**RELATÓRIO FINAL**

**RELATÓRIO FINAL DE ATIVIDADES DE INICIAÇÃO CIENTÍFICA**

**Sistema de Visão Computacional para Reconhecimento Inteligente de Objetos**

**vinculado ao projeto Sistemas Computacionais Inteligentes controle avançado com ênfase em Mapas Cognitivos Fuzzy Dinâmicos aplicado áreas circunscritas, em especial robótica**

**Mariana Gasparotto Palácios**

**Bolsista CNPq**

**Engenharia de Software**

**Data de ingresso no programa: 10/2024**

**Prof. Dr. Márcio Mendonça**

Área do Conhecimento: Sistemas de Computação

*CAMPUS* CORNÉLIO PROCÓPIO, 2025

**MARIANA GASPAROTTO PALÁCIOS**

**MÁRCIO MENDONÇA**

**Sistema Inteligente para Reconhecimento de Objetos Usando Visão Computacional**

Relatório de Pesquisa do Programa de Iniciação Científica ou Programa de Iniciação Tecnológica da Universidade Tecnológica Federal do Paraná.

*CORNÉLIO PROCÓPIO, 2025*

**SUMÁRIO**

RESUMO ...................................................................................................... 4

INTRODUÇÃO .............................................................................................. 5

METODOLOGIA ........................................................................................... 7

RESULTADOS E DISCUSSÕES .................................................................. 8

CONCLUSÕES ............................................................................................ 27

REFERÊNCIAS ............................................................................................ 28

| **RESUMO** |  |
| --- | --- |

Este trabalho propõe o desenvolvimento de um sistema de reconhecimento de objetos em trabalhadores com a utilização de visão computacional, tendo como foco identificar o uso de equipamentos de proteção individual (EPIs), como óculos de proteção, capacetes e máscaras, em trabalhadores da construção civil, onde o uso desses itens é essencial para a segurança. Foram utilizadas duas abordagens de redes neurais: Redes Neurais Convolucionais (CNNs) e Redes Neurais Artificiais (ANNs), para isso, a metodologia envolve a coleta de imagens de situações de uso de EPIs, a classificação dessas imagens e a comparação do desempenho das abordagens. Após a validação, notou-se a necessidade de um modelo mais robusto para detecção em tempo real e de múltiplos objetos simultaneamente, o que evidencia a necessidade de explorar em trabalhos futuros o modelo YOLOv9. Assim, os resultados atuais indicam a eficácia de ambas as técnicas e conclui-se o potencial do sistema para aplicações reais, com possibilidades de aprimoramentos futuros.

**Palavras-chave**: visão computacional, redes neurais, EPIs, segurança no trabalho.

**INTRODUÇÃO**

A segurança no ambiente de trabalho é uma preocupação fundamental, principalmente em setores de risco, assim, o uso de Equipamentos de Proteção Individual (EPIs) é essencial para garantir a integridade dos trabalhadores. EPIs comuns incluem capacetes, óculos de proteção e máscaras, itens que podem ser negligenciados por funcionários, colocando suas vidas em perigo (CHAVES LIMA et al., 2009).

Neste contexto, a aplicação de sistemas inteligentes para reconhecimento de objetos através de visão computacional é uma solução extremamente relevante para garantir o uso adequado desses equipamentos, visto que a visão computacional e as técnicas de aprendizado de máquina, como redes neurais, permitem desenvolver sistemas capazes de identificar se um trabalhador está utilizando os EPIs necessários em tempo real (REZENDE, 2003).

Redes neurais artificiais são técnicas computacionais inspiradas na estrutura neural de organismos inteligentes e são capazes de adquirir conhecimento por meio da experiência (RAUBER, 2005). Compostas por unidades computacionais chamadas de neurônios artificiais, essas redes apresentam características vantajosas como robustez, tolerância a falhas, paralelismo e a capacidade de processar informações incertas e aprender com dados incompletos (FERNEDA, 2006). Tais propriedades as tornam ideais para aplicações que exigem reconhecimento de padrões complexos, como a identificação de EPIs em imagens (NUNES; PRADO, 2002).

A visão computacional é a ciência que permite que máquinas simulem a visão humana, interpretando informações extraídas de imagens capturadas por câmeras e sensores de forma automatizada (DE MILANO; HONORATO, 2014). Ferramentas como a biblioteca OpenCV, desenvolvida pela Intel, tornam essa área ainda mais acessível, oferecendo centenas de funções voltadas à análise e ao processamento de imagens em tempo real, sendo amplamente utilizada em aplicações como robótica e interação homem-máquina (MARENGONI; STRINGHINI, 2009).

A proposta desta pesquisa trata-se da identificação de óculos de proteção, capacetes e máscaras, utilizando imagens capturadas, e, para isso, foram empregadas redes neurais artificiais, explorando tanto redes neurais convolucionais (CNNs) quanto redes neurais artificiais tradicionais (ANNs), visando comparar suas eficiências e limitações.

As redes neurais artificiais tradicionais (ANNs) são compostas por camadas de neurônios interconectados, onde cada neurônio de uma camada se conecta a todos os da próxima, sendo adequadas para tarefas gerais de classificação e regressão, não sendo tão eficientes no processamento de dados com estrutura espacial, como imagens (RAUBER, 2005). As redes neurais convolucionais (CNNs) foram projetadas especificamente para lidar com dados visuais, pois utilizam camadas de convolução capazes de identificar padrões locais, como formas e texturas, preservando a relação espacial entre os pixels, permitindo que as CNNs extraiam automaticamente características relevantes das imagens, tornando-as mais eficazes e precisas em reconhecimento de objetos em comparação com as ANNs (NUNES; PRADO, 2002).

A necessidade de uma solução mais eficiente e capaz de detectar múltiplos objetos em tempo real motivou a consideração de arquiteturas de detecção de objetos como o YOLOv9 (You Only Look Once), que permite identificar e localizar múltiplos itens simultaneamente em uma única imagem. Ao contrário de métodos mais antigos, que processam a imagem em múltiplas etapas, este utiliza uma abordagem unificada para prever classes de objetos de uma só vez. No contexto do reconhecimento de EPIs, o uso de uma arquitetura como o YOLOv9 será explorado em trabalhos futuros como uma alternativa promissora para superar as limitações das redes neurais clássicas, especialmente em cenários como a detecção de óculos, onde modelos anteriores apresentaram falhas.

Desse modo, o objetivo deste trabalho é desenvolver um sistema inteligente de reconhecimento de objetos usando redes neurais para identificar o uso de óculos, máscaras e capacetes em trabalhadores da construção civil, visando melhorar a segurança no ambiente de trabalho e proporcionando uma ferramenta eficaz para a gestão de EPIs e, consequentemente, reduzindo a incidência de acidentes laborais.

# METODOLOGIA

A etapa de coleta de dados utilizou um conjunto de 360 imagens, igualmente distribuídas entre as classes de interesse, compostas por óculos, capacete, máscara e as possíveis combinações entre estes itens, pré-processadas com redimensionamento para 224×224 px e normalização dos valores de pixel em [0,1]. As imagens foram armazenadas, para então dois modelos distintos serem implementados, um baseado em Redes Neurais Convolucionais (CNN) e outro em Redes Neurais Artificiais (ANN), os quais foram treinados por 70 épocas com monitoramento de loss e accuracy em treino e validação a cada época. Dessa forma, a performance dos modelos foi avaliada com base em gráficos gerados a partir das perdas e da acurácia coletadas de cada modelo.

