

UNIVERSIDADE ESTADUAL PAULISTA
“Júlio de Mesquita Filho”
INSTITUTO DE BIOCÊNCIAS DE BOTUCATU

**Análise de séries temporais biológicas de
eletromiografia de superfície (sEMG) e
acelerometria (ACC) para o diagnóstico de
Displasia Coxofemoral em Cães**

Eduardo Roldão Nonato Perondini

ORIENTADOR: Joel Mesa Hormaza

**BOTUCATU – SP
2025**

UNIVERSIDADE ESTADUAL PAULISTA
“Júlio de Mesquita Filho”
INSTITUTO DE BIOCÊNCIAS DE BOTUCATU

Análise de séries temporais biológicas de eletromiografia de superfície (sEMG) e acelerometria (ACC) para o diagnóstico de Displasia Coxofemoral em Cães

Eduardo Roldão Nonato Perondini

ORIENTADOR: Joel Mesa Hormaza

Trabalho de Conclusão de Curso apresentado ao Instituto de Biociências, Campus de Botucatu, UNESP, para obtenção do título de Bacharel em Física Médica.

BOTUCATU – SP
2025

P453a Perondini, Eduardo Roldão Nonato

Análise de Séries Temporais Biológicas de Eletromiografia de Superfície (sEMG) e Acelerometria (ACC) para o Diagnóstico de Displasia Coxofemoral em Cães / Eduardo Roldão Nonato Perondini. -- Botucatu, 2025

39 p.

Trabalho de conclusão de curso (Bacharelado - Física Médica) - Universidade Estadual Paulista (UNESP), Instituto de Biociências, Botucatu

Orientador: Joel Mesa Hormaza

1. Eletromiografia. 2. Alinhamento de sequências. 3. Autoregressão (Estatística). 4. Displasia Coxofemoral. 5. Acelerômetros. I. Título.

Eduardo Roldão Nonato Perondini

Análise de séries temporais biológicas de eletromiografia de superfície (sEMG) e acelerometria (ACC) para o diagnóstico de Displasia Coxofemoral em Cães

Trabalho de Conclusão de Curso, apresentado a Universidade Estadual Paulista, como parte das exigências para a obtenção do título de Bacharel, do curso de Graduação em Física Médica.

Botucatu, 4 de julho de 2025.

BANCA EXAMINADORA

Documento assinado digitalmente
 **JOEL MESA HORMAZA**
Data: 01/08/2025 14:23:33-0300
Verifique em <https://validar.iti.gov.br>

Prof. Dr. Joel Mesa Hormaza
Orientador

Documento assinado digitalmente
 **HELENICE DE OLIVEIRA FLORENTINO SILVA**
Data: 03/08/2025 01:01:49-0300
Verifique em <https://validar.iti.gov.br>

Prof. Dra. Helenice de Oliveira Florentino Silva
2ª Examinadora

Documento assinado digitalmente
 **DANIELA RENATA CANTANE**
Data: 01/08/2025 16:19:48-0300
Verifique em <https://validar.iti.gov.br>

Prof. Dra. Daniela Renata Cantane
3ª Examinadora

Dedicatória

Dedico esse trabalho à todos meus amigos, familiares e professores que me deram a força, o apoio moral e que depositaram sua devida confiança em mim ao longo desse processo de formação da universidade.

Agradecimentos

Gostaria de, primeiramente, agradecer à todos os professores do meu ensino médio que me motivaram, de alguma forma, a seguir na área acadêmica. Esses professores, mais do que tudo, são exemplos de amor pela ciência que me fizeram trilhar o caminho da pesquisa e da educação.

Agradeço, também, à todos os meus amigos do trabalho, de Ribeirão Preto e algumas amizades de Botucatu que fiz ao longo da graduação. Essas pessoas me inspiraram em diferentes momentos, pois cada um deles lutam todos os dias para realizar seus sonhos e vontades, me fazendo perceber o quão bonito é a vida. Todo esse processo de busca pelos seus objetivos é um período que vale a pena ser vivido, pois os aprendizados que derivam desses momentos são eternos e a troca de experiências entre as amizades são o que enriquecem a nossa experiência de vida.

Outra pessoa que foi muito importante durante a minha graduação foi o Prof. Dr. Joel Mesa Hormaza, meu orientador da iniciação científica e desse TCC. Agradeço imensamente à ele por toda sua orientação ao longo desses últimos anos da graduação. Por causa dele, decidi explorar mais a área acadêmica e mergulhar de cabeça na física. Lembro da primeira orientação que ele me deu, após uma reunião que havia feito com ele, expondo a minha insatisfação com o curso e minha insegurança como futuro Físico. Seus conselhos me marcam até hoje, tanto nas minhas escolhas quanto na minha rotina como pesquisador. Por mais que ele seja bem ocupado com os demais projetos e obrigações com o qual ele se envolve, sinto que há um respeito mútuo entre nós, devido a confiança que ele tem sobre o meu trabalho. O Joel como orientador, por mais que não totalmente presente na minha pesquisa, me fez tomar a postura de um verdadeiro cientista, num momento em que esse trabalho foi construído de forma quase-independente - ou seja, o Joel não pegou na minha mão em momento algum, apenas levantou hipóteses e reflexões para eu pensar ao longo do processo de pesquisa. Agradeço muito por isso, pois me mostrou o que um pesquisador enfrenta durante sua rotina. Muitos passos tomados nesse trabalho foram frutos de: muita leitura, releitura, busca de referências bibliográficas, questionamentos, reflexões, testes, hipóteses, confirmações e, por fim, muitas reuniões entre eu e o Joel, em que muitos debates foram feitos de como poderíamos solucionar os problemas enfrentados na pesquisa.

Gostaria de agradecer, também, toda a minha família. Agradecer a minha mãe, Rosana Roldão Nonato, por sempre confiar em mim e incentivar os meus estudos, onde todas as nossas longas conversas sempre me levaram a querer fazê-la orgulhosa pelo caminho que trilhei. Toda a minha pesquisa, estudos e dedicação são frutos do esforço da mesma em me manter estudando em Botucatu. Além disso, gostaria de agradecer ao meu pai, Eduardo Luciano Perondini, pelo exemplo de homem batalhador e esforçado que ele é, que serviu de exemplo para minha pessoa em momentos de grande dificuldade. Meu pai, desde sempre, conversava comigo e contava suas histórias que serviram de exemplo de vida, influenciando nas minhas escolhas acadêmicas e profissionais ao longo da graduação.

Além disso, gostaria de agradecer aos meninos da república Só K Fofó, por todos esses anos de vivência em conjunto e as aventuras proporcionadas que serão lembradas eternamente. Construir essa república foi uma das melhores ideias tolas que já pude ter, onde passamos por altos e baixos, mas sempre juntos e mantendo a personalidade "boba" da casa. Sinto que muito do que aprendi sobre convivência, num geral, foi por causa desse ambiente universitário, me possibilitando perceber a nuance da vida e o quão diferente somos uns dos outros, apesar de iguais.

Agradeço, também, aos meus eternos amigos Gustavo Flausino, Gabriel Faustino e Victor Arruda por me acompanharem durante os anos letivos. Nosso grupo foi imbatível ao longo da graduação, entregando os melhores trabalhos, produzindo coisas juntos e realizando sessões de estudo em grupo. Sinto que aprendi muito sobre companherismo com essas três pessoas e, facilmente, posso dizer que foram os meus melhores amigos da faculdade. Se algum dia eu precisei, eles estiveram lá por mim e, com toda certeza, eu estarei lá por eles algum dia.

Ademais, gostaria de agradecer os meus melhores amigos da vida, os Físicos Eron Vieira Trindade e João Vitor Fernandes, por me apoiarem em todos os meus projetos pessoais e de pesquisa. Desde o ensino médio estivemos juntos e, mesmo com a distância, parece que as coisas não mudam. Tenho um respeito enorme sobre os dois, visto que eles são duas grandes inspirações para mim na Física como um todo. Gosto muito de acompanhar o trabalho deles: Eron com seus estudos de Relatividade Geral e Matéria Escura e o João com seu recém projeto de mestrado envolvendo QCD. Nossos papos esporádicos sobre a vida, graduação, pesquisa etc, foram frutos de muitas inspirações que tive para seguir em frente. Um dia, quem sabe, trabalharemos juntos em algum problema e, posso expressar com toda certeza, que nesse dia trabalharei entre os futuros gigantes da Física.

Por último, mas não menos importante, agradeço à minha namorada Heloiza Alves Ferreira, por me apoiar em literalmente tudo! Desde que a conheci, soube que ela era diferente. Uma mulher dedicada, linda, estudiosa e, além disso tudo, com mentalidade de sempre querer mais. Todas essas características me fizeram, junto a ela, evoluir como pessoa e como profissional. Minha evolução acadêmica é graças a ela, que desde o início da faculdade, ela se mantém fiel a sua iniciação científica e ao seu laboratório, me servindo como inspiração sobre o que gostaria de me tornar profissionalmente. Sua relação com as pessoas do seu laboratório me ensinou que a pesquisa também é humana e que o ambiente acadêmico, acima de tudo, é coletivo. A troca de experiência entre as pessoas de um laboratório, ou departamento, ou de áreas similares ou diferentes, é crucial pro desenvolvimento científico brasileiro. E, tudo isso que percebi, foi diretamente causado por ela e por sua personalidade como recém pesquisadora. Definitivamente, a Heloiza terá um dos futuros mais brilhantes que já pude imaginar, apenas pelo seu jeito de ser, sua dedicação e seu amor pela profissão acadêmica que ela escolheu.

"And even when you've paid enough, been pulled apart or been held up. Every single memory of the good or bad, faces of luck! Don't lose any sleep tonight, I'm sure everything will end up alright. You may win or lose, but to be yourself is all that you can do"

Chris Cornell

Resumo

Acelerômetros (ACC) combinados com sensores de eletromiografia de superfície (sEMG) podem ser úteis na análise cinemática para a compreensão de problemas motores em cães, como é o caso da displasia coxofemoral. Neste estudo, validamos um método alternativo e não invasivo utilizando sinais de sEMG e ACC, capturados com os cães em uma esteira, com o objetivo de avaliar padrões no movimento e o esforço muscular do músculo bíceps femoral, pensando em futuras tarefas de classificação de sinais. A displasia coxofemoral, em casos mais graves, pode provocar instabilidade na locomoção e aumento da atividade muscular para manter a estabilidade do animal. O conjunto de dados de sinais de sEMG e ACC foi pré-processado, filtrado e analisado por meio do algoritmo Dynamic Time Warping (DTW) e do método Autorregressivo (AR), com o intuito de identificar padrões nos sinais e estabelecer uma referência para diagnósticos futuros. Apresentamos alguns padrões identificados com ambos os métodos aplicados a cada tipo de sinal, bem como a separação das distribuições entre as classes saudável e displásica por meio da análise DTW, indicando a existência de um possível limiar que distingue os dois grupos. O modelo AR revelou certos padrões nos coeficientes ϕ característicos de cada grupo, relacionados às séries temporais de sEMG modeladas, possibilitando a classificação por esse método.

