UNIVERSIDADE FEDERAL DO MARANHÃO - UFMA
CENTRO DE CIÊNCIAS EXATAS E TECNOLOGIAS - CCET
COORDENAÇÃO DO CURSO DE ENGENHARIA DA COMPUTAÇÃO

JOSÉ MARQUES CARDOSO SOUZA

UMA APLICAÇÃO MÓVEL PARA ACOMPANHAMENTO DO IDOSO COM ARTRITE UTILIZANDO DEEP LEARNING

SÃO LUÍS
2026

JOSÉ MARQUES CARDOSO SOUZA

UMA APLICAÇÃO MÓVEL PARA ACOMPANHAMENTO DO IDOSO COM ARTRITE UTILIZANDO DEEP LEARNING

Trabalho de Conclusão de Curso 2 apresentado ao Curso de Bacharelado em Engenharia da Computação da Universidade Federal do Maranhão Campus São Luís, como requisito parcial para obtenção do grau de Bacharel em Engenharia da Computação.

Orientador: Dr. Haroldo Gomes Barroso Filho

SÃO LUÍS
2026

Ficha gerada por meio do SIGAA/Biblioteca com dados fornecidos pelo(a) autor(a).
Diretoria Integrada de Bibliotecas/UFMA

Cardoso Souza, Jose Marques.
UMA APLICAÇÃO MÓVEL PARA ACOMPANHAMENTO DO IDOSO COM ARTRITE UTILIZANDO DEEP LEARNING / Jose Marques Cardoso Souza. - 2026.
53 f.

Orientador(a): Dr. Haroldo Gomes Barroso Filho.
Curso de Engenharia da Computação, Universidade Federal do Maranhão, Espaço Baites, 2026.

1. Artrite. 2. Idoso. 3. Deep Learning. 4. Visão Computacional. 5. Classificação de Exercícios. I. Gomes Barroso Filho, Dr. Haroldo. II. Título.

JOSE MARQUES CARDOSO SOUZA

UMA APLICAÇÃO MÓVEL PARA ACOMPANHAMENTO DO IDOSO COM ARTRITE UTILIZANDO DEEP LEARNING

Trabalho de Conclusão de Curso apresentado à Universidade Federal do Maranhão, em cumprimento às exigências institucionais para obtenção do grau de Bacharel em Engenharia da Computação, sob orientação do Prof. Dr. Haroldo Gomes Barroso Filho.

Aprovado em 22 de janeiro de 2026.

BANCA EXAMINADORA

Prof. Dr. Haroldo Gomes Barroso Filho
1º Membro da Comissão Avaliadora
Universidade Federal do Maranhão – UFMA

Prof. Dr. Pedro Baptista Fernandes
2º Membro da Comissão Avaliadora (Orientador)
Universidade Federal do Maranhão – UFMA

Prof. Me. Marcio Mendes Cerqueira
3º Membro da Comissão Avaliadora
Universidade Federal do Maranhão – UFMA

RESUMO

O envelhecimento populacional tem ampliado a incidência de doenças crônicas que comprometem a autonomia do idoso, entre elas a artrite, especialmente quando afeta as articulações das mãos. Nesses casos, atividades simples do cotidiano tornam-se difíceis, e a realização correta de exercícios terapêuticos passa a ser essencial para preservar a funcionalidade e reduzir desconfortos. No entanto, muitos idosos encontram dificuldades para manter essa prática fora do ambiente clínico, seja por insegurança na execução dos movimentos, seja pela ausência de acompanhamento contínuo. Diante desse cenário, este trabalho propõe o desenvolvimento de uma aplicação móvel voltada ao acompanhamento de idosos com artrite, utilizando técnicas de *deep* learning aplicadas à visão computacional. A solução é baseada em um pipeline que integra detecção da mão, segmentação, estimação de pose e classificação automática dos exercícios terapêuticos, empregando a arquitetura *YOLO (You Only Look Once)*. A partir de vídeos capturados pelo próprio usuário, o sistema identifica o exercício realizado, apresenta informações sobre seus benefícios terapêuticos e disponibiliza um vídeo processado com a identificação visual do movimento. O processamento das informações é realizado em um servidor, enquanto a aplicação móvel atua como interface de interação, tornando o uso do sistema simples e acessível. Os resultados obtidos demonstram que a abordagem proposta é viável para o reconhecimento automático de exercícios da mão em condições não controladas, contribuindo para maior segurança, autonomia e adesão ao tratamento. Dessa forma, a aplicação se apresenta como uma ferramenta de apoio complementar ao acompanhamento profissional, com potencial para auxiliar na promoção da qualidade de vida de idosos com artrite.

Palavras-chave: Artrite; Idoso; *Deep Learning*; Visão Computacional; Classificação de Exercícios; *YOLO*.

ABSTRACT

The advancement of population aging has increased the prevalence of chronic conditions that compromise the autonomy of older adults, among which arthritis stands out, especially when it affects the joints of the hands. In such cases, simple daily activities become difficult, and the correct practice of therapeutic exercises becomes essential to preserve functionality and reduce discomfort. However, many older adults face difficulties in maintaining this practice outside the clinical environment, either due to insecurity in performing the movements or the lack of continuous monitoring. In this context, this work proposes the development of a mobile application aimed at monitoring older adults with arthritis, using deep learning techniques applied to computer vision. The proposed solution is based on a processing pipeline that integrates hand detection, segmentation, pose estimation, and automatic classification of therapeutic exercises, employing the YOLO architecture. From videos captured by the user, the system identifies the performed exercise, provides information about its therapeutic benefits, and makes available a processed video with visual identification of the movement. The processing is carried out on a server, while the mobile application acts as an interaction interface, ensuring simple and accessible use. The results indicate that the proposed approach is feasible for the automatic recognition of hand exercises in uncontrolled environments, contributing to greater safety, autonomy, and adherence to treatment. Thus, the application presents itself as a complementary support tool to professional monitoring, with potential to assist in promoting the quality of life of older adults with arthritis.

Keywords: Arthritis; Older Adults; Deep Learning; Computer Vision; Exercise Classification; YOLO.

Sumário

1 INTRODUÇÃO

1.1 Motivação

O envelhecimento populacional tem ampliado a incidência de doenças crônicas que comprometem a funcionalidade e a autonomia, entre elas a artrite, especialmente quando afeta as articulações das mãos. Nessa condição, atividades simples do dia a dia — como segurar talheres, abotoar roupas, abrir recipientes ou manusear objetos — podem se tornar difíceis e dolorosas, o que impacta diretamente a independência do idoso e sua qualidade de vida. Estudos indicam que doenças crônicas musculoesqueléticas estão entre os principais fatores associados à perda funcional na velhice (DUARTE; ANDRADE; LEBRÃO, 2020).

Apesar de a prática regular de exercícios terapêuticos para as mãos ser amplamente recomendada e apresentar benefícios clínicos, muitos idosos enfrentam barreiras para manter a rotina de forma consistente. Entre os principais obstáculos estão a insegurança em executar os movimentos corretamente sem supervisão, a dificuldade de lembrar a sequência dos exercícios e a falta de acompanhamento contínuo fora do ambiente clínico. Como consequência, a adesão ao tratamento pode cair justamente quando a continuidade é mais necessária.

Nesse cenário, soluções digitais baseadas em visão computacional e *deep learning* surgem como uma alternativa promissora para apoiar o idoso durante a realização dos exercícios, oferecendo orientação e *feedback* de forma acessível. Ao reconhecer automaticamente o exercício que está sendo realizado e apresentar informações claras sobre sua execução e benefícios terapêuticos, uma aplicação móvel pode contribuir para aumentar a confiança do usuário, reduzir o medo de “fazer errado” e favorecer a manutenção do cuidado no cotidiano.

1.2 Justificativa

A escolha do tema deste trabalho justifica-se pela convergência entre três fatores relevantes: o envelhecimento populacional, a alta prevalência da artrite em idosos e o avanço recente das tecnologias baseadas em *deep learning* aplicadas à saúde. Embora a literatura aponte benefícios claros da realização de exercícios terapêuticos para as mãos, observa-se que muitos idosos têm dificuldade em manter a prática de forma correta e contínua, especialmente fora do ambiente clínico, onde não há supervisão direta do profissional de saúde.

Além disso, grande parte das soluções tecnológicas existentes voltadas à reabilitação física apresenta foco em movimentos corporais amplos ou em exercícios genéricos, não contemplando de maneira específica os movimentos finos da mão, que são justamente os mais afetados pela artrite. Essa lacuna torna-se ainda mais relevante quando se considera que pequenas variações na execução desses exercícios podem comprometer seus benefícios terapêuticos ou gerar insegurança no usuário.

Do ponto de vista tecnológico, o uso de modelos de *deep learning*, em especial arquiteturas voltadas à detecção, estimação de pose, segmentação e classificação, permite analisar movimentos com maior precisão e sensibilidade a detalhes. Integrar essas técnicas em uma aplicação móvel amplia o potencial de acesso, uma vez que *smartphones* são dispositivos amplamente difundidos e de uso familiar para uma parcela crescente da população idosa.

Assim, este trabalho se justifica por propor uma solução que alia tecnologia e cuidado em saúde, buscando apoiar o idoso com artrite na execução de exercícios da mão de forma mais segura, clara e autônoma. Ao oferecer reconhecimento automático do exercício realizado e informações sobre seus benefícios, a aplicação pretende contribuir para a adesão ao tratamento, a promoção da autonomia funcional e a melhoria da qualidade de vida, sem substituir o acompanhamento profissional, mas atuando como um suporte complementar.

1.3 Objetivos

Desenvolver uma aplicação móvel baseada em técnicas de *deep learning* capaz de reconhecer exercícios terapêuticos da mão realizados por idosos com artrite, oferecendo suporte à prática correta dos movimentos e contribuindo para a autonomia e a adesão ao tratamento.

1.3.2 Objetivos Específicos

Estudar e selecionar técnicas de visão computacional e *deep learning* adequadas para a análise de movimentos da mão, com ênfase em detecção, estimação de pose, segmentação e classificação.

Construir um conjunto de dados representativo contendo exercícios terapêuticos da mão, considerando variações naturais de execução.

Treinar e avaliar modelos baseados na arquitetura *YOLO* para o reconhecimento dos exercícios selecionados.

Implementar um *pipeline* de processamento que integre detecção da mão, análise do movimento e identificação do exercício realizado.

Desenvolver uma aplicação móvel que apresente, de forma clara e acessível, o exercício identificado e seus benefícios terapêuticos.

Avaliar o desempenho da arquitetura de processamento baseada em *deep learning*, utilizando métricas apropriadas para verificar a precisão dos modelos de detecção, segmentação, estimação de pose e classificação.

1.4 Estrutura do trabalho

Este trabalho está organizado em seis capítulos, estruturados de forma a conduzir o leitor da contextualização do problema até a apresentação dos resultados e conclusões.

O Capítulo 1 apresenta a introdução do estudo, abordando a motivação, a justificativa, os objetivos e a organização geral do trabalho.

O Capítulo 2 contempla a fundamentação teórica, na qual são discutidos os principais conceitos relacionados à artrite, ao envelhecimento e à exclusão social do idoso, bem como os fundamentos de *deep learning* e da arquitetura YOLO aplicados ao reconhecimento de movimentos. Também são apresentados os trabalhos correlatos que embasam e contextualizam a proposta desenvolvida.

No Capítulo 3, é descrita a metodologia adotada, incluindo o *pipeline* do sistema, a caracterização do conjunto de dados, os exercícios da mão considerados, as etapas de detecção, estimação de pose, segmentação e classificação, além da implementação da aplicação móvel.

O Capítulo 4 trata das métricas de avaliação utilizadas para analisar o desempenho do sistema, apresentando os conceitos, as formulações e os critérios empregados.

O Capítulo 5 apresenta os resultados obtidos a partir dos experimentos realizados, discutindo o desempenho do modelo e sua adequação ao objetivo proposto.

Por fim, o Capítulo 6 apresenta as conclusões do trabalho, destacando as contribuições alcançadas, as limitações identificadas e sugestões para trabalhos futuros.

