

ESTUDO DE INTERAÇÃO HUMANO COMPUTADOR SEM TOQUE UTILIZANDO CÂMERAS COMO SENSORES E APRENDIZADO DE MÁQUINA

João Vitor Contim Mantovani (PIBIC/CNPq/FA/UEM), Heloise Manica Paris Teixeira (Orientadora). E-mail: hmpeteixeira@uem.br.

Universidade Estadual de Maringá, Centro de Tecnologia Maringá, PR.

Área e subárea do conhecimento: Ciência da Computação, Metodologia e Técnicas da Computação.

Palavras-chave: Interação Humano-Computador; Comandos Gestuais; Aprendizado de Máquina.

RESUMO

A área de Interação Humano-Computador (IHC) busca continuamente formas mais naturais de comunicação entre usuários e sistemas. A interação por gestos se destaca como uma alternativa promissora, especialmente em contextos em que o toque deve ser evitado, como em ambiente hospitalar onde o controle de infecções é essencial. No entanto, soluções que exigem hardware como o Leap Motion, se aplicado em larga escala podem ter um custo elevado. Este trabalho apresenta o desenvolvimento e a avaliação de um sistema para reconhecimento de gestos utilizando uma webcam e técnicas de visão computacional. O sistema proposto emprega o framework *MediaPipe* para extrair os pontos de referência das mãos e utiliza uma engenharia de features baseada em distâncias euclidianas entre pontos-chave da mão, garantindo robustez à rotação. Foram treinados e comparados diferentes modelos de aprendizado de máquina (SVM, KNN e Random Forest) com a biblioteca *Scikit-learn*. Os resultados demonstram a viabilidade da abordagem, alcançando boas taxas de acurácia e oferecendo uma alternativa acessível para a criação de interfaces sem toque.

INTRODUÇÃO

A Interação Humano-Computador (IHC) busca otimizar a comunicação com sistemas (Barbosa et al., 2021), sendo os gestos uma forma de interação natural e intuitiva (Wu et al., 2023), como demonstrado por tecnologias como o Leap Motion, um sensor suporta movimentos de mãos e dedos como entrada, de forma análoga a um mouse, mas não requer contato ou toque manual, método utilizado no trabalho de Felizardo et al. (2024). Contudo, a necessidade de hardware especializado representa uma barreira de custo, principalmente em aplicações de interface gestual em grande escala, como por exemplo em hospitais da rede pública que possuem recursos financeiros limitados.

Com aos avanços em visão computacional (Zhao et al., 2024), hoje é viável utilizar webcams comuns para capturar e interpretar gestos, uma possibilidade já explorada em trabalhos como os de Ansari e Singh (2019) e Coimbra et al. (2014), tornando a tecnologia mais acessível. A solução proposta neste estudo pode ser aplicada em diferentes contextos, como ambientes críticos hospitalares, centros cirúrgicos e UTIs; cabines de voo; sistemas de metrô; estações de bombeiros, e demais setores críticos em que a comunicação e interação com os sistemas precisam ser simples e efetivas de modo a evitar acidentes e/ou contaminações.

O objetivo desta pesquisa é desenvolver um sistema de reconhecimento de gestos de baixo custo utilizando uma webcam padrão. O projeto emprega técnicas de visão computacional e aprendizado de máquina para permitir que um computador seja operado por meio de gestos, eliminando a necessidade de equipamentos extra como sensores.

MATERIAIS E MÉTODOS

O sistema foi desenvolvido em Python, utilizando um conjunto de ferramentas de código aberto. A captura de vídeo foi realizada com a biblioteca OpenCV, enquanto o framework MediaPipe, do Google, foi o núcleo da abordagem, sendo usado para extrair os 21 pontos de referência da mão em tempo real. A partir desses pontos, foi criado um dataset personalizado com 3.600 amostras de 3 gestos distintos. A principal inovação, inspirada em trabalhos como o de Felizardo et al. [2024], foi uma engenharia de features que calcula as distâncias euclidianas entre pontos-chave da mão, gerando uma representação do gesto que demonstrou ser robusta à rotação.

RESULTADOS E DISCUSSÃO

A Figura 1 ilustra o vocabulário de gestos proposto. Com os dados tratados, foram treinados e comparados três modelos de aprendizado de máquina (SVM, KNN e Random Forest) com a biblioteca Scikit-learn. O modelo escolhido foi então implementado em um sistema de controle final que traduz os gestos em comandos de computador usando PyAutoGUI. Este sistema inclui um limiar de confiança de 90% para criar um estado "neutro" e evitar ações acidentais.

