

INTELIGÊNCIA ARTIFICIAL NA ANÁLISE BIOMECÂNICA DO MOVIMENTO DE SAQUE DO TÊNIS DE CAMPO

Pedro Nicolau Backes Jovenal¹, Marlon de Matos de Oliveira²

Resumo: O presente trabalho propôs a elaboração de uma base de dados a partir de gravações de atletas de tênis da região de Criciúma/SC, utilizada para treinar modelos de inteligência artificial voltados à análise biomecânica do saque no tênis de campo. Foram comparadas duas abordagens: a rede recorrente GRU e a Spatio-Temporal Graph Convolutional Network (ST-GCN), cada uma com estratégias de pré-processamento distintas: a GRU foi alimentada com vetores temporais de coordenadas corporais, enquanto a ST-GCN utilizou grafos representando articulações e conexões biomecânicas. O processo metodológico envolveu a extração de pontos corporais, a organização dos dados em sequências temporais e o treinamento supervisionado dos modelos. A aplicação desenvolvida analisa e classifica vídeos em tempo real, identificando saques corretos e incorretos por fase. Nos resultados, embora a GRU tenha alcançado acurácia global de 98,64%, observou-se que parte desse valor decorre de classificações inadequadas, indicando limitações na detecção de falhas técnicas. Já a ST-GCN apresentou acurácia de 91,12%, porém com maior consistência na identificação de erros biomecânicos ao longo das fases do movimento. Conclui-se que técnicas de visão computacional e aprendizado de máquina podem ser aplicadas como ferramentas de apoio ao treinamento esportivo, permitindo análises biomecânicas mais objetivas e auxiliando no diagnóstico técnico do saque no tênis de campo.

Palavras-chave: análise biomecânica; aprendizado de máquina; GRU; ST-GCN; visão computacional; saque no tênis.

¹Curso de Ciência da Computação, Universidade do Extremo Sul Catarinense (Unesc), Criciúma - Santa Catarina - Brasil. backespedro2016@gmail

²Orientador, Curso de Ciência da Computação, Universidade do Extremo Sul Catarinense (Unesc), Criciúma - Santa Catarina - Brasil. marlon.oliveira@unesc.net

ABSTRACT: The present work proposed the development of a database from recordings of tennis athletes from the region of Criciúma/SC, used to train artificial intelligence models aimed at the biomechanical analysis of the tennis serve. Two approaches were compared: the recurrent GRU network and the Spatio-Temporal Graph Convolutional Network (ST-GCN), each with distinct preprocessing strategies. The GRU was fed with temporal vectors of body coordinate data, while the ST-GCN used graphs representing joints and biomechanical connections. The methodological process involved extracting body landmarks, organizing the data into temporal sequences, and performing supervised training of the models. The developed application analyzes and classifies videos in real time, identifying correct and incorrect serves by phase. In the results, although the GRU achieved an overall accuracy of 98.64%, it was observed that part of this value resulted from inadequate classifications, indicating limitations in detecting technical errors. The ST-GCN, in turn, achieved an accuracy of 91.12%, but with greater consistency in identifying biomechanical errors throughout the phases of the movement. It is concluded that computer vision and machine learning techniques can be applied as support tools for sports training, enabling more objective biomechanical analyses and assisting in the technical diagnosis of the tennis serve.

Keywords: biomechanical analysis; machine learning; GRU; ST-GCN; computer vision; tennis serve.

1 INTRODUÇÃO

O tênis de campo é um dos maiores esportes do mundo contando com aproximadamente 1,17% da população global como praticantes ativos (Federation, 2019). Um jogo de tênis é disputado por dois ou quatro jogadores dentro de uma quadra e basicamente o esporte consiste em rebater uma bola utilizando uma raquete e rebatendo-a para o lado do adversário até que a bola saia dos limites da quadra ou realize 2 toques seguidos ao chão, caracterizando um ponto. Todos os pontos são iniciados a partir de um movimento especial conhecido como saque (Fantoura, 2003).

O saque é o movimento mais importante e mais complexo dentro do esporte, exigindo um grande empenho dos atletas para desenvolver a técnica. Ele envolve uma sequência coordenada de ações biomecânicas, nas quais o atleta arremessa a bola para o alto, enquanto prepara a raquete com a outra mão para realizar o golpe. Esse movimento exige sincronização entre membros superiores e inferiores, controle postural, transferência

de peso e precisão no contato com a bola, sendo fundamental para dar início ao ponto de forma eficaz e estratégica (Ishizaki; Castro, 2006).

Há uma demanda significativa para elaborar planos específicos de desenvolvimento de treinamento dentro do tênis de campo. É ideal que esses treinamentos atuem com ênfase em algumas características como a prevenção de lesões, ganho de velocidade, atividade muscular, efetividade, entre outros. Podendo então, serem aplicados tanto para atletas profissionais e principalmente para atletas amadores que na maioria das vezes não possuirá um especialista ao seu lado (Valim, 2021).

A análise biomecânica do saque no tênis pode ser compreendida a partir um conceito conhecido como cadeia cinética. A cadeia cinética refere-se a um sistema mecânico no qual diferentes segmentos corporais atuam de forma sequencial e coordenada para a realização de movimentos complexos, como o saque no tênis. A cinemática da cadeia cinética pode ser dividida em fases, que representam os segmentos individuais do movimento de forma cronológica (Kovacs; Ellenbecker, 2012).

O cadeia cinética divide-se em três fases principais: preparação, aceleração e finalização. Na preparação, o corpo se alinha e armazena energia potencial, com destaque para a flexão dos joelhos, rotação simétrica do tronco, estabilização do core e posicionamento do braço dominante. A fase de aceleração ocorre da rotação externa máxima do ombro até o contato com a bola, gerando energia cinética principalmente pelas pernas e tronco, enquanto a rotação interna do ombro e a flexão do punho determinam a velocidade da raquete. Já a finalização envolve a desaceleração etapa de maior sobrecarga excêntrica, exigindo ação do manguito rotador posterior e a desaceleração, no qual o contato com o solo influencia equilíbrio e recuperação, variando conforme a técnica utilizada (Kovacs; Ellenbecker, 2012).

A compreensão da cadeia cinética é essencial para a análise biomecânica do saque, pois cada fase do movimento envolve padrões específicos de coordenação e transferência de energia entre segmentos corporais. Nesse contexto, técnicas computacionais capazes de identificar padrões complexos em sequências de movimento tornam-se adequadas para automatizar essa avaliação, o que justifica a aplicação de métodos de Machine Learning na análise biomecânica.