O experimento inicial utilizou as imagens para implementar uma topologia de rede base para analisar as duas arquiteturas-base, ANN e CNN. Para aprofundar a análise de robustez e capacidade de generalização, cada arquitetura-base foi estendida com quatro topologias alternativas, com variações de complexidade, profundidade, normalização e pooling, totalizando cinco configurações testadas por família de rede:

* Variações de rede para arquitetura ANN
  1. *ReduzComplexidade* (menos neurônios por camada)
  2. *RedeMaisProfunda* (camadas ocultas adicionais)
  3. *BatchNormalization* (acelera a convergência e estabiliza o treinamento)
  4. *FunilInvertido* (ordem de camadas em “funil” invertido)
* Variações de rede para arquitetura CNN
  1. *BatchNormalization* (normaliza e acelera o aprendizado)
  2. *AumentaProfundidade* (uma rede mais profunda melhora a extração de padrões)
  3. *ArquiteturaRasa* (menos camadas convolucionais)
  4. *GlobalAveragePooling* (reduz o overfitting e simplifica a rede)

Para cada configuração foi gerado um conjunto de três gráficos, a curva de acurácia em treino/validação, curva de loss em treino/validação e matriz de confusão no teste, que permitem comparar diretamente desempenho e tendência de overfitting.

Após o treinamento e validação dos modelos, foram conduzidos testes práticos com imagens reais capturadas em cenários simulados de uso de EPIs, com o objetivo de avaliar o desempenho dos modelos em condições mais próximas da aplicação real. Foram testadas diversas combinações de equipamentos (capacetes de diferentes cores, máscaras, óculos e suas combinações), totalizando 10 amostras por classe, tanto para a arquitetura ANN quanto para a CNN. Esses testes práticos permitiram uma avaliação mais abrangente da capacidade de generalização dos modelos e evidenciaram limitações específicas, principalmente na detecção de óculos transparentes, o que motivou a consideração de modelos mais avançados como o YOLOv9 em trabalhos futuros, uma arquitetura de detecção de objetos em tempo real que permite identificar e localizar múltiplos EPIs simultaneamente em uma única imagem.

**RESULTADOS E DISCUSSÕES**

Ambos os modelos, CNN e ANN, apresentaram resultados relativamente satisfatórios em termos de precisão na identificação do uso de EPIs nas imagens, entretanto, as CNNs, como esperado, mostraram desempenho superior na extração de características visuais complexas, como formatos e texturas dos EPIs, enquanto as ANNs, embora com desempenho levemente inferior, também se mostraram viáveis para aplicações mais simples. Essa comparação demonstra a aplicabilidade dos dois métodos, considerando diferentes níveis de complexidade e recursos computacionais disponíveis. Foram obtidos resultados comparativos entre as duas abordagens a partir dos testes de treinamento tanto com a topologia de rede base, quanto das variações, conforme a Figura 1 e a Figura 2, que apresentam exemplos de imagens do dataset utilizado para as classes “capacete” e “capacete e máscara”, respectivamente.

No ANN base, a curva de acurácia evoluiu de cerca de 14 % na primeira época até estabilizar em torno de 36 % na época 70 (Figura 3), enquanto o loss decresceu de aproximadamente 6,3 para 1,1 (Figura 4). A matriz de confusão no conjunto de teste (54 imagens) indicou acertos de 12 óculos, 14 capacetes e 13 máscaras (Figura 5), resultando em val\_accuracy de 54,2 % e val\_loss final de 1,3060. Entre as quatro variações testadas — *ReduzComplexidade*, *RedeMaisProfunda*, *BatchNormalization* e *FunilInvertido* — a topologia com BatchNormalization destacou-se, atingindo val\_accuracy de 73,6 % e val\_loss de 1,0717, conforme exibido em Figura 12 (acurácia), Figura 13 (loss) e Figura 14 (matriz de confusão).

Na família CNN, o modelo base rapidamente alcançou acurácia de treino próxima a 100 % por volta da época 30 e manteve-a até o fim, com val\_accuracy de 93,06 % e val\_loss de 0,3538 (Figura 18 e Figura 19). Sua matriz de confusão (Figura 20) mostrou perfeita detecção em quase todas as classes, com apenas um leve desequilíbrio na classe de capacetes refletivos. Das quatro variações avaliadas — *BatchNormalization*, *AumentaProfundidade*, *ArquiteturaRasa* e *GlobalAveragePooling* — a configuração AumentaProfundidade foi a melhor, registrando val\_accuracy de 97,22 % e val\_loss de 0,1164 (ver Figura 24, Figura 25 e Figura 27), além de reduzir o gap entre treino e validação para menos de 5 pp.

A Tabela 1 mostra que, entre as ANNs, apenas a variação com BatchNormalization superou a base (val\_acc subiu de 55,6 % para 73,6 % e val\_loss caiu de 1,08 para 1,07), enquanto as demais tiveram desempenho inferior. Já a Tabela 2 evidencia que todas as CNNs melhoraram a base (93,1 % de val\_acc, loss 0,35), com destaque para AumentaProfundidade (97,2 % e 0,12), seguida por ArquiteturaRasa (94,4 %, 0,27) e GlobalAveragePooling (93,1 %, 0,25), confirmando a maior eficácia de redes convolucionais mais profundas.

Em síntese, embora a ANN com BatchNormalization tenha melhorado significativamente sua generalização em relação à versão base, foi a CNN AumentaProfundidade que obteve o melhor desempenho geral, equilibrando alta acurácia e baixa perda sem evidentes sinais de overfitting.

Além das avaliações em conjuntos de validação, testes práticos foram realizados com imagens reais representando diferentes configurações de uso de EPIs. Foram testadas 10 imagens para cada categoria, totalizando 90 exemplos por modelo (ANN e CNN). Os resultados reforçam as observações anteriores: a CNN, apesar de excelente desempenho na validação (val\_accuracy = 97,2%), apresentou falhas consistentes ao detectar óculos isoladamente (0 acertos em 10) e combinações com óculos, como "óculos + máscara" e "tudo" (1 ou nenhum acerto). Já a ANN, embora menos precisa no geral, teve desempenho excelente em capacetes e máscaras isoladas, com 10/10 acertos para capacete azul, branco e capacete + máscara. Entretanto, ambas as redes apresentaram dificuldades com óculos, o que motivou a análise de alternativas mais robustas.

Modelos como o YOLOv9 mostram-se promissores para superar essas limitações e deverão ser aplicados em projetos futuros para permitir a detecção de múltiplos objetos com maior robustez e em tempo real, oferecendo perspectivas de avanços significativos na detecção de EPIs em cenários reais.



Figura 1: Exemplo de imagem do dataset (classe capacete).



Figura 2: Exemplo de imagem do dataset (classe capacete e máscara).

