Journal of Health Informatics

J Health Inform. 2026;18:e1852
DOI: https://doi.org/10.59681/2175-4411.v18.iEspecial.2026.1852

ARTIGO COMPLETO
CBIS 2026

Uma análise comparativa entre funções kernel no algoritmo ALDKRLS para filtragem em ECG

A comparative analysis of kernel functions in the ALDKRLS algorithm for ECG filtering

Pedro Augusto Pereira de Góes[1], Fábio Pires Itturriet[2], Renata Coelho Borges[2], Wemerson Delcio Parreira[1] ✉

ORCID: 0009-0008-5293-8013 · 0000-0003-3140-2259 · 0000-0003-3259-5646 · 0000-0003-1896-0520

[1] Pontifícia Universidade Católica de Campinas (PUC – Campinas), Campinas (SP), Brasil.

[2] Programa de Pós-Graduação em Engenharia Biomédica (PPGEB-CT), Universidade Tecnológica Federal do Paraná (UTFPR), Curitiba (PR), Brasil.

FICHA DO ARTIGO

TIPO
Artigo completo · CBIS 2026

IDENTIFICADOR
e1852

DOI
10.59681/2175-4411.v18.iEspecial.2026.1852

RECEBIDO
17/05/2026

APROVADO
20/07/2026

PROCESSO DE AVALIAÇÃO
Double Blind Review

EDITOR RESPONSÁVEL
Juliano de Souza Gaspar

ACESSO
Aberto, CC BY 4.0

AUTOR CORRESPONDENTE
Wemerson Delcio Parreira
wemerson.delcio@puc-campinas.edu.br

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: Este estudo apresenta uma análise comparativa entre as funções kernel Polinomial e Gaussiana aplicadas ao algoritmo Approximate Linear Dependency Kernel Recursive Least Squares (ALDKRLS) para filtragem de sinais de eletrocardiograma (ECG). **Método:** A metodologia consistiu em uma pesquisa experimental in silico utilizando registros da base MIT-BIH contaminados por ruídos reais e sintéticos em cenários críticos de 0 dB e 6 dB. **Resultados:** Os resultados demonstram que o kernel Gaussiano superou o polinomial em preservação morfológica, obtendo melhores métricas em 6 dB e maior correlação em 0 dB, cenário em que o polinomial foi melhor nas outras 3 métricas, enquanto o polinomial apresentou esparsidade extrema, mas com menor fidelidade diagnóstica. **Conclusão:** Conclui-se que, embora o mapeamento Gaussiano seja superior para diagnósticos de alta precisão, o kernel polinomial de baixa ordem permanece como uma solução pragmática para dispositivos com severas limitações de hardware e eficiência energética.

Descritores: Eletrocardiografia; Ruído; Algoritmos.

ABSTRACT

Objective: This study presents a comparative analysis between Polynomial and Gaussian kernel functions applied to the Approximate Linear Dependency Kernel Recursive Least Squares (ALDKRLS) algorithm for electrocardiogram (ECG) signal filtering. **Method:** The methodology consisted of in silico experimental research using records from the MIT-BIH database contaminated by real and synthetic noise in critical scenarios of 0 dB and 6 dB. **Results:** Results demonstrate that the Gaussian kernel outperformed the polynomial in morphological preservation, obtaining better metrics at 6 dB and higher correlation at 0 dB, where the polynomial kernel was better on the other 3 metrics, while the polynomial presented extreme sparsity but lower diagnostic fidelity. **Conclusion:** It is concluded that, although Gaussian mapping is superior for high-precision diagnostics, the low-order polynomial kernel remains a pragmatic solution for devices with severe hardware limitations and energy efficiency.

Keywords: Electrocardiography; Noise; Algorithms.

CBIS 2026 XXI CONGRESSO BRASILEIRO DE INFORMÁTICA EM SAÚDE

Artigo apresentado no XXI Congresso Brasileiro de Informática em Saúde (CBIS 2026), realizado de 23 a 25 de setembro de 2026, em Brasília (DF), e publicado na seção CBIS 2026 do Journal of Health Informatics.

COMO CITAR ESTE ARTIGO: de Góes PAP, Itturriet FP, Borges RC, Parreira WD. Uma análise comparativa entre funções kernel no algoritmo ALDKRLS para filtragem em ECG. J Health Inform 2026;18:e1852. doi: https://doi.org/10.59681/2175-4411.v18.iEspecial.2026.1852

1 INTRODUÇÃO

A eletrocardiografia (ECG) é uma ferramenta essencial na análise da atividade elétrica do coração, sendo amplamente utilizada para o diagnóstico de doenças cardiovasculares [1]. No entanto, a qualidade do sinal ECG pode ser expressivamente comprometida por diversos tipos de ruídos e artefatos, incluindo ruído aditivo Gaussiano branco (do inglês *Additive white Gaussian noise*, AWGN), interferência de linha de energia (PLI), artefatos de movimento de eletrodos (EM) e artefatos musculares (EMG) [2]. A remoção eficaz desses componentes indesejados é fundamental para garantir a precisão na interpretação clínica e na detecção de padrões cardíacos anormais [3]. Dentre as técnicas mais recentes para *denoising* de sinais de ECG, os filtros adaptativos baseados em *kernel* têm se mostrado promissores devido à sua capacidade de modelar relações não lineares presentes nos dados biomédicos [2].

