventions for acute pain relief in premature infants. **Revista Latino-Americana de Enfermagem**, v. 25, n. e2917, p. 1–9, 2017.
- 4 SANTIAGO, A. D. et al. Neonatal Morbidity and Mortality in Intensive Care Unit. **Tempus**, v. 11, n. 1, p. 141–151, 2017.
- 5 SMITH, G. C. et al. Neonatal Intensive Care Unit Stress Is Associated with Brain Development in Preterm Infants. **Ann Neurol**, v. 70, p. 541–549, 2011.
- 6 BRUMMELTE, S. et al. Procedural Pain and Brain Development in Premature Newborns. **American Neurological Association**, v. 71, n. 3, p. 385–96, 2012.
- 7 GUINSBURG, R. Assessing and treating pain in the newborn. **Jornal de Pediatria**, v. 75, n. 3, p. 149–60, 15 maio 1999.
- 8 BALDA, R. DE C. X.; GUINSBURG, R. Avaliação da dor no período neonatal. *In*: KOPELMAN I.B. **Diagnóstico e tratamento em neonatologia**. São Paulo: Atheneu, 2004.
- 9 DOESBURG, S. M. et al. Neonatal pain-related stress, functional cortical activity and visual-perceptual abilities in school-age children born at extremely low gestational age. **Pain**, v. 154, n. 10, p. 1946–1952, 2013.
- 10 WILLIAMS, M. D.; LASCELLES, B. D. X. Early Neonatal Pain—A Review of Clinical and Experimental Implications on Painful Conditions Later in Life. **Frontiers in Pediatrics**, v. 8, n. February, p. 1–18, 2020.
- 11 QUARESMA, M. E. et al. Factors associated with hospitalization during neonatal period. **Jornal de Pediatria**, v. 94, n. 4, p. 390–398, 2018.
- 12 LEAPE, L. L. Reporting of adverse events. **The new england journal of medicine**, v. 347, n. 20, p. 1633–1638, 2002.
- 13 VENTURA, C. M. U.; ALVES, J. G. B.; MENESES, J. DO A. Eventos adversos em Unidade de Terapia Intensiva Neonatal. **Revista Brasileira de Enfermagem**, v. 65, n. 1, p. 49–55, 2012.
- 14 SPIRONELLO, R. A.; CUMAN, R. K. N. Caracterização de eventos adversos em uma unidade de terapia intensiva neonatal. **Revista Recien - Revista Científica de Enfermagem**, v. 9, n. 28, p. 131–136, 2019.

- 15 CARVALHO, M. DE; VIEIRA, A. A. Erro médico em pacientes hospitalizados. **Jornal de Pediatria**, v. 78, n. 4, p. 261–268, 2002.
- 16 ANDREAZZA, M. G. et al. Percepção da dor em neonatos pela equipe de enfermagem de unidade de terapia intensiva neonatal. **Brazilian Journal of Health Research**, v. 19, out. 2017.
- 17 PERRY, M. et al. Neonatal Pain: Perceptions and Current Practice. **Critical care nursing clinics of North America**, v. 30, n. 4, p. 549–561, 2018.
- 18 BRASIL. OBM BRASIL. **Os Objetivos de Desenvolvimento do Milênio**. Disponível em: <http://www.odmbrasil.gov.br/os-objetivos-de-desenvolvimento-do-milenio>. Acesso em: 29 mar. 2022.
- 19 BRASIL. NAÇÕES UNIDAS. **Os Objetivos do Desenvolvimento Sustentável**. Disponível em: <https://brasil.un.org/pt-br/sdgs>. Acesso em: 29 mar. 2022.
- 20 FRANÇA, E. B. et al. Leading causes of child mortality in Brazil, in 1990 and 2015: Estimates from the Global Burden of Disease study. **Revista Brasileira de Epidemiologia**, v. 20, n. 25000192049, p. 46–60, 2017.
- 21 LEAL, M. D. C. et al. Reproductive, maternal, neonatal and child health in the 30 years since the creation of the Unified Health System (SUS). **Ciencia e Saude Coletiva**, v. 23, n. 6, p. 1915–1928, 2018.
- 22 BRASIL. NAÇÕES UNIDAS. **Relatório sobre Mortalidade Infantil 2023**. Disponível em: <https://brasil.un.org/pt-br/263674-relat%C3%B3rio-sobre-mortalidade-infantil-2023>. Acesso em: 16 out. 2024.
- 23 BERNARDINO, F. B. S. et al. Tendências da mortalidade neonatal no Brasil de 2007 a 2017. **Ciencia e Saude Coletiva**, v. 27, n. 2, p. 567–578, 2022.
- 24 BRASIL. MINISTÉRIO DA SAÚDE. **Mortalidade infantil e fetal por causas evitáveis no Brasil é a menor em 28 anos**. Disponível em: <https://www.gov.br/saude/pt-br/assuntos/noticias/2024/marco/mortalidade-infantil-e-fetal-por-causas-evitaveis-no-brasil-e-a-menor-em-28-anos#:~:text=O%20Brasil%20registrou%20em%202023,uma%20s%C3%A9rie%20hist%C3%B3rica%20desde%201996>. Acesso em: 16 out. 2024.
- 25 PREZOTTO, K. H. et al. Trend of preventable neonatal mortality in the States of Brazil. **Revista Brasileira de Saúde Materno Infantil**, v. 21, n. 1, p. 291–299, mar. 2021.
- 26 BRASIL. MINISTÉRIO DA SAÚDE. **Saúde Brasil 2017: Uma análise da situação de saúde e os desafios para o alcance dos objetivos de desenvolvimento sustentável**. Brasília - DF: Ministério da Saúde, Secretaria de Vigilância em Saúde, Departamento de Vigilância de Doenças e Agravos não Transmissíveis e Promoção da Saúde, 2018.

