

COMPARAÇÃO DE TÉCNICAS DE APRENDIZADO DE MÁQUINA PARA ANÁLISE DE MOVIMENTOS BIOMECÂNICOS EM EXERCÍCIOS FÍSICOS

Patrick Burin Rodriguez¹, Marlon de Matos de Oliveira²

Resumo: Este trabalho investiga alternativas para a avaliação de movimentos biomecânicos por meio da visão computacional e do aprendizado de máquina, utilizando três exercícios: agachamento, flexão de joelhos e extensão de quadril. Foram utilizados três modelos com diferentes abordagens de pré-processamento e aprendizado de máquina, visando a avaliação da capacidade dos modelos em identificar o tipo de exercício e validar sua execução com diferentes divisões de treinamento e teste. Como diferencial, este estudo propõe a combinação das abordagens MediaPipe e LSTM, escolhidas por suas características complementares, que se alinham aos requisitos da análise de exercícios físicos, com o objetivo de desenvolver um protótipo funcional capaz de identificar e validar a execução dos exercícios com maior precisão. Os três modelos foram implementados conforme o esperado, atendendo aos critérios de viabilidade e apresentando desempenho adequado com base na acurácia, o que permitiu a comparação entre as abordagens utilizadas e a continuidade no desenvolvimento do protótipo.

Palavras-chave: Visão Computacional; MediaPipe; OpenPose; LSTM; Biomecânica; Aprendizado de Máquina; Inteligência Artificial.

¹Curso de Ciência da Computação, Universidade do Extremo Sul Catarinense (Unesc), patrickburin17@gmail.com

²Curso de Ciência da Computação, Universidade do Extremo Sul Catarinense (Unesc), marlon.oliveira@unesc.net

ABSTRACT: This study investigates alternative approaches for evaluating biomechanical movements through computer vision and machine learning, using three exercises: squats, knee flexion, and hip extension. Three models were used with different preprocessing and machine learning approaches, aiming to evaluate the models' ability to identify the type of exercise and validate its execution using different training and testing splits. As a distinguishing feature, this study proposes the combination of MediaPipe and LSTM approaches, chosen for their complementary characteristics, which align with the requirements of physical exercise analysis, with the goal of developing a functional prototype capable of identifying and validating exercise execution with greater precision. The three models showed satisfactory results, both in terms of implementation feasibility and the performance obtained, measured by accuracy, thus enabling the comparison between the approaches used and also the development of the prototype.

Keywords: Computer Vision; MediaPipe; OpenPose; LSTM; Biomechanics; Machine Learning; Artificial Intelligence.

1 INTRODUÇÃO

O avanço das tecnologias vem proporcionando novas possibilidades para a área da saúde, principalmente com o desenvolvimento da inteligência artificial e das técnicas de aprendizado de máquina. Com a popularização de sensores e câmeras, seus custos diminuíram, tornando possível a implementação de soluções profissionais tanto na área da saúde quanto em indivíduos que estão em processo de reabilitação, como a fisioterapia, por exemplo. Entre essas soluções, destaca-se a análise biomecânica na área de reabilitação, visto que ela busca entender os movimentos do corpo humano, levando em consideração conceitos como força, velocidade e rotação, com o objetivo de não só tratar lesões, mas também preveni-las e até melhorar o desempenho físico (Mesquita; Goes, 2023).

A realização de exercícios fisioterapêuticos é essencial para a recuperação de pacientes, e a execução incorreta dos movimentos pode retardar a reabilitação, comprometer o processo e até gerar novas lesões (Mesquita; Goes, 2023). Para auxiliar esse processo de análise do movimento, novas ferramentas têm sido cada vez mais utilizadas, como o Vicon, o Qualisys e o Microsoft Kinect, que oferecem recursos avançados na captura e análise de movimento. Apesar da sua precisão, esses sistemas possuem um custo elevado e necessitam de uma infraestrutura especializada, o que

limita sua adoção em ambientes clínicos com orçamentos mais reduzidos (Zsarnoczky-Dulhazi et al., 2023; Bernardina, 2016).

Como alternativa para os desafios relacionados à infraestrutura e ao orçamento, surgem ferramentas de código aberto que se destacam por sua acessibilidade, como OpenPose e o MediaPipe. O OpenPose é uma biblioteca que, a partir de vídeos comuns, consegue detectar as articulações humanas sem a necessidade de sensores sofisticados ou marcadores físicos (Cao et al., 2021). Já o MediaPipe é uma biblioteca desenvolvida pelo Google, que oferece estimativa de pose corporal em tempo real, tendo como foco a eficiência e compatibilidade com hardwares de menor desempenho. Ambas têm ganhado espaço na área da reabilitação, principalmente devido à maior acessibilidade e à capacidade de adaptação a diferentes contextos (Lin; Jiao; Zhao, 2023).

