Journal of Health Informatics

J Health Inform. 2026;18:e1535
DOI: https://doi.org/10.59681/2175-4411.v18.2026.1535

ARTIGO ORIGINAL

Avaliação de representações simbólicas de movimentos articulares humanos úteis para sistemas de telemedicina

Evaluation of symbolic representations of human joint movements helpful for telemedicine systems

Evaluación de representaciones simbólicas de movimientos articulares humanos útiles para sistemas de telemedicina

Ana Paula Merencia[1], Huei Diana Lee[1], Weber Shoity Resende Takaki[1], Newton Spolaôr[1], Alexandre Peiter Ferraz[1], Wu Feng Chung[1]

ORCID: 0009-0004-0782-4667 · 0000-0002-2189-1047 · 0000-0002-3540-5732 · 0000-0003-0748-3693 · 0009-0000-4641-7707 · 0000-0001-6032-9260

[1] Laboratório de Bioinformática (LABI), Universidade Estadual do Oeste do Paraná (UNIOESTE), Foz do Iguaçu (PR), Brasil.

FICHA DO ARTIGO

TIPO
Artigo Original

IDENTIFICADOR
e1535

DOI
10.59681/2175-4411.v18.2026.1535

RECEBIDO
21/08/2025

APROVADO
22/02/2026

PUBLICADO
01/04/2026

PROCESSO DE AVALIAÇÃO
Double Blind Review

EDITOR RESPONSÁVEL
Juliano de Souza Gaspar

ACESSO
Aberto, CC BY 4.0

AUTOR CORRESPONDENTE
Ana Paula Merencia
anamerencia06@gmail.com

LICENÇA

Este é um artigo publicado em acesso aberto sob a licença Creative Commons CC BY 4.0, que permite uso, distribuição e reprodução em qualquer meio, desde que o trabalho original seja corretamente citado.

CC BY 4.0

RESUMO

Objetivo: Avaliar o desempenho dos métodos *Symbolic Aggregate approXimation* e *Symbolic Fourier Approximation* na representação simbólica de séries temporais de movimentos articulares humanos. Foram considerados três tipos de movimento (flexão/extensão, abdução/adução e circundução) e três tamanhos de alfabeto (três, sete e dez símbolos). **Métodos:** Oito voluntários saudáveis realizaram os movimentos com um *smartphone* fixado ao punho para coleta de aceleração. Os sinais foram suavizados com filtro Gaussiano (σ = 2), segmentados e convertidos em palavras simbólicas pelas abordagens avaliadas. **Resultados:** O desempenho dos métodos variou conforme o movimento. Para abdução/adução, *Symbolic Fourier Approximation* apresentou menores distâncias médias de Hamming, com diferença estatisticamente significativa. Para flexão/extensão, *Symbolic Aggregate approXimation* foi significativamente superior. Para circundução, *Symbolic Fourier Approximation* apresentou menor média, porém sem superioridade estatística. Na síntese entre as melhores configurações por movimento, a menor distância média global foi observada em uma configuração do *Symbolic Fourier Approximation* (circundução). **Conclusão:** A avaliação indica que não há um método universalmente superior; a escolha do método e da parametrização deve considerar o tipo de movimento e a consistência entre sujeitos. Esses achados apoiam a seleção de configurações específicas por movimento em aplicações de saúde digital, como monitoramento remoto, mHealth, Telessaúde e apoio à decisão clínica.

Descritores: Estudos de Séries Temporais; Reconhecimento Automatizado de Padrão; Telemedicina.

ABSTRACT

Objective: To evaluate the performance of the Symbolic Aggregate approXimation and Symbolic Fourier Approximation methods in the symbolic representation of time series derived from human joint movements. Three movement types (flexion/extension, abduction/adduction, and circumduction) and three alphabet sizes (three, seven, and ten symbols) were considered. **Methods:** Eight healthy volunteers performed the movements with a smartphone attached to the wrist to collect acceleration data. Signals were smoothed using a Gaussian filter (σ = 2), segmented, and converted into symbolic words using the evaluated approaches. **Results:** Method performance varied across movements. For abduction/adduction, Symbolic Fourier Approximation yielded lower mean Hamming distances, with a statistically significant difference. For flexion/extension, Symbolic Aggregate approXimation was significantly superior. For circumduction, Symbolic Fourier Approximation showed a lower mean, but without statistical superiority. When synthesizing the best configurations for each movement, the lowest overall mean distance was observed in a Symbolic Fourier Approximation configuration (circumduction). **Conclusion:** The evaluation indicates that there is no universally superior method; the choice of method and parameterization should consider the movement type and consistency across subjects. These findings support selecting movement-specific configurations in digital health applications, such as remote monitoring, mHealth, telehealth, and clinical decision support.

