

## **Sistema para detecção de sonolência e fadiga com inferência visual em dispositivo embarcado utilizando *Edge Impulse* e *ESP32-CAM***

System for drowsiness and fatigue detection with visual inference on embedded device using Edge Impulse and ESP32-CAM.

Sistema para la detección de somnolencia y fatiga con inferencia visual en dispositivo embebido utilizando Edge Impulse y ESP32 CAM

Leonardo de Souza Camargo <sup>1</sup>  
Arnaldo de Carvalho Junior <sup>2</sup>

RECEBIDO EM 07/08/2025  
ACEITO EM 17/10/2025

### **RESUMO**

Este trabalho apresenta o desenvolvimento de um sistema computacional embarcado para detecção de sonolência e fadiga em condutores, com base em visão computacional e aprendizado de máquina. Motivado pelos altos índices de acidentes causados por cansaço e desatenção ao volante, o estudo propõe uma solução de baixo custo e fácil implementação, com foco na prevenção de comportamentos de risco. O principal objetivo foi criar um sistema capaz de identificar, de forma autônoma, sinais visuais de sonolência e acionar alertas em tempo real. A metodologia envolveu a coleta e refinamento de imagens da região do rosto, utilizadas para treinar uma rede neural convolucional na plataforma *Edge Impulse*. O modelo resultante, com acurácia de 94,7%, foi embarcado em um microcontrolador *ESP32-CAM*, responsável por capturar as imagens e realizar inferência local. Em testes práticos, o sistema se mostrou eficaz ao detectar

<sup>1</sup> Instituto Federal de Educação, Ciência e Tecnologia de São Paulo, Campus Cubatão, Cubatão, SP, Brasil.  
leonardo.scamargo04@gmail.com - <https://orcid.org/0009-0006-4102-3155>

<sup>2</sup> Instituto Federal de Educação, Ciência e Tecnologia de São Paulo, Campus Cubatão, Cubatão, SP, Brasil.  
adecarvalhojr@ifsp.edu.br - <https://orcid.org/0000-0002-3417-0062>

padrões de sonolência com base em critérios como tempo contínuo de olhos fechados e frequência de fechamento em determinado intervalo. Os resultados demonstraram a viabilidade da proposta, evidenciando sua aplicabilidade como ferramenta de apoio à segurança veicular.

**PALAVRAS-CHAVE:** sonolência ao volante; visão computacional embarcada; segurança veicular; rede neural convolucional.

## ABSTRACT

This work presents the development of an embedded computer system for detecting driver drowsiness and fatigue using computer vision and machine learning techniques. Motivated by the high rate of traffic accidents caused by tiredness and driver inattention, the study proposes a low-cost and easily implementable solution focused on preventing risky behavior. The main objective was to create a system capable of autonomously identifying visual signs of drowsiness and triggering real-time alerts. The methodology involved collecting and refining images of the facial region, which were used to train a convolutional neural network on the Edge Impulse platform. The resulting model, with an accuracy of 94.7%, was deployed on an ESP32-CAM microcontroller responsible for image capture and local inference. In practical tests, the system effectively detected drowsiness patterns based on criteria such as continuous eye closure time and frequency of closure within a specific time window. The results demonstrated the feasibility of the proposal, highlighting its applicability as a support tool for vehicle safety.

**KEYWORDS:** driver drowsiness; embedded computer vision; vehicle safety; convolutional neural network.

## RESUMEN

Este trabajo presenta el desarrollo de un sistema informático embebido para la detección de somnolencia y fatiga en conductores, utilizando técnicas de visión por computadora y aprendizaje automático. Motivado por las altas tasas de accidentes de tránsito causados por el cansancio y la falta de atención del conductor, el estudio propone una solución de bajo costo y fácil implementación, enfocada en la prevención de comportamientos de riesgo. El objetivo principal fue crear un sistema capaz de identificar de forma autónoma señales visuales de somnolencia y activar alertas en tiempo real. La metodología incluyó la recolección y el refinamiento de imágenes de la región facial, que fueron utilizadas

para entrenar una red neuronal convolucional en la plataforma Edge Impulse. El modelo resultante, con una precisión del 94,7%, fue implementado en un microcontrolador ESP32-CAM responsable de capturar imágenes y realizar inferencias localmente. En pruebas prácticas, el sistema detectó eficazmente patrones de somnolencia basados en criterios como el tiempo continuo con los ojos cerrados y la frecuencia de cierre en un intervalo específico. Los resultados demostraron la viabilidad de la propuesta, evidenciando su aplicabilidad como herramienta de apoyo a la seguridad vehicular.

**PALABRAS CLAVE:** somnolencia al volante; visión por computadora embebida; seguridad vehicular; redes neuronales convolucionales.

## 1 Introdução

Acidentes de trânsito representam uma das principais causas de mortalidade global. A Organização Mundial da Saúde estima que cerca de 1,4 milhão de pessoas morrem todos os anos em acidentes viários, com uma morte a cada 24 segundos, em média (World Health Organization, 2018). No Brasil, a situação é igualmente problemática: entre 2010 e 2019, aproximadamente 392 mil pessoas morreram em acidentes de transporte terrestre (Instituto de Pesquisa Econômica Aplicada, 2023). Além disso, em 2018, o país registrou mais de 41 mil óbitos causados por acidentes em vias públicas (Ministério dos Transportes, 2018). Com uma frota estimada em um veículo para cada três habitantes, a segurança viária se torna um desafio crescente e urgente (Instituto Brasileiro de Geografia e Estatística, 2022).