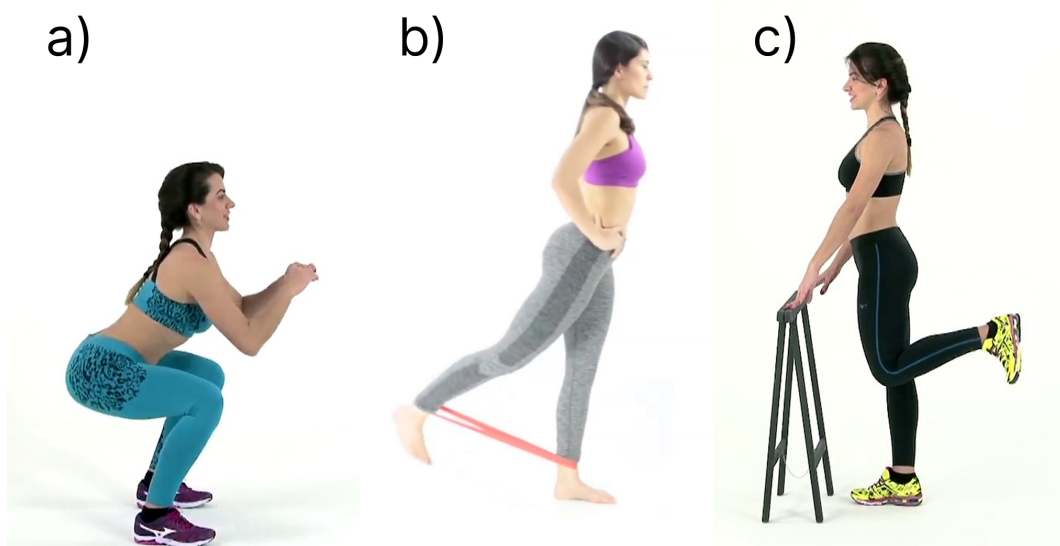
Diante da grande diversidade de ferramentas e a baixa acessibilidade a essas tecnologias, surge a necessidade de investigar alternativas mais viáveis, que mantenham bons níveis de precisão e usabilidade sem elevar significativamente os custos. A questão norteadora deste trabalho é: é possível obter uma melhora significativa na eficiência de um sistema de avaliação biomecânica baseado em inteligência artificial ao testar novas abordagens para a biblioteca de análise de movimento, a rede neural utilizada e a proporção de dados no treinamento?

Considerando a relevância do tema e a possibilidade de aprimoramento de soluções existentes, foi utilizada a base de dados desenvolvida por Francisco e Rodrigues (2022), a partir de vídeos de reabilitação física disponíveis na internet, com o objetivo de treinar e validar redes neurais para o reconhecimento e a avaliação biomecânica de três exercícios ilustrados na Figura 1: agachamento (a), extensão de quadril (b) e flexão de joelho (c). Cada exercício possui 12 vídeos gravados por uma câmera posicionada em ângulo reto em relação ao plano de movimento, favorecendo a identificação das articulações. Os vídeos foram pré-processados com a biblioteca OpenPose, que os transforma em frames (30 por segundo) e extrai 25 articulações por frame, sendo que cada ponto possui três valores: x, y e um índice de confiança *confidence score*, que indica o nível de certeza do modelo sobre a posição detectada.

O conjunto inicial de dados, denominado Base-Original, continha apenas execuções corretas. Para torná-lo mais robusto, foram aplicados ruídos artificiais nos pontos articulares, simulando variações na execução dos movimentos. Isso resultou em duas novas bases: Original-Oscillation,

que recebeu oscilações de 25% para representar variações leves que ainda são consideradas corretas, e Oscillation-Errors, com oscilações de 50% simulando execuções incorretas. As oscilações foram geradas de forma controlada, aplicando variações aleatórias nas coordenadas articulares de cada frame. A lógica adotada foi multiplicar o valor de cada coordenada por um fator aleatório, variando aproximadamente entre 0,75 e 1,25 para as oscilações leves (25%) e entre 0,5 e 1,5 para as mais severas (50%).

Figura 1 - Ilustração do três exercícios utilizados na base dados.



Fonte: Francisco e Rodrigues (2022).

