v7.n9.2026

REVISTA CIENTÍFICA - RECIMA21 ISSN 2675-6218

DETECÇÃO DE QUEDAS EM AMBIENTES RESIDENCIAIS POR VISÃO COMPUTACIONAL E MODELAGEM TEMPORAL CNN-LSTM

FALL DETECTION IN RESIDENTIAL ENVIRONMENTS USING COMPUTER VISION AND CNN-LSTM TEMPORAL MODELING

DETECCIÓN DE CAÍDAS EN ENTORNOS RESIDENCIALES MEDIANTE VISIÓN POR COMPUTADOR Y MODELADO TEMPORAL CNN-LSTM

Gabriela Ramos de Jesus[1], Lucas dos Passos Martins[1], Kalebe Silva de Vasconcelos[1], Glória Faustine Costa de Oliveira[1], Keventon Rian Guimarães Gonçalves[2], Armando José de Sá Santos[3], Fabricio de Souza Farias[4]

e798966
https://doi.org/10.47820/recima21.v7i9.8966

PUBLICADO: 09/2026

RESUMO

O aumento da expectativa de vida tem trazido à tona novos desafios relacionados à saúde e à segurança da população. Entre os riscos enfrentados, estão as quedas e acidentes domésticos, frequentes causadores de lesões e internações hospitalares. Diante deste contexto, este trabalho propõe um sistema para detecção de quedas de pessoas por meio do uso de técnicas de Inteligência Artificial e Visão Computacional. A solução proposta integra as ferramentas *You Only Look Once* (YOLOv3) para detecção de pessoas, AlphaPose para estimativa de poses corporais e um modelo de rede neural recorrente *Long Short-Term Memory* (LSTM) para classificação de movimentos suspeitos de queda. O sistema foi treinado a partir de vídeos de simulações de quedas coletados em um ambiente residencial controlado, os quais foram processados e convertidos em pontos-chave, permitindo o aprendizado de padrões de postura e transição. Os resultados no ambiente controlado atingiram acurácia global de 95% e sensibilidade (*recall*) de 98% para a classe de quedas, enquanto o teste com o conjunto externo de generalização em cenário residencial distinto alcançou 80% de acurácia e 85% de *recall*, indicando potencial de aplicação, mas ainda condicionado à validação em cenários reais e mais diversos. Conclui-se que a integração entre técnicas de visão computacional e inteligência artificial é uma alternativa promissora na detecção de acidentes domésticos, contribuindo para o avanço de tecnologias assistivas voltadas à saúde e bem-estar.

PALAVRAS-CHAVE: Inteligência Artificial. Visão Computacional. *You Only Look Once* (YOLOv3). AlphaPose. *Long Short-Term Memory* (LSTM). Acidentes domésticos.

ABSTRACT

The increase in life expectancy has brought new challenges related to the health and safety of the population to the forefront. Among the risks faced are falls and domestic accidents, which are frequent causes of injuries and hospital admissions. In this context, this study proposes a system for detecting human falls through the use of Artificial Intelligence and Computer Vision techniques. The proposed solution integrates the You Only Look Once (YOLOv3) framework for person detection, AlphaPose for body pose estimation, and a Long Short-Term Memory (LSTM) recurrent neural network model for classifying suspicious fall movements. The system was trained using videos of simulated falls collected in a controlled residential environment, which were processed and converted into keypoints, enabling the learning of posture and transition patterns.

[1] Universidade do Estado do Pará (UEPA), Graduanda(o) em Engenharia de Software.
[2] Universidade Federal do Pará (UFPA), Mestrando em Computação Aplicada.
[3] Universidade do Estado do Pará (UEPA), Doutorando em Engenharia Elétrica.
[4] Universidade Federal do Pará (UFPA), Doutor em Engenharia Elétrica.

Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

The results in the controlled environment achieved an overall accuracy of 95% and a sensitivity (recall) of 98% for the fall class, while testing with an external generalization dataset in a different residential setting achieved 80% accuracy and 85% recall, indicating application potential, although still subject to validation in real-world and more diverse scenarios. It is concluded that the integration of computer vision and artificial intelligence techniques is a promising alternative for detecting domestic accidents, contributing to the advancement of assistive technologies aimed at health and well-being.

KEYWORDS: *Artificial Intelligence. Computer Vision. You Only Look Once (YOLOv3). AlphaPose. Long Short-Term Memory (LSTM). Domestic Accidents.*

RESUMEN
El aumento de la esperanza de vida ha puesto de manifiesto nuevos desafíos relacionados con la salud y la seguridad de la población. Entre los riesgos enfrentados se encuentran las caídas y los accidentes domésticos, que son causas frecuentes de lesiones e ingresos hospitalarios. En este contexto, este trabajo propone un sistema para la detección de caídas de personas mediante el uso de técnicas de Inteligencia Artificial y Visión por Computador. La solución propuesta integra el modelo You Only Look Once (YOLOv3) para la detección de personas, AlphaPose para la estimación de poses corporales y un modelo de red neuronal recurrente Long Short-Term Memory (LSTM) para la clasificación de movimientos sospechosos de caída. El sistema fue entrenado a partir de vídeos de simulaciones de caídas recopilados en un entorno residencial controlado, los cuales fueron procesados y convertidos en puntos clave, permitiendo el aprendizaje de patrones de postura y transición. Los resultados en el entorno controlado alcanzaron una precisión global del 95% y una sensibilidad (recall) del 98% para la clase de caídas, mientras que la prueba con un conjunto de datos externo de generalización en un escenario residencial diferente alcanzó una precisión del 80% y un recall del 85%, lo que indica un potencial de aplicación, aunque todavía condicionado a la validación en escenarios reales y más diversos. Se concluye que la integración de técnicas de visión por computador e inteligencia artificial constituye una alternativa prometedora para la detección de accidentes domésticos, contribuyendo al avance de tecnologías asistivas orientadas a la salud y el bienestar.

PALABRAS CLAVE*: Inteligencia Artificial. Visión por Computador. You Only Look Once (YOLOv3). AlphaPose. Long Short-Term Memory (LSTM). Accidentes domésticos.*

INTRODUÇÃO

Atualmente, percebe-se o crescimento do número de domicílios unipessoais. Na América Latina, o terceiro maior arranjo de domicílios se dá por lares unipessoais (CAVALCANTE *et al.*, 2022). Esta transformação social traz novos desafios, especialmente quando se percebe também um aumento crescente na quantidade de pessoas idosas morando sozinhas. No Brasil, 41,8% das pessoas que moram sozinhas são pessoas idosas, resultado da busca por maior independência por parte desta população (CHUNG, 2025).

Todavia, dentre os inúmeros desafios impostos por essa nova realidade social, destacam-se as quedas e os acidentes domiciliares (CAETANO *et al.*, 2023). A ausência de supervisão ou suporte imediato pode retardar o atendimento médico, o que eleva o risco de

Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

complicações e desfechos graves (KALIL *et al.*, 2021). Esta realidade é ainda mais preocupante quando se trata de pessoas idosas, uma vez que, segundo dados do Instituto Nacional de Traumatologia e Ortopedia (INTO), 30% a 60% da população com mais de 65 anos cai anualmente e, dessas quedas, de 40% a 60% resultam em algum tipo de lesão, dificuldade motora, redução de mobilidade ou internações hospitalares prolongadas. Essas complicações podem evoluir para quadros ainda mais críticos (INTO, 2015; CAETANO *et al.*, 2023).

Diante desta problemática de saúde pública, torna-se indispensável o desenvolvimento de tecnologias assistivas que proporcionem uma rede de segurança eficiente e auxiliem na rápida detecção de acidentes. No cenário tecnológico atual, existem opções de segurança na forma de dispositivos vestíveis e não vestíveis, sendo o primeiro representado por dispositivos acoplados ao corpo, e o segundo dispositivos que utilizam visão computacional para identificar quedas (GONÇALVES; BERTOLIN; GABRIEL, 2025).

Os dispositivos vestíveis, todavia, enfrentam obstáculos, especialmente quando se trata da pessoa idosa, uma vez que, para serem utilizados, o usuário precisa ter interações diretas com o dispositivo como vestir, configurar e realizar recarga frequente (SCHREIBER, 2025). A população idosa costuma apresentar maior resistência em interagir com equipamentos eletrônicos, o que pode facilmente acarretar o uso incorreto do dispositivo e a anulação da sua eficácia. Além disso, barreiras culturais também são encontradas, o que gera, em muitos casos, resistência por parte do usuário em utilizar o dispositivo, fator crucial para o seu funcionamento (CHUNG, 2025).

Quando se trata de dispositivos não vestíveis que utilizam visão computacional como recurso para identificação de quedas, a interação entre o usuário e o sistema apresenta-se consideravelmente menor (GONÇALVES; BERTOLIN; GABRIEL, 2025). Este tipo de tecnologia funciona de forma automática, sendo praticamente independente da intervenção externa, mas também apresenta limitações relacionadas ao cenário em que a imagem for capturada: ângulo, iluminação e oclusões influenciam diretamente no reconhecimento da pose performada pelo indivíduo (LUBNA; MUFTI; SHAH, 2021). Apesar destas limitações, estas abordagens tecnológicas têm sido consideradas promissoras, o que abre espaço para novos estudos sobre o tema (GONÇALVES; BERTOLIN; GABRIEL, 2025).

Diante da lacuna identificada, este trabalho propõe uma abordagem que utiliza tecnologias de Inteligência Artificial (IA) e Visão Computacional para analisar padrões biomecânicos e identificar comportamentos anômalos em ambientes domésticos. A aplicação combinada de algoritmos de detecção de objetos e algoritmos de estimativa de pose baseados

Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

em Redes Neurais Convolucionais (CNNs), com uma estrutura de Rede Neural Recorrente (RNN) possui o intuito de criar e avaliar uma alternativa às demandas supracitadas.