Dentre os fatores que mais contribuem para esses acidentes, destacam-se a desatenção, fadiga e a sonolência ao volante. Estima-se que entre 2% e 20% das mortes no trânsito estejam diretamente associadas ao sono ou fadiga dos motoristas (National Highway Traffic Safety Administration - NHTSA, 2016). Entre 2009 e 2013, mais de 72 mil acidentes nos Estados Unidos envolveram condutores sonolentos, causando mais de 41 mil feridos e 800 mortes (NHTSA,

2016). Em 2015, a *AAA Foundation for Traffic Safety* relatou que motoristas em estado de sonolência ou fadiga estiveram envolvidos em cerca de 7% de todos os acidentes e 16,5% dos acidentes fatais, resultando em mais de 5 mil mortes naquele ano (NHTSA, 2016).

No Brasil, a Associação Brasileira de Medicina do Tráfego (ABRAMET), em conjunto com o Conselho Regional de Medicina (CRM) e a Academia Brasileira de Neurologia (ABN), aponta que 42% dos acidentes em vias urbanas e rodovias são causados por distração, enquanto 18% resultam diretamente de fadiga ao volante (MAIS DE 40%..., 2019). Campanhas como “Não dê carona ao sono!” vêm alertando sobre os perigos da privação de sono em condutores que passam longos períodos sem descanso.

Com o avanço da tecnologia, sistemas de assistência ao condutor (*Advanced Driver-Assistance Systems - ADAS*) estão sendo implementados em veículos modernos, como a frenagem automática de emergência. De acordo com estudos realizados pela Bosch (2018), na Alemanha, até 72% das colisões traseiras poderiam ser evitadas se todos os veículos estivessem equipados com esse tipo de sistema. No entanto, esses recursos ainda são inacessíveis para grande parte da população, principalmente para motoristas de veículos mais antigos ou sem tecnologia embarcada de fábrica.

Nesse cenário, o uso de soluções baseadas em Inteligência Artificial (IA), especialmente as técnicas de visão computacional, utilizando redes neurais convolucionais (*Convolutional Neural networks - CNN*), surgem como uma alternativa viável, econômica e eficiente. As CNN representam um dos pilares da detecção de padrões visuais e são amplamente utilizadas em aplicações que envolvem imagens, como diagnóstico médico, segurança pública e, principalmente, monitoramento de condutores. As CNN são projetadas para extrair automaticamente características relevantes de imagens por meio de camadas convolucionais, *pooling* e funções de ativação.

O estudo de Rocha e Domingues (2022), por exemplo, destacou o uso bem-sucedido de CNN na detecção de expressões faciais de fadiga, validando o uso desse tipo de arquitetura como principal mecanismo de classificação nesse tipo de sistema. Modelos de CNN, como o MobileNetSSDv1 e o SSD com ResNet-10, vêm se destacando por sua capacidade de detecção em tempo real em dispositivos embarcados de baixo custo (Dipu *et al.*, 2021). Esses modelos possibilitam o reconhecimento de padrões visuais como olhos fechados, bocejos e expressões faciais associadas à sonolência.

A literatura técnica ainda fornece critérios objetivos para a caracterização do estado de sonolência ao volante. A detecção de sonolência em motoristas é um tema amplamente explorado pela comunidade científica, especialmente no contexto da prevenção de acidentes causados por fadiga ao volante. Alguns estudos identificaram critérios objetivos que permitem inferir, a partir de imagens, se um condutor apresenta sinais de fadiga ou mesmo se encontra em estado de microssono.

Um dos principais parâmetros utilizados para a detecção de fadiga ou microssono é o tempo de fechamento ocular contínuo. Segundo Heitmann (2021) e Wierwille (1994), o fechamento dos olhos por um período maior que seis segundos consecutivos pode indicar o início de um episódio de microssono, condição crítica para a condução segura. Essa métrica é amplamente adotada por sistemas de monitoramento, pois oferece um limite mensurável e prático para o desenvolvimento de soluções automatizadas baseadas em visão computacional. Com base nesse critério, é possível calcular o número de quadros necessários para detectar esse estado, levando em conta a taxa de quadros por segundo (*Frames Per Second - FPS*) da câmera. A equação 1 expressa essa relação:

$$N = FPS \times \tau \quad (1)$$

Onde: “N” é o número de quadros consecutivos com os olhos fechados, “ $\tau$ ” é o tempo de fechamento ocular em segundos (neste caso, 6 segundos), e “FPS” é a taxa de quadros por segundo da câmera. Por exemplo, em um sistema com câmera de 30 FPS, deve-se observar 180 quadros consecutivos com os olhos fechados para se identificar a condição de sonolência. Esse cálculo foi incorporado ao sistema proposto neste trabalho como um dos critérios automáticos de disparo de alerta.

Contudo, o tempo de fechamento não é o único parâmetro relevante. Outro indicador robusto é o *PERCLOS* (*Percent Eye Closure*), que representa o percentual de tempo em que os olhos permanecem fechados durante uma janela de análise. De acordo com Dinges e Grace (1998), esse índice é um dos mais confiáveis para avaliar o nível de alerta de um indivíduo em tarefas contínuas de atenção como a direção. Ademais, para completar esse parâmetro, Masala e Grosso (2013) afirmam: “A sonolência é detectada quando os olhos permanecem fechados em 80% ou mais dos quadros analisados dentro de um intervalo de tempo, geralmente um minuto”.

A idealização deste estudo foi embasada por diversos trabalhos acadêmicos que trataram da detecção de sonolência e do uso de aprendizado profundo no setor automotivo. São eles: Bovolato e Dominguez (2023), que utilizaram redes neurais artificiais para inferência de sonolência via expressões faciais; Hossain (2022) que explorou CNN profundas para detecção de distração, alcançando alta acurácia; e Chengula (2024) que destacou o uso de IA Explicável (Explainable Artificial Intelligence - XAI) no sistema avançado de assistência ao condutor (Advanced Driver-Assistance System - ADAS), trazendo novas abordagens para interpretação de decisões.

