



UNIVERSIDADE FEDERAL DE SANTA CATARINA
CENTRO DE CIÊNCIAS, TECNOLOGIAS E SAÚDE DO CAMPUS ARARANGUÁ
CURSO DE GRADUAÇÃO EM ENGENHARIA DE COMPUTAÇÃO

Jayor Nesi Teixeira

**Classificação de Golpes de Karate através de Técnicas de Visão
Computacional**

Araranguá
2024

Jayor Nesi Teixeira

**Classificação de Golpes de Karate através de Técnicas de Visão
Computacional**

Trabalho de Conclusão de Curso do Curso de Graduação em Engenharia de Computação do Centro de Ciências, Tecnologias e Saúde do Campus Araranguá da Universidade Federal de Santa Catarina para a obtenção do título de Bacharel em Engenharia de Computação.

Orientador: Prof. Antonio Carlos Sobieranski, Dr.

Araranguá

2024

Ficha catalográfica gerada por meio de sistema automatizado gerenciado pela BU/UFSC.
Dados inseridos pelo próprio autor.

TEIXEIRA, JAYOR NESI

Classificação de Golpes de Karate através de Técnicas de
Visão Computacional / JAYOR NESI TEIXEIRA ; orientador,
ANTONIO CARLOS SOBIERANSKI, 2024.

46 p.

Trabalho de Conclusão de Curso (graduação) -
Universidade Federal de Santa Catarina, Campus Araranguá,
Graduação em Engenharia de Computação, Araranguá, 2024.

Inclui referências.

1. Engenharia de Computação. 2. VISÃO COMPUTACIONAL. 3.
KARATE. I. SOBIERANSKI, ANTONIO CARLOS. II. Universidade
Federal de Santa Catarina. Graduação em Engenharia de
Computação. III. Título.

Jayor Nesi Teixeira

**Classificação de Golpes de Karate através de Técnicas de Visão
Computacional**

Este Trabalho de Conclusão de Curso foi julgado adequado para obtenção do Título de “Bacharel em Engenharia de Computação” e aprovado em sua forma final pelo Curso de Graduação em Engenharia de Computação.

Araranguá, 29 de novembro de 2024.

Prof. Jim Lau, Dr.
Coordenador do Curso

Banca Examinadora:

Prof. Antonio Carlos Sobieranski, Dr.
Orientador

Prof. Anderson Luiz Fernandes Perez, Dr.
Avaliador
Universidade Federal de Santa Catarina

Prof. Rodrigo Vinícius Mendonça Pereira,
Dr.
Avaliador
Universidade Federal de Santa Catarina

Este trabalho é dedicado aos meus colegas de classe e aos meus queridos pais.

AGRADECIMENTOS

A conclusão deste trabalho só foi possível graças ao apoio e contribuição de diversas pessoas, às quais expresso minha mais profunda gratidão.

Primeiramente, agradeço a Deus, pela força, sabedoria e bênçãos concedidas ao longo desta jornada, permitindo-me superar os desafios e alcançar este objetivo. Aos meus pais, pelo amor, paciência e por sempre me incentivarem a buscar meus objetivos com dedicação e perseverança. À minha equipe de Karate, incluindo minha esposa, minha irmã e minha cunhada, que colaboraram diretamente em uma etapa importante deste projeto. Ao meu orientador, pelo suporte técnico, pela orientação valiosa e por acreditar no potencial deste trabalho desde o início.

Por fim, à minha esposa, pelo amor incondicional, pela paciência nos momentos mais desafiadores e por ser minha maior fonte de apoio e inspiração. A todos que, direta ou indiretamente, contribuíram para a realização deste projeto, deixo aqui o meu mais sincero agradecimento.

*“Karate não é apenas uma arte de combate,
mas um estudo constante sobre si mesmo
e o mundo ao redor.
(Gichin Funakoshi, 1935)*

RESUMO

O *Karate* é uma modalidade esportiva de combate na qual dois atletas buscam o maior número de pontos entre si enquanto lutam em um tatame de área 8mx8m, sendo caracterizado como uma luta que contém chutes e socos de variadas direções e ângulos. Com a participação do *Karate* como modalidade Olímpica em Tóquio 2020, esta arte marcial ganhou destaque, recebendo alto investimento em todo o mundo através de governos e confederações. O processo de arbitragem consiste na análise visual dos participantes e atribuição de pontos e faltas que acontecem dentro da área delimitada do tatame. No entanto, dado que o processo de análise pode apresentar ambiguidade, viu-se a necessidade de utilização de tecnologia, assim como já ocorre em outros esportes, para auxiliar no processo de tomada de decisão de arbitragem. As tecnologias podem ser caracterizadas com o simples uso de vídeo, como já ocorre no *Video Review* (VR), ou sistemas computacionais de processamento de imagem para a interpretação de ações na imagem. Dados os recentes avanços nas áreas de visão computacional e inteligência artificial, o presente trabalho possui como finalidade propor tecnologias para o auxílio da tomada de decisão no processo de arbitragem e treinamento, coletando dados e analisando o desempenho dos atletas por vídeo. Foram utilizadas redes neurais convolucionais (CNNs) associadas com técnicas de reconhecimento de padrões para a extração das características e detecção dos *keypoints* em vídeo dos competidores, um dataset contendo 4235 anotações foi especificamente construído a partir de situações reais do esporte em determinados golpes selecionados. Entre os modelos comparados (*yolov8n-pose*, *yolov8s-pose*, *yolov8m-pose*, *yolov8l-pose*, *yolov8x-pose*), o *yolov8n-pose* obteve os melhores índices de eficiência, com 7.16 FPS e uma latência média por quadro de 0,1166 segundos, tornando-se uma opção adequada para aplicações em tempo real. Entretanto, o modelo *yolov8m-pose*, embora mais lento com 2.08 FPS e 0.4561 segundos de latência por quadro, apresentou uma combinação equilibrada entre desempenho e precisão, destacando-se como uma tecnologia promissora de classificação de padrões para o *Karate*, capaz de lidar com tarefas mais complexas de detecção de pose.

Palavras-chave: Visão Computacional. Artes Marciais. Reconhecimento de Ações Humanas. Análise de Vídeos Esportivos. *Karate*.

ABSTRACT

The *Karate* is a combat sport in which two athletes aim to score the highest number of points against each other while competing on an 8x8m tatami mat. It is characterized as a fight involving kicks and punches from various directions and angles. With the inclusion of *Karate* as an Olympic sport at Tokyo 2020, this martial art gained prominence, receiving significant investments worldwide through governments and federations. The arbitration process involves visually analyzing participants and assigning points and penalties for actions occurring within the defined tatami area. However, since this analysis process can present ambiguities, the use of technology has become necessary to assist in decision-making during arbitration, as is already the case in other sports. Technologies can range from the simple use of video, as seen with *Video Review* (VR), to computational image processing systems for interpreting actions in the video. Given recent advancements in computer vision and artificial intelligence, this study aims to propose technologies to assist in decision-making during arbitration and training processes, collecting data and analyzing athletes' performance through video. Convolutional Neural Networks (CNNs) were used in conjunction with pattern recognition techniques to extract features and detect *keypoints* in videos of competitors. A dataset containing 4,235 annotations was specifically built based on real-life situations in the sport, focusing on selected moves. Among the compared models (*yolov8n-pose*, *yolov8s-pose*, *yolov8m-pose*, *yolov8l-pose*, *yolov8x-pose*), *yolov8n-pose* achieved the highest efficiency scores, with 7.16 FPS and an average frame latency of 0.1166 seconds, making it a suitable option for real-time applications. However, the *yolov8m-pose* model, while slower with 2.08 FPS and 0.4561 seconds of frame latency, offered a balanced combination of performance and accuracy, standing out as a promising technology for pattern classification in *Karate*, capable of handling more complex pose detection tasks.

Keywords: Computer Vision. Martial Arts. Human Action Recognition. Sports Video Analysis. Karate.

LISTA DE FIGURAS

Figura 1	–	<i>Tabela Comparativa de Métodos de Reconhecimento de Ações.</i>	17
Figura 2	–	Uso da visão computacional em partidas da NBA	19
Figura 3	–	Estágios e Camadas de uma Rede Neural Convolutacional.	20
Figura 4	–	Exemplo de separação de classes usando SVM.	22
Figura 5	–	Ilustração do <i>Grid Search</i> para Otimização de Hiperparâmetros.	23
Figura 6	–	<i>YOLOv8 keypoints.</i>	24
Figura 7	–	Fluxo computacional da abordagem proposta.	25
Figura 8	–	Golpes de Karate desempenhados para composição do Dataset	26
Figura 9	–	Vídeos Gravados.	27
Figura 10	–	Aplicação para anotação dos golpes.	28
Figura 11	–	Desempenho dos Melhores Modelos SVM.	29
Figura 12	–	Distribuição de Amostras por Golpes no Dataset Produzido.	33
Figura 13	–	<i>Keypoints</i> Detectados durante execução do Sistema.	34
Figura 14	–	Distribuição por faixa etária.	35
Figura 15	–	Tempo de Treinamento.	36
Figura 16	–	Graduações dos Árbitros.	36
Figura 17	–	Classificação dos Árbitros.	37
Figura 18	–	Tempo de Experiência em Arbitragem.	37
Figura 19	–	Precisão na Identificação dos Golpes.	38
Figura 20	–	Utilizariam o Sistema.	38
Figura 21	–	Influência do Sistema na Arbitragem.	39

LISTA DE TABELAS

Tabela 1 – Critérios de Exclusão	15
Tabela 2 – Resultados da busca e seleção de estudos nas principais bases de dados	16
Tabela 3 – Tabela de teclas e seus significados.	31
Tabela 4 – Tabela reduzida com valores de pontos e classes.	31
Tabela 5 – Descrição dos Golpes no Dataset Produzido	32
Tabela 6 – Métricas de Desempenho no Treinamento e Teste	34

SUMÁRIO

1	INTRODUÇÃO	12
2	TRABALHOS CORRELATOS	15
3	FUNDAMENTAÇÃO TEÓRICA	18
3.1	<i>O KARATE</i>	18
3.2	VISÃO COMPUTACIONAL	19
3.3	REDES NEURAS CONVOLUCIONAIS (CNNS)	19
3.4	SVMS E <i>GRID SEARCH</i>	22
3.5	RECONHECIMENTO DE AÇÕES HUMANAS	23
4	METODOLOGIA PROPOSTA	25
4.1	GRAVAÇÃO DE VÍDEOS	25
4.2	DETECÇÃO DOS <i>KEYPOINTS</i> E ELABORAÇÃO DO CON- JUNTO DE DADOS	27
4.3	TREINAMENTO COM SVM	28
4.4	VALIDAÇÃO DOS GOLPES	29
5	RESULTADOS E AMBIENTE EXPERIMENTAL	30
5.1	FERRAMENTA DE ANOTAÇÃO DE VÍDEOS	30
5.1.1	ESTRUTURA E FUNCIONALIDADES	30
5.1.2	PROCESSO DE OPERAÇÃO	30
5.2	DATASET PRODUZIDO	32
5.3	RESULTADOS EXPERIMENTAIS	33
5.4	VALIDAÇÃO QUALITANTITATIVA	34
5.4.1	PERFIL DOS RESPONDENTES	35
5.4.2	AVALIAÇÃO DO SISTEMA	37
6	DISCUSSÃO, CONCLUSÃO E TRABALHOS FUTUROS	40
6.1	DISCUSSÃO	40
6.2	CONCLUSÃO	40
6.3	TRABALHOS FUTUROS	41
	REFERÊNCIAS	42

1 INTRODUÇÃO

O *Karate* é uma arte marcial japonesa, desenvolvida na ilha de *Okinawa*, ao sul do Japão, por volta do século XVIII (KARATE DO BRASIL, 2024). A palavra *Karate* tem como significado “mãos vazias”, devido ao uso do próprio corpo como arma de defesa pessoal. O *karate* moderno possui um sistema de luta com regras bem definidas para atribuição de pontos nos combates, permitindo que qualquer pessoa possa praticá-lo e competí-lo.