2 FUNDAMENTAÇÃO TEÓRICA

Este capítulo apresenta a fundamentação teórica que dá suporte ao desenvolvimento da aplicação proposta. Inicialmente, são discutidos trabalhos correlatos que exploram o uso de visão computacional e *deep learning* como ferramentas de apoio à reabilitação física. Em seguida, abordam-se aspectos relacionados à artrite e ao processo de envelhecimento, destacando seus impactos funcionais e sociais na vida do idoso. Por fim, são apresentados os principais conceitos de *deep learning* e da arquitetura *YOLO*, assim como as técnicas de detecção, estimação de pose, segmentação e classificação, que sustentam tecnicamente a solução desenvolvida neste trabalho.

2.1 Trabalhos correlatos

Nos últimos anos, diferentes pesquisas têm investigado como a visão computacional e o aprendizado profundo podem auxiliar na reabilitação física, principalmente no reconhecimento automático de movimentos. De maneira geral, esses estudos mostram que modelos baseados em *deep learning* conseguem identificar gestos e padrões de movimento com uma precisão elevada o suficiente para oferecer um apoio complementar ao usuário durante a execução dos exercícios — sem substituir a orientação profissional, que permanece indispensável no processo terapêutico.

Chen et al. (2020), por exemplo, apresentaram um sistema voltado ao reconhecimento de gestos da mão utilizando arquiteturas convolucionais. O trabalho mostrou que movimentos pequenos e sutis podem ser diferenciados com bastante precisão, o que abre espaço para aplicações fora do ambiente clínico, como acompanhamento remoto ou suporte ao paciente durante atividades diárias. Em outra perspectiva, Li e Zhao (2021) utilizaram pontos-chave articulares *(keypoints)* para analisar movimentos terapêuticos e observaram que os usuários demonstram maior envolvimento quando recebem uma interpretação clara e imediata do gesto executado.

Huang et al. (2022) direcionaram sua pesquisa para exercícios realizados pelos membros superiores e testaram modelos de estimação de pose como ferramenta de classificação desses movimentos. Um dos resultados mais relevantes foi a percepção de segurança relatada por usuários idosos, que afirmaram cometer

menos erros ao receber feedback visual sobre a execução. Já Santos e Melo (2021), ao analisarem tecnologias digitais aplicadas à artrite, destacaram que soluções de apoio tendem a reduzir o receio de praticar exercícios longe do terapeuta, contribuindo para a continuidade do tratamento. De forma complementar, Wang et al. (2023) demonstraram que algoritmos de reconhecimento de gestos são capazes de identificar corretamente movimentos simples da mão, favorecendo a autonomia do idoso em tarefas do cotidiano.

Em conjunto, esses trabalhos deixam claro que ferramentas baseadas em aprendizado profundo têm potencial para apoiar o processo terapêutico, oferecendo mais segurança, confiança e regularidade na prática dos exercícios. No entanto, nota-se que a maioria das pesquisas aborda movimentos corporais de maneira ampla ou exercícios gerais de membros superiores, sem um foco específico nos exercícios da mão voltados ao idoso com artrite. Essa lacuna justifica a proposta do presente estudo, que busca desenvolver um sistema capaz de reconhecer esses movimentos específicos e reforçar seus benefícios terapêuticos de forma simples, acessível e não invasiva. O Quadro 1 apresenta uma síntese comparativa dos principais trabalhos correlatos analisados.

Quadro 1 – Comparação dos trabalhos correlatos

Autor / Ano	**Técnica Utilizada**	**Objetivo do Estudo**	**Público-alvo**	**Principais Resultados**	**Limitações**	**Relação com o Presente Estudo**
Chen et al. (2020)	*CNN*	Reconhecer gestos da mão	Adultos	Alta precisão na identificação de gestos	Não foca em idosos	Viabilidade do reconhecimento de movimentos da mão
Li e Zhao (2021)	*Keypoints + DL*	Identificar movimentos terapêuticos	Adultos	Maior engajamento com *feedback* automático	Não específico para mãos	Reforça a importância do *feedback* ao usuário
Huang et al. (2022)	Estimação de pose	Classificar exercícios de membros superiores	Idosos	Redução de erros e maior segurança	Não aborda exercícios manuais	Mostra benefícios do *feedback* visual
Santos e Melo (2021)	Revisão bibliográ-fica	Analisar tecnologias na artrite	Idosos	Aumento da adesão ao tratamento	Não propõe sistema prático	Justifica relevância social do tema
Wang et al. (2023)	DL para gestos manuais	Apoiar autocuidado do idoso	Idosos	Boa precisão em gestos simples	Não foca exercícios terapêuti-cos	Evidencia a lacuna abordada

Fonte: Elaborado pelo autor, 2026.

2.2 Artrite

A artrite, especialmente em suas formas degenerativas, está entre as condições crônicas mais frequentes no envelhecimento e figura como uma das principais causas de dor persistente, limitação funcional e perda gradual de autonomia. Como destacam Duarte et al. (2020), o processo inflamatório que caracteriza a doença compromete articulações essenciais para tarefas cotidianas, particularmente as pequenas articulações das mãos, responsáveis pela mobilidade fina, pela manipulação de objetos e por grande parte das ações de autocuidado.

Com a evolução do quadro, torna-se comum que a pessoa idosa apresente rigidez ao despertar, dificuldade de preensão e perda de força para realizar movimentos simples, como abotoar uma peça de roupa, segurar utensílios ou manusear embalagens. Esses desafios, segundo Oliveira e Lima (2020), ultrapassam a dimensão física e afetam diretamente a autoestima, a sensação de independência e a percepção de capacidade funcional — aspectos fundamentais para uma vida ativa e socialmente integrada.

A literatura aponta que exercícios específicos para as mãos — como movimentos de pinça, aproximação e afastamento dos dedos, arranhar e fechar completamente a mão — podem trazer benefícios consistentes, incluindo aumento da amplitude de movimento, redução da dor e fortalecimento da musculatura intrínseca (Barbosa et al., 2021). Entretanto, tais resultados dependem da repetição adequada e da execução correta dos movimentos, o que nem sempre se mantém quando o idoso realiza os exercícios de forma independente.

Nesse cenário, recursos tecnológicos podem desempenhar um papel importante. Mendes (2022) observa que ferramentas acessíveis, capazes de orientar passo a passo e esclarecer dúvidas durante a prática, diminuem o receio de realizar o exercício “de maneira errada” e tendem a aumentar a adesão ao tratamento. Um sistema que reconhece automaticamente qual exercício está sendo executado e apresenta seus benefícios terapêuticos contribui para uma experiência mais segura, clara e motivadora — sobretudo para idosos com artrite, que frequentemente enfrentam incertezas quanto à forma correta de realizar os movimentos.

2.2.1 Idoso como excluído socialmente

O envelhecimento no Brasil ocorre em um contexto marcado por desigualdades sociais que, somadas a estereótipos culturais, podem colocar a pessoa idosa em posição de vulnerabilidade. Debert (2019) destaca que a exclusão social desse grupo não se resume à ausência de políticas públicas, mas é reforçada por representações que associam velhice à incapacidade, limitando a participação social e alimentando um ciclo de dependência.

A artrite tende a intensificar esse quadro. Para Santos et al. (2021), a dor crônica e a dificuldade em realizar movimentos funcionais levam muitos idosos a reduzir suas interações sociais e evitar atividades que exigem esforço manual. Além

disso, não é incomum que o medo de piorar os sintomas ou de executar incorretamente um exercício recomendado pelo terapeuta acabe diminuindo a adesão ao tratamento. Essa insegurança é ainda maior quando a prática ocorre no ambiente doméstico, sem supervisão direta.

Mendes (2022) argumenta que tecnologias simples, intuitivas e capazes de oferecer retorno visual imediato podem funcionar como mediadoras da autonomia, ajudando o idoso a se sentir apoiado mesmo quando está sozinho. A orientação clara reduz dúvidas, reforça a autoconfiança e contribui para a continuidade da prática terapêutica.

Assim, um sistema que identifica automaticamente o exercício da mão realizado e explica seu propósito terapêutico vai além de um apoio técnico: ele atua como facilitador da inclusão. Ao proporcionar segurança e clareza no momento da prática, a tecnologia reduz medos, diminui barreiras sociais e possibilita que o idoso mantenha sua rotina de cuidados com mais independência. Encarar o idoso como sujeito em risco de exclusão permite compreender a relevância social deste trabalho, que não se limita ao desenvolvimento de uma ferramenta digital, mas busca promover autonomia funcional e favorecer um envelhecimento mais ativo, seguro e participativo.

2.3 Deep Learning

O *deep learning* tornou-se uma das abordagens mais influentes na área de visão computacional devido à sua capacidade de identificar padrões complexos em grandes volumes de dados (GOODFELLOW; BENGIO; COURVILLE, 2016; LECUN; BENGIO; HINTON, 2015). Ao contrário de métodos tradicionais, que dependem de descritores manuais, modelos de *deep learning* aprendem representações diretamente a partir dos exemplos, refinando suas próprias características internas durante o treinamento. Essa capacidade de extrair automaticamente informações relevantes torna a técnica especialmente adequada para o reconhecimento de movimentos finos da mão, que envolve variações sutis de posição, ângulo e articulação — aspectos fundamentais em exercícios terapêuticos para artrite (CHEN; LI; ZHANG, 2020; HUANG; LIU; WANG, 2022) .

No nível matemático, a base do *deep learning* são as redes neurais artificiais, compostas por camadas de neurônios que realizam transformações sucessivas

sobre um vetor de entrada. Cada neurônio computa uma combinação linear de suas entradas e aplica uma função de ativação não linear, conforme as Equações (1) e (2) (HAYKIN, 2009; NIELSEN, 2015):

$$z = W.x + b \quad (1)$$

$$a = \phi(z) \quad (2)$$

onde:

- W representa os pesos aprendidos,
- b é o termo de viés,
- $\phi(\cdot)$ é a função de ativação, responsável por introduzir não linearidade no modelo.

A função ReLU (Rectified Linear Unit), uma das mais utilizadas, é definida como na Equação (3) (GOODFELLOW; BENGIO; COURVILLE, 2016):

$$ReLU(z) = max(0, z) \quad (3)$$

Ela acelera o aprendizado ao evitar saturações e facilita a propagação do gradiente nas camadas profundas.

Em modelos aplicados à visão computacional, como os utilizados neste trabalho, destaca-se o papel das camadas convolucionais, que realizam operações de convolução entre a imagem e um conjunto de filtros treináveis. Cada filtro aprende a detectar padrões específicos por meio da operação de convolução, conforme descrito na Equação (4) — como bordas, texturas ou pequenas regiões características dos gestos da mão (KRIZHEVSKY; SUTSKEVER; HINTON, 2012; RAWAT; WANG, 2017):

$$(f * g)(i,j) = \sum_m \sum_n f(m,n) \cdot g(i - m, j - n) \quad (4)$$

Essas camadas atuam como extratoras automáticas de características, tornando possível distinguir variações sutis entre um movimento de pinça, arranhar ou lateralização dos dedos. Complementando esse processo, camadas de *pooling* reduzem a dimensionalidade dos mapas de ativação, preservando padrões essenciais e diminuindo o custo computacional, o que é especialmente relevante em aplicações móveis (LECUN; BENGIO; HINTON, 2015).

O treinamento de modelos de *deep learning* envolve um processo iterativo de otimização. O objetivo é minimizar uma função de perda — como a entropia cruzada

— que quantifica a diferença entre a saída prevista e o rótulo correto, conforme apresentado na Equação (5) (RUMELHART; HINTON; WILLIAMS, 1986):

$$L = -\sum_{c=1}^{c} y_c \log log \left(\hat{y}_c\right) \quad (5)$$

A atualização dos pesos ocorre por meio do método do gradiente descendente e sua generalização, o *backpropagation*, que calcula o gradiente da perda em relação a cada parâmetro da rede. Otimizadores modernos, como o *Adam*, ajustam dinamicamente a taxa de aprendizado, acelerando a convergência e proporcionando maior estabilidade durante o treinamento.

Essa arquitetura baseada em múltiplas transformações sucessivas permite que redes profundas capturem relações complexas entre posições articulares e movimentos funcionais. Por isso, o *deep learning* vem sendo amplamente adotado em tarefas de detecção, segmentação, estimação de pose e classificação — etapas essenciais para reconhecer exercícios terapêuticos executados pela mão. No contexto da artrite, essa capacidade se torna ainda mais relevante: pequenas alterações de ângulo, amplitude ou rigidez podem indicar dificuldades específicas, e modelos bem treinados conseguem representar essas diferenças de maneira precisa e consistente.