Palavras Chave: Eletromiografia de superfície, Algoritmo DTW, Modelo $AR(p)$, Displasia Coxofemoral, Acelerometria

Abstract

Accelerometers (ACC), when combined with surface electromyography (sEMG) sensors, can be valuable tools for kinematic analysis aimed at understanding motor disorders in dogs, such as hip dysplasia. In this study, we propose and validate a non-invasive alternative method using sEMG and ACC signals collected while dogs walked on a treadmill. The goal was to evaluate movement patterns and muscular effort of the biceps femoris muscle, with a view toward future signal classification tasks. In more severe cases, hip dysplasia can lead to locomotor instability and increased muscle activity to maintain the animal's balance. The dataset composed of sEMG and ACC signals was preprocessed, filtered, and analyzed using the Dynamic Time Warping (DTW) algorithm and the Autoregressive (AR) model, in order to identify recurring patterns and provide a reference framework for future diagnostic tools. We present some of the characteristic signal patterns identified by both methods and highlight the separation between healthy and dysplastic classes based on the DTW analysis, suggesting the presence of a potential threshold for distinguishing the two groups. The AR model also revealed group-specific trends in the ϕ coefficients derived from the sEMG time series, indicating the feasibility of classification based on this approach.

Keywords: Surface Electromyography, DTW Algorithm, Autoregressive Model ($AR(p)$), Hip Dysplasia, Accelerometry

Sumário

1	Introdução	8
1.1	Entendendo as Séries Temporais e os Fenômenos Biológicos	8
1.2	Displasia Coxofemoral em Cães	10
1.3	Método Não-Invasivo e Alternativo para o Diagnóstico da DCF	11
2	Metodologia	14
2.1	Captura e Pré-processamento dos Sinais de sEMG e ACC	14
2.1.1	Detalhando o método de pré-processamento dos sinais de ACC . . .	16
2.2	Algoritmo <i>Dynamic Time Warping</i> (DTW)	17
2.3	Modelo Auto Regressivo ($AR(p)$)	22
2.4	Análise dos Dados e o Teste de Tukey	23
2.5	Escolha do Sinal de ACC de Referência	23
3	Resultados e Discussão	25
3.1	Análise do Sinal de ACC	25
3.2	Teste de Tukey	26
3.3	Cálculo da Incerteza Intrínseca ao Sinal de ACC	28
3.4	Análise do Sinal de sEMG	29
4	Conclusão	31
4.1	Conclusões Preliminares	31
4.2	Procedimentos Futuros	31
5	Referências Bibliográficas	33

Introdução

No ramo veterinário, existe uma necessidade voltada para o diagnóstico da Displasia Coxofemoral quanto à introdução de métodos alternativos e não-invasivos na rotina clínica - onde tudo isso é abordado nas Seções 1.2 e 1.3. A dificuldade do diagnóstico da Displasia Coxofemoral junto com a falta de ferramentas para a extração das características dos sinais capturados pelos sensores de cinemática (descritos nas Seções 1.3, 2.2 e 2.3), leva numa grande problemática que vem sido recentemente estudada para validação do método para compor um diagnóstico mais preciso. Entretanto, visto que existem poucas referências bibliográficas que conduzem o estudo pela análise de séries temporais através de métodos computacionais. Um emaranhado de artigos de inúmeras áreas diferentes foram utilizados como base de construção dessa monografia, visando validar o método proposto e sugerir meios para a análise das séries temporais capturadas por sensores cinemáticos. Nas próximas Seções e Capítulos, todo o fundamento teórico sobre as séries temporais, a displasia, os métodos alternativos e não-invasivos e os algoritmos utilizados serão explicados, com o intuito de direcionar a importância desse trabalho ao leitor. Por último, com base nos resultados, uma reflexão sobre as possibilidades de análises e implementações futuras é feita na Seção 4.2, mostrando que há um longo caminho a ser explorado através desses sinais. Sendo assim, o reconhecimento de padrões nos sinais cinemáticos é uma peça chave para auxiliar no diagnóstico da Displasia Coxofemoral em cães, com grande potencial de resultados e melhorias.

1.1 Entendendo as Séries Temporais e os Fenômenos Biológicos

O uso de ferramentas matemáticas e analíticas na medicina clínica como um todo têm sido amplamente usadas. Seja através da extração de características de um dado sistema biológico pelos dados obtidos ou pelo próprio diagnóstico de uma condição que foi interpretada através de *dados quantitativos*. O processamento e a interpretação de sinais médicos são cruciais, uma vez que fornecem uma óptica diferente das análises clínicas, permitindo encontrar padrões característicos de determinada patologia em uma classe de indivíduos [1]. Além da importância do diagnóstico, o estudo de processamento de sinais nos leva à predição de padrões em um próximo intervalo de tempo t_{n+1} , conforme as características do fenômeno biológico estudado [2]. A predicabilidade de um padrão depende de uma série de fatores, como o conhecimento das contribuições deste fenômeno biológico, a quantidade de dados disponíveis para a análise e quando a predição afeta o evento que queremos prever [2].

A ritmicidade no movimento dos animais é uma área de estudo que utiliza de sensores que extraem informações que envolvem o movimento, para a identificação dos ciclos cinemáticos em conjunto com o esforço muscular exercido pelo animal, buscando avaliar a fadiga produzida pelo músculo durante o exercício [3]. Esse tipo de dados observáveis

são chamados de séries temporais, onde que se tem por definição que uma série temporal é um conjunto de pontos coletados durante um intervalo de tempo (Equação 1.1), representando uma série de observações feitas de um fenômeno de interesse [3, 4]. Cada informação pontual representa um estado em um momento específico, e através da observação desses pontos de informação, pode-se entender vários tipos de padrões a longo prazo, como a tendência dos dados (T_t), a sua sazonalidade (S_t), a ciclicidade de comportamentos temporais, as irregularidades no sinal (I_t) e os ruídos (R_t) [4, 5].

$$Z_t = (t_1, t_2, \dots, t_{n-1}, t_n) \quad (1.1)$$

Em que a função temporal Z_t é dita como a série temporal que carrega um conjunto de pontos $t_1, t_2, \dots, t_{n-1}, t_n$ de um certo fenômeno a ser observado. Além disso, como mencionando anteriormente, a série temporal Z_t é formada por diferentes características que compõem o sinal, ou seja, pela Equação 1.2 vemos que as observações temporais possuem uma tendência, sazonalidade, irregularidades e ruídos que formam esse sinal [4-6].

$$Z_t = T_t + S_t + I_t + R_t \quad (1.2)$$

Essa decomposição de Z_t é essencial para a análise da série temporal, uma vez que essas quatro características constroem a base do sinal [6]. Através da Figura 1.1, pode-se observar uma descrição mais precisa dessas variáveis que formam os sinais.

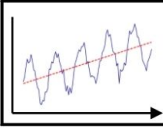
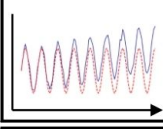
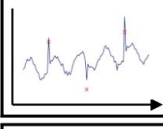
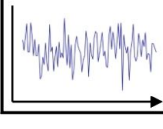
Tendência	Esse atributo representa mudanças de longo prazo, fornecendo uma série de padrões onde os dados aumentam, diminuem ou permanecem constantes ao longo do tempo.	
Sazonalidade	Representa padrões de curto prazo que se repetem em um período fixo, explicando flutuações em momentos específicos.	
Ruído	Refere-se a pontos de dados anormais que se desviam dos padrões gerais, representando flutuações aleatórias que comprometem a consistência dos dados e dificultam a análise, exigindo remoção ou minimização durante a modelagem.	
Irregularidades	Elementos de flutuação aleatória que são difíceis de prever em dados de séries temporais, refletindo a incerteza inerente dos dados.	

Figura 1.1: Principais características que podem ser observadas nas séries temporais, em que essas variáveis constroem o sinal Z_t e fornecem alguns *insights* sobre o comportamento do dado fenômeno estudado. **Fonte:** Adaptado de [4].

No estudo sobre as séries temporais, existe a distinção entre séries com apenas uma variável e processos multivariados [7]. Cada tipo de série temporal possui técnicas diferentes de análises, principalmente dependendo do número de variáveis [8]. As principais ferramentas para o estudo das séries são as funções de autocovariância (Equação 1.3) e autocorrelação (Equação 1.4).

$$K_{XX}(t_1, t_2) = \text{cov}[X_{t_1} X_{t_2}] = E[X_{t_1} X_{t_2}] - \mu_{t_1} \mu_{t_2} \quad (1.3)$$

$$R_{XX}(t_1, t_2) = E[X_{t_1} \bar{X}_{t_2}] \quad (1.4)$$

Onde E é o valor esperado entre os valores, X_{t_1} e X_{t_2} são os dados no tempo, μ_{t_1} e μ_{t_2} são as funções médias ($\mu_t = E[X_t]$). Além dessas funções, existem outros modelos que permitem a captação de *features* que formam ou expressam a série temporal estudada. Alguns desses modelos serão citados na Seção 2.3, onde um dos métodos utilizados neste estudo utilizam de modelos Auto Regressivos ($AR(p)$) para extração de coeficientes que realizam uma regressão linear da série temporal [9]. A garantia desses coeficientes podem nos fornecer padrões no comportamento das séries, principalmente quando essa análise é orientada para o *clustering* dos sinais [9, 10].

1.2 Displasia Coxofemoral em Cães

A Displasia Coxofemoral (DCF) é uma patologia que afeta a condição biomecânica do quadril dos cães [11]. Essa doença é herdável, não congênita e possui uma incidência maior de casos em cães de grande porte [12]. Além disso, a expressão dos genes da DCF pode ser influenciada por inúmeros fatores ambientais [13]. Uma das problemáticas encontradas em uma rotina clínica de diagnóstico da DCF é pela dificuldade no diagnóstico dessa patologia, por causa da vasta maioria dos casos demonstrarem poucos sinais de displasia ou sinal clínico algum [14]. Normalmente, a displasia de quadril nos cães é examinada através de uma radiografia (Figura 1.3), onde a característica principal a ser observada é uma anormalidade no ângulo de Norberg (AN) (Figura 1.2), sendo esse utilizado para mensurar o deslocamento da cabeça do fêmur em relação ao acetábulo, permitindo avaliar o nível de frouxidão da articulação [15].