A principal modalidade competitiva de *Karate*, o *Kumite* (luta), é disputada entre 2 atletas com um limite de tempo de 3 minutos, onde o vencedor é o que estiver com maior número de pontos ao fim do tempo estabelecido. Nele, existem regras específicas para que um golpe seja válido, e certas restrições também são impostas para evitar lesões nos atletas. Golpes com força excessiva e abaixo da linha da cintura são penalizados, por exemplo. Além das restrições especificadas anteriormente, os golpes de *karate* devem possuir uma série de critérios para que seja considerado pontuável: boa forma, atitude esportiva, vigor, tempo e distância corretos e estado de alerta durante e após o golpe (WORLD KARATE FEDERATION, 2024a). O combate é avaliado por cinco juízes, dos quais quatro ficam dispostos nos cantos do tatame e são responsáveis pela atribuição dos pontos, e o quinto medeia a luta confirmando os pontos, atribuindo penalidades, entre outras funções.

No entanto, embora as pontuações sejam obtidas de forma objetiva através da análise de golpes executados corretamente e com o impacto no oponente, existe uma subjetividade inerente ao processo de julgamento. Mais de um árbitro em geral analisa a execução do golpe para obtenção de consenso, e em competições o número de áreas de combate pode variar de quatro a até mais de dez. Em competições locais e em algumas competições internacionais de atletas de base, o número de árbitros disponíveis não atinge o mínimo aceitável e adaptações acabam sendo feitas, como a redução de cinco para três árbitros em um combate (WORLD KARATE FEDERATION, 2024b). Isto reduz a precisão da anotação de pontos, visto que não há uma cobertura de todos os ângulos de visão da luta para analisar os golpes desferidos.

Como tentativa de melhorar a subjetividade na arbitragem em esportes, não somente exclusivas ao *karate*, recursos adicionais tecnológicos vêm sendo utilizados nos últimos anos para melhorar o processo de julgamento. Sensores têm sido utilizados para captar os impactos dos golpes durante lutas de taekwondo (INTERNATIONAL OLYMPIC COMMITTEE, 2024), sistemas eletrônicos de desempenho e rastreamento (EPTS) baseados em câmeras e vestíveis foram adotados pela FIFA para melhorar o desempenho dos atletas (FIFA, 2024). Esses sistemas, que monitoram em tempo real a movimentação, a distância percorrida e os dados fisiológicos dos jogadores, não apenas foram adotados, mas também estabelecidos como uma parte fundamental do futebol moderno. As ligas e competições organizadas pela FIFA utilizam amplamente esses recursos, e o uso dos

EPTS está amplamente aceito tanto por equipes técnicas quanto por jogadores, sendo considerado indispensável para a preparação tática e física, além de contribuir para a prevenção de lesões e a análise detalhada de desempenho (FIFA, 2024). O sistema oferece transparência e precisão, permitindo que decisões sejam tomadas com base em dados objetivos, o que reforça a credibilidade dessas ferramentas no esporte.

Com o incremento da capacidade computacional obtido nos últimos anos e o advento de novas técnicas de inteligência artificial e visão computacional, formas de automatizar a arbitragem por imagem vêm sendo utilizadas em vários esportes (NAIK; HASHMI; BOKDE, 2022). Novas arquiteturas computacionais foram desenvolvidas recentemente, possibilitando a execução eficiente de instruções de máquina próximo ou em tempo real, permitindo a análise local dos eventos esportivos. Arquiteturas de processamento gráfico, tais como as disponibilizadas pelas placas de vídeo, possibilitam a execução paralela de subfluxos computacionais, e desta forma sendo utilizadas como processadores paralelos GPGPU (OWENS *et al.*, 2008). Da mesma forma, incrementos significativos foram obtidos na questão de *software*, especialmente na área de inteligência artificial, com as redes neurais convolucionais (CNNs), as quais vêm sendo utilizadas como solucionadores gerais de problemas das mais diversas naturezas (LECUN; BENGIO; HINTON, 2015). Anteriormente, em métodos clássicos, a solução do problema era obtida através do design da solução de forma específica e isolada, por meio da compreensão dos aspectos que permeiam o problema. Nos modelos atuais de IA, treina-se um modelo computacional para extração direta das características do problema, e estas características são diretamente embarcadas em um reconhecedor (GOODFELLOW; BENGIO; COURVILLE, 2016).

Com o intuito de fornecer suprimentos tecnológicos para facilitar o processo de arbitragem no karatê, na literatura é possível verificar alguns trabalhos especificamente elaborados para este propósito. Embora escassos e limitados ao escopo que se propõem, os trabalhos em geral realizam a leitura do ambiente por meio de câmeras bi-dimensionais ou por profundidade (simples ou múltiplas), e realizam a interpretação do golpe performado (HACHAJ; OGIELA; KOPTYRA, 2015). No entanto, é considerado complexo implementar um sistema capaz de interpretar movimentos psicomotores entre duas ou mais pessoas no mesmo ambiente (SONG *et al.*, 2021). Desta forma, verifica-se que ainda trata-se de um problema em aberto, e novas metodologias ou técnicas computacionais são requeridas para melhorar a performance do processo decisório.

O presente trabalho descreve o desenvolvimento de um sistema que reconhece e classifica golpes de *Karate* entre dois lutadores, utilizando uma CNN (*Convolutional Neural Network*) especificamente designada para a captação dos *keypoints* do esqueleto através do *YOLOv8*. Para a classificação dos golpes, foi utilizado o algoritmo *Support Vector Machines* (SVM) com o *kernel* RBF e os hiperparâmetros ajustados ($C = 100$ e $\text{gamma} = 1$). Todos os pontos-chave do esqueleto foram utilizados, incluindo articulações como cabeça, ombros, cotovelos, punhos, quadris, joelhos e tornozelos, sendo posteriormente

passados para o SVM, que realiza o reconhecimento e classificação dos golpes performados. Para auxiliar no processo, foi construído um *dataset* próprio composto de dez vídeos, 4.345 marcações e cinco golpes, separados em 6 classes, para auxiliar no processo de treinamento dos modelos computacionais. Os resultados experimentais indicaram o modelo SVM com o *kernel* RBF como o mais preciso, obtendo 99,45% pela métrica de precisão durante o treinamento.

O restante deste trabalho está organizado da seguinte forma: A Seção 2 apresenta uma revisão da literatura atual, discutindo algumas estratégias de reconhecimento de ações humanas em esportes de combate. A Seção 3 discute a base teórica, detalhando os conceitos fundamentais necessários para compreender este estudo. Na Seção 4, é descrito o método empregado neste trabalho, detalhando as etapas para criar o *dataset* utilizado e identificar e classificar golpes de *karate* por meio da análise de imagens e redes neurais convolucionais. A Seção 5 apresenta e discute os resultados obtidos com a aplicação do modelo desenvolvido na detecção de golpes, avaliando a eficácia dos recursos computacionais. Finalmente, na Seção 6, são expostas as principais conclusões do estudo, reflexões sobre as implicações dos resultados e sugestões para pesquisas futuras.

2 TRABALHOS CORRELATOS

Para o presente estudo, as revisões bibliográficas foram realizadas utilizando as bases de dados *IEEE Xplore*[®], *Science Direct*[®], *Scopus*[®]. Na literatura foram encontrados artigos científicos nos idiomas inglês e chinês, mas para melhor entendimento e facilidade de leitura consideramos apenas os trabalhos na língua inglesa. Em todas as bases de dados foram utilizadas duas strings de busca em conjunto: (String 1) Todos os campos (“*image processing*” OR “*computer vision*”) AND (String 2) Título, resumo e palavras-chave (“*punch*” OR “*mixed martial arts*” OR “*boxing*” OR “*wrestling*” OR “*karate*” OR “*mma*” OR “*taekwondo*”) AND (“*action recognition*” OR “*pose estimation*”). Estes termos foram escolhidos para que fossem incluídos estudos abrangendo artes marciais com estilos de golpes semelhantes que utilizassem técnicas de visão computacional e processamento de imagem no reconhecimento de poses e ações dos atletas. Os critérios aplicados para a exclusão de estudos no processo de seleção bibliográfica estão resumidos na Tabela 1.

Tabela 1 – Critérios de Exclusão

Critérios de Exclusão	Descrição
Critério 1: Tipo de Artigo	Foram considerados apenas artigos publicados em periódicos.
Critério 2: Idioma	Na literatura foram encontrados artigos científicos nos idiomas inglês e chinês, mas para melhor entendimento e facilidade de leitura consideramos apenas os trabalhos na língua inglesa.
Critério 3: Data	A pesquisa considerou trabalhos publicados entre 1 de janeiro de 2015 e 7 de junho de 2023, data de início da pesquisa.
Critério 4: Natureza do Método	Somente foram selecionados estudos que propuseram métodos originais, excluindo aqueles que se limitam a realizar análises comparativas de métodos já existentes.
Critério 5: Duplicidade	Foram removidos artigos duplicados encontrados em duas ou mais bases de dados, mantendo-se apenas um.

Fonte: Elaborado pelo autor (2024)

Foram removidos 5 artigos por duplicidade. A tabela 2 a seguir mostra a quantidade de estudos obtidos através da busca inicial e a quantidade após a filtragem pelos critérios da tabela anterior.

Após a aplicação dos critérios de exclusão descritos no Quadro 1, começou-se a avaliação por meio de uma leitura das introduções e das conclusões dos estudos restantes. Esse processo levou a uma redução significativa da amostra inicial, de 21 para 12 estudos. Em seguida, foram analisados os objetivos gerais e as metodologias descritas nos trabalhos restantes. Esse processo resultou na seleção final de cinco estudos que apresentavam metodologias detalhadas. Vale ressaltar que as Bases de Dados *IEEE Xplore* e *Science Direct* resultam em 0 estudos selecionados devido à remoção de artigos duplicados, ou

seja, os artigos selecionados nas respectivas bases de dados também eram encontrados na base de dados *Scopus*.

Tabela 2 – Resultados da busca e seleção de estudos nas principais bases de dados

Base de Dados	Busca inicial	Estudos selecionados
<i>IEEE Xplore</i> ®	52	0
<i>Science Direct</i> ®	5	0
<i>Scopus</i> ®	54	21
Total	106	21

Fonte: Elaborado pelo autor (2024)

No estudo de Echeverria e Santos (2021) sobre estimativa de pose humana em combates de *Karate* utilizando *OpenPose*, foi desenvolvida uma abordagem sistemática que incluiu a seleção do *OpenPose* devido à sua robustez e suporte para múltiplas pessoas, além do uso de câmeras de alta velocidade e sistemas computacionais com GPUs para processamento em tempo real. O estudo envolveu a coleta de vídeos de combates, a conversão de quadros de vídeo em imagens, a rotulagem das partes do corpo e a extração de características significativas como velocidade dos movimentos e ângulos das articulações. Quatro algoritmos (*BayesNet*, *J48*, *MLP* e *DeepLearning4J*) foram aplicados para classificação e análise dos dados, e a avaliação do modelo utilizou métricas como precisão e *recall*. O sistema desenvolvido forneceu *insights* valiosos sobre o desempenho psicomotor dos praticantes de karate, com potencial para melhorar técnicas e desempenho através de *feedback* em tempo real e análise detalhada pós-sessão.

No estudo de Lee e Jung (2020) foi abordada a necessidade de ferramentas de avaliação quantitativa nas competições de *poomsae* de *taekwondo* (luta imaginária), onde atualmente há uma falta de padrões objetivos de pontuação. Para resolver isso, os autores desenvolveram o *TUHAD*, um conjunto de dados multimodal de sequências de imagens de ações de *poomsae*, e propuseram uma arquitetura de rede neural convolucional (CNN) baseada em quadros-chave para reconhecimento de ações de *Taekwondo*. O estudo demonstrou uma precisão de reconhecimento de até 95,833%, contribuindo significativamente para a pesquisa e desenvolvimento de métodos de reconhecimento de ações no contexto do *Taekwondo*.