Keywords: Time Series Studies; Automated Pattern Recognition; Telemedicine.

RESUMEN

Objetivo: Evaluar el desempeño de los métodos Symbolic Aggregate approXimation y Symbolic Fourier Approximation en la representación simbólica de series temporales provenientes de movimientos articulares humanos. Se consideraron tres tipos de movimiento (flexión/extensión, abducción/aducción y circunducción) y tres tamaños de alfabeto (tres, siete y diez símbolos). **Métodos:** Ocho voluntarios sanos realizaron los movimientos con un smartphone fijado a la muñeca para la recolección de datos de aceleración. Las señales se suavizaron con un filtro gaussiano (σ = 2), se segmentaron y se convirtieron en palabras simbólicas mediante los enfoques evaluados. **Resultados:** El desempeño de los métodos varió según el movimiento. Para abducción/aducción, Symbolic Fourier Approximation presentó menores distancias medias de Hamming, con diferencia estadísticamente significativa. Para flexión/extensión, Symbolic Aggregate approXimation fue significativamente superior. Para circunducción, Symbolic Fourier Approximation presentó una media menor, pero sin superioridad estadística. En la síntesis entre las mejores configuraciones por movimiento, la menor distancia media global se observó en una configuración de Symbolic Fourier Approximation (circunducción). **Conclusión:** La evaluación indica que no existe un método universalmente superior; la elección del método y de la parametrización debe considerar el tipo de movimiento y la consistencia entre sujetos. Estos hallazgos respaldan la selección de configuraciones específicas por movimiento en aplicaciones de salud digital, como el monitoreo remoto, mHealth, telesalud y el apoyo a la decisión clínica.

Descriptores: Estudios de Series Temporales; Reconocimiento Automatizado de Patrones; Telemedicina.

COMO CITAR ESTE ARTIGO: Merencia AP, Lee HD, Takaki WSR, Spolaôr N, Ferraz AP, Chung WF. Avaliação de representações simbólicas de movimentos articulares humanos úteis para sistemas de telemedicina. J Health Inform 2026;18:e1535. doi: https://doi.org/10.59681/2175-4411.v18.2026.1535

1 INTRODUÇÃO

Nos últimos anos, observa-se um crescente interesse no monitoramento e no reconhecimento de movimentos humanos aplicados a sistemas de informação em saúde. Motivos incluem o avanço de tecnologias portáteis e a necessidade de soluções eficientes em áreas como reabilitação, ergonomia, esportes, condicionamento físico e diagnóstico precoce de lesões. Adicionalmente, no Brasil, as Doenças Osteomusculares Relacionadas ao Trabalho (DORT) representam uma das principais causas de afastamento laboral, com mais de 67 mil casos registrados entre 2007 e 2016, refletindo o impacto dessas lesões sobre a saúde e a produtividade da população[1,2]. Nesse contexto, sensores inerciais, como acelerômetros, destacam-se por sua portabilidade, custo acessível e capacidade de capturar dados relevantes, como aceleração, velocidade, deslocamento e vibrações, originando Séries Temporais (ST) que descrevem padrões de movimento ao longo do tempo[2].

Entretanto, a análise direta dessas séries apresenta desafios significativos, como a alta dimensionalidade, a variabilidade inter e intraindivíduo e a complexidade intrínseca dos sinais registrados em múltiplos eixos do espaço 3D[3-5]. Tais fatores dificultam a extração de informações consistentes e a integração automatizada desses dados a sistemas computacionais de apoio (i) à prevenção e ao diagnóstico de lesões em articulações e partes do corpo, e (ii) à decisão clínica. Nesse cenário, a aplicação de métodos simbólicos de representação de ST é uma alternativa promissora de pré-processamento de dados para promover a compressão, o reconhecimento inteligente e o monitoramento automático de padrões motores em sistemas de saúde digital.

Esses métodos convertem sinais contínuos em sequências discretas de símbolos, preservando suas características essenciais e permitindo análises posteriores eficientes, com baixo custo computacional[6]. Entre os principais algoritmos simbólicos propostos, destacam-se duas abordagens complementares: (i) os métodos baseados em aproximação no domínio do tempo, como o *Piecewise Aggregate Approximation* (PAA) e o *Symbolic Aggregate approXimation* (SAX) [7], e (ii) os métodos baseados em decomposição no domínio da frequência, como a *Discrete Fourier Transform* (DFT) e o *Symbolic Fourier Approximation* (SFA)[8].

O SAX é amplamente usado por sua simplicidade, eficiência e aplicabilidade em séries longas, permitindo o uso de medidas de similaridade categórica, como a distância de Hamming, em tarefas de análise inteligente de sinais, como classificação e detecção de anomalias. Já o SFA, ao aplicar a Transformada Rápida de Fourier (FFT) e discretizar os coeficientes dominantes com base na energia do sinal, tende a capturar informações no domínio da frequência relevantes, fornecendo retratos mais consistentes para a análise e a indexação de movimentos cíclicos e complexos[8].