- 27 COSTA, M. DE F. DOS S.; MAGLUTA, C.; GOMES JUNIOR, S. C. DOS S. Profile of service providers in high-risk neonatal care in the databases of the Brazilian Unified National Health System. **Cadernos de Saúde Pública**, v. 36, n. 4, p. 1–13, 2020.
- 28 BRASIL. MINISTÉRIO DA SAÚDE. **Portaria n. 930, de 10 de maio de 2012**. Brasil, 2012. Disponível em: <https://www.as.saude.ms.gov.br/wp-content/uploads/2016/07/Portaria-GM-MS-n%C2%BA1.020-do-dia-29-de-maio-de-2013.pdf>. Acesso em: 16 out. 2024.
- 29 BRASIL. MINISTÉRIO DA SAÚDE. **Estratégia QualiNEO**. Disponível em: <https://www.gov.br/saude/pt-br/assuntos/saude-de-a-a-z/s/saude-da-crianca/cuidado-neonatal/qualineo>. Acesso em: 16 out. 2024.
- 30 BRASIL. MINISTÉRIO DA SAÚDE. **Portaria GM/MS nº 889, de 17 de julho de 2023**. Disponível em: [https://bvsmms.saude.gov.br/bvs/saudelegis/gm/2023/prt0889_18_07_2023.html#:~:text=a%20esta%20Portaria.,Art.,\(oitava\)%20parcela%20de%202023](https://bvsmms.saude.gov.br/bvs/saudelegis/gm/2023/prt0889_18_07_2023.html#:~:text=a%20esta%20Portaria.,Art.,(oitava)%20parcela%20de%202023). Acesso em: 16 out. 2024.
- 31 FUNDAÇÃO OSWALDO CRUZ. INSTITUTO NACIONAL DE SAÚDE DA MULHER, DA C. E DO A. F. F. **Dor em recém-nascidos: como avaliar, prevenir e tratar**. Disponível em: <https://portaldeboaspraticas.iff.fiocruz.br/atencao-recem-nascido/manejo-da-dor-e-do-estresse/>. Acesso em: 16 out. 2024.
- 32 GUINSBURG, R.; CUENCA A, M. C. C. A linguagem da dor no recém-nascido. **Documento Científico do Departamento de Neonatologia Sociedade Brasileira de Pediatria**, v. 3, n. 9, p. 19–24, 2019.
- 33 HEIDERICH, T. M. **Desenvolvimento de software para identificar a expressão facial de dor do recém-nascido**. 2013. Tese (Doutorado em Pediatria e Ciências Aplicadas à Pediatria). Universidade Federal de São Paulo, São Paulo, 2013. Disponível em: <https://repositorio.unifesp.br/server/api/core/bitstreams/88284b0c-d4c7-4ead-8790-da163f3c4b02/content>. Acesso em: 5 out. 2019.
- 34 CARLSEN MISIC, M. et al. Nurses' perception, knowledge, and use of neonatal pain assessment. **Paediatric and Neonatal Pain**, v. 3, n. 2, p. 59–65, 2021.
- 35 CHERMONT, A. G. et al. O que os pediatras conhecem sobre avaliação e tratamento da dor no recém-nascido? **Jornal de Pediatria**, v. 79, n. 3, p. 265–72, 2003.
- 36 PRESTES, A. et al. The frequency of pharmacological pain relief in university neonatal intensive care units Ana. **Jornal de pediatria**, v. 81, p. 405–410, 2005.
- 37 BRASIL. MINISTÉRIO DA SAÚDE. **Manual de vigilância do óbito infantil e fetal e do Comitê de Prevenção do Óbito Infantil e Fetal**. 2ª ed. Brasília - DF: Secretaria de Vigilância em Saúde, Departamento de Análise de Situação em Saúde, 2009.
- 38 HEIDERICH, T. M.; LESLIE, A. T. F. S.; GUINSBURG, R. Neonatal procedural pain can be assessed by computer software that has good sensitivity and specificity to detect facial movements. **Acta Paediatrica**, v. 104, n. 2, p. e63–e69, fev. 2015.

- 39 ZAMZMI, G. et al. An approach for automated multimodal analysis of infants' pain. *In: 23rd INTERNATIONAL CONFERENCE ON PATTERN RECOGNITION (ICPR)*, 2016. **Anais eletrônicos...** IEEE, 2016. Disponível em: <http://ieeexplore.ieee.org/document/7900284/>. Acesso em: 5 out. 2019.
- 40 TERUEL, G. F. et al. Analysis And Recognition Of Pain In 2d Face Images Of Full Term And Healthy Newborns. *In: XV ENCONTRO NACIONAL DE INTELIGÊNCIA ARTIFICIAL E COMPUTACIONAL (ENIAC 2018)*, 2018. **Anais eletrônicos...** Sociedade Brasileira de Computação – SBC, 2018. Disponível em: <http://portaldeconteudo.sbc.org.br/index.php/eniac/article/view/4419>. Acesso em: 5 out. 2019.
- 41 ZHI, R. et al. Automatic Infants' Pain Assessment by Dynamic Facial Representation: Effects of Profile View, Gestational Age, Gender, and Race. **Journal of Clinical Medicine**, v. 7, n. 7, p. 173, 11 jul. 2018.
- 42 BRAHNAM, S. et al. Neonatal pain detection in videos using the iCOPEvid dataset and an ensemble of descriptors extracted from Gaussian of Local Descriptors. **Applied Computing and Informatics**, p. 1–22, 17 jul. 2020.
- 43 ORONA, P. A. S. S. et al. Atlas of neonatal face images using triangular Meshes. *In: ANAIS DO XV WORKSHOP DE VISÃO COMPUTACIONAL (WVC 2019)*, 2019. **Anais eletrônicos...** Sociedade Brasileira de Computação - SBC, 2019. Disponível em: <https://sol.sbc.org.br/index.php/wvc/article/view/7622>. Acesso em: 5 out. 2019.
- 44 ZAMZMI, G. et al. Convolutional Neural Networks for Neonatal Pain Assessment. **IEEE Transactions on Biometrics, Behavior, and Identity Science**, v. 1, n. 3, p. 192–200, jul. 2019.
- 45 SUN, Y. et al. Video-based discomfort detection for infants. **Machine Vision and Applications**, v. 30, n. 5, p. 933–944, 13 jul. 2019.
- 46 SILVA, P. A. S. O. **Interpretação e reconhecimento de padrões para avaliação de dor em imagens faciais de recém-nascidos**. 2020. 103 f. Dissertação (Mestrado em Engenharia Elétrica) - Centro Universitário FEI, São Bernardo do Campo, 2020. Disponível em: <https://repositorio.fei.edu.br/bitstreams/41f27d34-6ed6-42da-8cc2-fa547e6d2cd2/download>. Acesso em: 18 set. 2024.
- 47 SALEKIN, M. S. et al. Future roles of artificial intelligence in early pain management of newborns. **Paediatric and Neonatal Pain**, v. 3, n. 3, p. 134–145, 5 set. 2021.
- 48 HOTI, K.; CHIVERS, P. T.; HUGHES, J. D. Assessing procedural pain in infants: a feasibility study evaluating a point-of-care mobile solution based on automated facial analysis. **The Lancet Digital Health**, v. 3, n. 10, p. e623–e634, out. 2021.
- 49 ZAMZMI, G. et al. A Comprehensive and Context-Sensitive Neonatal Pain Assessment Using Computer Vision. **IEEE Transactions on Affective Computing**, v. 13, n. 1, p. 28–45, 1 jan. 2022.

