



UNIVERSIDADE FEDERAL DO TOCANTINS
CÂMPUS DE PALMAS
PROGRAMA DE PÓS-GRADUAÇÃO EM GOVERNANÇA E
TRANSFORMAÇÃO DIGITAL

Starley do Nascimento Lobo

Plataforma de Monitoramento Comportamental Baseado em Visão
Computacional para Apoio ao Acompanhamento Pedagógico no Ensino Básico

Palmas/TO
2026

Starley do Nascimento Lobo

Plataforma de Monitoramento Comportamental Baseado em Visão
Computacional para Apoio ao Acompanhamento Pedagógico no Ensino Básico

Dissertação apresentada ao Programa de Pós-Graduação
em Governança e Transformação Digital. Foi avaliada
para obtenção do título de Mestre em PPGTD em sua
forma final pelo orientador e pela Banca Examinadora.

Orientador: Dr. Humberto Xavier de Araújo

Palmas/TO
2026

Dados Internacionais de Catalogação na Publicação (CIP)
Sistema de Bibliotecas da Universidade Federal do Tocantins

D676p do Nascimento Lobo, Starley.

Plataforma de Monitoramento Comportamental Baseado em Visão Computacional para Apoio ao Acompanhamento Pedagógico no Ensino Básico. / Starley do Nascimento Lobo. – Palmas, TO, 2026.

75 f.

Dissertação (Mestrado Profissional) - Universidade Federal do Tocantins – Câmpus Universitário de Palmas - Curso de Pós-Graduação (Mestrado Profissional) em Governança e Transformação Digital - PPGTD, 2026.

Orientador: Humberto Xavier de Araújo

1. Inteligência Artificial. 2. Visão Computacional. 3. Monitoramento Comportamental. 4. Análise Pedagógica. I. Título

CDD 004

TODOS OS DIREITOS RESERVADOS – A reprodução total ou parcial, de qualquer forma ou por qualquer meio deste documento é autorizado desde que citada a fonte. A violação dos direitos do autor (Lei nº 9.610/98) é crime estabelecido pelo artigo 184 do Código Penal.

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

FOLHA DE APROVAÇÃO

Starley do Nascimento Lobo

Plataforma de Monitoramento Comportamental Baseado em Visão Computacional para Apoio ao Acompanhamento Pedagógico no Ensino Básico

Dissertação apresentada ao Programa de Pós-Graduação em Governança e Transformação Digital da Universidade Federal do Tocantins. Foi avaliada para obtenção do título de Mestre em Governança e Transformação Digital e aprovada em sua forma final pelo orientador e pela Banca Examinadora.

Orientador: Dr. Humberto Xavier de Araújo

Data de aprovação: 13 / 05 / 2026

Banca Examinadora

Prof. Dr. Humberto Xavier de Araújo, UFT

Prof. Dr. David Nadler Prata, UFT

Prof. Dr. Luciano Vieira de Araújo, UFT

Palmas/TO, 2026

AGRADECIMENTOS

Agradeço, primeiramente, a Deus, por ter me concedido saúde, força, sabedoria e perseverança para superar os desafios ao longo desta caminhada. Sem Sua presença e direção, esta conquista não seria possível. Aos meus pais, Valdeci e Dorivalda, minha eterna gratidão pelos ensinamentos, pelo exemplo de honestidade, trabalho e dedicação à família. Tudo o que sou hoje tem como base os valores que recebi de vocês. À minha esposa, Damiana, pelo amor, companheirismo, compreensão e incentivo durante todo este percurso. Às minhas filhas, Alice e Mariana, que foram fonte constante de inspiração, alegria e motivação para que eu nunca desistisse dos meus objetivos. Aos meus irmãos, Ranyere e Starlyne, pelo apoio, amizade e por compartilharem comigo tantos momentos importantes desta jornada. Às minhas sobrinhas, Geovanna, Isis e Liz, que, com sua alegria, sempre trouxeram leveza aos dias mais difíceis. Expresso também minha sincera gratidão ao meu orientador, Professor Dr. Humberto Xavier de Araújo, pela confiança, orientação, disponibilidade e valiosas contribuições ao longo do desenvolvimento desta pesquisa. Seus conhecimentos e incentivo foram fundamentais para a conclusão deste trabalho. Por fim, agradeço a todos que, direta ou indiretamente, contribuíram para a realização deste sonho. Esta conquista não é apenas minha, mas de todos aqueles que caminharam ao meu lado e acreditaram que este momento seria possível.

RESUMO

Este trabalho apresenta o desenvolvimento de uma plataforma de monitoramento comportamental em sala de aula baseada em técnicas de Inteligência Artificial e Visão Computacional. A solução integra reconhecimento facial, por meio da biblioteca InsightFace, e análise postural utilizando o modelo YOLOv11-Pose, permitindo a identificação automática de comportamentos observáveis dos alunos em tempo real.

A plataforma é capaz de classificar padrões comportamentais relevantes ao contexto educacional, como atenção, distração, participação, agitação e sonolência, a partir da análise de imagens capturadas em ambiente de sala de aula. Os dados coletados são processados e organizados em um banco de dados estruturado, possibilitando o acompanhamento ao longo do tempo e a visualização por meio de uma interface interativa desenvolvida em Streamlit.

A proposta tem como objetivo oferecer suporte à análise pedagógica, auxiliando professores na identificação de padrões comportamentais e na compreensão da dinâmica da turma durante as aulas. O sistema foi concebido como uma prova de conceito funcional, não tendo como finalidade a realização de diagnósticos clínicos, mas sim o fornecimento de subsídios objetivos para apoio à prática educacional.

Os resultados obtidos demonstram a viabilidade técnica da solução e seu potencial como ferramenta de apoio à tomada de decisão pedagógica, respeitando os limites éticos e interpretativos do uso de tecnologias de visão computacional no ambiente escolar.

Palavras-chaves: Inteligência Artificial; Visão Computacional; Monitoramento Comportamental; Análise Pedagógica; Educação Básica.

ABSTRACT

This work presents the development of a classroom behavioral monitoring platform based on Artificial Intelligence and Computer Vision techniques. The proposed solution integrates facial recognition using the InsightFace library and posture analysis through the YOLOv11-Pose model, enabling the automatic identification of observable student behaviors in real time.

The platform is capable of classifying behavioral patterns relevant to the educational context, such as attention, distraction, participation, agitation, and drowsiness, based on the analysis of images captured in the classroom environment. The collected data are processed and organized in a structured database, allowing longitudinal monitoring and visualization through an interactive interface developed in Streamlit.

The main objective of this study is to support pedagogical analysis by assisting teachers in identifying behavioral patterns and understanding classroom dynamics during lessons. The system is designed as a functional proof of concept and does not aim to perform clinical diagnoses, but rather to provide objective insights to support educational practice.

The results demonstrate the technical feasibility of the proposed solution and its potential as a tool to support pedagogical decision-making, while respecting the ethical and interpretative limitations associated with the use of computer vision technologies in educational environments.

Keywords: Artificial Intelligence; Computer Vision; Behavioral Monitoring; Pedagogical Analysis; Basic Education.

LISTA DE ILUSTRAÇÕES

Figura 1 - Estratégias para desenvolver o máximo das capacidades estudantes ao longo dos seus marcos de desenvolvimento. Fonte: (BELLO, MACHADO; CAPELLINE, 2022).	14
Figura 2 - Representação dos marcos do desenvolvimento infantil, destacando a evolução das capacidades motoras, cognitivas, linguísticas e socioemocionais ao longo da infância. Fonte: (BELLO, MACHADO; CAPELLINE, 2022).	15
Figura 3 - Domínios fundamentais do desenvolvimento infantil, incluindo aspectos motores, cognitivos, socioemocionais e da linguagem, utilizados como referência para identificação de progressos e possíveis atrasos. Fonte: (BELLO, MACHADO; CAPELLINE, 2022).	16
Figura 4 - Fluxo de Visão Computacional para Detecção de Comportamento em Sala de Aula. (YUVARAJ; PRINCE; MURUGAPPAN, 2024)	17
Figura 5 - Arquitetura híbrida baseada em 2D CNN e modelos recorrentes para classificação de ações estudantis. Fonte: Adaptado de Yuvaraj, Prince e Murugappan (2024).	18
Figura 6 - Arquitetura do sistema de detecção de ações, emoções e reconhecimento facial para análise de atenção estudantil. Fonte: Adaptado de Singh, Gururani e Bhatnagar (2023).	20
Figura 7 - Filtro de Imagem – Fonte: (Autor, 2026).	21
Figura 8 - Conjunto de <i>keypoints</i> corporais utilizados em modelos de estimativa de pose humana. Fonte: Adaptado de Dai, Xie e Yu (2025).	22
Figura 9 - Processo de reconhecimento facial: detecção, alinhamento do rosto e geração do <i>embedding</i> vetorial. Fonte: Adaptado de Schroff, Kalenichenko e Philbin (2015).	23
Figura 10 - Processo de conversão da imagem facial em <i>embedding</i> vetorial e armazenamento em banco de dados.	24
Figura 11 - Exemplo de diferenciação de identidades faciais por meio de vetores gerados por rede neural convolucional (CNN). (Autor, 2025).	24
Figura 12 - Representação do espaço de <i>embeddings</i> , mostrando a separação entre diferentes identidades faciais. Fonte: Adaptado de Schroff, Kalenichenko e Philbin (2015).	25
Figura 13 - Exemplo de sistema de sala de aula inteligente com sensores de ambiente e comportamento, processadores em nuvem/edge e atuadores para resposta automática. Fonte: Adaptado de Zhang et al. (2024).	27
Figura 14 - Reconhecimento dos <i>Keypoints</i> do Corpo Humano Utilizando Modelo OpenPose. (DAI; XIE; YU, 2025).	29
Figura 15 - Classificação de Status Comportamentais em Sala de Aula. (LIU; JIANG; JIANG, 2025).	30
Figura 16 - Fluxo de Detecção de Comportamento em Sala de Aula. (LIU, 2023).	31
Figura 17 - Arquitetura multimodal de monitoramento comportamental, integrando câmeras individuais, câmera central, sensores fisiológicos e modelos de <i>deep learning</i> para estimar postura, emoções e atividades dos estudantes. Fonte: Adaptado de Márquez-Carpintero et al 2023.	32
Figura 18 - Estrutura de rede para reconhecimento de comportamentos em sala de aula, incorporando super-resolução (SRGAN) e um YOLOv8 aprimorado para detectar ações como levantar a mão, leitura, escrita, uso de celular, cabeça baixa e apoio sobre a mesa. Fonte: Zhang et al., 2024.	33
Figura 19 - Arquitetura do <i>InsightFace</i> com <i>ArcFace Loss</i> . (DENG, 2019).	34

Figura 20 - Representação geométrica da margem angular introduzida pelo <i>ArcFace</i> em comparação a outras perdas de reconhecimento facial (<i>Softmax</i> , <i>SphereFace</i> e <i>CosFace</i>).	35
Figura 21 - Estrutura de um bloco residual utilizado na arquitetura <i>ResNet</i> , com conexões de atalho que permitem o treinamento de redes profundas para extração de embeddings faciais. Fonte: Adaptado de He et al. (2016).	36
Figura 22 - Comparação visual entre diferentes funções de perda aplicadas ao reconhecimento facial, destacando a separação intra e interclasse e o efeito da margem angular. Fonte: Adaptado de Deng et al. (2019).	38
Figura 23 - Fluxo de registro e verificação com extração de <i>embeddings</i> , criptografia homomórfica e armazenamento seguro em nuvem. Fonte Song et al. 2025.	42
Figura 24 - Câmera Hikivision. (HIKVISION, 2025).	45
Figura 25 - Cluster com múltiplas GPUs utilizado nos testes de processamento em larga escala. Fonte: (Autor, 2025).	47
Figura 26 - Detecção dos pontos chave utilizando <i>YOLOv11-Pose</i> . (ULTRALYTICS, 2025).	49
Figura 27 - Representação esquemática dos cinco comportamentos detectados pela plataforma a partir da análise geométrica dos <i>keypoints</i> corporais estimados pelo modelo <i>YOLOv11-Pose</i> . Fonte: (Autor, 2025).	50
Figura 28 - Versões <i>YOLOv11-Pose</i> . (ULTRALYTICS, 2025).	51
Figura 29 - Tela de cadastro dos alunos na plataforma. Fonte: (Autor, 2026).	52
Figura 30 - Pipeline de identificação facial utilizado no sistema. (Autor, 2025).	53
Figura 31- Fluxo de reconhecimento facial – Fonte: (Autor, 2025).	54
Figura 32 - Modelo relacional do banco de dados da plataforma e monitoramento comportamental. (Autor, 2026).	57
Figura 33 - Estrutura da tabela <i>face_embeddings</i> no PostgreSQL, exibindo o armazenamento dos vetores faciais (<i>embeddings</i>) utilizando a extensão <i>pgvector</i> . (Autor, 2026).	58
Figura 34 - Fluxo geral de implementação do sistema de monitoramento comportamental. (Autor, 2025).	59
Figura 35 - Exemplo integrado de reconhecimento facial e detecção de comportamentos. Fonte: (Autor, 2026).	64
Figura 36 - Distribuição percentual de permanência do aluno em cada comportamento. Fonte: (Autor, 2026).	65
Figura 37 - Tempo total, em minutos, dedicado a cada estado comportamental identificado pela plataforma durante a aula. Fonte: (Autor, 2026).	66
Figura 38 - Gráfico de contagem temporal dos comportamentos do aluno. Fonte: (Autor, 2026).	68
Figura 39 - Comparação percentual dos comportamentos observados em diferentes tipos de aula, evidenciando variações no nível de atenção e distração entre atividades expositivas e avaliativas. (Autor, 2026).	69

LISTA DE ABREVIATURAS E SIGLAS

Sigla	Significado
ADHD	<i>Attention Deficit Hyperactivity Disorder</i> (Transtorno de Déficit de Atenção e Hiperatividade)
BN	<i>Batch Normalization</i>
CNN	<i>Convolutional Neural Network</i> (Rede Neural Convolucional)
CPU	<i>Central Processing Unit</i>
CSV	<i>Comma-Separated Values</i>
CUDA	<i>Compute Unified Device Architecture</i>
cuDNN	<i>CUDA Deep Neural Network Library</i>
FPS	<i>Frames Per Second</i> (Quadros por Segundo)
GPU	<i>Graphics Processing Unit</i>
IA	Inteligência Artificial
ISO	<i>International Organization for Standardization</i>
LGPD	Lei Geral de Proteção de Dados
NVR	<i>Network Video Recorder</i>
PDF	<i>Portable Document Format</i>
PGVector	Extensão do PostgreSQL para armazenamento e busca vetorial
PostgreSQL	Sistema gerenciador de banco de dados objeto-relacional
RGB	<i>Red, Green, Blue</i>
RTSP	<i>Real Time Streaming Protocol</i>
SQL	<i>Structured Query Language</i>
TDAH	Transtorno de Déficit de Atenção e Hiperatividade
TLS	<i>Transport Layer Security</i>
YOLO	<i>You Only Look Once</i> (modelo de detecção de objetos)
YOLO-Pose	Variante do YOLO para estimativa de pose humana

SUMÁRIO

1. INTRODUÇÃO	10
1.1 Problema de pesquisa.....	11
1.2 Objetivo geral	12
1.3 Objetivos específicos	12
2. FUNDAMENTAÇÃO TEÓRICA.....	13
2.1 Transtornos de Aprendizagem e Dificuldade de Aprendizagem na Sala de Aula.....	13
2.2 Visão Computacional Aplicada à Educação	16
2.2.1 Monitoramento do Engajamento e Classificação de Interações.....	17
2.2.2 Suporte à Governança e Personalização Pedagógica com Análise de Dados	19
2.2.3 Fundamentos do Processamento de Imagens	20
2.2.4 Segmentação, Detecção e Extração de Regiões de Interesse	21
2.2.5 Reconhecimento Facial e Embeddings na Educação	22
2.2.6 Aplicações da Visão Computacional no Contexto Educacional	25
2.3 Estimativa de Pose Humana para Análise Comportamental	28
2.4 Reconhecimento Facial e Expressões no Contexto Educacional	31
2.5 Modelo de Fusão Espaço-Temporal para Reconhecimento de Comportamento.....	38
3. METODOLOGIA.....	43
3.1 Coleta de imagens	44
3.1.1 Câmera.....	44
3.2 Software	45
3.2.1 Detecção do Comportamento.....	48
3.2.2 Processo de Cadastro e Coleta das Imagens dos Alunos	51
3.2.3 Identificação do Aluno.....	52
3.2.4 Demonstração das Informações	54
3.2.5 Registro e Armazenamento Estruturado dos Dados no PostgreSQL	56
3.3 Fluxo de Implementação.....	58
3.4 Avaliação da Viabilidade da Plataforma.....	59
3.5 Procedimentos de Coleta de Dados	60
3.6 Considerações Éticas.....	61
4. RESULTADOS PARCIAIS.....	62
4.1 Implementação da Estrutura da Plataforma.....	62
4.2 Reconhecimento Facial com <i>InsightFace</i>	62
4.3 Detecção de Comportamentos com YOLOv11-Pose.....	63
4.4 Integração com Banco de Dados e Visualização em Streamlit	64
4.5 Análise Temporal dos Dados	66
5. CONSIDERAÇÕES FINAIS.....	69
REFERÊNCIAS BIBLIOGRÁFICAS	71
APÊNDICE A – Relatório observacional pedagógico gerado pela plataforma.....	73

1. INTRODUÇÃO

O ambiente escolar, especialmente no ensino básico, apresenta desafios significativos relacionados ao acompanhamento contínuo do comportamento e do engajamento dos alunos em sala de aula. Professores frequentemente precisam lidar com turmas numerosas, múltiplas demandas simultâneas e tempo limitado para observações individuais, o que torna difícil identificar, de forma sistemática, padrões comportamentais relevantes ao processo de aprendizagem.

Nesse contexto, aspectos como atenção, participação, distração, agitação ou sinais de cansaço tendem a ser percebidos de maneira pontual e subjetiva, podendo variar de acordo com a experiência do docente e as condições do ambiente. Essa limitação dificulta a construção de uma visão mais abrangente e contínua sobre o comportamento dos estudantes ao longo das aulas, o que pode impactar a tomada de decisão pedagógica e o planejamento de intervenções educacionais mais adequadas.

Diante desse cenário, o uso de tecnologias baseadas em Inteligência Artificial (IA) e Visão Computacional surge como uma alternativa promissora para apoiar o acompanhamento do comportamento discente. A possibilidade de analisar imagens em tempo real e extrair informações estruturadas permite transformar observações visuais em dados organizados, favorecendo uma compreensão mais objetiva da dinâmica da sala de aula.

Este trabalho propõe o desenvolvimento de uma plataforma de monitoramento comportamental aplicada ao contexto educacional, capaz de identificar padrões observáveis de comportamento a partir da análise de imagens capturadas em sala de aula. A solução integra o modelo YOLOv11-Pose para estimativa de pose humana e a biblioteca InsightFace para reconhecimento facial, possibilitando associar comportamentos como “atento”, “distraído”, “perguntando”, “agitado” e “dormindo” aos respectivos alunos.

A proposta não tem como objetivo substituir a observação do professor, nem realizar qualquer tipo de diagnóstico clínico, mas sim oferecer uma ferramenta de apoio à análise pedagógica. Ao registrar e organizar dados comportamentais ao longo do tempo, o sistema permite identificar padrões, tendências e variações no comportamento dos alunos, contribuindo para uma compreensão mais estruturada do ambiente educacional.

Por meio de uma interface interativa desenvolvida em Streamlit, a plataforma possibilita a visualização de informações em forma de gráficos, indicadores e sínteses, facilitando a interpretação dos dados por parte dos professores. Dessa forma, busca-se fornecer subsídios que auxiliem na tomada de decisão pedagógica, respeitando os limites éticos e interpretativos do uso de tecnologias de visão computacional no ambiente escolar.

Diante disso, esta dissertação apresenta a concepção, o desenvolvimento e a avaliação de uma plataforma de monitoramento comportamental baseada em Inteligência Artificial, com foco no apoio à prática educacional e na análise de padrões comportamentais em sala de aula.

1.1 Problema de pesquisa

Apesar dos avanços no uso de tecnologias educacionais, o acompanhamento do comportamento dos alunos em sala de aula ainda depende, em grande parte, da observação subjetiva dos professores. Em turmas numerosas e com múltiplas demandas simultâneas, torna-se difícil realizar uma análise contínua e individualizada dos estudantes, o que pode limitar a identificação de padrões comportamentais relevantes ao processo de aprendizagem.

Além disso, a ausência de ferramentas que permitam registrar, organizar e analisar esses comportamentos de forma estruturada dificulta a compreensão da dinâmica da sala de aula ao longo do tempo. Sistemas tradicionais de avaliação educacional não contemplam, de maneira sistemática, aspectos comportamentais observáveis, como atenção, participação, distração, agitação ou sinais de cansaço.

Diante desse contexto, emerge a seguinte questão de pesquisa:

Como uma plataforma baseada em visão computacional e inteligência artificial pode contribuir para o monitoramento comportamental de alunos do ensino básico, apoiando a análise pedagógica por meio da identificação de padrões comportamentais observáveis em sala de aula?

1.2 Objetivo geral

Desenvolver uma plataforma de monitoramento comportamental para salas de aula do ensino básico, utilizando técnicas de visão computacional e inteligência artificial, com o propósito de identificar padrões observáveis de comportamento dos estudantes, como atenção, participação, distração, agitação e sonolência, a fim de apoiar a análise pedagógica e a tomada de decisão educacional baseada em dados.

1.3 Objetivos específicos

Para alcançar o objetivo geral, estabelecem-se os seguintes objetivos específicos:

- Implementar o reconhecimento facial dos alunos em sala de aula utilizando a biblioteca InsightFace, permitindo a identificação individual durante o monitoramento.
- Desenvolver a detecção de comportamentos a partir da estimativa de pose humana com o modelo YOLOv11-Pose, considerando padrões observáveis no contexto educacional.
- Associar os comportamentos identificados (atento, distraído, perguntando, agitado e dormindo) aos respectivos alunos reconhecidos em tempo real.
- Estruturar o armazenamento dos dados comportamentais em banco de dados, possibilitando consultas por aluno, turma e período.
- Desenvolver uma interface interativa em Streamlit para visualização dos dados por meio de gráficos, indicadores e sínteses.
- Analisar os dados gerados pela plataforma com foco na identificação de padrões comportamentais que possam apoiar a interpretação pedagógica.