Guo *et al.* (2021) propõe um modelo com o mecanismo de atenção *long short-term memory* (LSTM) com convolução de grafos espaço-temporal aprimorado por atenção (ASTGC-LSTM) para reconhecimento de ações no *Karate*. Utilizando o algoritmo de *Delaunay* para construir uma topologia espaço-temporal, o modelo captura melhor as características dos movimentos técnicos dos atletas. Comparado com métodos de última geração, o ASTGC-LSTM demonstrou desempenho superior em datasets gerados através de *Kinect* e NTU-RGB+D, além de um dataset específico de *Karate*, evidenciando uma melhoria significativa na precisão do reconhecimento de ações.

O estudo de Hachaj, Ogiela e Koptyra (2015) propõe um algoritmo para segmentação e classificação online de técnicas de *Karate* usando dados de captura de movimento. O método combina um classificador de Linguagem de Descrição de Gestos, treinado com aprendizado não supervisionado, e Modelos Ocultos de Markov (HMM) de densidade contínua. Avaliado com dados de seis artistas marciais de *Karate Oyama-ryu*, o sistema obteve taxas de reconhecimento entre 81% e 100%, demonstrando alta eficácia. Esta abordagem inovadora tem aplicações potenciais em sistemas de treinamento de esportes, oferecendo uma análise precisa de movimentos complexos em tempo real.

Kasiri *et al.* (2017) desenvolveram um sistema robusto para a classificação automática de socos de boxe utilizando imagens de profundidade capturadas de cima, minimizando problemas de oclusão. O método inclui rastreamento de partes do corpo e uso de classificadores SVM multiclases e *Random Forest*, com um classificador hierárquico SVM baseado em conhecimento prévio dos tipos de socos. Testado com 14 boxeadores de elite, o sistema alcançou 97,3% de precisão, demonstrando eficácia e robustez na análise de ações finas, superando métodos anteriores de reconhecimento de ações em esportes.

Esses estudos demonstram o avanço significativo na aplicação de técnicas de visão computacional e aprendizado profundo para a análise detalhada e precisa de movimentos em esportes de combate, oferecendo novas perspectivas para a melhoria contínua no treinamento e na avaliação de atletas.

Figura 1 – Tabela Comparativa de Métodos de Reconhecimento de Ações.

Autor(es)	Características	Dataset Utilizado	% Precisão	Método de Extração de Features/ Keypoints
Kasiri et al., 2017	Reconhecimento de golpes de boxe com imagens de profundidade; classificador hierárquico baseado em SVM.	Dados de 14 boxeadores profissionais	97,30%	Trajetórias corporais e descritores de forma
Hachaj et al., 2015	Segmentação e classificação de técnicas de Karate Oyama usando HMM; baseado em dados MoCap.	1236 amostras de MoCap de seis artistas marciais	81% a 100%	Dados de articulações corporais obtidos de MoCap
Guo et al., 2021	Modelo de grafos espaciais-temporais com mecanismo de atenção para reconhecer técnicas de Karate.	Karate Dataset personalizado	Alta precisão	Topologia espacialtemporal usando Delaunay
Lee & Jung, 2020	Reconhecimento baseado em frames chave para técnicas de Taekwondo; utiliza CNN para análise de frames.	TUHAD: 1936 amostras de 8 técnicas Taekwondo	74,49% a 95,83%	Detecção de frames chave a partir de imagens RGB-D
Echeverria & Santos, 2021	Modelagem de desempenho psicomotor em combates de Karate com estimativa de pose utilizando visão computacional.	KUMITRON Karate Dataset	Alta precisão	Estimativa de pose corporal com Open-Pose

Fonte: Elaborado pelo autor.

3 FUNDAMENTAÇÃO TEÓRICA

Para fundamentar as escolhas metodológicas deste trabalho, esta seção revisa as principais contribuições da literatura sobre reconhecimento de ações humanas usando visão computacional, detalha as arquiteturas de redes neurais convolucionais para detecção de keypoints, e explora a aplicação de modelos de classificação, como SVM, no contexto de artes marciais (GUO *et al.*, 2021; KASIRI *et al.*, 2017). A revisão abrange também estudos comparativos de técnicas de última geração e discute os desafios e oportunidades na aplicação dessas tecnologias para auxiliar a arbitragem e o treinamento no *Karate*.

3.1 O KARATE

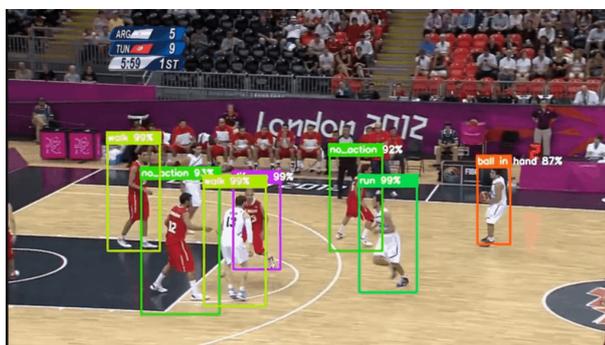
O *Karate* é uma arte marcial de origem japonesa, desenvolvida na ilha de *Okinawa*, com raízes em técnicas chinesas de luta e princípios filosóficos zen budistas. Tradicionalmente, o *Karate* é caracterizado por golpes diretos, rápidos e potentes, utilizando principalmente socos, chutes, bloqueios e golpes com o cotovelo e joelho. O *Karate* não é apenas uma técnica de combate, mas também um caminho de autodesenvolvimento e disciplina mental. Os princípios filosóficos do *Karate* estão profundamente enraizados no respeito, autocontrole, humildade e perseverança. O treinamento contínuo visa não apenas o desenvolvimento físico, mas também a formação de caráter, moldando os praticantes para que se tornem indivíduos mais equilibrados e resilientes. Com o passar dos anos, o *Karate* evoluiu e se disseminou por todo o mundo, resultando no surgimento de diferentes estilos, como o *Shotokan*, *Goju-Ryu*, *Shito-Ryu* e *Wado-Ryu*, cada um com suas particularidades técnicas e filosóficas.

Esta arte marcial se fundamenta em três pilares principais: *Kihon* (fundamentos), *Kata* (luta imaginária) e *Kumite* (luta). O *Kihon* refere-se aos fundamentos e técnicas básicas do *Karate*. Ele é o alicerce para o desenvolvimento das habilidades do praticante, com foco na execução correta de socos, chutes, defesas e posturas. O treinamento contínuo do *Kihon* visa não apenas a perfeição técnica, mas também o desenvolvimento de disciplina, concentração e força mental. Já o *Kata* é uma forma pré-estabelecida de combate imaginário, onde o praticante executa uma série de movimentos coordenados contra adversários fictícios. Cada *Kata* tem um significado simbólico e estratégico, ensinando ao praticante os princípios de defesa, ataque, deslocamento e ritmo. Além do aprimoramento técnico, o *Kata* também é um exercício mental, pois exige concentração e domínio do próprio corpo. Por fim, o *Kumite* é a aplicação prática das técnicas de *Karate* em combate. Existem diferentes formas de *Kumite*, desde o combate básico, onde os movimentos são pré-determinados, até o combate livre, onde os praticantes têm liberdade de aplicar diferentes técnicas em tempo real. No *Kumite* de competição, a ênfase está na velocidade, precisão e controle dos golpes, pois a segurança do praticante e do oponente é uma prioridade.

3.2 VISÃO COMPUTACIONAL

A visão computacional é uma subárea da inteligência artificial que tem como objetivo interpretar e compreender o mundo visual da mesma forma que os seres humanos, através de computadores. Este campo envolve a extração de informações significativas de imagens e vídeos digitais para executar tarefas como reconhecimento de ações humanas, detecção de objetos, reconhecimento facial, segmentação de imagens, reconstrução 3D, entre outras. A visão computacional combina conceitos de cálculo, estatística, aprendizado de máquina, e processamento de imagens para atingir seus objetivos (SZE *et al.*, 2017).

Figura 2 – Uso da visão computacional em partidas da NBA



Fonte: FRANCI, Simone, 2024.

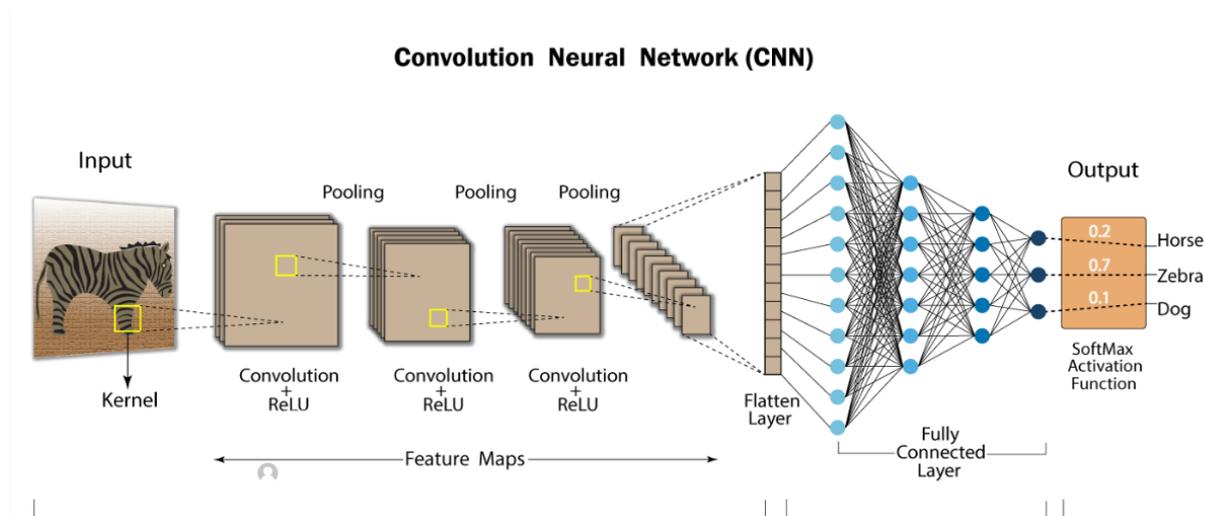
O desenvolvimento de técnicas de aprendizado profundo, especialmente Redes Neurais Convolucionais (CNNs), revolucionou a visão computacional ao permitir que computadores reconheçam padrões complexos em imagens e vídeos de forma mais eficaz do que os métodos tradicionais baseados em características manuais. CNNs são particularmente úteis na detecção de objetos em ambientes complexos, como cenas esportivas, onde a precisão e a rapidez de resposta são essenciais mesmo em meio a muito ruído. No contexto de esportes de combate, como o *Karate*, a visão computacional pode ser usada para detectar golpes, movimentos ilegais e rastreamento fornecendo informações valiosas para a arbitragem e o treinamento de atletas (KRIZHEVSKY; SUTSKEVER; HINTON, 2017; LECUN; BENGIO; HINTON, 2015; GOODFELLOW; BENGIO; COURVILLE, 2016). (KRIZHEVSKY; SUTSKEVER; HINTON, 2017; LECUN; BENGIO; HINTON, 2015; GOODFELLOW; BENGIO; COURVILLE, 2016)

3.3 REDES NEURAI CONVOLUCIONAIS (CNNs)

As Redes Neurais Convolucionais (CNNs) consistem em neurônios artificiais interconectados, cada um com um peso e um viés ajustáveis. As CNNs possuem uma ampla gama de aplicações, como reconhecimento de texto em imagens, reconhecimento facial, detecção e localização de objetos, classificação e segmentação de imagens, processamento de vídeo e linguagem natural, e reconhecimento de fala (DHILLON; VERMA, 2019). Essas

redes têm demonstrado excelente desempenho em competições de visão computacional e processamento de imagens (KHAN *et al.*, 2020).

