

Detecção de Posturas Humanas em Superfícies de Repouso com o Uso de Redes Neurais Convolucionais

Daniel B. Maia¹

¹Programa de Pós-Graduação em Computação Aplicada (PPCA) – Universidade Federal do Pará (UFPA) – Campus Tucuruí
Tucuruí – PA – Brasil

{daniel.maia@tucurui.ufpa.br}

Resumo. Este artigo visa desenvolver uma rede neural convolucional que seja capaz de detectar e classificar duas posturas humanas enquanto o indivíduo se encontrar em uma superfície de repouso. O objetivo consiste em diferenciar as posturas de uma pessoa sentada das posturas de uma pessoa deitada, para poder auxiliar no desenvolvimento de tecnologias assistivas que possam prevenir possíveis quedas. Para realizar o treinamento da rede neural convolucional, foi produzido e preparado um banco de imagens próprio de forma a alimentar a CNN com imagens de indivíduos deitados e sentados. Essa base de dados será composta de imagens de pessoas reais em diversas posições e posturas e de bonecos representando outros tipos de posturas e contextos. Ao final, a rede é avaliada obtendo resultados satisfatórios de 95% de acurácia. Além disso, para a classe deitada, os resultados são 91% de precisão, 98 % de revocação e 95% de F1 score. Enquanto que para a classe sentada os resultados são 99% de precisão, 92% de revocação e 95 % de F1 Score.

1. Introdução

O presente trabalho tem como foco treinar uma rede neural convolucional através de métodos de classificação supervisionada em um modelo binário de posturas humanas: pessoas deitadas e sentadas em uma superfície de repouso. O principal objetivo desta detecção de posturas é formular uma possível ferramenta de tecnologia assistiva que possa auxiliar um sistema de alerta e monitoramento de pessoas com mobilidade reduzida.

As quedas dentro do público idoso, principalmente acima de 80 anos, costumam ser desencadeadores de diversos outros problemas de saúde. Foi constatado que, em decorrência das quedas, há um forte declínio na qualidade de vida e, consequentemente, na expectativa de vida daquele idoso [3]. Assim, torna-se relevante desenvolver tecnologias assistivas que sejam capazes de prevenir que essas quedas ocorram.

2. Trabalhos Relacionados

Boa parte dos estudos voltados para a área de detecção de posturas utilizam redes neurais convolucionais, a qual é a mesma técnica adotada por este estudo, porém na maioria deles a CNN é combinada com outras técnicas e métodos, como o uso de sensores de pressão[4].

Após a pesquisa em trabalhos relacionados, ficou evidente uma lacuna na literatura quanto a detecção com uso exclusivo de CNNs para a identificação específica da postura de indivíduos sentados em superfícies de repouso e a diferenciação dessa postura da de um indivíduo deitado na mesma superfície. A maioria dos trabalhos analisados que detectam esta posição específica fizeram uso de outras técnicas além das

CNNs, o que obteve resultados eficientes, porém com uma complexidade mais alta, seja na logística ou no custo dos projetos[5][6]. Além disso, nenhum dos trabalhos analisados possui o foco em detectar posturas humanas durante o período de repouso, ou com foco específico nas posturas deitadas e sentadas.

3. Metodologia

3.1 Banco de Imagens

Inicialmente, foi criada uma maquete simples, com a utilização de um boneco manequim para representar o sujeito monitorado e uma caixa de papelão para representar a cama. Um total de 150 imagens foram utilizadas nesta etapa, sendo 75 da categoria de pessoa deitada e 75 da categoria de pessoa sentada. Porém, apenas imagens de maquetes não são suficientes para resultados reais, já que os resultados obtidos foram insuficientes.

Assim, também foram realizados testes com pessoas reais que aceitaram participar do experimento, onde suas diferentes posturas foram adicionadas tanto ao banco de dados de entrada quanto ao banco de dados de teste. Cinco sujeitos participantes gravaram vídeos, com durações entre 30 segundos e 1 minuto, de cada um se movendo em uma cama, passando da postura deitada para a postura sentada. Cada vídeo possui ângulos, cenários e iluminações diferentes. Isso se deve para tentar abranger um máximo possível de situações que a rede neural possa detectar.