Assim, o *deep learning* oferece a base computacional necessária para que o sistema proposto identifique corretamente o movimento realizado pelo idoso, fornecendo apoio técnico para uma prática mais segura, clara e orientada, sem a necessidade de supervisão constante.

2.3.1 *YOLO*

família de modelos *YOLO* consolidou-se como uma das abordagens mais eficientes para tarefas de visão computacional em tempo real. Diferentemente de métodos que realizam várias etapas para detectar objetos — como proposta de regiões, classificação e refinamento — o *YOLO* trata todo o processo como um único problema de regressão, prevendo simultaneamente as classes e as coordenadas dos objetos presentes na imagem. Essa estratégia torna o modelo extremamente rápido, característica essencial para aplicações móveis e interativas, como o reconhecimento de exercícios da mão proposto neste estudo.

O princípio básico da *YOLO* consiste em dividir a imagem em uma grade espacial e, para cada célula dessa grade, prever um conjunto de caixas delimitadoras (*bounding boxes*), suas respectivas probabilidades e os rótulos das classes. No nível matemático, cada predição combina termos conforme a formula (6) (REDMON; FARHADI, 2018):

- coordenadas normalizadas (y)do centro da caixa,
- largura e altura (h),
- confiança da detecção (C),
- probabilidade da classe $(p(c))$.

A predição final é dada conforme a Equação:

$$score = C \cdot p(c) \quad (6)$$

O modelo aprende essas estimativas ajustando seus pesos para minimizar uma função de perda que integra diferentes componentes, como erro de localização, confiança da caixa e classificação da classe. Uma forma simplificada dessa função de perda pode ser escrita conforme a Equação (7) (WANG et al., 2022):

$$L = \lambda coord \cdot MSE(bbox) + \lambda obj \cdot BCE(C) + \lambda cls \cdot CE(p(c)) \quad (7)$$

onde *MSE*, *BCE* e *CE* representam diferentes tipos de erro (quadrático, binário e entropia cruzada), cada um controlado por hiperparâmetros que equilibram sua contribuição.

A versão utilizada neste trabalho, a *YOLOv11*, incorpora melhorias significativas em relação às gerações anteriores, principalmente na extração de características e na eficiência de inferência. Essas melhorias envolvem arquiteturas mais profundas, blocos convolucionais mais leves e mecanismos internos que aumentam a precisão sem comprometer a velocidade — fator crucial quando o processamento precisa ocorrer diretamente em dispositivos móveis ou em ambientes com recursos computacionais limitados.

Além disso, a estrutura modular da *YOLOv11* permite integrar facilmente cabeçalhos adicionais (*heads*) para tarefas específicas, como detecção, segmentação, estimação de pose e classificação, o que a torna especialmente adequada para o reconhecimento de exercícios terapêuticos da mão. A detecção das regiões relevantes, a identificação das articulações e a classificação do exercício executado — distinguindo movimentos como pinça, arranhar ou

lateralização dos dedos — podem ser tratadas dentro de um único arcabouço, mantendo coerência, estabilidade e eficiência nas previsões.

Outro aspecto importante é que a *YOLO* opera de maneira totalmente supervisionada, o que possibilita treinar o modelo com exemplos reais dos exercícios realizados pelos usuários. À medida que novos dados são incorporados, o modelo se ajusta e se torna mais sensível a variações sutis de movimento, postura e amplitude — características essenciais no contexto da artrite, em que pequenas diferenças na articulação dos dedos podem indicar limitações funcionais relevantes.

Assim, a *YOLOv11* funciona como o núcleo do sistema desenvolvido neste trabalho, permitindo que a aplicação reconheça, com rapidez e precisão, qual exercício está sendo realizado. Essa capacidade fornece ao idoso um retorno imediato sobre o movimento executado, contribuindo para uma prática mais segura e para um acompanhamento terapêutico mais claro e acessível.

2.3.1.4 Classificação

A classificação consiste no processo de atribuir um rótulo a uma entrada visual com base em padrões aprendidos durante o treinamento do modelo (GOODFELLOW; BENGIO; COURVILLE, 2016). Diferentemente da detecção e da segmentação, que se concentram na localização espacial dos objetos, a classificação tem como objetivo identificar a qual classe uma imagem pertence, considerando suas características globais. Em aplicações terapêuticas, essa técnica é especialmente relevante para reconhecer automaticamente qual exercício está sendo realizado pelo usuário.

Em modelos baseados em *deep learning*, a classificação é geralmente realizada por redes neurais convolucionais, que aprendem representações hierárquicas da imagem por meio de sucessivas camadas de convolução e *pooling*. Essas camadas extraem características visuais discriminantes — como formas, padrões de movimento e configurações articulares — que são posteriormente utilizadas por camadas totalmente conectadas para a tomada de decisão.

Do ponto de vista matemático, a saída do modelo de classificação corresponde a um vetor de probabilidades associado às classes possíveis. Essa saída é comumente obtida por meio da função *softmax*, definida na Equação (8) (NIELSEN, 2015):

$P(c_i) = \frac{e^{z_i}}{\sum_{j=1}^{K} e^{z_j}}$ (8) em que z_i representa a ativação associada à classe c_i e K corresponde ao número total de classes. O modelo atribui à imagem a classe que apresenta a maior probabilidade estimada.

A arquitetura *YOLO*, além de suas aplicações consolidadas em detecção, segmentação e estimação de pose, também disponibiliza um módulo específico para classificação. Esse módulo permite treinar o modelo para distinguir diferentes exercícios terapêuticos a partir de imagens da mão, considerando padrões globais do movimento. Durante o treinamento, o ajuste dos pesos é realizado minimizando uma função de perda, geralmente baseada na entropia cruzada, que penaliza discrepâncias entre a classe prevista e o rótulo correto.

No contexto deste trabalho, a classificação é utilizada para identificar automaticamente o exercício terapêutico executado pelo idoso. Essa etapa complementa as demais técnicas empregadas, permitindo que o sistema associe o movimento observado a um exercício específico e forneça ao usuário informações claras sobre sua execução e seus benefícios terapêuticos.

2.3.1.1 Detecção

A detecção de objetos consiste em identificar automaticamente a presença e a localização de elementos de interesse em uma imagem ou sequência de imagens (REDMON et al., 2016). Diferentemente de tarefas puramente classificatórias, nas quais o modelo apenas indica a classe predominante, a detecção envolve também a estimativa da posição espacial do objeto, geralmente representada por meio de caixas delimitadoras (*bounding boxes*). Em aplicações de reabilitação, essa etapa é fundamental para isolar regiões relevantes do corpo, como a mão, antes da análise mais detalhada dos movimentos executados (HUANG; LIU; WANG, 2022).

Nos modelos da família YOLO, a detecção é tratada como um problema único e integrado. A imagem de entrada é processada por uma rede convolucional profunda que extrai características hierárquicas e, ao final, produz diretamente as coordenadas das caixas delimitadoras, a confiança da detecção e a classe associada. Essa abordagem elimina etapas intermediárias de proposta de regiões,

reduzindo o tempo de processamento e tornando o modelo adequado para aplicações em tempo real (REDMON; FARHADI, 2017).

Do ponto de vista matemático, cada detecção é descrita por um conjunto de parâmetros que representam a posição do objeto no plano da imagem. As coordenadas do centro da caixa (x | y), bem como sua largura w e altura h, são previstas de forma normalizada em relação às dimensões da imagem. Além disso, o modelo estima um valor de confiança que indica a probabilidade de existência do objeto naquela região. A combinação desses elementos permite calcular a pontuação final da detecção, utilizada para selecionar apenas os resultados mais relevantes (REDMON; FARHADI, 2018).

Durante o treinamento, o modelo ajusta seus pesos minimizando uma função de perda que penaliza erros de localização, classificação incorreta e falsas detecções. Um componente importante desse processo é a utilização da métrica *Intersection over Union* (IoU), que mede o grau de sobreposição entre a caixa prevista pelo modelo e a caixa real anotada no conjunto de dados. Valores elevados de IoU indicam maior precisão na localização do objeto detectado, sendo esse um critério amplamente adotado para avaliar o desempenho de modelos de detecção (WANG et al., 2022).

No contexto deste trabalho, a etapa de detecção tem como objetivo identificar corretamente a região da mão durante a execução dos exercícios terapêuticos. A partir dessa identificação, o sistema consegue concentrar o processamento apenas na área relevante da imagem, reduzindo ruídos e aumentando a confiabilidade das análises subsequentes. Esse enfoque é particularmente importante no caso de idosos com artrite, em que pequenas variações de postura ou movimento podem ser determinantes para a correta identificação do exercício realizado (SANTOS; MELO, 2021).

Assim, a detecção baseada na YOLOv11 fornece a base estrutural para o funcionamento do sistema proposto, permitindo localizar a mão de forma rápida e precisa e viabilizando as etapas seguintes de segmentação, estimação de pose e classificação do exercício, essenciais para a interpretação adequada dos movimentos terapêuticos (REDMON et al., 2016; WANG et al., 2022).

2.3.1.2 Estimação de pose

A estimação de pose tem como objetivo identificar a posição espacial de pontos-chave do corpo, conhecidos como *keypoints*, que representam articulações ou regiões anatômicas relevantes (CAO et al., 2017). No caso da mão, esses pontos correspondem, por exemplo, às articulações dos dedos, à palma e ao punho. Diferentemente da detecção, que apenas localiza a região de interesse, a estimação de pose permite compreender como o movimento está sendo executado, fornecendo uma descrição mais detalhada da postura e da dinâmica do gesto (HUANG; LIU; WANG, 2022).

Do ponto de vista computacional, a estimação de pose baseia-se na predição das coordenadas bidimensionais (ou tridimensionais, em alguns casos) de cada *keypoint* em relação à imagem de entrada. Em modelos modernos, essa tarefa é realizada por redes neurais profundas que aprendem relações espaciais entre diferentes partes do corpo, explorando padrões geométricos e temporais. Assim, o modelo não analisa cada articulação de forma isolada, mas considera o conjunto de pontos como um sistema articulado (CAO et al., 2017).

Matematicamente, cada *keypoint* pode ser representado por um par ordenado $\left(y_i\right)$, onde $i$indica a articulação correspondente. O treinamento do modelo consiste em minimizar a distância entre os pontos previstos e os pontos reais anotados no conjunto de dados, geralmente utilizando funções de perda baseadas no erro quadrático médio. Esse processo permite que a rede aprenda a localizar com precisão articulações mesmo em cenários com variações de iluminação, ângulo de câmera ou oclusão parcial (NIELSEN, 2015).

No contexto da reabilitação, a estimação de pose apresenta vantagens importantes. Ao identificar a posição relativa dos dedos e da mão, o sistema consegue distinguir movimentos que, à primeira vista, podem parecer semelhantes, mas que possuem finalidades terapêuticas diferentes. Exercícios como pinça, arranhar ou lateralização dos dedos diferem principalmente na configuração articular e na amplitude do movimento — aspectos que são capturados de forma eficaz pela análise dos *keypoints* (HUANG; LIU; WANG, 2022).

Para idosos com artrite, essa abordagem é particularmente relevante. Pequenas limitações de mobilidade ou rigidez articular podem alterar

significativamente a execução de um exercício. A estimação de pose permite identificar essas variações e fornece uma base mais precisa para reconhecer o movimento realizado, mesmo quando ele não é executado de forma perfeita. Dessa forma, o sistema se adapta à realidade do usuário, evitando interpretações rígidas que poderiam gerar frustração ou desmotivação (SANTOS; MELO, 2021).

No sistema proposto, a estimação de pose atua como uma camada intermediária entre a detecção da mão e a classificação final do exercício. A partir dos pontos-chave extraídos, o modelo consegue representar o movimento de maneira estruturada, possibilitando uma análise mais fiel da execução e fornecendo subsídios para a correta identificação do exercício realizado, contribuindo para um acompanhamento terapêutico mais sensível às limitações individuais do idoso.

2.3.1.3 Segmentação

A segmentação consiste no processo de separar, de forma precisa, as regiões relevantes de uma imagem, atribuindo a cada pixel uma classe específica (RAWAT; WANG, 2017). Diferentemente da detecção, que delimita o objeto por meio de uma caixa, a segmentação permite identificar com maior detalhamento o contorno e a forma do elemento de interesse. No contexto deste trabalho, essa abordagem é especialmente útil para isolar a mão do restante da cena, reduzindo interferências do fundo e aumentando a precisão da análise dos movimentos (WANG et al., 2022).