Figura 3 – Estágios e Camadas de uma Rede Neural Convolutiva.



Fonte: Adaptado de Kategaru (2020).

As CNNs utilizam pesos compartilhados e conexões locais para aproveitar ao máximo os dados de entrada 2D, como imagens, reduzindo o número de parâmetros necessários, simplificando o treinamento e aumentando a eficiência da rede. Esse processo é semelhante ao funcionamento das células no córtex visual, que processam apenas partes específicas de uma cena para detectar padrões como bordas e texturas (ALZUBAIDI *et al.*, 2021). Um modelo comum de CNN inclui várias camadas convolucionais seguidas por camadas de subamostragem (pooling) e termina em camadas totalmente conectadas (ALZUBAIDI *et al.*, 2021). Como mostrado na Figura 3, as CNNs são compostas por múltiplas camadas em diferentes estágios, cada uma composta por mapas de características que servem como entradas e saídas.

As CNNs são caracterizadas por três camadas principais: convolucional, de *pooling* e totalmente conectada. Essas camadas formam os estágios da rede, onde camadas convolucionais e de *pooling* são seguidas por camadas totalmente conectadas. Elas utilizam conexões limitadas, que processam apenas regiões específicas da imagem de entrada, e aproveitam o compartilhamento de peso e o *pooling* para reconhecer características automaticamente, exigindo menos pré-processamento em comparação com técnicas tradicionais (SINGH; SABROL, 2021).

As camadas convolucionais têm como objetivo principal extrair características da entrada, como bordas, linhas e cantos nas primeiras camadas, e aparências de objetos nas camadas mais profundas. Cada camada contém vários filtros, ou *kernels*, que realizam a multiplicação de matrizes para gerar um mapa de características. Os kernels dividem a imagem em pequenas seções, extraindo padrões característicos que são combinados

para calcular a saída da camada convolucional (SINGH; SABROL, 2021). A operação de convolução, que envolve o compartilhamento de peso ao deslizar o kernel pela imagem, permite que as CNNs sejam eficientes em termos de parâmetros ((LECUN; BENGIO; HINTON, 2015)).

Dada uma matriz $I \in \mathbb{R}^{n \times m}$ e um kernel $K \in \mathbb{R}^{k \times k}$, o produto de convolução $I * K$ gera uma matriz de tamanho $(n - k + 1) \times (m - k + 1)$, definida como:

$$(I * K)_{ij} = \sum_{a=0}^{k-1} \sum_{b=0}^{k-1} I_{(i-a)(j-b)} K_{ab}$$

Em uma CNN, a entrada de cada camada é uma matriz tridimensional composta por altura, largura e profundidade, representada por $m \times m \times r$, onde r indica o número de canais, como em uma imagem RGB onde $r = 3$. Os *kernels* convolucionais, que também são tridimensionais, possuem dimensões menores que as da entrada e são utilizados para realizar conexões locais. Esses *kernels* compartilham parâmetros, como pesos e viés, para gerar mapas de características que são convoluídos com a entrada. Após a operação de convolução, uma função de ativação, como *ReLU*, é aplicada para produzir a saída final da camada (ALZUBAIDI *et al.*, 2021).

A expressão para a saída da operação pode ser escrita como:

$$h_k = f(W_k \cdot x + b_k)$$

A ocorrência de características extraídas pela operação de convolução pode variar em diferentes pontos de uma imagem. Após a extração das características, a localização exata torna-se menos importante, desde que a posição relativa entre elas seja mantida (KHAN *et al.*, 2020). A camada de *pooling*, ou agrupamento, é a segunda camada na hierarquia das CNNs e conecta individualmente os mapas de características da camada convolucional anterior. Essa camada reduz o tamanho dos dados ao processar cada mapa de características separadamente, antes de transmitir os resultados para a camada subsequente (SINGH; SABROL, 2021):

$$Z_k^l = g_p(F_k^l)$$

A operação de *pooling*, como *max pooling* ou *average pooling*, é local e resume informações semelhantes na vizinhança de um campo receptivo, retornando a resposta mais forte dentro daquela região (KHAN *et al.*, 2020). Essa operação ajuda a criar combinações de características que são robustas a pequenas variações na entrada, reduzindo o risco de *overfitting* e melhorando a generalização do modelo (KHAN *et al.*, 2020). O *pooling* diminui os parâmetros e acelera o treinamento, permitindo que a rede aprenda mais eficientemente (ALZUBAIDI *et al.*, 2021).

As camadas totalmente conectadas representam a fase final das CNNs, onde características combinadas de forma não linear são utilizadas para classificação. Elas produzem

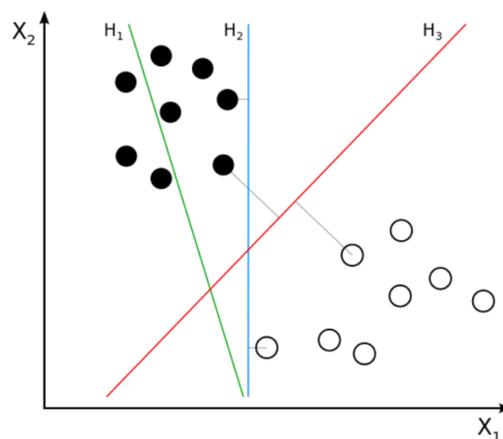
pontuações de classificação que indicam a probabilidade de uma entrada pertencer a uma classe específica, utilizando métodos como máquinas de vetores de suporte (SVMs) ou *softmax* (ALZUBAIDI *et al.*, 2021). No entanto, o uso de muitos parâmetros nessas camadas pode causar *overfitting*, um problema frequentemente mitigado com a técnica de *dropout* (SINGH; SABROL, 2021).

Considerando isso, é possível entender a eficácia das CNNs na detecção de objetos, uma tarefa central da visão computacional que envolve identificar e localizar objetos específicos em imagens ou vídeos. Esse processo é vital em diversas aplicações, incluindo sistemas de segurança e veículos autônomos, destacando a importância das CNNs em avanços tecnológicos e desafios na pesquisa aplicada.

3.4 SVMs E GRID SEARCH

As Máquinas de Vetores de Suporte (SVMs) são uma técnica de aprendizado supervisionado amplamente utilizada para problemas de classificação e regressão. O princípio central das SVMs é encontrar um hiperplano em um espaço de alta dimensionalidade que consiga separar os dados de diferentes classes com a maior margem possível. O hiperplano ideal é aquele que maximiza a distância mínima entre qualquer ponto de dados de qualquer classe e o hiperplano, o que ajuda a garantir a robustez do modelo ao lidar com novos dados (CORTES; VAPNIK, 1995).

Figura 4 – Exemplo de separação de classes usando SVM.



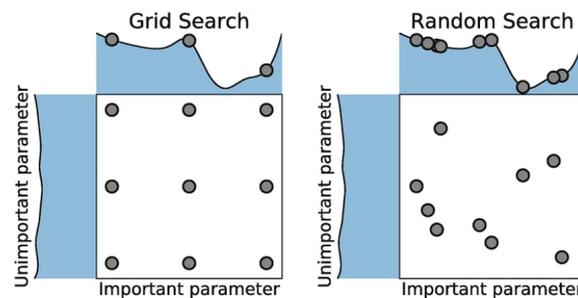
Fonte: SVM separating hyperplanes (SVG). Wikimedia Commons, 2008.

Para otimizar o desempenho de uma SVM, é fundamental ajustar seus hiperparâmetros, como o parâmetro de regularização C , que controla a penalidade para erros de classificação, e o parâmetro γ (gamma) do *kernel* RBF, que define a largura da curva de decisão. A escolha adequada desses hiperparâmetros é crucial, pois influencia diretamente a capacidade do modelo de generalizar a partir dos dados de treinamento ((CHANG;

LIN, 2011)). Ajustar esses parâmetros corretamente pode ser complexo, e uma abordagem sistemática é necessária para garantir que o modelo seja otimizado.

O *Grid Search* é uma técnica de otimização de hiperparâmetros que envolve a definição de uma grade de possíveis valores para cada hiperparâmetro e a avaliação de todas as combinações para identificar a que resulta no melhor desempenho do modelo. Essa abordagem é particularmente útil para SVMs, onde a escolha dos hiperparâmetros C e γ pode ter um impacto significativo na acurácia do modelo. A combinação de *Grid Search* com validação cruzada permite que o modelo final seja ajustado para obter o melhor desempenho possível nos dados de teste (BERWICK, 2019).

Figura 5 – Ilustração do *Grid Search* para Otimização de Hiperparâmetros.



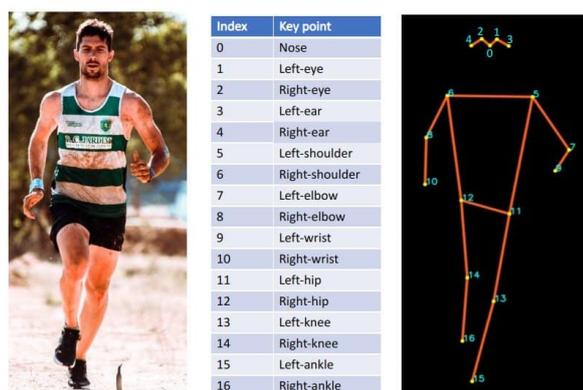
Fonte: Bergstra e Bengio (2012)

3.5 RECONHECIMENTO DE AÇÕES HUMANAS

O reconhecimento de ações humanas é um problema desafiador em visão computacional, que envolve a detecção e categorização de movimentos humanos em sequências de vídeo. Em esportes de combate, como o *Karate*, onde os movimentos são rápidos e complexos, técnicas avançadas são necessárias para alcançar uma classificação precisa. O *YOLO* (*You Only Look Once*) é uma das arquiteturas de redes neurais convolucionais mais populares para a detecção de objetos em tempo real devido à sua capacidade de fazer previsões rápidas e precisas (REDMON *et al.*, 2016). A versão mais recente durante a pesquisa, *YOLOv8*, oferece melhorias substanciais em termos de velocidade e precisão de detecção.

A principal vantagem do *YOLOv8* é sua capacidade de detectar múltiplos objetos e pontos-chave de esqueletos humanos em uma única passagem através da rede neural, o que é especialmente útil para o reconhecimento de ações humanas em vídeos esportivos. A arquitetura *YOLOv8* incorpora camadas convolucionais aprimoradas e técnicas de regularização que permitem um aprendizado mais eficiente das características espaciais e temporais dos movimentos de *Karate*, criando uma base robusta para a classificação subsequente usando SVMs (ULTRALYTICS, 2024).

Figura 6 – *YOLOv8 keypoints*.



Fonte: LEARN OPENCV.

4 METODOLOGIA PROPOSTA

O método apresentado neste trabalho consiste na análise *keypoints* de pose humana para identificação de golpes de *karate* através da construção de um *dataset* próprio. No geral, a metodologia consiste em cinco etapas, ilustradas na Figura 7. O primeiro estágio consiste na aquisição de vídeos de atletas realizando os golpes selecionados para estudo em um ambiente controlado, com quantidade distribuída homoganeamente, já o segundo na criação e execução de um sistema para anotação manual de imagens em vídeo utilizando CNNs para obtenção dos *keypoints* dos atletas, garantindo maior qualidade e menor tempo de trabalho. O terceiro estágio se dedica à elaboração do conjunto de dados e definição dos parâmetros do modelo SVM escolhido. No quarto estágio é realizada a definição dos parâmetros do modelo SVM escolhido e a implementação dos processos de treinamento, gerando o modelo de detecção. Por fim, no quinto estágio é executado o modelo para classificação com base em vídeos de golpes de *karate*.

Figura 7 – Fluxo computacional da abordagem proposta.



Fonte: Elaborado pelo autor (2024).

