



UFOP

Universidade Federal
de Ouro Preto

**Universidade Federal de Ouro Preto
Instituto de Ciências Exatas e Aplicadas
Departamento de Computação e Sistemas**

**Uso de redes neurais profundas para
classificação dos movimentos de flexão
plantar e dorsiflexão**

Gabriel Oliveira Silva

**TRABALHO DE
CONCLUSÃO DE CURSO**

**ORIENTAÇÃO:
Gilda A. de Assis**

**Setembro, 2025
João Monlevade—MG**

Gabriel Oliveira Silva

Uso de redes neurais profundas para classificação dos movimentos de flexão plantar e dorsiflexão

Orientador: Gilda A. de Assis

Monografia apresentada ao curso de Engenharia da Computação do Instituto de Ciências Exatas e Aplicadas, da Universidade Federal de Ouro Preto, como requisito parcial para aprovação na Disciplina “Trabalho de Conclusão de Curso II”.

Universidade Federal de Ouro Preto

João Monlevade

Setembro de 2025

SISBIN - SISTEMA DE BIBLIOTECAS E INFORMAÇÃO

S586u Silva, Gabriel Oliveira.

Uso de redes neurais profundas para classificação dos movimento de flexão plantar e dorsiflexão. [manuscrito] / Gabriel Oliveira Silva. - 2025. 43 f.: il.: color., gráf., tab..

Orientadora: Profa. Dra. Gilda Aparecida de Assis.

Monografia (Bacharelado). Universidade Federal de Ouro Preto. Instituto de Ciências Exatas e Aplicadas. Graduação em Engenharia de Computação .

1. Aprendizado de máquina. 2. Inteligência artificial - Locomoção humana. 3. Redes neurais (Computação). 4. Redes Neurais Convolucionais. I. Assis, Gilda Aparecida de. II. Universidade Federal de Ouro Preto. III. Título.

CDU 004.85

Bibliotecário(a) Responsável: Flavia Reis - CRB6-2431



FOLHA DE APROVAÇÃO

Gabriel Oliveira Silva

Uso de redes neurais profundas para classificação dos movimento de flexão plantar e dorsiflexão

Monografia apresentada ao Curso de Engenharia de Computação da Universidade Federal de Ouro Preto como requisito parcial para obtenção do título de Bacharel em Engenharia de Computação

Aprovada em 03 de setembro de 2025

Membros da banca

Dra. Gilda Aparecida de Assis - Orientadora - Universidade Federal de Ouro Preto
Dr. Alexandre Magno de Sousa - Universidade Federal de Ouro Preto
Dr. Ronan Silva Ferreira - Universidade Federal de Ouro Preto

Gilda Aparecida de Assis, orientadora do trabalho, aprovou a versão final e autorizou seu depósito na Biblioteca Digital de Trabalhos de Conclusão de Curso da UFOP em 18/09/2025



Documento assinado eletronicamente por **Gilda Aparecida de Assis, PROFESSOR DE MAGISTERIO SUPERIOR**, em 18/09/2025, às 16:13, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



A autenticidade deste documento pode ser conferida no site http://sei.ufop.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **0980920** e o código CRC **EC4E724B**.

Dedico este trabalho aos meus pais, que são a base de tudo o que sou. Pai, obrigado por cada noite de trabalho, por cada esforço silencioso e por ter me proporcionado oportunidades que o senhor mesmo não teve. Sua dedicação foi o que me trouxe até aqui. Mãe, obrigado pelo cuidado constante, pelas ligações só para saber se eu estava bem, pelo carinho de sempre e até pelos segredos guardados (mesmo quando eu não concordava), apenas para me proteger e me ver chegar onde cheguei. Sem o amor, o esforço e a confiança de vocês, nada disso seria possível.

Agradecimentos

Chegar até aqui não foi exatamente um passeio no parque, talvez mais parecido com uma longa maratona. Mas como toda boa jornada, não foi trilhada sozinho.

Agradeço, primeiramente, à minha família e em especial aos meus pais pelo apoio incondicional, carinho, incentivo e, principalmente, pela paciência ao longo de toda a minha trajetória acadêmica. Sem vocês, este capítulo da minha vida não teria sequer começado.

À minha namorada Ana Luiza, por estar sempre ao meu lado inclusive nos momentos em que nem eu queria estar comigo. Obrigado pelo apoio constante, pelo carinho nos momentos difíceis e pelas palavras certas nas horas incertas.

À minha orientadora, professora Gilda de Assis, por ser inspiração, porto seguro e, acima de tudo, humana. Sua paciência (em nível quase divino) e atenção fizeram toda a diferença neste trabalho e em cada uma das suas aulas.

Ao professor Alexandre Magno, que, como o nome sugere, deixou uma marca importante na minha formação. Sua disciplina foi desafiadora, mas essencial para minha evolução assim como suas dicas para apresentação final deste trabalho.

Aos professores do curso de Engenharia da Computação da Universidade Federal de Ouro Preto — Instituto de Ciências Exatas e Aplicadas por transmitirem seus conhecimentos técnicos e pessoais. Cada disciplina foi uma peça importante nesta jornada.

Aos amigos de caminhada, em especial ao Luiz Marques e aos membros da República BlackJack. Obrigado pelo companheirismo, pelas conversas existenciais às 3 da manhã e pelas risadas que desfrutamos, foram fundamentais nos momentos mais importantes.

E, finalmente, a todos que contribuíram, direta ou indiretamente, com este trabalho: meu sincero e caloroso agradecimento. Cada gesto, por menor que parecesse, ajudou a tornar este projeto realidade. E por isso, deixo aqui não apenas um “obrigado”, mas um “até logo”, porque a jornada, felizmente, ainda continua.

Resumo

Neste trabalho foi desenvolvido uma Rede Neural Convolucional (*do inglês, Convolutional Neural Network - CNN*) com MobileNetV2 com o objetivo de classificar movimentos de flexão plantar e dorsiflexão. Esses movimentos articulares são essenciais para locomoção humana e os mesmos são realizados predominante no tornozelo. Logo, um modelo de aprendizado de máquina (*do inglês, Machine Learning - ML*) capaz de fazer esta classificação é um avanço tecnológico para área. Foram utilizados oito vídeos de diferentes participantes no total, sendo quatro destinados ao conjunto de treinamento e quatro para confrontar com os dados adquiridos pelo trabalho anterior. Isto foi importante para garantir variabilidade e diversidade na base de dados. Para o conjunto de treinamento e teste, os vídeos tem duração total de 32 minutos e 13 segundos, enquanto o conjunto de vídeos para confrontar com dados do trabalho anterior totalizou 34 minutos e 12 segundos. A rotulação das imagens foram feitas manualmente, onde: a cada frame exibido era informado a posição correta do movimento e para este processo, um algoritmo foi implementado para extrair automaticamente frames múltiplos de 15 para reduzir redundâncias e padronizar a seleção dos dados. Os resultados demonstraram que a CNN desenvolvida alcançou uma precisão de 89%, recall 85% e o F1-score de 85%, evidenciando um bom desempenho na tarefa de classificação. Portanto esses resultados demonstram o potencial das técnicas de aprendizado profundo como alternativa ou complemento aos sensores físicos, e isso, contribui para gerar soluções mais acessíveis, dinâmicas e escaláveis.

Palavras-chaves: Rede Neural Convolucional. Aprendizado de Máquina. Redes Neurais Profundas.

Abstract

In this work, a Convolutional Neural Network (CNN) based on MobileNetV2 was developed with the objective of classifying plantar flexion and dorsiflexion movements. These joint movements are essential for human locomotion and are predominantly performed at the ankle. Therefore, a Machine Learning (ML) model capable of performing this classification represents a technological advancement in the field. A total of eight videos from different participants were used, four assigned to the training set and four to be compared with the data obtained from previous work. This procedure was important to ensure variability and diversity in the dataset. For the training and testing set, the videos had a total duration of 32 minutes and 13 seconds, while the set of videos used for comparison with previous work totaled 34 minutes and 12 seconds. The image labeling was performed manually, where for each displayed frame the correct movement position was indicated. For this process, an algorithm was implemented to automatically extract every 15th frame in order to reduce redundancies and standardize data selection. The results showed that the developed CNN achieved an accuracy of 89%, recall of 85%, and an F1-score of 85%, demonstrating good performance in the classification task. Therefore, these results highlight the potential of deep learning techniques as an alternative or complement to physical sensors, contributing to the development of more accessible, dynamic, and scalable solutions.

