

**UNIVERSIDADE FEDERAL DE UBERLÂNDIA**  
**FACULDADE DE ENGENHARIA ELÉTRICA**  
**PROGRAMA DE PÓS-GRADUAÇÃO EM**  
**ENGENHARIA BIOMÉDICA**

**Proposta de Sistema Integrado Para Detecção de**  
**Temperatura Corporal e Expressões Faciais**

AMANDA ROSA FERREIRA JORGE

Uberlândia, MG, Brasil

Junho de 2023

AMANDA ROSA FERREIRA JORGE

## **Proposta de Sistema Integrado Para Detecção de Temperatura Corporal e Expressões Faciais**

Dissertação apresentada ao Programa de Pós-graduação em Engenharia Biomédica da Universidade Federal de Uberlândia como requisito parcial para obtenção do título de mestre em Engenharia Biomédica.

Área de concentração: Engenharia Biomédica  
Orientador: Prof. Alcimar Barbosa Soares, PhD  
Coorientador: Prof. Dr. Márcio José da Cunha

Banca Examinadora:

Prof. Alcimar Barbosa Soares, PhD – UFU (Orientador)  
Prof. Dr. Marcelo Barros de Almeida - FEELT UFU  
Rodrigo Antonio Faccioli, PhD - MI4U

Uberlândia, MG, Brasil

Junho de 2023

Ficha Catalográfica Online do Sistema de Bibliotecas da UFU  
com dados informados pelo(a) próprio(a) autor(a).

J82  
2023 Jorge, Amanda Rosa Ferreira, 1995-  
Proposta de Sistema Integrado Para Detecção de  
Temperatura Corporal e Expressões Faciais [recurso  
eletrônico] / Amanda Rosa Ferreira Jorge. - 2023.

Orientador: Alcimar Barbosa Soares.  
Coorientador: Márcio José da Cunha.  
Dissertação (Mestrado) - Universidade Federal de  
Uberlândia, Pós-graduação em Engenharia Biomédica.  
Modo de acesso: Internet.  
Disponível em: <http://doi.org/10.14393/ufu.di.2023.333>  
Inclui bibliografia.  
Inclui ilustrações.

1. Engenharia biomédica. I. Soares, Alcimar Barbosa,  
1965-, (Orient.). II. Cunha, Márcio José da, 1978-,  
(Coorient.). III. Universidade Federal de Uberlândia.  
Pós-graduação em Engenharia Biomédica. IV. Título.

CDU: 62:61

Bibliotecários responsáveis pela estrutura de acordo com o AACR2:  
Gizele Cristine Nunes do Couto - CRB6/2091  
Nelson Marcos Ferreira - CRB6/3074



**UNIVERSIDADE FEDERAL DE UBERLÂNDIA**  
Coordenação do Programa de Pós-Graduação em Engenharia Biomédica  
Av. João Naves de Ávila, 2121, Bloco 3N, Sala 115 - Bairro Santa Mônica, Uberlândia-MG, CEP 38400-902  
Telefone: (34) 3239-4761 - www.ppegb.feelt.ufu.br - ppegb@feelt.ufu.br



### ATA DE DEFESA - PÓS-GRADUAÇÃO

Programa de Pós-Graduação em:	Engenharia Biomédica				
Defesa de:	Dissertação de Mestrado Acadêmico, 094, PPGEB				
Data:	Trinta de junho de dois mil e vinte e três	Hora de início:	09 horas e 10 minutos	Hora de encerramento:	11:00 horas
Matrícula do Discente:	12112EBI001				
Nome do Discente:	Amanda Rosa Ferreira Jorge				
Título do Trabalho:	Proposta de Sistema Integrado Para Detecção de Expressões Faciais e Temperatura Corporal				
Área de concentração:	Engenharia Biomédica				
Linha de pesquisa:	Engenharia de Reabilitação e Biomecânica				
Projeto de Pesquisa de vinculação:	A Comprehensive System for Clinical Rehabilitation of Upper-Limb Amputees in Virtual Reality				

Reuniu-se em web conferência pela plataforma RNP em conformidade com a RESOLUÇÃO CONPEP Nº 17, DE 09 DE JUNHO DE 2022, a Banca Examinadora, designada pelo Colegiado do Programa de Pós-graduação em Engenharia Biomédica, assim composta: Doutor Rodrigo Antonio Faccioli - MI4U Ltda; Prof. Doutor Marcelo Barros de Almeida - FEELT/UFU, e Prof. Doutor Alcimar Barbosa Soares - PPGEB/UFU orientador da candidata.

Iniciando os trabalhos o presidente da mesa, Prof. Alcimar Barbosa Soares, apresentou a Comissão Examinadora e a candidata, agradeceu a presença do público, e concedeu à Discente a palavra para a exposição do seu trabalho. A duração da apresentação da Discente e o tempo de arguição e resposta se deram conforme as normas do Programa.

A seguir o senhor presidente concedeu a palavra, pela ordem sucessivamente, aos examinadores, que passaram a arguir a candidata. Ultimada a arguição, que se desenvolveu dentro dos termos regimentais, a Banca, em sessão secreta, atribuiu o resultado final, considerando a candidata:

Aprovada.

Esta defesa faz parte dos requisitos necessários à obtenção do título de Mestre.

O competente diploma será expedido após cumprimento dos demais requisitos, conforme as normas do Programa, a legislação pertinente e a regulamentação interna da UFU.

Nada mais havendo a tratar foram encerrados os trabalhos. A presente ata foi lavrada e, após lida e achada conforme, foi assinada pela Banca Examinadora.



Documento assinado eletronicamente por **Alcimar Barbosa Soares, Professor(a) do Magistério Superior**, em 30/06/2023, às 11:03, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).

---



Documento assinado eletronicamente por **Marcelo Barros de Almeida, Professor(a) do Magistério Superior**, em 30/06/2023, às 11:04, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).

---



Documento assinado eletronicamente por **Rodrigo Antonio Faccioli, Usuário Externo**, em 14/07/2023, às 14:23, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).

---



A autenticidade deste documento pode ser conferida no site [https://www.sei.ufu.br/sei/controlador\\_externo.php?acao=documento\\_conferir&id\\_orgao\\_acesso\\_externo=0](https://www.sei.ufu.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0), informando o código verificador **4614363** e o código CRC **241A14B0**.

---

Dedico este trabalho a Deus, à minha mãe e à  
todos que contribuíram com a minha  
caminhada.

## **AGRADECIMENTOS**

Agradeço a Deus, por sempre ter me guiado, me abençoado, e ter me dado forças em inúmeras vezes. Agradeço à minha mãe, Alessandra Luzia Rosa, que sempre esteve ao meu lado segurando a minha mão. Agradeço ao meu avô e pai, Archimedes Rosa Netto, que sempre acreditou em mim. Agradeço ao meu namorado e melhor amigo, Lukas Gabriel, por todo o apoio e parceria. Agradeço ao professor Dr. Alcimar Soares pela orientação e confiança. Agradeço ao professor Dr. Márcio Cunha pela paciência e auxílio. Agradeço também ao Rodrigo Faccioli juntamente com a empresa MI4U que me motivaram e me apoiaram nessa caminhada. Agradeço aos voluntários por seu tempo e disponibilidade. Agradeço aos meus amigos por sempre estarem ao meu lado, até nos momentos mais difíceis. Agradeço também à Universidade Federal de Uberlândia pela oportunidade, bem como o suporte das agências de fomento à pesquisa CNPq, CAPES E FAPEMIG. Mais este passo não seria possível sem a contribuição de todos, muito obrigada.

“A vida são as incessantes oportunidades que surgem pela frente, jamais os insucessos que ocorreram no passado.”

(Joanna de Ângelis)



## RESUMO

As unidades de terapia intensiva ou centros de terapia intensiva são áreas hospitalares especializadas que fornecem tratamento específico e com monitoramento contínuo mediante a condição física dos pacientes internados. Os sinais vitais majoritariamente utilizados para monitoramento são o eletrocardiograma, a frequência cardíaca, a pressão arterial, a saturação de oxigênio no sangue e a temperatura corporal. Além de ter como causa o esgotamento físico do profissional intensivista, há a perda de informações aparentemente irrelevantes do paciente. Fato o qual gera atraso no diagnóstico e imprecisão na execução de intervenções médicas. Estudos apontam que a análise isolada dos sinais vitais não expõe o real estado fisiológico do paciente, defende-se que é necessário a implementação de um “sistema de sistemas” no qual a análise global de sinais agrupados aparentemente secundários pode ajudar a revelar importantes condições físicas ocultas. Por exemplo, a febre pode afetar até 75% dos pacientes da UTI, causando agitação e delírio. Também é verificado que mais de 50% dos pacientes internados apresentam relatos de dor moderada a severa durante a intubação em estado consciente e em repouso, podendo gerar aumento da carga de trabalho cardíaco. O ponto em que a agitação, ansiedade, dor, delírio e sedação possuem em comum são as potenciais expressões faciais do paciente. Assim, além de a temperatura corporal anormal ser caracterizada como febre, as expressões faciais dos pacientes em unidades de tratamento intensivo podem ser indícios da necessidade de intervenção médica.

A literatura atual pouco avança sobre a monitoração da expressão facial e da temperatura corporal sem contato de pacientes em geral, mas com especial destaque para aqueles não-comunicativos. Ainda que eficientes, os modelos presentes no estado-da-arte sobre expressões faciais não são robustos, são projetados para aplicações em sistemas offline e possuem alto custo de treinamento do algoritmo. Diante disso, o presente estudo tem como objetivo principal desenvolver modelos para captura da temperatura corporal e para classificação de expressões faciais de forma automatizada e eficiente a partir da aquisição de dados por câmera RGB. O modelo de classificação de emoções é baseado na análise de distâncias entre pontos específicos da face. O modelo foi desenvolvido para classificação de sete emoções faciais: raiva, nojo, feliz, neutro, triste, surpresa e medo. Para seleção de características utilizou-se a técnica de algoritmos genéticos e para classificação foram utilizadas máquinas de vetores de suporte. Os resultados demonstram que a técnica desenvolvida é capaz de realizar a classificação a partir de dados com acurácias médias acima de 80% para todas as classes, utilizando-se imagens dos bancos de dados WSEFEP, CK+ e JAFFE. As emoções surpresa, feliz e neutro apresentaram a melhor acurácia. Para predição de temperatura corporal sem contato, foram utilizadas imagens

capturadas de câmara convencional, no espectro visível, em RGB. Modelos de regressão foram utilizados para estimar a temperatura corporal, tomando-se como base três pontos na face: bochechas, subclaviculares e testa. Os resultados mostram que a temperatura média estimada varia de acordo com a região (bochechas, subclaviculares e testa) com  $R^2$  de 0,79 a 0,94. Os erros médios quadráticos se situam na faixa de 0,15-0,20 °C. Conclui-se assim, que as técnicas desenvolvidas são capazes de predizer com bom grau de acurácia as emoções e a temperatura corporal dos sujeitos, sem contato, e podem ser integradas em um único sistema de monitoração por câmara RGB com foco em pacientes acamados e não-comunicativos.

**Palavras-chave:** pacientes UTI; temperatura corporal; expressões faciais; sistema integrado.

## ABSTRACT

Intensive care units or intensive care centers are specialized hospital areas that provide specific treatment and continuous monitoring of the physical condition of hospitalized patients. The vital signs mostly used for monitoring are the electrocardiogram, heart rate, blood pressure, blood oxygen saturation and body temperature. In addition to being caused by the physical exhaustion of the intensivist professional, the loss of apparently irrelevant information from the patient generates delay in diagnosis and inaccuracy in the execution of medical interventions. Studies indicate that the isolated analysis of vital signs does not expose the patient's real physiological state, it is argued that it is necessary to implement a "system of systems" in which the global analysis groups apparently secondary signs that can help to reveal important hidden physical conditions . For example, fever can affect up to 75% of ICU patients, causing agitation and delirium. It is also verified that more than 50% of hospitalized patients present reports of moderate to severe pain during intubation in a conscious state and at rest, which may generate an increase in cardiac workload. The point at which agitation, anxiety, pain, delirium and sedation have in common is the patient's potential facial expressions. Thus, in addition to abnormal body temperature being characterized as fever, the facial expressions of patients in intensive care units may be indications of the need for medical intervention.

The current literature makes little progress on the monitoring of facial expression and body temperature without contact of patients in general, but with special emphasis on non-communicative patients. Although efficient, the state-of-the-art models on facial expressions are not robust, are designed for applications in offline systems and have a high algorithm training cost. Therefore, the main objective of the present study is to develop models for capturing body temperature and for classifying facial expressions in an automated and efficient way from data acquisition by RGB camera. The emotion classification model is based on the analysis of distances between specific points on the face. The model was developed to classify seven facial emotions: anger, disgust, happy, neutral, sad, surprise and fear. For feature selection, the genetic algorithm technique was used and for classification, support vector machines were used. The results demonstrate that the developed technique is able to perform the classification from data with average accuracies above 80% for all classes, using images from the WSEFEP, CK+ and JAFFE databases. Surprise, happy and neutral emotions showed the best accuracy. For non-contact body temperature prediction, images captured from a conventional camera were used, in the visible spectrum, in RGB. Regression models were used to estimate body temperature, based on three points on the face: cheeks, subclavicle and

forehead. The results show that the estimated average temperature varies according to the region (cheeks, subclavicular and forehead) with  $R^2$  from 0.79 to 0.94. Mean squared errors are in the range of 0.15-0.20 °C. It is thus concluded that the developed techniques are able to predict with a good degree of accuracy the emotions and body temperature of the subjects, without contact, and can be integrated into a single RGB camera monitoring system focusing on bedridden patients and non-communicative.

**Keywords:** ICU patients; body temperature; facial expressions; integrated system.

## LISTA DE ILUSTRAÇÕES

Figura 1 - Espectro eletromagnético.....	19
Figura 2 - Exemplos de imagens por diferentes frequências de radiação eletromagnética. De (A) a (D) respectivamente, domínio da luz visível, ultrassom, ondas de tv e ressonância magnética.....	20
Figura 3 - Aquisição da câmera para apresentação em duas dimensões. ....	21
Figura 4 - Formação da imagem em uma câmera digital. ....	21
Figura 5 - Espaço de cores em cubo, representação do espaço de cores com valores de intensidade de cores vermelho, verde e azul plotados como coordenadas x, y e z, respectivamente. ....	23
Figura 6 - Exemplo de um histograma em RGB. ....	24
Figura 7 - Representação de vetores considerando características globais e locais a partir de uma imagem. ....	25
Figura 8 - Conceitos de algoritmos genéticos como população, indivíduo e gene. ....	27
Figura 9 - Ilustração do classificador SVM considerando a margem que melhor separa grupos de atributos semelhantes.....	30
Figura 10 - Arquitetura do modelo MLP.....	32
Figura 11 - Apresentação de algumas unidades de ação (UA) facial de acordo com movimentos faciais estabelecidos por Paul Ekman.....	36
Figura 12 - Anatomia da expressão. ....	37
Figura 13 - Relação de unidades de ação faciais com a sensação de dor pelo SCAF. ....	38
Figura 14 - Vetores de medição de movimentos faciais.....	40
Figura 15 - No eixo X estão os bancos de dados analisados e no eixo Y acurácias alcançadas na literatura.....	41
Figura 16 - Regiões faciais comumente analisadas na literatura. ....	44
Figura 17 – Regiões corporais comumente analisadas na literatura.....	45
Figura 18 - Temperaturas de diferentes regiões corporais para sensação de desconforto frio, conforto e desconforto quente. ....	46
Figura 19 - Região facial apresentada pela câmera infravermelho. ....	47
Figura 20 - Regiões faciais analisadas por Ernst et al em imagens em domínio visível. ....	50
Figura 21 - Framework para monitoração de temperatura corporal e classificação de expressões faciais por câmera RGB.....	53

Figura 22 - Visão ilustrativa da sala de coleta durante aquisição de dados. Temperaturas corporais coletadas a partir dos sensores NTC durante a aquisição de dados. Apresentação de cinco sensores no plano anterior: (1) frente/testa, (2) região temporal da face, e (3) região subclavicular.....	55
Figura 23 - Apresentação do hardware em (A) LED vermelho e em (B) agrupamentos dos sensores térmicos utilizados. ....	56
Figura 24 - Diagrama de blocos do protocolo realizado. ....	58
Figura 25 - Diagrama de desenvolvimento e análise de dados.....	60
Figura 26 - Em azul escuro é demarcado as áreas de interesse calculadas automaticamente a partir das coordenadas dos olhos, nariz e cantos da boca. As regiões estão indicadas na imagem como (1) Glabella, (2) Bochechas e (3) Nariz e (4) é apresentado a diferença entre as regiões nariz e glabella.....	61
Figura 27 - Apresentação do retorno da ferramenta Dlib - 68 pontos da face; os pontos em azul claro são considerados no estudo.....	63
Figura 28 - Temperatura subclavicular esquerda - GM.....	67
Figura 29 - Temperatura subclavicular direita - GM.....	67
Figura 30 - Temperatura subclavicular esquerda - GF.....	68
Figura 31 - Temperatura subclavicular esquerda - GF.....	68
Figura 32 - Temperatura Ambiente e temperaturas corporais de um voluntário do GM. ....	69
Figura 33 - Temperatura Ambiente e Temperaturas corporais de uma voluntária do GF.....	70
Figura 34 - Temperaturas Subclavicular Esquerda e Direita, Média da Região da Glabella, Região do Nariz e Média da diferença entre regiões dos Pixels para o Canal Verde de um voluntário do GM. ....	71
Figura 35 - Temperaturas Subclavicular Esquerda e Direita e Média da Região da Bochecha Direita e Esquerda dos Pixels para o Canal Verde de um voluntário do GM. ....	71
Figura 36 - Matriz de correlação considerando variáveis do ambiente, temperaturas corporais e variáveis dos pixels do canal verde de um voluntário do GM.....	72
Figura 37 - Temperaturas Subclavicular Esquerda e Direita, Média da Região da Glabella, Região do Nariz e Média da diferença entre regiões dos Pixels para o Canal Verde de uma voluntária do GF.....	73
Figura 38 - Temperaturas Subclavicular Esquerda e Direita e Média da Região da Bochecha Direita e Esquerda dos Pixels para o Canal Verde de uma voluntária do GF.....	73
Figura 39 - Matriz de correlação considerando variáveis do ambiente, temperaturas corporais e variáveis dos pixels do canal verde de uma voluntária do GF. ....	74

Figura 40 - Matriz de correlação considerando as variáveis temperatura ambiente, temperaturas corporais e variações dos pixels do canal verde do GM.....	75
Figura 41 - Matriz de correlação média considerando as variáveis temperatura ambiente, temperaturas corporais e variações dos pixels do canal verde do GF.....	75
Figura 42 - Erro médio quadrático versus número de componentes principais – Bochechas GM. ....	77
Figura 43 - Abrangência da variância em (%) pelo número de componentes principais - Bochechas GM. ....	78
Figura 44 - Análise de predição de temperatura média das bochechas para o GM - curva em azul: valor real; curva em vermelho: valor predito.....	78
Figura 45 - Apresentação do desvio padrão entre valor real e valor predito para 100 amostras aleatórias, considerando a temperatura média das bochechas do GM. ....	79
Figura 46 - Erro médio quadrático pelo número de componentes principais – Bochechas GF. ....	79
Figura 47 - Apresentação da abrangência da variância em (%) pelo número de componentes principais – Bochechas GF.....	80
Figura 48 - Análise de predição de temperatura média das bochechas para o GF - curva em azul: valor real; curva em vermelho: valor predito. ....	80
Figura 49 - Apresentação do desvio padrão entre valor real e valor predito para 100 amostras aleatórias, considerando a temperatura média das bochechas do GF.....	80
Figura 50 - Erro médio quadrático pelo número de componentes principais – Bochechas GM. ....	82
Figura 51 - Apresentação da abrangência da variância em (%) pelo número de componentes principais – Bochechas GM.....	83
Figura 52 - Análise de predição de temperatura média das bochechas para o GM - curva em azul: valor real; curva em laranja: valor predito.....	83
Figura 53 - Apresentação do desvio padrão entre valor real e valor predito para 100 amostras aleatórias, considerando a temperatura média das bochechas do GM. ....	84
Figura 54 - Erro médio quadrático pelo número de componentes principais - Bochechas GF. ....	84
Figura 55 - Apresentação da abrangência da variância em (%) pelo número de componentes principais - Bochechas GF.....	85
Figura 56 - Análise de predição de temperatura média das bochechas para o GF - curva em azul: valor real; curva em laranja: valor predito.....	85

Figura 57 - Apresentação do desvio padrão entre valor real e valor predito para 100 amostras aleatórias, considerando a temperatura média das bochechas do GF.....	86
Figura 58 - Banco de dados WSEFEP com seleção de atributos e treinamento por MLP.....	87
Figura 59 - Banco de dados WSEFEP com seleção de atributos e treinamento por SVM. ....	87
Figura 60 - Banco de dados CK+ com seleção de atributos e treinamento por MLP. ....	88
Figura 61 - Banco de dados CK+ com seleção de atributos e treinamento por SVM. ....	89
Figura 62 - Banco de dados JAFFE com seleção de atributos e treinamento por MLP.....	89
Figura 63 - Banco de dados JAFFE com seleção de atributos e treinamento por SVM. ....	90
Figura 64 - Banco de dados misto com seleção de atributos e treinamento por MLP. ....	91
Figura 65 - Banco de dados misto com seleção de atributos e treinamento por SVM.....	91
Figura 66 – Framework do modelo proposto. ....	94



## LISTA DE TABELAS

Tabela 1 - Relação da emoção humana com as respectivas unidades de ativação estabelecido por Paul Ekman. ....	36
Tabela 2 - Apresentação do estado-da-arte com alguns estudos mais relevantes sendo nome do primeiro autor, ano de publicação, dados de entrada, técnica de classificador, base de dados e a média da acurácia. ....	42
Tabela 3 - Estado-da-arte em monitoramento de temperatura corporal sem contato.....	49
Tabela 4 - Estado-da-arte com foco em monitoramento de sinais vitais sem contato em domínio visível. ....	52
Tabela 5 – Idade, altura e peso dos voluntários.....	55
Tabela 6 - Análise de RCP com 32 variáveis dependentes incluindo temperatura e umidade ambiente. Ao todo 25 componentes principais foram selecionadas (D.P.: Desvio Padrão).....	76
Tabela 7 - Análise de RCP com 31 variáveis dependentes incluindo temperatura e umidade ambiente. Ao todo 25 componentes principais foram selecionadas. D.P. : Desvio Padrão. ....	82
Tabela 8 - Base de Dados, arquitetura de classificador utilizado, taxa de acurácia por emoção e a média da acurácia.....	92
Tabela 9 - Comparação estado-da-arte e o presente trabalho.....	103
Tabela 10 - Comparação estado-da-arte e o presente trabalho. Desempenho da predição de temperatura corporal quantificada pela pontuação $R^2$ e classificação de expressões faciais pela acurácia em (%). ....	105

## LISTA DE ABREVIATURAS E SIGLAS

UTI	UNIDADE DE TERAPIA INTENSIVA
CTI	CENTRO DE TERAPIA INTENSIVA
TRF	TECNOLOGIA DE RECONHECIMENTO FACIAL
cbPPG	FOTOPLETISMOGRAFIA BASEADOS EM CÂMERA
GPU	UNIDADE DE PROCESSAMENTO GRÁFICO
SVM	SUPPORT VECTOR MACHINE
CNN	CONVOLUTIONAL NEURAL NETWORK
MLP	MULTI-LAYER PERCEPTRON
AG	ALGORITMO GENÉTICO
AE	ALGORITMO EVOLUTIVO
RCP	REGRESSÃO DE COMPONENTES PRINCIPAIS
UA	UNIDADE DE AÇÃO
SCAF	SISTEMA DE CÓDIGO DE AÇÃO FACIAL
RDVB	RASTREADOR DE DEFORMAÇÃO DE VOLUME DE BÉZIER
REF	RECONHECIMENTO DE EXPRESSÃO FACIAL
WSEFEP	WARSAW SET OF EMOTIONAL FACIAL EXPRESSION PICTURES
CK+	EXTENDED COHN-KANADE DATASET
JAFFE	JAPANESE FEMALE FACIAL EXPRESSION

## SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO.....</b>	<b>12</b>
1.1	HIPÓTESE.....	15
1.2	OBJETIVO GERAL .....	15
1.3	OBJETIVOS ESPECÍFICOS.....	15
1.4	ESTRUTURA DA DISSERTAÇÃO .....	16
<b>2</b>	<b>REVISÃO DA LITERATURA.....</b>	<b>17</b>
2.1	VISÃO COMPUTACIONAL.....	17
2.1.1	CAPTURE E PROCESSAMENTO DE IMAGENS.....	18
2.1.1.1	LUZ E COMPRIMENTOS DE ONDA.....	18
2.1.1.2	AQUISIÇÃO DE IMAGENS E SENSORES.....	20
2.1.1.3	CARACTERÍSTICAS ESTRUTURAIS DAS IMAGENS .....	22
2.1.1.4	MÉTODOS DE APRENDIZADO DE MÁQUINA.....	25
2.1.1.4.1	ALGORITMOS GENÉTICOS .....	27
2.1.1.4.2	MÁQUINA DE VETORES DE SUPORTE (SVM) .....	29
2.1.1.4.3	PERCEPTRON DE MÚLTIPLAS CAMADAS (MLP) .....	31
2.1.1.4.4	REGRESSÃO DE COMPONENTES PRINCIPAIS (RCP) .....	33
2.2	EXPRESSÕES FACIAIS E EMOÇÕES HUMANAS.....	35
2.2.1	MÉTODOS DE RECONHECIMENTO DE EXPRESSÕES FACIAIS .....	39
2.3	TEMPERATURA CORPORAL .....	42
2.3.1	TÉCNICAS PARA MONITORAMENTO DE SINAIS VITAIS SEM CONTATO .....	45
<b>3</b>	<b>PROPOSTA DE MODELO INTEGRADO PARA MONITORAÇÃO DE TEMPERATURA CORPORAL E RECONHECIMENTO DE EXPRESSÕES.....</b>	<b>53</b>
3.1	METODOLOGIA.....	54
3.1.1	PARTICIPANTES .....	54
3.1.2	CRITÉRIOS DE INCLUSÃO DE VOLUNTÁRIOS.....	54
3.1.3	AQUISIÇÃO DE IMAGENS EM DOMÍNIO RGB .....	55
3.1.3.1	COLETA DE DADOS.....	56
3.1.4	BANCO DE DADOS .....	59
3.1.5	PREDIÇÃO DE TEMPERATURA CORPORAL .....	59
3.1.5.1	LOCALIZAÇÃO DE REGIÕES DE INTERESSE .....	61
3.1.6	CLASSIFICAÇÃO DE EXPRESSÕES FACIAIS .....	63

3.1.6.1	LOCALIZAÇÃO DE PONTOS-CHAVE DA REGIÃO FACIAL.....	63
3.1.6.2	MODELO DE CLASSIFICAÇÃO DE EXPRESSÕES FACIAIS .....	64
4	RESULTADOS .....	66
4.1	PREDIÇÃO DE TEMPERATURA CORPORAL .....	66
4.2	CLASSIFICAÇÃO DE EXPRESSÕES FACIAIS .....	86
4.3	FRAMEWORK DO MODELO PROPOSTO .....	93
5	DISCUSSÃO.....	95
5.1	PREDIÇÃO DE TEMPERATURA CORPORAL .....	95
5.2	CLASSIFICAÇÃO DE EXPRESSÕES FACIAIS .....	99
5.3	DISCUSSÃO GERAL .....	104
6	CONCLUSÃO.....	106
6.1	PREDIÇÃO DE TEMPERATURA CORPORAL .....	106
6.2	CLASSIFICAÇÃO DE EXPRESSÕES FACIAIS .....	106
6.3	CONCLUSÃO GERAL .....	107
7	TRABALHOS FUTUROS .....	107
8	PRODUÇÃO CIENTÍFICA .....	108
9	REFERÊNCIAS.....	109

# 1 INTRODUÇÃO

Unidades de terapia intensiva (UTI) ou centros de terapia intensiva (CTI) são áreas hospitalares especializadas que fornecem tratamento específico e com monitoramento contínuo mediante a condição física dos pacientes internados. Sendo assim, a UTI e o CTI oferecem um suporte avançado de vida em razão do estado de saúde de pacientes gravemente enfermos. Nestas unidades hospitalares encontram-se equipamentos com rigoroso desempenho à aquisição de dados fisiológicos do paciente. Como o monitoramento é contínuo, há grandes quantidades de dados adquiridos e avaliados pelos intensivistas com foco em auxiliar a equipe médica a executar as devidas intervenções para a recuperação do paciente [DING, 2017].

As causas mais frequentes de internações nestas unidades são infecções como pneumonia e sepse, doenças autoimunes como esclerose múltiplas, acidentes traumáticos e pós-operatórios em estado delicado. O monitoramento individual exige dos profissionais a constante atenção de informações alarmantes, pois é uma demanda humanamente impossível avaliar ininterruptamente todos os dados dos pacientes monitorados. Por este fato, nestas alas de internação, há uma central de monitoração na qual é advertido qualquer anormalidade nos sinais vitais.

Os sinais vitais majoritariamente utilizados para monitoramento são o eletrocardiograma, a frequência cardíaca, a pressão arterial, a saturação de oxigênio no sangue e a temperatura corporal. O valor destes parâmetros para a análise médica é alto pois são índices sensíveis que revelam, tradicionalmente, a condição de saúde do paciente [KIEKKAS, 2007]. Porém, o estudo de Bliss et al. (2015) defende que a análise isolada destes sinais não representa o estado geral do paciente, por exemplo, a análise do eletrocardiograma afere a frequência cardíaca e ritmo cardíaco mas não afere sobre a qualidade do bombeamento do coração [BLISS, 2015].

Além de ter como causa o esgotamento físico do intensivista, a perda de informações aparentemente irrelevantes gera atraso no diagnóstico e imprecisão na execução de intervenções médicas. Quando comparado com um sistema de aquisição de dados contínuo, pode-se agrupar todas as variáveis adquiridas e compor uma análise global de desempenho vital do paciente [BLISS, 2015].

Bliss et al. (2015) afirma que nos últimos anos, o número de técnicas de monitoramento para pacientes em UTI e CTI tem aumentado exponencialmente. Porém, como a análise isolada dos sinais vitais não expõe o real estado fisiológico do paciente, o estudo defende que é

necessária a implementação de um “sistema de sistemas” no qual a análise global agrupa sinais aparentemente díspares e pode ajudar a revelar condições físicas ocultas [BLISS, 2015].

A partir disso, embora as informações do paciente sejam coletadas ao longo do dia, raramente são direcionadas ao clínico supervisor, exceto quando surgem valores críticos ou o clínico interroga o estado de saúde do paciente. De modo que, quando há necessidade de uma intervenção, o clínico deve realizar uma revisão e sintetização de dados isolados para construir uma avaliação específica do paciente sobre o desempenho fisiológico. Com isso, o risco é alto de intervenções atrasadas, falhas em diagnóstico e ausência de informações em relatórios manuais [BLISS, 2015].

A febre é o aumento da temperatura corporal acima da faixa normal, temperatura de homeostase, em resposta a uma perturbação fisiológica ou doença. Esta resposta de defesa imunológica é caracterizada como componente de fase aguda executada pelo sistema metabólico, sistema endócrino e sistema nervoso autônomo. A elevação da temperatura corporal fornece vantagens adaptativas para o hospedeiro, visando a inibição do crescimento do parasita [KIEKKAS, 2007]. A literatura apresenta que a febre interfere diretamente em todos os sinais vitais apresentados como padrões de análise. Com a elevação da temperatura corporal, verifica-se um aumento significativo da frequência cardíaca, interferindo proporcionalmente na magnitude da febre. Verifica-se também a diminuição da média da pressão arterial e diminuição da saturação de oxigênio, fatos que se agravam com magnitudes que excedam 39 °C [KIEKKAS, 2007]. A febre é particularmente comum em pacientes de UTI e pode afetar até 75% dos pacientes da unidade médico-cirúrgica. Kiekkas et al. (2010) ressalta que a febre é o segundo fator de risco mais importante para a manifestação de agitação e delírio do paciente. A febre pode agravar qualquer disfunção mental pré-existente, principalmente em pacientes idosos pois são mais susceptíveis à elevação da temperatura corporal [KIEKKAS, 2010].

A agitação é um distúrbio psicomotor caracterizado por aumento acentuado nas atividades motoras e psicológicas, muitas vezes acompanhada por uma perda de controle da ação e uma desorganização do pensamento. Este problema é conduzido por situações frequentes na UTI como ansiedade, sensação de medo e confusão mental. Assim, como principais causas da agitação pode-se destacar distúrbios metabólicos, medicamentos, encefalopatia associada à sepse e também a fatores externos como ruído, desconforto, dor, entre outros [CHEVROLET, 2007]. Em contraposição à agitação e sensação de dor manifestada pelo paciente, também pode-se levantar a necessidade de sedação do paciente mediante intervenção médica. Assim, o nível de sedação e o objetivo terapêutico a partir da medicação devem ser rigorosamente ajustados e periodicamente avaliados por meio de pontuações e escalas, a fim de evitar os

perigos de sedação insuficiente ou excessiva [CHEVROLET, 2007]. O estudo de Rahu et al. (2010) apresenta uma revisão da literatura sobre as expressões faciais intimamente relacionadas com a dor de paciente seriamente doentes em estado não-comunicativo. Neste estudo, afirma-se que mais de 50% dos pacientes internados apresentam relatos de dor moderada a severa durante a intubação em estado consciente e em repouso. E, também, 80% dos pacientes relatam dor em ações nociceptivas, como alteração de posição no leito.

Pacientes inconscientes ou sedados não conseguem comunicar seu nível de dor usando escalas numéricas e, portanto, correm o risco de serem inadequadamente medicados. As avaliações imprecisas e o consequente tratamento inadequado da dor em adultos criticamente doentes podem levar a consequências fisiológicas significativas, como aumento da carga de trabalho miocárdica, que pode levar à isquemia miocárdica ou troca gasosa prejudicada, que pode resultar em insuficiência respiratória [RAHU, 2010].

O ponto em que a agitação, ansiedade, dor, delírio e sedação possuem em comum é a presença ou ausência de expressões faciais. Como apresentado por Ekman et al. (1992), as expressões faciais podem ser classificadas a partir de unidade de ação, ou seja, pontos da face característicos na realização do movimento facial. Assim, além de um alto nível de temperatura corporal anormal ser caracterizado como febre, as expressões faciais realizadas por pacientes de unidade de tratamento intensiva podem ser indícios de necessidade de intervenção médica.

Considerando a literatura, a temperatura corporal é um parâmetro essencial na monitoração do estado fisiológico do paciente, investigado na seção 2.3. Atualmente, a literatura apresenta abordagens em domínio infravermelho que exige equipamentos específicos, caros e processamentos offline.

Ao longo de presente trabalho serão abordadas de forma mais aprofundada o estado-da-arte com relação à monitoração de temperatura corporal sem contato e com relação à classificação de emoções faciais. Porém de antemão, é importante ressaltar que segundo a revisão bibliográfica de Molinaro et al. (2020) que apresenta os métodos mais recentes de sinais vitais sem contato, não abordou a temperatura corporal porque expressa que não há padrão-ouro na literatura a respeito do tema. Esse fato apresenta a escassez de estudo com esta temática [MOLINARO, 2020].

Segundo os estudos de Lopes et al. (2017), Minaee et al. (2021), Pu et al. (2021) e Ramis et al. (2022) são apresentadas abordagens de classificação de emoção facial com relevante acurácia porém necessitam de treinamento do algoritmo por redes neurais convolucionais, de uma grande base de dados e de um alto poder computacional, ou seja, na maioria dos estudos o

processamento é realizado de forma offline e demanda arquiteturas complexas [LOPES, 2017; MINAEE, 2021; PU, 2021; RAMIS, 2022].

Assim, as técnicas de aferição de temperatura corporal e classificação de expressões faciais são pouco abordadas em aplicações em tempo-real para o cenário de pacientes não-comunicantes em monitoração de terapia intensiva. O reconhecimento de padrões de doença ou a decisão da intervenção médica necessária pode envolver a coleta de dados a princípio discordantes e sem correlação. A falta de correlação entre os dados faz com que clínicos não tenham conhecimento de informações relevantes por não reconhecer um padrão de risco emergente de fontes de dados díspares. As técnicas acima mencionadas podem ser utilizadas para minimização destes problemas. Como são fontes de extração de dados que podem auxiliar a intervenção médica, neste trabalho propomos um sistema integrado de análise de temperatura corporal e expressões faciais de forma automática com aquisição de dados por câmera RGB com eficiência de desempenho comparável ao estado-da-arte.

## **1.1 HIPÓTESE**

Baseado no estado-da-arte, admite-se a possibilidade de extrair informações por visão computacional sobre a temperatura corporal e expressões faciais com menor quantidade de dados de entrada quando comparado à literatura, permitindo aplicações integrativas, em tempo real, mais robustas e com equipamentos mais acessíveis, como por exemplo, câmeras fotográficas comuns.

## **1.2 OBJETIVO GERAL**

Considerando a importância da implementação de um sistema composto de diferentes sinais adquiridos no monitoramento de pacientes não-comunicativos, o presente estudo tem como objetivo principal criar um framework capaz de propiciar a monitoração de temperatura corporal e classificar expressões faciais de forma automatizada e eficiente a partir da aquisição de dados por câmera RGB, ou seja, sem necessidade de equipamentos caros, específicos e de difícil manipulação.

## **1.3 OBJETIVOS ESPECÍFICOS**

Para a análise de temperatura corporal os objetivos específicos são:



- Avaliar o estado-da-arte;
- Realizar coleta de dados e criar o banco de dados original;
- Desenvolver técnica de pré-processamento de imagens;
- Desenvolver técnica de processamento dispondo de algoritmos de aprendizado de máquina; e
- Avaliar desempenho do método desenvolvido mediante a literatura.

Para a classificação de emoções humanas os objetivos específicos são:

- Avaliar o estado-da-arte;
- Diminuição do volume de dados de entrada;
- Simplificação do treinamento do algoritmo; e
- Avaliação da acurácia alcançada mediante a literatura.

## **1.4 ESTRUTURA DA DISSERTAÇÃO**

A estrutura da dissertação é dividida em cinco principais partes. A primeira parte é apresentado uma revisão da literatura englobando a área de visão computacional, os métodos de aprendizado de máquina sedimentados na literatura, aplicações de interface homem-máquina e os principais estudos atualmente. Na segunda parte da dissertação, o foco é a apresentação da proposta de modelo integrado para monitoração de temperatura corporal e reconhecimento de expressões faciais. Na terceira parte, é realizado a apresentação dos resultados encontrados no presente estudo. E, por fim, é apresentado a discussão e a conclusão geral para o presente trabalho.