Filtros digitais convencionais e técnicas baseadas em *wavelets* apresentam limitações inerentes ao lidar com a natureza não estacionária dos biossinais [4]. Em contrapartida, os filtros adaptativos baseados em *kernel* oferecem uma abordagem robusta ao mapearem o sinal de entrada para um espaço de características de alta dimensão, conhecido como Espaço de Hilbert de núcleo reprodutor (*Reproducing Kernel Hilbert Space*, RKHS) [5]. Esse mapeamento permite que problemas não lineares complexos sejam tratados em uma estrutura linear simplificada no espaço transformado.

Em particular, o algoritmo *Kernel Recursive Least Squares* (KRLS) com Dependência Linear Aproximada (ALD), denominado ALDKRLS, tem sido empregado para mitigar artefatos e ruídos em ECGs de forma eficiente, preservando características importantes do sinal original [2]. No entanto, a escolha adequada da função *kernel* desempenha um papel fundamental na performance do método, impactando diretamente a capacidade do modelo de capturar padrões complexos nos sinais [6].

Entretanto, o elevado custo computacional historicamente limitou o uso dessas técnicas em tempo real [5]. O algoritmo ALDKRLS soluciona essa questão ao introduzir o critério de esparsidade ALD, que seleciona apenas as amostras geometricamente independentes para compor um dicionário de dados [7]. Essa estratégia reduz drasticamente a demanda de memória e processamento, viabilizando a aplicação do filtro em sistemas embarcados e dispositivos de monitoramento contínuo [5,7].

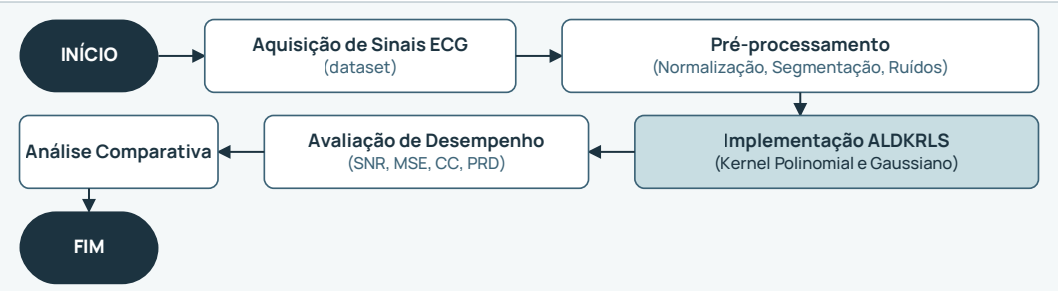
Diante disso, o presente trabalho propõe avaliar o comportamento dos *kernels* Polinomial e Gaussiano na estrutura do algoritmo ALDKRLS. A escolha do *kernel* representa uma decisão central em sistemas de aprendizado de máquina, pois determina como o modelo generaliza os padrões dos dados, influenciando tanto sua expressividade quanto sua viabilidade computacional [5]. No contexto do Sistema Único de Saúde (SUS), essa escolha ganha caráter estratégico: configurações inadequadas podem comprometer a confiabilidade dos resultados e a segurança do paciente, enquanto as restrições orçamentárias e a heterogeneidade de infraestrutura, sobretudo na Atenção Primária à Saúde (APS) e em regiões de cobertura limitada, tornam a definição do *kernel* determinante para a viabilidade técnica e financeira do monitoramento cardíaco em larga escala.

Este artigo está estruturado como se segue. A Seção 2 detalha a metodologia, englobando o pré-processamento dos sinais do *dataset*, a modelagem dos *kernels* Polinomial e Gaussiano, a arquitetura do ALDKRLS e as métricas de desempenho. A Seção 3 apresenta os resultados experimentais obtidos sob cenários com diferentes Relações Sinal-Ruído (SNR). A Seção 4 discute a análise comparativa focada na preservação morfológica do sinal e no custo computacional atrelado à esparsidade dos dicionários. Por fim, as conclusões do estudo e as diretrizes para aplicações em sistemas de monitoramento contínuo são apresentadas na Seção 5.

2 MÉTODOS

Nesta seção, detalham-se os procedimentos técnicos, a fundamentação das ferramentas utilizadas e a evolução cronológica dos experimentos. O fluxo metodológico seguiu o *pipeline* estruturado na Figura 1.

Figura 1. *Pipeline* estruturado da metodologia adotada neste projeto



Fonte: Elaborado pelos autores (2026).

2.1 Caracterização da Pesquisa e Ambiente de Desenvolvimento

O presente estudo caracteriza-se como uma pesquisa experimental *in silico*, com simulações quantitativas baseadas no processamento digital de biossinais. Todos os algoritmos, testes e validações foram desenvolvidos na linguagem de programação Python, utilizando um ambiente virtualizado. Para a manipulação matemática dos vetores e matrizes inerentes ao algoritmo adaptativo, empregou-se a biblioteca numpy; para a importação e leitura padronizada dos registros médicos, utilizou-se a biblioteca wfdb (*Waveform Database*); e para a geração das análises visuais, a biblioteca matplotlib [8].

2.2 Base de Dados (dataset) e Aspectos Éticos

Em relação aos aspectos éticos e à amostra da pesquisa, ressalta-se que foram utilizados registros de eletrocardiogramas previamente anonimizados e de domínio público, providos pelo repositório PhysioNet (https://physionet.org/). Especificamente, utilizou-se a base *MIT-BIH Arrhythmia Database* [9] (sinais limpos de referência) combinada com a *MIT-BIH Noise Stress Test Database* (NSTDB), de onde foram extraídos ruídos reais de variação de linha de base (BW), artefatos de movimento de eletrodo (EM) e contrações musculares (MA) [10]. Para a execução das análises estatísticas finais, definiu-se uma amostragem composta por 10 registros distintos (Registros 100 a 109), garantindo variabilidade morfológica. Por se tratar de um banco de dados aberto internacionalmente, com aprovação ética institucional prévia, este trabalho dispensa submissão autônoma a um Comitê de Ética em Pesquisa (CEP).

