



Universidade Federal do Piauí
Centro de Ciências da Natureza
Programa de Pós-Graduação em Ciência da Computação

Sentinela: Metodologia de Acompanhamento e Controle de Avaliação de Aprendizado em Plataforma de Ensino Virtual

Wellington Pacheco Silva

Teresina-PI, Agosto de 2022

Wellington Pacheco Silva

**Sentinela: Metodologia de Acompanhamento e Controle
de Avaliação de Aprendizado em Plataforma de Ensino
Virtual**

Dissertação de Mestrado apresentada ao Programa de Pós-Graduação em Ciência da Computação da UFPI (área de concentração: Sistemas de Computação), como parte dos requisitos necessários para a obtenção do Título de Mestre em Ciência da Computação.

Universidade Federal do Piauí – UFPI

Centro de Ciências da Natureza

Programa de Pós-Graduação em Ciência da Computação

Orientador: André Castelo Branco Soares

Teresina-PI

Agosto de 2022

FICHA CATALOGRÁFICA
Universidade Federal do Piauí
Sistema de Bibliotecas da UFPI – SIBi/UFPI
Biblioteca Setorial do CCN

S586s Silva, Wellington Pacheco.
Sentinela: metodologia de acompanhamento e controle de
avaliação de aprendizado em plataforma de ensino virtual /
Wellington Pacheco Silva – 2022.
51 f.: il.

Dissertação (Mestrado) – Universidade Federal do Piauí,
Centro de Ciências da Natureza, Pós-Graduação em Ciência da
Computação, Teresina, 2022.
“Orientador: Prof. Dr. André Castelo Branco Soares.”

1. Sistema de Avaliação Virtual. 2. Rastreamento ocular. 3.
Aprendizagem de máquina. I. Soares, André Castelo Branco. II.
Título.

CDD 004.54

Bibliotecária: Caryne Maria da Silva Gomes – CRB3/1461

“Sentinela: Metodologia de Acompanhamento e Controle de Avaliação de Aprendizado em Plataforma de Ensino Virtual”

WELLINGTON PACHECO SILVA

Dissertação apresentada ao Programa de Pós-Graduação em Ciência da Computação do Centro de Ciências da Natureza da Universidade Federal do Piauí, como parte integrante dos requisitos necessários para obtenção do grau de Mestre em Ciência da Computação.

Aprovada por:

Documento assinado digitalmente
 ANDRE CASTELO BRANCO SOARES
Data: 06/09/2022 17:34:53-0300
Verifique em <https://verificador.iti.br>

Prof. André Castelo Branco Soares
(Presidente da banca examinadora)

Documento assinado digitalmente
 MARIO ANTONIO MEIRELES TEIXEIRA
Data: 13/09/2022 11:14:14-0300
Verifique em <https://verificador.iti.br>

Prof. Mário Antonio Meireles Teixeira
(Examinador externo à instituição)

Documento assinado digitalmente
 KELSON ROMULO TEIXEIRA AIRES
Data: 09/09/2022 11:33:00-0300
Verifique em <https://verificador.iti.br>

Prof. Kelson Rômulo Teixeira Aires
(Examinador interno)

Documento assinado digitalmente
 RAIMUNDO SANTOS MOURA
Data: 12/09/2022 10:05:55-0300
Verifique em <https://verificador.iti.br>

Prof. Raimundo Santos Moura
(Examinador interno)

Teresina, 30 de agosto de 2022

*Aos meus pais, Aristides e Elmira,
pelo amor, apoio e esforço incondicionais.*

Agradecimentos

Agradeço a Deus por ter abençoado todos os dias da minha vida, por iluminar meu caminho e me dar forças para seguir sempre em frente.

Aos meus pais pelo amor, carinho e suporte.

Ao meu irmão, Wilson, pela amizade e companheirismo.

À minha querida esposa, Samilla, e minha amada filha, Heaven, pelo amor e carinho. Mesmo distantes, foram minha fonte de inspiração e motivação nesta caminhada.

Ao meu orientador, André Castelo Branco Soares, pela oportunidade, orientações, ensinamentos e competência.

Aos colegas do grupo de trabalho orientados pelo prof. André Soares, em especial, ao amigo Martony, pela troca de conhecimentos.

Aos professores da UFPI, primordialmente, ao prof. Ricardo Lira, pelo profissionalismo e conhecimentos adquiridos.

Aos colegas de turma, sobretudo, aos colegas de trabalho da STI/UFPI pela parceria.

Ao amigo, Gilberto Nunes, pelos aconselhamentos acadêmicos.

Aos demais amigos e familiares por partilharem dessa conquista.

*“Libertem-se da escravidão mental,
ninguém além de nós mesmos pode libertar nossas mentes.”*

Bob Marley

Resumo

Nos últimos anos, os exames on-line têm sido amplamente utilizados para avaliar o desempenho dos alunos. Durante a pandemia de covid-19, essa situação se intensificou e o uso de novas tecnologias possibilitou a diversificação de formas de ensino e aprendizagem a distância. O isolamento e a falta de interação face a face fizeram com que o monitoramento desse tipo de prova se tornasse um grande desafio. Diante deste cenário, o presente trabalho apresenta uma metodologia denominada Sentinela, referência à figura de um guarda, na qual se utiliza da combinação do rastreamento ocular do estudante e do uso da janela ativa do navegador durante a realização de exames on-line para gerar artefatos que possam auxiliar o avaliador na checagem de possíveis trapagens. O processo de rastreamento dos olhos baseia-se no uso de apenas a *webcam* do computador de cada estudante para a captura de fotos às quais são processadas e classificadas por meio de Rede Neural Convolutiva, atestando a posição do olhar (dentro ou fora da tela) no momento de cada foto. Já a dinâmica de verificação do uso da janela do navegador identifica se houve a troca de tela e até mesmo quais caracteres foram digitados. Nos experimentos, foi alcançada uma taxa de acerto médio de 98,75% e um desvio padrão médio de 0,104%. Os resultados obtidos demonstram que o Sentinela é capaz de municiar o avaliador com informações seguras envolvendo a confrontação do rastreamento ocular e a checagem da janela do navegador, ajudando a identificar algum tipo de trapagem durante o processo de avaliação.

Palavras-chaves: Ambientes de Aprendizado Virtual. Supervisão de Exames On-line. Aprendizagem de Máquina.

Abstract

In recent years, online exams have been widely used to assess student performance. During the covid-19 pandemic, this situation intensified and the use of new technologies made it possible to diversify forms of distance teaching and learning. The isolation and lack of face-to-face interaction made monitoring this type of evidence a major challenge. Given this scenario, the present work presents a methodology called Sentinela, a reference to the figure of a guard, which uses a combination of eye tracking of the student and the use of the active browser window during online exams to generate artifacts. that can help the evaluator in checking for possible cheating. The eye tracking process is based on the use of just the *webcam* of each student's computer to capture photos which are processed and classified by means of a Convolutional Neural Network, attesting to the position of the gaze (within or out of the screen) at the time of each photo. The dynamic verification of the use of the browser window identifies whether there was a screen change and even which characters were typed. The experiments were achieved an average success rate of 98.75% and an average standard deviation of 0.104%. The results obtained demonstrate that Sentinela is capable of providing the evaluator with secure information involving eye tracking confrontation and browser window checking, helping to identify any kind of cheating during the evaluation process.

Keywords: Virtual Learning Environments. Online Exam Proctoring. Machine Learning.

Lista de ilustrações

Figura 1 – Principais navegadores utilizados atualmente.	2
Figura 2 – Tela de Login do protótipo.	9
Figura 3 – Tela de gerenciamento de atividades.	9
Figura 4 – Tela de captura de imagens do estudante.	10
Figura 5 – Posição do olhar de cada participante.	11
Figura 6 – Coleta de imagens antes do início do procedimento.	12
Figura 7 – Coleta de imagens em execução.	12
Figura 8 – Fotos aptas x Fotos inaptas.	13
Figura 9 – Estrutura da arquitetura <i>LeNet</i>	14
Figura 10 – Operações de <i>pooling</i> na <i>LeNet</i>	14
Figura 11 – <i>LeNet</i> com entrada de uma imagem colorida.	15
Figura 12 – Fluxo de processamento das imagens.	15
Figura 13 – Demonstração das coordenadas em uma face.	16
Figura 14 – Tela de prova com mais de uma aba.	18
Figura 15 – Etapas da metodologia proposta.	19
Figura 16 – Status de prova do Sentinela.	21
Figura 17 – Tela de prova do protótipo em andamento.	21
Figura 18 – Arquitetura <i>LeNet</i> utilizada nesta pesquisa.	21
Figura 19 – Participantes do primeiro experimento.	23
Figura 20 – Posição do olhar de cada participante durante a prova.	27
Figura 21 – Média do número de capturas por posição do olhar do participante.	27
Figura 22 – Acurácia por participante.	28
Figura 23 – Duração duração de prova x quantidade de alertas.	28
Figura 24 – Quantidade de alertas x Duração total dos alertas.	29
Figura 25 – Gráficos de acurácia e perda no modelo de treino e teste.	31
Figura 26 – Confrontação na Prova 1: rastreador ocular X classificação pela CNN.	31
Figura 27 – Confrontação na Prova 2: rastreador ocular X classificação pela CNN.	32
Figura 28 – Confrontação na Prova 3: rastreador ocular X classificação pela CNN.	33

Lista de tabelas

Tabela 1 – Trabalhos sobre supervisão de provas on-line.	7
Tabela 2 – Matriz de confusão e suas variáveis.	16
Tabela 3 – Validação cruzada: K-fold, com k=5.	17
Tabela 4 – Especificação do computador para coleta do segundo experimento.	24
Tabela 5 – Total de fotos dos participantes do primeiro experimento.	26
Tabela 6 – Total de fotos dos participantes do segundo experimento.	30
Tabela 7 – Resultados do modelo da rede do segundo experimento.	30
Tabela 8 – Resultados da Validação Cruzada.	30
Tabela 9 – Resultados da classificação pela rede para o segundo experimento.	31

Lista de abreviaturas e siglas

API	Application Programming Interface - Interface de Programação de Aplicação
CNN	Convolutional Neural Network - Rede Neural Convolutacional
Covid-19	Coronavirus disease 2019 - Doença do Coronavírus 2019
CPU	Central Process Unit - Unidade Central de Processamento
CSS	Cascading Style Sheets - Folha de Estilo em Cascatas
DDR3	Double Data Rate 3rd generation - Taxa Dupla de Transferência terceira geração
DDR4	Double Data Rate 4th generation - Taxa Dupla de Transferência quarta geração
GB	Gigabyte
HTML	HyperText Markup Language — Linguagem de Marcação de Hipertexto
IP	Internet Protocol - Protocolo da Internet
MIT	Massachusetts Institute of Technology - Instituto de Tecnologia de Massachusetts
MOOCs	Massive Open Online Courses - Cursos On-line Abertos e Massivos
OMS	Organização Mundial de Saúde
RAM	Random Access Memory - Memória de Acesso Aleatório
RGB	Red, Green and Blue - Vermelho, Verde e Azul
VGA	Video Graphics Array - Matriz Gráfica para Vídeo
SSD	Solid State Drive - Unidade em Estado Sólido

Sumário

1	INTRODUÇÃO	1
1.1	Motivação	3
1.2	Objetivo Geral	3
1.2.1	Objetivos Específicos	3
1.3	Hipótese de Pesquisa	4
1.4	Trabalho publicado	4
1.5	Estrutura do Trabalho	4
2	PESQUISA BIBLIOGRÁFICA	5
3	MATERIAIS E MÉTODOS	8
3.1	Contextualização do trabalho de Mestrado	8
3.1.1	Ambiente Digital	8
3.1.2	Rastreamento ocular	10
3.1.3	Captura de imagens e formação da base de dados	11
3.1.4	Processamento e classificação das imagens	13
3.1.5	Métricas de Avaliação	16
3.1.6	Monitoramento do uso do navegador	18
3.2	Metodologia Proposta	19
3.2.1	Etapa 1	19
3.2.2	Etapa 2	19
3.2.3	Etapa 3	20
3.2.4	Etapa 4	20
3.2.5	Etapa 5	20
3.2.6	Etapa 6	20
3.2.7	Etapa 7	20
3.3	Experimentos	22
3.3.1	Primeiro experimento	22
3.3.2	Segundo experimento	23
4	RESULTADOS E DISCUSSÃO	26
4.1	Resultados do primeiro experimento	26
4.2	Resultados do segundo experimento	29
5	CONCLUSÕES E TRABALHOS FUTUROS	34
5.1	Trabalhos Futuros	34

REFERÊNCIAS	36
--------------------------	-----------

1 Introdução

A disseminação da internet tem propiciado o uso de tecnologias nos mais diversos cenários do dia a dia. Os ambientes de ensino virtuais possibilitam a realização de aulas que podem ocorrer tanto em tempo real através de transmissões ao vivo, como também por meio gravações, facilitando a interação entre professores e alunos que estão distantes geograficamente.

