



UNIVERSIDADE FEDERAL DO PARÁ
FACOMP – FACULDADE DE COMPUTAÇÃO
CURSO DE BACHARELADO EM SISTEMAS DE INFORMAÇÃO

**UMA REVISÃO SISTEMÁTICA SOBRE ESTUDOS DE DETECÇÃO DE
EXPRESSÕES FACIAIS A PARTIR DO EMPREGO DE MODELOS DE
INTELIGÊNCIA ARTIFICIAL**

ALANA MIRANDA MEDEIROS

Castanhal-PA
2022



UNIVERSIDADE FEDERAL DO PARÁ
FACOMP – FACULDADE DE COMPUTAÇÃO
CURSO DE BACHARELADO EM SISTEMAS DE INFORMAÇÃO

ALANA MIRANDA MEDEIROS

**UMA REVISÃO SISTEMÁTICA SOBRE ESTUDOS DE DETECÇÃO DE
EXPRESSÕES FACIAIS A PARTIR DO EMPREGO DE MODELOS DE
INTELIGÊNCIA ARTIFICIAL**

Trabalho de Conclusão de Curso
submetido ao colegiado da Faculdade de
Computação da Universidade Federal do
Pará, como requisito parcial para a
obtenção do grau de bacharel em
Sistemas de Informação.

Orientador: Profa. Dra. Yomara Pinheiro
Pires

Castanhal-PA
2022

UNIVERSIDADE FEDERAL DO PARÁ
FACOMP – FACULDADE DE COMPUTAÇÃO
CURSO DE BACHARELADO EM SISTEMAS DE INFORMAÇÃO

**UMA REVISÃO SISTEMÁTICA SOBRE ESTUDOS DE DETECÇÃO DE
EXPRESSÕES FACIAIS A PARTIR DO EMPREGO DE MODELOS DE
INTELIGÊNCIA ARTIFICIAL**

Trabalho de Conclusão de Curso apresentado ao Colegiado da Faculdade de Computação (FACOMP) da Universidade Federal do Pará do campus de Castanhal, como requisito parcial para a obtenção do Grau de Bacharel em Sistemas de Informação.

Profa. Dra. Yomara Pinheiro Pires
Orientadora-UFPA/FACOMP

Prof. Dr. Igor Ruiz Gomes
Membro da Banca – UFPA/FACOMP

Prof. Dr. Marcos César da Rocha Seruffo
Membro da Banca – UFPA/FCT

Prof. Dr. João Claudio Chamma Carvalho
Diretor (a) da Faculdade de Computação - FACOMP

Castanhal-PA
2022

AGRADECIMENTOS

Primeiramente, gostaria de agradecer à Universidade Federal do Pará (UFPA) e aos coordenadores da Faculdade de Computação (FACOMP).

À minha orientadora Dra. Yomara Pires pela confiança depositada em mim para o desenvolvimento deste Trabalho, não medindo esforços e sendo sempre muito paciente nos momentos de discussão sobre o tema.

Aos bolsistas do Projeto GIPSTIL pela imensa ajuda com a Revisão Sistemática da Literatura.

À família que faço parte: minha mãe Ana Laura e meu padrasto Gilzibene, pelo investimento na minha educação e por todo o amparo e apoio em minhas escolhas, sendo sempre meu alicerce nos momentos de dificuldade ao longo dessa caminhada.

A todos que contribuíram, direta ou indiretamente, para a conclusão dessa etapa.

*“O coração instável é a única constante
nesse mundo”
(Hayao Miyazaki)*

SUMÁRIO

INTRODUÇÃO	14
Contextualização	14
Justificativa	15
Objetivos	16
Objetivo geral	16
Objetivos Específicos	16
Estrutura do trabalho	16
ASPECTOS TEÓRICOS	17
Introdução	17
Violência Contra Mulheres	17
Revisão Sistemática da Literatura	19
Questão da Pesquisa e estratégia PICOC	19
Estratégia de Busca	20
Estratégia de Seleção	21
Inteligência Artificial: Computação Afetiva	22
Reconhecimento de Emoção Facial (FER)	23
Redes Neurais Convolucionais	25
Contextualização: Redes Neurais Artificiais	25
Rede Neural Convolucional	26
Arquiteturas de CNN	28
Conclusão	29
METODOLOGIA DA PESQUISA	30
Introdução	30
Revisão Sistemática no Parsifal	30
Planejamento	31
Condução	34
Relato da Revisão	35
Conclusão	35
RESULTADOS E DISCUSSÕES	36

Introdução	36
Resultados Obtidos	36
Discussão sobre os resultados obtidos	38
CNN	38
DATABASE	40
Sumarização e discussão dos artigos selecionados	41
Conclusão	43
CONSIDERAÇÕES FINAIS	44
Dificuldades Encontradas	44
Trabalhos Futuros	44
REFERÊNCIAS	46

LISTA DE ILUSTRAÇÕES

Figura 2.1- Microexpressão gerada pela combinação das AUs 9 e 17	25
Figura 2.2 - Ilustração do modelo de uma Rede Neural de Aprendizado Profundo	26
Figura 2.3 - Representação de uma imagem colorida separadas pelos seus três planos de cores	27
Figura 3.1- Metodologia geral da Revisão	30
Figura 4.1- Resultados da busca de artigos	36
Figura 4.2- Quantidade de artigos selecionados	37
Figura 4.3- Distribuição dos artigos por fontes	37
Figura 4.4- Bases de dados mais utilizadas	40

LISTA DE TABELAS

Tabela 2.1- Exemplo de String de busca	21
Tabela 2.2- Bases de dados comumente utilizadas em Revisões da Literatura	22
Tabela 2.3- Questionário proposto para a avaliação de qualidade de artigos em uma SRL	23
Tabela 3.1- Questões de Pesquisa utilizada	31
Tabela 3.2- Termos-chave utilizados	32
Tabela 4.1- Técnicas e performance dos 13 artigos obtidos na avaliação qualitativa	41

LISTA DE QUADROS

Quadro 3.1- Termos chave utilizados	32
-------------------------------------	----

RESUMO

A compreensão de emoções é uma das grandes capacidades adquiridas pelos seres humanos, pois permite o entendimento de expressões faciais que facilitam a captura de informações importantes sobre outros indivíduos, as quais são utilizadas para a percepção de estados mentais ou emocionais. Os avanços na área de Inteligência Artificial (IA), mais especificamente em Aprendizado Profundo com o advento das Redes Neurais Artificiais (RNA), possibilitaram às máquinas a capacidade de inferir emoções humanas através de análise de imagens. Dentro deste contexto, esse estudo propõe uma Revisão Sistemática da Literatura com o intuito de explorar as diversas técnicas e algoritmos baseados em Inteligência Artificial para o reconhecimento de emoções pela detecção de expressões faciais. A iniciativa faz parte do projeto WHOT (Women's Health Observer Tool) 2.0, que propõe uma otimização no módulo de inferência de emoções da ferramenta, cuja finalidade é identificar, por meio da combinação de reconhecimento de expressões faciais e entrevista com questionário digital, padrões psicocomportamentais denotadores de violência em mulheres. Os resultados sugerem que usos de base de dados com imagens controladas em laboratório, combinadas com Redes Neurais Convolucionais (CNN's) como VVG e RESnet, têm excelentes desempenhos em seus testes.

PALAVRAS-CHAVES: Reconhecimento de Emoções, Inteligência Artificial, Visão Computacional, Violência contra a mulher, Computação Afetiva

ABSTRACT

Understanding emotions is one of the great abilities acquired by human beings, as it allows the understanding of facial expressions that facilitate the capture of important information about other individuals, which are used for the perception of mental or emotional states. Advances in Artificial Intelligence (AI), more specifically in Deep Learning with the advent of Artificial Neural Networks (ANN), have enabled machines to infer human emotions through image analysis. Within this context, this study proposes a Systematic Review of the Literature in order to explore the various techniques and algorithms based on Artificial Intelligence for the recognition of emotions by detecting facial expressions. The initiative is part of the WHOT (Women's Health Observer Tool) 2.0 project, which proposes an optimization of the tool's emotions inference module, whose purpose is to identify, through a combination of facial expression recognition and interview with a digital questionnaire, psychobehavioral patterns denoters of violence against women. The results suggest that database uses with controlled images in the laboratory, combined with Convolutional Neural Networks (CNN's) such as VVG and RESnet, have excellent performances in their tests.

KEYWORDS: Emotion Recognition, Artificial Intelligence, Computer Vision, Violence against women, Affective Computing

LISTA DE SIGLAS

IA- Inteligência Artificial

RNA- Redes Neurais Artificiais

CNN- Rede Neural Convolutacional

WHOT - Women's Health Observer Tool

ML- Machine Learning

SLR- Revisão Sistemática da Literatura (SLR, do inglês Systematic Literature Review)

PICOC- Population, Intervention, Comparison, Outcome, Context

FER - Reconhecimento de Emoção Facial (FER, do inglês *Facial Emotion Recognition*)

RQ - Questões de Pesquisa (RQ, do inglês Research Question)

CA - Computação Afetiva

FACS - Facial Action Coding System

GIPSTIL - Grupo Interdisciplinar de Pesquisa sobre Sociedade, Tecnologia, Inovação e Linguagens

PPGEAA - Programa de Pós-Graduação em Estudos Antrópicos na Amazônia

FACOMP - Faculdade de Computação de Castanhal

ONU - Organização das Nações Unidas

MMFDH - Ministério da Mulher, da Família e dos Direitos Humanos

OMS - Organização Mundial da Saúde

IST - Infecção Sexualmente Transmissíveis

OPAS - Organização Pan-Americana de Saúde

FC - Fully Connected

LFFW - Light Field Faces in the Wild

LFCC - Light Field Faces Constrained

1 INTRODUÇÃO

1.1 Contextualização

Os impactos dos avanços na área da tecnologia são observados em inúmeras esferas da sociedade, tendo sua contribuição principalmente na resolução de problemas e na minimização das suas consequências. Neste contexto, destaca-se a abrangente área da IA, cujo objetivo é estudar técnicas capazes de reproduzir artificialmente a estrutura e o funcionamento do cérebro humano (RAUBER, 2005) com o intuito de empregá-las na automatização de tarefas sejam elas simples ou mais complexas.

A Visão Computacional é o ramo da IA responsável pelo estudo de métodos computacionais que simulam a visão humana, através do processamento e análise de imagens (MARENGONI, 2009). Tais métodos permitem que o computador possa identificar e classificar objetos e rostos, por exemplo, e com isso tomar decisões. Estudos baseados no reconhecimento de expressões faciais a partir da detecção de rostos em imagens podem contribuir em diferentes campos de atuação profissional, tais como Pedagogia (HAQUE, 2018), (SIDHU, 2022) e Psicologia (ZHOU, 2018).