Em modelos baseados em *deep learning*, a segmentação é geralmente realizada por redes neurais que produzem máscaras binárias ou multiclasse, nas quais cada pixel indica a probabilidade de pertencer ao objeto segmentado. Essas redes aprendem a distinguir padrões visuais como textura, cor e bordas, combinando informações locais e globais da imagem. O resultado é uma representação mais fiel da geometria do objeto, o que se mostra vantajoso em tarefas que exigem maior sensibilidade a detalhes, como a análise de movimentos finos da mão (RAWAT; WANG, 2017).

Do ponto de vista matemático, a segmentação pode ser interpretada como um problema de classificação pixel a pixel. Durante o treinamento, o modelo busca minimizar uma função de perda que compara a máscara prevista com a máscara real anotada no conjunto de dados, frequentemente utilizando métricas como a

entropia cruzada ou variações baseadas no coeficiente de Dice. Essas métricas avaliam o grau de sobreposição entre as regiões segmentadas, incentivando o modelo a produzir contornos mais precisos (GOODFELLOW; BENGIO; COURVILLE, 2016).

No âmbito da reabilitação, a segmentação oferece benefícios adicionais quando combinada à detecção e à estimação de pose. Ao isolar a mão com maior exatidão, o sistema reduz ruídos causados por elementos externos, como objetos ao redor ou partes do corpo não relacionadas ao exercício. Isso é particularmente importante para idosos com artrite, que podem executar os movimentos em ambientes variados, sem controle rigoroso de iluminação ou fundo (SANTOS; MELO, 2021).

No sistema proposto, a segmentação contribui para uma análise mais robusta da execução dos exercícios da mão. Ao delimitar claramente a região analisada, o modelo consegue interpretar melhor a abertura, o fechamento e a disposição dos dedos, mesmo quando há limitações de mobilidade ou variações individuais na forma de executar o movimento. Dessa maneira, a segmentação atua como um refinamento final do processo de reconhecimento, fortalecendo a confiabilidade do sistema e ampliando sua capacidade de adaptação às condições reais de uso.

3 METODOLOGIA

Este capítulo descreve a metodologia adotada para o desenvolvimento do sistema proposto, apresentando as etapas seguidas desde a organização dos dados até a aplicação dos modelos de *deep learning* para a análise dos exercícios terapêuticos da mão. A abordagem adotada possui caráter experimental e foi estruturada de modo a refletir condições reais de uso, considerando as limitações naturais de execução dos movimentos por indivíduos com artrite.

O foco principal da metodologia é a análise automática de exercícios da mão por meio de técnicas de detecção, estimação de pose, segmentação e classificação, permitindo identificar o exercício realizado e fornecer suporte ao acompanhamento terapêutico.

3.1 *Pipeline* do Sistema

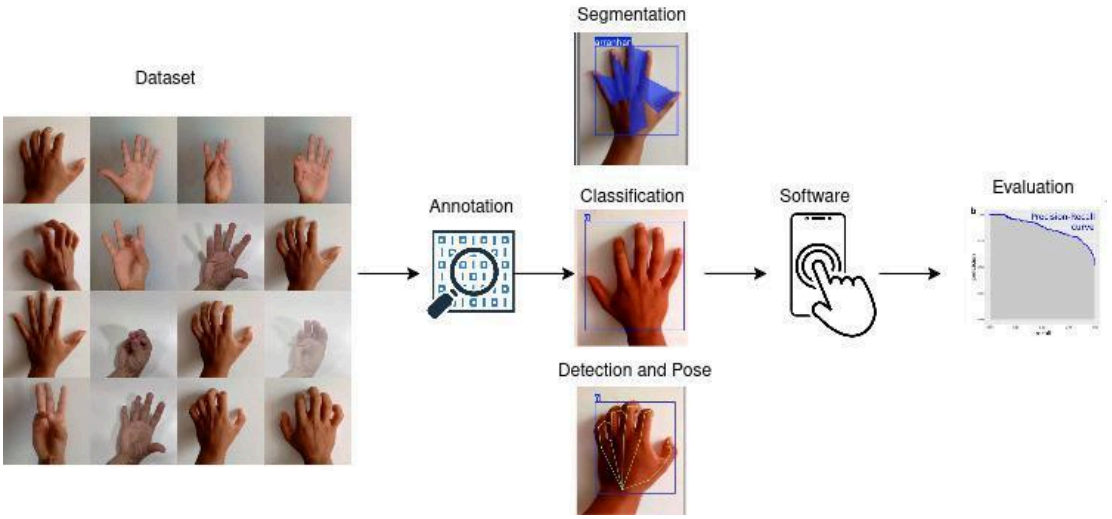
O *pipeline* do sistema foi definido para organizar, de forma clara e sequencial, todas as etapas envolvidas no processamento dos dados e no reconhecimento dos exercícios da mão. O fluxo inicia-se com a aquisição dos dados visuais e segue até a obtenção do resultado final, que corresponde à identificação do exercício executado por meio de um processo de classificação automática.

Inicialmente, os vídeos contendo a execução dos exercícios são capturados e organizados em conjuntos de treinamento e validação. A partir desses vídeos, quadros individuais são extraídos e preparados para serem utilizados como entrada dos modelos de *deep learning*. Esse preparo inclui a adequação do formato das imagens às exigências da arquitetura utilizada.

Após essa etapa, o *pipeline* é composto por quatro processos principais: detecção da mão, segmentação, estimação de pose e classificação do exercício. Cada um desses processos é tratado como uma tarefa específica, com modelos treinados de forma independente. Essa separação permite maior controle experimental, facilita a análise individual de cada etapa e contribui para a robustez do sistema como um todo.

Por fim, as informações obtidas nessas etapas são integradas com o objetivo de identificar automaticamente o exercício realizado e oferecer suporte ao acompanhamento terapêutico. A representação gráfica completa desse *pipeline* será apresentada conforme a Figura 1.

Figura 1 – Visão geral do *pipeline* do sistema proposto



Fonte: Elaborado pelo orientador, 2026.

3.2 *Dataset*

O conjunto de dados utilizado neste trabalho foi construído pelo próprio autor a partir da gravação de vídeos reais contendo a execução de exercícios terapêuticos da mão, que foram validados por um fisioterapeuta, com o objetivo de garantir a correta execução dos movimentos. Os vídeos foram capturados com resolução *Full* HD (1080p) e taxa de 30 quadros por segundo (30 fps), assegurando qualidade visual adequada para a extração das características necessárias às técnicas de visão computacional empregadas. A partir desses vídeos, foram extraídos quadros individuais (frames), gerando um conjunto de imagens estáticas no formato RGB, utilizadas como entrada para os modelos de detecção, segmentação, estimação de pose e classificação. Cada vídeo contribuiu com dezenas a centenas de imagens, dependendo de sua duração, resultando em um conjunto de dados composto por milhares de imagens ao final do processo de extração. Buscou-se contemplar variações naturais de execução, como mudanças de posicionamento da mão, diferenças de iluminação e variações no ritmo dos movimentos, de modo a aproximar o treinamento das condições reais de uso do sistema. A Figura 2 apresenta uma visão geral do *dataset*, composta por uma sequência de miniaturas

extraídas dos vídeos, ilustrando a diversidade de enquadramentos e execuções dos movimentos considerados.

Figura 2 – Sequência de miniaturas representativas do conjunto de dados utilizado

Fonte: Elaborado pelo autor, 2026.

Para o desenvolvimento do modelo principal descrito neste trabalho, foi selecionado o exercício arranhar, por se tratar de um movimento amplamente utilizado na reabilitação das mãos e relevante para indivíduos com artrite. A partir dos vídeos gravados desse exercício, foram extraídos quadros individuais que compõem o conjunto de imagens utilizado no treinamento das etapas de detecção, segmentação e estimação de pose. O treinamento do modelo *YOLO* foi realizado utilizando imagens extraídas de três vídeos do exercício arranhar, enquanto a etapa de validação foi conduzida com imagens provenientes de um vídeo distinto, garantindo a separação adequada entre os conjuntos e evitando sobreposição de dados.

Além do exercício de arranhar, os exercícios de fechamento em cúpula e oposição do polegar também compõem o *dataset*. No entanto, os três exercícios são utilizados conjuntamente apenas no treinamento do modelo final de classificação. As

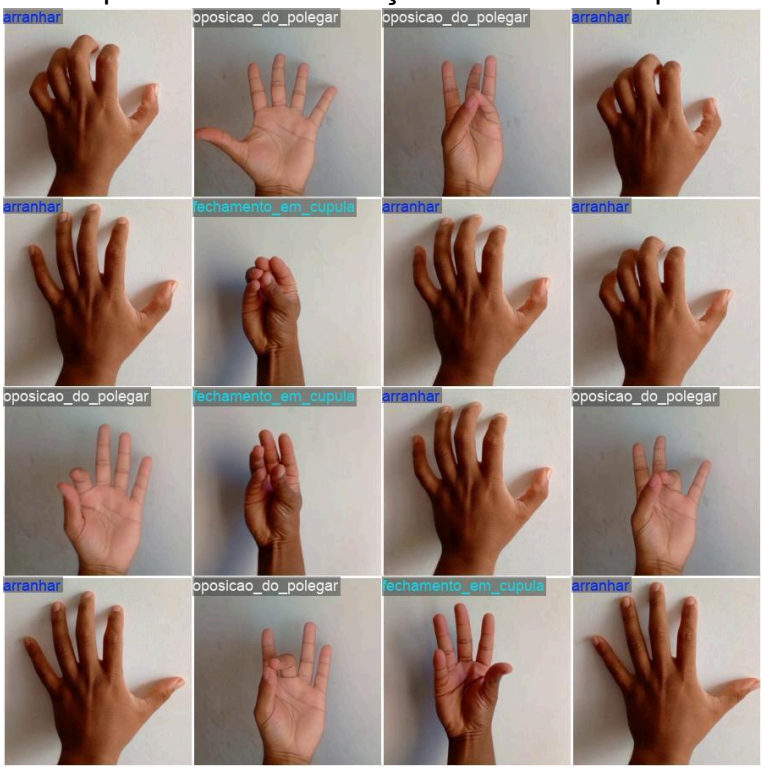
imagens extraídas dos vídeos desses exercícios foram empregadas exclusivamente nessa etapa, permitindo que o sistema aprendesse a distinguir entre diferentes movimentos terapêuticos da mão. Para as demais etapas do pipeline, os exercícios adicionais são apresentados e discutidos em apêndice, como forma de complementar o estudo e indicar possibilidades de ampliação futura do sistema.

Os dados foram organizados seguindo o padrão exigido pela ferramenta *Ultralytics YOLO*, com separação adequada entre imagens e anotações, permitindo a reprodução dos experimentos realizados.

3.6 Classificação de Exercícios

A etapa de classificação foi adotada com o objetivo de identificar automaticamente qual exercício terapêutico da mão está sendo executado pelo usuário. A Figura 3 apresenta um exemplo do experimento de classificação utilizando a arquitetura *YOLO*, ilustrando a atribuição do rótulo correspondente ao exercício identificado durante o processamento do vídeo.

Figura 3 – Experimento classificação utilizando a arquitetura *YOLO*



Fonte: Elaborado pelo autor, 2026.

Para essa tarefa, foi utilizado o módulo de classificação da *YOLO*, treinado com imagens representativas dos exercícios considerados no sistema. O modelo

recebe como entrada imagens previamente processadas pelas etapas de detecção, segmentação e estimação de pose, garantindo que a análise seja concentrada exclusivamente na região da mão.

A partir dessas imagens, o sistema analisa padrões visuais globais associados à configuração e ao movimento da mão, atribuindo um rótulo correspondente ao exercício realizado. Essa abordagem permite reconhecer o exercício mesmo diante de variações naturais de execução, comuns em indivíduos com artrite, como limitações de amplitude ou diferenças no ritmo do movimento.