4.1 GRAVAÇÃO DE VÍDEOS

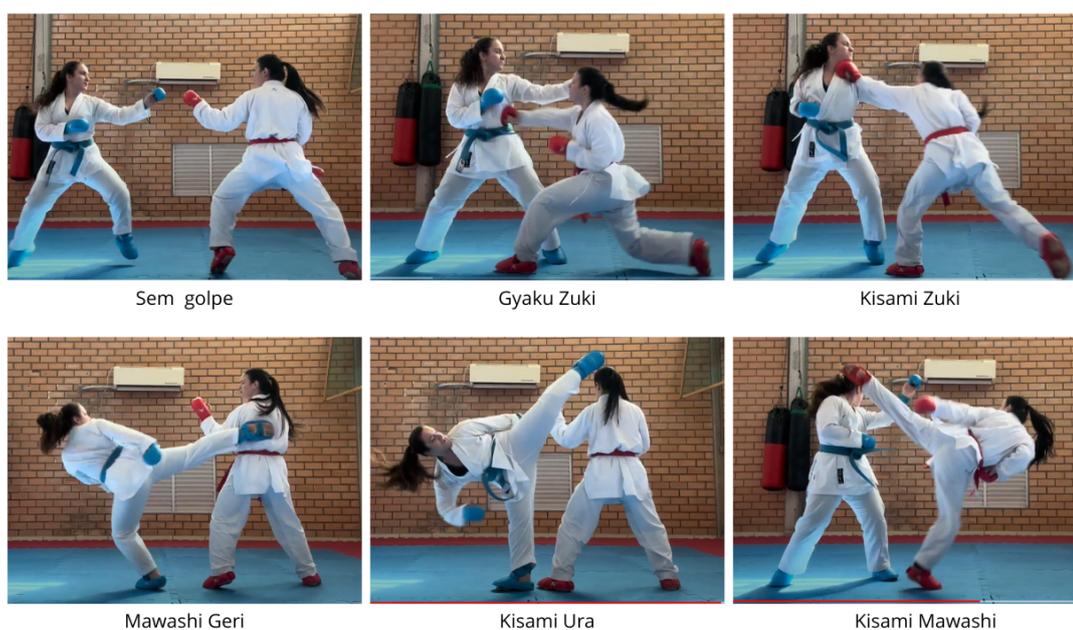
O estágio de coleta de imagens foi planejado para garantir a máxima eficiência do método, utilizando atletas de alto rendimento em atividade e uma variação equilibrada de golpes. A coleta foi realizada com a participação de cinco atletas, sendo dois homens e três mulheres. Dentre esses, três são membros ativos da seleção brasileira de karate, um é ex-integrante da seleção brasileira, e outro faz parte da seleção catarinense.

As gravações foram realizadas em uma academia de karate, equipada com um tatame que reproduz as características das áreas de competição, incluindo cores (azul e vermelha) e dimensões oficiais. Essa ambientação visou simular as condições reais de competição, garantindo a relevância dos dados coletados.

Durante as gravações, foram executados cinco tipos de golpes: *gyaku zuki* (soco invertido no abdômen), *mawashi geri* (chute circular nas costas), *kisami mawashi* (chute

circular na cabeça), *kisami ura* (chute circular na cabeça com a sola do pé) e *kizami zuki* (soco com a mão da frente no peito). A distribuição dos golpes foi homogênea entre os participantes, respeitando as posições e ângulos de combate (guarda), garantindo uma amostragem representativa de cada movimento.

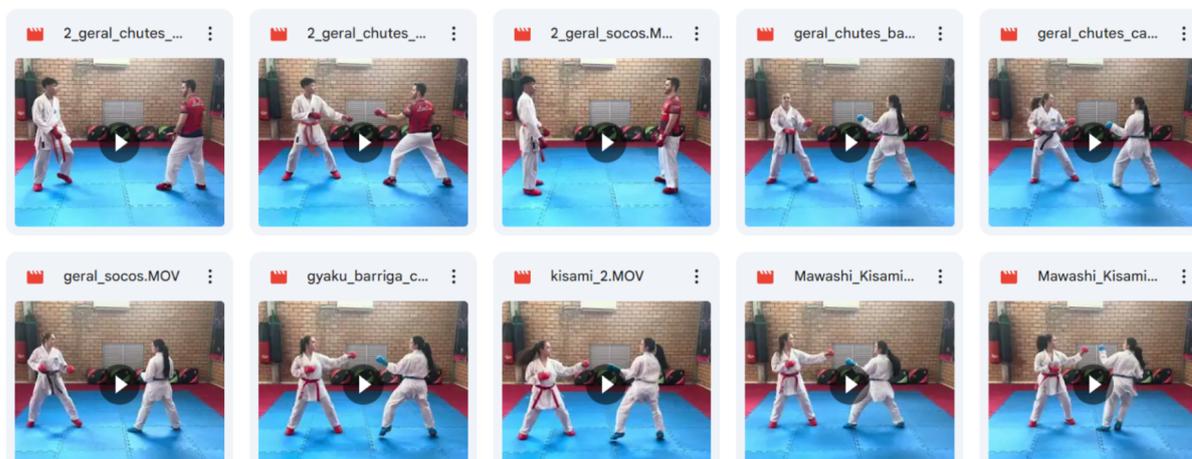
Figura 8 – Golpes de Karate desempenhados para composição do Dataset



Fonte: Elaborado pelo autor (2024).

No total, foram gravados 10 vídeos, sendo que 7 deles foram destinados à criação do dataset de treinamento do modelo, enquanto 3 foram reservados para o estágio de validação. Em cada vídeo, os atletas realizaram os mesmos golpes de forma mútua, assegurando uma quantidade equilibrada de ações de cada tipo, o que contribuiu para a consistência e qualidade dos vídeos para geração posterior do conjunto dados utilizados à serem utilizados no treinamento do modelo.

Figura 9 – Vídeos Gravados.



Fonte: Elaborado pelo autor (2024).

Com essa abordagem, buscou-se não apenas a representatividade dos movimentos e dos diferentes perfis de atletas, mas também a fidelidade às situações de combate, garantindo que o modelo treinado possa ser aplicado em contextos reais de competições de *karate*.

4.2 DETECÇÃO DOS *KEYPOINTS* E ELABORAÇÃO DO CONJUNTO DE DADOS

No estágio de detecção de *keypoints*, foi desenvolvida uma aplicação exclusiva em *Python*, dedicada à anotação precisa de cada golpe executado pelos atletas durante as gravações. Essa aplicação utilizou o modelo *YOLOv8m-pose* para detectar os *keypoints* de cada atleta em cada frame dos vídeos. Ao todo, até 17 pontos chave do corpo foram detectados. A detecção foi realizada de forma minuciosa, analisando *frame a frame* os vídeos capturados. Para facilitar o processo de anotação, foram configuradas teclas especiais na aplicação, permitindo que cada frame fosse classificado de acordo com a ação ou golpe realizado. Assim, cada vídeo foi reproduzido de forma detalhada, e cada ação dos atletas foi registrada, com a devida categorização dos *keypoints* identificados.

Figura 10 – Aplicação para anotação dos golpes.



Fonte: Elaborado pelo autor (2024).

Os conjuntos de pontos detectados pelo modelo *YOLO* foram então classificados em diferentes categorias, baseadas na disposição espacial dos *keypoints* ao longo dos movimentos dos atletas. Essa abordagem permitiu distinguir as diferentes posturas e gestos, oferecendo uma análise detalhada dos padrões de movimento para cada tipo de golpe. Ao término desse processo de anotação, foi gerado um arquivo CSV contendo todas as informações relevantes sobre as coordenadas dos *keypoints* de cada atleta, bem como suas respectivas classificações manuais. Esse arquivo estruturado serviu como uma base de dados essencial para a etapa subsequente de treinamento, permitindo uma análise precisa do comportamento dos atletas e a criação de um modelo mais robusto e adaptado às variações dos golpes.

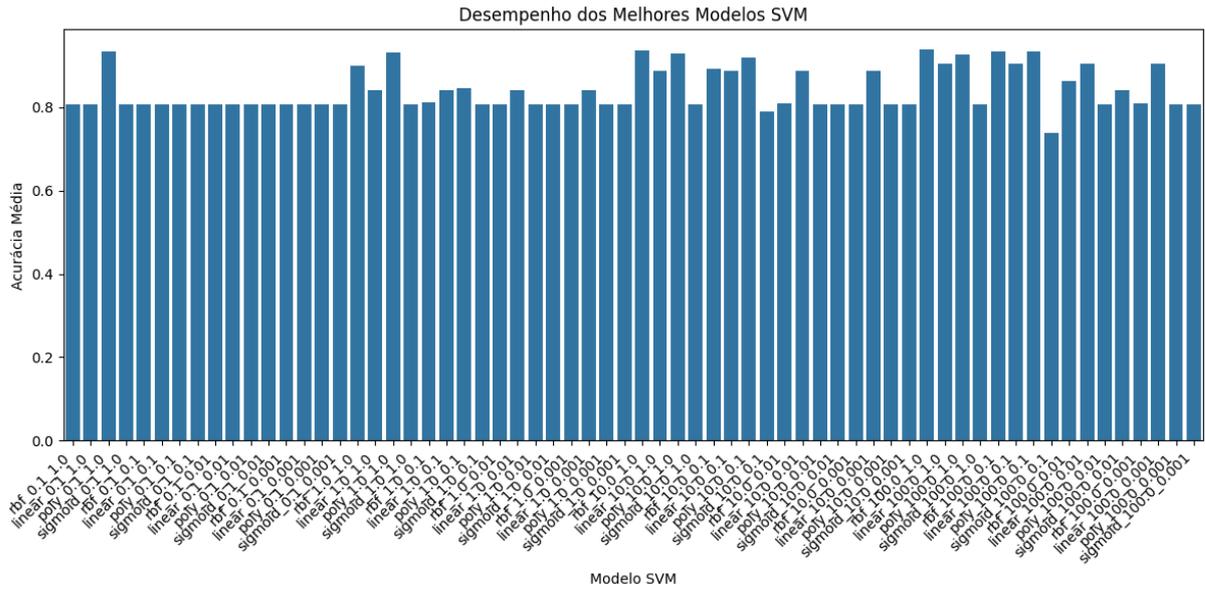
4.3 TREINAMENTO COM SVM

Para o estágio de treinamento, o dataset foi carregado e processado, utilizando as coordenadas como variáveis preditoras e as classes dos golpes como rótulos. Os tipos de golpes identificados incluíam movimentos como *gyaku*, *kisami*, *kisami mawashi*, *kisami ura*, *mawashi*, além de uma classe sem golpe para frames neutros. Os dados foram divididos em conjuntos de treinamento e teste com a proporção 80/20, assegurando que o modelo fosse avaliado com informações não vistas durante o aprendizado.

O modelo selecionado para a tarefa de classificação foi o SVM (*Support Vector Machine*), uma técnica amplamente reconhecida por sua eficácia na resolução de problemas

de classificação, especialmente em cenários de alta dimensionalidade, como os dados representados pelas coordenadas dos *keypoints*. A otimização dos hiperparâmetros do SVM foi realizada por meio da técnica de *GridSearchCV*, resultando na escolha do *kernel* RBF (*Radial Basis Function*), com o parâmetro C ajustado para 100 e gamma definido como 1, visando maximizar a acurácia do modelo.

Figura 11 – Desempenho dos Melhores Modelos SVM.



Fonte: Elaborado pelo autor (2024).

4.4 VALIDAÇÃO DOS GOLPES

A avaliação do desempenho do classificador foi conduzida por meio de métricas fundamentais como acurácia, *f1-score*, *recall* e *precisão*, tanto para os dados de treinamento quanto para os de teste. Durante o treinamento, as métricas indicaram um ajuste muito eficaz do modelo aos dados, refletindo uma alta capacidade de identificação correta das diferentes classes de golpes de karate. Na etapa de teste, que mede a habilidade do modelo em generalizar para novos dados, os resultados mostraram um desempenho consistente e preciso. A proximidade entre os resultados do treinamento e do teste sugere que o modelo foi bem ajustado, evitando problemas como o *overfitting* e mantendo sua robustez ao ser aplicado em dados não vistos anteriormente.