Nesse contexto, este trabalho tem como objetivo geral investigar melhorias na eficiência da avaliação de movimentos biomecânicos, adotando novas abordagens tanto para o pré-processamento dos dados quanto para o treinamento de máquina, com a finalidade de aprimorar a precisão, flexibilidade e acessibilidade do sistema.

Além do objetivo geral, este artigo apresenta objetivos específicos que orientam o desenvolvimento do trabalho, como a identificação dos principais pontos na análise do movimento biomecânico, a compreensão das tecnologias de visão computacional aplicáveis ao contexto e a aplicação de técnicas de inteligência artificial e aprendizado de máquina para identificar posturas incorretas durante a execução do movimento, por meio de um protótipo funcional, permitindo assim a avaliação dos resultados obtidos com cada abordagem.

O trabalho desenvolvido por Francisco e Rodrigues (2022), publicado na IEEE Transaction on Neural Systems and Rehabilitation Engineering, aplicou técnicas de aprendizado de máquina e visão computacional para

realizar a detecção e análise de movimentos por meio da biblioteca OpenPose, utilizando a arquitetura de rede neural modular. A base de dados foi construída a partir de vídeos públicos de pessoas executando três exercícios: agachamento, flexão de joelho e extensão de quadril. A partir da extração dos ângulos articulares, esses dados foram utilizados como entrada para os modelos de rede neural, aplicando ruídos artificiais para simular movimentos incorretos. Para a avaliação dos resultados, foram utilizadas matrizes de confusão e área sob a curva (ROC), que representa a capacidade do modelo em diferenciar corretamente entre as classes. Os resultados indicaram precisão superior a 90% tanto na detecção do exercício quanto na avaliação da execução.

O estudo Latyshev et al. (2024) teve como objetivo a análise da precisão de tecnologias de visão computacional, especificamente entre as bibliotecas OpenPose e MediaPipe, voltadas à estimativa de pose humana. A pesquisa envolveu estudantes que realizaram o Teste da Cegonha, utilizado para avaliar o equilíbrio, no qual a pessoa se mantém em uma perna só, com a outra apoiada na coxa e os braços ao lado do corpo. Os resultados indicaram uma maior precisão do MediaPipe em relação ao OpenPose, apresentando menores desvios percentuais em comparação à anotação manual na maioria dos pontos-chave. Ambos os modelos demonstraram competência na identificação dos pontos-chave, porém apresentaram desvios significativos especificamente na detecção dos pontos dos pés e dos pulsos. A análise revelou que, nesses pontos, mais de 70% das detecções do OpenPose e mais de 30% das do MediaPipe estavam acima do limite de precisão aceitável. O estudo também propõe o desenvolvimento de uma aplicação para monitoramento autônomo do equilíbrio corporal, contribuindo para a prática de exercícios físicos de forma mais segura e eficiente.

O artigo Gonçalves et al. (2023) teve como objetivo o desenvolvimento de uma aplicação que utiliza visão computacional para detectar e corrigir a postura durante a execução de exercícios físicos. O estudo incluiu a criação de uma base de dados voltada à detecção de três exercícios: agachamento, flexão e abdominal, além da avaliação do desempenho do modelo YOLOv8, utilizado para detecção de objetos em tempo real. Para o desenvolvimento da base de dados, foram gravados vídeos e posteriormente extraídas imagens utilizando a biblioteca MediaPipe, totalizando cerca de trinta mil quadros. Em seguida, foi realizado o treinamento do modelo YOLOv8 para a detecção e avaliação dos exercícios. Por fim, desenvolveu-se o aplicativo com a implementação da interface do usuário e integração do

modelo treinado. O modelo foi avaliado em diferentes subversões, sendo que a versão YOLOv8x apresentou a melhor precisão, com 88,8%. O protótipo, denominado AiSweat, cumpriu seu propósito, permitindo a detecção e correção postural em tempo real por meio de uma interface intuitiva.

Nesse contexto, compreender os fundamentos da biomecânica torna-se essencial, uma vez que ela fornece as bases para interpretar corretamente os movimentos analisados por esse tipo de aplicação. Trata-se de uma área que estuda os movimentos do corpo humano com base nos princípios da mecânica. É frequentemente utilizada na fisioterapia para compreender como os movimentos são realizados, identificar padrões incorretos e prevenir lesões. A execução incorreta de exercícios físicos pode não só atrasar como até mesmo comprometer a reabilitação de pacientes, tornando essencial o acompanhamento desses movimentos (Mesquita; Goes, 2023).