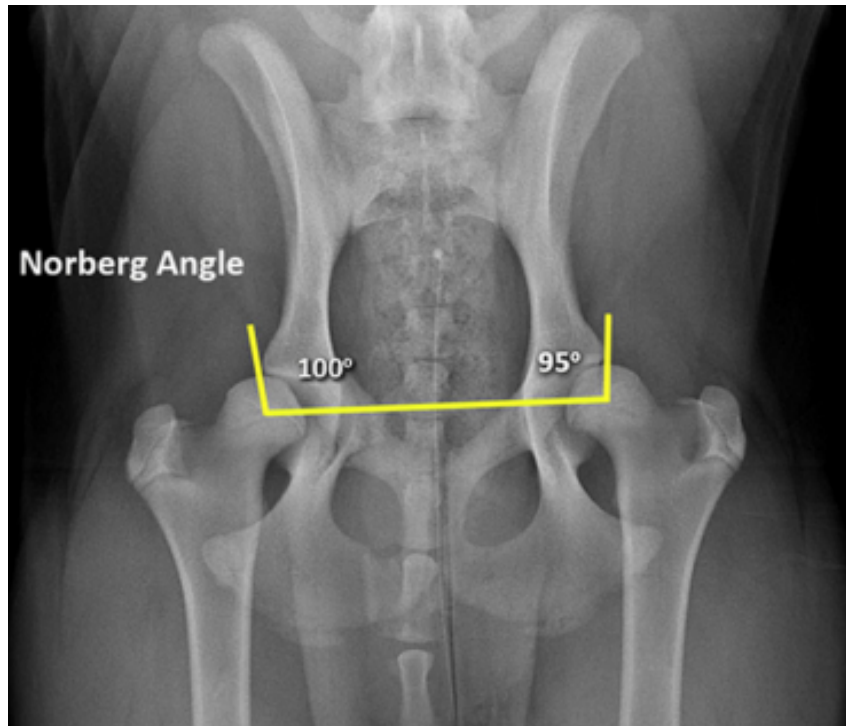


Figura 1.2: Visualização do AN através da radiografia convencional [17].

Esse ângulo pode ser impreciso se os pontos anatômicos não forem claramente visíveis, o que limita a confiabilidade do diagnóstico com base em uma única radiografia [16]. O processo de evolução da DCF pode ocasionar em uma subluxação, onde eventualmente pode resultar em vários graus de achatamento da cabeça femoral e do acetábulo, levando a progressão de uma condição chamada osteoartrose (OA), isto é, o desgaste da cartilagem da articulação (Figura 1.4) [17]. Os parâmetros radiográficos da DCF também possuem níveis, onde essa nuance depende da medida de angulação AN, permitindo um diagnóstico mais preciso sobre a gravidade da patologia [18]. Num geral, existem três organizações internacionais que organizam esquemas para a classificação dos níveis da displasia de quadril canina: a Fédération Cynologique Internationale (FCI), a Orthopedic Foundation for Animals (OFA) e a British Veterinary Association/The Kennel Club (BVA/KC) [18]. Na Tabela 1.1 são apresentados os critérios que o sistema da FCI utiliza para nivelar a DCF, onde podemos observar a presença de cinco níveis (A até E). Entretanto, os sistemas de classificação através da radiografia e da medida do ângulo de Norberg não são totalmente precisos, fazendo com que esses métodos necessitem de pequenos ajustes baseado em outras métricas [19]. Dessa forma, além dos métodos convencionais para a detecção da DCF, o emprego de métodos alternativos se tornam cruciais para o diagnóstico da patologia em questão.



Figura 1.3: Radiografia convencional com a presença da DCF [17].



Figura 1.4: Radiografia convencional de um caso severo da DCF, onde é visível a presença da condição de OA, levando ao desgaste da cartilagem da articulação [17].

1.3 Método Não-Invasivo e Alternativo para o Diagnóstico da DCF

Em uma série de estudos, foi analisado a possibilidade do uso de sensores de cinemática e ativação muscular, bem como a eletromiografia de superfície (sEMG) e a acelerometria (ACC), para o monitoramento do início e fim ciclo de passada canino e a fadiga muscular, pensando no diagnóstico da DCF [20]. Na biomecânica, os sensores de sEMG são amplamente utilizados para identificação de força muscular produzida pela contração dos

músculos, sendo um indicativo de fadiga [21]. Pensando no dano estrutural que cães portadores de DCF podem sofrer, existe um aumento de atividade muscular considerável, devido a instabilidade no movimento que é causado pela patologia [21]. Dessa forma, o uso de sensores de movimento e atividade muscular, se tornam cruciais para a detecção de padrões que entreguem o perfil da doença, por conta da análise cinemática de ambos os sensores [22, 23]. A eletromiografia é a técnica de detecção e captura de atividades musculares [24]. Sensores de eletromiografia conseguem captar potenciais elétricos gerados pelas células musculares durante o processo de contração muscular ou relaxamento [25]. No caso da eletromiografia de superfície (sEMG), a atividade muscular captada é superficial, através de eletrodos, se tornando um procedimento não invasivo ao indivíduo [26]. Já os sensores de acelerometria (ACC) são sensores do tipo inercial, ou seja, capturam sinais de aceleração de um objeto de massa m [27]. O sinal de acelerometria é utilizado para estimar a aceleração de um corpo devido a aceleração da gravidade, dentro dos limites do próprio sensor [28]. O sensor ACC é extremamente sensível e muito suscetível a ruídos, principalmente em casos de variações angulares, quando não aplicado junto à outro sensor de giroscopia [28]. Neste trabalho, sinais de sEMG e ACC foram capturados, pré-processados e analisados através de algoritmos e modelos, como o *Dynamic Time Warping* (DTW) e pelo modelo Auto Regressivo (AR), com o objetivo de identificar padrões nos sinais de cães saudáveis e em cães com displasia coxofemoral.

Tabela 1.1: Sistema de classificação FCI de 5 graus para DCF [18].

Grau	Descrição
Grau A	Sem sinais de displasia coxofemoral A cabeça do fêmur e o acetábulo são congruentes, e o ângulo acetabular segundo Norberg é de aproximadamente 105° (como referência). O espaço articular é estreito e uniforme. O rebordo craniolateral aparece bem definido e ligeiramente arredondado. Não há sinais de osteoartrose.
Grau B	Articulações quase normais A cabeça do fêmur e o acetábulo são ligeiramente incongruentes e o ângulo acetabular segundo Norberg é de aproximadamente 105°, ou a cabeça do fêmur e o acetábulo são congruentes e o ângulo de Norberg é de cerca de 100°. O centro da cabeça do fêmur está medial à borda dorsal do acetábulo. Não há sinais de osteoartrose.
Grau C	Displasia coxofemoral leve A cabeça do fêmur e o acetábulo são incongruentes, o ângulo acetabular segundo Norberg é de cerca de 100° e/ou há um leve achatamento do rebordo craniolateral. Irregularidades ou sinais leves de alterações osteoartríticas no margo acetabular cranial, caudal ou dorsal, ou na cabeça e no colo do fêmur, podem estar presentes.
Grau D	Displasia coxofemoral moderada Incongruência óbvia entre a cabeça do fêmur e o acetábulo com subluxação. O ângulo acetabular segundo Norberg é maior que 90° (apenas como referência). Achatamento do rebordo craniolateral e/ou sinais de osteoartrose.
Grau E	Displasia coxofemoral severa Alterações displásicas marcantes das articulações do quadril, como luxação ou subluxação distinta, ângulo acetabular segundo Norberg menor que 90°, achatamento óbvio do margo acetabular cranial, deformação da cabeça do fêmur (em forma de cogumelo, achatamento) ou outros sinais de osteoartrose.

Metodologia

Antes de tudo, este estudo foi aprovado pelo Comitê de Ética Institucional para o Uso de Animais (CEUA – nº 0121/2017). A captura dos sinais de sEMG e ACC foram realizadas pela Faculdade de Medicina Veterinária e Zootecnia (FMVZ), onde dois grupos de cães foram separados entre um grupo controle (saúdável) e um grupo com displasia coxofemoral (displásico). Diversas medidas captadas pelos sensores foram coletadas e convertidas em pequenos conjuntos de dados para cada cão, possibilitando a organização e a análise de cada indivíduo dos grupos, através dos métodos citados nas Seções 2.2 e 2.3. A seleção dos animais foi baseada nas raças, e todos os indivíduos selecionados para esta análise eram cães da raça Labrador Retriever. Esses grupos de animais foram divididos em cinco cães saudáveis e quatro cães com displasia, sendo que um dos indivíduos do grupo controle foi designado como sinal de referência, o qual será importante posteriormente para o método *Dynamic Time Warping* (DTW). A idade e o peso dos indivíduos não foram utilizados, uma vez que ambos os métodos não consideram esses parâmetros em suas análises.

2.1 Captura e Pré-processamento dos Sinais de sEMG e ACC

O equipamento utilizado foi um sEMG acoplado a um acelerômetro pertencente ao sistema (Trigno™ Wireless EMG Mini Sensor; Delsys, Massachusetts, EUA), e o posicionamento da mini-cabeça do sensor foi baseado em estudos anteriores [22]. Para a obtenção das medições de sEMG e ACC, os cães foram avaliados em uma esteira (GAIT4Dog; CIR Systems Inc., Franklin, NJ, EUA) a uma velocidade de 0,9 m/s, na qual os sinais foram obtidos por meio do software (EMGworks Acquisition, Delsys, Massachusetts, EUA). A taxa de amostragem foi de 2.000 Hz para os sinais de sEMG e 148,148 Hz para os sinais de ACC, conforme os valores recomendados pelo fabricante do sistema [21]. Os bancos de dados de cada grupo de cães passaram por diferentes procedimentos, em que ambos os sinais de sEMG e ACC foram devidamente visualizados, filtrados e tratados, visando a minimização do ruído para evitar quaisquer efeitos indesejados produzidos em futuros modelos de classificação, assim como demonstrado pelo esquema da Figura 2.1 e pela Seção 2.1.1. Para os sinais de acelerometria, os três eixos dos sensores foram levados em consideração, num momento em que a aceleração total (ACC_{Total}) pós-processada evita com que haja ruídos produzidos por variações angulares em quaisquer dos três eixos $\vec{x}$, $\vec{y}$ e $\vec{z}$, pela falta de um giroscópio intrínseco ao sensor de acelerometria. Dessa forma, a Equação 2.1 foi calculada para cada *dataset* dos cães, computando a aceleração total produzida pelo movimento dos bíceps femorais dos indivíduos. Em seguida, os sinais de ACC foram filtrados utilizando um filtro passa-alto de 4 Hz para capturar os pulsos produzidos pela marcha dos cães.

$$ACC_{Total} = \sqrt{(ACC_x)^2 + (ACC_y)^2 + (ACC_z)^2} \quad (2.1)$$

Para garantir a uniformidade dos dados, todos os cães incluídos neste estudo são da raça Labrador Retriever e ambos os grupos, de controle e displásico, possuem a mesma quantidade de indivíduos, exceto pelo cão de referência que foi utilizado para as comparações através do algoritmo DTW. A partir disso, 9 sinais foram analisados, sendo 8 sinais definidos como candidatos e 1 sinal de referência. Os *datasets* coletados das medidas de eletromiografia e acelerometria de cada músculo e cada lado do corpo dos cães continham $n = 4514$ linhas por $m = 16$ colunas (4514×16), em que cada coluna representa uma série temporal diferente, como esquematizado na Tabela 2.1.