2.3 Pré-processamento e Cenário de Ruído

A metodologia de pré-processamento considera um janelamento temporal dos sinais, com os dados segmentados em vetores de tamanho de 5.400 amostras, operando sob uma frequência de amostragem fixa de 360 Hz. Para garantir o cálculo das distâncias euclidianas no RKHS, procedeu-se à normalização estatística dos sinais para manter a proporcionalidade da amplitude original das ondas do ECG. Tanto o sinal limpo de referência $s_{\text{limpo}}[n]$ quanto o sinal contaminado $s_{\text{cont}}[n]$ foram submetidos a esta padronização para o cálculo preciso das métricas de erro.

Para impor condições de estresse realistas e robustas ao filtro adaptativo, o cenário de ruído foi composto por uma combinação de artefatos (ruídos). O vetor de interferência total foi gerado pela soma de ruídos fisiológicos extraídos de bases de dados (Artefatos de Movimento de Eletrodo, EM; Flutuação de Linha de Base, BW; e Artefatos Musculares, MA) [10] com a adição de Ruído Branco Gaussiano Aditivo (AWGN) gerado sinteticamente. O ruído AWGN foi gerado sinteticamente a partir de uma amostragem de uma distribuição normal padrão, possuindo média nula (μ = 0) e variância unitária (σ^2 = 1), expressa matematicamente por $n_{\text{AWGN}} \sim N(0, 1)$. A composição do vetor de ruído total não escalonado consistiu na sobreposição linear ponto a ponto dessas quatro fontes de interferência ao longo da janela de amostras, o que pode ser descrito por:

$$ruido_{total}[n] = n_{EM}[n] + n_{BW}[n] + n_{MA}[n] + n_{AWGN}[n]. \quad \text{(a)}$$

A contaminação do sinal original foi realizada de forma algebricamente controlada para atingir valores exatos de Relação Sinal-Ruído (SNR), testando o algoritmo sob cenários variados, desde níveis moderados (6 dB) até condições de ruído severo (0 dB). O dimensionamento do ruído foi calculado analiticamente a partir da potência do sinal original. Primeiramente, determinou-se a potência média do sinal limpo (P_s), calculada pela média dos quadrados de suas amplitudes, e, a partir do SNR desejado em decibéis, definiu-se a potência alvo do ruído (P_{n_alvo}), usando:

$$P_{n_alvo} = P_s / 10^{\left(\frac{SNR_{dB}}{10}\right)}. \quad \text{(b)}$$

Para escalar o vetor de ruído combinado à potência alvo, um fator de escala (σ_{ruido}) foi calculado extraindo a raiz quadrada da razão entre a potência alvo e a potência atual do vetor de ruído ($P_{\text{ruido_base}}$), expresso por:

$$\sigma_{ruido} = \sqrt{\frac{P_{n_alvo}}{P_{ruido_base}}}. \quad \text{(c)}$$

Por fim, o sinal contaminado (s_{cont}) submetido à filtragem foi obtido pela injeção das interferências no sinal original, feita através da soma do sinal limpo com o vetor de ruído total previamente escalonado, resultando em:

$$s_{cont}[n] = s_{limpo}[n] + ruido_{total}[n]. \quad \text{(d)}$$

Essa abordagem matemática garantiu a reprodutibilidade exata das condições de estresse para os testes subsequentes.

2.4 Kernel Polinomial

A modelagem não linear intrínseca aos filtros KRLS é viabilizada por meio de funções de *kernel*, que projetam os vetores de entrada num espaço de características de dimensão superior *RKHS* [5]. Neste estudo, um dos *kernels* adotados é o *Kernel* Polinomial, que mapeia os vetores de entrada considerando tanto suas interações cruzadas quanto informações de ordem superior no espaço de características, de acordo com a Equação (e):

$$k(x, y) = (x^T y + c)^d \quad \text{(e)}$$

em que c atua como uma constante que pondera a influência dos termos de ordem inferior, enquanto o

parâmetro d representa o grau do polinômio. Sob a perspectiva da implementação algorítmica utilizada nas simulações, o processamento ocorre de forma amostral unidimensional, onde x e y representam as amplitudes normalizadas do ECG em instantes de tempo discretos. Uma propriedade intrínseca a este cenário é que, para entradas escalares mapeadas por um *kernel* polinomial de grau d, a dimensão do espaço de características subjacente, e consequentemente o número máximo de bases estritamente independentes, é matematicamente limitado a $d + 1$ [5,11].

