# Pré-processamento de sinais de eletromiografia para extração de características e treinamento de modelos preditivos de fadiga muscular\*

1<sup>st</sup> Ian Antonio Fonseca Araújo Dep. de Eng. de Computação e Automação UFRN Natal, Brasil ian.fonseca.115@ufrn.edu.br

4th Vívian dos Santos Pereira Dep. de Eng. Biomédica UFRN Natal, Brasil vivian.pereira.016@ufrn.edu.br

Dep. de Eng. Biomédica UFRN Natal, Brasil yasminsantana647@gmail.com

5<sup>th</sup> Amanda Geórgia Diniz de Campos Inst. Int. de Neurociências E. e L. Safra ISD Natal. Brasil amanda.campos@edu.isd.org.br

2<sup>nd</sup> Yasmin Maria Santana de Melo 3<sup>rd</sup> Jorge Richard dos Santos Rocha Dep. de Eng. Biomédica **UFRN** Natal, Brasil jorge.rocha,116@ufrn.edu.br

> 6<sup>th</sup> Ernano Arrais Junior Dep. de Eng. Biomédica **UFRN** Natal, Brasil ernano.arrais@ufrn.br

Resumo-O estudo investiga o pré-processamento de sinais de EMG para detectar fadiga muscular (FM) e a eficácia de modelos de aprendizado de máquina na sua classificação. Foram extraídas características dos sinais, incluindo energia, frequência média (MNF) e frequência mediana (MDF), além de coeficientes da Transformada Wavelet Discreta (TWD) para captura de variações temporais e espectrais. Modelos K-Nearest Neighbors (KNN) e Random Forest (RF) foram treinados, variando hiperparâmetros para otimizar a classificação entre sinais com e sem FM.

Index Terms-Fadiga, EMG, TWD, KNNs, RF

### I. INTRODUÇÃO

A fadiga muscular (FM) pode ser definida como a redução temporária da capacidade do músculo de realizar ações físicas [1] ou, para alguns autores, como a incapacidade de sustentar a força máxima [2]. Ela se manifesta em algumas doenças neurodegenerativas de forma mais acentuada, como o Parkinson, no qual há um aumento da FM durante contrações musculares em comparação com pessoas saudáveis [3], e na esclerose lateral amiotrófica (ELA), na qual, além da fraqueza muscular, os pacientes também sofrem de FM [2].

Para quantificar a FM, é necessário utilizar ferramentas que permitam analisar a atividade muscular de forma detalhada. A eletromiografia (EMG) é uma técnica amplamente empregada para esse fim, pois possibilita a extração de informações relevantes sobre a função muscular [4]. No entanto, o processamento de sinais EMG é essencial para extrair informações úteis ocultas, eliminando interferências e dados irrelevantes [4]. A partir desse processamento, é possível obter características temporais e de frequência que permitem a mensuração da FM [5].

A escolha adequada das características extraídas é essencial para garantir um bom desempenho na análise e classificação da FM nos sinais, pois a inclusão de atributos não correlacionados com o estado de FM pode comprometer a eficiência dos modelos preditivos. Geralmente, as características extraídas dos sinais de EMG são classificadas em três principais grupos: domínio do tempo, domínio da frequência e representação tempo-frequência [6].

No domínio do tempo, o valor eficaz, conhecido como Root Mean Square (RMS), destaca-se por medir a amplitude do sinal de EMG, refletindo a atividade muscular [6]. No domínio da frequência, métricas como a Frequência Média (MNF) e a Frequência Mediana (MDF) são amplamente utilizadas na análise desses sinais [6]. Já na representação tempo-frequência, a Transformada Wavelet Discreta (TWD) permite decompor o sinal em diferentes escalas, capturando simultaneamente padrões transitórios e variações temporais e espectrais. Essa abordagem tem sido aplicada com sucesso na análise de diversos sinais biomédicos, como eletrocardiograma (ECG) [7], e se destaca na classificação da FM em sinais de EMG [8].

Este trabalho tem como objetivo analisar métodos de préprocessamento de sinais de EMG de superfície, a fim de identificar características que diferenciem sinais musculares normais daqueles indicativos de FM. Além disso, serão explorados possíveis modelos preditivos de inteligência artificial para a avaliação da FM, buscando aprimorar a precisão na sua detecção e classificação.

## II. METODOLOGIA

# A. Dataset

No presente estudo, foram utilizados exclusivamente os dados referentes ao movimento de flexão de cotovelo, extraídos do dataset "A Dataset of sEMG and Self-Perceived Fatigue Levels for Muscle Fatigue Analysis"[9]. Esse movimento foi

1

escolhido, pois, nesta etapa inicial do estudo, o foco é analisar apenas movimentos que envolvem a ativação predominante do bíceps braquial.