Tabela 2.1: Esquematização dos conjuntos de dados de cada cão estudado, demonstrando que cada coluna se refere a uma série temporal e a separação entre os lados dos membros inferiores.

Índices	Bíceps Femoral Direito				...	Bíceps Femoral Esquerdo			
	sEMG	ACCX	ACCY	ACCZ		sEMG	ACCX	...	ACCZ
1	x_{11}	x_{12}	x_{13}	x_{14}	...	$x_{1(m-3)}$	$x_{1(m-2)}$	...	x_{1m}
2	x_{21}	x_{22}	x_{23}	x_{24}	...	$x_{2(m-3)}$	$x_{2(m-2)}$	...	x_{2m}
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
$n - 1$	$x_{(n-1)1}$	$x_{(n-1)2}$	$x_{(n-1)3}$	$x_{(n-1)4}$	...	$x_{(n-1)(m-3)}$	$x_{(n-1)(m-2)}$	...	$x_{(n-1)m}$
n	x_{n1}	x_{n2}	x_{n3}	x_{n4}	...	$x_{n(m-3)}$	$x_{n(m-2)}$	...	x_{nm}

Para a organização dos dados dos sinais de acelerometria, após os cálculos das acelerações totais pela Equação 2.1, todas as séries temporais foram subdivididas em sequências curtas de 100 pontos cada, para o janelamento dos pulsos produzidos pela marcha canina. Através desse procedimento, cada indivíduo produziu 45 sequências para serem alinhadas e analisadas pelo algoritmo DTW. Este comprimento das sequências foi escolhido com base em uma análise empírica dos dados, que revelou que 100 pontos capturavam adequadamente a duração típica de um ciclo de passada durante a caminhada na esteira a $0,9m/s$. Além disso, a análise espectral dos sinais do acelerômetro indicou que os principais componentes de frequência relacionados à dinâmica da marcha foram preservados dentro desse intervalo de tempo. No entanto, para a análise da eletromiografia de superfície, dados brutos de sEMG foram utilizados, mas convertidos em um envelope após procedimentos como ajuste de *offset*, filtragem (filtro passa-banda de 20–450 Hz e passa-baixa de 5 Hz) e pela retificação do sinal, conforme mostrado na Figura 2.1. Na sequência, os dados de sEMG pós-processados foram separados em outro *dataset*, em que essa divisão é importante para a aplicação do modelo Auto Regressivo descrito na Seção 2.3.

Após todos os pré-processamentos dos sinais, os mesmos foram devidamente analisados através do algoritmo DTW e modelo $AR(p)$. Os resultados das análises foram plotados e apresentados nas Seções 3.2 e 3.4, em que discussões acerca da potencial eficiência dos modelos como uma ferramenta de diagnóstico foram realizadas. Em um total, para os dados de acelerometria, foram analisados $45 \times 8 = 360$ custos de alinhamento comparados pelo DTW. Já para os sinais de eletromiografia de superfície, foram extraídos $12 \times 9 = 108$

coeficientes do modelo Auto Regressivo, em que essas *features* possibilitam o processo de clusterização dos sinais de sEMG, para estudos futuros.

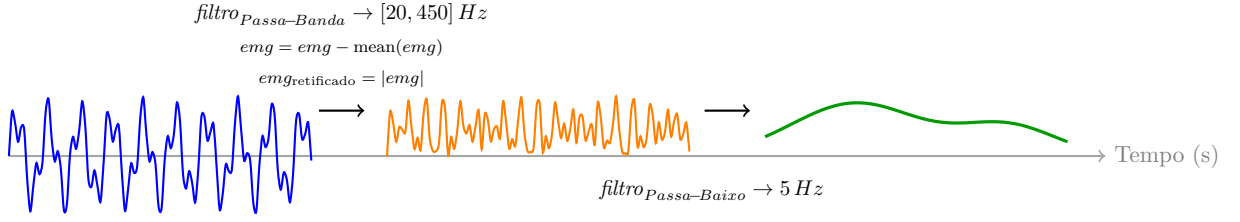


Figura 2.1: Processo de filtragem dos sinais de sEMG, evidenciando as etapas e os tipos de filtros utilizados para a geração dos envelopes de atividades musculares. Em que a amplitude do sinal, em Volts [V], indica a potência produzida pelo exercício muscular ao longo de uma marcha canina, onde a fadiga muscular pode ser avaliada conforme a evolução temporal do sinal capturado. **Fonte:** Elaborado pelo próprio autor.

2.1.1 Detalhando o método de pré-processamento dos sinais de ACC

Analisando a Tabela 2.1, podemos demonstrar matematicamente como o cálculo da Equação 2.1 funciona em um *dataframe* real. Vamos imaginar que cada eixo do acelerômetro, isto é, ACC_x , ACC_y e ACC_z , seja um conjunto de pontos x_{ij} com $i = 1, 2, \dots, n$ e $j = 1, 2, \dots, m$. Ou seja, a partir dessa definição, temos que cada eixo possui o conjunto de pontos:

$$\begin{cases} ACC_x = \{x_{12}, x_{22}, x_{32}, \dots, x_{(n-1)2}, x_{n2}\} \\ ACC_y = \{x_{13}, x_{23}, x_{33}, \dots, x_{(n-1)3}, x_{n3}\} \\ ACC_z = \{x_{14}, x_{24}, x_{34}, \dots, x_{(n-1)4}, x_{n4}\} \end{cases}$$

A partir dessa análise, podemos observar um comportamento padrão na evolução da posição dos elementos desse conjunto de dados. Dessa forma, é possível criar um algoritmo que calcule cada aceleração total, em cada ponto, onde que um processo iterativo é necessário. Outra característica interessante dessa correção, é que há uma redução de dimensionalidade considerável do sistema, isto é, ao invés de aplicarmos o algoritmo DTW em cada coluna, apenas uma acaba sendo necessária. Observando o esquema abaixo, e através da Equação 2.2, pode-se aplicar a correção da aceleração total em um *dataset* qualquer de dados de acelerometria.

$$\begin{aligned} ACC_{Total_1} &= \sqrt{x_{12}^2 + x_{13}^2 + x_{14}^2} = a_{Total_1} \\ ACC_{Total_2} &= \sqrt{x_{22}^2 + x_{23}^2 + x_{24}^2} = a_{Total_2} \\ ACC_{Total_3} &= \sqrt{x_{32}^2 + x_{33}^2 + x_{34}^2} = a_{Total_3} \\ &\vdots \\ ACC_{Total_n} &= \sqrt{x_{n2}^2 + x_{n3}^2 + x_{n4}^2} = a_{Total_n} \end{aligned}$$

Assim, fica claro que o método computa um novo conjunto de valores de aceleração total, ou seja, $ACC_{Total} = [a_{Total_1}, a_{Total_2}, \dots, a_{Total_{n-1}}, a_{Total_n}]^t$. A partir desse padrão observado nos cálculos acima, pode-se construir um algoritmo que siga a mesma linha de contas, generalizando os passos através de um processo iterativo (Equação 2.2).

$$ACC_{Total_i} = \left(\sum_{j=1}^3 (x_{i(j+1)})^2 \right)^{\frac{1}{2}} = \sqrt{\sum_{j=1}^3 (x_{i(j+1)})^2} \quad (2.2)$$

Onde o intervalo de iteração para as colunas é de $j = 1$ até $j = 3$, dependendo de cada caso, pois as colunas dos dados de acelerometria se misturam com os dados de sEMG. Entretanto, as colunas dos três eixos dos sinais de ACC sempre estarão em sequência, possibilitando a escolha do valor mínimo e máximo de j , isto é, a escolha do intervalo $[j_{min}, j_{max}]$. Ou seja, a partir disso, podemos generalizar a Equação 2.2 ainda mais para a Equação 2.3.

$$ACC_{Total_i} = \left(\sum_{j_{min}=a}^{j_{max}=b} (x_{i(j+1)})^2 \right)^{\frac{1}{2}} = \sqrt{\sum_{j_{min}=a}^{j_{max}=b} (x_{i(j+1)})^2} \quad (2.3)$$

Em que $b = a + 2$, isto é, considerando que o sensor possui três eixos e sabendo que as colunas de cada eixo do acelerômetro se organizam em sequência. Portanto, estabelece-se uma regra geral (Algoritmo 1) para o cálculo dos valores de acelerometria totais dentro da base de dados.

Algorithm 1 Cálculo do ACC_{Total} dentro do *dataframe* dos cães. **Fonte:** Elaborado pelo próprio autor.