Segundo [Joye, Moreira e Rocha \(2020\)](#), o conceito de Educação a Distância (EaD) é bem mais simples: professores e alunos que estão mediando seu conhecimento por meio de interação síncrona e/ou assíncrona em espaços e tempos distintos, com ou sem uso de artefatos digitais. O termo “a distância” explica suas principais características: a separação física de professores e alunos no espaço, mas não impede o contato direto dos alunos entre si ou com os professores, a partir do uso da tecnologia.

Para [Bacich, Neto e Trevisani \(2015\)](#), o uso das tecnologias digitais no ambiente escolar oferece aos seus participantes diferentes possibilidades para um trabalho educativo mais significativo. A personalização da aprendizagem não é um substituto das aulas presenciais, mas sim um complemento no processo educacional, que além de melhorar a interação entre esses agentes da educação, favorece o uso de plataformas e/ou navegadores que disponibilizem o conteúdo que será consumido a partir de um computador ou um dispositivo móvel (*tablet, smartphone etc*).

As novas tecnologias digitais possibilitaram a ampliação da gama de formas de ensino e da aprendizagem a distância. Nesse contexto, a educação remota revolucionou o mercado educacional, especialmente com a inserção de plataformas específicas para modalidades: cursos on-line abertos e massivos (*MOOCs*), *Coursera, EdX, Udacity, Moodle*, entre outras ([INDI et al., 2021](#)). Essas ferramentas ou portais possibilitam o acesso a cursos de qualidade, tanto em universidades pelo Brasil como em instituições de renome internacional (*MIT, Stanford University*).

A pandemia de covid-19 afetou a população mundial no início de 2020 e exigiu uma ação emergencial dos governos nos mais diversos setores da sociedade para conter a propagação da doença. As medidas de distanciamento social sugeridas pela Organização Mundial de Saúde (OMS) e adotadas na maioria dos países, causaram o fechamento temporário das instituições de ensino, o que demandou um novo modelo educacional, sustentado pelas tecnologias digitais e pautado nas metodologias da educação remota ([VIEIRA; SILVA, 2020](#)). Assim, o número de cursos on-line e de estudantes nessas plataformas cresceu exponencialmente ([INDI et al., 2021](#)).

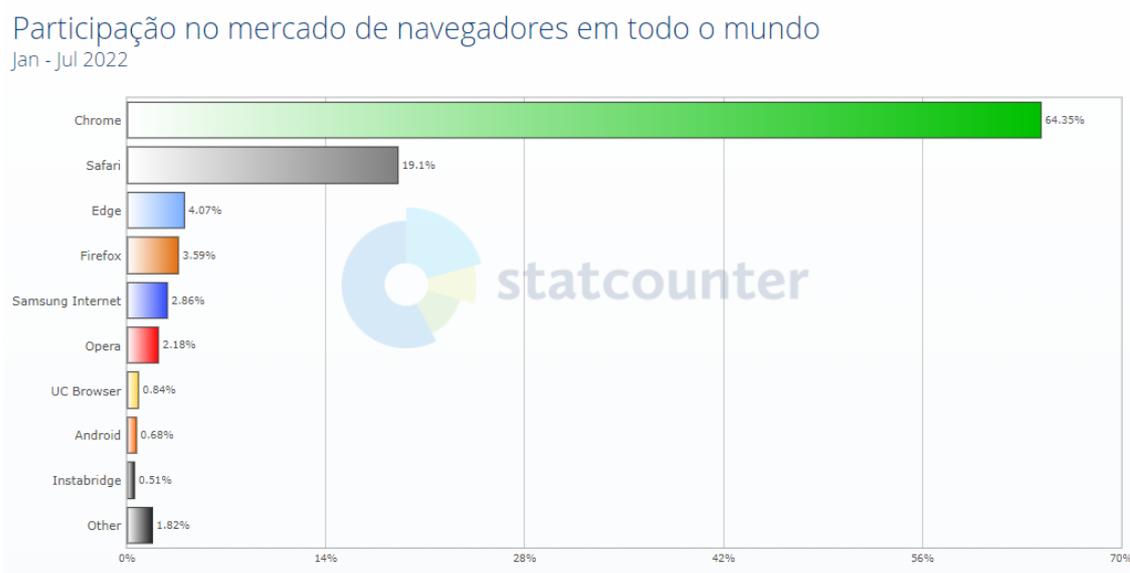
Os cursos na modalidade on-line também exigem que os alunos sejam avaliados

e certificados para a realização do treinamento. Portanto, torna-se necessário utilizar mecanismos de fiscalização na aplicação dos exames on-line. Tendo em vista que não há supervisão presencial, esse ainda é um desafio a ser superado uma vez que pode gerar tentativas de fraude, que exigem a verificação efetiva da identidade do aluno e o controle das tentativas de trapaça (PRATHISH; BIJLANI et al., 2016). Além disso, devido ao grande número de alunos em uma determinada turma, todo esse acompanhamento carece de algum tipo de automatização.

Nessa perspectiva, o objetivo deste trabalho é apresentar uma estratégia que auxilie o professor no gerenciamento da aplicação de exames no contexto do ensino remoto. Para isso, é apresentado um ambiente digital para controle do comportamento do estudante durante a realização do teste. Tal supervisão é composta por: i) monitoramento da janela do navegador que consta a prova, e ii) monitoramento do olhar do aluno durante a realização do exame.

A *webcam* será utilizada durante a prova para capturar as imagens que serão classificadas em uma Rede Neural Convolutiva (CNN) e será feita uma confrontação com as informações extraídas do monitoramento do navegador. O diferencial neste estudo é que o exame on-line pode ser utilizado nos navegadores mais comuns (Google Chrome, Safari, Microsoft Edge e Mozilla Firefox), conforme Figura 1, sem a necessidade de instalação de *plug-ins* ou outros softwares. Ademais, podemos embutir várias ferramentas de grande aceitação como o *Google Forms*, bastando apenas informar o URL no cadastro das avaliações do Sentinela. Por fim, esta pesquisa conduziu um teste de ambiente digital proposto a estudantes universitários em uma prova real e remota.

Figura 1 – Principais navegadores utilizados atualmente.



Fonte: StatCounter disponível em <https://gs.statcounter.com>. Acesso: 01/08/2022.

1.1 Motivação

Os exames on-line têm sido amplamente utilizados para a avaliação do desempenho dos alunos. Com a pandemia de covid-19 e o consequente isolamento social, esse cenário de provas remotas foi ampliado e trouxe consigo a necessidade de aumentar aspectos de controle e fiscalização. Uma alternativa interessante é o rastreamento ocular que permite a captura de informações importantes dos usuários, porém, tanto a aquisição desse tipo de dispositivo como as licenças de uso, possuem custos bastante elevados. A motivação deste trabalho é apresentar uma alternativa no monitoramento ocular de baixo custo por meio da própria *webcam* embutida nos computadores, e que empreenda o rastreamento das ações na tela do navegador durante a realização das provas.

O Plenário da Câmara dos Deputados do Brasil aprovou¹ no dia 04 de agosto de 2022, durante sessão deliberativa virtual, proposição que estabelece regras para os concursos públicos para contratação de servidores federais, referente ao Projeto de Lei nº 252, de 2003. A proposta autoriza a realização de provas a distância, de forma on-line ou por plataforma eletrônica com acesso individual seguro. O texto seguiu para análise pelo Senado. Este fato recente reforça ainda mais a relevância e a necessidade de alternativas que visem auxiliar o monitoramento de provas on-line.

1.2 Objetivo Geral

Este trabalho tem como objetivo principal propor uma metodologia que ajude no controle e fiscalização da aplicação de exames on-line.

1.2.1 Objetivos Específicos

- Criar um protótipo para monitorar e coletar informações a partir de avaliações on-line;
- Avaliar o comportamento da CNN utilizada juntamente com os ajustes realizados;
- Confrontar e atestar resultados por meio da utilização de um rastreador ocular vendido comercialmente;
- Avaliar a metodologia proposta por meio de um protótipo, em uma prova real;
- Validar a proposta de técnicas e mecanismos que auxiliem a integridade e lisura do processo de avaliação on-line;
- Avaliar a acurácia do método por meio de um dispositivo comercial de rastreamento ocular.

¹ <https://www.camara.leg.br/noticias/901670-plenario-aprova-proposta-que-cria-norma-geral-para-concursos-publicos-acompanhe/>

1.3 Hipótese de Pesquisa

A hipótese que norteia a presente pesquisa consiste em responder as seguintes questões: É possível inferir se o aluno está olhando para dentro ou fora da tela durante a realização de uma prova on-line? Os movimentos de cabeça e as ações do usuário na tela podem ser monitorados por meio de um computador com *webcam* usando apenas um navegador e sem a necessidade de softwares adicionais? A metodologia proposta auxilia na integridade e lisura do processo de avaliação remota?

No âmbito deste trabalho, pretende-se realizar uma prova real e remota, sem ambiente controlado, ou seja, sem que ocorram ajustes prévios nos equipamentos dos alunos, com o intuito de identificar algum tipo de trapaça. A metodologia proposta deve ser capaz de atuar nos mais diversos cenários, trazendo consigo uma fonte confiável de acompanhamento e controle no processo de avaliação.

1.4 Trabalho publicado

Durante o desenvolvimento das atividades de Mestrado, elaborou-se o artigo completo intitulado “Ambiente de Acompanhamento e Controle de Avaliação de Aprendizado em Plataforma de Ensino Virtual”, apresentado e publicado no XXXII Simpósio Brasileiro de Informática na Educação (SBIE 2021), ISSN: 2316-6533, como parte do X Congresso Brasileiro de Informática na Educação (CBIE 2021). Este trabalho ajudou a definir as estratégias utilizadas na elaboração da metodologia proposta pelo Sentinela.

Informações sobre a publicação:

- Autores: Wellington Pacheco, Martony Demes, James da L. Dias, Tardelly Cavalcante, André Soares, Jordão Frazão, Anselmo Paiva;
- DOI: <<https://doi.org/10.5753/sbie.2021>>.
- Qualis: B1

1.5 Estrutura do Trabalho

Este trabalho está organizado em 5 capítulos. No Capítulo 2 são expostos os trabalhos relacionados ao tema proposto. O Capítulo 3 apresenta materiais e métodos, trazendo o aparato teórico empregado na metodologia proposta, as etapas da metodologia em questão e os experimentos realizados. O Capítulo 4 mostra os resultados alcançados através da metodologia proposta. Ao final, é exposto o Capítulo 5 com as conclusões e trabalhos futuros.

2 Pesquisa Bibliográfica

O método de monitoramento durante a aplicação de provas on-line pode ser empregado no movimento dos olhos e realizado através do rastreamento da tela ativa de um computador ou dispositivo móvel. A partir deste conjunto de informações, sinais de comportamento indesejado podem ser inferidos no contexto desse tipo de prova. A seguir, apresenta-se um levantamento de pesquisas científicas que exploram alguns desses controles.

O trabalho de [Prathish, Bijlani et al. \(2016\)](#) apresentou uma proposta dividida em três módulos que incluem a captura da janela ativa registrando as mudanças de tela, processamento de entrada de vídeo, e captura de áudio. Esse estudo também utilizou a gravação de vídeo para extração de *frames* (quadros individuais) para determinar a posição da cabeça e o rosto do aluno. A acurácia desta pesquisa foi de 80%, e trouxe como diferencial a captura de áudio, onde pode-se detectar as variações de ruído nas proximidades do exame.

A pesquisa de [Hu, Jia e Fu \(2018\)](#) propôs um software que possibilitou a obtenção de informações sobre a face, postura da cabeça e boca do aluno durante o exame on-line, e a verificação do comportamento indevido do estudante. Para tanto, utilizou-se uma combinação dos algoritmos *AdaBoost* e *Haar Cascade* para detectar a face a partir das imagens capturadas. Somado a isso, realizou-se a estimativa da posição da cabeça por meio de CNN. Ademais, foi rastreada a posição da boca, feita a conversão das imagens capturadas para escala de cinza e determinado se a boca estava aberta ou fechada. O autor não executa o monitoramento de captura de tela e não relaciona a abertura da boca com a reprodução de áudio. A taxa de monitoramento e reconhecimento foi de 90%.

Outra pesquisa relacionada, foi o estudo de [Asep e Bandung \(2019\)](#). Nesse trabalho, foi colocado um sistema proprietário que também pode detectar os movimentos dos olhos. A proposta foi de realizar a verificação do rosto sem aplicar processamento adicional como equalização de histograma ou alinhamento de face para superar as variações de pose e iluminação. No entanto, ele não detectou a tela de janela ativa. O diferencial foi o uso do ambiente voltado para dispositivos móveis (*Android*).

O estudo de [Muzaffar et al. \(2021\)](#) forneceu um levantamento bibliográfico de trabalhos relacionados ao assunto entre 2016 e 2020, ilustrando as principais pesquisas sobre as técnicas utilizadas para supervisionar o exame on-line. A pesquisa apresentou 16 algoritmos de redes neurais mais utilizados, 11 conjuntos de dados e 21 ferramentas de exames on-line. [Muzaffar et al. \(2021\)](#) apontou ainda quatro fatores que são cruciais para que os exames on-line sejam adotados de maneira global: Infraestrutura de rede, requisitos

de hardware, complexidade de implementação e requisitos de treinamento. Dessa forma, cada país pode buscar a forma mais adequada de trabalhar sua infraestrutura de educação a distância existente e levando em conta a sua situação econômica.