## 2 REVISÃO DA LITERATURA

Primeiramente, durante todo o desenvolvimento da pesquisa foi realizado uma revisão da literatura para entender o estágio tecnológico atual associado aos tópicos de análise de temperatura corporal sem contato e classificação de expressões humanas.

### 2.1 VISÃO COMPUTACIONAL

A visão computacional é um tópico de estudo científico e matemático que aborda como os computadores (máquinas) podem obter entendimento de alto nível a partir de imagens ou vídeos digitais. O objetivo desta área de estudo em visão computacional é criar programas (*softwares*) que possam interpretar e analisar imagens, dando ao programa a habilidade de interpretação uma vez que uma imagem é apresentada como dado de entrada [KUDRAYVTSEV, 2020].

Tendo em vista a grande quantidade de dados atualmente, encontramos um elevado custo computacional para processar o grande volume de dados. Neste interím, os avanços na extração de características obteve relevante importância no processamento de imagens. O estado da arte estabelece a visão computacional como sendo a construção de sistemas artificiais que analisam e interpretam qualquer conjunto de dados multidimensionais [ANSARI, 2020].

A área de visão computacional engloba processos como: Classificação de Imagens, Detecção de Objetos, Segmentação de Imagens, Segmentação Semântica, Reconstrução de Imagens, Localização de Objetos e Transferência de Estilos de Imagens. Assim, uma ampla variedade de campos de aplicação pode ser seguida desde o pré-processamento até o algoritmo de treinamento de dados adquirido por visão computacional [BALTRUSAITIS, 2016].

A tecnologia de reconhecimento facial (TRF) encontrou rapidamente um lugar em vários contextos como um meio de reconhecer, verificar, identificar e categorizar entidades únicas, como objeto, pessoa ou organização, graças aos avanços tecnológicos em visão computacional e algoritmos cada vez mais eficientes. A produção em massa de equipamentos de detecção facial relativamente baratos tornou disponíveis aplicativos TRF sofisticados para uso nos setores público e privado [SELWYN, 2022]. Durante a pandemia de Covid-19, o uso de TRF resultou na diminuição do contato humano e da interação social e, assim, limitou a propagação da doença [ZHONG, 2021]. Em comparação com outras tecnologias biométricas, como

reconhecimento de impressão digital, reconhecimento de íris e reconhecimento de voz, a aplicação de TRF é a mais escolhida por ser a mais eficiente [CHEN, 2023].

Com os avanços na pesquisa de visão computacional, tornou-se possível capturar com precisão as emoções de forma automática dos sujeitos-alvo. A TRF tem sido empregada em pesquisas clínicas em áreas como depressão, autismo e esquizofrenia. Atualmente, a TRF é dividida principalmente em extração de recursos artesanais e extração de recursos aprendidos. Recursos artesanais incluem recursos geométricos e de aparência, como recursos de padrão binário local, histograma de recursos de gradientes orientados e recursos de unidades de ação facial. Já os recursos aprendidos é o emprego da inteligência artificial e aprendizado de máquina [LIU, 2023].

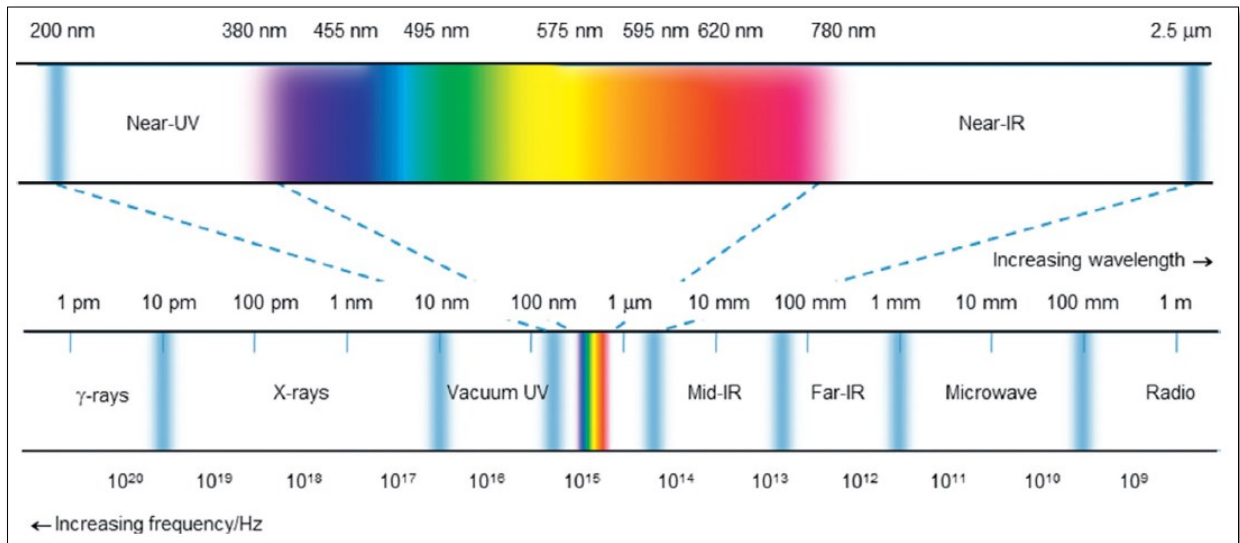
Assim, pode-se utilizar a visão computacional para localização e reconhecimento facial em imagens e trabalhar a partir desta localização diversas informações advindas do indivíduo. Estas informações podem posteriormente virem a ser dados de entrada de algoritmos classificadores cuja finalidade é a prevenção, tratamento, bem como a reabilitação mental, física e social [LIU, 2023; DEBNATH, 2022].

## **2.1.1 CAPTURA E PROCESSAMENTO DE IMAGENS**

### **2.1.1.1 LUZ E COMPRIMENTOS DE ONDA**

Radiação eletromagnética é o termo técnico para luz, sendo não apenas luz visível, mas qualquer radiação com comprimento de onda de 1pm a 1m, variando desde frequências de raios gama ao rádio, conforme apresentado na Figura 1. Como o nome sugere, a luz de todos os tipos é irradiada através de campos elétricos e magnéticos, como mostra Maxwell em meados do século XIX. Para apreciar plenamente a natureza da radiação eletromagnética, no entanto, temos que considerar tanto sua natureza ondulatória quanto fotônica [ANDREW, 2017].

Figura 1 - Espectro eletromagnético.



Fonte: ANDREW et al. (2017).

No caso de radiação monocromática (frequência única), a luz se propaga como uma onda com um comprimento de onda repetido bem definido, conforme descrito na Equação (I). Sendo velocidade da luz no meio estudado ( $v$ ), a velocidade da luz no vácuo ( $c$ ) e o índice de refração ( $n$ ). Viajando na velocidade da luz,  $c$ , essas ondas oscilam a uma frequência característica  $n$  dada por  $c/\lambda$ . Para a luz viajar no espaço livre, o campo elétrico, o campo magnético e a direção de propagação são considerados mutuamente perpendiculares, e por conveniência pode ser escolhido para definir um conjunto de eixos cartesianos  $x$ ,  $y$  e  $z$ , respectivamente [ANDREW, 2017].

$$v = c/n \quad (I)$$

A luz visível, por sua vez, é uma energia em um espectro de comprimentos de onda que os fótons podem sensibilizar as células da retina do olho. Assim, para cada comprimento de onda dentro dessa faixa visível está associada a percepção de uma cor de 370 nm (violeta) a 750 nm (vermelho) [BISI, 2013]. A radiação infravermelha, por sua vez, tem um comprimento de onda acima do vermelho, não sensibilizando as células da retina.

As câmeras, por sua vez, são dispositivos ópticos que conseguem captar desde o domínio infravermelho até o domínio de luz visível a depender da finalidade na qual é utilizada. Atualmente, câmeras podem ser, por exemplo, dispositivos independentes (câmeras fotográficas) e fazem parte de outros equipamentos como computadores, celulares, carros e

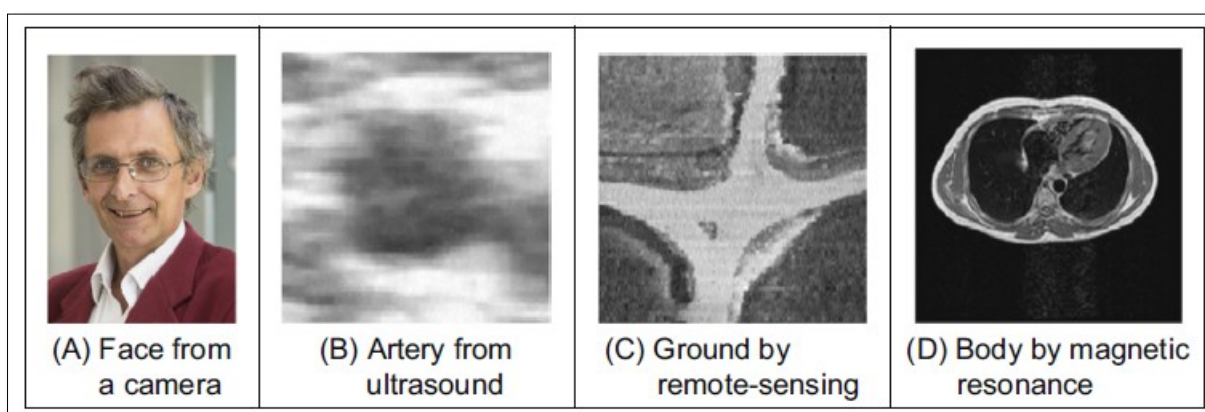
aviões, que registram imagens com pixels alimentados por fótons dentro do espectro de sinal captado pelos sensores eletrônicos [BISI, 2013; MA, 2017].

### 2.1.1.2 AQUISIÇÃO DE IMAGENS E SENSORES

A imagem fotográfica é uma representação digital de um objeto, uma pessoa ou um cenário. A digitalização significa convertê-la em uma série de números e armazená-los em um sistema de armazenamento de dados (sistema de memória não volátil) [ANSARI, 2020]. As imagens estão altamente presentes nos dias atuais em diversos contextos de rotina. Alguns exemplos vão desde a presença de imagens em redes sociais até diagnóstico de doenças por exames de Raios-X, Tomografia e Ressonância Magnética [FARDOULY, 2018; HORSTMAN, 2010].

Tal presença das imagens na rotina do homem moderno torna facilmente a acessibilidade em aquisição de imagens [FARDOULY, 2018]. A Figura 2 abaixo apresentada pelo livro de Mark Nixon, da esquerda (A) até a direita (D), respectivamente, uma fotografia de uma pessoa tomada por câmera comum, uma artéria por ultrassom, foto de satélite e imagem do corpo humano por ressonância magnética [NIXON, 2020].

Figura 2 - Exemplos de imagens por diferentes frequências de radiação eletromagnética. De (A) a (D) respectivamente, domínio da luz visível, ultrassom, ondas de tv e ressonância magnética.

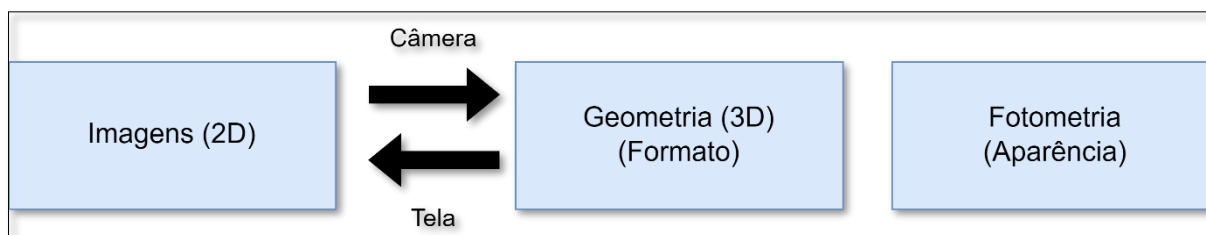


Fonte: NIXON et al. (2020).

A aquisição e a apresentação de imagens segue a lógica de repassar a informação presente em um espaço 3D para um espaço 2D, ou seja, para a tela de um celular, monitor ou displays de dispositivos. Assim, conforme apresentado na Figura 3 é possível verificar que a

partir da fotometria as informações do espaço são registradas e é feita a redução dimensional da informação [KUDRAYVTSEV, 2020]. As imagens são geradas a partir da combinação de uma fonte de “iluminação” e consequente reflexão e absorção de energia dessa fonte pelos elementos da “cena” cuja imagem está sendo gerada [GONZALEZ, 2010].

Figura 3 - Aquisição da câmera para apresentação em duas dimensões.

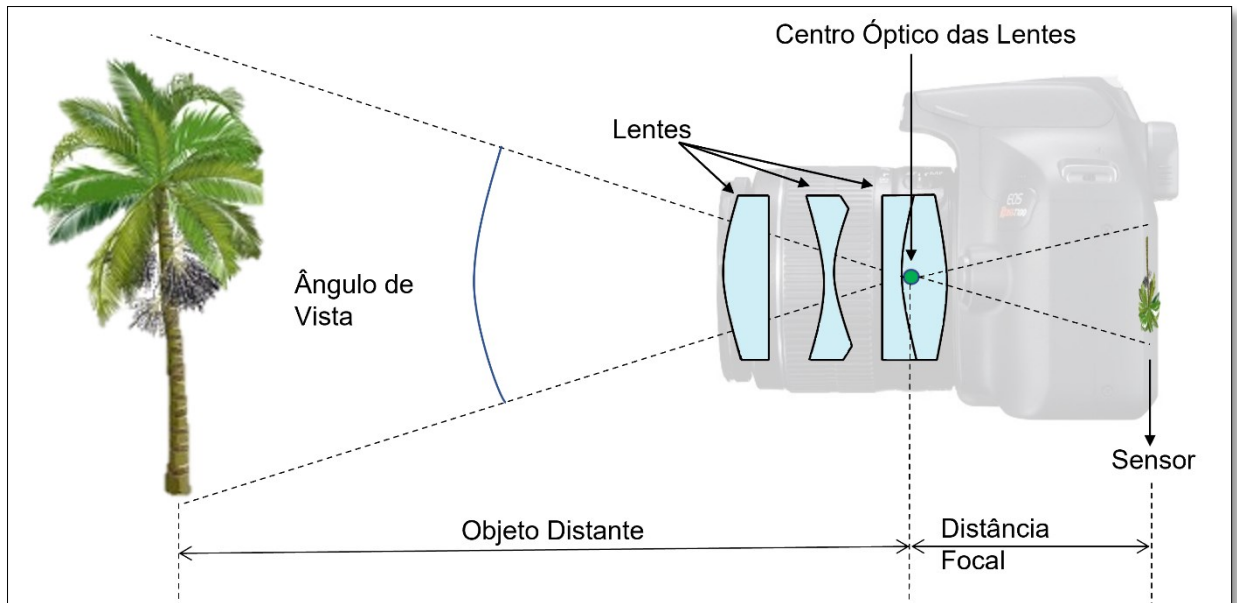


Fonte: Própria.

Dependendo da natureza do material estudado, a energia da fonte é refletida ou transmitida através do material. Como exemplo da reflexão temos a luz refletida em uma superfície plana. Um exemplo da transmissão através do material temos quando raios X passam através do corpo humano para gerar uma imagem radiográfica para diagnóstico. Em exemplificação de outras aplicações, pode-se mencionar a microscopia eletrônica quando a energia refletida ou transmitida é direcionada para um fotoconversor (por exemplo, uma tela fluorescente), que converte a energia recebida em luz visível [GONZALEZ, 2010].

A história da câmera fotográfica começa com a câmara obscura (quarto escuro) que é experimentada no século XVI. Câmara escura é o fenômeno de a luz refletida de objetos passando por um pequeno orifício e então projetada como a imagem do objeto em uma tela do outro lado. Com o desenvolvimento de lentes côncavas e convexas, esse fenômeno é usado em um quarto escuro ou em uma caixa com uma lente como auxílio para desenho e pintura. Somente durante o século XIX as imagens começaram a ser capturadas em material sensível à luz. Após a invenção da eletrônica foram criados sensores e câmeras digitais. O mesmo fenômeno físico é usado em câmeras digitais até atualmente, diferenciando apenas na questão da luz ser projetada em um sensor cuja intensidade de luz, em diferentes partes do sensor, são lidas eletronicamente e armazenadas digitalmente, conforme mostrado na Figura 4 [TANIŞ, 2019].

Figura 4 - Formação da imagem em uma câmera digital.



Fonte: Própria.

A intensidade de cada pixel (número de elétrons) é proporcional a radiação captada pelo sensor segundo a cena observada. Com relação à aquisição da captação da luz em seus diferentes comprimentos de onda, podemos destacar as imagens nos domínios do infravermelho e da luz visível. As imagens no domínio infravermelho, ou também chamadas de imagens termográficas, são construídas a partir da energia de irradiação de comprimentos de onda de 0,9 - 14  $\mu\text{m}$ , sendo assim energia térmica [LARKIN, 2011]. Já as imagens em domínio da luz visível são construídas a partir da irradiação com os comprimentos de onda de 380 -780 nm [NIXON, 2020].

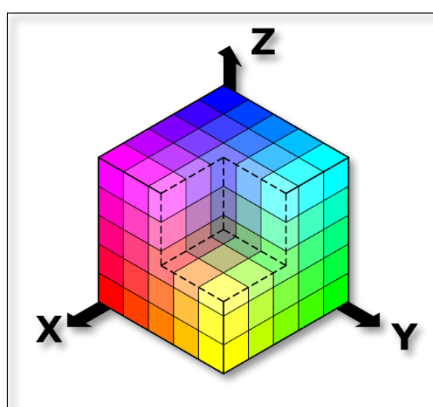
### 2.1.1.3 CARACTERÍSTICAS ESTRUTURAIS DAS IMAGENS

O pixel é a menor estrutura de uma imagem e consiste em uma pequena área de iluminação em uma tela de exibição, uma das muitas a partir das quais uma imagem é composta. Os pixels podem ser apresentados em escala de cinza ou coloridos, onde há presente de mais de um canal de sinal [GONZALEZ, 2010]. Os pixels e os canais utilizados em que a imagem será apresentada ao usuário final e definido pelo espaço de cores. O espaço de cores é uma organização específica de cores. Em combinação com perfis de cores suportados por vários dispositivos físicos [ERNST, 2021].

Alguns exemplos são HSV, YUV, RGB, entre outros. O espaço de cores abordado no presente trabalho é RGB sendo Red (vermelho), Green (verde) e Blue (azul). O modelo RGB

descreve o componente de cromaticidade de uma determinada cor excluindo a variação da luminância [BABU, 2010]. A Figura 5 ilustra tridimensionalmente como é descomposto o modelo de cores em um cubo RGB.

Figura 5 - Espaço de cores em cubo, representação do espaço de cores com valores de intensidade de cores vermelho, verde e azul plotados como coordenadas x, y e z, respectivamente.



Fonte: BABU et al. (2010).

Segunda uma escala de cinza, em uma imagem com um único canal e em 8 bits, cada pixel assume um valor entre 0 e 255. O valor 0 representa o preto e 255 representa o branco. Os valores intermediários são vários tons de cinza. Os valores próximos a 0 são tons de cinza mais escuros e os valores próximos a 255 são tons de cinza mais claros [ANSARI, 2020].

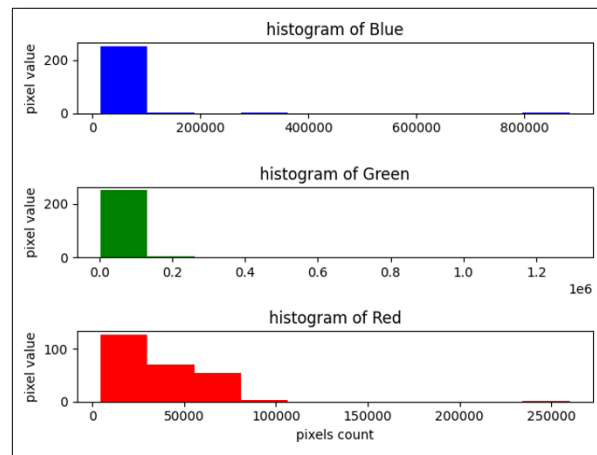
O modelo de cores RGB é uma das representações de cores mais populares de um pixel [ANSARI, 2020]. A maioria dos sistemas de fotopletismografia baseados em câmera (cbPPG) inclui câmeras que fornecem um canal vermelho (R), verde (G) e azul (B) para cada pixel, as chamadas câmeras RGB [ERNST, 2021].

As séries temporais de intensidade dos canais RGB diferem devido a: 1) a composição espectral da iluminação, 2) parâmetros ópticos dependentes do comprimento de onda do tecido da pele e do sangue e 3) as características de sensibilidade espectral específicas do sensor da câmera. No modelo RGB, cada pixel é representado como um conjunto de três valores, geralmente representados da seguinte forma: (valor do componente vermelho, valor do componente verde, valor do componente azul). Cada uma das três cores é representada por números inteiros que variam de 0 a 255 [ANSARI, 2020]. Seguindo alguns exemplos os quais (0,0,0) é uma cor preta, (255,0,0) é uma cor vermelha pura e (0,255,0) é uma cor verde pura.



As estruturas das imagens digitais são matrizes de números que representam intensidades de cores [LAKSHMANAN, 2021]. Uma operação bastante realizada é a análise das propriedades estatísticas da imagem usando um histograma, que mostra essencialmente a distribuição de intensidades em uma imagem. Isso segue perfeitamente a noção de tratar imagens simplesmente como funções (discretas): mapeamentos de coordenadas (x, y) para valores de intensidade. A Figura 6 apresenta um exemplo do histograma de uma imagem.

Figura 6 - Exemplo de um histograma em RGB.



Fonte: Própria

O processamento de imagens é a técnica de manipulação de uma imagem digital para obter uma imagem com realce de detalhes ou extrair informações específicas. Um vídeo é uma série de imagens ou quadros. Portanto, a técnica de processamento de imagem também se aplica ao processamento de vídeo [ANSARI, 2020]. A extração de características é um campo da computação que abrange o fato de que existe um grande conjunto de dados a serem analisados.

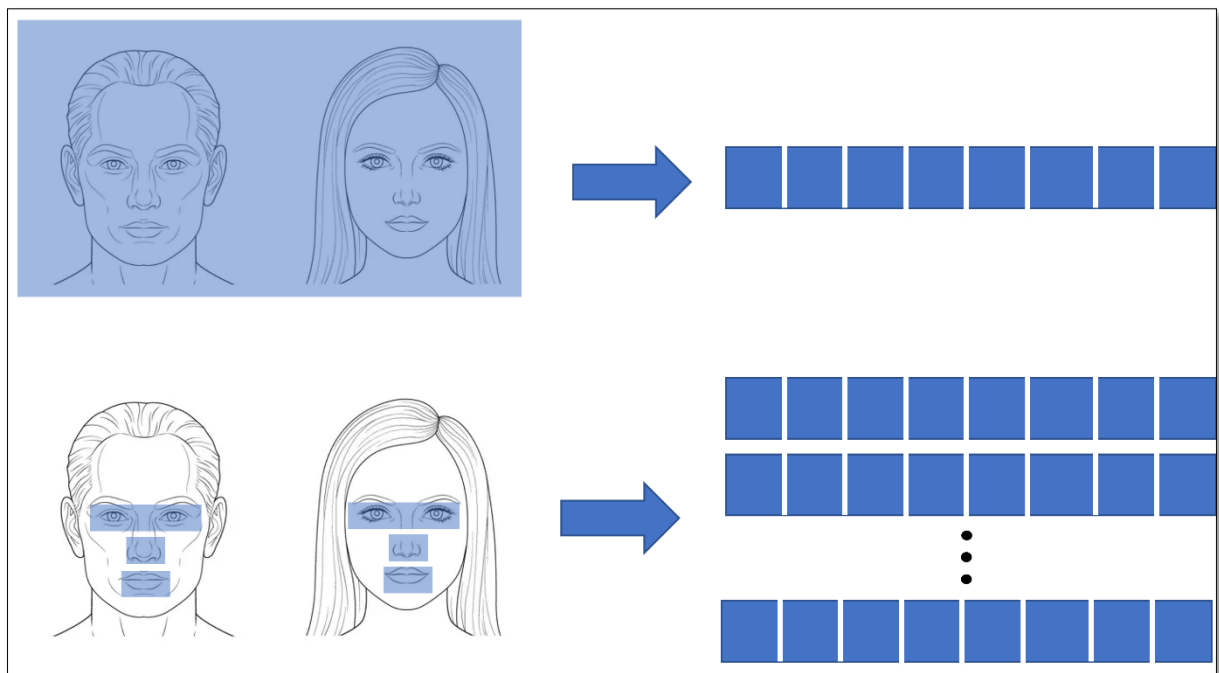
Esta técnica é útil para reduzir o número de recursos, variáveis no total, ou seja, reduzir a dimensionalidade dos dados sem perder nenhuma informação relevante [MA, 2017]. De fato, um conjunto de dados mais enxuto leva a execução de modelos de aprendizado de máquina mais rápidos, aumenta a velocidade de aprendizado do algoritmo e generalização com menor custo computacional e alta precisão [ROSEBROCK, 2017].

Bem como o ser humano extrai informações da imagem a partir do processo visual feito pelo cérebro, a máquina deve realizar e aprender da mesma forma, ou seja, a partir de informações da imagem. Estas informações podem ser categorizadas como características globais e características locais da imagem [AWAD, 2017]. As características globais visam descrever informações da imagem como um todo e são consideradas uma propriedade particular

referente a todos os pixels. Os exemplos de recursos globais são o espaço de cores, histograma, texturas, arestas e, no caso, a possibilidade de aplicação de filtro/kernel pela imagem [AWAD, 2017].

Por outro lado, as características locais são estritamente dependentes das regiões de interesse na imagem desde a detecção ou regiões salientes invariantes à iluminação e mudanças de ponto de vista. Essas estruturas locais extraídas de descritores de recursos locais são consideradas pontos-chave [AWAD, 2017]. A Figura 7 ilustra como os recursos globais ou locais impactam nas estruturas de dados, como por exemplo, número de vetores ou na dimensão dos dados. É apresentada a ilustração que uma imagem completa pode ser representada por um único vetor de dados o que caracteriza um recurso global. Já com relação aos recursos locais é possível que diferentes estruturas da imagem sejam extraídas como diferentes vetores de informações.

Figura 7 - Representação de vetores considerando características globais e locais a partir de uma imagem.



Fonte: Própria.

#### 2.1.1.4 MÉTODOS DE APRENDIZADO DE MÁQUINA

O aprendizado de máquina é um ramo da inteligência artificial (IA) que se concentra no uso de dados e algoritmos para imitar a maneira como os humanos aprendem, melhorando

gradualmente sua precisão [SOORI, 2023]. No que se refere a esta pesquisa, diversas técnicas de aprendizado de máquina podem ser utilizadas em vários sistemas para predição de emoção e/ou temperatura. Assim, descrevemos a seguir os elementos centrais que definem tais técnicas e que são utilizados na pesquisa.

Em geral, os algoritmos de aprendizado podem ser classificados em dois subgrupos: clássicos e baseados em redes neurais, segundo o estudo de Canal et al. (2022) [CANAL, 2022].

As abordagens clássicas são caracterizadas como técnicas que utilizam métodos de regressão, análises probabilísticas, algoritmos capazes de gerar, a partir de dados de entrada, funções de mapeamento em sua saída. Com o desenvolvimento das unidades de processamento gráfico (GPU), muitas arquiteturas de aprendizado de máquina foram aprimoradas com o tempo de execução próximo ao tempo real [CANAL, 2022].

Especificamente para o problema de reconhecimento de emoções a partir de imagens faciais, essas melhorias computacionais também forneceram um novo conjunto de ferramentas, abrindo caminho para o desenvolvimento de várias abordagens baseadas nos métodos clássicos. Uma abordagem considerada clássica bastante utilizada é a Máquina de Vetores de Suporte (SVM) pois corresponde a 40% dos métodos utilizados em reconhecimento de emoções faciais segundo a revisão bibliográfica por Canal et al. (2022) [CANAL, 2022].

As redes neurais artificiais também são muito populares. A arquitetura deste tipo de rede é baseada em redes neurais biológicas que constroem a estrutura do cérebro humano. Semelhante a um cérebro humano que possui neurônios interconectados entre si, as redes neurais artificiais também possuem “neurônios” que são interconectadas por relações ponderadas em várias camadas das redes. Esses neurônios são conhecidos como nós. As arquiteturas de *backpropagation* e *feed-forward* possibilitam a estas arquiteturas de redes um desempenho notavelmente alto. Como exemplo temos as Redes Neurais Convolucionais (CNN) e o modelo Perceptron Multicamadas (MLP) [KUBAT, 2017].

O aprendizado de máquina, atualmente, manipula basicamente objetos chamados de tensores. O tensor, relacionado à matemática, é um objeto algébrico que descreve uma relação multilinear entre dados no espaço multidimensional. Na visão computacional, todas as informações são controladas como arranjos de vetores ou matrizes. Desta forma, o campo investigativo lida com tensores, que são um único vetor ou matrizes de vetores de n-dimensões que expressam todo tipo de informação. Resumindo, o tensor pode ser definido em linha com o construtor de arrays ou vetores como uma lista de listas que constroem dados multidimensionais [ROSEBROCK, 2017].

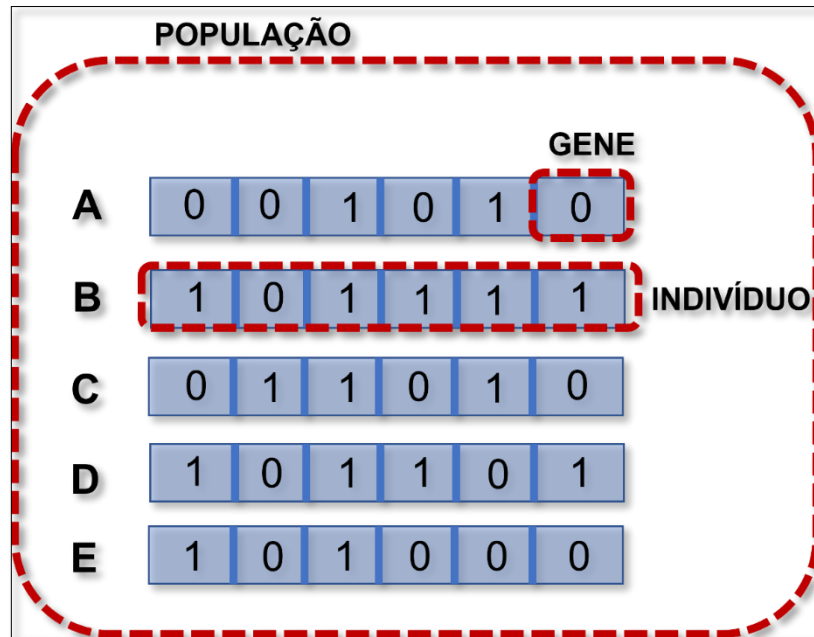
O aprendizado profundo de máquina emergiu em vários campos como processamento de imagens e rotulação de objetos em visão computacional. Este fato é consequência da alta capacidade de modelar equações complexas, utilizando redes neurais, para abstração de problemas de alto nível como semântica e reconhecimento [MA, 2017]. A seguir serão abordados os tópicos sobre algoritmos genéticos, máquina de vetores de suporte, perceptron de múltiplas camadas e regressão de componentes principais uma vez que estas arquiteturas de aprendizado de máquina são utilizadas ao longo do trabalho.

#### ***2.1.1.4.1 ALGORITMOS GENÉTICOS***

O algoritmo genético (AG) é uma meta-heurística inspirada no processo de seleção natural que pertence à uma classe maior de algoritmos chamados algoritmos evolutivos (AE). Algoritmos genéticos são comumente usados para gerar soluções de alta qualidade para otimização e problemas de busca, contando com operadores inspirados biologicamente, como mutação, cruzamento e seleção [MITCHELL, 1998].

Neste tipo de estratégia tem-se uma população de soluções candidatas (chamadas de indivíduos) para um problema de otimização. As populações vão direcionando seus descendentes para aqueles que apresentam as melhores soluções para o problema. Cada solução candidata possui um conjunto de propriedades (seus cromossomos ou genótipo) que podem ser mutados e alterados; tradicionalmente, as soluções são representadas em binário como strings de 0s e 1s [MITCHELL, 1998]. A Figura 8 apresenta a ilustração dos conceitos de população, indivíduo e gene.

Figura 8 - Conceitos de algoritmos genéticos como população, indivíduo e gene.



Fonte: Própria

A evolução geralmente começa a partir de uma população de indivíduos gerados aleatoriamente e é um processo iterativo, com a população em cada iteração chamada de geração. A cada geração, a aptidão de cada indivíduo da população é avaliada; a aptidão é geralmente o valor da função objetivo no problema de otimização que está sendo resolvido, chamada de função fitness. Os indivíduos mais aptos são selecionados estocasticamente da população atual e o genoma de cada indivíduo é modificado (recombinado e possivelmente mutado aleatoriamente) para formar uma nova geração. A nova geração de soluções candidatas é então usada na próxima iteração do algoritmo. Normalmente, o algoritmo termina quando um número máximo de gerações foi produzido ou um nível de aptidão satisfatório foi alcançado para a população [MITCHELL, 1998].

Uma representação padrão de cada solução candidata é como uma matriz de bits, também chamada de conjunto de bits ou sequência de bits. Vetores de outros tipos e estruturas podem ser usados essencialmente da mesma maneira. A principal propriedade que torna essas representações genéticas convenientes é que suas partes são facilmente alinhadas devido ao seu tamanho fixo, o que facilita operações simples de cruzamento. Representações de comprimento variável também podem ser usadas, mas a implementação de cruzamento é mais complexa neste caso.

As representações em forma de árvore são exploradas na programação genética e as representações em forma de gráfico são exploradas na programação evolutiva; uma mistura de

cromossomos lineares e árvores é explorada na programação da expressão gênica. Uma vez que a representação genética e a função de aptidão são definidas, um AG procede para inicializar uma população de soluções e então melhorá-la através da aplicação repetitiva dos operadores de mutação, cruzamento, inversão e seleção [MITCHELL, 1998].

O AG foi utilizado no presente estudo para análise de distâncias euclidianas entre pontos faciais. Assim, o AG exerceu o papel de um seletor de atributos de entrada para aprendizado de máquina. Dependendo do nível de correlação positiva ou negativa dos atributos considerados, a precisão do classificador pode ser melhorada ou prejudicada. O emprego do AG pode nos responder quais são os principais atributos, ou seja, as principais distâncias euclidianas que sejam capazes de nos fornecer a melhor acurácia média entre todas as emoções classificadas.

#### ***2.1.1.4.2 MÁQUINA DE VETORES DE SUPORTE (SVM)***

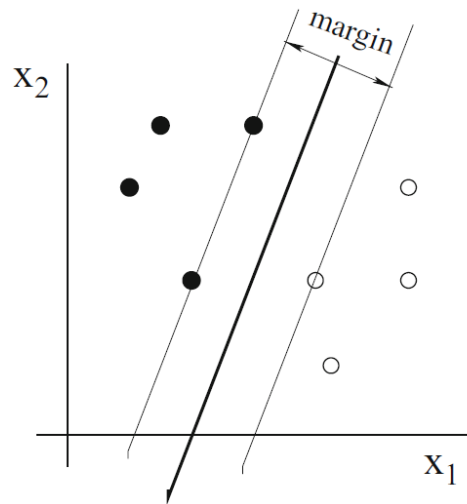
A máquina de vetores de suporte (SVM), por sua vez, é um poderoso modelo de aprendizado de máquina supervisionado, comumente utilizado para problemas de classificação e regressão, bem como para detecção de outliers. As SVMs são particularmente úteis quando há necessidade de agrupamento de dados e dispõem de kernels para mapear pontos de dados em um espaço dimensional superior, permitindo que eles encontrem o hiperplano ideal que pode classificar os dados. Além disso, as SVMs também podem ser usadas para seleção de recursos, o que pode reduzir a dimensionalidade dos dados e melhorar o desempenho.

O reconhecimento de padrões é uma área da ciência cujo objetivo é a classificação de objetos dentro de um número de categorias ou classes e, neste projeto, dar-se-á pelo método SVM, cujo funcionamento constitui como entrada um conjunto de dados e prediz, para cada entrada dada, qual das possíveis classes a entrada faz parte. Tal característica torna o SVM um classificador binário não probabilístico. Dado um conjunto de exemplos de treinamento, cada amostra marcada como pertencente a uma das categorias, o algoritmo de treinamento constrói e/ou reconstrói o modelo que atribui novos pontos de dados a uma das categorias com base em sua semelhança com os pontos de dados classificados anteriormente. Basicamente a máquina de vetor de suporte é uma máquina linear possui como principal fonte de desempenho a construção de um hiperplano de superfície de decisão de tal forma que a margem de separação entre as categorias de informação seja máxima [HAYKIN, 2001].

A técnica das máquinas de vetores de suporte é ilustrada na Figura 9. A linha sólida, denominada como margem, é o melhor classificador. O gráfico mostra também duas linhas mais finas, paralelas ao classificador, cada uma à mesma distância. Pode-se ver que elas

analisam as amostras mais próximas do classificador. Essas amostras são chamadas de vetores de suporte, porque, afinal, cada amostra é um vetor de atributos. A tarefa do aprendizado de máquina é identificar os vetores de suporte que maximizam a margem [KUBAT, 2017].

Figura 9 - Ilustração do classificador SVM considerando a margem que melhor separa grupos de atributos semelhantes.



Fonte: HAYKIN et al. (2001).

Classificadores lineares e polinomiais definem uma superfície de decisão que separa as amostras positivas dos negativos. Especificamente, os classificadores lineares rotulam os exemplos de acordo com o sinal da expressão (II).

$$w_0 + w_1x_1 + \dots + w_nx_n \quad (\text{II})$$

O comportamento do classificador é determinado pelos pesos  $w_i$ . A tarefa para a máquina “aprender” é encontrar valores apropriados para esses pesos. Geralmente, várias épocas de treinamento são necessárias. Em domínios com mais de duas classes, pode-se considerar o uso de um classificador específico para cada classe. Um “classificador mestre” então escolhe a classe cujo classificador teve o maior valor de  $\sum_{i=0}^n w_i x_i$ .

O tipo de aprendizado abordado deste método é o supervisionado, no qual o objetivo é que a representação gerada seja capaz de produzir saídas corretas para novas entradas não apresentadas previamente. Neste caso, dado um conjunto de exemplos rotulados na forma  $(x_i, y_i)$ , em que  $x_i$  representa um exemplo e  $y_i$  denota o seu rótulo, deve-se produzir um classificador, também denominado modelo, preditor ou hipótese, capaz de prever

precisamente o rótulo de novos dados. Esse processo de indução de um classificador a partir de uma amostra de dados é denominado treinamento. O classificador obtido também pode ser visto como uma função  $f$ , a qual recebe um dado  $x$  e fornece uma predição  $y$  [LORENA, 2007].