2.5 Kernel Gaussiano

O *Kernel* Gaussiano, do grupo *Radial Basis Function* (RBF), é a função de mapeamento mais referenciada na literatura de filtragem adaptativa para o processamento de biossinais não lineares e complexos, tais como o sinal de ECG de acordo com [4,5]. Sua ampla adoção justifica-se pela capacidade de projetar os dados em um *RKHS* de dimensão infinita, o que permite a linearização de dinâmicas temporais que apresentam alta variabilidade e irregularidade morfológica [11]. A função *kernel* Gaussiana entre duas amostras, x e y, é definida matematicamente por:

$$\kappa(x, y) = \exp\left(-\frac{\|x - y\|^2}{2\xi^2}\right) \quad \text{(f)}$$

em que ξ representa a largura de banda (*bandwidth*) do *kernel*. Este hiperparâmetro é crítico para o desempenho do algoritmo ALDKRLS, pois controla a suavidade do mapeamento: valores de σ muito reduzidos podem levar ao *overfitting*, onde o filtro interpreta ruídos de alta frequência como componentes do sinal; já valores excessivamente altos tendem a linearizar o modelo, perdendo a capacidade de capturar as nuances morfológicas complexas do complexo QRS do sinal ECG [5].

2.6 Algoritmo ALDKRLS: Arquitetura e Procedimento de Filtragem

Para viabilizar o processamento de biossinais em tempo real e contornar o crescimento ilimitado do dicionário no algoritmo KRLS tradicional, a metodologia empregada adota o critério ALD. Esta estratégia de esparsificação assegura que apenas as amostras de ECG que tragam inovação geométrica ao modelo sejam armazenadas, garantindo eficiência computacional sem perda de fidelidade morfológica [7].

A dinâmica de filtragem inicia-se com a configuração da matriz de covariância inversa (P), inicializada como uma matriz identidade escalonada pelo fator de esquecimento lambda (λ), e o vetor de pesos (h), inicializado com zeros. Este parâmetro atua como um fator de regularização, sendo responsável por estabilizar numericamente a inversão da matriz de Gram e ponderar a influência das amostras passadas na atualização do modelo. Tal configuração é essencial para que o filtro preserve sua capacidade adaptativa frente à natureza não estacionária dos sinais [6,7].

A cada nova iteração n, correspondente à chegada da amostra $x(n)$, o vetor de *kernel* atualizado $k(n)$ é calculado em relação a todos os L vetores de suporte $\tilde{x}_j$ já alocados no dicionário $\mathcal{D} = \{\tilde{x}_1, \tilde{x}_2, ..., \tilde{x}_L\}$, conforme:

$$k(n) = [\kappa(x(n), \overline{x_1}), \kappa(x(n), \overline{x_2}), \ldots, \kappa(x(n), \overline{x_L})]^T. \quad \text{(g)}$$

A função *kernel* $\kappa(\cdot, \cdot)$ define a métrica de similaridade no RKHS. Na sequência, o algoritmo avalia o nível de dependência linear da nova amostra projetada. Utiliza-se a matriz de Gram inversa do dicionário atual, $\tilde{K}^{-1}(n-1)$, para determinar o vetor de coeficientes de reconstrução $a(n) = \tilde{K}^{-1}(n-1)k(n)$. O erro de reconstrução geométrica, ou distância quadrada no espaço de características (δ), é obtido por:

$$\delta = \kappa(x(n), x(n)) - k^T(n)a(n) \quad \text{(h)}$$

O procedimento de decisão é regido por um limiar de esparsidade fixo v. Se $\delta > v$, a amostra $x(n)$ é classificada como estritamente inovadora e é inserida no dicionário ($\mathcal{D} \leftarrow \mathcal{D} \cup \{x(n)\}$), disparando a expansão da matriz de Gram inversa e dos vetores de peso associados. Caso contrário ($\delta \leq v$), a amostra é considerada redundante; a estrutura do dicionário permanece inalterada e o algoritmo executa apenas a atualização recursiva dos pesos $h(n)$ via equações clássicas do RLS para minimizar o erro de filtragem instantâneo [4,6].

2.7 Cenário Experimental e Métricas de Desempenho

A validação quantitativa da fidelidade diagnóstica e da capacidade de supressão de artefatos foi estruturada por meio de quatro métricas de desempenho:

i. Erro Quadrático Médio (*Mean Squared Error*, MSE): Avalia a precisão global do processo de predição. Representa a média dos quadrados das diferenças entre o sinal de referência limpo padronizado ($s[n]$) e o sinal filtrado ($\hat{s}[n]$). Valores de MSE próximos de zero são desejáveis, pois indicam que a estimativa do filtro converge para o sinal original com erro residual mínimo [1,3]:

$$MSE = \frac{1}{N}\sum_{n=1}^{N}(s[n] - \hat{s}[n])^2 \quad \text{(i)}$$

ii. Coeficiente de Correlação (CC): Mensura a preservação da morfologia fisiológica do eletrocardiograma. Verifica o grau de dependência linear entre as amplitudes do sinal de saída e o de referência, em que $\bar{s}$ e $\bar{\hat{s}}$ denotam as médias aritméticas amostrais do sinal original e filtrado, respectivamente, calculadas no mesmo intervalo. Resultados próximos a um (≈ 1,0) são considerados ideais, sinalizando que a integridade das ondas P, complexo QRS e onda T foi mantida sem distorções morfológicas expressivas [1,3], conforme:

$$CC = \frac{\sum_{n=1}^{N} (s[n] - \bar{s})(\hat{s}[n] - \bar{\hat{s}})}{\sqrt{\sum_{n=1}^{N} (s[n] - \bar{s})^2 \sum_{n=1}^{N} (\hat{s}[n] - \bar{\hat{s}})^2}} \quad \text{(j)}$$