Key-words: Convolutional Neural Network, Machine Learning, Deep Neural Networks

Lista de Figuras

Figura 1 – Fluxograma da metodologia aplicada	13
Figura 2 – Representação das três posições do tornozelo: a)Flexão Plantar, b) Descanso e c) Dorsiflexão.	17
Figura 3 – Arquitetura de uma CNN	19
Figura 4 – Representação da matriz de confusão	20
Figura 5 – Hardware e Dispositivo Papete Inteligente	21
Figura 6 – Posição para Gravações	22
Figura 7 – Representação dos jogos Acelera e Chute ao Gol	22
Figura 8 – Ilustração dos movimentos	23
Figura 9 – Curvas de acurácia (esquerda) e perda (direita) para os conjuntos de treino e teste	30
Figura 10 – Curvas de acurácia (esquerda) e perda (direita) para os conjuntos de treino e teste após aplicação de EarlyStopping e Fine-tuning	30
Figura 11 – Matriz de confusão obtida na teste	31
Figura 12 – Ilustração de classificação incorreta dos movimentos	32

Lista de tabelas

Tabela 1	–	Resumo dos vídeos em frames e a divisão dos dados em treino e teste .	23
Tabela 2	–	Relatório de classificação final do modelo.	33
Tabela 3	–	Resultados dos vídeos por participante e jogo.	36
Tabela 4	–	Comparativo entre os dados obtidos na nova construção.	36

Lista de abreviaturas e siglas

CNN	Convolutional Neural Network - Rede Neural Convolucional.
ML	Machine Learning - Aprendizado de Máquina
RGB	Red, Green, Blue - Vermelho, Verde, Azul

Sumário

	Lista de Figuras	7
1	INTRODUÇÃO	11
1.1	O problema de pesquisa	12
1.2	Objetivos	12
1.3	Metodologia	13
1.4	Organização do trabalho	14
2	REVISÃO BIBLIOGRÁFICA	16
2.1	Introdução	16
2.2	Movimentos de Flexão e Dorsiflexão do Tornozelo	16
2.3	Aprendizado de Máquina	18
2.3.1	Redes Neurais Convolucionais	18
2.3.1.1	Arquitetura MobileNetV2	18
2.3.2	Métricas de Avaliação	19
3	DESENVOLVIMENTO	21
3.1	Ferramentas	21
3.2	Estudo Trabalho Anterior	21
3.3	Aquisição e Organização dos Dados	22
3.4	Pré-Processamento dos Dados	24
3.5	Arquitetura do Modelo	24
3.6	Estratégia de Treinamento	25
3.7	Redução de Features	27
3.8	Avaliação do Modelo	28
4	RESULTADOS	29
4.1	Desempenho Durante o Treinamento	29
4.2	Análise Matriz de Confusão	31
4.3	Relatório de Classificação	32
4.3.1	Interpretação e Possíveis Melhorias	34
4.4	Comparativo entre os dados dos vídeos e a nova construção analítica	35
5	CONCLUSÃO	38
	REFERÊNCIAS	40

1 Introdução

A identificação e análise de movimentos humanos é muito importante na área de reabilitação, esportes e robótica assistiva. Os movimentos articulares, como a flexão e a dorsiflexão do tornozelo são fundamentais para a locomoção e tem grande relevância na qualidade de vida das pessoas. Tipicamente, o estudo e classificação destes movimentos dependem de sensores físicos que podem ser caros, complexos e de pouca usabilidade. Por isso, neste trabalho, o objetivo é oferecer uma alternativa mais prática e tecnológica, como: o uso de modelos de aprendizado de máquina.

A flexão plantar ocorre quando o pé é direcionado para baixo, com a planta do pé afastando-se da canela, enquanto a dorsiflexão é o movimento contrário, no qual a parte superior do pé é elevada em direção à canela. A flexão plantar é mediada pelos músculos gastrocnêmio e sóleo, enquanto a dorsiflexão é controlada pelos músculos tibial anterior, extensor longo dos dedos e extensor longo do hálux, todos essenciais para a absorção de impacto e a flexibilidade necessária na fase de apoio da marcha. Assim sendo, os movimentos são fundamentais para atividades como caminhar, correr e saltar, pois influenciam diretamente a mecânica da marcha e a postura corporal ([CRESSWELL, 2015](#); [BERGMANN, 2018](#); [KONG, 2018](#)).

Logo, a execução e interação eficiente entre esses movimentos traz grandes impactos na mobilidade humana. Problemas na execução destes movimentos podem trazer dificuldades de locomoção, aumentar o risco de entorses e desgastes na articulação do tornozelo ([ROYER, 2015](#)). Por isso, é essencial a análise da flexão e dorsiflexão do tornozelo, para entender a dinâmica da marcha e criar programas eficientes de reabilitação.

O objetivo deste estudo é explorar a aplicação de métodos de ML para identificação dos movimentos de flexão plantar e dorsiflexão a partir de vídeos coletados no estudo anterior ([ASSIS, 2024](#)). Para substituir as papetes, adotadas em estudo anterior, este trabalho propõe o desenvolvimento de um modelo que seja capaz de classificar os movimentos com maior precisão, utilizando diretamente as sequências de imagens extraídas dos vídeos. Para este trabalho, foram avaliados dois modelos: CNN (Rede Neural Convolutacional) e MobileNetV2, sendo o último também uma CNN pré treinada destacado pela sua alta capacidade de generalização. Além disso, busca-se avaliar a acuracidade dos dados gerados e investigar como a precisão dos movimentos pode melhorar a eficácia dos jogos digitais sérios, empregados como ferramentas de reabilitação motora.

Segundo [Zhang \(2019\)](#) e [Chen \(2021\)](#) o uso de CNN's possibilita a classificação eficiente de grandes volumes de dados de movimento, superando as limitações de métodos tradicionais, como uso de sensores e equipamentos sofisticados para esse análise. Além

de permitir uma avaliação mais detalhada e precisa, difíceis de alcançar com observações manuais.

E outro benefício importante dessa abordagem é a possibilidade de integração de *feedback* em tempo real, essencial para o sucesso da reabilitação. Possibilitando ajustes imediatos nos exercícios e tratamentos personalizados a cada paciente (LIU, 2020; NAVARRO, 2018). E de acordo com, Lee (2019), a coleta contínua de dados facilita o monitoramento da recuperação. Gerando uma abordagem mais eficiente, escalável e acessível no tratamento de distúrbios motores.

Portanto, o aprendizado de máquina se apresenta como uma poderosa ferramenta para ser aplicada em conjunto com terapias de reabilitação. Isto significa, que a classificação dos movimentos de flexão e dorsoflexão com ML juntamente com reabilitação proporciona benefícios relacionados a eficiência e personalização do tratamento, além de torna-lós mais acessíveis e escaláveis no contexto de saúde digital.

1.1 O problema de pesquisa

A identificação e análise dos movimentos de flexão plantar e dorsiflexão são fundamentais para entender a biomecânica da marcha humana e com isso, desenvolver estratégias eficazes de reabilitação motora. A precisão no reconhecimento desses movimentos permite intervenções terapêuticas mais adequadas, personalizadas e com maior potencial de sucesso na recuperação. Então o problema desta pesquisa envolve estudar e entender qual o método mais adequado de aprendizado de máquina para classificar os movimentos descritos anteriormente. Logo, substituindo e dispensando sensores físicos como papetes ou marcadores corporais. A substituição desses dispositivos por modelos de aprendizado de máquina trás uma alternativa aos equipamentos físicos e oferece vantagens, como: redução de custos, maior acessibilidade e escalabilidade da solução.

Este trabalho, vem como alternativa e continuidade do estudo apresentado por Assis (2024), que trouxe a classificação do movimento de flexão e dorsiflexão usando papetes com sensores acoplados. Portanto, o problema central abordado neste trabalho consiste em: desenvolver, testar e refinar uma CNN que seja capaz de classificar movimentos a partir de frames extraídos de vídeos.

1.2 Objetivos

O objetivo principal deste trabalho é desenvolver um modelo de rede neural convolucional que seja capaz de classificar os movimentos de flexão plantar e de dorsiflexão, e assim, trazer uma alternativa ao trabalho apresentado por Assis (2024).

Este trabalho possui aos seguintes objetivos específicos:

- Implementar e documentar o fluxo metodológico completo, incluindo: etapas de preparação dos dados, tratamento de imagens, arquitetura do modelo, treinamento e teste;
- Otimizar parâmetros e estrutura para melhorar acuracidade e capacidade de generalização;
- Realizar testes com dados não vistos e vídeos adicionais, avaliando a capacidade dos modelos em identificar corretamente os movimentos em contextos diversos;
- Analisar o desempenho do modelo em termos de: tempo de processamento e viabilidade de execução em ambientes de hardware limitado;
- Comparar os resultados alcançados com o trabalho de [Assis \(2024\)](#).

1.3 Metodologia

A metodologia deste trabalho foi estruturada em duas etapas principais: (i) descrição do estudo anterior, que serviu como base de dados e comparação, e (ii) desenvolvimento do modelo de uma rede neural convolucional. A [Figura 1](#) explicita com mais detalhes o fluxo adotado para o desenvolvimento deste trabalho. Então o primeiro passo foi estudar o trabalho de ([ASSIS, 2024](#)) para entender como os dados foram gerados e a partir desses dados foi estruturado o dataset empregado nesse estudo. Esse dataset foi pré-processado, e dividido em dados de treino (80%) e dados de teste (20%). A CNN foi desenvolvida e ensinada com os dados de treinamento e por fim, a avaliação do modelo foi feita com os dados de teste.

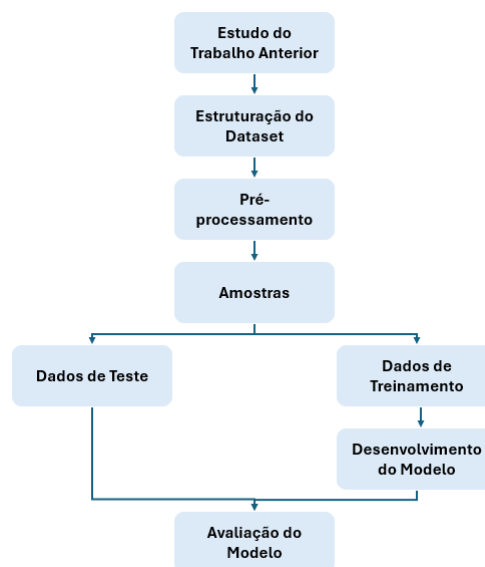


Figura 1 – Fluxograma da metodologia aplicada

Fonte: dos Autores

Desse modo, o trabalho presente é uma continuação da pesquisa intitulada: “*Jogos com tecnologia vestível como estímulo à saúde dos pés – avaliação de usabilidade*” de (ASSIS et al., 2024). Nesse estudo, foram capturados os movimentos de flexão plantar e dorsiflexão por meio de um calçado (papete) equipado com sensores. O hardware desenvolvido utilizou-se de um acelerômetro acoplado a um microcontrolador Arduino Uno. Os sinais gerados foram processados em tempo real por um algoritmo embarcado, que foi responsável por identificar o movimento realizado pelos participantes. Logo, essa abordagem possibilitou a classificação dos movimentos, fornecendo uma base de dados para análise e servindo como referência. E, também, foram gerados vídeos com a execução dos movimentos articulares realizados por diferentes participantes. Portanto, esses vídeos foram utilizados como insumo principal para o presente trabalho, visando treinar e testar modelos de aprendizado profundo.