O SVM implementa um mapeamento, executado por um produto interno kernel escolhido a priori, dos dados de entrada para um espaço de características de alta dimensão, em que um hiperplano ótimo é construído para separar os dados linearmente em duas ou mais classes. Quando os dados de treinamento são separáveis, o hiperplano ótimo no espaço de características é aquele que apresenta a máxima margem de separação [LORENA, 2007].

### 2.1.1.4.3 PERCEPTRON DE MÚLTIPLAS CAMADAS (MLP)

O perceptron de múltiplas camadas (MLP) é uma rede neural artificial popular usada para aprendizado de máquina supervisionado. Consiste em uma camada de entrada, uma ou mais camadas ocultas e uma camada de saída. Cada camada é composta de neurônios, que são conectados uns aos outros de maneira *feed-forward*. O MLP é treinado usando *backpropagation*, ou seja, com retroalimentação. Nesta técnica os pesos das conexões entre os neurônios são ajustados até que a saída do MLP coincida com a saída desejada dos dados de treinamento. Os MLPs podem ser usados para uma variedade de tarefas, como reconhecimento de imagem, análise de sentimento e processamento de linguagem natural.

A função de um neurônio, a unidade básica de um perceptron multicamada, é bastante simples. Uma soma ponderada de sinais que chegam à entrada é submetida a uma transferência função. Várias funções de transferência diferentes podem ser usadas; aquele que é preferido neste capítulo é o chamado sigmóide, definido pela fórmula (III) onde  $\Sigma$  é a soma ponderada das entradas.

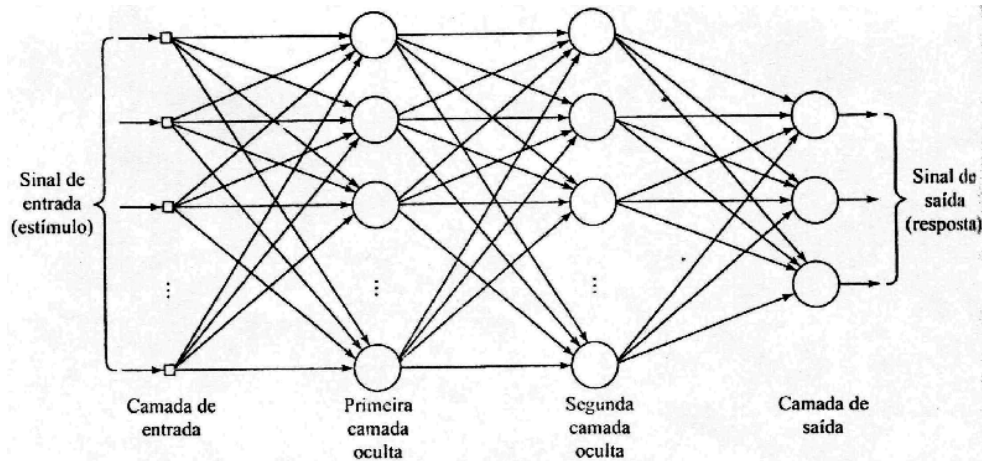
$$f(\Sigma) = \frac{1}{1 + e^{-\Sigma}} \quad (\text{III})$$

A função de transferência é usada para transformar a entrada em uma saída desejada. Ela é usada para transformar a entrada em um valor de saída que é proporcional à sua importância. É geralmente uma função não linear, como a função sigmoideal. A função de transferência ajuda a regular a saída do neurônio, permitindo que a rede aprenda a partir dos dados de entrada. A rede neural na Figura 10 é conhecida como perceptron de multicamadas. Os neurônios, representados por nós, estão dispostos em camadas de saída e camadas ocultas.



Para simplificar, encontra-se a camada de entrada, seguida por duas camadas ocultas e por fim a camada de saída retornando a resposta do classificador [KUBAT, 2017].

Figura 10 - Arquitetura do modelo MLP.



Fonte: HAYKIN et al. (2001).

O perceptron de múltiplas camadas é um conjunto de unidades denominados nós de fonte que constituem a camada de entrada, uma ou mais camadas ocultas de nós computacionais e uma camada de saída de nós computacionais. O sinal de entrada se propaga para frente através da rede, camada por camada. Este tipo de arquitetura tem sido aplicado a problemas e através do treinamento de forma supervisionada, com o método popular conhecido como retropropagação de erro (error back-propagation). Este método é baseado na regra de aprendizagem por correção de erro, generalização do algoritmo de filtragem adaptativa mínimo quadrado médio (LMS) para o caso especial de único neurônio linear [HAYKIN, 2001].

Basicamente, a aprendizagem por retropropagação de erro consiste de dois passos: a propagação (passo a frente) e a retropropagação (passo para trás). Durante a propagação, um padrão de informações (vetor de entrada) é aplicado aos nós sensoriais da rede e seu efeito se propaga através da rede, camada a camada. Então, um conjunto de saídas é produzido como resposta real da rede. Durante a propagação, os pesos sinápticos da rede são valores fixos [HAYKIN, 2001].

Durante a retropropagação, em contra-ponto, os pesos sinápticos são ajustados de acordo com a regra de correção de erro. Especificamente, a resposta real da rede é subtraída de uma resposta desejada (alvo) para produzir o sinal de erro. Este sinal de erro encontrado é propagado através da rede em direção oposta às conexões sinápticas. Assim, os pesos sinápticos

são ajustados para fazer com que a resposta real da rede se mova cada vez mais próxima à resposta desejada a cada iteração [HAYKIN, 2001].

Assim, a retropropagação tem a importância de treinar a rede e ajustar os pesos a partir do feedback dos resultados das saídas ao longo das iterações. Considerando o estudo de Canal et. Al (2022), a literatura utiliza para classificações de emoções humanas algoritmos tanto considerados clássicos, baseados em regressão, como algoritmos baseados em redes neurais [CANAL, 2022].

#### **2.1.1.4.4 REGRESSÃO DE COMPONENTES PRINCIPAIS (RCP)**

Por fim, a regressão de componentes principais é uma combinação de análise de componentes principais (ACP) somado ao método de regressão linear múltipla. Formalmente, a ACP é uma técnica estatística para reduzir a dimensionalidade dos dados, compressão de dados, extração de características e visualização de dados. Realizado pela transformação linear dos dados em um novo sistema de coordenadas onde (a maior parte) da variação nos dados pode ser descrita com menos dimensões do que os dados iniciais. Muitos estudos usam as duas primeiras componentes principais para apresentar os dados em duas dimensões e para identificar visualmente grupos de pontos de dados intimamente relacionados [HAYKIN, 2001].

Para um conjunto de dados, cada componente principal é tratada como um vetor de dados. A quantidade de autovetores determina a dimensões de dados com que se está manipulando, esta quantidade de vetores é proporcional a quantidade de variáveis independentes. Para cada autovetor tem-se um autovalor, o autovetor determina a direção no plano espacial e o autovalor determina o valor da variância que há nos dados em sua determinada direção. Para a seleção das componentes principais deve-se considerar qual o número de componentes necessário que atingem o mais próximo de 100% de toda a variância dos dados originais, ou seja, variância acumulada das variáveis independentes. Por este fato, a ACP é uma técnica de redução dimensional de dados [HAYKIN, 2001].

Dado um vetor aleatório (I) cada variável independente é uma variável aleatória. Calcula-se então o vetor média  $\mu$  de  $\bar{X}$  (IV) e a matriz de covariância  $V$  (V) entre todas variáveis independentes  $N$ . A direção de maior variabilidade dos dados é dada pelos autovetores da matriz de covariância. Assim, os autovetores são ordenados de forma decrescente a partir de seus respectivos autovalores  $k$ . Assim, a ACP é a solução de um sistema de equações lineares que a partir da estimativa de autovetores da matriz de covariância produz um conjunto de

vetores ortogonais entre si (componentes principais) no sentido de maior variabilidade dos dados e menor erro médio quadrático.

$$\vec{X} = (X_1, X_2, \dots, X_n) \quad (\text{IV})$$

$$V = \begin{bmatrix} \sum \frac{x_1^2}{N} & \sum \frac{x_1 x_2}{N} & \dots & \sum \frac{x_1 x_n}{N} \\ \sum \frac{x_2 x_1}{N} & \sum \frac{x_2^2}{N} & \dots & \sum \frac{x_2 x_n}{N} \\ \dots & \dots & \dots & \dots \\ \sum \frac{x_n x_1}{N} & \sum \frac{x_n x_2}{N} & \dots & \sum \frac{x_n^2}{N} \end{bmatrix} \quad (\text{V})$$

A regressão linear é um método usado para prever o valor de uma variável com base no valor de outra variável. A variável que se deseja prever é chamada de variável dependente. A variável utilizada para prever o valor da outra variável é chamada de variável independente [HAYKIN, 2001]. Essa forma de análise estima os coeficientes da equação linear, envolvendo uma (regressão linear simples) ou mais variáveis independentes (regressão linear múltipla) que melhor predizem o valor da variável dependente. A regressão linear ajusta uma linha reta ou superfície que minimiza as discrepâncias entre os valores de saída previstos e reais [HAYKIN, 2001].

O modelo de regressão múltipla possui a forma geral conforme a equação (VI) inclui variáveis independentes  $x$  para predição de uma variável dependente  $y$ . O modelo de regressão múltipla pode se caracterizar linear ou polinomial. O modelo simples apresenta os coeficientes  $\beta$  que acompanham a modelagem da variáveis independentes sempre de primeira ordem. Para modelagem polinomial os coeficientes  $\beta$  podem ser de segunda ou maiores ordens.

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k + \epsilon \quad (\text{VI})$$

A regressão de componentes principais (RCP) é uma técnica de regressão que atende ao mesmo objetivo da regressão linear padrão, ou seja, modelar a relação entre a variável de dependente e as variáveis independentes. O foco da RCP é a utilização de autovetores (componentes principais) selecionados a partir da ACP como variáveis preditoras para análise de regressão em vez dos recursos originais. Portanto, ao invés de considerar todos os recursos

originais para regressão, utiliza-se apenas o subconjunto dos componentes principais que melhor representam a variabilidade dos dados.

## 2.2 EXPRESSÕES FACIAIS E EMOÇÕES HUMANAS

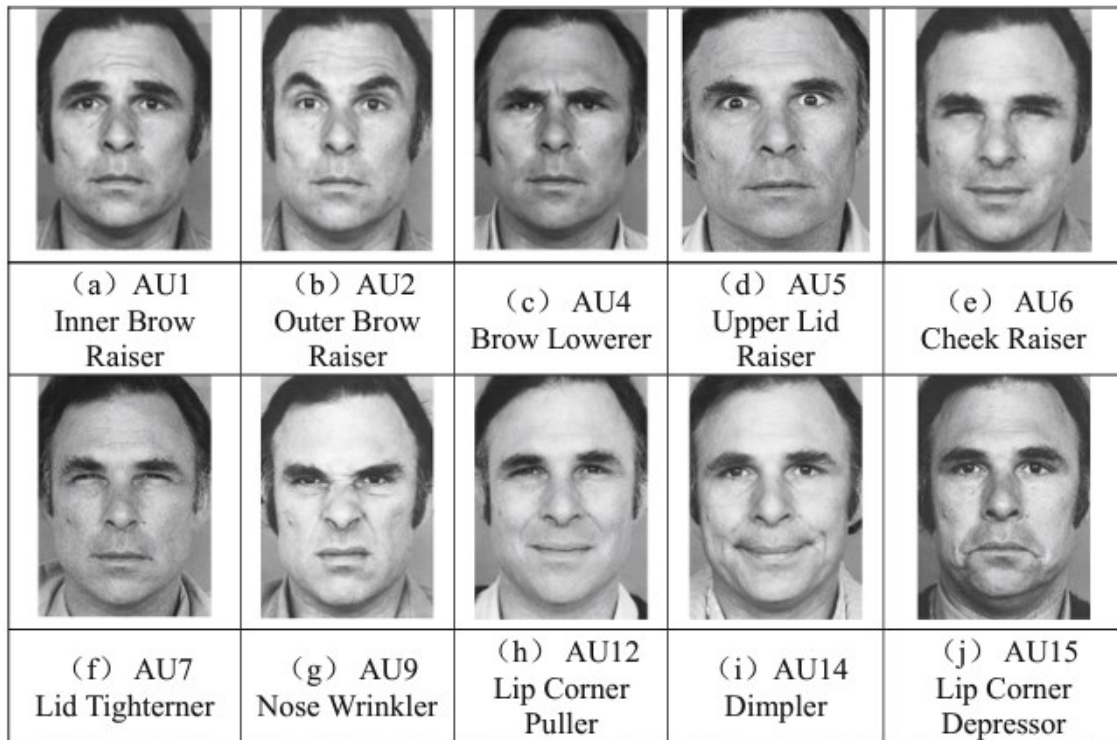
As emoções são caracterizadas como sensações físicas ou emocionais sentidas pelo ser humano, provocadas por algum estímulo, como um sentimento ou um acontecimento. Reconhecer e interpretar a emoção é natural para a interação humana, pois ela desempenha um papel importante na comunicação não-verbal. Este tipo de comunicação é caracterizado principalmente pela transmissão de mensagens sem a necessidade da fala ou da escrita [KUMAR, 2021; SAJJAD, 2020].

A partir da interação social, o indivíduo receptor deve ter a habilidade de reconhecer a expressão facial, interpretar o seu significado, entender a sua motivação e responder a ela. O rosto humano expressa informações sobre sentimentos, sobre o estado mental e emocional, de modo que as informações capturadas a partir da face podem ser úteis em várias aplicações a partir de interfaces homem-máquina com ferramentas de visão computacional como ciência comportamental, educação, segurança e na área da medicina [SLIMANI, 2018; THUSEETHAN, 2022].

O importante psicólogo Paul Ekman, em 1992, padronizou 6 expressões faciais básicas a partir do seu estudo em Papua Nova Guiné em uma tribo Fore. Pelo fato de esta tribo nunca ter tido contato com outra civilização, Ekman registrou as expressões faciais realizadas a partir de diferentes acontecimentos cotidianos da tribo. Estas emoções são instintivas e reconhecidas pela grande maioria das pessoas, as quais são alegria, tristeza, raiva, medo, nojo e surpresa [EKMAN, 1992].

Além da investigação sobre as expressões faciais das emoções, Ekman e o pesquisador Friesen criaram o manual de análise da Face Humana chamado de Sistema de Código de Ação Facial (SCAF) o qual foi publicado pela primeira vez em 1978 por Ekman e Friesen e foi revisado mais recentemente em 2003. Este manual consiste em uma metodologia validada cientificamente para análise e medição dos movimentos faciais em seres humanos [EKMAN, 1992]. É possível verificar que há enumerações para diferentes microexpressões conforme exemplo apresentado na Figura 11.

Figura 11 - Apresentação de algumas unidades de ação (UA) facial de acordo com movimentos faciais estabelecidos por Paul Ekman.



Fonte: EKMAN et al. (1992).

De acordo com o manual de unidades de ativação, há combinações de AUs as quais estão presentes nas sete expressões básicas estudadas. Estas combinações estão descritas na Tabela 1.

Tabela 1 - Relação da emoção humana com as respectivas unidades de ativação estabelecido por Paul Ekman.

<b>Emoção</b>	<b>Unidades de ativação</b>
Felicidade	AU6 + AU12
Tristeza	AU1 + AU4 + AU15
Surpresa	AU1 + AU2 + AU5B + AU26
Medo	AU1 + AU2 + AU4 + AU5 + AU7 + AU20 + AU26
Raiva	AU4 + AU5 + AU7 + AU23
Nojo	AU9 + AU15 + AU17

No entanto, de acordo com vários estudos, observa-se que a precisão do reconhecimento de emoções é baixa considerando ruídos na análise como várias pessoas em uma mesma

imagem, baixo dimensionamento do rosto e baixa qualidade da imagem [THUSEETHAN, 2022]. Uma das principais causas da baixa capacidade de reconhecimento é negligência da pose corporal. No entanto, numerosos estudos no passado revelaram que a pose do corpo e o cenário da foto também podem ser usados para estimar as emoções, pois elas geralmente afetam os estados emocionais [THUSEETHAN, 2022].

Os músculos faciais, por sua vez, trabalham juntos para controlar as partes que compõem a face humana. Os músculos da face podem ser decompostos em subáreas faciais sendo bochechas, queixo, sobrancelhas, testa, lábios e nariz. O rosto apresenta aproximadamente 20 músculos esqueléticos planos que se ligam a diferentes lugares do crânio. Os músculos se originam do osso ou da fáscia e se inserem na pele. A Figura 12 apresenta uma visão frontal dos músculos superficiais da face realizando as 7 emoções básicas e posição neutra da face [ZARINS, 2017].

Figura 12 - Anatomia da expressão.



Fonte: ZARINS et al. (2017).

Assim, musculatura craniofacial possui duas principais funções realização de ações e expressões. Ações como mastigar, conversar, cantar, beber líquidos e expressões faciais como sorrir ou levantar as sobrancelhas de surpresa [ZARINS, 2017]. Os músculos faciais estão localizados em todo o rosto e podem ser categorizados mediante sua localização:

- Músculos bucolabiais, dentro e ao redor da boca.
- Músculos nasais, ao redor do nariz.
- Músculos epicranianos da testa, crânio e pescoço.
- Músculos auriculares, ao redor das orelhas.
- Músculos orbitais, ao redor dos olhos.

Para realização de movimentos, os músculos faciais recebem sinais do cérebro por meio do nervo facial. A assimetria muscular do rosto pode expor condições ou comprometimentos decorrentes de sequelas após algumas doenças. Portanto, quando os músculos faciais não conseguem receber sinais cerebrais adequadamente, isso pode causar parilisia facial, dificuldades em realizar ações e expressões [ZARINS, 2017]. A assimetria facial é mais um fator para diagnóstico médico para doenças que comprometem o nervo facial e causam problemas aos músculos da face: doenças autoimunes (Síndrome Guillain-Barré ou Esclerose Múltipla), Acidente Vascular Cerebral (AVC), Infecções, entre outras.



Rahu et. Al (2010) apresenta que a avaliação da dor é um desafio significativo em adultos gravemente doentes, especialmente naqueles que não conseguem comunicar o nível de dor. Infelizmente, nos cuidados intensivos, muitos fatores alteram a comunicação verbal com os pacientes, incluindo intubação traqueal, redução do nível de consciência, sedação e administração de drogas paralisantes [RAHU, 2010]. Portanto, é importante uma avaliação precisa dos comportamentos não verbais de dor, como a expressão facial, especialmente em pacientes críticos. Assim, este estudo caracteriza as unidades de movimento facial do SCAF que compõem a sensação de dor, conforme apresentado na Figura 13.

Figura 13 - Relação de unidades de ação faciais com a sensação de dor pelo SCAF.



Facial expression correlated with pain using the Facial Action Coding System	
Description	Action Unit
Brow lower	AU4
Cheek raised	AU6
Lid tightened	AU7
Nose wrinkle	AU9
Upper lip raiser	AU10
Lip corner puller	AU12
Lip stretcher	AU20
Lip presser	AU24
Lips parted	AU25
Jaw drop	AU26
Mouth stretched	AU27
Eyes closure	AU43
Blink	AU 45

	
Natural Expression	Pain Expression

Fonte: Rahu et al. (2010).

### 2.2.1 MÉTODOS DE RECONHECIMENTO DE EXPRESSÕES FACIAIS

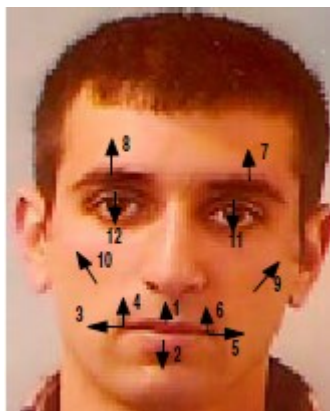
Durante os últimos anos, os pesquisadores têm convergido suas atenções para o reconhecimento de expressões faciais nas áreas de processamento de imagens e visão computacional, onde as expressões faciais são consideradas um meio significativo para transmitir o estado emocional do indivíduo, para comunicar suas emoções, propósitos e caracterizar seu comportamento [SLIMANI, 2020].

Cohen et al. (2003) apresenta o desenvolvimento de um sistema com uma rede de classificadores neurais bayesianos para reconhecimento de expressões rotuladas e não-rotuladas. Este estudo utilizou o banco de dados CK e posteriormente validou o sistema em um banco de dados de imagens faciais de paciente internado na UTI. O estudo utilizou uma ferramenta de rastreamento facial chamado Rastreador de Deformação de Volume de Bézier (RDVB). E como entrada dos seus classificadores extraiu a mudança de direção de vetores



faciais a partir das unidades de movimento estabelecidas por Paul Ekman, conforme apresentado na Figura 14 [COHEN, 2003].

Figura 14 - Vetores de medição de movimentos faciais.



Fonte: Cohen et al. (2003).

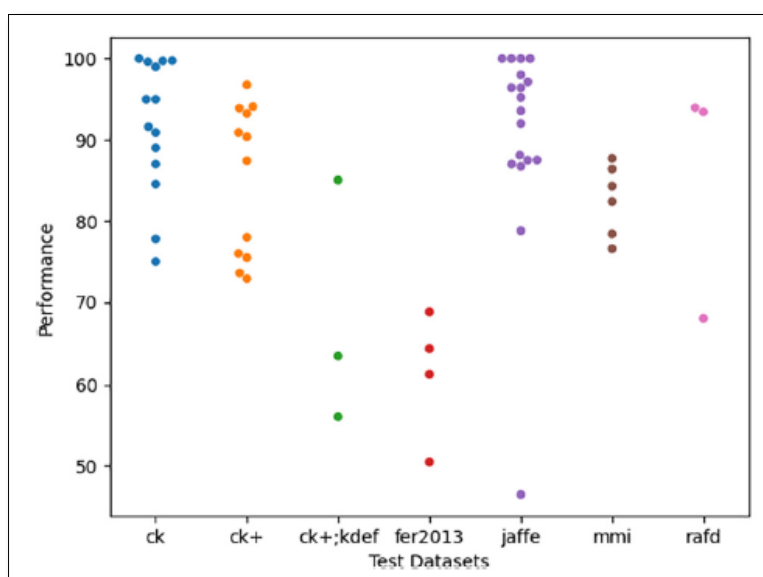
Os estudos apresentados a seguir são relacionados a técnicas mais atuais a partir de reconhecimento de emoções faciais em ambientes cotidianos. O estudo de Sajjad et al. (2020) apresenta que no estado-da-arte há diversos métodos propostos para REF em imagens e vídeos. Os métodos mais utilizados são classificadores como máquina de vetor de suporte (SVM), K-vizinhos mais próximos, árvore de decisão, redes neurais, rede bayesiana e classificadores baseados em regras. O foco principal da literatura a partir das técnicas mencionadas é fornecer uma relação ótima entre o tempo de processamento e a precisão do algoritmo [SAJJAD, 2020].

O principal problema atualmente diante dos métodos de reconhecimento de expressões faciais na literatura é que utilizam demasiadamente treinamentos de classificadores cujas entradas apresentam dados bidimensionais ou com mais dimensões. Todo este volume de dados gera uma impossibilidade em aplicações mais rápidas e eficientes, e exigindo assim, equipamentos potentes para utilização final pelo usuário. Como os métodos de treinamento e reconhecimento facial necessitam de uma grande base de dados inicial, atualmente os projetos não apresentam um estudo mais focado ao usuário final [SLIMANI,2018; KAUSHIK, 2020].

O estudo de Li et al. (2020) apresenta um sistema de redes neurais de convolução para reconhecimento automático de emoções composto por etapas desenvolvidas para extração de características, realce de informação, reconstrução de informação e classificação de informação. Alguns dos conjunto de dados utilizados foram o CK+ e o JAFFE. Para o CK+, o sistema desenvolvido atingiu acurácia média de 98,68% e para o JAFFE alcançou 98,52%. Não há citações para aplicações em tempo-real [LI, 2020].

O estudo de Canal, 2022, apresentou uma revisão do estado-da-arte a respeito de técnicas de reconhecimento de emoção facial. Ao todo foram considerados 51 artigos e a verificação de quase 100 métodos distintos [CANAL, 2022]. As acurácias dos métodos analisados estão na faixa de 80 a 100% de acurácia média para 6 classes de emoções básicas anteriormente citadas considerando 7 bancos de dados diferentes conforme apresentados na Figura 15.

Figura 15 - No eixo X estão os bancos de dados analisados e no eixo Y acurácias alcançadas na literatura.



Fonte: Canal et al. (2022).

Os estudo de Lopes et al. (2017), Minaae et al. (2021), Pu et al. (2021 e Ramis et al. (2022) utilizaram arquiteturas convolucionais próprias, ou seja, técnicas de compressão e reconstrução de dados que podem ser demasiadamente lentas no treinamento do algoritmo e/ou na classificação dos dados pela rede [LOPES, 2017; MINAE, 2021; PU, 2021; RAMIS, 2022]. Já os estudos de Yao et al, Drivekar et al e Liu utilizaram análises clássicas em que se utiliza métodos estatísticos. Yao et al. (2021) utilizou aprendizado ativo para detectar unidades de ação faciais e classificação por máquina de vetor de suporte [YAO, 2021]. E Drivekar et al. (2021) utilizou descritores locais para selecionar regiões faciais de interesse e algoritmos de classificação como floresta randômica, máquina de vetor de suporte e o classificador XGBoost [DIVEKAR, 2021].

Em resumo, com relação ao reconhecimento de emoções faciais, a Tabela 2 apresenta o estudo citado, o método de extração de característica, o método aprendido de máquina

utilizado, a base de dados utilizada e a acurácia média apresentada pelo estudo. Pode-se verificar que as acurácias médias, entre as emoções básicas mais a emoção neutra, estão acima de 80%.

Tabela 2 - Apresentação do estado-da-arte com alguns estudos mais relevantes sendo nome do primeiro autor, ano de publicação, dados de entrada, técnica de classificador, base de dados e a média da acurácia.

<b>Autores</b>	<b>Ano</b>	<b>Dados de Entrada</b>	<b>Classificador</b>	<b>Base de Dados</b>	<b>Acurácia Média</b>
Cohen et al.	2003	Unidades de Ação	Bayesiano	CK	81,80%
Lopes et al.	2017	Imagens	CNN	CK+	96,76%
Li et al.	2020	Unidades de Ação	Descritores Locais	CK+	98,68%
Divekar et al.	2021	Pontos de Interesse	SVM	ADFES and WSEFEP	95%
Minaae et al.	2021	Imagens	CNN	CK+	98%
Pu et al.	2021	Imagens	CNN	RAF-DB	81%
Yao et al.	2021	Unidades de Ação	SVM	CK+	93,42%
Ramis et al.	2022	Imagens	CNN	CK+	93,23%
Liu et al.	2023	Unidades de Ação	Correlação	FERPlus	91,5%

## 2.3 TEMPERATURA CORPORAL

Diante das informações vitais relevantes para a área da saúde, podemos destacar a temperatura corporal (TC), destacar pressão arterial (PA), frequência cardíaca (FC), frequência respiratória (FR) e saturação de oxigênio (SpO2) [HARFORD, 2017; HASAN, 2021]. A revisão da literatura atual aponta que os principais estudos desenvolvidos foram voltados para o monitoramento da FC, PA, FR e SpO2, por meio de análise de imagem [HASSAN, 2020; ERNST, 2021; CHEN, 2021].

A homeostase corporal denota a manutenção, ou regulação, de variáveis corporais internas em um estado de relativa constância. Homeostase é uma característica de todos os organismos que possuem ambientes internos, desde o citoplasma de indivíduos células para

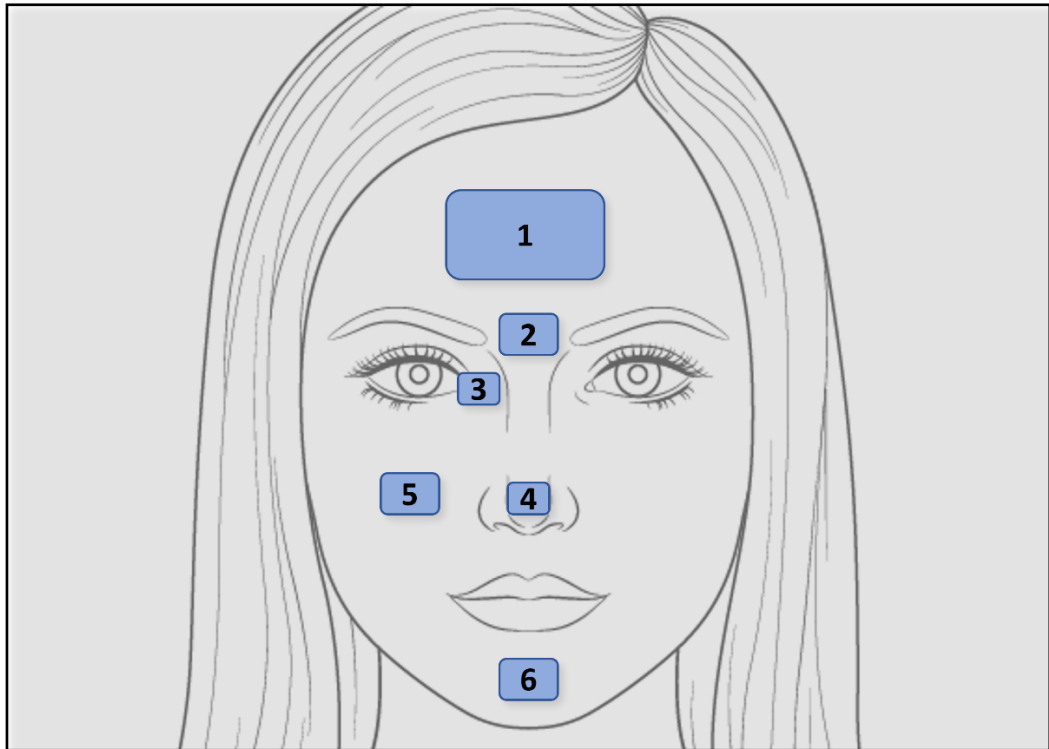
organismos unicelulares até para os múltiplos ambientes internos "sistêmicos" dos mamíferos [ASARIAN, 2012]. Assim, definimos como temperatura de homeostase a condição de equilíbrio entre todos os sistemas do corpo necessários para que o corpo sobreviva e funcione corretamente.

O estudo de Protsiv et al. (2020) apresenta uma revisão da temperatura corporal central ao longo das décadas na raça humana e na linha do tempo de vivência desde o nascimento. Este artigo apresenta que a temperatura média corporal é de 36,5 °C sendo considerada sob normalidade, em um intervalo de confiança de temperatura corporal entre 36,0–37,5 °C [PROTSIV, 2020]. Neste estudo, temperaturas corporais de todos os três conjuntos de dados, quaisquer valores extremos de temperatura (<35 °C e >39 °C) foram excluídos da análise porque eram implausíveis ou porque indicavam o diagnóstico de febre. Xuan et al. (2021) afirma que a temperatura nas regiões localizadas profundamente no corpo apresenta em normalidade dentro de variações no intervalo de 36,5 – 38,5 °C [XUAN, 2021].

Com relação à termorregulação da temperatura corporal, a temperatura da pele está em constante contato com a temperatura ambiente e serve como um sinal de retroalimentação para o sistema de termorregulação. Os mamíferos apresentam então um mecanismo de autorregulação que independente da temperatura do meio externo, o sistema biológico está restrito a um intervalo de temperatura para seu correto funcionamento. Se a temperatura do seu corpo ficar muito fria ou quente, isso pode levar a sintomas graves e até à morte [ROMANOVSKY, 2014].

De acordo com a literatura é importante ressaltar regiões tanto faciais quanto corporais específicas que estão constantemente sob análise dos experimentos apresentados no estado-da-arte. Conforme apresentado na Figura 16, consegue-se verificar ao todo 6 pontos da região facial sendo eles enumerados de 1 a 6. Em ordem crescente temos a região da testa, glabella, canto interno dos olhos, nariz, bochechas e queixo.

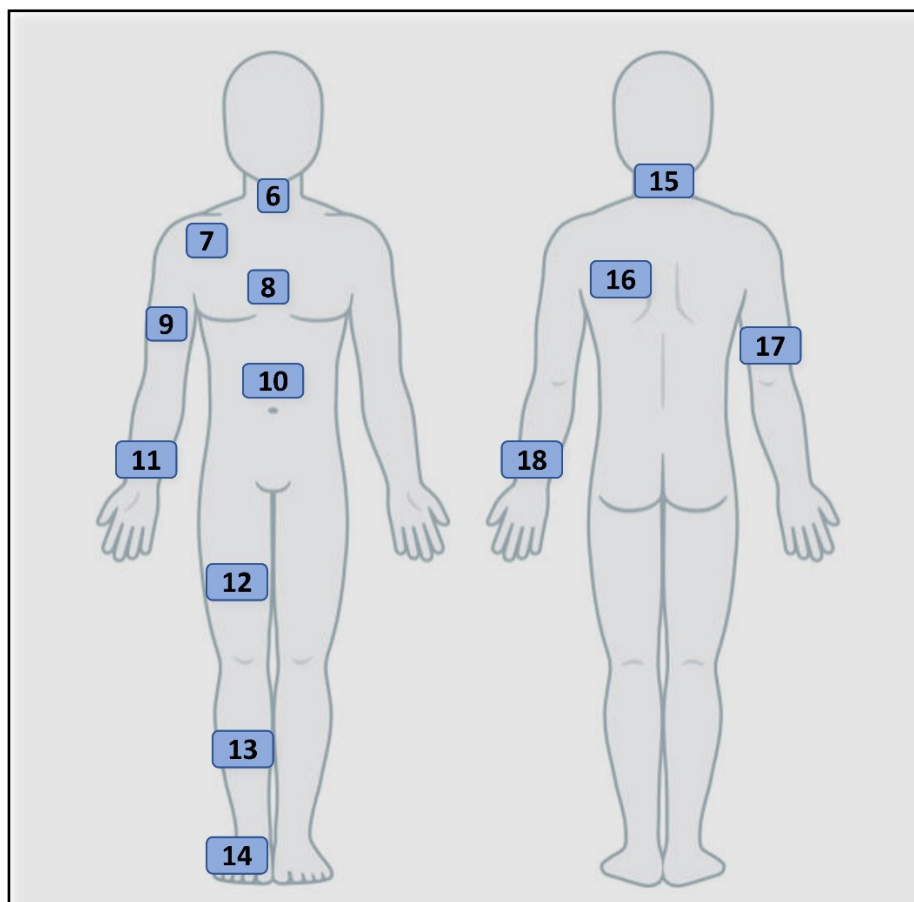
Figura 16 - Regiões faciais comumente analisadas na literatura.



Fonte: Própria

De acordo com as regiões corporais, conforme apresentado na Figura 17, podemos verificar mais 12 pontos também utilizados em estudos de medição de sinais vitais sem contato. São estes em ordem crescente de enumeração no plano frontal anterior: pescoço, região subclavicular, tórax, antebraço, abdômen, braço (punho), coxa, região da tíbia e pé. No plano frontal posterior tem-se a região do pescoço, região da escápula, antebraço e braço (punho).

Figura 17 – Regiões corporais comumente analisadas na literatura.



Fonte: Própria

### 2.3.1 TÉCNICAS PARA MONITORAMENTO DE SINAIS VITAIS SEM CONTATO

Molinaro et al. (2022) realizaram uma revisão bibliográfica que fornece uma visão geral dos aspectos fisiológicos e das técnicas atualmente utilizadas para a predição de sinais vitais sem contato como frequência respiratória, frequência cardíaca, saturação de oxigênio no sangue e pressão arterial a partir de imagens digitais em domínio visível, bem como os campos potenciais de aplicação dessas tecnologias. Este estudo apresenta que a temperatura corporal não foi abordada em sua revisão porque não há evidências literárias de padrão ouro para medição de temperatura corporal no domínio de luz visível [MOLINARO, 2022]. Por este fato, como presente trabalho aborda predição de temperatura corporal em visível os estudos encontrados no estado-da-arte foram divididos pelo foco da pesquisa sendo:

1. Monitoramento de temperatura corporal
2. Monitoramento de sinais vitais sem contato em domínio visível

Para cada categoria é apresentado a seguir uma tabela com informações específicas sobre cada campo de pesquisa e seus respectivos obstáculos. Sendo assim, com relação à primeira categoria “Monitoramento de temperatura corporal” apresenta-se a Tabela 3. Estes estudos abrangem análise de temperatura corporal em domínio infravermelho.

O estudo de James et al, em 2014, apresentou como objetivo a investigação da validade e da confiabilidade das medições de temperatura da pele usando um termistor de telemetria e uma câmera térmica durante o exercício em um ambiente quente. Este estudo concluiu que podem ser usadas como ferramentas úteis imagens de indivíduos estáticos para medir a temperatura da pele em ambientes controlados, porém seu uso não é recomendado para monitoramento de indivíduos durante a atividade de exercício físico [JAMES, 2014].

Em Cosma et al. (2018) foi avaliado o uso de câmeras termográficas como um método não invasivo para modelar automaticamente conforto térmico em condições ambientes de temperaturas transitórias. Usando dados de 30 indivíduos saudáveis testados em uma configuração de ambiente com temperaturas entre 21,11 °C e 27,78 °C. As variáveis temperatura ambiente, umidade relativa, quantidade de porções de pele exposta a temperatura (tipo de vestimenta) foram medidas automaticamente em aproximadamente 27 minutos por voluntário, usando sensores remotos e evitando qualquer contato com os voluntários [COSMA, 2018].

Os níveis de conforto térmico foram avaliados usando feedback dos indivíduos, registrado a cada minuto durante todo o experimento. A temperatura média da pele foi adquirida de cinco locais diferentes: mão, cotovelo, ombro, tórax e cabeça (ambilateral), com temperaturas médias de 33,5 °C, 34,5 °C e 35,6 °C correspondentes ao frio conforto, conforto e desconforto quente, respectivamente. A Figura 18 apresenta os valores de cada região corporal e respectivamente os níveis de conforto térmico. Sendo as localidades mão esquerda (LH), mão direita (RH), cotovelo esquerdo (LE), cotovelo direito (RE), ombro esquerdo (LS), ombro direito (RS), cabeça e tórax [COSMA, 2018].

Figura 18 - Temperaturas de diferentes regiões corporais para sensação de desconforto frio, conforto e desconforto quente.

Average surface temperature(°C) for all 30 subjects grouped by TCV.