O presente trabalho tem sua procedência em 2015, com os primeiros estudos sobre Computação Afetiva no projeto de extensão intitulado “Computing Emotions: Estudo dos Aspectos Afetivos em Ambientes Educacionais Computacionais”, na Faculdade de Computação (Facomp) do Campus Castanhal da Universidade Federal do Pará (UFPA), que posteriormente serviram como base para a criação da ferramenta WHOT, em 2019, que auxilia na construção de perfis psicocomportamentais de mulheres vítimas de violência a partir de três recursos: i) reconhecimento de expressões faciais para inferir emoções; ii) fornecimento de questionários digitais específicos sobre situações de violência; iii) geração de relatórios individuais com cruzamentos de análises estatísticas entre os dados obtidos em cada entrevista (COSTA, 2021).

Dito isso, este estudo propõe identificar as técnicas e algoritmos de Inteligência Artificial, tais como Aprendizado Profundo, mais utilizados na detecção de expressões faciais, para que esses possam ser aplicados para superar limitações da ferramenta WHOT, que atualmente é capaz de identificar apenas quatro emoções básicas: alegria, tristeza, raiva e surpresa.

Para tal objetivo, realizou-se uma extensa busca e triagem de trabalhos científicos através de uma Revisão Sistemática da Literatura. Para a importação de

artigos e gerenciamento das demais etapas que envolvem uma Revisão, foi utilizado o Parsifal (KEELE, 2007), uma ferramenta online desenvolvida para pesquisadores conduzirem, de forma colaborativa, revisões da literatura. Na plataforma foi possível planejar e definir os parâmetros de pesquisa, strings de buscas, bases de dados, critérios de inclusão e exclusão de seleção dos artigos, assim como a sua avaliação de qualidade.

1.2 Justificativa

Este trabalho é justificado em função da necessidade de continuidade dos estudos desenvolvidos no projeto PIBIC 2021 - *WHOT 2.0: Uma ferramenta de Computação Afetiva Aplicada à Situação de Violência Contra Mulher*, desenvolvido pelo Grupo Interdisciplinar de Pesquisa sobre Sociedade, Tecnologia, Inovação e Linguagens (GIPSTIL) do Programa de Pós-Graduação em Estudos Antrópicos na Amazônia (PPGEAA) e pela Faculdade de Computação de Castanhal (FACOMP) sobre inferência de emoções, especificamente nas áreas da Computação Afetiva (CA), Inteligência Artificial e Aprendizado de Máquina aplicados em estudos de caso voltados a um problema social em escala mundial, à Situação de Violência Contra Mulher, que resultou em uma ferramenta com a habilidade de realizar reconhecimento de emoções por meio de expressões faciais, de fornecer questionário digital a mulheres sobre situações de violência e também de gerar relatórios em que se agregam as respostas dos questionários com quatro emoções ao longo da entrevista (alegria, tristeza, raiva e surpresa). A ferramenta desenvolvida foi intitulada de *Women's Health Observer Tool* (WHOT).

Os estudos e a ferramenta, até o momento desenvolvida, podem ser adaptados às mais diversas situações enfrentadas na sociedade do momento presente. Desta forma, faz-se necessário a adequação da ferramenta já desenvolvida; bem como a definição de uma metodologia para reconhecimento de emoções, por meio de expressões faciais, através de Aprendizado de Máquina.

Desta feita, a SLR aqui apresentada proporcionará o incremento do *know-how* em computação afetiva, especificamente em estudos sobre reconhecimento facial e inferência de emoções, colaborando para o estado da arte nessa temática e oportunizando a maior inserção do Campus, da FACOMP e do PPGEAA nessa área em que tantos avanços tem sido verificados.

1.3 Objetivos

1.3.1 Objetivo geral

Este trabalho tem por objetivo realizar uma Revisão Sistemática da Literatura para mapear, sumarizar e eleger técnicas e métodos de IA no processamento de imagens para o reconhecimento de expressões faciais. Pretende-se, através desse processo, identificar os modelos mais utilizados que possam ser aplicados em futuras implementações na ferramenta WHOT, no que diz respeito ao seu módulo de inferência de emoções, para que forneça diagnósticos mais precisos, auxiliando na tomada de decisões na prevenção e proteção de mulheres vítimas de violência.

1.3.2 Objetivos Específicos

- Realizar uma revisão sistemática da literatura a fim de mapear e sumarizar estudos que empregam técnicas, métodos e algoritmos de reconhecimento de emoções baseados em IA, como ML e Aprendizado Profundo;
- Através dos resultados da Revisão, identificar as Bases de Dados mais utilizadas no treinamento dos modelos encontrados.

1.4 Estrutura do trabalho

Este trabalho está organizado em cinco capítulos, sendo que o restante deste trabalho está organizado como segue:

Capítulo 2: Apresenta os aspectos teóricos dos principais temas presentes na execução da pesquisa, tais como: violência contra a mulher, Revisão Sistemática da Literatura, Inteligência Artificial e computação afetiva, reconhecimento de emoções faciais e redes neurais convolucionais.

Capítulo 3: A metodologia, incluindo o relato da Revisão Sistemática e demais etapas realizadas, estão incluídos neste capítulo.

Capítulo 4: Apresenta a discussão dos resultados.

Capítulo 5: O capítulo cinco apresenta as considerações finais para este trabalho, destacando a sua contribuição para as pesquisas da área.

2 ASPECTOS TEÓRICOS

Neste capítulo são abordados os aspectos teóricos dos principais temas discutidos no trabalho, incluindo a problemática do contexto do projeto: a violência contra mulher.

2.1 Violência Contra Mulheres

De acordo com (DE ALMEIDA, 2017) o conceito de violência de gênero deve ser entendido como “uma relação de poder de dominação do homem e de submissão da mulher”, em que os papéis atribuídos aos dois gêneros ao longo da história, induzem relações violentas entre os sexos. A Organização das Nações Unidas (ONU) classifica como violência contra a mulher qualquer ato de violência baseado em gênero, que resulte em dano físico, sexual ou mental às mulheres, podendo afetar negativamente a saúde física, mental, sexual e reprodutiva das mulheres, além de aumentar a vulnerabilidade à Infecção Sexualmente Transmissíveis (IST) e, em situações extremas, levar ao homicídio ou suicídio.

De acordo com a Organização Pan-americana de Saúde (OPAS/OMS)¹, as principais causas associadas à violência contra mulheres são a baixa escolaridade, maltrato infantil ou exposição à violência na família, uso nocivo do álcool, atitudes violentas e desigualdade de gênero.

Uma análise conduzida pela OMS junto à *London School of Hygiene and Tropical Medicine* e ao *Medical Research Council*, baseada em dados de 80 países, descobriu que, em todo o mundo, quase um terço (30%) de todas as mulheres que estiveram em um relacionamento sofreram violência física e/ou sexual ou por parte de seu parceiro. As estimativas de prevalência variam de 23,2% nos países de alta renda e 24,6% na região do Pacífico Ocidental para 37% na região do Mediterrâneo Oriental da OMS e 37,7% na região do Sudeste Asiático. Além disso, 38% de todos os assassinatos de mulheres são cometidos por parceiros em todo mundo. Além da violência perpetrada por parceiros, 7% das mulheres em todo o mundo relatam terem sido assediadas sexualmente por terceiros, embora os dados para essa questão sejam mais limitados. (OPAS/OMS disponível em <https://www.paho.org/pt/topics/violence-against-women>)

¹ <https://www.paho.org/pt/topics/violence-against-women>

Segundo dados do relatório de Violência contra Mulheres de 2021 (FBSP, 2022) houve um aumento de 4% nos casos de estupros entre 2020 e 2021, isto significa que a cada 10 minutos uma menina ou mulher é vítima deste tipo de violência sexual; a cada um dia três mulheres são vítimas de femicídio (FBSP, 2019); um travesti ou mulher trans é assassinada a cada dois dias (ANTRA, 2021). Dados do (FBSP, 2020b) indicam que são registrados 266.310 casos de lesão corporal dolosa contra mulheres em decorrência de violência doméstica e familiar, o que corresponde a média de 729 agressões cometidas diariamente em 2019.

A Organização Mundial da Saúde (OMS) alerta que o cenário de violência contra mulheres é ainda mais alarmante durante crises humanitárias, evidenciando a pandemia causada pelo Sars-CoV-2 e o consequente isolamento social, que foi definido como medida de prevenção. (SOUSA, 2021) aponta como fator contribuinte a condição de distanciamento social, em que as vítimas passaram a ter um maior tempo de convivência estando no mesmo ambiente que os seus agressores, e com isso se restringiram em realizar denúncias.

O Fórum Brasileiro de Segurança Pública (FBSP), com o objetivo de atualizar os dados sobre violência doméstica durante a pandemia de Covid-19, publicou uma nota técnica que apontou a redução, em vários estados brasileiros, de uma série de crimes contra as mulheres com exceção da violência letal, caracterizada como feminicídio, o que é um indicativo de que as mulheres estão encontrando mais dificuldades em denunciar a(s) violência(s) sofridas no período de pandemia. Os dados referentes aos registros de lesão corporal dolosa mostraram uma redução de 27,2% no período cumulativo entre março e maio de 2020, em comparação com o mesmo período no ano anterior. Em contraste, houve um aumento de 2,2% na taxa de feminicídio para os mesmos períodos citados anteriormente (FBSP, 2020a)

Diante do exposto, ressalta-se a importância de pesquisas e implementações de soluções computacionais que visem auxiliar no enfrentamento do quadro da violência de gênero, seja por meio de divulgação, diagnóstico ou prevenção. A exemplo, destacam-se as seguintes soluções: “Estamos Juntas” (SILVEIRA, 2018), que caracteriza um sistema especialista em uma aplicação web que apoia mulheres na identificação de relacionamentos abusivos, além de alertar sobre diferentes tipos de violência, informando os procedimentos para a notificação. (SOUSA, 2020) descreve a elaboração e validação do “Alertômetro”, uma solução de tecnologia educacional para o

público feminino para atuar na divulgação do conhecimento acerca das manifestações de violência, permitindo à mulher analisar os atos cometidos pelos seus parceiros e a partir disso refletir sobre possíveis cenários de violência em que esteve ou pode ser inserida.

2.2 Revisão Sistemática da Literatura

A Revisão Sistemática da Literatura (SLR, Systematic Literature Review) é uma metodologia de busca utilizada para identificar, avaliar e interpretar estudos que são relevantes para uma determinada questão de pesquisa, tópico ou fenômeno de interesse (KEELE, 2007). A própria natureza dessa modalidade indica que os seus resultados tendem a apresentar maior qualidade, ser mais abrangentes e menos tendenciosos quando comparados com outros tipos de revisões da literatura (SIDDAWAY, 2019).

Diferente de uma revisão da literatura de conveniência, geralmente utilizada apenas para reunir e discutir trabalhos científicos para o tratamento de uma determinada temática, sem apresentar critérios explícitos sobre o seu planejamento ou execução, uma Revisão Sistemática da Literatura exige a adoção de protocolos específicos para a elaboração de suas etapas, além de ter o caráter de reprodutibilidade para outros pesquisadores, uma vez que todo o processo de pesquisa é documentado (GALVÃO, 2019).