5 RESULTADOS E AMBIENTE EXPERIMENTAL

5.1 FERRAMENTA DE ANOTAÇÃO DE VÍDEOS

A ferramenta de anotação de vídeos desenvolvida foi projetada para facilitar a marcação de *keypoints* em sequências de vídeo capturadas de combates de *Karate*. O objetivo principal foi automatizar a coleta e categorização dos movimentos dos atletas, otimizando o processo de criação de datasets para treinamento de modelos de aprendizado de máquina, como ilustrado na Figura 10.

5.1.1 ESTRUTURA E FUNCIONALIDADES

A aplicação foi desenvolvida em *Python*, utilizando bibliotecas modernas de visão computacional, como *OpenCV* e o framework *YOLOv8*, especializado na detecção de poses humanas. Essa combinação garantiu uma solução robusta e eficiente para atender às necessidades do projeto. As principais funcionalidades da ferramenta incluem:

- **Detecção de Keypoints:** Implementação do modelo *YOLOv8m-pose*, capaz de identificar até 17 pontos-chave do corpo humano em cada frame dos vídeos processados, incluindo articulações como cabeça, ombros, joelhos e tornozelos.
- **Visualização Gráfica:** Os *keypoints* são sobrepostos diretamente aos quadros do vídeo em tempo real, proporcionando uma visualização clara e intuitiva que auxilia na validação e correção de detecções pelo operador.
- **Normalização de Coordenadas:** As coordenadas dos *keypoints* são ajustadas automaticamente para corresponder à resolução do vídeo original, garantindo consistência ao trabalhar com diferentes fontes e formatos de mídia.
- **Interatividade do Usuário:** A ferramenta oferece uma interface interativa, permitindo ao operador navegar facilmente pelos frames do vídeo usando comandos de teclado. Essa funcionalidade inclui avançar, retroceder e classificar os golpes de forma prática e eficiente.
- **Geração Automatizada de *Dataset*:** Após a anotação, cada frame é armazenado em um arquivo CSV estruturado, contendo as coordenadas normalizadas dos *keypoints* e a classe do golpe identificado. Esse arquivo constitui uma base de dados essencial para o treinamento de modelos de aprendizado de máquina.

5.1.2 PROCESSO DE OPERAÇÃO

O funcionamento da ferramenta desenvolvida é estruturado em etapas que garantem eficiência e precisão na coleta e anotação dos dados. As etapas são descritas a seguir:

1. **Carregamento do Vídeo:** O vídeo de combate é carregado no sistema, permitindo o processamento frame a frame. Isso assegura que cada movimento capturado seja analisado de forma detalhada.
2. **Deteção e Exibição dos Keypoints:** O modelo *YOLOv8m-pose* é utilizado para detectar até 17 pontos-chave do corpo humano em cada frame. Esses *keypoints* são exibidos visualmente sobre os quadros do vídeo em tempo real, facilitando a validação das detecções.
3. **Anotação Manual:** Utilizando comandos de teclado intuitivos, o operador categoriza os movimentos e golpes registrados em cada frame. A Tabela 3 apresenta as teclas configuradas e seus significados.

Tecla	Significado
a	Voltar um frame no vídeo
d	Avançar um frame no vídeo
i	Kisami Zuki (cabeça)
j	Kisami Mawashi (cabeça)
k	Mawashi Geri (costas)
l	Gyaku Zuki (barriga)
s	Sem golpe
q	Sair

Tabela 3 – Tabela de teclas e seus significados.

4. **Exportação de Dados Estruturados:** Após a anotação manual, as informações coletadas, incluindo as coordenadas dos *keypoints* e as classes atribuídas, são exportadas para um arquivo CSV estruturado. A Tabela 4 ilustra um exemplo dos dados exportados.

Tabela 4 – Tabela reduzida com valores de pontos e classes.

Classe	P1XPt1	P1YPt1	...	P2XPt17	P2YPt17
sem golpe	0,51723176	0,41986278	...	0,23464422	0,8591554
sem golpe	0,4997954	0,42519462	...	0,23565593	0,8435355
sem golpe	0,4813713	0,4384002	...	0,23196249	0,83080274
gyaku	0,33978522	0,3334381	...	0,6158761	0,87043524
gyaku	0,34006837	0,34571064	...	0,60332084	0,8729345
gyaku	0,3374456	0,3528667	...	0,5878788	0,87683445
gyaku	0,3341469	0,36514038	...	0,5850365	0,8728805
gyaku	0,43765342	0,46413517	...	0,21917987	0,8516296
gyaku	0,44253522	0,4618194	...	0,2190868	0,8508933
sem golpe	0,32447034	0,35015047	...	0,3829204	0,8565539
sem golpe	0,45134878	0,44842643	...	0,21396662	0,84810734
sem golpe	0,4581069	0,43730485	...	0,21560025	0,8530044
...

A ferramenta desenvolvida se destaca por sua precisão e eficiência, oferecendo uma solução robusta para a criação de *datasets* específicos voltados para a análise de golpes de *Karate*. Sua flexibilidade e interatividade tornam o processo de anotação mais intuitivo, garantindo maior qualidade e confiabilidade nos dados coletados.

5.2 DATASET PRODUZIDO

O dataset desenvolvido neste trabalho foi construído a partir da análise de 9 vídeos gravados, nos quais os atletas executaram movimentos específicos em um ambiente controlado. Durante a etapa de anotação, foram identificados 4235 amostras, com as ações distribuídas em 6 classes, sendo uma delas destinada aos frames sem golpes e as demais correspondendo aos golpes selecionados para o estudo:

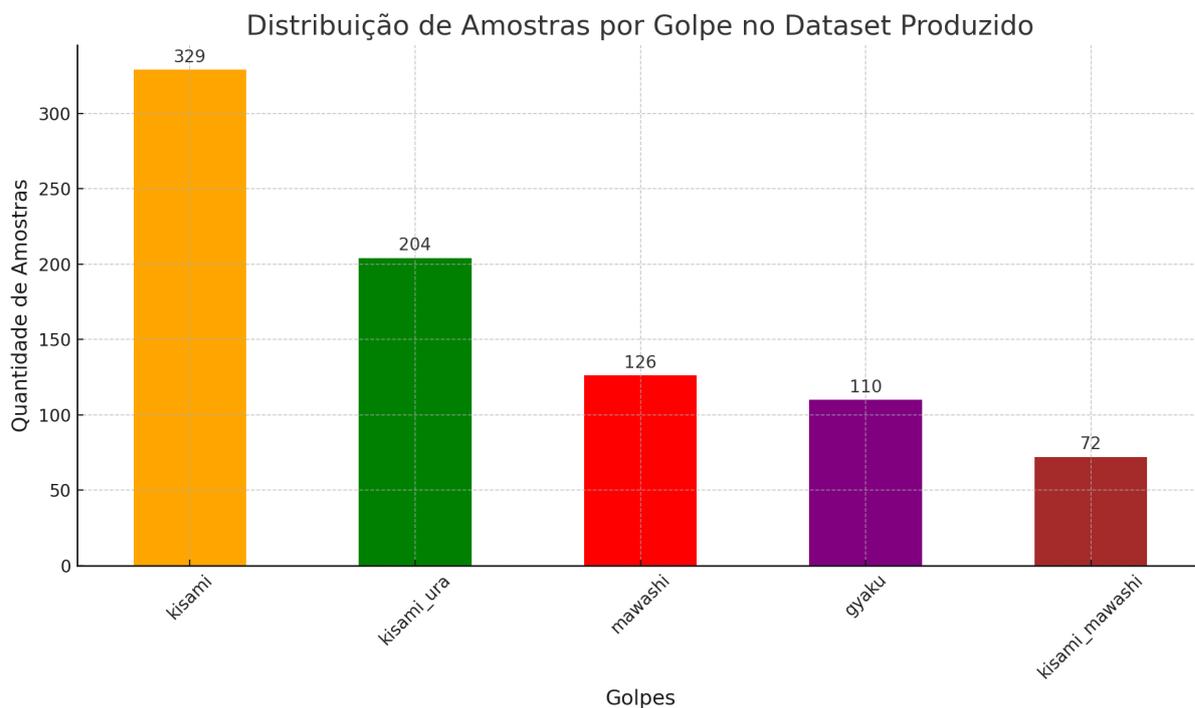
Tabela 5 – Descrição dos Golpes no Dataset Produzido

Golpe	Descrição
Gyaku Zuki	Soco direto invertido.
Mawashi Geri	Chute circular no tronco.
Kisami Zuki	Soco frontal realizado com a mão da frente.
Kisami Mawashi Geri	Chute circular frontal no rosto.
Kisami Ura Mawashi Geri	Chute circular frontal inverso realizado com a sola do pé, no rosto.

Cada amostra corresponde a uma imagem extraída dos vídeos, na qual foram detectados 17 *keypoints* utilizando o modelo *YOLOv8m-pose*. Esses pontos foram anotados manualmente para associar cada frame à respectiva classe. Essa abordagem resultou em um dataset robusto, que serviu como base para o treinamento e validação do modelo de classificação.

Embora a classe "Sem Golpe" domine em número de amostras, a representatividade dos golpes foi mantida ao balancear as execuções de cada atleta durante a coleta. Esse balanceamento é essencial para garantir que o modelo seja capaz de identificar golpes menos frequentes com precisão.

Figura 12 – Distribuição de Amostras por Golpes no Dataset Produzido.



Fonte: Elaborado pelo autor (2024).

Esse dataset, por sua composição e metodologia de construção, constitui uma importante contribuição para a análise de ações em combates de karate, oferecendo uma base sólida para o treinamento e validação de modelos preditivos. Há a intenção de publicá-lo e torná-lo público, já tendo a autorização de imagem dos atletas participantes.

5.3 RESULTADOS EXPERIMENTAIS

Os resultados experimentais deste trabalho avaliaram a eficiência do modelo na classificação dos golpes de Karate com base no *dataset* produzido. Para isso, foi utilizado um modelo de SVM (*Support Vector Machine*) ajustado e validado em amostras representativas do conjunto de dados.

O modelo foi configurado com o *kernel* RBF, após ajuste dos hiperparâmetros via *Grid Search*, resultando em $C = 100$ e $\gamma = 1$. O *dataset* foi dividido em 80% para treinamento e 20% para teste. Cada entrada do conjunto representa um frame extraído dos vídeos, contendo as coordenadas dos *keypoints* corporais e o rótulo correspondente ao golpe identificado.

O modelo foi avaliado utilizando as métricas de *acurácia*, *precisão*, *recall* e *F1-score*. Os resultados obtidos são apresentados na Tabela 6.

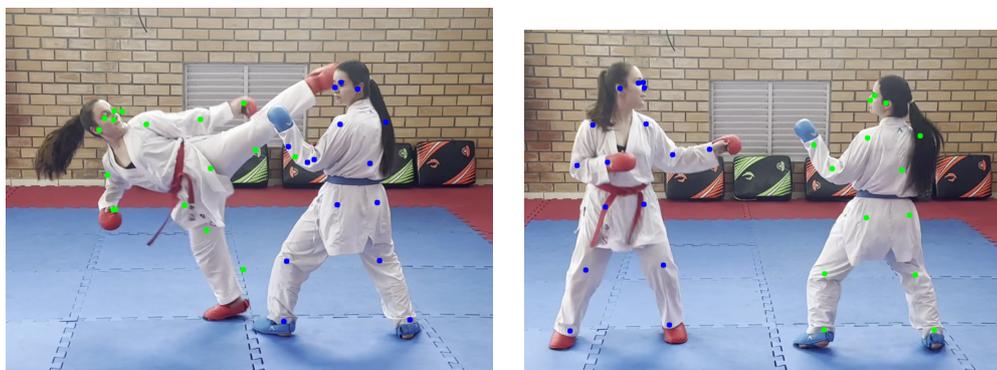
Esses resultados indicam que o modelo alcançou um alto desempenho, sendo capaz de aprender os padrões dos movimentos e generalizar com precisão para amostras não

Tabela 6 – Métricas de Desempenho no Treinamento e Teste

Métrica	Treinamento (%)	Teste (%)
Acurácia	99,42	95,62
F1-score	99,43	95,59
Recall	99,42	95,62
Precisão	99,45	95,64

vistas. A diferença entre as métricas de treinamento e teste demonstra que não houve ocorrência de *overfitting*, evidenciando a robustez do modelo.