Com base nesse referencial, este trabalho propõe o desenvolvimento de um sistema embarcado para identificação de sonolência em motoristas, utilizando como plataforma principal o microcontrolador ESP32-CAM, uma solução de

baixo custo, com capacidade de processamento local e câmera integrada. O sistema utiliza um modelo de IA treinado via plataforma “no-code” Edge Impulse (EDGE IMPULSE, 2024), sendo capaz de realizar inferência visual em tempo real diretamente no dispositivo.

Este artigo detalha as etapas do estudo, desde a definição do problema até o treinamento do modelo, integração do *hardware*, testes práticos e resultados obtidos com o protótipo funcional. A proposta visa a contribuir para a segurança no trânsito por meio de uma solução de baixo custo, acessível e adaptável a diferentes contextos veiculares.

## 2 Material e Métodos

O desenvolvimento deste trabalho foi realizado em etapas sequenciais, com foco na criação de um sistema de identificação de sonolência veicular baseado em inferência visual utilizando CNN, embarcado em um microcontrolador ESP32-CAM. As etapas envolvidas vão desde a concepção e treinamento do modelo de aprendizado de máquina pelo Edge Impulse, até a implementação embarcada e validação funcional em um protótipo físico.

### 2.1 Critérios de detecção de sonolência

O sistema desenvolvido neste trabalho considera dois gatilhos principais para identificar sonolência: fechamento ocular contínuo por 6 segundos ou mais e *PERCLOS* maior ou igual a 80% dentro de uma janela de 60 segundos. A adoção simultânea desses critérios confere maior confiabilidade ao sistema, reduzindo a chance de falsos positivos e aumentando a assertividade dos alertas gerados. A abordagem proposta prioriza decisões baseadas em dados visuais objetivos, com embarcamento leve em ESP32-CAM, permitindo sua aplicação prática em protótipos de baixo custo voltados à segurança veicular.

## 2.2 Modelo de Aprendizado de Máquina (*Machine Learning* – *ML*)

O ponto principal para um bom desempenho de uma CNN é ter um banco de dados realista e diversificado com relação ao tipo de análise a ser feito. Para isso, utilizou-se neste projeto um banco de dados retirado do site Kaggle (Ali, 2024). Porém, até se chegar à definição final de qual banco de dado utilizar, outros dois bancos de dados foram testados (Ali, 2024; Djerarda, 2023). Com isso, totalizou-se três ciclos distintos de treinamento: a) Banco de dados com imagens do rosto completo dentro de um carro – mesmo sendo a versão mais realista, foi pouco eficaz por não identificar os olhos com exatidão; b) Segunda versão - com foco apenas nos olhos, visando a reduzir ruído visual que problematizou o primeiro banco de dados – ela foi descartada porque houve dificuldade em localizar padrões, ou seja, identificava erroneamente os olhos; c) Terceira versão - e a mais eficiente, composta por recortes do rosto centralizado e limpo, padronizado quanto à posição e iluminação.

A terceira abordagem (Ali, 2024) apresentou os melhores resultados em testes reais, tanto com a *webcam* do computador, quanto com o *ESP32-CAM*, sendo, portanto, a adotada na solução final. Os testes com os diferentes bancos de dados serão melhores explanados posteriormente.

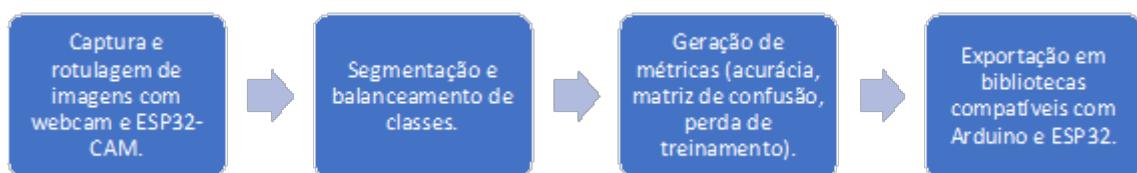
No caso deste projeto, o modelo de *ML* com *CNN* foi treinado para classificar imagens do rosto do motorista como “*AWAKE*” (acordado) ou “*SLEEPING*” (dormindo). O modelo foi construído utilizando a plataforma *Edge Impulse* e exportado de forma otimizada para rodar diretamente no microcontrolador *ESP32-CAM*.

## 2.3 Criação e Treinamento da IA

A plataforma *Edge Impulse* foi utilizada para o desenvolvimento do modelo de inferência visual. Suas funcionalidades e etapas podem ser verificadas no fluxograma da Figura 1.

Ao longo do processo, foram testadas diferentes taxas de aprendizado, arquiteturas, resoluções de imagem e técnicas de aumento de dados (*data augmentation*). A própria plataforma forneceu *insights* importantes sobre *overfitting* nos primeiros testes, levando à necessidade de ajustes no número de épocas e na diversidade do *dataset*.

FIGURA 1 - Fluxograma de desenvolvimento de um identificador de imagem pelo *Edge Impulse*.



Fonte: elaborado pelos autores (2025).

Além disso, a função de live *classification* do *Edge Impulse* permitiu testes em tempo real da IA diretamente no navegador, acelerando a validação antes da exportação para o microcontrolador.

## 2.4 Desenvolvimento do *Hardware*

O ESP32-CAM foi a plataforma de *hardware* escolhida para embarcar o modelo de IA e realizar todo o ciclo de inferência, do processamento da imagem ao acionamento dos alertas. Sua seleção se deu com base nos seguintes fatores: a) Compatibilidade com bibliotecas otimizadas para inferência de imagem; b) Baixo consumo energético e custo acessível; c) Presença de câmera integrada

(OV2640) e conectividade *Wi-Fi*. d) Suporte a armazenamento local (*SD Card*) e comunicação com periféricos (*buzzers*, *LEDs*, sensores e atuadores).