A frequência de amostragem utilizada para a aquisição dos sinais foi de 1259 Hz. A colocação dos sensores de EMG foi realizada seguindo as diretrizes do SENIAM (Surface ElectroMyoGraphy for the Non-Invasive Assessment of Muscles), que estabelece padrões para a aquisição de sinais de EMG de superfície [10].

No protocolo experimental, os voluntários repetiram o movimento de flexão de cotovelo diversas vezes até atingirem a fadiga [10]. O sinal EMG foi dividido em três etapas distintas: fase inicial (sem fadiga), na qual o músculo apresenta desempenho máximo na execução da tarefa; fase de transição, caracterizada pelo aquecimento do músculo e pelo aumento da dificuldade em manter a intensidade do exercício; e fase de fadiga, onde ocorrem compensações no movimento e a tarefa deixa de ser executada corretamente, refletindo a redução da capacidade contrátil do músculo [10].

Durante a análise dos dados, observou-se que o voluntário 6 (S06) apresentou inconsistências nos labels que indicavam os níveis de fadiga. Diante disso, optou-se por excluir os dados desse voluntário do conjunto final.

#### B. Extração de características de sinais de EMG

1) Energia do sinal: A energia de um sinal EMG é definida como a soma dos quadrados de cada valor de amplitude do sinal [6]. Sua formulação matemática é dada por:

$$E = \sum_{i=1}^{N} x_i^2 \tag{1}$$

onde E representa a energia do sinal,  $x_i$  é a amplitude do sinal EMG no instante i, e N é o número total de amostras do sinal.

2) Frequência média: A MNF, também conhecida como frequência central, é calculada como a soma dos produtos do espectro de potência pela frequência do sinal, dividida pela soma da intensidade do espectro [6]. Sua formulação matemática é dada por:

$$MNF = \frac{\sum_{i=1}^{N} f_i P_i}{\sum_{i=1}^{N} P_i}$$
(2)

onde MNF representa a frequência média,  $f_i$  é a frequência correspondente ao componente *i* do espectro,  $P_i$  é a densidade espectral de potência nesse componente, e N é o número total de componentes de frequência.

*3) Frequência mediana:* A MDF é definida como a frequência na qual o espectro de potência é dividido em duas regiões de mesma soma de amplitudes [6]. Sua formulação matemática é dada por:

$$\sum_{i=1}^{k} P_i = \sum_{i=k+1}^{N} P_i$$
(3)

onde  $P_i$  representa a densidade espectral de potência na frequência  $f_i$ , N é o número total de componentes de frequência, e k é o índice correspondente à frequência mediana MDF, que satisfaz essa condição de equilíbrio no espectro de potência.

#### C. Transformada Wavelet

A Transformada Wavelet é uma ferramenta amplamente utilizada na análise de sinais biomédicos, como o sinal de EMG [11]. Ela permite a decomposição do sinal em diferentes faixas de frequência, preservando simultaneamente as informações temporais e espectrais [7].

A TWD realiza a decomposição de um sinal discreto em duas partes principais: os coeficientes de aproximação, que representam a versão suavizada do sinal e são obtidos por meio de um filtro passa-baixa, capturando as componentes de baixa frequência e preservando a forma do sinal; e os coeficientes de detalhe, que capturam as informações de alta frequência do sinal, representando variações bruscas, são extraídos através de um filtro passa-alta [7]. Os coeficientes podem ser obtidos pelas seguintes equações:

$$s_j(k) = \sum_{n=-\infty}^{\infty} g(n-2k)s_{j-1}(n)$$
 (4)

$$\omega_j(k) = \sum_{n=-\infty}^{\infty} h(n-2k)s_{j-1}(n)$$
(5)

Os coeficientes de aproximação  $s_j(k)$  no nível de resolução *j* são obtidos pela aplicação do filtro passa-baixa g(n), enquanto os coeficientes de detalhe  $\omega_j(k)$  no nível *j* são extraídos por meio do filtro passa-alta h(n). Os coeficientes  $s_{j-1}(n)$  representam o nível anterior da decomposição. As funções g(n) e h(n) correspondem às respostas ao impulso dos filtros passa-baixa e passa-alta, respectivamente. O fator 2kindica a operação de subamostragem, reduzindo pela metade a quantidade de coeficientes em cada nível de decomposição.

### D. Modelos de Classificação

1) K-nearest neighbors (KNN): é um algoritmo de aprendizado supervisionado para classificação e regressão que classifica novos dados com base nos k vizinhos mais próximos em um espaço de características. Na análise de EMG, esse espaço é formado pelas características extraídas e processadas do sinal.