- 50 BUZUTI, L. et al. Neonatal Pain Assessment From Facial Expression Using Deep Neural Networks. *In: ANAIS DO XVI WORKSHOP DE VISÃO COMPUTACIONAL (WVC 2020)*, 2020. **Anais eletrônicos...** Sociedade Brasileira de Computação - SBC, 2020. Disponível em: <https://sol.sbc.org.br/index.php/wvc/article/view/13486>. Acesso em: 18 set. 2024.
- 51 HEIDERICH, T. M. et al. Face-based automatic pain assessment: challenges and perspectives in neonatal intensive care units. **Jornal de Pediatria**, v. 99, n. 6, p. 546–560, 1 nov. 2023.
- 52 BATTON, D.; BARRINGTON, K.; WALLMAN, C. Prevention and Management of Pain in the Neonate: An Update. **Pediatrics**, v. 118, n. 5, p. 2231–2241, 1 nov. 2006.
- 53 RAUPP, A. J. F.; SCHERER, C. M.; KRAUSE, G. C. Avaliação e manejo da dor em unidade de terapia intensiva neonatal: Perspectiva da equipe de enfermagem. **Research, Society and Development**, v. 12, n. 11, p. e28121143566, 22 out. 2023.
- 54 ERIKSSON, M.; CAMPBELL-YEO, M. Assessment of pain in newborn infants. **Seminars in Fetal and Neonatal Medicine**, v. 24, n. 4, p. 101003, ago. 2019.
- 55 DOMINGUES, P. H. S. et al. Neonatal Face Mosaic: An areas-of-interest segmentation method based on 2D face images. *In: XVII WORKSHOP DE VISÃO COMPUTACIONAL (WVC 2021)*, 2021. **Anais...** Sociedade Brasileira de Computação - SBC, 2021. Disponível em: <https://sol.sbc.org.br/index.php/wvc/article/view/18914>. Acesso em: 18 set. 2024.
- 56 CARLINI, L. P. et al. Human vs machine towards neonatal pain assessment: A comprehensive analysis of the facial features extracted by health professionals, parents, and convolutional neural networks. **Artificial Intelligence in Medicine**, v. 147, p. 102724, 1 jan. 2024.
- 57 FONTAINE, D. et al. Artificial intelligence to evaluate postoperative pain based on facial expression recognition. **European Journal of Pain**, v. 26, n. 6, p. 1282–1291, 6 jul. 2022.
- 58 OTHMAN, E. et al. Automatic vs. Human Recognition of Pain Intensity from Facial Expression on the X-ITE Pain Database. **Sensors**, v. 21, n. 9, p. 3273, 10 maio 2021.
- 59 GRUNAU, R. V. E.; CRAIG, K. D. Pain expression in neonates: facial action and cry. **Pain**, v. 28, n. 3, p. 395–410, mar. 1987.
- 60 SOUTO, S. A Dor no Recém-nascido: O desafio da avaliação. **Rev Nurs**, p. 6–12, 2008.
- 61 SÉ, A. C. S. et al. Adesão de protocolo assistencial sobre fixação de cateter nasoenteral. **Brazilian Journal of Enterostomal Therapy**, v. 20, n. 2022, p. 1–10, 2022.

- 62 BRASIL - RESOLUÇÃO COFEN Nº 0619/2019. **Normas para atuação da equipe de enfermagem na sondagem oro/nasogástrica e nasoentérica**. Brasília, 2019.
- 63 SUBRAMANIAM, P.; HO, J. J.; DAVIS, P. G. Prophylactic nasal continuous positive airway pressure for preventing morbidity and mortality in very preterm infants. **Cochrane Database of Systematic Reviews**, v. 2016, n. 6, 14 jun. 2016.
- 64 WYLLIE, J. P. Neonatal endotracheal intubation. **Archives of Disease in Childhood: Education and Practice Edition**, v. 93, n. 2, p. 44–49, abr. 2008.
- 65 FIOCRUZ - PORTAL DE BOAS PRÁTICAS. **Principais Questões sobre Prevenção de Extubação Acidental no Recém-nascido em Ventilação Mecânica**. Disponível em: <https://portaldeboaspraticas.iff.fiocruz.br/atencao-recem-nascido/principais-questoes-sobre-prevencao-de-extubacao-acidental-no-recem-nascido-em-ventilacao-mecanica/>. Acesso em: 18 set. 2024.
- 66 SILVA, É. H. A. DA et al. Cuidados de enfermagem com a fototerapia em recém-nascidos com icterícia. **Rev Bras Interdiscip Saúde - ReBIS**, v. 3, n. 4, p. 49–57, 2021.
- 67 YAN, J. et al. FENP: A Database of Neonatal Facial Expression for Pain Analysis. **IEEE Transactions on Affective Computing**, v. 14, n. 1, p. 245–254, 1 jan. 2023.
- 68 LUCEY, P. et al. Painful data: The UNBC-McMaster shoulder pain expression archive database. *Face and Gesture*, 2011. In: **Anais eletrônicos...IEEE**, 2011. Disponível em: <http://ieeexplore.ieee.org/document/5771462/>. Acesso em: 18 set. 2024.
- 69 BRAHNAM, S. et al. Machine recognition and representation of neonatal facial displays of acute pain. **Artificial Intelligence in Medicine**, v. 36, n. 3, p. 211–222, mar. 2006.
- 70 EGEDÉ, J. et al. Automatic Neonatal Pain Estimation: An Acute Pain in Neonates Database. In: 8TH INTERNATIONAL CONFERENCE ON AFFECTIVE COMPUTING AND INTELLIGENT INTERACTION (ACII), 2019. **Anais eletrônicos... IEEE**, 2019. Disponível em: <https://ieeexplore.ieee.org/document/8925480/>. Acesso em: 18 set. 2024.
- 71 HARRISON, D. et al. Too many crying babies: A systematic review of pain management practices during immunizations on YouTube. **BMC Pediatrics**, v. 14, n. 1, 29 maio 2014.
- 72 SALEKIN, M. S. et al. Multimodal neonatal procedural and postoperative pain assessment dataset. **Data in Brief**, v. 35, p. 106796, 1 abr. 2021.
- 73 HUANG, B. et al. A neonatal dataset and benchmark for non-contact neonatal heart rate monitoring based on spatio-temporal neural networks. **Engineering Applications of Artificial Intelligence**, v. 106, p. 104447, 1 nov. 2021.