2. FUNDAMENTAÇÃO TEÓRICA

2.1 Transtornos de Aprendizagem e Dificuldade de Aprendizagem na Sala de Aula

A compreensão dos fenômenos que afetam o processo de aprendizagem exige distinguir, de maneira clara, as dificuldades de aprendizagem dos transtornos de aprendizagem, distinção essencial para o uso ético e pedagógico de qualquer plataforma de monitoramento comportamental.

Os transtornos de aprendizagem como Dislexia, Disortografia, Discalculia, TDAH e condições do Espectro Autista possuem origem neurobiológica, relacionando-se a diferenças funcionais no processamento cerebral. São condições persistentes, caracterizadas por padrões recorrentes de desatenção, hiperatividade, impulsividade, dificuldades específicas de leitura e escrita, processamento fonológico ou organização cognitiva, demandando acompanhamento especializado e intervenções contínuas (BELLO; MACHADO; CAPELLINE, 2022).

Por outro lado, as dificuldades de aprendizagem apresentam natureza temporária e geralmente resultam de fatores ambientais, emocionais, sociais ou pedagógicos. Entre esses fatores incluem-se: metodologias inadequadas, baixa estimulação prévia, conflitos familiares, ansiedade, privação de sono, mudanças bruscas de rotina ou falta de vínculos escolares. Nesse sentido, muitos casos de baixo rendimento não decorrem de transtornos, mas de condições contextuais que afetam momentaneamente a capacidade de concentração e engajamento do estudante.

A compreensão do desenvolvimento infantil exige considerar que o processo de aprendizagem ocorre de forma gradual e multidimensional ao longo da infância. Cada etapa do crescimento está associada a marcos específicos relacionados ao desenvolvimento motor, cognitivo, social e emocional da criança. Nesse contexto, estratégias pedagógicas adequadas devem buscar estimular essas capacidades de forma integrada, respeitando o ritmo individual de cada estudante. A Figura 1 apresenta uma síntese das estratégias voltadas ao desenvolvimento máximo das capacidades dos estudantes ao longo de seus marcos evolutivos, evidenciando como diferentes estímulos educacionais podem contribuir para o fortalecimento

das competências cognitivas e socioemocionais durante o processo formativo (BELLO; MACHADO; CAPELLINE, 2022).

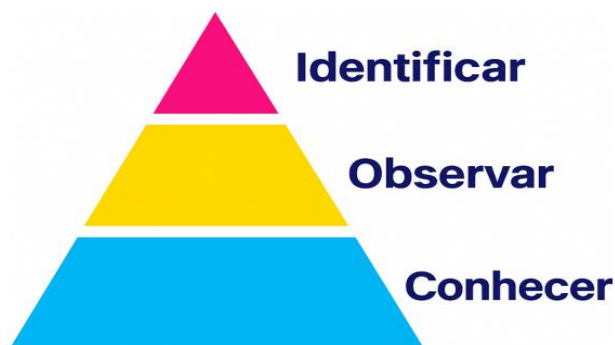


Figura 1 - Estratégias para desenvolver o máximo das capacidades estudantes ao longo dos seus marcos de desenvolvimento. Fonte: (BELLO, MACHADO; CAPELLINE, 2022).

Diversos estudos apontam que a baixa aprendizagem costuma ser precedida por sinais comportamentais sutis, como inquietação, queda progressiva da atenção, retraimento social, postura corporal de cansaço, movimentos repetitivos das mãos ou evitação de tarefas escolares. Entretanto, a identificação desses sinais depende quase exclusivamente da percepção do professor, que enfrenta limitações naturais: turmas numerosas, múltiplas demandas simultâneas, tempo reduzido para observações individuais e forte carga emocional diária (LIU; JIANG; JIANG, 2025).

Nessa realidade, sinais relevantes podem passar despercebidos ou ser interpretados de forma equivocada por exemplo, uma criança ansiosa pode ser vista como desinteressada, um aluno com TDAH pode ser interpretado como indisciplinado, um estudante com dificuldades temporárias pode ser confundido com um caso clínico.

O desenvolvimento infantil ocorre por meio de transformações progressivas que envolvem diferentes dimensões do crescimento humano, incluindo aspectos motores, cognitivos, linguísticos e socioemocionais. Esses marcos evolutivos são amplamente utilizados na literatura educacional e psicológica como referências para avaliar o progresso das crianças e identificar possíveis atrasos ou dificuldades no processo de aprendizagem. A compreensão desses estágios é fundamental para educadores e pesquisadores, pois permite interpretar comportamentos observados no ambiente escolar à luz das características típicas de cada fase do desenvolvimento. A Figura 2 ilustra os principais marcos do desenvolvimento

infantil, destacando a evolução dessas capacidades ao longo das diferentes etapas da infância (BELLO; MACHADO; CAPELLINE, 2022).



Figura 2 - Representação dos marcos do desenvolvimento infantil, destacando a evolução das capacidades motoras, cognitivas, linguísticas e socioemocionais ao longo da infância. Fonte: (BELLO, MACHADO; CAPELLINE, 2022).

Além disso, dificuldades de aprendizagem não identificadas podem desencadear impactos emocionais significativos, como sentimentos de incapacidade, vergonha, ansiedade escolar ou retraimento social. Esses efeitos podem se somar ao baixo desempenho e perpetuar um ciclo de autopercepção negativa, prejudicando o engajamento do estudante nas atividades pedagógicas.

Nesse sentido, torna-se essencial que a escola disponha de ferramentas que permitam observar o aluno de forma longitudinal não apenas em episódios pontuais possibilitando identificar padrões como:

- repetição de comportamentos atípicos;
- variação do nível de atenção ao longo das aulas;
- mudanças bruscas de humor ou postura;
- agitação motora persistente;
- retraimento ou baixa participação.

O acompanhamento contínuo, quando bem conduzido, evita interpretações isoladas e contribui para decisões pedagógicas mais assertivas.

Além da identificação dos marcos de desenvolvimento, é igualmente importante compreender os domínios que estruturam o crescimento infantil e que influenciam diretamente o desempenho acadêmico e social das crianças. Esses domínios englobam

aspectos motores, cognitivos, socioemocionais e linguísticos, os quais interagem continuamente no processo de aprendizagem e adaptação ao ambiente escolar. A análise desses componentes permite identificar padrões de desenvolvimento esperados, bem como possíveis sinais de dificuldades ou transtornos que podem impactar o processo educacional. Nesse sentido, a Figura 3 apresenta os principais domínios do desenvolvimento infantil, destacando sua importância como referência para a análise do progresso e para a identificação de possíveis atrasos no desenvolvimento das crianças (BELLO; MACHADO; CAPELLINE, 2022).



Figura 3 - Domínios fundamentais do desenvolvimento infantil, incluindo aspectos motores, cognitivos, socioemocionais e da linguagem, utilizados como referência para identificação de progressos e possíveis atrasos. Fonte: (BELLO, MACHADO; CAPELLINE, 2022).

Embora a literatura apresente relações entre comportamento e possíveis dificuldades de aprendizagem, este trabalho não tem como objetivo realizar qualquer tipo de diagnóstico clínico. A abordagem proposta limita-se à identificação de padrões comportamentais observáveis em sala de aula, com foco no apoio à análise pedagógica.

2.2 Visão Computacional Aplicada à Educação

A visão computacional é um campo da inteligência artificial que permite aos computadores interpretar e compreender imagens e vídeos, extraindo informações relevantes para apoiar a tomada de decisões. Segundo (LIU; JIANG; JIANG, 2025), essa área busca simular a capacidade humana de percepção visual por meio de métodos computacionais, possibilitando aplicações que variam desde inspeção industrial até o reconhecimento de padrões comportamentais em contextos educacionais.

A aplicação da visão computacional no contexto educacional tem ganhado destaque nos últimos anos, permitindo a análise automatizada de comportamentos em sala

de aula a partir de fluxos de vídeo. De acordo com (YUVARAJ; PRINCE; MURUGAPPAN, 2024), o processo envolve etapas de pré-processamento, como extração e seleção de quadros relevantes e redimensionamento das imagens, seguidas pela utilização de modelos híbridos, como redes neurais convolucionais bidimensionais (2D CNN) integradas a redes recorrentes (RNN). Essa arquitetura possibilita a classificação de diferentes ações realizadas pelos alunos, como levantar a mão, explicar, ler, escrever no quadro ou utilizar o telefone celular, por meio de camadas de decisão como o *Softmax*. A Figura 4 ilustra esse fluxo, evidenciando a importância da visão computacional para a identificação e monitoramento de comportamentos, contribuindo para análises pedagógicas e para a melhoria do ambiente escolar.

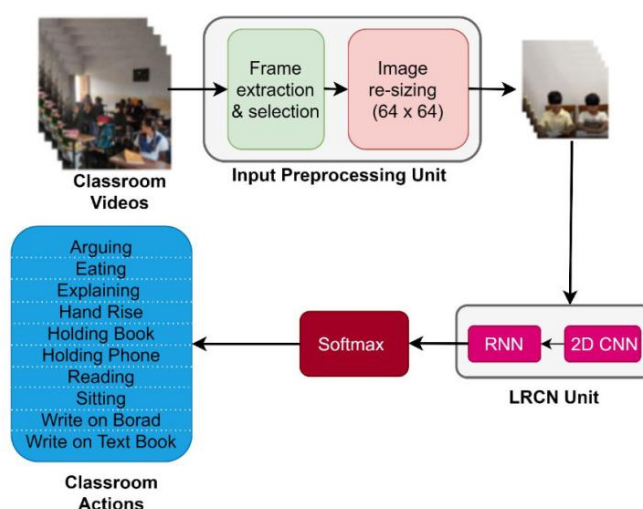


Figura 4 - Fluxo de Visão Computacional para Detecção de Comportamento em Sala de Aula.
Fonte: (YUVARAJ; PRINCE; MURUGAPPAN, 2024)

A aplicação dessas técnicas no contexto educacional permite um acompanhamento contínuo e objetivo, servindo de suporte para a personalização das estratégias pedagógicas. Quando combinada a modelos de análise de comportamento, a visão computacional também pode auxiliar na identificação precoce de sinais de transtornos.

2.2.1 Monitoramento do Engajamento e Classificação de Interações

Um dos principais benefícios da Visão Computacional na educação é sua capacidade de transformar observações subjetivas em registros objetivos e contínuos. Modelos baseados em análise postural e reconhecimento de ações são capazes de classificar comportamentos específicos, como:

- levantar a mão;
- escrever;
- ler;
- conversar;
- usar celular;
- inclinar a cabeça;
- virar o tronco repetidamente;
- demonstrar sinais de sonolência.

A análise automática de comportamentos em ambientes educacionais tem sido amplamente investigada por meio da integração de modelos de aprendizado profundo capazes de interpretar padrões espaciais e temporais presentes nas ações dos estudantes. Nesse contexto, arquiteturas híbridas que combinam redes neurais convolucionais bidimensionais (2D CNN) com modelos recorrentes, como as redes Long Short-Term Memory (LSTM), têm se mostrado particularmente eficazes para reconhecer sequências de ações ao longo do tempo. Enquanto as CNNs são responsáveis pela extração de características visuais relevantes em cada quadro da imagem, os modelos recorrentes capturam a dependência temporal entre essas observações, permitindo identificar padrões comportamentais mais complexos. A Figura 5 apresenta uma arquitetura híbrida baseada nessa combinação de modelos, evidenciando o fluxo de processamento desde a extração de características visuais até a classificação final das ações observadas no ambiente educacional (YUVARAJ; PRINCE; MURUGAPPAN, 2024).

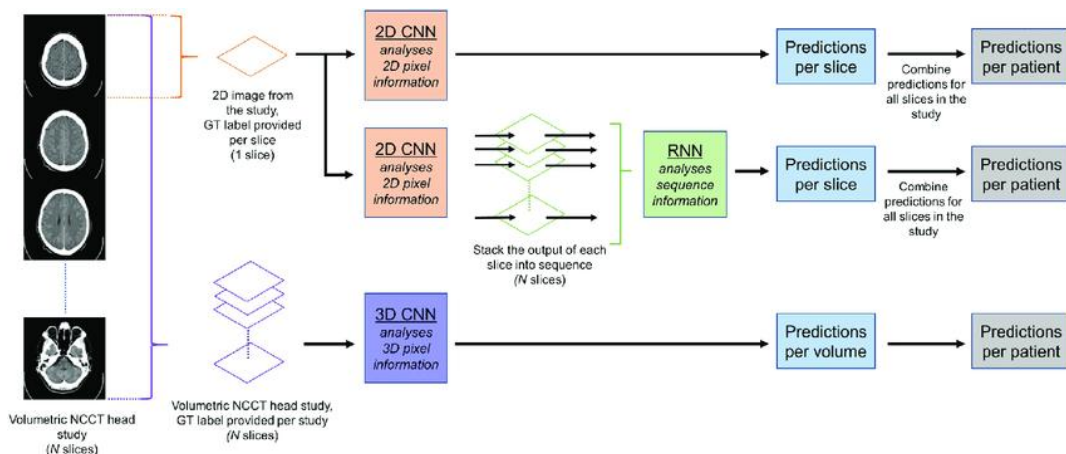


Figura 5 - Arquitetura híbrida baseada em 2D CNN e modelos recorrentes para classificação de ações estudantis. Fonte: Adaptado de Yuvaraj, Prince e Murugappan (2024).

2.2.2 Suporte à Governança e Personalização Pedagógica com Análise de Dados

A Visão Computacional tem impacto direto nos processos de gestão educacional baseada em evidências. Ao registrar dados objetivos de comportamento, a plataforma possibilita:

- acompanhamento contínuo do estudante;
- identificação de padrões de engajamento;
- diferenciação entre comportamentos ocasionais e recorrentes;
- suporte à inclusão e ao atendimento especializado;
- intervenções pedagógicas personalizadas.

Conforme Liu, Jiang e Jiang (2025), a ausência de instrumentos que permitam esse monitoramento sistemático limita a capacidade das escolas de detectar precocemente sinais de dificuldades ou transtornos. A introdução de plataformas automatizadas fortalece estratégias de governança, planejamento pedagógico e acompanhamento individualizado.

A literatura recente sobre monitoramento educacional tem explorado arquiteturas integradas que combinam diferentes técnicas de visão computacional e aprendizado profundo para interpretar o comportamento dos estudantes em sala de aula. Esses sistemas geralmente utilizam múltiplos módulos responsáveis por detectar ações, reconhecer emoções faciais e identificar indivíduos, permitindo uma análise mais abrangente do nível de atenção e engajamento durante as atividades escolares. A integração dessas tecnologias possibilita que diferentes tipos de informação sejam processados simultaneamente, gerando indicadores mais confiáveis sobre o comportamento discente. A Figura 6 apresenta uma arquitetura de sistema capaz de realizar detecção de ações, análise emocional e reconhecimento facial de estudantes, evidenciando o fluxo de processamento das informações capturadas pelas câmeras até a classificação dos estados de atenção e engajamento no ambiente educacional (SINGH; GURURANI; BHATNAGAR, 2023).

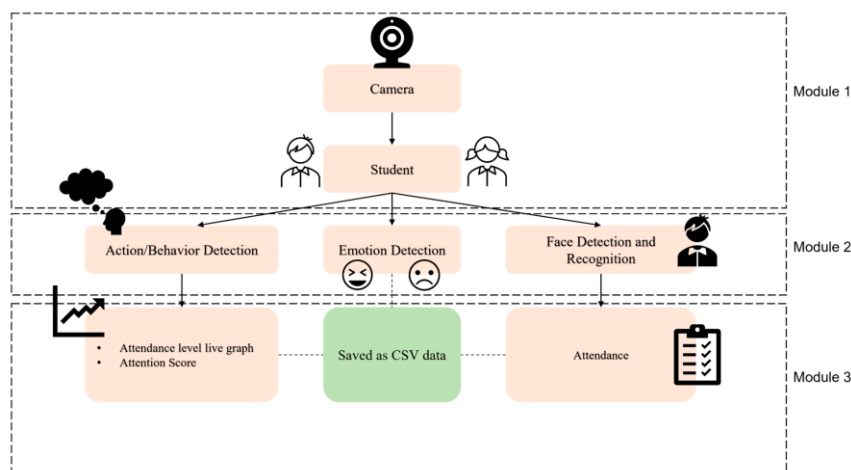


Figura 6 - Arquitetura do sistema de detecção de ações, emoções e reconhecimento facial para análise de atenção estudantil. Fonte: Adaptado de Singh, Gururani e Bhatnagar (2023).

2.2.3 Fundamentos do Processamento de Imagens

Embora a Visão Computacional aplicada à educação esteja fortemente baseada em modelos de *Deep Learning*, grande parte das plataformas depende de etapas clássicas do Processamento Digital de Imagens (PDI), que transformam imagens em dados manipuláveis pelos algoritmos. Uma imagem é composta por uma matriz de *pixels*, cada um contendo informação de cor e intensidade. No formato mais comum, o modelo RGB, cada *pixel* possui três canais (vermelho, verde e azul), representando uma tríade de valores que variam entre 0 e 255. Essa estrutura permite a codificação de milhões de cores, possibilitando que algoritmos identifiquem variações sutis entre regiões da imagem (GONZALEZ; WOODS, 2018).

Em muitas plataformas, incluindo modelos de comportamento estudantil, a conversão para *grayscale* (tons de cinza) é utilizada para reduzir a complexidade computacional, preservando apenas informações essenciais de intensidade luminosa. Essa conversão facilita operações como detecção de bordas, realce de contornos e eliminação de ruídos, etapas cruciais para garantir que algoritmos como *YOLO-Pose* ou *OpenPose* identifiquem corretamente os *keypoints* corporais.

Outro componente central do PDI são as máscaras e filtros (*kernels*), pequenas matrizes que percorrem a imagem para extrair padrões. Filtros como Sobel, Laplaciano ou Canny são amplamente utilizados para detectar bordas, enquanto a filtragem Gaussiana reduz ruídos, favorecendo a segmentação posterior (YUVARAJ; PRINCE; MURUGAPPAN, 2024). Esses filtros são essenciais para melhorar a qualidade das informações enviadas à rede

neural, especialmente em ambientes escolares com iluminação irregular ou sombras, situações comuns durante o funcionamento da plataforma.

No processamento digital de imagens, uma das operações fundamentais consiste na aplicação de filtros espaciais sobre a matriz de pixels que compõe a imagem. Esses filtros utilizam pequenas matrizes, conhecidas como *kernels* ou máscaras de convolução, que percorrem a imagem original realizando operações matemáticas entre os valores dos *pixels* e os coeficientes definidos no filtro. Esse procedimento permite destacar características específicas da imagem, como bordas, texturas ou regiões de interesse, além de reduzir ruídos e melhorar a qualidade dos dados que serão utilizados por algoritmos de visão computacional. A Figura 7 ilustra o funcionamento desse processo, demonstrando como uma matriz de entrada (*image matrix*) é combinada com um *kernel* de convolução para gerar uma nova matriz de saída (*output matrix*), resultado da aplicação do filtro sobre os pixels da imagem.

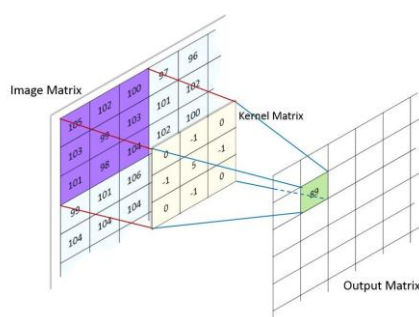


Figura 7 - Filtro de Imagem – Fonte: (Autor, 2026).

2.2.4 Segmentação, Detecção e Extração de Regiões de Interesse

A segmentação de imagens consiste em dividir a imagem em áreas significativas para o algoritmo, como o corpo do estudante, as mãos ou o rosto. Essa separação permite que a plataforma concentre sua análise apenas nas regiões relevantes, ignorando o fundo da sala ou objetos sem importância. Estudos recentes mostram que técnicas como Mask R-CNN e YOLO-Seg podem melhorar a precisão na identificação de gestos, direção do olhar e postura corporal (LIU; JIANG; JIANG, 2025).

Na plataforma proposta, embora a segmentação explícita não seja realizada como uma etapa isolada, ela ocorre implicitamente quando o *YOLOv11-Pose* detecta o corpo e extrai

keypoints estruturais (cabeça, ombros, cotovelos, punhos, quadris etc.). Cada um desses pontos é tratado como uma Região de Interesse (ROI), permitindo inferir comportamentos como:

- cabeça abaixada → possível sonolência ou leitura;
- braços erguidos → participação ou pergunta;
- curvatura excessiva do tronco → cansaço, desengajamento ou postura inadequada;
- movimentos repetitivos das mãos → possíveis estereotípias.

A Figura 8 apresenta um exemplo do conjunto de keypoints corporais utilizados em modelos de estimativa de pose humana.

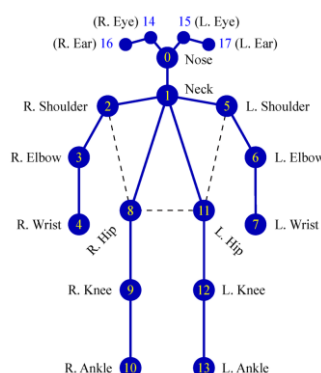


Figura 8 - Conjunto de keypoints corporais utilizados em modelos de estimativa de pose humana.
Fonte: Adaptado de Dai, Xie e Yu (2025).

A combinação dessas informações espaciais com a análise temporal discutida posteriormente (WANG, 2023) fornece uma estrutura robusta para o reconhecimento comportamental. Assim, mesmo sem uma etapa tradicional de segmentação, a plataforma consegue isolar regiões relevantes e interpretá-las em tempo real, apoiando a classificação de comportamentos associados ao engajamento e à atenção dos estudantes.

2.2.5 Reconhecimento Facial e Embeddings na Educação

O reconhecimento facial é uma das tecnologias mais utilizadas em plataformas de Visão Computacional e tem se consolidado como uma ferramenta relevante no contexto educacional para identificação automática de estudantes, controle de presença e

personalização de estratégias pedagógicas. O processo baseia-se na capacidade de um algoritmo extrair características únicas de um rosto humano e representá-las matematicamente de forma compacta e discriminativa. Essa representação recebe o nome de *embedding* facial, que funciona como uma “impressão digital” do rosto, permitindo comparar indivíduos de maneira rápida e precisa.