Especificamente, o trabalho tem como objetivo desenvolver um sistema para a detecção de quedas de pessoas a partir da integração das ferramentas *You Only Look Once version 3* (YOLOv3) e AlphaPose à RNN *Long Short-Term Memory* (LSTM). Para isso, busca-se construir uma base de dados representativa para o treinamento e teste do algoritmo, realizar o aprendizado da rede LSTM e validar o sistema com dados coletados em um ambiente residencial controlado. Além disso, busca-se avaliar o desempenho do modelo em um cenário distinto daquele empregado no treinamento, verificando sua capacidade de generalização diante de variações contextuais. Por fim, pretende-se quantificar a eficiência e a robustez do sistema proposto, fornecendo subsídios e direcionamentos para aprimoramentos em aplicações futuras.

A relevância e a justificativa deste estudo fundamentam-se no seu impacto direto na preservação da vida e na promoção do bem-estar humano, visando respostas rápidas em emergências. Sob a ótica acadêmica e tecnológica, o trabalho oferece uma contribuição ao integrar conceitos de Engenharia de *Software*, Visão Computacional e Saúde Digital para enfrentar um problema de grande impacto social. Espera-se que este trabalho sirva como base para futuras pesquisas dedicadas ao monitoramento assistivo e à consolidação da tecnologia como aliada da saúde.

O restante deste trabalho está organizado da seguinte forma: a Seção 1 apresenta o Referencial Teórico, detalhando a fundamentação conceitual que embasa a pesquisa. A Seção 2 descreve a Metodologia adotada para o desenvolvimento e treinamento do sistema. Na Seção 3, são discutidos os Resultados obtidos e, por fim, a Seção 4 traz as Considerações Finais do estudo.

1. REFERENCIAL TEÓRICO

Esta seção apresenta a fundamentação teórica para a construção deste projeto. As contribuições deste trabalho estão atreladas à Engenharia de *Software*, à Visão Computacional e à Saúde Digital, tópicos que serão aprofundados nesta seção. Também serão explorados os conceitos da metodologia Knowledge Discovery in Databases (KDD), utilizada na construção metodológica deste trabalho.

Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

1.1. Engenharia de *software*

A área de atuação deste projeto está diretamente ligada à Engenharia de *Software*, uma vez que este campo de pesquisa trata o estudo do *software* e seu desenvolvimento, considerando métodos, regras e dados quantitativos (WAZLAWICK, 2019). Dentro desta engenharia podem ser listadas inúmeras subáreas, destacando-se para esta pesquisa a engenharia de requisitos, que é essencial para a construção de sistemas (BRITTO *et al.*, 2024).

1.1.1. Engenharia de requisitos

A Engenharia de Requisitos é uma disciplina da Engenharia de *Software* responsável por definir as diretrizes necessárias para o desenvolvimento de um sistema (VAZQUEZ; SIMÕES, 2016; BRITTO *et al.*, 2024). Para isso, analisa-se o contexto no qual o *software* será aplicado, bem como as necessidades dos usuários e do próprio sistema. A partir desta análise, elabora-se o levantamento de requisitos, subdividido em Requisitos Funcionais (RF), que representam os serviços e funcionalidades a serem entregues, e Requisitos Não Funcionais (RNF), que estabelecem as restrições e qualidades inerentes ao sistema (VAZQUEZ; SIMÕES, 2016; BRITTO *et al.*, 2024).

Aplicado a este trabalho, primeiramente é necessário definir as funcionalidades gerais do sistema. Para isso, a Tabela 1 apresenta os principais RFs e RNFs implementados na solução para identificação de quedas. Os requisitos levantados nortearam as decisões para a escolha das arquiteturas de Inteligência Artificial e definiram os parâmetros para a construção da base de dados em vídeo, garantindo que o algoritmo fosse treinado com imagens que representassem simulações de quedas em atividades rotineiras. Também na Tabela 1, estão demonstrados os RNFs definidos para esta aplicação de forma sintetizada.

Tabela 1. Detalhamento dos principais RFs e RNFs necessários para o sistema

Requisito	Descrição
RF01	Delimitar corpos humanos nos quadros de vídeo através de caixas delimitadoras (*bounding boxes*).
RF02	Extrair os pontos-chave das articulações corporais das pessoas identificadas na imagem.
RF03	Analisar o histórico temporal e sequencial das coordenadas articulares para identificar e classificar eventos de queda.

Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

RF04	Categorizar a pose performada em N para normal (pose segura) e V para variação (queda).
RNF01	Operar de forma totalmente passiva, sem exigir configuração, recargas ou intervenção direta do usuário de forma recorrente.
RNF02	Apresentar alta sensibilidade (*recall*) na classificação dos movimentos para minimizar falhas e reduzir os falsos negativos.
RNF03	Apresentar acurácia geral acima de 70%.
RNF04	Utilizar parada precoce no treinamento da CNN-LSTM para que o sistema não vicie.

Estes requisitos foram levantados com base nos trabalhos de Huang *et al.* (2024) e de Zhao *et al.* (2022), que também utilizam visão computacional para detecção de quedas.

1.2. Visão computacional

De maneira geral, a visão computacional é uma subárea da IA que permite aos computadores interpretar e analisar dados visuais, extraindo deles informações significativas para um contexto pré-estabelecido (SZELISKI, 2022). Esta é uma tecnologia que advém da interseção entre a visão humana e a IA, de modo que as máquinas aprendem a mimetizar processos biológicos da visão (LINDSAY, 2021). Este campo de pesquisa da computação conta com diversas soluções algorítmicas, sendo neste trabalho apresentados o YOLOv3 e o AlphaPose.

1.2.1. YOLOv3

O YOLOv3 é um *software* de código aberto que tem por objetivo identificar objetos em uma imagem passando por ela apenas uma vez utilizando CNNs. Por conta disso, ele é rápido e mantém o equilíbrio entre latência e acurácia, sendo ideal para projetos que exigem visão computacional em tempo real e análise de documentos em vídeo (TIAN; YE; DOERMANN, 2025).

O processo de refinamento das caixas delimitadoras (*bounding boxes*) no YOLOv3 ocorre mediante o cálculo de deslocamento aplicado nas dimensões âncora pré-estabelecidas. A transformação das predições brutas geradas pela rede neural nas coordenadas finais da caixa b está representada pelas Equações (1) a (4) (REDMON; FARHADI, 2018):

$$b_x = \sigma(t_x) + c_x \quad (1)$$

Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

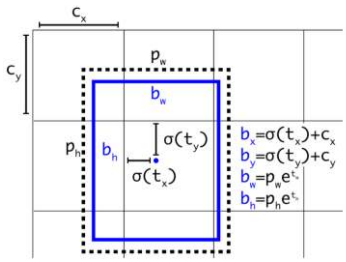
$$b_y = \sigma(t_y) + c_y \quad (2)$$

$$b_w = p_w\, e^{t}_{w} \quad (3)$$

$$b_h = p_h\, e^{t}_{h} \quad (4)$$

Conforme ilustrado na Figura 1, as variáveis b_x e b_y correspondem às coordenadas centrais finais (x, y) da caixa delimitadora ajustada, enquanto b_w e b_h representam a sua largura e altura finais, respectivamente. Os termos t_x, t_y, t_w e t_h constituem os valores brutos de deslocamento e escala previstos pela rede neural. Adicionalmente, c_x e c_y indicam as coordenadas do canto superior esquerdo da célula da grade em análise, ao passo que p_w e p_h representam a largura e a altura originais da caixa âncora. A função de ativação sigmoide, denotada por $\sigma(\cdot)$, é aplicada para restringir o centro da caixa aos limites da célula da grade no intervalo [0, 1], enquanto os fatores exponenciais e^{t}_{w} e e^{t}_{h} ajustam a escala de largura e altura da âncora pré-definida.

Figura 1. Representação gráfica da lógica aritmética da extração de coordenadas do YOLOv3 (REDMON; FARHADI, 2018)



Passada a etapa de localização das caixas delimitadoras, o YOLOv3 precisa extrair as características presentes na imagem delimitada, e para isso a rede base “Darknet-53” é aplicada. Esta corresponde a uma CNN com 52 camadas de convolução inerentes e 1 camada de convolução totalmente conectada, usada para a saída de classificação da rede original (REDMON; FARHADI, 2018). Feito o processamento de localização e identificação do objeto, são gerados tensores tridimensionais que codificam as coordenadas da caixa delimitadora, a probabilidade da presença do objeto e a classe à qual aquele objeto pertence (REDMON; FARHADI, 2018). Estes dados estruturados são então consumidos por soluções que possuem a arquitetura *top-down*, como, por exemplo, o sistema AlphaPose (FANG *et al.*, 2022).

Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

1.2.2. AlphaPose

O AlphaPose é um sistema de código aberto que atua como uma ferramenta de estimativa multipessoa de pose humana (HUANG *et al.*, 2024). Trata-se de um sistema veloz e com performance comprovada na identificação de movimentos e no tratamento de bancos de dados de larga escala (YANG *et al.*, 2024). Este algoritmo utiliza mapas de calor (*heatmaps*) e pontos-chave (*keypoints*) para realizar a identificação do corpo humano e a suposição da pose performada pelo corpo encontrado. Ademais, conta com um *pipeline* de cinco etapas entre captação de dados, detecção, transformação, suposição e pós-processamento (FANG *et al.*, 2022).

O AlphaPose é um algoritmo que possui a arquitetura *top-down*, ou seja, o corpo analisado é primeiramente localizado e delimitado em caixas (*bounding boxes*) para, em seguida, serem definidos os pontos-chave para a previsão da pose (FANG *et al.*, 2022). O sistema responsável por identificar e dividir o corpo em caixas neste trabalho é o YOLOv3 (REDMON; FARHADI, 2018). A saída deste processamento é transformada em um mapa de calor que prevê os locais em que cada parte do corpo identificada na imagem provavelmente está (FANG *et al.*, 2022).