Require: n, j_{min}, j_{max}

Ensure: ACC_{Total_i}

$j_{min} \leftarrow a$

▷ Inserir os valores de j_{min} e j_{max}

$j_{max} \leftarrow b = a + 2$

$n \leftarrow n_{Linhas}$

while $i \leq n_{Linhas}$ **do**

while $j_{min} \leq j_{max}$ **do**

$ACC_{Total_i} = \left(\sum_{j_{min}=a}^{j_{max}=b} (x_{i(j+1)})^2 \right)^{\frac{1}{2}} = \sqrt{\sum_{j_{min}=a}^{j_{max}=b} (x_{i(j+1)})^2}$

end while

end while

2.2 Algoritmo *Dynamic Time Warping* (DTW)

O algoritmo *Dynamic Time Warping* (ou DTW) é amplamente conhecido e utilizado em diferentes áreas da ciência da computação [29]. Ele foi, primeiramente, introduzido nos anos 60, mas extensivamente explorado nos anos 70 para a aplicação de reconhecimento de falas, visto a capacidade do algoritmo de reconhecer padrões sonoros pelo espectro dos áudios [29]. Mais tarde, o algoritmo foi reconhecido e aplicado para classificação e agrupamento de séries temporais em inúmeras áreas de interesse: como na análise em sinais de eletrocardiograma (ECG), *clustering* de perfis de expressão gênica, biometria, monitoramento de processos etc [30, 31]. Seja o conjunto de pontos $X = \{x_1, x_2, \dots, x_{n-1}, x_n\}$

de tamanho $N \in \mathbb{N}$ e $Y = \{y_1, y_2, \dots, y_{m-1}, y_m\}$ de tamanho $M \in \mathbb{N}$ sinais discretizados distribuídos equidistantemente ao longo do tempo, o objetivo principal do DTW é calcular o alinhamento entre duas sequências de pontos. Esse alinhamento é calculado através de uma medida de custo local c que pode ser mapeada dentro de um espaço de características $\mathcal{F}$. Este espaço é, geralmente, utilizado para caracterizar os dados, ou seja, valores dentro desse espaço podem assumir características α, β, γ etc [41]. Então, para comparar duas características $\alpha, \gamma \in \mathcal{F}$, é necessário uma medida de custo local que é definida como:

$$c : \mathcal{F} \times \mathcal{F} \rightarrow \mathbb{R} \geq 0,$$

Nesse contexto, se X e Y possuem características similares, então o custo $c(X, Y)$ é baixo. Em casos onde há diferença entre as características das sequências, o custo $c(X, Y)$ é alto [41]. Dito isso, se avaliarmos a medida de custo local para cada par de elementos das sequências X e Y , obtem-se a matriz de custo local $c \in \mathbb{R}^{N \times M}$, onde pode ser definida como:

$$C(N, M) := c(x_n, y_m),$$

A ideia, a partir desses conceitos, é encontrar entre esse arranjo de alinhamentos de pontos o custo mínimo geral. Intuitivamente, esse baixo custo percorre um "vale" de valores mapeados em uma matriz de custo C [41].

Definição 3.1. Um caminho distorcido (N, M) é uma sequência de caminhos $p = (p_1, p_2, \dots, p_L)$ com $p_l = (n_l, m_l) \in [1, N] \times [1, M]$ para $l \in [1, L]$ satisfazendo as três condições:

- **Condição de Contorno:** $p_1 = (1, 1)$ e $p_L = (N, M)$.
- **Condição de Monotonicidade:** $n_1 \leq n_2 \leq \dots \leq n_L$ e $m_1 \leq m_2 \leq \dots \leq m_L$.
- **Tamanho do Passo:** $p_{l+1} - p_l \in \{(1, 0), (0, 1), (1, 1)\}$ para $l \in [1, L - 1]$.

Assim, nota-se que a condição do tamanho do passo implica na condição de monotonicidade. Já a condição de contorno força que os primeiros elementos das sequências X e Y , assim como os últimos elementos, sejam de fato alinhados entre si. Isto é, através do tamanho do passo, expressa-se um tipo de continuidade, evitando que quaisquer elementos das sequências X e Y sejam omitidos e que não haja replicações dos mesmos [41]. Dessa forma, para calcular um custo total $c_p(X, Y)$ de todo o arranjo de alinhamentos entre as sequências com respeito à medida de custo local $c(X, Y)$, é feito uma soma desses custos pela Equação 2.4.

$$c_p(X, Y) = \sum_{l=1}^L c(x_{n_l}, y_{m_l}) \quad (2.4)$$

Agora, tendo em vista o objetivo do algoritmo de encontrar o melhor alinhamento entre as sequências através da **minimização** deste custos, isto é, encontrando o **custo de alinhamento ótimo** entre as sequências, pode-se encontrar um caminho p^* que minimiza a Equação 2.4. A partir daí, define-se a **distância DTW** pela Equação 2.5, onde essa

distância encontra o custo mínimo global de alinhamento entre duas sequências $\forall X, Y$ [41].

$$DTW(X, Y) := c_{p^*}(X, Y) = \min\{c_p(X, Y)\} = \min \sum_{l=1}^L c(x_{n_l}, y_{m_l}) \quad (2.5)$$

A distância DTW é uma medida simétrica, desde que a função de custo local c também seja. Entretanto, mesmo que a matriz de custo C seja positiva definida¹, em geral a distância DTW não é - ou seja, pode assumir valor zero mesmo para sequências distintas. Além disso, essa distância não satisfaz a desigualdade triangular, ainda que c seja uma métrica [41].

Para encontrar um caminho ótimo de alinhamento p^* , seria possível testar todos os caminhos mínimos locais possíveis até encontrar o mínimo global, entre as sequências X e Y . Mas, tal procedimento afetaria a complexidade computacional do algoritmo [41], tornando-a uma exponencial na ordem de N e M , como expresso pela Figura 2.2.

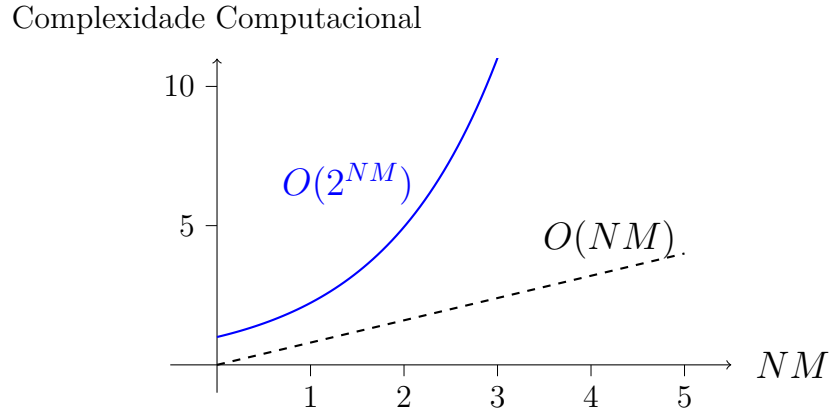


Figura 2.2: Gráfico das curvas da complexidade computacional do algoritmo DTW nos piores casos. A curva azul representa o cenário em que o DTW realiza o alinhamento considerando todas as combinações possíveis entre os pontos de duas sequências. Já a curva preta demonstra o desempenho computacional real do algoritmo DTW, refletindo na sua eficiência na prática. **Fonte:** Elaborado pelo próprio autor.

Seja as sequências finitas $X = \{x_1, \dots, x_n\}$ para $N \in [1 : N]$ e $Y = \{y_1, \dots, y_m\}$ para $M \in [1 : M]$. Para contornarmos o problema da complexidade computacional, pode-se fazer algumas alterações no algoritmo para encontrar a **matriz de custo acumulada** [41], definida pela Equação 2.6. Onde essa formulação, define uma matriz $D(N, M)$ de ordem $N \times M$ que satisfaz o Teorema 3.1.

$$D(N, M) = DTW(X[1 : N], Y[1 : M]) = DTW(X, Y) \quad (2.6)$$

Teorema 3.1. A matriz de custo acumulada D satisfaz as seguintes identidades:

- $D(N, 1) = \sum_{k=1}^N c(x_k, y_1)$

¹Uma matriz positiva definida é uma matriz simétrica no qual todos seus autovalores são positivos. Aprofundando, para uma matriz ser considerada positiva definida exige-se a satisfação de algumas condições: $A = A^T$, $x^T A x > 0, \forall x \neq 0$, ter todos os autovalores de A positivos etc.

- $D(1, M) = \sum_{k=1}^M c(x_1, y_k)$

Assim, temos que para N e M o algoritmo DTW pode ser montado de forma mais eficiente pela Equação 2.7. Fazendo com que a distância DTW seja calculada tendo $O(NM)$ operações, em que matematicamente, aplicando o limite de $N \rightarrow \infty$ e $M \rightarrow \infty$, o custo computacional leva mais tempo para explodir de forma assintótica conforme o aumento dos dados $N \times M$ [41].

$$D(N, M) = \min\{D(N, M-1), D(N-1, M), D(N-1, M-1)\} + c(x_n, y_m) \quad (2.7)$$

Em que o termo $\min\{D(N, M-1), D(N-1, M), D(N-1, M-1)\}$ conferem três opções de alinhamento entre os pontos das sequências, a intersecção $D(N, M-1)$, a correspondência $D(N-1, M)$ e a deleção $D(N-1, M-1)$ [41].

Exemplo: Sejam duas sequências de pontos $X = \{x_1, x_2, \dots, x_{n-1}, x_n\}$ de tamanho N e $Y = \{y_1, y_2, \dots, y_{m-1}, y_m\}$ de tamanho M , em que os valores de cada conjunto assumem valores $\mathbb{R}$ aleatórios. Para duas sequências aleatórias e de mesmo tamanho, a matriz de custo local $C(N, M)$ tem o seguinte perfil:

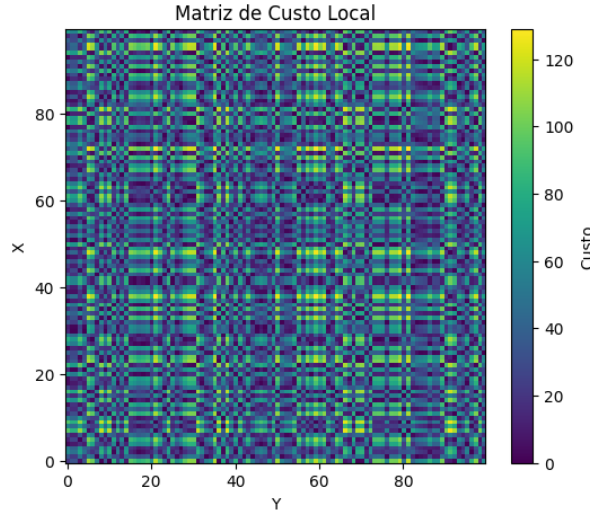


Figura 2.3: Matriz de custo local $C(N, M)$ de duas sequências X e Y aleatórias. **Fonte:** Elaborado pelo próprio autor.

Em um sinal capturado por um sensor no mundo real, as sequências obtidas não são totalmente aleatórias, apesar dos sinais ditos como biológicos serem inerentemente estocásticos. Ou seja, para fins de observação, um exemplo mais condizente com a realidade seria caso houvesse duas sequências $\cos(X) = \{x_1 + \varepsilon_1, \dots, x_n + \varepsilon_n\}$ e $\sin(Y) = \{y_1 + \varepsilon_1, \dots, y_m + \varepsilon_m\}$ onde os dados tivessem um perfil ondulatório com um ruído associado a cada ponto. Na Figura 2.4, duas sequências oscilatórias com ruído foram analisadas através do DTW, mostrando a matriz de custo local com o alinhamento ótimo (mínimo global) sendo destacado. Após isso, a matriz de custo acumulada foi gerada (Figura 2.5) para apresentar e discutir a diferença entre as duas matrizes, quando consideradas como ferramentas de análise.