A estrutura da base de dados adotada é composta por oito participantes distintos, onde gerou-se oito vídeos, sendo metade utilizada no treinamento e teste da CNN (32 minutos e 13 segundos) e o restante empregada para confrontar os dados capturados pelos sensores no trabalho anterior (34 minutos e 12 segundos). Essa divisão garantiu diversidade na amostra e maior robustez para a etapa de classificação.

Os vídeos foram segmentados em quadros (frames) utilizando um algoritmo que selecionava automaticamente imagens múltiplas de 15, reduzindo redundâncias e padronizando o conjunto de dados. Cada frame foi rotulado manualmente, de acordo com o movimento correspondente: flexão plantar, dorsiflexão ou repouso. E além disso, foram aplicadas técnicas como: *earlin stop* para evitar o *overfitting*, *batchNormalization* com objetivo de estabilizar o treinamento e *GlobalAveragePooling2D* para reduzir a dimensionalidade das imagens. Todas essas técnicas foram fundamentais para melhorar o desempenho da rede neural convolucional.

O modelo implementado foi uma rede neural convolucional para classificar os movimentos articulares. A arquitetura da CNN é composta por três camadas convolucionais, seguidas por camadas de *max pooling*, funções de ativação ReLU e uma camada densa final com função *softmax*. Para o *dataset* do modelo, usou-se 4 vídeos, sendo 80% separado para treino e o restante para teste. Durante o processo, foram monitoradas métricas de desempenho, como: acurácia, *recall* e *F1-score*, para avaliar a eficácia do modelo implementado.

1.4 Organização do trabalho

Para facilitar a compreensão deste trabalho, sua estrutura foi organizada da seguinte maneira: no Capítulo 2 apresenta-se a fundamentação teórica. Abordando conceitos de aprendizado de máquina, redes neurais convolucionais e sobre os movimentos do tornozelo.

Além disso, é apresentada uma análise dos trabalhos relacionados que foram usados para aprendizado e desenvolvimento deste estudo.

No Capítulo 3 é descrito todo o processo de desenvolvimento do modelo de classificação, desde a arquitetura do modelo até métricas de avaliação empregadas.

No Capítulo 4 são apresentados os resultados obtidos pelo modelo e se discute o desempenho em comparação com trabalho desenvolvido por Assis (2024).

Por fim, no Capítulo 5 aborda-se sobre a conclusão, com os principais achados, limitações do estudo e sugestões para trabalhos futuros.

2 Revisão bibliográfica

A revisão bibliográfica está organizada em três eixos principais: primeiro, a descrição dos movimentos de flexão plantar e dorsiflexão do tornozelo, fundamentais para compreender a biomecânica envolvida; em seguida, a apresentação dos conceitos de aprendizado de máquina aplicados à análise de imagens e vídeos; e, por fim, a discussão sobre redes neurais convolucionais e métricas de avaliação, que fundamentam a proposta metodológica deste trabalho.

2.1 Introdução

A revisão da literatura e a análise dos trabalhos relacionados visaram entender as bases teóricas da flexão plantar, dorsiflexão e as tecnologias de aprendizado de máquina aplicadas a esses movimentos. A pesquisa focou na biomecânica do tornozelo e pé, na análise de movimento usando aprendizado de máquina e no uso de vídeos e imagens para monitoramento de reabilitação motora. Estudos como os de ([BERGMANN, 2018](#)), ([CRESSWELL, 2015](#)), ([KONG, 2018](#)), ([ROYER, 2015](#)) e ([ASSIS et al., 2024](#)) forneceram insights valiosos sobre a mecânica dos movimentos e as técnicas de aprendizado de máquina na análise de movimento.

Adicionalmente, foram exploradas aplicações de redes neurais convolucionais e, em especial, da arquitetura MobileNetV2 para classificação e reconhecimento de padrões com ênfase no uso de vídeos, em vez de sensores, para a captura de dados. Diversas técnicas foram estudadas com o intuito de aprimorar a acurácia dos modelos, e autores como ([SANDLER, 2018](#)) e ([PEREZ; WANG, 2017](#)) ressaltam a relevância dessas abordagens para aumentar a eficácia da classificação. Nesse sentido, o uso de ML aplicado a vídeos mostra-se uma ferramenta eficaz no reconhecimento de padrões, especialmente no contexto da reabilitação, como evidenciam os estudos de ([BAUMGARTL; BUETTNER, 2021](#)) e ([HUONG et al., 2025](#)).

Portanto, neste capítulo o objetivo é expor os principais conceitos e contribuições que fundamentaram o desenvolvimento deste trabalho.

2.2 Movimentos de Flexão e Dorsiflexão do Tornozelo

Os movimentos realizados pelo tornozelo tem um grande impacto na locomoção e por consequência na qualidade de vida das pessoas. Desse modo, algum problema ou dificuldade na execução desses movimentos afetam atividades essenciais como caminhar,

correr e saltar. Os principais movimentos dessa articulação são o de flexão plantar e a dorsiflexão. Na Figura 4 tem-se a representação destes movimentos.



Figura 2 – Representação das três posições do tornozelo: a) Flexão Plantar, b) Descanso e c) Dorsiflexão.

Fonte: (BARBAN, 2025)

Segundo Bergmann (2018) e Royer (2015), a flexão plantar é caracterizada pelo movimento em que a ponta do pé se afasta da tíbia, direcionando-se para baixo. Ou seja, bem parecido como ficar na ponta dos pés (observe a Figura 4 e sua ilustração à esquerda). E este movimento é diretamente responsável a impulsão da marcha e se relaciona com o desempenho em atividades esportivas.

E como observa Cresswell (2015) e Kong (2018) a dorsiflexão corresponde ao movimento oposto. Isto é, a ponta do pé se aproxima da tíbia, direcionando-se para cima (observe a Figura 4 e sua ilustração à direita). E este movimento é fundamental para o controle da postura e para fase de apoio da marcha.

Logo, diversos estudos têm enfatizado a relevância desses movimentos para biomecânica do tornozelo e do pé, além de sua importância em processos de reabilitação. As pesquisas de Bergmann (2018) e Royer (2015) analisaram e discutiram como a flexão plantar gera força propulsora no caminhar das pessoas. Já Cresswell (2015) e Kong (2018) destacam a dorsiflexão como fundamental para estabilidade articular e o equilíbrio. Além disso, Assis et al. (2024) discute o monitoramento desses movimentos no contexto de reabilitação motora assistida por tecnologias digitais.

Então, compreender esses movimentos é essencial para o desenvolvimento de ferramentas computacionais baseadas em aprendizado de máquina voltadas à reabilitação. E isso é reforçado em trabalhos recentes que visam integrar conceitos da biomecânica com técnicas de visão computacional (BAUMGARTL; BUETTNER, 2021; HUONG et al., 2025).

2.3 Aprendizado de Máquina

O aprendizado de máquina (ML) é uma subárea da inteligência artificial que tem como objetivo modelar algoritmos capazes de aprender padrões de dados e realizar previsões ou classificações. Tudo sem depender de regras explicitamente programadas, baseando-se apenas em estatística (PEREZ; WANG, 2017).

Além disso, os modelos de ML podem ser de aprendizado supervisionado, não supervisionado e por reforço. Isso depende da forma como os dados são utilizados no aprendizado. Também existem diferentes algoritmos, tanto para regressão quanto para classificação. No contexto de análise de imagens e vídeos, as redes neurais convolucionais se destacam, pois tem grande capacidade de extrair características e proporcionar maior precisão (HARRISON, 2020).

Segundo Baumgartl e Buettner (2021), Huong et al. (2025) o ML é uma poderosa e eficiente ferramenta para reconhecimento de padrões de movimento humano. O que permite uma análise automatizada de tarefas como: marcha, flexão plantar e dorsiflexão. Substituindo a necessidade de uso de sensores físicos para captar estes movimentos.

2.3.1 Redes Neurais Convolucionais

As CNN's são arquiteturas de aprendizado profundo, foram projetadas para trabalhar com dados estruturados espacialmente, como imagens e vídeos. Nelas temos camadas convolucionais para identificar as características relevantes, como: formas, texturas e contornos. Depois são processadas em camadas que estão totalmente conectadas (SANDLER, 2018; PEREZ; WANG, 2017).

A Figura 3 ilustra a arquitetura de uma CNN, nota-se que, nas camadas convolucionais, as características da imagem são extraídas e, posteriormente, processadas nos neurônios totalmente conectados. Por fim, obtém-se o resultado da classificação.

Os trabalhos apresentados por Baumgartl e Buettner (2021), Huong et al. (2025) demonstraram que as redes neurais convolucionais oferecem resultados superiores na análise de movimento humano quando comparadas a métodos tradicionais de aprendizado de máquina, quando os dados são provenientes de imagens ou vídeos.

2.3.1.1 Arquitetura MobileNetV2

A MobileNetV2 é uma arquitetura otimizada de CNN projetada para ser leve e eficiente, permitindo implementação em dispositivos com recursos computacionais limitados sem comprometer a precisão (SANDLER, 2018).