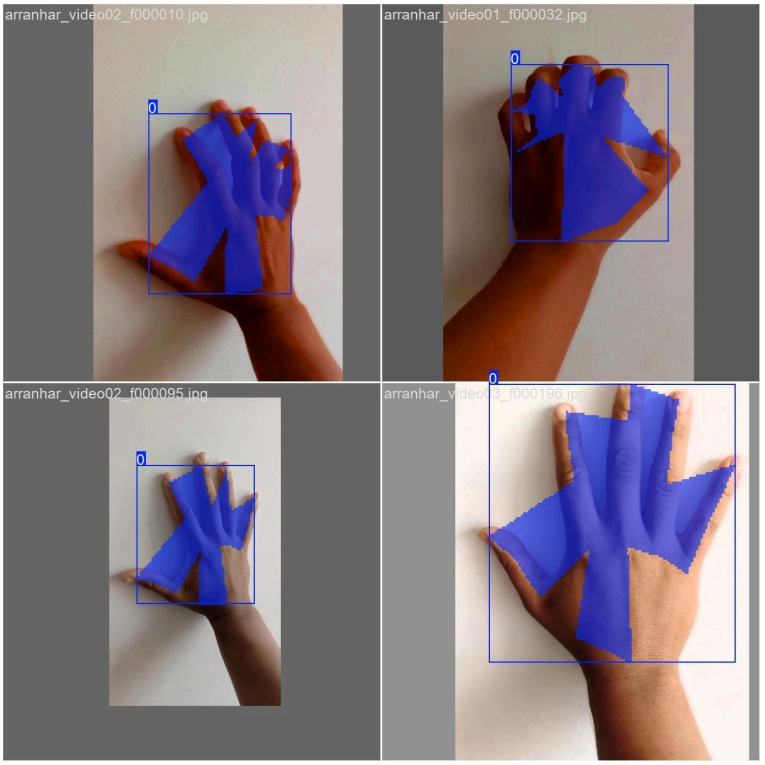
O resultado da classificação é integrado ao fluxo do sistema e encaminhado à aplicação móvel, que apresenta ao usuário o nome do exercício reconhecido e um texto explicativo sobre seus benefícios terapêuticos. Dessa forma, a etapa de classificação conclui o *pipeline*

proposto neste trabalho, permitindo a identificação clara do exercício realizado e fornecendo suporte ao acompanhamento terapêutico.

3.3 Segmentação de exercícios

A etapa de segmentação foi adotada com o objetivo de isolar a região da mão em relação ao fundo da imagem, reduzindo interferências visuais e permitindo uma análise mais precisa dos movimentos executados. A Figura 4 apresenta um exemplo do resultado do experimento de segmentação utilizando a arquitetura *YOLO*, no qual é possível observar a máscara gerada para delimitar a mão durante a execução do exercício terapêutico.

Figura 4 – Experimento segmentação utilizando a arquitetura *YOLO*



Fonte: Elaborado pelo autor, 2026.

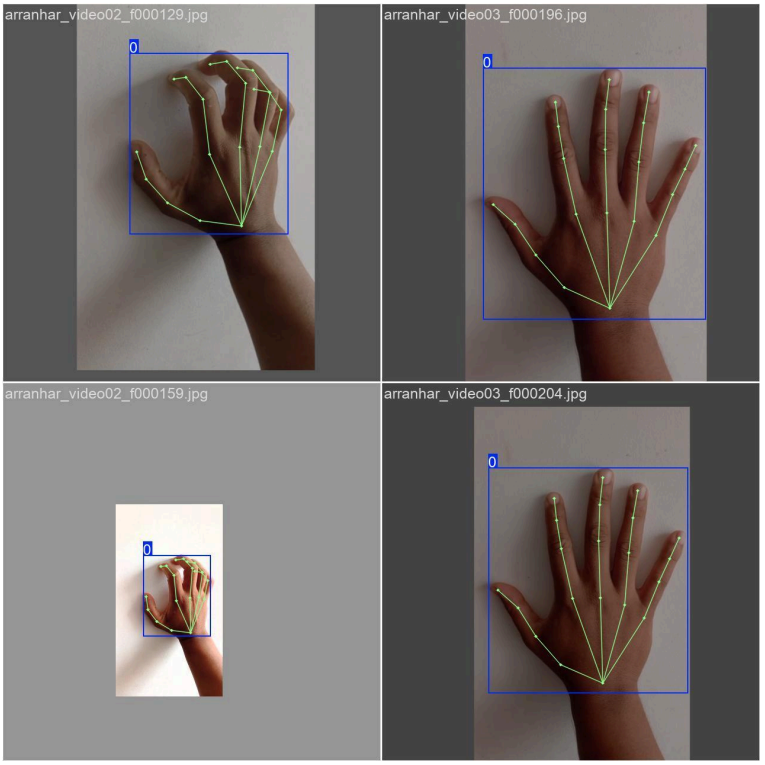
Essa abordagem torna-se especialmente relevante em ambientes não controlados, nos quais objetos externos ou variações de iluminação podem dificultar a interpretação correta do movimento. Para essa tarefa, foi utilizado um modelo de segmentação baseado na arquitetura *YOLO*, treinado para gerar máscaras que delimitam a mão em cada imagem, fornecendo uma representação mais detalhada da forma da mão quando comparada apenas ao uso de caixas delimitadoras.

Ao restringir a análise à região segmentada, o sistema torna-se mais robusto a ruídos visuais e mais sensível às variações reais do movimento terapêutico, contribuindo para a confiabilidade das etapas subsequentes do *pipeline*.

3.4 Estimação de Exercícios

A estimação dos exercícios baseia-se na análise da configuração articular da mão durante a execução dos movimentos. A Figura 5 ilustra um exemplo do experimento de estimação de pose realizado com a *YOLO*, destacando os pontos-chave identificados nas articulações dos dedos, da palma e do punho ao longo do exercício.

Figura 5 – Experimento estimação de pose realizado com a *YOLO*



Fonte: Elaborado pelo autor, 2026.

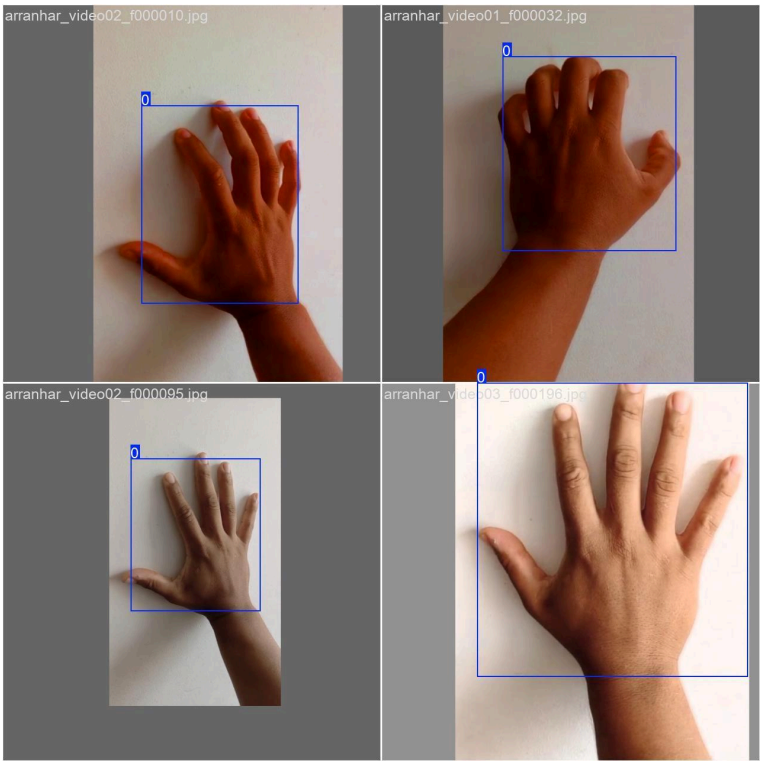
Neste trabalho, a estimação de pose considerou 21 pontos-chave da mão, fornecendo uma descrição detalhada da postura e da dinâmica do movimento. A partir desses pontos, torna-se possível diferenciar exercícios que apresentam variações sutis de posicionamento, mesmo quando executados de forma não padronizada ou com limitações articulares decorrentes da artrite.

Essa etapa contribui para uma interpretação mais flexível do movimento, permitindo que o sistema reconheça o exercício realizado mesmo diante de pequenas variações individuais na execução, o que é especialmente relevante no contexto da reabilitação.

3.5 Detecção de Exercícios

A detecção de exercícios corresponde à identificação automática da mão e de sua localização na imagem, sendo responsável por delimitar a região de interesse que será analisada pelas demais etapas do *pipeline*. A Figura 6 apresenta um exemplo do experimento de detecção realizado com a *YOLO*, no qual a mão é identificada por meio de caixas delimitadoras durante a execução do exercício.

Figura 6 – Experimento detecção realizado com a *YOLO*



Fonte: Elaborado pelo autor, 2026.

Para essa etapa, foi utilizado um modelo de detecção baseado na arquitetura *YOLO*, treinado para localizar a mão e estimar simultaneamente a posição da região detectada e a confiança associada à detecção. Dessa forma, apenas as regiões relevantes da imagem são encaminhadas para processamento nas etapas seguintes.

A detecção correta da mão é fundamental para o funcionamento do sistema, pois garante que a análise seja concentrada exclusivamente na região de interesse, aumentando a precisão e a confiabilidade dos resultados obtidos.

3.6 Implementação em Software

A implementação da solução proposta foi realizada por meio de uma arquitetura híbrida, composta por uma aplicação móvel e um servidor responsável pelo processamento dos vídeos. Essa abordagem foi adotada com o objetivo de manter a aplicação leve no dispositivo do usuário, enquanto as etapas de maior custo computacional são executadas em ambiente controlado no servidor.

A aplicação móvel foi desenvolvida utilizando o *framework Flutter*, com apoio do ambiente *Android Studio* para configuração, compilação e testes. A estrutura do aplicativo foi organizada em telas independentes, permitindo uma navegação simples e intuitiva. O fluxo de utilização inicia-se em uma tela inicial, conforme ilustrado na Figura 7, na qual o usuário é convidado a iniciar a execução dos exercícios terapêuticos.

Figura 7 – Tela inicial da aplicação móvel

Fonte: Elaborado pelo autor, 2026.

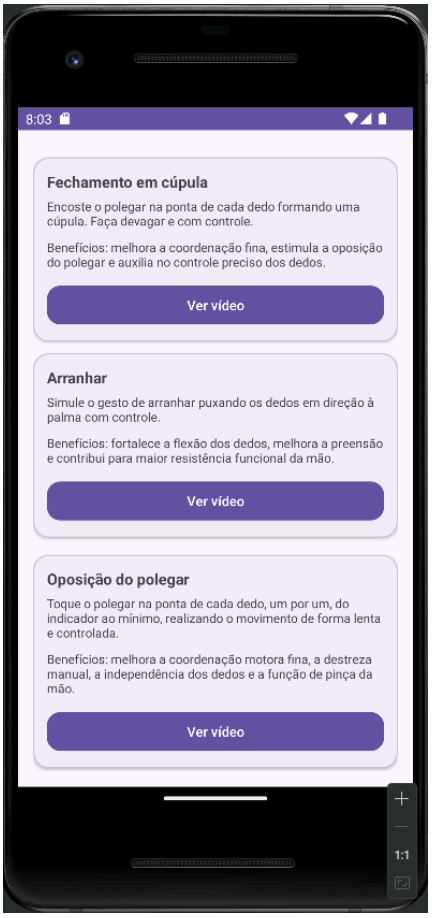
A partir dessa tela, o usuário tem acesso às principais funcionalidades do sistema, incluindo a opção de gravar um vídeo diretamente pela câmera do dispositivo ou selecionar um vídeo previamente armazenado na galeria, conforme apresentado na **Figura 8**. Essa flexibilidade permite que o sistema seja utilizado em diferentes contextos de uso, respeitando as preferências e limitações do usuário.

Figura 8 – Tela de seleção de funcionalidades

Fonte: Elaborado pelo autor, 2026.

Além das funcionalidades de captura e seleção de vídeos, a aplicação disponibiliza uma interface dedicada à visualização dos exercícios terapêuticos disponíveis, contendo descrições textuais e vídeos demonstrativos. Essa interface, apresentada na Figura 9, permite que o usuário compreenda corretamente a execução de cada exercício antes de realizá-lo, contribuindo para uma prática mais segura e eficaz.

Figura 9 – Interface de visualização dos exercícios terapêuticos



Fonte: Elaborado pelo autor, 2026.

Após a seleção ou gravação do vídeo, o arquivo é enviado ao servidor para processamento. Concluída essa etapa, a aplicação apresenta ao usuário o exercício identificado, acompanhado de um texto explicativo sobre seus benefícios terapêuticos, bem como a opção de visualizar o vídeo processado, que contém a identificação visual do exercício por meio de sua respectiva *label*. A Figura 10 ilustra a tela de apresentação do resultado final, incluindo as opções de reprodução em tela cheia e download do vídeo anotado.