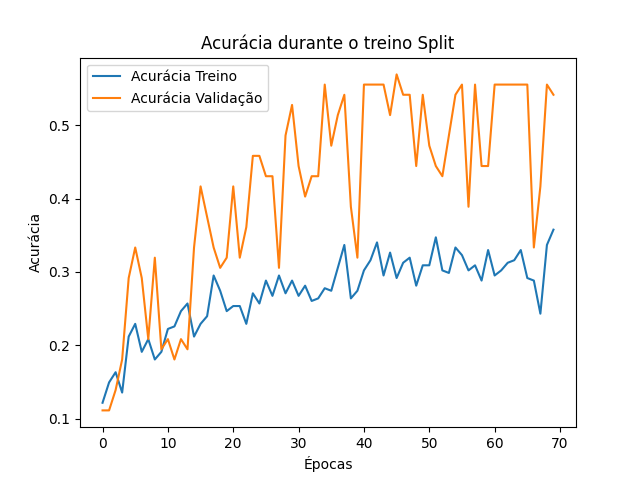


Figura 3: Acurácia da ANN base

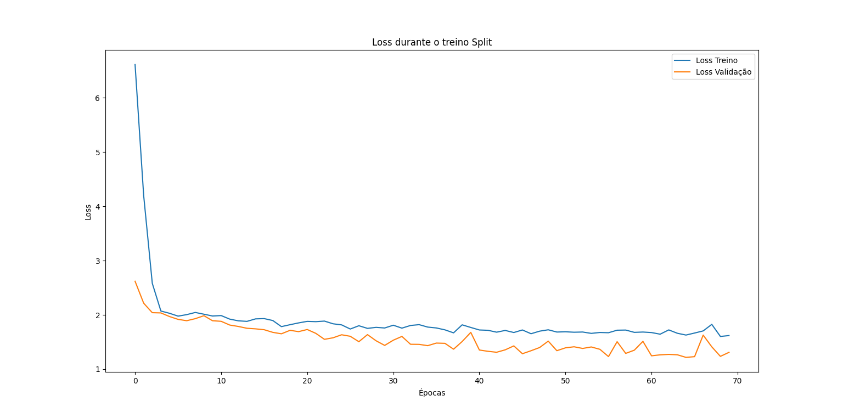


Figura 4: Loss da ANN base

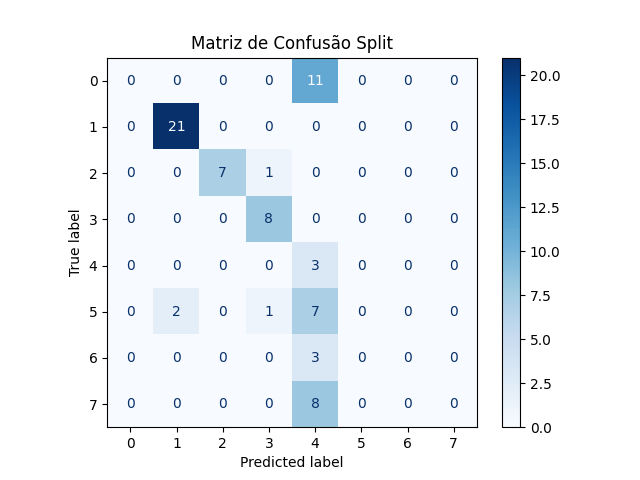


Figura 5: Matriz de confusão da ANN base

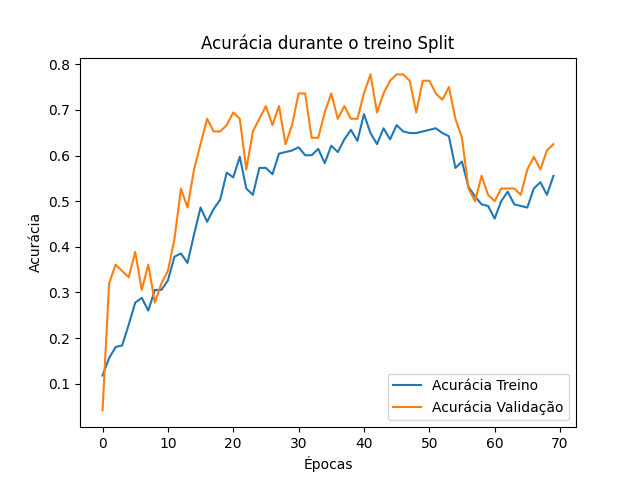


Figura 6: Acurácia da ANN ReduzComplexidade

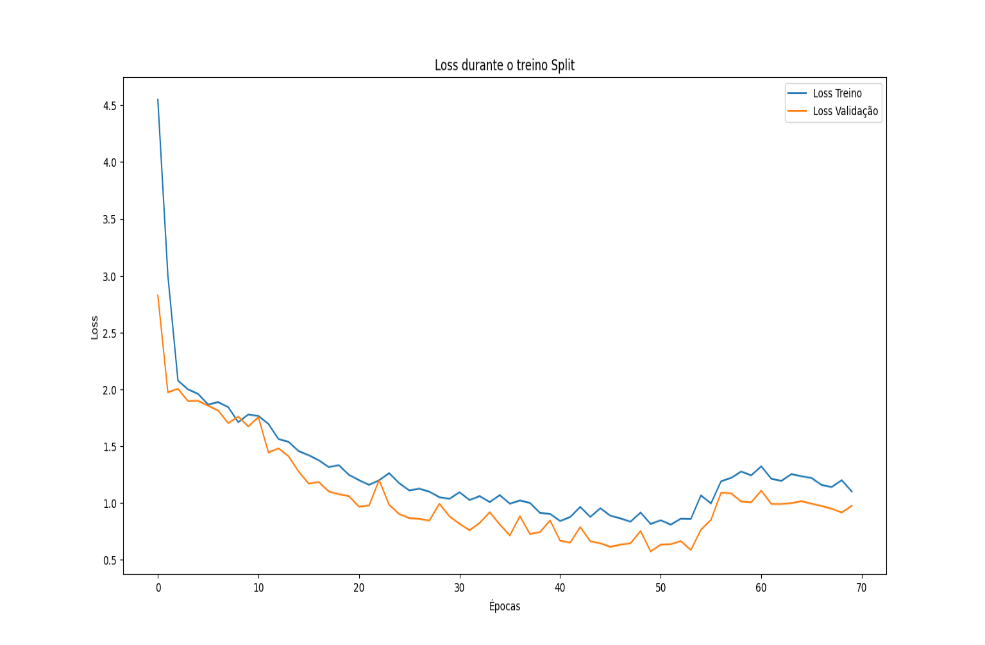


Figura 7: Loss da ANN ReduzComplexidade

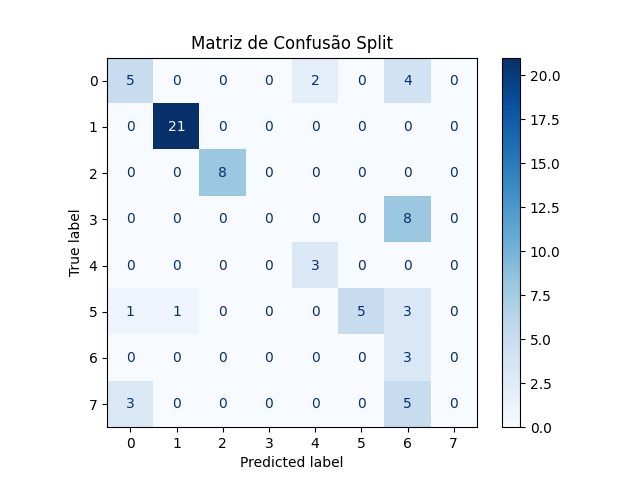


Figura 8: Matriz de confusão da ANN ReduzComplexidade

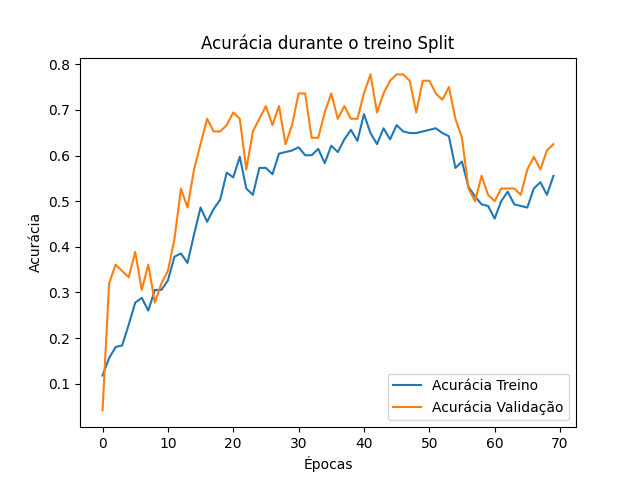


Figura 9: Acurácia da ANN RedeMaisProfunda

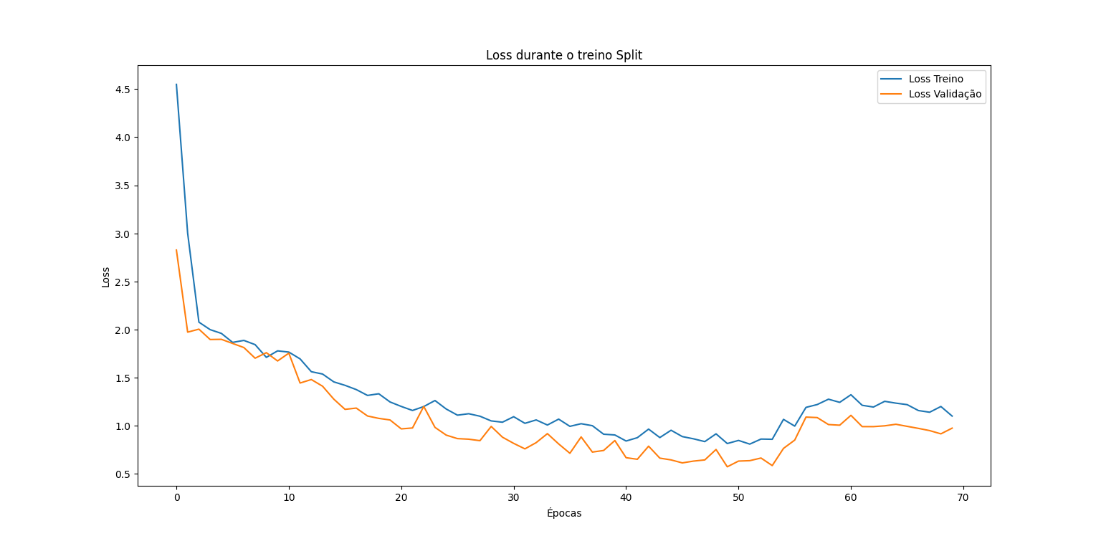


Figura 10: Loss da ANN RedeMaisProfunda

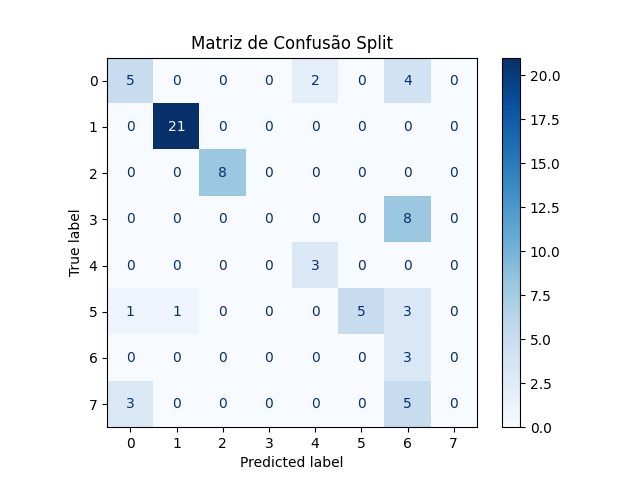


Figura 11: Matriz de confusão da ANN RedeMaisProfunda

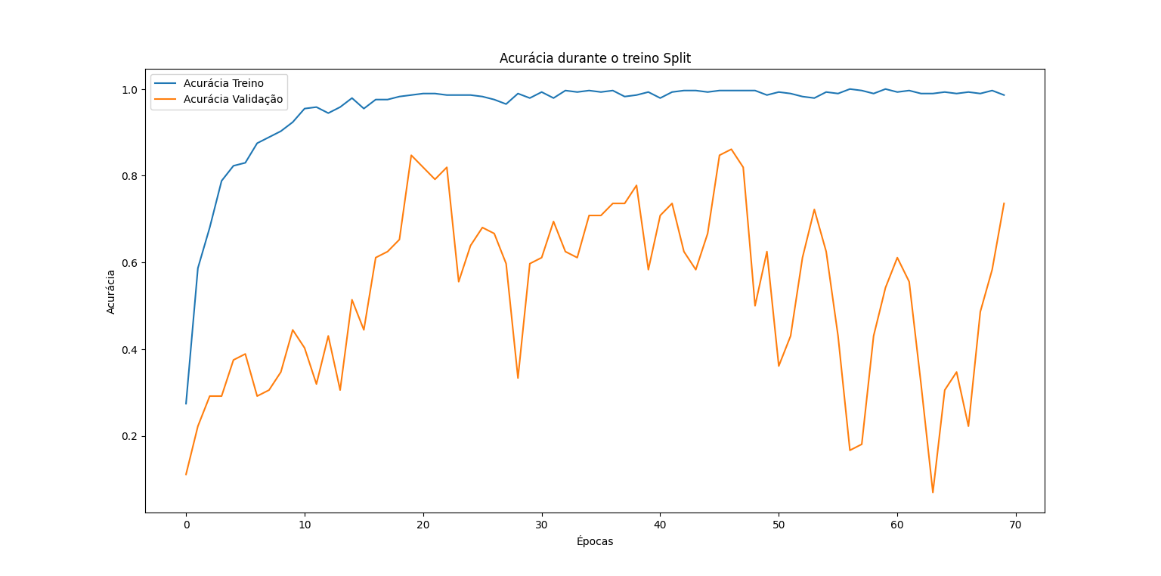


Figura 12: Acurácia da ANN com BatchNormalization

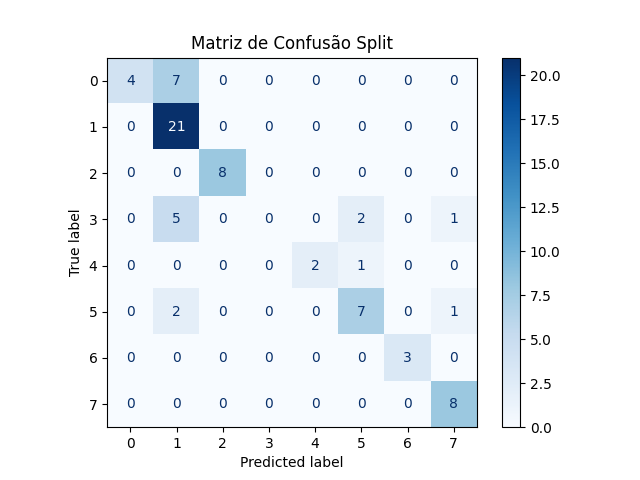


Figura 13: Matriz de confusão da ANN com BatchNormalization

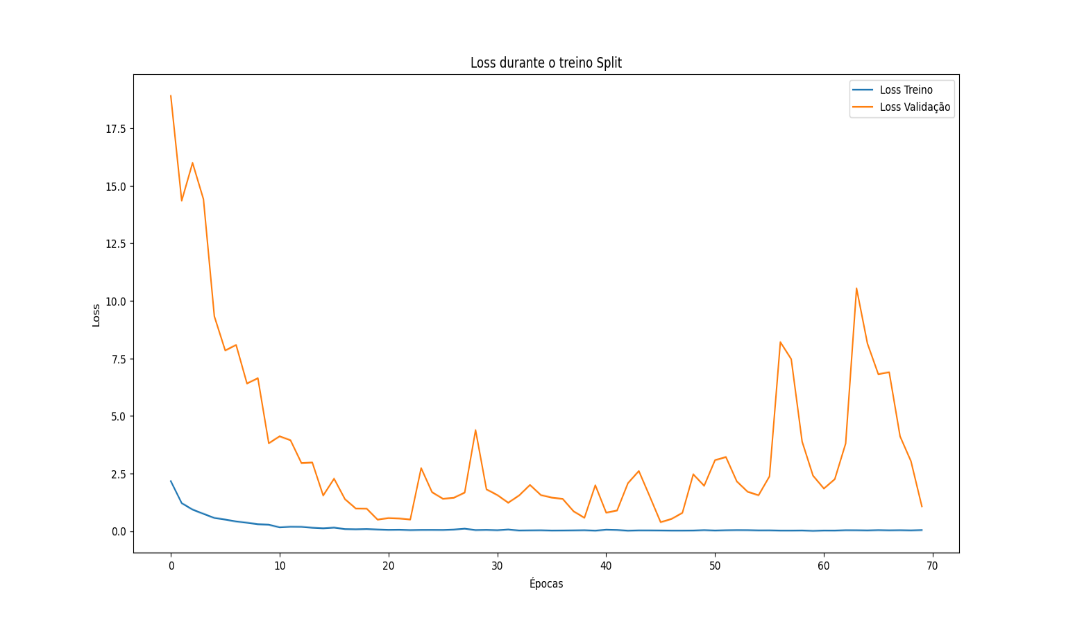


Figura 14: Loss da ANN com BatchNormalization

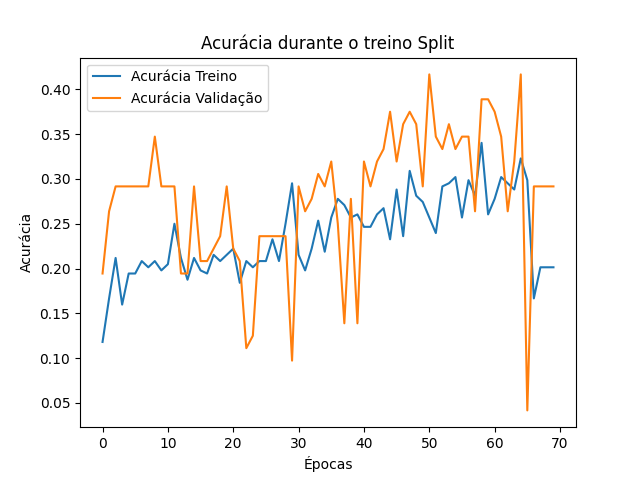


Figura 15: Acurácia da ANN FunilInvertido

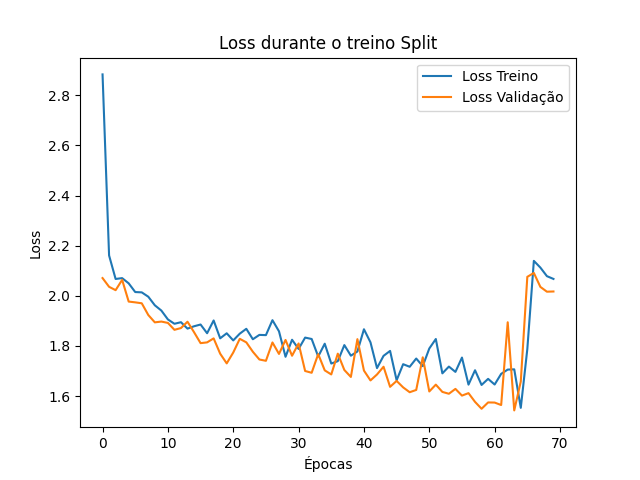


Figura 16: Loss da ANN FunilInvertido

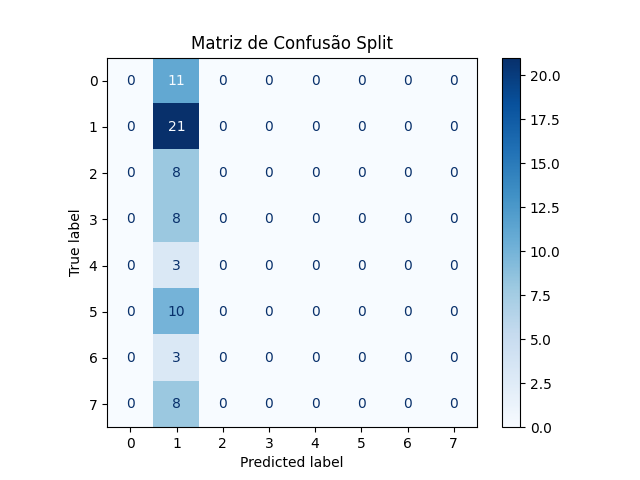


Figura 17: Matriz de confusão da ANN FunilInvertido