- 74 BUZUTI, L. et al. **Generative AI for Neonatal Pain Assessment: A Sound Approach to Improve Data-Driven Automatic Recognition in Intensive Care Unit.** , 2024. Disponível em: <https://www.ssrn.com/abstract=4994076>
- 75 VAHDAT, A.; KAUTZ, J. NVAE: A Deep Hierarchical Variational Autoencoder. *In: 34th CONFERENCE ON NEURAL INFORMATION PROCESSING SYSTEMS, 2021, Canada (Vancouver).* **Anais eletrônicos...** Canada, 2021. Disponível em: <http://arxiv.org/abs/2007.03898>. Acesso em: 01 ago. 2024.
- 76 HJELMÅS, E.; LOW, B. K. Face detection: A survey. **Computer Vision and Image Understanding**, v. 83, n. 3, p. 236–274, 2001.
- 77 TIAN et al. Recognizing LowerFace Action Units for Facial Expression Analysis. **Proceedings of the 4th IEEE International Conference on Automatic Face and Gesture Recognition (FG'00)**, v. 23, n. 2, p. 484–490, 2000.
- 78 EKMAN, P.; FRIESEN, W. V. **Facial Action Coding System: A Technique for the Measurement of Facial Movement.** [s.l.] Consulting Psychologists Press, 1978.
- 79 CHEN, Z. **FACS-Based Automated Pain Detection From Spontaneous Facial Expressions.** Philosophy in Electrical and Computer Engineering—Chicago, Illinois: University of Illinois at Chicago, 2020.
- 80 WILKIE, D. J. Facial Expressions of Pain in Lung Cancer. **Analgesia**, v. 1, n. 2, p. 91–99, 1 jan. 1995.
- 81 DENG, J. et al. Retinaface: Single-shot multi-level face localisation in the wild. **Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition**, p. 5202–5211, 2020.
- 82 SAGONAS, C. et al. 300 Faces In-The-Wild Challenge: database and results. **Image and Vision Computing**, v. 47, p. 3–18, mar. 2016.
- 83 LIU, Y. et al. Grand challenge of 106-point facial landmark localization. **Proceedings - 2019 IEEE International Conference on Multimedia and Expo Workshops, ICMEW 2019**, v. 1, p. 613–616, 2019.
- 84 LUGARESI, C. et al. MediaPipe: A Framework for Building Perception Pipelines. 14 jun. 2019.
- 85 BAZAREVSKY, V. et al. BlazeFace: Sub-millisecond Neural Face Detection on Mobile GPUs. 11 jul. 2019.
- 86 SCHIAVENATO, M. et al. Neonatal pain facial expression: Evaluating the primal face of pain. **Pain**, v. 138, n. 2, p. 460–471, 31 ago. 2008.
- 87 AMARAL, V. et al. Normalização Espacial de Imagens Frontais de Face em Ambientes Controlados e Não-Controlados. **Periodico Científico Eletronico da FATEC São Caetano do Sul (FaSci-Tech)**, v. 1, out. 2009.

- 88 CHENG, D. et al. Current state of science in machine learning methods for automatic infant pain evaluation using facial expression information: study protocol of a systematic review and meta-analysis. **BMJ Open**, v. 9, n. 12, p. e030482, 11 dez. 2019.
- 89 FONTANA, É. **Introdução aos Algoritmos de Aprendizagem Supervisionada**. Paraná: Universidade Federal do Paraná (UFPR), 2020.
- 90 AKIBA, T. et al. Optuna: A Next-generation Hyperparameter Optimization Framework. **KDD '19: Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining**. Disponível em: <https://doi.org/10.1145/3292500.333070125> jul. 2019. Acesso em: 01 ago. 2024.
- 91 BRAHNAM, S. et al. Machine assessment of neonatal facial expressions of acute pain. **Decision Support Systems**, v. 43, n. 4, p. 1242–1254, ago. 2007.
- 92 VAPNIK, V.; CHERVONENKIS, A. On a Class of Perceptrons. **Automation and Remote Control**, v. 25, n. 1, p. 103–109, 1964.
- 93 LECUN, Y. et al. Gradient-based learning applied to document recognition. **Proceedings of the IEEE**, v. 86, n. 11, p. 2278–2324, 1998.
- 94 CARLINI, L. P. et al. A Convolutional Neural Network-based Mobile Application to Bedside Neonatal Pain Assessment. *In*: 34th SIBGRAPI CONFERENCE ON GRAPHICS, PATTERNS AND IMAGES, 2021. **Anais eletrônicos ...IEEE**, 2021. Disponível em: <https://ieeexplore.ieee.org/document/9643144/>. Acesso em: 01 ago. 2024.
- 95 VASWANI, A. et al. Attention Is All You Need. *In*: 31st CONFERENCE ON NEURAL INFORMATION PROCESSING SYSTEMS (NIPS), 2017, Long Beach, CA, USA. **Anais... USA**, 2017.
- 96 DOSOVITSKIY, A. et al. An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. *In*: THE NINTH INTERNATIONAL CONFERENCE ON LEARNING REPRESENTATIONS, 2021. **Anais eletrônicos...ICLR**, 2021. Disponível em: <https://openreview.net/pdf?id=YicbFdNTTy>. Acesso em: 29 ago. 2024.
- 97 SARMADI, A. et al. Comparative analysis of vision transformers and convolutional neural networks in osteoporosis detection from X-ray images. **Scientific Reports**, v. 14, n. 1, p. 18007, 3 ago. 2024.
- 98 DOMINGUES, P. H. **Segmentação de faces parcialmente ocluídas para avaliação da expressão de dor neonatal**. 2024. 69 f. Dissertação (Mestrado em Engenharia Elétrica) - Centro Universitário FEI, São Bernardo do Campo, 2024. Disponível em: <https://repositorio.fei.edu.br/bitstreams/99a0b123-fccf-4894-9d66-918610b4ea46/download>
- 99 ANAND, K. J. S.; HICKEY, P. R. Pain and Its Effects in the Human Neonate and Fetus. **New England Journal of Medicine**, v. 317, n. 21, p. 1321–1329, 1987.