Todavia, para o bom funcionamento da identificação da pose, é necessária maior precisão nas predições de localização. Logo, são aplicadas mais duas camadas de processamento: a primeira é o operador matemático *soft-argmax*, responsável por normalizar os dados oferecidos pelos mapas de calor em uma escala entre 0 e 1, e calcular a média ponderada destes valores para encontrar as coordenadas em um gradiente assimétrico; e a segunda é o *Symmetric Integral Keypoints Regression* (SIKR), utilizado para corrigir a assimetria do gradiente gerado pelo *soft-argmax* e ajustar os pesos para que a predição da coordenada não seja influenciada por nenhum viés gerado nos gradientes anteriores (FANG *et al.*, 2022; SAUVALLE; LA FORTELLE, 2023; ZENG; JIANG, 2025).

Uma vez que as coordenadas estão alinhadas, o sistema já é capaz de supor a pose performada pelo corpo. Junto à suposição, de forma simultânea, um número de identificação (*re-ID*) é associado a cada corpo identificado, e o AlphaPose determina qual a pose performada pelo *re-ID* em questão (HUANG *et al.*, 2024).

A Figura 2 ilustra uma amostra de pontos-chave oriundos da análise de um *frame*. É possível observar que os dados são organizados no formato *JSON*, separando as informações em relação ao ID da imagem, à categoria e aos pontos-chave. Cada ponto-chave representa as coordenadas espaciais (x,y) de uma articulação anatômica específica do corpo humano acompanhada pelo respectivo grau de confiança da detecção efetuada pelo modelo.

Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

Figura 2. Amostra de extração de pontos-chave no formato JSON


```
[{"image_id": "0.jpg",
"category_id": 1,
"keypoints":
[550.892333984375, 255.289794921875, 0.8831897974014282, 555.8969116210938, 248.6170196533203, 0.9155845046043396,
549.22412109375, 246.94882202148438, 0.8785461187362671, 590.9290161132812, 243.6124267578125, 0.8774386048316956,
590.9290161132812, 243.6124267578125, 0.5412012934684753, 630.9656982421875, 295.32647705078125, 0.4785458445549011,
584.2562255859375, 281.9809265136719, 0.3782893419265747, 629.2974853515625, 352.04510498046875, 0.30537453293800354,
565.9060668945312, 348.7087097167969, 0.2848048210144043, 579.2516479492188, 347.04052734375, 0.16610093414783478,
545.8877563476562, 317.0130310058594, 0.0767018124461174, 595.93359375, 352.04510498046875, 0.08301152288913727,
560.9014892578125, 352.04510498046875, 0.23349331319332123, 569.242431640625, 335.3631591796875, 0.056702859699726105,
565.9060668945312, 335.3631591796875, 0.044610388576984406, 572.578857421875, 353.71331787109375, 0.06068561598658562,
565.9060668945312, 352.04510498046875, 0.2027868777513504]}]
```


Já a Figura 3 ilustra a saída visual, com um exemplo do *frame* de entrada e o *frame* processado com a representação gráfica das marcações feitas pelo AlphaPose. Nota-se que após a aplicação das CNNs, o *frame* ilustrado na Figura 3 (a) é coberto pelos dados presentes no *JSON* que contém os pontos-chave, sendo gerado em seguida o *frame* ilustrado na Figura 3 (b).

Figura 3. Demonstração da aplicação do AlphaPose no *frame* analisado. (a) Imagem original. (b) Imagem após a marcação obtida a partir do processamento da imagem



1.2.3. Rede LSTM

Na análise de vídeos, o processamento exige tanto a extração de características espaciais (forma e postura) quanto temporais (movimentos em sequência). Neste contexto, a arquitetura adotada integra as Redes Neurais Convolucionais (CNNs), aqui representadas pelas etapas prévias do YOLOv3 e AlphaPose na extração espacial das coordenadas corporais, a uma rede *Long Short-Term Memory* (LSTM). A LSTM é uma arquitetura de rede neural recorrente desenvolvida especificamente para modelar dependências temporais e mitigar problemas como o desaparecimento do gradiente no

Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

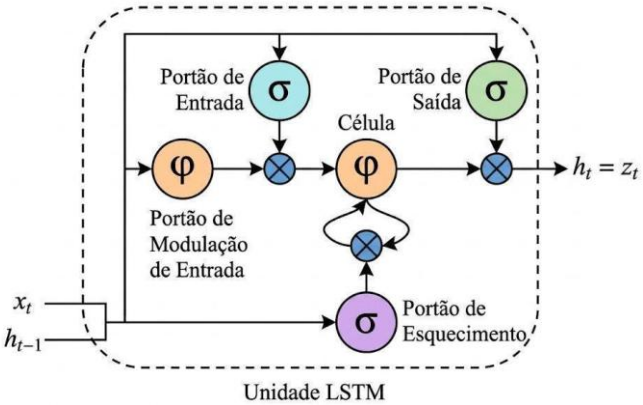
processamento de sequências (BEHERA *et al.*, 2021). O funcionamento da célula LSTM é regido por três portões, sendo eles portão de entrada, portão de esquecimento e portão de saída (AL-SELWI *et al.*, 2024).

Uma das representações matemáticas do portão de esquecimento se dá pela Equação (5):

$$f_t = \delta\left(W_f\left[h_{t-1}, X_t\right] + b_f\right) \quad (5)$$

onde W_f é o peso, h_{t-1} é o estado prévio, X_t é a entrada atual e b_f é o viés. Levados em conta todos estes parâmetros, o algoritmo consegue realizar a diferenciação entre o que interfere e o que não interfere na análise dos dados para obtenção de uma resposta através da IA construída (WEN; LI, 2023). Tratando todos os portões e o funcionamento geral da célula LSTM, considera-se a entrada X_t como inicial, onde, em sua primeira passagem pela célula, os dados são processados de forma individual e torna-se o estado prévio, representado por h_{t-1}, que volta ao início da célula a ser processado juntamente com a nova entrada (WEN; LI, 2023). O portão de esquecimento controla quanto do estado de memória anterior será preservado ou atenuado na atualização da célula (HUANG *et al.*, 2024). A Figura 4 ilustra uma célula do LSTM representada na forma de diagrama.

Figura 4. Diagrama da célula do LSTM elaborado pelos autores



De acordo com Huang et al. (2024), essa capacidade de retenção seletiva da informação temporal torna a arquitetura LSTM ideal para o processamento de dados sequenciais complexos, otimizando o treinamento e a eficiência do tratamento de dados ao longo do tempo (AL-SELWI *et al.*, 2024).

Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

1.3. Saúde Digital

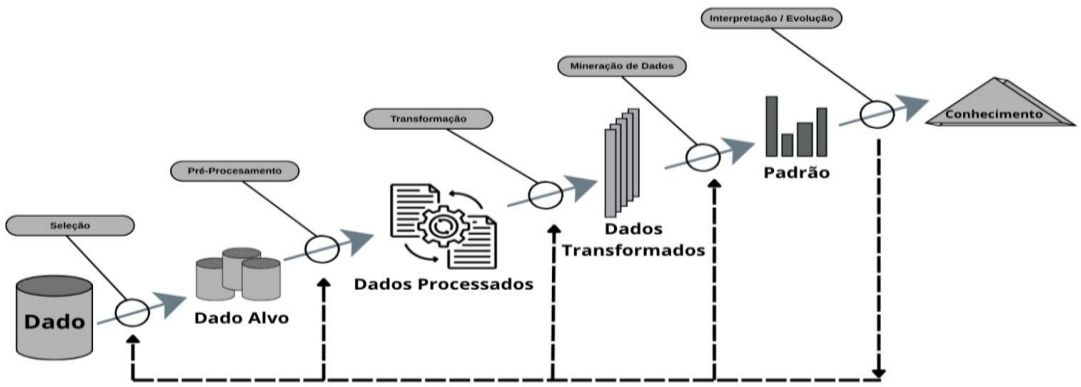
Mediante a transformação digital enfrentada pelo mundo, a aplicação da tecnologia em áreas diversas foi se tornando cada vez mais uma realidade irreversível (SILVA*et al.*, 2024). A saúde digital é um conceito em constante ascensão, especialmente após o período de pandemia da COVID-19, em que medidas de distanciamento foram necessárias e o uso da tecnologia tornou-se indispensável em atividades cotidianas (WOSIK *et al.*, 2020).

Em termos conceituais, a saúde digital pode ser compreendida como um conjunto de ferramentas tecnológicas destinadas a apoiar procedimentos clínicos e realizar o monitoramento preventivo da saúde populacional (RACHID *et al.*, 2023). Na literatura, está diretamente relacionada à padronização de dados e a telemedicina; todavia, a saúde digital pode alcançar maior abrangência sendo também utilizada para monitoramento e tratamento preventivo (CORREA *et al.*, 2026). A exemplo do que foi citado, mostram-se os sistemas de monitoramento de quedas, que permitem não prevenir a queda em si, mas diminuir o tempo entre queda e socorro, facilitando a redução das complicações causadas pela queda (HUANG *et al.*, 2024).

1.4. KDD

O processo de Descoberta de Conhecimento em Bancos de Dados (*Knowledge Discovery in Databases* - KDD) é um modelo de ciclo de vida de dados estruturado, analítico e iterativo. Segundo Fayyad *et al.* (1996), o KDD tem como objetivo identificar padrões válidos, novos, potencialmente úteis e compreensíveis a partir de um grande volume de dados. A abordagem envolve uma série de técnicas divididas em diferentes fases, indo desde a concepção inicial da base de dados até a extração de conhecimento e avaliação final.