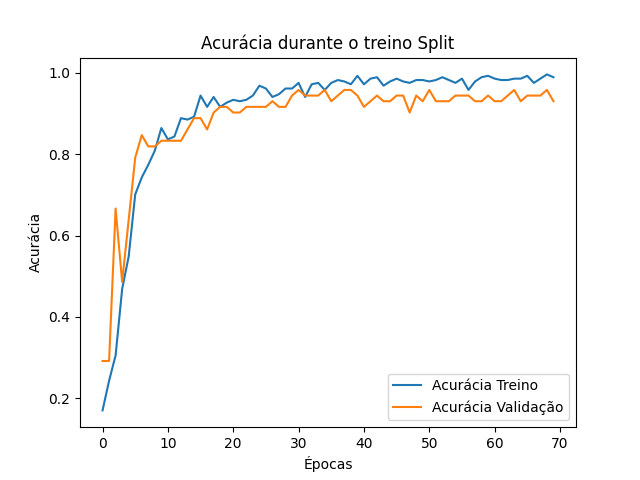


Figura 18: Acurácia da CNN base

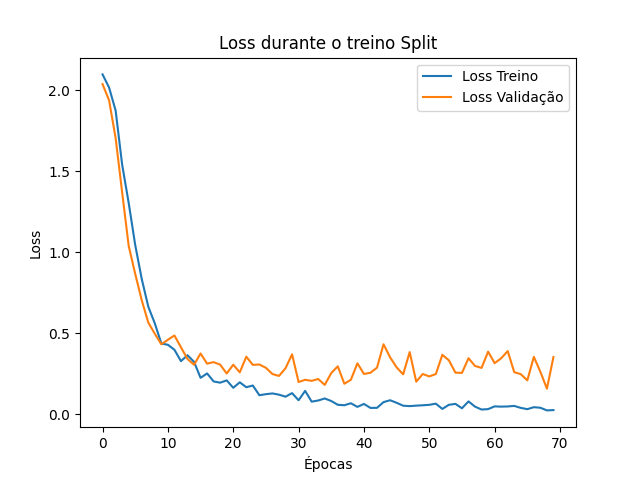


Figura 19: Loss da CNN base

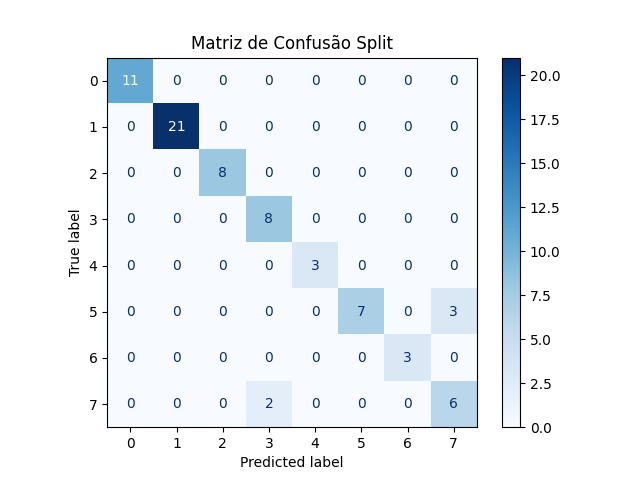


Figura 20: Matriz de confusão da CNN base

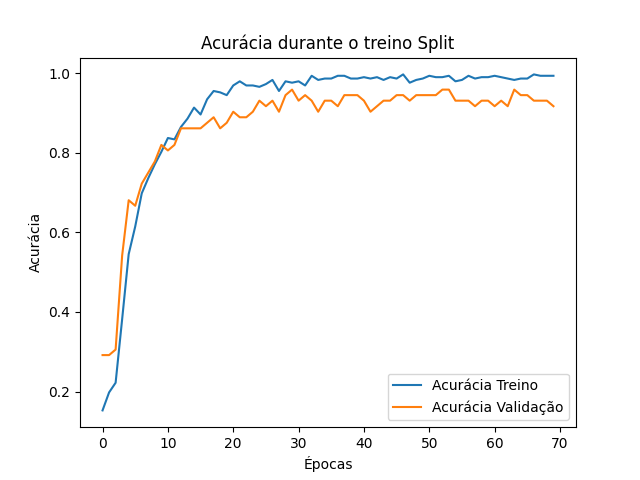


Figura 21: Acurácia da CNN com BatchNormalization

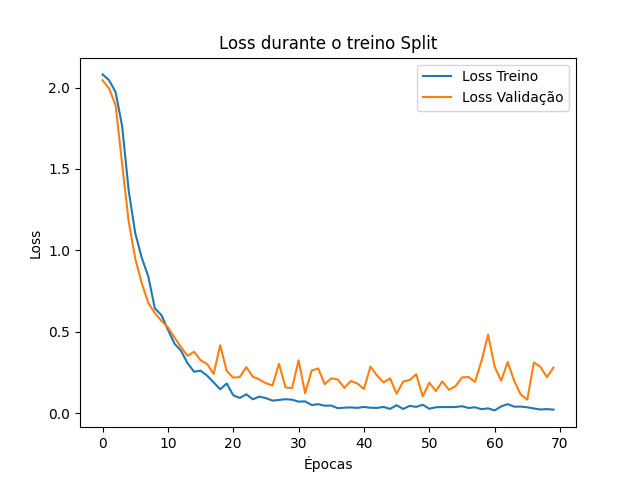


Figura 22: Loss da CNN com BatchNormalization

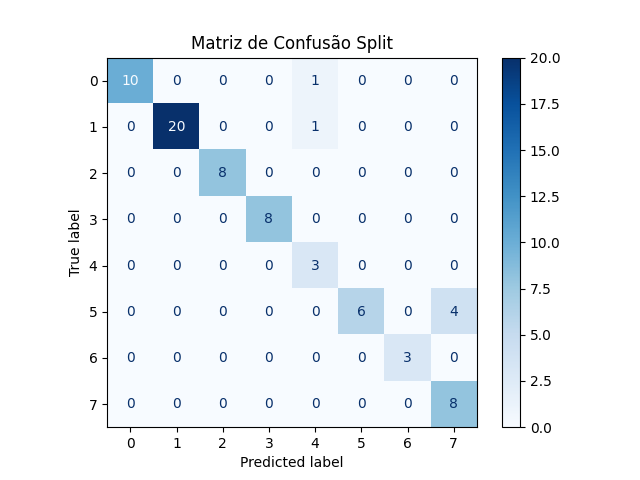


Figura 23: Matriz de confusão da CNN com BatchNormalization

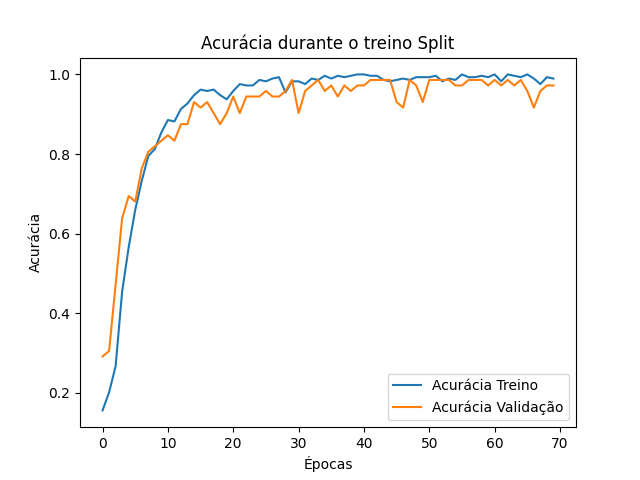


Figura 24: Acurácia da CNN AumentaProfundidade

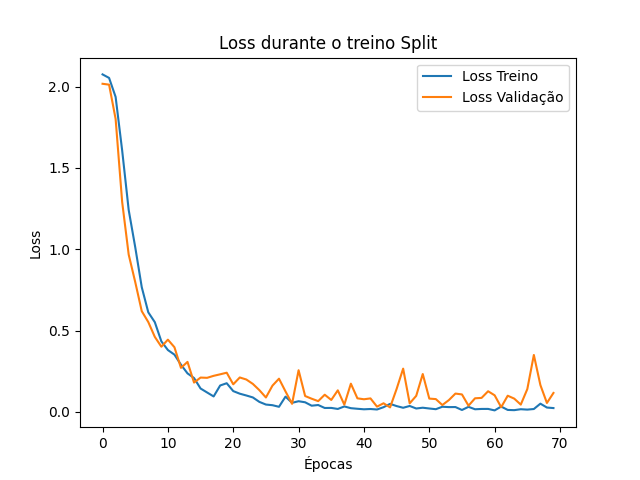


Figura 25: Loss da CNN AumentaProfundidade

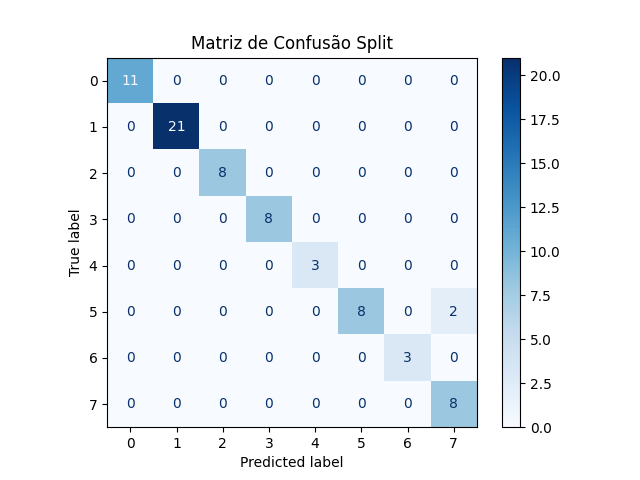


Figura 26: Matriz de confusão da CNN AumentaProfundidade

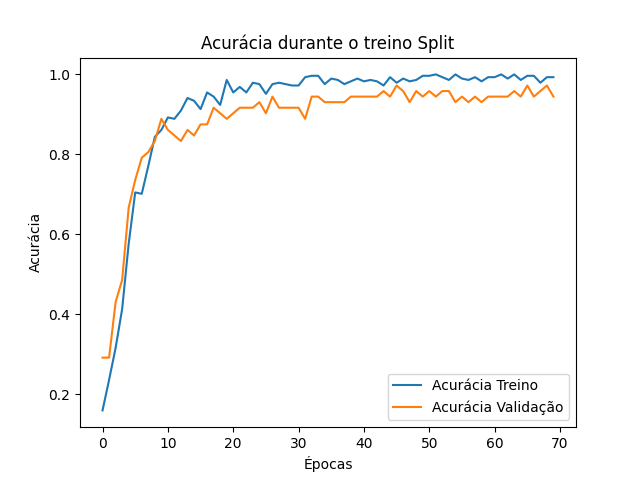


Figura 27: Acurácia da CNN ArquiteturaRasa

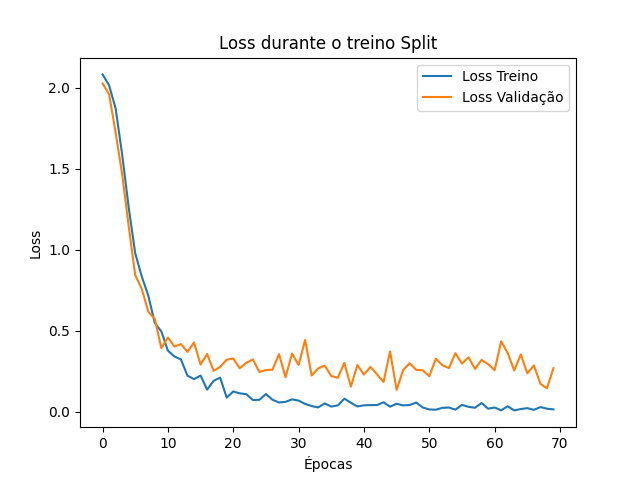


Figura 28: Loss da CNN ArquiteturaRasa

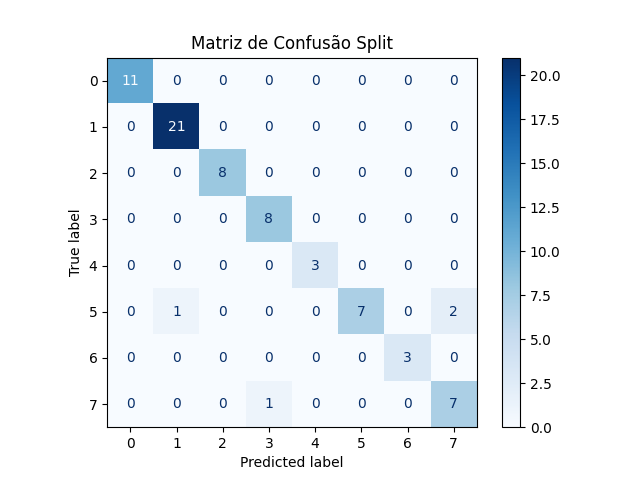


Figura 29: Matriz de confusão da CNN ArquiteturaRasa

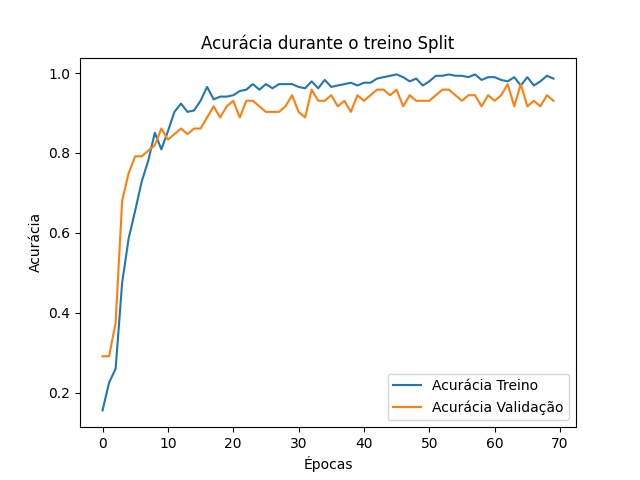


Figura 30: Acurácia da CNN GlobalAveragePooling

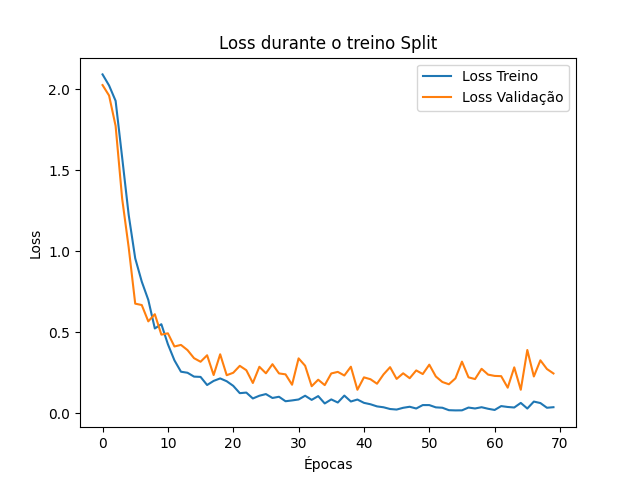


Figura 31: Loss da CNN GlobalAveragePooling

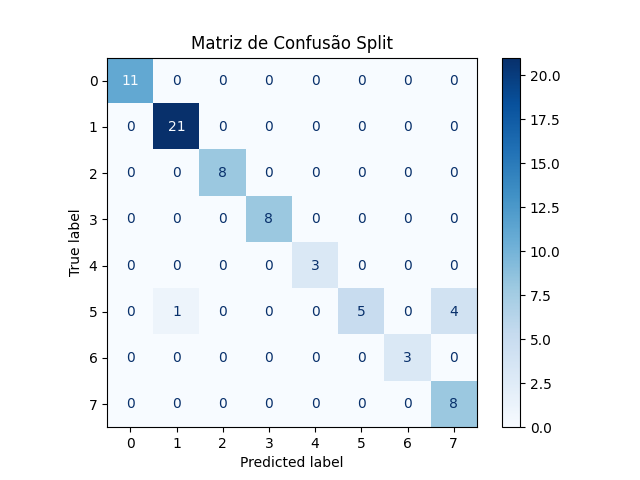