- 100 GOLIANU, B. et al. Pediatric acute pain management. **Pediatric Clinics of North America**, v. 47, n. 3, p. 559–587, 2000.
- 101 AYMAR, C. L. G. DE; COUTINHO, S. B. Fatores relacionados ao uso de analgesia sistêmica em neonatologia. **Revista Brasileira de Medicina Intensiva**, v. 20, n. 4, p. 405–410, 2008.
- 102 HARRISON, D.; BUENO, M. Translating evidence: pain treatment in newborns, infants, and toddlers during needle-related procedures. **Pain Reports**, v. 8, n. 2, p. E1064, 16 mar. 2023.
- 103 THE JOINT COMMISSION ON ACCREDITATION OF HEALTHCARE ORGANIZATIONS; THE NATIONAL PHARMACEUTICAL COUNCIL. Pain: Current Understanding of Assessment, Management, and Treatments. **National Pharmaceutical Council, 2001.**, 2001.
- 104 MAXWELL, L. G.; FRAGA, M. V.; MALAVOLTA, C. P. **Assessment of Pain in the Newborn: An Update.** Clinics in Perinatology. W.B. Saunders, 1 dez. 2019.
- 105 MCPHERSON, C.; ORTINAU, C. M.; VESOULIS, Z. Practical approaches to sedation and analgesia in the newborn. **Journal of Perinatology**, Springer Nature, 1 mar. 2021.
- 106 GIORDANO, V. et al. Pain and Sedation Scales for Neonatal and Pediatric Patients in a Preverbal Stage of Development: A Systematic Review. **JAMA Pediatrics**, v. 173, n. 12, p. 1186–1197, 2019.
- 107 STEVENS, B. J. et al. The Premature Infant Pain Profile-Revised (PIPP-R). **The Clinical Journal of Pain**, v. 30, n. 3, p. 238–243, 2014.
- 108 SARKARIA, E. Assessing Neonatal Pain with NIPS and COMFORT-B: Evaluation of NICU 's Staff Competences. **Pain Research and Management**, v. 2022, p. 1–9, 2022.
- 109 VAN DIJK, M. et al. The reliability and validity of the COMFORT scale as a postoperative pain instrument in 0 to 3-year-old infants. **Pain**, v. 84, n. 2, p. 367–377, 1 fev. 2000.
- 110 HUMMEL, P. et al. N-PASS: Neonatal Pain, Agitation, and Sedation Scale - reliability and validity. *In: PEDIATRIC ACADEMIC SOCIETIES' ANNUAL MEETING*, 2003, Seattle. **Anais...**, Seattle, WA, 2003.
- 111 GRUNAU, R. E. et al. Bedside application of the Neonatal Facial Coding System in pain assessment of premature neonates. **Pain**, v. 76, p. 277–286, 1998.
- 112 VAART, M. et al. Multimodal pain assessment improves discrimination between noxious and non-noxious stimuli in infants. **Paediatric and Neonatal Pain**, v. 1, n. 1, p. 21–30, 9 set. 2019.

- 113 VAN DER VAART, M. et al. Premature infants display discriminable behavioral, physiological, and brain responses to noxious and nonnoxious stimuli. **Cerebral Cortex**, v. 32, n. 17, p. 3799–3815, 22 ago. 2022.
- 114 BARROS, M. C. DE M. et al. Identification of pain in neonates: the adults' visual perception of neonatal facial features. **Journal of Perinatology**, v. 41, n. 9, p. 2304–2308, 12 set. 2021.
- 115 SILVA, G. V. T. DA et al. What Facial Features Does the Pediatrician Look to Decide That a Newborn Is Feeling Pain? **American Journal of Perinatology**, v. 40, n. 08, p. 851–857, 30 jun. 2023.
- 116 TAMANAKA, F. G. et al. Neonatal pain assessment: A Kendall analysis between clinical and visually perceived facial features. **Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization**, v. 11, n. 3, p. 331–340, 4 maio 2022.
- 117 SERPA, A. B. M. et al. Multidimensional pain assessment of preterm newborns at the 1st, 3rd and 7th days of life. **Sao Paulo Medical Journal**, v. 125, n. 1, p. 29–33, jan. 2007.
- 118 SAKAI, T.; NAGAO, M.; KANADE, T. Computer analysis and classification of photographs of human faces. In: FIRT USA-JAPAN COMPUTER CONFERENCE, 1972. **Anais...**1972.
- 119 SCHIAVENATO, M. et al. Neonatal pain facial expression: Evaluating the primal face of pain. **Pain**, v. 138, n. 2, p. 460–471, ago. 2008.
- 120 BERGAMASCO, L. et al. Pain Assessment in Neonatal Clinical Practice via Facial Expression Analysis and Deep Learning. Em: [s.l: s.n.]. p. 249–263.
- 121 ZHAO, F. et al. Neonatal Pain Detection from Facial Expressions Using Deep Learning. **Research Square**, p. 1–16, 2024.
- 122 BEAM, K. et al. Artificial intelligence in the neonatal intensive care unit: the time is now. **Journal of Perinatology**, v. 44, n. 1, p. 131–135, 13 jan. 2024.
- 123 BRAHNAM, S. et al. SVM Classification of Neonatal Facial Images of Pain. In: 6TH INTERNATIONAL WORKSHOP ON FUZZY LOGIC AND APPLICATIONS, 2005. **Anais...**Crema, Italy: LNCS Vol. 3849, 2005.
- 124 BRAHNAM, S.; NANNI, L.; SEXTON, R. Introduction to Neonatal Facial Pain Detection Using Common and Advanced Face Classification Techniques. In: YOSHIDA, H., JAIN, A., ICHALKARANJE, A., JAIN, L.C., ICHALKARANJE, N. **Advanced Computational Intelligence Paradigms in Healthcare – 1. Studies in Computational Intelligence**. vol 48. Springer, 2007, Berlin, Heidelberg. Disponível em: https://doi.org/10.1007/978-3-540-47527-9_9. Acesso em: 01 ago. 2024.