2.2.1 Questão da Pesquisa e estratégia PICOC

Entre as principais etapas que devem compor uma SLR, destaca-se a importância da definição de uma ou mais questões de pesquisa, que irão conduzir as buscas para o estudo. A fim de que a questão de pesquisa esteja em concordância com os objetivos da Revisão, (PETTICREW, 2008) sugere “quebrá-la” em sub questões, estruturadas pela estratégia PICOC (*Population, Intervention, Comparison, Outcome, Context*), descrita a seguir, de acordo com o ponto de vista da Engenharia de software (WOHLIN, 2012):

a) Population: a população na qual a evidência é coletada. Qual grupo de pessoas, áreas de aplicações ou negócios são de interesse para a revisão?

b) Intervention: é a intervenção aplicada no estudo, baseada em uma tecnologia, ferramenta ou procedimento;

c) Comparison: é a metodologia, tecnologia, ferramenta ou procedimento com a qual a intervenção está sendo comparada;

d) Outcome: deve estar relacionado a fatores de importância para profissionais, tais como medidas de aperfeiçoamento a partir da intervenção utilizada.

e) **Context:** deve ser definido como uma visão ampliada da população.

2.2.2 Estratégia de Busca

Quando se conduz uma Revisão da Literatura, o objetivo é encontrar todos os estudos publicados e não publicados que correspondam às suas questões de pesquisa, através de strings de buscas (SIDDAWAY, 2019). A formulação da string, juntamente com a escolha das bases de dados em que os artigos serão procurados, compõem a estratégia de busca da Revisão que deve ser seguida.

As strings de busca são compostas por palavras-chave (*keywords*), que resumem o que o estudo busca abordar. *Keywords* podem ser extraídas da estratégia PICOC e relacionadas a uma lista de sinônimos, abreviações e alternativas ortográficas. Outras terminologias podem ser encontradas considerando o conteúdo dos títulos usados em revistas e jornais (KEELE, 2007).

As strings de busca podem ser estruturadas com o uso dos operadores lógicos OR, para integrar uma palavra-chave e seus sinônimos, e AND, para integrar as diferentes palavras-chave. Observe o exemplo na Tabela 2.1:

Tabela 2.1- Exemplo de *String* de busca

(“palavra-chave1” OR “sinônimo” OR “sinônimo” OR “sinônimo”) AND
 (“palavra-chave2” OR “sinônimo” OR “sinônimo”) AND (“palavra-chave3” OR
 “sinônimo”)

Fonte: Autora.

Outro passo importante a ser seguido é a escolha das bases de dados, que devem ser compatíveis com a temática abordada na Revisão. Para tal, devem-se selecionar bases que são relevantes na área. Galvão (2019) destaca algumas bases que têm se consolidado no que tange ao desenvolvimento de Revisões da Literatura, expostas na Tabela 2.2:

Tabela 2.2- Bases de dados comumente utilizadas em Revisões da Literatura

Fonte de Dados	Abrangência
	Várias áreas do conhecimento, incluindo: análise

SCOPUS	bibliométrica, história, educação, psicologia, direito, religião, linguística e literatura
SciELO	Artigos produzidos em vários países da América Latina
LISTA	Literatura Internacional nas áreas de Ciência e Tecnologia da Informação
ERIC	Artigos e relatórios em inglês na área da educação
ACM Digital Library	Trabalhos da área de Tecnologia da Informação
PubMed	Literatura Internacional de todas as especialidades médicas
IEEE Explore	Trabalhos na área de tecnologia publicados pelo Institute of Electrical and Electronic Engineers

Fonte: (Adaptado de GALVÃO, 2019)

Tendo em vista que as bases de dados bibliográficas internacionais priorizam a língua inglesa como idioma de busca, é importante ressaltar a necessidade de traduzir corretamente os termos-chave para o idioma (GALVÃO, 2019).

2.2.3 Estratégia de Seleção

Outra etapa importante a ser realizada antes da leitura dos artigos é definir os critérios de seleção e inclusão, que fazem parte da primeira seleção de estudos baseada nos seus resumos ou títulos. Critérios exclusivos podem determinar a exclusão de artigos duplicados ou cuja data de publicação do trabalho não esteja em um intervalo de tempo específico, por exemplo. De modo geral, esses critérios devem refletir o que é desejado para que um trabalho seja eleito na revisão.

Kitchenham (KEELE, 2007) ressalta ainda a importância do planejamento de uma segunda etapa de seleção dos artigos para fornecer critérios de inclusão e exclusão ainda mais detalhados para a elegibilidade dos artigos. Para a segunda seleção, que compreende uma avaliação da qualidade a partir da leitura do artigo como um todo, (WEN, 2012) sugere uma metodologia baseada em uma checkbox, com um questionário para avaliar o rigor e a relevância dos estudos, atribuindo um índice de qualidade a eles. As perguntas são apresentadas na Tabela 2.3 e para cada pergunta há três possíveis respostas: Sim, Parcialmente e Não. As três respostas possuem os respectivos pesos: 1, 0.5 e 0.

Tabela 2.3- Questionário proposto para a avaliação de qualidade de artigos em uma SRL

ID	Questão
QA1	Os objetivos da pesquisa estão claramente definidos?
QA2	O contexto de estimativa está descrito adequadamente?
QA3	Os métodos de estimativa são bem definidos e deliberados?
QA4	O projeto do experimento é apropriado e justificável?
QA5	O experimento é aplicado em conjuntos de dados projetados suficientes?
QA6	A precisão da estimativa é medida e relatada?
QA7	O método de estimativa proposto é comparado com outros métodos?
QA8	Os resultados do estudo são claramente especificados e apoiados por relatórios de resultados?
QA9	As limitações dos estudos são analisadas explicitamente?
QA10	O estudo agrega valor à comunidade acadêmica ou industrial?

Fonte: (WEN, 2012).

2.3 Inteligência Artificial: Computação Afetiva

A formação e o desenvolvimento de diversas esferas da sociedade como conhecemos atualmente são frutos da inteligência humana, com seu processo evolutivo e adaptativo de aprendizagem. (AZEVEDO, 2001) define a aprendizagem como “Uma das atividades mais ricas desempenhadas pelo ser humano, proporcionando-lhe a assimilação de novos conhecimentos.”

A capacidade de comunicação dos humanos se dá de forma escrita, verbal e não verbal (DE SOUZA et al, 2012). Esta última, também conhecida como linguagem corporal, é composta por um conjunto de gestos, expressões, atitudes e imagens que resultam numa combinação de expressões corporais e faciais capazes de serem percebidas e interpretadas, gerando os mais diversos estímulos em um processo de interação e convívio social.

Portanto, a comunicação não verbal está fortemente ligada às nossas emoções, nos emocionamos por qualquer coisa, a qualquer momento, em qualquer lugar, em qualquer situação que nos encontremos. Cada emoção gera sensações únicas em nosso

corpo, a exemplo a face, onde podemos expressar emoções como alegria, tristeza, medo, raiva, surpresa, desprezo e nojo. Estas expressões faciais são universais e demonstram genuinamente o que sentimos (EKMAN, 2011).

Os estudos que desempenham o desenvolvimento de técnicas computacionais para reproduzir artificialmente o processo de aprendizagem humana são contemplados pelos campos da Inteligência Artificial. Dentre eles, destaca-se o campo da Computação Afetiva (CA), cujo desafio é fornecer ao computador a habilidade de detectar emoções humanas (PICARD, 2000). (BERCHT, 2006) faz uma distinção entre dois grandes grupos de sistemas relacionados à CA: a) o primeiro grupo considera e estuda as emoções do usuário, a partir da sua interação com o ambiente computacional; e b) que considera a emoção da máquina (síntese de emoções), através da simulação de estados emocionais do próprio sistema para a construção e aperfeiçoamento de robôs que pareçam mais reais, por exemplo.

O contexto computacional em que a pesquisa se insere diz respeito ao primeiro grupo, uma vez que visa melhorias no módulo de inferência de emoções da ferramenta WHOT, que tem o foco nas emoções do usuário (mulheres entrevistadas), realizando reconhecimento de expressões faciais a partir da sua interação com os questionários digitais disponibilizado pela ferramenta.

2.4 Reconhecimento de Emoção Facial (FER)

Para compreender melhor o funcionamento de um Reconhecimento de Emoção Facial (FER, do inglês *Facial Emotion Recognition*) feito por uma máquina computacional, faz-se necessário entender alguns conceitos básicos sobre emoção e microexpressões faciais, bem como os estudos que possibilitaram o surgimento e aperfeiçoamento desse tipo de tarefa de Inteligência Artificial.

Uma microexpressão facial é um evento momentâneo na face em que determinados músculos se contraem formando uma expressão a partir de uma emoção sentida (SANTOS, 2022). Entende-se por isso, que as emoções são expressas através dessas microexpressões. O entendimento e uso das expressões faciais pelos humanos desempenham papel fundamental nas suas interações sociais, uma vez que estas podem ser conduzidas por essa percepção.

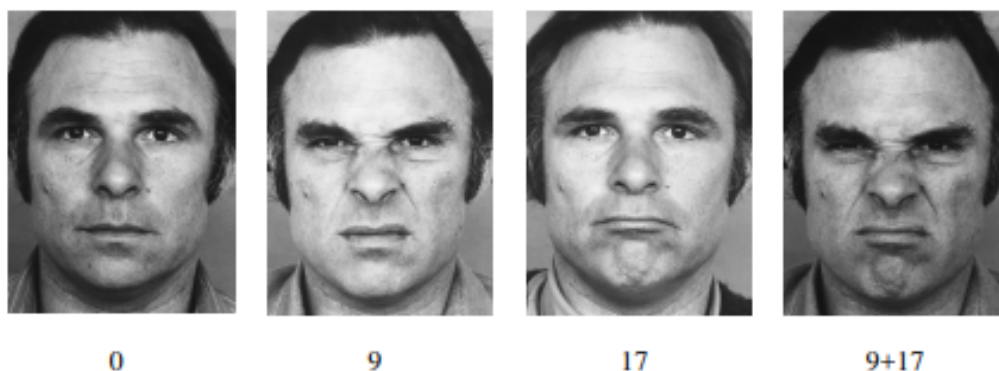
A base teórica dos estudos sobre microexpressões faciais teve inúmeras contribuições com as pesquisas desenvolvidas pelo Psicólogo Paul Ekman. Destacam-se, principalmente, os seus estudos sobre a universalidade das emoções

(EKMAN, 1971), (EKMAN, 1971a) e (EKMAN, 1979), que validaram a teoria de que as emoções são universais, ou seja, independente da nacionalidade dos indivíduos e conseqüentemente influências culturais, as suas contrações faciais para um determinado sentimento ou contexto serão as mesmas. Em seus estudos, Ekman ainda classifica as emoções como sete emoções primárias: felicidade, tristeza, medo, nojo, raiva, surpresa e desprezo. Com a compreensão do comportamento facial e da universalidade das emoções, Ekman e Friesen (EKMAN, 1978) desenvolveram o Facial Action Coding System (FACS), um sistema para categorizar as microexpressões faciais humanas.

As Unidades de Ação (AU, do inglês *Action Units*) são essenciais para o entendimento do funcionamento do FACS, pois elas fazem referência às contrações faciais. Cada AU representa uma contração facial, por exemplo: a AU 9 - franzir o nariz e a Au 17 - Elevar o queixo.