Figura 5. Demonstração dos passos que compõem o processo de Descoberta de Conhecimento em Bancos de Dados (*Knowledge Discovery in Databases* - KDD)



Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

A primeira etapa do processo de KDD é a seleção dos dados, na qual são identificadas as informações relevantes para o problema investigado. Em seguida, realiza-se o pré-processamento, etapa destinada ao preparo dos dados para as análises posteriores. Essas etapas são fundamentais para que os dados utilizados no processo apresentem condições adequadas para a aplicação das técnicas de mineração de dados (STEINER *et al.*, 2006).

Na sequência, ocorre a transformação dos dados, etapa em que os dados previamente preparados são adequados às necessidades da análise. Essa transformação pode envolver a reformatação dos dados e outras modificações necessárias para sua utilização no processo de mineração (GALVÃO; MARIN, 2009).

A quarta etapa corresponde à mineração de dados (*Data Mining*). De acordo com Galvão e Marin (2009), a mineração de dados constitui uma das etapas mais importantes do KDD e tem como finalidade identificar padrões e relações nos dados. A quinta etapa consiste na avaliação e interpretação dos resultados, momento em que os padrões e modelos obtidos são analisados de acordo com os objetivos estabelecidos para o estudo. Dessa forma, o processo de KDD permite que os dados sejam progressivamente transformados em informação e, posteriormente, em conhecimento, contribuindo para o processo de tomada de decisão (GALVÃO; MARIN, 2009).

Nesse contexto, Silva (2020) também destaca que a descoberta de conhecimento envolve uma série de procedimentos destinados a transformar dados em informações relacionadas ao contexto da investigação. Para o autor, os dados utilizados no processo KDD devem ser limpos e organizados de modo a possibilitar um processamento adequado. O autor também apresenta o KDD como um processo composto por identificação do problema, preparação dos dados, mineração de dados e pós-processamento, ressaltando a importância da seleção e preparação dos dados para a qualidade dos resultados obtidos (SILVA, 2020).

Assim, o KDD apresenta-se como uma abordagem estruturada para a transformação de grandes volumes de dados em conhecimento, permitindo não apenas maior agilidade no processo de análise, mas também a identificação de relações e padrões entre diferentes conjuntos de informações, os quais podem contribuir para o desenvolvimento de pesquisas futuras (SILVA, 2020). A aplicação sistemática de suas etapas possibilita organizar o processo de análise, desde a seleção e preparação dos dados até a mineração e avaliação dos resultados, favorecendo a obtenção de informações relevantes para a compreensão do problema investigado e para a tomada de decisões.

Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

2. METODOLOGIA

Esta seção apresenta a metodologia aplicada e os procedimentos técnicos adotados para o desenvolvimento do sistema. Nela, é detalhada a abordagem aplicada para a análise de dados e a implementação prática, bem como as ferramentas de visão computacional, as bibliotecas utilizadas e as práticas de Engenharia de *Software* utilizadas na solução.

2.1. Natureza e abordagem

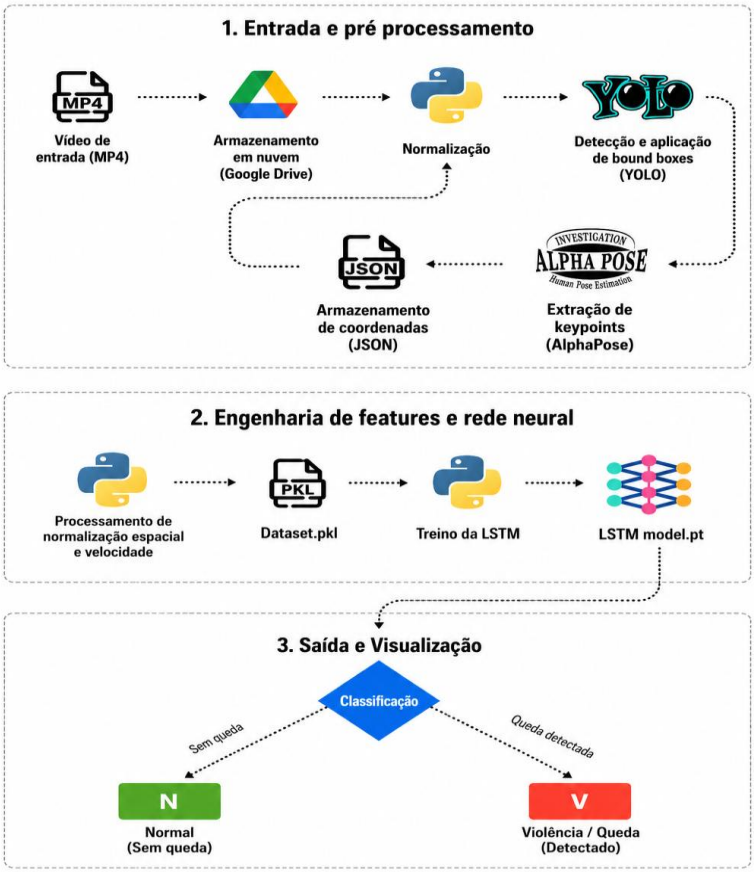
A presente pesquisa é de natureza aplicada, pois busca desenvolver uma solução prática para o problema de detecção de quedas de pessoas em ambiente doméstico. A abordagem adotada é experimental, uma vez que envolve a implementação, o treinamento e a validação de uma rede LSTM, com o objetivo de observar seu comportamento em diferentes cenários e conjuntos de dados. Além disso, o estudo tem caráter quantitativo, visto que os resultados são analisados com base em métricas numéricas (como acurácia e precisão) que permitem avaliar de forma objetiva o desempenho do sistema.

Para o desenvolvimento metodológico deste trabalho, inspirou-se no modelo KDD para gerenciar o ciclo de vida dos dados, visando escalabilidade e eficiência. Esta abordagem envolve uma série de etapas, que vão desde a concepção inicial e seleção da base de dados até a extração de conhecimento e sua avaliação final (FAYYAD *et al.*, 1996). O *pipeline* resultante opera a partir da aplicação sequencial dessas fases.

A integração sistemática das CNNs YOLOv3 e AlphaPose com a rede LSTM constitui um *pipeline* caracterizado por sua arquitetura híbrida CNN-LSTM, na qual as CNNs detectam a pessoa em movimento e extraem os pontos-chave e a rede LSTM processa esses dados para reconhecer os padrões temporais entre os quadros de vídeo, diferenciando ações normais de quedas. As etapas e a circulação dos dados ao longo do sistema estão descritas na Figura 6.

Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

Figura 6. Arquitetura do *pipeline* de processamento e fluxo de dados

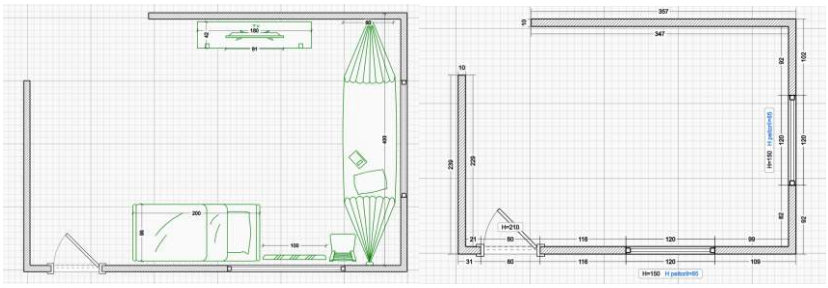


2.2. Seleção (coleta de dados)

Para a realização do treinamento de qualquer ferramenta de IA, é necessária a utilização de um banco de dados, e o bom funcionamento do sistema depende diretamente da sua robustez e da qualidade dos dados ofertados (BARBOSA; PORTES, 2023). Por esta razão, na etapa de seleção, o banco de dados utilizado para realizar o treinamento do *software* foi desenvolvido e analisado pelos próprios autores. Estruturou-se um ambiente controlado provido de iluminação natural com uma disposição específica de móveis, conforme ilustrado na Figura 7 (a). O ambiente possui uma área de 14,3 m², planta representada na Figura 7 (b).

Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

Figura 7. Planta baixa do cenário de gravação e testes elaborado pelos autores. (a) Planta sem mobília. (b) Planta com mobília



Com o objetivo de capturar imagens em ambiente controlado, utilizou-se um *smartphone* (*Samsung Galaxy S25 FE)* acoplado a um tripé. Cada registro coletado possui duração média de 15 segundos e foi gravado em resolução *Full HD* (1080p) com uma taxa de 30 *frames* por segundo (FPS). As capturas visaram simular tanto acidentes domésticos (tropeços, quedas e escorregões) quanto comportamentos cotidianos (sentar-se, andar e coletar objetos no chão). No total original do ambiente, foram coletados 318 vídeos (totalizando 79,5 minutos). Por fim, os arquivos obtidos foram armazenados no repositório em nuvem *Google Drive* para assegurar a integridade dos dados até as etapas posteriores.

A fim de compor os conjuntos de treinamento, testes e generalização, as gravações foram protagonizadas exclusivamente por autores do artigo que consentiram em utilizar sua imagem para popular o banco, uma vez que apenas suas coordenadas corporais seriam armazenadas. Para compor o conjunto destinado ao treinamento e aos testes internos da aplicação, participaram dois autores do sexo masculino, ambos com 21 anos de idade e 166 cm de estatura, apresentando padrões corporais distintos (um de biotipo esguio e outro de constituição robusta). Posteriormente, visando distinguir formalmente as etapas de teste e avaliar a robustez do sistema, estruturou-se um conjunto externo de generalização em um cenário inédito. Este conjunto foi protagonizado por uma participante distinta (sexo feminino, 21 anos, 167 cm de estatura e perfil corporal esguio).