iii. Ganho de SNR (*SNR Gain*): Quantifica o nível efetivo de ruído extraído pelos algoritmos, calculando a diferença em dB entre a SNR do sinal após a filtragem e a SNR de entrada da simulação. Em que σ_s^2 representa a variância do sinal de referência limpo e σ_e^2 denota a variância do sinal de erro residual, definido por $e[n] = s[n] - \hat{s}[n]$. Quanto maior o valor do ganho, mais eficiente é a capacidade do *kernel* de isolar as componentes ruidosas da atividade elétrica cardíaca [2], de acordo com:

$$SNR_{Saída} = 10\log_{10}\left(\frac{\sigma_s^2}{\sigma_e^2}\right) \quad \text{(k)}$$

$$Ganho_{SNR} = SNR_{Saída} - SNR_{Entrada} \quad \text{(l)}$$

iv. Percentual da Raiz da Diferença Quadrática Média (*Percent Root Mean Square Difference*, PRD): Expressa a distorção global do sinal em uma escala percentual. Valores reduzidos de PRD são desejáveis, uma vez que indicam que as modificações impostas pelo filtro não alteraram a informação diagnóstica essencial para a análise clínica [1,2], conforme:

$$PRD = \sqrt{\frac{\sum_{n=1}^{N} (s[n] - \hat{s}[n])^2}{\sum_{n=1}^{N} s[n]^2}} \times 100. \quad \text{(m)}$$

Visando assegurar o rigor estatístico da análise e neutralizar o viés inerente ao período transiente (de adaptação) dos coeficientes do filtro, o cálculo de todos os indicadores foi restrito aos últimos 33% da janela de amostras da simulação. Essa estratégia metodológica garante que a avaliação de performance reflita o comportamento do algoritmo estritamente em regime permanente, ou seja, após a estabilização do dicionário e a convergência dos pesos [6,7].

3 RESULTADOS

A avaliação de desempenho do algoritmo ALDKRLS foi feita considerando diferentes cenários de testes para fins comparativos. O critério de esparsidade foi fixado em ν = 0,01, o fator de esquecimento em λ = 0,001 e a janela de amostragem em 5.400 amostras. Para o *kernel* polinomial, a constante c da função foi fixada em 0,01. O objetivo foi quantificar a eficácia das funções *kernel* Polinomial e Gaussiana na recuperação da morfologia do sinal sob condições críticas de ruído (SNR de 0 dB e 6 dB). Os indicadores apresentados nas Tabelas 1 e 2 constituem a média aritmética dos resultados obtidos nos dez primeiros registros da base MIT-BIH (registros 100 a 109) [9]. Essa abordagem estatística justifica a ocorrência de valores decimais na métrica de pontos do dicionário (m).

3.1 Análise do Kernel Polinomial

Os testes com o *kernel* polinomial (Tabela 1) focaram na variação do grau (d ∈ {2, 3, 5, 7}). Os dados indicam que esta função favorece a esparsidade extrema, mantendo dicionários com até 8 pontos de suporte. Esse comportamento, contudo, reflete uma limitação intrínseca da natureza de mapeamento global deste *kernel* como já supracitado [5,11].

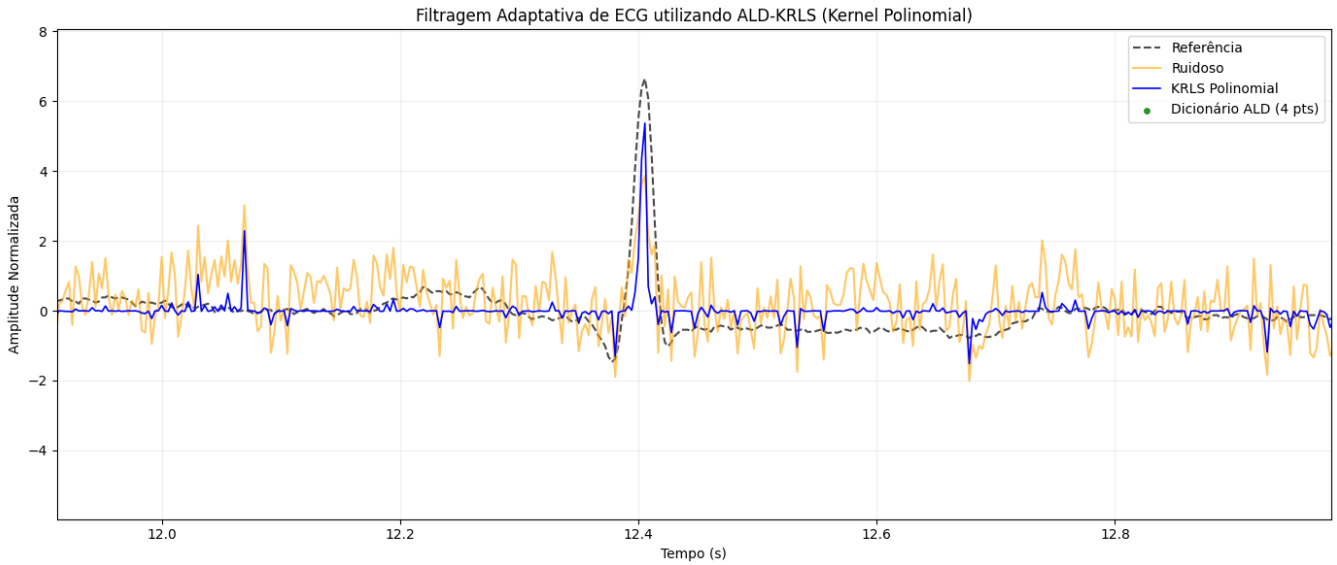
Observa-se na Tabela 1 que o aumento do grau polinomial não resultou em ganho de fidelidade. No cenário de 0 dB, o melhor resultado envolvendo todas as 4 métricas ocorreu para d = 3 (teste 2.P). No cenário de 6 dB, o grau d = 2 (teste 5.P) apresentou os melhores resultados para as 4 métricas, evidenciando que modelos polinomiais de ordem inferior são mais estáveis para a extração do complexo QRS em ambientes ruidosos.