A pesquisa de [Indi et al. \(2021\)](#) sugeriu um modelo de sistema de atenção visual do aluno que detecta movimentos oculares considerados inadequados no exame on-line, no qual alcançou uma acurácia de 96,04%. O aplicativo proposto extrai e usa *frames* (quadros individuais) de um vídeo ao vivo. Cada quadro extraído é alimentado em um classificador, que determina se um rosto foi detectado ou não. O classificador utilizou o algoritmo *Haar Cascade* que trata o aprendizado de máquina para detecção de objetos em uma imagem ou vídeo. Observamos que este trabalho não envolveu o monitoramento de janela ativa, nem utilizou um navegador da web, muito menos conduziu experimentos em alunos.

O estudo [Li et al. \(2021\)](#) propôs um método para supervisionar provas on-line por meio de captura de face, olho, tela e mouse. A diferença neste trabalho foi a utilização de um ambiente com *plugin* (*) instalado através de um navegador web que foi aplicado numa prova envolvendo 24 alunos. Durante a aplicação da prova foram gerados vídeos de 30 *frames* por segundo para analisar a posição da cabeça e se o rosto estava presente na filmagem. No processo de identificar a face, usou-se um detector de aprendizagem profunda, chamado Dockerface, que não requer treinamento, porém requer instalação. Já no processo de captura do mouse foram rastreados movimentos específicos como: foco, perda do foco, copiar, colar, mouse em movimento e mouse rolando a tela. O uso do monitoramento da tela ativa foi relacionado apenas ao uso do mouse, deixando margem para o usuário tentar fraudes trocando a janela por meio do teclado.

Nessa perspectiva, esta pesquisa confronta esses trabalhos citados, com o objetivo de comparar características e evidenciar as diferenças exploradas nesta apuração. Neste levantamento, verificou-se quais trabalhos fazem uso da internet e integram a tecnologia do *Google Forms*, um mecanismo largamente aceito que varia desde a coleta de dados até a avaliação do aluno e que, para [Sandhya et al. \(2020\)](#), há uma série de vantagens como: número ilimitado de formulários, criação e incorporação intuitivos, fácil compartilhamento, notificação eficiente, visualização também em dispositivos móveis e possibilidade de customização através do *Google Apps Script*, que é uma plataforma do *Google* para desenvolvimento de aplicativos.

Diante desse cenário, a Tabela 1 reúne as principais características exploradas nos trabalhos citados no contexto de supervisão de provas on-line.

Tabela 1 – Trabalhos sobre supervisão de provas on-line.

TRABALHOS	DMOI	MJA	UNW	UGF	APR
(PRATHISH; BIJLANI et al., 2016)	✓	✓		✓	
(HU; JIA; FU, 2018)	✓				
(ASEP; BANDUNG, 2019)	✓				
(INDI et al., 2021)	✓				
(LI et al., 2021)	✓	✓	✓*		✓
Proposta desta pesquisa	✓	✓	✓	✓	✓

DMOI (Detecção de Movimento Ocular Inapropriado); MJA (Monitoramento de Janela Ativa); UNW (Uso de Navegador Web); UGF (Uso do *Google Forms*); APR (Aplicação de uma Prova Real). textitPlug-in*.

Ante o exposto, observa-se que todos os trabalhos detectaram movimentos oculares. No entanto, apenas duas pesquisas investigaram o monitoramento de janela (ou tela) ativa. É importante notar que apenas uma pesquisa avaliou os alunos num cenário real, sendo que foi necessária a instalação de textitplug-in. Ademais, nenhum trabalho usou a tecnologia do *Google Forms*.

3 Materiais e Métodos

3.1 Contextualização do trabalho de Mestrado

Este trabalho propõe criar uma metodologia que possa ser implementada na web com independência de plataforma e que a partir de um computador com *webcam* seja capaz de realizar o acompanhamento de provas on-line confrontando o rastreamento ocular do estudante e as ações na tela da prova. Nesta seção, serão apresentados: o protótipo utilizado como ambiente digital, o processo de rastreamento ocular, a captura e formação da base de dados, o processamento e classificação das imagens, métricas de avaliação, e por fim, o monitoramento do uso do navegador. Além disso, será detalhada a metodologia proposta e os experimentos realizados.

3.1.1 Ambiente Digital

Foi desenvolvido um protótipo que permite o gerenciamento de turmas, relacionando alunos e atividades. Cada atividade possui um link de formulário do *Google Forms* (criado previamente pelo docente) e um prazo de realização (data/hora de início e término). Através desse ambiente, foi possível inspecionar a janela do navegador e os movimentos faciais e oculares dos alunos durante a realização das avaliações.

A ferramenta foi projetada usando a linguagem de programação PHP (versão 7.1.33). Utilizou-se também o *framework* Bootstrap que permitiu acelerar o processo de programação envolvendo HTML, CSS e JavaScript, além de tornar o layout do projeto responsivo. Dessa forma, pode-se adaptar a execução da ferramenta em qualquer tamanho de tela.

Além da responsividade para o layout, buscou-se criar uma solução que não dependesse da instalação de *plug-ins* ou outros softwares, e que pudesse ser utilizada nos navegadores mais comuns, conforme Figura 1 citada no capítulo de introdução.

A solução está hospedada num ambiente Linux por meio de um servidor web Apache e realiza conexão com o banco de dados MySQL (versão 5.7.36). Por conta da necessidade de processamento e armazenamento, o projeto foi alocado numa máquina com 3 núcleos de CPU, 160 GB de armazenamento e 4 GB de memória RAM DDR4.

O arquetipo possui uma área restrita exclusiva aos docentes para que possam administrar suas atividades. Os docentes poderão cadastrar seus alunos em suas respectivas turmas, bem como criar tarefas individuais com data e hora de realização pré-definidas. A Figura 2 mostra as credenciais necessárias para o acesso.

termo de consentimento livre e esclarecido para participar desta pesquisa. Em seguida, é apresentada uma tela, conforme Figura 4, onde é acionada a *webcam* do computador e o aluno deve se posicionar adequadamente para iniciar a prova. Ao clicar em "Iniciar a Avaliação", é apresentado o formulário do *Google* relacionado à atividade. Após concluir a avaliação, o aluno deve clicar em "Finalizar Avaliação" para encerrar o acesso à prova.

Durante todo o espaço de tempo entre o início e o término da prova, a *webcam* faz o registro da foto do aluno a cada 1/4 de segundo, ou seja, são registradas 4 fotos por segundo durante todo o tempo da avaliação. Essas fotos são armazenadas em tempo real no servidor.

Figura 4 – Tela de captura de imagens do estudante.



O protótipo foi desenvolvido para aceitar links de outras plataformas além dos formulários *Google*, pois essas aplicações externas podem ser embutidas nos corpos das atividades.

3.1.2 Rastreamento ocular

O rastreamento ocular é um mecanismo usado definir o ponto onde a pessoa está olhando através da medida do movimento ocular (POOLE; BALL, 2006). Para Lai et al. (2013) a metodologia de rastreamento ocular vem sendo usada largamente em pesquisa de aprendizagem, onde o docente ajuda os alunos a encontrar soluções para problemas na educação através do movimento do olho.

Utilizou-se nesta pesquisa o rastreamento ocular através da *webcam* do computador para avaliar o comportamento dos alunos através dos olhos e ajudar a inferir possíveis tentativas de trapaceas. Durante a fase de experimentação, realizamos testes por meio de um rastreador ocular vendido comercialmente e que segundo o fabricante, são emitidos feixes de luz na direção dos olhos do usuário e um conjunto de câmeras capturam imagens dos olhos que são processadas e apontadas informações precisas do olhar.

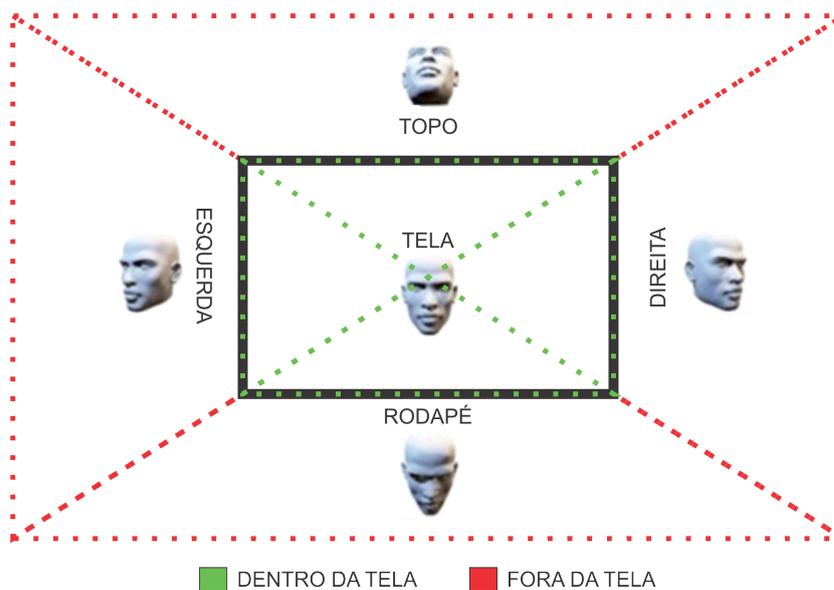
3.1.3 Captura de imagens e formação da base de dados

O processo de captura das imagens visa formar uma base de dados a ser treinada. Após uma busca em diversas plataformas por bases públicas que pudessem atender ao requisito de fotos de pessoas olhando para dentro e fora da tela do computador. Como não encontramos, optamos por gerar uma base própria.

Como citado na subseção anterior, as fotos dos participantes foram coletadas a cada 1/4 de segundo. Foram realizados experimentos no sentido de diminuir esse tempo, proporcionando mais fotos a serem capturadas, porém esse processo provocou lentidão no upload de fotos ao servidor, ou em alguns casos, as fotos registradas ficaram borradas pela tentativa exagerada de captura.

Os participantes foram orientados a posicionar a cabeça para regiões pré-definidas e a realizar movimentos suaves mantendo o foco do olhar na região alvo. Estas regiões foram separadas em dentro e fora da tela, com isso estabelecemos as seguintes classes: 0-TELA, 1-TOPO, 2-DIREITA, 3-RODAPÉ e 4-ESQUERDA, ilustradas na Figura 5.

Figura 5 – Posição do olhar de cada participante.



Para cada região foi organizado um intervalo de tempo padrão e ao final de cada período, foi emitido um sinal sonoro, auxiliando o usuário na transição do olhar e evitando que tivesse que olhar para a tela para saber quando deveria fazer a troca de região.

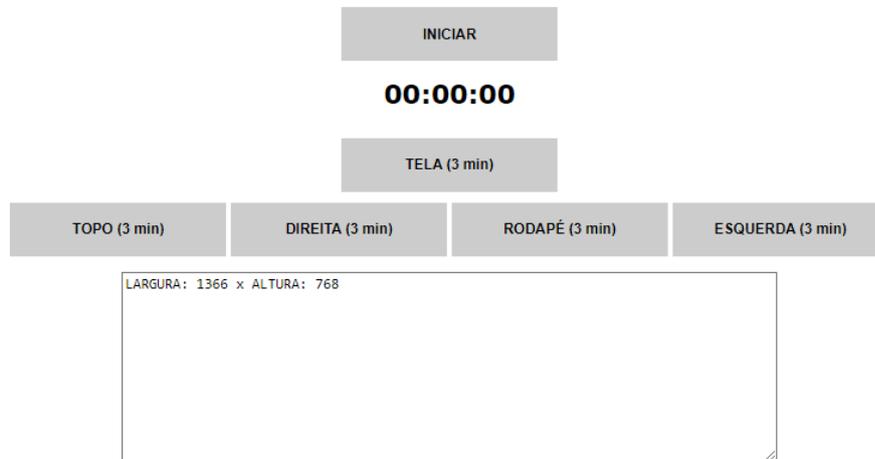
É importante ressaltar que realizamos diversas simulações quanto ao posicionamento do olhar nas regiões de borda dos monitores, onde ao olhar para o extremo de um dos lados da tela, não havia a possibilidade desta foto ser apontada como de outra região.

Já nas regiões de fora da tela, as coordenadas horizontais e verticais do ponto onde o olho estava focado, possibilitaram definir se ao olhar para o lado superior direito, por exemplo, estava realmente olhando para o topo ou para a direita. Essas comprovações

ocorreram por meio do rastreador ocular vendido comercialmente.