Conforme Sandler (2018), Huong et al. (2025) essa arquitetura emprega blocos de resíduos invertidos, que primeiro expandem o número de canais, aplicam convoluções

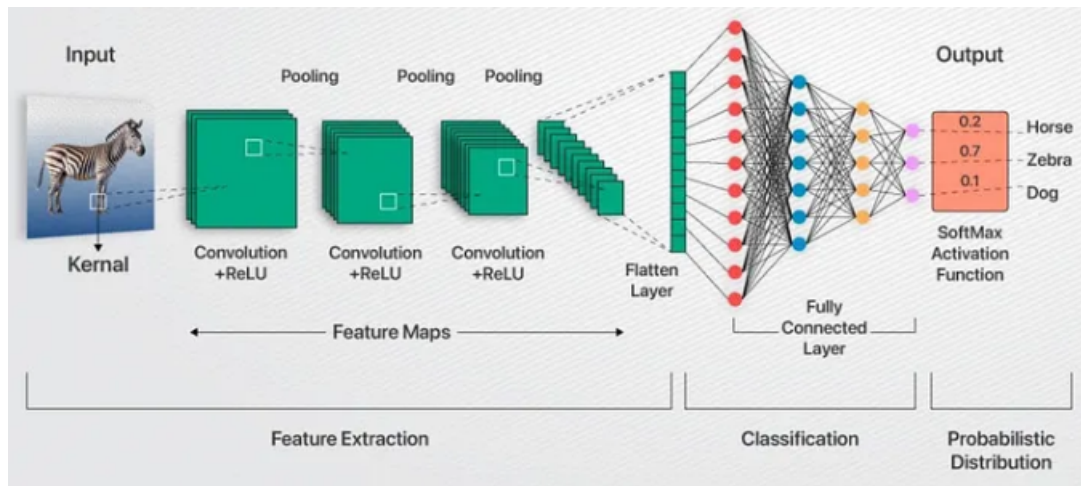


Figura 3 – Arquitetura de uma CNN

Fonte: (SINGH, 2024)

leves e depois projetam de volta para um espaço de dimensão menor, adicionando a entrada original. Toda essa combinação permite manter um alto desempenho e reduzindo a complexidade computacional, o que torna essa arquitetura adequada para dispositivos de recursos limitados.

2.3.2 Métricas de Avaliação

As métricas de avaliação de uma rede neural convolucional são essenciais para verificar a qualidade das previsões do modelo e sua capacidade de generalização.

Como descrito por Harrison (2020) as principais métricas são:

- **Matriz de Confusão:** é uma tabela que descreve o desempenho de um modelo, mostrando a relação entre as classes previstas e as classes reais. Por meio dela, é possível identificar tipos específicos de erro, como:
 - Verdadeiro Positivo (TP): instâncias positivas corretamente classificadas;
 - Falso Positivo (FP): instâncias negativas incorretamente classificadas como positivas;
 - Verdadeiro Negativo (TN): instâncias negativas corretamente classificadas;
 - Falso Negativo (FN): instâncias positivas incorretamente classificadas como negativas.
- **Precisão:** mede a proporção de instâncias classificadas como positivas que realmente pertencem à classe positiva. É uma métrica importante quando o custo de um falso positivo é alto, por exemplo: em diagnósticos médicos ou detecção de fraudes;

Matriz de confusão		Classe predita	
		Positiva	Negativa
Classe original	Positiva	VP	FN
	Negativa	FP	VN

Figura 4 – Representação da matriz de confusão

Fonte: (MARIANO, 2021)

$$Precisão = \frac{TP}{TP + FP} \quad (2.1)$$

- **Recall:** calcula a proporção de instâncias positivas corretamente identificadas pelo modelo;

$$Recall = \frac{TP}{TP + FN} \quad (2.2)$$

- **F1-score:** é a média entre precisão e recall, oferecendo uma medida equilibrada que considera tanto a capacidade do modelo de evitar falsos positivos quanto de detectar corretamente os positivos. É uma métrica muito importante, especialmente em conjuntos de dados desbalanceados (classes são menos representadas do que outras) e, portanto, mais suscetíveis a erros de classificação.

$$F1-score = 2 \times \frac{Precisão \cdot Recall}{Precisão + Recall} \quad (2.3)$$

Então o uso combinado dessas métricas permite uma avaliação completa do desempenho de uma CNN. Com todas essas métricas é possível analisar a capacidade do modelo de lidar com diferentes classes e minimizar erros.

3 Desenvolvimento

Este capítulo tem como objetivo apresentar todo o processo e métodos utilizados no desenvolvimento da CNN. Ou seja, descrever desde as ferramentas e bibliotecas utilizadas até a etapa de avaliação do modelo.

3.1 Ferramentas

Utilizou-se diferentes ferramentas de software e recursos computacionais para implementação, treinamento e a teste do modelo de aprendizado de máquina.

- Linguagem de Programação: *Python 3.10*;
- Frameworks: *TensorFlow 2.15*, *Keras*, *Scikit-learn*, *Pandas* e *Seaborn*;
- Ambiente de desenvolvimento: Google Colab com suporte a GPU;
- Armazenamento: Google Drive.

3.2 Estudo Trabalho Anterior

A aquisição dos dados utilizados neste estudo foi realizada a partir do trabalho de [Assis \(2024\)](#), que desenvolveu jogos com tecnologia vestível voltados ao estímulo da saúde dos pés. Foram selecionados dois jogos principais: o Acelera e o Chute ao Gol. Para estes experimentos participaram 10 voluntários, distribuídos em dois grupos. Para a interação, foi utilizado o dispositivo denominado Papete Inteligente, equipado com acelerômetros acoplados e conectado a um microcontrolador Arduino Uno que foi responsável pela interpretação dos sinais. A [Figura 5](#) mostra esse dispositivo.



Figura 5 – Hardware e Dispositivo Papete Inteligente

Fonte: ([ASSIS, 2024](#))

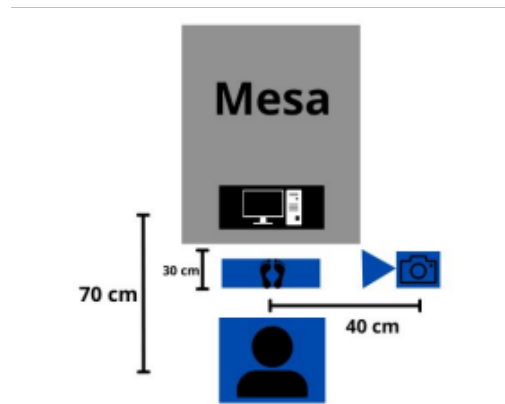


Figura 6 – Posição para Gravações

Fonte: Assis (2024)

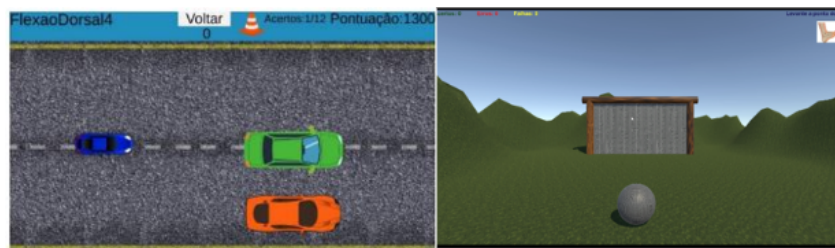


Figura 7 – Representação dos jogos Acelera e Chute ao Gol

Fonte: Assis (2024)

Durante as sessões, além dos dados captados pelos sensores, foram gravados vídeos dos participantes executando os movimentos dos pés e o posicionamento do participante e câmera de gravação é apresentada na Figura 6. Também houve os registros da tela dos jogos e a Figura 7 mostra os dois jogos, "Acelera" e "Chute ao Gol", respectivamente.

Esses vídeos formam a base de dados gerada por Assis (2024), mas para o trabalho presente não foi feito uso de todos estes dados. Foi utilizado somente os vídeos que gravaram a movimentação dos pés dos participantes, usando ao todo oito vídeos. Sendo metade deles destinados ao treinamento e teste.

3.3 Aquisição e Organização dos Dados

Os vídeos empregados neste trabalho foram produzidos por Assis (2024). E nesses vídeos há três tipos de movimentos: dorsiflexão, flexão e repouso. Observa a Figura 8 que expõem os movimentos descritos acima.



Figura 8 – Ilustração dos movimentos

Fonte: dos Autores

As imagens foram extraídas e salvas no Google Drive, com nomes padronizados e classificados visualmente Direita como Dorsiflexão, Esquerda como Flexão Plantar e Meio com repouso, a classificação de saída nos arquivos txt utilizados foram de -1, 1 e 0 respectivamente. Um arquivo texto (.txt) foi utilizado para relacionar cada imagem com sua respectiva classe. Ao todo, foram utilizados 4 vídeos para o treinamento, com a seguinte quantidade total de frames:

- V1: 10.548 frames;
- V2: 19.043 frames;
- V3: 19.020 frames;
- V4: 10.630 frames.

Destes, foi realizada a seleção de 1 a cada 15 frames, totalizando 4.228 frames classificados manualmente para o treinamento do modelo. Desse conjunto, 80% foram destinados ao treinamento e 20% à teste. A Tabela 1 resume as informações descritas anteriormente.

Tabela 1 – Resumo dos vídeos em frames e a divisão dos dados em treino e teste

Vídeos/Frames	Total	Selecionados	Para Treino (80%)	Para Teste (20%)
V1	10.548	703	562	141
V2	19.043	1.269	1.015	254
V3	19.020	1.268	1.014	254
V4	10.630	688	550	138
Total	59.241	4.228	3.141	837

Fonte: dos Autores

Além disso, após o treinamento, foram utilizados quatro vídeos adicionais, V5 (22.137 frames) e V6 (10.597 frames), V7 (10.488 frames), V8 (19.295 frames), para avaliação

do modelo. Estes vídeos foram classificados integralmente pelo modelo e utilizados para comparação com dados do trabalho apresentado por [Assis \(2024\)](#).

3.4 Pré-Processamento dos Dados

Após a organização dos dados, é necessário fazer o pré-processamento dos dados. Essa etapa é fundamental em projetos de ML, para garantir a padronização das entradas e melhorar a qualidade treinamento.