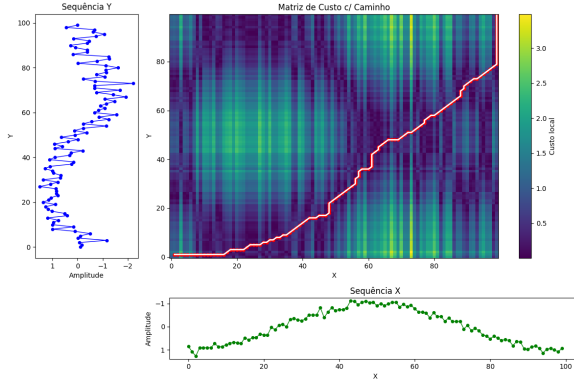


Figura 2.4: Matriz de custo local $C(\cos(X), \sin(Y))$. **Fonte:** Elaborado pelo próprio autor.

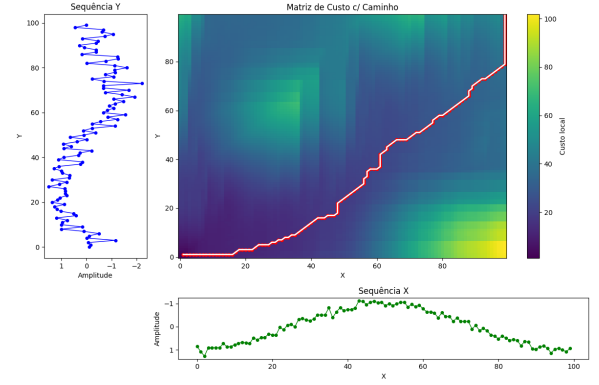


Figura 2.5: Matriz de custo acumulada $DTW(\cos(X), \sin(Y))$. **Fonte:** Elaborado pelo próprio autor.

Na Figura 2.6, as duas sequências tiveram seus conjuntos de pontos alinhados pelo DTW. Dessa forma, apresentando como o algoritmo alinha as sequências ponto a ponto, onde uma de suas vantagens é que o DTW é sensível a padrões transladados entre as sequências, sendo invariante à deslocamentos. O Algoritmo 2 descreve o passo a passo a construção do DTW, o pseudo-código também mostra as variáveis que constroem o método, tendo valores de retorno como o caminho ótimo p^* e a matriz de custo acumulada D .

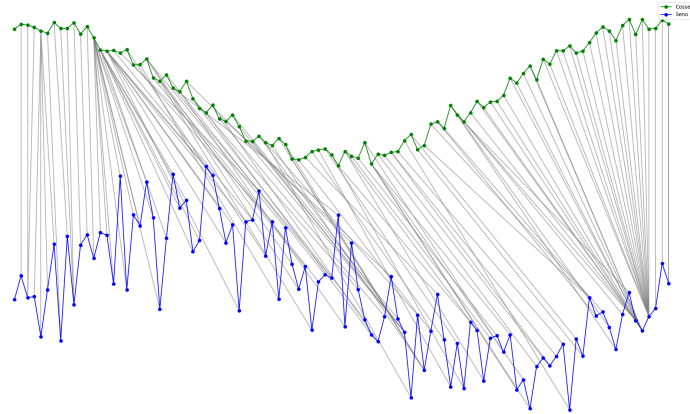


Figura 2.6: Sequências de pontos $\cos(X)$ e $\sin(Y)$ alinhadas pelo DTW, evidenciando os alinhamentos realizados e os padrões capturados pelo algoritmo. **Fonte:** Elaborado pelo próprio autor.

Algorithm 2 DTW: *Dynamic Time Warping* **Fonte:** Inspirado nos trabalhos de [30], [40] e [41].

Require: Sequências $X = (x_1, x_2, \dots, x_N)$ e $Y = (y_1, y_2, \dots, y_M)$

Ensure: p^*, D

```

1: Início  $D[0, 0] \leftarrow 0$ 
2: for  $i \leftarrow 1$  para  $N$  do
3:    $D[i, 0] \leftarrow \infty$ 
4: end for
5: for  $j \leftarrow 1$  para  $M$  do
6:    $D[0, j] \leftarrow \infty$ 
7: end for
8: for  $i \leftarrow 1$  para  $N$  do
9:   for  $j \leftarrow 1$  para  $M$  do
10:     $cost \leftarrow \text{local\_cost}(x_i, y_j)$ 
11:     $D[i, j] \leftarrow cost + \min(D[i-1, j], D[i, j-1], D[i-1, j-1])$ 
12:   end for
13: end for
14:  $i \leftarrow N, j \leftarrow M$ 
15:  $p^* \leftarrow []$ 
16: while  $i > 0$  ou  $j > 0$  do
17:   Juntar  $(i, j)$  para  $p^*$ 
18:    $steps \leftarrow [(D[i-1, j-1], i-1, j-1), (D[i-1, j], i-1, j), (D[i, j-1], i, j-1)]$ 
19:    $(cost, i, j) \leftarrow \text{argmin}(steps)$ 
20: end while
21: Reverter  $p^*$ 
22: return  $p^*, D[1 : N, 1 : M]$ 

```

2.3 Modelo Auto Regressivo ($AR(p)$)

O modelo Autoregressivo (AR) é utilizado em diferentes contextos para a modelagem de séries temporais [32]. Para garantir características que permitam que uma série temporal seja analisada, o modelo Autoregressivo pode ser usado [33]. Sendo esse modelo matemático originado durante os anos de 1920, sendo primeiramente aplicado por Udney Yule, enquanto analisava o comportamento de séries temporais de manchas solares [34]. Agora, pensando nos sinais de eletromiografia de superfície (sEMG), entender a biomecânica do movimento e criar um modelo do sistema muscular é um problema complexo devido à não-linearidades, pela falta de dinâmicas musculares não modeladas, ruídos e interferências [35].

Definição 3.2. O modelo Autoregressivo $AR(p)$ determina um valor de um processo em um tempo arbitrário de passo t utilizando a combinação linear dos p -valores anteriores.

$$y_t = \phi_1 y_{t-1} + \dots + \phi_p y_{t-p} + \varepsilon_t = \sum_{i=1}^p \phi_i y_{t-i} + \varepsilon_t \quad (2.8)$$

Onde o número p é chamado de ordem do modelo AR , os coeficientes ϕ_i da combinação linear são os parâmetros do modelo e, além disso, o modelo AR assume que esse

processo é composto por um ruído, isto é, o termo ε_t é dito como não correlacionado e identicamente distribuído ao longo do tempo, com um valor esperado igual à zero e variância finita. Muitas definições exigem que o processo descrito pelo modelo Autoregressivo seja estacionário. A aplicação desse modelo é feito para descrever séries temporais do tipo:

$$Z_t = \{z_1, z_2, \dots, z_{t-1}, z_t\}$$

A ordem do modelo AR e seus parâmetros são comumente determinados utilizando da perspectiva estatística e estocástica [2, 7]. Sendo esses caminhos, normalmente, baseados na maximização da função de probabilidade e pela obtenção de parâmetros do modelo através da solução das equações de Yule-Walker [36] ou do método dos Mínimos Quadrados [37]. Quanto a aplicação do modelo Autoregressivo nos sinais de eletromiografia de superfície, é interessante a tentativa de extração dos parâmetros dos sinais de sEMG para a validação de processos fisiológicos, como a fadiga muscular [35]. Dessa forma, conhecendo o conjunto de pontos da série temporal $Z_t = \{z_1, \dots, z_t\}$ e o ruído ε_t , pode-se calcular cada coeficiente através da manipulação dos termos.

2.4 Análise dos Dados e o Teste de Tukey

Após a aplicação dos métodos de extração de características dos dois tipos de sinais, ACC e sEMG, um conjunto de análises foram realizadas sobre os dados obtidos. Dessa forma, buscando algum tipo de interpretação de valor biológico que possa indicar uma separação entre os indivíduos dos grupos saudáveis e displásicos. A distribuição dos dados foram analisadas, na tentativa de determinar um limiar de separação entre as classes e também foi calculado a incerteza presente na distribuição, que dificulta a classificação de um indivíduo que se encontra no espectro observado.

Outra abordagem considerada foi o Teste de Tukey, onde esse teste é utilizado para múltiplas comparações entre mais de duas médias de distribuições [38]. Quando temos diversas distribuições para serem comparadas, é incorreto realizar o teste t-Student para cada par de distribuições, por causa do problema de múltiplos testes [39]. Esse problema acontece quando em uma mesma distribuição, realiza-se diversos testes, aumentando a probabilidade cometer um erro de tipo I, ou seja, conduzindo dois testes t-Student na mesma base de dados pode dobrar a probabilidade de erro [39]. Assim, o teste de Tukey *Honestly Significant Difference* (HSD) foi conduzido a fim de comparar as médias entre as distribuições considerando um nível de significância de $p - valor = 0.05$. Dessa forma, o teste foi conduzido entre todos os 9 indivíduos, de ambos os grupos, visando a extração de informações que distingam cada cão entre os grupos de controle e displásicos.

Para os sinais de sEMG, após a aplicação do método $AR(p)$, os coeficientes dos sinais foram obtidos e comparados, visando identificar padrões entre os sinais de mesmo grupo. Outros tipos de visualizações foram empregadas, de modo a facilitar o agrupamento entre os dados quantitativos calculados.

2.5 Escolha do Sinal de ACC de Referência

Antes de qualquer análise, deve-se selecionar o sinal que será levado como referência para o alinhamento das sequências, isto é, que servirá como molde para encontrar as similaridades entre os sinais. Para isso uma série de procedimentos foram realizados para computar uma comparação que determine qual sinal deverá ser utilizado como referência,

como múltiplas comparações envolvendo o arranjo de sinais, a fim de encontrar a melhor distribuição com menor custo possível. Através do gráfico da Figura 2.8, nota-se que dentre todas as possibilidades de alinhamento, apenas uma distribuição possui o menor intervalo de custo, indicando que ela se adequa melhor para servir como molde de comparações. Isso faz sentido, pois num momento em que a sequência definida realiza um ótimo alinhamento entre todas as demais sequências, quer dizer que essa sequência representa todos os padrões possíveis inclusos no conjunto de sinais de ACC. Dentre os 5 sinais de cães saudáveis, o sinal de referência foi escolhido através de múltiplos alinhamentos, tomando cada sinal de cada cão como referência e alternando os candidatos até encontrar uma distribuição de custos com menor intervalo. O esquema na Figura 2.7 ilustra bem os procedimentos para realizar as comparações.