A visão computacional tem ganhado cada vez mais espaço na área da saúde, principalmente por sua capacidade de automatizar a análise de imagens e auxiliar no diagnóstico com maior precisão e agilidade. Ferramentas como o OpenPose e o MediaPipe têm sido frequentemente aplicadas para análise corporal, por dispensarem o uso de sensores sofisticados ou marcadores físicos. O OpenPose é capaz de detectar até 25 articulações principais do corpo humano em tempo real, com base em vídeos gravados com câmeras comuns, como as de celulares, tornando-se uma escolha frequente em estudos de reabilitação devido à sua precisão e acessibilidade (Cao et al., 2021). Já o MediaPipe é uma biblioteca desenvolvida pelo Google, capaz de detectar 33 articulações do corpo humano a partir de vídeos RGB, em tempo real e com baixo consumo de recursos computacionais, sendo considerada uma opção viável em contextos clínicos com infraestrutura limitada (Lin; Jiao; Zhao, 2023).

O trabalho de Francisco e Rodrigues (2022) é uma das principais referências para este estudo, ao propor uma metodologia que integra visão computacional e aprendizado de máquina voltados para tratamentos fisioterapêuticos. No pré-processamento, os autores utilizaram a biblioteca *OpenPose* para obter os pontos articulares do corpo humano. Para o treinamento de máquina, foi utilizada uma variação de *Multilayer Perceptron* (MLP), na qual foram aplicadas algumas técnicas com o objetivo de aumentar a robustez do modelo treinado.

No estudo de Nguyen et al. (2024), foi desenvolvido um conjunto de dados clínicos voltado à reabilitação de pacientes com dores lombares, com o objetivo de avaliar automaticamente a execução de exercícios por

meio da análise de movimentos corporais. Os autores utilizaram uma rede do tipo LSTM para realizar tarefas de avaliação e classificação de erros nos movimentos, aproveitando sua capacidade de lidar com sequências temporais. Essa escolha se justifica pela necessidade de capturar a dinâmica dos movimentos ao longo do tempo, o que é essencial para a interpretação correta de uma execução, que não pode ser compreendida apenas por imagens isoladas.

Considerando essas contribuições, este trabalho busca explorar abordagens que combinem visão computacional e aprendizado de máquina, avaliando alternativas como o uso do MediaPipe em relação ao OpenPose, devido à sua portabilidade, eficiência em tempo real e bom desempenho mesmo em hardwares mais simples (Lin; Jiao; Zhao, 2023). Além disso, será explorada uma nova abordagem com o uso de redes do tipo LSTM, em vez da MLP, com o objetivo de capturar melhor as dependências temporais dos movimentos analisados. A partir dessas possibilidades, pretende-se contribuir com soluções mais acessíveis e eficazes para a avaliação automática de movimentos na área da reabilitação.

Para organizar essa proposta, este trabalho está estruturado em quatro seções principais. A Seção 1 apresenta a introdução ao tema, contextualizando a importância da análise biomecânica na reabilitação física, destacando as tecnologias utilizadas e justificando os objetivos do estudo. A Seção 2 descreve os materiais e métodos adotados, detalhando a base de dados utilizada, o pré-processamento aplicado com as bibliotecas OpenPose e MediaPipe, a arquitetura dos modelos de rede neural e os critérios de validação. Na Seção 3, são discutidos os resultados obtidos nos experimentos realizados com os três modelos desenvolvidos, com foco na detecção dos exercícios e na avaliação da execução dos movimentos. Por fim, a Seção 4 apresenta as conclusões do estudo, destacando as contribuições, limitações observadas e sugestões para trabalhos futuros.

2 MATERIAIS E MÉTODOS

Para a realização deste estudo, foram utilizados os mesmos vídeos da base de dados descrita na introdução, desenvolvida por Francisco e Rodrigues (2022), composta por execuções corretas de três exercícios de reabilitação física. A partir dessa base, foram aplicadas novas abordagens de pré-processamento, resultando na criação de dois modelos distintos, denominados Modelo 1 e Modelo 2. No Modelo 1, manteve-se a biblioteca OpenPose para o processamento dos vídeos. Já no Modelo 2, utilizou-se

a biblioteca MediaPipe, mantendo a taxa de 30 frames por segundo para garantir condições semelhantes de análise. Diferentemente do OpenPose, no MediaPipe, cada pessoa detectada em um frame gera um conjunto de 33 pontos corporais, com três coordenadas (x, y, z) e um atributo adicional chamado *visibility*, que indica a visibilidade e a confiabilidade da detecção.