No manual do FACS, citado anteriormente, Ekman descreve e ilustra com imagens todas as AUs do sistema, bem como as suas combinações, que resultam em microexpressões faciais. A imagem seguir ilustra, com foto do próprio autor, a microexpressão gerada pela combinação das AUs 9 e 17:

Figura 2.1- Microexpressão gerada pela combinação das AUs 9 e 17



Fonte: (Adaptada de EKMAN, 1978)

A primeira foto representa uma expressão neutra, usada para comparações com as demais; a segunda foto é a ilustração da AU 9; a terceira é a AU 17; e por fim, a quarta representa a microexpressão gerada pela combinação de ambas.

Os estudos sobre o FACS foram expandidos, resultando na criação de um segundo modelo, chamado EM-FACS (FRIESEN, 1983), que categoriza as Unidades de

Ações mais presentes em cada tipo de expressão emocional. Os estudos das emoções e microexpressões faciais, como os citados nesta seção, são amplamente utilizados em pesquisas sobre o reconhecimento computacional de emoções (DE OLIVEIRA, 2008), da área de Computação Afetiva, como pode ser visto em (GIUNTINI, 2019) e (JAQUES, 2007).

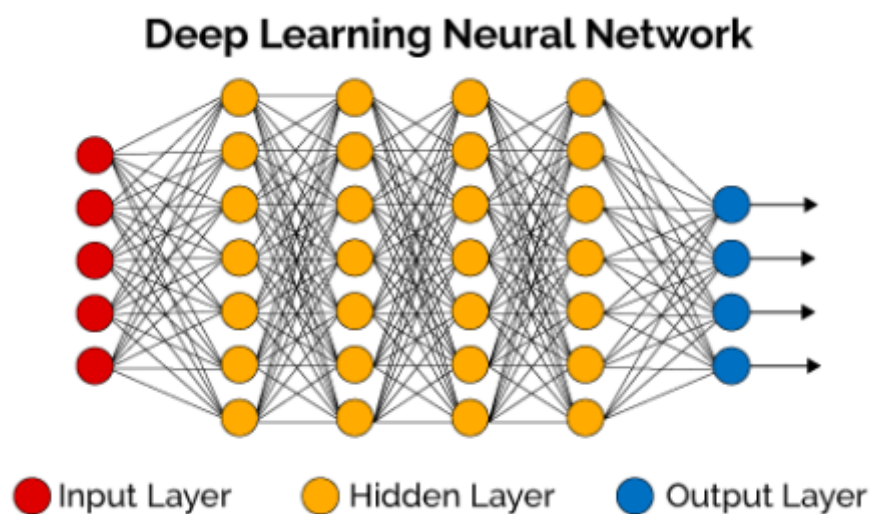
2.5 Redes Neurais Convolucionais

As Redes Neurais Convolucionais são conhecidas pelo bom desempenho em tarefas de detecção de objetos na área de Visão Computacional.

2.5.1 Contextualização: Redes Neurais Artificiais

Entre os principais avanços da área de Inteligência Artificial, destacam-se os estudos de Redes Neurais Artificiais, que compreendem o uso de algoritmos cuja arquitetura é baseada no modelo do neurônio biológico (RAUBER, 2005). As redes neurais são compostas por neurônios matemáticos responsáveis por processar os dados de entrada que são passados através de cada camada da rede, com a saída da camada anterior fornecendo entrada para a próxima camada. A primeira camada de uma rede é chamada de camada de entrada, a última é a camada de saída e entre elas há as chamadas camadas ocultas (*Data Science Academy*, 2022), como representado na imagem:

Figura 2.2: Ilustração do modelo de uma Rede Neural de Aprendizado Profundo



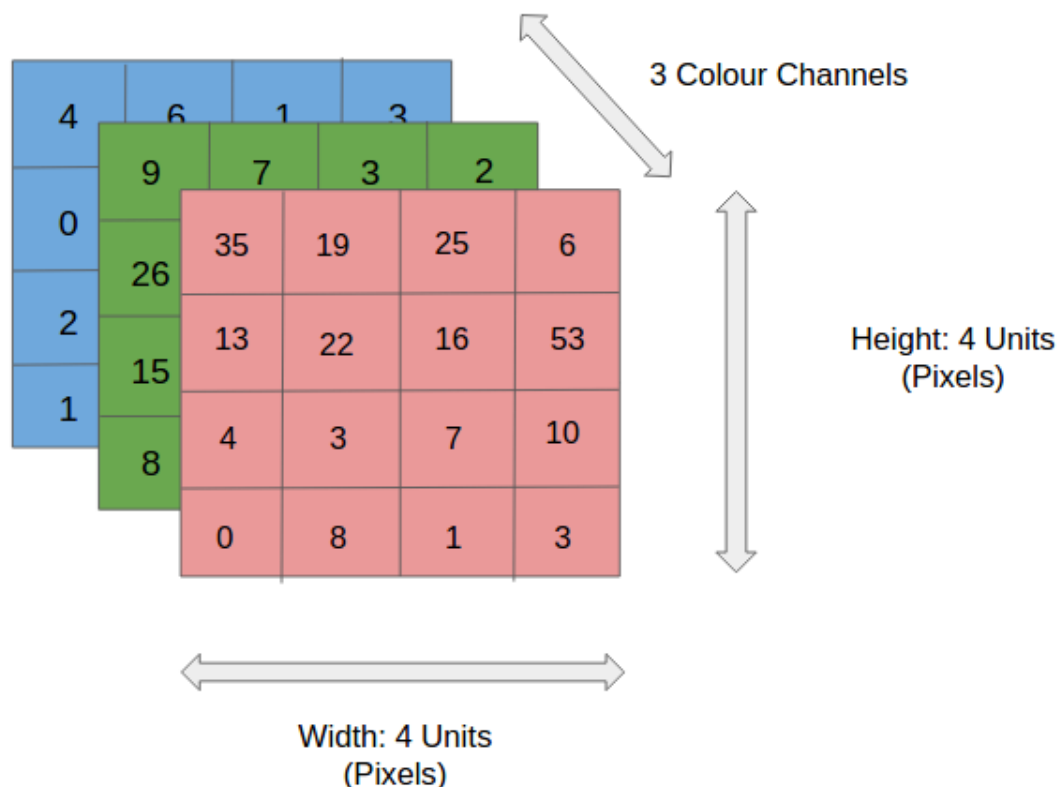
Fonte: (Adaptado de *Data Science Academy*)

2.5.2 Rede Neural Convolucional

A Rede Neural Convolucional (CNN, do inglês Convolutional Neural Network) é um tipo de arquitetura de Redes Neurais Artificiais baseada no mecanismo de percepção visual animal, com o conceito de neurônios localmente sensíveis e seletivos para a projeção de imagens, estudados por (HUBEL, 1962). Nos modelos neurais artificiais para processamento de imagens, com campos receptivos locais, os neurônios seletivos podem extrair características visuais, tais como bordas e outras extremidades (LECUN, 1998), importantes para a assimilação de um determinado objeto ou pessoa. Em outras palavras, a arquitetura de um algoritmo de CNN permite que o seu processamento de imagem atribua importância a características que são essenciais para a sua classificação final.

A imagem é processada pelas Redes Neurais Artificiais como uma matriz de três dimensões (também chamado de Tensor): altura x largura (dimensões da imagem) x profundidade, correspondente aos canais de cores que usualmente são valores RGB (Red, Green e Blue, para imagens coloridas) ou valores em escala de cinza, para imagens em preto e branco (WU, 2017). A figura a seguir ilustra como uma imagem colorida é processada em uma rede neural:

Figura 2.3- Representação de uma imagem colorida separadas pelos seus três planos de cores



Fonte: (Data Science Academy, 2022)

A imagem em questão possui dimensões de altura e largura 4x4, considerada relativamente pequena. Sendo assim, o seu processamento em uma RNA seria feito de tal maneira: $4 \times 4 = 16 \times 3 = 48$ (que correspondem aos neurônios na camada de entrada). Imagens com pequenas resoluções não exigem um alto grau de desempenho computacional, porém, dependendo da aplicação ou do formato de captura das imagens, estas podem ter dimensões 32x32, 64x64 ou ainda maiores, como as imagens de resolução 8K (7680×4320), resultando em um maior tempo de processamento, se considerar ainda a quantidade de imagens a serem classificadas.

O uso de CNN reduz as dimensões das imagens para diminuir o seu tempo de processamento, sem perder recursos importantes para obter uma boa previsão (Data Science Academy, 2022). O modelo matemático tradicional de uma CNN é composto por três camadas, em que as duas primeiras correspondem ao processo de extração de *features* (características) da imagem e a terceira, que é uma camada do tipo Fully Connected (FC) que mapeia os recursos extraídos como dados de saída na tarefa de

classificação (YAMASHITA, 2018). As propriedades de cada camada são abordadas a seguir:

a) Camada de Convolução

É a principal camada de processamento em uma CNN, que realiza a extração de recursos aplicando *kernels*, uma pequena matriz de números, nas entradas de uma imagem, que também é uma matriz (YAMASHITA, 2018), que geram um mapa de características (*feature map*). Basicamente, a convolução consiste em operações entre matrizes para extrair recursos de imagem e a partir disso, gerar uma nova imagem.

b) Camada de Pooling

São camadas de agrupamento que recebem as saídas dos *feature maps*, gerados na camada de convolução, e preparam um mapa de características condensadas, simplificando ainda mais os dados de saída (Data Science Academy, 2022). Ao reduzir os *features maps*, as camadas de pooling reduzem também o número de parâmetros na saída e, conseqüentemente, acelera o tempo de processamento.

c) Camada totalmente conectada

A saída dos *feature maps* gerados na etapa de extração de características (camadas de convolução e pooling), é transformada em uma matriz unidimensional de números (um vetor) e conectada a uma ou mais camadas do tipo Fully Connected, nas quais cada entrada é conectada a cada saída por um peso aprendível (YAMASHITA, 2018).

2.5.3 Arquiteturas de CNN

- a) VGG (Visual Geometry Group)** é uma arquitetura de CNN profunda proposta por Karen Simonyan e Andrew Zisserman em (SIMONYAN, 2014). A profundidade está relacionada com o número de camadas. Por padrão, as entradas são em RGB medindo 224x224 pixels. Além disso, os autores usam dois tipos de camadas convolucionais: uma camada Conv3 com filtros (kernel) 3x3 e uma camada Conv1, com filtros 1x1.

- b) ResNet** é a arquitetura de CNN proposta por (HE, 2015), utilizando-se dos conceitos de Blocos Residuais para melhorar a precisão de modelos de CNN profundos, que podem apresentar problemas de “Descida de gradiente”, como os relatados em (HE, 2015) e (SRIVASTAVA, 2015), que levam a erros de treinamento.

Neste capítulo foram apresentados os principais aspectos teóricos utilizados neste trabalho, tais como, o conceito, contextualização e alguns dados nacionais e internacionais sobre a temática violência contra mulher. Também foram descritos os protocolos adotados para a execução da SLR. Os aspectos computacionais utilizados na pesquisa foram apresentados através dos conceitos de FER, IA e Computação Afetiva. No capítulo a seguir serão descritos os procedimentos metodológicos adotados no trabalho ora apresentado.