Figura 10 – Apresentação do resultado do processamento

Fonte: Elaborado pelo autor, 2026.

O servidor foi implementado utilizando a linguagem *Python* e o *framework FastAPI*, sendo responsável por receber os vídeos enviados pela aplicação, realizar o processamento quadro a quadro e disponibilizar os resultados. O processamento segue o *pipeline* definido neste trabalho, envolvendo as etapas de detecção da mão, segmentação, estimação de pose e classificação do exercício.

Como saída, o servidor disponibiliza a classificação do exercício identificado, acompanhada do texto explicativo sobre seus benefícios terapêuticos. Adicionalmente, é gerado um vídeo do exercício processado, contendo a identificação visual do exercício por meio de sua respectiva *label*. Esse arquivo é armazenado em um diretório público e disponibilizado por meio de uma rota específica para visualização e *download*, sendo então retornado à aplicação móvel, concluindo o fluxo de interação do sistema.

4 MÉTRICAS DE AVALIAÇÃO

Este capítulo apresenta as métricas utilizadas para avaliar o desempenho dos modelos que compõem o *pipeline* do sistema proposto. Conforme descrito na metodologia, as etapas de detecção da mão, segmentação e estimação de pose foram treinadas individualmente utilizando apenas um exercício terapêutico, com o objetivo de analisar o comportamento dos modelos em um cenário controlado. Já a etapa de classificação dos exercícios foi treinada considerando os três exercícios terapêuticos definidos neste trabalho.

Para as etapas de detecção, segmentação e estimação, foram adotadas as métricas de Precisão, Revocação (*Recall*) e *F1-Score*, amplamente utilizadas em tarefas de visão computacional. Essas métricas são analisadas por meio de gráficos em função do limiar de confiança do modelo, permitindo avaliar a estabilidade e o equilíbrio entre confiabilidade e sensibilidade.

A etapa de classificação, por sua vez, foi avaliada por meio da matriz de confusão, que permite analisar a capacidade do modelo em distinguir corretamente entre os diferentes exercícios terapêuticos considerados.

4.1 Precisão

A precisão mede a proporção de predições corretas em relação ao total de predições realizadas pelo modelo. No contexto das tarefas de detecção, segmentação e estimação, essa métrica indica o quanto as predições aceitas pelo modelo correspondem corretamente à região da mão, aos pontos-chave ou à localização do objeto de interesse.

Matematicamente, a precisão é definida como a Equação (9):

$$\widetilde{Precisão} = \frac{VP}{VP+FP} \quad (9)$$

em que VP representa os verdadeiros positivos e FP os falsos positivos.

Valores elevados de precisão indicam que o modelo apresenta baixa incidência de predições incorretas.

4.2 Revocação (Recall)

A revocação avalia a capacidade do modelo em identificar corretamente todas as ocorrências reais do objeto de interesse presentes na imagem. Essa métrica é

particularmente relevante em aplicações de reabilitação, nas quais a omissão de partes da mão ou de articulações pode comprometer a análise do movimento.

A revocação é definida conforme a Equação (10):

$$Recall = \frac{VP}{VP+FN} \quad (10)$$

em que FNrepresenta os falsos negativos.

Valores elevados de revocação indicam maior sensibilidade do modelo na identificação das estruturas relevantes.

4.3 F1-Score

O F1-Score combina as métricas de precisão e revocação em um único indicador, permitindo uma avaliação equilibrada do desempenho do modelo. Essa métrica é especialmente útil quando se deseja analisar simultaneamente a confiabilidade e a sensibilidade das predições.

O *F1-Score* é definido conforme a Equação (11):

$$F1 - Score = 2 \cdot \frac{\widetilde{Precisao} \cdot Recall}{\widetilde{Precisao} + Recall} \quad (11)$$

Valores elevados de *F1-Score* indicam bom equilíbrio entre precisão e revocação.

4.4 Matriz de Confusão

A matriz de confusão foi utilizada para avaliar o desempenho da etapa de classificação dos exercícios terapêuticos, treinada com os três exercícios considerados neste estudo. Essa métrica permite analisar, de forma detalhada, a relação entre as classes reais e as classes previstas pelo modelo.

Cada linha da matriz representa a classe real do exercício, enquanto cada coluna corresponde à classe prevista. Os valores na diagonal principal indicam acertos de classificação, ao passo que os valores fora da diagonal representam confusões entre exercícios distintos.

A análise da matriz de confusão possibilita identificar padrões de erro e avaliar a capacidade do modelo em distinguir movimentos semelhantes, aspecto essencial no contexto da reabilitação da mão.

5 RESULTADOS

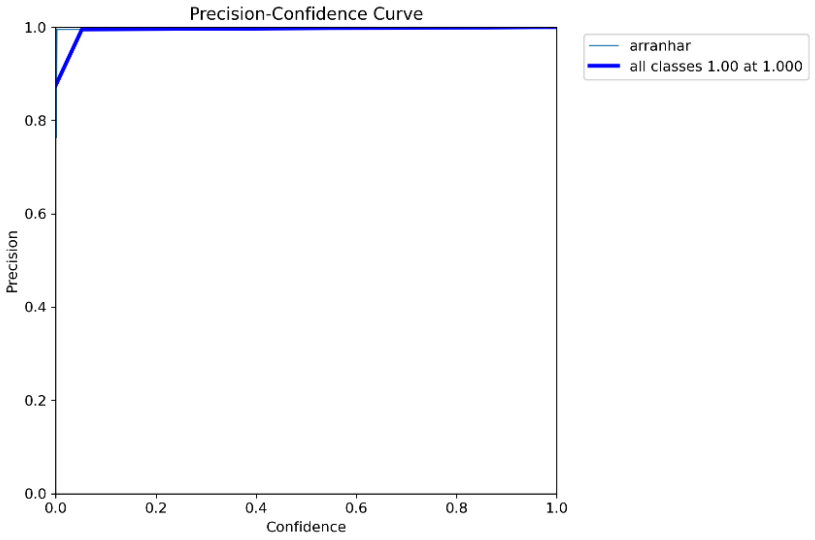
Este capítulo apresenta os resultados obtidos a partir da avaliação dos modelos que compõem o *pipeline* do sistema proposto. A análise contempla as etapas de segmentação da mão, estimação dos exercícios e detecção da mão, considerando os gráficos das métricas de Precisão, Revocação (*Recall*) e *F1-Score* em função do limiar de confiança. Conforme definido na metodologia, essas três etapas foram treinadas individualmente utilizando apenas o exercício terapêutico arranhar.

Na sequência, são apresentados os resultados da classificação dos exercícios, cuja avaliação foi realizada por meio da matriz de confusão, considerando o treinamento com os três exercícios terapêuticos definidos neste trabalho.

5.1 Resultados da Segmentação

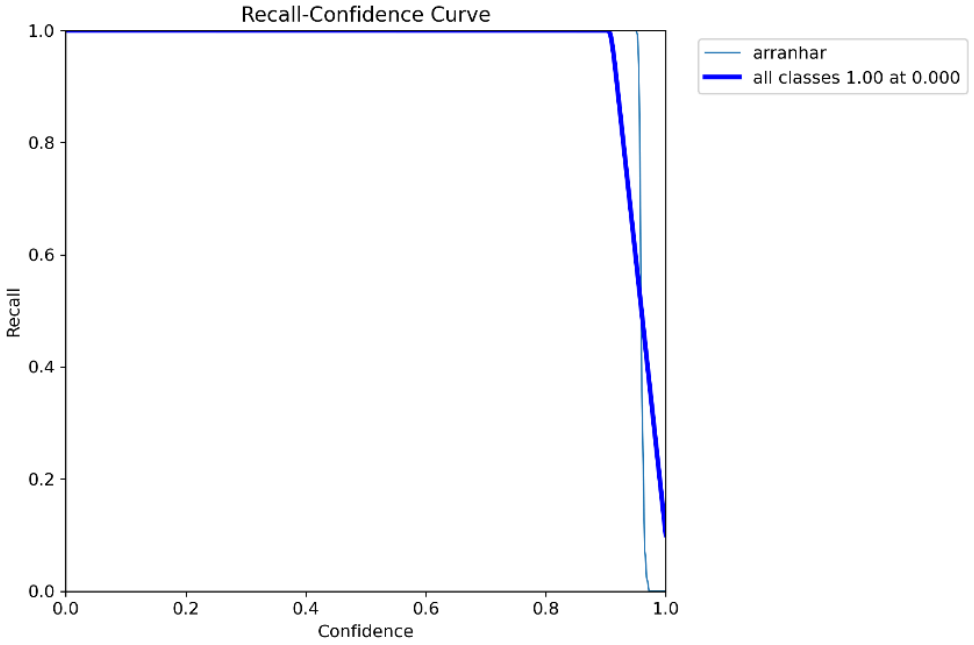
Os resultados da etapa de segmentação são apresentados por meio das curvas Precisão–Confiança, Revocação–Confiança e F1–Confiança, ilustradas nas Figuras 11, 12 e 13, respectivamente. Essas curvas permitem analisar o comportamento do modelo de segmentação da mão em diferentes níveis de confiança.

Figura 11 – Curva da Precisão em função da confiança



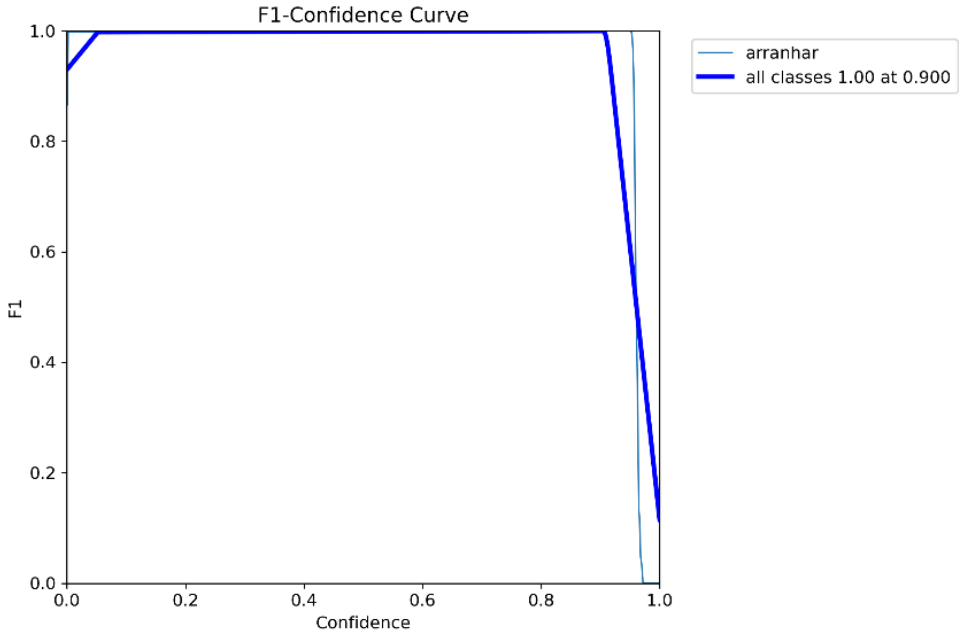
Fonte: Elaborado pelo autor, 2026.

Figura 12 – Curva da Revocação em função da confiança



Fonte: Elaborado pelo autor, 2026.

Figura 13 – Curva F1 × Confiança



Fonte: Elaborado pelo autor, 2026.

A curva Precisão–Confiança indica que o modelo alcança valores elevados de precisão desde limiares baixos, mantendo-se estável ao longo de praticamente toda

a faixa analisada. Esse comportamento evidencia que as regiões segmentadas correspondem, em sua maioria, à área correta da mão, com baixa incidência de segmentações incorretas.