Em todas as fases de coleta, os indivíduos utilizaram vestimentas casuais e leves para não restringir a liberdade e a naturalidade dos movimentos. Cabe ressaltar que, uma vez que as simulações foram conduzidas exclusivamente por adultos jovens, o presente estudo caracteriza-se como uma prova de conceito técnica para a detecção de quedas. Indivíduos na terceira idade possuem características biomecânicas específicas, não sendo possível inferir automaticamente

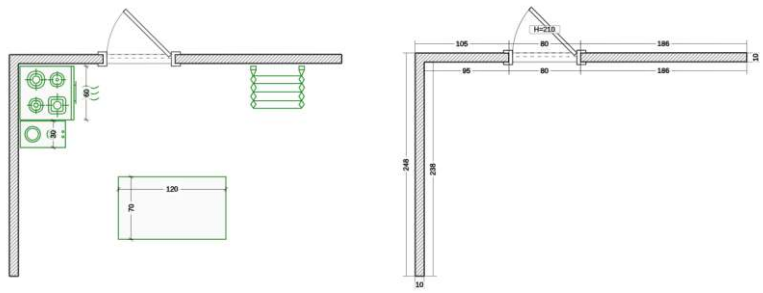
Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

um desempenho clínico equivalente na população idosa real, o que configura uma oportunidade para aprimoramento e validação futura.

2.2.1. Testes de generalização do sistema (coleta em cenário inédito)

Com o intuito de investigar a capacidade de generalização do algoritmo treinado, estabeleceu-se um protocolo de coleta de dados em um cenário residencial inédito. O objetivo principal desta etapa foi apurar a robustez do modelo CNN-LSTM e sua aptidão para manter a eficácia preditiva em um ambiente desconhecido, avaliando o comportamento do modelo diante de variáveis físicas e espaciais não presentes no conjunto de treinamento original, tais como nova disposição da mobília, diferentes distâncias focais e iluminação distinta. A Figura 8 ilustra a planta baixa do novo ambiente adotado para este teste.

Figura 8. Planta baixa do cenário de gravação para os testes de generalização do sistema elaborado pelos autores. (a) Planta com mobília. (b) Planta sem mobília



A partir deste protocolo, estruturou-se um conjunto externo de generalização composto por 40 amostras em formato de vídeo inéditas. Este *corpus* foi distribuído de maneira equitativa entre 20 simulações de quedas e 20 atividades cotidianas normais, totalizando 10 minutos e 12 segundos de gravação.

2.3. Pré-processamento (tratamento dos vídeos)

A fase inicial de pré-processamento consistiu na triagem e organização dos arquivos capturados. Os arquivos foram inicialmente separados nas categorias “Queda” e “Normal”, segregados em diretórios distintos. Logo em seguida foi executado um *script* desenvolvido na linguagem *Python* utilizando as bibliotecas *os* e *shutil* para interagir com o sistema operacional, gerenciar caminhos, pastas e arquivos. O módulo *os* realiza funções do sistema de arquivos e

Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

comandos básicos, enquanto o *shutil* lida com operações complexas de alto nível, como copiar e apagar pastas inteiras (PYTHON SOFTWARE FOUNDATION, s.d.).

No contexto desta aplicação as bibliotecas foram usadas para a padronização automática de nomenclatura: adotou-se o prefixo “N” para atividades normais e “V” para variações, seguidos de um identificador numérico (ex: N01, V01). Posteriormente, elaborou-se um arquivo de anotações centralizado no formato *JSON* para catalogar o nome de cada vídeo e vinculá-lo à sua respectiva classe rotulada.

Com o intuito de assegurar a integridade da avaliação, a divisão metodológica ocorreu de modo que, antes de qualquer processamento ou segmentação em clipes do corpus original, destinou-se 80% dos vídeos (254 vídeos) exclusivamente para o conjunto de treinamento da rede, enquanto os 20% restantes (64 vídeos) compuseram o conjunto de validação utilizado durante o ajuste do modelo. O conjunto de dados resultou em matrizes processadas que foram separadas estrategicamente, dessa forma, não há compartilhamento de vídeos, *frames* ou sequências temporais sobrepostas entre as bases de treinamento, validação e teste de generalização.

2.4. Transformação (extração de características)

Na fase de transformação, os dados brutos são convertidos e estruturados para alimentar a rede neural. O processamento individual das imagens foi realizado por meio da integração das ferramentas YOLOv3 e AlphaPose, utilizando uma infraestrutura computacional composta por um processador *Ryzen 5 2600X*, *16 GB de RAM DDR4*, placa de vídeo GPU *RTX 4060 Ti* com *8GB* de *VRAM* e o sistema operacional *Linux Mint 22.3*. O sistema processou cada vídeo e armazenou os resultados em diretórios específicos, gerando os arquivos *JSON* contendo os pontos-chave capturados em cada *frame*.

Na sequência, o arquivo de anotações prévio foi submetido a um *script* de estruturação de dados que mapeou as classes e cruzou com a localização dos *JSONs* do AlphaPose. A transformação seguiu parâmetros estritos para garantir a reprodutibilidade temporal, gerando as sequências por meio da técnica de janela deslizante (*sliding window*). Cada sequência temporal extraída possui um comprimento exato de 32 quadros consecutivos, com um *stride* (salto de sobreposição) fixado em 8 quadros. Para cada quadro, o sistema estruturou 17 *keypoints* anatômicos, resultando em 4 dimensões paramétricas espaciais e temporais (as coordenadas x, y e as métricas de velocidade v_x, v_y). A normalização espacial foi aplicada subtraindo-se as coordenadas do ponto central (quadril) de todas as demais articulações e dividindo-se o resultado pela métrica de comprimento do tronco. Para o tratamento de oclusões, instâncias de baixa

Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

confiança do AlphaPose ou ausência de *keypoints* que resultassem em vetores próximos a zero, estabeleceu-se um limiar de segurança matemático: valores de distância inferiores a $1{,}0 \times 10^{-6}$ foram automaticamente substituídos por $1{,}0$, prevenindo divisões por zero e garantindo a estabilidade dimensional do tensor. Por fim, todo esse conjunto de vetores padronizados foi estruturado e exportado em um banco de dados no formato *.pkl* (*Pickle*), deixando a base pronta para alimentar a rede neural.

2.5. Mineração de dados (treinamento do modelo CNN-LSTM)

A fase de Mineração de Dados constitui o núcleo do aprendizado da arquitetura, momento em que a rede assimila os padrões temporais e espaciais das sequências de movimento. A arquitetura exata implementada consiste em uma rede neural recorrente do tipo LSTM padrão. O modelo foi estruturado com duas camadas ocultas compostas por 128 unidades neurais cada, aplicando-se uma taxa de *dropout* (desativação aleatória de neurônios) de 0,3 para regularização. A camada de saída utiliza uma camada totalmente conectada para a classificação em duas classes, operando sobre parâmetros treináveis ajustados dinamicamente para as 68 entradas por quadro da rede (provenientes de 17 *keypoints* multiplicados pelas 4 dimensões analíticas: coordenadas espaciais x, y e métricas de velocidade v_x, v_y).

Uma vez estruturados, os tensores foram injetados no algoritmo de treinamento sob um limite máximo de 100 épocas. A otimização foi conduzida pelo algoritmo *Adam*, com taxa de aprendizado de 0,001 e processamento em lotes (batch size) de 16 amostras por iteração para otimização de memória. Para assegurar a reprodutibilidade, fixou-se a *seed* (semente) aleatória em 42. A função de perda adotada foi a *Cross-Entropy Loss* (função de perda de entropia cruzada), incorporando pesos de classe desbalanceados (1,0 para atividades normais e 1,78 para quedas), forçando o modelo a priorizar a detecção de acidentes. Sob o critério clínico de maximizar a Sensibilidade (*recall*), o limiar de decisão (*threshold*) foi calibrado para 0,35.

A evolução do modelo foi analisada época a época mediante o monitoramento contínuo das curvas de perda e acurácia. A estratégia de seleção da "melhor configuração" e a prevenção da otimização indireta basearam-se exclusivamente no conjunto de validação interno (20% dos dados de origem). Aplicou-se a técnica de *Early Stopping* com tolerância de 5 épocas, interrompendo o treinamento automaticamente no momento em que o modelo atingiu seu ápice de generalização. Simultaneamente a essa interrupção, o arquivo contendo os pesos otimizados da rede (*lstm_best.pt*) foi salvo, permitindo sua posterior aplicação no script de inferência. O conjunto de teste final manteve-se em rigoroso isolamento (*hold-out*), não participando sob

Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

nenhuma hipótese de avaliações repetidas para ajuste do modelo ou escolha do threshold, garantindo um reflexo imparcial de seu desempenho no mundo real.

2.6. Interpretação / avaliação

Na finalização do processo de treinamento, o sistema emite um relatório contendo métricas de classificação, como *Recall*, Precisão e *F1-Score*. Os testes de inferência são realizados com vídeos inéditos e é utilizada uma Matriz de Confusão para mapear a distribuição dos acertos e erros, com foco especial na proporção de Falsos Positivos e Falsos Negativos, sendo a minimização deste último um fator crítico em aplicações voltadas à saúde digital e ao monitoramento de pacientes vulneráveis, pois os erros de predição possuem custos assimétricos.

Segundo Igual, Medrano e Plaza (2013), o desenvolvimento de sistemas de detecção de quedas enfrenta o desafio crítico de equilibrar o limiar entre Falsos Positivos e Falsos Negativos. No entanto, a literatura médica e de bioengenharia prioriza a maximização da sensibilidade. Falhar ao registrar uma queda real pode resultar em longos períodos sem socorro, agravando severamente o quadro clínico do indivíduo. Por outro lado, um Falso Positivo resulta apenas em um incômodo operacional temporário.