Até o momento, não foram encontrados na literatura científica estudos que realizassem uma comparação sistemática entre os métodos simbólicos SAX e SFA aplicados à representação de movimentos articulares humanos. Embora existam trabalhos que empregam outros métodos simbólicos para análise de movimentos[9,10], observa-se uma lacuna de conhecimento quanto à avaliação comparativa do desempenho das tradicionais abordagens SAX e SFA nesse domínio.

Essa lacuna, atacada neste trabalho, é especialmente relevante para profissionais da área de informática em saúde. De fato, a identificação de um algoritmo adequado para representar e comparar movimentos de forma simbólica e eficiente pode ser útil na criação de sistemas e políticas de saúde digital para várias finalidades, tais como telemonitoramento fisioterápico de pacientes, apoio a especialistas na decisão clínica, avaliação motora remota via Internet e acompanhamento de alunos de academias públicas e privadas.

Em particular, métodos investigados no presente trabalho têm potencial para contribuir na caracterização de dados temporais de movimentações de indivíduos. Esses dados podem ser enviados como entrada para ferramentas computacionais em saúde já existentes ou que venham a ser propostas.

Vale ressaltar que a acessibilidade desses sistemas, em termos de redução de custos e ampliação do público-alvo, pode ser promovida por duas ações exploradas neste trabalho: (i) coletar ST de movimentos usando sensores embarcados em *smartphones*, e (ii) acessar o Sistema de Monitoramento de Movimentos e Análise Remota em Telemedicina (SMMAR-T)[11] para aquisição, transmissão e centralização segura de dados adquiridos. Assim, a integração entre (i) técnicas de representação simbólica de séries e (ii) sistemas de informação em saúde oferece uma perspectiva inovadora para o monitoramento automatizado de parâmetros biomecânicos e funcionais de sujeitos ao longo do tempo.

Este estudo tem como objetivo avaliar o desempenho dos métodos SAX e SFA na representação simbólica de ST provenientes de movimentos articulares humanos, considerando três tipos de movimento e três tamanhos de alfabeto. Em particular, esses métodos convertem sinais contínuos em sequências discretas de símbolos de um alfabeto com tamanho específico, preservando suas características essenciais e permitindo análises posteriores eficientes, com baixo custo computacional. O ambiente experimental controlado considerou séries reais de oito indivíduos saudáveis, coletadas neste estudo via *smartphone*.

Tabela 1. Dados descritivos e demográficos dos oito indivíduos.

Sujeito	Idade	Sexo	Profissão	Atividades físicas por semana	Consumo de álcool por semana	Peso (kg)	Altura (m)	Comprimento (cm)	
								Membro Superior Direito	Membro Superior Esquerdo
1	25	M	Programador	6	0	72	1,71	76	76
2	25	F	Estudante	6	0	64	1,64	73	73
3	24	F	Enfermeira	5	1	63	1,59	69	69
4	19	M	Estudante	6	1	86	1,80	78	78
5	25	F	Enfermeira	5	2	63	1,58	67	67
6	18	M	Estudante	5	1	90	1,75	72	72
7	24	M	Estudante	5	0	130	1,81	80	80
8	18	F	Estudante	5	0	64	1,70	73	73

Fonte: Elaborado pelos autores (2026).

O foco do presente estudo é verificar se as representações simbólicas geradas por SAX e SFA apresentam consistência e produzem sequências similares quando o mesmo movimento é executado sob condições equivalentes, com repetições do protocolo e perfil homogêneo de participantes hígidos, priorizando a validade interna para comparação entre métodos.

A contribuição deste trabalho na representação simbólica de dados temporais de movimentos sadios é relevante à literatura porque padrões encontrados nesses sujeitos formam referências que podem ser comparadas, futuramente, com padrões de pacientes com anormalidades musculoesqueléticas. Portanto, essa investigação inovadora é um passo inicial essencial para a futura aplicação dos melhores métodos identificados em sistemas de monitoramento, telemedicina e análise de séries de movimentos do corpo humano aplicados à saúde digital.

2 MÉTODOS

Oito voluntários saudáveis, não tabagistas, com informações antropomórficas e funcionais mostradas na Tabela 1, executaram três tipos de movimentos: flexão/extensão dos cotovelos, abdução/adução e circundução dos ombros[12]. Os dados de aceleração foram obtidos por um acelerômetro com frequência de amostragem de 10Hz, em um *smartphone* Samsung Galaxy A51 fixado ao punho deles e conectado ao SMMAR-T, conforme descrito na dissertação que originou este estudo[13]. Para cada movimento articular, cada participante realizou dez ciclos por período, em três períodos distintos, totalizando 30 ciclos por participante em cada movimento. Considerando os oito voluntários e os três movimentos avaliados, o conjunto de dados construído neste trabalho resultou em 720 ciclos no total (8 participantes × 30 ciclos × 3 movimentos).