TCV	LH	LE	LS	RH	RE	RS	Head	Torso
Cold Discomfort	32.6	33.7	32.7	32.5	33.9	32.7	35.1	31.5
Comfortable	33.7	34.5	33.8	33.5	34.6	33.9	36.2	33.3
Warm Discomfort	35.3	35.3	34.9	35.0	35.5	35.1	37.0	34.8

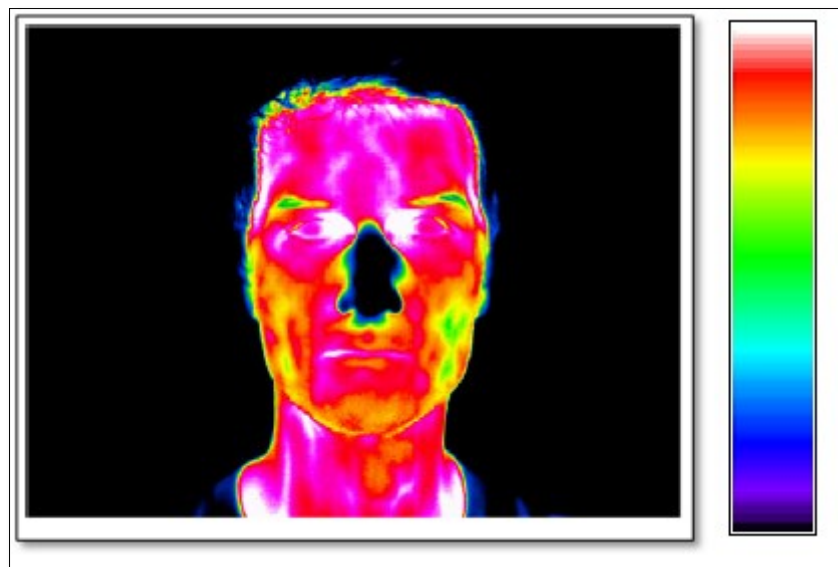
Fonte: Metzmacher et al. (2018).

Segundo Metzmacher et al. (2018) apresenta um framework de código aberto e extensível que suporta entrada de sensor multimodal com foco na fusão de dados ópticos e sensor convencional como entrada para análise também de conforto térmico. Neste estudo utiliza-se a combinação de duas câmeras: uma operando em domínio visível para rastreamento facial e outra câmera térmica, operando em domínio infravermelho conforme apresentado na Figura 19, para extração de informação de temperatura corporal [METZMACHER, 2018]. A partir da escala pode-se verificar temperaturas mínimas na escala de cor preto até a temperatura corporal na escala de cor branca.

O reconhecimento de imagem é usado para identificar e rastrear regiões pré-definidas do corpo humano em uma imagem térmica. Este estudo apresenta que é observado que as câmeras térmicas de alta resolução são equipamentos de alto custo e estão sujeitas a mudanças de temperatura, assim, um rastreamento automático de regiões corpo humano é um processo bem complexo. Então aplicaram na metodologia um sistema de duas câmeras que são utilizadas com diferentes tarefas de detecção de região e rastreamento de temperatura. Os testes são realizados somente em ambiente controlado. Neste estudo concluiu-se que a temperatura do ambiente, umidade do ar e a velocidade do ar são fatores que interferem na aquisição de temperatura do indivíduo por imagem [METZMACHER, 2018].

Figura 19 - Região facial apresentada pela câmera infravermelho.





Fonte: Metzmacher et al. (2018).

Aryal et al. (2019) apresenta em sua pesquisa a comparação da precisão do uso de sensoriamento ambiental com um sensor de temperatura do ar, sensoriamento fisiológico com um dispositivo de pulso para monitorar a temperatura da pele do punho e uma câmera térmica para monitorar a temperatura da pele facial para prever a sensação térmica individual e a satisfação. Neste estudo, os pesquisadores alteraram também a temperatura ambiente para investigação. Semelhante ao estudo de Metzmacher et al, também utilizam imagens em domínio visível para rastreamento de regiões faciais e imagens térmicas para análise de temperatura corporal [ARYAL, 2019].

No ano de 2021, Cardone et al. (2021) analisa vídeos em domínio visível e infravermelho dos rostos de seis sujeitos os quais foram gravados simultaneamente durante movimento da cabeça. Assim, abordam a combinação da região facial em imagem de luz visível com a imagem simultânea em infravermelho. Analisaram conseguinte diferentes métodos de fusão e medições de temperatura da pele em aplicações offline. Este estudo alcançou a estimativa de temperatura regional da pele na ordem de RMSE de  $0.09\text{ }^{\circ}\text{C}$  [CARDONE, 2021].

No mesmo ano, tem-se também em destaque o estudo de Xuan et al, o qual utiliza uma técnica de pré-processamento especialmente desenvolvida para ampliar características do rosto e, posteriormente, melhorar o rastreamento nas imagens térmicas. As temperaturas da pele facial foram comparadas com as temperaturas do tímpano. Este estudo apresenta como temperatura corporal de homeostase o intervalo entre  $36.5\text{-}38.5\text{ }^{\circ}\text{C}$  [XUAN,2022].

E por fim, em 2022, Wu et al. (2022) estuda a relação entre a temperatura da pele humana com diferentes temperaturas de ambientes fechados. Eles também utilizam imagens

infravermelho para análise e verificam a eficiência de imagens IR para predição de parâmetros de conforto térmico de indivíduos. É importante ressaltar que para ambientes configurados em 16 °C, as estruturas faciais apresentaram os intervalos temperaturas de 33-34 °C na região da testa, 31,5-32 °C na região das bochechas e 26-31 °C na região do nariz. Já para o ambiente configurado a 30 °C, as regiões apresentaram o intervalo de temperatura 36,2-37 °C para a testa, 36-36,7 °C para as bochechas e 34,2-37 °C para o nariz [WU,2022].

Mediante os estudos anteriormente apresentados podemos destacar que todos utilizam imagens IR para aferição térmica, realizam fusão entre imagens de diferentes domínios de frequência, necessitam de equipamentos específicos e complexos. Geralmente, os estudos trabalham em aplicações offline devido ao alto custo computacional, as regiões de interesse da face são selecionadas manualmente quando considerados apenas imagens IR e possuem foco geral em predição de conforto térmico em ambientes controlados.

Tabela 3 - Estado-da-arte em monitoramento de temperatura corporal sem contato.

<b>Autores</b>	<b>Ano</b>	<b>Título</b>
James et al.	2014	Reliability and validity of skin temperature measurement by telemetry thermistors and a thermal camera during exercise in the heat
Cosma et al.	2018	Thermal comfort modeling in transient conditions using real-time local body temperature extraction with a thermographic camera
Metzmacher et al.	2018	Real-time human skin temperature analysis using thermal image recognition for thermal comfort assessment
Aryal et al.	2019	A comparative study of predicting individual thermal sensation and satisfacting using wrist-worn temperature sensor, thermal camera and ambient temperature sensor
Cardone et al.	2021	Automated warping procedure for facial thermal imaging based on features identification in the visible domain
Xuan et al.	2021	Thermal infrared imaging based facial temperature in comparison to ear temperature during a real-driving scenario
Varshini et al.	2021	IoT-Enabled smart doors for monitoring body temperature and face mask detection

Wu et al.	2022	Recognition and prediction of individual thermal comfort requirement based on local skin temperature
-----------	------	--

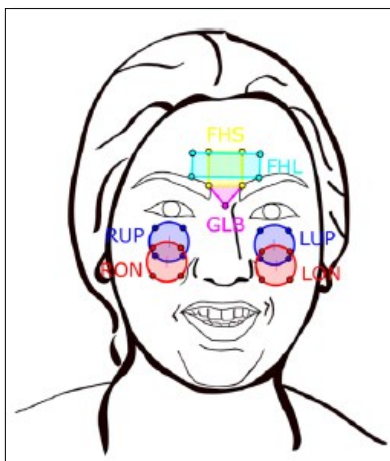
Com relação à segunda categoria são apresentados os trabalhos que abordam “Monitoramento de sinais vitais em domínio visível”, ou seja, estudos que abordam somente sinais em frequência da luz visível (domínio RGB).

O estudo de Fernandez et al. (2020) apresentou um método de medição de taxa de batimentos cardíacos (HRV) baseado em smartphone. O método foi originalmente desenvolvido e validado pelos pesquisadores. Eles utilizaram a própria câmera do celular. Como comparação tem-se a coleta do sinal eletrocardiograma pelo equipamento clínico validado como padrão ouro na área médica [FERNANDEZ, 2020].

No mesmo ano, Hassan et al. (2020) apresentou também um estudo de eficiência de medição de taxa de batimentos cardíacos utilizando uma câmera digital para monitoração de saúde. Este experimento alcançou 45 voluntários e investigou os efeitos de movimentação corporal, iluminação, tom de pele e variação da distância do indivíduo em relação à câmera. Este estudo apresentou que indivíduos em até 5 metros de distância da câmera de aquisição de imagens conseguem ser monitorados. O estudo defende que a aplicação do monitoramento de saúde baseado em câmeras RGB seria encorajadora, pois desenvolver um método de monitoramento de saúde remoto para monitorar o estado fisiológico das pessoas em espaços públicos evitaria desde fatalidades até a propagação de doenças transmissíveis [HASSAN, 2020].

Em 2021, Ernst et al. (2021) apresenta que conforme a frequência cardíaca é um sinal vital essencial que pode ser medido remotamente com fotopletismografia baseada em câmera (cbPPG). Os sistemas para cbPPG normalmente usam câmeras que fornecem canais vermelho, verde e azul (RGB). Neste estudo consideraram as seguintes regiões faciais, conforme apresentado na Figura 20 sendo a glabella (GLB), região pequena da testa (FHS), região larga da testa (FHL), região alta da bochecha direita (RUP) e esquerda (LUP), bochecha direita (RON) e bochecha esquerda (LON).

Figura 20 - Regiões faciais analisadas por Ernst et al em imagens em domínio visível.



Fonte: ERNST et al. (2021).

As regiões de interesse da glabella e da testa forneceram acurácia de até 74,1% e uma razão sinal-ruído menor do que -3 dB mais altos com o canal Hue do espaço de cores HSV e o canal de crominância Q do espaço de cores NTSC. Ernst et al. (2021) conseguiu apresentar, por meio de buscas sistemáticas em grade de crescimento ou decréscimo de 0.01 em coeficientes para cada canal de cores do RGB. A pesquisa em grade revelou uma combinação ótima global e linear para os canais RGB atingindo acurácia de 79,2%. Esta combinação independe do tom de pele para medir a frequência cardíaca com cbPPG. Considerando seus resultados para a predição de FC, provaram também que é possível a imersão em pesquisas para análises de outros sinais vitais em domínio visível [ERNST, 2021].

As evidências concluídas pelo estudo de Molinaro et al. (2022) mostram uma premissa de que as câmeras digitais são uma tecnologia discreta e fácil de usar para monitoramento de sinais fisiológicos. Mas que ainda é necessário mais pesquisas para promover os avanços da tecnologia, e futuramente permitir sua aplicação em uma ampla gama de população e vida cotidiana, promovendo uma abordagem biométrica holística do corpo humano [MOLINARO, 2022].

Ainda no ano de 2022, Przybylo et al. (2022) destaca que o monitoramento remoto de idosos ou pacientes em isolamento domiciliar é parte essencial da medicina moderna. Eles apresentam um método de medição da frequência cardíaca baseado na Rede Neural Profunda de Long Short Term Memory (LSTM) tendo o canal verde (Green) dos canais RGB como componente principal de manipulação e análise. Também utilizam outras técnicas de pré-processamentos nos canais de cores como ICA, POS e ExG. Os banco de dados utilizado apresentam informações em domínio visível e em infravermelho, e concluíram que o canal

verde é altamente capaz de carrear informações bem como presentes no domínio infravermelho [PRZYBYLO, 2022].

Zheng et al. (2022) também aborda medição de taxa de batimentos cardíacos de forma remota a partir de vídeos gravados por câmeras comerciais. Obtiveram como foco os distúrbios e ruídos causados pela movimentação do indivíduo durante a medição e pela variação de iluminação. E apresentaram um framework capaz de eliminar movimentações corporais intrínsecas e predizer de forma eficiente a taxa de batimentos cardíacos [ZHENG, 2022].

Neste segundo grupo de estudos apresentados foram utilizados então câmeras digitais que operam em domínio de luz visível para medição e predição de sinais vitais com foco especificamente em taxa de batimentos cardíacos pela literatura, conforme resumido pela Tabela 4. Foi apresentado que o estado-da-arte não abordou especificamente predição de temperatura corporal com câmeras comerciais em domínio da luz visível. Visto isso, especificamente para o sinal fisiológico temperatura corporal não é sedimentado em vista da apresentação de obstáculos por muitos estudos como iluminação, movimento corporal, tons de pele e variação da distância indivíduo-câmera.

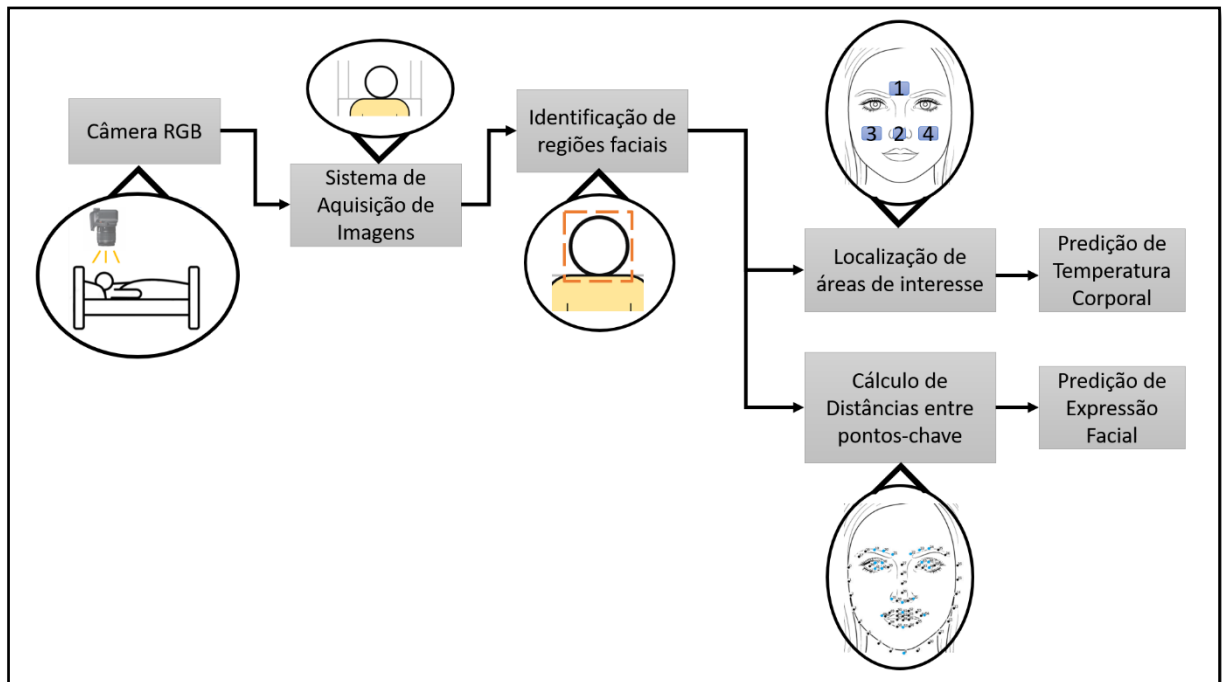
Tabela 4 - Estado-da-arte com foco em monitoramento de sinais vitais sem contato em domínio visível.

<b>Autores</b>	<b>Ano</b>	<b>Título</b>
Mcduff et al	2018	Camera measurement of physiological vital signs
Fernandéz et al	2020	A photoplethysmography smartphone-based method for heart rate variability assessment: device model and breathing influences
Hasan et al	2020	Towards health monitoring using remote heart rate measurement using digital camera: A feasibility study
Ernst et al	2021	Optimal color channel combination across skin tones for remote heart rate measurement in camera-based photoplethysmography
Molinaro et al	2022	Contactless vital signs monitoring from videos recorded with digital cameras: an overview
Przybylo et al	2022	A deep learning approach for remote heart rate estimation
Zheng et al	2022	Remote measurement of heart rate from facial video in different scenarios

### 3 PROPOSTA DE MODELO INTEGRADO PARA MONITORAÇÃO DE TEMPRATURA CORPORAL E RECONHECIMENTO DE EXPRESSÕES

Considerando as vantagens da utilização da visão computacional, o presente estudo aborda duas técnicas de aquisição de dados com foco em monitoração de pacientes não-comunicativos compondo um sistema integrado para avaliação do paciente. A Figura 21 apresenta a lógica de desenvolvimento do trabalho para monitoração de temperatura corporal e classificação de expressões faciais por câmera RGB desenvolvidos neste trabalho. Este passos consistem na seguinte sequência, a partir da aquisição de dados por uma câmera RGB é realiado a aquisição de imagens. Em seguida, é realizado a detecção de regiões faciais. A partir desta detecção, há a localização de áreas de interesse para a predição de temperatura corporal e em paralelo é realizado o cálculo de distâncias euclidianas entre pontos-chave da face para posterior predição de expressão facial.

Figura 21 - Framework para monitoração de temperatura corporal e classificação de expressões faciais por câmera RGB.



Fonte: Própria.

Assim, este trabalho apresenta dois métodos de análise de dados faciais: um com foco na predição de temperatura corporal e outro com foco na classificação de expressões faciais. Ambas as técnicas têm sido abordadas na literatura, mas até o presente não se observam modelos suficientemente robustos para aplicação em tempo-real no cenário de terapia intensiva, nem tampouco possuem modelos para aquisição de dados automática por apenas uma câmera RGB com eficiência e desempenho adequados.

## **3.1 METODOLOGIA**

Para a abordagem de predição de temperatura corporal foi analisado o comportamento de alterações de temperaturas corporais, bem como, imagens na frequência de luz visível (fotos) da região facial. Assim, é caracterizado como um estudo controlado do tipo analítico experimental no qual foram selecionados e recrutados um grupo de pessoas. O projeto foi aprovado pelo comitê de ética da UFU, com número de protocolo 60135722.5.0000.5152. A metodologia é descrita pelos tópicos participantes, critérios de inclusão de voluntários, aquisição e processamento de dados, hardware e software.

### **3.1.1 PARTICIPANTES**

Ao todo, foram recrutados 53 indivíduos, sem lesões na epiderme, não sendo permitido o uso de maquiagem, óculos em geral ou qualquer outro obstáculo que obstruísse o campo visual da região facial do indivíduo. Os sujeitos participaram do estudo por livre decisão e disponíveis para desistir da participação a qualquer momento. Por meio da assinatura do termo de consentimento de participação, os indivíduos participaram do estudo.

### **3.1.2 CRITÉRIOS DE INCLUSÃO DE VOLUNTÁRIOS**

Como critérios de inclusão, os voluntários deveriam possuir idade de 18 a 60 anos, de ambos os sexos, sem apresentar lesões ou problemas na epiderme. Os voluntários deveriam possuir um biotipo semelhante, com uma estrutura corporal mediana e índice de massa corporal (IMC) entre 18,5 e 24,9. Foram excluídos da pesquisa os voluntários que faltaram a sessão experimental, não conseguiram realizar o procedimento experimental, que passaram a sentir

desconforto ou dor em excesso e os que fizeram uso de substâncias estimulantes e psicoativas no dia do procedimento experimental.

Os voluntários do grupo de estudo foram recrutados pela autora do presente trabalho na Faculdade de Engenharia Elétrica (FEELT) da Universidade Federal de Uberlândia (UFU), a partir do convite direto aos estudantes da graduação e pós-graduação. Foram convidados a participar do estudo participantes na faixa etária de 18 a 60 anos, que atendessem todos os critérios de inclusão e não apresentassem nenhum dos critérios de exclusão identificados no protocolo do estudo. A Tabela 5 apresenta as características dos voluntários como idade, com média de 22 anos e desvio padrão de 7,58, altura média de 1,75 m com desvio padrão de 0,13 e peso com média de 75 kg com desvio padrão de 15,70.

Tabela 5 – Idade, altura e peso dos voluntários.

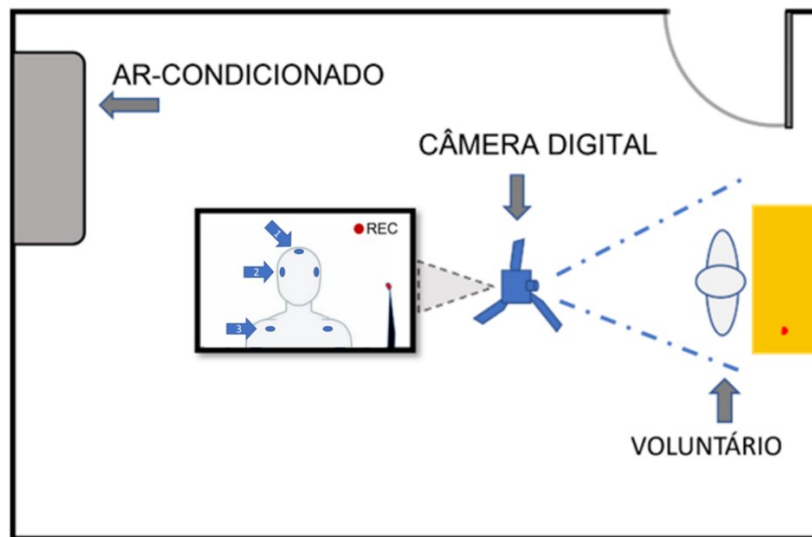
<b>Variável</b>	<b>Média</b>	<b>Desvio Padrão</b>
<b>Idade</b>	22	7.58
<b>Altura</b>	1.75	0.13
<b>Peso</b>	75	15.70

### 3.1.3 AQUISIÇÃO DE IMAGENS EM DOMÍNIO RGB

O dispositivo de aquisição de imagens fotográficas utilizado foi a câmera Canon modelo T7i. 18-55 stm. A aquisição de dados ocorreu com o voluntário alocado à 70 cm da câmera digital, e assim, o indivíduo permaneceu sentado, ao longo de toda a coleta, em uma cadeira confortável posicionado frontalmente à câmera. Conforme apresentado na Figura 22 com a ilustração da disposição da sala de coleta e seus componentes.

Figura 22 - Visão ilustrativa da sala de coleta durante aquisição de dados. Temperaturas corporais coletadas a partir dos sensores NTC durante a aquisição de dados. Apresentação de cinco sensores no plano anterior: (1) frente/testa, (2) região temporal da face, e (3) região subclavicular.





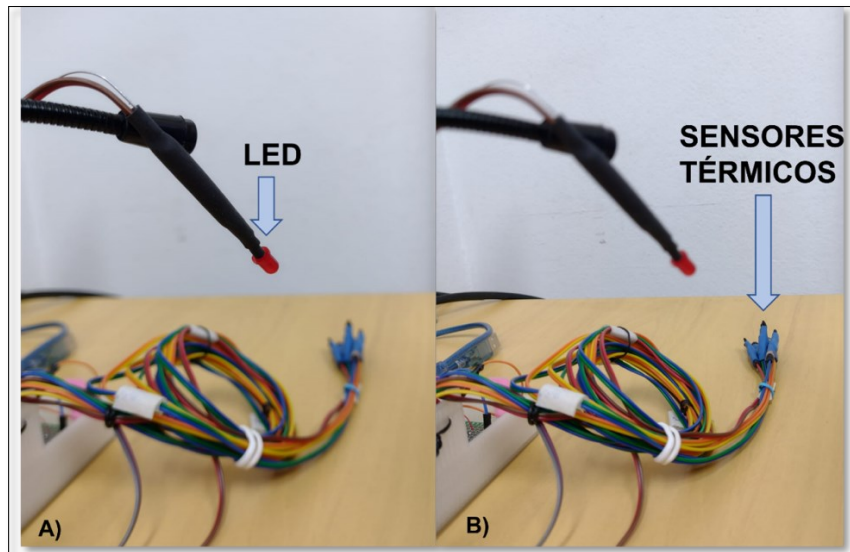
Fonte: Própria

### 3.1.3.1 COLETA DE DADOS

Para coleta das temperaturas corporais foram utilizados os seguintes componentes: termistores NTC 10K 3mm, resistores 10K, sensor de temperatura e umidade do ambiente DHT22, LED vermelho e um Arduino Mega 2560 R3. As temperaturas corporais foram coletadas pelos termistores fixados com esparadrapo. As regiões corporais analisadas foram frente/testa, regiões temporais da face e regiões subclaviculares. Com relação às regiões temporais e subclaviculares foram coletados em regiões homólogas: direito e esquerdo.

Com relação ao termistores, a partir da Equação Steinhart-Hart (III), é possível realizar a calibração dos sensores superficiais. A estrutura de aquisição de temperatura pelo termistor é envolto pelo termoretrátil com o devido isolamento para evitar perda de calor. A Figura 23 apresenta em (A) a disposição do LED utilizado como *trigger* nas gravações de vídeo e posterior combinação de tempo frame-temperatura e em (B) a disposição dos sensores térmicos que foram fixados no indivíduo nas regiões previamente apresentadas.

Figura 23 - Apresentação do hardware em (A) LED vermelho e em (B) agrupamentos dos sensores térmicos utilizados.



Fonte: Própria

A equação de Steinhart – Hart é uma expressão empírica que mapeia o valor da resistência com a temperatura onde o dispositivo está inserido. Quando os coeficientes  $A$ ,  $B$  e  $C$  são obtidos com instrumentação de precisão, a equação resulta em um erro inferior a 0,02 graus centígrados em uma faixa de cerca de 200 graus. A curva de resposta de um termistor NTC não é linear e as constantes que determinam essa curva não são fornecidas pelo fabricante. Uma aproximação linearizada do valor resistivo do termistor, de acordo com a temperatura, pode ser obtida através da equação (VI).

$$T = \frac{1}{A + B \cdot \ln(R) + C \cdot (\ln(R))^3} \quad (\text{VI})$$

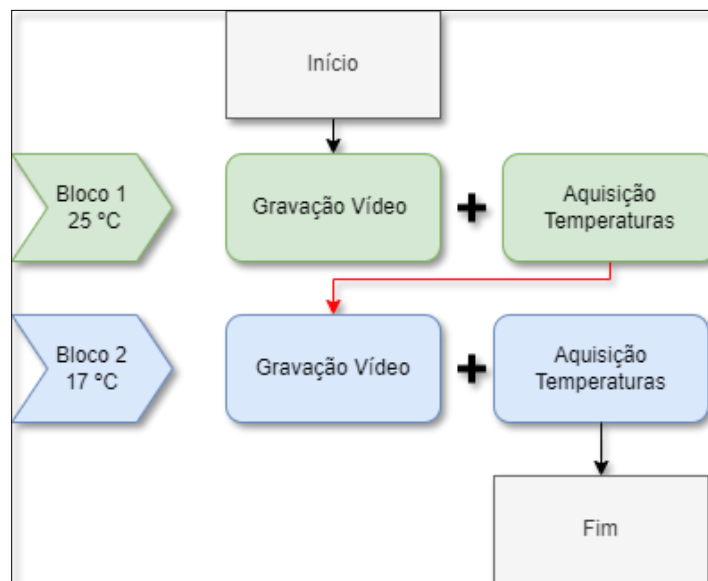
A variável  $T$  é a temperatura em Kelvin.  $R$  é a resistência em Ohms.  $A$ ,  $B$  e  $C$  = constantes fornecidas pelo fabricante ou determinadas experimentalmente. Para executar a calibração dos sensores NTC, foram feitas 10 medições e consequente aferição do valor expresso pelo sensor e pelo termômetro digital, aparelho comercializado em estabelecimentos de saúde. A partir da curva de calibração, foi realizado uma regressão linear e o termistores foram calibrados.

Os algoritmos desenvolvido para o presente projeto englobou a linguagem C++ para as aplicações em Arduino, Python para o processamento dos dados e para análise estatística. A coleta segue o protocolo apresentado na Figura 24, no qual são realizados dois blocos de aquisição de dados por cada voluntário. Durante o bloco 1, o ambiente está aquecido a 25°C. A

gravação do vídeo da região facial do voluntário possui duração de 5 minutos em que o voluntário está estático, sem movimentação com a cabeça e olhando em direção à câmera. Simultaneamente, a coleta das temperaturas é realizada.

Após um intervalo de 15 minutos, representado pela linha em vermelho, o ar-condicionado é configurado a 17 °C e a temperatura ambiente decaía para 22°C e novamente a coleta de vídeo e temperaturas corporais é realizada. Uma vez que os vídeos apresentam 5 minutos de duração e foram adquiridos a 30 FPS, tem-se como base de dados 18.000 frames para processamento por voluntário. Uma vez que é dada o início da gravação das temperaturas corporais, o LED vermelho é acessado durante 5 minutos no vídeo, considerado *trigger*, e assim sabe-se o instante exato de início e término. Posteriormente o vídeo é recortado para apenas os instantes de coleta.

Figura 24 - Diagrama de blocos do protocolo realizado.



Fonte: Própria

Como a taxa de amostragem utilizada para aquisição de temperatura foi de 0.5 Hz, foi necessário um ajuste para realizar o cruzamento de hora, minuto e segundo entre os frames e as temperaturas adquiridas. Uma vez coletado o tempo do frame, checava-se os instantes de temperaturas adquiridas na casa dos segundos. Caso não houvesse temperatura adquirida na mesma unidade de segundo, considerava-se então a casa de dezena dos segundos. Se a análise da dezena do segundo não houvesse combinação com o tempo da temperatura analisada, passava-se a análise de combinação para a amostra subsequente na

aquisição da temperatura corporal. Este processo ocorreu para todos os frames em combinação com suas respectivas temperaturas corporais. Assim, considerou-se como aceitável uma diferença máxima de 10 segundos entre frame e instante temperatura corporal adquirida.

### **3.1.4 BANCO DE DADOS**

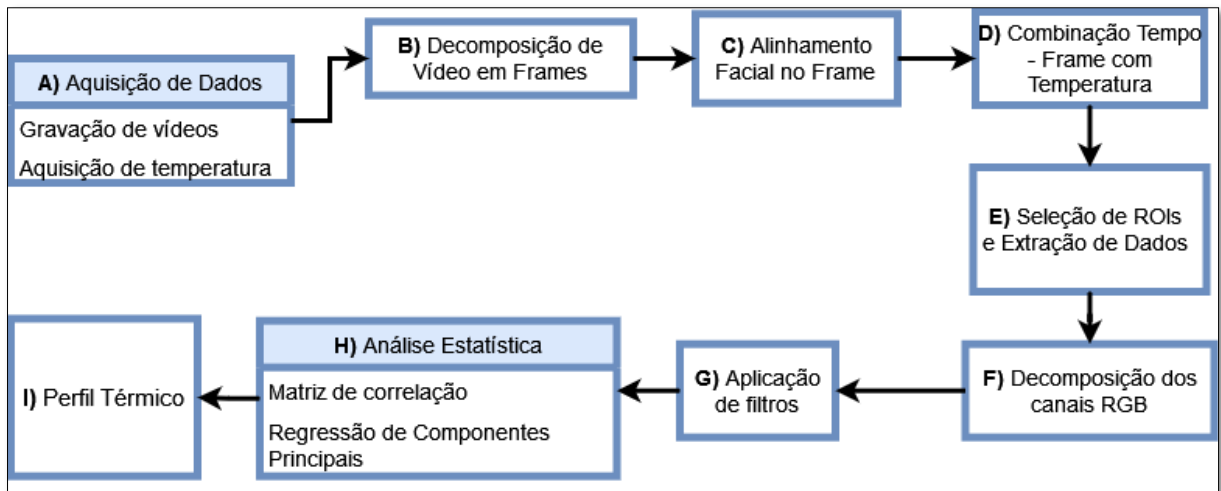
Para a análise de expressões faciais foram utilizados os bancos de dados mais recorrentes diante dos estudos analisados. Desta forma, será possível averiguar a eficiência do modelo desenvolvido. Ao todo foram utilizados três bancos de dados privados: Conjunto de Imagens de Expressão Facial Emocional de Varsóvia (WSEFEP), contendo 210 imagens de 30 indivíduos, Extended Cohn-Kanade Dataset (CK+), contendo 902 imagens, e The Japanese Female Facial Expression (JAFPE), contendo 188 imagens. Deve-se realizar uma solicitação para obter acesso aos os bancos de dados em seus respectivos repositórios. As classes de emoções presentes nos conjuntos de dados são: Neutro, Triste, Feliz, Medo, Surpresa, Nojo e Raiva. As imagens faciais foram coletadas com indivíduos estáticos realizando diferentes expressões faciais.

### **3.1.5 PREDIÇÃO DE TEMPERATURA CORPORAL**

A Figura 25 apresenta o diagrama de desenvolvimento e análise dos dados. Foram coletados os dados de 24 voluntários com a gravação de vídeo e aquisição de temperaturas corporais simultaneamente. Em seguida, os vídeos foram decompostos em frames e a partir da detecção facial foram alinhados no plano cartesiano bidimensional. Em seguida, é realizado o cruzamento de informação, ou seja, horas, minutos e segundos do frame com as respectivas temperaturas adquiridas naquele instante.

Adiante, é realizado a seleção das regiões de interesse em cada frame, ou seja, a extração de característica de cada região facial para todos os frames do indivíduo. A partir disso são decompostos os canais RGB em análises individuais. Em seguida, são aplicados 2 filtros: média-móvel e passa-baixa tipo Butterworth. Por fim, foram realizadas análises em matrizes de correlação para cada indivíduo e por grupo geral. Em seguida, foi utilizada a técnica de Regressão de Componentes Principais (RCP) para apresentação do perfil térmico aproximado do indivíduo.

Figura 25 - Diagrama de desenvolvimento e análise de dados.



Fonte: Própria

A apresentação dos resultados para aplicação da RCP é também dividido para o grupo de voluntários masculino (GM) e feminino (GF). O cálculo do Erro Médio Quadrático (EMQ) para as análise ocorreu a partir da aplicação de validação cruzada k-Fold sendo a média de 10 splits com 3 iterações ao todo.

Em suma, durante a análise de resultados, temos como inserção no modelo os vetores das variáveis calculadas com os sinais já filtrados. São estas:

- Umidade do ambiente\*
- Temperatura do ambiente
- Vetores de Pixels para cada canal Verde, Vermelho e Azul:
  - Variância da região do nariz
  - Variância da região da glabella
  - Variância da região da bochecha esquerda
  - Variância da região da bochecha direita
  - Variância da diferença entre glabella e nariz
  - Média da região do nariz
  - Média da região da glabella
  - Média da região da bochecha esquerda
  - Média da região da bochecha direita
  - Média da diferença entre glabella e nariz

A umidade do ambiente apresenta um (\*) porque são feitas duas análises. A primeira análise considera todas as 32 variáveis independentes, e a segunda análise considera-se 31 variáveis independentes excluindo a umidade ambiente. Apenas a título de informação, análises considerando apenas a umidade e desconsiderando a temperatura do ambiente não surgiram resultados eficientes. Com relação as variáveis dependentes, foram realizados três análises separadamente para predição de temperatura:

- Temperatura Média da região das Bochechas
- Temperatura Média da região Subclaviculares
- Temperatura da região da Testa

### 3.1.5.1 LOCALIZAÇÃO DE REGIÕES DE INTERESSE

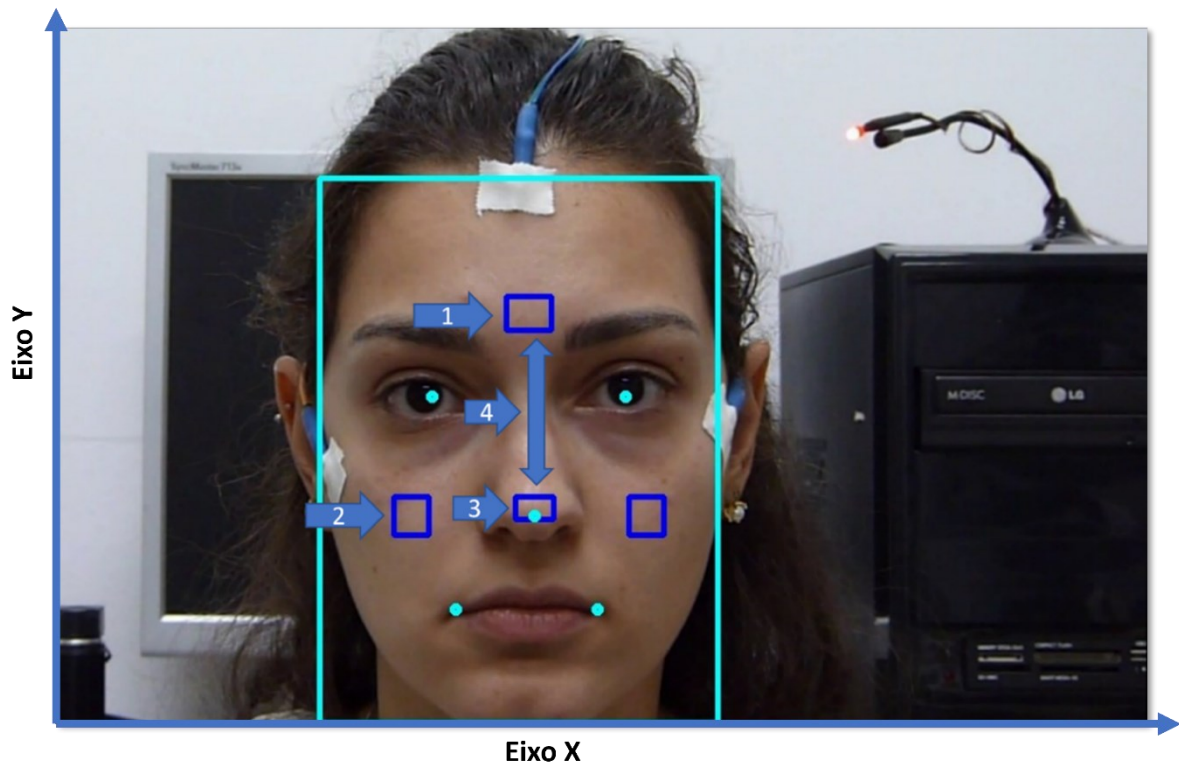
A detecção facial foi realizada pela ferramenta open-source RetinaFace apresentada pelo estudo de [DENG, 2019]. A partir da combinação de redes neurais convolucionais combinados com técnicas de regressão, esta biblioteca retorna as coordenadas dos olhos direito e esquerdo, do nariz e dos cantos direito e esquerdo da boca. Em azul claro são representadas as informações retornadas pela ferramenta, e em azul escuro são apresentadas as regiões de segmentação que tem-se interesse em análise, conforme apresentado na Figura 26.

O alinhamento facial é realizado a partir das coordenadas dos olhos direito e esquerdo e, caso não apresentem valores semelhantes no eixo vertical, a imagem foi rotacionada considerando seu ponto central. Como sendo a hipotenusa temos a distância euclidiana entre os olhos e como o cateto adjacente temos a diferença absoluta entre as coordenadas X dos olhos direito e esquerdo. Realizando operações trigonométricas é calculado o ângulo de rotação.