2) Random Forest (RF): combina múltiplas árvores de decisão para obter resultados mais robustos e precisos em tarefas de classificação e regressão. Um hiperparâmetro importante é o número de estimadores, que define a quantidade de árvores na floresta. Assim como o KNN, o RF opera no espaço de características extraídas dos sinais de EMG.

#### E. Pré-processamento

Inicialmente, os sinais de EMG foram separados em dois grupos: sem FM e com FM. No total, foram utilizados sinais de 11 voluntários com registros de ambos os braços. Dessa forma, a base de dados contou com 22 sinais na condição sem FM e 22 sinais na condição com FM. Vale destacar que os atributos de entrada correspondem às tensões elétricas registradas durante os experimentos, sendo, portanto, atributos numéricos que representam a atividade muscular medida pelos eletrodos.

Além do cálculo das características MNF, MDF e energia para os sinais fadigados e não fadigados, foi realizada a análise da correlação de Pearson entre essas características e o estado de FM. Essa correlação permitiu avaliar a relação entre cada métrica extraída e a presença de FM, auxiliando na identificação das características mais relevantes para a distinção entre os dois grupos.

Com o objetivo de ampliar a quantidade de dados disponíveis e permitir a classificação com base em trechos mais curtos do sinal EMG, reduzindo assim a quantidade de contrações necessárias para a identificação da FM, os sinais foram particionados em segmentos de 10 segundos. Esse particionamento resultou em um total de 287 segmentos correspondentes a sinais com FM e 186 segmentos correspondentes a sinais sem FM. As características e correlações foram extraídas de cada um desses segmentos particionados. Posteriormente, essas características foram utilizadas no treinamento do modelo de classificação binária, com o propósito de distinguir entre sinais com e sem FM.

# F. Treinamento dos modelos

Foi criado um espaço de características com base nas métricas extraídas dos sinais particionados: energia, MNF e MDF. Para otimizar o desempenho, o modelo KNN foi treinado com diferentes valores de k, variando entre 1 e 10. O algoritmo foi executado uma vez para cada valor, identificando aquele com melhor performance. De forma similar, o número de estimadores do modelo RF foi ajustado entre 1 a 100, buscando a configuração mais eficaz.

Os dados de treinamento e teste vieram dos trechos particionados dos sinais de EMG, com um vetor de características contendo energia, MNF e MDF para cada partição. As amostras foram divididas aleatoriamente: 80% para treinamento e 20% para teste.

Adicionalmente, foram treinados modelos utilizando a TWD para extração de características em faixas específicas de frequência. Foram aplicadas wavelets de Daubechies de segunda ordem, com decomposições até o quarto nível. As características extraídas dos coeficientes de aproximação de ordem 4, que melhor se correlacionaram com o estado de FM, compuseram o vetor de atributos, substituindo as métricas originais sem adição dos atributos anteriores.

A acurácia dos modelos foi avaliada diretamente no conjunto de teste, sem utilização de validação cruzada neste estudo.

# III. RESULTADOS E DISCUSSÃO

A análise dos sinais mostrou uma correlação positiva entre a energia do sinal e o estado de FM e uma correlação negativa da MNF e da MDF do sinal com o estado FM como listada na Tabela I. Uma correlação menor, porém de mesmo sentido, se manteve após o sinal ser particionado como mostrado na Tabela II.

Tabela I: Correlações sinais não particionados

	Correlação de Pearson			
	Energia	MNF	MDF	
orrelação	0,3413	-0,3124	-0,3494	

C

Tabela II: Correlações sinais particionados

	Correlação de Pearson			
	Energia	MNF	MDF	
Correlação	0,2688	-0,2325	-0,2842	

A Figura 1 mostra a acurácia do modelo de RF de acordo com a quantidade de estimadores. O melhor desempenho foi alcançado com o número de estimadores de 8 e 12, ambos com a mesma acurácia de 86,31%.

A Figura 2 representa a acurácia do modelo de KNN para k variando de 1 até 10. A melhor versão do modelo foi a com k = 3, que obteve 84,21% de acurácia.

Apesar da similaridade no desempenho, o modelo de RF obteve a melhor performance em comparação com o KNN.

Analisando a Tabela III, observa-se que os coeficientes de aproximação do terceiro nível apresentaram correlações ligeiramente superiores para o parâmetro de Energia em comparação aos do quarto nível. Contudo, optou-se pela utilização dos coeficientes do quarto nível no treinamento dos modelos, uma vez que estes apresentaram correlações mais consistentes e equilibradas entre todos os parâmetros, especialmente para o MDF, proporcionando uma representação mais uniforme das características do sinal.