3. RESULTADOS E DISCUSSÃO

Esta seção apresenta os resultados obtidos a partir do sistema desenvolvido. Nela, detalham-se o processo de treinamento da rede LSTM, as métricas que demonstram a eficácia do modelo CNN-LSTM e a descrição dos testes realizados, fornecendo subsídios para a análise e a discussão sobre a qualidade da solução proposta.

3.1. Análise de treinamento e convergência do modelo

A avaliação do treinamento e a seleção do estado ideal do modelo basearam-se no monitoramento das métricas da rede perante o conjunto de validação interno. Conforme estabelecido na configuração metodológica, os ajustes de aprendizado e o acionamento do mecanismo de parada precoce (*Early Stopping*) ocorreram de forma isolada do conjunto de teste final, garantindo ausência de otimização indireta sobre dados inéditos. O treinamento que demonstrou a melhor capacidade de generalização e estabilidade ao longo das épocas gerou as curvas de convergência e os resultados de classificação discutidos a seguir.

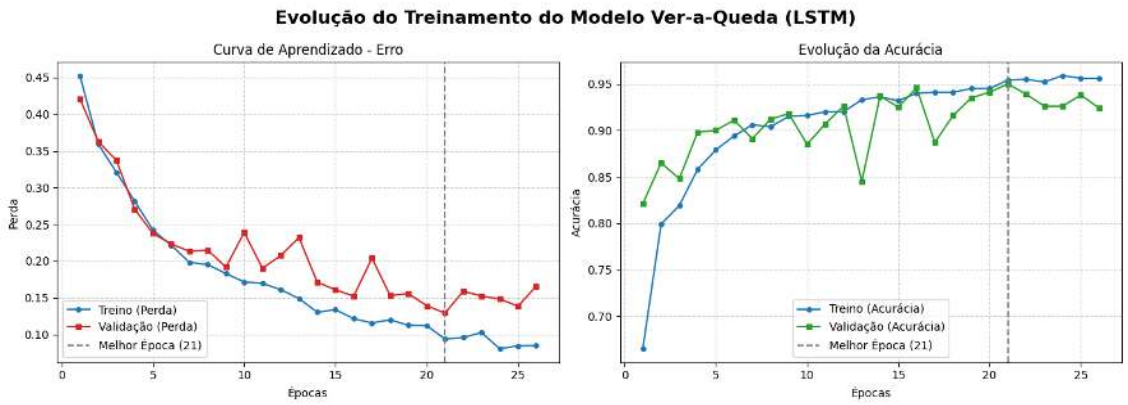
Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

O comportamento da rede na sua melhor configuração é quantificado pelas métricas de função de perda (*loss*) e acurácia, cujas evoluções temporais ao longo das 26 épocas executadas estão ilustradas na Figura 9.

Conforme observado nas curvas de aprendizagem na Figura 9 (a), os conjuntos de treinamento (*tr loss*) e validação (*te loss*) apresentaram uma convergência descendente expressiva nas épocas iniciais, com a perda reduzindo de 0,45 na primeira época para patamares inferiores a 0,20 a partir da 9ª época. Este decaimento acentuado demonstra a capacidade robusta da rede em extrair e assimilar as características espaço-temporais inerentes à dinâmica das poses corporais. O valor mínimo da função de perda de validação foi registrado na 21ª época (0,1291), caracterizando o ponto ótimo de generalização do algoritmo, momento em que os pesos otimizados da rede foram salvos. Nas épocas subsequentes, a ocorrência de discretas oscilações ascendentes ativou o mecanismo de *early stopping* na 26ª época, interrompendo o treinamento para mitigar o *overfitting*.

Em consonância com a minimização do erro, o gráfico de acurácia na Figura 9 (b) evidencia a evolução contínua da capacidade preditiva do modelo final. As taxas experimentaram uma rápida ascensão, partindo de 66,5% no treinamento e 82,1% na validação, para estabilizarem-se em valores consistentes acima de 94%, atestando a eficácia do aprendizado estruturado.

Figura 9. Curvas de aprendizado (perda e acurácia) do modelo LSTM durante o treinamento



3.2. Avaliação de desempenho no conjunto de testes

Para avaliar a robustez e a capacidade de generalização do sistema em cenários reais, o modelo otimizado foi testado em um conjunto de inferência inédito composto por 1.713

Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

amostras (clipes de vídeo processados). Os resultados das métricas de classificação globais e por classe estão sintetizados na Tabela 2.

Tabela 2. Relatório de classificação final do modelo LSTM para detecção de quedas em ambiente controlado

Classe	Precisão	*Recall*	*F1-Score*	Suporte
Normal	0,97	0,92	0,95	826
Queda	0,93	0,98	0,95	887
Acurácia Global			**0,95**	**1713**
Média Macro	0,95	0,95	0,95	1713
Média Ponderada	0,95	0,95	0,95	1713

O modelo proposto alcançou uma Acurácia Global de 95%, acompanhada por um *F1-Score* balanceado de 0,95, indicando desempenho elevado na discriminação entre as classes “Normal” e “Queda”, nas condições experimentais avaliadas. No contexto de aplicações de visão computacional voltadas para o monitoramento de quedas, o foco primário recai sobre o *recall* do algoritmo na identificação de eventos críticos. Neste quesito, o sistema registrou uma taxa de *recall* de 98% para a detecção de quedas, assegurando que a grande maioria dos eventos de queda sejam identificados.

Adicionalmente, o modelo manteve uma precisão de 93% para a classe de interesse, indicando uma baixa incidência de alarmes falsos. A métrica confirma que o modelo é capaz de discernir de forma consistente atividades cotidianas que envolvem transições posturais abruptas (como sentar-se ou curvar-se rapidamente), não comprometendo a rotina do usuário com notificações incorretas. Desta forma, os resultados demonstram desempenho elevado nas condições experimentais avaliadas para a detecção de acidentes domésticos.

3.3. Testes e validação do sistema

Para avaliar o desempenho, a robustez e a capacidade de generalização da rede LSTM desenvolvida, o sistema foi submetido a dois protocolos de testes distintos. Esta divisão metodológica visa comprovar a eficácia da rede tanto nas condições experimentais conhecidas (conjunto de teste interno) quanto em cenários reais desconhecidos (conjunto de generalização).

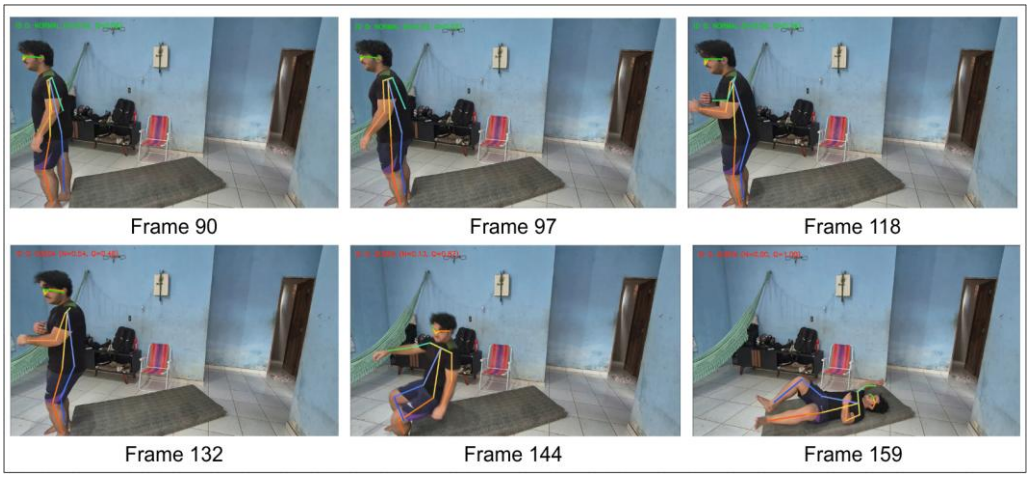
Com o intuito de avaliar o algoritmo em um ambiente controlado, utilizou-se o conjunto de validação correspondente a 20% do *corpus* original coletado. Este subconjunto é composto por 64 vídeos, totalizando aproximadamente 16 minutos de gravação. Após o janelamento

Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

temporal, essas amostras resultaram em 1.713 sequências de clipes processados, apresentando uma distribuição de classes equilibrada: 826 amostras de atividades cotidianas (classe “Normal”) e 887 simulações de acidentes domésticos (classe “Queda”).

A Figura 10 ilustra a progressão da inferência do modelo ao longo do tempo. No *frame* 90, o algoritmo classifica a postura da pessoa como normal. Com o avanço temporal (*frames* 97 e 118), observa-se um aumento contínuo no nível de confiança para a classe “queda”, refletindo o momento exato em que o corpo inicia sua curvatura e o centro de massa é projetado para trás. A partir do *frame* 132, a transição do movimento é concluída e o modelo passa a classificar o evento como queda com alta confiança, mantendo o alerta estabilizado nos *frames* 144 e 159.