Assim como no modelo proposto por Francisco e Rodrigues (2022), também foram aplicadas oscilações artificiais aos dados utilizados nos Modelos 1 e 2, simulando variações leves, com oscilações de 25%, e execuções incorretas, com oscilações de 50%, seguindo os mesmos critérios de intensidade e geração controlada.

Nos dois modelos, as coordenadas extraídas dos frames foram armazenadas em arquivos do tipo JSON, onde cada arquivo representava a sequência temporal de um vídeo, formando uma série de vetores numéricos contendo os dados de posição e também a distância entre as articulações. Após a obtenção dos arquivos JSON, a base de dados foi dividida em dois módulos:

- Módulo de Detecção: tem como objetivo identificar qual exercício estava sendo executado entre os três exercícios utilizados;
- Módulo de Medição: tem como objetivo identificar se a execução do exercício estava correta ou não.

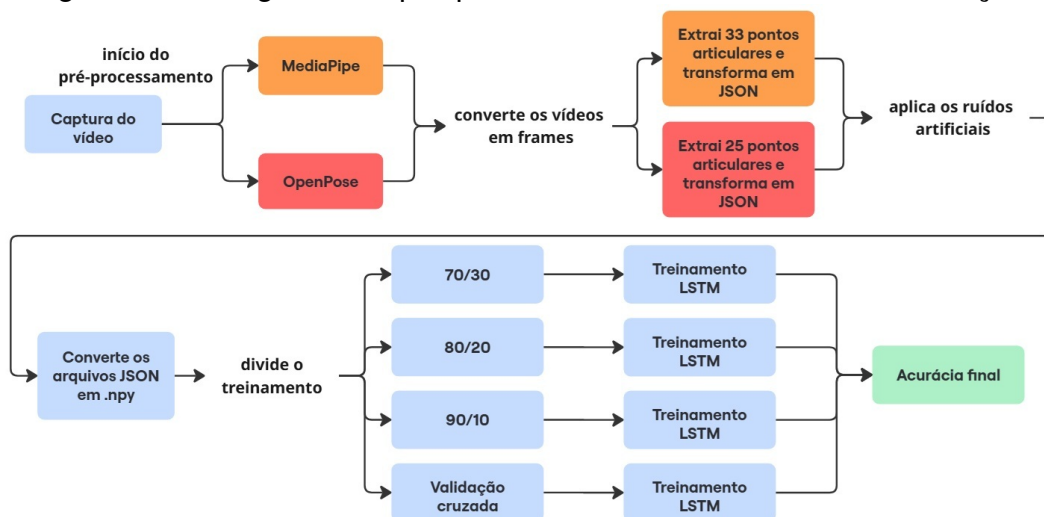
Assim como no modelo desenvolvido por Francisco e Rodrigues (2022), foram aplicados ruídos artificiais nos pontos articulares para simular diferentes variações de execução. Foi utilizada uma oscilação de 25% para representar variações pequenas, em que os movimentos ainda são considerados corretos. Já para simular execuções com erros evidentes, foi utilizada uma oscilação de 50%. As oscilações foram geradas de forma controlada, aplicando variações aleatórias nas coordenadas em cada frame. A lógica consiste em multiplicar o valor de cada coordenada por um fator aleatório. Por exemplo, para uma oscilação de 25%, cada ponto teve sua posição alterada por um fator entre aproximadamente 0,75 e 1,25.

No Modelo 2, foi introduzida uma quarta classe no módulo de detecção do exercício, representando vídeos aleatórios nos quais nenhum exercício estava sendo realizado. O objetivo dessa classe era tornar o modelo mais robusto, especialmente diante de vídeos que não pertenciam a nenhuma das três categorias iniciais. Sem essa classe adicional, o modelo tendia a classificar qualquer vídeo como sendo um dos três exercícios, mesmo quando o conteúdo não possuía relação com eles, resultando em

classificações incorretas com acurácia relativamente alta. Com a inclusão da quarta classe, o modelo passou a reconhecer situações fora do escopo dos exercícios avaliados, contribuindo para uma maior precisão e redução de falsos positivos. Além disso, adotou-se a lógica de que, quando a confiabilidade da detecção fosse inferior a 75%, a execução não era avaliada, considerando que o movimento não se enquadrava adequadamente em nenhum dos três exercícios esperados.