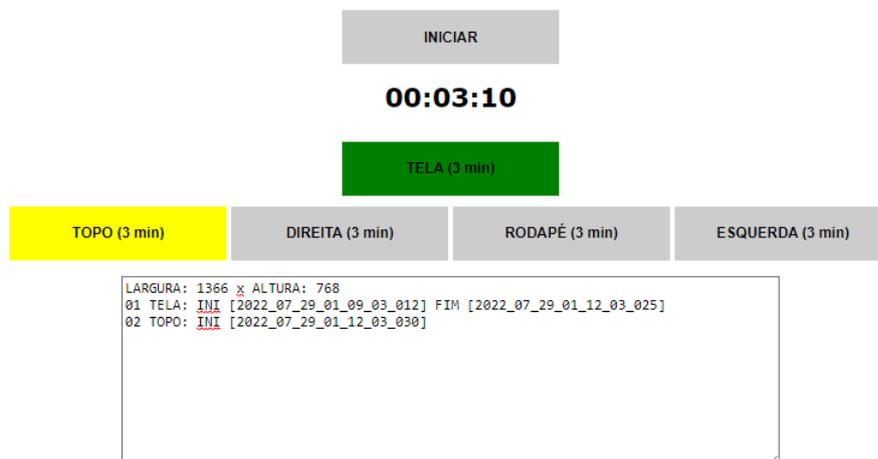
A Figura 6 apresenta a tela inicial utilizada para a coleta de imagens.

Figura 6 – Coleta de imagens antes do início do procedimento.



A Figura 7 mostra o momento exato em que foi realizada a coleta da primeira região (tela), e que a região do topo está em andamento. A cada região foram feitos os registros de início e término, e cada foto é armazenada com o rótulo de sua respectiva região.

Figura 7 – Coleta de imagens em execução.



Ao final do processo de captura, são contabilizadas as fotos e descartadas as fotos tidas como inaptas, ou seja, aquelas que não se enquadram nas regiões definidas. A Figura 8 mostra situações de imagens onde 1A e 2A o usuário está olhando para o rodapé, e 1B e 2B o usuário está olhando para a direita.

Observe que nas imagens aptas é possível identificar os olhos. A partir dessas informações poderemos trabalhar a segmentação dessas imagens. Essa temática será abordada na próxima subseção. Já as imagens inaptas são colocadas como descarte. Ao final, checka-se qual região possui a menor quantidade de fotos aptas e são excluídas das

Figura 8 – Fotos aptas x Fotos inaptas.



demais regiões a porção necessária para que fiquem com a mesma fração em cada região, dessa forma garantimos o balanceamento da base.

Outra observação importante sobre a Figura 8, é que nas imagens 1A e 2A, podemos distinguir quando o usuário está olhando para o teclado (1A) e quando o usuário está olhando para uma região mais abaixo (2A), como no caso da utilização de um celular para consultar alguma informação.

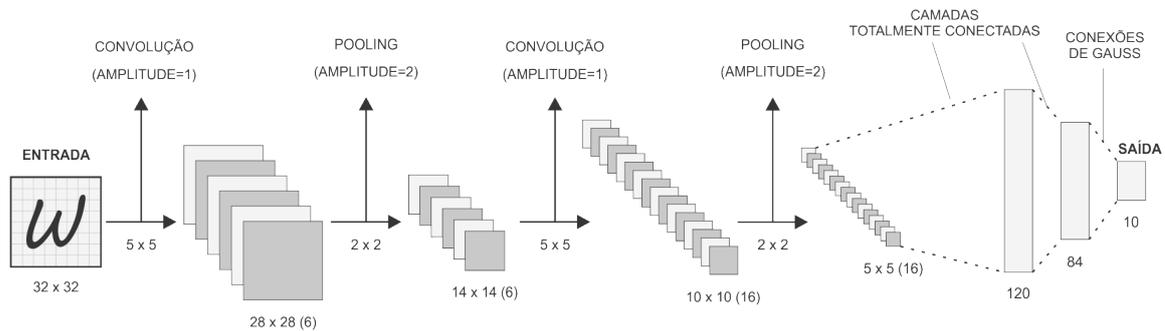
3.1.4 Processamento e classificação das imagens

O reconhecimento de imagem é um clássico problema de classificação, e as Redes Neurais Convolucionais (CNN) possuem um histórico de alta acurácia para esse tipo de problema. Optou-se por usar a LeNet5 (LECUN et al., 1998), com sete camadas entre convoluções e totalmente conectadas, por possibilitar um retorno no processo de classificação e também por ser uma rede bastante estudada no nosso grupo de trabalho, Disnel/UFPI.

A *LeNet* foi uma das primeiras arquiteturas de redes convolucionais verdadeiramente bem sucedidas e foi usada para classificar dígitos manuscritos (LECUN et al., 1998). A Figura 9 mostra a composição da rede em sete camadas. A primeira é uma camada convolucional com formato 28×28 pixels. Esse formato ocorre pelo fato dos campos receptivos terem 5×5 pixels com uma amplitude de 1, portanto a rede considera a perda, na imagem de entrada, de 2 pixels do lado direito e 2 pixels da parte inferior, resultando em 28 pixels na camada seguinte. A segunda camada é uma camada de agrupamento (*pooling*). A cada 2×2 , ou seja, 4 ativações de neurônios são reduzidas a uma única ativação, não permitindo sobreposição. Ao final, o tamanho do mapa é dividido por dois, dessa forma, das seis características da primeira camada, são gerados 6 mapas para a segunda, reduzindo de 28×28 para 14×14 pixels.

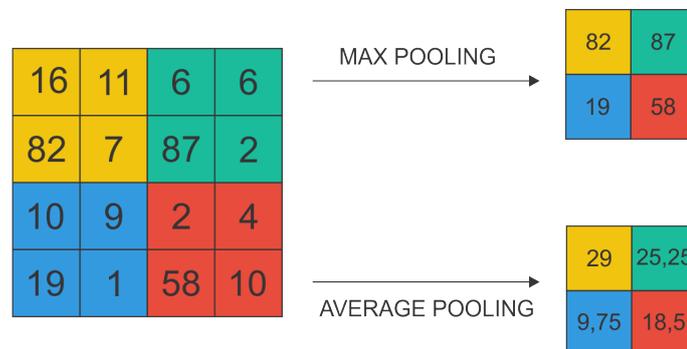
O agrupamento pode ocorrer de algumas formas. Na operação de *Max Pooling*

Figura 9 – Estrutura da arquitetura *LeNet*.



(RANZATO et al., 2007), o maior valor num grupo de 2x2 se torna o valor na próxima camada. Já na *Average Pooling* (LECUN et al., 1998) calcula-se a média de cada grupo 2x2. O que importa nessa fase é que seja obtida a redução da dimensionalidade tomando grupos de *pixels* ou saídas de neurônios, reduzindo-os ao final a uma única saída. A Figura 10 ilustra ambas as operações.

Figura 10 – Operações de *pooling* na *LeNet*.



O processamento das imagens pela *Lenet* não é limitado a preto e branco, onde são trabalhados os *pixels* a partir da intensidade. A Figura 11 mostra o processamento de uma imagem colorida com entrada de dimensões 32 x 32 x 3, onde o 3 denota os três canais de cores RGB (vermelho, verde e azul). A operação matemática segue o mesmo padrão de uma imagem preto e branco, a diferença ocorre no filtro que deve conter os três canais e ao final restará também apenas um número (XU et al., 2016).

A *Lenet* possui uma abordagem voltada para a extração de recursos brutos de uma imagem de entrada usando convoluções, reduzindo a dimensionalidade com camadas de agrupamento. A partir daí, realiza-se um processamento final com camadas densas totalmente conectadas, alcançando dessa forma o reconhecimento dos dados. Somado a isso, devido ao grande número de imagens a serem processadas optou-se por ser uma arquitetura leve e de rápida curva de aprendizado, exigindo pouco poder de processamento.

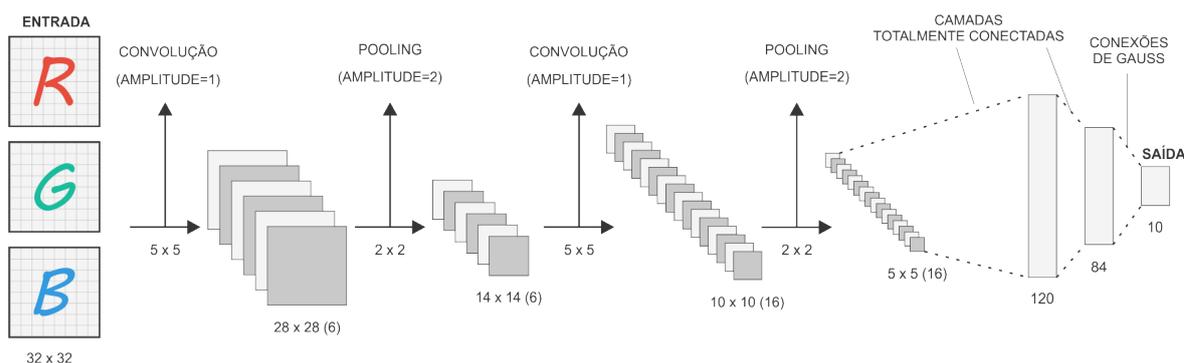


Figura 11 – LeNet com entrada de uma imagem colorida.

O processamento das imagens capturadas foi realizado para a identificação e segmentação das faces e olhos, redimensionando-as para o formato de 72×72 pixels, conforme mostra a Figura 12.

Figura 12 – Fluxo de processamento das imagens.



O fluxo apresentado na Figura 12 se inicia com cada uma das imagens capturadas sendo transformadas por meio de Python OpenCV para escalas de cinza para que possam ser interpretadas como uma matriz 2D, formato necessário para a CNN.

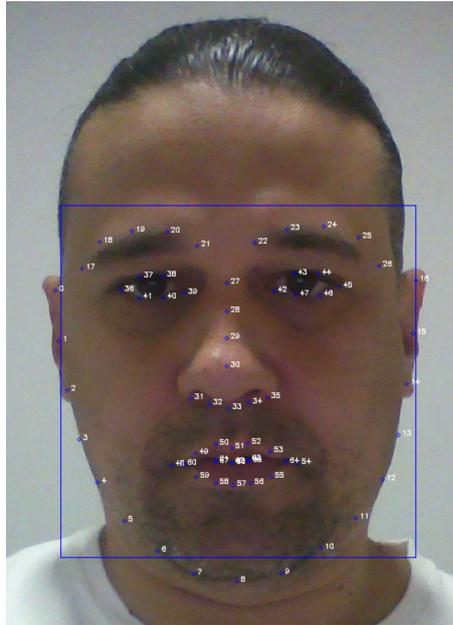
Em seguida, utiliza-se a biblioteca DLIB em Python para realizar a detecção facial. É realizada uma estimativa da localização de 68 coordenadas (x, y) que mapeiam os pontos faciais no rosto de uma pessoa. A Figura 13 demonstra o mapeamento em uma face.

Por meio dessas coordenadas é extraída a área da face, sendo descartadas as fotos sem a face encontrada e garantindo que as fotos restantes sejam segmentadas adequadamente.

A partir das fotos com faces, são extraídas as áreas dos olhos através da biblioteca Python OpenCV, caso não detecte, ocorre o respectivo descarte. Os olhos são justapostos numa imagem de 72×72 pixels conforme mostra saída da Figura 12.

Nos trabalhos apresentados na pesquisa bibliográfica, observou-se o uso de aumento de dados, uma técnica que se utiliza de rotações, translações e inversões das imagens para criar um novo conjunto de dados, isso reduz o *overfitting* e aumenta a capacidade de generalização, incorrendo em previsões mais acertadas em dados não observados a partir de dados observados (PEREZ; WANG, 2017).

Figura 13 – Demonstração das coordenadas em uma face.



Neste trabalho, utilizamos a Keras ImageDatagenerator, da biblioteca TensorFlow, para o aumento dos dados. O código abaixo foi usado para aumentar os dados da base de treinamento.

```
ImageDataGenerator(rescale=1./255)
```

Nesse caso, foram realizados: (rescale) para redimensionar 1./255 e transformar cada valor de pixel do intervalo [0,255] -> [0,1].

3.1.5 Métricas de Avaliação

Para avaliar a CNN adotada, utilizou-se as métricas oriundas da matriz de confusão e exposta na Tabela 2.

Tabela 2 – Matriz de confusão e suas variáveis.

		Valor Predito	
		Sim	Não
Real	Sim	Verdadeiro Positivo (VP)	Falso Negativo (FN)
	Não	Falso Positivo (FP)	Verdadeiro Negativo (VN)

A partir da Tabela 2, pode-se identificar:

- Verdadeiro Positivo (VP): imagens classificadas corretamente e que pertencem à região indicada;
- Falso Positivo (FP): imagens classificadas incorretamente e que não pertencem à região indicada;

- Verdadeiro Negativo (VN): imagens classificadas corretamente e que não pertencem à região indicada;
- Falso Negativo (FN): imagens classificadas corretamente e que não pertencem à região indicada.

No primeiro experimento utilizou-se apenas a acurácia para nortear o nível de acerto da classificação. Já no segundo experimento levou-se em conta as seguintes métricas:

- Acurácia (A): número de acertos

$$A = \frac{VP + VN}{VP + VN + FP + FN} \quad (3.1)$$

- Precisão (P): porcentagem de acertos

$$P = \frac{VP}{VP + FP} \quad (3.2)$$

- *Recall* (R) ou sensibilidade: o quanto o modelo consegue diferenciar as classes

$$R = \frac{VP}{VP + FN} \quad (3.3)$$

- *F1-score*: média harmônica entre precisão e recall

$$F1 - score = 2 * \frac{P * R}{P + R} \quad (3.4)$$

Além das métricas oriundas da matriz de confusão, descritas acima, optou-se pela validação cruzada K-fold, com o intuito de avaliar os resultados de segmentação dos olhos por todo o conjunto de dados.