Figura 32: Matriz de confusão da CNN GlobalAveragePooling

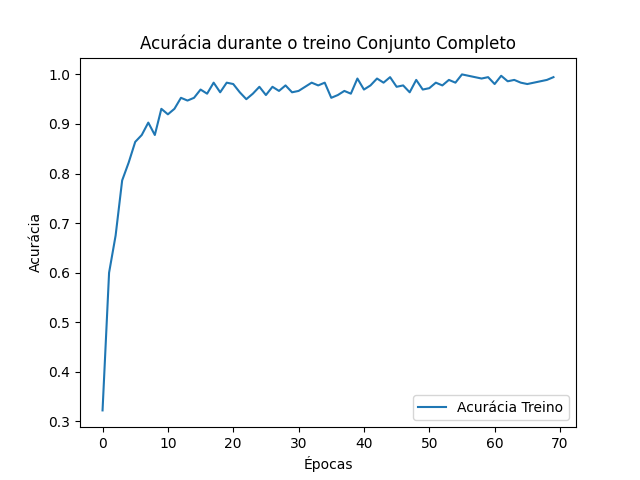


Figura 33: Acurácia do teste real da ANN

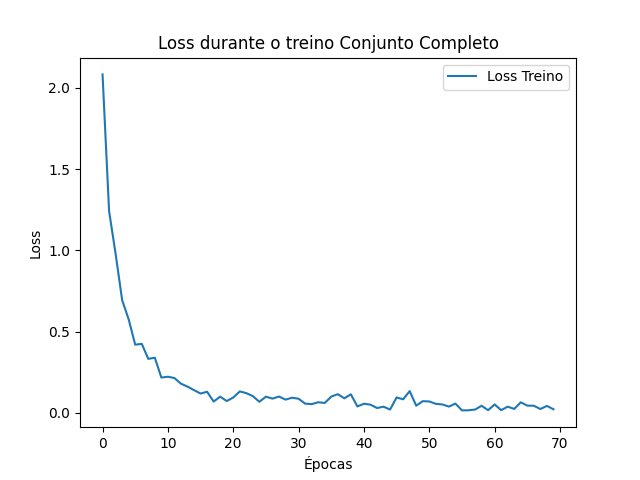


Figura 34: Loss do teste real da ANN

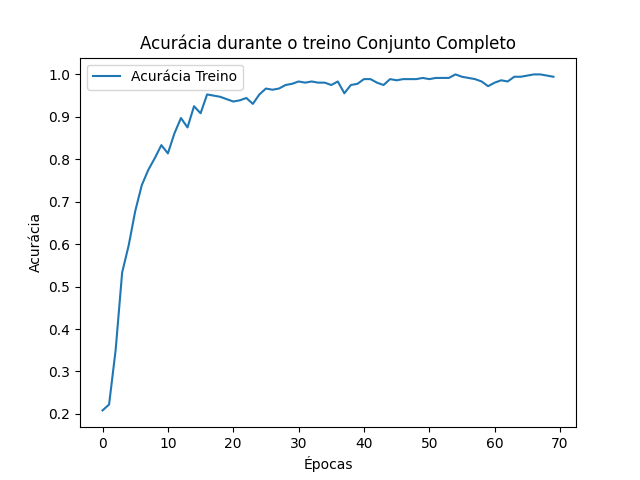


Figura 35: Acurácia do teste real da CNN

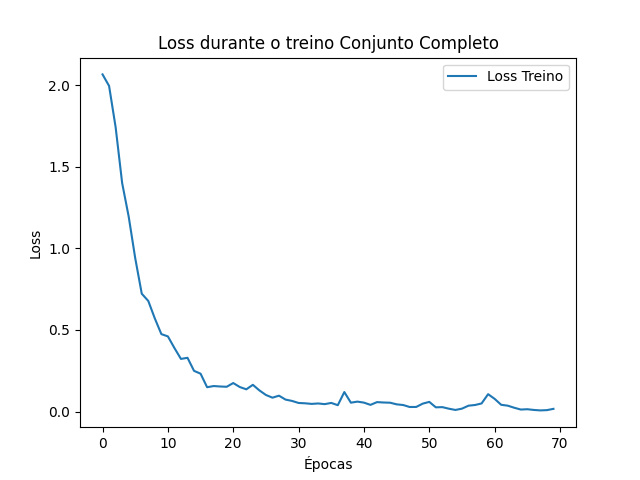


Figura 36: Loss do teste real da CNN

Tabela 1: Resultados da ANN

| Topologia | Loss Final | Acurácia Final |
| --- | --- | --- |
| Modelo Base Inicial | 0,3538 | 55,6 |
| ReduzComplexidade | 0,9752 | 62,50 |
| RedeMaisProfunda | 1,4956 | 33,33 |
| BatchNormalization | 1,0717 | 73,61 |
| FunilInvertido | 2,0174 | 29,17 |

Tabela 2: Resultados da CNN

| Topologia | Loss Final | Acurácia Final |
| --- | --- | --- |
| Modelo Base Inicial | 0,3538 | 93,06 |
| BatchNormalization | 0,2796 | 91,67 |
| AumentaProfundidade | 0,1164 | 97,22 |
| ArquiteturaRasa | 0,2685 | 94,44 |
| GlobalAveragePooling | 0,2462 | 93,06 |

Tabela 3: Acertos por classe em teste real da ANN

|  | Erros | Acertos |
| --- | --- | --- |
| Capacete azul | 0 | 10 |
| Capacete branco | 0 | 10 |
| Máscara | 3 | 7 |
| Óculos | 10 | 0 |
| Capacete e máscara | 0 | 10 |
| Capacete e óculos | 9 | 1 |
| Óculos e máscara | 4 | 10 |
| Tudo | 8 | 2 |
| Sem nada | 10 | 0 |