A partir da imagem rotacionada, as segmentações são automáticas da região de interesse. Elas são importantes para se extrair informações locais da imagem e com foco em uma região facial apenas de modo que a partir da decomposição da informação possa se obter uma análise mais específica e localizada. As regiões de interesse facial analisadas neste estudo são a região da glabella (1), bochechas (2) e nariz (3). A partir destas localidades também foram feitas combinações: análise de diferença da glabella com relação ao nariz, também apresentado na Figura 26.

Figura 26 - Em azul escuro é demarcado as áreas de interesse calculadas automaticamente a partir das coordenadas dos olhos, nariz e cantos da boca. As regiões estão

indicadas na imagem como (1) Glabella, (2) Bochechas e (3) Nariz e (4) é apresentado a diferença entre as regiões nariz e glabella.



Fonte: Própria

O cálculo automático das regiões de interesse (ROIs) seguiu a proporção facial, a região do nariz segue a coordenada retornada pela ferramenta RetinaFace, com altura de  $1/3$  da altura facial. A região da glabella conta como sendo  $2/3$  de toda a altura do rosto, apresentando a mesma coordenada do eixo X do nariz. As regiões das bochechas apresentam as mesmas coordenadas do eixo X dos pontos retornados de cada olho e coordenadas Y da região do nariz.

Os canais RGB são decomposto após a extração de informações em cada ROI. Neste caso os frames são retornados na sequência BGR pelo OpenCV da linguagem Python. Cada sinal, por canal, é salvo separadamente para posterior análise. Uma vez que tem-se a região de interesse delimitada, tem-se consequentemente o vetor de pixels. E, para este vetor de pixels, é calculado a média e a variância dos pixels. Assim, são consideradas as seguintes informações: média da glabella canal verde, média nariz canal verde, média bochecha esquerda canal verde, média bochecha direita canal verde e média diferença glabella-nariz canal verde. O mesmo é repetido para a variável variância e para os outros canais vermelho e azul como apresentado anteriormente na seção 3.5.2.

Por fim, os sinais de temperatura e sinais advindos dos pixels foram submetidos ao filtro de média-móvel para pacotes de 30 amostras, que condiz a 1 segundo de aquisição, e ao filtro de passa-baixa tipo butterworth de 3ª ordem com frequência de corte em 0,015 Hz.

### **3.1.6 CLASSIFICAÇÃO DE EXPRESSÕES FACIAIS**

Os algoritmos foram desenvolvidos na linguagem Python. As imagens de cada banco de dados foram agrupadas em sete classes de classificação de 0 a 6. Os alvos qualitativos foram alterados para números: 0-raiva, 1-nojo, 2-feliz, 3-neutro, 4-triste, 5-surpresa, 6-medo. O algoritmo de pré-processamento percorre então os diretórios com as imagens e cria planilhas com os valores analíticos extraídos das imagens, além da categoria de emoção. A biblioteca de código aberto Dlib, lançada em 2015 e disponibilizado pelo OpenCV, é utilizada para detectar e recortar a face, além de coletar as coordenadas 68 pontos-chave da região facial [KING, 2015].

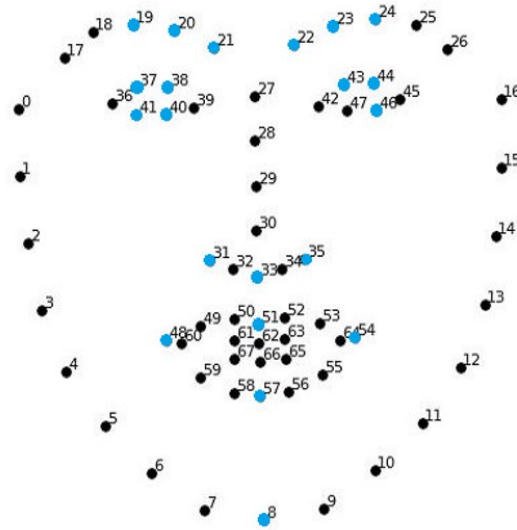
#### **3.1.6.1 LOCALIZAÇÃO DE PONTOS-CHAVE DA REGIÃO FACIAL**

A Figura 27 apresenta 68 pontos identificados e retornados pela biblioteca open-source Dlib no OpenCV em linguagem Python. Esta ferramenta é baseada na análise de descritores locais da imagem em combinação com técnicas de aprendizado de máquina. Na Figura, os pontos em azul são aqueles selecionados para análise. Ao todo considerou-se, então, 20 pontos da região facial: 8, 19, 20, 21, 22, 23, 24, 31, 33, 35, 37, 38, 41, 43, 44, 46, 48, 51, 54 e 57. As distâncias euclidianas foram calculadas para 15 duplas, compreendendo: 19-33, 24-33, 37-41, 44-46, 48-54, 51-57, 21-22, 8-33, 33-48, 33-54, 20-21, 22-23, 20-38, 23-43, 31-33, 33-35.

Assim, a seleção dos pontos e as distâncias calculadas tiveram como base as unidades de ativação (UA) de expressões faciais estabelecidos por Paul Ekman em 1992 citadas na seção 2.2 [EKMAN, 1992]. Os pontos selecionados também foram baseados no estudo de Sajjad et al. (2020) no qual afirma que estudos psicológicos provaram que características de expressões faciais estão localizadas ao redor da boca, nariz e olhos, e por isso, são pontos importantes para o REF [SAJJAD, 2020].

Figura 27 - Apresentação do retorno da ferramenta Dlib - 68 pontos da face; os pontos em azul claro são considerados no estudo.





Fonte: Própria

A partir da localização dos pontos-chave da face, em um espaço bidimensional da imagem tem-se as coordenadas  $x$  e  $y$  de cada ponto. Assim, é possível calcular a distância euclidiana ( $d$ ) de cada dupla de pontos, conforme apresentado na equação (VII) considerando ponto A e ponto B. Estas distâncias foram os dados de entrada para o desenvolvimento do modelo de classificação de expressões faciais.

$$d_{AB} = \sqrt{(X_B - X_A)^2 + (Y_B - Y_A)^2} \quad (\text{VII})$$

### 3.1.6.2 MODELO DE CLASSIFICAÇÃO DE EXPRESSÕES FACIAIS

O desenvolvimento do modelo de classificação de expressões faciais utilizou o emprego de Algoritmos Genéticos (AG) para seleção de características e treinamento de algoritmos classificadores Máquina de Vetor de Suporte (SVM) e Perceptron Multi-camadas (MLP). O algoritmo genético realizou uma seleção dos componentes para a classificação da emoção considerados, ou seja, entre as 15 distâncias euclidianas quais eram de fato as mais relevantes. Desta forma, tem-se então um vetor com 15 posições. A população do algoritmo genético continha também 15 genes com valores randômicos de 0s e 1s. Como exemplo, considerando apenas 1 indivíduo dessa população, somente as posições em que havia 1s, eram as distâncias euclidianas consideradas, e as posições que havia o número 0, as distâncias eram nulas.

Ao todo foi considerado uma população com 1.000 indivíduos criados randomicamente com posições aleatórias de 0s e 1s. A elite, ou seja, os melhores indivíduos de cada geração

foram considerados 200 ao todo. O restante dos indivíduos para formar uma nova população de 1.000 indivíduos, completou-se com torneios de indivíduos 3 a 3, prosseguindo o vencedor do torneio. Com a geração nova formada, também foi aplicado com 80% de probabilidade a ocorrência de crossover e 50% de probabilidade de mutação. Considerando o número máximo de 30 interações, que já apresentava uma alta convergência.

Dentro deste algoritmo genético, a função fitness é considerada a acurácia do classificador escolhido, ou seja, desempenho do SVM ou do MLP. Portanto, quanto maior a acurácia do classificador, melhor o indivíduo. Como há os mecanismos de perpetuação dos melhores indivíduos ao longo das gerações, com algumas variabilidades entre os indivíduos, é possível melhorar ainda mais o desempenho e chegar ao máximo global procurado. Ao final, o algoritmo apresenta o melhor indivíduo, que realizou a seleção das principais distâncias euclidianas, mediante as 15 calculadas, e que apresentou a maior acurácia mediante o classificador escolhido.

Considerando os classificadores utilizados, para ambos o grupo de teste dos dados foi 25% do total da base de dados analisada. Com relação ao perceptron multicamadas, foram consideradas 4 camadas escondidas sendo o número de neurônios de 35, 35, 35 e 13, com máximo de 3500 iterações. Para a arquitetura da máquina de vetores de suporte utilizou-se o kernel sigmoid. As acurácias retornadas pelo respectivo classificador avaliava o desempenho do indivíduo considerado no algoritmo genético.

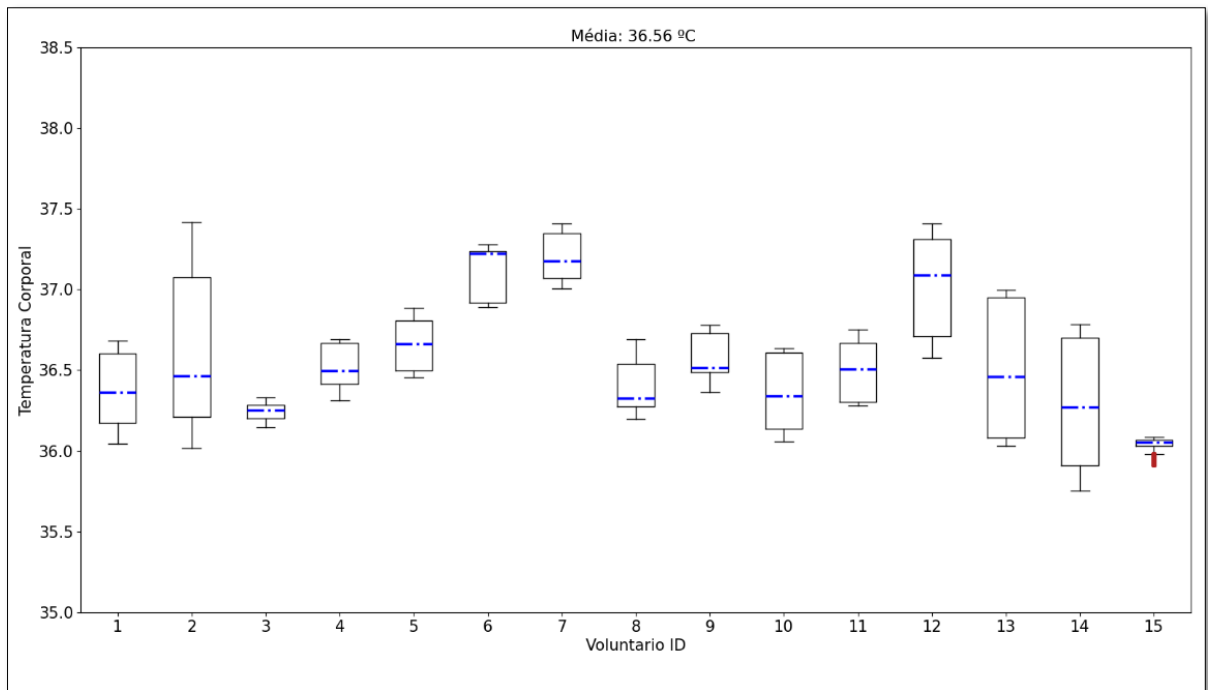
## 4 RESULTADOS

### 4.1 PREDIÇÃO DE TEMPERATURA CORPORAL

Os resultados foram divididos entre grupo de voluntários masculino (GM) e feminino (GF). Esta divisão para análise foi baseada no estudo de Jia et al. (2022) pois apresenta que há diferenças nas temperaturas corporais de homeostase conforme o gênero. O protocolo durante o desenvolvimento da pesquisa foi alterado algumas vezes, desde a quantidade de blocos, até a transição de registro de fotos em extensão *raw* para o formato de vídeo com frames em *jpg*. Ao todo, foram recrutados 53 voluntários. No entanto, apenas 24 sujeitos completaram os protocolos de coleta, sendo 15 do grupo masculino e 9 do grupo feminino. Os dados coletados de todos os participantes que finalizaram as coletas foram processados, como descrito anteriormente. A seguir, para ilustrar os achados, são apresentados os resultados de parte dos sujeitos de cada grupo.

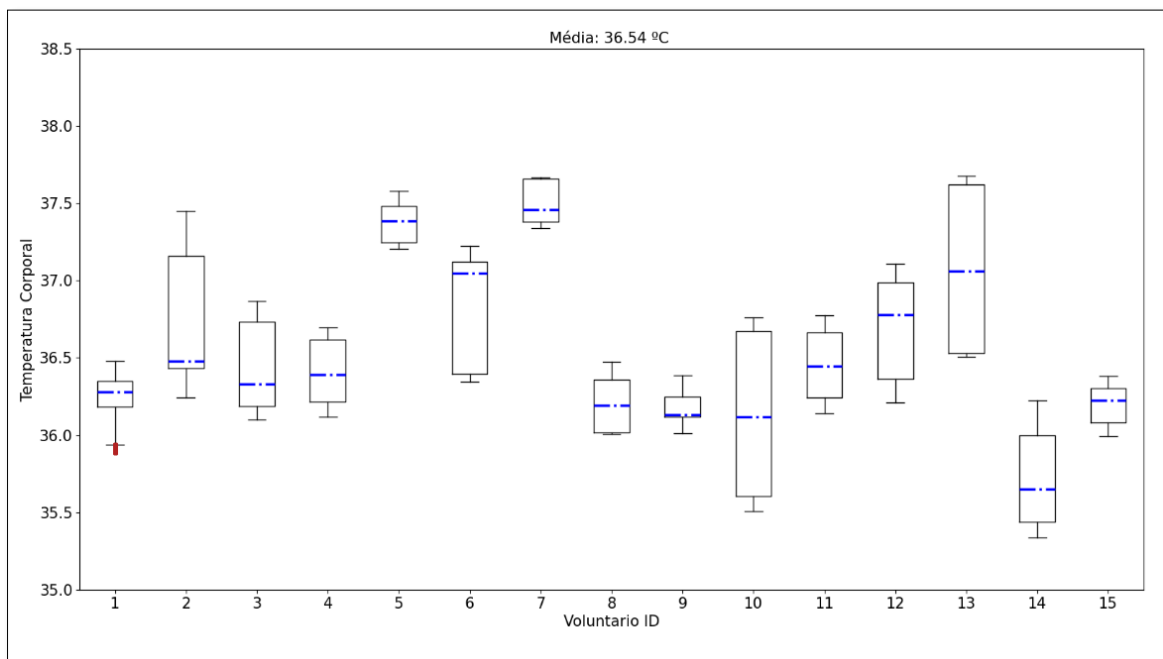
Em geral, as temperaturas corporais do GM obteve, para temperatura subclaviculares direita e esquerda, os valores médios 36.56 °C e 36.54 °C, respectivamente. Quando a temperatura ambiente decaía para 22 °C, a alteração das temperaturas subclaviculares dependia exclusivamente de cada indivíduo e sua capacidade de regulação da temperatura corporal. Entretanto, as temperaturas na região facial como testa e bochechas, como são regiões mais expostas ao ambiente externo, obtiveram maiores alterações conforme a temperatura ambiente. As Figuras 28 e 29 apresentam os gráficos boxplots do indivíduos do GM, de acordo com as temperaturas subclavicular esquerda e direita, respectivamente. As temperaturas corporais do GF obteve para temperaturas subclavicular esquerda e direita, os valores 36.50 °C e 36.46 °C, respectivamente, durante a aquisição de dados. As Figuras 30 e 31 apresentam os gráficos boxplots das variações de temperatura dos grupos de voluntários.

Figura 28 - Temperatura subclavicular esquerda - GM.



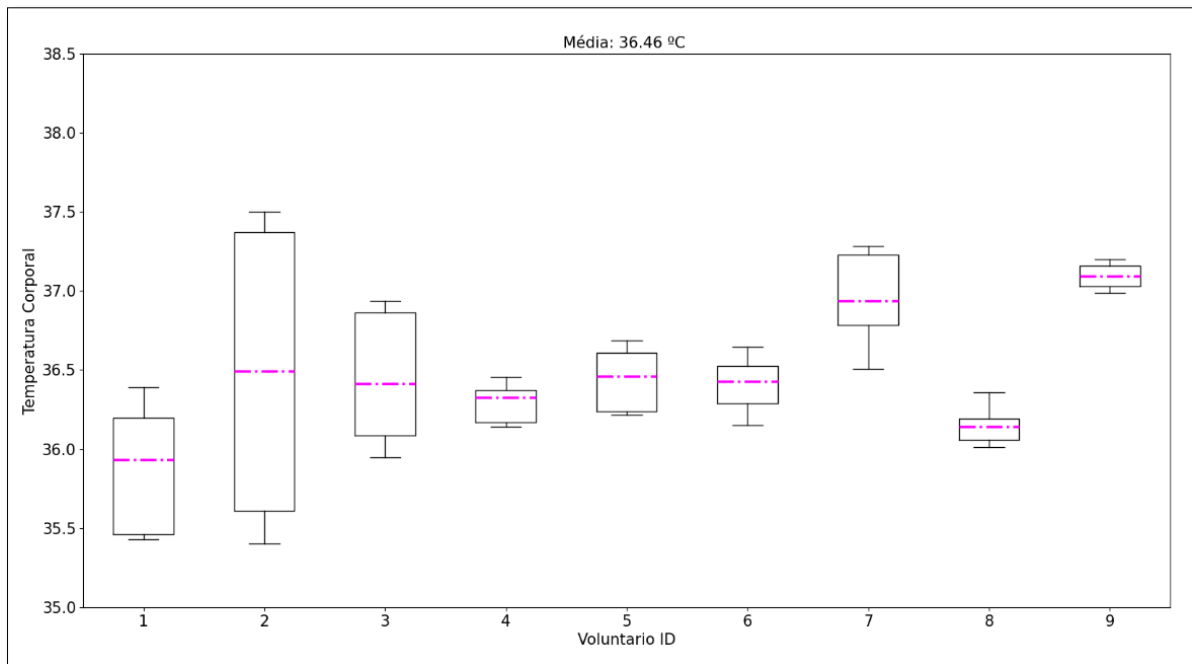
Fonte: Própria

Figura 29 - Temperatura subclavicular direita - GM.



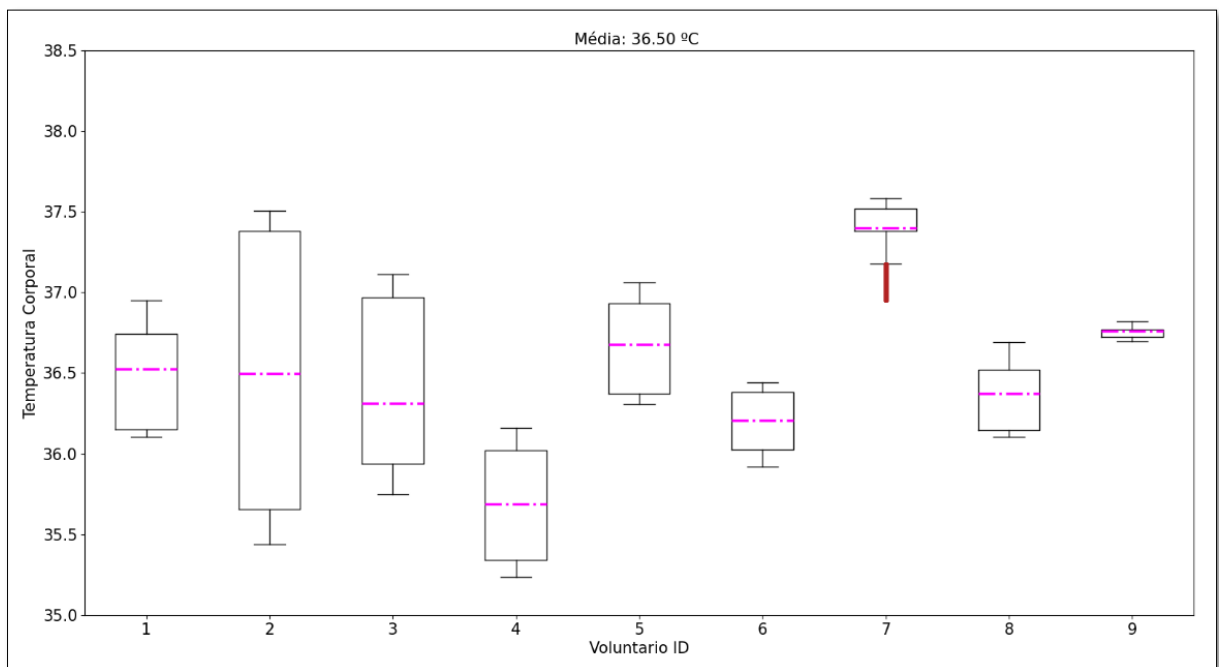
Fonte: Própria

Figura 30 - Temperatura subclavicular esquerda - GF.



Fonte: Própria.

Figura 31 - Temperatura subclavicular esquerda - GF.



Fonte: Própria.

Foram selecionados um voluntário do GM e uma voluntária do GF para representar a abordagem utilizada para os cálculos e seus resultados. As Figuras 32 e 33 destacam as

variações de temperatura ao longo do tempo, para o ambiente e para um indivíduo do GM e do GF, respectivamente. Ao todo são 18.000 amostras de temperatura foram tomadas, sendo 9.000 para o Bloco 1 e 9.000 para o Bloco 2. Na amostra número 9.000 é possível verificar o decaimento da temperatura ambiente de acordo com a alteração realizada na configuração do ar-condicionado.

De forma individual, cada voluntário apresentou variações de temperatura corporais que correspondiam aproximadamente às alterações da temperatura ambiente. As regiões da testa e das bochechas apresentaram maiores alterações pelo fato de estarem em contato direto com o meio externo sem a sobreposição de vestimentas ou acessórios.

Figura 32 - Temperatura Ambiente e temperaturas corporais de um voluntário do GM.

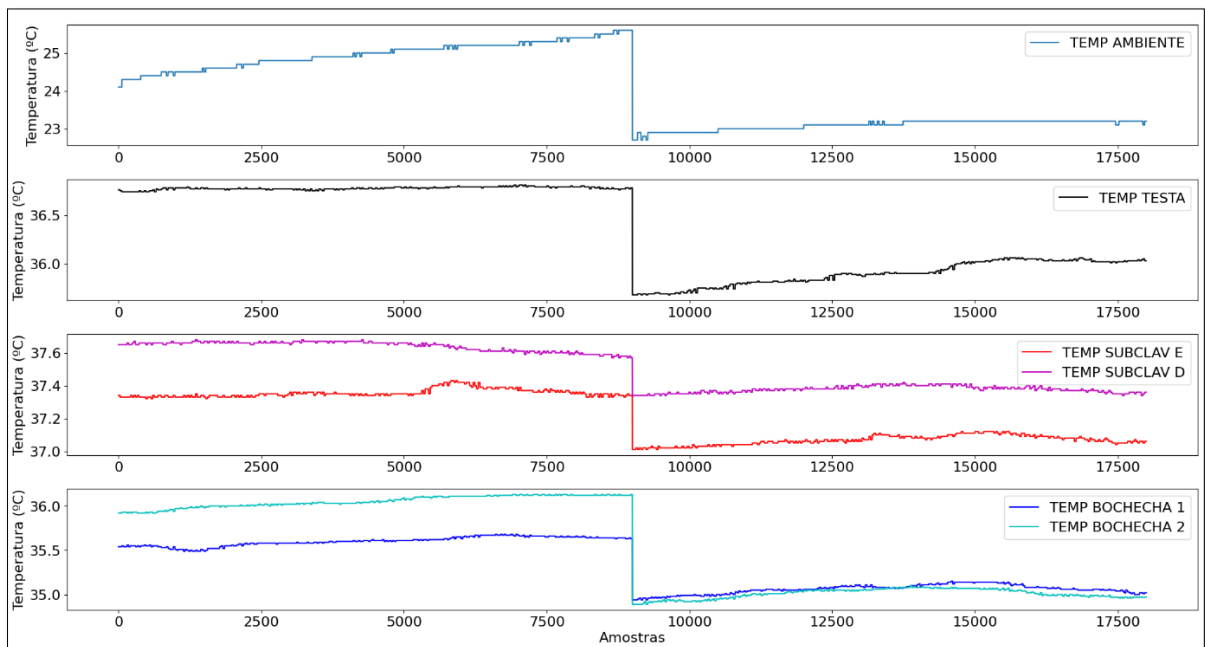
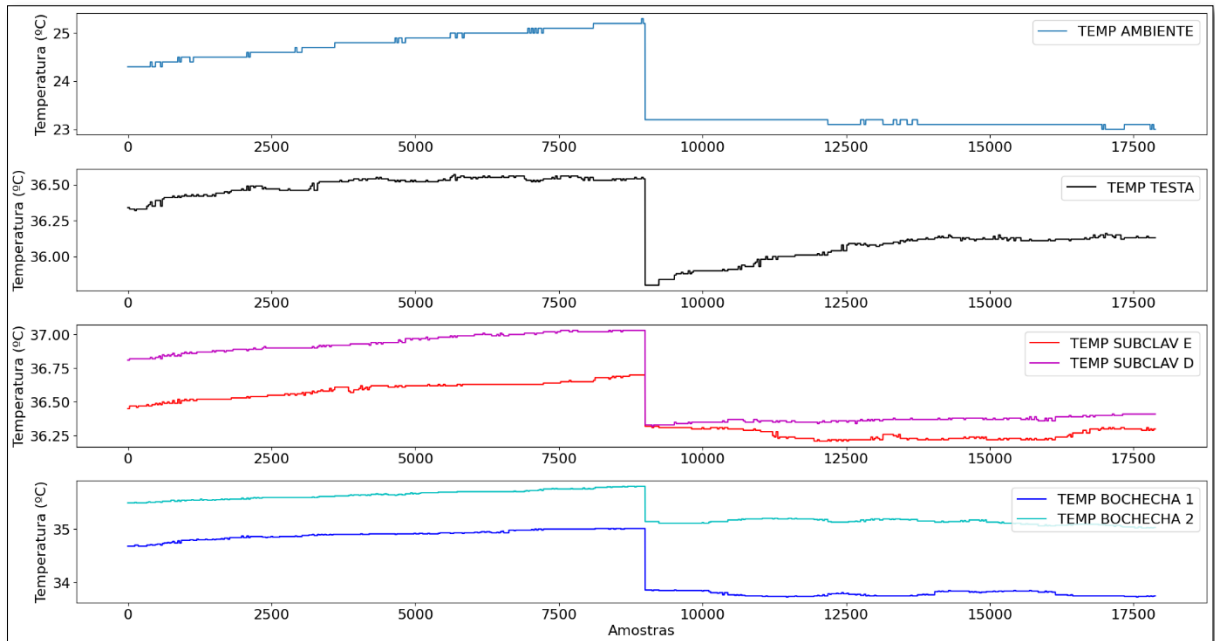


Figura 33 - Temperatura Ambiente e Temperaturas corporais de uma voluntária do GF.



Posteriormente, é apresentado o comportamento da temperaturas subclaviculares de acordo com a média do vetor dos pixels do canal verde nas diferentes regiões faciais após filtragem pelo filtro de média-móvel e pelo filtro passa-baixa. É possível verificar que existem alterações nas informações do canal verde entre o Bloco 1 e 2, tanto para o GM quanto para o GF, à medida a temperatura corporal se altera. Como esperado, foi encontrado uma resposta fisiológica à temperatura ambiente específica e variável entre sujeitos. Entretanto, percebe-se diferenças importantes entre os dois blocos de coleta, quente e frio, em valor absoluto da variável e também na oscilação do sinal dos pixels a partir de alteração de temperatura corporal.

A seguir é apresentado os gráficos de variação de temperatura e de intensidade do canal verde para cada região de um voluntário do GM e uma voluntária do GF. As correlações entre aquelas informações foram calculadas a partir do coeficiente de Pearson, considerando-se significativas para as análises aqueles coeficientes de correlação maior ou igual a 0,8. Primeiramente são apresentados os resultados do GM (Figuras 34 a 36) e, em seguida, do GF (Figuras 37 a 39). Por fim, as Figuras 53 e 54 apresentam as matrizes de correlação generalizada para o GM e o GF, respectivamente.

Figura 34 - Temperaturas Subclavicular Esquerda e Direita, Média da Região da Glabela, Região do Nariz e Média da diferença entre regiões dos Pixels para o Canal Verde de um voluntário do GM.

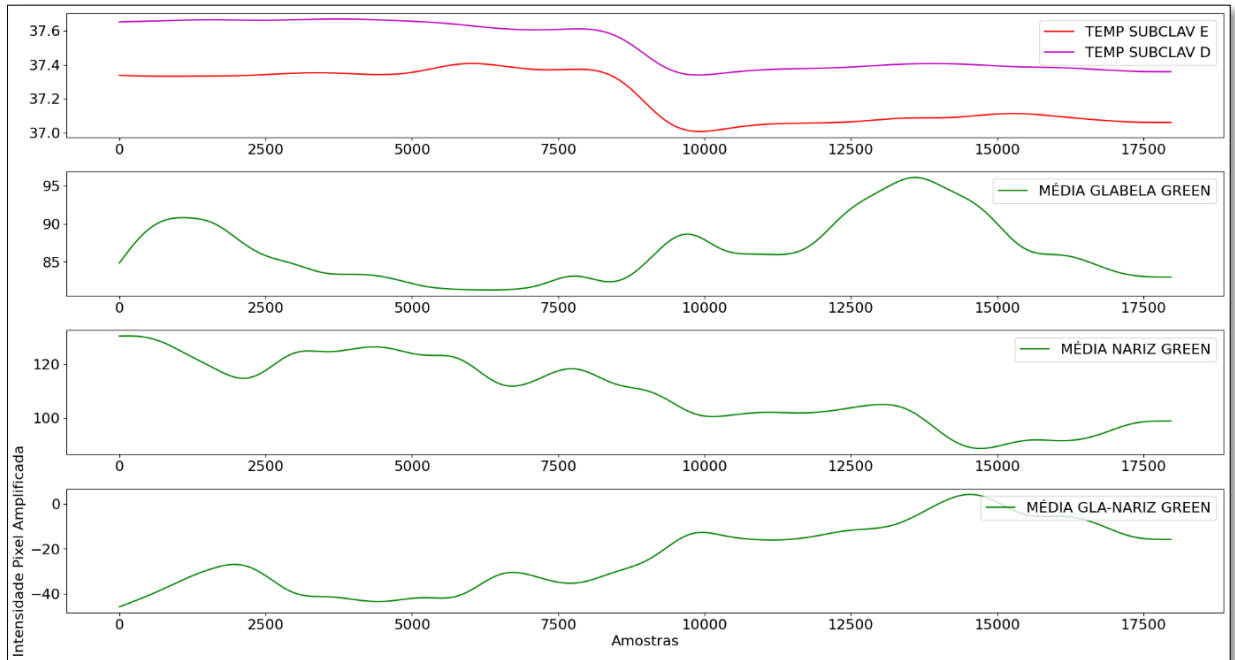


Figura 35 - Temperaturas Subclavicular Esquerda e Direita e Média da Região da Bochecha Direita e Esquerda dos Pixels para o Canal Verde de um voluntário do GM.

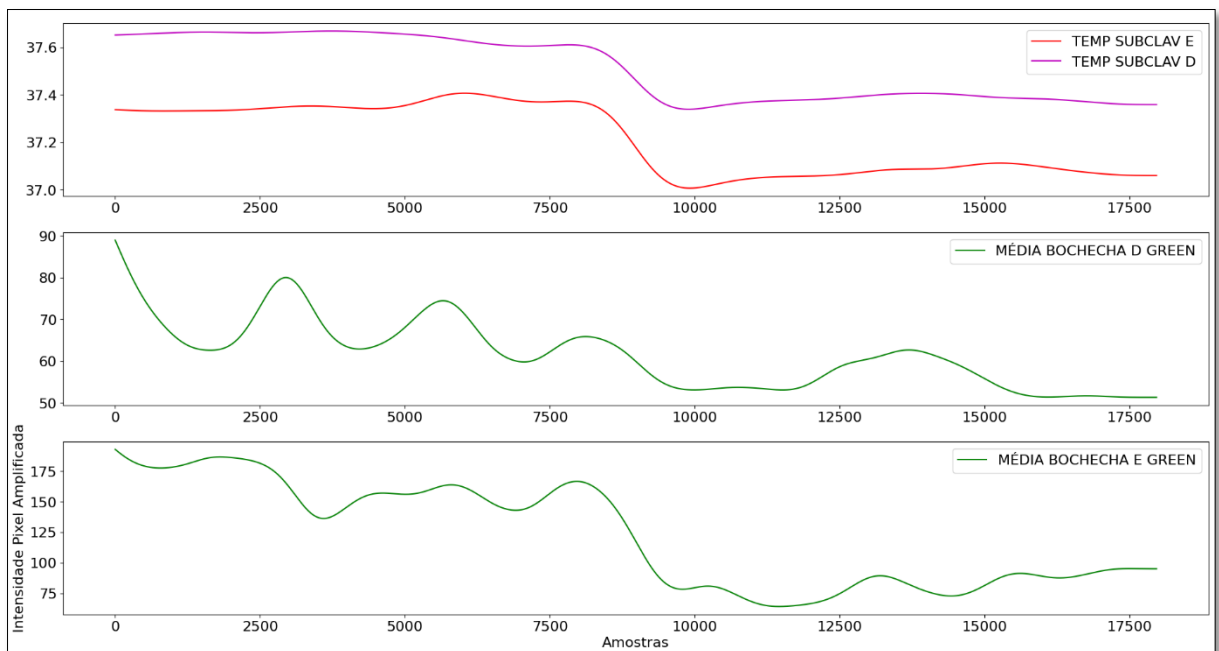
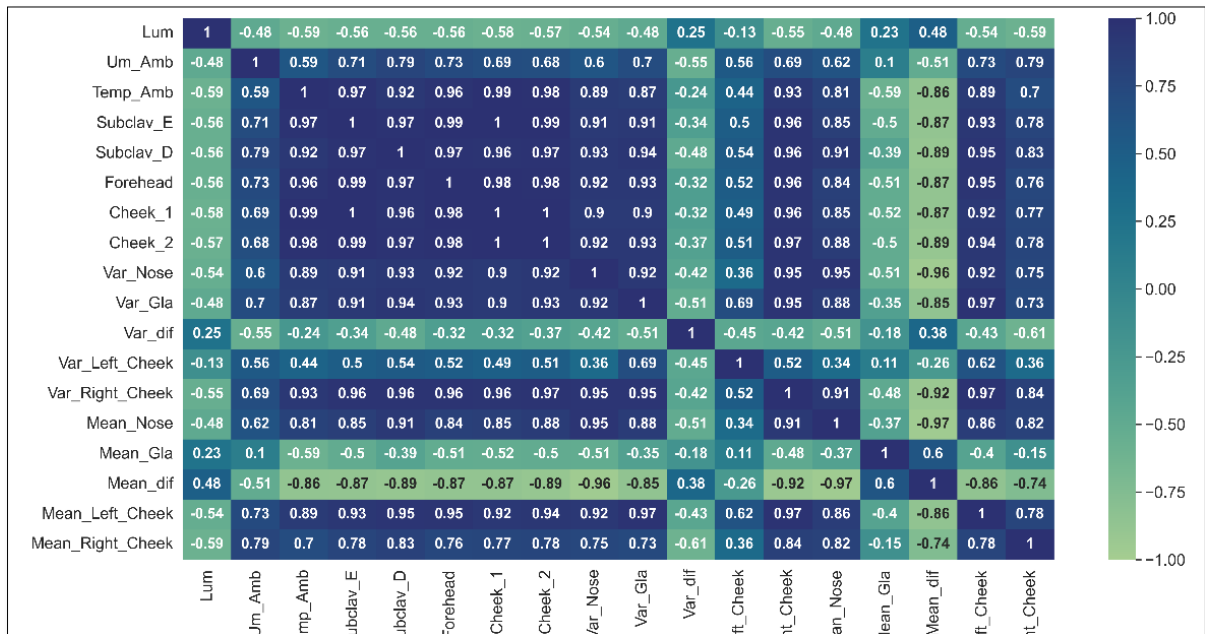




Figura 36 - Matriz de correlação considerando variáveis do ambiente, temperaturas corporais e variáveis dos pixels do canal verde de um voluntário do GM.



Para o voluntário do GM verificou-se um alto coeficiente de correlação da variável temperatura do ambiente com as temperaturas corporais. Segundo a matriz de correlação da Figura 36, podemos observar que a variância do nariz, a variância da glabella e variância da bochecha direita apresentaram coeficiente correlação intimamente forte com as temperaturas do ambiente, temperatura da região subclavicular esquerda e direita, testa e bochechas 1 e 2. Para tanto, tal relação também está presente de acordo com a variável média do nariz e média da bochecha esquerda com as temperaturas do ambiente, temperatura da região subclavicular esquerda e direita, testa e bochechas 1 e 2.

Figura 37 - Temperaturas Subclavicular Esquerda e Direita, Média da Região da Glabela, Região do Nariz e Média da diferença entre regiões dos Pixels para o Canal Verde de uma voluntária do GF.

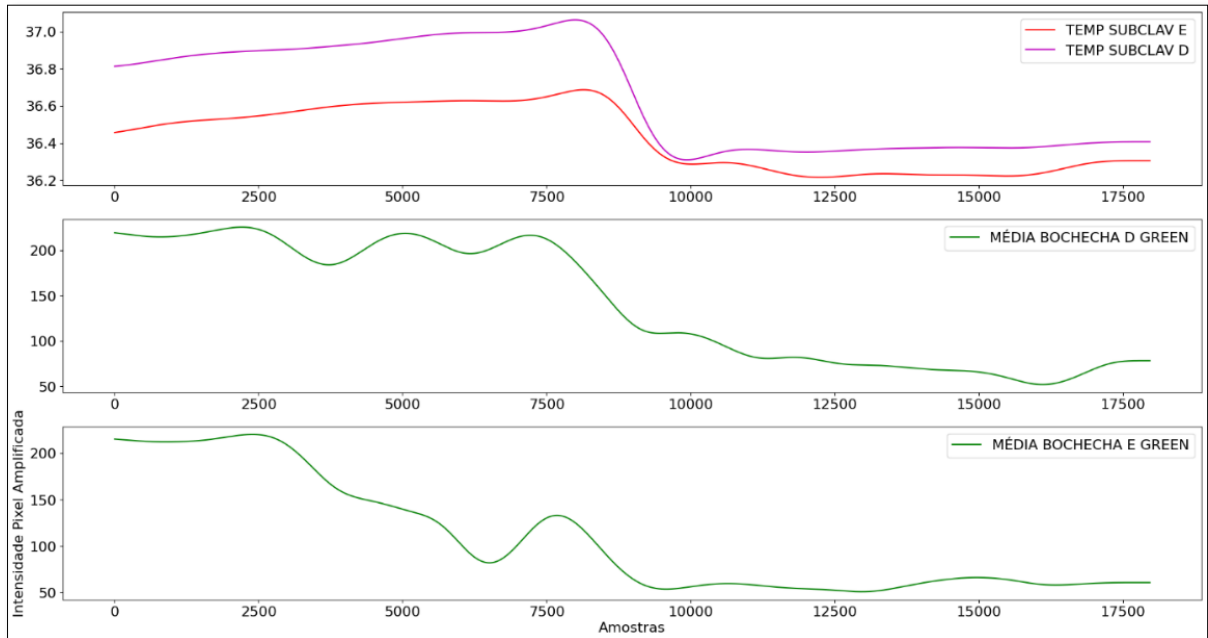


Figura 38 - Temperaturas Subclavicular Esquerda e Direita e Média da Região da Bochecha Direita e Esquerda dos Pixels para o Canal Verde de uma voluntária do GF.