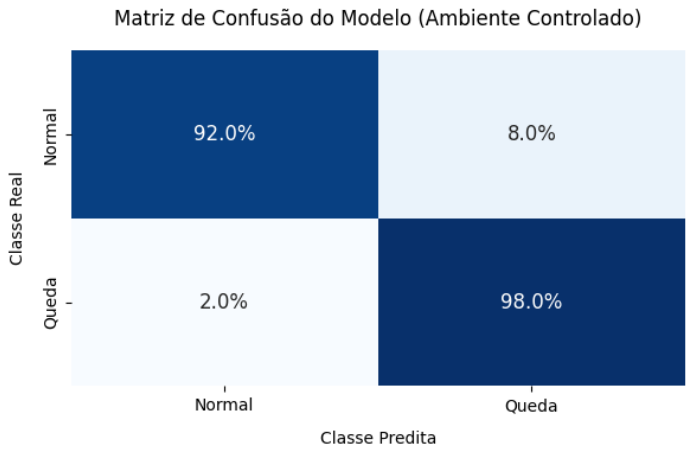
Figura 10. Processo de classificação do algoritmo



Para a consolidação dos resultados, recorreu-se à construção de uma Matriz de Confusão (Figura 11). O modelo demonstrou desempenho elevado, atingindo uma Acurácia Global de 95% e um F1-Score balanceado de 0,95. O grande destaque do sistema está no *recall* para a detecção de eventos críticos: a rede alcançou a marca de 98% de acerto para a classe "Queda". Isso indica que o sistema foi capaz de identificar quase a totalidade das quedas reais submetidas, minimizando drasticamente a ocorrência de Falsos Negativos. Em relação às amostras de movimentos rotineiros, o modelo identificou corretamente 92% das atividades normais, gerando uma taxa muito baixa de alarmes falsos (Falsos Positivos). No cenário das aplicações voltadas para a saúde digital, prioriza-se fortemente a minimização de falsos negativos (KALIL *et al.*, 2021). Logo, ao garantir que a ocorrência de incidentes graves não passe despercebida, o sistema cumpre seu objetivo primordial de segurança, ao mesmo tempo em que mantém os alarmes falsos em um limiar estritamente aceitável.

Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

Figura 11. Matriz de Confusão gerada após a submissão das amostras de teste



Para fins de comparação e avaliação da qualidade do sistema proposto, trabalhos correlatos que utilizam arquiteturas diferentes da proposta foram analisados e comparados na Tabela 3. Os trabalhos analisados foram os de Zhao *et al.* (2022) e o de Huang et al. (2024), ambos têm por objetivo desenvolver um sistema de detecção de quedas.

Tabela 3. Comparação de métricas e metodologias entre trabalhos correlatos

Autor (Ano)	**Arquitetura**	**Tamanho do *Dataset***	**Ambiente**	**Acurácia**	***Recall* (Sensibilidade)**
Zhao *et al.* (2022)	Modelo otimizado de AlphaPose acelerado por YOLOv4-tiny e algoritmo de julgamento baseado em velocidade linear de pontos articulares e ângulos de bissetrizes	104 vídeos de quedas humanas do *dataset* público UR *Fall Dataset*	Residencial (sala de estar), escritório e casa de chá	89,4% (para o modelo proposto YOLOv4--tiny + AlphaPose)	Não especificado explicitamente como métrica de porcentagem final (foco em acurácia e FPS)
Huang *et al.* (2024)	Plataforma NVIDIA Jetson utilizando a biblioteca jetson-inference com o modelo PoseNet (ResNet-18) e lógica matemática de proporção largura/altura	40 amostras (30 tentativas de quedas reais e 10 falsas)	Residencial (quarto e sala de estar)	82,5%	86,66%

Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

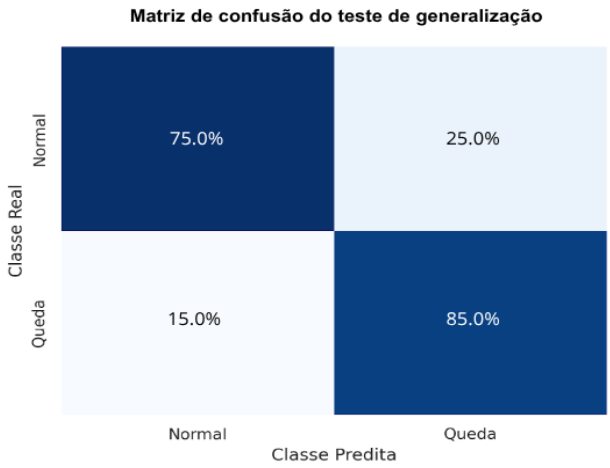
Este Artigo	Modelo CNN-LSTM: YOLOv3 e AlphaPose conectados a uma rede LSTM treinada com *keypoints*	318 vídeos (140 vídeos de ações cotidianas e 178 vídeos de queda)	Residencial (sala de estar)	95%	98%

Os desafios encontrados nos testes de validação dos sistemas citados foram bastante semelhantes aos deste artigo. Tópicos como baixa luminosidade, ângulos desfavoráveis à captação do esqueleto e oclusões são abordados em ambos os trabalhos correlatos. Todavia, um problema relatado nos trabalhos de Huang et al. (2024) e Zhao *et al.* (2022) está relacionado aos falsos negativos e à velocidade da queda; segundo os autores o sistema encontrou dificuldade em diferenciar ações cotidianas, como sentar-se e deitar, de quedas. Neste artigo, a utilização da rede LSTM mostrou eficiência nas condições experimentais avaliadas em fazer tal diferenciação, considerando a velocidade do movimento para realizar a classificação.

3.4. Testes de generalização do sistema

Após a submissão das 40 amostras de vídeo inéditas coletadas no novo cenário residencial, o sistema alcançou uma acurácia global de 80%. Destaca-se o *recall* de 85% na detecção de quedas, identificando corretamente 17 dos 20 acidentes simulados. Em relação às atividades rotineiras, o sistema obteve 75% de taxa de acerto (15 de 20 amostras).

Figura 12. Matriz de Confusão gerada após a análise de testes de generalização



Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

É importante promover uma discussão crítica sobre a redução observada nestas métricas em relação aos 95% de acurácia obtidos no conjunto de validação interno inicial. Longe de diminuir o mérito da arquitetura proposta, essa variação de desempenho revela uma significativa sensibilidade do modelo à mudança de domínio. Esse resultado evidencia cientificamente a necessidade estrita de validação ambiental para sistemas de monitoramento baseados em visão computacional. O teste de generalização demonstra que a introdução de um cenário inédito e novos ângulos de filmagem impõem variáveis físicas não mapeadas pelo treinamento original, provando que, em aplicações no mundo real, a eficácia do algoritmo não depende exclusivamente da capacidade da rede recorrente em aprender o padrão da queda, mas também de sua robustez diante de interferências do novo ambiente.

A análise detalhada dos erros evidenciou que as limitações na generalização estão atreladas ao processamento do AlphaPose diante de novos arranjos físicos. O modelo apresentou dificuldades de classificação quando partes do corpo ficaram sobrepostas por objetos e móveis não vistos previamente (oclusão). Esse fator comprova que a precisão da extração de *keypoints* depende da adequação do ambiente, tornando de suma importância o posicionamento estratégico da câmera em ângulos que possibilitem a visão completa do indivíduo. Adicionalmente, observou-se que a variabilidade na iluminação natural e artificial do novo cenário influencia ativamente a assertividade da classificação temporal.

Tais fatores apontam o caminho para aprimoramentos futuros, sugerindo que a implantação da tecnologia em ambiente doméstico exigirá calibrações ambientais prévias ou a aplicação de técnicas de aumento de dados visando tornar a detecção invariante a oclusões e oscilações de luminosidade.

4. CONSIDERAÇÕES FINAIS

Com base nos estudos realizados para a construção deste trabalho, pode-se perceber que no cenário atual existe uma crescente tendência aos lares unipessoais, especialmente entre a população idosa (CAVALCANTE *et al.*, 2022; CHUNG, 2025). Esta transformação traz consigo desafios e necessidades, dentre elas, os impactos causados pelo socorro tardio em casos de quedas e acidentes domésticos. Partindo dessa realidade, o presente trabalho buscou desenvolver um sistema de monitoramento com a premissa de identificar a ocorrência de quedas em ambientes domésticos de forma assertiva, utilizando técnicas de engenharia de *software*, inteligência artificial e visão computacional.

Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

Este estudo conclui que a integração entre Engenharia de *Software*, Visão Computacional e Saúde Digital na construção de soluções eficientes apresenta-se como uma ferramenta promissora para a detecção de quedas em ambientes domésticos. As métricas apresentadas neste trabalho fornecem evidências preliminares de bom desempenho nas condições avaliadas, indicando a eficiência do modelo em cumprir seus objetivos iniciais.

Foi necessária a construção de uma base de dados em vídeo própria. A captação das imagens, englobando tanto atividades cotidianas normais quanto diversas simulações de acidentes domésticos, permitiu criar uma estrutura adequada para o treinamento, teste e validação das arquiteturas de Inteligência Artificial adotadas na pesquisa.

No âmbito da visão computacional, o projeto obteve êxito ao integrar de forma fluida as ferramentas YOLOv3 e AlphaPose. Neste sentido, o YOLOv3 cumpriu o papel de detectar pessoas nos quadros com alta velocidade, enquanto o AlphaPose demonstrou elevada eficiência na estimativa e na extração dos pontos-chave das articulações corporais. A partir dos dados extraídos, avançou-se para o desenvolvimento e o treinamento da rede LSTM.

A arquitetura alcançou acurácia global de 95% e demonstrou capacidade de analisar sequencialmente os pontos-chave ao longo dos quadros de vídeo. Dessa forma, o modelo conseguiu aprender e classificar a diferença entre padrões de movimentos, atingindo também um *recall* de 98% para a classe de interesse no conjunto de testes internos. Com o intuito de investigar a capacidade de adaptação da solução, o sistema foi submetido também a um teste de generalização em um cenário residencial inédito. Diante de características não vistas durante o treinamento, como nova disposição de móveis e iluminação distinta, o modelo se mostrou eficiente, alcançando 80% de acurácia global e 85% de *recall* na detecção de quedas. Contudo, a introdução de novos elementos físicos evidenciou desafios práticos em que o desempenho do AlphaPose se mostrou sensível a oclusões e a baixa luminosidade.