Ainda no pré-processamento, os arquivos JSON foram convertidos para arrays NumPy (.npy), para serem utilizados no treinamento com redes LSTM. Esses arrays possuem uma estrutura tridimensional no seguinte formato: (número de sequências, número de frames por sequência, número de atributos por frame). Para cada abordagem, foram gerados dois conjuntos principais, um para o módulo de detecção e outro para o módulo de medição, todos armazenados em arquivos no formato .npy. O processo completo descrito até aqui pode ser visualizado no fluxograma apresentado na Figura 2, que resume as etapas desde a captura dos vídeos até a avaliação final dos modelos.

Figura 2 - Fluxograma do pré-processamento, treinamento e validação.



Fonte: Elaborado pelo autor.

O treinamento foi realizado com a biblioteca Keras e o auxílio do TensorFlow. A estrutura das redes foi mantida semelhante, compostas por camadas LSTM, seguidas de Dropout para evitar *overfitting* (quando o modelo aprende demais os dados de treino e perde capacidade de generalização) e, por fim, camadas responsáveis pela classificação. No módulo de detecção, a rede era configurada para classificar entre quatro categorias

(três exercícios e a classe "outro"), enquanto no módulo de medição, a classificação era entre execução correta ou incorreta.

Para realizar a validação dos modelos, foram utilizados dois tipos de validação:

- Holdout com divisões de 70% treino e 30% teste, 80% treino e 20% teste e 90% treino e 10% teste.
- Validação cruzada (cross-validation) com 5 *fold*s, onde cada *fold* 80% dos dados foram usados para treino e 20% para teste, alternando as partições e fazendo com que todos os dados fossem utilizados no treinamento e no teste. A Figura 3 ilustra visualmente como funciona esse processo.

Figura 3 - Ilustração do funcionamento da validação cruzada.



Fonte: Elaborado pelo autor.

Essa estratégia foi utilizada para garantir uma avaliação mais robusta, especialmente devido ao baixo volume de dados para treinamento e teste.

Para validar o desempenho dos modelos, foi utilizada como métrica a acurácia, onde é representada a porcentagem de acertos do modelo em relação ao total de previsões feitas, onde, por exemplo, uma acurácia de 0.95 indica que o modelo acertou em 95% dos casos. Essa métrica foi aplicada tanto nas divisões de treino e teste do tipo *holdout* quanto na validação cruzada. Em ambos os modelos, além da acurácia na detecção do exercício, também foi analisada a confiabilidade da execução, que representa o quanto o modelo considera que o movimento está correto.

3 DISCUSSÃO E RESULTADOS

Como resultado dos experimentos realizados com o Modelo 1 (OpenPose e LSTM), o Modelo 2 (MediaPipe e LSTM) e o Modelo 3 (OpenPose e MLP), proposto por (Francisco; Rodrigues, 2022)), a análise foi

dividida em dois pontos principais: a identificação do tipo de exercício físico realizado e a validação da execução do movimento.

Também foram considerados alguns ajustes aplicados aos modelos desenvolvidos, como no caso do Modelo 2, no qual foi adicionada a classe “Outro”. Essa classe permite que a aplicação final reconheça quando um vídeo não corresponde a nenhum exercício. Com isso, caso o modelo identifique que não se trata de um exercício físico válido, o vídeo é classificado como “Outro” e não avança para a etapa de validação da execução, evitando a análise de vídeos que não correspondem aos exercícios utilizados no trabalho.

A Tabela 1 mostra os resultados de acurácia do módulo de detecção e validação do exercício do Modelo 1 e Modelo 2. Ambos os modelos utilizaram a rede neural LSTM, sendo a biblioteca de pré-processamento e a classe "outro" no modelo 2 as únicas coisas que os diferem. O Modelo 1 utilizou OpenPose e o Modelo 2 MediaPipe.

Tabela 1 - Comparação entre Modelo 1 e Modelo 2

Divisão	Modelo 1	Modelo 2
	Detecção/Validação	Detecção/Validação
70/30	0.8617 / 0.7603	0.9842 / 0.9784
80/20	0.8644 / 0.7797	0.9846 / 0.9891
90/10	0.8652 / 0.8090	0.9982 / 1.000
Cross-validation	0.8237 / 0.8339	0.9484 / 0.9796

Fonte: Elaborado pelo autor.

Na Tabela 1, observa-se que o Modelo 2 apresentou as maiores acurácias em todos os cenários, superando significativamente os resultados obtidos com o Modelo 1. A diferença de desempenho é especialmente notável na validação cruzada, em que o Modelo 2 alcançou uma média de 94,84% na detecção e 97,96% na validação, enquanto o Modelo 1 obteve apenas 82,37% e 83,39%.