Recorreu-se a classe *ImageDataGenerator* do *TensorFlow* para aplicar o pré-processamento e aumentar a variabilidade do conjunto de dados com as seguintes transformações:

- **Normalização:** todos os pixels foram reescalados para o formato RGB (*do inglês: Red, Gren, Blue*) no intervalo de $[0, 1]$;
- **Transformações:** rotação aleatória (até 15°), zoom, brilho, shear e flip horizontal;
- **Separação:** o dataset foi dividido em 80% para treinamento e 20% para teste.

Por fim, as imagens foram redimensionadas para 224×224 pixels, em formato RGB, conforme exigido pela arquitetura MobileNetV2.

3.5 Arquitetura do Modelo

O modelo desenvolvido foi uma rede neural convolucional com MobileNetV2, pré-treinada na base ImageNet. A camada de classificação original do MobileNetV2 foi removida para permitir a adaptação da rede ao conjunto de classes específico deste trabalho. Essa escolha se justifica pelo bom desempenho da MobileNetV2 em tarefas de visão computacional, aliada à sua leveza e eficiência computacional, permitindo treinamento mais rápido mesmo em equipamentos com recursos limitados.

As seguintes camadas foram adicionadas:

- *GlobalAveragePooling2D* para reduzir a dimensionalidade das features;
- *BatchNormalization* para estabilizar o treinamento;
- *Camada Dense* com 128 unidades e ativação *ReLU*;
- Nova camada *Dense* com ativação *softmax* e 3 unidades (número de classes).

Resumindo, a camada de classificação original do modelo foi removida para permitir a adaptação da rede ao conjunto de classes específico deste trabalho. Sobre a base convolucional foram adicionadas novas camadas para aplicações específicas. A camada *GlobalAveragePooling2D* foi utilizada para reduzir a dimensionalidade das features e assim, evitar a necessidade camadas densas e diminuir o risco de *overfitting*. Em seguida, foi aplicada a camada de *BatchNormalization*, que normaliza as saídas das camadas anteriores, estabilizando o treinamento e acelerando a convergência do modelo. Posteriormente, foi adicionada uma camada densa com 128 unidades e função de ativação *ReLU*, responsável por aprender combinações não lineares das características extraídas, permitindo que o modelo capture padrões complexos presentes nas imagens. Por fim, a camada de saída é uma camada densa com três unidades e ativação *softmax*, que gera as probabilidades para cada classe.

De forma adicional, é bom lembrar que a MobileNetV2 é composta por blocos residuais com convoluções *depthwise* separáveis, que permitem reduzir o número de parâmetros da rede sem comprometer a capacidade de extração de características. Logo, o modelo foi treinado utilizando as imagens pré-processadas, aproveitando o *transfer learning* da base *ImageNet*, o que possibilita ao modelo aprender padrões visuais genéricos e adaptá-los à classificação específica dos movimentos de interesse.

3.6 Estratégia de Treinamento

O modelo foi treinado em duas fases, seguindo a abordagem de *transfer learning*, que consiste em aproveitar o conhecimento prévio de uma rede neural previamente treinada em um grande conjunto de dados (neste caso, a *ImageNet*) e adaptá-lo a uma nova tarefa específica.

Na primeira fase, a base pré-treinada do MobileNetV2 foi mantida congelada, pois isso permite que apenas as camadas adicionadas para a classificação dos movimentos de flexão plantar, dorsiflexão e repouso fossem treinadas. Esse congelamento impede que os pesos, previamente aprendidos, sejam alterados, preservando representações genéricas de características visuais, como bordas, texturas e formas, que são úteis para a nova tarefa. O otimizador Adam foi utilizado com a taxa de aprendizado padrão, garantindo uma convergência eficiente durante as primeiras iterações. Essa fase inicial é importante porque o objetivo é adaptar o modelo às novas classes sem comprometer o conhecimento prévio, acelerando o treinamento.

Na segunda fase, executa-se o *fine-tuning*, em que as últimas 20 camadas convolucionais do MobileNetV2 são descongeladas para permitir um refinamento das representações aprendidas. O *fine-tuning* consiste em ajustar os pesos de camadas mais avançadas para que o modelo se adapte melhor às particularidades do novo conjunto de dados. Uma taxa

de aprendizado reduzida (1×10^{-5}) foi utilizada para evitar alterações bruscas nos pesos e gerar ajustes finos mais precisos, melhorando a capacidade de generalização do modelo.

Em ambas as fases, a função de perda utilizada foi a *categorical crossentropy*. Para prevenir *overfitting*, foi implementado o *early stopping* com paciência de 3 épocas, interrompendo o treinamento caso não houvesse melhora na performance do conjunto de teste. Os melhores pesos obtidos durante o treinamento foram salvos, garantindo que o modelo final atingisse o desempenho ótimo.

Então os hiperparâmetros são parâmetros definidos antes do treinamento que controlam o processo de aprendizado do modelo, como taxa de aprendizado, número de épocas, tamanho do batch, entre outros. Diferente dos pesos da rede, que são aprendidos, os hiperparâmetros precisam ser ajustados para otimizar o desempenho do modelo. Resumindo, nesse trabalho foram utilizados os seguintes hiperparâmetros:

- Otimizador: Adam, utilizado para ajustar os pesos da rede de forma eficiente;
- Taxa de aprendizado: 0,001 na fase inicial e 1×10^{-5} no *fine-tuning*, controlando a velocidade de atualização dos pesos;
- Número de épocas: 10 na fase inicial e 5 no *fine-tuning*, definindo quantas vezes o modelo passa pelo conjunto de treinamento;
- Batch size: 32, indicando a quantidade de imagens processadas antes de cada atualização de pesos;
- Função de perda: Categorical Crossentropy, usada para medir o erro entre a previsão do modelo e o rótulo real;
- Camadas congeladas: base do MobileNetV2 inicialmente congelada. As últimas 20 camadas descongeladas durante o *fine-tuning*, controlando quais camadas são treináveis;
- Número de neurônios na camada Dense: 128, camada intermediária com ativação ReLU;
- Early stopping: paciência de 3 épocas, interrompendo o treinamento caso não haja melhora no teste.

Portanto, a estratégia de treinamento em duas fases combinou a eficiência do *transfer learning* com a precisão do *fine-tuning*, resultando em um modelo capaz de generalizar bem para novos dados.

3.7 Redução de Features

As imagens utilizadas no presente estudo foram divididas em 3.386 amostras de treinamento e 846 amostras de teste, distribuídas em três classes de movimentos: direita, esquerda e meio. Cada imagem de entrada foi redimensionada para $(224 \times 224 \times 3)$, o que corresponde a aproximadamente:

$$224 \times 224 \times 3 = 150.528, \quad (3.1)$$

features originais (valores de pixel em escala RGB).

Após passarem pela base convolucional da MobileNetV2, essas representações de alta dimensionalidade foram transformadas em mapas de ativação de dimensão $7 \times 7 \times 1280$, totalizando:

$$7 \times 7 \times 1280 = 62.720 \text{ features}. \quad (3.2)$$

Em seguida, a aplicação da camada *GlobalAveragePooling2D* promoveu um resumo ainda mais eficiente, convertendo cada canal em um único valor numérico, resultando em apenas 1280 features finais. Essa transformação representa uma redução de aproximadamente:

$$\text{Redução}(\%) = \frac{150.528 - 1280}{150.528} \times 100 \approx 99,15\%, \quad (3.3)$$

em relação ao espaço original de entrada. Essa redução significativa da dimensionalidade mantém as informações e características mais relevantes para a tarefa de classificação, ao mesmo tempo em que diminui redundâncias e o risco de *overfitting*.

Portanto, a aplicação do *GlobalAveragePooling2D* proporcionou uma redução de aproximadamente 99,85% em relação ao espaço original de entrada, mantendo informações semânticas relevantes ao mesmo tempo em que diminuiu redundâncias e o risco de *overfitting*, favorecendo o processo de classificação.

O modelo final foi estruturado com a adição de uma camada totalmente conectada Dense(128, ReLU) e, posteriormente, uma camada de saída Dense(3, Softmax), correspondendo ao número de classes. O treinamento foi realizado em duas fases: (i) com a base convolucional congelada por 10 épocas, e (ii) em regime de *fine-tuning*, liberando as últimas 20 camadas da MobileNetV2 para ajuste com taxa de aprendizado reduzida, por mais 5 épocas.

O desempenho foi avaliado por meio de métricas de acurácia, matriz de confusão, relatório de classificação (precisão, recall e f1-score por classe) e F1-score ponderado.

3.8 Avaliação do Modelo

Para avaliar o desempenho do modelo, foram utilizadas as métricas clássicas de classificação. Todas calculadas sobre o conjunto de teste, que corresponde a 20% dos frames.

Essas métricas permitem quantificar a capacidade do modelo de identificar corretamente os movimentos de flexão plantar e dorsiflexão, bem como analisar padrões de erro. Além disso, os vídeos V5 e V6, que não foram utilizados durante o treinamento, foram integralmente classificados, possibilitando uma avaliação da generalização do modelo para dados completamente novos.

Após o treinamento, o modelo foi avaliado com as seguintes métricas:

- **Acurácia do teste:** foi obtida a cada época, permitindo monitorar a evolução do aprendizado durante o treinamento;
- **Matriz de Confusão:** representa graficamente o desempenho do modelo em cada classe, evidenciando onde ocorrem os erros;
- **F1-score ponderado:** métrica robusta para avaliação de modelos em conjuntos de dados desbalanceados;
- **Relatório de Classificação:** gerado com a função *classification report*, forneceu medidas detalhadas de precisão (2.1), recall (2.2) e F1-score (2.3) para cada classe individualmente, permitindo uma análise mais completa do desempenho do modelo.

4 Resultados

Neste capítulo a finalidade é discutir sobre os resultados obtidos a partir da implementação da CNN. Inicialmente, descrevem-se os indicadores de desempenho durante o processo de treinamento e teste. Em seguida, é exposto e analisado o desdobramento das métricas de avaliação. Também é apresentada uma interpretação crítica dos resultados, destacando os pontos fortes, as limitações observadas e possíveis estratégias de melhoria. Por último, é feita uma comparação entre os dados obtidos neste trabalho com os dados do trabalho de [Assis \(2024\)](#).