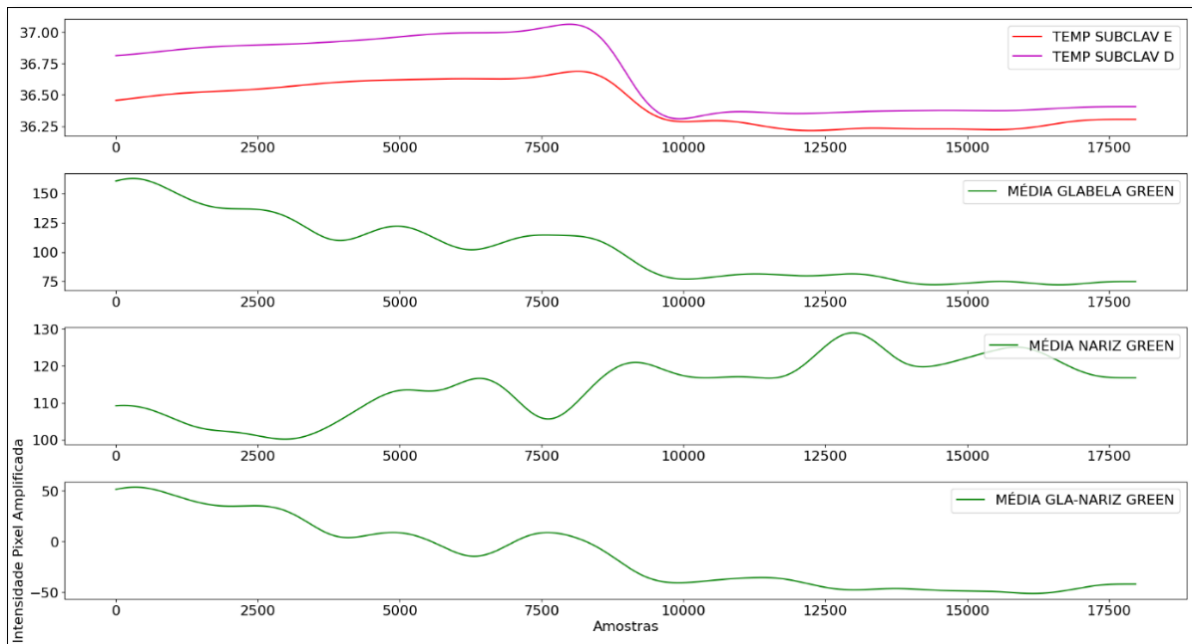
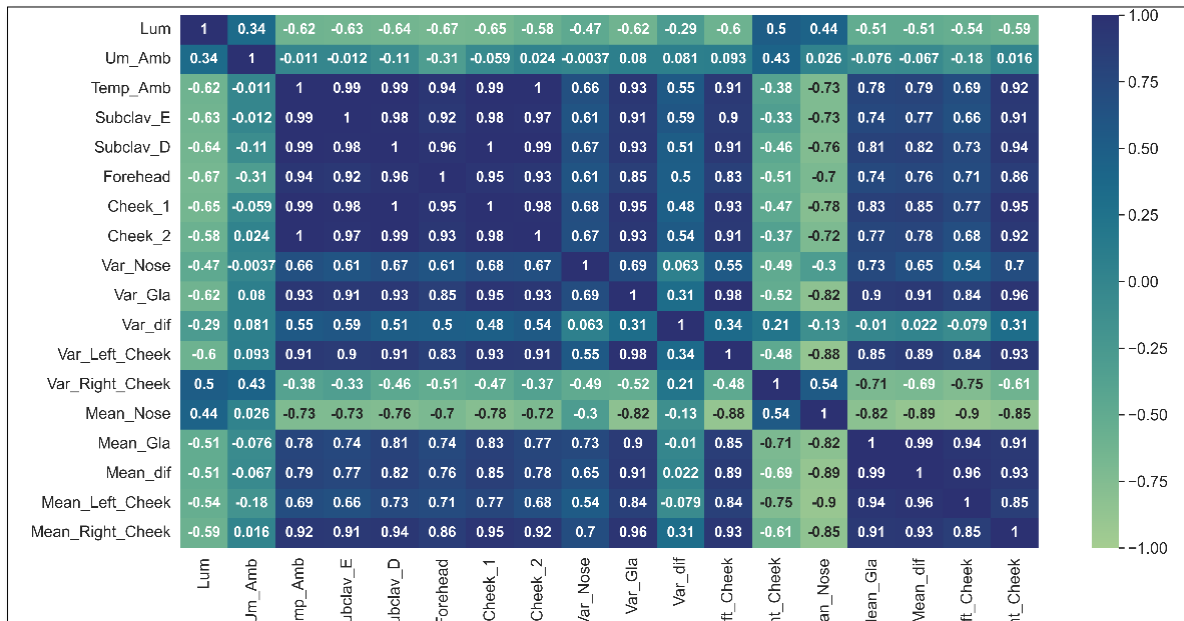


Figura 39 - Matriz de correlação considerando variáveis do ambiente, temperaturas corporais e variáveis dos pixels do canal verde de uma voluntária do GF.



Para a voluntária do GF pode-se verificar que a variável do ambiente temperatura apresenta forte relação com as temperaturas corporais adquiridas. Pode-se observar na Figura 39, na matriz de correlação, que a variância da glabella, a variância da bochecha esquerda e a média da bochecha direita apresentaram alta correlação com a temperatura ambiente e com as temperaturas corporais coletadas.

A partir dos dados de todos os voluntários do GM, os 15 voluntários foram agrupados e foi calculada a matriz de correlação geral na Figura 40. Para tanto, o mesmo resultado é verificado para o grupo do total de 9 voluntárias do GF, conforme a Figura 41. Como pode-se observar, não houve correlações estatísticas significativas de forma generalizada para os grupos, o que evidencia que não é possível a análise de temperatura corporal considerando unicamente uma variável extraída ao longo do tempo da região facial.

Figura 40 - Matriz de correlação considerando as variáveis temperatura ambiente, temperaturas corporais e variações dos pixels do canal verde do GM.

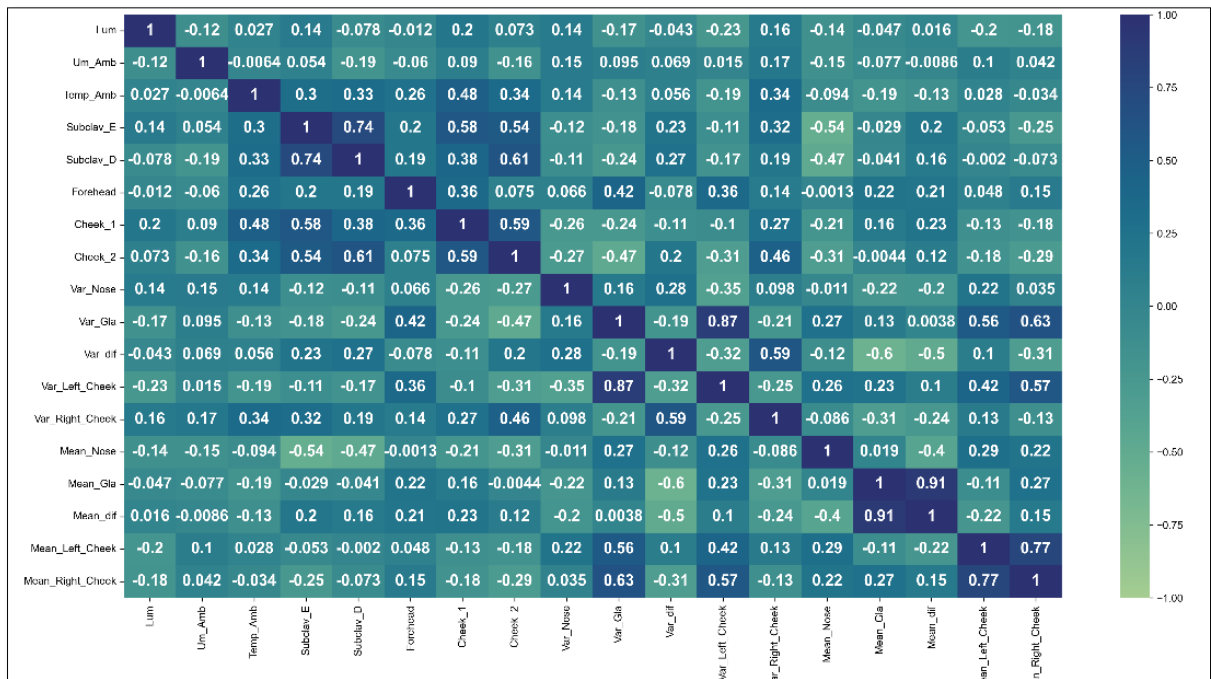
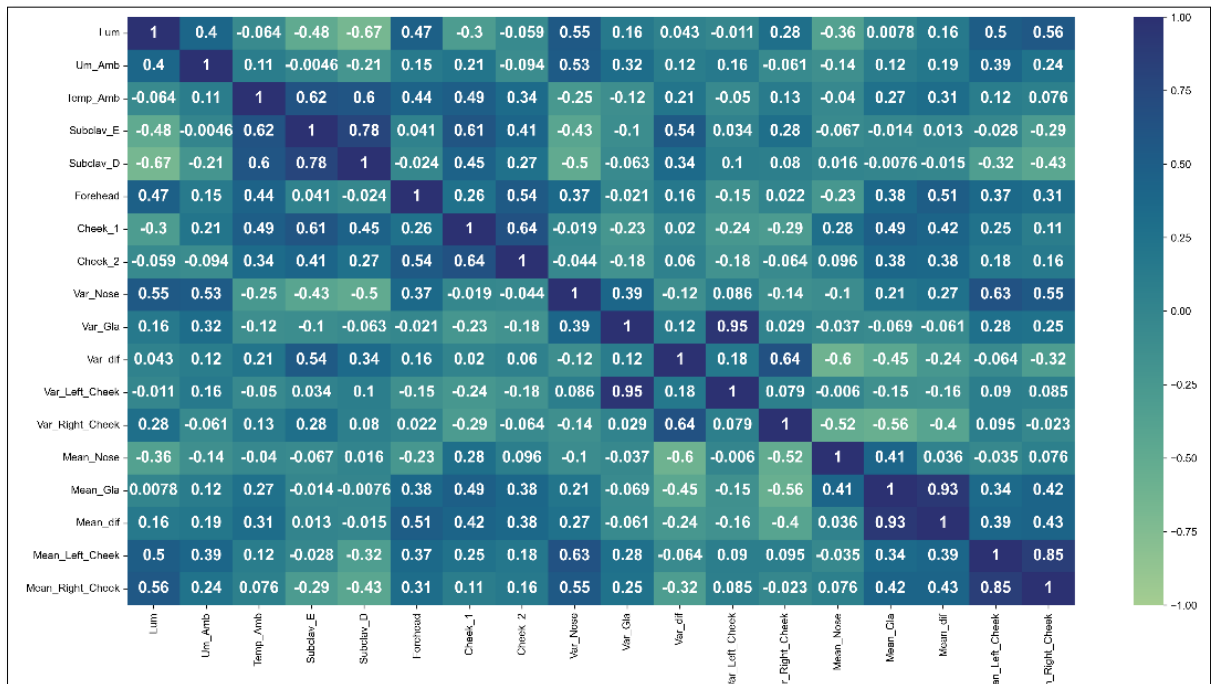


Figura 41 - Matriz de correlação média considerando as variáveis temperatura ambiente, temperaturas corporais e variações dos pixels do canal verde do GF.



A fim de sobrepor os obstáculos apresentados pelas matrizes de correlação generalizadas do GM e GF, considerou-se então, a utilização do método de Regressão de

Componentes Principais (RCP). Assim, a partir da análise de raio de variância e do erro médio quadrático verifica-se a possibilidade de utilização da extração de componentes principais a análise da regressão linear múltipla. O método de treinamento, teste e validação do modelo foi composto por 70% dos dados para treinamento, 30% para teste, e validação cruzada K-fold com 10 interações e 3 splits.

Ao todo foram consideradas dois grupos de variáveis independentes sendo 32 variáveis possíveis para análise e usadas como entrada para o RCP. Com relação à variável dependente foi feita uma análise empírica considerando três casos: a temperatura média da região das bochechas, temperatura média subclaviculares e temperatura da região da testa. Para todas as análises foram consideradas 25 componentes principais visto que esta quantidade apresentou valores de erro médio quadrático abaixo de 0.2 e raio de variância dos dados próximo a 100%.

Conforme a tabela 6 pode-se verificar que para o GM a escolha da temperatura média das bochechas como variável dependente para a predição apresentou desvio padrão médio de 0.118, pontuação  $R^2$  de 0.840 e erro médio quadrático de 0.210 °C. Escolhendo a temperatura média subclaviculares resultada em desvio padrão médio de 0.109,  $R^2$  igual a 0.788 e erro médio quadrático de 0.205 °C. Por fim, para a temperatura da região da testa como variável predita, tem-se o valor de desvio padrão de 0.207,  $R^2$  igual a 0.744 e erro médio quadrático de 0.379. O melhor resultado apresentou-se então para predição da temperatura média das bochechas com maior  $R^2$  considerando 32 variáveis independentes de análise.

Para o GF a escolha da temperatura média das bochechas como variável dependente para a predição apresentou desvio padrão médio de 0.080, pontuação  $R^2$  de 0.938 e erro médio quadrático de 0.154 °C. Escolhendo a temperatura média subclaviculares obtem-se desvio padrão médio de 0.081, pontuação  $R^2$  em 0.921 e erro médio quadrático de 0.146 °C. Por fim, para a temperatura da região da testa como variável predita, tem-se o valor de desvio padrão de 0.102, pontuação  $R^2$  de 0.929 e erro médio quadrático de 0.188. O melhor resultado apresentou-se então para predição também é a temperatura média das bochechas com maior pontuação  $R^2$  considerando 32 variáveis independentes de análise também apresentado na Tabela 6.

Tabela 6 - Análise de RCP com 32 variáveis dependentes incluindo temperatura e umidade ambiente. Ao todo 25 componentes principais foram selecionadas (D.P.: Desvio Padrão).

<b>Grupo Vol.</b>	<b>Var. Dependente</b>	<b>D.P. Médio</b>	<b><math>R^2</math> Score</b>	<b>Erro Médio (°C)</b>
GM	T. Média Bochechas	0.118	0.840	0.210

	T. Média Subclaviculares	0.109	0.788	0.205
	T. Testa	0.207	0.744	0.379
GF	T. Média Bochechas	0.080	0.938	0.154
	T. Média Subclaviculares	0.081	0.921	0.146
	T. Testa	0.102	0.929	0.188

Considerando 32 variáveis independentes, o modelo apresentou melhor desempenho para a predição da temperatura média das bochechas, assim, as Figuras 42 a 49 apresentam os gráficos de erro médio quadrático, abrangência da variância em (%) a partir do número de componentes principais, a análise de predição de temperatura média da bochecha e o desvio padrão em 100 amostras aleatórias para o GM e o GF, respectivamente.

Figura 42 - Erro médio quadrático versus número de componentes principais – Bochechas GM.

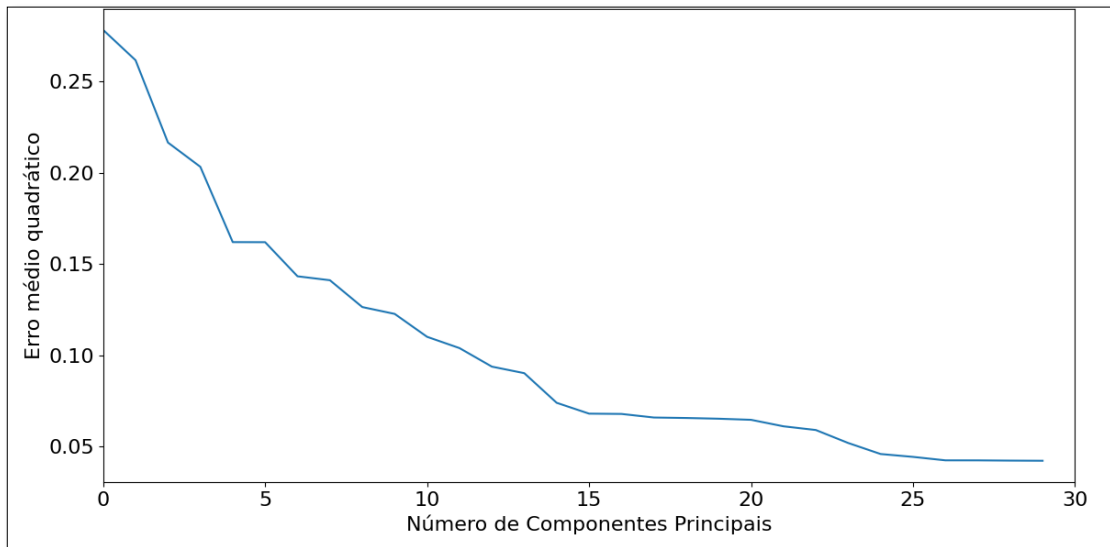


Figura 43 - Abrangência da variância em (%) pelo número de componentes principais - Bochechas GM.

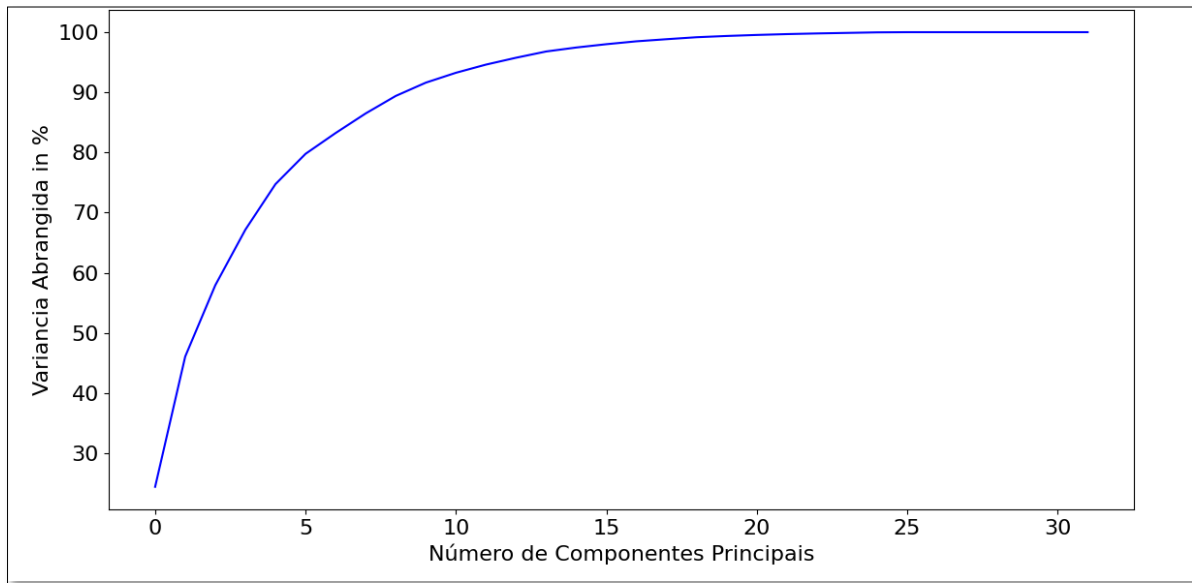


Figura 44 - Análise de predição de temperatura média das bochechas para o GM - curva em azul: valor real; curva em vermelho: valor predito.

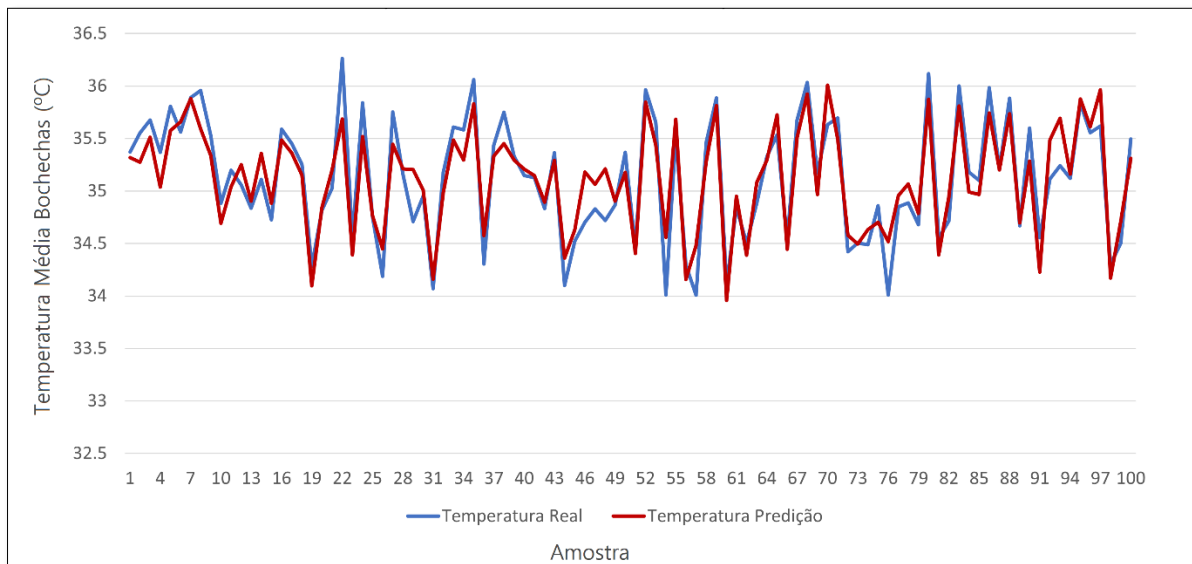


Figura 45 - Apresentação do desvio padrão entre valor real e valor predito para 100 amostras aleatórias, considerando a temperatura média das bochechas do GM.

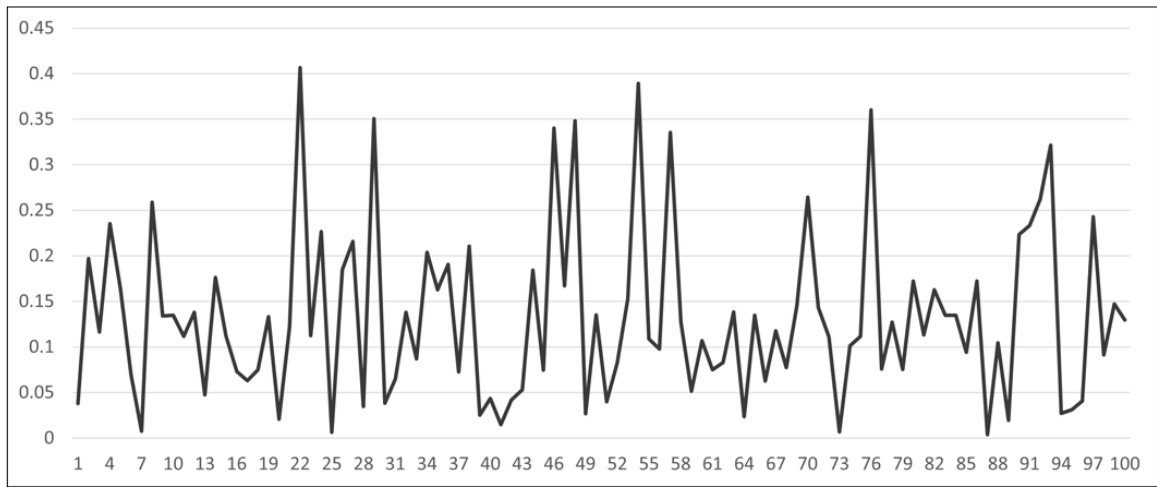


Figura 46 - Erro médio quadrático pelo número de componentes principais – Bochechas GF.

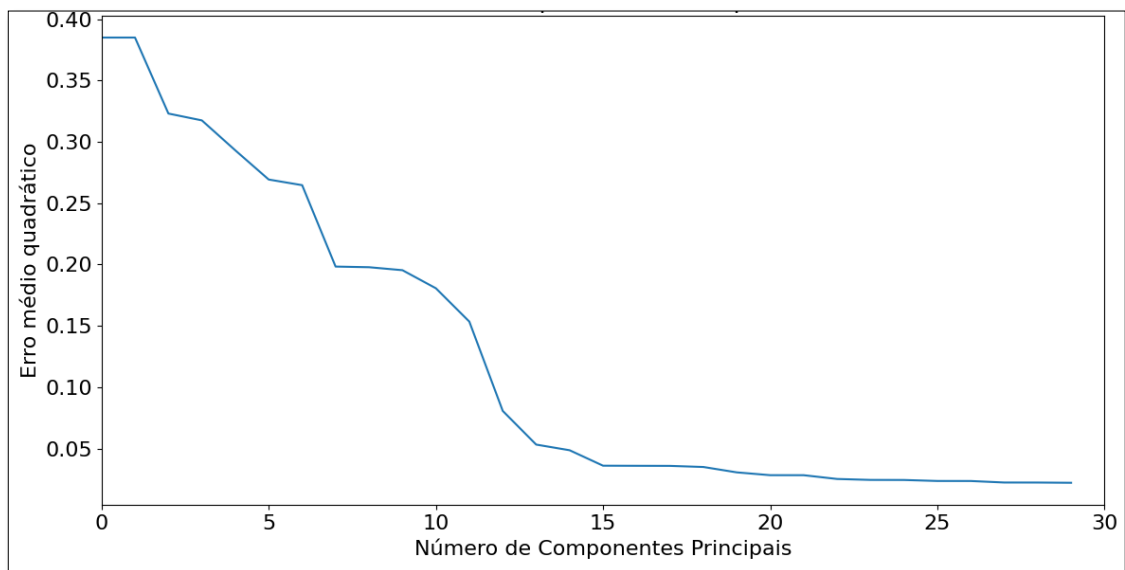




Figura 47 - Apresentação da abrangência da variância em (%) pelo número de componentes principais – Bochechas GF.

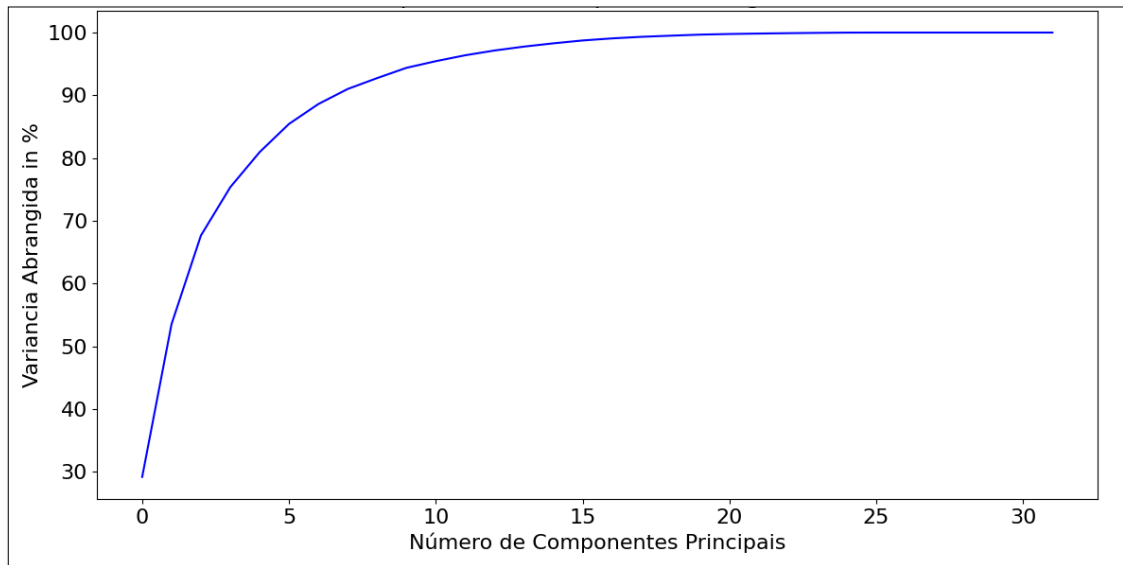


Figura 48 - Análise de predição de temperatura média das bochechas para o GF - curva em azul: valor real; curva em vermelho: valor predito.

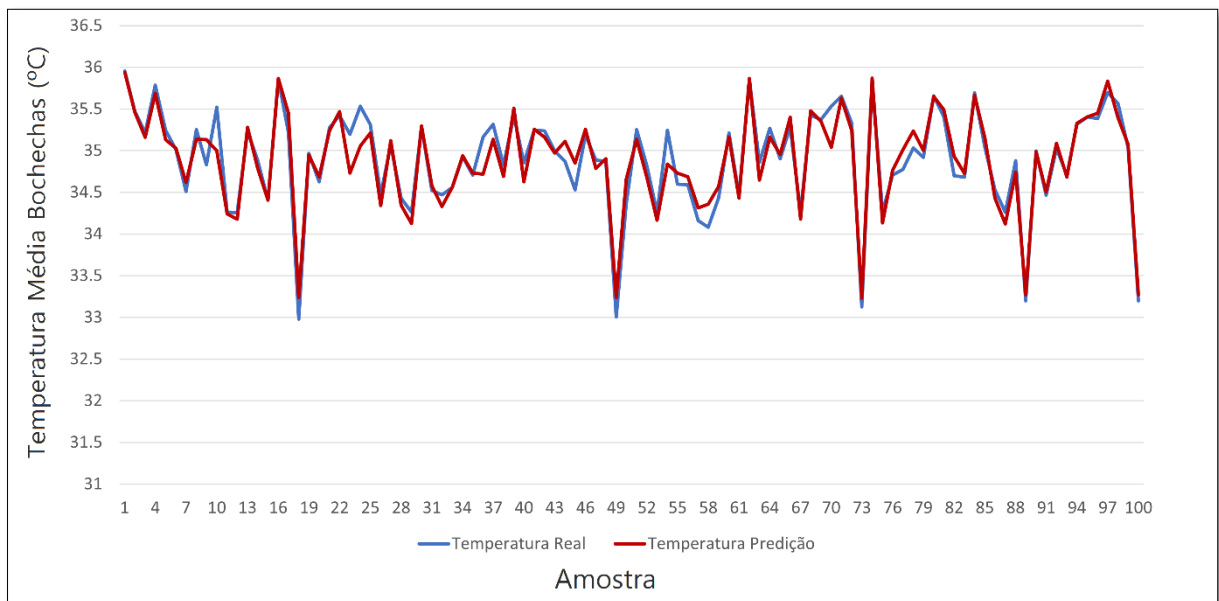
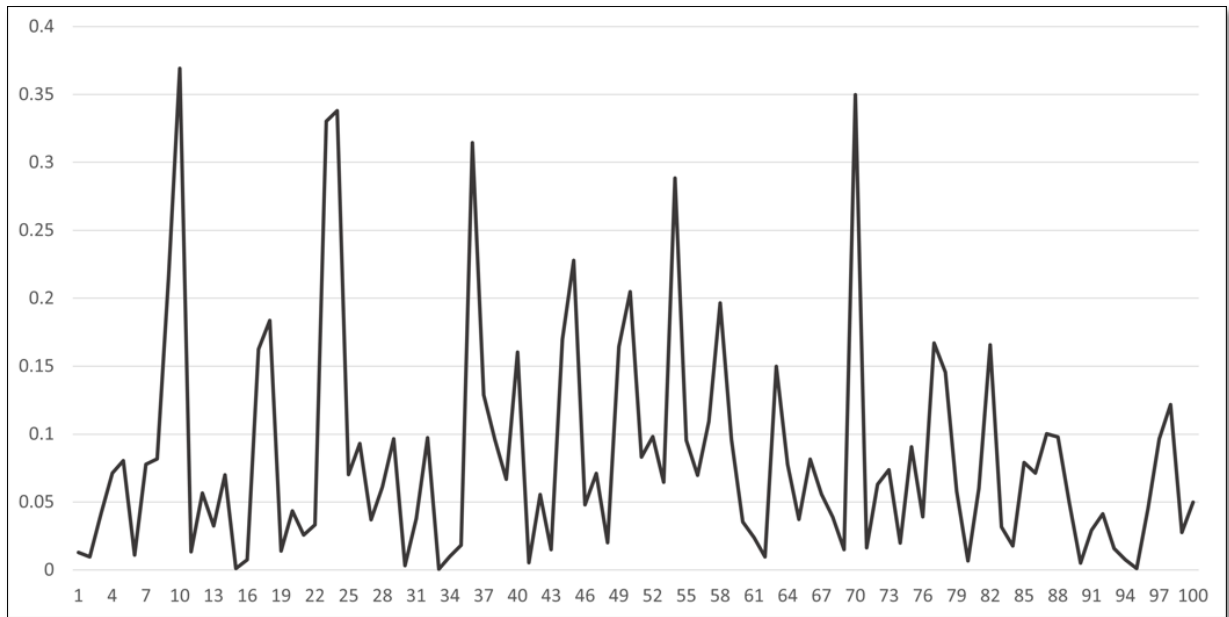


Figura 49 - Apresentação do desvio padrão entre valor real e valor predito para 100 amostras aleatórias, considerando a temperatura média das bochechas do GF.



A fim de verificar a importância da umidade ambiente na análise pelo modelo, foi realizado novamente a análise de dados, desta vez excluindo este dado, ou seja, utilizando 31 variáveis independentes. Conforme a Tabela 7, verifica-se que para o GM a escolha da temperatura média das bochechas como variável dependente para a predição apresentou desvio padrão médio de 0.117, pontuação  $R^2$  de 0.844 e erro médio quadrático de 0.207 °C. Escolhendo a temperatura média subclaviculares obtem-se desvio padrão médio de 0.108, pontuação  $R^2$  em 0.784 e erro médio quadrático de 0.207 °C. Por fim, para a temperatura da região da testa como variável predita, tem-se o valor de desvio padrão de 0.212, pontuação  $R^2$  de 0.732 e erro médio quadrático de 0.388. O melhor resultado apresentou-se então para predição da temperatura média das bochechas com maior pontuação  $R^2$  considerando 31 variáveis independentes de análise.

Para o GF a escolha da temperatura média das bochechas como variável dependente para a predição apresentou desvio padrão médio de 0.080, pontuação  $R^2$  de 0.938 e erro médio quadrático de 0.154 °C. Escolhendo a temperatura média subclaviculares obtem-se desvio padrão médio de 0.086, pontuação  $R^2$  em 0.910 e erro médio quadrático de 0.156 °C. Por fim, para a temperatura da região da testa como variável predita, tem-se o valor de desvio padrão de 0.099, pontuação  $R^2$  de 0.932 e erro médio quadrático de 0.184. O melhor resultado apresentou-se então para predição também é a temperatura média das bochechas com maior pontuação  $R^2$  considerando 31 variáveis independentes de análise, também apresentado na Tabela 7.

Tabela 7 - Análise de RCP com 31 variáveis dependentes incluindo temperatura e umidade ambiente. Ao todo 25 componentes principais foram selecionadas. D.P. : Desvio Padrão.

Grupo Vol.	Var. Dependente	D.P. Médio	R <sup>2</sup> Score	Erro Médio (°C)
GM	T. Média Bochechas	0.117	0.844	0.207
	T. Média Subclaviculares	0.108	0.784	0.207
	T. Testa	0.212	0.732	0.388
GF	T. Média Bochechas	0.080	0.938	0.154
	T. Média Subclaviculares	0.086	0.910	0.156
	T. Testa	0.099	0.932	0.184

Considerando 31 variáveis independentes, o modelo também apresentou melhor desempenho para a predição da temperatura média das bochechas e é demonstrado nas Figuras 50 a 57 os gráficos de erro médio quadrático, abrangência da variância em (%) a partir do número de componentes principais, a análise de predição de temperatura média da bochecha e o desvio padrão em 100 amostras aleatórias para o GM e o GF, respectivamente.

Figura 50 - Erro médio quadrático pelo número de componentes principais – Bochechas GM.

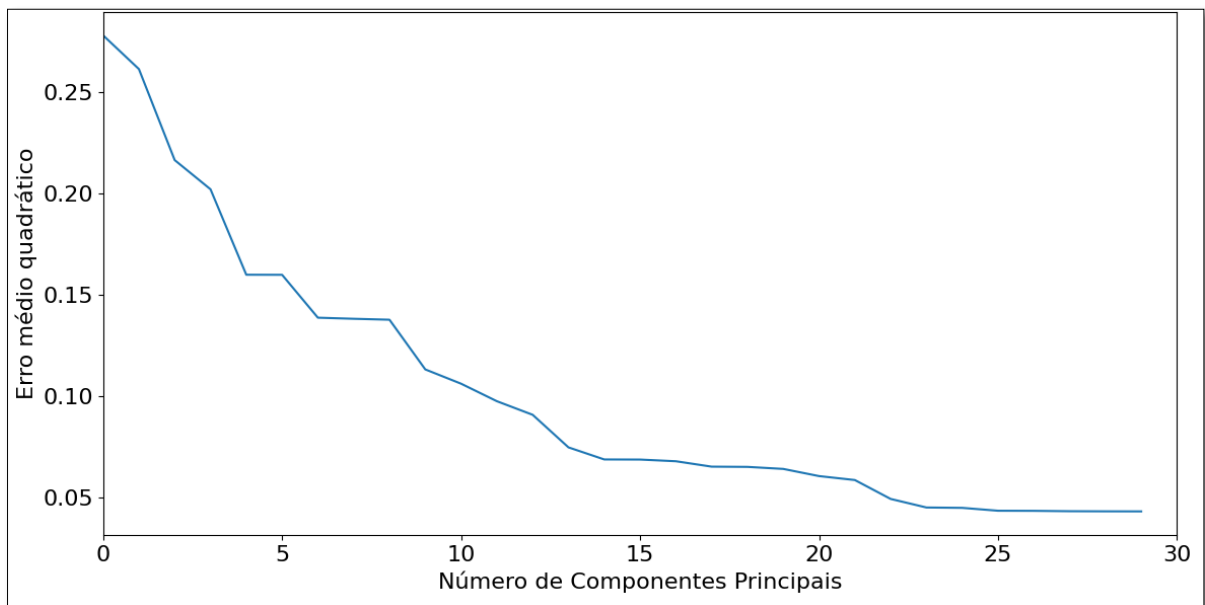


Figura 51 - Apresentação da abrangência da variância em (%) pelo número de componentes principais – Bochechas GM.