Tabela 1. Desempenho do ALDKRLS com Kernel Polinomial.

Teste	SNR (dB)	d	m	MSE	CC	*SNR Gain* (dB)	PRD (%)
1.P	0	2	3	0,727	0,564	+1,391	85,21
2.P	0	3	4	0,659	0,585	+1,798	81,72
3.P	0	5	6	0,842	0,379	+0,736	92,33
4.P	0	7	8	0,831	0,416	+0,809	91,66
5.P	6	2	3	0,428	0,746	-1,718	63,23
6.P	6	3	4	0,499	0,720	-2,964	71,03
7.P	6	5	6	0,591	0,650	-3,447	75,58
8.P	6	7	8	0,633	0,602	-3,730	79,10

Fonte: Elaborado pelos autores (2026).

Figura 2. Gráfico comparativo sinal ECG filtrado vs. ruidoso (Teste 2.P)



Registro 100 do MIT-BIH.

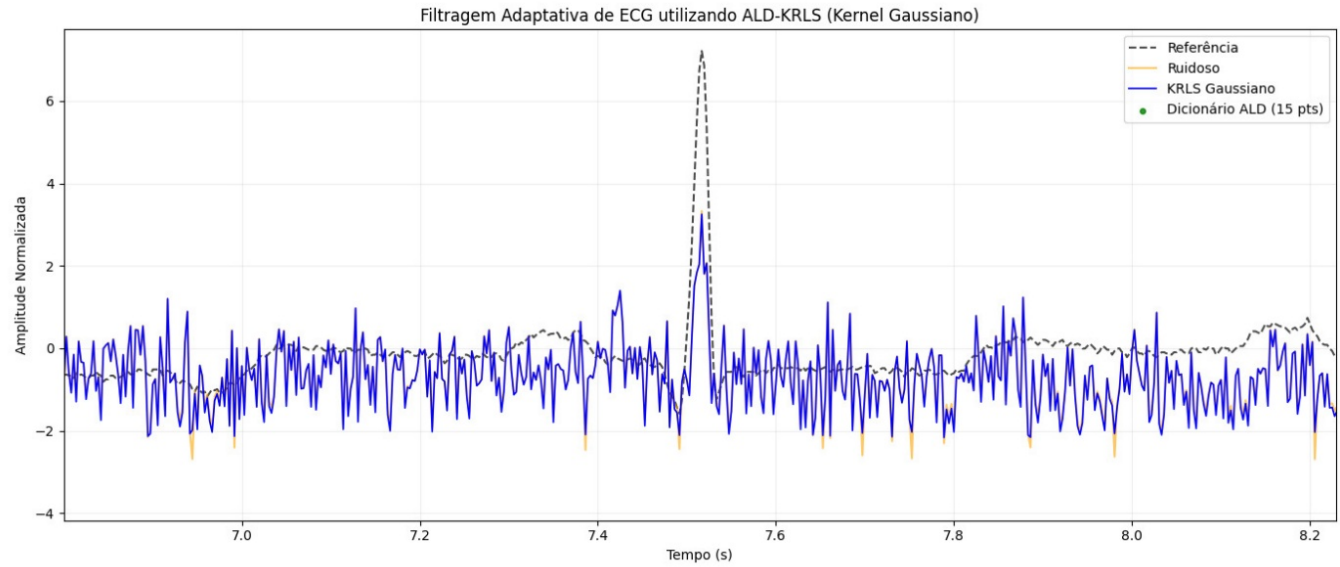
Fonte: Elaborado pelos autores (2026), a partir de [9].

Tabela 2. Desempenho do ALDKRLS com Kernel Gaussiano (RBF).

Teste	SNR (dB)	ξ	m	MSE	CC	*SNR Gain* (dB)	PRD (%)
1.G	0	2,0	8	0,764	0,588	+1,098	87,70
2.G	0	1,0	14	0,752	0,600	+1,134	87,17
3.G	0	0,5	23	0,734	0,600	+1,301	85,83
4.G	0	0,25	42	0,740	0,600	+1,245	86,24
5.G	6	2,0	8	0,335	0,821	-1,233	57,73
6.G	6	1,0	14	0,327	0,828	-1,109	56,98
7.G	6	0,5	25	0,323	0,828	-1,079	56,72
8.G	6	0,25	44	0,394	0,800	-1,717	61,68

Fonte: Elaborado pelos autores (2026).

Figura 3. Gráfico comparativo sinal ECG filtrado vs. ruidoso (Teste 6.G)



Registro 100 do MIT-BIH.

Fonte: Elaborado pelos autores (2026), a partir de [9].

3.2 Análise do Kernel Gaussiano

Os ensaios com o *kernel* Gaussiano (Tabela 2) exploraram a sensibilidade do algoritmo à redução da largura de banda ($\xi \in \{2{,}0; 1{,}0; 0{,}5; 0{,}25\}$).