Utilizou-se $k = 5$, onde a base de treinamento é dividida em 5 partes de mesmo tamanho. Em seguida, são realizadas rodadas de treino e validação, em que uma das cinco partes, ou seja, 20% é utilizada para validação enquanto os outros 80% vão para treino (FIGUEIREDO; OLIVEIRA; SOUTO, 2020). A Tabela 3 demonstra essa execução.

Tabela 3 – Validação cruzada: K-fold, com $k=5$.

	BASE DE TREINAMENTO				
k=1	Teste	Treino			
k=2	Treino	Teste	Treino		
k=3	Treino		Teste	Treino	
k=4	Treino			Teste	Treino
k=5	Treino				Teste

Para Figueiredo, Oliveira e Souto (2020), a vantagem da validação cruzada se encontra no fato de que são utilizadas todas as imagens da base de treinamento, tanto no treino como na validação, trazendo ao final resultados que podem ser demonstrados através de métricas estatísticas como acurácia e desvio padrão aplicadas durante cada rodada de execução.

3.1.6 Monitoramento do uso do navegador

Dentre os trabalhos relacionados realizamos um apanhado das principais técnicas usadas para tentar impedir trapaças na tela do computador.

Verificou-se o uso de controles isolados para monitoramento da tela de prova. A fiscalização das mudanças de tela utilizada trata da mudança de tela e da verificação do uso da entrada USB.

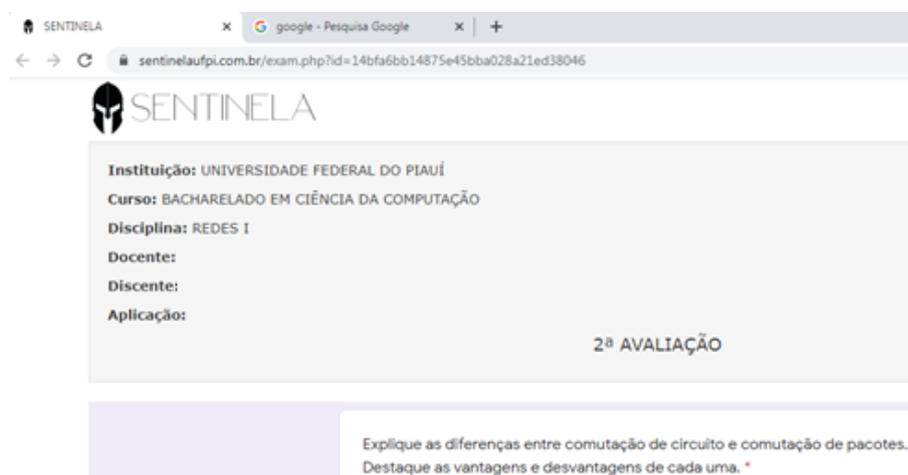
Outra inspeção ocorreu no mouse do aluno, através da checagem das coordenadas dos cliques, porém isso abre margem pra a prática de trapaças por meio do uso apenas do teclado durante a prova.

Um aspecto bem interessante foi o processamentos de entrada de áudio, afinal pode haver repasse de informações por meio de uma segunda pessoa escondida ao lado do aluno. A captura pelo microfone do computador pode ser passível de fraude ao passo que o usuário pode desativar do computador essa funcionalidade.

No protótipo desenvolvido neste trabalho, buscou-se criar um controle eficiente na aba de prova, dessa forma ao sair da tela de prova, seja por clique ou por teclas, ativa-se uma contagem do número de ocorrências e o intervalo de duração entre a saída e retorno à tela de prova. Dessa forma, podemos quantificar essas informações e apresentar ao professor/avaliador para tomada de decisão.

A Figura 14 mostra uma tela de prova e uma segunda aba acessando o *Google*. Nesse caso, ao pressionar ctrl+tab ou clicar na aba do *Google*, será ativada uma nova ocorrência e ao retornar à tela de prova, teremos o tempo em que ficou fora da tela de prova.

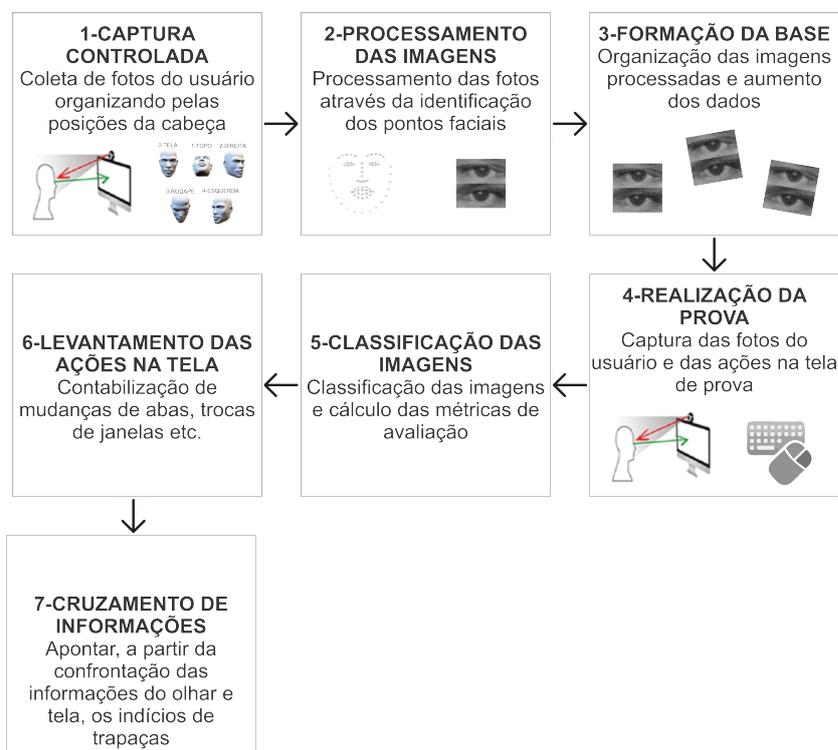
Figura 14 – Tela de prova com mais de uma aba.



3.2 Metodologia Proposta

Neste capítulo, a metodologia proposta neste trabalho de mestrado é descrita em detalhes para validar o comportamento ocular de alunos utilizando computador e webcam sem a necessidade de dispositivo rastreador ocular, somado ao levantamento das ações na tela de prova, proporcionando um cruzamento de informações e indicando a possibilidade de trapaças durante as atividades. A Figura 15 apresenta o conjunto das etapas envolvidas nesta metodologia que foi chamada de Sentinela.

Figura 15 – Etapas da metodologia proposta.



3.2.1 Etapa 1

A primeira etapa trata da captura controlada das fotos dos usuários, onde ocorre a coleta das imagens por meio do protótipo citado na seção anterior. Essas fotos são encaminhadas ao servidor da aplicação considerando as posições de cabeça com olhar para dentro ou fora da tela.

3.2.2 Etapa 2

Na segunda etapa, é realizado o processamento das imagens a partir das posições faciais, alcançando dessa forma a segmentação dos olhos e a sobreposição de cada par em formato 72×72 pixels. As fotos que não se ajustam a esse processamento são descartadas.

3.2.3 Etapa 3

Nesta fase, as imagens processadas são organizadas em 5 regiões e em seguida é realizado o aumento dos dados por meio de reescala. Ao final, temos a base de treinamento que será utilizada nas etapas seguintes para classificação na CNN.

Considerando que tenhamos uma base formada previamente, a execução desta metodologia só se fará necessária a partir da etapa 4 a seguir, que é quando realmente são processadas as novas avaliações.

3.2.4 Etapa 4

Na quarta fase, ocorre a realização da atividade através do protótipo. A captura de fotos acontece de maneira intermitente do momento em que o aluno assinala o início e o término da avaliação. Somado a isso, o monitoramento da tela decorre paralelamente. São armazenadas informações de teclas pressionadas, saídas e retornos à tela de prova, seja por meio de mouse seja apenas pelo teclado.

3.2.5 Etapa 5

Nesta etapa é realizada a classificação das imagens através de CNN, apontando os momentos em que olhou para dentro e para fora da tela de prova. A cada foto enviada ao servidor realizamos também o armazenamento da data e hora em que ela fora processada, assim após a classificação podemos levantar os tempos de cada conjunto de imagens e contabilizar os tempos de olhar para cada uma das regiões.

3.2.6 Etapa 6

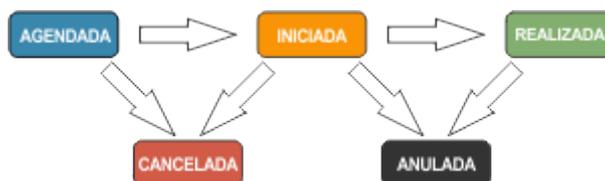
Na sexta etapa o foco é a tela de prova. São levantadas informações como: se saiu da tela de avaliação e quanto tempo despreendeu em cada saída e retorno à tela de prova. Averiguações como cliques realizados e os caracteres pressionados serão importantes na última etapa.

3.2.7 Etapa 7

A etapa final é responsável por compilar as informações das etapas 5 e 6 para confrontar olhar e tela, podendo inferir se num determinado intervalo de tempo estava olhando para fora da tela. Nesse mesmo instante poderão ser cheçadas informações sobre a tela de prova, apontando se estava noutra janela, por exemplo.

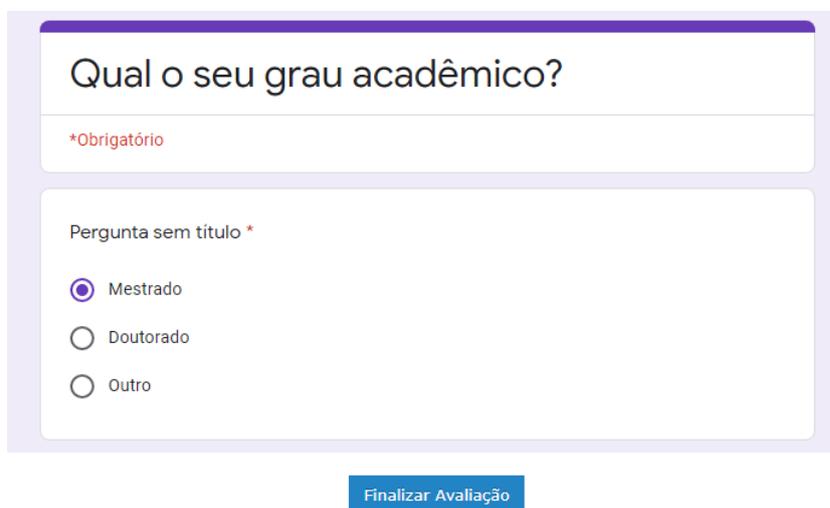
A Figura 16 mostra os status de prova. Por meio do protótipo, o professor pode cancelar uma atividade ou até mesmo anular uma prova, a partir das informações a ele repassadas durante a realização de uma atividade.

Figura 16 – Status de prova do Sentinela.



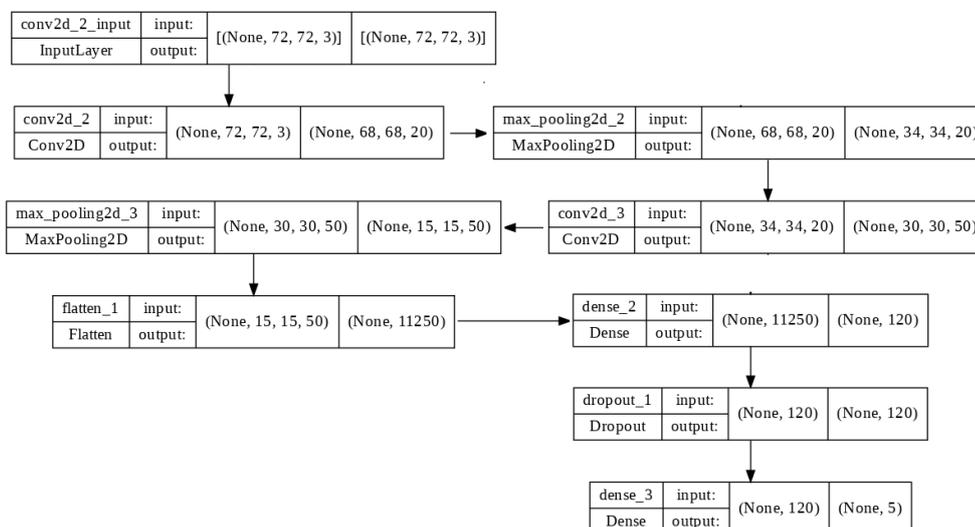
A Figura 17 mostra um caso de atividade em andamento. Neste caso, mesmo que o aluno atualize a página só restará a opção de finalizar a avaliação, fazendo com que o status passe para "realizada". Se durante a prova o avaliador achar pertinente poderá cancelar ou até mesmo anular a avaliação, impedindo que o aluno prossiga com a atividade.

Figura 17 – Tela de prova do protótipo em andamento.



Para classificar as imagens, conforme a etapa 5 desta metodologia, trabalhou-se a CNN conforme a modelagem apresentada na Figura 18.