A utilização do ESP32-CAM com Módulo adaptador ESP32 Cam MB (*Mother Board*) eliminou a necessidade de dispositivos auxiliares mais caros, como *Raspberry Pi* ou *Jetson Nano*. Em comparação, microcontroladores como o controlador de Interface Programável (*Programmable Interface Controller – PIC*) foram considerados para controle de atuadores, mas descartados para processamento visual devido à limitação computacional. No trabalho de Silva (2009), por exemplo, a automação de frenagem do veículo foi feita com base no acionamento de motores de corrente continua (*Direct Current – DC*) por um microcontrolador PIC. No presente projeto, optou-se por focar no alerta visual e sonoro no protótipo inicial, com a possibilidade de integrar atuadores mecânicos em versões futuras.

## 2.5 Coleta e preparação dos dados

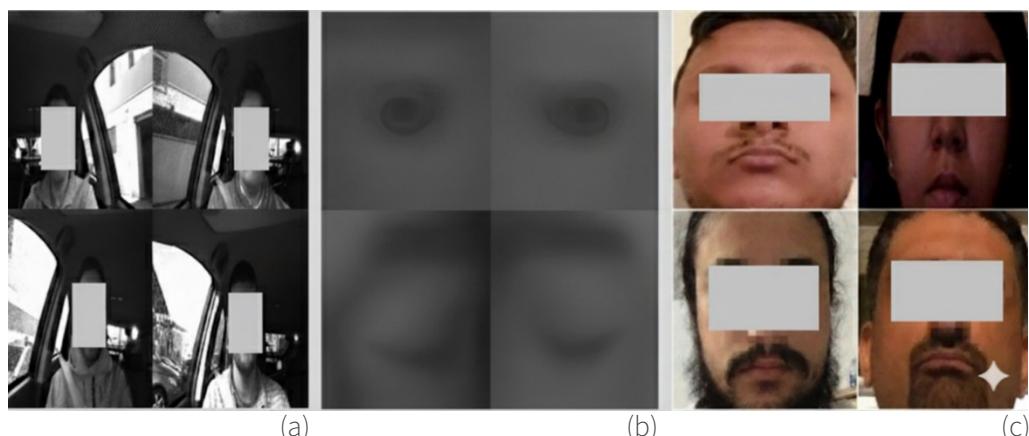
O primeiro desafio foi a formação de um banco de imagens representativas que pudessem treinar adequadamente a IA. Três versões de datasets foram criadas e avaliadas ao longo do projeto:

Na versão 1 (Figura 2a), foram utilizadas fotos de pessoas em diferentes posições dentro de um carro, mostrando um pouco acima da cabeça na parte superior, e logo abaixo dos ombros na borda inferior da imagem. Contudo, os testes iniciais mostraram baixa acurácia nas classificações reais, devido à variação de iluminação, posicionamento da cabeça e baixa resolução. Além disso, por conta de se utilizar uma câmera de baixa qualidade (2mb), sabia-se que os resultados no dispositivo embarcado seriam ainda menores, então já nos testes de *Webcam*, desistiu-se de usar um *dataset* mais próximo do real por enquanto.

Na segunda tentativa, a versão 2 (Figura 2b), buscou-se isolar a região ocular para reduzir ruído e variações externas. Embora o modelo tenha alcançado melhor desempenho em testes locais (com *webcam*), sua eficiência caiu ao ser executado na ESP32-CAM, devido à perda de resolução e limitações de foco da câmera embarcada.

Já na versão 3 (Figura 2c), utilizou-se imagens frontais com recorte de toda a área do rosto, centralizadas e bem iluminadas, como pode ser observado na Figura 4. Essa foi a base de dados com melhor desempenho prático, pois conciliava boa capacidade de generalização com compatibilidade ao campo de visão da ESP32-CAM. O modelo foi treinado com duas classes: “awake” (acordado) e “*sleeping*” (dormindo).

FIGURA 2 - Exemplo de imagens (anormalizadas) do *Dataset* da versão 1 (a), versão 2 (b) e versão 3 (c).



Fonte: adaptado de AL (2024); Djerarda (2023); Ali (2024).

## 2.6 Treinamento do modelo de IA no *Edge Impulse*

O treinamento da CNN foi realizado utilizando a plataforma *Edge Impulse*, voltada ao desenvolvimento de modelos de IA otimizados para microcontroladores. As etapas no *Edge Impulse* incluíram: a) *upload* e balanceamento de dados entre as duas classes; b) geração automática de features, usando blocos de processamento de imagem e *flattening*, e configuração do processamento

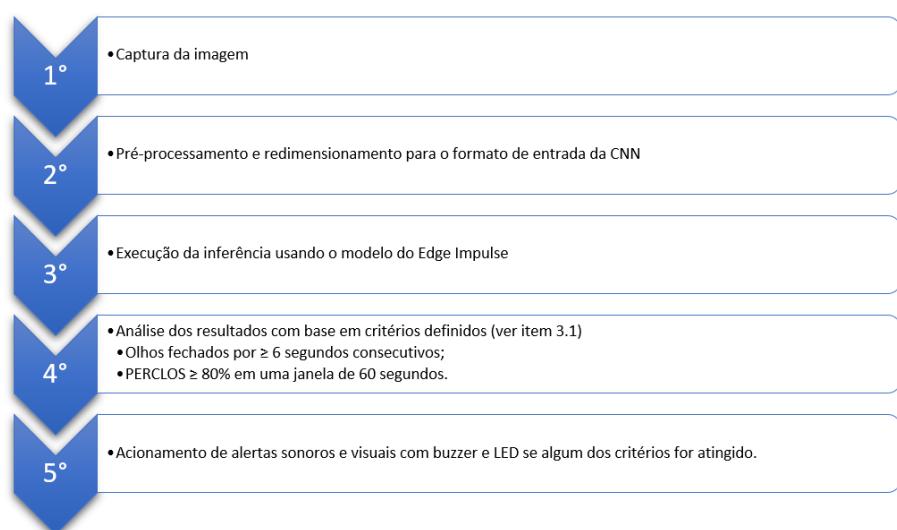
de imagens (escala de cinza, tamanho 96x96 e enquadramento de imagem *Fit Shortest*, no caso do projeto em questão); c) escolha da arquitetura de rede como *Transfer Learning* com *MobileNetV2* (otimizada para dispositivos embarcados); c) treinamento com validação cruzada e ajustes de parâmetros até atingir acurácia acima de 90% (no primeiro treinamento, obteve-se menos de 70%, porém com o aumento do *dataset*, refino do mesmo, e modificação de parâmetros do *transfer learning*, chegou-se em quase 95% de acurácia); d) exportação do modelo em formato *Arduino Library*, para ser integrado diretamente ao *firmware* da ESP32.