Apesar dessa escolha metodológica, os modelos não apresentaram um desempenho tão satisfatório quanto aqueles treinados com as características extraídas sem o uso da TWD. A acurácia obtida pelo modelo RF foi de 74,73%, enquanto o modelo KNN atingiu 66,32%, conforme ilustrado nas Figuras 3 e 4.



Figura 1: Acurácia por quantidade de estimadores RF usando características originais.



Figura 2: Acurácia por número de vizinhos KNN usando características originais.



Figura 3: Acurácia por quantidade de estimadores RF usando coeficientes de aproximação de nível 4.



Figura 4: Acurácia por número de vizinhos KNN usando coeficientes de aproximação de nível 4.

Tabela III: Correlações dos coeficientes de aproximação da wavelet por parâmetro

	Correlação de Pearson				
Parâmetro	Coef. 1	Coef. 2	Coef. 3	Coef. 4	
Energias	0,2708	0,2919	0,3348	0,3264	
MNF	-0,2915	-0,3080	-0,3078	-0,3018	
MDF	-0,2909	-0,2937	-0,2572	-0,2995	

## IV. CONCLUSÃO

Os processamentos realizados indicaram que a TWD não apresentou eficácia satisfatória na extração de características relevantes para a análise de FM. Esse resultado sugere que a escolha da ordem da wavelet e/ou a faixa de frequência determinada pela decomposição wavelet podem não ter sido adequadas para decompor os segmentos do sinal que contêm informações significativas relacionadas à FM. Essa limitação pode impactar diretamente a capacidade de discriminar padrões ou a precisão nas classificações subsequentes, apontando para a necessidade de parâmetros mais refinados ou a adoção de técnicas complementares para otimizar a decomposição.

Como perspectivas futuras, planeja-se explorar outras transformadas wavelets (packet e redundante), diferentes famílias além da Daubechies, e modelos de predição mais avançados para aprimorar a análise de FM.

#### REFERÊNCIAS

- ENOKA, Roger M.; DUCHATEAU, Jacques. Muscle fatigue: what, why and how it influences muscle function. The Journal of physiology, v. 586, n. 1, p. 11-23, 2008.
- [2] SANJAK, M. et al. Quantitative assessment of motor fatigue in amyotrophic lateral sclerosis. Journal of the neurological sciences, v. 191, n. 1-2, p. 55-59, 2001.
- [3] FRIEDMAN, Joseph H. et al. Fatigue in Parkinson's disease: a review. Movement disorders: official journal of the Movement Disorder Society, v. 22, n. 3, p. 297-308, 2007.
- [4] OLMO, Manuel del; DOMINGO, Rosario. EMG characterization and processing in production engineering. Materials, v. 13, n. 24, p. 5815, 2020.
- [5] NAGATA, S. et al. EMG power spectrum as a measure of muscular fatigue at different levels of contraction. Medical and Biological Engineering and Computing, v. 28, p. 374-378, 1990.
- [6] PHINYOMARK, Angkoon; PHUKPATTARANONT, Pornchai; LIM-SAKUL, Chusak. Feature reduction and selection for EMG signal classification. Expert systems with applications, v. 39, n. 8, p. 7420-7431, 2012.
- [7] ARRAIS JÚNIOR, Ernano. Sistema de análise de sinal cardíaco para aplicação em telecardiologia. 2016.
- [8] PHINYOMARK, Angkoon; LIMSAKUL, Chusak; PHUKPATTARA-NONT, Pornchai. Application of wavelet analysis in EMG feature extraction for pattern classification. Measurement Science Review, v. 11, n. 2, p. 45, 2011.
- [9] CERQUEIRA, S.; VILAS-BOAS, R.; FIGUEIREDO, J.; SANTOS, C. A Dataset of sEMG and Self-Perceived Fatigue Levels for Muscle Fatigue Analysis (Version v2) [conjunto de dados]. Zenodo, 2024. Disponível em: https://doi.org/10.5281/zenodo.14182446. Acesso em: 20 fev. 2025.
- [10] CERQUEIRA, Sara M. et al. A Comprehensive Dataset of Surface Electromyography and Self-Perceived Fatigue Levels for Muscle Fatigue Analysis. Sensors, v. 24, n. 24, p. 8081, 2024.
  [11] UNSER, Michael; ALDROUBI, Akram. A review of wavelets in bio-
- [11] UNSER, Michael; ALDROUBI, Akram. A review of wavelets in biomedical applications. Proceedings of the IEEE, v. 84, n. 4, p. 626-638, 1996.