Então, foi feita uma conversão, transformando os *frames* dos vídeos em imagens para serem agregadas ao banco de imagens final. Ao fim, o banco de imagens ficou com um total de 700 imagens, sendo 327 de posturas deitadas, e 373 de posturas sentadas. Na Figura 1, há exemplos das imagens utilizadas.



Figura 1. Exemplos de imagens com pessoas reais utilizadas no banco de dados: (a) indivíduo deitado e (b) indivíduo sentado.

3.2 Preparação das Imagens

Cada uma das duas classes (Sentada e Deitada) deverá ter suas imagens divididas em três grupos principais: Imagens de Treinamento, Imagens de Validação e Imagens de Teste.

A recomendação da literatura sobre a divisão de forma a evitar o *overfitting*, é definir em torno de 70% das imagens disponíveis para o grupo de imagens de treinamento.[1] As imagens restantes serão divididas entre os grupos de validação e teste. Assim, foi definido um total de 10% para as imagens de validação e 20% para as imagens de teste.

Alguns parâmetros básicos são definidos: o padrão escolhido foi o tamanho 50 x 50 pixels, um tamanho considerado suficiente para a análise posturas corporais e a escala cinza, pois é uma escala que torna a definição de características exclusivas de cada figura mais simples [4].

O formato das imagens foram reajustados para que suas informações fossem armazenadas em um array (x), contendo o índice de cada imagem, as dimensões (50x50) e o valor 1 (que representa apenas 1 escala de cor, nesse caso, cor cinza). Foi gerado também um segundo array (y) relacionado ao primeiro através dos índices, em que serão armazenados os rótulos de cada imagem, ou seja, a qual casa cada imagem pertence.

Também foi feita uma normalização dos pixels da imagem, ou seja, os pixels que variavam entre 0 e 255 foram convertidos para valores entre -0,1 e 1,175. Isso ocorreu para garantir que a média dos valores seja 0 e o desvio padrão 1, o que proporcionou uma redução relevante no tempo de treinamento.

3.3 Treinamento da Rede Neural Convolucional

A rede neural utilizada é baseada na arquitetura LeNet e é ideal para classificações binárias, sendo utilizada neste estudo para diferenciar as posturas deitadas de sentadas. No caso da CNN deste trabalho, foi utilizada a função de perda *sparse_categorical_crossentropy*. Foi definido um otimizador, que será o responsável por reajustar os pesos utilizados de forma a diminuir a quantidade de erros. O selecionado foi o otimizador Adam. O processo de treinamento foi realizado por 20 épocas, agregando as imagens de 50 por lote. Foram utilizadas as seguintes métricas de avaliação dos resultados: acurácia, precisão, revocação, F1 Score e a Matriz de Confusão.

4. Análise dos Resultados

De um total de 700 imagens que compunham o banco de dados, foram reservadas um total de 140 imagens para fazer parte do conjunto de teste do experimento, o que representa 20 % do total. A Tabela 1 apresenta os resultados obtidos.

Tabela 1 – Resultados do Experimento

CNN	Acurácia	Precisão	Revocação	F1 Score
Classe Deitada (65 imagens)	95%	91%	98%	95%
Classe Sentada (75 imagens)	95%	99%	92%	95%

A Figura 2 apresenta a matriz de Confusão obtida.

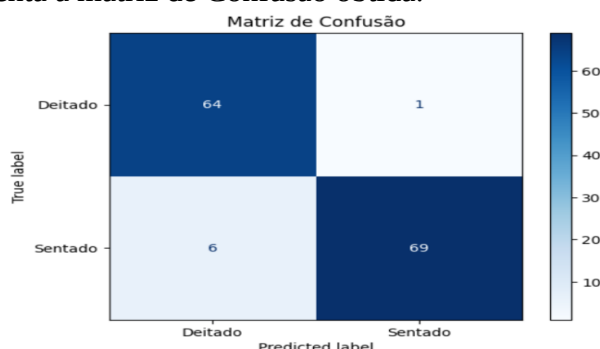


Figura 2. Matriz de Confusão

A Figura 3 apresenta os gráficos que evidenciam que a partir da quinta época o treinamento já pôde ser considerado bem-sucedido já que as taxas de acurácia e de perda se mantiveram estáveis até o final do mesmo.