Machine Learning (ML), ou aprendizado de máquina, é um subcampo da inteligência artificial voltado ao desenvolvimento de métodos que permitem aos computadores aprender automaticamente a partir de dados,

sem necessidade de programação explícita para cada tarefa. Trata-se de uma abordagem que automatiza a construção de modelos analíticos, utilizando algoritmos que identificam padrões de informações para gerar conhecimento e tomar decisões (Monard; Baranauskas, 2003).

De forma geral, os algoritmos de aprendizado de máquina aprendem a partir de experiências passadas, chamadas de casos ou exemplos, que ajudam na tomada de decisão e contribuem para melhorar seu desempenho ao longo do tempo. A principal diferença entre os tipos de aprendizado está na presença ou ausência de rótulos nos dados utilizados. No aprendizado supervisionado, os dados de treinamento já vêm com rótulos, ou seja, cada exemplo está associado a uma resposta conhecida. Por outro lado, no aprendizado não supervisionado, esses rótulos não existem, e o algoritmo precisa encontrar padrões ou estruturas escondidas nos dados (Chapelle; Schölkopf; Zien, 2009; Oliveira, 2018).

Dentro desse contexto, o *deep learning* surge como uma vertente do aprendizado de máquina que utiliza redes neurais profundas, compostas por múltiplas camadas, capazes de extrair representações hierárquicas e mais complexas dos dados. Essa abordagem é especialmente eficiente para problemas que envolvem grandes volumes de informações e alta complexidade, como visão computacional e análise de movimentos humanos (Liu et al., 2019).

O modelo de *deep learning Gated Recurrent Unit* (GRU) é uma variação das redes neurais recorrentes projetadas para lidar com dados sequenciais, como séries temporais e movimentos corporais. Sua principal característica está na capacidade de capturar dependências de longo prazo em sequências por meio de mecanismos de portas (gate update e gate reset), que regulam o fluxo de informações relevantes e evitam problemas comuns de RNNs tradicionais, como o desaparecimento do gradiente (Cho et al., 2014a). No contexto da análise de movimentos no tênis, as GRUs permitem modelar a dinâmica temporal do saque, considerando a ordem e a evolução dos gestos executados.

Já a *Spatio-Temporal Graph Convolutional Network* (ST-GCN) é um modelo que combina convoluções em grafos com a dimensão temporal, sendo amplamente aplicada no reconhecimento de ações humanas a partir de sequências de poses extraídas por técnicas de visão computacional (Yan; Xiong; Lin, 2018). Nessa arquitetura, cada articulação do corpo é representada como um nó em um grafo, e as conexões entre elas são modeladas como arestas, permitindo capturar relações espaciais entre jun-

tas e também a variação desses padrões ao longo do tempo. Além disso, a ST-GCN possibilita a criação de subgrafos que representam partes específicas do corpo, como tronco, braços ou pernas, permitindo uma análise segmentada e detalhada dos movimentos. Essa característica torna a ST-GCN especialmente adequada para analisar gestos complexos, como o saque no tênis, em que tanto a coordenação espacial das articulações quanto a evolução temporal do gesto são determinantes para a qualidade do desempenho.

A complexidade do golpe, aliada à necessidade de capturar informações espaciais e temporais precisas, torna o tema desafiador e pouco explorado. Nesse contexto, trabalhos correlatos que investigam o uso de algoritmos de aprendizado de máquina e visão computacional em esportes de movimento contínuo servem de base teórica e metodológica para a presente pesquisa, oferecendo meios para compreender como diferentes abordagens computacionais.

Valim (2021) teve como objetivo a avaliação do movimento de saque de tênis de campo, por meio de processamento de imagens em vídeo. O autor desenvolve um software utilizando a linguagem de programação C++ que utiliza de bibliotecas como o OpenCV e outros softwares auxiliares. Com isso, seu programa foi capaz de mensurar os pontos-chaves articulares do corpo humano com base em registros de vídeo, possibilitando a identificação de falhas mecânicas e a geração de *feedbacks* analíticos sobre o movimento de saque. O estudo se destaca por apresentar uma abordagem técnica e detalhada, voltada para a obtenção de dados precisos sobre a movimentação dos atletas. Ainda que forneça informações relevantes, o trabalho foca mais na coleta e análise de medidas do corpo durante o movimento, sem explorar interpretações mais amplas ou contextuais sobre como esses movimentos são realizados ou compreendidos pelo praticante.

Oliveira (2018) teve como foco a comparação do uso de técnicas de aprendizado de máquina aplicados na análise biomecânica da corrida, com o objetivo de entender o corpo em movimento a fim de melhorar a performance e prevenir lesões usando ML. A autora investigou algoritmos de classificação, como o *k-Nearest Neighbors* (kNN) e o *Support Vector Machine* (SVM), aplicando-os a dados biomecânicos previamente tratados e organizados. Para isso, foi realizado um pré-processamento dos dados brutos com o intuito de extrair informações relevantes, transformando-as em atributos que pudessem ser analisados por meio de aprendizado su-

pervisionado. Esse trabalho representa uma contribuição importante para o uso de inteligência artificial no contexto da biomecânica, ao demonstrar que algoritmos de ML podem ser eficazes na identificação de padrões que antecedem lesões.

Além dos estudos acadêmicos que aplicam técnicas de visão computacional ou aprendizado de máquina à biomecânica, iniciativas do setor esportivo também apontam para o potencial da utilização de sistemas baseados em IA na identificação de padrões de movimento que não são facilmente perceptíveis a olho nu. Essa perspectiva destaca a possibilidade de tornar o treinamento mais eficiente, oferecendo análises contínuas, feedback imediato e suporte ao desenvolvimento técnico de atletas de diferentes níveis, inclusive fora do ambiente de treino ou na ausência de especialistas da área (Kinetec, 2024).

Com o uso das bibliotecas de algoritmos de ML, como o Pose Landmark Detection, a aplicação será capaz de analisar a biomecânica do movimento através de pontos específicos espalhados pelo corpo e classificar conforme as características chave do saque de tênis mencionadas anteriormente (Knudson; Luedtke; Faribault, 1994), comparando a imagem de vídeo a uma base de dados elaborada junto a especialistas regionais da área de tênis.

Apesar dos avanços observados nos trabalhos correlatos, nota-se uma lacuna significativa quanto à aplicação de modelos de aprendizado profundo voltados especificamente à análise biomecânica do saque no tênis de campo (Toledo; Marocolo, 2019). Enquanto pesquisas anteriores concentraram-se em algoritmos clássicos de ML ou em medições articulares obtidas por bibliotecas tradicionais de visão computacional, a literatura ainda carece de estudos que integrem informações espaciais e temporais de forma conjunta, especialmente em movimentos complexos como o saque.