- 125 KARTYNNIK, Y. et al. Real-time Facial Surface Geometry from Monocular Video on Mobile GPUs, 2019. *In: WORKSHOP ON COMPUTER VISION FOR AUGMENTED AND VIRTUAL REALITY, 2019. Anais eletrônicos...*, Long Beach, CA, USA, 2019. Disponível em: <https://doi.org/10.48550/arXiv.1907.06724> . Acesso em: 01 ago. 2024.
- 126 R CORE TEAM. **R: A Language and Environment for Statistical Computing**. Vienna, Austria R Foundation for Statistical Computing, , 24 abr. 2024. Disponível em: <https://www.R-project.org/>. Acesso em: 01 out. 2024.
- 127 MCNEMAR, Q. Note on the sampling error of the difference between correlated proportions or percentages. **Psychometrika**, v. 12, n. 2, p. 153–157, jun. 1947.
- 128 LANDIS, J. R.; KOCH, G. G. The measurement of observer agreement for categorical data. **Biometrics**, v. 33, n. 1, p. 159–74, mar. 1977.
- 129 RIDNIK, T. et al. ImageNet-21K Pretraining for the Masses. **CoRR**, 22 abr. 2021. Disponível em: <https://dblp.org/rec/journals/corr/abs-2104-10972.bib>. Acesso em: 01 out. 2024.
- 130 SHORTEN, C.; KHOSHGOFTAAR, T. M. A survey on Image Data Augmentation for Deep Learning. **Journal of Big Data**, v. 6, n. 1, p. 60, 6 dez. 2019.
- 131 HEIDERICH, T. M.; GUINSBURG, R. Fisioterapia sob o contexto da dor na Unidade de Cuidados Intensivos Neonatal. *In: SARMENTO, G. Fisioterapia Hospitalar em Pediatria*, São Paulos: Manole, 2017.
- 132 HEIDERICH, T. M.; BARROS, M. C. DE M.; GUINSBURG, R. Inter-observer agreement in the identification of pain faces in full-term and late preterm newborns: cross-sectional study. **Brazilian Journal Of Pain**, v. 3, n. 3, p. 1–6, 2020.
- 133 CARLINI, L. P. et al. A Convolutional Neural Network-based Mobile Application to Bedside Neonatal Pain Assessment. *In: 34th SIBGRAPI CONFERENCE ON GRAPHICS, PATTERNS AND IMAGES, 2021. Anais eletrônicos...IEEE*, 2021. Disponível em: <https://ieeexplore.ieee.org/document/9643144/>. Acesso em: 01 ago. 2024.
- 134 SILVA, E. S. et al. Pediatricians' focus of sight at pain assessment during a neonatal heel puncture. **Revista Paulista de Pediatria**, v. 42, 2024.
- 135 POLO, T. C. F.; MIOT, H. A. Aplicações da curva ROC em estudos clínicos e experimentais. **Jornal Vascular Brasileiro**, v. 19, 2020.
- 136 BORTEN, J. B. L. et al. Looking through Providers' Eyes: Pain in the Neonatal Intensive Care Unit. **American Journal of Perinatology**, v. 41, n. S 01, p. e3242–e3248, 16 maio 2024.
- 137 ORSI, R. N. et al. Visual attention during neonatal pain assessment: A 2-s exposure to a facial expression is sufficient. **Electronics Letters**, v. 59, n. 6, 15 mar. 2023.

- 138 SOARES, J. DO C. A. et al. Looking at neonatal facial features of pain: do health and non-health professionals differ? **Jornal de Pediatria**, v. 98, n. 4, p. 406–412, 1 jul. 2022.

**ANEXO A - TRABALHOS PUBLICADOS COM A PARTICIPAÇÃO DA AUTORA
DA TESE**

- L. P. Carlini, G. A. S. Coutrin, L. A. Ferreira, J. C. A. Soares, G. V. T. Silva, T. M. Heiderich, R. C. X. Balda, M. C. M. Barros, R. Guinsburg, and C. E. Thomaz. Human vs machine towards neonatal pain assessment: A comprehensive analysis of the facial features extracted by health professionals, parents, and convolutional neural networks, *Artificial Intelligence in Medicine*, vol. 147, 10pp., January 2024.
- E. S. Silva, M. C. M. Barros, J. B. L. Borten, L. P. Carlini, R. C. X. Balda, R. N. Orsi, T. M. Heiderich, C. E. Thomaz, and R. Guinsburg. Pediatricians' focus of sight at pain assessment during a neonatal heel puncture, *Revista Paulista de Pediatria*, vol. 42, 8pp., 2024.
- J. B. L. Borten, M. C. M. Barros, E. S. Silva, L. P. Carlini, R. C. X. Balda, R. N. Orsi, T. M. Heiderich, A. Sanudo, C. E. Thomaz, and R. Guinsburg. Looking through providers' eyes: pain in the Neonatal Intensive Care Unit, *American Journal of Perinatology*, vol. 41, 7pp., 2024.
- P. H. S. Domingues, T. M. Heiderich, M. C. M. Barros, R. Guinsburg, and C. E. Thomaz. Neonatal Face Segmentation with and without Clinical Devices using SAM. In proceedings of the 36th SIBGRAPI, Conference on Graphics, Patterns and Images, Workshop of Works in Progress (WiP), pp. 101-104, Rio Grande, Rio Grande do Sul, Brazil, November 6th-9th 2023.
- T. M. Heiderich, L. P. Carlini, L. F. Buzuti, R. C. X. Balda, M. C. M. Barros, R. Guinsburg, and C. E. Thomaz. Face-based automatic pain assessment: challenges and perspectives in neonatal intensive care units, *Jornal de Pediatria*, 15pp., June 2023.
- R. N. Orsi, L. P. Carlini, T. M. Heiderich, G. V. T. Silva, J. C. A. Soares, R. C. X. Balda, M. C. M. Barros, R. Guinsburg and C. E. Thomaz. Visual attention during neonatal pain assessment: A 2-s exposure to a facial expression is sufficient, *Electronics Letters*, vol. 59, no. 6, 3pp., March 2023.
- L. F. Buzuti, T. M. Heiderich and C. E. Thomaz. Spatial distribution analysis of images in GAN training with small datasets. In proceedings of the XIX Encontro Nacional de Inteligencia Artificial, ENIAC 2022, pp. 775-786, Campinas, Sao Paulo, Brazil, November 28th - December 1st 2022.

- G. A. S. Coutrin, L. P. Carlini, L. A. Ferreira, T. M. Heiderich, R. C. X. Balda, M. C. M. Barros, R. Guinsburg and C. E. Thomaz. Convolutional neural networks for newborn pain assessment using face images: A quantitative and qualitative comparison. In proceedings of the 3rd International Conference on Medical Imaging and Computer-Aided Diagnosis, MICAD 2022, Springer LNEE, vol. 810, pp. 503-513, Leicester, UK, November 20th-21st 2022.
- J. C. A. Soares, M. C. M. Barros, G. V. T. Silva, L. P. Carlini, T. M. Heiderich, R. N. Orsi, R. C. X. Balda, P. A. S. O. Silva, C. E. Thomaz and R. Guinsburg. Looking at neonatal facial features of pain: do health and non-health professionals differ?, *Jornal de Pediatria*, vol. 98, no. 4, pp. 406-412, July-August 2022.
- F. G. Tamanaka, L. P. Carlini, T. M. Heiderich, R. C. X. Balda, M. C. M. Barros, R. Guinsburg and C. E. Thomaz. Neonatal pain assessment: A Kendall analysis between clinical and visually perceived facial features, *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization*, 10pp., March 2022.
- P. H. S. Domingues, R. M. M. Silva, I. J. Orra, M. E. Cruz, T. M. Heiderich and C. E. Thomaz. Neonatal face mosaic: An areas-of-interest segmentation method based on 2D face images. In proceedings of the XVII Workshop de Visao Computacional, WVC 2021, pp. 201-205, virtual, November 22nd-23rd 2021.
- M. C. M. Barros, C. E. Thomaz, G. V. T. Silva, J. C. A. Soares, L. P. Carlini, T. M. Heiderich, R. N. Orsi, R. C. X. Balda, P. A. S. Orona, A. Sanudo, S. Andreoni and R. Guinsburg. Identification of pain in neonates: the adults' visual perception of neonatal facial features, *Journal of Perinatology*, July 2021.
- G. V. T. Silva, M. C. M. Barros, J. C. A. Soares, L. P. Carlini, T. M. Heiderich, R. N. Orsi, R. C. X. Balda, C. E. Thomaz and R. Guinsburg. What facial features does the pediatrician look to decide that a newborn is feeling pain?, *American Journal of Perinatology*, June 2021.