A análise foi conduzida em quatro níveis: ciclo, período, sujeito e grupo. Após a segmentação automática, cada ciclo foi representado por SAX e SFA, com alfabetos de 3, 7 e 10 símbolos. Em seguida, as palavras simbólicas foram comparadas por distância de Hamming entre ciclos. Depois, as distâncias dos 30 ciclos de cada participante foram agregadas em uma média por sujeito, gerando um valor de desempenho por método e configuração. Por fim, os métodos foram comparados no nível do grupo, sintetizando os resultados dos participantes.

O protocolo de coleta de dados completo, incluindo critérios de inclusão e informações complementares, está descrito em[13]. Este estudo foi aprovado pelo Comitê de Ética em Pesquisa com Seres Humanos da Universidade Estadual do Oeste do Paraná, sob o número de protocolo 34323820.9.0000.0107. Todos os participantes assinaram o Termo de Consentimento Livre e Esclarecido.

Após suavização dos sinais com filtro Gaussiano (σ=2) [14] e segmentação automática dos ciclos baseada nos cruzamentos de zero da derivada, cada ciclo foi representado pelos métodos SAX e SFA, utilizando palavras simbólicas de dez letras e três diferentes tamanhos de alfabeto: três, sete e dez.

2.1 SAX

O método SAX em sua forma Padrão (SAXP), proposto em[7], atua no domínio do tempo. SAXP é constituído pelas cinco etapas principais descritas a seguir:

Normalização z-score

Inicialmente, a ST $S = \{s_1, s_2, \ldots, s_n\}$ é normalizada para média zero e desvio padrão um[7]. A Etapa 1 garante que diferentes séries sejam comparáveis independentemente de amplitude ou tendência.

Segmentação da série

A série normalizada é dividida em w segmentos de igual tamanho. Para cada segmento j, calcula-se a média $\bar{s}_j$, formando uma representação intermediária

chamada PAA, que é calculada como definido na Equação 1.

$$\bar{s}_j = \frac{1}{l} \sum_{i=(j-1)l+1}^{jl} \hat{s}_i, \text{ com } l = \frac{n}{w} \quad (1)$$

onde l representa a quantidade de amostras por segmento e w representa a quantidade total de segmentos em que a ST normalizada $\hat{S}$ será dividida.

Discretização das médias

Cada média segmentada $\bar{s}_j$ é transformada em um símbolo pertencente a um alfabeto de tamanho a. Utilizam-se *breakpoints* β_i baseados na distribuição normal padrão e definidos na Equação 2.

$$\beta_i = \Phi^{-1}\left(\frac{i}{a}\right), \; i = 1, 2, \ldots, a-1 \quad (2)$$

onde Φ^{-1} é a função inversa da Função de Distribuição Acumulada da normal padrão.

Mapeamento simbólico

Cada valor é mapeado para um símbolo com base no intervalo entre *breakpoints* em que ele se encontra, como definido na Equação 3.

$$s_j = símbolos_k \text{ se } \beta_{k-1} < \bar{s}_j \leq \beta_k \quad (3)$$

com $\beta_0 = -\infty$ e $\beta_a = +\infty$. O alfabeto simbólico *símbolos* é constituído por a letras consecutivas do alfabeto latino, como $\{A, B, C, \ldots, Z\}$, e o índice k varia de 1 a a, indicando o intervalo de *breakpoint* no qual $\bar{s}_j$ se encontra.

Formação da palavra SAX

A sequência $\{s_1, s_2, \ldots, s_w\}$ constitui a palavra SAX correspondente à série S, fornecendo uma representação simbólica de comprimento fixo que mantém a estrutura temporal e simplifica a comparação entre diferentes movimentos. Neste trabalho, SAXP foi aplicado com um tamanho fixo de palavra w = 10, *i.e.*, cada ciclo de movimento foi representado por uma sequência de 10 símbolos. O tamanho do alfabeto a, que determina o nível de granularidade da discretização, foi variado com os seguintes valores, pois a exploração de diferentes tamanhos é recomendada[9,10]:

- a = 3: representação de mais alto nível, dividindo os valores em apenas três faixas – baixo, médio e alto;
- a = 7: discretização intermediária;
- a = 10: representação mais detalhada, permitindo maior sensibilidade às variações sutis nos ciclos de movimento.