O objetivo geral deste trabalho é comparar métodos de aprendizado de máquina aplicados à visão computacional para a identificação de falhas biomecânicas no movimento do saque de tênis. Para alcançar esse propósito, busca-se compreender a biomecânica envolvida no gesto esportivo, de modo a estabelecer uma base teórica sólida para a análise. Além disso, pretende-se elaborar uma base de dados rotulada e pública contribuindo para a pesquisa científica. Com isso será possível treinar dois modelos diferentes de *deep learning* e realizar a comparação de desempenho dos modelos aplicados, visando avaliar sua eficácia na detecção de

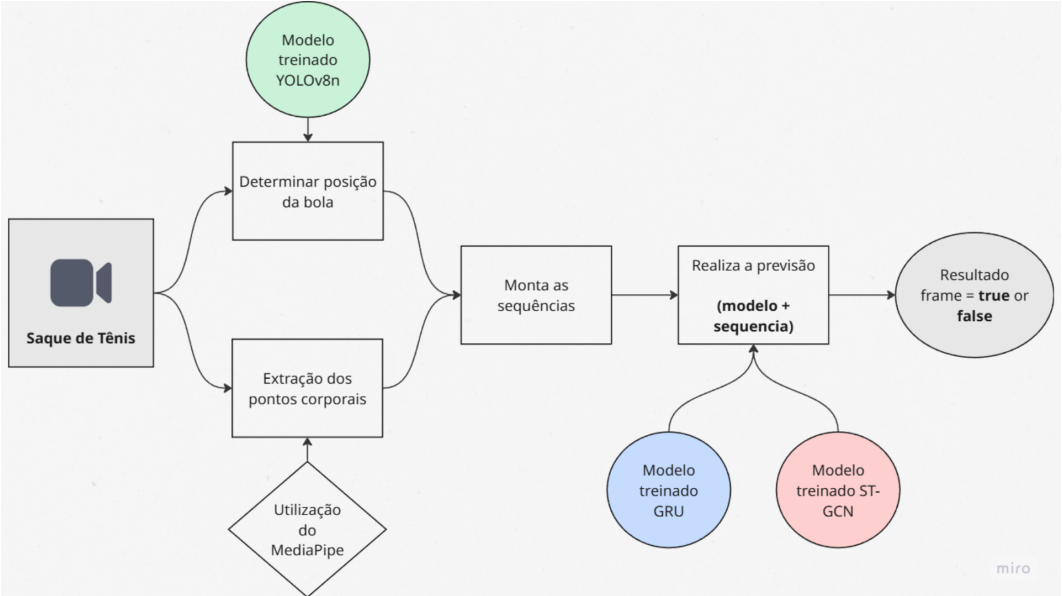
erros. Por fim, espera-se contribuir para diagnósticos técnicos mais precisos e para o avanço da análise biomecânica do tênis de campo.

A solução será capaz de captar os movimentos do atleta por meio de vídeos, extrair dados biomecânicos utilizando ferramentas como o MediaPipe e comparar esses dados com uma base de movimentos previamente cadastrada, a fim de detectar possíveis desvios ou falhas técnicas. A proposta visa contribuir com a democratização de análises técnicas que, até então, dependem de profissionais especializados e equipamentos de alto custo.

2 MATERIAIS E MÉTODOS

A presente pesquisa caracteriza-se por ser aplicada e de base tecnológica. Desenvolveu-se uma aplicação capaz de analisar biomecanicamente o movimento do saque no tênis de campo, utilizando dois algoritmos distintos de aprendizado de máquina supervisionado. Para o treinamento dos modelos, foi criada uma base de dados composta por gravações de atletas de tênis da região de Criciúma/SC. A arquitetura geral da aplicação está representada no diagrama apresentado na Figura 1, que ilustra as principais etapas do sistema e a interação entre seus componentes.

Figura 1 - Diagrama da aplicação



Fonte: Elaborado pelo autor.

A seleção dos algoritmos teve como objetivo comparar abordagens complementares para a análise de movimento. A rede GRU (*Gated Recurrent Unit*) foi escolhida por ser um modelo consolidado em tarefas de

séries temporais, oferecendo desempenho estável na detecção de padrões ao longo de sequências. Já o ST-GCN (*Spatial-Temporal Graph Convolutional Network*) representa uma abordagem mais recente aplicada à análise de movimentos humanos, pois opera diretamente sobre grafos que representam articulações e conexões corporais, possibilitando explorar simultaneamente as relações espaciais e temporais presentes na cadeia cinética.

2.1 GRAVAÇÃO DA BASE DE DADOS

Foram gravados 65 vídeos de seis atletas distintos executando o movimento de saque com o braço direito. O grupo foi composto por três atletas juvenis e por três professores de tênis atuantes da região de Criciúma/SC. Todas as filmagens foram realizadas mediante autorização prévia dos participantes, seguindo os procedimentos éticos estabelecidos para o uso de imagem em pesquisa acadêmica.

As filmagens foram realizadas utilizando um iPhone 15 Pro Max, com resolução 4K e taxa de 60 *frames* por segundo (fps). Cada atleta executou, em média, de oito a doze tentativas de saque, com duração aproximada de 4 a 7 segundos por vídeo. A câmera foi posicionada à direita da linha de base, no mesmo lado da quadra do sacador, a uma distância entre quatro e oito metros do atleta, conforme ilustrado na Figura 6 do Apêndice A. Essa disposição visou capturar adequadamente o plano de movimento e minimizar oclusões corporais durante a execução.

2.2 PRÉ-PROCESSAMENTO DOS VÍDEOS

Os vídeos gravados não passaram por nenhum tipo de tratamento ou filtragem de imagem. Não foram aplicadas técnicas de correção de cor, estabilização, recorte de quadros ou normalização de iluminação, visto que o objetivo principal era preservar as condições reais de execução do movimento. Assim, os arquivos de vídeo originais foram utilizados integralmente nas etapas seguintes de análise e extração de dados.