Tabela 4: Acertos por classe em teste real da CNN

|  | Erros | Acertos |
| --- | --- | --- |
| Capacete azul | 9 | 1 |
| Capacete branco | 5 | 5 |
| Máscara | 1 | 9 |
| Óculos | 10 | 0 |
| Capacete e máscara | 2 | 8 |
| Capacete e óculos | 3 | 7 |
| Óculos e máscara | 10 | 0 |
| Tudo | 9 | 1 |
| Sem nada | 5 | 5 |

**CONCLUSÕES**

A pesquisa confirmou que sistemas de visão computacional baseados em redes neurais são viáveis para detecção automatizada de EPIs em ambientes de trabalho. Entre as ANNs, a variação que incluiu BatchNormalization elevou a acurácia de validação de 55,6 % para 73,6 % e reduziu o loss de 1,08 para 1,07, mas ainda ficou atrás das CNNs. Na família convolucional, todas as variações superaram o modelo base (93,1 % de val\_acc e loss 0,35), com destaque para a topologia AumentaProfundidade (97,2 % de val\_acc e loss 0,12). Esses resultados confirmam que, embora ANNs simples possam atender a cenários de baixo custo computacional, são as CNNs mais profundas, especialmente com camadas adicionais, que oferecem o melhor equilíbrio entre precisão, robustez e generalização. O sistema final pode ser ainda aprimorado, estendendo sua aplicabilidade para monitoramento real-time e garantindo maior segurança aos trabalhadores e a fiscalização eficiente do uso de EPIs.

