

# Análise de Desempenho em Combates de Espada Utilizando Visão Computacional e Machine Learning

Alexandre T. Covolan  
Pontifícia Universidade Católica de  
Campinas  
Campinas - SP, Brasil  
alexandre.tc1@puccampinas.edu.br

Benito T. Marino  
Pontifícia Universidade Católica de  
Campinas  
Campinas - SP, Brasil  
benito.tm@puccampinas.edu.br

João V. D. Oresco  
Pontifícia Universidade Católica de  
Campinas  
Campinas - SP, Brasil  
joao.vdo@puccampinas.edu.br

Leonardo G. Quartaroli  
Pontifícia Universidade Católica de  
Campinas  
Campinas - SP, Brasil  
leonardo.gq@puccampinas.edu.br

Leonardo M. Alves  
Pontifícia Universidade Católica de  
Campinas  
Campinas - SP, Brasil  
leonardo.ma1@puccampinas.edu.br

Mateus M. A. de Lima  
Pontifícia Universidade Católica de  
Campinas  
Campinas - SP, Brasil  
mateus.mal@puccampinas.edu.br

Saullo H. G. de Oliveira  
Pontifícia Universidade Católica de  
Campinas  
Campinas - SP, Brasil  
saullo.haniell@puc-campinas.edu.br

## Abstract

Performance analysis in fencing is still largely dependent on manual observation and subjective interpretation by coaches and referees, which limits scalability, reproducibility, and data-driven decision-making. This work presents a computer vision-based system designed to automatically analyze fencing bouts from video data. The system applies human pose estimation to track athletes' movements and extracts kinematic features such as displacement, velocity, vertical position of the weapon's guard, and movement intensity. These features are processed to generate performance metrics and structured visualizations. The results indicate that automated video-based analysis can reduce subjectivity, increase analytical consistency, and serve as a practical support tool for athletes and coaches in fencing training and performance evaluation.

## Keywords

Pose Estimation, Machine Learning, Sport Performance Analysis

## 1 INTRODUÇÃO

A esgrima é uma prática corporal que combina velocidade, precisão e tomada de decisão em frações de segundo. Acredita-se que a ponta da arma da esgrima é o segundo objeto mais rápido das Olimpíadas, atrás apenas da munição do tiro esportivo [1].

Por simular combates com armas letais, cada golpe recebido por um atleta tem muito a ensinar ao praticante, sendo a movimentação do atirador (termo usado para descrever o atleta em combate) um parâmetro importante da análise. Entender como cada movimento contribui para o desempenho do atleta é essencial para o seu treinamento e desenvolvimento, e esse entendimento atualmente é feita principalmente com auxílio de um especialista, extremamente escasso por se tratar de um esporte de nicho.

Utilizando algoritmos de Visão Computacional (VC), Estimativa de Pose e Aprendizado de Máquina (ML), este artigo relata

o desenvolvimento de uma ferramenta que extrai informações relevantes de atletas presentes em vídeos de combate, e exibe métricas e análises diversas para servir de apoio a atletas e treinadores. Os objetivos para o desenvolvimento da ferramenta foram:

- i) automatizar parte da análise de combates, permitindo identificar eventos relevantes sem a intervenção manual;
- ii) extrair dados e informações de combates de esgrima em dados estruturados, fornecendo informações objetivas que apoiem decisões de treinamento;
- iii) extrair métricas relevantes do desempenho do atleta, incluindo postura, distância, padrões de movimento e indicadores táticos.

## 2 TRABALHOS RELACIONADOS

Comparado a esportes como vôlei, futebol e golfe, a esgrima não possui um vasto campo de estudo com utilização de novas tecnologias. Em geral, os sistemas existentes concentram-se na detecção dos toques e na pontuação do combate.

No trabalho [2], é proposto o método FACTS (Fine-Grained Action Classification for Tactical Sports), baseado em *transformers*, que aprende a reconhecer padrões temporais e espaciais diretamente dos quadros do vídeo. Este método não necessita de esqueleto 3D ou pontos corporais para efetuar a classificação e reconhecimento detalhado e preciso de ações complexas em esportes táticos de combate, como a esgrima e o boxe.