2.3 ROTULAÇÃO DA BASE DE DADOS

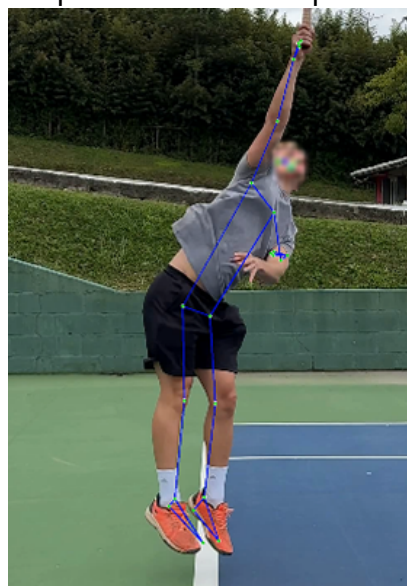
Para realizar a rotulação, foi utilizado um algoritmo em Python que, por meio da ferramenta de detecção de pontos corporais do MediaPipe, extraiu as coordenadas dos principais marcos do corpo em todos os *frames* dos vídeos, conforme ilustrado na Figura 7 do Apêndice A. Para a manipulação e armazenamento dos dados, foram utilizadas bibliotecas como pandas, para leitura e escrita de arquivos CSV, Numpy, para ope-

rações numéricas, e OS e Json para organização e gestão dos arquivos. Essa preparação dos dados permitiu estruturar o conjunto de informações necessário para o treinamento dos modelos de aprendizado de máquina implementados em PyTorch. Além disso, a biblioteca OpenCV foi empregada para processar os vídeos e extrair os *frames* individuais. Desta forma foi possível extrair todas as coordenadas dos principais marcos do corpo em todos os *frames* dos vídeos, conforme ilustrado na Figura 2.

Figura 2 - Comparação de pontos corporais no MediaPipe



(a) Pontos corporais no contato.



(b) Pontos corporais na pré-aceleração.

Fonte: Elaborado pelo autor.

Buscando incorporar o conhecimento técnico já consolidado na literatura esportiva, foram aplicadas uma série de regras propostas por Kovacs e Ellenbecker (2012), baseadas em seu modelo de cadeia cinética. O primeiro passo consistiu em reconhecer o início do movimento do saque e identificar cada um dos quadros correspondentes às posições: início, soltura, carregamento, pré-aceleração, contato, desaceleração e finalização. Para aumentar a precisão nessa detecção, foi utilizado um modelo adicional de aprendizado profundo específico para a identificação da posição da bola de tênis durante a execução do golpe.

2.3.1 Detecção da bola de tênis

A detecção de objetos em visão computacional pode ser realizada por diferentes técnicas. Entre as abordagens tradicionais testadas nesta pesquisa, destacam-se: métodos baseados em cor, como a limiarização no espaço HSV (*Hue, Saturation, Value*), métodos baseados em mo-

vimento, como a diferença entre *frames* e a segmentação por movimento (*Motion Segmentation*), e estratégias híbridas que combinam cor e movimento para reduzir falsos positivos. Foram realizados testes práticos, cada um avaliando uma dessas abordagens, utilizando vídeos de saque gravados previamente. Apesar disso, devido à grande variação de iluminação, enquadramento e qualidade dos vídeos, essas técnicas não apresentaram resultados consistentes.

Diante dessas limitações, optou-se por uma abordagem baseada em aprendizado profundo. Para isso, utilizou-se o YOLOv8 (*You Only Look Once*), um algoritmo de detecção de objetos baseado em *Convolutional Neural Networks* (CNN), otimizado para detecção de objetos pequenos (Ultralytics, 2025). O modelo escolhido para esta pesquisa foi a variante YOLOv8n, consolidada e adequada para bases de dados relativamente pequenas.

Antes do treinamento do modelo, foi necessário criar uma base de dados específica para a detecção da bola de tênis. Essa base de dados foi composta por 506 imagens. Para realizar as anotações, utilizou-se a ferramenta online e gratuita Roboflow, que permite a marcação manual de objetos em imagens. No total, foram demarcadas 506 imagens contendo a bola de tênis, gerando um conjunto de dados pronto para treinamento com o YOLOv8n.

O treinamento do modelo foi realizado utilizando estas 506 imagens, com 50 épocas, tamanho de lote igual a 16 e taxa de aprendizado de 0,01.

2.3.2 Atribuição dos rótulos na base de dados

Já com todas as coordenadas do corpo do atleta e da bola de tênis, é possível identificar as posições-chave desejadas. O objetivo inicial era trabalhar com todas as posições previstas na literatura. Entretanto, como só está sendo analisado um ângulo do movimento, não é possível captar todos os dados necessários para todos os *frames* do vídeo, como por exemplo o movimento de pré-aceleração, que o cotovelo direito do atleta acaba se posicionando atrás do ombro direito prejudicando a identificação de sua posição por parte do MediaPipe conforme Figura 2.

Além disso a posição de desaceleração e finalização também não foi analisada uma vez que cada atleta possui uma execução diferente para essas posições e a literatura não apresenta regras notáveis para elas. Por isso, nesta aplicação, foram utilizadas 4 posições-chave do saque de

tênis. São elas: a posição de preparação, posição de soltura, posição de carregamento ou troféu e posição de contato, conforme Figura 3.

Figura 3 - Fases analisadas



Fonte: Elaborado pelo autor.

A fim de detectar o *frame* exato de cada posição dos saques e realizar a aplicação das regras disponibilizadas nos estudos de cadeia cinética por Martin Caroline Touzard e Fourel (2024), foi criado um algoritmo em Python que percorre cada um dos vídeos e realiza comparações para detectar cada uma das posições.

Para identificar o *frame* da posição de preparação, é observado se o momento em que a mão esquerda do atleta ultrapassa a altura do nariz. Neste momento, é removido sessenta *frames* (dois segundos de vídeo em trinta *frames* por segundo), e armazenado como posição inicial. A posição inicial não tem regra, cada atleta tem uma posição própria. Essa posição se faz necessária para recortar o vídeo no momento exato em que o saque se inicia sem alimentar a base de dados com *frames* desnecessários.

Já a posição de soltura, é definida no momento em que a bola de tênis sai da mão esquerda do atleta. E para essa posição foram observados e comparados dois ângulos. O ângulo formado pelos nós 11, 13 e 15 (cotovelo esquerdo) que deve maior ou igual a cento e cinquenta graus no momento do lançamento. E o ângulo dos nós 13, 11 e 23 (ombro esquerdo), que deve ser entre 90° e 135° seguindo as recomendações do estudo de Kovacs e Ellenbecker (2012).

Para detectar a posição troféu no momento do saque, é realizado uma comparação do ângulo máximo de flexão dos joelhos, calculados através dos nós 24, 26 e 28 onde é identificado o ponto máximo de flexão quando a bola de tênis e a mão esquerda do atleta estiverem acima do

nó 0 (nariz). Se por quaisquer motivos não seja possível identificar a essa posição. É atribuído para essa posição o *frame* da posição de soltura adicionado mais vinte *frames*. Que é uma margem média de segurança entre o momento do arremesso até a posição de troféu. Nesta posição são aplicadas três comparações: o ângulo dos joelhos flexionados deve ser maior que quinze graus. A mão que realiza a soltura deve estar acima do nariz do atleta e o ombro direito abaixo do ombro esquerdo caracterizando a posição de troféu do saque de tênis.