2.2 SFA

O método SFA em sua forma Padrão (SFAP) foi proposto em[8] e utiliza informações do domínio da frequência, extraídas por meio da FFT. SFAP é constituído pelas seguintes etapas:

Normalização z-score

Inicialmente, a ST é normalizada para média zero e desvio padrão um[7], de modo análogo ao método SAX.

Transformada de Fourier e seleção de coeficientes

Aplica-se a FFT à série normalizada, obtendo-se um conjunto de coeficientes complexos $C = \{c_1, c_2, \ldots, c_n\}$ que representam a contribuição de diferentes frequências em S. A energia total E_{total} foi calculada pela soma do quadrado do módulo de cada coeficiente em C, como definido em[13].

Para reduzir a dimensionalidade mantendo as informações mais relevantes, foram selecionados os primeiros n_{coef} coeficientes que, juntos, representavam cumulativamente 90% da energia total da série. A energia acumulada foi computada de modo crescente, e o ponto de corte n_{coef} foi definido na Equação 4.

$$n_{coef} = \min\left\{ k \in \mathbb{N} \mid \sum_{i=1}^{k} |c_i|^2 \geq 0{,}9 . E_{total} \right\} \quad (4)$$

onde k indica o número mínimo de coeficientes necessários para que a soma da energia acumulada seja superior ou igual a 90% da energia total.

Representação vetorial dos coeficientes

Após a seleção dos coeficientes que acumulam 90% da energia, cada coeficiente complexo c_i foi convertido em um escalar representativo pela média entre suas partes real e imaginária[13]. Ressalta-se que o uso da média constitui uma questão de projeto dentro do processo de engenharia de atributos avaliado neste trabalho. Embora alternativas como magnitude e potência também sejam possíveis, elas produzem valores não negativos. Neste estudo, optou-se por manter uma combinação simples que preserva o sinal[13].

Discretização dos valores médios

Para a discretização são utilizados os *breakpoints* baseados na distribuição normal padrão, conforme definido na Equação 2.

Mapeamento simbólico e formação da palavra SFA

Cada valor médio foi mapeado para um símbolo do alfabeto latino, de acordo com o intervalo delimitado pelos *breakpoints* em que ele se encontrava (Equação 4). Neste trabalho, os parâmetros w e a foram variados igualmente no SFA e no SAX.

Distância de Hamming

Para quantificar a diferença entre as representações simbólicas dos ciclos analisados, foi usada a distância de Hamming[15]. Matematicamente, dadas duas sequências simbólicas $S = \{s_1, s_2, \ldots, s_n\}$ e $T = \{t_1, t_2, \ldots, t_n\}$, ambas de comprimento n, a distância de Hamming $d_H(S, T)$ é definida na Equação 5.

$$d_H(S,T) = \sum_{i=1}^{n} \delta(s_i, t_i), \text{ onde } \delta(s_i, t_i) = \begin{cases} 0, & \text{se } s_i = t_i \\ 1, & \text{se } s_i \neq t_i \end{cases} \quad (5)$$

2.3 Sequência padrão do movimento

Para cada movimento (e respectiva configuração de lateralidade e alfabeto), definimos uma sequência padrão (palavra simbólica de referência) a partir da frequência de ocorrência dos símbolos em cada posição. Especificamente, em cada posição, selecionou-se o símbolo mais recorrente, resultando em uma palavra representativa do padrão observado sob condições controladas. A construção da referência usada nos cálculos foi realizada conforme a estratégia de validação *Leave One Subject Out* (LOSO): em cada iteração, um participante foi mantido como teste e a sequência padrão foi construída apenas com os dados dos demais participantes (treino). Em seguida, os ciclos do participante excluído foram comparados à sequência padrão do treino por meio da distância de Hamming, e os valores foram agregados no nível do participante.

2.4 Análise estatística

Aplicou-se LOSO para construir a sequência simbólica de referência sem incluir o participante avaliado em cada iteração dessa estratégia. As distâncias de Hamming foram agregadas por sujeito (n=8), por configuração (movimento × lateralidade × alfabeto) e por método (SAX e SFA). A comparação SAX *vs* SFA foi pareada por configuração. O resultado de Shapiro–Wilk[16] determinou o uso do teste t pareado[16] ($p \geq 0{,}05$) ou Wilcoxon pareado[16] ($p < 0{,}05$), com ajuste de p por Holm[16] ($\alpha = 0{,}05$). O "melhor por movimento" foi definido pela configuração com menor média, priorizando as combinações que apresentaram significância após Holm. O "melhor dos melhores" (uma configuração por movimento) foi identificado entre movimentos com ANOVA de medidas repetidas[16] ou Friedman[16], com pós-teste pareado e correção por Holm[17].