Allez Go [3, 4] é uma *startup* que tem como principal objetivo tornar os combates de esgrima mais acessíveis e visuais ao público e, possivelmente, auxiliar a arbitragem. A abordagem que eles utilizam no *software* é, a partir do vídeo e do áudio de um combate, detectar eventos como impactos, movimentos da lâmina, deslocamento dos atiradores, detectar qual atleta possui a prioridade (fator primordial para o sabre e o florete), e identificar quem marcou cada ponto.

### 3 METODOLOGIA

O desenvolvimento da ferramenta considerou diversas etapas. Primeiro, a arquitetura do sistema será descrita. Em seguida, os sistemas de classificação e segmentação são apresentados, bem como as métricas de avaliação e o *dashboard* de desempenho individual, relatados na seção de resultados.

#### 3.1 Arquitetura do Sistema

A arquitetura do sistema segue o fluxo a seguir: O usuário faz o upload do vídeo de um combate, selecionando se este vídeo possui ou não o *overlay* de placar virtual. O vídeo, então, segue as seguintes etapas de extração de informação:

- estimação dos pontos corporais por meio do modelo MMPose;
- diferenciação dos atiradores por meio de um modelo de rotulação;
- segmentação de momentos “em combate” e “fora de combate”;
- os pontuadores de cada ponto são estimados seguindo dois métodos: i) se possuir o placar virtual, o vídeo passará por um modelo de identificação dos pontos de forma visual, identificando este placar no próprio vídeo; ii) se não possuir placar virtual, o vídeo passará por um modelo de detecção de pontuador com base nos últimos 75 quadros do combate.

Por fim, as informações extraídas são apresentadas em um *dashboard* interativo e configurável das métricas de combate do vídeo. Cada uma dessas etapas é detalhada nas sub-seções a seguir.

#### 3.2 Identificação de Atiradores

Em vídeos de competições de esgrima, além dos atletas, também estão presentes outras pessoas no ambiente. No vídeo, então, é preciso diferenciar quem são os atiradores. Para isso, criamos um modelo de rotulação de atiradores.

**3.2.1 Anotação da Base de Dados.** Separamos 4211 imagens retiradas de momentos aleatórios de 15 diferentes vídeos de combates e rotulamos os *bounding boxes* dos atletas em cada uma das imagens. A rotulação foi feita por meio do *software* CVAT e treinada usando o modelo YOLOv8 1.1 (*ultralytics*).

**3.2.2 Classificação de Atiradores.** Para a identificação dos pontos corporais dos esgrimistas no vídeo, foi utilizado o modelo MMPose, que consegue identificar mais de uma pessoa e é recomendado para situações não laboratoriais [5].

**3.2.3 Utilização do MMPose + Rotulação.** O resultado do modelo de rotulação foi um detector capaz de identificar exclusivamente os atiradores em cada quadro do vídeo, retornando apenas os pontos dos *bounding boxes* deles, usado como parâmetro para o modelo do MMPose na estimativa de pose 2D, para realizar somente a análise das estimativas de pose dentro daquelas áreas.

**3.2.4 Métricas de Análise.** Para os experimentos quantitativos do modelo MMPose, foi calculada a Média de Precisão (AP), em três limiares de OKS (*Object Keypoint Similarity*), e a Média de Revocação (AR) em dois limiares de OKS.

Para os experimentos quantitativos do modelo de rotulação de atiradores, foram calculadas a Precisão e Revocação, além da Média da Precisão Média em diferentes limiares de IoU.

#### 3.3 Identificação do Atleta Pontuador

A implementação da identificação do pontuador foi realizada de duas formas distintas no nosso sistema:

- Utilizando o placar virtual que o vídeo possui.
- Replicando o projeto [6], disponível no Kaggle.

**3.3.1 Identificação por Placar Virtual.** Para a identificação via placar virtual, utilizamos o Tesseract OCR para o reconhecimento de caracteres na imagem, configurando-o de modo a analisar exclusivamente os números do placar em tempo real. A implementação foi realizada por meio da biblioteca pytesseract, em Python, configurando para os dois tipos de placar que estão presentes na maioria dos vídeos de esgrima. Então, criamos um modelo que consegue analisar qual dos dois placares está presente no vídeo, e aplicar filtros específicos para cada um.