Por fim, a posição de contato que é determinada a partir de uma lógica que identifica o momento em que a bola de tênis é atingida. Para essa posição são analisados três ângulos: o ângulo formado pelos nós 14, 12 e 24 (ombro direito) que deverá ser entre cinquenta e cinco e oitenta e cinco graus. Para os nós 14, 12 e 16 (cotovelo direito) é esperado que o ângulo seja menor que quinze graus. Além disso é esperado que a inclinação do tronco, observada através dos nós 11 e 13 em relação ao eixo Y seja entre 125° e 145° graus. E por fim é realizado uma comparação com a coordenada Y do nó 27 (pé esquerdo) com o *frame* da posição de troféu e contato identificando se houve a impulsão esperada para a realização do golpe, ou seja, uma diferença nessa coordenada indica um salto que é esperado na execução correta do movimento

Nota-se que, durante a criação da base de dados foi identificado uma variação muito grande para a inclinação do tronco esperada com a inclinação identificada nos atletas, o que prejudicaria o treinamento dos modelos. Logo foi realizado um ajuste junto à um profissional da área que altera a inclinação esperada entre 131° até 145° para 125° até 145°

Essas regras são aplicadas em todos os vídeos da base de dados e armazenadas em arquivos JSON, permitindo então que outros algoritmos manipulem essa informação gerando os *dataframes* direcionados para os dois modelos de *deep learnings* comparados na presente pesquisa.

2.4 TREINAMENTO E VALIDAÇÃO DOS MODELOS

Esta etapa descreve o processo de treinamento e validação dos modelos de aprendizado profundo empregados na pesquisa. O objetivo foi avaliar o desempenho das arquiteturas selecionadas na classificação das fases do movimento do saque, considerando diferentes estratégias de representação e processamento dos dados.

2.4.1 Criação dos Modelos

Para o treinamento do modelo GRU, os dados coletados foram primeiramente organizados em um único arquivo CSV, contendo informações de *frames* e coordenadas dos nós corporais, juntamente com os rótulos correspondentes às fases do movimento. As colunas utilizadas para o treinamento incluem: *video_id*, *frame_id*, *fase*, *label*, seguidas pelas coordenadas (x, y, z) de cada nó corporal.

O modelo GRU foi implementado utilizando o *framework* PyTorch, configurado com uma camada oculta contendo 64 neurônios, função de ativação ReLU e otimizador Adam com taxa de aprendizado inicial de 0,001. O treinamento ocorreu por 100 épocas, com *batch size* de 8 e sequências de 30 *frames*, empregando a função de perda *Binary Cross-Entropy Loss* (BCELoss), adequada à classificação binária das fases do movimento. Foram utilizados 80% dos dados para treinamento e 20% para validação. As entradas do modelo correspondem às sequências temporais de coordenadas normalizadas, permitindo que a rede capture a dinâmica do movimento ao longo dos *frames*.

O modelo ST-GCN segue uma abordagem distinta, baseada na análise de grafos. Para isso, foram definidas as arestas entre os nós do corpo humano de acordo com os *landmarks* do MediaPipe Pose, formando a estrutura de grafo completa. Além disso, foram criados dois subgrafos específicos para análise detalhada de regiões relevantes para o saque: a região do quadril (*landmarks* 11, 12, 23 e 24) e o braço direito (*landmarks* 12, 14, 16 e 24). Essa segmentação permite examinar com maior precisão a relação entre movimentos do tronco e do membro superior.

O modelo ST-GCN foi implementado com duas camadas de convolução espaço-temporal, *kernel* temporal igual a 9, função de ativação ReLU e otimização via Adam com taxa de aprendizado de 0,001, durante 100 épocas e *batch size* igual a 8. A função de perda utilizada foi a *CrossEntropyLoss*, adequada à classificação multiclasse. O particionamento dos dados seguiu a mesma proporção de 80% para treinamento e 20% para validação.

O treinamento dos modelos foi realizado em um computador com processador Intel i5 de 13ª geração, 16 GB de RAM, 1 TB de HD e 128 GB de SSD, equipado com placa de vídeo GeForce RTX 3060 Ti, utilizando o sistema operacional Windows 10.

2.5 DESENVOLVIMENTO DA APLICAÇÃO

A aplicação desenvolvida tem como objetivo realizar a análise biomecânica do saque a partir de sequências de *frames* extraídos de vídeos do atleta. Cada sequência passa por um processo de pré-processamento, no qual são obtidas as coordenadas tridimensionais normalizadas dos pontos corporais detectados pelo MediaPipe. Essas coordenadas, correspondentes aos *landmarks* corporais, são então organizadas no formato exigido por cada modelo de aprendizado de máquina.

No caso do modelo GRU, cada sequência é composta por 30 *frames* consecutivos, sendo convertida em um vetor temporal no formato (T, C), onde T representa o número de *frames* e C o número de características (coordenadas x, y, z de cada ponto corporal). Esse vetor é então fornecido à rede recorrente, que processa a sequência ao longo do tempo e realiza a classificação das fases do movimento.

Para o modelo ST-GCN, os mesmos *frames* são convertidos em uma estrutura de grafo, onde cada nó representa uma articulação do corpo e cada aresta representa uma conexão biomecânica. A entrada do modelo segue o formato tridimensional (N, C, T, V), em que N é o número de amostras, C o número de canais (coordenadas x, y, z), T o número de *frames* e V o número de vértices (articulações). Essa representação permite que a rede extraia informações tanto espaciais das relações entre articulações quanto temporais da evolução do movimento ao longo do tempo.

A aplicação foi implementada em Python, o sistema recebe um vídeo de entrada, realiza o pré-processamento e executa a inferência no modelo escolhido GRU e ST-GCN. Ao final, é gerado um novo vídeo com a reprodução do movimento analisado, exibindo em tempo real o resultado da classificação em cada *frame*, indicando se a execução está correta ou incorreta. Essa abordagem permite comparar o desempenho e a eficiência de ambas as arquiteturas, fornecendo uma avaliação detalhada da execução técnica do atleta.

2.5.1 Validação dos Modelos

A validação dos modelos teve como objetivo avaliar a capacidade de generalização e identificar possíveis sobreajustes (*overfitting*). Para isso, foram utilizadas métricas como acurácia, precisão, sensibilidade e F1-score, obtidas a partir do conjunto de validação.