3 METODOLOGIA DA PESQUISA

Neste capítulo serão descritos os aspectos metodológicos utilizados no trabalho. Especificamente, as etapas SLR utilizadas na plataforma Parsifal, que consistem em Planejamento, Condução e Relato da Revisão.

3.1 Revisão Sistemática no Parsifal

A importação de artigos e gerenciamento das demais etapas que envolvem uma Revisão Sistemática da Literatura foram executadas na plataforma gratuita Parsifal², desenvolvida para realizar revisões da literatura de forma colaborativa, que possibilitou uma boa execução de todas as etapas descritas na seção 2.3 deste trabalho, incluindo o planejamento da revisão e o gerenciamento dos artigos importados.

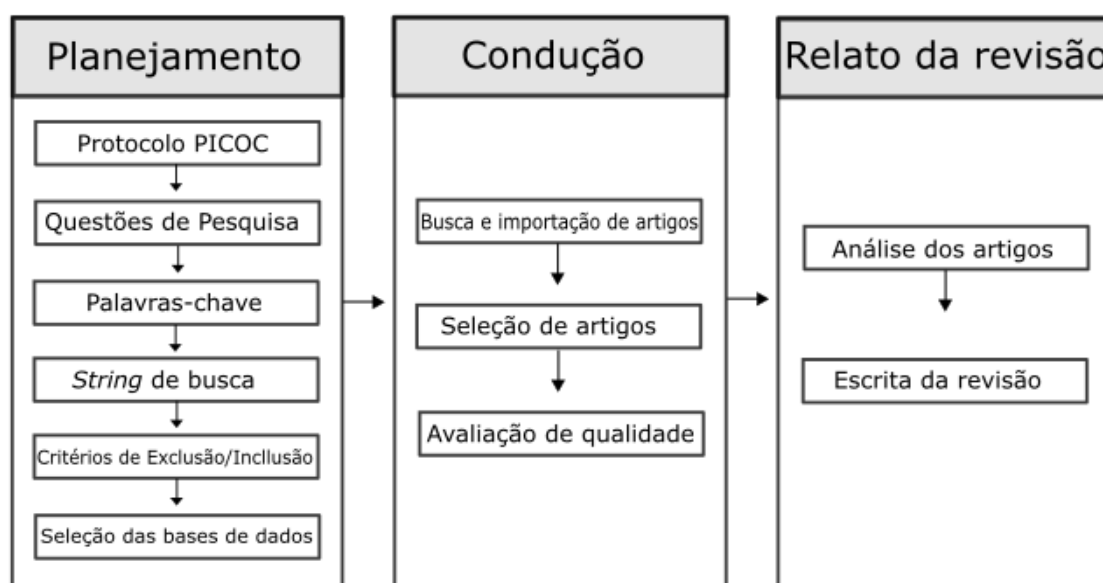
O uso da ferramenta auxiliou na identificação de trabalhos pertinentes ao tema central da pesquisa, mediante a definição dos objetivos específicos, estratégia PICOC, questões de pesquisa, palavras-chave e sinônimos e, conseqüentemente, uma *string* de

² Parsifal é uma plataforma poderosa que oferece suporte total para pesquisadores geograficamente distribuídos para trabalhar na mesma Revisão Sistemática de Literatura. Não há limite de autores por revisão. Disponível em: <https://parsif.al>

busca, além dos critérios de inclusão e exclusão e critérios para a avaliação de qualidade.

A Revisão foi realizada seguindo a divisão de etapas definidas pela ferramenta Parsifal, que por sua vez foi desenvolvida a partir dos estudos de Kitchenham (KEELE, 2007), que propõe que uma Revisão Sistemática deva envolver as seguintes fases: Planejamento, Condução e Relato da Revisão. Essa metodologia está representada na Figura 3.1

Figura 3.1- Metodologia geral da Revisão



Fonte: Autora

3.1.1 Planejamento

No planejamento descreve-se toda a estratégia para a condução da Revisão, incluindo os principais parâmetros de busca e seleção que estruturam o estudo. Para este trabalho, o objetivo da SLR foi identificar estudos que abordassem técnicas ou algoritmos utilizados em reconhecimento de emoções baseados em Inteligência Artificial.

A metodologia PICOC, como descrita na seção 2.3.1, apresenta cinco elementos fundamentais: *population*, *intervention*, *comparison*, *outcomes* e *context*, que são utilizados para descrever todos os componentes relacionados ao objetivo da revisão e

auxiliar na composição das questões de pesquisa. Sendo assim, a estratégia PICOC foi definida da seguinte forma:

- *Population*: reconhecimento de emoção, detecção facial;
- *Intervention*: técnicas de inteligência artificial;
- *Outcomes*: acurácia;
- *Context*: computação afetiva.

Como não houve intenção de comparar técnicas, o elemento *Comparison* não foi definido.

Com base na estratégia PICOC e em concordância com o objetivo da revisão, as questões de pesquisa (RQ, do inglês Research Question), norteadoras para a busca, foram assim definidas na Tabela 3.1.

Tabela 3.1- Questões de Pesquisa utilizadas

ID	Questão de Pesquisa	Razão
RQ1	Quais as técnicas de Inteligência Artificial mais utilizadas em reconhecimento de emoções?	Identificar as técnicas de Inteligência Artificial mais utilizadas nos estudos selecionados
RQ2	Quais as bases de dados mais utilizadas nas pesquisas de reconhecimento facial?	Identificar as bases de dados mais utilizadas nos estudos selecionados

Fonte: Autora.

Em seguida, definiram-se as palavras-chave e os seus respectivos sinônimos, necessários para a composição da *string* de busca. A Tabela 3.2 apresenta as palavras-chave e sinônimos usados.

Tabela 3.2- Termos-chave utilizados.

Palavra-chave	Sinônimos
técnicas de inteligência artificial	aprendizado de máquina, aprendizado profundo, inteligência artificial, redes neurais artificiais
artificial intelligence technique	artificial intelligence, artificial neural networks, deep learning, machine learning
affective computing	emotion recognition, face detection

computação afetiva	reconhecimento de emoção, reconhecimento facial
reconhecimento de imagem	image recognition

Fonte: Autora.

Como a busca foi feita em periódicos nacionais e internacionais, os termos-chave e sinônimos deveriam estar nos idiomas português e inglês.

É importante mencionar que a *string* de busca original, teve algumas alterações e adaptações para retornar corretamente as publicações nas bases de dados. A configuração final é vista no quadro abaixo:

Quadro 3.1 - Termos-chave utilizados.

("técnicas de inteligência artificial" OR "aprendizado de máquina" OR "aprendizado profundo" OR "inteligência artificial" OR "redes neurais artificiais" OR "artificial intelligence techniques" OR "artificial intelligence" OR "artificial neural networks" OR "deep learning" OR "machine learning") AND ("affective computing" OR "emotion recognition" OR "computação afetiva" OR "reconhecimento de emoção") AND ("reconhecimento de imagem" OR "image recognition")

Fonte: Autora.

Os operadores lógicos OR foram empregados para integrar keywords e sinônimos, e AND e os parênteses para integrar logicamente as diferentes keywords. As aspas duplas são utilizadas para palavras compostas.

Em relação a escolha das bases de dados, optou-se pela pesquisa de trabalhos publicados em periódicos nacionais e internacionais das áreas de computação, visto que são os principais instrumentos de veiculação de trabalhos desta natureza. A partir de uma pesquisa inicial na literatura (GALVÃO, 2019), elegeram-se as principais fontes de pesquisa: ACM Digital Library³, IEEE Digital Library⁴, SciELO Citation Index⁵, Scopus⁶ e para garantir uma busca mais abrangente, o Google Scholar⁷.

³ <https://dl.acm.org/>

⁴ <https://ieeexplore.ieee.org/Xplore/home.jsp>

⁵ <https://blog.scielo.org/>

⁶ <https://www.scopus.com/search/form.uri?display=basicbasic>

⁷ <https://scholar.google.com.br/>

Ainda na etapa de planejamento, foi necessário definir os critérios de inclusão e exclusão para a primeira seleção, que é baseada apenas nos resumos de cada trabalho. Era desejável filtrar apenas artigos que abordassem as técnicas, ou os desafios em aplicá-las, no contexto especificado. Além disso, estudos duplicados, incompletos e que não estivessem escritos nos idiomas Português ou Inglês, deveriam ser excluídos. Para delimitar o tempo de pesquisa, optou-se por incluir apenas trabalhos publicados entre 2018 e 2021 (ano anterior à execução da Revisão) e ainda fez-se necessário especificar que o reconhecimento de emoções deveria ser através de detecção facial, tendo em vista que há técnicas de inferência de emoções através de análise de voz, de texto e outras.

Sendo assim, os critérios foram definidos da seguinte forma:

Critérios Inclusivos: a) Estudos que abordam o uso de técnicas de Inteligência Artificial para reconhecimento de emoções; b) Estudos que abordam os desafios tecnológicos para ferramentas de reconhecimento de emoções.

Critérios Exclusivos: a) O estudo está duplicado; b) O estudo não aborda técnicas de Inteligência Artificial para reconhecimento de emoções através de detecção facial; c) O estudo não está nos idiomas Português ou Inglês; d) O estudo não foi publicado nos últimos quatro anos e) O estudo veio com informações incompletas.

Para a segunda seleção de artigos, que prevê uma avaliação de qualidade, definiram-se perguntas avaliativas com base nas recomendações de (WEN, 2012), apresentadas na seção (2.3.3) deste trabalho. As perguntas estão descritas abaixo:

- 1) O objetivo da pesquisa está claramente descrito?
- 2) O estudo realizou um experimento bem descrito para avaliar a proposta?
- 3) O estudo identifica um experimento para avaliar a técnica utilizada?
- 4) O estudo mostra a eficácia ou eficiência da técnica utilizada?
- 5) A Base de Dados utilizada é informada?
- 6) A pesquisa apresenta uma aplicabilidade descrita?
- 7) A base de dados informada é de domínio público?

Essas perguntas são feitas após a leitura completa de cada trabalho. Para respondê-las também definem-se possíveis respostas, com pesos atribuídos, para no final comparar os resultados.

Na revisão em questão, as respostas definidas foram: Sim, Parcialmente e Não. Foram atribuídos os respectivos pesos: 1.0, 0.5 e 0.0. Sendo assim, a pontuação máxima para a avaliação de cada artigo era de 7 (sete) pontos.

3.1.2 Condução

Após conclusão da fase de planejamento, deu-se início ao processo de importação e seleção dos artigos das bases de dados. O uso da string de busca nas bases especificadas no planejamento resultou na identificação de 349 (trezentos e quarenta e nove) artigos.

Para estes 349 artigos, foi realizada a primeira seleção a partir da leitura dos seus resumos, a fim de realizar uma triagem dos trabalhos mais relevantes para a revisão, levando em consideração os critérios de inclusão e exclusão definidos previamente na etapa de planejamento.

A partir dos trabalhos aceitos na primeira seleção, deu-se início a uma segunda seleção de artigos, porém avaliando a qualidade do trabalho como um todo. Para essa avaliação foram aplicadas as perguntas e respostas de qualidade, descritas no final da seção 3.2.1.

Para cada artigo atribuiu-se uma nota de 0 (zero) a 7 (sete).