4.1 Desempenho Durante o Treinamento

Na [Figura 9](#) tem-se a evolução da acurácia e da perda durante as primeiras doze épocas de treinamento. Observa-se que a acurácia no conjunto de treino manteve-se estável e elevada, acima de 95%, enquanto a acurácia de teste apresentou oscilações mais acentuadas, alcançando valores próximos a 85%. A perda de teste, por sua vez, também apresentou variações, sobretudo entre a 3ª e a 9ª épocas, sugerindo que o modelo encontrou dificuldade em manter consistência na generalização.

Mesmo não sendo uma métrica de avaliação real da taxa de acertos e erros para uma CNN que classifica imagens, é possível usar tal ponto para avaliar possíveis *overfitting* no modelo. A [Figura 9](#) ajuda a identificar que há um leve sobreajuste nos dados, pois o modelo consegue memorizar padrões do conjunto de treino, mas apresenta instabilidade ao lidar com dados não vistos. Os fatores que podem justificar essas oscilações são: o desequilíbrio entre classes, a sobreposição de padrões visuais semelhantes e o tamanho relativamente reduzido do conjunto de teste.

Portanto, para mitigar esse efeito, foi aplicada a técnica de *EarlyStopping* em conjunto com o *fine-tuning*. A [Figura 10](#) mostra o comportamento após essa etapa, em que parte das camadas convolucionais da base pré-treinada foi descongelada e ajustada. Após esse processo, tanto a acurácia quanto a perda de teste apresentaram curvas mais estáveis, refletindo maior capacidade de generalização e melhor equilíbrio entre aprendizado e teste.

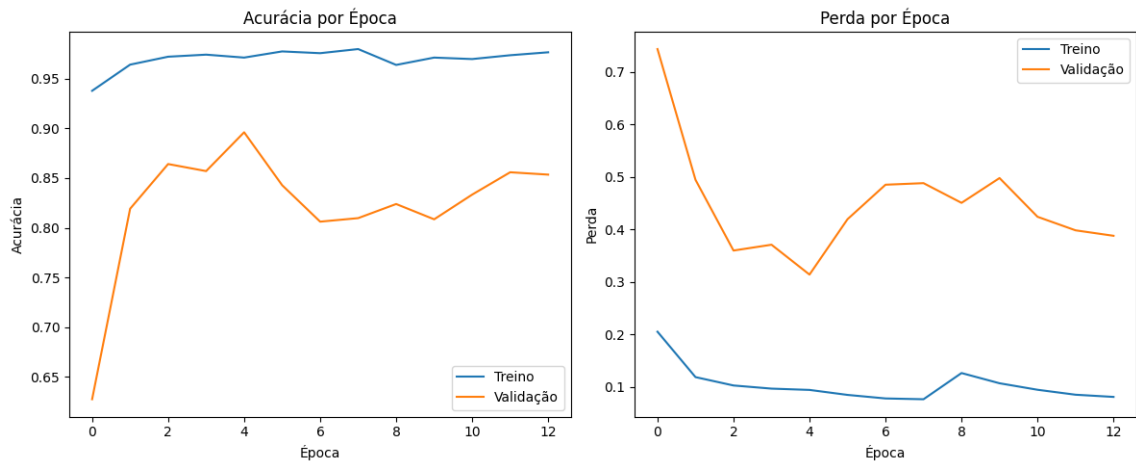


Figura 9 – Curvas de acurácia (esquerda) e perda (direita) para os conjuntos de treino e teste

Fonte: dos Autores

Em resumo, a etapa de treinamento evidenciou que o modelo alcançou alta acurácia no conjunto de treino, mas apresentou oscilações no teste, que foram corrigidas com o uso de *EarlyStopping* e *fine-tuning*. Assim sendo, os ajustes foram fundamentais para permitir maior estabilidade e capacidade de generalização. A seguir, apresenta-se a análise da matriz de confusão, que possibilita uma compreensão mais detalhada dos acertos e erros por classe.

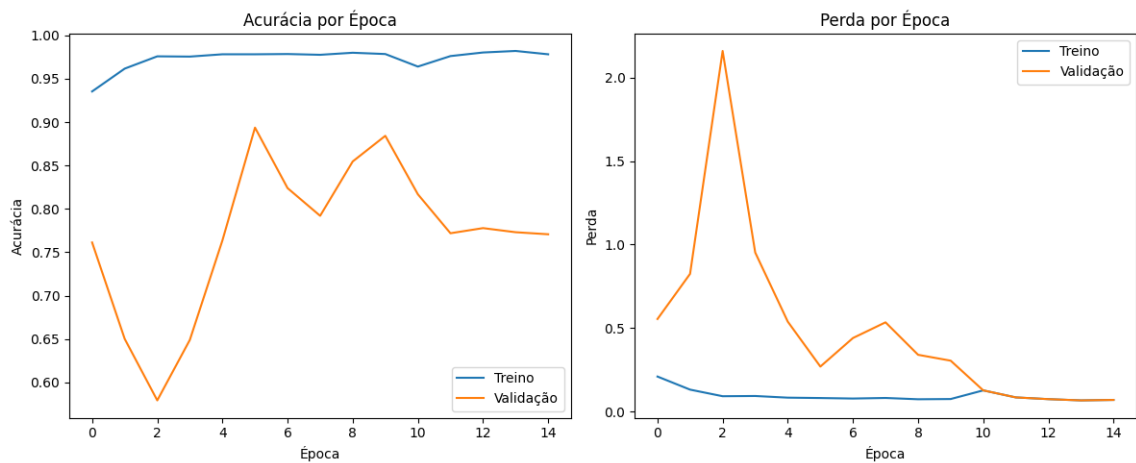


Figura 10 – Curvas de acurácia (esquerda) e perda (direita) para os conjuntos de treino e teste após aplicação de EarlyStopping e Fine-tuning

Fonte: dos Autores

4.2 Análise Matriz de Confusão

Na fase de avaliação final, o modelo atingiu um **F1-score** ponderado de 0.8547, com acurácia global de aproximadamente 85,3%. A matriz de confusão é apresentada na Figura 11.

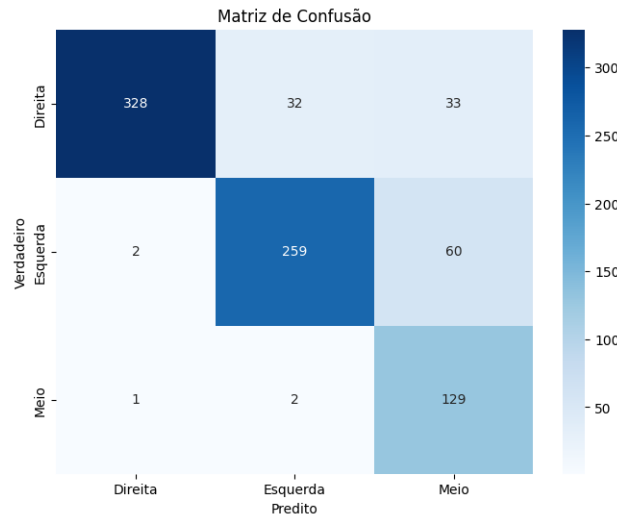


Figura 11 – Matriz de confusão obtida na teste

Fonte: dos Autores

Uma observação importante é que para facilitar no desenvolvimento do modelo, adotou-se classe direita como dorsiflexão, classe esquerda como flexão plantar e meio como repouso.

A partir da matriz, é possível avaliar em detalhe o desempenho em cada classe:

- **Classe Direita (Dorsiflexão):** O modelo obteve 328 classificações corretas, com 32 instâncias confundidas com a classe “Esquerda” e 33 com a classe “Meio”. Isso representa cerca de 16,5% de erros, mas ainda assim com um desempenho robusto, mantendo alta precisão e recall;
- **Classe Esquerda (Flexão):** Foram corretamente classificadas 259 amostras. Mas, houve 60 erros ao serem confundidas com a classe “Meio”, além de apenas 2 instâncias incorretamente preditas como “Direita”. Isso sugere que há maior sobreposição entre os padrões de movimento de “Esquerda” e “Meio”;
- **Classe Meio (Repouso):** Apresentou 129 acertos em 132 amostras, resultando em recall muito elevado (98%). No entanto, essa classe recebeu diversas predições incorretas oriundas de “Direita” (33) e “Esquerda” (60), o que comprometeu sua precisão (58%). Isso indica que, embora o modelo tenha identificado quase todas as instâncias de “Meio”, ele também superestimou essa classe.

De forma geral, a matriz de confusão evidencia que o modelo apresenta bom equilíbrio entre as classes. Com melhor desempenho em “Direita” (dorsiflexão) e “Esquerda” (flexão plantar), mas ainda enfrenta desafios na diferenciação da classe “Meio” (repouso). Isso pode ser possível devido a sua sub-representação no conjunto de treinamento e devido sua similaridade visual com as demais classes. Os erros apresentado pelo modelo podem ser observados na Figura 12. Nota-se, na imagem a), que a classe repouso (meio) foi classificada incorretamente como flexão plantar (esquerda). Já, na imagem b), a classe prevista é repouso mas o movimento é de flexão plantar. E por fim, na imagem c), o previsto é meio e o movimento real é de dorsiflexão. Portanto, há uma dificuldade do modelo em identificar a posição de repouso.