3 RESULTADOS E DISCUSSÃO

Cabe ressaltar que a escolha do limiar de 90% da energia para SFA equilibrou a preservação da informação significativa dos sinais e a redução da dimensionalidade. Assim, apenas os coeficientes com maior contribuição energética — e, portanto, maior representatividade da dinâmica do movimento — foram usados para representação, minimizando o impacto de componentes de baixa amplitude ou ruído[13].

SAXP apresenta complexidade de comportamento linear, proporcional ao tamanho da série ($O(n)$), uma vez que realiza apenas operações de normalização, cálculo de médias e discretização no domínio do tempo. Já o SFAP possui custo $O(n \log n)$. Logo, SAXP é mais eficiente para aplicações em tempo real e telemonitoramento.

Na Figura 1 são exibidas as representações simbólicas (*heatmaps*) geradas pelo SAXP para o movimento de flexão e extensão do cotovelo direito do Sujeito 1, com alfabeto de três símbolos. O realce em âmbar indica a frequência máxima de ocorrência de um símbolo em determinada posição da sequência, representando o padrão mais comum identificado entre os ciclos de movimento. Cada ciclo realizado em cada um dos três períodos é identificado por "C_i", com i variando de 1 a 10. No material suplementar disponível em: https://doi.org/10.5281/zenodo.18645054 estão todos os *heatmaps* gerados por SAXP e SFAP neste trabalho.

Os resultados mostraram que não há um método superior em todos os casos: o SFAP apresentou melhor desempenho em abdução/adução, com diferença estatisticamente significativa, enquanto o SAXP foi superior em flexão/extensão, também com significância estatística. Na circundução, o SFAP apresentou menor média de distância de Hamming, porém sem significância estatística após a correção, sendo esse achado interpretado apenas como tendência. Esse resultado é relevante, pois sugere que movimentos mais complexos podem reduzir o tamanho do efeito entre métodos e aumentar a sobreposição de desempenho, como observado na circundução.

Em seguida, foram comparadas as três configurações selecionadas como representativas de cada movimento, com base nas médias de distância de Hamming agregadas por participante, a fim de obter uma síntese comparativa entre as configurações vencedoras. Nessa

Figura 1. Uma representação simbólica gerada pelo método SAXP para o Sujeito 1.

C1	C	C	B	A	A	A	A	B	C	C
C2	C	C	B	A	A	A	A	B	C	C
C3	C	C	B	A	A	A	A	B	C	C
C4	C	C	B	A	A	A	A	B	C	C
C5	C	C	A	A	A	A	B	C	C	C
C6	C	C	B	A	A	A	A	B	C	C
C7	C	C	B	A	A	A	A	B	C	C
C8	C	C	B	A	A	A	A	B	B	C
C9	C	B	A	A	A	A	A	B	C	C
C10	C	C	A	A	A	A	A	B	C	C

C	C	B	A	A	A	A	B	C	C
C	C	B	A	A	A	A	B	C	C
C	C	B	A	A	A	A	B	C	C
C	C	B	A	A	A	A	B	C	C
C	C	B	A	A	A	A	B	C	C
C	C	B	A	A	A	A	B	C	C
C	C	B	A	A	A	A	B	C	C
C	C	B	A	A	A	A	B	C	C
C	C	B	A	A	A	A	B	C	C
C	C	B	A	A	A	A	B	C	C

C	C	B	A	A	A	A	B	C	C
C	C	B	A	A	A	A	B	C	C
C	C	B	A	A	A	A	B	C	C
C	C	B	A	A	A	A	B	C	C
C	C	B	A	A	A	A	B	C	C
C	C	B	A	A	A	A	B	C	C
C	C	B	A	A	A	A	B	C	C
C	C	B	A	A	A	A	B	C	C
C	C	B	A	A	A	A	B	C	C
C	C	B	A	A	A	A	B	C	C

Fonte: Elaborado pelos autores (2026).

etapa, a configuração de circundução (SFAP, alfabeto 3, lado esquerdo) apresentou a menor distância média global entre as três, seguida pela configuração de abdução/adução. Essa comparação deve ser interpretada como uma síntese entre configurações previamente selecionadas e em conjunto com os resultados inferenciais específicos de cada movimento, sobretudo nos casos em que não houve significância após a correção para múltiplas comparações.

Convém ressaltar que a análise de indivíduos saudáveis foi fundamental para a consolidação de uma sequência padrão de cada classe de movimento. Essa sequência constitui, neste trabalho, uma linha de referência biomecânica inovadora, a qual servirá como parâmetro para identificar desvios em populações com lesões musculoesqueléticas. Essa etapa é indispensável antes da aplicação clínica dos métodos para sistemas, pesquisas e políticas de saúde digital, pois permite caracterizar o comportamento normal das ST e verificar a consistência das representações sob condições controladas. As comparações estatísticas completas por configuração (incluindo p-valores brutos e ajustados) são apresentadas nas Tabelas 1, 2 e 3 (Material Suplementar).