Além dessas métricas, a matriz de confusão foi empregada para analisar detalhadamente o desempenho dos modelos em cada classe. A

matriz de confusão permite visualizar a quantidade de acertos e erros de classificação, identificando os casos de falso positivo (quando o modelo classifica erroneamente uma instância como pertencente a uma classe) e falso negativo (quando o modelo deixa de identificar uma instância que realmente pertence à classe) (Monard; Baranauskas, 2003). Essa análise é essencial para compreender quais fases do movimento apresentam maior dificuldade de classificação e orientar ajustes futuros nos modelos.

Os códigos fontes e a base de dados deste trabalho estão disponíveis no repositório do GitHub³.

3 RESULTADOS E DISCUSSÃO

A presente seção tem como objetivo apresentar e discutir os resultados obtidos durante o treinamento e a validação dos modelos desenvolvidos, bem como a análise prática realizada por meio da aplicação implementada. Foram avaliadas métricas de desempenho, curvas de aprendizado e resultados qualitativos obtidos a partir da aplicação dos modelos em vídeos reais de execução do saque no tênis.

3.1 DESEMPENHO DOS MODELOS

Durante o treinamento, foram registradas as métricas de acurácia, precisão, sensibilidade e *F1-score* para ambos os modelos, possibilitando a comparação de desempenho entre as abordagens, conforme Tabela 1.

Tabela 1 - Tabela de dados do treinamento

	GRU	ST-GCN
Acurácia (<i>accuracy</i>)	0,9864	0,9112
Precisão (<i>precision</i>)	0,8412	0,9130
Sensibilidade (<i>recall</i>)	0,8707	0,9302
<i>F1-score</i>	0,8390	0,9302

Fonte: Elaborado pelo autor.

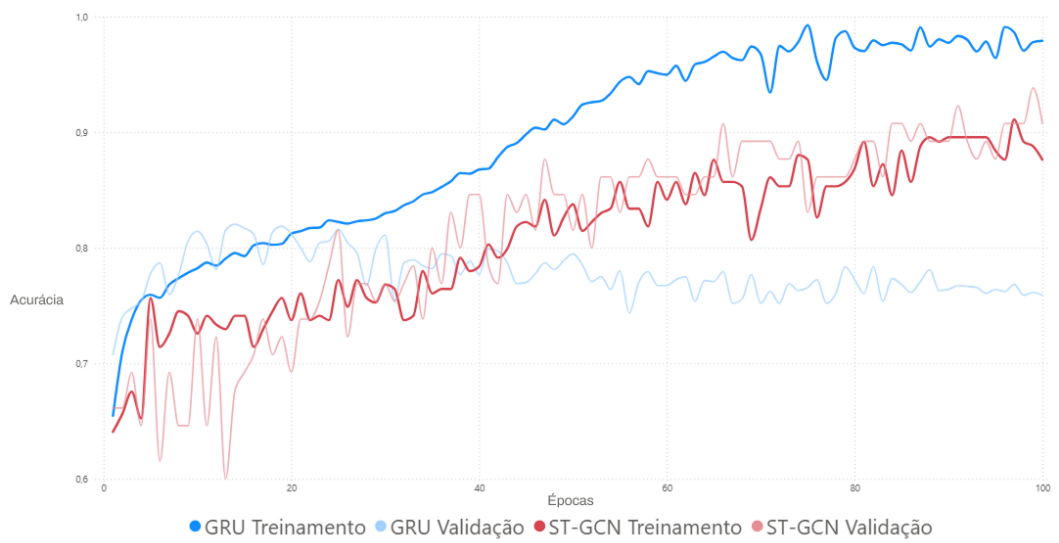
As curvas de aprendizado apresentaram convergência estável, evidenciando a capacidade de ambos os modelos em generalizar os padrões de movimento após um número adequado de épocas. A figura 4 ilustra o gráfico de acurácia obtido no treinamento dos modelos GRU e ST-GCN.

Ao comparar o desempenho dos modelos, observa-se que a

³<https://github.com/nicolauuu/tennis-serve-biomechanics-ai>

rede GRU obteve a maior acurácia global, atingindo 0,9864 após 100 épocas. No entanto, esse valor sugere uma menor capacidade de generalização e indícios de *overfitting* dado a complexidade do problema.

Figura 4 - Gráfico de Acurácia



Fonte: Elaborado pelo autor.

Observando as matrizes de confusão na Tabela 2, nota-se que o modelo GRU apresenta uma quantidade considerável de predições incorretas, especialmente de falsos negativos, ou seja, casos em que saques corretos foram classificados como incorretos. Esse comportamento impacta diretamente o desempenho prático do modelo, refletindo-se nas métricas complementares, que registram precisão máxima de 0,8412, *F1-score* de 0,8390 e sensibilidade de 0,8707. Apesar de demonstrar bom desempenho geral, o modelo ainda revela instabilidade na distinção entre as classes.

Tabela 2 - Matriz de confusão do modelo GRU

	Predito Negativo (0)	Predito Positivo (1)
Real Negativo (0)	3	17
Real Positivo (1)	3	42

Fonte: Elaborado pelo autor.

Tabela 3 - Matriz de confusão do modelo ST-GCN

	Predito Negativo (0)	Predito Positivo (1)
Real Negativo (0)	17	5
Real Positivo (1)	1	42

Fonte: Elaborado pelo autor.

Em contrapartida, o modelo ST-GCN alcançou uma acurácia máxima de 0,9112, mas apresentou maior estabilidade durante a validação, confirmando sua eficácia na captura de relações espaço-temporais e na representação dos padrões biomecânicos do gesto técnico. A matriz de confusão observada na Tabela 3, demonstra um equilíbrio expressivo entre acertos e erros, reforçado pelos valores de precisão, sensibilidade e *F1-score* iguais a 0,9302, indicando desempenho consistente e balanceado entre as classes. A utilização dos subgrafos do quadril e do braço direito contribuiu para uma melhor representação dos segmentos corporais mais relevantes, aprimorando a sensibilidade do modelo às variações técnicas do movimento.

3.2 ANÁLISE DOS TESTES COM A APLICAÇÃO

Após o treinamento, a aplicação foi testada utilizando 12 vídeos distintos que não fizeram parte da etapa de treinamento. Esses vídeos incluíram execuções variadas do movimento de saque, contemplando tanto tentativas tecnicamente corretas quanto erros comuns, como posicionamento inadequado dos pés, desalinhamento do quadril e trajetória incorreta do braço. Cada um dos 12 vídeos foi processado separadamente pelos dois modelos avaliados (GRU e ST-GCN), totalizando 24 análises.