Segundo Deng et al. (2019), modelos modernos de reconhecimento facial, como os implementados na biblioteca *InsightFace*, aprendem a mapear fotografias de rostos para um espaço vetorial de alta dimensão, onde rostos semelhantes se situam próximos e rostos distintos se afastam. Esse espaço, conhecido como *embedding space*, permite verificar identidade por meio de distâncias matemáticas, normalmente utilizando métricas como distância euclidiana ou cosseno. Essa abordagem tornou-se o padrão da área após os trabalhos de Schroff, Kalenichenko e Philbin (2015), responsáveis pelo *FaceNet*, um dos primeiros modelos a demonstrar que *embeddings* bem treinados superavam métodos tradicionais de reconhecimento facial.

A Figura 9 ilustra esse fluxo, destacando as três etapas essenciais: captura da imagem contendo um rosto, alinhamento facial e geração do *embedding*. Esse processo é a base de plataformas modernas como *InsightFace*, que utilizam estratégias avançadas de normalização angular para aumentar a separabilidade entre identidades.

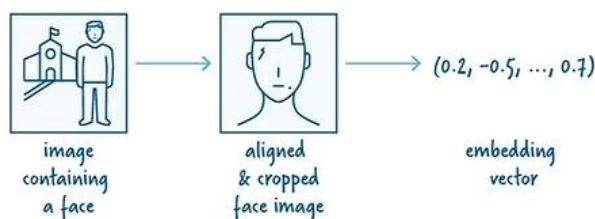


Figura 9 - Processo de reconhecimento facial: detecção, alinhamento do rosto e geração do *embedding* vetorial. Fonte: Adaptado de Schroff, Kalenichenko e Philbin (2015).

A geração de *embeddings* é realizada por meio de uma sequência estruturada de etapas que envolve detecção facial, alinhamento geométrico e extração vetorial. Esse processo transforma a imagem do rosto em um conjunto de valores numéricos que representam apenas suas características essenciais, sem manter informações visuais diretamente identificáveis. A Figura 10 ilustra esse fluxo, demonstrando como a fotografia inicial é convertida em um vetor de alta dimensão e posteriormente armazenada em banco de dados para consultas rápidas por similaridade.

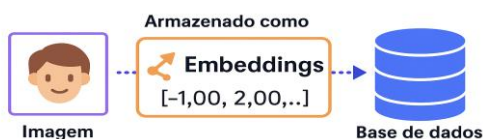


Figura 10 - Processo de conversão da imagem facial em *embedding* vetorial e armazenamento em banco de dados. Fonte: (Autor, 2026).

Após gerados, os *embeddings* são projetados em um espaço vetorial de alta dimensão, onde rostos semelhantes ficam próximos e rostos distintos se afastam. Essa organização espacial permite que a plataforma compare identidades de forma matemática, utilizando métricas como distância euclidiana ou cosseno. A capacidade de separar diferentes indivíduos nesse espaço é o que garante a alta precisão dos modelos contemporâneos, especialmente aqueles baseados em arquiteturas como *ArcFace* e *InsightFace*.

A Figura 11 ilustra esse fluxo, demonstrando como a fotografia inicial é convertida em um vetor de alta dimensão e posteriormente armazenada em banco de dados para consultas rápidas por similaridade.

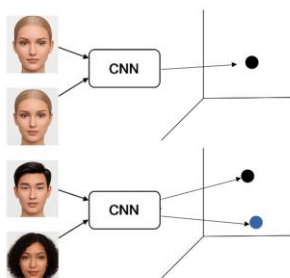


Figura 11 - Exemplo de diferenciação de identidades faciais por meio de vetores gerados por rede neural convolucional (CNN). Fonte: (Autor, 2025).

A Figura 12 apresenta uma visualização desse espaço, onde *embeddings* de vários indivíduos são agrupados em clusters distintos. Cada cor representa uma pessoa diferente, demonstrando como o modelo aprende a separar identidades de forma clara e consistente, mesmo em condições reais de variação de ângulo, iluminação ou expressão facial.

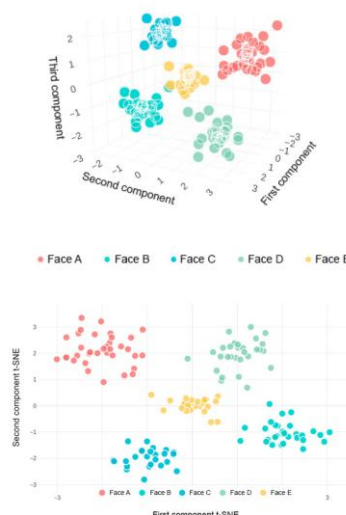


Figura 12 - Representação do espaço de *embeddings*, mostrando a separação entre diferentes identidades faciais. Fonte: Adaptado de Schroff, Kalenichenko e Philbin (2015).

Essa capacidade de representar e diferenciar identidades faciais de forma robusta é particularmente útil no ambiente educacional. Ao vincular o *embedding* de cada aluno às ações detectadas por modelos de estimativa de pose, a plataforma consegue identificar quem está participando, quem está distraído, quem apresenta sinais de sonolência ou inquietação e quem demonstra engajamento elevado. Além disso, a utilização de *embeddings* em vez de imagens brutas reduz riscos de privacidade, uma vez que os vetores numéricos não permitem reconstruir o rosto original, tornando o processo mais alinhado às exigências da Lei Geral de Proteção de Dados (LGPD) e às boas práticas de segurança mencionadas por Wang (2023).

2.2.6 Aplicações da Visão Computacional no Contexto Educacional

A Visão Computacional tem se consolidado como uma das tecnologias mais promissoras para apoiar processos educacionais, especialmente em ambientes que demandam acompanhamento contínuo e análise detalhada do comportamento discente. A capacidade de interpretar automaticamente expressões, gestos, posturas corporais e padrões de engajamento possibilita a criação de ferramentas capazes de fornecer dados objetivos sobre a dinâmica da sala de aula. Essa transição de um modelo puramente observacional para um modelo orientado por dados representa um avanço significativo no campo da gestão educacional e da personalização do ensino (YUVARAJ; PRINCE; MURUGAPPAN, 2024).

Entre as aplicações mais relevantes, destaca-se o monitoramento do engajamento e da atenção. Modelos de estimativa de pose como o *YOLOv11-Pose* utilizado nesta pesquisa permitem identificar comportamentos como cabeça abaixada, corpo inclinado, mão levantada, movimentos repetitivos e direcionamento do olhar. Essas ações, quando analisadas ao longo do tempo, fornecem indicadores importantes sobre participação, foco, cansaço, desatenção e até mesmo sinais de sobrecarga cognitiva. Segundo Liu, Jiang e Jiang (2025), esses indicadores permitem compreender não apenas momentos isolados, mas padrões recorrentes que podem sinalizar dificuldades de aprendizagem ou necessidade de intervenção pedagógica.

Outra aplicação relevante é o controle automatizado de presença, possibilitado pelo reconhecimento facial através de *embeddings*. Plataformas desse tipo reduzem significativamente o tempo gasto em tarefas administrativas, eliminando a chamada manual e permitindo que docentes concentrem seus esforços em atividades instrucionais. Além disso, ao registrar automaticamente entradas, saídas e períodos de permanência, tais plataformas podem auxiliar na identificação de alunos com alta frequência de atrasos, ausências súbitas ou comportamentos de evasão. Esses dados, quando integrados a plataformas pedagógicas maiores, tornam-se instrumentos valiosos para o acompanhamento individualizado e intervenções precoces.

O uso da Visão Computacional também tem avançado em áreas relacionadas ao bem-estar emocional dos alunos. Embora o presente trabalho não implemente classificadores de emoção, pesquisas recentes têm explorado modelos capazes de reconhecer expressões faciais que indicam alegria, surpresa, tristeza, frustração ou ansiedade. Esses dados podem ser úteis para identificar estudantes em sofrimento emocional, apoiar equipes multiprofissionais e orientar ações de acolhimento escolar (SINGH; GURURANI; BHATNAGAR, 2023). No entanto, é essencial reforçar que tais tecnologias devem ser utilizadas com responsabilidade ética, respeitando privacidade, consentimento e limites legais, especialmente no que se refere à LGPD.

O conceito de sala de aula inteligente tem sido amplamente explorado na literatura recente como uma abordagem capaz de integrar tecnologias de sensoriamento, processamento computacional e análise de dados para apoiar o processo educacional. Nesse modelo, diferentes tipos de sensores são utilizados para coletar informações do ambiente e do comportamento dos estudantes, como temperatura, qualidade do ar, níveis de ruído e padrões de interação em sala de aula. Esses dados são processados por sistemas computacionais que

podem operar tanto em infraestrutura de nuvem (*cloud computing*) quanto em dispositivos de borda (*edge computing*), possibilitando análises em tempo real e respostas automatizadas do sistema. A Figura 13 apresenta um exemplo de arquitetura de sala de aula inteligente, na qual sensores capturam dados do ambiente e do comportamento dos estudantes, processadores analisam essas informações e atuadores executam respostas automáticas, como alertas sonoros, indicadores luminosos ou atualizações em painéis de monitoramento (ZHANG et al., 2024).

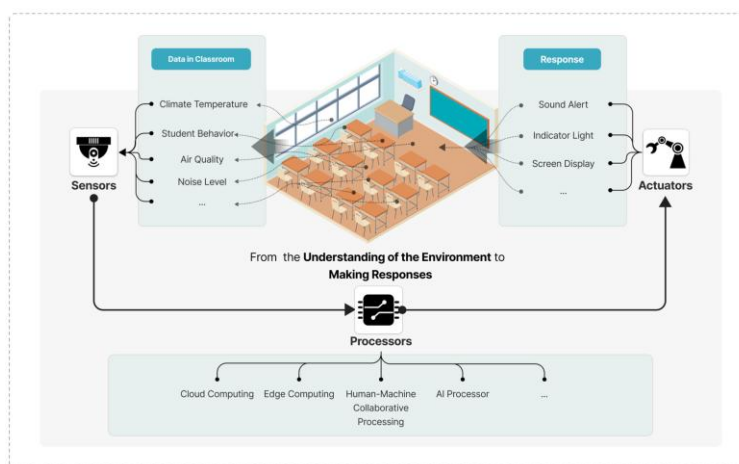


Figura 13 - Exemplo de sistema de sala de aula inteligente com sensores de ambiente e comportamento, processadores em nuvem/edge e atuadores para resposta automática. Fonte: Adaptado de Zhang et al. (2024).

Além das aplicações diretamente relacionadas aos estudantes, a Visão Computacional também pode auxiliar professores na análise de suas próprias práticas pedagógicas. Plataformas baseadas em gravações de aula podem detectar padrões de mobilidade do docente, tempo de exposição no quadro, quantidade de interações com determinados grupos de alunos e até mesmo o equilíbrio entre metodologias expositivas e participativas. Esses dados ativos de observação tornam-se dados estruturados que ajudam o professor a refletir sobre sua prática e adotar estratégias de ensino mais eficientes.

Por fim, a integração entre Visão Computacional, bancos de dados educacionais e plataformas de análise estatística abre caminho para a criação de painéis preditivos capazes de identificar alunos em risco de queda de desempenho, sugerir intervenções personalizadas e gerar insights para políticas públicas educacionais baseadas em evidências. Esse tipo de abordagem, defendido por autores como Wang (2023), representa uma convergência entre

tecnologia e pedagogia que pode transformar o modo como escolas compreendem, acompanham e apoiam seus estudantes.

2.3 Estimativa de Pose Humana para Análise Comportamental

A Estimativa de Pose Humana (*Human Pose Estimation* – HPE) consiste em identificar, em uma imagem ou vídeo, a posição dos principais pontos articulares do corpo como cabeça, ombros, cotovelos, mãos e quadris permitindo compreender a postura, os movimentos e as interações do indivíduo no ambiente. Avanços recentes em visão computacional permitiram que algoritmos de HPE atingissem níveis elevados de precisão e velocidade, tornando seu uso viável em contextos educacionais, especialmente para monitoramento de comportamentos durante as aulas (CAO et al., 2017; LIN et al., 2023).

A análise de pose é particularmente relevante na educação porque muitos comportamentos pedagógicos importantes são expressos por meio de gestos e posturas corporais. Por exemplo, levantar a mão indica intenção de participação; apoiar a cabeça nos braços pode sugerir sonolência; inclinar-se sobre a mesa pode indicar leitura; movimentos repetitivos das mãos podem estar associados a inquietação ou sinais de estresse. Assim, a estimativa de pose atua como uma ponte entre a percepção visual automatizada e a interpretação pedagógica do comportamento discente.

A estimativa de pose humana é uma técnica amplamente utilizada em visão computacional para identificar e rastrear pontos anatômicos do corpo humano a partir de imagens ou sequências de vídeo. Esses pontos, conhecidos como *keypoints*, representam articulações ou regiões relevantes do corpo, como cabeça, ombros, cotovelos, punhos, quadris, joelhos e tornozelos. A partir da identificação dessas coordenadas, torna-se possível reconstruir a estrutura corporal do indivíduo e analisar padrões de postura e movimento. Essa representação estruturada do corpo humano constitui a base para diversos sistemas de reconhecimento de ações e análise comportamental em ambientes educacionais. A Figura 14 apresenta a estrutura de *keypoints* utilizada em modelos de estimativa de pose humana, ilustrando os principais pontos corporais detectados e suas conexões na formação do esqueleto corporal estimado por algoritmos como o *OpenPose* (DAI; XIE; YU, 2025).

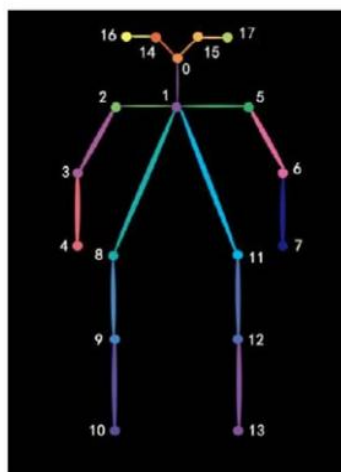


Figura 14 - Reconhecimento dos *Keypoints* do Corpo Humano Utilizando Modelo OpenPose.

Fonte: (DAI; XIE; YU, 2025).

Modelos modernos de HPE podem ser divididos em abordagens *top-down* e *bottom-up*. Métodos *top-down* primeiro detectam a pessoa na imagem para depois estimar sua pose, enquanto métodos *bottom-up* identificam todos os pontos corporais de todos os indivíduos e posteriormente agrupam esses pontos em esqueletos completos. Abordagens *bottom-up*, como *OpenPose*, eram tradicionais (CAO et al., 2017), mas modelos mais recentes como o *YOLO-Pose* e variantes do *YOLOv11-Pose* têm se destacado pela velocidade, simplicidade e robustez características essenciais para aplicações em tempo real, como o monitoramento em sala de aula (LIN et al., 2023).

O modelo utilizado nesta pesquisa, *YOLOv11-Pose*, integra detecção de objeto e estimativa de pose em uma única arquitetura, reduzindo processamento redundante e aumentando a taxa de quadros por segundo (FPS). Essa eficiência permite que a plataforma analise continuamente a postura dos alunos, produzindo dados comportamentais de forma instantânea. Esse diferencial é crucial em ambientes escolares, onde múltiplos estudantes estão em movimento simultâneo e onde atrasos podem prejudicar a interpretação do comportamento coletivo.

Para ilustrar esse processo, a Figura 15 apresenta como diferentes estruturas corporais detectadas em uma mesma cena são segmentadas e classificadas em estados comportamentais distintos. A partir da análise dos esqueletos, a plataforma é capaz de identificar simultaneamente múltiplos alunos e atribuir rótulos comportamentais em tempo real, refletindo o funcionamento prático de modelos baseados em *YOLOv11-Pose*.

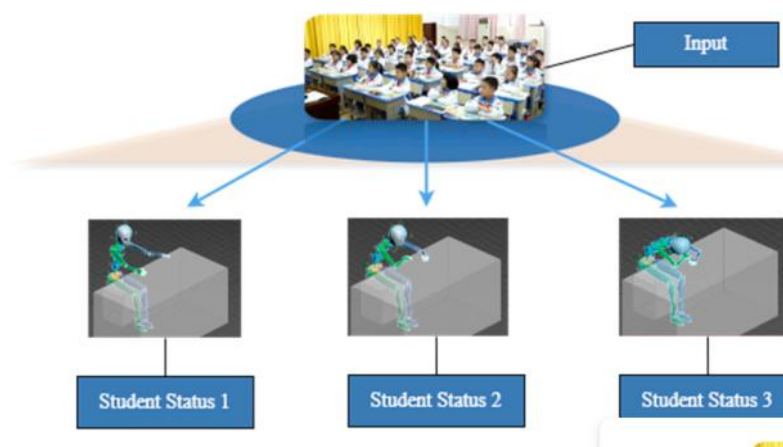


Figura 15 - Classificação de Status Comportamentais em Sala de Aula. (LIU; JIANG; JIANG, 2025)

A partir dos pontos corporais detectados, é possível derivar informações semânticas que descrevem comportamentos relevantes ao processo de ensino. Segundo Yuvaraj, Prince e Murugappan (2024), a análise temporal da pose possibilita distinguir entre comportamentos de curto prazo como levantar a mão ou mudar de direção do olhar e padrões persistentes como desatenção prolongada ou inquietação recorrente. Essa capacidade de segmentar comportamentos ao longo do tempo possibilita intervenções educacionais mais embasadas e sensíveis às necessidades dos estudantes.

Na plataforma proposta, cada ponto detectado funciona como uma Região de Interesse (ROI) para o algoritmo de análise comportamental. A cabeça é observada para identificar leitura, sonolência ou desatenção; os braços são monitorados para detectar participação ou agitação; a coluna e a inclinação do tronco ajudam a inferir cansaço, desconforto ou engajamento. Assim, a combinação de pose + regras comportamentais fornece uma interpretação rica do ambiente da sala de aula e auxilia no acompanhamento individual dos estudantes.

A Figura 16 complementa essa estrutura ao apresentar um fluxo completo de detecção de comportamento, desde a captura das imagens até a geração dos esqueletos e posterior reconhecimento de ações. Esse pipeline demonstra como diferentes módulos detecção, estimativa de pose e classificação atuam de maneira integrada para identificar estados como sonolência, atenção, leitura ou uso de celular, reforçando o potencial da abordagem para análise pedagógica contínua.

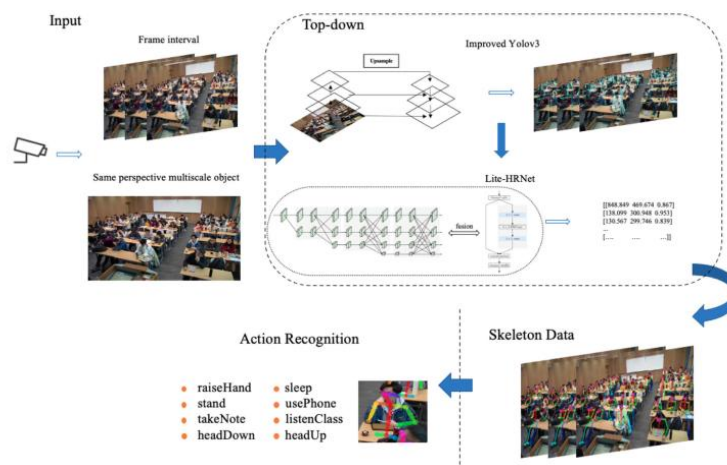


Figura 16 - Fluxo de Detecção de Comportamento em Sala de Aula. (LIU, 2023).

A utilização da pose humana como base para a análise comportamental também reduz ambiguidades presentes em classificadores puramente visuais. Por exemplo, um aluno inclinando o rosto para ler um livro pode ser erroneamente classificado como “distraído” em plataformas que não analisam articulações corporais. A HPE permite diferenciar leitura intencional de desatenção, oferecendo maior precisão contextual. Estudos recentes reforçam que essa abordagem estruturada aumenta significativamente a confiabilidade de plataformas de monitoramento escolar (ZHANG et al., 2024; LIU; JIANG; JIANG, 2025).

Por fim, a integração entre estimativa de pose, reconhecimento facial e análise temporal cria um ecossistema capaz de mapear comportamentos de forma contínua e personalizada. Esse tipo de técnica, quando aplicado com responsabilidade ética e alinhado à LGPD, pode apoiar professores e equipes pedagógicas na identificação de dificuldades, no acompanhamento de padrões de engajamento e na promoção de ambientes de aprendizagem mais inclusivos e responsivos.

2.4 Reconhecimento Facial e Expressões no Contexto Educacional

O reconhecimento facial aplicado ao ambiente educacional tem ganhado relevância à medida que escolas e instituições buscam ferramentas capazes de automatizar processos, analisar o engajamento dos alunos e apoiar práticas pedagógicas baseadas em evidências. Em contraste com métodos tradicionais de acompanhamento, a visão computacional permite identificar estudantes individualmente, mapear padrões de comportamento e inferir estados

emocionais de forma contínua, não intrusiva e altamente escalável (SINGH; GURURANI; BHATNAGAR, 2023).

Além da identificação, o reconhecimento de expressões faciais tem sido utilizado como indicador complementar de envolvimento cognitivo. Expressões como interesse, frustração, surpresa ou confusão podem refletir o estado emocional do estudante, auxiliando professores e plataformas inteligentes a compreenderem em quais momentos a aprendizagem está ocorrendo com maior ou menor fluidez. Embora a interpretação emocional não deva ser vista como diagnóstica, modelos modernos combinam expressões faciais, pose humana e análise temporal para identificar padrões mais robustos (YUVARAJ; PRINCE; MURUGAPPAN, 2024; LIN et al., 2023).

Figura 17 apresenta uma arquitetura multimodal de monitoramento comportamental que sintetiza o funcionamento de plataformas avançadas utilizadas em pesquisas atuais. Nesse modelo, múltiplas fontes de dados como câmeras individuais, câmera central e sensores vestíveis são combinadas para captar informações complementares sobre expressão facial, postura corporal, movimentos e sinais fisiológicos dos estudantes. Cada fluxo de dados é processado por módulos especializados de aprendizado de máquina, responsáveis por estimar direção da cabeça, emoções, pose 2D e atividades comportamentais. Em seguida, essas estimativas são integradas em um modelo profundo, que produz indicadores consolidados exibidos ao professor por meio de uma interface visual. Essa abordagem evidencia o potencial da visão computacional aliada a sensores embarcados para oferecer uma compreensão mais completa, contínua e contextualizada do comportamento discente em sala de aula.