Esses resultados sugerem que a abordagem utilizando o MediaPipe, em comparação ao OpenPose, proporcionou ganhos expressivos na tarefa de detecção e na validação do exercício físico. Por extrair uma quantidade maior de pontos por frame e apresentar maior estabilidade nos pontos articulares, o MediaPipe demonstra certa vantagem nesse tipo de tarefa. Outro aspecto relevante no Modelo 2 foi a inclusão da classe “Outro”, que representa situações em que a detecção não corresponde a nenhum dos três exercícios físicos avaliados. Essa classe atua como um mecanismo

de controle durante o treinamento do modelo, permitindo que a aplicação identifique vídeos que não contenham exercícios físicos pertencentes ao grupo avaliado, melhorando assim a acurácia do sistema.

Na Tabela 2 é apresentado os resultados da acurácia do Modelo 1 e Modelo 3. Ambos os modelos utilizaram OpenPose como biblioteca de pré-processamento, sendo a rede neural a única coisa que os diferem. O Modelo 1 utilizou LSTM e o Modelo 3 MLP.

Tabela 2 - Comparação entre Modelo 1 e Modelo 3

Divisão	Modelo 1	Modelo 3
	Detecção/Validação	Detecção/Validação
70/30	0.8617 / 0.7603	-
80/20	0.8644 / 0.7797	-
90/10	0.8652 / 0.8090	~0.90 / ~0.90
Cross-validation	0.8237 / 0.8339	-

Fonte: Elaborado pelo autor.

A Tabela 2 mostra um melhor desempenho do Modelo 3 na divisão 90/10, sendo a única divisão utilizada por Francisco e Rodrigues (2022). Apesar da LSTM ter suas características voltadas para a análise dos frames em sequência (Nguyen et al., 2024), a MLP obteve um desempenho superior em comparação à LSTM, tanto na detecção quanto na validação do exercício. Isso aconteceu por conta dos diversos refinamentos realizados por Francisco e Rodrigues (2022). Os aprimoramentos tiveram como foco justamente a melhoria na detecção e na validação do exercício, tendo como principais ajustes feitos a alteração no número de camadas ocultas e neurônios e a utilização de funções de ativação. Com essa série de ajustes, a MLP passou a ter um desempenho superior à LSTM, mesmo que a LSTM tenha como sua principal característica a análise dos frames como uma sequência, o que é primordial para a detecção e validação da execução do exercício.

Em todos os três modelos foram aplicados dois níveis de ruídos artificiais nos dados de entrada: ruídos de 25% para simular pequenas variações no movimento ainda consideradas corretas, e ruídos de 50% para representar desvios mais significativos, caracterizando execuções incorretas. Essa estratégia foi originalmente adotada por Francisco e Rodrigues (2022) no Modelo 3, uma vez que a base de dados continha apenas execuções corretas. Com o objetivo de realizar comparações em condições equivalentes, a mesma técnica foi aplicada nos Modelos 1 e 2 deste trabalho.

Os resultados obtidos nos três modelos demonstram que a utilização do MediaPipe proporcionou ganhos significativos tanto na detecção quanto na validação da execução dos exercícios. O MediaPipe mostrou-se mais eficiente por oferecer uma maior quantidade de pontos articulares por frame e também maior estabilidade na detecção, o que contribui diretamente para a qualidade do modelo. Esses fatores ficam evidentes quando comparados ao Modelo 1, que utilizou a mesma arquitetura de rede LSTM, porém com o OpenPose. Com isso, fica claro que a escolha da ferramenta de extração dos pontos articulares tem grande impacto na robustez do modelo.

Quando comparamos os dois modelos que usaram OpenPose, é possível observar que ambos apresentaram desempenhos distintos, mesmo utilizando a mesma ferramenta de extração de pontos. O Modelo 3 atingiu cerca de 90% de acurácia na detecção e validação do movimento na divisão 90/10, enquanto o Modelo 1 apresentou resultados significativamente inferiores, com acurácia de aproximadamente 80%. Esses dados sugerem que, apesar das vantagens citadas sobre a rede LSTM em lidar com sequências temporais, a forma como as redes modulares foram ajustadas no Modelo 3 pode ter sido mais eficiente para aquele conjunto de dados. Ainda assim, vale ressaltar que o Modelo 1 foi testado em múltiplas divisões e com validação cruzada, tornando sua avaliação mais completa em comparação ao Modelo 3. Apesar disso, algumas limitações devem ser consideradas na interpretação desses resultados.