Os testes práticos reforçaram os padrões observados na validação: a CNN demonstrou acurácia elevada em classes como “máscara” e “capacete + máscara”, mas apresentou desempenho inferior em identificar óculos isoladamente, mesmo com boa generalização aparente no treinamento. A ANN, embora com menor desempenho geral, foi mais confiável para capacetes.

O YOLOv9 foi identificado como uma alternativa promissora para superar as limitações encontradas — em especial na detecção de óculos, prejudicada pela transparência — e será aplicado em trabalhos futuros para expandir as capacidades do sistema, permitindo detecção de múltiplos objetos com alta precisão e em tempo real.

# REFERÊNCIAS

[1] NUNES, L.E.; PRADO, P.P. RECONHECIMENTO DE OBJETOS CONTIDOS EM IMAGENS ATRAVÉS DE REDES NEURAIS. Periódicos Unitau, Taubaté, v. 5.8, p. 77-88, 1999-2002.

[2] RAUBER, T.W.; Redes Neurais Artificiais. Universidade Federal do Espírito Santo, v. 29, p. 39, 2005.

[3] REZENDE, Solange Oliveira. Sistemas inteligentes: fundamentos e aplicações. Editora Manole Ltda, 2003.

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

[5] DE MILANO, Danilo; HONORATO, Luciano Barrozo. Visao computacional. UNICAMP Universidade Estadual de Campinas FT Faculdade de Tecnologia, 2014.