3.1.3 Relato da Revisão

Nesta fase foi realizada a análise e a discussão dos resultados relevantes para o estudo. Dessa forma, os trabalhos selecionados na última etapa foram analisados visando a extração e mapeamento de dados correspondentes às questões de pesquisa. O detalhamento desta etapa metodológica está descrito no capítulo 4, em Resultados e Discussões.

3.2 Conclusão

Este capítulo abordou as fases da SLR conduzidas pela plataforma Parsifal. No capítulo a seguir serão descritos os resultados obtidos na aplicação dos procedimentos metodológicos aqui descritos.

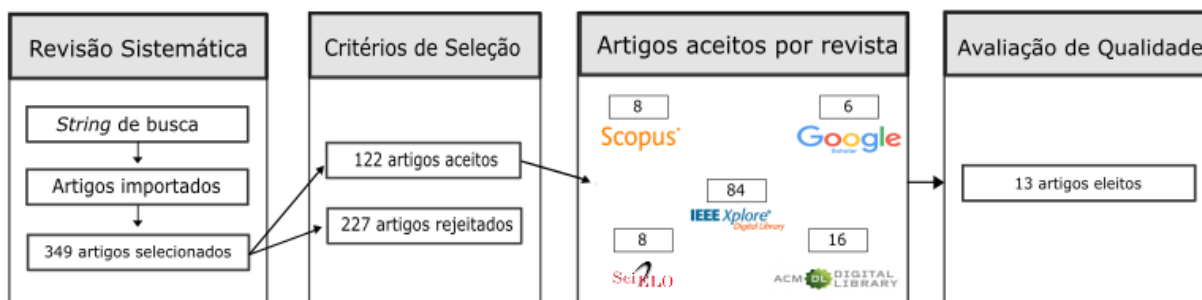
4 RESULTADOS E DISCUSSÕES

Este capítulo aborda os resultados obtidos na aplicação do protocolo SLR para mapear e sumarizar estudos que empregam técnicas, métodos e algoritmos de reconhecimento de emoções baseados em IA, como ML e Aprendizado Profundo.

4.1 Resultados Obtidos

Conforme pode ser visualizado na Figura 4.1, a partir da estratégia apresentada e com base na string de busca, foram selecionados inicialmente 349 (trezentos e quarenta e nove) artigos, que foram publicados entre os anos de 2018 e 2021, nos idiomas Inglês e Português.

Figura 4.1- Resultados da busca de artigos



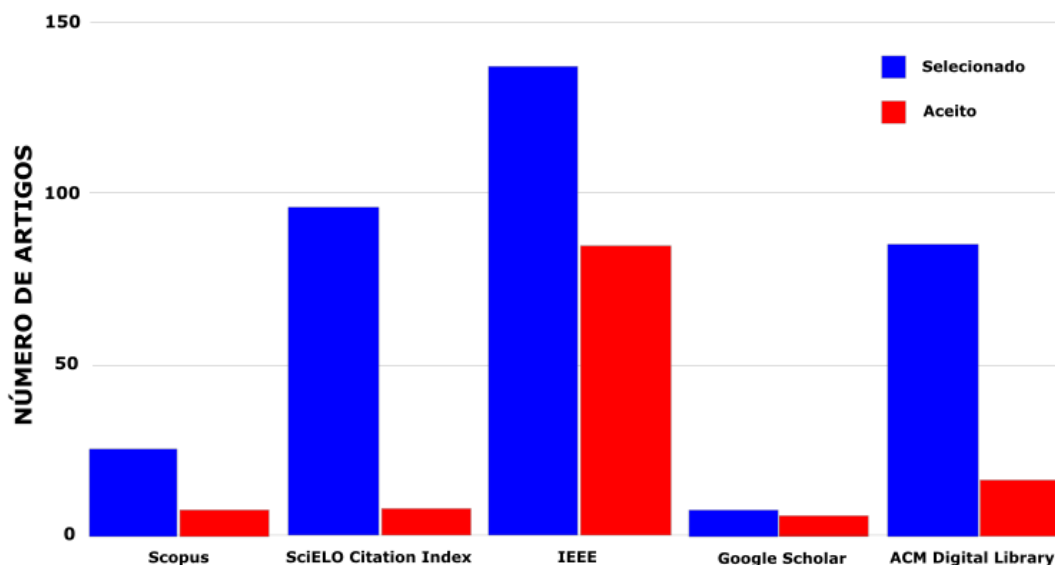
Fonte: Autora

Deste total de 349, 85 são da ACM Digital Library, 25 do SciELO Citation, 136 do IEEE Digital Library, 95 do SciELO Citation e apenas 8 do Google Scholar.

Após a primeira seleção, com a leitura dos resumos baseada nos critérios de inclusão e exclusão, 122 trabalhos foram selecionados. Entre os 227 descartados, predominaram trabalhos repetidos e que não abordavam técnicas para o reconhecimento de emoções através de detecção facial. Dentre os 122 artigos aceitos, 84 foram da IEEE Digital Library, 16 da ACM Digital Library, 8 do Scopus, 8 da SciELO Citation Index e 6 do Google Scholar.

A Figura 4.2 destaca a quantidade de artigos selecionados, na cor azul, e aceitos, na cor vermelha, considerando cada biblioteca.

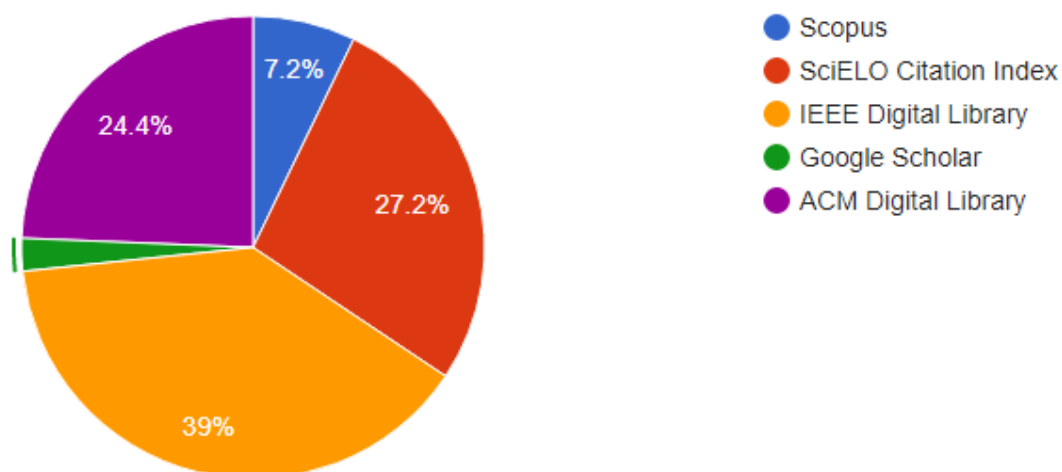
Figura 4.2- Quantidade de artigos selecionados



Fonte: Autora

A biblioteca com o maior índice de trabalhos aceitos na primeira seleção da pesquisa foi o IEEE Digital Library. A Figura 4.3 detalha a distribuição dos artigos selecionados por fontes bibliográficas.

Figura 4.3 - Distribuição dos artigos por fontes.



Fonte: Autora

O segundo processo de seleção feito na revisão sistemática resultou na eleição de 13 artigos, que completaram o método avaliativo de qualidade. Os 122 artigos aceitos na primeira seleção foram avaliados de acordo com as sete perguntas de qualificação,

definidas durante o planejamento da revisão e descritas na seção 3.2.1. Sendo assim, cerca de 76,5% dos trabalhos avaliados ficaram com notas acima de 5.5 e 23,5% com notas abaixo ou igual a 5. Os 13 artigos eleitos, obtiveram nota máxima (7.0) na avaliação, o que corresponde a 17,3% de todos os 122 artigos.

Para os artigos eleitos, foram filtradas as técnicas e as bases de dados de imagens utilizadas. Com o objetivo de classificar os melhores métodos, foram separados os principais recursos utilizados para reconhecer expressões faciais em imagens.

4.2 Discussão sobre os resultados obtidos

Discute-se os resultados obtidos de acordo com as Questões de Pesquisas descritas na seção 3.2.1

RQ1- Quais as técnicas de Inteligência Artificial mais utilizadas em reconhecimento de emoções?

Nos artigos selecionados, o reconhecimento de faces em imagem foi um dos fatores mais relevantes para a identificação de expressões faciais, onde as imagens eram reconhecidas a partir de uma base de dados. Dentre as bases de dados utilizadas, vale mencionar que databases como CK+⁸, JAFFE⁹ e FER2013¹⁰ se destacam dos demais, tendo melhor rendimento em CNN's pelo fato das imagens serem tratadas em laboratório; desta maneira, não possuindo ruído de imagem, sombras e interferências externas, fazendo com que os algoritmos de reconhecimento de imagem tenham um melhor rendimento em sua acurácia, como reforçado em (JIN, 2019).

4.2.1 CNN

Observou-se que todos os artigos selecionados utilizam alguma arquitetura de CNN para desenvolver um modelo de reconhecimento facial. Isso se dá porque as CNNs têm demonstrado bons resultados para segmentação de imagens e análise de rostos por regiões, sendo considerada como boa estratégia para identificar diferentes expressões faciais, conforme descrito em (SEPAS-MOGHADDAM, 2021), (DAI, 2017) e (CHEN, 2018).

⁸ <https://paperswithcode.com/dataset/ck>

⁹ <https://paperswithcode.com/dataset/jaffe>

¹⁰ <https://paperswithcode.com/dataset/fer2013>

O modelo de CNN VGG-16¹¹ está presente em aproximadamente 30% dos artigos. A VGG-16 foi originalmente desenvolvida para classificação em larga escala de imagens naturais, possuindo 13 camadas de convolução e 3 camadas totalmente conectadas (FC). Uma VGG-16 pré treinada é muito poderosa para diversas aplicações em segmentação de imagens, sendo comumente treinada com a base de dados CK+.

Outra rede neural convolucional recorrente é a ResNet50¹², utilizada pelos artigos (SEPAS-MOGHADDAM, 2021) e (AZA, 2020), sendo cerca de 15% dos trabalhos selecionados. Em relação ao número de camadas, a ResNet50 é muito mais profunda do que o VGG-16, com 50 camadas no total. A profundidade é crucial para as redes neurais, mas quanto mais profunda a rede, maior é a dificuldade de treiná-la, ou seja, exige um maior esforço computacional. Assim como a VGG-16, a ResNet50 é comumente treinada com a base de dados CK+. Outros modelos citados nos trabalhos selecionados são DCNN¹³ (7%), VGGNet¹⁴(7%), MTCNN¹⁵ (7%), CNN-LSTM¹⁶ + RNN¹⁷ (7%) , Senet50¹⁸ (7%), SVM¹⁹ () (7%) e CNN Fernet²⁰(7%).

Conforme a avaliação, as CNN's como VGG-16, VGG-2D e RESnet50 apresentaram maior eficiência para o reconhecimento de expressões faciais. Como mostrado nos artigos selecionados, CNN's com métodos de otimização possuem mais acurácia do que CNN's sem otimização. Destarte, algumas técnicas mais utilizadas são: 'Prediction', 'Face cropping' e 'Landmarks'. Ambas técnicas citadas procuram otimizar as bases de dados para um melhor reconhecimento de expressões faciais.