**3.3.2 Identificação com Redes Neurais (CNN + LSTM).** Em vídeos que não possuem placar virtual, fazemos a análise de quem marcou os pontos replicando o projeto do Alexandre Paugeaud disponível no Kaggle [6]. Ele obteve um resultado de 70% de acurácia com dados extraídos de vídeos usando o modelo de estimativa de pose OpenPose, combinando uma CNN e uma arquitetura LSTM para o treinamento e teste. Os dados foram segmentados em cliques de 75 quadros (2,5 segundos), onde cada clipe significa o fim de um combate (momento em que ocorre o ponto). Cada clipe possui uma classificação, se foi ponto do esgrimista da esquerda, da direita, ou um toque simultâneo. Cada linha do dataset possui o id do clipe e as coordenadas dos pontos corporais dos dois esgrimistas. Na nossa utilização, foram removidos os dados da classe de toque simultâneo, tanto pela ambiguidade de identificação quanto pelo desbalanceamento de classes.

#### 3.4 Segmentação de Momentos “Em Combate”

Outro problema na análise de combates é a segmentação do vídeo nos momentos em que o combate está de fato ocorrendo. Vídeos de esgrima são constantemente paralisados. Este problema também foi resolvido com um sistema de classificação, detalhado a seguir.

**3.4.1 Anotação da Base de Dados.** Momentos fora de combate, se analisados, interferem na análise estatística das métricas, diluindo a relevância dos dados. Por isso, foi desenvolvido um sistema de classificação temporal para distinguir os estados “em combate” e “fora de combate”. Inicialmente, foi criada uma base de dados contendo 9 vídeos (aproximadamente 270 mil quadros) rotulados manualmente com o estado de combate (0 = fora de combate, 1 = em combate). A partir dessas amostras, realizamos uma engenharia de atributos, reduzindo a dimensionalidade da base de dados - que tinha originalmente 34 atributos - para cinco:

- média do ângulo do braço da frente;
- média do ângulo da perna da frente;
- média do movimento corporal;
- diferença de distância entre atletas;
- distância absoluta entre atletas.

3.4.2 *Classificação “Em Combate” × “Fora de Combate”*. Os dados foram padronizados, centralizando todas as features para média 0 e desvio padrão 1, e utilizados para ajustar uma rede neural com arquitetura LSTM, que é capaz de considerar dependências temporais entre quadros [7].

3.4.3 *Métricas de Análise*. Para avaliar quantitativamente o modelo de segmentação de combate, foram calculadas as métricas AUC, F1-score, Acurácia (ACC) e a Função de Perda (Loss).

## 4 RESULTADOS OBTIDOS

Apresentamos: i) os resultados das avaliações quantitativas, que consistiram em medir a capacidade dos modelos de classificar corretamente os dados não vistos na base de treinamento; ii) as principais visualizações desenvolvidas no *dashboard*; e iii) a avaliação qualitativa constituiu em uma análise completa do sistema, realizada por um Mestre D’Armas (título dado a um profissional com conhecimentos técnicos e pedagógicos para formar esgrimistas, aplicar regras e arbitrar competições).

### 4.1 Experimentos Quantitativos

AP	AP50	AP75	AR	AR50
0.756	0.908	0.826	0.809	0.945

**Tabela 1: Resultados do modelo MMPose, que extrai a pose dos atiradores nos vídeos.**

Precision	mAP50	mAP50-95	Recall
0.994	0.992	0.995	0.905

**Tabela 2: Resultados do modelo de classificação de atiradores.**

AUC	F1	ACC	Loss
0.903	0.798	0.822	0.392

**Tabela 3: Resultados do modelo de segmentação de momentos “Em Combate” contra “Fora de Combate”.**

Na tabela 1, percebe-se que, com limiares mais baixos de OKS, tanto a precisão média como a revocação média (AP50, AR50) apresentam valores acima de 90%. Com a tabela 2, nota-se que todas as métricas relacionadas à precisão obtiveram valores acima de 99%, com revocação de 90%. Já na tabela 3 vemos que a acurácia (ACC) obtida foi de 82%, com uma ROC-AUC de 90,3%. Os erros observados ocorreram em curtos intervalos (geralmente < 10 quadros). Para mitigá-los, implementamos uma etapa de pós-processamento que corrige previsões isoladas com duração inferior a 30 quadros.