Considerando as limitações identificadas nas fases de teste, como a dificuldade em identificar as quedas em ambientes com iluminação desfavorável ou oclusões e o tamanho limitado do banco de dados, é possível encontrar diferentes possibilidades de aprimoramento para trabalhos futuros. Dentre elas, destaca-se o aumento do banco de vídeos mediante a aplicação de técnicas de data augmentation e a inclusão de maior variabilidade de ambientes. Sugere-se também a avaliação de modelos de estimativa de pose mais recentes, o uso de múltiplas câmeras simultâneas para diminuir o risco de oclusões, e a realização de testes com pessoas idosas em ambientes cotidianos. Outra contribuição futura dá-se na implementação do algoritmo de inferência em tempo real para realização de testes em ambientes reais,

Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

possibilitando a comprovação da eficiência do algoritmo em situações cotidianas e o prosseguimento do estudo até a sua aplicação prática.

REFERÊNCIAS

AL-SELWI, S. M. *et al.* RNN-LSTM: From applications to modeling techniques and beyond Systematic review. Journal of King Saud University - Computer and Information Sciences, v. 36, n. 5, p. 102068, 2024.

BARBOSA, L. M.; PORTES, L. A. F. A Inteligência Artificial. Revista Tecnologia Educacional, v. 52, n. 236, p. 16-27, 2023.

BEHERA, R. K. *et al.* Co-LSTM: Convolutional LSTM model for sentiment analysis in social big data. Information Processing & Management, v. 58, n. 1, p. 102435, 2021.

BRITTO, N. P. *et al.* Potencializando o Engajamento no Ensino de Engenharia de Software e Requisitos: Relatos de Experiências com a Dinâmica Fábrica de Aviões. In: ANAIS do XXX Workshop de Informática na Escola (WIE 2024). Porto Alegre, RS, Brasil: Sociedade Brasileira de Computação, 2024. p. 91-100. DOI: 10.5753/wie.2024.242241.

CAETANO, G. M. *et al.* Risco de quedas e seus fatores associados na pessoa idosa hospitalizada. Revista Brasileira de Geriatria e Gerontologia, v. 26, e230155, 2023.

CAVALCANTE, P. P. S. *et al.* O Mercado Emergente de Lares de Pessoas que Moram Sozinhas: Estudo Introdutório Acerca das Variáveis que Explicam esse Fenômeno Mundial. In: ANAIS do 8º Encontro dos Programas de Pós-Graduação Profissionais em Administração (EMPRAD). São Paulo: FEAUSP, 2022. p. 1-16.

CHUNG, M. Uso de visão computacional para detecção de quedas em tempo real. 2025. Trabalho de Conclusão de Curso (Engenharia de Controle e Automação) – Instituto Federal de Educação, Ciência e Tecnologia do Amazonas, Campus Manaus Distrito Industrial, Manaus, 2025. Disponível em: https://www.ifam.edu.br/.

CORREA, J. P. C. *et al.* Saúde digital, inovação e cuidado: contribuições das tecnologias para o fortalecimento do SUS. Lumen et Virtus, [S. l.], v. 17, n. 61, p. e13475, 2026. DOI: 10.56238/levv17n61-037. Disponível em: https://periodicos.newsciencepubl.com/LEV/article/view/13475. Acesso em: 7 ago. 2026.

FANG, H.-S. *et al.* Alphapose: Whole-body regional multi-person pose estimation and tracking in real-time. IEEE Transactions on Pattern Analysis and Machine Intelligence, v. 45, n. 6, p. 7157-7173, 2022.

FAYYAD, U.; PIATETSKY-SHAPIRO, G.; SMYTH, P. From Data Mining to Knowledge Discovery in Databases. AI Magazine, v. 17, n. 3, p. 37, 1996.

GALVÃO, N. D.; MARIN, H. F. Técnica de mineração de dados: uma revisão da literatura. Acta Paulista de Enfermagem, São Paulo, v. 22, n. 5, p. 686-690, 2009. DOI: 10.1590/S0103-21002009000500014.

Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

GONÇALVES, A. C. L.; BERTOLIN, D. C.; GABRIEL, S. A. Dispositivos digitais para monitoramento de quedas em idosos. Revista Corpus Hippocraticum, v. 1, n. 2, 2025.

HUANG, T. *et al.* Fall detection model based on AlphaPose combined with LSTM and Lightgbm. In: MIPPR 2023: PATTERN RECOGNITION AND COMPUTER VISION. [S.l.]: SPIE, 2024. v. 13086, p. 130860g.

IGUAL, C.; MEDRANO, C.; PLAZA, I. Challenges, issues and trends in fall detection systems. BioMedical Engineering OnLine, v. 12, n. 1, p. 66, 2013.

INTO. Lançada campanha do INTO contra as quedas. [S.l.: s.n.], 2015. Disponível em: https://www.into.saude.gov.br/area-de-imprensa/noticias/140-into-lanca-campanha-contra-as-quedas. Acesso em: 9 maio 2026.

KALIL, K. *et al.* Quedas de pacientes idosos atendidos em uma unidade de pronto atendimento na cidade de Goianésia - GO. Revista Científica da Escola Estadual de Saúde Pública de Goiás "Cândido Santiago", v. 7, e7000045, 2021.

LINDSAY, G. W. Convolutional neural networks as a model of the visual system: past, present, and future. Journal of Cognitive Neuroscience, v. 33, n. 10, p. 2017-2031, 2021.

LUBNA, N.; MUFTI, N.; SHAH, S. A. A. Automatic Number Plate Recognition: A Detailed Survey of Relevant Algorithms. Sensors, v. 21, n. 9, p. 3028, 2021. Disponível em: https://doi.org/10.3390/s21093028. Acesso em: 30 ago. 2026.

PYTHON SOFTWARE FOUNDATION. shutil — Operações de arquivo de alto nível. Documentação Python 3.8.20. [S. l.], [s.d.]. Disponível em: https://docs.python.org/pt-br/3.8/library/shutil.html. Acesso em: 31 jul. 2026.

RACHID, R. *et al.* Saúde digital e a plataformização do Estado brasileiro. Ciência & Saúde Coletiva, v. 28, p. 2143-2153, 2023. Disponível em: https://doi.org/10.1590/1413-81232023287.14302022. Acesso em: 6 ago. 2026.

REDMON, J.; FARHADI, A. YOLOv3: An Incremental Improvement. [S.l.: s.n.], 2018. arXiv: 1804.02767 [cs.CV].

SAUVALLE, B.; LA FORTELLE, A. de. Unsupervised Multi-Object Segmentation Using Attention and Soft-Argmax. In: PROCEEDINGS OF THE IEEE/CVF WINTER CONFERENCE ON APPLICATIONS OF COMPUTER VISION (WACV). [S.l.: s.n.], 2023. p. 3267-3276.

SCHREIBER, G. Aplicação de aprendizado de máquina a sistemas embarcados para identificação de quedas de idosos. 2025. Dissertação (Mestrado em Ciência da Computação) – Pontifícia Universidade Católica do Rio Grande do Sul, Programa de Pós-Graduação em Ciência da Computação, Escola Politécnica, Porto Alegre, 2025. Disponível em: https://tede2.pucrs.br/tede2/handle/tede/11946.

SILVA, C. R. D. V. *et al.* Conceito de saúde digital na atenção primária à saúde (2020-2022): um estudo baseado no método evolucionário de Rodgers. Boletim de Conjuntura (BOCA), Boa Vista, v. 17, n. 49, p. 432-454, 2024. DOI: 10.5281/zenodo.10565467. Disponível em: https://revista.ioles.com.br/boca/index.php/revista/article/view/3156. Acesso em: 19 ago. 2026.

Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

SILVA, L. R. A. B. da. Inteligência artificial em processos de extração de conhecimento KDD e KDT. REU – Revista de Estudos Universitários, Sorocaba, v. 46, n. 1, p. 161-180, jun. 2020.

STEINER, M. T. A. *et al.* Abordagem de um problema médico por meio do processo de KDD com ênfase à análise exploratória dos dados. Gestão & Produção, São Carlos, v. 13, n. 2, p. 325-337, maio/ago. 2006. DOI: 10.1590/S0104-530X2006000200013.

SZELISKI, R. Computer vision: algorithms and applications. 2. ed. Cham: Springer, 2022.

TIAN, Y.; YE, Q.; DOERMANN, D. YOLOv12: Attention-Centric Real-Time Object Detectors. [S.l.: s.n.], 2025. arXiv: 2502.12524 [cs.CV].

VAZQUEZ, C. E.; SIMÕES, G. S. Engenharia de Requisitos: Software Orientado ao Negócio. [S.l.]: Brasport, 2016. ISBN 9788574527901.

WAZLAWICK, R. Engenharia de Software: Conceitos e Práticas. 2. ed. [S.l.]: Elsevier Editora Ltda., 2019. ISBN 9788535292732.

WEN, X.; LI, W. Time Series Prediction Based on LSTM-Attention-LSTM Model. IEEE Access, v. 11, p. 48322-48331, 2023. DOI: 10.1109/ACCESS.2023.3276628.

WOSIK, J. *et al.* Telehealth transformation: COVID-19 and the rise of virtual care. Journal of the American Medical Informatics Association, v. 27, n. 6, p. 957-962, 2020. DOI: 10.1093/jamia/ocaa067. Disponível em: https://pubmed.ncbi.nlm.nih.gov/32311034/.

YANG, Y. *et al.* Lightweight Attention-Guided YOLO With Level Set Layer for Landslide Detection From Optical Satellite Images. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, v. 17, p. 3543–3559, 2024. DOI: 10.1109/JSTARS.2024.3351277.

ZENG, P.; JIANG, S. Semi-parametric Functional Classification via Path Signatures Logistic Regression. [S.l.: s.n.], 2025. arXiv: 2507.06637 [stat.ML].

ZHAO, X. *et al.* An Alphapose-Based Pedestrian Fall Detection Algorithm. In: SUN, Xingming *et al.* (ed.). Artificial Intelligence and Security. Cham: Springer International Publishing, 2022. p. 650–660. ISBN 978-3-031-06794-5.

Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.