## 2.7 Implementação embarcada (ESP32-CAM)

A Figura 3 apresenta as etapas para o ciclo de inferência no dispositivo.

O modelo exportado foi incorporado a um código em C++ utilizando a IDE Arduino, que controla a captura de imagens pela câmera e executa a inferência localmente. O sistema foi programado para fazer a inferência em loop, com intervalo de 5 a 6 frames entre cada ciclo, respeitando as limitações de processamento da placa.

FIGURA 3 - Fluxograma do ciclo de inferência do dispositivo.

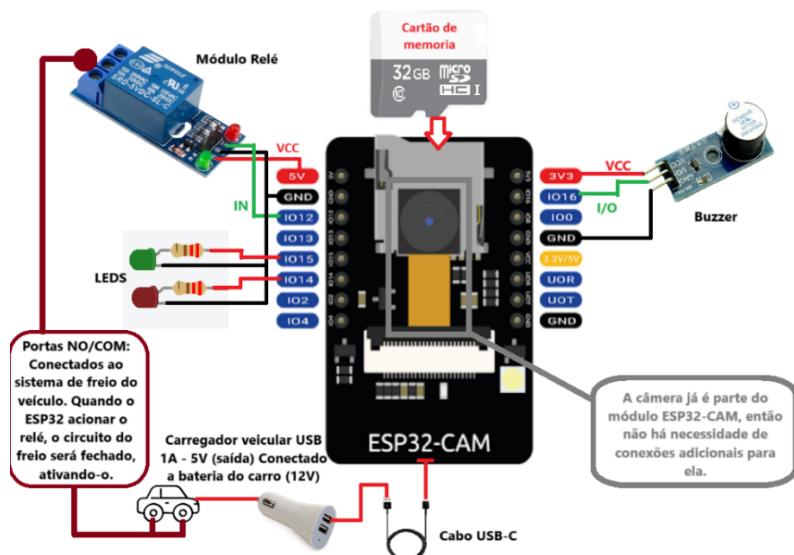


Fonte: elaborado pelos autores (2025).

## 2.8 Montagem física e prototipagem

A Figura 4 apresenta o diagrama de conexões do microcontrolador. A montagem do protótipo considerou os seguintes aspectos: a) Fixação da ESP32-CAM em uma case personalizada, próxima ao painel do carro ou simulador de direção, com foco direto no rosto do condutor; b) Conexão de um *buzzer* e um LED de alta intensidade, simulando os alertas sonoros e visuais, acionados por GPIOs; c) Alimentação feita via USB, podendo futuramente ser substituída por fonte automotiva de 5VDC regulada; d) Caixa (case) utilizada para acondicionar o protótipo, conforme Figura 5; e) A prototipagem foi feita em bancada e validada em ambiente controlado com diferentes condições de iluminação.

FIGURA 4 - Diagrama de conexões inicial do projeto com outras possibilidades de periféricos inseridos.



Fonte: elaborado pelos autores (2025).

FIGURA 5 - Desenho 3D da case.



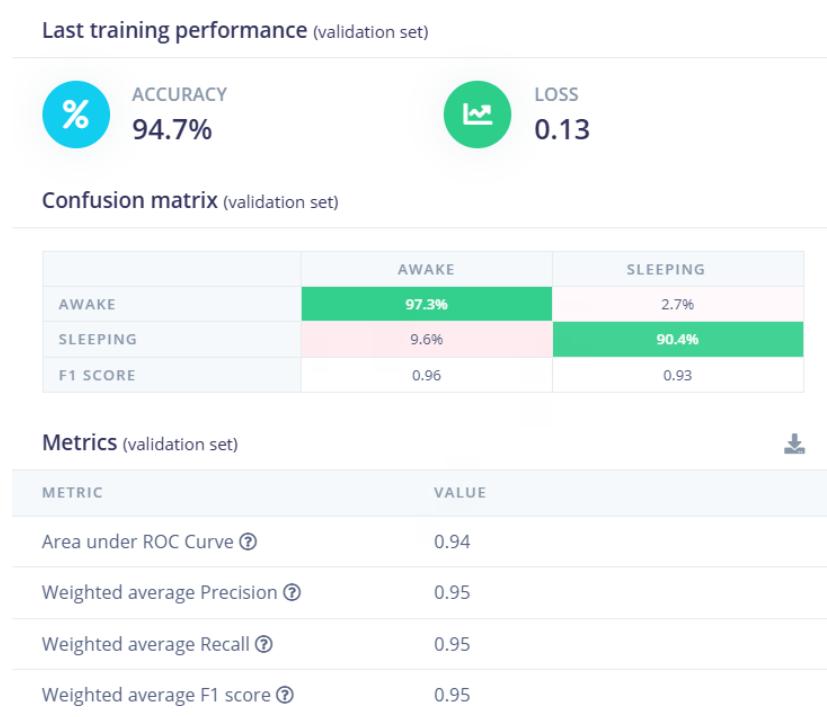
Fonte: Adaptado 877USER (2021).

### 3 Resultados e Discussão

Após obter as ferramentas e dados necessários, executar um treinamento com a mais alta precisão possível e desenvolver um projeto físico, o próximo passo foi realizar testes reais tanto do modelo de IA, quanto do protótipo. Portanto, esta seção apresenta os resultados obtidos e o impacto deles para o objetivo do projeto.