Figura 12 – Ilustração de classificação incorreta dos movimentos

Fonte: dos Autores

No contexto deste trabalho, é fundamental reduzir tanto os falsos positivos quanto os falsos negativos, uma vez que o objetivo é garantir uma classificação precisa e confiável dos movimentos. Observando a matriz de confusão, percebe-se que os falsos positivos já se encontram em valores muito baixos, próximos de 0, 1 ou 2 ocorrências, o que indica que o modelo raramente classifica incorretamente um movimento inexistente. Portanto, o aspecto que requer maior atenção é a redução dos falsos negativos, ou seja, casos em que um movimento real não é identificado pelo modelo. A melhoria nesse ponto contribuirá diretamente para aumentar a acurácia geral e o F1-score, que são as métricas de interesse para este estudo, garantindo que o modelo seja robusto tanto na detecção correta dos movimentos quanto na minimização de erros de classificação.

4.3 Relatório de Classificação

A análise detalhada do relatório de classificação (Tabela 2) evidencia diferenças importantes no desempenho do modelo entre as classes. A classe “Direita” (dorsiflexão) apresentou elevada precisão (0,99), indicando que a grande maioria das previsões para essa classe estava correta; no entanto, seu recall (0,83) sugere que algumas instâncias dessa classe foram confundidas com outras, possivelmente com a classe “Meio” (repouso).

Por outro lado, a classe “Meio” demonstrou recall muito elevado (0,98), indicando que quase todas as instâncias dessa categoria foram corretamente identificadas, mas sua precisão baixa (0,58) evidencia que muitas previsões atribuídas a essa classe eram, na verdade, de outras categorias, principalmente da classe “Esquerda”(flexão plantar). Já a classe “Esquerda” apresentou um equilíbrio maior entre precisão (0,88) e recall (0,81), resultando em um F1-score consistente (0,84).

Tabela 2 – Relatório de classificação final do modelo.

Classe	Precisão	Recall	F1-score	Suporte
Direita (Dorsiflexão)	0.99	0.83	0.91	393
Esquerda (Flexão Plantar)	0.88	0.81	0.84	321
Meio (Repouso)	0.58	0.98	0.73	132
Média macro	0.82	0.87	0.83	846
Média ponderada	0.89	0.85	0.85	846

Fonte: dos Autores

Portanto, é possível inferir que os resultados corroboram a análise da matriz de confusão, na qual se observa que a maior fonte de erro do modelo está na distinção entre as classes “Meio”(repouso) e “Esquerda”(flexão). A avaliação conjunta do relatório de classificação e da matriz de confusão é, essencial, pois permite compreender o desempenho geral do modelo, e identificar erros específicos. Também é possível identificar as áreas que podem ser aprimoradas, como ajustes nos dados de treinamento ou estratégias de balanceamento de classes.

De maneira geral, a análise das métricas do modelo evidencia um comportamento típico em tarefas de classificação onde a prioridade é não perder eventos importantes. Observa-se que, mesmo com variações na precisão entre as classes, o recall elevado garante que a maioria dos casos reais seja corretamente identificada.

Em contextos críticos, como a classificação de movimentos do pé, é mais relevante capturar todos os eventos ou períodos específicos, ainda que algumas previsões incorretas ocorram. A perda de eventos reais (baixo recall) comprometeria a confiabilidade da análise, impactando exercícios de reabilitação, monitoramento clínico ou interpretações comportamentais. Por outro lado, falsos positivos (baixa precisão) podem gerar notificações redundantes, mas não comprometem a integridade do processo de monitoramento.

Portanto, o modelo demonstra adequação ao minimizar falsos negativos e assegurar que praticamente todos os eventos relevantes sejam detectados, mantendo a efetividade da classificação, mesmo quando a precisão apresenta valores mais baixos em determinadas

classes.

4.3.1 Interpretação e Possíveis Melhorias

É possível concluir que o modelo apresenta desempenho robusto em termos gerais, com destaque para a classe "Direita"(dorsiflexão), que alcançou altos valores de precisão e recall. Em contraste, a classe "Meio"(repouso) mostrou um recall excelente, mas sua baixa precisão evidencia que muitas predições atribuídas a essa classe são, na verdade, de outras categorias. Esse padrão de erro pode ser explicado por fatores como:

- Desequilíbrio entre as classes no conjunto de dados, com menor representatividade da classe "Meio";
- Similaridade visual entre os movimentos da classe “Meio” e outras direções em determinados quadros;
- Sub-representação da classe “Meio” nas imagens de treino, dificultando a aprendizagem de padrões distintos.

A partir disso algumas estratégias podem ser adotadas para aprimorar o desempenho do modelo, que são:

- Aplicação de **técnicas de balanceamento**, como *class weighting* ou *oversampling*, para reduzir o impacto das classes majoritárias;
- Ajuste da **função de perda**, utilizando, por exemplo, *Focal Loss*, que penaliza de forma mais eficaz erros em classes menos representadas;
- Realização de **fine-tuning** em camadas mais profundas do backbone, permitindo que o modelo aprenda características mais específicas das classes;
- Revisão e aprimoramento da **rotulagem das amostras**, garantindo consistência e qualidade nos critérios de classificação.

Essas estratégias visam não apenas aumentar a precisão da classe "Meio", mas também reduzir erros sistêmicos e melhorar a generalização. A combinação de ajustes no conjunto de dados, na arquitetura e na função de perda constitui um caminho promissor para otimizar o desempenho global da rede.

4.4 Comparativo entre os dados dos vídeos e a nova construção analítica

Nesta seção, apresenta-se uma análise comparativa entre os resultados de classificação por uma rede neural convolucional e os dados provenientes de sensores físicos, acoplados na papete. Então, a meta é avaliar a consistência e confiabilidade da CNN e verificar em que medida a classificação automática por vídeo se aproxima dos padrões identificados por instrumentos de medição direta.

Durante o processo de classificação dos vídeos V5 a V8 e na análise dos dados disponibilizados pelo trabalho anterior, foi identificado que o vídeo V6 possuía um arquivo de log atípico, contendo mais de 700 mil linhas, significativamente maior que os demais. Por esse motivo, o vídeo V6 foi excluído da análise comparativa dos resultados, a fim de manter a consistência dos dados avaliados.

Para os demais vídeos, observou-se o seguinte número de linhas de informações de log: V5 com 1.461 linhas, V7 com 703 linhas e V8 com 1.272 linhas. Considerando a quantidade de frames verificados pela metodologia de aprendizado de máquina (ML), tornou-se necessária uma normalização dos dados para possibilitar a comparação entre as bases.

Foi adotada uma proporção de aproximadamente 15 para 1, onde foram utilizados os 703 frames da base do trabalho anterior como referência, enquanto na nossa análise com CNN foram considerados os frames múltiplos de 15. Essa equivalência corresponde aproximadamente a meio segundo de vídeo por frame selecionado na classificação.

A seguir, apresentam-se as razões entre a quantidade de frames da nova análise e as linhas de log da base anterior para os vídeos considerados:

$$\frac{22137}{1461} \approx 15,15 \quad (4.1)$$

$$\frac{10488}{703} \approx 14,92 \quad (4.2)$$

$$\frac{19295}{1272} \approx 15,17 \quad (4.3)$$

Esses valores confirmam a consistência da normalização adotada, evidenciando que a seleção de um frame a cada 15 aproxima-se da equivalência temporal esperada.

A análise comparativa entre os dados apresentados nas duas tabelas evidencia diferenças importantes em termos de percentual de flexão plantar, dorsiflexão e quantidade de sinais analisados. Na Tabela 3, que compila os resultados dos vídeos por participante e

Tabela 3 – Resultados dos vídeos por participante e jogo.

Participante	Jogo	% Flexão Plantar	% Dorsiflexão	Total de Sinais	Vídeo
5	Acelera	23	77	1461	V5
9	Acelera	23	73	1272	V8
4	Chute ao Gol	46	54	703	V7

Fonte: adaptado de [Assis \(2024\)](#)

Tabela 4 – Comparativo entre os dados obtidos na nova construção.

Participante	Jogo	Sinais	Flexão Plantar (%)	Descanso (%)	Dorsiflexão (%)
5	Acelera	22137	6,72	65,29	27,99
9	Acelera	19295	26,98	61,75	11,27
4	Chute ao Gol	10488	42,60	16,03	41,37

Fonte: dos Autores

jogo, os percentuais de flexão plantar para os participantes do jogo “Acelera” (participantes 5 e 9) são muito próximos, com valores de 23,77% e 23,73%, respectivamente. Já o participante 4, do jogo “Chute ao Gol”, apresenta um percentual significativamente maior de flexão plantar (46,54%). Os valores de dorsiflexão permanecem baixos e estáveis, entre 1,91% e 2,00%, enquanto o total de sinais por vídeo varia de 703 a 1461.

A Tabela 4, que traz o comparativo com a nova construção analítica. Nota-se uma ampliação expressiva na quantidade de sinais, o que sugere uma análise mais abrangente. Para o participante 5 (jogo “Acelera”), o percentual de flexão plantar diminui para 6,72%, enquanto para o participante 9, também do mesmo jogo, o percentual sobe para 26,98%. No jogo “Chute ao Gol”, o participante 4 apresenta percentual de flexão plantar (42,60%) semelhante ao obtido inicialmente, reforçando a consistência dos dados para este caso.

Além disso, a nova análise incorpora o percentual de repouso, que revela períodos consideráveis em que o pé estava em neutralidade, especialmente para os participantes do jogo “Acelera”. Por exemplo, o participante 5 apresenta 65,29% de descanso, enquanto o participante 9 apresenta 61,75%. O participante 4, do “Chute ao Gol”, possui um percentual de descanso mais baixo (16,03%), alinhado com seus valores de flexão plantar e dorsiflexão.

Um ponto importante à destacar é que o participante V6, do jogo “Chute ao Gol”, aparece na nova construção com 10.597 sinais, mas não foi possível identificar o log correspondente para comparação direta com os dados da primeira tabela, impossibilitando uma análise mais completa para esse caso.