Diferente da variante polinomial, o *kernel* Gaussiano apresenta maior capacidade de preservação morfológica, atingindo métricas melhores no cenário de 6 dB e maior correlação no de 0 dB, cenário em que o polinomial foi melhor em MSE, ganho de SNR e PRD, como mostrado nas Tabelas 1 e 2. Entretanto, o custo computacional, representado pela função e pelo tamanho do dicionário, apresentou crescimento acentuado em comparação ao polinomial, resultando em um incremento de aproximadamente 4x na alocação do dicionário. O ponto de operação ótimo para este *kernel* concentrou-se em $\xi = 0{,}5$ (testes 3.G e 7.G), em que se obtiveram os melhores resultados das 4 métricas para ambos os cenários ruidosos. A Figura 3 representa o comportamento referente ao Teste 6.G.

4 DISCUSSÃO

Os resultados demonstram que a eficácia do algoritmo ALDKRLS no processamento adaptativo de biossinais é estritamente dependente da geometria topológica da função *kernel*. A comparação entre os mapeamentos Polinomial e Gaussiano (RBF) evidencia uma relação entre a fidelidade da reconstrução morfológica e a eficiência algorítmica por esparsificação [7].

4.1 Resiliência Estocástica em Diferentes Relações Sinal-Ruído

A avaliação do *kernel* polinomial revela uma limitação intrínseca frente ao aumento da potência do ruído. No cenário de 6 dB, o modelo de menor complexidade (Teste 5.P) obteve o melhor desempenho deste cenário. Contudo, sob a condição crítica de 0 dB, tornou-se mais explícito que a elevação do grau do polinômio (Testes 3.P e 4.P, $d \geq 5$) resultou em uma degradação severa da fidelidade morfológica, com a correlação decrescendo para patamares inferiores a 0,42. Este fenômeno é explicado pela natureza de mapeamento global da função: graus elevados tendem a ajustar o modelo às componentes de alta frequência do ruído, resultando em *overfitting* e na distorção do complexo QRS [5].

Em contrapartida, o *kernel* Gaussiano exibiu maior estabilidade estrutural. Para SNR em 6 dB, as configurações de largura de banda intermediária (Testes 6.G e 7.G) atingiram a correlação máxima de 0,828. Ao transitar para o SNR de 0 dB (Testes 2.G e 3.G), a métrica CC apresentou um decréscimo controlado para 0,6, mantendo a integridade da convergência independentemente da variação de σ. Tal resiliência decorre da propriedade de localidade do RBF, em que o decaimento exponencial da função atenua a influência de amostras ruidosas temporalmente distantes no espaço de Hilbert, preservando a matriz de Gram contra flutuações estocásticas.

No confronto direto entre as arquiteturas, a superioridade morfológica do Gaussiano é ratificada. Ao comparar os picos de performance em 6 dB (Teste 5.P *vs.* Teste 7.G), o RBF supera o Polinomial tanto na preservação do formato da onda (CC de 0,828 contra 0,746) quanto na redução do erro residual (MSE de 0,323 contra 0,428). Teoricamente, enquanto o *kernel* polinomial restringe o sinal a um subespaço de dimensão finita, o Gaussiano opera em um Espaço de Hilbert de dimensão infinita.

4.2 Preservação Morfológica e Valor Diagnóstico

A integridade morfológica do ECG, mensurada pelos índices de CC e PRD, é o critério determinante para a viabilidade clínica do processamento. O confronto direto entre os melhores desempenhos em 6 dB revela a superioridade do *kernel* Gaussiano sobre o Polinomial; em 0 dB, o Gaussiano tem CC maior (0,600 contra 0,585), e o Polinomial, PRD menor (81,72% contra 85,83%). Enquanto o mapeamento Gaussiano atingiu um CC de 0,828, o polinomial limitou-se a uma correlação de 0,746.

Sob o ponto de vista diagnóstico, o PRD apresentado pelo *kernel* Gaussiano (56,72%) demonstra uma preservação superior dos intervalos temporais e amplitudes em relação ao Polinomial (63,23%). A natureza de mapeamento global da função polinomial impõe uma rigidez estrutural que compromete a detecção de variações súbitas de potencial elétrico, essenciais para a identificação de arritmias complexas. Em contrapartida, a capacidade de aproximação universal do *kernel* Gaussiano assegura que a filtragem não suprima componentes fisiológicos vitais, mesmo sob regime de ruído severo [5,11]. Dessa forma, a seleção metodológica deve ser pautada pela criticidade da necessidade de implementação.

4.3 Custo Computacional e Viabilidade de Implementação

A disparidade no esforço computacional entre as funções deriva diretamente de suas naturezas matemáticas fundamentais. O *kernel* polinomial, ao projetar o sinal em um subespaço de dimensão finita, induz uma esparsidade extrema no dicionário ALD. Essa estrutura reduz a ocupação de memória e o volume de operações matemáticas mais complexas, consolidando-se como a arquitetura ideal para sistemas onde a economia de recursos, processamento e a autonomia da bateria são os requisitos críticos [7].

Em oposição, o *kernel* Gaussiano impõe um ônus de processamento e memória expressivamente superior. Por operar em um Espaço de Hilbert de dimensão infinita, a sensibilidade do critério de dependência linear é amplificada, resultando em dicionários com

cargas massivas de cálculos para cada iteração adaptativa [5,7].