[6] FERNEDA, Edberto. Redes neurais e sua aplicação em sistemas de recuperação de informação. Ci. Inf., Brasília, v. 35, n. 1, p. 25-30, jan./abr. 2006.

[7] CHAVES LIMA, Sonia Cristina; MARTINS DE LEÃO, Inez Cristina; SOUSA SANTANA, Vilma; NOGUEIRA DE SANTANA, Jusiene; ARAGÃO DE ALMEIDA LACERDA, Lívia Maria. Determinantes da implantação de um programa de segurança e saúde no trabalho. Rev Panam Salud Publica/Pan Am J Public Health 25(3), 2009.

[8] ZHAO, L.; SHEN, Y.; WANG, Y. Real-Time Helmet Detection Based on YOLOv3. IEEE Access, v. 7, p. 165394–165401, 2019.

[9] HARISH, M.; MAHABOOB BASHA, S. A Lightweight Convolutional Neural Network for Hard Hat Detection. International Journal of Advanced Computer Science and Applications, v. 10, n. 7, p. 243–249, 2019.

[10] DENG, J.; GUO, Z.; XU, T.; WANG, Y. Construction Site Safety Helmet Wearing Detection Based on Improved Faster R-CNN. Sensors, v. 20, n. 3, 2020.

[11] LIU, W.; ANGUELOV, D.; ERHAN, D.; SZEGE DY, C.; REED, S.; FU, C.-Y.; BER G, A.C. SSD: Single Shot MultiBox Detector. In: European Conference on Computer Vision (ECCV), LNCS 9905, p. 21–37, 2016.

| \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ | \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ |
| --- | --- |
| Márcio Mendonça  Orientador | Mariana Gasparotto Palácios  Bolsista |