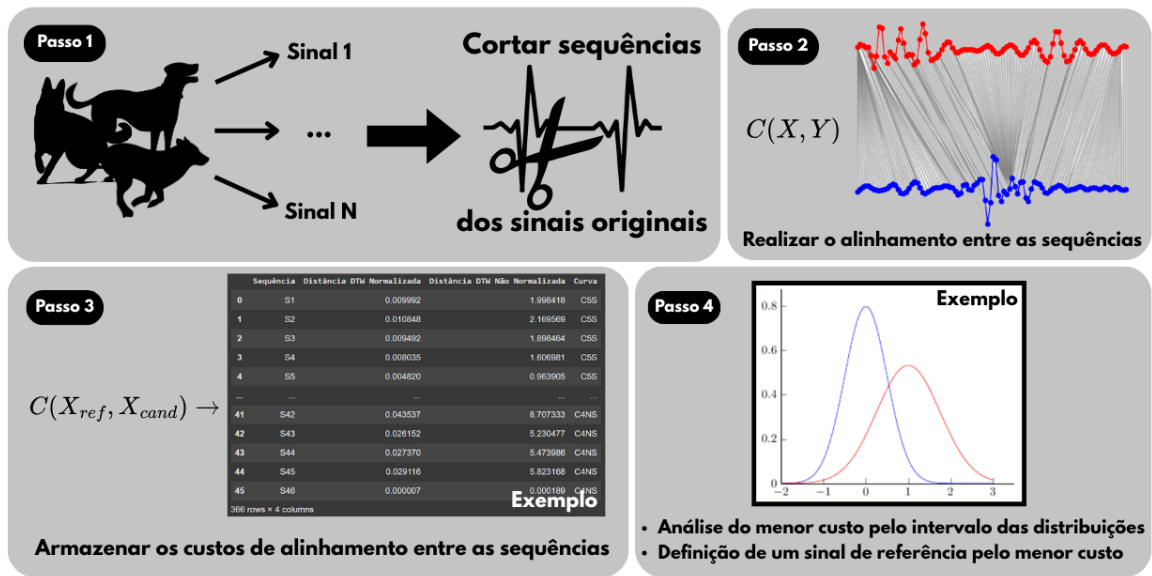


Figura 2.7: Ilustração do processos de comparação entre os sinais candidatos e o sinal de referência utilizado para selecionar o indivíduo com menor custo de alinhamento. **Fonte:** Elaborado pelo próprio autor.

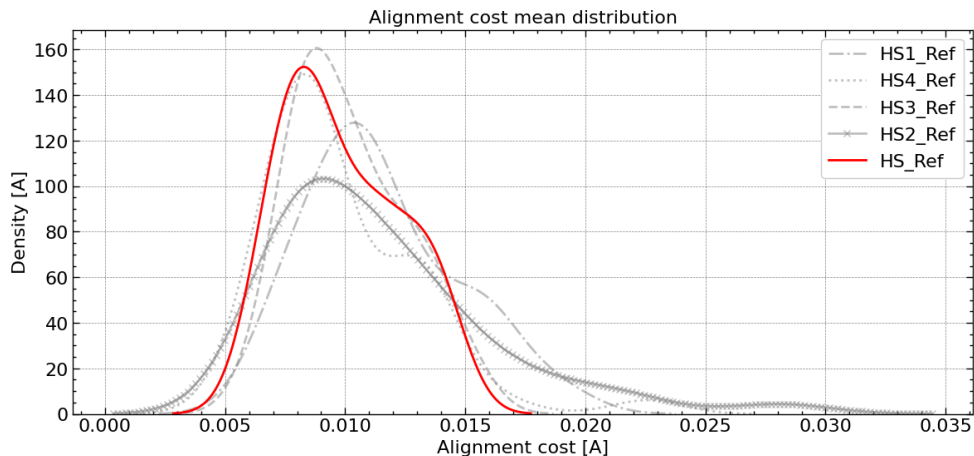


Figura 2.8: Comparação das distribuições de custo para escolha do sinal de referência. **Fonte:** Elaborado pelo próprio autor.

Resultados e Discussão

3.1 Análise do Sinal de ACC

Inicialmente, após o pré-processamento de todos os conjuntos de dados de ACC e sEMG utilizando a metodologia mencionada na Seção 2.1, uma série de análises foram realizadas através dos métodos de alinhamento e autoregressão. Para os sinais de acelerometria, o algoritmo DTW foi aplicado a cada sequência, realizando um alinhamento entre todos os indivíduos com o sinal de referência. Cada indivíduo, pertencente ao grupo saudável ou displásico, produziu um sinal de 30 segundos com aproximadamente 4500 pontos capturados pelos sensores. Os sinais foram cortados em 45 sequências de 100 pontos cada, evidenciando apenas o pulso produzido pelo vetor de ACC_{Total} durante o processo de caminhada dos cães. Além disso, o sinal de referência escolhido, através do critério explicado na Seção 2.5, pertence ao grupo dos cães saudáveis, ou seja, espera-se que o algoritmo DTW compute um custo maior para o conjunto de sinais dos cães displásicos. Outra avaliação importante é o distanciamento entre as distribuições de custo de ambos os grupos, pois pensando na viabilidade de classificação dos indivíduos apoiando-se em técnicas de *Machine Learning*, é necessário averiguar as distribuições dos dados e determinar a possibilidade de aplicação dos métodos de aprendizado de máquina.

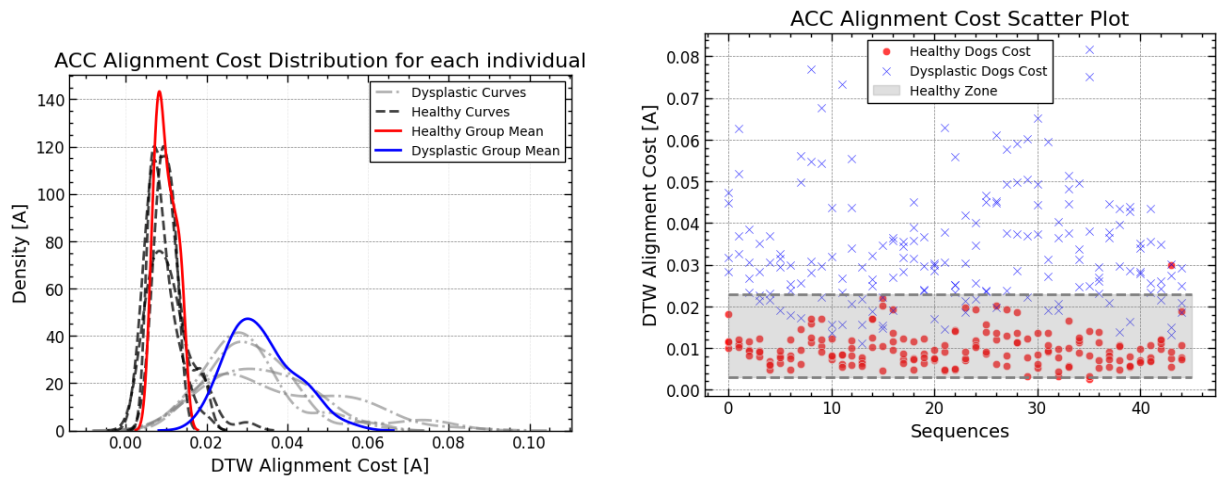


Figura 3.1: Distribuições de custo de cada grupo de sinais gerados através dos alinhamentos dos candidatos com a referência.

Fonte: Elaborado pelo próprio autor.

Figura 3.2: Gráfico de dispersão dos custos do grupo controle e do grupo displásico, evidenciando o intervalo de custo considerado como saudável através do agrupamento dos dados. **Fonte:** Elaborado pelo próprio autor.

Como ilustrado na Figura 3.1, o algoritmo DTW calcula um custo para cada sequência

de cada indivíduo, ou seja, cada cão recebe um conjunto de 45 comparações, representando uma distribuição de custos de alinhamento ótimo. Dessa forma, nota-se padrões de distribuição entre cada grupo. Assim, permitindo tirar conclusões sobre o comportamento dos grupos dos cães. Além disso, através das Figuras 3.1 e 3.2, confirma-se a viabilidade de utilizar o algoritmo DTW para extrair informações biológicas, como padrões e comportamentos, dos sinais de ACC. Adiante, na Figura 3.2, os custos do grupo de cães com displasia apresentaram grandes flutuações, com custos de alinhamento baixos e altos, permitindo uma análise de incerteza entre os conjuntos de dados. Outra informação importante que pode ser extraída da Figura 3.2 é o intervalo do grupo controle, visto que a distribuição dos dados se concentra entre o intervalo $[0.003, 0.025]$. Esse resultado é relevante, pois se torna um possível parâmetro de diagnóstico da DCF, ou seja, cães com distribuições de custo majoritariamente fora desse intervalo podem pertencer ao grupo displásico. Além disso, alguns cálculos como a média (μ) e o desvio padrão (σ) foram feitos, visando obter resultados quantitativos sobre como essas distribuições diferem entre si, assim como observado na Tabela 3.1.

Tabela 3.1: Análise descritiva de todas as distribuições de custo de alinhamento.

HS	HS_σ	HS_μ	DS	DS_σ	DS_μ
HS1	0.003785	0.008538	DS1	0.015396	0.036048
HS2	0.003675	0.010970	DS2	0.016458	0.037548
HS3	0.005540	0.010886	DS3	0.010397	0.029618
HS4	0.002546	0.009018	DS4	0.010292	0.031201
HS_{Médio}	0.003887	0.009853	DS_{Médio}	0.013136	0.033604

Em que os acrônimos *HS* e *DS* significam, respectivamente, Sinal Saudável e Sinal Displásico - traduzindo do Inglês. Analisando o desvio padrão de ambos os grupos de distribuições, observa-se maiores valores para os sinais de cães com displasia, confirmando as flutuações dos custos de alinhamento descritas anteriormente. Esse resultado indica que os sinais displásicos possuem padrões diferentes ou irregularidades em relação ao grupo controle, assim mostrando uma maior instabilidade no movimento dos cães portadores de DCF. Analisando a média e o desvio padrão dos sinais saudáveis através da Tabela 3.1, nota-se uma maior homogeneidade nos dados de custo de alinhamento. Ou seja, isso mostra que a distribuição dos custos do grupo controle está muito mais concentrado do que o grupo com DCF, assim como observado na Figura 3.2. Entretanto, apesar da separabilidade entre as distribuições dos grupos, existe um pequeno intervalo que contém uma intersecção de valores, ocasionando em possíveis incertezas no diagnóstico dos cães. Esse resultado será melhor discutido na Seção 3.3, onde o cálculo da incerteza será demonstrado e interpretado.

3.2 Teste de Tukey

Através das distribuições de custos de cada grupo, o Teste de *Tukey HSD* foi aplicado, visando realizar múltiplas comparações entre as distribuições dos dados e garantir a similaridade (ou a dissimilaridade) entre os conjuntos (Tabela 3.2). Esse teste é importante, pois permite comparar o quão próximo são os conjuntos de dados entre si, através das médias das distribuições. Dessa forma, confirmando o agrupamento dos dados discutidos na Seção 3.1.