Figura 18 – Arquitetura LeNet utilizada nesta pesquisa.



A arquitetura utilizada recebe como entrada uma imagem de 72x72 pixels. Inicialmente possui duas camadas convolucionais seguidas de duas camadas de *MaxPooling*. Após as sequências de convolução e *pooling*, foi utilizada uma camada *flatten*, transformando as saídas em um vetor unidimensional. Em prosseguimento, processou-se uma camada densa para conectar os neurônios da camada anterior e inseriu-se uma camada de dropout, visando impedir o *overfitting*. Ao final, uma camada de saída com a quantidade de classes trabalhadas nesta pesquisa.

3.3 Experimentos

Foram realizados dois experimentos. O primeiro apresentou algumas deficiências que foram trabalhadas no segundo. Além disso, buscou-se comprovar a precisão do método de classificação das imagens e balancear corretamente a base a ser treinada.

3.3.1 Primeiro experimento

O primeiro experimento tratou de uma prova real da disciplina de Redes de Computadores do curso de pós-graduação em Ciência da Computação.

A base de dados criada para treinamento foi composta por 5 participantes estudantes de pós-graduação em Ciências da Computação (dois de mestrado e três de doutorado), com média de idade de 32,6 anos, e que fazem parte do grupo de trabalho desta pesquisa.

O processo de obtenção das imagens para a base de treinamento da rede foi realizado através do protótipo, onde o participante foi orientado a olhar por 25 segundos para cada um dos quatro lados fora da tela (topo, direita, rodapé e esquerda) e 100 segundos para a região de dentro da tela, conforme mostra a Figura 5 citada anteriormente.

A cada término de coleta dessas fotos por região, a ferramenta emitia um sinal sonoro indicando o fim da captura. O processo de registro das fotos ocorreu ao passo de 4 fotos por segundo. Ao final da execução, foram registradas 400 fotos do participante olhando para fora da tela e 400 fotos do participante olhando para a tela. Essas fotos foram rotuladas manualmente nas 5 regiões, revisando dessa forma possíveis descartes por não estarem na classe adequada.

Foram realizados os procedimentos ligados ao processamento das fotos, identificando os pontos faciais e realizando as devidas segmentações, alcançando as imagens com os olhos sobrepostos em tons de cinza.

A modelagem da rede assumiu o padrão da *LeNet* proposta pela metodologia. Em seguida, foi compilada, processado o aumento dos dados e por fim o treinamento propriamente dito. A base foi organizada em 60% treino, 20% validação e 20% teste, e

utilizou-se o treinamento em 1000 épocas. Dessa forma, a rede ficou apta para classificar as imagens da prova real e realizar os cálculos de acurácia.

Para realização da prova real foram selecionados 13 estudantes (quatro de mestrado e nove de doutorado). A Figura 19 mostra a composição dos participantes envolvidos neste experimento.

Figura 19 – Participantes do primeiro experimento.

GRUPO	#	IDENTIFICAÇÃO	IDADE	CURSO	PLATAFORMA	SISTEMA OPERACIONAL	NAVEGADOR
BASE DE TREINO	1	#01	39	Mestrado	Desktop	Windows 7	Google Chrome
	2	#02	38	Doutorado	Desktop	Windows 10	Google Chrome
	3	#03	26	Mestrado	Desktop	Windows 10	Google Chrome
	4	#04	31	Doutorado	Desktop	Windows 10	Google Chrome
	5	#05	29	Doutorado	Desktop	Windows 10	Google Chrome
			MÉDIA DE IDADE:	32,60			
PROVA REAL	1	PART. 23	54	Doutorado	Desktop	Windows 10	Google Chrome
	2	PART. 25	48	Doutorado	Desktop	Windows 10	Google Chrome
	3	PART. 27	33	Mestrado	Desktop	Windows 10	Google Chrome
	4	PART. 28	29	Doutorado	Desktop	Mac OS X / Windows 10	Google Chrome
	5	PART. 29	39	Doutorado	Desktop	Windows 10	Google Chrome
	6	PART. 30	26	Doutorado	Desktop	Windows 10	Google Chrome
	7	PART. 31	35	Mestrado	Desktop	Linux / Windows 10	Google Chrome / Mozilla Firefox
	8	PART. 32	29	Doutorado	Desktop	Windows 10 / Windows 7	Opera / Google Chrome
	9	PART. 34	28	Mestrado	Desktop	Windows 10 / Windows 7	Google Chrome
	10	PART. 36	25	Doutorado	Desktop	Windows 10 / Windows 8.1	Google Chrome
	11	PART. 38	24	Mestrado	Desktop	Windows 10	Google Chrome
	12	PART. 39	30	Doutorado	Desktop	Windows 10	Google Chrome
	13	PART. 42	48	Doutorado	Desktop	Windows 10	Google Chrome
		MÉDIA DE IDADE:	34,46				

A prova apresentou 6 questões discursivas para cada aluno e a execução ocorreu através do protótipo. Todos os participantes receberam por email a notificação com o link da prova e a avaliação transcorreu com tranquilidade, sem imprevistos. Antes do início do exame, todos os estudantes foram elucidados quanto aos trâmites de realização.

As imagens dos alunos foram registradas no servidor e preliminarmente, foram separadas aleatoriamente apenas 1000 fotos de cada participante para que fossem rotuladas manualmente nas 5 regiões. Em seguida, foram classificadas pela CNN e alcançou-se a acurácia individual de cada estudante. As informações de uso da tela de prova também foram armazenadas para contabilização e posterior confrontação.

Ao final do processo, foram reunidas informações de tela (teclas digitadas e trocas de tela) e do olhar dos alunos importantes para detectar algum tipo de trapaça.

3.3.2 Segundo experimento

O segundo experimento tratou de um conjunto de 3 provas envolvendo questões discursivas das camadas do modelo OSI.

A base de dados criada para treinamento foi composta por 6 participantes estudantes de graduação e pós-graduação em Ciências da Computação (três de mestrado e três de graduação), com média de idade de 32,2 anos. Todos os participantes utilizaram o mesmo computador com as especificações na Tabela 4.

Tabela 4 – Especificação do computador para coleta do segundo experimento.

TIPO	DESCRIÇÃO
Hardware	<p>Notebook DELL Inspiron 3421 com as seguintes especificações:</p> <ul style="list-style-type: none"> - Processador: Intel i3 1.80 GHz - RAM: 8 GB DDR3 - VGA: Intel HD Graphics 4000 - Armazenamento: SSD 200 GB - Sistema operacional: Windows 10 Home - Webcam: 0,92 MP (resolução máxima 1280x720)
Software	<ul style="list-style-type: none"> - Apache Netbeans IDE - Google Colab, Python, Keras e TensorFlow

Para a formação do nosso conjunto de dados, utilizamos um dispositivo comercial voltado para o rastreador ocular, evitando dessa forma a rotulação manual ocorrida no primeiro experimento e atestando os resultados de alta acurácia esperados.

O rastreador foi posicionado na parte frontal da tela e por meio da reflexão de luzes nos olhos do usuário permitiu estimar o olhar.

Para manusear o rastreador, utilizou-se uma aplicação desenvolvida em Unity que aciona as funções disponíveis por meio da API do rastreador. Um protótipo foi criado e executado em cima dessa aplicação que emula um navegador. Dessa forma, o protótipo tratou de alertar através de sinal sonoro a mudança de região do olhar e o navegador ficou responsável de coletar as coordenadas X e Y do olhar.

A cada 0,16 milissegundo ocorreu a anotação das coordenadas e a cada 1,5 segundos aconteceu o registro da foto associada a coordenadas. O protótipo tratou de limitar a 3 minutos o tempo total de cada região (dentro e fora da tela), fazendo com que cada execução durasse 15 minutos.

Foram realizados os procedimentos ligados ao processamento das fotos, identificando os pontos faciais e realizando as devidas segmentações, alcançando as imagens com os olhos sobrepostos em tons de cinza.

A modelagem da rede assumiu o padrão da *LeNet* proposta pela metodologia. Em seguida, foi compilada, processado o aumento dos dados e por fim o treinamento propriamente dito. A base foi organizada em 60% treino, 20% validação e 20% teste, e utilizou-se o treinamento em 200 épocas. Dessa forma, a rede ficou apta para classificar as imagens da prova controlada e realizar os cálculos de acurácia.

A prova controlada foi planejada da seguinte forma:

- Uma prova com 2 questões em que o participante deveria realizar sem nenhuma trapaça;
- Uma prova com 2 questões em que o participante utiliza um celular posicionado abaixo da mesa de prova para realizar consultas indevidas;
- Uma prova com 2 questões em que o participante utiliza um segundo monitor posicionado à direita da tela de prova para realizar consultas indevidas.

As provas foram aplicadas a um mesmo participante e seguiram para classificação pela CNN alcançando a acurácia individual por prova. As informações de uso da tela de prova também foram armazenadas para contabilização e posterior confrontação.

Ao final do processo, foram reunidas informações de tela (teclas digitadas e trocas de tela) e do olhar dos alunos importantes para detectar algum tipo de trapaça.

4 Resultados e Discussão

Neste capítulo serão expostos e discutidos os resultados obtidos a partir dos experimentos realizados, detalhando possíveis indícios dos resultados que possam validar a continuidade do presente trabalho.

4.1 Resultados do primeiro experimento

No primeiro experimento, que trata de uma prova real, atingimos um total de 141.047 fotos, imagens estas que foram extraídas durante o intervalo de início e término de cada avaliação individual dos alunos. A Tabela 5 apresenta esses totais.

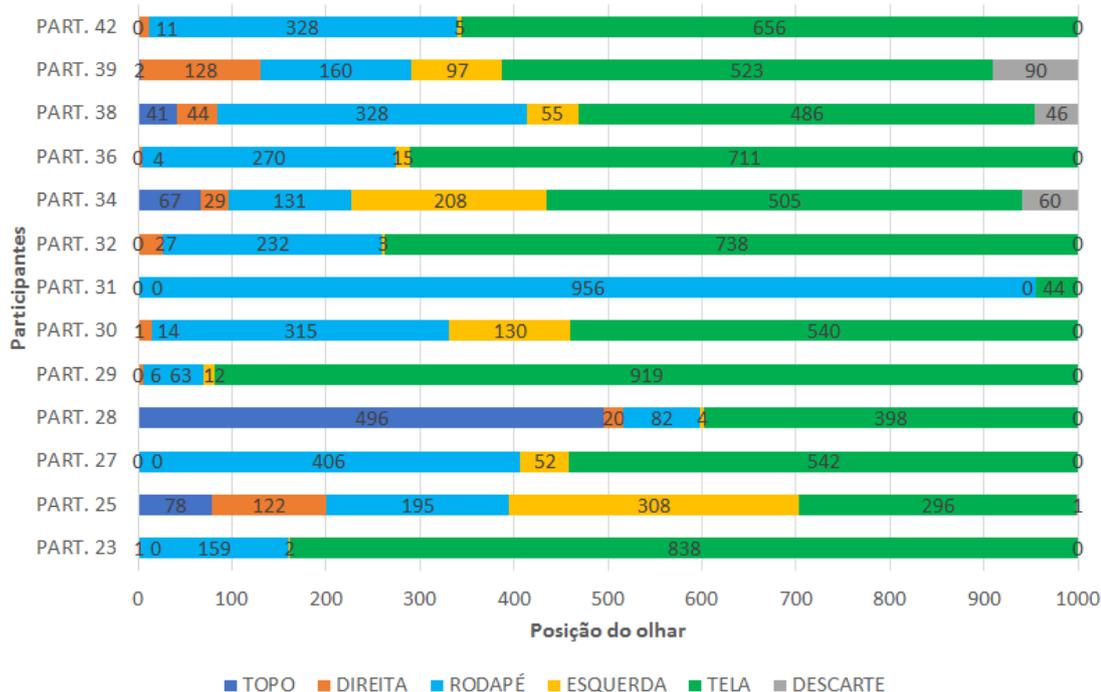
Tabela 5 – Total de fotos dos participantes do primeiro experimento.

Aluno	Núm. de fotos
PART. 23	11.197
PART. 25	16.146
PART. 27	11.084
PART. 28	6.386
PART. 29	5.496
PART. 30	10.477
PART. 31	10.799
PART. 32	9.825
PART. 34	11.740
PART. 36	11.162
PART. 38	12.972
PART. 39	9.618
PART. 42	14.145
	141.047

O primeiro resultado relacionado, é apresentado na Figura 20. O gráfico mostra a posição do olhar do participante em relação aos pontos na tela. Para tanto, 1.000 imagens faciais dos participantes foram obtidas de forma aleatória, e em seguida rotuladas manualmente, indicando para cada imagem, se o usuário estava olhando para dentro ou fora da tela do computador. Para cada foto mantivemos os registros de histórico, dessa forma saberíamos qual o exato momento em que foi registrada.