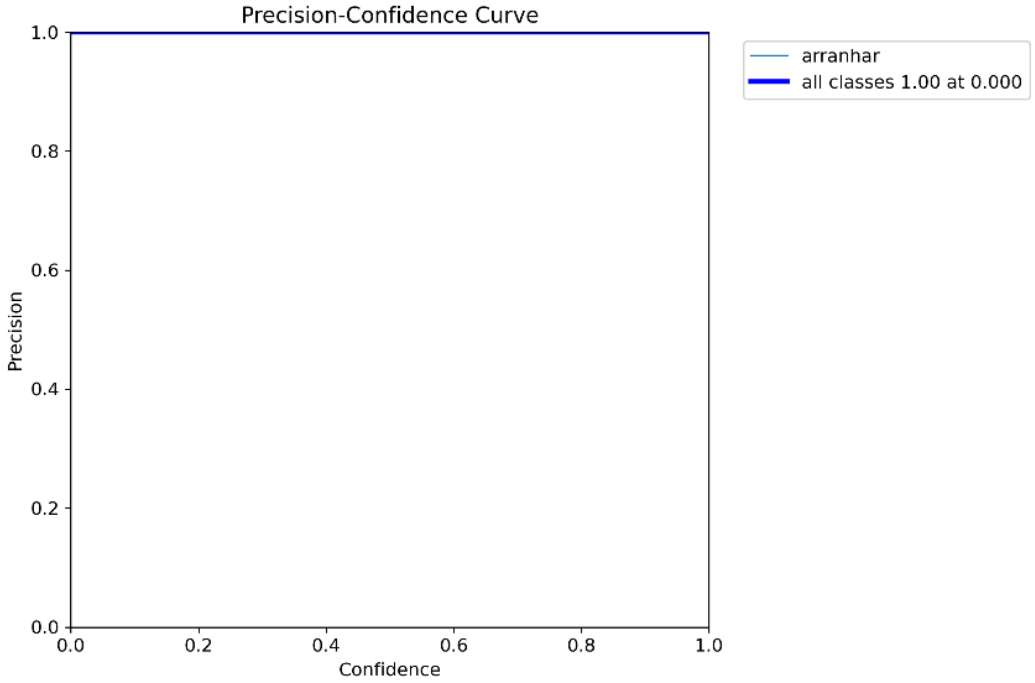
A curva Revocação–Confiança demonstra que o modelo apresenta elevada capacidade de identificar a região da mão em uma ampla faixa de limiares, com redução mais acentuada apenas em valores muito elevados de confiança. Esse comportamento reflete o aumento do rigor na aceitação das predições, resultando na exclusão de algumas segmentações válidas.

A curva F1–Confiança sintetiza o equilíbrio entre precisão e revocação, apresentando valores elevados ao longo da maior parte do intervalo analisado. O ponto de máximo *F1-Score* ocorre em um limiar intermediário-alto de confiança, indicando desempenho consistente do modelo de segmentação.

5.2 Resultados da Estimação dos Exercícios

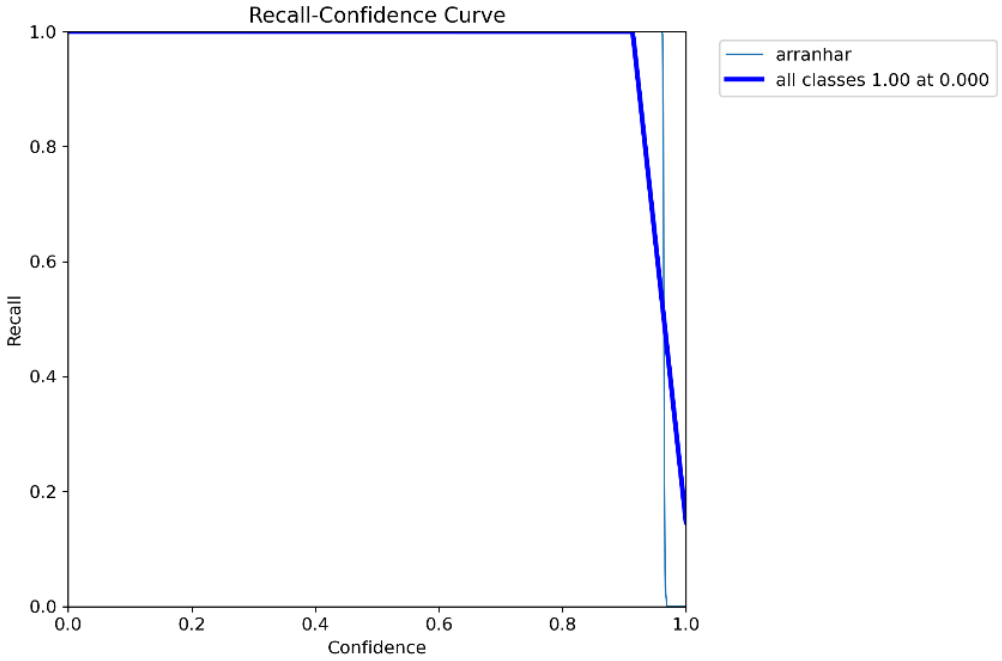
A avaliação da etapa de estimação dos exercícios foi realizada por meio das curvas Precisão–Confiança, Revocação–Confiança e F1–Confiança, apresentadas nas Figuras 14, 15 e 16. Essa análise considera a capacidade do modelo em identificar corretamente os pontos-chave da mão durante a execução do exercício arranhar.

Figura 14 – Curva da Precisão em função da confiança



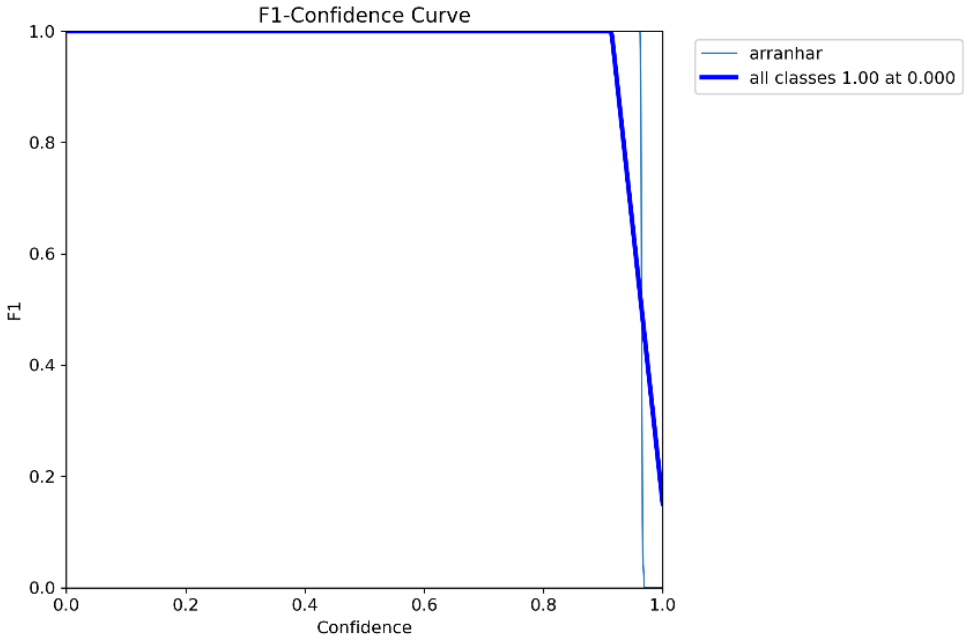
Fonte: Elaborado pelo autor, 2026.

Figura 15 – Curva da Revocação em função da confiança



Fonte: Elaborado pelo autor, 2026.

Figura 16– Curva da Medida F1 em função da confiança



Fonte: Elaborado pelo autor, 2026.

A curva Precisão–Confiança mostra que o modelo mantém valores elevados de precisão ao longo de praticamente toda a faixa de confiança, indicando que os pontos-chave estimados correspondem, em sua maioria, às articulações reais da mão.

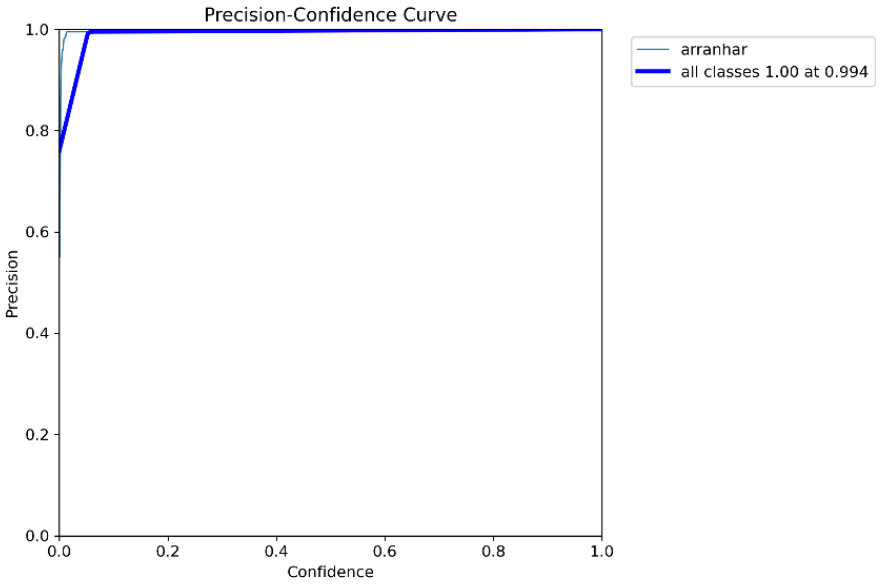
A curva Revocação–Confiança evidencia elevada sensibilidade do modelo na identificação dos pontos-chave, com redução mais significativa apenas em limiares muito elevados de confiança. Esse comportamento é esperado em modelos de estimação de pose, uma vez que limiares mais restritivos tendem a descartar predições com menor grau de certeza.

A curva F1–Confiança confirma o bom equilíbrio entre precisão e revocação, apresentando valores elevados ao longo da maior parte do intervalo analisado. Esses resultados indicam que o modelo de estimação fornece uma representação estável e consistente da configuração articular da mão.

5.3 Resultados da Detecção

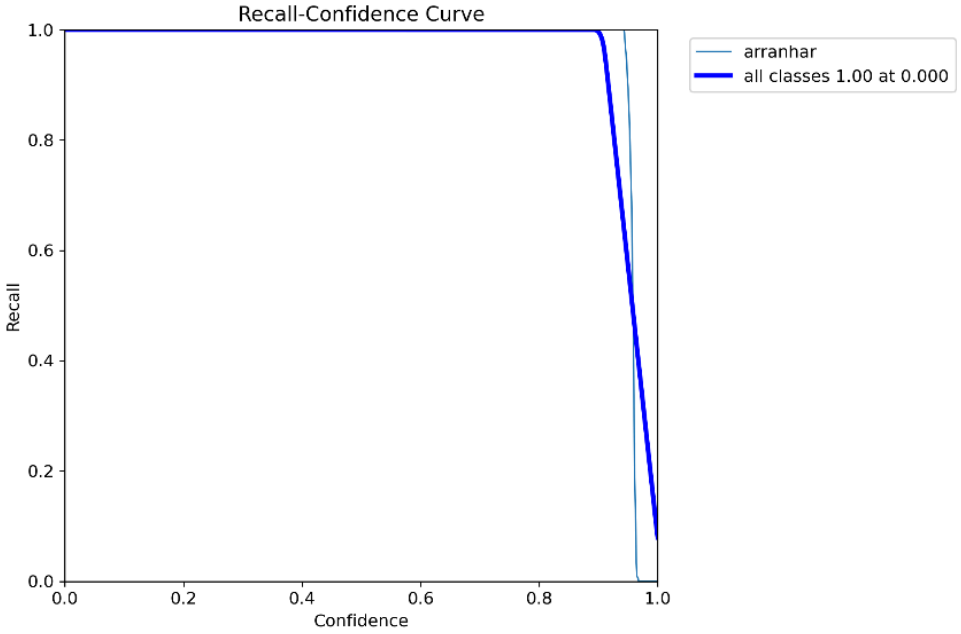
Os resultados da etapa de detecção da mão são apresentados por meio das curvas Precisão–Confiança, Revocação–Confiança e F1–Confiança, ilustradas nas Figuras 17, 18 e 19. Essas curvas permitem avaliar o desempenho do modelo na localização da mão em diferentes níveis de confiança.