- L. P. Carlini, F. G. Tamanaka, J. C. A. Soares, G. V. T. Silva, T. M. Heiderich, R. C. X. Balda, M. C. M. Barros, R. Guinsburg and C. E. Thomaz. Neonatal pain scales and human visual perception: An exploratory analysis based on facial expression recognition and eye-tracking. In proceedings of the 25th International Conference on Pattern Recognition, ICPR 2020, Workshop on Computational and Affective Intelligence in Healthcare Applications, 15pp., Milan, Italy, 10th January 2021.
- L. F. Buzuti, T. M. Heiderich, M. C. M. Barros, R. Guinsburg and C. E. Thomaz. Neonatal pain assessment from facial expression using Deep Neural Networks. In proceedings of the XVI Workshop de Visão Computacional, WVC 2020, pp. 87-92, Uberlandia, Minas Gerais, Brazil, October 7th-8th 2020.
- P. A. S. O. Silva, D. A. D. Fabbro, T. M. Heiderich, M. C. M. Barros, R. C. X. Balda, R. Guinsburg and C. E. Thomaz. Interpretação e reconhecimento de padrões para avaliação de dor em imagens faciais de recém-nascidos. In proceedings of the 20th Simpósio Brasileiro de Computação Aplicada a Saúde, SBCAS 2020, pp. 285-296, Salvador, Bahia, Brazil, September 15th-18th 2020.
- L. P. Carlini, J. C. A. Soares, G. V. T. Silva, T. M. Heiderich, R. C. X. Balda, M. C. M. Barros, R. Guinsburg and C. E. Thomaz. A visual perception framework to analyse neonatal pain in face images. In proceedings of the 17th International Conference on Image Analysis and Recognition, ICIAR 2020, Springer LNCS, pp. 233-243, Povia de Varzim, Portugal, June 24th-26th 2020.
- L. P. Carlini, T. M. Heiderich, M. C. M. Barros, R. C. X. Balda, R. Guinsburg and C. E. Thomaz. Visual perception of pain in neonatal face images. In proceedings of the XV Workshop de Visao Computacional, WVC 2019, pp. 37-42, Sao Bernardo do Campo, Sao Paulo, Brazil, September 9th-11 2019.
- P. A. S. Orona, D. A. D. Fabbro, T. M. Heiderich, M. C. M. Barros, R. C. X. Balda, R. Guinsburg and C. E. Thomaz. Atlas of neonatal face images using triangular meshes. In proceedings of the XV Workshop de Visão Computacional, WVC 2019, pp. 19-24, Sao Bernardo do Campo, Sao Paulo, Brazil, September 9th-11 2019.

G. F. Teruel, T. M. Heiderich, R. Guinsburg and C. E. Thomaz. Analysis and recognition of pain in 2D face images of full term and healthy newborns. In proceedings of the XV Encontro Nacional de Inteligencia Artificial, ENIAC 2018, pp. 228-239, Sao Paulo, Sao Paulo, Brazil, October 22nd-25th 2018.40

ANEXO B - SOLICITAÇÃO DE AUTORIZAÇÃO PARA USO DE IMAGENS – ICOPE

SOLICITAÇÃO DE UTILIZAÇÃO DA BASE DE DADOS iCOPE

Eu, **Tatiany Marcondes Heiderich**, RG. 41.169.619-1, solicito autorização para utilização das imagens da base de dados iCOPE:

Consentimento solicitado:

O(s) investigador(es) concorda(m) com as seguintes restrições na Base de Dados iCOPE:

1. Redistribuição: Sem aprovação prévia do Investigador Principal, a Base de Dados iCOPE, em no todo ou em parte, **não serão distribuídos, publicados, copiados ou divulgados de qualquer forma ou forma, seja com ou sem fins lucrativos**. Isto inclui uma maior distribuição, cópia ou divulgação para uma instalação diferente ou unidade organizacional dentro da universidade, organização, ou empresa requerente.

2. Modificação e utilização comercial: Sem aprovação prévia do Investigador Principal, a base de dados iCOPE, no todo ou em parte, **não pode ser modificada ou utilizada para fins comerciais**. A licença concedida aqui é especificamente para fins de investigação interna do investigador, e o investigador não deve duplicar ou utilizar as bases de dados iCOPE divulgadas ou o seu conteúdo para fabricar ou vender produtos ou tecnologias (ou porções do mesmo) direta ou indiretamente para comercialização ou qualquer outra finalidade direta com fins lucrativos sem a autorização prévia por escrito do Investigador Principal da Base de Dados do iCOPE.

3. Pedidos para as bases de dados da iCOPE: Todos os pedidos para a Base de Dados de COPE Infantil serão encaminhados para o Investigador Principal.

4. Requisitos de publicação: Em caso algum devem as imagens faciais ser utilizadas de uma forma que possa causar ao sujeito original ou família do sujeito embaraçoso ou angústia mental.

5. Citação: Todos os documentos e trabalhos que relatam pesquisas que utilizam as bases de dados iCOPE devem reconhecer a utilização da base de dados através da inclusão de citações apropriadas; se não estiverem listadas na página seguinte, elas ser fornecido a pedido do Investigador Principal.

6. Indemnização: O investigador concorda em indemnizar, defender, e manter inofensivo o Dr. Brahnam, Hospital St John's, e a Universidade Estatal do Missouri e o seu Conselho de Curadores, funcionários, empregados e agentes, individualmente e colectivamente, de todas e quaisquer perdas, despesas, danos, reivindicações e/ou reclamações baseadas em qualquer um desses danos ou danos (reais ou alegados) e pagará todos os danos, reclamações, sentenças ou despesas resultantes da utilização por parte do Investigador das bases de dados iCOPE.

Citation information for pilot Infant Cope database:

Sheryl Brahnam, Chao-Fa Chuang, Frank Y. Shih, Melinda R. Slack. 'Svm classification of neonatal facial images of pain.' In Proceedings of the WILF 2005 6th International Workshop on Fuzzy Logic and Applications, Crema, Italy, 2005, LNCS Vol. 3849.

Sheryl Brahnam, Chao-Fa Chuang, Frank Y. Shih, Melinda R. Slack. 'Machine recognition and representation of neonatal facial displays of acute pain.' In Artificial Intelligence in Medicine, 36 (3): 211-222, 2006.