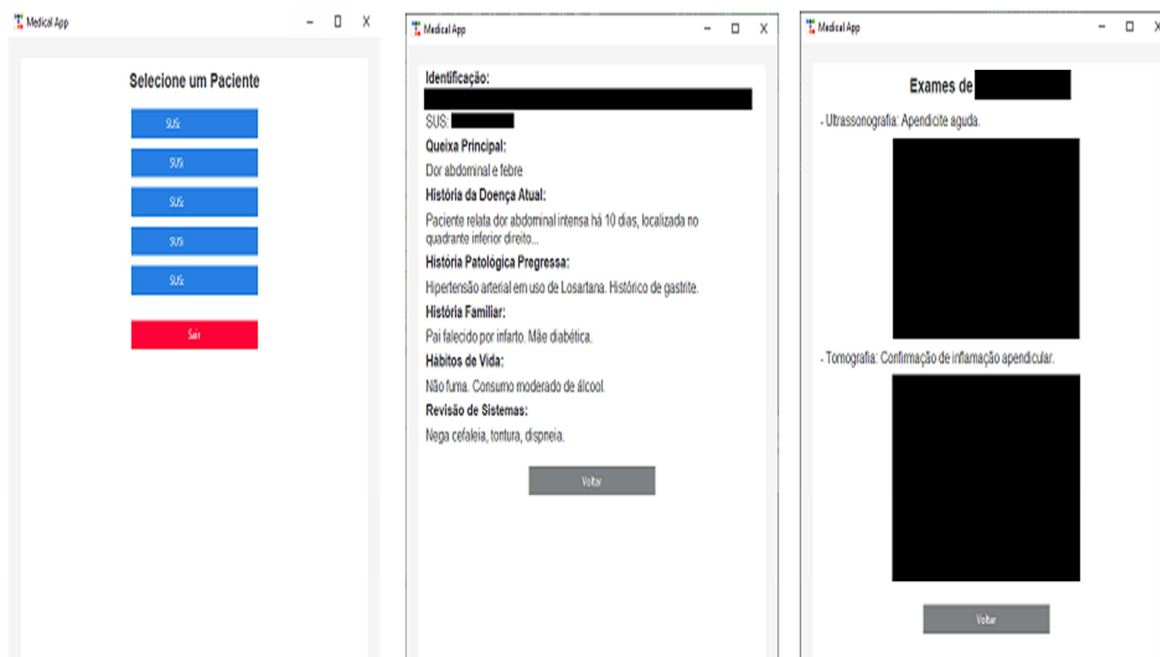
Imagem 1 – Vocabulário de gestos implementado no sistema e as ações correspondentes.



Fonte: Autor

O experimento prático foi realizado por meio de um protótipo que simula a manipulação de uma ficha de paciente em um sistema médico, demonstrando sua aplicabilidade em ambientes que exigem interação sem toque. A Figura 2 ilustra algumas interfaces do sistema simulado.

Imagem 2 – Interfaces do sistema médico simulado.



Fonte: Autor

Após o treinamento, os três modelos foram avaliados. O modelo SVM mostrou-se estável, mas menos preciso em gestos rápidos. O KNN demonstrou rapidez, mas sensibilidade a pequenas variações, enquanto o Random Forest apresentou maior precisão e robustez geral. A Tabela 1 compila os resultados de acurácia obtidos no conjunto de teste.

Tabela 1 - Tabela Comparativa de Acurácia dos Modelos

Modelo	Acurácia de Teste
SVC/SVM	91.39%
KNN	96.39%
Random Forest	98.06%

Fonte: Autor.

CONCLUSÕES

O presente estudo demonstra a viabilidade do uso de câmeras Web como uma alternativa acessível para a criação de interfaces sem toque. Os resultados demonstram a alta viabilidade da abordagem, com todos os modelos alcançando acurácia superior a 91%. O modelo Random Forest destacou-se com o melhor desempenho, atingindo 98.06% de acurácia, indicando uma capacidade superior de generalização.

AGRADECIMENTOS

Agradecemos à Fundação Araucária pela bolsa concedida por meio do Programa PIBIC/CNPq-FUNDAÇÃO ARAUCÁRIA-UEM.

REFERÊNCIAS

ANSARI, M. A.; SINGH, D. K. An approach for human machine interaction using dynamic hand gesture recognition. *In*: IEEE CONFERENCE ON INFORMATION AND COMMUNICATION TECHNOLOGY, 2019, Allahabad. **Anais** [...]. Allahabad: IEEE, 2019. p. 1-6.

BARBOSA, S. D. J. et al. **Interação humano-computador e experiência do usuário**. São Paulo: Câmara Brasileira do Livro, 2021.

COIMBRA, V. et al. Implementação do jogo rock-paper-scissors-lizard-spock utilizando reconhecimentos de gestos via k-means e redes neurais artificiais. *In*: SIMPÓSIO BRASILEIRO DE JOGOS E ENTRETENIMENTO DIGITAL, 13., 2014, Porto Alegre. **Anais** [...]. Porto Alegre: SBC, 2014.

FELIZARDO, H. S. et al. **Machine learning for gesture recognition using leap motion controller**. 2024. Relatório Técnico – Universidade Estadual de Maringá, Maringá, 2024.

34º Encontro Anual de Iniciação Científica
14º Encontro Anual de Iniciação Científica Júnior



12 e 13 de novembro de 2025

WU, D. et al. Virtual model interaction based on single rgb camera. *In: IEEE INTERNATIONAL CONFERENCE ON MULTIMEDIA AND EXPO, 2023, Brisbane. Anais [...].* Brisbane: IEEE, 2023.

ZHAO, X. et al. A review of convolutional neural networks in computer vision. **Artificial Intelligence Review**, v. 57, n. 4, 2024.