Na Figura 13, são apresentados exemplos de frames anotados do *dataset*, com os *keypoints* corporais destacados e os golpes classificados pelo modelo. Essa visualização ilustra a capacidade do modelo em capturar as características dos movimentos, mesmo diante de variações na execução e estilos dos atletas.

Figura 13 – *Keypoints* Detectados durante execução do Sistema.

Os golpes apresentados incluem: *Gyaku Zuki*, *Kisami Zuki*, *Kisami Mawashi Geri*, *Kisami Ura Mawashi Geri* e *Mawashi Geri*. Cada um dos movimentos foi identificado corretamente pelo modelo, confirmando sua eficácia em capturar diferentes padrões de movimento, mesmo em condições de variabilidade.

Os resultados experimentais confirmam a viabilidade do sistema proposto para a classificação de golpes de Karate. O modelo de SVM demonstrou desempenho robusto, mesmo diante de variações no estilo e na execução dos movimentos. A precisão alcançada reforça o potencial do sistema como uma ferramenta prática para suporte à arbitragem e treinamento no contexto esportivo.

5.4 VALIDAÇÃO QUALITANTITATIVA

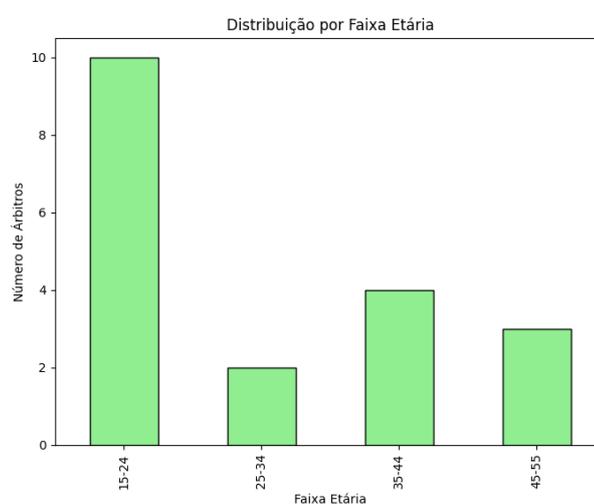
A validação do sistema desenvolvido foi realizada por meio de um questionário direcionado a árbitros de Karate com experiência prática em competições. O objetivo principal foi avaliar qualitativamente e quantitativamente a percepção dos profissionais em relação à precisão, utilidade e aplicabilidade do sistema no contexto de arbitragem. As perguntas abrangeram diversos aspectos, como precisão do sistema na identificação de

golpes, experiência prática dos árbitros, graduações, classificações, aceitação do sistema e impacto nas decisões de arbitragem.

5.4.1 PERFIL DOS RESPONDENTES

Os respondentes incluíram árbitros com idades variando entre 15 e 55 anos, representando uma ampla faixa etária. A maior concentração de participantes estava na faixa de 15 a 24 anos, indicando uma predominância de árbitros mais jovens, mais abertos à utilização de tecnologias.

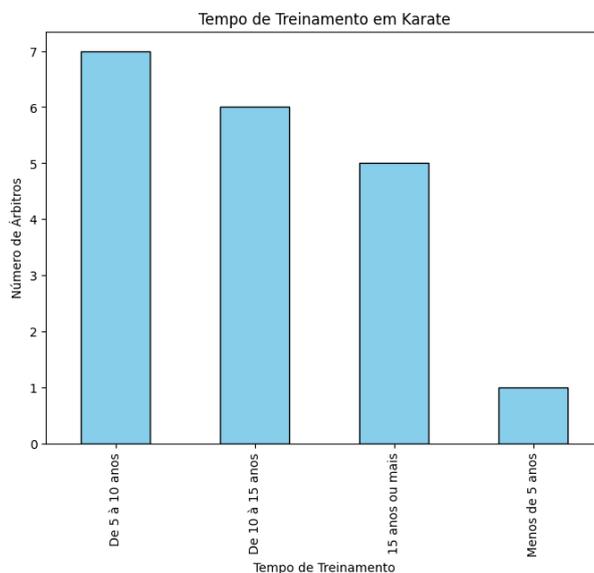
Figura 14 – Distribuição por faixa etária.



Fonte: Elaborado pelo autor (2024).

No que diz respeito ao tempo de treinamento em Karate, a maioria dos respondentes indicou possuir entre 5 e 10 anos de prática contínua, evidenciando um nível avançado de experiência. Alguns árbitros relataram mais de 15 anos de treinamento, demonstrando um alto grau de envolvimento e especialização.

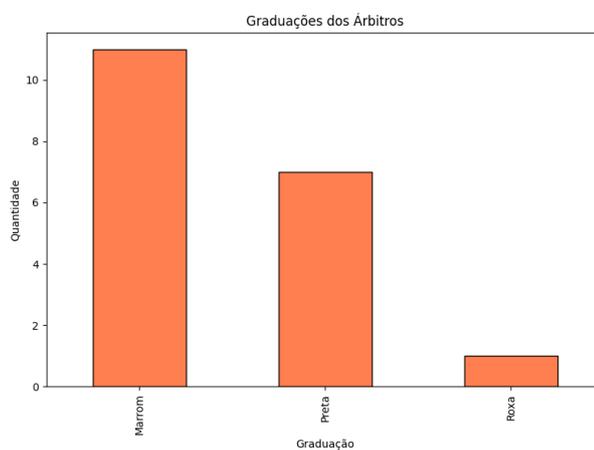
Figura 15 – Tempo de Treinamento.



Fonte: Elaborado pelo autor (2024).

As graduações dos árbitros incluíram 11 faixas marrons, 7 faixas pretas e 1 faixa roxa. Essa distribuição reforça a diversidade de níveis de expertise entre os respondentes, possibilitando uma análise abrangente sobre a aceitação do sistema em diferentes contextos.

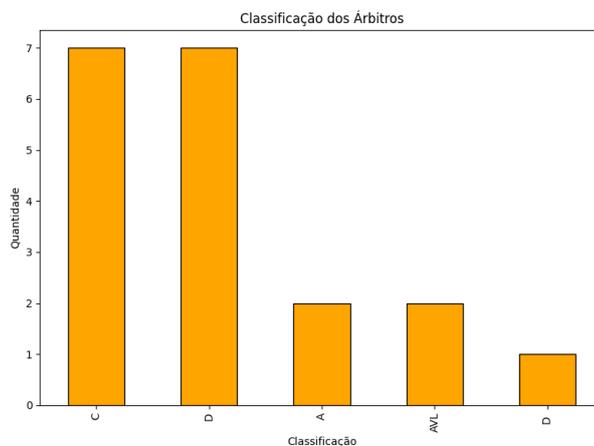
Figura 16 – Graduações dos Árbitros.



Fonte: Elaborado pelo autor (2024).

Com relação à classificação de árbitros, os dados indicaram que 7 eram da classe C, 8 da classe D, 2 da classe A e 2 estavam em avaliação (AVL). Esse aspecto demonstra a presença de árbitros em diferentes estágios de formação e atuação.

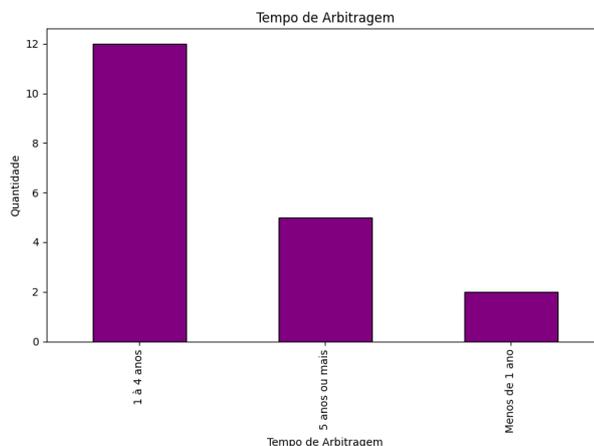
Figura 17 – Classificação dos Árbitros.



Fonte: Elaborado pelo autor (2024).

Quanto ao tempo de experiência em arbitragem, os participantes apresentaram variações significativas, abrangendo desde 1 a 4 anos até mais de 5 anos de atuação. Essa diversidade de tempo de experiência agrega valor à avaliação do sistema, oferecendo múltiplas perspectivas.

Figura 18 – Tempo de Experiência em Arbitragem.

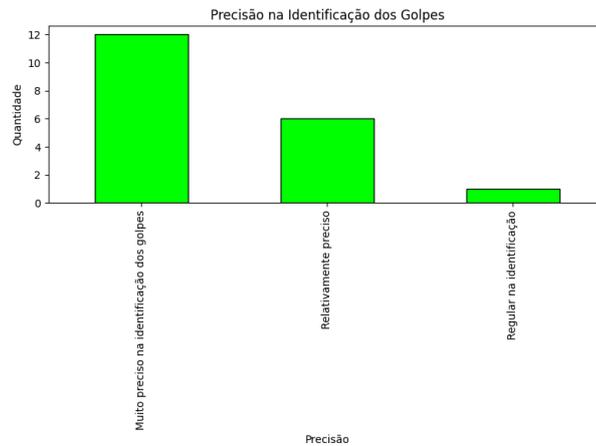


Fonte: Elaborado pelo autor (2024).

5.4.2 AVALIAÇÃO DO SISTEMA

Os árbitros foram convidados a avaliar a precisão do sistema na identificação automática de golpes. A maioria, equivalente a 63%, classificou o sistema como "muito preciso", enquanto 32% o consideraram "relativamente preciso" e apenas 5% avaliaram como "regular". Esses resultados evidenciam a confiança dos árbitros na funcionalidade do sistema.

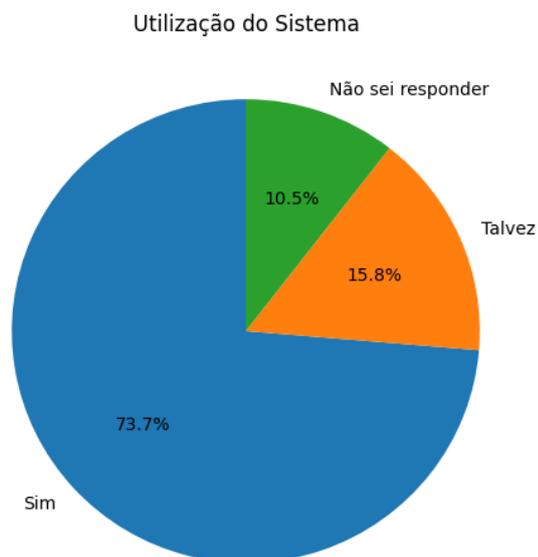
Figura 19 – Precisão na Identificação dos Golpes.



Fonte: Elaborado pelo autor (2024).

Em relação à utilização do sistema como ferramenta de apoio, 74% dos árbitros indicaram que o utilizariam, 16% consideraram a possibilidade, enquanto 10% permaneceram indecisos. Esses resultados reforçam o potencial de aplicação prática do sistema.