Primeiro, é importante entender como se saíram os resultados de treinamento do modelo de IA, pois ele foi determinante para a continuação do projeto. Na versão final do modelo de *ML no Edge Impulse*, com imagens de rosto recortado, o treinamento apresentou acurácia total de 94,7%, precisão (*sleeping*) de 90,4%, precisão (*awake*) de 97,3%, perda de validação (*val/loss*) de < 0,14 e tempo de inferência na ESP32-CAM de ~0,7s por imagem (Figura 6). Esses resultados foram satisfatórios, sobretudo considerando as limitações de processamento da plataforma ESP32-CAM, que possui recursos restritos para aplicações de visão computacional embarcada (Sanchez-Iborra; Skarmeta, 2020).

FIGURA 6 - Resultado final do treinamento da IA via *Edge Impulse*.

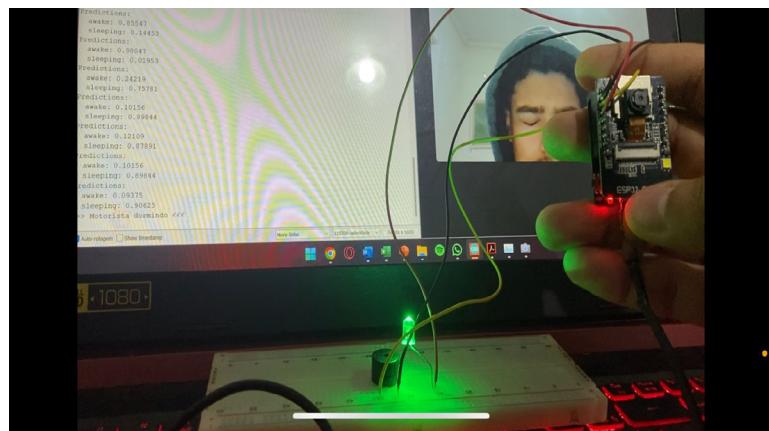


Fonte: elaborado pelos autores (2025).

Os resultados foram satisfatórios, especialmente considerando as limitações de processamento da ESP32-CAM. A rede conseguiu generalizar bem as imagens mesmo com variações leves de iluminação, postura e distância do rosto.

Na sequência, foram realizados experimentos utilizando imagens capturadas por *webcam* de *notebook*, câmera de celular e pela própria câmera embarcada da ESP32-CAM. Os testes confirmaram que o modelo manteve desempenho consistente quando a face estava bem iluminada e posicionada (Figura 7). Apesar disso, a necessidade de ajuste preciso do ângulo da câmera no interior do veículo mostrou-se um desafio para aplicações em cenários reais, conforme também apontado em estudos de sistemas embarcados de monitoramento de motoristas (Picot; Charbonnier; Caplier, 2012).

FIGURA 7 - Protótipo em funcionamento.



Fonte: elaborado pelos autores (2025).

Os testes com a ESP32-CAM confirmaram que o modelo mantinha um bom desempenho quando a face do condutor estava corretamente posicionada e iluminada. Mesmo que ainda a questão de posicionamento do dispositivo seja bem pontual e específica, se bem ajustado dentro do carro, pode se obter um bom desempenho, longe do ideal para ser considerada uma *ADAS*, mas um ponto de partida interessante.

Nos testes de simulação de sonolência, o sistema acionou os alertas com base nos cenários conforme planejado, mas no código foi colocado para que o primeiro item, em vez de 6 segundos, fosse analisado com 5 segundos, pois já se contou com um atraso de inferência por conta das limitações de processamento do ESP32. Em ambos os casos, o sistema acionou *buzzer* e LED por 5 segundos intermitentemente com efeito imediato.

Ainda, analisando-se a Figura 7, é possível observar alguns pontos importantes. Primeiro, está sendo utilizado um computador, porém o projeto nesse estágio não dependia do mesmo para processamento e inferência, estava apenas sendo usado para alimentação do microcontrolador e visualização de inferência via porta serial. Ou seja, o projeto mesmo que ainda montado em Placa de Circuito Impresso (PCI), já estava embarcado, e a utilização do computador serviu apenas como orientador quanto ao posicionamento da

câmera, pois o processador do ESP32 não suportou realizar a inferência e ao mesmo tempo gerar imagens ao vivo, e por isso se utilizou a porta serial (que estava gerando em tempo real as porcentagens da inferência para olho aberto e olho fechado). A imagem do rosto é advinda da *webcam* do *notebook*, e foi ligada apenas para realizar o registro fotográfico para comprovar que o dispositivo estava emitindo alerta ao identificar olho fechado por mais de cinco segundos, conforme esperado. O produto também identificou corretamente o segundo gatilho programado, o PERCLOS.

Depois de todos os testes feitos e constatado o bom funcionamento da prototipagem em bancada, o dispositivo foi inserido na case construída por impressora 3D, como pode ser visualizado na Figura 8.

FIGURA 8 - Dispositivo final montado na case e funcional.



Fonte: elaborado pelos autores (2025).

Foram observadas, durante o desenvolvimento, algumas limitações no projeto, as quais deverão ser mitigadas ou corrigidas em estudos subsequentes. Primeiro, a qualidade da câmera OV2640 (utilizada com o ESP32-CAM), apesar de barata, sofre com foco fixo e distorções sob baixa iluminação. Outro ponto observado foi a precisão do reconhecimento, que em alguns ângulos (rosto parcialmente oculto ou inclinado), o dispositivo não conseguiu identificar corretamente. Por último, é importante citar também a velocidade de infe-

rência, que embora seja aceitável ( $\sim 0,7$  segundos), o tempo entre capturas limita o número de frames por segundo e exige mais tempo para validar o critério de seis segundos.