### 4.2 Dashboard de Desempenho Individual

A partir das funcionalidades implementadas na ferramenta, desenvolvemos então um *dashboard* onde os usuários podem interagir com as informações extraídas dos combates de espada. Poderá obter informações referentes à velocidade do punho em comparação ao adversário, intensidade do movimento, ângulo da perna da frente, porcentagem do tempo em que manteve o copo acima, etc. A superioridade do copo em momentos de ataque está muito associada ao sucesso do ataque. O objetivo principal desta funcionalidade é extrair ao máximo os dados e informações relevantes do vídeo processado, para proporcionar uma análise automatizada que apoie a compreensão dos movimentos de atiradores nos combates.

## 4.3 Análise Qualitativa

A avaliação qualitativa do nosso sistema consistiu num relato de percepção individual. O Mestre D’armas Maicon Pereira, atleta em atividade, multicampeão e atleta destaque do ano de 2025 pela Confederação Brasileira de Esgrima (CBE), utilizou a ferramenta com o apoio dos autores. Em seguida, solicitamos um registro de sua percepção sobre as funcionalidades implementadas, bem como sua utilidade prática. Foi ponderado nesta avaliação que o nosso sistema tem bastante funcionalidade e usabilidade no meio da esgrima. Hoje, é muito difícil para praticantes do esporte fora da federação conseguirem fazer uma análise tão precisa dos treinos e combates, algo que é disponibilizado na ferramenta.

“Este trabalho é revolucionário no mundo da esgrima. A esgrima não é nem um pouco democrático, diversos estados do Brasil não possuem um Mestre D’armas, o que torna a introdução de novos alunos muito difícil, e pensando tanto do ponto de vista pedagógico quanto do ponto de vista do auto rendimento, este trabalho, que é de nível de seleção brasileira, ajudará muito”.

## 5 CONSIDERAÇÕES FINAIS

Os resultados obtidos demonstram que a integração de técnicas de Estimativa de Pose e ML representa uma abordagem viável e eficiente para a análise técnica na esgrima. Os resultados evidenciam que pequenas variações individuais de estilo entre atletas não comprometem significativamente a robustez do modelo. Contudo, a expansão da base de dados deve aumentar a capacidade de generalização e a precisão temporal.

A eliminação da dependência de sensores corporais e sistemas IoT torna a análise mais acessível e não intrusiva. O sistema proposto pode ser implementado utilizando câmeras convencionais, incluindo dispositivos móveis como smartphones, reduzindo custos de aplicação e facilitando sua adoção por atletas e treinadores.

Como trabalhos futuros, é possível explorar combates de outras armas da esgrima esportiva. Para isso, será necessário aprofundar os temas de prioridade de ataque, ponto primordial para a detecção de pontuadores em sabres e floretes, permitindo que o sistema alcance maior eficiência nessa detecção. Uma direção mais complexa seria o desenvolvimento de módulos para classificação de tipos de movimento e ataque, possibilitando a identificação específica de cada ação executada pelos esgrimistas, viabilizando análises técnicas mais detalhadas e precisas de cada ponto do combate.

## Referências

- [1] Factbox: Fencing at the Olympics – Chicago Tribune.
- [2] Christopher Lai, Jason Mo, Haotian Xia, and Yuan-fang Wang. FACTS: Fine-Grained Action Classification for Tactical Sports, December 2024. arXiv:2412.16454 [cs] version: 1.
- [3] Jason Mo. Allez Go: Computer Vision and Audio Analysis for AI Fencing Referees. *Journal of Student Research*, 11(4), November 2022.
- [4] Allez Go.
- [5] Jelena Aleksic, Dmitry Kanevsky, David Mesaroš, Olivera Knezevic, Dimitrije Cabarkapa, Branislav Božović, and Dragan Mirkov. Validation of Automated Countermovement Vertical Jump Analysis: Markerless Pose Estimation vs. 3D Marker-Based Motion Capture System. *Sensors*, 24:6624, October 2024.
- [6] Alexandre Pageaud. Sport : Fencing Matches AI.
- [7] Yaxuan Kong, Zepu Wang, Yuqi Nie, Tian Zhou, Stefan Zohren, Yuxuan Liang, Peng Sun, and Qingsong Wen. Unlocking the Power of LSTM for Long Term Time Series Forecasting, February 2025. arXiv:2408.10006 [cs].