4 CONCLUSÃO

Este trabalho contribui para a literatura ao explorar o uso de novas ferramentas aplicadas à detecção e validação de movimentos biomecânicos. As abordagens baseadas em visão computacional apresentaram bons resultados quando aplicadas à fisioterapia. Tanto a biblioteca MediaPipe quanto a rede neural LSTM, de forma conjunta ou isolada, demonstraram ser alternativas promissoras nesse campo.

Entre os modelos propostos, o que combinou MediaPipe com a LSTM obteve o melhor desempenho. Esse resultado se deve, principalmente, à superioridade do MediaPipe em comparação ao OpenPose, uma vez que o MediaPipe não apenas superou o desempenho do OpenPose, como também compensou a desvantagem da LSTM frente à MLP refinada. Apesar da alta acurácia, os resultados devem ser interpretados com cautela, visto que a base de dados era limitada. A simulação de erros com oscilações artificiais, embora válida, não representa com total fidelidade a complexidade dos

movimentos incorretos reais.

Como continuidade deste trabalho, sugere-se ampliar a base de dados com vídeos que incluam execuções incorretas e novos tipos de exercícios, visando melhorar a generalização do modelo. Também se recomenda investigar cenários com múltiplas pessoas no enquadramento, garantindo a análise apenas do usuário principal. Por fim, propõe-se o desenvolvimento de uma aplicação capaz de realizar a análise e fornecer *feedback* em tempo real, facilitando o uso em ambientes clínicos e domiciliares.

REFERÊNCIAS

BERNARDINA, G. R. D. **Utilização de câmeras de ação para análise cinemática tridimensional do movimento humano**. Universidade Federal de Viçosa, 2016. Acesso em: 10 mai. 2024. Disponível em: <<https://locus.ufv.br/server/api/core/bitstreams/c12344ca-9625-4cdc-9a44-13b31889523e/content>>.

CAO, Z. et al. **Openpose: Realtime multi-person 2d pose estimation using part affinity fields**. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2021, v. 43, n. 1, p. 172–186. Acesso em: 28 ago. 2025. Disponível em: <<https://ieeexplore.ieee.org/abstract/document/8765346>>.

FRANCISCO, J. A.; RODRIGUES, P. S. **Computer vision based on a modular neural network for automatic assessment of physical therapy rehabilitation activities**. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 2022, v. 31, p. 2174–2183. Acesso em: 14 ago. 2024. Disponível em: <<https://ieeexplore.ieee.org/abstract/document/9969647>>.

GONÇALVES, J. et al. **Aplicação para detectar e corrigir a postura em exercícios físicos**. *Revista de Sistemas e Computação - RSC*, 2023, v. 13, n. 3. Acesso em: 20 mar. 2025. Disponível em: <<https://revistas.unifacs.br/index.php/rsc/article/view/8480>>.

LATYSHEV, M. et al. **Computer vision technologies for human pose estimation in exercise: accuracy and practicality**. *Society. Integration. Education*, 2024, v. 2, p. 626–636. Acesso em: 20 mar. 2025. Disponível em: <<https://journals.ru.lv/index.php/SIE/article/view/7842/6627>>.

LIN, Y.; JIAO, X.; ZHAO, L. **Detection of 3d human posture based on improved mediapipe**. *Journal of Computer and Communications*, 2023, Scientific Research Publishing, v. 11, n. 2, p. 102–121. Acesso em: 24 mar. 2025. Disponível em: <<https://www.scirp.org/journal/paperinformation?paperid=123355>>.

MESQUITA, G. S. R.; GOES, A. L. B. **Alterações biomecânicas durante execução do Y-Balance Test e Star Excursion Balance Test em homens hígidos praticantes de esporte**. 2023. Acesso em: 29 ago. 2024. Disponível em: <<https://repositorio.bahiana.edu.br>>.

NGUYEN, S. M. et al. **A medical low-back pain physical rehabilitation dataset for human body movement analysis**. 2024. ArXiv preprint, CHRU Brest, Brest, França. ArXiv:2407.00521. Acesso em: 2 mar. 2025. Disponível em: <<https://arxiv.org/pdf/2407.00521>>.

ZSARNOCZKY-DULHAZI, F. et al. **Software de análise de movimento baseado em ia para avaliação de esporte e fisioterapia**. Revista Brasileira de Medicina do Esporte, 2023, SciELO Brasil, v. 30, p. e2022_0020. Acesso em: 18 out. 2024. Disponível em: <<https://www.scielo.br/j/rbme/a/nZdcCvptq99Qmzp7FF6ds5F/format=html&lang=pt>>.