Tais limitações estão alinhadas com desafios já documentados em sistemas de visão embarcados, como no detector desenvolvido por Alshaqaqi (2013) em sua pesquisa, reforçando a necessidade de ajustes futuros em *hardware* e *software*.

## 4 Considerações Finais

Este trabalho apresentou o desenvolvimento de um sistema embarcado de prevenção de acidentes veiculares por meio da detecção de sonolência em motoristas, utilizando técnicas de aprendizado de máquina e visão computacional, com a implementação prática em uma ESP32-CAM.

O processo contemplou a criação e reformulação de um banco de imagens, treinamentos sucessivos de modelos em Edge Impulse, seleção de uma arquitetura leve e compatível com dispositivos embarcados (MobileNetV2), e a implementação de critérios validados pela literatura científica para definição de estados de sonolência, como os 6 segundos com olhos fechados e o *Percent Eye Closure (PERCLOS)* acima de 80% em uma janela de um minuto. O protótipo desenvolvido demonstrou desempenho robusto para um sistema de baixo custo, conseguindo detectar com precisão comportamentos indicativos de fadiga e emitir alertas sonoros e visuais de maneira eficiente. Os testes práticos confirmaram que o sistema pode operar de forma embarcada e autônoma, tornando-o uma alternativa viável para veículos que não possuem ADAS.

Como possibilidades de trabalhos futuros, destaca-se a ampliação do conjunto de classes detectadas (como distração e bocejo), melhorias no sistema óptico da câmera, integração com sistemas reais de frenagem automática, armazenamento em nuvem para análise de comportamento e testes

mais amplos em ambientes reais com diferentes perfis de motoristas. Essas evoluções permitirão transformar o protótipo em uma solução mais robusta, aplicável em larga escala e com maior impacto na segurança veicular.

Em suma, o projeto validou a viabilidade de se desenvolver uma solução funcional, acessível e eficaz para a prevenção de acidentes causados por sonolência ao volante. Os resultados obtidos reforçam o potencial do uso de IA embarcada para promover segurança viária proativa, contribuindo para salvar vidas e reduzir custos com sinistros no trânsito.

## Agradecimentos

Os autores agradecem ao Instituto Federal de Educação, Ciência e Tecnologia de São Paulo – *Campus Cubatão* pelo suporte institucional, ao Prof. Dr. Enzo Bertazini, pelo acompanhamento no desenvolvimento do trabalho, e à empresa TECHSCAN, pelo suporte na montagem física do protótipo.

## Referências

- 877USER. **ESP32 CAM case for AI Thinker**, Thingverse, 2021. Disponível em: <https://www.thingverse.com/thing:4714663>. Acesso em 10 out. 2024.
- AL, H. **Driver Fatigue Detection Sleep**, Yawn, Kaggle, 2024. Disponível em: <https://www.kaggle.com/datasets/mhdalhafez/driver-fatigue-detection-sleepyawn>. Acesso em: 10 out. 2024.
- ALI, Y. Y. **Driver Drowsiness Detection**, Kaggle, 2024. Disponível em: <https://www.kaggle.com/code/youssefyasserali/driver-drowsiness-detection-pytorch/notebook>. Acesso em: 22 nov. 2024.
- ALSHAQAQI, B. et al. Driver drowsiness detection system. In: International Workshop on Systems, Signal Processing and their Applications (WoSSPA), 8., 2013, Algiers, Algeria: IEEE, 2013. Disponível em: <https://ieeexplore.ieee.org/document/6602353>. Acesso em: 22 nov. 2024
- BOSCH. **Bosh analysis**: driver assistance systems continue their strong advance, 2018. Disponível em: <https://www.bosch-presse.de/pressportal/de/en/bosch-analysis-driver-assistance-systems-continue-their-strong-advance-148032.html>. Acesso em: 10 out. 2024.

BOVOLATO, G. G.; DOMINGUEZ, G. A. R. **Desenvolvimento de sistema para identificação de nível de fadiga e sonolência em condutores de veículos por redes neurais artificiais.** 2023.

Trabalho de Conclusão de Curso (Graduação em Engenharia e Controle de Automação) - Instituto Federal de São Paulo, Cubatão, 2023. Disponível em: <https://repositorio.ifsp.edu.br/items/f254df9-fdbf-452c-b612-fe272b8ea26e>. Acesso em: 10 out. 2024.

CHENGULA, T. J. *et al.* Enhancing advanced driver assistance systems through explainable artificial intelligence for driver anomaly detection. **Machine Learning with Applications**, v. 17, p. 100580, 2024. Disponível em: <https://www.sciencedirect.com/science/article/pii/S2666827024000562>. Acesso em: 10 out. 2024.

DINGES, D. F.; GRACE, R. **PERCLOS: A valid psychophysiological measure of alertness as assessed by psychomotor vigilance.** Washington, D.C.: U.S. Department of Transportation, Federal Highway Administration, 1998. Disponível em: <https://rosap.ntl.bts.gov/view/dot/113>. Acesso em: 23 out. 2025.

DIPU, M. T. A. *et al.* Real-time driver drowsiness detection using deep learning. **International Journal of Advanced Computer Science and Applications**, v. 12, n. 7, 2021. Disponível em: <https://thesai.org/Publications/ViewPaper?Volume=12&Issue=7&Code=IJACSA&SerialNo=94>. Acesso em: 22 nov. 2024.

DJERARDA, I. E. **MRL Eye Dataset, kaggle**, 2023. Disponível em: <https://www.kaggle.com/datasets/imadeddinedjerrada/mrl-eye-dataset>. Acesso em: 22 nov. 2024.

EDGE IMPULSE. **Edge Impulse** – development platform for edge AI [s.d.]. Disponível em: <https://www.edgeimpulse.com>. Acesso em: 6 ago. 2025.