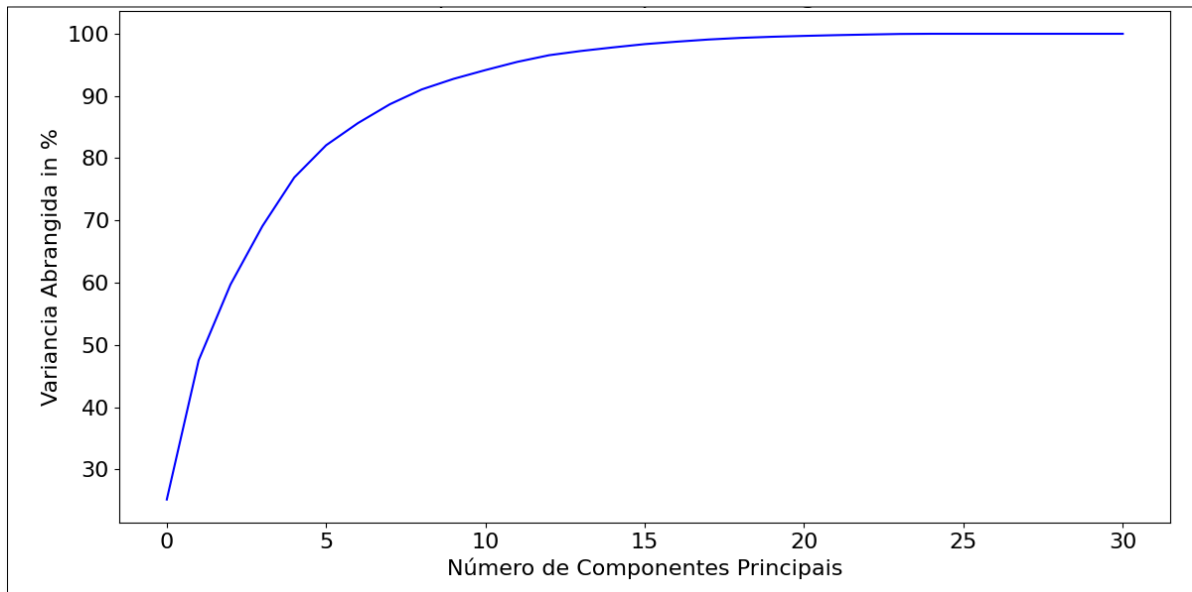


Figura 52 - Análise de predição de temperatura média das bochechas para o GM - curva em azul: valor real; curva em laranja: valor predito.

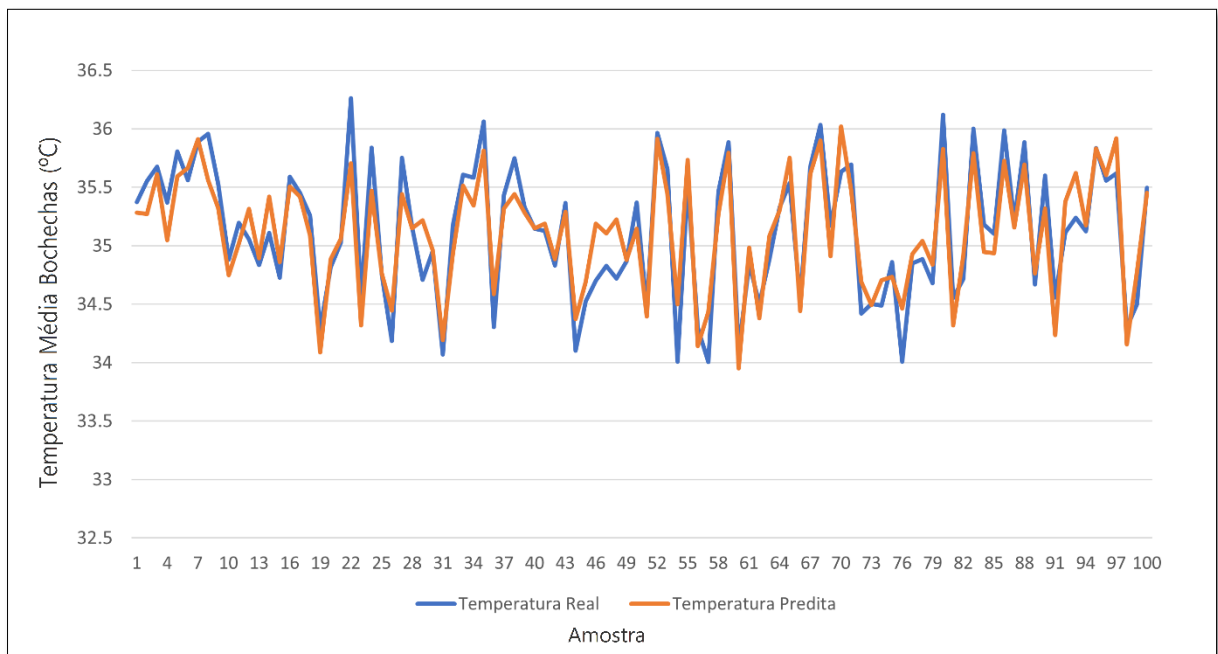


Figura 53 - Apresentação do desvio padrão entre valor real e valor predito para 100 amostras aleatórias, considerando a temperatura média das bochechas do GM.

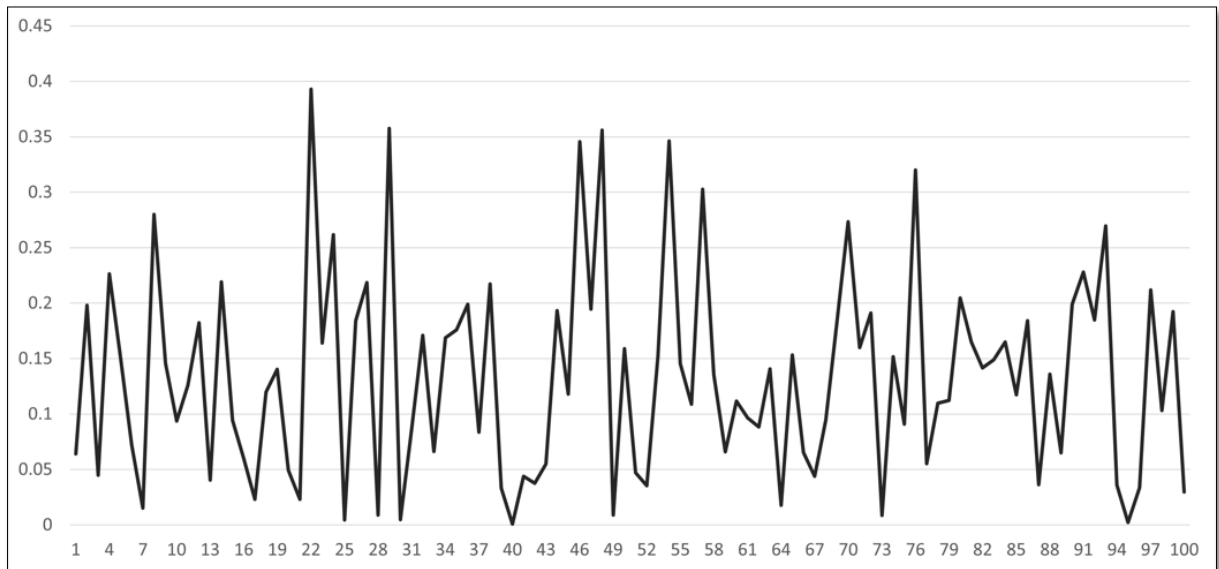


Figura 54 - Erro médio quadrático pelo número de componentes principais - Bochechas GF.

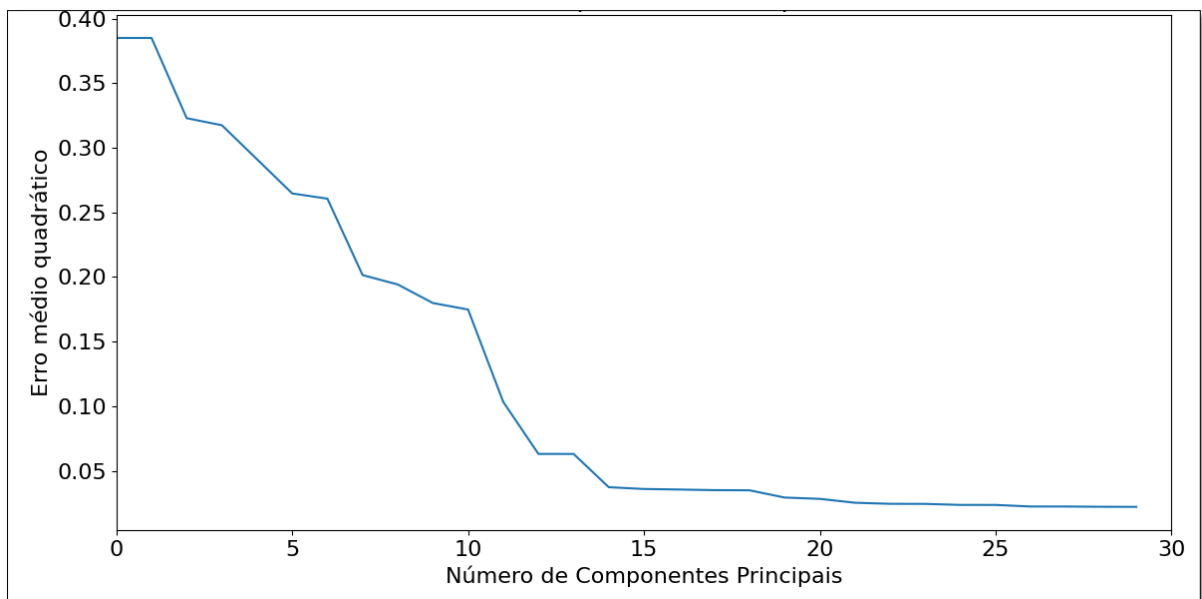


Figura 55 - Apresentação da abrangência da variância em (%) pelo número de componentes principais - Bochechas GF.

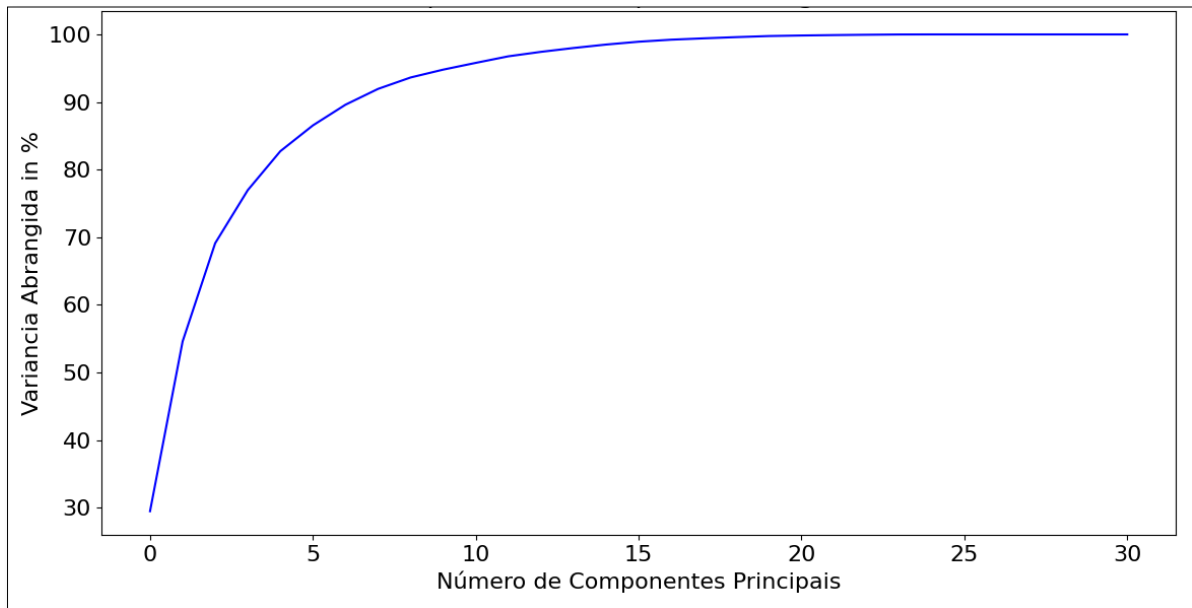


Figura 56 - Análise de predição de temperatura média das bochechas para o GF - curva em azul: valor real; curva em laranja: valor predito.

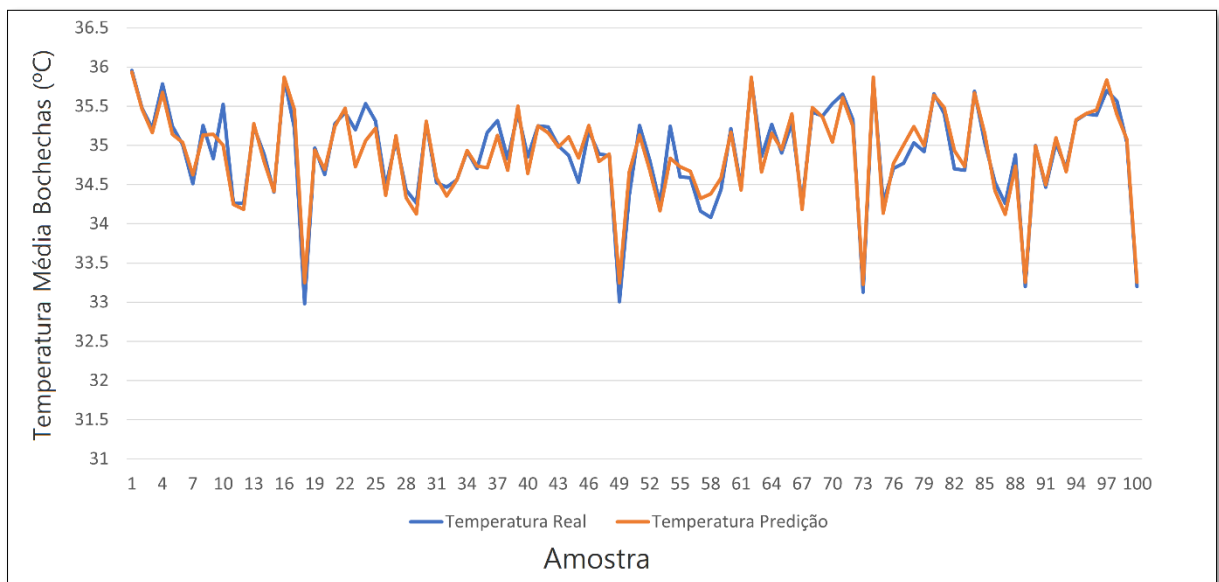
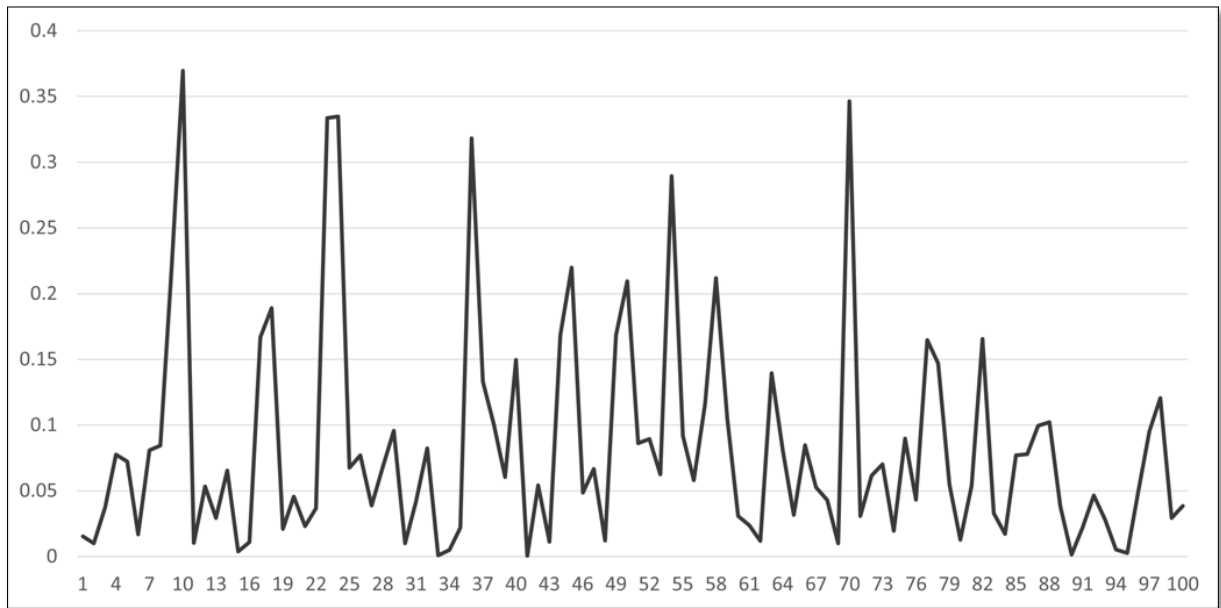


Figura 57 - Apresentação do desvio padrão entre valor real e valor predito para 100 amostras aleatórias, considerando a temperatura média das bochechas do GF.



## 4.2 CLASSIFICAÇÃO DE EXPRESSÕES FACIAIS

Os gráficos a seguir apresentam matrizes de confusão respectivamente para cada banco de dados utilizado e para o classificador MLP e SVM de forma separada. A linha diagonal principal das matrizes demonstram a porcentagem da taxa de precisão, ou seja, verdadeiro positivo dividido pela soma dos resultados verdadeiro positivo com falso positivo.

Os resultados são apresentados em sequência para o banco de dados WSEFEP, CK+, JAFFE e por último uma base de dados mista considerando a união de todos os bancos de dados anteriores. O banco de dados misto foi considerado para avaliar o comportamento do algoritmo em diferentes cenários, gêneros, etnias e idades.

Conforme apresentado abaixo, as imagens 58 e 59 são matrizes com relação ao banco de dados WSEFEP para arquiteturas MLP e SVM, respectivamente. As imagens 60 e 61 são com relação ao banco de dados CK+, as 28 e 29 são com relação ao banco de dados JAFFE, e as imagens 62 e 63 foram feitas classificações a partir da união de todos os bancos de dados, chamado de misto. De acordo com a Imagem 58, as emoções nojo, feliz, surpresa e medo obtiveram acurácia de 100%. Em ordem decrescente, a emoção neutro obteve 75% de precisão, seguido pela emoção raiva com 67% e a emoção triste com 60%. Os atributos selecionados nesta arquitetura e para este banco de dados foram as duplas de pontos 37-41, 44-46, 48-54, 51-57, 21-22, 08-33, 33-54, 20-38 e 31-33. A partir da matriz de confusão é possível verificar quais





De acordo com as imagens 60 e 61, consideramos o banco de dados CK+. Para a arquitetura MLP, encontrou-se 100% de acurácia para a emoção surpresa, 97% para a emoção neutro, 92% para a emoção nojo, 90% para a emoção feliz, 73% para a emoção raiva, 71% para triste e, por fim, 62% para a emoção medo. Os dados selecionados para esta arquitetura foram 24-33, 37-41, 44-46, 48-54, 51-57, 21-22, 08-33, 33-48, 33-54, 23-43, 20-21, 31-33 e 33-35. Para este banco de dados, houve conflitos entre as classes 0, 4 e 6 com a classe 3, sendo respectivamente, as classes reais raiva, triste e medo com a predição da emoção neutro.

Para arquitetura SVM, encontrou-se uma precisão de 100% para as emoções neutro e surpresa. Seguido de 91% para e emoção raiva e 90% para a emoção feliz. Obteve-se 69% de acurácia para a emoção nojo e 50% para a emoção medo. Podemos verificar que houve 100% de confusão na classificação da emoção triste que foi considerado pelo algoritmo como neutro. Os atributos selecionados na arquitetura SVM para o CK+ foram 24-33, 48-54, 51-57, 21-22, 08-33, 33-54, 23-43, 20-38, 20-21, 22-23, 31-33 e 33-35.

Figura 60 - Banco de dados CK+ com seleção de atributos e treinamento por MLP.



Figura 61 - Banco de dados CK+ com seleção de atributos e treinamento por SVM.



De acordo com as imagens 62 e 63, consideramos o banco de dados JAFFE. Para a arquitetura MLP, encontrou-se 100% de acurácia para a emoção nojo, 89% para a emoção neutro, 86% para as emoções raiva e feliz, 83% para a emoção triste, 80% para a emoção medo, e, por fim, 75% para a emoção surpresa. Os atributos selecionados por essa arquitetura foram 24-33, 37-41, 44-46, 48-54, 51-57, 21-22, 08-33, 33-48, 33-54, 23-43, 20-38, 20-21 e 22-23. Conforme na imagem 18, pode-se verificar que houver mais confusões entre as classes do que nos bancos de dados anteriores. Os erros de classificação mais relevantes estão presente entre a classificação de surpresa e medo e entre feliz e medo.

Já para arquitetura SVM, encontrou-se uma precisão de 100% para as emoções raiva, feliz e triste. Seguido de 78% para a emoção neutro e 75% para a emoção surpresa. Obteve-se 60% de acurácia para a emoção nojo e 40% para a emoção medo. Os atributos selecionados neste caso foram 24-33, 37-41, 44-46, 48-54, 51-57, 21-22, 08-33, 33-48, 33-54 e 23-43. Ainda pode-se verificar confusões entre as emoções raiva e nojo, e entre predições de feliz, triste e surpresa com a classe real medo.

Figura 62 - Banco de dados JAFFE com seleção de atributos e treinamento por MLP.



Figura 63 - Banco de dados JAFFE com seleção de atributos e treinamento por SVM.



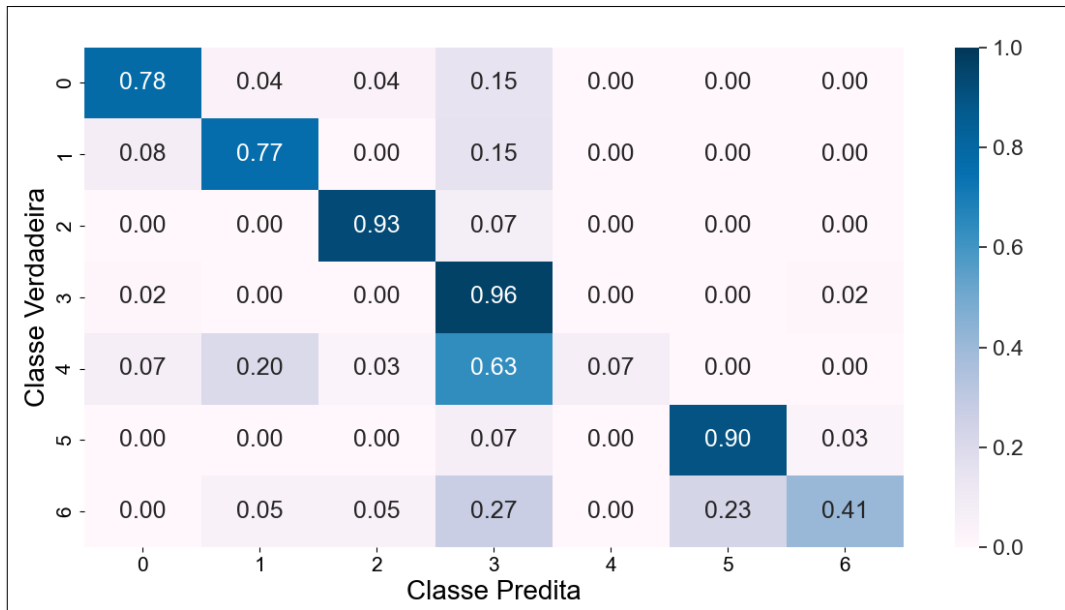
De acordo com as imagens 64 e 65, consideramos o banco de dados misto, ou seja, a união de todas as bases de dados utilizadas neste trabalho. Para a arquitetura MLP, encontrou-se 93% de acurácia para as emoções feliz, neutro e surpresa, 81% para a emoção raiva, 73% para a emoção nojo, 41% para a emoção medo e 37% para a emoção triste. Os atributos selecionados por essa arquitetura foram 19-33, 24-33, 37-41, 44-46, 48-54, 51-57, 21-22, 08-33, 33-48, 33-54, 23-43, 20-38, 20-21, 22-23, 31-33 e 33-35. Nesta união entre todos os bancos de dados, destaca-se principalmente confusões do algoritmo entre predição da classe neutro e classe real triste, e predição da classe surpresa e classe real era medo.

Para arquitetura SVM, encontrou-se uma precisão de 96% para a emoção neutro. Seguido de 93% para a emoção feliz e 90% para a emoção surpresa. Obteve-se 78% de acurácia para a emoção raiva e 77% para a emoção nojo. Por fim, encontra-se 41% para a emoção medo e 7% para a emoção triste, que novamente foi confundida pelo algoritmo com a emoção neutro. Os atributos selecionados neste caso foram 24-33, 37-41, 44-46, 48-54, 51-57, 21-22, 08-33, 33-48, 33-54, 23-43, 20-38, 20-21, 22-23, 31-33 e 33-35. Ainda permanece as confusões de classificação entre predição de neutro sendo classe real triste, e, predição de neutro e surpresa sendo que a classe real é medo.

Figura 64 - Banco de dados misto com seleção de atributos e treinamento por MLP.



Figura 65 - Banco de dados misto com seleção de atributos e treinamento por SVM.



A Tabela 8 apresenta de forma resumida as informações anteriormente citadas com relação ao banco de dados utilizado, o seletor de atributos e algoritmo de treinamento, as emoções e suas respectivas acurácias e a acurácia média entre todas as classes. As emoções com as melhores taxas de precisão ao longo dos testes foram feliz, neutro e surpresa. Por outro lado as emoções que obtiveram menor precisão na classificação foram triste e medo.

A melhor acurácia encontrada foi com relação a arquitetura MLP, com relação ao banco de dados CK+, neste caso, a emoção surpresa apresentou 100% de acerto. Enquanto as emoções nojo, feliz e neutro apresentaram acurácia maior ou igual a 90%. É importante ressaltar que entre as categorias de emoções, para o CK+, não possui a mesma quantidade de fotos entre as emoções, há emoções que apresentam mais imagens para treinamento do que outras.

Tabela 8 - Base de Dados, arquitetura de classificador utilizado, taxa de acurácia por emoção e a média da acurácia.

Base de Dados	Classif.	Raiva	Nojo	Feliz	Neutro	Triste	Surpresa	Medo	Acc Média
WSEFEP	MLP	67%	100%	100%	75%	60%	100%	100%	81,13%
	SVM	78%	83%	100%	100%	70%	83%	100%	86,79%
CK +	MLP	73%	92%	90%	97%	71%	100%	62%	<b>93,80%</b>
	SVM	91%	69%	90%	100%	0%	100%	50%	92,48%
JAFFE	MLP	86%	100%	86%	89%	83%	75%	80%	85,11%
	SVM	100%	60%	100%	78%	100%	75%	40%	80,85%

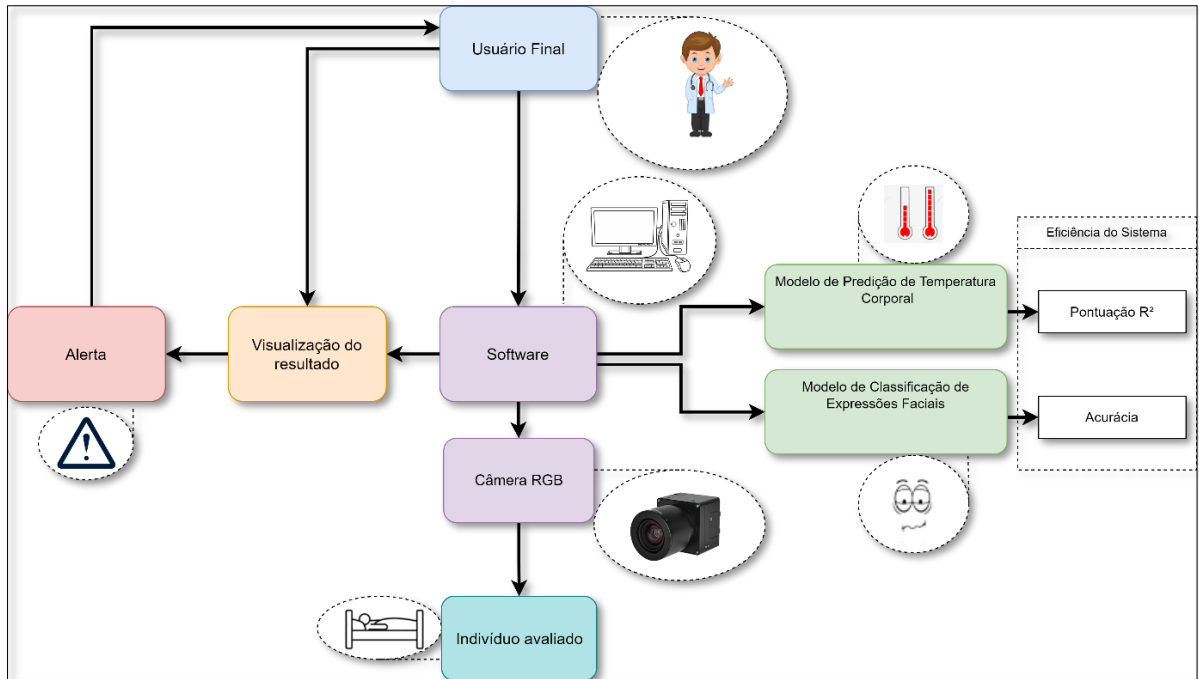
Misto	MLP	81%	73%	93%	93%	37%	93%	41%	81,53%
	SVM	78%	77%	93%	96%	7%	90%	41%	80,31%

Em resumo, as principais distâncias selecionadas em todas as execuções do algoritmo AG com MLP foram 37-41, 44-46, 48-54, 51-57, 21-22, 08-33, 33-54. As distâncias selecionadas em todas as execuções para o modelo SVM foram 24-33, 48-54, 21-22, 08-33, 33-54, 23-43, 20-38. Considerando a análise para MLP e SVM as distâncias selecionadas em todas as execuções foram 48-54, 21-22, 08-33, 33-54.

### 4.3 FRAMEWORK DO MODELO PROPOSTO

A seguir é apresentado o framework do sistema integrado proposto pelo presente trabalho na Figura 66. O framework fornece uma estrutura que posteriormente é utilizado como base para o desenvolvimento de softwares, aplicativos e projetos digitais para utilização por usuários finais. Assim, o usuário final utiliza o software com o sistema proposto. O software por sua vez adquire os dados a partir de uma câmera RGB e processa os sinais de entrada do usuário avaliado. Em seguida, realiza-se a predição de temperatura corporal e a classificação de expressões faciais mediante desempenho dos modelos desenvolvidos. As métricas para apresentação do resultado são apresentação do valor do perfil térmico do indivíduo com relação a pontuação  $R^2$  alcançada pelo modelo e acurácia da emoção melhor classificada. Por fim, a partir do software o usuário final consegue visualizar o resultado apresentado pelo sistema integrado e receber alertas do sistema caso os resultados apresentados exijam atenção maior ao indivíduo monitorado.

Figura 66 – Framework do modelo proposto.



## 5 DISCUSSÃO

### 5.1 PREDIÇÃO DE TEMPERATURA CORPORAL

Molinaro et al. (2020) apresenta uma revisão bibliográfica sobre a utilização de câmeras digitais em análises de sinais vitais sem contato. Entretanto, este estudo não apresentou nenhum estudo com relação a análise de temperaturas corporais pois afirmou que não há evidências literárias de um método padrão ouro para medição de temperatura corporal no domínio visível [MOLINARO, 2020]. O presente estudo então, visto a escassez do estado-da-arte em abordagens com relação à temperatura corporal, visa como pesquisa e enfrentamento a análise deste tópico da classe de sinais vitais em domínio visível sem contato.

O estudo de Ernst et al. (2021) analisa a taxa de batimentos cardíacos baseados em câmera por diferentes espaços de cores. As regiões analisadas foram glabella (GLB), região pequena da testa (FHS), região larga da testa (FHL), região alta da bochecha direita (RUP) e esquerda (LUP), bochecha direita (RON) e bochecha esquerda (LON). O canal de cor verde, da cartela de cores RGB, apesar de não ter apresentado a melhor acurácia entre todos os espaços de cores investigados no estudo, apresentou uma acurácia de 50% para a região da glabella e pouco mais de 40% para considerando as outras regiões. É interessante apresentar que Ernst afirma que a maioria dos sistemas tecnológicos atuais apresentam uma aquisição de sinais com os canais de cores RGB para cada pixel, pois apresentam câmeras RGB embarcadas nos dispositivos.

O estudo de Wu et al. (2022) apresenta uma investigação sobre reconhecimento e previsão do conforto térmico individual com base na temperatura local da pele e utilizaram regiões faciais como testa, olhos, nariz, orelhas e bochechas. Neste estudo, os pesquisadores variaram a temperatura do ambiente entre 16 e 30 °C e captaram por sensores térmicos as temperaturas corporais das regiões anteriormente citadas. Durante a variação da temperatura ambiente, em 16°C, a região da testa se alterou entre 33-34°C. A região das bochechas se alteraram entre 31,5-32°C, a temperatura do nariz variou de 26-31°C e a região das orelhas variou de 27.5-31°C. Com o aumento da temperatura ambiente para 30°C, a região da testa alterou entre 36.2-37°C, a temperatura das bochechas se alterou entre 36-36.7°C, a temperatura da região do nariz variou entre 36.2 e 36.4°C e a temperatura das orelhas variou entre 34.2-36.4°C [WU, 2022].

As diferentes variações de temperatura entre as regiões faciais apresentadas destacam que regiões com estrutura anatômicas com maior presença de cartilagem são mais propícias a



interferência da temperatura do ambiente, do que regiões com maior circulação sanguínea sendo glabella, testa e bochechas. Conforme apresentado em Wu et al, as variações de temperatura apresentada na região do nariz e na região das orelhas obtiveram maior intervalo variação de temperatura de acordo com a temperatura do ambiente [WU, 2022].

Em Xuan et al, 2022, é apresentado uma abordagem em temperatura facial de acordo com imagens infravermelhas térmicas em comparação com a temperatura da superfície da orelha em um cenário configurado para 23 °C. De acordo com este estudo, comparou-se a temperatura predita da região da testa, da região do nariz, e os 20 pixels com maior intensidade a partir de imagens infravermelho, com a temperatura adquirida com sensor térmico na região da orelha.

Assim, encontraram que a testa e o nariz apresentaram temperaturas de 35,17°C e 30.97 °C, respectivamente, e os 20 pixels mais quentes eram 36.34 °C, ou seja, a temperatura de homeostase corporal do indivíduo. É importante ressaltar que neste estudo, os pesquisadores consideram a média do vetor de pixels como uma medida de acurácia e representatividade das mudanças superficiais na pele humana. E, ainda, realizam a aplicação de regressão linear como a melhor método para análise dos dados.

No presente trabalho, as temperaturas adquiridas entre os voluntários se apresentaram em torno de 36 °C para a região da testa, e 35 °C para a região das bochechas. Apesar de não ter sido adquirida a temperatura do nariz, visto que iria interferir da aquisição dos pixels da região, a alteração na média dos valores dos pixels da cor verde são presentes nos gráficos apresentados na seção 4.1 em análise dos resultados.

O estudo de Jia et al. (2022) apresentou o desenvolvimento de um sistema automático baseado na temperatura da pele e conforto térmico do indivíduo. Neste artigo os pesquisadores dividiram os grupos de estudos entre voluntários homens e voluntárias mulheres e concluíram que a temperatura da pele superficial da região facial indicou diferenças entre os sexos. As mulheres têm menor temperatura da pele facial em condições de frio e calor. Assim, o gênero do voluntário é um fator importante na avaliação da percepção térmica, e portanto, para o presente estudo, na avaliação de temperatura corporal. Por esse fator, o presente trabalho realizou a divisão entre grupo de voluntários homens e voluntárias mulheres.

Outros achados do estudo de Jia et al. (2022) são apresentados que, primeiramente, diferentes partes faciais podem ter diferentes temperaturas de pele nas mesmas condições de conforto/sensação térmica para um mesmo indivíduo. Em segundo, diferentes indivíduos apresentam mudanças distintas na temperatura da pele facial e apresentam diferentes tendências

em diferentes situações térmicas, ou seja, de frio para calor ou calor para frio configura uma variabilidade às repostas de temperaturas corporais.

Assim, Jia et al (2022) conclui que quando a sensação térmica muda de frio para quente, a temperatura da pele facial tende a aumentar, mas não de forma linear. Alguns indivíduos apresentam a temperatura mais alta da pele na testa em situações de frio, mas em situações quentes, a temperatura mais alta da pele muda para a região da bochecha. Apresentam também que alguns indivíduos possuem a mesma temperatura de pele mais alta na região da testa tanto em situações de frio quanto em situações de calor [JIA, 2022]. Assim, é notório que a variação da temperatura ambiente reflete de forma imprecisa nas temperaturas corporais de cada indivíduo.

O artigo de Zheng et al. (2022) apresenta a análise de taxa de batimentos cardíacos a partir da análise da região facial de forma remota. Neste estudo, os autores utilizam as coordenadas faciais para automatização da seleção e demarcação de regiões de interesse na face, que no caso deles, seleção da região da testa. A partir da decomposição dos canais, Zheng et al, utilizou o canal verde como foco de filtragem e processamento do sinal e detectar o pulso de volume de sangue, então chamado de BVP, ao longo do tempo [ZHENG, 2022]. Assim, o presente estudo também considerou as regiões da glabella, testa, bochechas e nariz de acordo com as abordagens dos estudos de Ernst, Metzmacher, Wu e Zheng [ERNST, 2021; METZMACHER, 2018; WU, 2022; ZHENG, 2022].

No estudo de Wu et al. (2022) utilizaram-se técnicas de aprendizado de máquina regressão logística (LR), máquinas de vetor de suporte (SVM) e floresta aleatória (RF) para averiguação dos resultados de conforto térmico. Em adição, no estudo de Przybylo et al. (2022) apresenta uma proposta de abordagem de estimativa de taxa de batimentos cardíacos em imagens em domínio visível, com foco no canal verde, utilizando memória de longo prazo rede neural profunda (LSTM) de rede de aprendizagem profunda. Expõe, também, que o estado-da-arte dispõe de análises mediante o canal verde como análise de componentes independentes (ICA) e plano ortogonal à pele (POS) [PRZYBYLO, 2022].

No presente estudo, bem como apresentado na literatura, também são utilizadas técnicas de aprendizado de máquina para investigação dos resultados. Utiliza-se a regressão linear múltipla com seleção de componentes principais (RCP). A análise de componentes principais tem como base a análise de grandes conjuntos de dados contendo um grande número de dimensões/características por observação, aumentando a interpretabilidade dos dados enquanto preserva a quantidade máxima de informações e permite a visualização de dados multidimensionais.