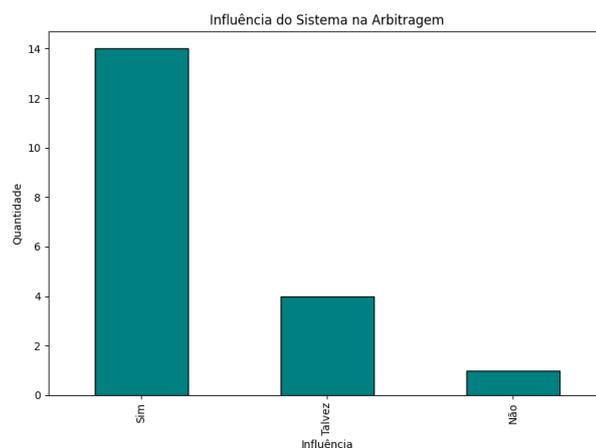
Figura 20 – Utilizariam o Sistema.



Fonte: Elaborado pelo autor (2024).

Por fim, avaliou-se o impacto do sistema nas decisões de arbitragem. Dos respondentes, 74% acreditam que o sistema pode influenciar positivamente suas decisões, 21% consideraram que talvez houvesse impacto, enquanto 5% não identificaram influência significativa.

Figura 21 – Influência do Sistema na Arbitragem.



Fonte: Elaborado pelo autor (2024).

Os resultados da validação qualitativa e quantitativa reforçam a viabilidade do sistema como uma ferramenta inovadora no contexto de arbitragem esportiva. Apesar de algumas limitações pontuais, como a adaptação para golpes mais complexos, o feedback recebido destaca a aceitação e o potencial impacto positivo do sistema em competições de Karate.

6 DISCUSSÃO, CONCLUSÃO E TRABALHOS FUTUROS

6.1 DISCUSSÃO

Os resultados obtidos neste trabalho confirmam o potencial do sistema proposto para classificação de golpes de *Karate*, com precisão acima de 95% nas métricas avaliadas. No entanto, algumas questões técnicas e operacionais merecem atenção:

A latência observada durante a execução de modelos robustos, como o *YOLOv8m-pose*, pode restringir a aplicação do sistema em contextos de tempo real, como competições de grande escala. A redução dessa latência, sem comprometer a precisão do modelo, é essencial para aumentar sua usabilidade.

A variabilidade nos estilos de execução dos golpes também foi identificada como um desafio. Movimentos mais complexos ou menos frequentes apresentaram maior dificuldade de classificação, indicando que a diversidade e representatividade do *dataset* são fatores cruciais para a generalização do modelo. Além disso, o processo de anotação manual, embora eficaz, demonstrou ser trabalhoso e suscetível a erros, sugerindo a necessidade de métodos mais automatizados.

Outro ponto relevante é a aceitação do sistema pelos árbitros, que destacaram sua utilidade em competições, mas também apontaram a necessidade de melhorias na interface para facilitar sua integração com sistemas existentes. A aplicabilidade do sistema pode ser ampliada com ajustes que considerem essas demandas práticas.

6.2 CONCLUSÃO

Este trabalho apresentou o desenvolvimento e a validação de um sistema inovador para classificação de golpes de *Karate*, combinando visão computacional com aprendizado de máquina. A integração do modelo *YOLOv8m-pose* para extração de *keypoints* com o algoritmo SVM resultou em um sistema altamente preciso e eficiente, capaz de identificar e classificar diferentes golpes com taxas de acurácia superiores a 95%.

Os resultados experimentais demonstraram a viabilidade e robustez do sistema, enquanto a avaliação qualitativa com árbitros evidenciou sua aceitação e potencial como ferramenta de apoio à arbitragem esportiva. O sistema contribui para reduzir a subjetividade nas decisões de arbitragem e pode ser um aliado importante na modernização de competições e treinamentos.

As contribuições deste trabalho abrem caminho para novas aplicações e estudos no campo de visão computacional e aprendizado de máquina no esporte. Além disso, o sistema desenvolvido possui flexibilidade para adaptações futuras, permitindo sua expansão para outras modalidades esportivas. Dessa forma, este estudo estabelece uma base sólida para avanços tecnológicos que promovam maior justiça, precisão e eficiência em contextos competitivos.

6.3 TRABALHOS FUTUROS

Com base nos resultados e limitações identificadas, as seguintes direções são propostas para trabalhos futuros:

1. **Ampliação do *Dataset*:** Expandir o conjunto de dados com registros de competições reais, diferentes estilos de *Karate* e variáveis adicionais, como iluminação, ângulos de câmera e condições ambientais. Isso aumentará a robustez do modelo em cenários reais.
2. **Integração de Sensores Adicionais:** Combinar dados de visão computacional com sensores de impacto e movimento, como acelerômetros e giroscópios, para fornecer uma análise multimodal que considere tanto a precisão visual quanto a intensidade dos golpes.
3. **Otimização da Latência:** Explorar alternativas para reduzir a latência em aplicações de tempo real, como a utilização de modelos mais leves ou estratégias de inferência acelerada, mantendo o alto desempenho do sistema.
4. **Melhoria da Interface:** Desenvolver uma interface intuitiva e amigável para operadores, árbitros e organizadores de competições. A integração com sistemas de pontuação automatizados pode aumentar significativamente a aceitação prática.
5. **Adaptação para Outras Modalidades Esportivas:** Validar e adaptar o sistema para outras artes marciais, como *Taekwondo*, *MMA* ou *Judô*, ampliando seu impacto em diferentes contextos esportivos.
6. **Automação do Processo de Anotação:** Implementar aprendizado semi-supervisionado ou não supervisionado para reduzir a dependência de anotações manuais e acelerar a expansão do *dataset*.

Essas melhorias visam não apenas aprimorar o desempenho técnico do sistema, mas também expandir sua aplicabilidade e aceitação em cenários competitivos e de treinamento.

REFERÊNCIAS

ALZUBAIDI, Laith *et al.* Review of deep learning: Concepts, CNN architectures, challenges, applications, future directions. **Computer Science Review**, Elsevier, v. 40, p. 100394, 2021.

BERWICK, Andrew. **Machine Learning Essentials: A Practical Guide for Beginners**. [S.l.]: TechPress Publishing, 2019.

CHANG, Chih-Chung; LIN, Chih-Jen. LIBSVM: A library for support vector machines. **ACM Transactions on Intelligent Systems and Technology (TIST)**, ACM, v. 2, n. 3, p. 1–27, 2011.

CORTES, Corinna; VAPNIK, Vladimir. Support-vector networks. **Machine learning**, Springer, v. 20, n. 3, p. 273–297, 1995.

DHILLON, Amandeep; VERMA, Gurpreet. Convolutional neural networks: A comprehensive review. **Computational Intelligence and Neuroscience**, Hindawi, 2019.

ECHEVERRIA, J.; SANTOS, O. C. Toward Modeling Psychomotor Performance in Karate Combats Using Computer Vision Pose Estimation. **Sensors**, v. 21, p. 8378, 2021. DOI: 10.3390/s21248378. Disponível em: <https://doi.org/10.3390/s21248378>.

FIFA. **Football Technology: Standards for Electronic Performance and Tracking Systems (EPTS)**. [S.l.: s.n.], 2024. Acesso em: 18 nov. 2024. Disponível em: <https://inside.fifa.com/technical/football-technology/standards/epts>.

GOODFELLOW, Ian; BENGIO, Yoshua; COURVILLE, Aaron. **Deep Learning**. Cambridge, MA: MIT Press, 2016. Disponível em: <https://www.deeplearningbook.org>.

GUO, J. *et al.* An Attention Enhanced Spatial–Temporal Graph Convolutional LSTM Network for Action Recognition in Karate. **Applied Sciences**, v. 11, p. 8641, 2021. DOI: 10.3390/app11188641. Disponível em: <https://doi.org/10.3390/app11188641>.

HACHAJ, T.; OGIELA, M. R.; KOPTYRA, K. Application of Assistive Computer Vision Methods to Oyama Karate Techniques Recognition. **Symmetry**, v. 7,

p. 1670–1698, 2015. DOI: 10.3390/sym7041670. Disponível em:
<https://doi.org/10.3390/sym7041670>.

INTERNATIONAL OLYMPIC COMMITTEE. **Know Your Sport: Taekwondo - Rules, Scoring, Equipment**. [S.l.: s.n.], 2024. Acesso em: 29 jul. 2024. Disponível em:
<https://olympics.com/en/news/know-your-sport-taekwondo-rules-scoring-equipment>.

KARATE DO BRASIL. **História do Karate**. [S.l.: s.n.], 2024. Acesso em: 22 jul. 2024. Disponível em: <https://www.karatedobrasil.com/historia>.

KASIRI, S. *et al.* Fine-grained action recognition of boxing punches from depth imagery. **Computer Vision and Image Understanding**, v. 159, p. 143–153, 2017. DOI: 10.1016/j.cviu.2017.04.007. Disponível em:
<https://doi.org/10.1016/j.cviu.2017.04.007>.

KHAN, Salman *et al.* A survey on the recent developments in vision transformers. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, IEEE, 2020.

KRIZHEVSKY, Alex; SUTSKEVER, Ilya; HINTON, Geoffrey E. Imagenet classification with deep convolutional neural networks. **Communications of the ACM**, ACM, v. 60, n. 6, p. 84–90, 2017.

LECUN, Yann; BENGIO, Yoshua; HINTON, Geoffrey. Deep Learning. **Nature**, v. 521, n. 7553, p. 436–444, 2015. DOI: 10.1038/nature14539. Disponível em:
<https://www.nature.com/articles/nature14539>.

LEE, J.; JUNG, H. TUHAD: Taekwondo Unit Technique Human Action Dataset with Key Frame-Based CNN Action Recognition. **Sensors**, v. 20, p. 4871, 2020. DOI: 10.3390/s20174871. Disponível em: <https://doi.org/10.3390/s20174871>.

NAIK, Banoth Thulasya; HASHMI, Mohammad Farukh; BOKDE, Neeraj Dhanraj. A Comprehensive Review of Computer Vision in Sports: Open Issues, Future Trends and Research Directions. **arXiv preprint arXiv:2203.02281**, 2022. Disponível em:
<https://arxiv.org/abs/2203.02281>.

OWENS, John D. *et al.* GPU Computing. **Proceedings of the IEEE**, v. 96, n. 5, p. 879–899, 2008. DOI: 10.1109/JPROC.2008.917757. Disponível em:
<https://escholarship.org/uc/item/0cv1p1nc>.

REDMON, Joseph *et al.* You Only Look Once: Unified, Real-Time Object Detection. *In: PROCEEDINGS of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. [S.l.: s.n.], 2016. P. 779–788.

SINGH, Neha; SABROL, Roopesh. A comprehensive review of convolutional neural networks. **Journal of Computational Intelligence and Electronic Systems**, Springer, v. 12, n. 3, p. 12–24, 2021.

SONG, Liangchen *et al.* Human pose estimation and its application to action recognition: A survey. **Journal of Visual Communication and Image Representation**, v. 76, p. 103055, 2021. Acesso em: 1 out. 2024. ISSN 1047-3203. DOI: 10.1016/j.jvcir.2021.103055. Disponível em: <https://doi.org/10.1016/j.jvcir.2021.103055>.

SZE, Vivienne *et al.* Efficient processing of deep neural networks: A tutorial and survey. **Proceedings of the IEEE**, IEEE, v. 105, n. 12, p. 2295–2329, 2017.

ULTRALYTICS. **YOLOv8 Documentation: Advanced Real-Time Object Detection and Classification**. [S.l.], 2024. Available at: <https://ultralytics.com>.

WORLD KARATE FEDERATION. **WKF Kumite Competition Rules 2024**. [S.l.: s.n.], 2024. Acesso em: 22 jul. 2024, p. 56, Apêndice 8.1. Disponível em: https://www.wkf.net/pdf/WKF_Kumite_Competition_Rules_2024.pdf.

_____. _____. [S.l.: s.n.], 2024. Acesso em: 22 jul. 2024, p. 23, Artigo 8.1. Disponível em: https://www.wkf.net/pdf/WKF_Kumite_Competition_Rules_2024.pdf.