HEITMANN, A. *et al.* Technologies for the monitoring and prevention of driver fatigue. In: INTERNATIONAL DRIVING SYMPOSIUM ON HUMAN FACTORS IN DRIVER ASSESSMENT, TRAINING AND VEHICLE DESIGN. I., 2021. **Proceedings...** Aspen, Colorado, Iowa City, Public Policy Center of Iowa, 2021. Disponível em: <https://core.ac.uk/download/pdf/129642767.pdf>. Acesso em: 10 out. 2024.

HOSSAIN, M. D. U. *et al.* Automatic driver distraction detection using deep convolutional neural networks. **Intelligent Systems with Applications**, v. 14, p. 200075, 2022. Disponível em: <https://doaj.org/article/86bf287d3e404143a03313e06f47e9e9>. Acesso em: 10 de outubro de 2024.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. **Censo Demográfico**. Rio de Janeiro: IBGE, 2022. Disponível em: <https://www.ibge.gov.br/estatisticas/sociais/rendimento-despesa-e-consumo/22827-censo-demografico-2022.html>. Acesso em: 10 out. 2024.

INSTITUTO DE PESQUISA ECONÔMICA APLICADA. **Estudo aponta aumento de 13,5% em mortes no trânsito.** 2023. Disponível em: <https://www.ipea.gov.br/portal/categorias/45-todas-as-noticias/noticias/13899-estudo-aponta-aumento-de-13-5-em-mortes-no-transito>. Acesso em: 22 nov. 2024.

MAIS DE 40% DOS ACIDENTES DE TRÂNSITO ACONTECEM POR SONOLÊNCIA, AFIRMA A ABRAMET. Autopapo, 6 set. 2019. Disponível em: <https://autopapo.uol.com.br/noticia/mais-de-40-dos-acidentes-de-transito-acontece-por-sonolencia-afirma-a-abramet/>. Acesso em: 10 out. 2024.

MASALA, G. L.; GROSSO, E. Detecting driver inattention by rough iconic classification. In: IEEE INTELLIGENT VEHICLES SYMPOSIUM, 4., 2012. **Proceedings**...Sassari, IEEE, 2013. Disponível em: [https://iris.uniss.it/retrieve/e1dc1a2c-f62c-1507-e053-3a05fe0ac7a3/Masala\\_G\\_Detecting\\_driver\\_inattention\\_by.pdf](https://iris.uniss.it/retrieve/e1dc1a2c-f62c-1507-e053-3a05fe0ac7a3/Masala_G_Detecting_driver_inattention_by.pdf). Acesso em: 10 out. 2024.

MINISTÉRIO DOS TRANSPORTES. **Estudo aponta que mais de 50% dos acidentes de trânsito são causados por falhas humanas.** 2018. Disponível em: <https://www.gov.br/transportes/pt-br/assuntos/noticias/ultimas-noticias/estudo-aponta-que-mais-de-50-dos-acidentes-de-transito-sao-causados-por-falhas-humanas765>. Acesso em: 10 out. 2024.

NATIONAL HIGHWAY TRAFFIC SAFETY ADMINISTRATION. **Drowsy Driving Research and Program Plan.** U.S. Dept. of Transportation, 2016. Disponível em: [https://www.nhtsa.gov/sites/nhtsa.gov/files/drowsydriving\\_strategicplan\\_030316.pdf](https://www.nhtsa.gov/sites/nhtsa.gov/files/drowsydriving_strategicplan_030316.pdf). Acesso em: 10 out. 2024.

PICOT, A., CHARBONNIER, S., CAPLIER, A. On-line detection of drowsiness using brain and visual information. **IEEE Transactions on Systems, Man, and Cybernetics**, v. 42, n. 3, p. 764-775, 2012. Disponível em: <https://doi.org/10.1109/TSMCA.2011.2164242>. Acesso em: 20 out. 2024.

ROCHA, M. T.; DOMINGUES, D. W. B. P. **Aplicação de detecção de sonolência em motoristas utilizando Deep Learning.** Trabalho de Conclusão de Curso (Graduação em Engenharia Mecatrônica) – Universidade de São Paulo, São Paulo, 2022. Disponível em: <https://bdta.abcd.usp.br/directbitstream/7a44fe00-e260-480a-9791-cc222d5eab45/MARCELO%20TEIXEIRA%20ROCHA%20TCC-PMR.pdf>. Acesso em: 10 out. 2024.

SANCHEZ-IBORRA, R.; SKARMETA, A. F. TinyML- enabled frugal smart objects: challenges and opportunities. **IEEE Circuits and Systems Magazine**, v. 20, n. 3, p. 4-18, 2020. Disponível em: <https://scispace.com/papers/tinyml-enabled-frugal-smart-objects-challenges-and-3aznu26pld>. Acesso em: 15 out. 2024.

SILVA, T. M. **Controle automático do mecanismo de frenagem de um veículo não tripulado.** Projeto Final de Curso (Graduação em Engenharia de Controle e Automação) - Universidade Federal de Minas Gerais, Belo Horizonte, 2009. Disponível em: [http://www.coro.cpdee.ufmg.br/publications/tiago\\_mendonca.pdf](http://www.coro.cpdee.ufmg.br/publications/tiago_mendonca.pdf). Acesso em: 10 out. 2024.

WIERWILLE, W. W. *et al.* **Research on vehicle-based driver status/performance monitoring; development, validation, a refinement of algorithms for detection of driver drowsiness.** Technical report n. DOT HSs 808 247. Virginia: Department of Transportation, 1994. Disponível em: <https://rosap.ntl.bts.gov/view/dot/2578>. Acesso em: 22 nov. 2024.

WORLD HEALTH ORGANIZATION. **Death on the roads:** based on the WHO Global Status Report on Road Safety 2018. Brazil: WHO, 2018. Disponível em: [https://extranet.who.int/roadsafety/death-on-the-roads/#country\\_or\\_area/BRA](https://extranet.who.int/roadsafety/death-on-the-roads/#country_or_area/BRA). Acesso em: 22 nov. 2024.