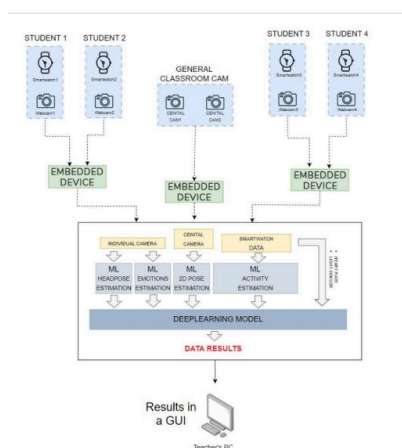


Figura 17 - Arquitetura multimodal de monitoramento comportamental, integrando câmeras individuais, câmera central, sensores fisiológicos e modelos de *deep learning* para estimar postura, emoções e atividades dos estudantes. Fonte: Adaptado de Márquez-Carpintero et al 2023.

A integração entre reconhecimento facial, expressões e pose humana permite uma leitura mais completa do comportamento discente. Expressões faciais isoladas podem ser ambíguas, mas, quando combinadas com inclinação corporal e movimentos das mãos, ajudam a distinguir estados como atenção concentrada, leitura profunda, desatenção prolongada, sonolência ou agitação. Essa análise multimodal aumenta a precisão interpretativa e oferece uma perspectiva mais rica sobre a dinâmica da sala de aula (ZHANG et al., 2024).

A Figura 18 ilustra a arquitetura proposta por Zhang et al. (2024) para reconhecimento automático de comportamentos em sala de aula. Partindo de quadros de vídeo capturados em baixa resolução, o modelo aplica um módulo de super-resolução (SRGAN) para melhorar a nitidez dos estudantes e, em seguida, utiliza uma versão aprimorada do YOLOv8 para detectar ações como levantar a mão, ler, escrever, usar o celular, manter a cabeça baixa ou apoiar-se sobre a carteira. (ZHANG et al., 2024).

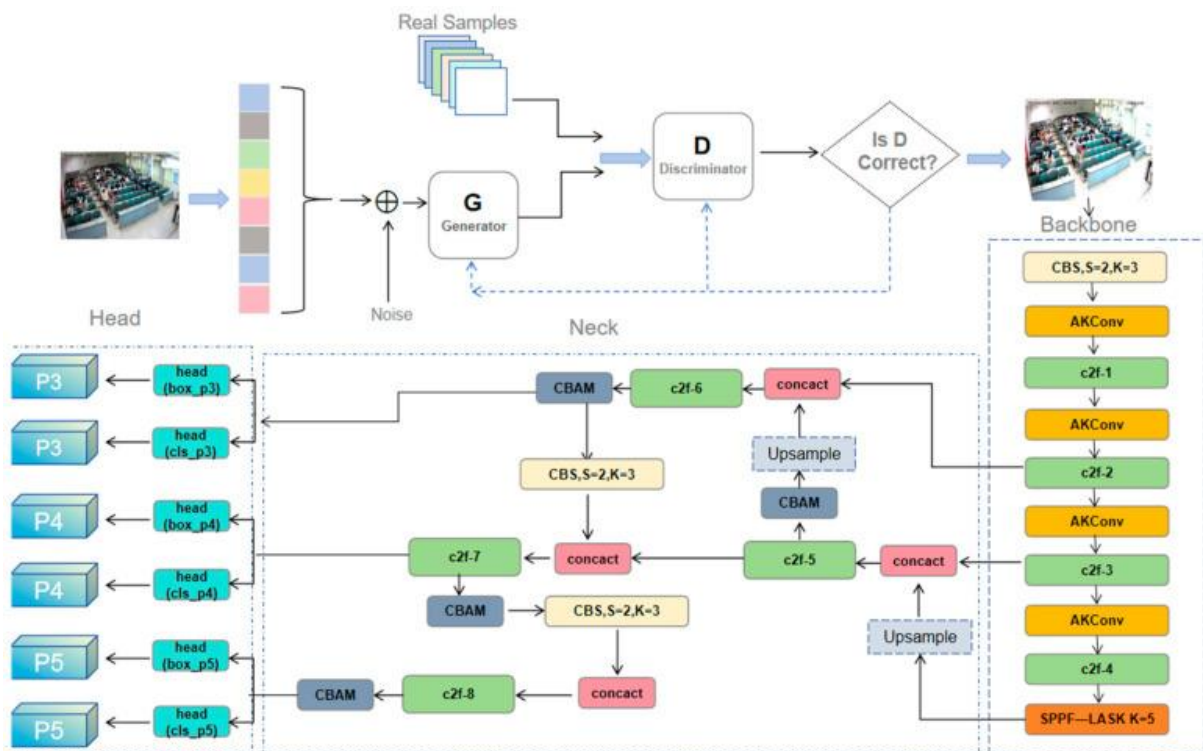


Figura 18 - Estrutura de rede para reconhecimento de comportamentos em sala de aula, incorporando super-resolução (SRGAN) e um YOLOv8 aprimorado para detectar ações como levantar a mão, leitura, escrita, uso de celular, cabeça baixa e apoio sobre a mesa. Fonte: Zhang et al., 2024.

No entanto, o uso educacional dessa tecnologia exige atenção aos aspectos éticos e legais. Uma vez que dados faciais são considerados sensíveis pela LGPD, é necessário aplicar práticas de minimização de dados, consentimento claro e uso de *embeddings* irrecuperáveis.

Ao transformar rostos em vetores matemáticos que não podem ser revertidos para imagens originais, plataformas como o *InsightFace* reduzem riscos de exposição e tornam o processo mais seguro e alinhado às boas práticas de proteção de dados (DENG et al., 2019; WANG, 2023).

Nesse contexto, o *InsightFace* destaca-se como uma das bibliotecas mais robustas para reconhecimento facial baseado em *embeddings* irreversíveis. Desenvolvido por Deng et al. (2019), o *framework* implementa redes neurais profundas combinadas com funções de perda angulares, como a *ArcFace*, capazes de extrair vetores altamente discriminativos e resilientes a variações de iluminação, pose e expressão. A *ArcFace Loss* introduz uma margem angular no espaço de características, aumentando a separabilidade entre identidades e fortalecendo a precisão da plataforma, especialmente em bases de grande escala.

A Figura 19 ilustra o fluxo de processamento empregado pelo *InsightFace*, desde a entrada da imagem até a etapa de classificação final, incluindo extração de características convolucionais, normalização em lote, regularização da distribuição e aplicação da *ArcFace Loss*. Essa arquitetura consolidou-se como referência em sistemas biométricos e aplicações educacionais, dada sua capacidade de gerar *embeddings* consistentes, comparáveis e alinhados às diretrizes de proteção de dados.

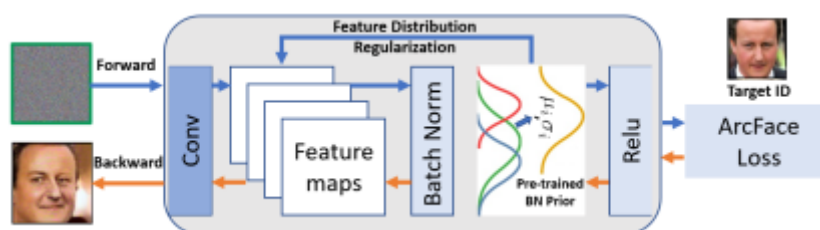


Figura 19 - Arquitetura do *InsightFace* com *ArcFace Loss*. (DENG, 2019).

Em arquiteturas modernas de reconhecimento facial, a similaridade cosseno torna-se o critério padrão de comparação, pois modelos treinados com perdas angulares projetam os *embeddings* na hipersfera, maximizando a separação angular entre identidades. O *ArcFace*, por sua vez, introduz uma margem angular aditiva no *softmax*, ampliando a discriminabilidade entre classes e favorecendo decisões baseadas em ângulo (cosseno) durante a inferência. Essa formulação apresentou desempenho superior em bases públicas de

larga escala, consolidando o *ArcFace* como estado da arte em verificação e identificação facial (DENG, 2019).

Por outro lado, quando o treinamento segue o paradigma do *triplet loss*, como no *FaceNet*, o objetivo é mapear imagens diretamente para um espaço euclidiano no qual as distâncias refletem a similaridade entre indivíduos. Nesses cenários, a métrica L2 torna-se a escolha mais apropriada para a verificação. Assim, a métrica de comparação deve estar alinhada ao critério de otimização utilizado: similaridade cosseno para perdas angulares (*ArcFace*, *CosFace*) e distância L2 para *embeddings* treinados com triplet loss (SCHROFF; KALENICHENKO; PHILBIN, 2015).

Para reforçar visualmente a distinção entre as funções de perda empregadas em reconhecimento facial, a Figura 20 apresenta a formulação geométrica das principais abordagens angulares. Nela observa-se como o *ArcFace* adiciona uma margem angular explícita entre classes, resultando em maior separação entre identidades no espaço de *embeddings* quando comparado ao *softmax* tradicional, *SphereFace* e *CosFace*. Essa regularização angular explica o desempenho superior obtido por modelos treinados com *ArcFace* em cenários de larga escala (DENG et al., 2019).

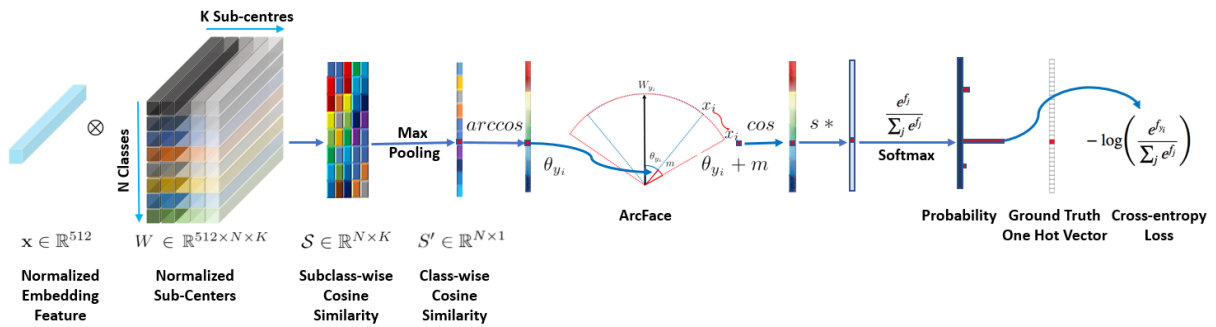


Figura 20 - Representação geométrica da margem angular introduzida pelo *ArcFace* em comparação a outras perdas de reconhecimento facial. Fonte: (Softmax, *SphereFace* e *CosFace*).

2.4.1. Fundamentos das Redes Neurais Convolucionais (CNNs) para Extração de *Embeddings*

As Redes Neurais Convolucionais (CNNs) constituem a base dos sistemas modernos de reconhecimento facial e são amplamente utilizadas em arquiteturas consolidadas como *VGGFace2*, *ResNet* e *ArcFace* (PARKHI; VEDALDI; ZISSERMAN, 2015; HE et al., 2016; DENG et al., 2019). Essas redes modelam, de forma aproximada, a capacidade humana de percepção visual ao extrair características hierárquicas da imagem. As primeiras camadas

identificam estruturas simples, como bordas e texturas, enquanto camadas mais profundas capturam padrões complexos relacionados à geometria facial tridimensional, tornando a representação mais robusta a variações de iluminação, posição e expressão.

O processamento inicial ocorre nas camadas convolucionais, responsáveis por aplicar filtros que detectam padrões locais da imagem. Após essa etapa, mecanismos como a *Batch Normalization* reduzem variações internas na distribuição dos dados e estabilizam o fluxo de gradientes, o que acelera e torna o treinamento mais eficiente (IOFFE; SZEGEDY, 2015). Na parte final da rede, camadas totalmente conectadas transformam o mapa de características resultante em um vetor de dimensão fixa, denominado embedding facial, que serve como base matemática para tarefas de identificação e verificação de identidade.

A evolução das CNNs foi marcada pela introdução dos blocos residuais por He et al. (2016). Esses blocos utilizam conexões de atalho que permitem à rede aprender mapeamentos mais estáveis, evitando problemas relacionados ao desaparecimento do gradiente em arquiteturas profundas. Esse avanço possibilitou a criação de redes com centenas de camadas sem perda de desempenho durante o treinamento e tornou a *ResNet* uma das arquiteturas mais utilizadas para extração de *embeddings* faciais em larga escala. A figura 21 a seguir ilustra a estrutura de um bloco residual, componente fundamental para compreender o funcionamento de modelos modernos empregados no reconhecimento facial.

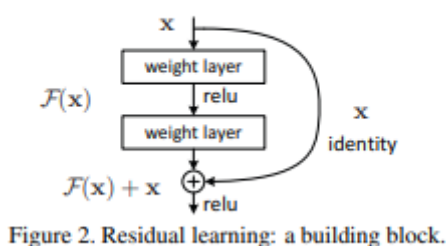


Figura 21 - Estrutura de um bloco residual utilizado na arquitetura *ResNet*, com conexões de atalho que permitem o treinamento de redes profundas para extração de embeddings faciais. Fonte: Adaptado de He et al. (2016).

Com a utilização desses blocos e de redes profundas, os *embeddings* faciais produzidos se tornam representações matemáticas altamente discriminativas em espaços de

alta dimensionalidade. Em modelos treinados com perdas angulares, como *ArcFace* e *CosFace*, esses *embeddings* são projetados sobre uma hipersfera, o que aumenta a separação angular entre identidades e aprimora a capacidade discriminativa do sistema (WANG et al., 2018; DENG et al., 2019). Essa abordagem reduz a sensibilidade a fatores como movimentos naturais, variações de iluminação e expressões faciais, condições típicas do ambiente escolar, garantindo maior consistência na identificação dos estudantes.

2.4.2. ArcFace Loss: O Algoritmo de Margem Angular Aditiva

O *ArcFace Loss* representa um dos avanços mais significativos no reconhecimento facial de última geração. Proposto por Deng et al. (2019), o método introduz uma margem angular aditiva no espaço de características, fazendo com que a decisão de classificação dependa diretamente do ângulo entre o *embedding* facial e o vetor de peso correspondente à classe. Esse ajuste geométrico aumenta a separabilidade entre identidades, produzindo fronteiras de decisão mais nítidas e altamente discriminativas.

Diferentemente do *softmax* tradicional, que opera essencialmente sobre intensidades de ativação, o *ArcFace* normaliza tanto os *embeddings* quanto os vetores de peso, projetando-os em uma hipersfera unitária. A partir dessa normalização, a identificação passa a ser governada exclusivamente por distâncias angulares, reduzindo a sensibilidade a variações de iluminação, pose ou expressão. Essa formulação também acelera a convergência durante o treinamento e melhora a generalização em bases de larga escala.

A introdução da margem angular resulta em maior distância entre classes no espaço vetorial, o que reduz ambiguidades entre identidades visualmente semelhantes. Abaixo podemos verificar a equação 1 que ilustra a função geométrica do *ArcFace* e sua comparação com outras perdas angulares, como *SphereFace* e *CosFace*, evidenciando o impacto da margem aditiva no aumento da separação entre classes.

$$L = -\frac{1}{N} \sum_{i=1}^N \log \left(\frac{e^{s \cdot \cos(\theta_{yi}) + m}}{e^{s \cdot \cos(\theta_{yi}) + m} + \sum_{j \neq y_i} e^{s \cdot \cos(\theta_j)}} \right) \quad (1)$$

Onde:

- N é o número de amostras de treinamento.
- s é o fator de escala (*scale factor*).
- θ_{yi} é o ângulo entre o vetor de características da amostra x_i e o peso W_{yi} da classe correta.

- m é a margem angular aditiva, que força uma maior distância angular entre as identidades, tornando os *embeddings* mais discriminativos.

Essa formulação garante que identidades diferentes apresentem maior distância angular entre si, tornando a plataforma mais robusta para aplicação em ambientes reais, como escolas, nas quais iluminação, expressões e micro variações faciais dificultam a identificação contínua dos estudantes.

A Figura 22 apresenta a arquitetura típica utilizada nos experimentos do *ArcFace*, mostrando como o *embedding* é calculado, normalizado e posteriormente classificado com a margem angular aditiva.

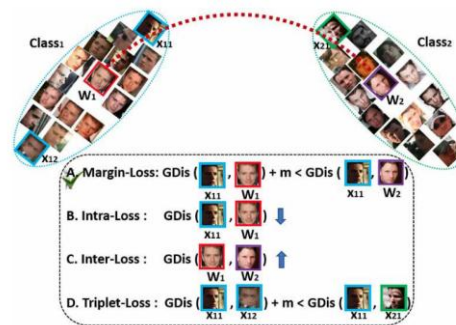


Figura 22 - Comparação visual entre diferentes funções de perda aplicadas ao reconhecimento facial, destacando a separação intra e interclasse e o efeito da margem angular. Fonte: Adaptado de Deng et al. (2019).

2.5 Modelo de Fusão Espaço-Temporal para Reconhecimento de Comportamento

O uso de modelos de fusão espaço-temporal tem se mostrado fundamental para aprimorar o reconhecimento de comportamentos em sala de aula. Segundo (WANG, 2023), a integração simultânea das informações espaciais, obtidas pela análise da posição e postura dos estudantes, com dados temporais, que capturam a sequência e a persistência desses comportamentos, permite identificar padrões recorrentes com maior precisão.

Essa abordagem não apenas reconhece eventos isolados, mas também mapeia tendências ao longo do tempo, o que é especialmente útil para análises mais assertivas. Por exemplo, ao detectar repetidamente períodos de desatenção de um aluno sempre no mesmo horário ou dia da semana, é possível estabelecer uma relação mais robusta entre o comportamento observado e possíveis causas externas ou internas, como fadiga, dificuldades cognitivas ou problemas de atenção.

Ao considerar múltiplas ocorrências e seu contexto temporal, o modelo descrito por (WANG, 2023) reduz a influência de ruídos e casos pontuais, fortalecendo a confiabilidade das informações fornecidas aos educadores e especialistas. Assim, a análise espaço-temporal não apenas amplia a acurácia técnica, mas também oferece *insights* valiosos para intervenções pedagógicas direcionadas.

2.6. Arquitetura de Dados, *Embeddings* e Proteção da Privacidade (LGPD)

O processamento de informações em plataformas de monitoramento comportamental baseados em visão computacional exige uma arquitetura de dados capaz de armazenar, consultar e proteger tanto registros históricos quanto vetores de características gerados em tempo real. *Embeddings* faciais, por serem representações matemáticas de alta dimensionalidade, demandam bancos de dados que ofereçam suporte eficiente a operações vetoriais, consultas por similaridade e escalabilidade horizontal.

O PostgreSQL destaca-se como uma das plataformas mais avançadas para armazenamento vetorial, especialmente após a introdução da extensão *pgvector*, que permite armazenar *embeddings* como vetores n-dimensionais e realizar operações como similaridade por cosseno, distância euclidiana e distância Manhattan diretamente no banco (ZAKIROV, 2023). Essas funcionalidades tornam o PostgreSQL adequado para aplicações em reconhecimento facial, busca semântica, recomendação e sistemas de recuperação de informação.

Além disso, o PostgreSQL oferece mecanismos avançados de segurança, incluindo controle de acesso baseado em função (RBAC), criptografia em trânsito (TLS), criptografia em repouso e monitoramento de acesso, características que o tornam alinhado às exigências da Lei Geral de Proteção de Dados (LGPD). No contexto educacional, onde dados sensíveis são manipulados, essas garantias são fundamentais para minimizar riscos operacionais e jurídicos.

Do ponto de vista da privacidade, a utilização de *embeddings* faciais reduz significativamente a exposição de dados sensíveis, uma vez que os vetores não podem ser revertidos para reconstruir o rosto original (DENG et al., 2019). Esse processo caracteriza uma forma de pseudonimização, que mitiga riscos em caso de acesso indevido e atende ao princípio da minimização de dados previsto na LGPD (BRASIL, 2018). Assim, a arquitetura

final combina três pilares essenciais: armazenamento vetorial eficiente, escalabilidade para operações em tempo real e proteção rigorosa da privacidade dos estudantes.

2.6.1. Bancos de Dados Relacionais com Suporte a Vetores (PostgreSQL e pgvector)

O uso de bancos de dados tradicionais é insuficiente para aplicações que dependem da comparação eficiente de *embeddings* de alta dimensionalidade. O PostgreSQL, aliado à extensão *pgvector*, tornou-se uma das soluções mais utilizadas em plataformas modernas de reconhecimento facial, permitindo armazenamento vetorial nativo e busca por similaridade com métricas otimizadas (ZAKIROV, 2023).

Essa estrutura é essencial porque cada rosto é convertido em um vetor numérico que representa características faciais discriminativas aprendidas por modelos como *FaceNet*, *ArcFace* e *InsightFace* (SCHROFF; KALENICHENKO; PHILBIN, 2015; DENG et al., 2019; GUO et al., 2021). A identificação ocorre comparando o *embedding* recém-capturado com *embeddings* de referência armazenados no banco de dados.

O *pgvector* implementa as principais métricas utilizadas pela literatura para mensurar a proximidade entre dois *embeddings*, como a distância Euclidiana e a similaridade de cosseno, que refletem diferentes perspectivas geométricas desse espaço vetorial.

1. Distância Euclidiana (mede a proximidade geométrica):

$$d(A, B) = \sqrt{\sum_{i=1}^n (A_i - B_i)^2} \quad (2)$$

A distância Euclidiana (Equação 2) mede a proximidade geométrica entre dois *embeddings* e foi amplamente empregada em arquiteturas como o *FaceNet* (SCHROFF; KALENICHENKO; PHILBIN, 2015).

Onde:

- A = vetor de *embedding* da imagem consultada (recém-capturada).
- B = vetor de *embedding* armazenado no banco de dados como referência.
- A_i, B_i = componentes individuais dos vetores A e B na dimensão i.
- n = dimensionalidade do *embedding* (geralmente 128, 256 ou 512).
- $d(A,B)$ = distância Euclidiana entre os *embeddings*, onde valores menores indicam maior similaridade.

2. Similaridade de Cosseno (mede o ângulo entre os vetores):

$$\text{Similaridade de Cosseno}(A, B) = \frac{A \cdot B}{|A||B|} \quad (3)$$

A Similaridade de Cosseno (Equação 3) é utilizada em modelos baseados em margens angulares, como o *ArcFace*, sendo a métrica padrão em sistemas de *embeddings* normalizados (DENG et al., 2019).