Reconhece-se que o número de participantes (n=8) constitui uma limitação e pode restringir a generalização dos achados para outras populações e contextos. Entretanto, o estudo foi conduzido como uma prova de conceito em condições controladas, com protocolo padronizado e múltiplas repetições por participante (30 ciclos por movimento). Isso permitiu estimar padrões intraindivíduo de forma mais estável e comparar, de maneira consistente, o desempenho das representações simbólicas entre configurações e métodos. Como trabalhos futuros, recomenda-se (1) ampliar a amostra e sua diversidade em variáveis como idade, sexo e nível de atividade e (2) incluir novos movimentos e avaliar o desempenho em cenários clínicos, como por exemplos indivíduos com disfunções musculoesqueléticas, de modo a promover a aplicabilidade da proposta em saúde digital.

Quanto às implicações práticas e o potencial de aplicação dos métodos avaliados, este estudo diferencia-se dos trabalhos anteriores por integrar a análise simbólica de ST à avaliação de movimentos articulares humanos, considerando tanto a análise estatística quanto a viabilidade computacional de SAXP e SFAP. Embora existam referências que aplicam algoritmos de Mineração de Dados (MD) em sinais motores[2,3], não foram encontradas investigações que avaliem sistemática e comparativamente o desempenho desses dois métodos de representação em movimentos e tamanhos de alfabeto distintos, conferindo originalidade à abordagem proposta. Os achados inspiram ainda trabalhos futuros no SMMAR-T e em outros sistemas de informática em saúde. Essas pesquisas podem usar o SAXP e SFAP para pré-processar ST clínicas, antes da realização de tarefas de MD que suportam decisões de especialistas, como a classificação e a diferenciação de movimentos articulares sadios e anormais.

Portanto, este estudo reforça o potencial de integração do SAXP e SFAP em sistemas de saúde digital, especialmente em plataformas de telessaúde, *mHealth* e reabilitação motora, nas quais é necessário combinar processamento em tempo real e baixo custo computacional. Em particular, a representação simbólica dos sinais de sensores inerciais inspira o desenvolvimento de ferramentas automáticas e inteligentes de análise de movimento, as quais são capazes de apoiar a decisão clínica, monitorar a evolução funcional e personalizar o acompanhamento terapêutico de pacientes.

4 CONCLUSÃO

Este trabalho avaliou o desempenho de SAXP e SFAP na representação simbólica de séries temporais de movimentos articulares, considerando diferentes movimentos e tamanhos de alfabeto, com foco em eficiência e aplicabilidade em saúde digital. O SMMAR-T foi utilizado para aquisição, transmissão e centralização segura dos dados de oito voluntários. Na síntese entre as melhores configurações por movimento, a menor distância média global foi observada em uma configuração do SFA (circundução), ainda que SAX também foi destacado em alguns casos.

Além do avanço metodológico, os resultados indicam o potencial de SAX e SFA em Informática em Saúde, especialmente em reabilitação digital e telessaúde, ao reduzir o volume de dados transmitidos e processados sem perder informação biomecânica essencial para monitoramento remoto e apoio à decisão clínica.

A escolha por analisar indivíduos saudáveis e a construção de uma base com os dados coletados destes sujeitos formam uma etapa estratégica para estabelecer padrões biomecânicos inovadores de normalidade, os quais servirão como referência em futuros estudos com populações com lesões motoras. Essas referências serão essenciais para que desvios nos padrões de movimento possam ser detectados e quantificados com precisão, por exemplo, por estratégias de MD, contribuindo para o desenvolvimento de sistemas inteligentes de apoio à reabilitação e a outras áreas.

Contribuição dos autores. Os autores contribuíram igualmente em todas as etapas de construção do artigo: concepção e delineamento do estudo; coleta, análise e interpretação dos dados; redação e revisão crítica do manuscrito; e aprovação da versão final.

Agradecimentos. Ao Programa de Pós-graduação em Engenharia Elétrica e Computação (PGEEC)/UNIOESTE.

Financiamento. Os autores não declararam fonte de financiamento.

Disponibilidade de dados e materiais. Os autores não declararam a disponibilidade dos dados e materiais da pesquisa.

Conflitos de interesse. Os autores não declararam conflitos de interesse.

Declaração de uso de IA. Os autores não declararam uso de ferramentas de Inteligência Artificial.