Para fins de ilustração e comparação qualitativa, dois vídeos foram selecionados como exemplos representativos: um contendo um saque tecnicamente correto e outro apresentando falhas biomecânicas evidentes. Ambos foram analisados por cada modelo para observar diferenças no comportamento das predições. O modelo ST-GCN demonstrou maior precisão na identificação de erros, destacando incorreções no alinhamento do quadril e na trajetória do braço no momento do contato. O modelo GRU apresentou bom desempenho geral, porém com menor sensibilidade para variações sutis entre fases consecutivas do movimento.

Esses resultados indicam que o ST-GCN se mostra mais robusto para o conjunto de dados utilizado, principalmente pela representação gráfica do corpo e pela capacidade de análise conjunta das dimensões espaciais e temporais. No entanto, o modelo GRU, por se basear em redes recorrentes profundas, apresenta um potencial de desempenho elevado para contextos com maior volume de dados, podendo identificar padrões complexos não perceptíveis visualmente o que é uma característica dos modelos de aprendizado profundo.

3.3 SÍNTESE DOS RESULTADOS

Os resultados obtidos demonstram que ambos os modelos são capazes de reconhecer padrões biomecânicos associados ao saque no tênis, ainda que com diferentes níveis de precisão. O modelo ST-GCN apresentou desempenho mais consistente com o volume de dados disponível, enquanto o GRU mostrou maior potencial de escalabilidade.

Conforme ilustrado na Figura 5 em um dos *frames* analisados, o GRU classificou incorretamente a execução como falsa, enquanto o ST-GCN identificou corretamente o movimento como verdadeiro. Essa comparação evidencia a capacidade do ST-GCN de capturar relações espaço-temporais mais detalhadas, particularmente em movimentos complexos, e reforça a importância da análise segmentada por subgrafos para aumentar a precisão do modelo.

Figura 5 - Comparação dos modelos ST-GCN e GRU



(a) Analisador do modelo ST-GCN.

(b) Analisador do modelo GRU.

Fonte: Elaborado pelo autor.

Os resultados alcançados reforçam o potencial das redes neurais profundas na análise biomecânica do saque, ampliando as possibi-

dades de interpretação do movimento em relação a abordagens baseadas em algoritmos clássicos, como os aplicados por (Oliveira, 2018) na análise da corrida. Enquanto técnicas tradicionais tendem a se concentrar em parâmetros isolados, o modelo ST-GCN permite capturar relações espaço-temporais mais complexas entre as articulações, oferecendo uma compreensão mais integrada da dinâmica corporal. Essa capacidade contribui para uma análise mais precisa e contextualizada dos gestos esportivos, especialmente em movimentos de maior complexidade técnica, como o saque no tênis.

Quando comparados aos achados de (Cho et al., 2014b), observa-se que o desempenho do modelo baseado em GRU no presente estudo segue a tendência descrita pelos autores: redes recorrentes são eficazes na captura de dependências temporais, mas apresentam limitações em representar relações estruturais complexas entre múltiplas articulações. No contexto do saque, que envolve grande variabilidade espacial o GRU demonstrou dificuldades em identificar padrões espaciais finos, resultando em menor precisão na detecção de erros biomecânicos.

Por outro lado, os resultados obtidos com o modelo ST-GCN alinham-se diretamente ao que foi reportado por (Yan; Xiong; Lin, 2018), que destacam o potencial das convoluções em grafos para capturar simultaneamente padrões espaciais e temporais do movimento humano. Assim como observado no estudo original, o ST-GCN empregado neste trabalho demonstrou maior robustez na identificação de inconsistências técnicas do saque, conseguindo mapear relações entre articulações distantes e variações sutis de postura ao longo da execução.

4 CONCLUSÃO

Os resultados obtidos demonstram que o uso de modelos baseados em aprendizado de máquina apresenta grande potencial para compreender os padrões biomecânicos do gesto. Tanto o modelo GRU quanto o ST-GCN foram capazes de extrair características relevantes das execuções analisadas, evidenciando a viabilidade dessas técnicas como ferramentas de apoio à análise esportiva.

A análise biomecânica do saque no tênis apresentou desafios significativos devido à complexidade da cadeia cinética do movimento, uma vez que parte do corpo do atleta pode ficar parcialmente oculto quando observado a partir de um único ângulo. Essa limitação dificultou a extração completa das informações biomecânicas necessárias à modelagem impos-

sibilitando que todas as fases do movimento sejam contempladas neste estudo.

Um dos principais avanços obtidos neste trabalho foi a criação de uma base de dados biomecânica, construída a partir de gravações de atletas de tênis. Essa base constitui uma contribuição inédita e relevante para futuras pesquisas, uma vez que oferece dados rotulados e contextualizados para o estudo do saque no tênis, o que até então era uma limitação observada na literatura.

A comparação entre os modelos GRU e ST-GCN permitiu cumprir o objetivo central deste trabalho, que consistiu em avaliar diferentes abordagens de aprendizado de máquina aplicadas à detecção de falhas biomecânicas no movimento do saque de tênis. Ambos os modelos foram treinados e analisados a partir da base de dados desenvolvida, possibilitando verificar suas capacidades de reconhecimento dos padrões de movimento e de identificação de inconsistências técnicas. Essa etapa experimental demonstrou a viabilidade das duas arquiteturas, evidenciando os pontos fortes e limitações de cada uma, bem como sua aplicabilidade prática no apoio ao treinamento esportivo e na prevenção de lesões.

O trabalho reforça o potencial da visão computacional como uma ferramenta de apoio à análise esportiva, abrindo caminho para aplicações futuras baseadas em redes neurais em contextos ainda mais amplos. As perspectivas de continuidade incluem o aprimoramento do sistema por meio da utilização de múltiplos ângulos de câmera, possibilitando uma análise mais completa do movimento de saque, bem como de outros gestos técnicos do tênis, como o *forehand* e o *backhand*. Além disso, abre-se a possibilidade para o desenvolvimento de um aplicativo móvel capaz de fornecer *feedback* em tempo real ao atleta durante o treino, tornando a análise biomecânica mais acessível e interativa.

REFERÊNCIAS

CHAPELLE, O.; SCHÖLKOPF, B.; ZIEN, A. **Semi-Supervised Learning**. Cambridge, Massachusetts; London, England: The MIT Press, 2009. ISBN 978-0262033589.

CHO, K. et al. Learning phrase representations using rnn encoder-decoder for statistical machine translation. In: **Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)**. [s.n.], 2014. p. 1724–1734. Acessado em: 02 dec. 2025. Disponível em: <<https://aclanthology.org/D14-1179>>.