Figura 17 – Curva da Precisão em função da confiança



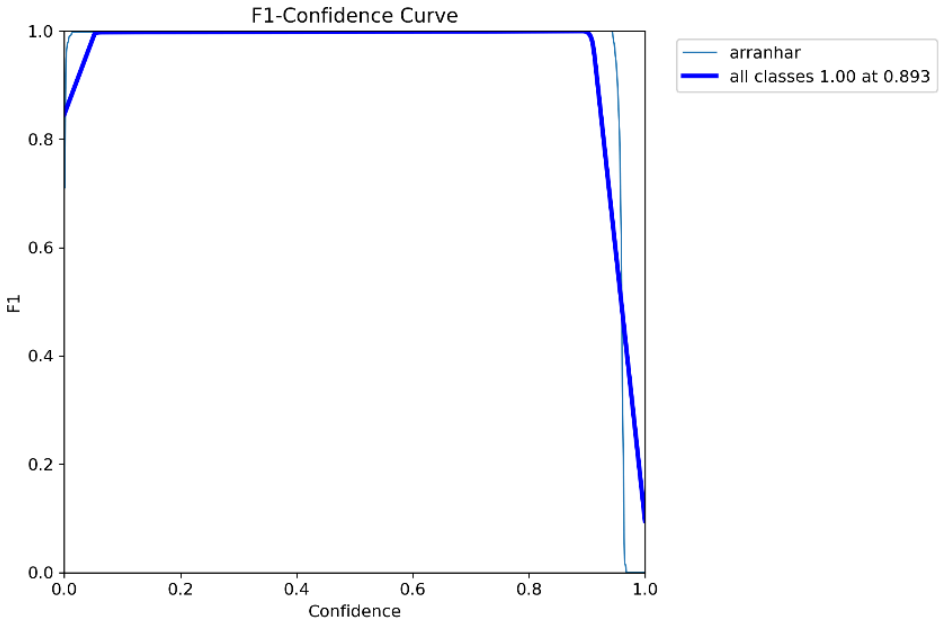
Fonte: Elaborado pelo autor, 2026.

Figura 18 – Curva da Revocação em função da confiança



Fonte: Elaborado pelo autor, 2026.

Figura 19 – Curva da Medida F1 em função da confiança



Fonte: Elaborado pelo autor, 2026.

A curva Precisão–Confiança indica que o modelo atinge valores elevados de precisão desde limiares baixos, mantendo-se próxima do valor máximo ao longo de

toda a faixa analisada. Esse comportamento sugere que as caixas delimitadoras previstas correspondem corretamente à localização da mão, com baixa ocorrência de falsas detecções.

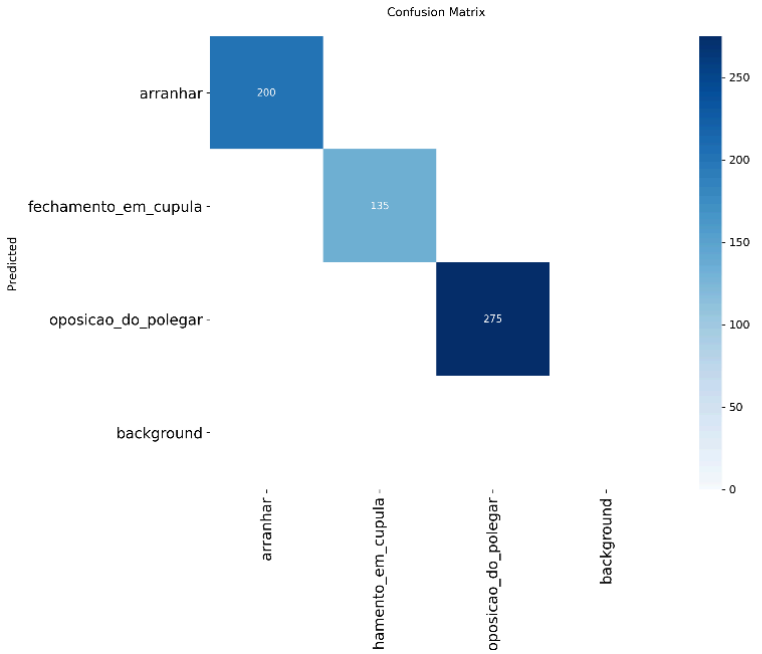
A curva Revocação–Confiança demonstra que o modelo apresenta elevada capacidade de identificar a presença da mão em uma ampla faixa de confiança, com queda mais acentuada apenas em valores muito elevados, em função do aumento do rigor na aceitação das detecções.

A curva F1–Confiança sintetiza o desempenho da etapa de detecção, apresentando valores elevados ao longo da maior parte do intervalo analisado. O ponto de máximo *F1-Score* ocorre em um limiar intermediário-alto, indicando equilíbrio entre confiabilidade e sensibilidade na detecção da mão.

5.4 Resultados da Classificação dos Exercícios

A avaliação da etapa de classificação dos exercícios foi realizada por meio da matriz de confusão, apresentada na Figura 20, considerando os três exercícios terapêuticos definidos neste trabalho: arranhar, fechamento em cúpula e oposição do polegar. Diferentemente das etapas anteriores, o modelo de classificação foi treinado de forma conjunta com todas as classes, conforme descrito na metodologia.

Figura 20 – Matriz de confusão da classificação



Fonte: Elaborado pelo autor, 2026.

A matriz de confusão evidencia uma forte concentração de valores na diagonal principal, indicando elevado número de acertos para todas as classes avaliadas. O exercício arranhar apresentou 200 classificações corretas, enquanto fechamento em cúpula e oposição do polegar registraram 135 e 275 acertos, respectivamente. A ausência de valores relevantes fora da diagonal principal indica que o modelo não apresentou confusões significativas entre os exercícios.

Esse comportamento sugere que o classificador foi capaz de discriminar adequadamente os padrões visuais e cinemáticos associados a cada exercício, mesmo considerando similaridades naturais entre movimentos terapêuticos da mão. Além disso, não foram observadas classificações incorretas relevantes associadas à classe *background*, o que indica robustez do modelo frente a ruídos ou frames sem execução válida do exercício.

Os resultados confirmam que a estratégia adotada para a etapa de classificação é eficaz e coerente com o *pipeline* proposto, complementando adequadamente as etapas de detecção, segmentação e estimação, e permitindo a identificação confiável do exercício executado pelo usuário.

6 CONCLUSÃO

Este trabalho apresentou o desenvolvimento de uma aplicação móvel voltada ao acompanhamento de idosos com artrite, utilizando técnicas de *deep learning* para a análise automática de exercícios terapêuticos da mão. A proposta surgiu da necessidade de oferecer suporte tecnológico acessível ao processo de reabilitação, considerando as limitações motoras, a variabilidade na execução dos movimentos e a importância do acompanhamento contínuo nesse público.

Ao longo do estudo, foi possível demonstrar que a utilização da arquitetura *YOLOv11*, explorando de forma modular as tarefas de detecção, segmentação, estimação de pose e classificação, mostrou-se adequada para o reconhecimento dos exercícios analisados. A separação das etapas permitiu maior controle experimental e facilitou a análise individual do desempenho de cada componente do sistema, contribuindo para a robustez da solução proposta.

Os resultados obtidos indicaram desempenho satisfatório nas tarefas de detecção, segmentação e estimação de pose, mesmo quando treinadas com um conjunto restrito de dados e focadas em um único exercício. No caso da classificação, treinada com três exercícios terapêuticos distintos, a matriz de confusão evidenciou uma boa capacidade de distinção entre as classes, reforçando o potencial do modelo para aplicação prática em cenários reais de reabilitação. Esses resultados demonstram que a abordagem adotada é promissora, mesmo diante das limitações naturais de dados e variabilidade inerentes ao contexto estudado.

A implementação do sistema em uma arquitetura híbrida, composta por uma aplicação móvel desenvolvida em *Flutter* e um servidor responsável pelo processamento dos vídeos, permitiu manter a aplicação leve no dispositivo do usuário, ao mesmo tempo em que garantiu flexibilidade e escalabilidade para o processamento computacional. A interface desenvolvida buscou priorizar simplicidade e clareza, aspectos fundamentais para o público-alvo, além de fornecer informações relevantes sobre os benefícios terapêuticos dos exercícios identificados.

Como limitações do trabalho, destaca-se o uso de um conjunto de dados reduzido e a avaliação restrita a um número limitado de exercícios. Ainda assim, essas restrições foram consideradas de forma consciente e alinhadas ao caráter

experimental da pesquisa, servindo como base para validação da proposta e para direcionar investigações futuras.

Como trabalhos futuros, sugere-se a ampliação do conjunto de dados, a inclusão de novos exercícios terapêuticos, a avaliação do sistema com usuários reais em ambiente clínico ou domiciliar, bem como a incorporação de métricas temporais que permitam analisar a qualidade da execução ao longo do tempo. Adicionalmente, a integração de mecanismos de feedback adaptativo pode contribuir para tornar o sistema ainda mais eficiente como ferramenta de apoio à reabilitação funcional do idoso.

Por fim, conclui-se que o sistema desenvolvido atende aos objetivos propostos, demonstrando que o uso de técnicas de *deep learning* aplicadas à visão computacional pode contribuir de forma significativa para o acompanhamento terapêutico de idosos com artrite, oferecendo uma alternativa tecnológica promissora, acessível e alinhada às demandas atuais da área da saúde digital.

REFERÊNCIAS

BARBOSA, Aline Lopes; FERREIRA, Marcos Coutinho; SILVA, Rafael Pereira. Efeitos de exercícios terapêuticos para mãos em idosos com artrite: uma revisão sistemática. *Revista Brasileira de Geriatria e Gerontologia*, Rio de Janeiro, 2021.

CAO, Zhe; SIMON, Tomas; WEI, Shih-En; SHEIKH, Yaser. Realtime multi-person 2D pose estimation using part affinity fields. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017.

CHEN, Xinyu; LI, Yong; ZHANG, Hao. Hand gesture recognition using convolutional neural networks for rehabilitation applications. *IEEE Access*, 2020.

DEBERT, Guita Grin. *A reinvenção da velhice: socialização e processos de reprivatização do envelhecimento*. São Paulo: Universidade de São Paulo, 2019.

DUARTE, Yeda Aparecida de Oliveira; ANDRADE, Cibele Lira; LEBRÃO, Maria Lúcia. O impacto das doenças crônicas na funcionalidade do idoso. *Revista de Saúde Pública*, São Paulo, 2020.

GOODFELLOW, Ian; BENGIO, Yoshua; COURVILLE, Aaron. *Deep learning*. Cambridge: MIT Press, 2016.

HAYKIN, Simon. *Neural networks and learning machines*. Upper Saddle River: Pearson, 2009.

HUANG, Yong; LIU, Xiaoyu; WANG, Jian. Upper limb exercise classification using pose estimation for elderly rehabilitation. *Sensors*, 2022.

KINGMA, Diederik P.; BA, Jimmy. Adam: a method for stochastic optimization. 2015.

KRIZHEVSKY, Alex; SUTSKEVER, Ilya; HINTON, Geoffrey E. ImageNet classification with deep convolutional neural networks. In: *Advances in Neural Information Processing Systems*, 2012.

LECUN, Yann; BENGIO, Yoshua; HINTON, Geoffrey. Deep learning. *Nature*, v. 521, p. 436–444, 2015.

LI, Jun; ZHAO, Yu. Therapeutic movement recognition based on keypoint detection and deep learning. *Pattern Recognition Letters*, 2021.

MENDES, Lucas Rodrigues. Tecnologias digitais como apoio à reabilitação funcional do idoso. 2022. Trabalho de Conclusão de Curso – Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2022.

NIELSEN, Michael A. *Neural networks and deep learning*. San Francisco: Determination Press, 2015.

OLIVEIRA, Andréa Silva; LIMA, Ricardo Mendes. Aspectos psicossociais da limitação funcional em idosos com artrite. *Ciência & Saúde Coletiva*, Rio de Janeiro, 2020.

RAWAT, Waseem; WANG, Zenghui. Deep convolutional neural networks for image classification: a comprehensive review. *Neural Computation*, 2017.

REDMON, Joseph; DIVVALA, Santosh; GIRSHICK, Ross; FARHADI, Ali.
You only look once: unified, real-time object detection. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016.

REDMON, Joseph; FARHADI, Ali. YOLO9000: better, faster, stronger. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017.

REDMON, Joseph; FARHADI, Ali. YOLOv3: an incremental improvement. 2018.

RUMELHART, David E.; HINTON, Geoffrey E.; WILLIAMS, Ronald J. Learning representations by back-propagating errors. *Nature*, 1986.

SANTOS, Fernando Alves; MELO, Rodrigo Silva. Tecnologias digitais aplicadas à reabilitação da artrite em idosos. São Paulo: Pontifícia Universidade Católica de São Paulo, 2021.

WANG, Chien-Yao et al. YOLOv7: trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. 2022.

WANG, Hao; LI, Xin; CHEN, Yu. Hand gesture recognition for elderly self-care assistance using deep learning. *IEEE Access*, 2023.