REFERÊNCIAS

1. Brasil. Ministério da Saúde. LER e DORT são as doenças que mais acometem os trabalhadores, aponta estudo [Internet]. Brasília: Ministério da Saúde; 2019 [citado 2025 nov 5]. Disponível em: https://www.gov.br/saude/pt-br/assuntos/noticias/2019/abril/ler-e-dort-sao-as-doencas-que-mais-acometem-os-trabalhadores-aponta-estudo
2. Wang X, Wang Y, Wu J. Position-aware indoor human activity recognition using multisensors embedded in smartphones. Sensors. 2024;24(11):3367. doi:10.3390/s24113367
3. Baroudi L, Barton K, Cain SM, Shorter KA. Classification of human walking context using a single-point accelerometer. Sci Rep. 2024;14:3039. doi:10.1038/s41598-024-53143-8
4. Ensina LA, Lee HD, Maciel M, Spolaôr N, Takaki WSR, Coy CSR, et al. Sistema computacional web para o monitoramento de movimentos em tempo real. J Health Inform. 2021;12:1-14.
5. Borges RC, Parreira WD, da Mata DL, Lage RM. Protótipo acessível para monitoramento cardíaco utilizando fotopletismografia: implementação e perspectivas. J Health Inform. 2024;16(Especial). doi: 10.59681/2175-4411.v16.iEspecial.2024.1324
6. Alanazi M, Aldahr RS, Ilyas M. Human activity recognition through smartphone inertial sensors with ML approach. Eng Technol Appl Sci Res. 2024;14(1):12780-7. doi:10.48084/etasr.6586
7. Lin J, Keogh E, Lonardi S, Chiu B. A symbolic representation of time series, with implications for streaming algorithms. In: Proceedings of the 8th ACM SIGMOD Workshop on Research Issues in Data Mining and Knowledge Discovery (DMKD '03); 2003 Jun 13; San Diego (CA), USA. New York: ACM; 2003. p. 2-11. doi:10.1145/882082.882086
8. Schäfer P, Högqvist M. SFA: a symbolic Fourier approximation and index for similarity search in high dimensional datasets. In: Proceedings of the 15th International Conference on Extending Database Technology (EDBT '12); 2012 Mar 27-30; Berlin, Germany. New York: ACM; 2012. p. 516-27. doi:10.1145/2247596.2247656
9. Shieh J, Keogh E. iSAX: indexing and mining terabyte-sized time series. In: Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '08); 2008 Aug 24-27; Las Vegas (NV), USA. New York: ACM; 2008. p. 623-31. doi:10.1145/1401890.1401966
10. Fulcher BD, Jones NS. Highly comparative feature-based time-series classification. IEEE Trans Knowl Data Eng. 2014;26(12):3026-37. doi:10.1109/TKDE.2014.2316504
11. Lee HD, Wu FC, Spolaôr N, Ensina LA, Takaki WSR, Coy CSR, et al. Sistema de monitoramento de movimentos e análise remota em telemedicina. BR 512020000583-7. Brasil: Instituto Nacional da Propriedade Industrial; 2020. Depósito em 7 jan 2020.
12. Hamill J, Knutzen KM, Derrick TR. Biomecânica básica dos movimentos humanos. 4ª ed. Barueri (SP): Manole; 2016.
13. Merencia AP. Avaliação de representações simbólicas de movimentos articulares humanos com SAX e SFA a partir de dados de acelerômetro [dissertação de mestrado]. Foz do Iguaçu: Universidade Estadual do Oeste do Paraná; 2025.
14. Souza HZ, Lee HD, Ferraz AP, Chung WF. Avaliação de método de segmentação automática de curvas de movimentos do corpo humano. In: Anais do Simpósio Internacional de Iniciação Científica e Tecnológica da USP (SIICUSP); 2025.
15. Hamming RW. Error detecting and error correcting codes. Bell Syst Tech J. 1950;29(2):147-60. doi:10.1002/j.1538-7305.1950.tb00463.x
16. Sheskin DJ. Handbook of parametric and nonparametric statistical procedures. 5th ed. Boca Raton (FL): CRC Press; 2011.
17. Holm S. A simple sequentially rejective multiple test procedure. Scand J Stat. 1979;6(2):65-70.

NOTAS EDITORIAIS

O Journal of Health Informatics é uma publicação científica mantida pela Sociedade Brasileira de Informática em Saúde, associação científica sem fins lucrativos dedicada ao desenvolvimento da Informática em Saúde e da Saúde Digital no Brasil. O processo editorial é conduzido com independência científica, sem interferência institucional, comercial ou de patrocinadores, e segue as boas práticas internacionais de ética em publicação científica.

Este artigo foi avaliado por pares, em sistema duplo-cego, por especialistas da área. A equipe editorial usou ferramentas de inteligência artificial, com supervisão humana, na triagem inicial, na conferência do texto, das referências e dos números e, quando necessário, no redesenho de figuras, quadros e tabelas, aprovado pelos autores na prova. As ferramentas e as salvaguardas estão na Política de uso de inteligência artificial da revista. As decisões editoriais foram tomadas por pessoas.

SBIS Sociedade Brasileira de Informática em Saúde