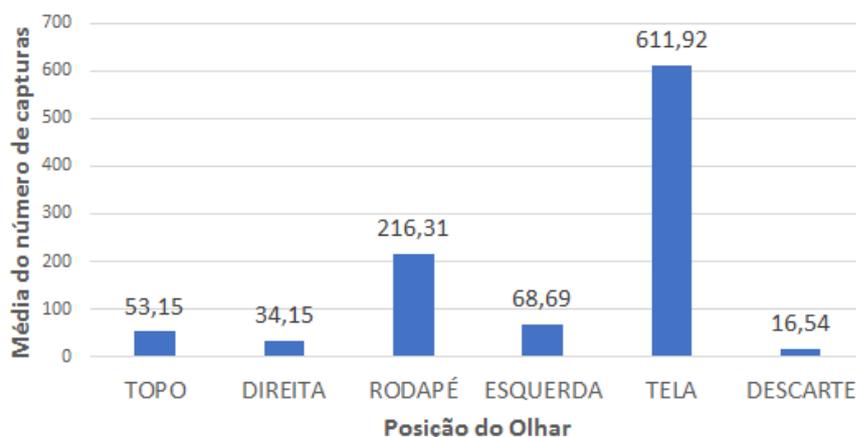
Conforme mostra a Figura 20, foram descartadas fotos que não continham imagens dos participantes, além de fotos desfocadas, por conta do rápido movimento da cabeça durante a captura. Podemos concluir que de maneira geral os alunos passaram mais tempo olhando para a tela de avaliação.

Figura 20 – Posição do olhar de cada participante durante a prova.



A Figura 21 evidencia o fato dos candidatos passarem mais tempo olhando para a tela. Em média, a cada 1000 fotos por candidato, cerca de 612 foram registros de imagens olhando para a tela. Outro ponto relevante é o baixo índice de descarte, demonstrando que a rotina de 4 fotos por segundo é capaz de realizar o armazenamento da foto com uma resolução em que a CNN tem potencial de realizar a tomada de decisão.

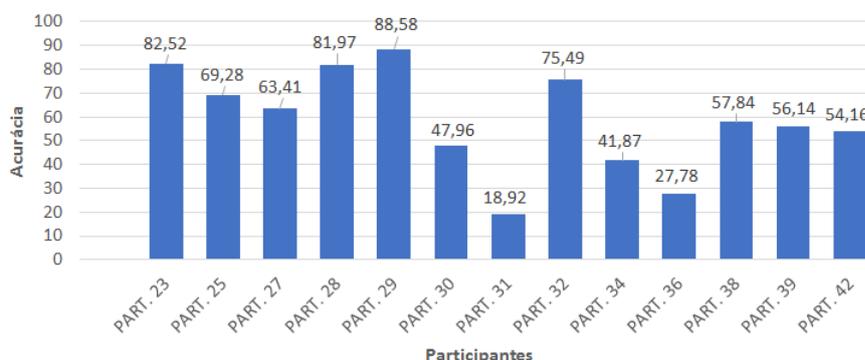
Figura 21 – Média do número de capturas por posição do olhar do participante.



A Figura 22 apresenta o resultado da classificação das imagens de cada estudante utilizando a CNN treinada. Nesse caso, percebe-se que o maior índice de acerto obtido é o do aluno *PART.29*, que é superior a 88%. Observa-se também alguns casos abaixo de 50%.

Essa baixa taxa de acerto (por volta de 20%) em dois participantes da Figura 22,

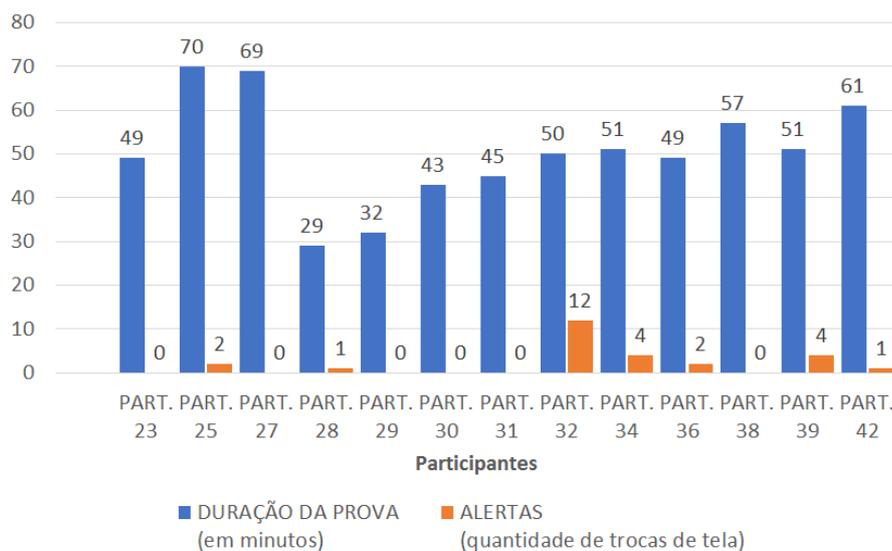
Figura 22 – Acurácia por participante.



serviu de motivação para que pudéssemos aperfeiçoar o processo de rotulação por meio do uso de um dispositivo comercial do tipo rastreador ocular para rotular as imagens da base de treinamento, e dessa forma tentar alcançar um experimento mais controlado onde a acurácia do método fosse melhorada.

Durante o processo de acompanhamento da prova pelo protótipo, quando ocorre alguma troca da tela de avaliação por uma outra, ou seja, o usuário acessa outra aba de navegação ou mesmo entra numa outra janela do computador, o sistema inicia um processo que chamamos de **Alerta**, no qual é feito o registro da data e hora de saída da tela de prova e a data e hora de retorno à essa tela. A cada saída e retorno, o sistema gera uma unidade de alerta e totaliza o tempo de duração. A Figura 23 apresenta a totalização dos alertas comparada ao tempo total de prova.

Figura 23 – Duração duração de prova x quantidade de alertas.

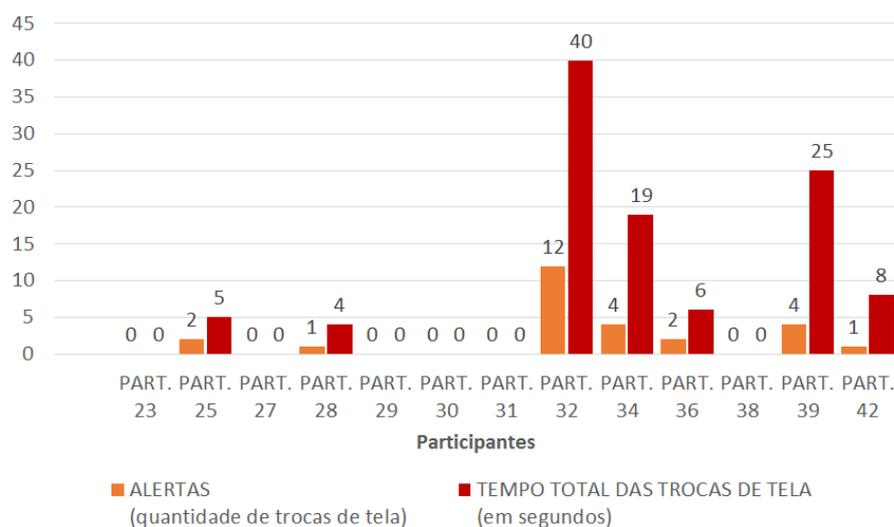


Nessa confrontação entre os dados de tempo de realização da prova e a quantidade de alertas na Figura 23, observa-se o caso do *PART.32*, no qual atingiu um total de 12 alertas num intervalo de 50 minutos de prova. Isso pode sugerir um comportamento

inoportuno no qual o aluno saiu e entrou da plataforma várias vezes podendo dessa forma ter feito algum tipo de consulta não autorizada.

Na Figura 24, podemos detalhar melhor o tempo em que cada um dos participantes levou durante as trocas de tela. A cada saída e retorno da tela era gerado um tempo que foi totalizado no gráfico abaixo.

Figura 24 – Quantidade de alertas x Duração total dos alertas.



Diante dos 13 participantes que realizaram a avaliação, conforme a Figura 24, podemos enfatizar a baixa proporção de tempo em que o aluno realizou a troca de janela ativa do navegador. Apresentamos como destaque novamente apenas o aluno *PART.32* que desprende 40 segundos de tempo total fora da tela de prova. Isso sugere que os alunos de maneira geral optaram por não tentar algo ilícito do que tange ao uso de outras janelas.

Com relação ao processo de rastreamento da tela, podemos ampliar o processo de acompanhamento, envolvendo controles no mouse e teclado dos alunos, além da captura de áudio. Dessa forma, podemos aumentar as informações que podem levar ao avaliador a confirmar algum tipo de trapaça.

Outra alternativa é a possibilidade do uso de tecnologias ligadas ao *Google Forms*, como o sistema de antiplágio, onde poderemos identificar se trechos das respostas dos alunos são parte de textos copiados da web.

4.2 Resultados do segundo experimento

Houveram algumas deficiências no primeiro experimento como: a rotulação manual da base de dados, o que pode ter ocasionado a índices baixos de acurácia em algumas avaliações. Além disso, não levamos em consideração a checagem de outras métricas, como por exemplo, a precisão. Por fim, realizamos uma validação cruzada dando ainda mais

embasamento à metodologia proposta. Com isso, realizamos um segundo experimento destacando situações específicas de trapaça para ressaltar o comportamento da metodologia frente a essas situações.

Neste experimento atingimos um total de 3.711 fotos, imagens estas que foram extraídas durante o intervalo de início e término de cada avaliação individual dos alunos. A Tabela 6 apresenta esses totais.

Tabela 6 – Total de fotos dos participantes do segundo experimento.

ATIVIDADE	ALUNO	NÚM. DE FOTOS
BASE DE TREINAMENTO	PART. 01	550
	PART. 02	585
	PART. 03	415
	PART. 04	595
	PART. 05	580
	PART. 06	590
PROVA 1	PART. 07	85
PROVA 2	PART. 07	149
PROVA 3	PART. 07	162
		3.711

A Tabela 7 demonstra percentuais acima de 99%, indicando que o modelo de CNN utilizado apresentou números elevados para as métricas verificadas.

Tabela 7 – Resultados do modelo da rede do segundo experimento.

CNN	A (%)	P (%)	R (%)	F1-score (%)
<i>LeNet utilizada no Sentinela</i>	99,60	99,61	99,60	99,60

A=Acurácia, P=Precisão, R=Recall

Com relação aos resultados alcançados através da validação cruzada, a Tabela 8 apresenta os indicadores a cada rodada de treinamento.

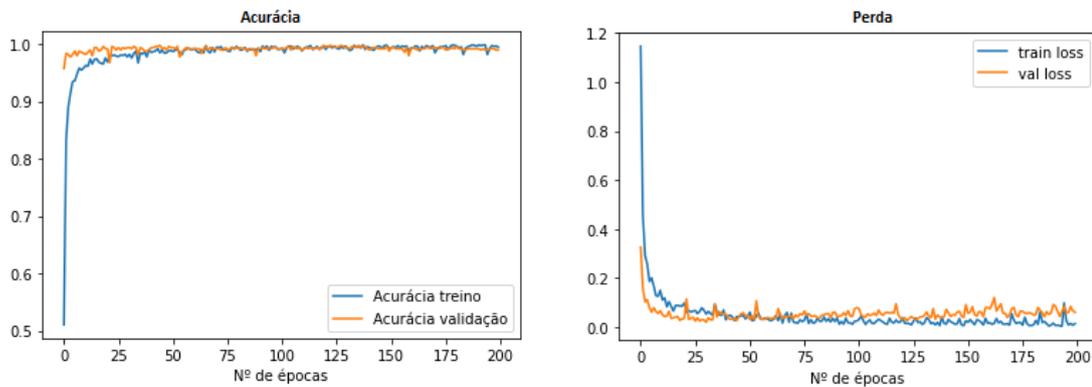
Tabela 8 – Resultados da Validação Cruzada.

Rodadas de treinamento	Acurácia	Desvio Padrão
k=1	98,82%	0,06%
k=2	98,82%	0,06%
k=3	98,75%	0,09%
k=4	98,75%	0,09%
k=5	98,62%	0,22%

A média de acurácia alcançada durante a validação cruzada foi de 98,75%, onde o pior caso foi de 98,62% com desvio padrão de 0,22%, demonstrando um desempenho bastante satisfatório.

Conforme mencionou-se anteriormente, o modelo utilizado interrompeu o treino em 200 épocas, plotando os gráficos de acurácia e perda, de acordo com a Figura 25. Observa-se um alto percentual de acurácia e um baixo índice de perda.

Figura 25 – Gráficos de acurácia e perda no modelo de treino e teste.



Com a rede treinada e apta a classificar as imagens de provas, executou-se a classificação das avaliações e os resultados são apresentados na Tabela 9.