Onde:

- A = embedding da imagem consultada.
- B = embedding de referência armazenado no banco de dados.
- $A \cdot B$ = produto escalar entre os vetores A e B.
- $\|A\|, \|B\|$ = norma (magnitude) dos vetores A e B.
- θ = ângulo entre os dois *embeddings* no espaço vetorial.
- $\cos(\theta)$ = grau de similaridade angular, onde valores próximos de 1 indicam maior compatibilidade entre *embeddings*.

Quando o modelo é treinado com perdas angulares, o espaço de *embeddings* é normalizado, fazendo da similaridade de cosseno a métrica mais adequada.

O reconhecimento ocorre quando a medida entre vetores atinge um limiar pré-definido, associando o embedding analisado ao aluno correspondente.

2.6.2. Armazenamento de *Embeddings* e Conformidade com a LGPD

O tratamento de dados biométricos em plataformas de monitoramento educacional baseados em visão computacional exige conformidade rigorosa com os princípios estabelecidos pela Lei Geral de Proteção de Dados (LGPD). Por serem classificados como dados sensíveis, as informações faciais demandam salvaguardas robustas que garantam confidencialidade, integridade e proteção contra qualquer forma de identificação indevida dos estudantes (BRASIL, 2018; DONEDA; MENDES, 2021). Nessa perspectiva, a forma como os dados são armazenados constitui um dos elementos centrais para assegurar a segurança e a legitimidade da plataforma.

Nos modelos contemporâneos de reconhecimento facial, o armazenamento exclusivo do *embedding* tem sido reconhecido como uma das práticas mais eficazes para a mitigação de riscos. O *embedding* é uma representação vetorial gerada por redes neurais profundas, contendo apenas características matemáticas necessárias para a distinção entre identidades.

Embora altamente discriminativo para fins de reconhecimento, esse vetor não permite a reconstrução da imagem facial original, o que configura uma forma de pseudonimização compatível com as exigências da LGPD e com diretrizes internacionais de proteção de dados biométricos, como a norma ISO/IEC 24745:2022 (DENG et al., 2019; ISO, 2022).

A adoção dessa abordagem é reforçada pela literatura especializada, que evidencia que a substituição do armazenamento de imagens por templates biométricos reduz significativamente a superfície de risco em caso de acesso indevido (JAIN; ROSS; NANDAKUMAR, 2021). Além disso, quando associada a técnicas complementares de segurança, como criptografia, controle de acesso e segregação por perfil de privilégio, a pseudonimização contribui para atender ao princípio da minimização de dados previsto na LGPD, assegurando que apenas os elementos estritamente necessários ao funcionamento da plataforma sejam conservados.

A Figura 23 apresenta um fluxo seguro de processamento e armazenamento de dados faciais utilizado em sistemas contemporâneos que buscam conciliar reconhecimento facial e proteção da privacidade. O diagrama mostra as etapas de registro e verificação de identidade, nas quais a imagem capturada passa pela extração das características faciais e pela geração do *embedding*, que em seguida é criptografado antes do envio para o ambiente de armazenamento. Os *templates* cifrados são então utilizados na etapa de comparação, garantindo que todo o processo ocorra sem a exposição direta das imagens originais. Essa arquitetura demonstra como práticas de pseudonimização e criptografia homomórfica podem reduzir riscos, atender às exigências da LGPD e assegurar que apenas informações matematicamente transformadas sejam manipuladas ao longo do fluxo de identificação.

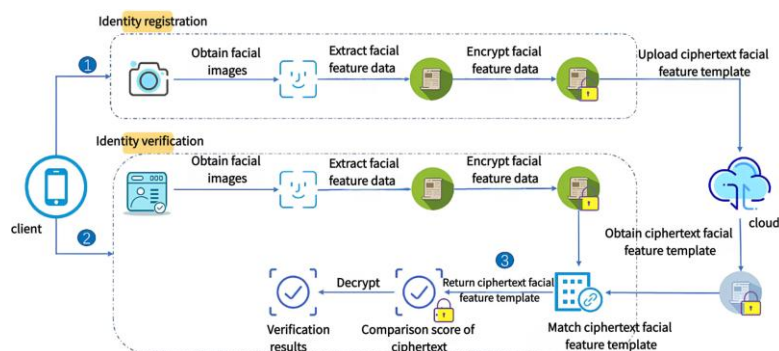


Figura 23 - Fluxo de registro e verificação com extração de *embeddings*, criptografia homomórfica e armazenamento seguro em nuvem. Fonte (Song et al. 2025).

3 METODOLOGIA

A metodologia adotada nesta pesquisa foi estruturada para garantir rigor científico, reprodutibilidade e aplicabilidade prática no contexto educacional. Trata-se de um estudo aplicado, com abordagem quantitativa e caráter experimental, cujo objetivo é validar uma plataforma de monitoramento comportamental de estudantes baseado em técnicas avançadas de visão computacional e inteligência artificial.

O processo metodológico contempla todas as etapas do ciclo de vida dos dados, desde a coleta contínua em tempo real por meio de câmeras instaladas na sala de aula até o processamento, armazenamento e análise computacional das informações registradas. Para a detecção e interpretação do comportamento discente, são utilizados dois modelos de referência: *YOLOv11-Pose*, empregado na estimativa de pose corporal, e *InsightFace*, responsável pelo reconhecimento facial e pela geração de *embeddings* discriminativos de cada aluno.

A arquitetura de dados da plataforma emprega o PostgreSQL em conjunto com a extensão *pgvector*, permitindo o armazenamento eficiente de vetores de características e a realização de buscas por similaridade, essenciais para a etapa de identificação facial. Essa infraestrutura garante escalabilidade, consultas vetoriais otimizadas e suporte a operações em tempo real, requisitos fundamentais para o funcionamento da plataforma em ambiente escolar.

A interface interativa foi desenvolvida em *Streamlit*, integrando captura de vídeo, operação dos modelos de IA, visualização dos resultados e registro automático dos comportamentos detectados durante as aulas. Toda a operação da plataforma é monitorada continuamente, permitindo avaliar o desempenho sob condições reais de uso.

A validação da solução será conduzida em ambiente educacional real, com análise de métricas como acurácia, precisão, *recall*, latência, estabilidade temporal das inferências e taxas de falsos positivos e negativos. Além do desempenho técnico, o processo metodológico considera diretrizes éticas e requisitos legais, garantindo o armazenamento seguro dos *embeddings* e conformidade com a LGPD, especialmente no que se refere à manipulação de dados biométricos sensíveis.

Esse delineamento metodológico busca não apenas validar a tecnologia proposta, mas também oferecer uma ferramenta prática de apoio pedagógico, auxiliando professores na compreensão de padrões comportamentais relevantes ao processo de ensino-aprendizagem.

3.1 Coleta de imagens

3.1.1 Câmera

A coleta de imagens constitui uma etapa fundamental no desenvolvimento da plataforma de monitoramento comportamental, pois fornece os dados visuais necessários para o reconhecimento facial e para a análise de posturas corporais dos alunos. Para essa finalidade, foi empregada uma câmera Hikvision modelo DS-2CD1123G0E-IC, instalada estrategicamente na sala de aula com o objetivo de garantir ampla cobertura do ambiente e captação estável dos fluxos de vídeo.

O dispositivo apresenta resolução de 2 megapixels e utiliza iluminadores infravermelhos (IR) com alcance aproximado de 30 metros, permitindo registrar imagens mesmo sob condições reduzidas de iluminação. Esses recursos asseguram maior consistência visual à plataforma de inteligência artificial, evitando perda de desempenho durante transições de luminosidade ao longo do dia. Além disso, o equipamento opera com baixo consumo energético, cerca de 5W, o que favorece sua utilização contínua em ambientes escolares.

Outro aspecto relevante é a robustez estrutural oferecida pelo equipamento, incluindo mecanismo antivandalismo e construção projetada para suportar variações ambientais típicas de salas de aula. Essas características contribuem para a estabilidade operacional da câmera e para a confiabilidade das imagens obtidas, garantindo que os modelos de IA utilizados na plataforma, como *InsightFace* para reconhecimento facial e *YOLOv11-Pose* para estimativa de pose, recebam dados adequados para realizar inferências com maior precisão.

No contexto da pesquisa, a qualidade das capturas fornecidas pela câmera é determinante para o desempenho geral da plataforma, pois imagens nítidas e consistentes reduzem erros de detecção, minimizam ruídos no processo de extração de *embeddings* e aumentam a acurácia das análises comportamentais subsequentes.

A coleta de imagens foi realizada em ambiente controlado, com foco na demonstração do funcionamento da plataforma. O objetivo não foi a constituição de uma base de dados em larga escala, mas sim a validação técnica dos módulos de reconhecimento e detecção comportamental. O uso das imagens respeita os princípios da Lei Geral de Proteção de Dados (LGPD), priorizando a minimização de dados e a utilização de representações vetoriais (embeddings) em vez de imagens brutas.



Figura 24 - Câmera Hikivision. Fonte: (HIKVISION, 2025).

3.2 Software

O monitoramento comportamental foi implementado e validado utilizando uma GPU NVIDIA RTX 3080 Ti, com 12 GB de memória dedicada, operando em um sistema Ubuntu 22.04 LTS. Essa configuração apresentou ser suficiente para processar fluxos de vídeo em tempo real, evidenciando que a solução proposta pode ser aplicada no ambiente escolar sem necessidade de infraestrutura de alto custo. A taxa de processamento obtida durante os experimentos garantiu estabilidade na detecção de poses e na identificação facial, permitindo que a plataforma acompanhe continuamente as interações dos alunos durante as aulas.

Os testes laboratoriais foram realizados também em um cluster composto por cinco GPUs NVIDIA RTX 3070 Ti, cada uma com 8 GB de VRAM. Esse conjunto adicional foi utilizado apenas para ensaios de comparação de desempenho, não sendo indispensável para os resultados apresentados nesta pesquisa. Dessa forma, a implementação prática em escolas pode ocorrer com uma única GPU de categoria intermediária, o que garante maior acessibilidade, escalabilidade e menor investimento inicial.

Para maximizar o desempenho dos algoritmos de visão computacional e das redes neurais profundas, foram utilizadas as bibliotecas *CUDA Toolkit* e *cuDNN*, fornecidas pela NVIDIA. Essas ferramentas possibilitam o processamento paralelo em GPU, reduzindo significativamente o tempo de inferência. Além disso, o *TensorRT* foi empregado para otimizar a execução dos modelos em tempo real, resultando em maior eficiência computacional e menor latência durante a análise dos comportamentos dos alunos (NVIDIA, 2023).

Tabela 1- Especificações do ambiente de processamento.

Componente	Especificação
Sistema Operacional	Ubuntu 22.04 LTS
GPU Utilizada	NVIDIA RTX 3080 Ti – 12GB VRAM
GPUs Adicionais	5 x NVIDIA RTX 3070 Ti – 8 GB VRAM
Frameworks	CUDA 12, cuDNN, TensorRT
Bibliotecas	<i>YOLOv11-Pose, InsightFace, Streamlit, SQLite</i>

Fonte: (Autor, 2025).

Para avaliar o desempenho da plataforma proposta em cenários de maior demanda computacional, foram realizados testes utilizando uma infraestrutura de processamento baseada em múltiplas unidades de processamento gráfico (GPUs). Esse ambiente permitiu explorar estratégias de paralelização para execução simultânea dos modelos de reconhecimento facial e estimativa de pose, ampliando a capacidade de processamento de fluxos de vídeo em larga escala. A utilização de clusters com múltiplas GPUs é uma abordagem comum em aplicações de visão computacional e aprendizado profundo, especialmente em tarefas que exigem alta capacidade de cálculo e processamento contínuo de dados visuais. A Figura 25 apresenta o cluster utilizado nos experimentos desta pesquisa, composto por várias GPUs dedicadas ao processamento paralelo dos modelos de inteligência artificial empregados na plataforma.



Figura 25 - Cluster com múltiplas GPUs utilizado nos testes de processamento em larga escala.
Fonte: (Autor, 2025).

Neste contexto, a escolha dos softwares utilizados na pesquisa baseou-se na necessidade de integrar, em um ambiente unificado, modelos de detecção de poses humanas e reconhecimento facial capazes de operar em tempo real. Para a estimativa de poses corporais, adotou-se o *YOLOv11-Pose*, que oferece alta precisão na identificação dos pontos articulares e na interpretação das ações dos alunos. O reconhecimento facial foi conduzido com o *InsightFace*, que possibilita a extração de *embeddings* robustos e invariantes a variações de iluminação, pose e expressão.

Para a construção da interface interativa, responsável pela visualização dos resultados e operação da plataforma durante o monitoramento, foi utilizado o *framework Streamlit*. Essa ferramenta permitiu o desenvolvimento de dashboards intuitivos, nos quais o fluxo de vídeo, as inferências e os registros comportamentais são apresentados de forma clara e acessível. O gerenciamento e armazenamento das informações foram realizados no PostgreSQL, associado à extensão *pgvector*, que oferece suporte nativo a operações vetoriais essenciais para a comparação de *embeddings* faciais.

O conjunto desses softwares compõe a base tecnológica da plataforma de monitoramento comportamental, possibilitando a integração entre captura de vídeo, processamento computacional em tempo real, registro estruturado das detecções e visualização dinâmica dos resultados. Essa arquitetura garante eficiência operacional, escalabilidade e aderência às exigências de privacidade e segurança necessárias ao contexto educacional.

3.2.1 Detecção do Comportamento

A detecção do comportamento dos alunos foi realizada por meio de técnicas avançadas de estimativa de pose humana, que permitem identificar e interpretar padrões corporais associados a diferentes estados de atenção, interação e participação em sala de aula. Para essa etapa, utilizou-se o *YOLOv11-Pose*, modelo baseado em redes neurais profundas que combina detecção de objetos com predição dos pontos-chave do corpo humano. Essa abordagem é reconhecida pela alta precisão na localização das articulações e pela capacidade de operar em tempo real, o que a torna adequada para aplicações educacionais dinâmicas (ZHANG et al., 2024).

A partir dos *keypoints* identificados pelo modelo, a plataforma aplica regras geométricas e relações espaciais para classificar comportamentos como atenção, desatenção, leitura, escrita, sonolência, mão levantada e agitação corporal. Estudos recentes demonstram que a estimativa de pose é um método eficiente para interpretar ações humanas em cenários complexos, já que padrões corporais tendem a ser mais estáveis e menos ambíguos do que expressões faciais isoladas (PAPADOPOULOS et al., 2023). Essa estratégia reduz falsos positivos e amplia a precisão na identificação de comportamentos acadêmicos relevantes.

Além das poses corporais, a plataforma integra a identificação facial, realizada pelo modelo *InsightFace*, a fim de associar cada ação detectada a um aluno específico. Essa etapa é fundamental para a construção de séries temporais de comportamento individual, permitindo análises longitudinais e identificação de padrões que auxiliem professores e gestores no entendimento do engajamento dos estudantes ao longo das aulas. A combinação entre reconhecimento facial e estimativa de pose reforça a robustez metodológica da plataforma, conforme proposto em pesquisas que tratam de análise multimodal do comportamento em ambientes educacionais (LI et al., 2023).

A execução da detecção comportamental ocorre de forma contínua, processando o fluxo de vídeo em tempo real e registrando automaticamente cada comportamento identificado no banco de dados PostgreSQL, que armazena *timestamps*, rótulos comportamentais e *embeddings* faciais anonimizado. Essa arquitetura possibilita análises imediatas e retrospectivas, além de garantir integridade e escalabilidade do processo de monitoramento.

Os modelos modernos de estimativa de pose humana são capazes de identificar automaticamente pontos anatômicos do corpo a partir de imagens ou fluxos de vídeo. Esses pontos, conhecidos como *keypoints*, representam articulações e regiões relevantes do corpo humano, como cabeça, ombros, cotovelos, punhos, quadris, joelhos e tornozelos. A partir da identificação dessas coordenadas espaciais, é possível reconstruir a estrutura corporal do indivíduo e analisar padrões de postura e movimento em tempo real. Modelos da família YOLO adaptados para estimativa de pose, como o YOLOv11-Pose, realizam simultaneamente a detecção da pessoa na imagem e a inferência dos pontos-chave do corpo, permitindo aplicações em monitoramento comportamental e análise de atividades humanas. A Figura 26 apresenta um exemplo do processo de detecção de pessoas e identificação dos *keypoints* corporais utilizando o modelo YOLOv11-Pose, evidenciando as caixas delimitadoras (*bounding boxes*) e as conexões que formam o esqueleto corporal estimado pelo modelo (ULTRALYTICS, 2025).

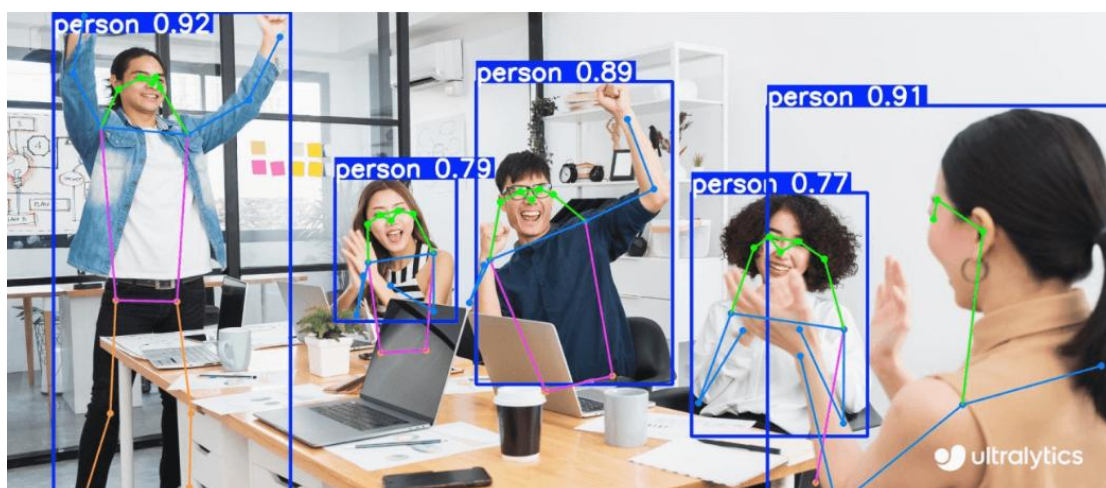


Figura 26 - Detecção dos pontos chave utilizando YOLOv11-Pose. Fonte: (ULTRALYTICS, 2025).

A detecção dos comportamentos é realizada a partir da análise geométrica dos *keypoints* fornecidos pelo modelo YOLOv11-Pose, considerando unicamente a região superior do corpo para reduzir interferências e manter a coerência com o ambiente de sala de aula. A Figura 27 ilustra os cinco comportamentos utilizados pela plataforma e suas respectivas relações espaciais. No comportamento “Atento”, o nariz (ponto 1) permanece alinhado ao centro dos ombros (ponto 2), sem inclinações significativas. O comportamento “Perguntando” é identificado quando o punho direito ou esquerdo (pontos 7 ou 8) se encontra acima da altura do respectivo ombro (pontos 3 ou 4), caracterizando o gesto de levantar a mão. No estado “Aagitado”, ambos os punhos ultrapassam simultaneamente a linha dos ombros, indicando

movimento excessivo dos braços. O comportamento “Dormindo” ocorre quando o nariz está posicionado abaixo dos ombros, evidenciando queda da cabeça em direção à mesa ou aos próprios braços. Por fim, o comportamento “Distraído” é detectado quando o nariz apresenta deslocamento horizontal significativo em relação ao ponto médio entre os ombros, indicando que o aluno está olhando para o lado. Essa modelagem vetorial permite uma interpretação clara e objetiva das posturas corporais, garantindo robustez ao processo de classificação.

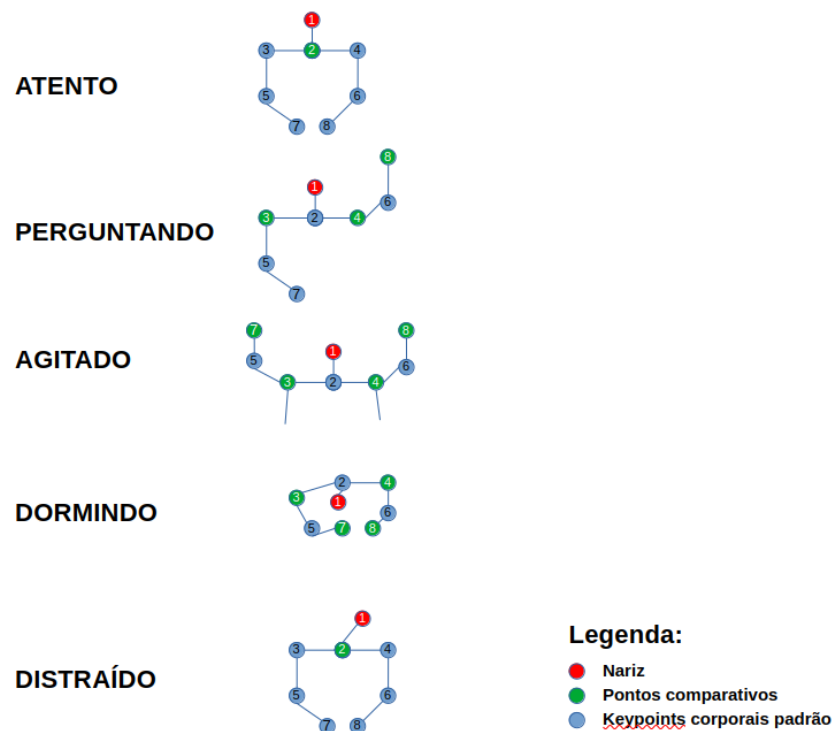


Figura 27 - Representação esquemática dos cinco comportamentos detectados pela plataforma a partir da análise geométrica dos *keypoints* corporais estimados pelo modelo *YOLOv11-Pose*.
Fonte: (Autor, 2025).

De acordo com a (ULTRALYTICS, 2023), os modelos da linha YOLO apresentam variações que se adaptam a diferentes cenários de processamento, desde versões leves para dispositivos de borda até versões robustas para servidores com GPU, oferecendo flexibilidade para aplicações acadêmicas e educacionais conforme a figura 28.