RQ2- Quais as bases de dados mais utilizadas nas pesquisas de reconhecimento de emoções?

¹¹ <https://neurohive.io/en/popular-networks/vgg16/>

¹² <https://www.kaggle.com/datasets/keras/resnet50>

¹³ <https://paperswithcode.com/method/dcn>

¹⁴ <https://viso.ai/deep-learning/vgg-very-deep-convolutional-networks/>

¹⁵ <https://github.com/ipazc/mtcnn>

¹⁶ <https://machinelearningmastery.com/cnn-long-short-term-memory-networks/>

¹⁷ <https://www.tensorflow.org/guide/keras/rnn>

¹⁸ <https://github.com/hujie-frank/SENet>

¹⁹ <https://scikit-learn.org/stable/modules/svm.html>

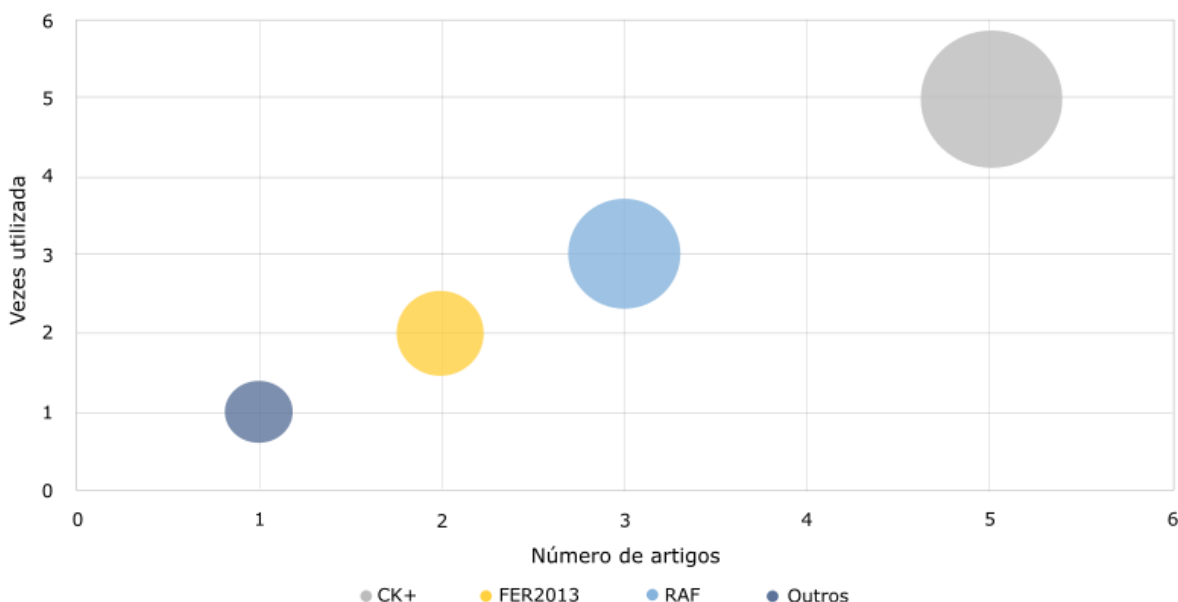
²⁰

<https://openbase.com/js/fernet/documentation>

4.2.2 DATABASE

Conforme um dos critérios de avaliação descritos na seção 3.2.1, todos os artigos eleitos utilizaram base de dados públicas para treinar seus modelos de CNN. Conforme ilustrado na Figura 4.4, cerca de 5 artigos, que correspondem a 38% dos trabalhos, citam a utilização da base de dados pública CK+ disponibilizada pela plataforma Kaggle²¹ e selecionada por ser uma base criada para aplicações de reconhecimento facial. O CK+ possui imagens, com resoluções de 640 x 490 ou 640 x 480 pixels, de pessoas na faixa etária de 18 a 50 anos executando 23 expressões faciais.

Figura 4.4 - Bases de dados mais utilizadas



Fonte: Autora

A segunda base de dados mais utilizada (23%) foi a RAF (Real-world Affective Faces)²², que tem destaque em tarefas de reconhecimento de expressões faciais, uma vez que as imagens nesse dataset possuem grande variância de características, tais como idade, sexo, etnia, condições de iluminação e outras. Essa variedade de características aumentam o aprendizado dos modelos, tornando possível a distinção de expressões em diversos ambientes e pessoas.

²¹

<https://www.kaggle.com>

²² <https://paperswithcode.com/dataset/raf-db>

Outra base de dados que teve destaque, utilizada em 15% dos artigos é a FER-2013²³. A base também está disponibilizada na plataforma Kaggle. FER-2013 possui aproximadamente 30.000 fotos nos canais RGB com imagens de 48 x 48 pixels, sendo 28,709 imagens para treino e 3,589 para teste. As imagens da Fer2013 são rotuladas em 7 tipos de expressões: raiva, alegria, tristeza, medo, surpresa, desgosto e neutra, e devido às dimensões das imagens não são necessárias muitas alterações para a utilização em um modelo de ML.

Outra base de dados bastante utilizada foi a JAFFE, presente em 7% dos artigos selecionados. A base de dados JAFFE (Japanese Female Facial Expression) conta com imagens de 10 indivíduos, todas são mulheres japonesas, expressando sete diferentes expressões faciais: neutra, alegria, raiva, desgosto, medo, tristeza e surpresa. Cada mulher possui de dois a quatro exemplos para cada expressão facial. No total, existem 213 imagens em preto e branco na base de dados Jaffe, cada uma com um tamanho de 256 x 256 pixels. Uma aplicação interessante para esta base é treinar modelos para reconhecer microexpressões faciais, pois é composta por japoneses cuja cultura utiliza expressões faciais mais sutis.

Alguns trabalhos optaram ainda por desenvolver a própria base de dados, como LFFW (Light Field Faces in the Wild) e LFFC (Light Field Faces Constrained), utilizadas pelo artigo (SEPAS-MOGHADDAM, 2021) e compostas completamente por imagens capturadas com uma câmera de campo de luz, com o objetivo de treinar modelos que reconhecem expressões faciais de vários ângulos diferentes, sem que seja necessário tirar diversas fotos.

Além das databases já citadas foi observado o uso de bases como AVEC 2013 e 2014 (7%), AffectNet (15%), FERF e MUG (7%), RAF (23%), KDEF (7%) e MMI (15%).

4.3 Sumarização e discussão dos artigos selecionados

A Tabela 4.1 sumariza as referências, técnicas e performance dos modelos de inteligência artificial, selecionados na revisão sistemática.

Tabela 4.1 - Técnicas e performance dos 13 artigos obtidos na avaliação qualitativa

²³

<https://paperswithcode.com/dataset/fer2013>

Autor	Técnica	Melhor performance
(HAQUE, 2018)	DCNN, VGGNet and LEAP Cluster to Train	Acurácia= 63,11%
(LI, 2019)	CNN-LSTM and RNN to capture contextual information and facial expressions	Acurácia=85,97%
(SEPAS-MOGHAD DAM, 2021)	VGG-16, Resnet50 and Capsnets for spatial and angular information. Light Field Toolbox v0.4	Média: 89.10%
(ACHARYA, 2018)	Multitasking Cascade (MTCNN), VGG-VD-16 and enhancement techniques, with: spdnet, reeig,logeig	Acurácia= 87%
(ESPINEL, 2021)	DCNN+ Autoencoder, DCNN1, DCNN2, DCNN3 AND DCNN4 based on VGG-16 deep neural network	Acurácia= 96%
(UEDA, 2019)	Dlib for facial points with Delaunay triangulation and ReLu function. Create a CNN Interception-v3	Não informado
(CHEN, 2020)	Action Unit Recognition (AUR) and face landmark detection (FLD) for facelandmark detection, KNN and created LDL-ALSG algorithm	Acurácia= 70.40%
(DAI, 2018)	Inspired by KMP and the combination of multiple stats	Média de Reconhecimento: 99.29%
(AZA, 2020)	VGG16, Resnet50, Senet50 retrained with Adam, RMSprop and Adagrad optimization algorithms were used. In addition to the YOLO v3 model for face detection.	Acurácia= 97%
(ZHANG, 2018)	CNN Fernet	Acurácia= 93.7%
(JIN, 2019)	Support Vector Machine (SVM), Bayesian networks and rule-based classifiers	Acurácia= 63.7%
(CHEN, 2018)	PPRL-VGAN, GAN, VAE, framework for learning an identity invariant representation for a face image	Acurácia=93.2
(ZHOU, 2018)	MR-DepressNet, a deep regression network that facilitates automated prediction of depression severity from facial images	RMSE: 8.28

Fonte: Autora.

Os artigos selecionados mostram suas performances e suas bases de dados. Pode-se destacar as excelentes performances que (SEPAS-MOGHADDAM, 2021), (DAI, 2017) e (CHEN, 2018) obtiveram em seus reconhecimentos de faces e expressões. Observa-se que (SEPAS-MOGHADDAM, 2021) usa como uma das CNN's, em destaque, a ResNET50+ sendo a principal CNN para reconhecer expressões em dados "in the wild", tendo a maior média de reconhecimento de face dentre as 3 CNN's destacadas nos resultados.

Em (ESPINEL, 2021), a precisão de reconhecimento foi a segunda mais alta de toda a revisão sistemática, usando o método de estatística MSL, para reconhecimentos de expressões faciais. Por último, (CHEN, 2018) obteve o maior desempenho de identificação de expressões faciais, sendo que o método com maior eficácia utilizava base de dados com ambientes controlados para obter resultados excelentes. Concomitante às CNN's, as databases têm um grande fator de impacto nos resultados de reconhecimento de expressões faciais.

Ressalta-se que os modelos de reconhecimento facial podem ser utilizados em diversas aplicações, como pôde ser observado nesta revisão sistemática. Dentre os artigos analisados, as áreas com aplicações são Computação Visual (70%), Saúde (15%), Segurança da Informação (7%) e Interação Humano-Robô (7%).

4.4 Conclusão

Foram apresentados e discutidos os resultados obtidos de acordo com as Questões de pesquisas apresentadas. O próximo capítulo apresenta as considerações finais.

5 CONSIDERAÇÕES FINAIS

Este trabalho apresenta o mapeamento e sumarização de estudos que abordam técnicas ou algoritmos mais utilizados no reconhecimento de expressão facial baseada em inteligência artificial, além de destacar as principais bases de dados.

Com isso, foi possível identificar os modelos de CNN's e databases mais utilizados e os seus desempenhos. Se sobressaem aqueles que fazem a leitura a partir de bancos de dados controlados em laboratório, sem interferência externa conforme (DAI, 2017), (CHEN, 2018) e (ACHARYA, 2018)

De modo geral, os estudos analisados abrangeram diversos aspectos das CNN's, desde erros mais comuns ao reconhecer expressões, até as dificuldades encontradas para o reconhecimento de faces em diversos ambientes, descrito no (ZHANG, 2018). Além disso, destaca-se o grande uso de CNN's como VGG e Resnet50, que combinadas com databases como a CK+, obtiveram um elevado desempenho em seus testes, usados em (ZHOU, 2018).