Dessa forma, a diretriz de escolha reside no equilíbrio entre fidelidade morfológica e restrição física. O *kernel* Gaussiano é a recomendação para instrumentação clínica de alta precisão diagnóstica. Já o polinomial de ordem reduzida permanece como a solução pragmática para o monitoramento onde a eficiência energética sobrepõe-se ao ganho marginal de desempenho, viabilizando equipamentos portáteis de ECG para a APS. No contexto do SUS, essa característica é particularmente relevante: dispositivos de baixo custo computacional podem ser implantados em Unidades Básicas de Saúde (UBS) e em programas de telemonitoramento cardiovascular sem demandar infraestrutura de processamento avançada, ampliando o acesso ao rastreamento de arritmias em regiões com recursos tecnológicos limitados. Por outro lado, a adoção do *kernel* polinomial em cenários de triagem inicial exige atenção redobrada: como demonstrado na Seção 4.1, configurações inadequadas do grau polinomial podem degradar severamente a fidelidade morfológica sob ruído intenso (CC < 0,42), o que reforça a necessidade de protocolos de validação técnica antes da implantação em massa, de modo a evitar que a economia de recursos comprometa a segurança diagnóstica dos pacientes atendidos pela rede pública. Cabe à decisão do projetista e ao propósito da aplicação avaliar se o incremento na qualidade do sinal justifica o sacrifício na autonomia operacional e na estabilidade térmica do sistema embarcado.

5 CONCLUSÕES

Este estudo apresentou uma análise comparativa entre as funções *kernel* Polinomial e Gaussiana aplicadas ao algoritmo ALDKRLS para a filtragem de sinais de ECG. Os resultados obtidos confirmam que a seleção da função de mapeamento é o fator determinante para o equilíbrio entre a fidelidade morfológica e a eficiência computacional em sistemas de monitoramento cardíaco.

Respondendo ao objetivo central da pesquisa, verificou-se uma dicotomia técnica clara entre os modelos avaliados. O *kernel* Gaussiano consolidou-se como estrutura de maior autoridade clínica, atingindo métricas superiores e preservando com precisão os intervalos temporais do complexo QRS. Em contrapartida, o *kernel* polinomial de baixa ordem demonstrou uma capacidade de esparsificação extrema, exigindo dicionários e custo computacional menores que a variante RBF. Essa característica ratifica o polinomial como a solução mais pragmática para dispositivos com limitações de *hardware*.

Em suma, a pesquisa evidencia que a parametrização do ALDKRLS não é universal, mas dependente da aplicação alvo. Enquanto o mapeamento local do *kernel* Gaussiano é indispensável para diagnósticos de alta precisão, a aplicação do mapeamento global polinomial é vantajosa para abordagens visando eficiência computacional, possibilitando a descentralização do monitoramento cardíaco por meio de dispositivos de baixo custo em UBS.

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.

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 declaram a utilização da ferramenta de Inteligência Artificial como suporte ao desenvolvimento deste trabalho. A ferramenta foi empregada de forma colaborativa nas seguintes etapas e finalidades: (i) Apoio na elaboração e refinamento textual para garantir clareza, correção gramatical e adequação ao estilo acadêmico; (ii) Auxílio no desenvolvimento, otimização e depuração dos scripts em linguagem Python. Os autores declaram-se integralmente responsáveis pelo conteúdo científico do manuscrito.

REFERÊNCIAS

1. Chatterjee S, Thakur RS, Yadav RN, Gupta L, Raghuvanshi DK. Review of noise removal techniques in ECG signals. IET Signal Processing. 2020;14(9):569-90. doi: https://doi.org/10.1049/iet-spr.2020.0104
2. Eltrass AS. Novel cascade filter design of improved sparse low-rank matrix estimation and kernel adaptive filtering for ECG denoising and artifacts cancellation. Biomedical Signal Processing and Control. 2022;77:103750. doi: https://doi.org/10.1016/j.bspc.2022.103750
3. Chen X, Lin J, Huang C, He L. A novel method based on Adaptive Periodic Segment Matrix and Singular Value Decomposition for removing EMG artifact in ECG signal. Biomedical Signal Processing and Control. 2020;62:102060. doi: https://doi.org/10.1016/j.bspc.2020.102060
4. Haykin S. Adaptive Filter Theory. 5th ed. Upper Saddle River (NJ): Prentice-Hall; 2014.
5. Liu W, Príncipe JC, Haykin S. Kernel Adaptive Filtering. [s.l.]: Wiley; 2010. doi: https://doi.org/10.1002/9780470608593

6. Parreira WD, Costa MH, Bermudez JCM. Stochastic behavior analysis of the Gaussian KLMS algorithm for a correlated input signal. Signal Processing. 2018;152:286-91. doi: https://doi.org/10.1016/j.sigpro.2018.06.013

7. Engel Y, Mannor S, Meir R. The Kernel Recursive Least-Squares Algorithm. IEEE Trans. Signal Process. 2004;52(8): 2275-85. doi: https://doi.org/10.1109/tsp.2004.830985

8. Hunter JD. Matplotlib: A 2D Graphics Environment. Comput. Sci. Eng. 2007;9(3):90-5. doi: https://doi.org/10.1109/mcse.2007.55

9. Moody GB, Mark RG. MIT-BIH arrhythmia database. Version 1.0.0. ed. [s.l.]: PhysioNet; 2005. doi: https://doi.org/10.13026/C2F305

10. Moody GB, Muldrow W, Mark RG. A noise stress test for arrhythmia detectors. Computers in Cardiology. 1984;11(3): 381-384.

11. Schölkopf B, Smola AJ. Learning with Kernels. [s.l.]: The MIT Press; 2001. doi: https://doi.org/10.7551/mitpress/4175.001.0001

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.