Modelo	tamanho (pixels)	mAP ^{pose} ₅₀₋₉₅	mAP ^{pose} ₅₀	Velocidade CPU ONNX (ms)	Velocidade T4 TensorRT10 (ms)	parâmetros (M)	FLOPs (B)
YOLO11n-pose	640	50.0	81.0	52,4 ± 0,5	1,7 ± 0,0	2.9	7.6
YOLO11s-pose	640	58.9	86.3	90,5 ± 0,6	2,6 ± 0,0	9.9	23.2
YOLO11m-pose	640	64.9	89.4	187,3 ± 0,8	4,9 ± 0,1	20.9	71.7
YOLO11l-pose	640	66.1	89.9	247,7 ± 1,1	6,4 ± 0,1	26.2	90.7
YOLO11x-pose	640	69.5	91.1	488,0 ± 13,9	12,1 ± 0,2	58.8	203.3

Figura 28 - Versões *YOLOv11-Pose*. Fonte: (ULTRALYTICS, 2025).

3.2.2 Processo de Cadastro e Coleta das Imagens dos Alunos

A plataforma conta com um módulo de cadastro de alunos responsável por registrar informações essenciais, como nome, matrícula, turma e disciplina, de modo a garantir a correta associação entre o estudante identificado pela IA e os comportamentos detectados em sala. Nessa etapa, a plataforma também conduz o aluno pelo processo de captura das imagens faciais que serão utilizadas para gerar seus *embeddings*.

Durante o procedimento de captura, a plataforma solicita que o aluno execute poses específicas, incluindo frontal, lateral esquerda, lateral direita e cabeça baixa. Essa variação garante a criação de um conjunto de imagens diversificado, permitindo que o modelo *InsightFace* gere *embeddings* robustos a diferentes ângulos, expressões e condições de iluminação.

O processo de cadastro dos estudantes constitui uma etapa fundamental para o funcionamento da plataforma de monitoramento comportamental, pois é nesse momento que são registradas as informações necessárias para associar cada aluno aos seus respectivos dados faciais e registros comportamentais. Durante esse procedimento, o sistema permite que o usuário informe dados básicos do estudante e realize a captura orientada das imagens faciais, que posteriormente serão utilizadas para gerar os *embeddings* empregados no processo de reconhecimento facial. A Figura 29 apresenta a *interface* de cadastro dos alunos na plataforma desenvolvida, destacando os campos utilizados para registro das informações e configuração da coleta das imagens faciais.

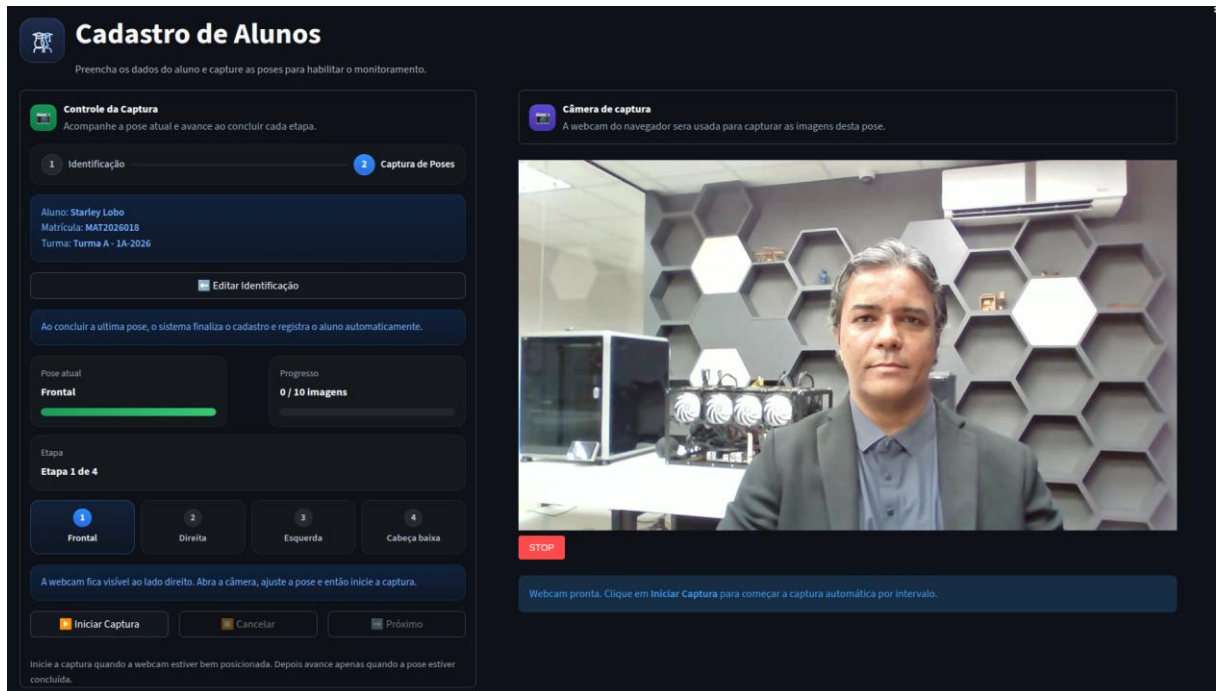


Figura 29 - Tela de cadastro dos alunos na plataforma. Fonte: (Autor, 2026).

3.2.3 Identificação do Aluno

A identificação facial dos alunos constitui uma etapa central no funcionamento da plataforma de monitoramento comportamental, pois permite associar cada comportamento detectado a um indivíduo específico. Essa etapa foi implementada utilizando o *framework InsightFace*, que emprega redes neurais profundas para a extração de *embeddings* faciais altamente discriminativos. O modelo utilizado para essa tarefa foi o *Buffalo-L*, integrante da família de modelos do *InsightFace*, reconhecido por gerar vetores faciais robustos mesmo em cenários com variações de iluminação, pose ou expressão (DENG et al., 2019).

O processo de reconhecimento ocorre em duas fases principais. A primeira consiste na detecção e alinhamento das faces presentes no quadro, realizada pelo modelo *RetinaFace*, responsável por localizar a face e corrigir sua orientação. Em seguida, o *Buffalo-L* extrai o *embedding* facial normalizado, projetando-o em um espaço vetorial de alta dimensionalidade. Esse vetor é então comparado aos registros armazenados no banco PostgreSQL por meio da extensão *pgvector*, que utiliza a métrica de similaridade de cosseno. Essa métrica é particularmente adequada para *embeddings* treinados com perdas angulares, como o *ArcFace*,

pois maximiza a separação entre identidades no espaço vetorial e melhora o desempenho em cenários reais (WANG et al., 2018).

Cada *embedding* previamente registrado está associado ao perfil de um aluno, possibilitando identificar o indivíduo presente no enquadramento da câmera com elevada precisão. Estudos recentes mostram que abordagens baseadas em *embeddings* vetoriais e margens angulares apresentam desempenho superior em ambientes desafiadores, como salas de aula, onde há constante movimentação, variações de postura e oclusões parciais (LI; SUN; ZHOU, 2023). A solidez desse método contribui para reduzir erros de identificação e permite acompanhar a evolução dos comportamentos de cada aluno ao longo das atividades escolares.

Os resultados da identificação facial são integrados ao sistema de detecção comportamental, garantindo que todas as ações reconhecidas, como atenção, leitura, desatenção ou mão levantada, sejam registradas de forma vinculada ao aluno correspondente. O armazenamento dessas informações no banco de dados possibilita análises posteriores, geração de indicadores temporais e avaliação longitudinal do comportamento dos estudantes, apoiando interpretações pedagógicas baseadas em evidências.

O processo de identificação facial implementado na plataforma segue uma sequência estruturada de etapas responsáveis por detectar o rosto do estudante, extrair suas características faciais e realizar a identificação por similaridade vetorial. Inicialmente, o sistema utiliza o modelo *RetinaFace* para realizar a detecção e o alinhamento da face presente na imagem capturada pela câmera. Em seguida, o modelo *Buffalo-L*, pertencente ao *framework InsightFace*, é responsável pela extração do *embedding* facial, convertendo a imagem do rosto em um vetor numérico de alta dimensionalidade que representa suas características discriminativas. Durante o treinamento desses modelos, a função de perda *ArcFace* é utilizada para maximizar a separação angular entre diferentes identidades no espaço vetorial, favorecendo a distinção entre indivíduos. A Figura 30 apresenta o pipeline de identificação facial adotado na plataforma, ilustrando as principais etapas envolvidas desde a detecção da face até a identificação do aluno no sistema.



Figura 30 - Pipeline de identificação facial utilizado no sistema. Fonte: (Autor, 2025).

Além do pipeline resumido apresentado anteriormente, a Figura 31 ilustra o fluxo completo do processo de reconhecimento facial implementado na plataforma, integrando as etapas de captura da imagem, detecção facial, extração do *embedding*, armazenamento vetorial no PostgreSQL por meio da extensão *pgvector* e decisão final baseada em similaridade. Esse fluxograma oferece uma representação detalhada da operação do módulo de reconhecimento facial, permitindo visualizar a interação entre os componentes de software e os mecanismos de comparação vetorial empregados na plataforma.

Fluxo de Reconhecimento Facial (InsightFace + pgvector)

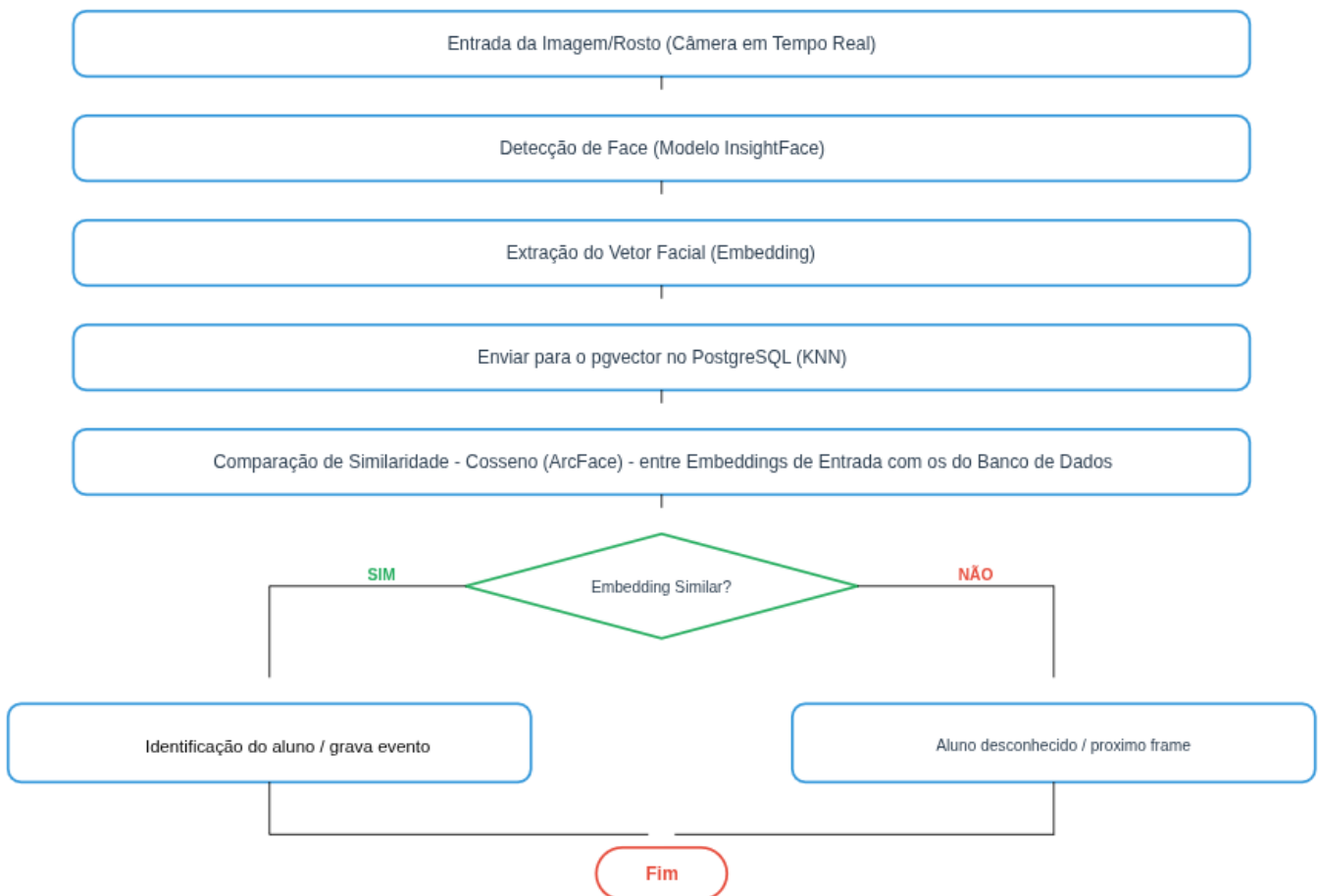


Figura 31- Fluxo de reconhecimento facial – Fonte: (Autor, 2025).

3.2.4 Demonstração das Informações

A etapa de demonstração das informações tem como objetivo apresentar, de forma clara e interpretável, os resultados obtidos pela plataforma de monitoramento

comportamental. Após a identificação facial do aluno e a detecção do comportamento correspondente, os registros são consolidados e organizados em uma interface interativa desenvolvida em *Streamlit*. Essa interface reúne elementos visuais que permitem ao usuário acompanhar, em tempo real, tanto o reconhecimento facial quanto a análise comportamental realizada pelo modelo *YOLOv11-Pose*.

A plataforma exibe simultaneamente o vídeo processado, a identificação do aluno e o comportamento detectado, permitindo ao observador verificar a precisão do modelo e compreender a dinâmica da sala de aula. Cada *frame* analisado gera uma anotação contendo a identidade do aluno, o comportamento classificado, o nível de confiança da predição e o horário exato da ocorrência. Essas informações são apresentadas em tabelas dinâmicas e gráficos atualizáveis, facilitando a análise temporal e a detecção de padrões comportamentais relevantes. A literatura destaca que a visualização estruturada das informações auxilia na interpretação pedagógica e reduz o esforço de análise por parte dos professores (KIM; CHEN; LI, 2024).

Além da visualização em tempo real, a plataforma organiza os registros históricos no banco PostgreSQL, possibilitando consultas posteriores e análises comparativas. Gráficos de linha exibem a evolução de comportamentos ao longo da aula, enquanto gráficos de barras permitem comparar diferentes alunos em termos de participação, atenção ou distração. Essas representações visuais auxiliam na identificação de tendências, como queda de atenção em determinados períodos ou aumento de interações durante atividades específicas. Estudos recentes indicam que dashboards educacionais podem melhorar a tomada de decisão docente, especialmente quando combinam dados comportamentais e indicadores temporais (NIAZI; ZHOU; FAROOQ, 2023).

A demonstração das informações também cumpre um papel metodológico essencial: validar o funcionamento integrado dos módulos de reconhecimento facial e detecção comportamental. A visualização permite identificar inconsistências, ajustar parâmetros e verificar a adequação da plataforma ao ambiente escolar. Dessa forma, a interface não apenas apresenta os resultados, mas contribui para a calibração contínua da ferramenta e para a confiabilidade dos dados utilizados nas análises posteriores.

3.2.5 Registro e Armazenamento Estruturado dos Dados no PostgreSQL

O armazenamento dos dados produzidos pela plataforma é realizado em um banco PostgreSQL, configurado para gerenciar tanto informações tabulares quanto vetores de *embeddings* faciais gerados pelo modelo *InsightFace*. A escolha pelo PostgreSQL deve-se à sua robustez, escalabilidade e compatibilidade com cargas de trabalho que combinam dados estruturados e operações vetoriais. Para esse fim, a plataforma utiliza a extensão *pgvector*, que possibilita o armazenamento de *embeddings* como vetores n-dimensionais, permitindo consultas eficientes baseadas em similaridade, essenciais para o reconhecimento facial em tempo real (ZAKIROV, 2023).

Cada registro armazenado inclui a identificação do aluno, o embedding facial extraído, o comportamento detectado, o nível de confiança da predição e o horário exato do evento. Essa estrutura possibilita a reconstrução precisa da linha temporal de atividades observadas, a auditoria dos comportamentos identificados e a vinculação segura entre reconhecimento facial e análise comportamental. Além disso, o PostgreSQL permite consultas avançadas para análise temporal, mineração de padrões e geração de relatórios agregados, contribuindo para o uso pedagógico dos dados em ambientes educacionais.

A Figura 32 apresenta o modelo relacional do banco de dados utilizado pela plataforma de monitoramento comportamental. A tabela *students* armazena os dados cadastrais dos alunos (ID, nome e matrícula) e se relaciona com a tabela *face_embeddings*, que registra os vetores faciais gerados pelo *InsightFace* e atualizados ao longo do tempo.

Já a tabela *behavior_episode* concentra os registros comportamentais identificados pela plataforma, vinculando cada episódio a um aluno por meio do campo *id_student*. Diferentemente de uma simples contagem de ocorrências, essa estrutura permite registrar episódios temporais completos de comportamento, armazenando informações como horário de início (*start_time*), horário de término (*end_time*) e a duração total do evento (*duration_seconds*). Além disso, a tabela também mantém dados contextuais relevantes, como data da aula, escola, disciplina, professor e o tipo de comportamento detectado. A tabela *users*, por sua vez, armazena os dados dos usuários da plataforma, como professores e gestores, possibilitando o controle de acesso ao sistema e permitindo futuras expansões de funcionalidades administrativas.

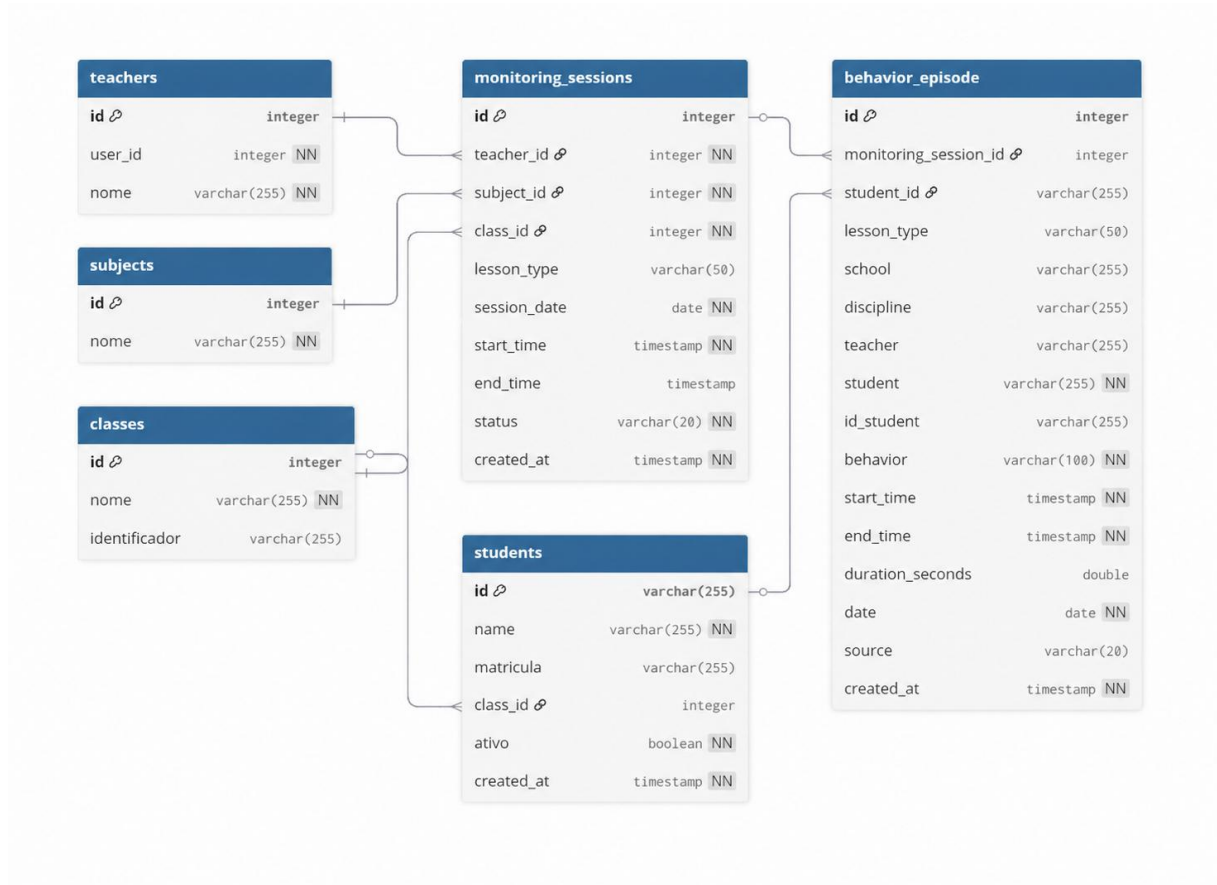
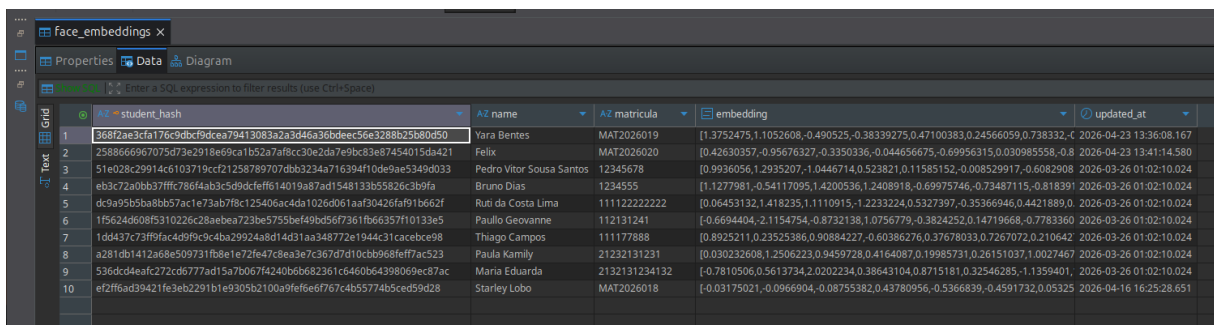


Figura 32 - Modelo relacional do banco de dados da plataforma e monitoramento comportamental.
Fonte: (Autor, 2026).