Sheryl Brahnam, Chao-Fa Chuang, Randal S. Sexton, Frank Y. Shih. 'Machine assessment of neonatal facial expressions of acute pain.' In Decision Support Systems, 43 1247-1254, 2007.

Sheryl Brahnam, Loris Nanni, Sexton Randall (2007). "Introduction to neonatal facial pain detection using common and advanced face classification techniques," Advanced computation Intelligence Paradigms in Healthcare - 1, Studies in Computational Intelligence (SCI) Series, Springer Berlin, Volume 48, pp.225-253, 2007.

**ANEXO C - SOLICITAÇÃO DE AUTORIZAÇÃO PARA USO DE IMAGENS -
UNIFESP-I**

São Paulo 11 de dezembro de 2024

SOLICITAÇÃO DE UTILIZAÇÃO DA BASE DE DADOS UNIFESP-I

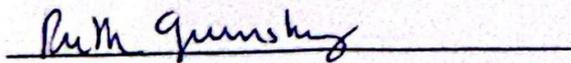
Eu, **Tatiany Marcondes Heiderich**, RG. 41.169.619-1, solicito autorização para utilização das fotos de faces de recém-nascidos obtidas na realização do projeto de doutorado da aluna Tatiany Marcondes Heiderich, intitulado "**DESENVOLVIMENTO DE SOFTWARE PARA IDENTIFICAR A EXPRESSÃO FACIAL DE DOR DO RECÉM-NASCIDO**", cujo projeto foi aprovado pelo Comitê de Ética em Pesquisa do Hospital São Paulo - Escola Paulista de Medicina/Unifesp, sob o CEP: 1299/2009, tendo sido obtido o Termo de Consentimento Livre e Esclarecido para a mãe ou responsável do recém-nascido.

Declaro que o uso destas imagens deverá ser utilizado para a realização da pesquisa de doutorado "**RECONHECIMENTO AUTOMATIZADO DA DOR POR MOVIMENTOS FACIAIS DE RECÉM-NASCIDOS INTERNADOS EM UNIDADE DE TERAPIA INTENSIVA NEONATAL**", do Centro Universitário FEI em parceria com Unidade Neonatal do Hospital São Paulo - Hospital Universitário da Escola Paulista de Medicina - Universidade Federal de São Paulo, Disciplina de Pediatria.

A pesquisa será por mim realizada, sob a orientação do **Prof. Carlos Eduardo Thomaz (FEI)**, e tem como objetivo desenvolver um método de reconhecimento automatizado da dor, por regiões da face do recém-nascido internado em Unidade de Terapia Intensiva Neonatal, que permanece com dispositivos fixados em sua face.

Autorizo

Não autorizo



Ruth Guinsburg

**ANEXO D - SOLICITAÇÃO DE AUTORIZAÇÃO PARA USO DE IMAGENS -
UNIFESP-II**

São Paulo 11 de dezembro de 2024

SOLICITAÇÃO DE UTILIZAÇÃO DA BASE DE DADOS UNIFESP-II

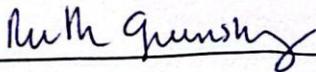
Eu, **Tatiany Marcondes Heiderich**, RG. 41.169.619-1, solicito autorização para utilização das fotos de faces de recém-nascidos obtidas na realização do projeto de pós-doutorado da aluna Tatiany Marcondes Heiderich, intitulado "**VALIDAÇÃO DE SOFTWARE QUE IDENTIFICA A EXPRESSÃO FACIAL DE DOR PARA SER UTILIZADO COM RECÉM-NASCIDOS PREMATUROS NA UNIDADE DE TERAPIA INTENSIVA**", cujo projeto foi aprovado pelo Comitê de Ética em Pesquisa do Hospital São Paulo - Escola Paulista de Medicina/Unifesp, sob o CEP: 0566/2015, número do parecer: 1.150.901, tendo sido obtido o Termo de Consentimento Livre e Esclarecido para a mãe ou responsável do recém-nascido.

Declaro que o uso destas imagens deverá ser utilizado para a realização da pesquisa de doutorado "**RECONHECIMENTO AUTOMATIZADO DA DOR POR MOVIMENTOS FACIAIS DE RECÉM-NASCIDOS INTERNADOS EM UNIDADE DE TERAPIA INTENSIVA NEONATAL** ", do Centro Universitário FEI em parceria com Unidade Neonatal do Hospital São Paulo - Hospital Universitário da Escola Paulista de Medicina - Universidade Federal de São Paulo, Disciplina de Pediatria.

A pesquisa será por mim realizada, sob a orientação do **Prof. Carlos Eduardo Thomaz (FEI)**, e tem como objetivo desenvolver um método de reconhecimento automatizado da dor, por regiões da face do recém-nascido internado em Unidade de Terapia Intensiva Neonatal, que permanece com dispositivos fixados em sua face.

Autorizo

Não autorizo



Ruth Guinsburg

**ANEXO E - SOLICITAÇÃO DE AUTORIZAÇÃO PARA USO DE IMAGENS –
SYNUNIFESP**

São Paulo 11 de Dezembro de 2024

SOLICITAÇÃO DE UTILIZAÇÃO DA BASE DE DADOS SynUnifesp

Eu, **Tatiany Marcondes Heiderich**, RG. 41.169.619-1, solicito autorização para utilização das fotos de faces de recém-nascidos obtidas na realização do projeto de doutorado do aluno Lucas Buzuti, intitulado "**INTELIGÊNCIA ARTIFICIAL GENERATIVA PARA AVALIAÇÃO DA DOR NEONATAL: UMA ABORDAGEM SÓLIDA PARA MELHORAR O RECONHECIMENTO AUTOMÁTICO BASEADO EM DADOS NA UNIDADE DE TERAPIA INTENSIVA**".

Declaro que o uso destas imagens deverá ser utilizado para a realização da pesquisa de doutorado "**RECONHECIMENTO AUTOMATIZADO DA DOR POR MOVIMENTOS FACIAIS DE RECÉM-NASCIDOS INTERNADOS EM UNIDADE DE TERAPIA INTENSIVA NEONATAL**", do Centro Universitário FEI em parceria com Unidade Neonatal do Hospital São Paulo - Hospital Universitário da Escola Paulista de Medicina - Universidade Federal de São Paulo, Disciplina de Pediatria.

A pesquisa será por mim realizada, sob a orientação do **Prof. Carlos Eduardo Thomaz (FEI)**, e tem como objetivo desenvolver um método de reconhecimento automatizado da dor, por regiões da face do recém-nascido internado em Unidade de Terapia Intensiva Neonatal, que permanece com dispositivos fixados em sua face.

Autorizo

Não autorizo



Carlos Eduardo Thomaz

Autorizo

Não autorizo



Lucas Buzuti