CHO, K. et al. Learning phrase representations using rnn encoder–decoder for statistical machine translation. In: **Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)**. [s.n.], 2014. p. 1724–1734. Acessado em: 02 dec. 2025. Disponível em: <<https://aclanthology.org/D14-1179>>.

FANTOURA, F. **Tênis para Todos**. [S.l.]: Ulbra, 2003.

FEDERATION, I. T. **ITF Global Tennis Report 2019**. 2019. Disponível em: <<http://itf.uberflip.com/i/1169625-itf-global-tennis-report-2019-overview/11>>. Acessado em: 6 abr. 2025.

GOOGLE. **Artificial Intelligence**. 2024. Disponível em: <https://ai.google.dev/edge/mediapipe/solutions/vision/pose_landmarker?hl=pt-br>. Acessado em: 28 out. 2024.

ISHIZAKI, M. T.; CASTRO, M. **Tênis: Aprendizagem e Treinamento**. 1. ed. São Paulo: Phorte, 2006.

KINETEC. **Impacto da Inteligência Artificial na Fisioterapia**. 2024. Disponível em: <<https://kinetec.com.br/new/impacto-da-inteligencia-artificial-na-fisioterapia/>>. Acessado em: 28 out. 2024.

KNUDSON, D.; LUEDTKE, D.; FARIBAULT, J. **How to analyze the serve**. Strategies, 1994, v. 7, n. 8, p. 19–22.

KOVACS, M.; ELLENBECKER, T. **An 8-stage model for evaluating the tennis serve**. PMCID, 2012, v. 5, n. 7, p. 2555–2565.

LIU, Z. et al. **Deep learning-based human motion prediction considering context awareness for human-robot collaboration in manufacturing**. Procedia CIRP, 2019, v. 81, p. 17–22. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S2212827119306948>>. Acessado em: 12 out. 2025.

MARTIN CAROLINE TOUZARD, P.; FOUREL, L. **The use of biomechanical analysis to help reduce serve-related injuries**. Aspetar Sports Medicine Journal, 2024, v. 13. Targeted Topic: Sports Medicine in Tennis.

MONARD, M. C.; BARANAUSKAS, J. A. **Conceitos sobre aprendizado de máquina**. Sistemas Inteligentes - Fundamentos e Aplicações, 2003, v. 1, n. 1.

OLIVEIRA, M. de. **Reconhecimento de padrões em Biomecânica de Corrida usando Aprendizado de Máquina**. 2018. Dissertação (Mestrado em Ciência da Computação) – Universidade Federal de Minas Gerais, MG.

TOLEDO, H. C.; MAROCOLO, M. **Ciência e tecnologia aliadas ao esporte**. Arq Cien Esp, 2019, v. 7, n. 2, p. 55–56. Disponível em: <<https://seer.uftm.edu.br/revistaeletronica/index.php/aces/article/download/4743/4541>>. Acessado em: 19 out. 2024.

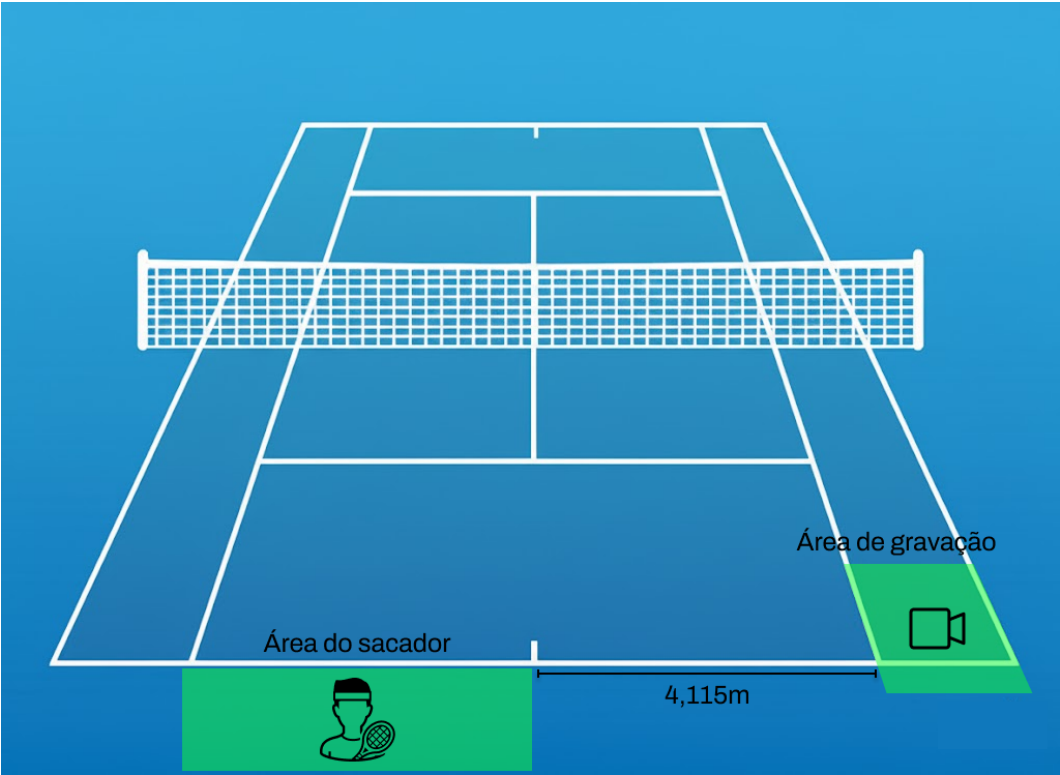
ULTRALYTICS. **YOLOv8: State-of-the-Art Computer Vision Model**. 2025. Disponível em: <<https://yolov8.com/>>. Acessado em: 31 out. 2025.

VALIM, B. **Análise do movimento do saque no tênis de campo por meio de processamento de imagem - diagnóstico visando à melhoria de desempenho**. Dissertação (Dissertação (Mestrado em Engenharia Elétrica)) — Universidade Federal de São Carlos, Centro de Ciências Exatas e de Tecnologia, São Carlos, 2021. 150 f.

YAN, S.; XIONG, Y.; LIN, D. Spatial temporal graph convolutional networks for skeleton-based action recognition. In: **Proceedings of the AAAI Conference on Artificial Intelligence**. [s.n.], 2018. v. 32, n. 1, p. 7444–7452. Acessado em: 02 dec. 2025. Disponível em: <<https://arxiv.org/abs/1801.07455>>.

A APÊNDICE

Figura 6 - Áreas da quadra



Fonte: Elaborado pelo autor.

Figura 7 - Pontos corporais MediaPipe