A integração com o *pgvector* torna o processo de comparação de *embeddings* altamente eficiente, pois utiliza métricas vetoriais como similaridade de cosseno ou distância euclidiana diretamente no banco de dados. Isso reduz a necessidade de processamento adicional na aplicação, garantindo respostas rápidas mesmo em cenários com grande volume de dados. Essa abordagem também assegura escalabilidade, permitindo que a plataforma seja estendida para turmas maiores ou múltiplas salas de aula sem perda significativa de desempenho.

Os *embeddings* faciais são armazenados no banco de dados PostgreSQL por meio da extensão *pgvector*, que permite o armazenamento e a comparação eficiente de vetores de alta dimensão. Cada vetor representa as características faciais de um aluno, possibilitando a identificação por meio de métricas de similaridade, como similaridade de cosseno. Essa abordagem permite realizar o reconhecimento facial de forma eficiente em tempo real, mantendo o desempenho e a escalabilidade da solução.

A Figura 33 apresenta a estrutura da tabela *face_embeddings* no PostgreSQL, responsável pelo armazenamento persistente dos vetores faciais extraídos pelo modelo *InsightFace*. Cada registro associa o *embedding* ao aluno correspondente por meio da matrícula e de um *hash* exclusivo. O uso da extensão *pgvector* permite que esses vetores sejam consultados diretamente por métricas de similaridade, como cosseno ou distância euclidiana, viabilizando consultas eficientes para processos de identificação facial em tempo real. Essa organização garante escalabilidade, rastreabilidade e integridade dos dados, além de atender aos requisitos de segurança previstos pela LGPD.



	student_hash	AZ name	AZ matricula	embedding	updated_at
1	368f2ae3cfa176c9dbcf0dca79413083a2a3d46a36bdeec56e3288b25b80d50	Yara Bentes	MAT2026019	[1.3752475,1.1052608,-0.490525,-0.38339275,0.47100383,0.24566059,0.738332,-0.2026-04-23 13:36:08.167	
2	2588666967075d73e2918e69ca1b52a7af8cc30e2da7e9bc83e67454015da421	Felix	MAT2026020	[0.42630357,-0.95676327,-0.3350336,-0.044656675,-0.69956315,0.030985558,-0.8 2026-04-23 13:41:14.580	
3	51e028c29914c6103719ccf21258789707dbb3234a716394f10de9ae5349d033	Pedro Vitor Sousa Santos	12345678	[0.9936056,1.2935207,-1.0446714,0.523821,0.11585152,-0.008529917,-0.6082908 2026-03-26 01:02:10.024	
4	eb3c72a0bb37ffcf786f4ab3c5d9dfeff614019a87ad1548133b55826c3b9fa	Bruno Dias	12345555	[1.1277981,-0.54117095,1.4200536,1.2408918,-0.69975746,-0.73487115,-0.818391 2026-03-26 01:02:10.024	
5	dc9a95b5ba8b57ac1e73ab7f8c125406ac4da1026d061aaf30426af91b662f	Ruti da Costa Lima	111122222222	[0.06453132,1.418235,1.1110915,-1.2233224,0.5327397,-0.35366946,0.4421889,0. 2026-03-26 01:02:10.024	
6	1f5624d608f5310226c28aeba723be5755bdf49bd56f7361fb66357f10133e5	Paulo Geovanne	112131241	[-0.6694404,-2.1154754,-0.8732138,1.0756779,-0.3824252,0.14719668,-0.7783360 2026-03-26 01:02:10.024	
7	1dd437c73f9f4c4d9bc9c4ba29924a8d14d31aa348772e1944c31cacebc98	Thiago Campos	111177888	[0.8925211,0.23525386,0.90884227,-0.60386276,0.37678033,0.7267072,0.210642 2026-03-26 01:02:10.024	
8	a281db1412a68e509731fb8e1e72fe47c8a3e7c367d7d10cbb968feffac523	Paula Kamilly	21232131231	[0.030232608,1.2506223,0.9459728,0.4164087,0.19985731,0.26151037,1.0027467 2026-03-26 01:02:10.024	
9	536dc4eafcf272cd777ad15a7b067f4240b6b682361c646064398069ec87ac	Maria Eduarda	2132131234132	[-0.7810506,0.5613734,2.0202234,0.38643104,0.8715181,0.32546285,-1.1359401, 2026-03-26 01:02:10.024	
10	ef2ffad39421fe3eb2291b1e9305b2100a9fef6e767c4b5577405ced59d28	Stanley Lobo	MAT2026018	[-0.03175021,-0.0966904,-0.08755382,0.43780956,-0.5366839,-0.4591732,0.05325 2026-04-16 16:25:28.651	

Figura 33 - Estrutura da tabela *face_embeddings* no PostgreSQL, exibindo o armazenamento dos vetores faciais (embeddings) utilizando a extensão *pgvector*. Fonte: (Autor, 2026).

3.3 Fluxo de Implementação

A plataforma foi organizada em um fluxo integrado que permite o processamento contínuo das imagens capturadas pela câmera instalada na sala de aula. Após a aquisição dos *frames*, ocorre a detecção das faces e a extração dos *embeddings* utilizando o *framework InsightFace*, responsável por representar matematicamente as características faciais dos alunos com alta discriminabilidade (DENG et al., 2019). Paralelamente, o modelo *YOLOv11-Pose* executa a estimativa da pose corporal, identificando comportamentos como “atento”, “distrato”, “escrevendo”, “perguntando” e “agitado”, de forma compatível com estudos recentes sobre análise comportamental por visão computacional (DUAN; LIU, 2023).

Os dados processados, incluindo identificação facial, comportamento detectado, níveis de confiança e marca temporal, são armazenados no banco PostgreSQL e indexados por meio da extensão *pgvector*, que possibilita consultas eficientes por similaridade entre *embeddings*. A interface de monitoramento foi desenvolvida em *Streamlit* e permite acompanhar, em

tempo real, o reconhecimento facial e a classificação comportamental, fornecendo um ambiente intuitivo para análise e exploração dos dados. Essa organização possibilita rastreabilidade, consistência e desempenho adequado para uso em contexto escolar.

A Figura 34 apresenta uma visão geral do fluxo de implementação da plataforma, destacando as etapas que compõem o processamento em tempo real, desde a captura da imagem até o armazenamento dos *embeddings* e a visualização dos dados no dashboard. Esse fluxo integrado garante a comunicação eficiente entre os módulos de detecção, reconhecimento e análise, permitindo que a plataforma opere de forma contínua e responsiva no ambiente escolar.

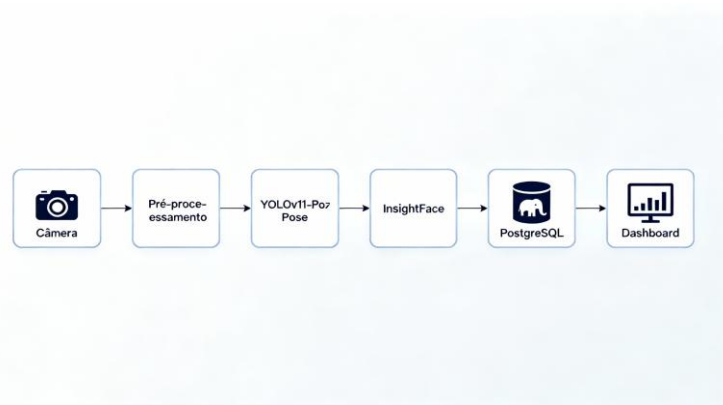


Figura 34 - Fluxo geral de implementação do sistema de monitoramento comportamental.
Fonte: (Autor, 2025).

3.4 Avaliação da Viabilidade da Plataforma

A avaliação da plataforma foi conduzida com base em três dimensões principais: desempenho do reconhecimento facial, qualidade da detecção de comportamentos e eficiência temporal para uso em tempo real. O objetivo desta etapa não foi realizar uma validação estatística em larga escala, mas verificar a viabilidade técnica e funcional da solução no contexto educacional.

Os experimentos foram realizados em ambiente controlado, simulando condições típicas de sala de aula, incluindo variações de iluminação, movimentação dos alunos e diversidade de posturas corporais. Essa abordagem permitiu analisar o comportamento da plataforma em situações próximas à realidade, sem caracterizar um estudo extensivo com população escolar.

O módulo de reconhecimento facial foi avaliado por meio da comparação entre *embeddings* previamente registrados e *embeddings* capturados durante a execução do sistema, utilizando a similaridade do cosseno como métrica de comparação. Essa abordagem é amplamente adotada em modelos baseados em perdas angulares, por proporcionar maior separabilidade entre identidades.

Para a análise comportamental, foram consideradas métricas clássicas como precisão, revocação e F1-score, utilizadas de forma complementar para observar o desempenho geral do sistema na classificação dos comportamentos. As comparações foram realizadas a partir de referências manuais, com o objetivo de verificar a consistência das respostas do modelo, e não de estabelecer um processo formal de validação estatística.

Além disso, foram analisados aspectos relacionados ao desempenho computacional, como o tempo médio de processamento por frame e a taxa de quadros por segundo (FPS), a fim de verificar a capacidade da plataforma de operar em tempo próximo ao real, requisito fundamental para sua aplicação prática.

Os resultados obtidos indicam que a plataforma apresenta comportamento consistente do ponto de vista técnico, sendo capaz de integrar os diferentes módulos propostos e gerar informações relevantes para análise. Dessa forma, a avaliação realizada reforça a viabilidade da solução como ferramenta de apoio ao acompanhamento comportamental em ambiente educacional, respeitando seus limites técnicos e interpretativos.

3.5 Procedimentos de Coleta de Dados

A coleta de dados ocorreu em sala de aula utilizando a câmera previamente instalada, possibilitando a captação contínua e natural do comportamento dos alunos durante atividades pedagógicas. As gravações foram realizadas em condições reais sem interferência externa, garantindo representatividade e autenticidade dos comportamentos analisados.

Os vídeos coletados foram processados em tempo real pela plataforma, gerando automaticamente *embeddings* faciais, classificações comportamentais e metadados associados, como horário e níveis de confiança. Para proteção da privacidade dos participantes, apenas os *embeddings* faciais foram armazenados, caracterizando uma prática de pseudonimização recomendada em tratamentos de dados biométricos. Estudos demonstram

que *embeddings* não podem ser revertidos para imagens originais, tornando esse formato compatível com os requisitos da LGPD e com normas internacionais de proteção de dados, como a ISO/IEC 24745 (DENG et al., 2019; ISO, 2022).

A coleta seguiu princípios éticos, garantindo o uso exclusivo dos dados para fins de pesquisa e respeitando integralmente a segurança, privacidade e integridade das informações tratadas. Esse procedimento assegura a conformidade metodológica necessária para análises confiáveis em avaliações posteriores.

Os procedimentos adotados possuem caráter exploratório e demonstrativo, não configurando um estudo clínico ou diagnóstico dos estudantes. A plataforma foi utilizada com o objetivo de analisar padrões comportamentais observáveis no contexto educacional, respeitando os limites éticos e interpretativos do uso da tecnologia. Dessa forma, os dados coletados devem ser compreendidos como suporte à análise pedagógica, cabendo ao professor a interpretação dos resultados no contexto da sala de aula.

3.6 Considerações Éticas

A condução desta pesquisa envolve o tratamento de dados sensíveis, especialmente aqueles relacionados à imagem e à identificação facial de indivíduos em ambiente educacional. Diante disso, foram adotadas medidas técnicas e metodológicas voltadas à proteção da privacidade, da segurança e da integridade das informações, em conformidade com a Lei Geral de Proteção de Dados (LGPD) e com boas práticas recomendadas na literatura.

Como estratégia de proteção, a plataforma não realiza o armazenamento de imagens faciais brutas, utilizando exclusivamente *embeddings* faciais, que consistem em representações vetoriais irreversíveis das características do rosto. Essa abordagem reduz significativamente os riscos associados à exposição de dados biométricos, uma vez que não permite a reconstrução da imagem original, estando alinhada ao princípio da minimização de dados (DENG et al., 2019; ISO, 2022).

O processamento das informações ocorre de forma local e controlada, sendo os dados utilizados exclusivamente para fins acadêmicos e de demonstração da solução proposta. O estudo possui caráter exploratório e não envolve a constituição de uma base de dados em larga

escala ou a aplicação institucional contínua, limitando-se à análise do funcionamento da plataforma em condições próximas ao ambiente educacional.

É importante destacar que a proposta não tem como finalidade a realização de diagnósticos clínicos, psicológicos ou comportamentais. A identificação de padrões observáveis, como atenção, distração ou agitação, deve ser compreendida como um suporte à análise pedagógica, cabendo ao professor a interpretação dos dados dentro do contexto da sala de aula.

Por fim, reforça-se que o uso de tecnologias de visão computacional em ambientes educacionais deve ser conduzido com responsabilidade, considerando não apenas aspectos técnicos, mas também éticos e sociais. A transparência no uso dos dados, o respeito à privacidade dos indivíduos e a limitação do uso das informações ao contexto educacional são princípios fundamentais para a adoção segura e responsável desse tipo de solução.

4 RESULTADOS PARCIAIS

4.1 Implementação da Estrutura da Plataforma

A implementação do módulo de cadastro e captura de imagens apresentou-se funcional e adequada ao ambiente escolar. Durante os testes, o procedimento de captura guiada das poses (frontal, laterais e cabeça baixa) permitiu a criação de um conjunto de imagens consistente, garantindo diversidade de ângulos e condições de iluminação. Essa variedade resultou em *embeddings* mais robustos, reduzindo falhas de identificação no módulo facial. O fluxo de cadastro foi concluído de forma rápida pelos alunos, com tempo médio inferior a 30 segundos por estudante, o que indica viabilidade prática para uso em turmas com grande número de participantes.

4.2 Reconhecimento Facial com *InsightFace*

Nos experimentos realizados em sala de aula, o módulo de reconhecimento facial alcançou taxas de identificação consistentes mesmo em cenários de variabilidade luminosa e posições não frontais. O modelo *Buffalo_L* apresentou estabilidade na geração dos *embeddings*, permitindo comparações eficientes via *pgvector*. A análise inicial indicou que a plataforma manteve alta precisão para distâncias inferiores a dois metros e continuou funcional com leves rotações da cabeça, o que é comum em ambientes educacionais. Casos de

falha ocorreram majoritariamente em situações de oclusão parcial do rosto, como mãos sobre o queixo ou uso de capuz, o que está alinhado com limitações relatadas na literatura sobre reconhecimento facial em ambientes dinâmicos.

4.3 Detecção de Comportamentos com YOLOv11-Pose

A detecção de comportamentos apresentou desempenho satisfatório nos testes iniciais. O modelo *YOLOv11-Pose* foi capaz de identificar corretamente estados como atenção, distração e leitura em tempo real, mantendo taxa média superior a 30 FPS na GPU utilizada, o que é aceitável para monitoramento contínuo. A plataforma mostrou capacidade de adaptação quando os alunos estavam em pé ou inclinados sobre a mesa, corrigindo adequadamente os *keypoints* corporais. Entretanto, alguns comportamentos sutis, como “atento” em ângulos extremos ou “distraído”, demonstraram maior sensibilidade a ruído visual, sugerindo a necessidade de calibração adicional nas regras de classificação. Esses resultados confirmam a viabilidade técnica do modelo para análises comportamentais contínuas.

A Figura 35 apresenta um exemplo da plataforma em funcionamento durante uma sessão de monitoramento em tempo real, integrando simultaneamente o reconhecimento facial dos estudantes ao módulo de análise comportamental baseado em estimativa de pose corporal. A partir dos *keypoints* identificados pelo modelo YOLOv11-Pose, o sistema interpreta características posturais e movimentos corporais, permitindo classificar comportamentos observados ao longo da aula.

Na interface apresentada, cada aluno identificado possui sua face reconhecida individualmente por meio do InsightFace, enquanto os comportamentos detectados, como Atento, Perguntando, Distraído e Dormindo, são exibidos diretamente sobre o vídeo de monitoramento. Essa visualização simultânea facilita a interpretação do contexto da sala de aula e permite acompanhar, de forma prática e intuitiva, a relação entre postura corporal, movimentação e padrões de engajamento observados durante a atividade pedagógica.

Além disso, o exemplo demonstra a capacidade da plataforma de operar em ambientes reais e não controlados, considerando fatores como múltiplos indivíduos em cena, diferentes distâncias em relação à câmera, variações de iluminação e diversidade de movimentos

corporais. Mesmo diante dessas condições, o sistema manteve a associação correta entre identidade e comportamento, evidenciando estabilidade operacional durante o monitoramento.

Do ponto de vista pedagógico, a integração entre reconhecimento facial e análise comportamental possibilita uma leitura mais ampla da dinâmica da sala de aula, permitindo identificar momentos de maior atenção, distração ou participação dos estudantes ao longo das atividades. Dessa forma, a plataforma demonstra potencial para auxiliar professores e gestores na observação de padrões comportamentais, contribuindo como ferramenta de apoio ao acompanhamento pedagógico em ambientes educacionais.

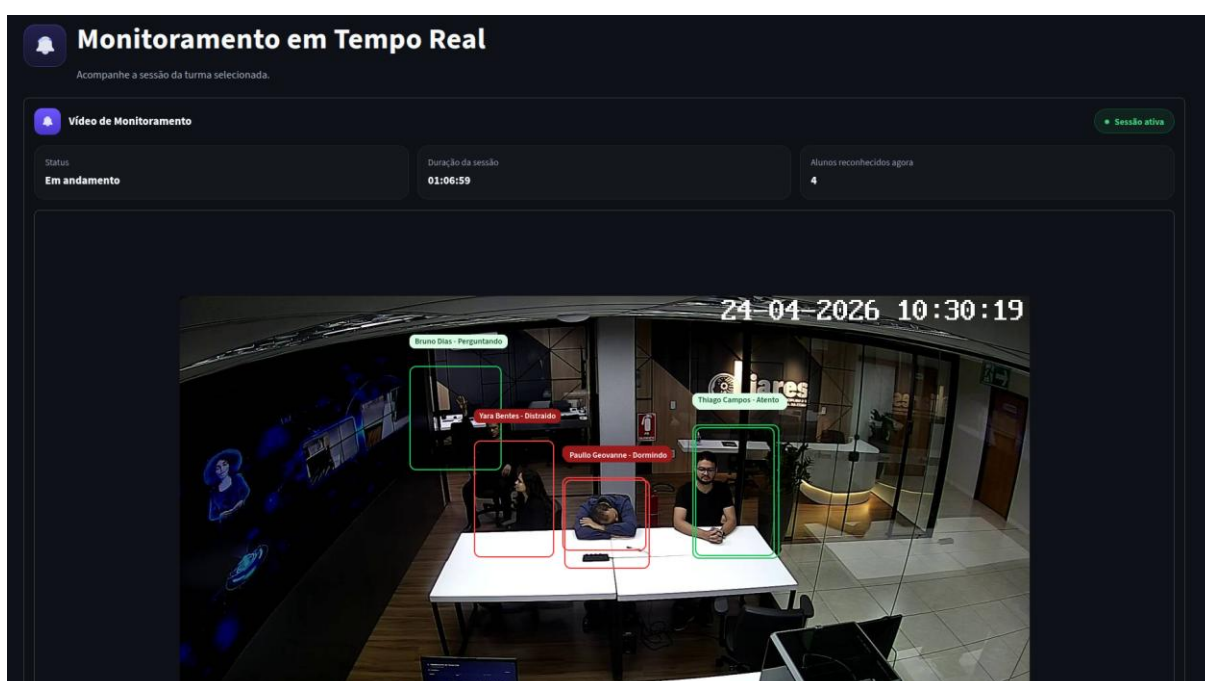


Figura 35 - Exemplo integrado de reconhecimento facial e detecção de comportamentos. Fonte: (Autor, 2026).

4.4 Integração com Banco de Dados e Visualização em Streamlit

A integração da plataforma com o banco de dados PostgreSQL apresentou-se eficiente para o armazenamento estruturado dos registros de comportamento e identificação facial. Durante a operação em tempo real, cada inferência comportamental foi consolidada em episódios temporais estáveis, permitindo a mensuração do tempo efetivo de permanência em cada estado identificado.

Diferentemente da contagem simples de ocorrências, a abordagem baseada em duração temporal possibilita uma análise mais precisa do engajamento ao longo da aula. O uso da extensão *pgvector* garantiu rapidez e estabilidade na comparação entre *embeddings*, mesmo em sessões prolongadas de monitoramento contínuo. Esses resultados confirmam a adequação do PostgreSQL como solução escalável para cenários educacionais com elevado volume de dados e necessidade de análises posteriores.

Na interface desenvolvida em *Streamlit*, os registros foram apresentados por meio de gráficos temporais que evidenciam tanto a distribuição percentual quanto o tempo absoluto dedicado a cada comportamento. Os gráficos de linha permitiram observar a dinâmica comportamental ao longo da aula, enquanto os gráficos de barras e setores facilitaram a interpretação da predominância e da duração relativa dos estados detectados.

A visualização organizada das informações contribuiu para identificar períodos prolongados de atenção, episódios pontuais de dispersão e variações comportamentais ao longo do tempo. Esses achados reforçam que a plataforma não apenas detecta eventos isolados, mas modela a continuidade comportamental, transformando dados brutos em indicadores relevantes para apoio ao acompanhamento pedagógico. A Figura 36 apresenta a proporção em que o aluno permaneceu em cada estado comportamental identificado.

Percentual do Tempo por Comportamento - Paulo Geovanne (24/04/26)

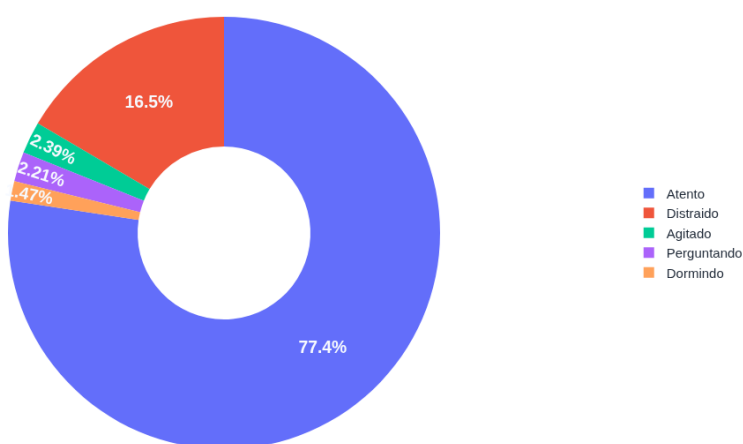


Figura 36 - Distribuição percentual de permanência do aluno em cada comportamento.
Fonte: (Autor, 2026).