É importante salientar que a comparação entre os resultados das duas abordagens apresenta certa complexidade, uma vez que os dados foram obtidos em contextos distintos: para a nova análise foram utilizados dois jogos diferentes na aquisição dos sinais, enquanto

os resultados anteriores consideram apenas os registros de cada vídeo individual. Além disso, a classificação original do dispositivo “papete inteligente” baseia-se exclusivamente nos dados de acelerômetro acoplados ao vestível, inferindo os movimentos do pé sem análise visual. Por isso, divergências nos percentuais de flexão plantar, dorsiflexão e descanso podem refletir tanto as diferenças nos jogos quanto nas metodologias de captura e classificação dos sinais. Essa consideração reforça a necessidade de cautela na interpretação direta das porcentagens, destacando que a coerência observada, mesmo com essas diferenças, é indicativa da robustez e consistência da modelo.

O comparativo demonstra que, apesar das diferenças na quantidade de dados analisados e na inclusão do percentual de descanso, os resultados da nova construção mantêm uma coerência geral com os dados iniciais, especialmente para os participantes 4 e 9. Isso reforça a robustez da metodologia utilizada e destaca a importância de uma maior quantidade de dados para uma análise mais detalhada dos sinais de flexão plantar, dorsiflexão e descanso.

5 Conclusão

Este trabalho teve como objetivo principal desenvolver e avaliar um modelo de aprendizado de máquina, baseado na arquitetura MobileNetV2, para a classificação automática de direções (direita, esquerda e meio) a partir de imagens de movimentos do pé, com foco na aplicação em contextos de interação motora. Adicionalmente, buscou-se realizar uma análise comparativa entre os dados obtidos por meio dessa abordagem e os resultados registrados em trabalhos anteriores baseados em logs manuais.

Os resultados alcançados demonstram a viabilidade e eficácia da solução proposta. O modelo treinado atingiu um desempenho global satisfatório, com **acurácia de 85,3%** e **F1-score ponderado de 0,8547**. A análise detalhada das métricas por classe revelou que a maior dificuldade do modelo encontra-se na correta classificação da classe “Meio”, que, embora apresente recall elevado (98%), possui baixa precisão (58%), indicando um alto número de falsos positivos. Por outro lado, as classes “Direita” e “Esquerda” apresentaram resultados mais equilibrados, com alta precisão e bom recall.

A utilização de técnicas como *fine-tuning*, *early stopping* e a divisão adequada dos dados de treino e teste contribuíram para mitigar os efeitos de sobreajuste (*overfitting*) e aprimorar a capacidade de generalização do modelo. Ainda assim, foi possível identificar oportunidades de melhorias, sobretudo em relação ao balanceamento das classes e à definição mais precisa dos critérios de rotulagem das imagens.

A comparação entre os resultados da CNN e os obtidos pela papete inteligente é complexa, devido à utilização de jogos diferentes e à natureza distinta das metodologias de captura de sinais. Mesmo assim, o modelo de CNN apresentou bom desempenho, mostrando-se consistente e passível de uso complementar à papete inteligente para a classificação automática das posições, ampliando a análise dos movimentos com maior detalhamento temporal.

Além disso, a normalização entre as diferentes fontes de dados, com base na proporção de quadros por linha de log, revelou-se uma estratégia eficaz para permitir uma comparação coerente entre abordagens distintas, reforçando a confiabilidade da metodologia empregada.

Portanto, os resultados demonstram que a classificação automática por visão computacional amplia a análise dos movimentos ao incorporar informações adicionais, como os períodos de descanso. Dessa forma, a abordagem proposta se mostra promissora para aplicações futuras na avaliação detalhada dos sinais de movimento do tornozelo, oferecendo maior abrangência e precisão na interpretação dos dados.

Limitações e Trabalhos Futuros

Apesar dos resultados positivos, algumas limitações foram identificadas. A principal delas diz respeito ao desequilíbrio entre as classes no conjunto de treinamento, o que influenciou negativamente a capacidade do modelo em classificar corretamente a classe “Meio”. Além disso, o número relativamente limitado de participantes e vídeos utilizados restringe a generalização dos resultados para outros contextos ou populações.

Como propostas para trabalhos futuros, sugerem-se:

- Ampliação do conjunto de dados, com inclusão de novos participantes e cenários;
- Aplicação de técnicas de balanceamento, como *oversampling*, *undersampling* ou o uso de pesos nas classes durante o treinamento;
- Utilização de arquiteturas mais recentes ou ensembles de modelos para ganho de robustez;
- Integração com sensores adicionais (como IMUs) para análise multimodal dos movimentos;
- Desenvolvimento de um sistema em tempo real para aplicação prática em ambientes de reabilitação ou jogos educativos.

Considerações Finais

Conclui-se que a abordagem adotada neste trabalho representa um avanço no uso de aprendizado de máquina para a análise automática de movimentos motores a partir de imagens. O modelo desenvolvido mostrou-se eficaz, com resultados consistentes tanto nas métricas de avaliação quanto na comparação com dados previamente coletados. A metodologia proposta apresenta potencial para ser aplicada em sistemas de reabilitação, educação motora e monitoramento automatizado de movimentos, contribuindo para a evolução de soluções tecnológicas voltadas à análise do comportamento humano.

Referências

- ASSIS, G. A. de et al. Jogos com tecnologia vestível como estímulo à saúde dos pés—avaliação de usabilidade. *Journal of Health Informatics*, v. 16, n. Especial, 2024. Citado 3 vezes nas páginas 14, 16 e 17.
- ASSIS, L. e. a. Estudo sobre a análise dos movimentos de flexão plantar e dorsiflexão utilizando sensores. *Journal of Rehabilitation Research*, v. 48, n. 1, p. 45–52, 2024. Citado 9 vezes nas páginas 11, 12, 13, 15, 21, 22, 24, 29 e 36.
- BARBAN, R. *Anatomia do pé*. 2025. Acesso em: 27 ago. 2025. Disponível em: <<https://www.rafaelbarban.com.br/anatomia-do-pe/>>. Citado na página 17.
- BAUMGARTL, H.; BUETTNER, R. Improving scene understanding for mobile robots using transfer learning and data augmentation. *arXiv preprint arXiv:2107.11187*, 2021. Disponível em: <<https://arxiv.org/abs/2107.11187>>. Citado 3 vezes nas páginas 16, 17 e 18.
- BERGMANN, H. e. a. Influência da flexão plantar e dorsiflexão na mecânica da marcha. *Journal of Biomechanics*, v. 51, n. 2, p. 203–210, 2018. Citado 3 vezes nas páginas 11, 16 e 17.
- CHEN, J. e. a. Análise de padrões de movimento usando aprendizado de máquina e visão computacional. *Computer Vision Journal*, v. 57, p. 115–124, 2021. Citado na página 11.
- CRESSWELL, A. e. a. Controle muscular da dorsiflexão durante a marcha. *Journal of Applied Physiology*, v. 119, n. 4, p. 511–519, 2015. Citado 3 vezes nas páginas 11, 16 e 17.
- HARRISON, M. *Machine Learning – Guia de Referência Rápida: Trabalhando com Dados Estruturados em Python*. São Paulo: Novatec, 2020. Este livro apresenta uma visão geral do processo de machine learning e da classificação com dados estruturados. ISBN 9788575228173. Citado 2 vezes nas páginas 18 e 19.
- HUONG, P. T. et al. Image classification of orange surface defects using transfer learning and data augmentation. *Journal of Low Frequency Noise, Vibration and Active Control*, v. 44, n. 2, p. 1–10, 2025. Disponível em: <<https://doi.org/10.1177/17483026241309070>>. Citado 3 vezes nas páginas 16, 17 e 18.
- KONG, P. e. a. Músculos envolvidos na flexão plantar e dorsiflexão: implicações para a reabilitação. *Clinical Biomechanics*, v. 60, p. 54–60, 2018. Citado 3 vezes nas páginas 11, 16 e 17.
- LEE, J. e. a. Análise quantitativa da recuperação motora com aprendizado de máquina. *Neuroscience and Biobehavioral Reviews*, v. 98, p. 135–146, 2019. Citado na página 12.
- LIU, X. e. a. Feedback em tempo real na reabilitação motora usando aprendizado de máquina. *Journal of Rehabilitation Engineering*, v. 47, n. 3, p. 189–199, 2020. Citado na página 12.

- MARIANO, D. *Métricas de avaliação em machine learning: acurácia, sensibilidade, precisão, especificidade e F-score*. 2021. Disponível em: <<https://bioinfo.com.br/metricas-de-avaliacao-em-machine-learning-acuracia-sensibilidade-precisao-especificidade-e-f-score/>>. Citado na página 20.
- NAVARRO, C. e. a. Sistema de reabilitação adaptativa com aprendizado de máquina. *International Journal of Rehabilitation Research*, v. 41, n. 2, p. 88–98, 2018. Citado na página 12.
- PEREZ, L.; WANG, J. The effectiveness of data augmentation in image classification using deep learning. *Convolutional Neural Networks for Visual Recognition*, 2017. Citado 2 vezes nas páginas 16 e 18.
- ROYER, T. e. a. Impactos de distúrbios na flexão plantar e dorsiflexão na marcha e prevenção de lesões. *Physical Therapy Journal*, v. 95, n. 6, p. 525–533, 2015. Citado 3 vezes nas páginas 11, 16 e 17.
- SANDLER, M. e. a. Mobilenetv2: Inverted residuals and linear bottlenecks. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, p. 4510–4520, 2018. Citado 2 vezes nas páginas 16 e 18.
- SINGH, R. *Decoding CNNs: A Beginner's Guide to Convolutional Neural Networks and their Applications*. 2024. Acesso em: 20 ago. 2025. Disponível em: <<https://ravjot03.medium.com/decoding-cnns-a-beginners-guide-to-convolutional-neural-networks-and-their-applications>>. Citado na página 19.
- ZHANG, Y. e. a. Utilização de aprendizado de máquina para análise de movimentos corporais em vídeos. *IEEE Transactions on Neural Networks*, v. 30, n. 10, p. 2681–2693, 2019. Citado na página 11.