Foi verificado ao longo dos gráficos de temperatura corporal por tempo e intensidade do pixel por tempo, que há variações significativas entre os três canais RGB para cada voluntário. Como o estudo de Ernst et al. (2021), apresenta uma combinação linear entre os canais RGB para medição de batimentos cardíacos, o presente estudo não se restringiu a apenas um canal de cor pois restringir a análise poderia causar perda de informação.

Os resultados apresentados para análise de componentes principais para os canais RGB considerando ao todo 25 componentes são bastante promissores. Para o GM a predição da temperatura média das bochechas apresentou uma pontuação  $R^2$  de 0.84, ou seja, quanto mais próximo a 1 melhor a adaptação da regressão aplicada à variabilidade dos dados, e a pontuação alcançada é muito favorável. Em segundo lugar como variável dependente, a temperatura média subclaviculares apresentou uma pontuação  $R^2$  de 0.79. Ambas variáveis analisadas apresentaram um desvio padrão médio para 100 amostras aleatórias na faixa de 0.1 °C e erro médio quadrático de 0.2 °C.

De acordo com a análise do GF, os resultados permaneceram semelhantes, sendo a temperatura média das bochechas a variável dependente com melhor desempenho. Para esta variável predita encontrou-se uma pontuação  $R^2$  também de 0.9380 e para a temperatura média subclaviculares tem-se uma pontuação  $R^2$  da regressão de 0.9101. Também, para ambas variáveis preditas o desvio padrão médio para 100 amostras aleatórias foi na faixa de 0.08 °C e erro médio quadrático de 0.15 °C. É importante ressaltar que para ambos grupos de gêneros, não é necessário considerar a umidade ambiente, mas é imprescindível a temperatura ambiente como variável independente.

Cardone et al. (2021) investigou a predição da temperatura corporal baseado em imagens em infravermelho alcançou a estimativa de temperatura regional da pele na ordem de RMSE de 0,09 °C [CARDONE, 2021]. O RMSE do presente trabalho é de 0,44 °C com a predição de temperatura média das bochechas, e 0,38 °C com a predição da temperatura média subclavicular, demonstrando uma proximidade nos resultados encontrados sem a necessidade de câmeras termográficas.

Como Kiekkas et al. (2007) apresenta que a febre interfere diretamente em sinais vitais como frequência cardíaca e saturação de oxigênio [KIEKKAS, 2007]. A alteração da temperatura corporal está intimamente relacionada, então, à disfunção de outros sinais vitais, tornando o registro contínuo essencial para o monitoramento de pacientes em terapia intensiva. Como a literatura não apresenta análise de temperatura corporal em domínio RGB, o presente estudo é o primeiro projeto com foco neste tipo de metodologia e que apresentou resultados

bastante encorajadores para o andamento da pesquisa e disponibilização em aplicações em unidades de terapia intensiva.

## 5.2 CLASSIFICAÇÃO DE EXPRESSÕES FACIAIS

O reconhecimento de expressões faciais desempenha um papel importante em diferentes áreas e aplicações, como compreensão do comportamento humano, sistemas de segurança, interações humanas e informáticas, sistemas de monitoramento facial do motorista e ensino à distância (e-learning). Essas aplicações se beneficiam do desenvolvimento alcançado no aprendizado de máquina e na computação afetiva que são vistas como uma disciplina fascinante para tratar e explorar as interações homem-máquina. Sistema de reconhecimento de expressões faciais baseado em computação afetiva oferece aos computadores a capacidade e habilidade de reconhecer, sintetizar e classificar as emoções [SLIMANI, 2018]. Porém, há uma escassez de aplicação de reconhecimento de expressões faciais em pacientes acamados de UTIs e CTIs.

Rahu et al. (2010) cita que mais de 50% dos pacientes conscientes intubados relataram dor enquanto em repouso e 80% durante a exposição nociceptiva, como virar o paciente de posição. Avaliações imprecisas da dor e o consequente tratamento inadequado da dor em adultos criticamente doentes podem levar a consequências fisiológicas significativas, como aumento da carga de trabalho miocárdica, que pode levar à isquemia miocárdica ou troca gasosa prejudicada, que pode resultar em insuficiência respiratória [RAHU, 2010].

Canal et al. (2022) apresentou uma revisão das arquiteturas dos algoritmos de reconhecimento de emoções no cenário de atividades cotidianas e os estudos em dois grupos: abordagens de aprendizado de máquina clássicas e abordagens baseadas em redes neurais [CANAL, 2022]. Com exceção das redes convolucionais, os métodos mais comuns para analisar os dados faciais são Máquina de Vetor de Suporte (SVM) e Perceptron Multi-camadas (MLP). Por esta razão, ambas as arquiteturas foram escolhidas e utilizadas neste estudo.

Os principais estudos na literatura abordam que há alguns anos já é possível usufruir de técnicas de aprendizados de máquina para realizar a classificação de emoções faciais com eficiência satisfatória. Porém, as arquiteturas apresentadas são complexas para extração de dados, realizam compressão e reconstrução de dados e necessitam de uma alta gama de imagens para o treinamento do algoritmo.

O artigo de Najah et al, 2017 propõe o uso de redes neurais convolucionais para estimar emoções a partir de imagens faciais. O estudo avaliou sete emoções básicas em seu estudo

sendo raiva, nojo, medo, alegria, tristeza, surpresa e neutro. Ao todo o modelo apresentou, para dois bancos de dados distintos, uma precisão de acurácia média entre todas as emoções classificadas de 55,9% e 57,2% nos conjuntos de dados CK+ e JAFFE, respectivamente. Os autores também descobriram que o modelo desenvolvido foi capaz de melhorar a precisão da estimativa de emoção em comparação com os métodos tradicionais. Assim, apresentam o potencial do uso de redes neurais profundas para estimativa de emoção a partir de imagens faciais [NAJAH, 2017].

Conforme apresentado em Najah et al, a emoção que alcançou a maior acurácia de classificação foi felicidade, com uma precisão de 62,8% no conjunto de dados CK+ e uma precisão de 55,8% no conjunto de dados JAFFE. Já a emoção que obteve maior confusão em sua classificação foi a emoção da tristeza com uma precisão de 44,5% no conjunto de dados CK+ e uma precisão de 42,1% no conjunto de dados JAFFE. Portanto estudo concluiu que as CNNs podiam ser usadas para estimar emoções de forma eficaz a partir de imagens faciais [NAJAH, 2017].

O artigo de Li et al. (2020) propõe o uso de uma rede neural convolucional baseada em mecanismo de atenção. O mecanismo de atenção permite que a rede se concentre em regiões importantes da face mais salientes e ressaltando as características das regiões dos pontos de interesse para assim tomar uma decisão de acordo com a probabilidade de classificação entre as categorias de emoções [LI, 2020].

Os resultados do estudo de Li et al. (2020) foram de acordo com a utilização dos bancos de dados JAFFE, CK+, FER2013 e Oulu-CASIA. Abrangendo, assim, a setes emoções básicas sendo: raiva, nojo, medo, felicidade, tristeza, surpresa e o neutro. O conjunto de dados FER 2013 contém 35.887 imagens, o conjunto de dados CK+ contém 593 imagens, o conjunto de dados Oulu-CASIA contém 6.887 imagens e o conjunto de dados JAFFE contém 213 imagens [LI, 2020].

Li et al. (2020) alcançou uma precisão média entre todas as emoções de 96,4% no conjunto de dados FER 2013, 98,68% no conjunto de dados CK+, 95,5% para o conjunto de dados Oulu-CASIA e 90,3% para o conjunto de dados JAFFE. A emoção que obteve melhor classificação foi a felicidade, com 98,3% para FER2013 e 98,2% para Oulu-CASIA, e surpresa com 99,3% para o CK+ e 93,7% para o JAFFE. As emoções que obtiveram menor acurácia na classificação foi a emoção surpresa com 92,2% na base de dados FER2013, a emoção medo com 96,6% para o CK+, também a emoção medo com 94,2% para o Oulu-CASIA e, por fim, a emoção nojo com precisão de 81,1% para a base de dados JAFFE [LI,2020].

Yao et al. (2021) utilizou aprendizado ativo para reconhecimento de unidades de ação (UAs) na face e as utilizou como entradas na máquina de vetor de suporte (SVM) para classificação de expressão facial. Os autores avaliaram a abordagem proposta utilizando também o conjunto de dados CK+. Os resultados mostram que a abordagem proposta superou o estado-da-arte para as emoções felicidade, tristeza, raiva, nojo e neutro, e apresentou pior desempenho para medo e surpresa. No geral, o algoritmo desenvolvido neste estudo alcançou uma precisão média de 93,7% no reconhecimento das expressões faciais [YAO, 2021].

O artigo de Pu et al. (2021) propõe o uso de um modelo de aprendizagem de representação com restrição de conhecimento, o qual é um tipo de rede convolucional. Apresentam então um framework para reconhecimento de emoções faciais nas imagens que combine as unidades de ação facial (AU) com aprendizado não-supervisionado da emoção. Apresentam que a literatura em geral utiliza somente dados de entrada em ambientes controlados e que o framework desenvolvido pelos pesquisadores contorna este problema visto que as UAs são intrínsecas a qualquer indivíduo saudável, em qualquer ambiente. Este estudo apresentou uma acurácia média entre todas as emoções de 81%. A emoção que obteve melhor classificação foi felicidade com 94,1% e a que obteve maior confusão foi nojo com 67,6% de precisão [PU, 2021].

O estudo de Minaae et al. (2021) propõe uma rede neural convolutiva por atenção (RNCA) o qual é um modelo híbrido de aprendizado profundo pois combina redes neurais convolucionais (RNN) e redes de memória de curto prazo (LTSM). A rede proposta consiste em três componentes principais sendo camadas convolucionais, camadas LTSM e camadas de atenção. A atenção neste caso se configura a um recurso aplicado a CNN que se destina a imitar a atenção cognitiva [MINAEE, 2021].

As camadas convolucionais são usadas, neste caso, para capturar características espaciais de imagens faciais. As camadas LTSM capturam informações temporais aprendendo as relações temporais entre as expressões faciais. Finalmente, as camadas de atenção são usadas para focar nas expressões faciais mais relevantes para cada expressão facial. Os resultados mostraram que o modelo proposto por Minaae et al. (2021) alcançou uma precisão média de 98% para o banco de dados CK+, superior aos modelos do estado-da-arte. Para a base de dados JAFFE, apresentou uma acurácia média de 92,8% entre as emoções e 99,3% para o banco de dados FERG [MINAEE, 2021].

O artigo de Ramis et al. (2022) propõe uma abordagem de aprendizado por transferência para adaptar modelos de reconhecimento de expressões faciais pré-treinados a diferentes bancos de dados. Os autores avaliaram o modelo em quatro conjuntos de dados sendo CK+, BU-4DFE,

JAFFE e WSEFEP. O modelo proposto atingiu uma acurácia média de 93,15% sendo a melhor acurácia para felicidade com 100% e a pior para a emoção triste com 84,88%. Para o banco de dados WSEFEP, o modelo proposto por Ramis et al. (2022) superou a literatura e atingiu acurácia média de 87,22% entre todas as classes de emoções básicas [RAMIS, 2022].

O estudo de Canal et al. (2022) apresenta uma revisão sistemática de 91 artigos a respeito das técnicas atualmente utilizadas para reconhecimento de emoções faciais. O estudo divide técnicas de treinamento de dados em dois grupos, métodos clássicos baseados em regressão e métodos baseados em redes neurais. Os métodos clássicos de aprendizado de máquina envolvem o uso de algoritmos de aprendizado supervisionado, como Máquina de Vetor de Suporte (SVMs), árvore de decisão e regressão logística. O emprego destas técnicas geralmente são empregadas como modelos supervisionados e dependem dos dados rotulados para treinar o algoritmo na classificação das expressões faciais. Apresenta, ainda, que estes modelos detêm menor precisão do que os modelos baseados em redes neurais [CANAL, 2022].

Os métodos baseados em redes neurais avaliados, segundo Canal et al. (2022) também são conhecidos como aprendizagem profunda e são algoritmos como redes neurais convolucionais (CNNs), redes neurais recorrentes (RNRs) e redes de memória de curto prazo (LTSMs). Esses modelos podem ser usados para extrair recursos de imagens, realizar a compressão e reconstrução dos dados e classificar expressões faciais. Em resumo, apresenta que considerando as sete emoções básicas, a felicidade obtem a maior precisão de classificação e a emoção medo, a menor precisão. E por fim, destaca a importância de combinar diferentes modelos de aprendizado de máquina para melhorar ainda mais a precisão e desempenho dos algoritmos [CANAL, 2022].

Considerando todos os estudos anteriormente citados, a emoção surpresa teve a maior precisão de predição com acurácia variando de 86 a 96,2%. A emoção de felicidade obteve a segunda melhor precisão de predição, com acurácias variando de 85 a 100%. As emoções com maiores confusões de classificação, no estado-da-arte, foram medo, nojo e tristeza. As emoções ordenadas de forma decrescente de acordo com a precisão da classificação nos estudos citados são surpresa, felicidade, raiva, triste, neutro, medo e nojo.

O presente trabalho apresenta resultados em consenso com a literatura e que confirmam a veracidade das classificações por meio de dados unidimensionais, ou seja, distâncias euclidianas dos pontos chave da face. Uma vez que as emoções com as melhores classificações foram felicidade, surpresa e neutra, e as emoções que apresentaram maiores dificuldades de classificação foram triste, medo e nojo.

Como sugerido como avanços futuros pelo estudo de Canal et al. (2022), o presente trabalho propõe uma abordagem neurogenética, ou seja, um modelo híbrido para extração e classificação de expressões. Assim, as informações filtradas pelo algoritmo genético foram utilizadas como dados de entrada para os classificadores SVM e MLP. Em adição, as acurácias apresentadas entre as sete emoções básicas estão no mesmo patamar de acurácia média dos estudos a partir de 2020.

É importante ressaltar que para diferentes banco de dados e classificadores utilizados, MLP ou SVM, as características selecionadas de entrada pelo algoritmo genético se alteram. Porém, como apresentado em Ramis et al. (2022), para um mesmo modelo desenvolvido encontra-se flutuações na acurácia média do classificador. Expondo, assim, uma variabilidade intrínseca ao problema proposto de acordo com as expressões realizadas de diferentes formas e ângulos de captação da imagem pela câmera. Portanto, os resultados encontrados são comparados com estudos que utilizaram os mesmos bancos de dados.

Segundo a Tabela 9 é possível verificar que os resultados apresentados no presente trabalho estão acima de 90% conforme a literatura para o banco de dados CK+. Para os bancos de dados WSEFEP e JAFFE estão acima de 80%, não superando a literatura porém alinhados com o estado-da-arte diante de arquiteturas que empregam redes convolucionais. Visto que as melhores acurácias do modelo proposta são para o estado neutro, emoção feliz e para a emoção surpresa.

Tabela 9 - Comparação estado-da-arte e o presente trabalho

<b>Autores</b>	<b>Ano</b>	<b>Banco de Dados</b>	<b>Acc Média</b>
Najah et al.	2017	JAFFE	77,46%
Li et al.	2020	CK+	98,68%
		JAFFE	98,52%
Yao et al.	2021	CK+	93,7%
Minaee et al.	2021	CK+	98%
		JAFFE	92,8%
Ramis et al.	2022	CK+	93,23%
		JAFFE	70,09%
		WSEFEP	91,91%



<b>Este trabalho</b>	<b>2023</b>	<b>CK+</b>	<b>93,8%</b>
		<b>JAFFE</b>	<b>85,11%</b>
		<b>WSEFEP</b>	<b>86,79%</b>

### 5.3 DISCUSSÃO GERAL

Bliss et al. (2015) afirma que nos últimos anos, o número de técnicas de monitoramento para pacientes em UTI e CTI tem aumentado exponencialmente. Porém, como análise isolada dos sinais vitais não expõe o real estado fisiológico do paciente, o estudo defende que é necessário a implementação de um “sistema de sistemas” no qual a análise global agrupa sinais aparentemente díspares e pode ajudar a revelar condições físicas ocultas [BLISS, 2015]. A monitoração de temperatura corporal e expressões faciais são dados pouco abordados no cenário de pacientes não-comunicantes. Além disso, são fontes de extração de informações do paciente que podem auxiliar na intervenção médica.

A revisão bibliográfica realizada pelo estudo de Molinaro et al. (2020) aborda a literatura mediante a análise dos principais sinais vitais de forma remota, sendo que a temperatura corporal não foi referida visto que não há uma técnica de análise padrão-ouro no estado-da-arte. A elevação da temperatura corporal em pacientes críticos é um tema muito relevante pois a febre pode afetar até 75% dos pacientes da UTI, causando agitação e delírio. Também é verificado que mais de 50% dos pacientes internados apresentam relatos de dor moderada a severa durante a intubação em estado consciente e em repouso, podendo gerar aumento da carga de trabalho cardíaco [KIEKKAS, 2010].

Ademais, pacientes não-comunicativos estão incapazes de expressar condições físicas e psicológicas como dor, desconforto, ansiedade e medo. No estudo de Rahu et al. (2010) mais de 50% dos pacientes internados relataram sensação de dor moderada a severa durante a intubação em estado consciente e em repouso, e mais de 80% relataram dor em ações nociceptivas como alteração de posição no leito.

A Tabela 10 apresenta, por fim, a comparação do melhor desempenho encontrado no estado-da-arte com o presente trabalho. Assim, considerado a monitoração de temperatura corporal não há estudos encontrados na literatura em domínio RGB durante o desenvolvimento deste trabalho, enquanto o presente trabalho atingiu a pontuação  $R^2$  de 0,938 para o GF com a predição da temperatura média das bochechas. Com relação à classificação de expressões faciais, o estudo de Li et al. (2020), com a análise do banco de dados CK+ atingiu acurácia do

modelo com 98,68%, enquanto o nosso modelo atingiu 93,8% considerando as distâncias entre os pontos 24-33, 37-41, 44-46, 48-54, 51-57, 21-22, 08-33, 33-48, 33-54, 23-43, 20-21, 31-33 e 33-35, ao todo 13 distâncias entre os pontos faciais. Ou seja, com apenas 13 distâncias euclidianas de pontos-chave da face foi possível apresentar uma acurácia média de classificação de expressões faciais comparável, com perda em 5% de eficiência, ao melhor estudo encontrado na área.

Tabela 10 - Comparação estado-da-arte e o presente trabalho. Desempenho da predição de temperatura corporal quantificada pela pontuação  $R^2$  e classificação de expressões faciais pela acurácia em (%).

<b>Dado adquirido</b>	<b>Autores</b>	<b>Melhor Desempenho</b>
Temperatura corporal	-	-
	Este trabalho	0,938
Expressão Facial	Li et al.	98,68%
	Este trabalho	93,8%

Em síntese, o estudo apresentou resultados promissores utilizando câmeras RGB comuns, com a utilização de uma baixa quantidade de dados de entrada para produzir resultados satisfatórios e eficientes em comparação com a literatura. Assim, o presente estudo contribui com o estado-da-arte em relação a novos caminhos de investigação na área de predição de temperatura corporal e expressões faciais por imagens em domínio visível.

## 6 CONCLUSÃO

### 6.1 PREDIÇÃO DE TEMPERATURA CORPORAL

Conforme apresentado, o presente trabalho aborda um método promissor de análise de temperatura corporal sem contato em domínio visível. Aplicou-se a técnica de RCP e com a seleção de 25 componentes principais a partir dos sinais da glabella, nariz, bochecha direita e esquerda e a diferença entre as regiões da glabella em comparação com o nariz, considerando os três canais RGB, a regressão linear foi aplicada e obteve resultado eficiente. Os resultados do presente estudo são animadores visto que as variáveis de predição temperatura média das bochechas e predição da temperatura média subclaviculares apresentaram pontuação  $R^2$  de 0.84 e 0.79 para o GM e pontuação  $R^2$  de 0.94 e 0.91 para GF, respectivamente. Os erros médios quadráticos se concentraram em 0.15-0.20 °C para ambas análises e ambos gêneros. A umidade ambiente é dispensável para análise, porém a temperatura do ambiente é imprescindível. A técnica desenvolvida é inovadora pois não há estudos semelhantes na literatura e por ser aplicada a partir da coleta de dados por uma câmera RGB comum.

### 6.2 CLASSIFICAÇÃO DE EXPRESSÕES FACIAIS

Os resultados demonstram que modelo proposto é capaz de classificar emoções humanas a partir de dados unidimensionais, resultando em acurácias médias de 93,80% para o banco de dados CK+, 85,11% para o JAFFE e 86,79% para a base de dados WSEFEP. As emoções surpresa, feliz e neutro apresentaram os melhores desempenhos entre as bases de dados analisadas. Os parâmetros de entrada por distância euclidianas de pontos faciais independe de iluminação, tons de pele e gênero, brilho ou contraste da imagem aquisitionada. As limitações deste estudo se concentram na estrutura do rosto que pode variar conforme a faixa etária do voluntário e a posição do indivíduo que deve estar frontal à câmera. As acurácias apresentadas como modelo se aproximam bastante do presente estado-da-arte e com a vantagem da diminuição do volume de dados de entrada, possibilitando caminhos de aplicações em tempo real com eficiência satisfatória a partir de uma câmera RGB. A sensação de dor, ansiedade, delírio e desconforto por pacientes incommunicáveis em unidades de terapia intensiva pode ser

detectada pelo treinamento do algoritmo de aprendizagem de máquina proposto a partir de pontos da face que classificam a sensação segundo o SCAF.

### **6.3 CONCLUSÃO GERAL**

Com o avanço da visão computacional diversas técnicas de extração de informações e manipulação de dados fisiológicos são apresentados. Atualmente há escassez de estudos com foco no cenário de pacientes críticos em regime de terapia intensiva hospitalar com aplicações em visão computacional. Sabe-se que os principais sinais vitais majoritariamente utilizados para monitoramento destes pacientes são o eletrocardiograma, a frequência cardíaca, a pressão arterial, a saturação de oxigênio no sangue e a temperatura corporal. Entretanto, a monitoração da temperatura corporal em domínio RGB ainda é pouco desenvolvido na literatura. Bem como a análise de expressões faciais em sistemas online com arquiteturas simplificadas, que não exijam um alto poder computacional e que não seja necessário uma elevada quantidade de dados pra treinamento do algoritmo. Sabendo que a análise eficiente de temperatura corporal e expressões faciais podem auxiliar intervenções médicas, o presente estudo apresentou um sistema integrado de dados adquiridos por câmera RGB de forma automática e com eficiência de resposta comparável à literatura. O presente estudo atingiu todos os objetivos apresentados e alcançou a eficiência de pontuação  $R^2$  de 0.938 para o modelo RCP com o GF mediante a predição de temperatura média das bochechas com 31 variáveis dependentes, sem a necessidade da umidade ambiente e acurácia de 93.8% para a base de dados CK+ considerando 13 distâncias entre pontos-chave da face. Portanto, o sistema desenvolvido permite aplicações em tempo-real por ser baseado em dados de entrada com baixa dimensão a partir de equipamento de aquisição de dados comum sendo uma câmera RGB. As possíveis aplicações deste sistemas são pacientes acamados, áreas clínicas de monitoração de pacientes não-comunicativos, pacientes neonatais e infantis e telemedicina.

## **7 TRABALHOS FUTUROS**

Uma das principais limitações deste trabalho é que o mesmo foi desenvolvido a partir de dados em ambiente controlado e foram utilizadas bases de dados sedimentadas na literatura. Como trabalhos futuros, pretende-se realizar a aplicação do framework no cenário médico-hospitalar. Para isso, será necessário um novo registro ao Comitê de Ética, bem como o

desenvolvimento de uma nova metodologia para aquisição de dados em nível de hardware e software. Assim, a partir da aplicação do sistema integrado em cenários não-controlado será possível realizar refinamentos no modelo a partir das respostas apresentadas pelo algoritmo.

## 8 PRODUÇÃO CIENTÍFICA

Ao longo do curso de mestrado, os seguintes trabalhos foram publicados:

- **FERREIRA-JORGE**, A. R.; **SILVA**, C. R. ; **SOARES**, A. B. . Impact of Architecture and Genetic Algorithm Application in Neural Pattern Recognition 2021 (Pôster - XIII Simpósio de Engenharia Biomédica UFU).
- **FERREIRA-JORGE**, A. R.; **CUNHA**, M. J. ; **SOARES**, A. B. . Input Parameters Selection for Emotion Recognition Using Genetic Algorithm and Machine Learning. 2022. (Apresentação de Trabalho/Simpósio). DOI: 10.5281/zenodo.7493085
- **FERREIRA-JORGE**, A. R.; **CUNHA**, M. J. ; **SOARES**, A. B. . Genetic Algorithms in Machine Learning Applied to Computer Vision: Facial Emotion Recognition. XXVIII Congresso Brasileiro de Engenharia Biomédica (CBEB), 2022.
- **SILVA**, C. R. ; **FERREIRA-JORGE**, A. R. ; **SOARES**, A. B. . Optimizing Hand Gesture Pattern Recognition with Differential Evolution Algorithm 2021 (Pôster - XIII Simpósio de Engenharia Biomédica UFU).
- **GOMES**, L. G. D. ; **JORGE**, A. R. F. ; **SOARES**, A. B. ; **ALMEIDA**, M. B. . 3D Printed Assistive Technology among Multiple Control Sites. In: XIV Simpósio de Engenharia Biomédica, 2022, Uberlândia. Anais do XIV Simpósio de Engenharia Biomédica, 2022. DOI: 10.5281/zenodo.7493325
- **Bessa**, P. ; **JORGE**, A. R. F. ; **SOARES**, A. B. Human Facial Emotion Classification: A Method Validation. In: XIV Simpósio de Engenharia Biomédica, 2022, Uberlândia. Anais do XIV Simpósio de Engenharia Biomédica, 2022. DOI: 10.5281/zenodo.7492331

Em processo de revisão por periódico científico:

- **JORGE**, A. R. F.; **DA CUNHA** M. J., ; **SOARES**, A. B. Contactless Body Temperature Prediction in RGB Domain. Journal of Health Informatics, 2023.

## 9 REFERÊNCIAS

ANSARI, S. *Building Computer Vision Applications Using Artificial Neural Networks*. Apress Berkeley, California, USA. Ed. 1, 2020. DOI: <https://doi.org/10.1007/978-1-4842-5887-3>

ARIF-RAHU, M.; GRAP, M. J. Facial expression and pain in the critically ill non-communicative patient: State of science review. **Intensive and Critical Care Nursing**, v. 26, n. 6, p. 343–352, 2010. DOI: 10.1016/j.iccn.2010.08.007

ARYAL A., GERBER, B. B. A comparative study of predicting individual thermal sensation and satisfaction using wrist-worn temperature sensor, thermal camera and ambient temperature sensor. **Building and Environment**, Elsevier, v. 160, 2019. DOI: 10.1016/j.buildenv.2019.106223

ASARIAN, L.; GLOY, V.; GEARY, N. Homeostasis. **Encyclopedia of Human Behavior: Second Edition**, p. 324–333, 2012. DOI: 10.1016/B978-0-12-375000-6.00191-9

AWAD, A. I.; HASSABALLAH, M. *Image Feature Detectors and Descriptors*, Springer, vol. 630, p. 438, 2016. DOI: 10.1007/978-3-319-28854-3

BABU, S. A.; STACH, J. E. M.; GOODFELLOW, M. Computer-assisted numerical analysis of colour-group data for dereplication of streptomycetes for bioprospecting and ecological purposes. **International Journal of General and Molecular Microbiology**, v. 97, n. 3, p. 231–239, 2010. DOI: 10.1007/s10482-009-9404-x

BALTRUSAITIS, T.; ROBINSON, P.; MORENCY, L. P. OpenFace: An open source facial behavior analysis toolkit. **IEEE Winter Conference on Applications of Computer Vision, WACV 2016**, 2016. DOI: 10.1109/WACV.2016.7477553

BISI, O. *Visible and Invisible: The Wonders of Light Phenomena*. Springer, ed. 1, Switzerland, 2011. DOI: <https://doi.org/10.1007/978-3-319-09825-8>

BLISS, D. Innovations and controversies in the monitoring of pediatric patients in the ICU. **Seminars in Pediatric Surgery**, v. 24, n. 1, p. 32–36, 2015. DOI: 10.1053/j.sempedsurg.2014.11.007

C.A. JAMES, A. J. R. Reliability and validity of skin temperature measurement by telemetry thermistors and a thermal camera during exercise in the heat. **Journal of Thermal Biology**, 2014. DOI: 10.1016/j.jtherbio.2014.08.010

CANAL, F. Z. et al. A survey on facial emotion recognition techniques: A state-of-the-art literature review. **Information Sciences**, v. 582, p. 593–617, 2022. DOI: 10.1016/j.ins.2021.10.005

CARDONE, D. et al. Automated warping procedure for facial thermal imaging based on features identification in the visible domain. **Infrared Physics and Technology**, v. 112, p. 103595, 2021. DOI: 10.1016/j.infrared.2020.103595

CHEN, H. et al. Predicting the thermophysical properties of skin tumor based on the surface temperature and deep learning. **International Journal of Heat and Mass Transfer**, v. 180, p. 121804, 2021. DOI: 10.1016/j.ijheatmasstransfer.2021.121804

CHENG, C.-H.; WONG, K.-L. Deep Learning Methods for Remote Heart Rate measurement: a review and future research agenda. **Sensors**, v. 21, p. 6296, 2021. DOI: <https://doi.org/10.3390/s21186296>

CHEVROLET, J. C.; JOLLIET, P. Clinical review: Agitation and delirium in the critically ill - Significance and management. **Critical Care**, v. 11, n. 3, p. 1–5, 2007. DOI: 10.1186/cc5787

COHEN, I. et al. Learning Bayesian network classifiers for facial expression recognition using both labeled and unlabeled data. **Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition**, v. 1, 2003. DOI: 10.1109/cvpr.2003.1211408

DENG, J. et al. RetinaFace: Single-stage Dense Face Localisation in the Wild. 2019. DOI: <https://doi.org/10.48550/arXiv.1905.00641>

DIVEKAR, A. V.; GHARPURE, D. C. Low-compute facial expression recognition using fiducial feature-sets. **IOP Conference Series: Materials Science and Engineering**, v. 1187, n. 1, p. 012025, 2021. DOI: 10.1088/1757-899x/1187/1/012025

EKMAN, P. An Argument for Basic Emotions. *Cognition and Emotion*, v. 6, n. 3–4, p. 169–200, 1992. DOI: 10.1080/02699939208411068

HARFORD, M. et al. Availability and performance of image/video-based vital signs monitoring methods: A systematic review protocol. **Systematic Reviews**, v. 6, n. 1, p. 1–7, 2017. DOI: 10.1186/s13643-017-0615-3

HASAN, Z.; RAMAMURTHY, S. R.; ROY, N. CamSense: A camera-based contact-less heart activity monitoring. **Smart Health**, v. 23, p. 100240, 2022. DOI: 10.1016/j.smhl.2021.100240

HAYKIN, S. *Redes Neurais - Princípios e Prática*. Bookman, São Paulo, Brasil, ed. 2, 2001. ISBN: 0-13-273350-1

HORSTMAN, J. *The Scientific American Brave New Brain*, Jossey-Bass, 1. ed. San Francisco, CA, 2010. ISBN: 978-85-7307-718-6

JAMES, C. A., RICHARDSON, A.J. Reliability and validity of skin temperature measurement by telemetry thermistors and a thermal camera during exercise in the heat. **Journal of Thermal Biology**, v. 45, p. 141-149, 2014. DOI: <https://doi.org/10.1016/j.jtherbio.2014.08.010>

JIA, M. et al. Development of facial-skin temperature driven thermal comfort and sensation modeling for a futuristic application. **Building and Environment**, v. 207, n. PB, p. 108479, 2022. DOI: 10.1016/j.buildenv.2021.108479

KIEKKAS, P. et al. Fever and agitation in elderly ICU patients: A descriptive study. **Intensive and Critical Care Nursing**, v. 26, n. 3, p. 169–174, 2010. DOI: 10.1016/j.iccn.2010.02.003

- KIEKKAS, P. et al. Fever and standard monitoring parameters of ICU patients: A descriptive study. **Intensive and Critical Care Nursing**, v. 23, n. 5, p. 281–288, 2007. DOI: 10.1016/j.iccn.2007.04.001
- KING, D. Dlib python toolkit. <https://pypi.org/project/dlib/> . Acessado em: 23 de abril de 2023.
- KUBAT, M. An Introduction to Machine Learning. 2. ed. Coral Glabes, FL, USA: Springer, 2017. DOI: 10.1007/978-3-319-63913-0
- KUDRAYVTSEV, G. Fundamentals of Computer Vision. 2020. ISBN: 9388511859
- KUMAR, A. et al. A Multimodal Emotion Recognition using affective biomarkers for smart prediction of emotional health for people analytics in smart industries. **Image and Vision Computing**, v. 123, p. 104483, 2022. DOI: 10.1016/j.imavis.2022.104483
- LAKSHMANAN, V.; GÖRNER, M.; GILLARD, R. Practical Machine Learning for Computer Vision. 1. ed. Sebastopol, CA: O’Reilly, 2021. ISBN: 1098102363
- LARKIN, P. J. Infrared and Raman Spectroscopy: Principles and Spectral Interpretation. Elsevier, ed. 1, Waltham, MA, USA: Elsevier, 2011. ISBN: 978-0-12-386984-5
- LI, J. et al. Attention mechanism-based CNN for facial expression recognition. **Neurocomputing**, v. 411, p. 340–350, 2020. DOI: 10.1016/j.neucom.2020.06.014
- LIU, H. et al. FEDA: Fine-grained emotion difference analysis for facial expression recognition. **Biomedical Signal Processing and Control**, v. 79, n. P2, p. 104209, 2023. DOI: 10.1016/j.bspc.2022.104209
- LOPES, A. T. et al. Facial expression recognition with Convolutional Neural Networks: Coping with few data and the training sample order. **Pattern Recognition**, v. 61, p. 610–628, 2017. DOI: 10.1016/j.patcog.2016.07.026
- LORENA, A. C. et al. Uma Introdução às Support Vector Machines. Revista de Informática Teórica e Aplicada, v. 14, n. 2, p. 43–67, 2007. DOI: <https://doi.org/10.22456/2175-2745.5690>
- MA, J. et al. Infrared and visible image fusion methods and applications: A survey. **Information Fusion**, vol. 45, pp. 153–178, 2019. DOI: <https://doi.org/10.1016/j.inffus.2018.02.004>
- METZMACHER, H. et al. Real-time human skin temperature analysis using thermal image recognition for thermal comfort assessment. **Energy and Buildings**, v. 158, p. 1063–1078, 2018. DOI: 10.1016/j.enbuild.2017.09.032
- MINAAE, S. Deep-Emotion: Facial Expression Recognition Using Attentional Convolutional Network, **Sensors**, v. 21, p. 3046, 2021. DOI: <https://doi.org/10.3390/s21093046>
- MITCHELL, M. An Introduction To Genetic Algorithms, Bradford Book, p. 221, 1998. DOI: 10.1007/978-3-319-63913-0



MOLINARO, N. et al. Contactless Vital Signs Monitoring From Videos Recorded With Digital Cameras: An Overview. **Frontiers in Physiology**, v. 13, 2022. DOI: 10.3389/fphys.2022.801709

NAJAH, G. Emotion Estimation From Facial Images. Atilim University, 2017. DOI: 10.13140/RG.2.2.25113.62565

NIXON M., AGUADO A. Feature Extraction and Image Processing for Computer Vision. 4. ed. San Diego, CA, USA: Elsevier, 2020. DOI: <https://doi.org/10.1016/C2011-0-06935-1>

PROTSIV, M. et al. Decreasing human body temperature in the United States since the industrial revolution. **eLife**, v. 9, p. 1–11, 2020. DOI: 10.7554/eLife.49555

PRZYBYLO, J. A deep learning approach for remote heart rate estimation. **Biomedical Signal Processing and Control**, v. 74, p. 103457, 2022. DOI: <https://doi.org/10.1016/j.bspc.2021.103457>

PU, T. et al. AU-Expression Knowledge Constrained Representation Learning for Facial Expression Recognition. **Proceedings - IEEE International Conference on Robotics and Automation**, v.2, p. 11154–11161, 2021. DOI: 10.1109/ICRA48506.2021.9561252

ROMANOVSKY, A. A. Skin temperature: Its role in thermoregulation. **Acta Physiologica**, v. 210, n. 3, p. 498–507, 2014. DOI: 10.1111/apha.12231

SAJJAD, M. et al. Human Behavior Understanding in Big Multimedia Data Using CNN based Facial Expression Recognition. **Mobile Networks and Applications**, v. 25, n. 4, p. 1611–1621, 2020. DOI: 10.1007/s11036-019-01366-9

SELWYN, N. et al. Autoroll: Scripting the emergence of classroom facial recognition technology. **Learning, Media and Technology**, v. 48, p. 1–14, 2022. DOI: <https://doi.org/10.1080/17439884.2022.2039938>

SLIMANI, K. et al. Facial emotion recognition: A comparative analysis using 22 LBP variants. **ACM International Conference Proceeding Series**, v. 2018- March, p. 88–94, 2018. DOI: 10.1145/3177148.3180092

SOORI, M. et al. Machine learning and artificial intelligence in CNC machine tools, A review. **Sustainable Manufacturing and Service Economics**, 2023. DOI: 10.1016/j.smse.2023.100009

Taniş C. M. Operational monitoring of snow cover using digital imagery, Aalto University, 2019. DOI: <https://doi.org/10.3390/data3030023>

THUSEETHAN, S. et al. EmoSeC: Emotion recognition from scene context. **Neurocomputing**, v. 492, p. 174–187, 2022. DOI: 10.1016/j.neucom.2022.04.019

WU, Y. Recognition and prediction of individual thermal comfort requirement based on local skin temperature. **Journal of Building Engineering**, v. 49, p. 104025, 2022. DOI: 10.1016/j.jobbe.2022.104025

XUAN P. et al. Thermal infrared imaging based facial temperature in comparison to ear temperature during a real-driving scenario. **Journal of Thermal Biology**, v. 96, p. 102806, 2021. DOI: 10.1016/j.jtherbio.2020.102806

YAO, L. et al. Action unit classification for facial expression recognition using active learning and SVM. **Multimedia Tools and Applications**, v. 80, n. 16, p. 24287–24301, 2021. DOI: 10.1007/s11042-021-10836-w

ZARINS, U. Anatomy of Facial Expression. EU: Exonicius Inc, ed. 1, p. 43, 2017. ISBN: 0990341119

ZHONG, Y. et al. Service transformation under industry 4.0: Investigating acceptance of facial recognition payment through an extended technology acceptance model. **Technology in Society**, v. 64, p. 101515, 2021. DOI: 10.1016/j.techsoc.2020.101515