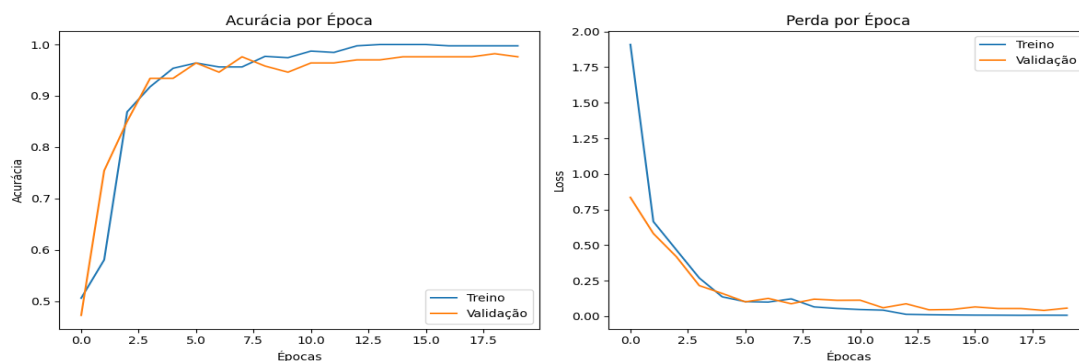


Figura 3. Gráficos de Acurácia e Perda por Época

Com os resultados obtidos, é possível assumir que a rede teve uma excelente eficiência nas classificações, pois com resultados de acurácia em 95%, fica demonstrado que em sua grande maioria as imagens das posturas foram corretamente identificadas. Os resultados das outras métricas ratificam esse dado.

5. Conclusão

O conteúdo exposto neste trabalho exemplificou a relevância da utilização de um banco de imagens próprio que possa contribuir na detecção de posturas humanas, ainda não tão amplamente detectadas, como a posição de uma pessoa sentada ou deitada em superfícies de repouso. Foi criada uma rede neural convolucional, que ao utilizar o banco de dados próprio desenvolvido neste estudo, conseguiu de forma bem sucedida classificar imagens entre pessoas sentadas e deitadas, com uma acurácia de 95%.

Os resultados obtidos neste artigo podem ser utilizados em trabalhos futuros para se desenvolver um novo método para aprimorar as tecnologias assistivas existentes, ou até mesmo, desenvolver novas tecnologias com o uso da inteligência artificial, através das redes neurais convolucionais, como ferramenta principal.

Referências

- [1] Data Science Academy. **Deep Learning Book**. Capítulo 6. Brasil, 2024. Disponível em: <https://www.deeplearningbook.com.br/o-perceptron-parte-1/>. Acesso em: 08 out. 2025.
- [2] Dive Into Deep Learning. **Convolutional Neural Networks**. Brasil, 2024. Disponível em: https://pt.d2l.ai/chapter_convolutional-neural-networks/index.html. Acesso em: 08 out. 2025.
- [3] Gonçalves, Ilana Carla Mendes; Dias, Priscila Oliveira; Gomes, Luciana Pedrosa et al. **Prevalência e fatores associados a quedas em idosos muito idosos: estudo SABE**. *Revista Brasileira de Epidemiologia*, [S. l.], v. 25, p. e220006, 2022. Disponível em: <https://www.scielo.br/j/rbepid/a/tyJzyFCQD8GJTYSk3ZvW3NR/?lang=pt>. Acesso em: 03 out. 2025.
- [4] Hsieh, Yi-Zeng, and Yu-Lin Jeng. “**Development of Home Intelligent Fall Detection IoT System Based on Feedback Optical Flow Convolutional Neural Network**.” *IEEE access* 6 (2018): 6048–6057.
- [5] Huang, Yueh-Min, An-Yen Cheng, and Ting-Ting Wu. “**Analysis of Learning Behavior of Human Posture Recognition in Maker Education**.” *Frontiers in psychology* 13 (2022): 868487.
- [6] Kulikajėvas, Audrius, Rytis Maskeliūnas, and Robertas Damaševičius. “**Detection of Sitting Posture Using Hierarchical Image Composition and Deep Learning**.” *PeerJ. Computer science* 7 (2021): e442.