A Figura 37 apresenta a distribuição do tempo acumulado para cada comportamento identificado durante o período analisado. Observa-se predominância significativa do comportamento atento, indicando que o aluno permaneceu engajado na maior parte da aula.

Também foram registrados momentos de distração, embora em menor proporção, sugerindo oscilações pontuais de atenção ao longo da atividade pedagógica. Já os comportamentos de agitação, sonolência e participação ativa por meio de perguntas ocorreram de forma reduzida, demonstrando baixa frequência durante a observação realizada.

De maneira geral, os resultados sugerem um padrão comportamental predominantemente estável, mas com pequenas variações de engajamento que podem estar relacionadas à dinâmica da aula, ao tipo de atividade desenvolvida ou ao momento pedagógico observado.

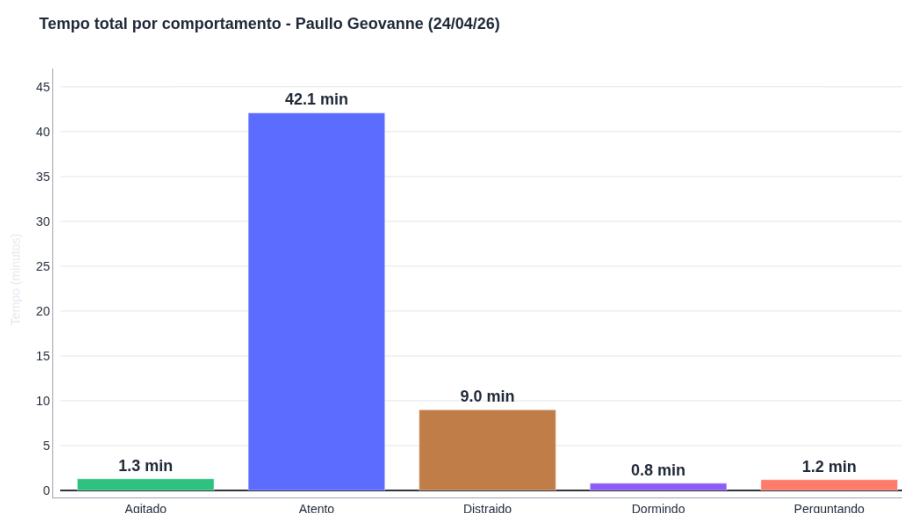


Figura 37 - Tempo total, em minutos, dedicado a cada estado comportamental identificado pela plataforma durante a aula. Fonte: (Autor, 2026)

4.5 Análise Temporal dos Dados

A análise temporal dos dados gerados pela plataforma revelou padrões importantes sobre o comportamento dos alunos ao longo das aulas. A coleta contínua permitiu visualizar a distribuição de comportamentos em gráficos de linha e barras, evidenciando momentos de maior concentração, períodos de queda de atenção e recorrência de ações como distração,

leitura ou agitação. Esse acompanhamento em escala temporal mostrou-se particularmente útil para identificar oscilações durante a aula, como picos de distração próximos ao final do período ou aumento de participação em atividades específicas.

Nos testes realizados, observou-se que determinados alunos apresentavam variações consistentes em seus níveis de atenção, enquanto outros demonstravam estabilidade ao longo da aula. Esses padrões reforçam achados da literatura que apontam que o engajamento não é uniforme e pode ser influenciado por fatores ambientais, emocionais ou pedagógicos (FREDRICKS; BLUMENFELD; PARIS, 2004). A repetição desses comportamentos ao longo de diferentes dias permitiu verificar que certos padrões não eram eventos isolados, mas sim tendências comportamentais, o que amplia a confiabilidade da interpretação dos dados.

A análise temporal também mostrou potencial para auxiliar professores na identificação de momentos críticos da aula, como quando há queda significativa de atenção ou aumento de comportamentos de dispersão. Esses *insights* podem orientar ajustes didáticos, como mudanças na dinâmica da aula, introdução de atividades interativas ou variações no ritmo de ensino. Assim, a análise temporal não apenas enriquece a visualização dos resultados, mas se apresenta como uma ferramenta estratégica para o acompanhamento pedagógico contínuo, contribuindo para intervenções mais precisas e fundamentadas no comportamento real dos alunos.

Para ilustrar essa dinâmica temporal dos comportamentos registrados durante a aula, a Figura 38 apresenta um exemplo da linha do tempo dos episódios comportamentais identificados pela plataforma ao longo do período monitorado. O gráfico evidencia não apenas a presença dos diferentes comportamentos detectados, mas também a sequência temporal e a duração relativa de estados como atenção, distração, agitação, sonolência e participação ativa por meio de perguntas.

Observa-se que o comportamento atento permaneceu predominante durante grande parte da aula, intercalado por momentos pontuais de distração e pequenas variações comportamentais ao longo do período analisado. Esse tipo de visualização permite compreender com maior clareza como os comportamentos se distribuem ao longo do tempo, possibilitando identificar trechos da aula em que ocorreram oscilações no nível de engajamento do aluno.

Além disso, a análise temporal contribui para uma interpretação mais contextualizada do comportamento estudantil, permitindo relacionar determinados episódios à dinâmica pedagógica desenvolvida naquele momento da aula. Dessa forma, a linha do tempo comportamental se apresenta como um recurso importante para apoiar professores e gestores na observação contínua dos padrões de participação e atenção dos estudantes em ambientes educacionais reais.

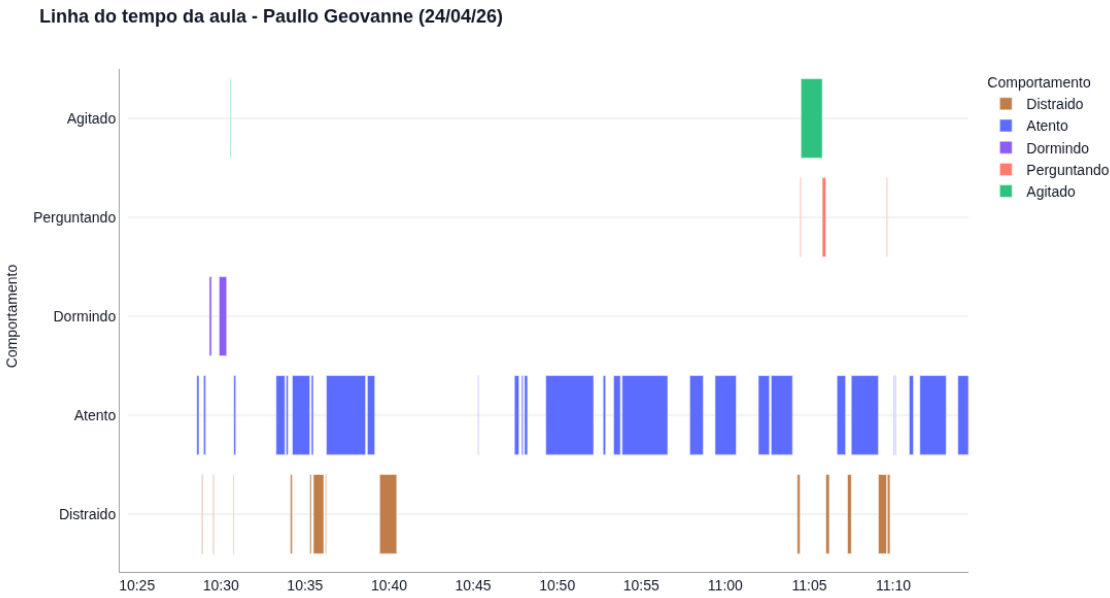


Figura 38 - Gráfico de contagem temporal dos comportamentos do aluno.
Fonte: (Autor, 2026).

A Figura 39 apresenta a comparação dos comportamentos observados de acordo com o tipo de aula analisado. Verifica-se que, durante as aulas expositivas, houve maior presença de episódios de distração, enquanto o comportamento atento apresentou percentual mais elevado no contexto de prova.

Esse resultado sugere que diferentes dinâmicas pedagógicas podem influenciar diretamente o nível de engajamento dos alunos ao longo da atividade. Nas aulas expositivas, a maior ocorrência de distração pode estar relacionada ao tempo prolongado de escuta e menor

interação ativa do estudante. Já no contexto avaliativo, observa-se aumento significativo do comportamento atento, indicando maior concentração e foco durante a realização da prova.

Além disso, os comportamentos de agitação, sonolência e participação por meio de perguntas apareceram de forma reduzida nos dois contextos analisados, demonstrando menor frequência durante o período observado. De maneira geral, os dados reforçam a importância de considerar o tipo de atividade pedagógica na interpretação dos padrões comportamentais registrados pela plataforma.

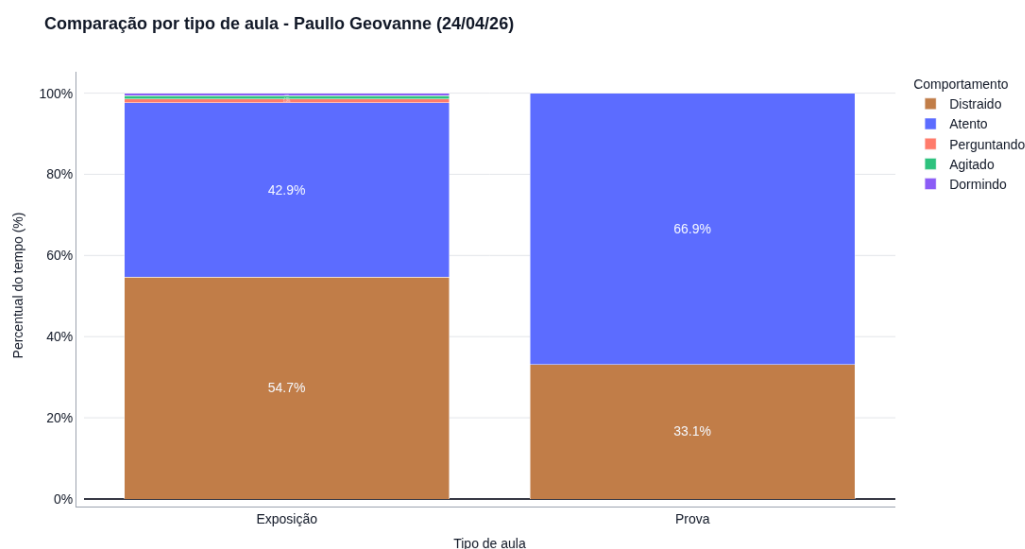


Figura 39 - Comparação percentual dos comportamentos observados em diferentes tipos de aula, evidenciando variações no nível de atenção e distração entre atividades expositivas e avaliativas. Fonte: (Autor, 2026).

5 CONSIDERAÇÕES FINAIS

O desenvolvimento da plataforma de monitoramento comportamental apresenta avanços significativos e consolida a viabilidade técnica da proposta. A integração entre os módulos de detecção de pose e reconhecimento facial apresentou resultados satisfatórios durante os testes iniciais. O modelo *YOLOv11-Pose* mostrou-se eficiente na identificação de posturas e movimentos corporais dos alunos em tempo real, possibilitando reconhecer comportamentos como atenção, distração, leitura e agitação. De forma complementar, o *InsightFace* apresentou desempenho consistente na identificação dos estudantes em diferentes

condições de ângulo e iluminação, assegurando a associação correta entre indivíduos e comportamentos registrados.

A implementação da interface em *Streamlit* contribuiu para tornar a plataforma mais acessível e funcional, permitindo visualizar gráficos, tabelas e registros de maneira intuitiva. Essa visualização facilitou a interpretação dos dados e possibilitou análises temporais mais detalhadas, evidenciando padrões relevantes ao contexto pedagógico, como oscilações de atenção ao longo da aula e momentos de maior dispersão. A integração com o banco de dados PostgreSQL, aliado ao uso da extensão *pgvector*, permitiu o armazenamento estruturado dos eventos e *embeddings* faciais, garantindo rastreabilidade e eficiência na recuperação das informações.

Apesar dos resultados promissores, algumas limitações permanecem evidentes. A acurácia da plataforma ainda pode ser comprometida em condições de luminosidade desfavorável, oclusões faciais ou movimentações bruscas, fatores que afetam tanto a estimativa de pose quanto o reconhecimento facial. Além disso, comportamentos mais sutis, como estados emocionais ou graus de engajamento, continuam sendo um desafio, exigindo aprimoramento das regras de classificação e calibração dos modelos. Essas limitações são coerentes com estudos recentes que destacam a complexidade de interpretar comportamentos humanos por meio de visão computacional em ambientes reais.

De maneira geral, os resultados obtidos indicam que a plataforma apresenta potencial para apoiar análises preliminares do comportamento estudantil e auxiliar o acompanhamento pedagógico em ambiente escolar. Apesar disso, ainda são necessários aprimoramentos técnicos e validações em diferentes cenários educacionais, visando ampliar a precisão, confiabilidade e aplicabilidade da solução.

REFERÊNCIAS BIBLIOGRÁFICAS

- AN, Xiang et al.** Partial FC: Training 10 Million Identities on a Single Machine. arXiv, 2020.
- BRASIL.** Lei nº 13.709, de 14 de agosto de 2018. Lei Geral de Proteção de Dados Pessoais (LGPD). Diário Oficial da União, Brasília, 2018.
- CAO, Zhe et al.** Realtime Multi-Person 2D Pose Estimation Using Part Affinity Fields. In: CVPR, 2017.
- DAI, Jianqiang; XIE, Aimin; YU, Deying.** Classroom Behavior Recognition Through Joint Improvement of Faster R-CNN Algorithm Under the Construction of Smart Education. KSII Transactions on Internet and Information Systems, 2025.
- DENG, Jiankang et al.** ArcFace: Additive Angular Margin Loss for Deep Face Recognition. In: CVPR, 2019.
- DONEDA, Danilo; MENDES, Laura Schertel.** Proteção de Dados Pessoais: A Função e os Limites do Consentimento. São Paulo: Thomson Reuters, 2021.
- FREDRICKS, J. A.; BLUMENFELD, P. C.; PARIS, A. H.** School engagement: potential of the concept, state of the evidence. Review of Educational Research, v. 74, n. 1, p. 59–109, 2004.
- GONZALEZ, Rafael C.; WOODS, Richard E.** Digital Image Processing. 4. ed. New York: Pearson, 2018.
- GUO, Jia et al.** InsightFace: An Open-Source 2D and 3D Deep Face Analysis Toolbox. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2021.
- HE, Kaiming et al.** Deep Residual Learning for Image Recognition. In: CVPR, 2016.
- HIKVISION.** DS-2CD1123G0E-I – 2 MP Fixed Dome Network Camera: Datasheet. 2024.
- IOFFE, Sergey; SZEGEDY, Christian.** Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. In: ICML, 2015.
- ISO.** ISO/IEC 24745:2022 – Biometric Information Protection. Geneva: International Organization for Standardization, 2022.
- JIA, Qingzheng; HE, Jialiang.** Student Behavior Recognition in Classroom Based on Deep Learning. Applied Sciences, 2024.
- KHANAM, Rahima; HUSSAIN, Muhammad.** YOLOv11: An Overview of the Key Architectural Enhancements. arXiv, 2024.

- LI, Meng; SUN, Zhenan; ZHOU, Jingqing.** Robust Face Recognition in Non-Cooperative Environments. Pattern Recognition Letters, 2023.
- LIN, Feng-Cheng et al.** Student Behavior Recognition System for the Classroom Environment Based on Skeleton Pose Estimation and Person Detection. Sensors, 2021.
- LIU, Qingtang; JIANG, Xinyu; JIANG, Ruyi.** Classroom Behavior Recognition Using Computer Vision: A Systematic Review. Sensors, 2025.
- MAJI, Debapriya et al.** YOLO-Pose: Enhancing YOLO for Multi-Person Pose Estimation. arXiv, 2022.
- MÁRQUEZ-CARPINTERO, Juan et al.** Multimodal Behavioral Monitoring in Educational Environments. Sensors, 2023.
- NAZI, I.; ZHOU, F.; FARROQ, M.** Performance metrics for real-time educational monitoring dashboards. IEEE Transactions on Learning Technologies, 2023.
- RUSSELL, Stuart; NORVIG, Peter.** Artificial Intelligence: A Modern Approach. 3. ed. Upper Saddle River: Pearson, 2013.
- (Você cita esse livro no início da fundamentação — confirmo que aparece no texto.)
- SCHROFF, Florian; KALENICHENKO, Dmitry; PHILBIN, James.** FaceNet: A Unified Embedding for Face Recognition and Clustering. In: CVPR, 2015.
- SHENG, Xiaojing; LI, Suqiang; CHAN, Sixian.** Real-Time Classroom Student Behavior Detection Based on Improved YOLOv8s. Scientific Reports, 2025.
- SINGH, Gururani; BHATNAGAR.** Emotion and Attention Recognition in Smart Classrooms. IEEE Access, 2023.
- STREAMLIT.** Streamlit: A Faster Way to Build and Share Data Apps. 2024.
- TRABELSI, Zouheir et al.** Real-Time Attention Monitoring System for Classroom: A Deep Learning Approach. Big Data and Cognitive Computing, 2023.
- ULTRALYTICS.** Ultralytics YOLOv11-Pose Documentation. 2024.
- WANG, Feng et al.** CosFace: Large Margin Cosine Loss for Deep Face Recognition. In: CVPR, 2018.
- WANG, Xiaoli et al.** A Spatio-Temporal Attention Fusion Model for Student Behavior Recognition in Classroom. EAI Transactions on Scalable Information Systems, 2023.
- ZAKIROV, Andrey.** pgvector: Vector Similarity Search Extension for PostgreSQL. 2023.
- ZHANG, Jian et al.** A YOLOv8-Based Behavior Detection Framework for Smart Classrooms. Scientific Reports, 2024.

APÊNDICE A – Relatório observacional pedagógico gerado pela plataforma

Relatório observacional pedagógico

Aluno: Paulo Geovanne
Período: 24-04-2026 a 24-04-2026 | Recorte: Diário | Emitido em: 29-04-2026

Comportamento predominante

Atento

Tempo total observado

54.4 min

Dias com registros

1

Resumo executivo

No período analisado, observou-se predominância do comportamento atento, indicando bom nível geral de engajamento com a atividade. Também foram registrados episódios de distração, o que pode indicar oscilações de atenção ou participação em parte da aula. A maior concentração de registros ocorreu no início da aula, indicando que esse trecho da aula merece atenção na leitura pedagógica. Como leitura inicial, os dados sugerem apoiar a decisão pedagógica com nova observação e análise do contexto da aula.

Leitura visual dos comportamentos



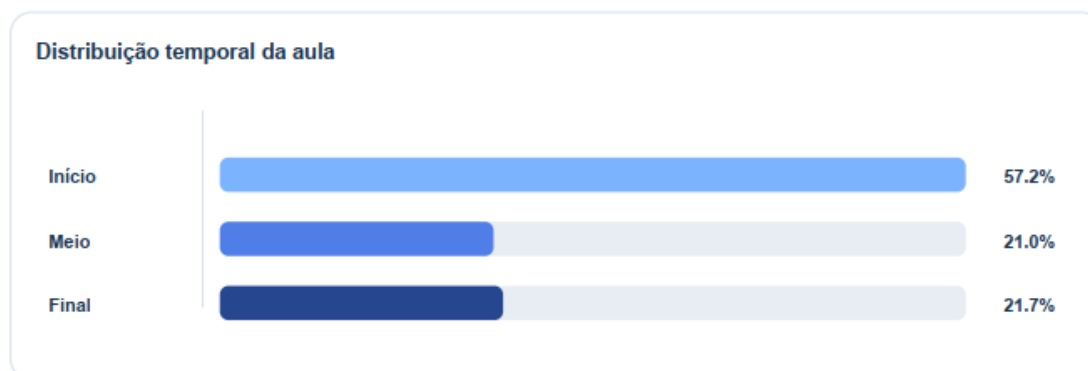
Análise interpretativa

O comportamento atento predominou no período observado (77.3% do tempo registrado), indicando bom nível geral de engajamento com a atividade.

A presença de distração em parte relevante da aula sugere oscilações que podem estar relacionadas à dinâmica da atividade, ao momento da aula ou ao tipo de interação proposta.

Essa leitura é pedagógica e observacional: ela orienta perguntas para a equipe escolar, mas não substitui a análise do professor nem produz diagnóstico.

Distribuição temporal



A maior concentração de registros ocorreu no início da aula. Nesse trecho, o comportamento mais presente foi atento. Esse resultado ajuda a gestão e o professor a observar quais condições desse trecho favoreceram ou concentraram os comportamentos registrados.

Recomendações pedagógicas

- Observar se a distração ocorre com maior frequência no início da aula ou durante transições de atividade.
- Testar estratégias de ativação inicial, como pergunta norteadora, atividade breve ou checagem de compreensão.
- Realizar nova observação em aulas futuras para verificar se o padrão se mantém, diminui ou se intensifica.

Indicação de prioridade observacional

Comportamentos que demandam atenção neste período

Distração: 16,6%
 Agitação: 2,4%
 Sonolência: 1,5%

Atenção observacional moderada

Essa classificação decorre da predominância do comportamento atento, combinada à presença de comportamentos que indicam redução de engajamento, principalmente distração (16,6%), além de ocorrências de agitação (2,4%) e sonolência (1,5%). Esses comportamentos estiveram mais concentrados no início da aula, sugerindo atenção a esse momento da aula. A leitura deve permanecer pedagógica e contextualizada, com nova observação para verificar se o padrão se mantém.

Recomenda-se observar especialmente:

- Momentos de distração no início da aula.
- Momentos de agitação no início da aula.
- Possíveis oscilações de atenção ao longo da atividade.

Comparação com o período anterior

Não houve registros suficientes no período anterior para uma comparação analítica consistente.

Síntese para tomada de decisão

O aluno apresentou predominância do comportamento atento (77,3%), porém foram observadas oscilações de engajamento, principalmente associadas à distração no início da aula. Recomenda-se acompanhamento pedagógico leve, com foco na estratégia do trecho de maior concentração e nova observação em aulas futuras.

Nota metodológica

Este relatório organiza dados observacionais para apoiar a leitura pedagógica. O tempo acumulado é uma estimativa baseada na duração dos episódios registrados e pode variar conforme captação, enquadramento, iluminação, oclusões e estabilidade do vídeo.

Limitações

A plataforma não realiza diagnóstico clínico nem identifica transtornos. Os resultados dependem das condições de captação e devem ser interpretados junto ao contexto da aula, ao planejamento pedagógico e à observação profissional da equipe escolar.