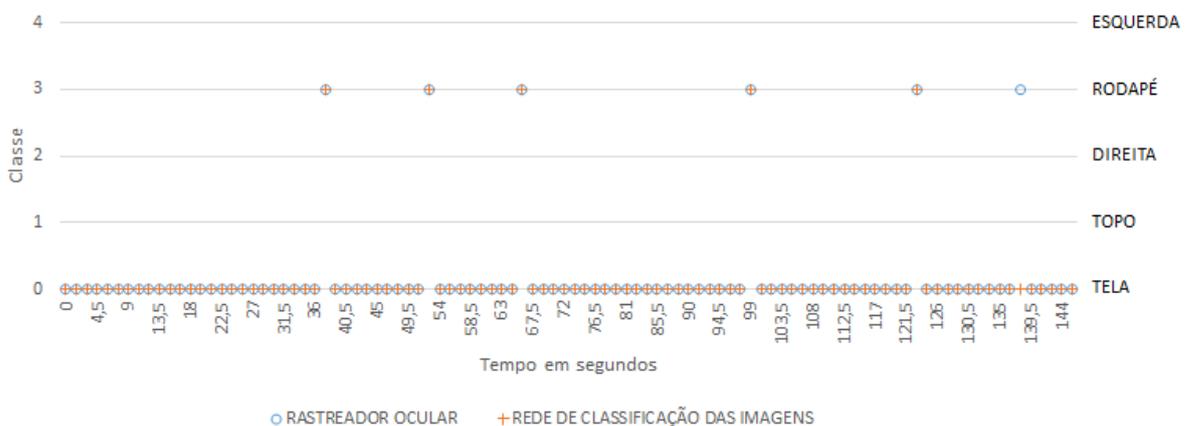
Tabela 9 – Resultados da classificação pela rede para o segundo experimento.

ATIVIDADE	A (%)	P (%)	R (%)	F1-score (%)
PROVA 1	98,89	98,90	98,89	98,85
PROVA 2	97,40	97,47	97,40	97,42
PROVA 3	97,70	99,23	97,70	98,22

A=Acurácia, P=Precisão, R=Recall

A Figura 26 mostra uma confrontação entre a catalogação do rastreador ocular utilizado e a classificação realizada pela rede para a Prova 1.

Figura 26 – Confrontação na Prova 1: rastreador ocular X classificação pela CNN.

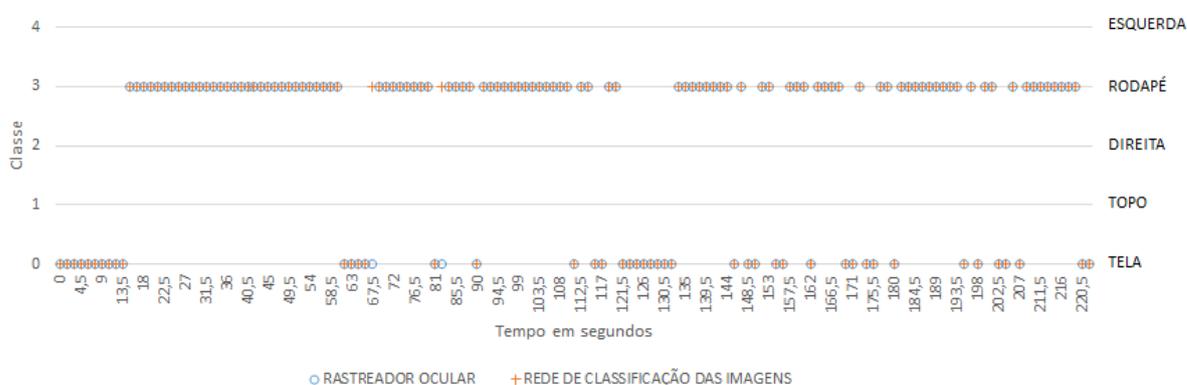


Verifica-se que na posição do tempo 138 da Figura 26, o rastreador indicou o olhar para a tela, enquanto que a classificação da rede inferiu como olhando para o rodapé.

A prova durou 149 segundos e o aluno se restringiu a olhar somente para a tela e o rodapé. Por se tratar de uma prova subjetiva espera-se que sejam os momentos em que ele olhou para o teclado para realização a digitação. Essa averiguação é feita através das informações coletadas pelo protótipo. Nessa caso, foram digitados 227 caracteres e não foram registrados alertas (ações de troca de aba ou tela) durante a realização da atividade, indicando que não houveram trapaças.

A confrontação na Prova 2 é exposta na Figura 27 e indica que a avaliação durou 224 segundos, e aos 81 segundos, o rastreador sugeriu o olhar para a tela, enquanto que a classificação da rede apontou como olhando para o rodapé.

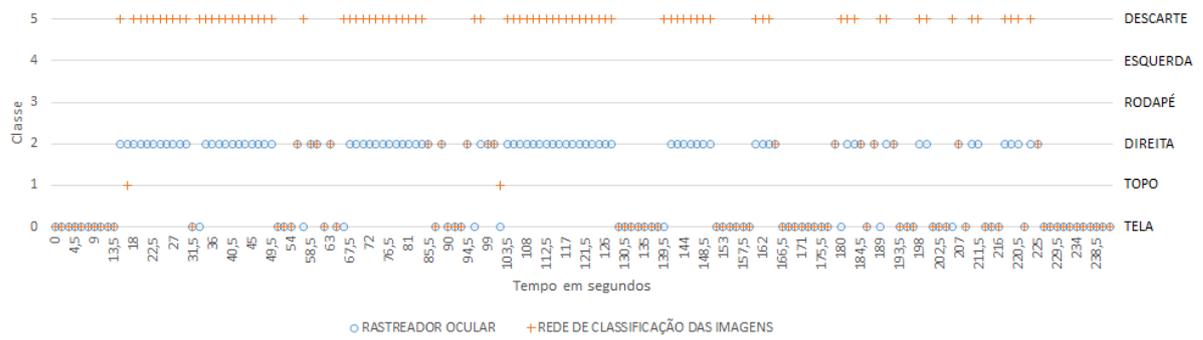
Figura 27 – Confrontação na Prova 2: rastreador ocular X classificação pela CNN.



A Figura 27 demonstra um grande volume de marcações do olhar do aluno para o rodapé. Surge nesse momento a necessidade de uma averiguação por parte do protótipo para atestar a possibilidade de trapaça. A solução apontou que foram digitados apenas 225 caracteres e que não houveram alertas. Ocorre que o aluno despreendeu muito tempo olhando para baixo sem interagir com o computador, indicando que ele estava se utilizando de algum dispositivo externo à prova. Nesse caso, foi um celular, conforme planejado para a execução da atividade.

Já a prova 3, durou 248 segundos e o rastreador indicou dois grandes volumes de informações do olhar: para a tela e para a direita. A parte relacionada à tela, foi em sua maioria confirmada pela classificação da rede, conforme mostra a Figura 28. Porém, a parte relacionada ao olhar para direita, foi na grande maioria descartada, indicando que não foi possível detectar os olhos para segmentação / classificação.

Figura 28 – Confrontação na Prova 3: rastreador ocular X classificação pela CNN.



Para demonstrar se o aluno, ao olhar para a direita, estava praticando de alguma trapaça na prova 3, foram recolhidas informações de tela, onde detectou-se que o aluno dos 248 segundos de prova, ele sofreu 3 alertas representando um total de 84 segundos fora da tela de prova. Isso atesta que o aluno estava praticando de trapaça por meio de um segundo monitor posicionado na parte direita da tela.

5 Conclusões e Trabalhos Futuros

Retomando os objetivos específicos, criou-se um protótipo para monitorar e coletar informações das provas on-line, que foi aplicado a dois experimentos, incluindo uma avaliação real.

Foi criada uma base de dados por meio da combinação de fotos captadas através de um protótipo e das informações registradas por um dispositivo comercial de rastreamento ocular. Essas imagens foram processadas e classificadas através de CNN, apresentando acurácia e precisão excelentes. Ademais, utilizou-se das informações de tela da prova dos alunos para auxiliar na tomada de decisão quanto à checagem da integridade e lisura das avaliações realizadas.

Diante do protótipo elaborado, foi possível inferir se o aluno estava olhando para dentro ou fora da tela durante a realização das atividades. Os movimentos de cabeça e as ações do usuário na tela foram monitorados por intermédio de um computador com *webcam* usando apenas um navegador e sem a necessidade de softwares adicionais. Portanto, conclui-se que a metodologia proposta auxilia na integridade e lisura do processo de avaliação remota. Dessa forma, alcançamos o objetivo geral que consiste em uma metodologia eficaz no controle e fiscalização da aplicação de exames on-line.

5.1 Trabalhos Futuros

Diante dos resultados apresentados neste trabalho, acredita-se que a metodologia proposta possa evoluir para uma ferramenta a ser utilizada na educação, possibilitando que os agentes da educação possam ter autonomia na confecção das provas a serem aplicadas remotamente.

Deve-se investigar outras técnicas de processamento de imagens, bem como a utilização de redes neurais mais modernas.

Outro aspecto que pode ser trabalhado é a implementação de uma etapa de calibração ou algum processo para obtenção de modelo ajustado a partir das imagens do estudante avaliado. Isso irá garantir que não tenhamos grande número de fotos descartadas.

Estudar a utilização de técnicas forenses no processo de armazenamentos das informações, garantindo que os artefatos funcionem como evidências dentro de um processo similar ao de um inquérito.

Finalmente, realizar estudos mais aprofundados da *API* do *Google Forms* visando uma integração eficiente, dando uma opção mais ampla aos avaliadores das atividades por

meio de funcionalidades antiplágio que a ferramenta apresenta.

Referências

- ASEP, H. S.; BANDUNG, Y. A design of continuous user verification for online exam proctoring on m-learning. In: IEEE. *2019 International Conference on Electrical Engineering and Informatics (ICEEI)*. [S.l.], 2019. p. 284–289. Citado 2 vezes nas páginas 5 e 7.
- BACICH, L.; NETO, A. T.; TREVISANI, F. de M. *Ensino híbrido: personalização e tecnologia na educação*. [S.l.]: Penso Editora, 2015. Citado na página 1.
- FIGUEIREDO, A. F.; OLIVEIRA, H. S. de; SOUTO, E. J. P. Redes neurais densas para classificação de estresse. *Journal of Health Informatics*, v. 12, 2020. Citado na página 17.
- HU, S.; JIA, X.; FU, Y. Research on abnormal behavior detection of online examination based on image information. In: IEEE. *2018 10th International Conference on Intelligent Human-Machine Systems and Cybernetics (IHMSC)*. [S.l.], 2018. v. 2, p. 88–91. Citado 2 vezes nas páginas 5 e 7.
- INDI, C. S. et al. Detection of malpractice in e-exams by head pose and gaze estimation. *International Journal of Emerging Technologies in Learning*, v. 16, n. 8, 2021. Citado 3 vezes nas páginas 1, 6 e 7.
- JOYE, C. R.; MOREIRA, M. M.; ROCHA, S. S. D. Educação a distância ou atividade educacional remota emergencial: em busca do elo perdido da educação escolar em tempos de covid-19. *Research, Society and Development*, v. 9, n. 7, p. e521974299–e521974299, 2020. Citado na página 1.
- LAI, M.-L. et al. A review of using eye-tracking technology in exploring learning from 2000 to 2012. *Educational research review*, Elsevier, v. 10, p. 90–115, 2013. Citado na página 10.
- LECUN, Y. et al. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, Ieee, v. 86, n. 11, p. 2278–2324, 1998. Citado 2 vezes nas páginas 13 e 14.
- LI, H. et al. A visual analytics approach to facilitate the proctoring of online exams. In: *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. [S.l.: s.n.], 2021. p. 1–17. Citado 2 vezes nas páginas 6 e 7.
- MUZAFFAR, A. W. et al. A systematic review of online exams solutions in e-learning: Techniques, tools, and global adoption. *IEEE Access*, IEEE, v. 9, p. 32689–32712, 2021. Citado na página 5.
- PEREZ, L.; WANG, J. The effectiveness of data augmentation in image classification using deep learning. *arXiv preprint arXiv:1712.04621*, 2017. Citado na página 15.
- POOLE, A.; BALL, L. J. Eye tracking in hci and usability research. In: *Encyclopedia of human computer interaction*. [S.l.]: IGI global, 2006. p. 211–219. Citado na página 10.
- PRATHISH, S.; BIJLANI, K. et al. An intelligent system for online exam monitoring. In: IEEE. *2016 International Conference on Information Science (ICIS)*. [S.l.], 2016. p. 138–143. Citado 3 vezes nas páginas 2, 5 e 7.

- RANZATO, M. et al. Sparse feature learning for deep belief networks. *Advances in neural information processing systems*, Vancouver, v. 20, p. 1185–1192, 2007. Citado na página 14.
- SANDHYA, S. et al. Adoption of google forms for enhancing collaborative stakeholder engagement in higher education. *Journal of Engineering Education Transformations*, v. 33, p. 283–289, 2020. Citado na página 6.
- VIEIRA, M.; SILVA, C. A educação no contexto da pandemia de covid-19: uma revisão sistemática de literatura. *Revista Brasileira de Informática na Educação*, v. 28, n. 0, p. 1013–1031, 2020. ISSN 2317-6121. Disponível em: <<http://br-ie.org/pub/index.php/rbie/article/view/v28p1013>>. Citado na página 1.
- XU, X. et al. Convolutional neural network for 3d object recognition using volumetric representation. In: *2016 First International Workshop on Sensing, Processing and Learning for Intelligent Machines (SPLINE)*. [S.l.: s.n.], 2016. p. 1–5. Citado na página 14.