Considerando o caráter de replicabilidade de uma Revisão Sistemática da Literatura, espera-se que o relato do estudo em questão possa servir como base para pesquisas futuras na área de reconhecimento de faces e expressões, expandindo o conhecimento sobre as técnicas e ajudando no desenvolvimento da área.

O alto desempenho das técnicas identificadas, juntamente com as bases de dados para treinamento das mesmas, evidencia potenciais melhorias para o módulo de inferência de emoções da ferramenta WHOT, que quando implementadas irão contribuir significativamente para diagnósticos mais precisos da ferramenta.

5.1 Dificuldades Encontradas

Durante a execução desse trabalho houveram algumas dificuldades na execução da Revisão Sistemática no que diz respeito aos seus protocolos de planejamento. Apesar disso, a Revisão atingiu o seu objetivo principal.

5.2 Trabalhos Futuros

Embora o método de avaliação tenha selecionado poucos artigos em relação à base inicial de 349 artigos, ressalta-se a importância dos dados obtidos para futuras pesquisas.

Além disso, sugerem-se novas publicações que deem continuidade à amplitude de reconhecimento de expressões faciais, com novos modelos de CNN's e databases, gerando novos desempenhos e novas formas de reconhecimento de expressões.

6 REFERÊNCIAS

(ANTRA). Dossiê Assassinatos e violência contra travestis e transexuais brasileiras em 2020. 2021.

ACHARYA, Dinesh et al. Covariance pooling for facial expression recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops. 2018. p. 367-374.

AZA, Marde Fasma'ul; SUCIATI, Nanik; HIDAYATI, Shintami Chusnul. Performance Study of Facial Expression Recognition Using Convolutional Neural Network. In: **2020 6th International Conference on Science in Information Technology (ICSITech)**. IEEE, 2020. p. 121-126.

AZEVEDO, Breno FT; TAVARES, Orivaldo de L. Um ambiente Inteligente para aprendizagem Colaborativa. In: **Brazilian Symposium on Computers in Education (Simpósio Brasileiro de Informática na Educação-SBIE)**. 2001. p. 331-339.

BERCHT, Magda. Computação Afetiva: vínculos com a psicologia e aplicações na educação. **Psicologia e informática: produções do III Psicoinfo e II jornada do NPPI. São Paulo: Conselho Regional de Psicologia 6a região, p. 106–115, 2006.**

CHEN, Jiawei; KONRAD, Janusz; ISHWAR, Prakash. Vgan-based image representation learning for privacy-preserving facial expression recognition. In: **Proceedings of the IEEE conference on computer vision and pattern recognition workshops**. 2018. p. 1570-1579.

CHEN, Shikai et al. Label distribution learning on auxiliary label space graphs for facial expression recognition. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2020. p. 13984-13993.

COSTA, Saulo William Da Silva et al. WHOT, a novel tool to assist women victims of violence: a case study in the Brazilian Amazon. **IEEE Access**, v. 9, p. 95046-95060, 2021.

DAI, Shuanglu; MAN, Hong. Mixture statistic metric learning for robust human action and expression recognition. **IEEE Transactions on Circuits and Systems for Video Technology**, v. 28, n. 10, p. 2484-2499, 2017

Data Science Academy. Deep Learning Book, 2022. Disponível em: <<https://www.deeplearningbook.com.br/>>. Acesso em: 13 de Julho. 2022

DE ALMEIDA TELES, Maria Amélia; DE MELO, Mônica. O que é violência contra a mulher. Brasiliense, p. 9, 2017.

DE OLIVEIRA, Eduardo; JAQUES, Patrícia Augustin. Inferindo as emoções do usuário pela face através de um sistema psicológico de codificação facial. In: **Proceedings of the VIII Brazilian Symposium on Human Factors in Computing Systems**. 2008. p. 156-165.

DE SEGURANÇA PÚBLICA, FBSP Anuário Brasileiro. Fórum Brasileiro de Segurança Pública, v. 14 (FBSP), 2022.

DE SEGURANÇA PÚBLICA, FBSP Anuário Brasileiro. Fórum Brasileiro de Segurança Pública (FBSP), v. 14, 2020b. 2021.

DE SEGURANÇA PÚBLICA, FBSP Anuário Brasileiro. Fórum Brasileiro de Segurança Pública (FBSP), 2019. 2019.

ESPINEL, Andrés et al. Face Gesture Recognition Using Deep-Learning Models. In: **2021 IEEE Colombian Conference on Applications of Computational Intelligence (ColCACI)**. IEEE, 2021. p. 1-6.

EKMAN, Paul; FRIESEN, Wallace V. Constants across cultures in the face and emotion. **Journal of personality and social psychology**, v. 17, n. 2, p. 124, 1971.

EKMAN, Paul. Universals and cultural differences in facial expressions of emotion. In: **Nebraska symposium on motivation**. University of Nebraska Press, 1971.

EKMAN, Paul; FRIESEN, Wallace V. Facial action coding system. **Environmental Psychology & Nonverbal Behavior**, 1978.

EKMAN, Paul; OSTER, Harriet. Facial expressions of emotion. **Annual review of psychology**, 1979.

EKMAN, Paul. *A Linguagem das Emoções*; São Paulo: Lua de Papel, 2011.

FÓRUM BRASILEIRO DE SEGURANÇA PÚBLICA (FBSP). Violência doméstica durante a pandemia de Covid-19, ed. 2, 2020a. Disponível em: <https://forumseguranca.org.br/wpcontent/uploads/2020/06/violencia-domesticacovid-19-ed02-v5.pdf>. Acesso em 13.07.2022.

FRIESEN, Wallace V.; EKMAN, Paul. EMFACS-7: Emotional Facial Action Coding System. **Unpublished manuscript**, University of California at San Francisco, 1983.

GALVÃO, Maria Cristiane Barbosa; RICARTE, Ivan Luiz Marques. Revisão sistemática da literatura: conceituação, produção e publicação. **Logeion: Filosofia da informação**, v. 6, n. 1, p. 57-73, 2019.

GIUNTINI, Felipe Taliar et al. How do I feel? Identifying emotional expressions on Facebook reactions using clustering mechanism. **IEEE Access**, v. 7, p. 53909-53921, 2019.

HAQUE, Md Inzamam Ul; VALLES, Damian. A facial expression recognition approach using DCNN for autistic children to identify emotions. In: **2018 IEEE 9th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON)**. IEEE, 2018. p. 546-551.

HE, Kaiming et al. Deep residual learning for image recognition. In: **Proceedings of the IEEE conference on computer vision and pattern recognition**. 2016. p. 770-778

HUBEL, David H.; WIESEL, Torsten N. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. **The Journal of physiology**, v. 160, n. 1, p. 106, 1962.

JAQUES, Patricia Augustin; VICARI, Rosa Maria. A BDI approach to infer student's emotions in an intelligent learning environment. **Computers & Education**, v. 49, n. 2, p. 360-384, 2007.

JIN, Xiaofang; XU, Ying. Research on Facial Expression Recognition Based on Deep Learning. In: 2019 3rd International Conference on Electronic Information Technology and Computer Engineering (EITCE). IEEE, 2019. p. 1144-1147.

LECUN, Yann et al. Gradient-based learning applied to document recognition. **Proceedings of the IEEE**, v. 86, n. 11, p. 2278-2324, 1998.

KEELE, Staffs et al. **Guidelines for performing systematic literature reviews in software engineering**. Technical report, ver. 2.3 ebse technical report. ebse, 2007.

LI, Tzoo-Hseng S. et al. CNN and LSTM based facial expression analysis model for a humanoid robot. **IEEE Access**, v. 7, p. 93998-94011, 2019.

MARENGONI, Maurício; STRINGHINI, Stringhini. Tutorial: Introdução à visão computacional usando opencv. **Revista de Informática Teórica e Aplicada**, v. 16, n. 1, p. 125-160, 2009.

PETTICREW, Mark; ROBERTS, Helen. **Systematic reviews in the social sciences: A practical guide**. John Wiley & Sons, 2008.

PICARD, Rosalind W. **Affective computing**. MIT press, 2000.

RAUBER, Thomas Walter. Redes neurais artificiais. **Universidade Federal do Espírito Santo**, v. 29, 2005.

SANTOS, Vitor. **Linguagem corporal: Guia prático para analisar e interpretar pessoas**. Fontanar, 2022.

SEPAS-MOGHADDAM, Alireza et al. Capsfield: light field-based face and expression recognition in the wild using capsule routing. **IEEE Transactions on Image Processing**, v. 30, p. 2627-2642, 2021.

SIDDAWAY, Andy P.; WOOD, Alex M.; HEDGES, Larry V. How to do a systematic review: a best practice guide for conducting and reporting narrative reviews, meta-analyses, and meta-syntheses. **Annual review of psychology**, v. 70, p. 747-770, 2019.

SIDHU, Prabhjot Kaur et al. Deep Learning Based Emotion Detection in an Online Class. In: **2022 IEEE Delhi Section Conference (DELCON)**. IEEE, 2022. p. 1-6.

SILVEIRA, Giulia Bordignon; DOS SANTOS, Daniela Scherer; DA MAIA, Gabriela Felten. Estamos Juntas: Expert system to support the identification and denunciation of violence against women. In: **Proceedings of the XIV Brazilian Symposium on Information Systems**. 2018. p. 1-9.

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

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

SOUSA, Ildenir Nascimento; DOS SANTOS, Fernanda Campos; ANTONIETTI, Camila Cristine. Fatores desencadeantes da violência contra a mulher na pandemia COVID-19: Revisão integrativa. **Revista de Divulgação Científica Sena Aires**, v. 10, n. 1, p. 51-60, 2021.

SOUSA, Elayne Kelly Sepedro et al. Elaboração e validação de uma tecnologia educacional acerca da violência contra a mulher. **Escola Anna Nery**, v. 24, 2020.

J. Ueda, K. Okajima, Face morphing using average face for subtle expression recognition, in: 2019 11th International Symposium on Image and Signal Processing and Analysis (ISPA), 2019. p. 187-192

WEN, Jianfeng et al. Systematic literature review of machine learning based software development effort estimation models. **Information and Software Technology**, v. 54, n. 1, p. 41-59, 2012.

WOHLIN, Claes et al. **Experimentation in software engineering**. Springer Science & Business Media, 2012.

WU, Jianxin. Introduction to convolutional neural networks. **National Key Lab for Novel Software Technology. Nanjing University. China**, v. 5, n. 23, p. 495, 2017.

YAMASHITA, Rikiya et al. Convolutional neural networks: an overview and application in radiology. **Insights into imaging**, v. 9, n. 4, p. 611-629, 2018.

ZHANG, Linhao et al. Research of facial expression recognition based on deep learning. In: 2018 IEEE 9th International Conference on Software Engineering and Service Science (ICSESS). IEEE, 2018. p. 1-4.

ZHOU, Xiuzhuang et al. Visually interpretable representation learning for depression recognition from facial images. **IEEE Transactions on Affective Computing**, v. 11, n. 3, p. 542-552, 2018.