Tabela 3.2: Múltiplas Comparações entre as Médias - Tukey HSD, FWER=0.05

group1	group2	meandiff	p-adj	lower	upper	reject
DS1	DS2	0.0015	0.9977	-0.0052	0.0081	False
DS1	DS3	-0.0063	0.0801	-0.0130	0.0004	False
DS1	DS4	-0.0047	0.3728	-0.0114	0.0019	False
DS1	HS1	-0.0269	0.0000	-0.0336	-0.0202	True
DS1	HS2	-0.0244	0.0000	-0.0311	-0.0177	True
DS1	HS3	-0.0246	0.0000	-0.0313	-0.0180	True
DS1	HS4	-0.0264	0.0000	-0.0331	-0.0197	True
DS2	DS3	-0.0078	0.0102	-0.0144	-0.0011	True
DS2	DS4	-0.0062	0.0882	-0.0129	0.0005	False
DS2	HS1	-0.0284	0.0000	-0.0350	-0.0217	True
DS2	HS2	-0.0259	0.0000	-0.0326	-0.0192	True
DS2	HS3	-0.0261	0.0000	-0.0327	-0.0194	True
DS2	HS4	-0.0278	0.0000	-0.0345	-0.0211	True
DS3	DS4	0.0015	0.9967	-0.0051	0.0082	False
DS3	HS1	-0.0206	0.0000	-0.0273	-0.0140	True
DS3	HS2	-0.0181	0.0000	-0.0248	-0.0114	True
DS3	HS3	-0.0183	0.0000	-0.0250	-0.0117	True
DS3	HS4	-0.0201	0.0000	-0.0268	-0.0134	True
DS4	HS1	-0.0222	0.0000	-0.0288	-0.0155	True
DS4	HS2	-0.0197	0.0000	-0.0264	-0.0130	True
DS4	HS3	-0.0199	0.0000	-0.0265	-0.0132	True
DS4	HS4	-0.0216	0.0000	-0.0283	-0.0149	True
HS1	HS2	0.0025	0.9504	-0.0042	0.0092	False
HS1	HS3	0.0023	0.9660	-0.0044	0.0090	False
HS1	HS4	0.0006	1.0000	-0.0061	0.0073	False
HS2	HS3	-0.0002	1.0000	-0.0069	0.0065	False
HS2	HS4	-0.0019	0.9885	-0.0087	0.0048	False
HS3	HS4	-0.0017	0.9934	-0.0084	0.0050	False

Para a aplicação deste teste, a biblioteca *MultiComparison* foi utilizada da linguagem Python, assumindo que todas as distribuições são normais. As hipóteses do teste são as seguintes:

$$\begin{cases} H_0 : & \text{Grupos Similares} \\ H_a : & \text{Grupos Não Similares} \end{cases}$$

Assim, se a comparação entre duas distribuições resulta na rejeição do teste como *True*, portanto ambas as distribuições são dissimilares. Caso a rejeição do teste se dê como *False*, portanto ambas as distribuições analisadas são similares. Na Tabela 3.2, pode-se observar que entre as distribuições de classes diferentes, o resultado do teste demonstra a rejeição da hipótese nula como **verdadeira**. Assim, confirmando a diferença entre as distribuições das classes. Além disso, o teste de *Tukey* também **não rejeita** a hipótese nula quando se compara distribuições de mesma classe entre si, ou seja, afirmando a similaridade entre as distribuições de mesma classe. Entretanto, apesar do sucesso do teste com a maioria das comparações, há uma falha na rejeição do teste entre as distribuições $DS2 \times DS3$, ou seja,

apesar de $DS2, DS3 \in$ Grupo DCF, o teste de *Tukey* computa a rejeição do teste como *True*. Dessa forma, classificando ambas as distribuições como diferentes entre si. Uma possibilidade para justificação deste erro se dá pela Tabela 3.1, havendo um contraste entre as médias $DS2$ e $DS3$, ou seja, o baixo valor da média $DS3$ pode ter levado a rejeição da hipótese nula do teste de *Tukey*.

3.3 Cálculo da Incerteza Intrínseca ao Sinal de ACC

Para compreender a incerteza entre as duas classes de distribuições, ambas as curvas de distribuições médias das classes foram integradas, visando encontrar a área de intersecção entre as curvas e especular a porcentagem de incerteza. Esse passo é importante, pois auxilia no processo de definição de um limite de distinção entre os grupos. Inicialmente, para calcular a intersecção entre as áreas, é necessário integrar ambas as curvas de distribuição de custo (A_{HS} e A_{DS} - Equação 3.1 e 4.2, respectivamente), onde após esse procedimento, calcula-se a área de intersecção ($A_{\text{intersecção}}$ - Equação 3.3). A partir desses resultados, um esquema de proporção pode ser feito para aproximar a porcentagem de incerteza de classificação entre os grupos.

$$A_{HS} = \frac{1}{\sigma_{C_{HS}}\sqrt{2\pi}} \int e^{-\frac{1}{2}\left(\frac{x_{C_{HS}} - \mu_{C_{HS}}}{\sigma_{C_{HS}}}\right)^2} dx_{C_{HS}} \quad (3.1)$$

$$A_{DS} = \frac{1}{\sigma_{C_{DS}}\sqrt{2\pi}} \int e^{-\frac{1}{2}\left(\frac{x_{C_{DS}} - \mu_{C_{DS}}}{\sigma_{C_{DS}}}\right)^2} dx_{C_{DS}} \quad (3.2)$$

$$A_{\text{intersecção}} = \frac{1}{\sqrt{2\pi}} \left(\frac{1}{\sigma_{C_{DS}}} \int e^{-\frac{1}{2}\left(\frac{x_{C_{DS}} - \mu_{C_{DS}}}{\sigma_{C_{DS}}}\right)^2} dx_{C_{DS}} - \frac{1}{\sigma_{C_{HS}}} \int e^{-\frac{1}{2}\left(\frac{x_{C_{HS}} - \mu_{C_{HS}}}{\sigma_{C_{HS}}}\right)^2} dx_{C_{HS}} \right) \quad (3.3)$$

Onde $\mu_{C_{DS}}$ e $\mu_{C_{HS}}$ são os valores médios de cada distribuição, $x_{C_{DS}}$ e $x_{C_{HS}}$ são os respectivos valores dos dados, e $\sigma_{C_{DS}}$ e $\sigma_{C_{HS}}$ são os desvios padrão. Para resolver as integrais numéricas, aplicou-se o método de 1/3 de Simpson em Python, permitindo calcular todas as áreas que serão utilizadas para especular a proporção de incerteza. Os resultados indicam que a área da intersecção é de $A_{\text{intersecção}} \approx 0,02103$, as demais áreas obtidas através das Equações 4.1 e 4.2 são $A_{HS} \approx 0,90905$ e $A_{DS} \approx 0,95478$. Assim, para calcular a incerteza, faremos:

$$\text{Incerteza} = \frac{A_{\text{intersecção}} \times 100\%}{A_{DS} + A_{HS}} \approx 1,13\%$$

Portanto, a distribuição dos dados mostra que entre as classes há uma incerteza de 1,13% de custos de sequências que são difíceis de classificar. Isso indica que existe uma diferença sólida entre as classes de cães, potencializando o diagnóstico através dos sinais de ACC. Esse nível de confiança na classificação demonstra o potencial das análises via algoritmo DTW, tornando-se uma ferramenta em potencial para o diagnóstico da DCF, principalmente quando integrada com outras modalidades de diagnóstico.

3.4 Análise do Sinal de sEMG

Além da perspectiva de movimento através dos sensores inerciais (ACC), os sensores de eletromiografia de superfície (sEMG) fornecem *insights* do movimento através do estresse dos membros musculares. Esses sinais medem o uso do músculo na escala de Voltagem [V], indicando a quantidade de estresse muscular que certo músculo produziu ao longo da caminhada. Essa métrica permite a visualização da instabilidade do movimento dos cães através do estresse muscular produzido, conforme a DCF afeta o movimento e exige uma maior compensação muscular dos indivíduos. Por exemplo, em casos mais severos de displasia coxofemoral, a habilidade dos cães de permanecerem em pé é comprometida. Após o pré-processamento dos sinais de sEMG, como apresentado anteriormente (Figura 2.1), padrões de envelopagem podem ser identificados nas Figuras 3.3 e 3.4. Indivíduos do grupo DCF apresentaram maior atividade muscular em contraste com o grupo controle, indicando que os cães saudáveis possuem maior estabilidade no movimento, dessa forma realizando menos esforço muscular durante a caminhada. Ou seja, cães pertencentes ao grupo DCF desenvolvem maior estresse muscular, aumentando o nível de fadiga, causado pela doença motora.

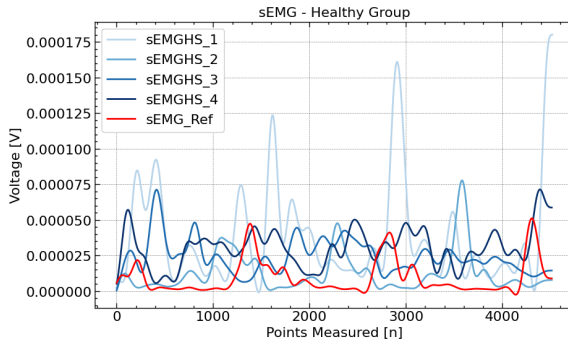


Figura 3.3: Sinais de sEMG do grupo controle pós-processados, nos quais é possível observar alguns padrões e baixos níveis de atividade muscular. **Fonte:** Elaborado pelo próprio autor.

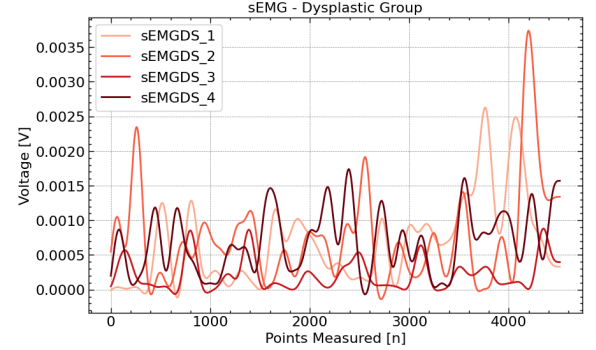


Figura 3.4: Sinais de sEMG do grupo displásico pós-processados, evidenciando uma maior atividade muscular associada à condição de displasia de quadril. **Fonte:** Elaborado pelo próprio autor.

Visando garantir uma análise mais rebuscada dos sinais de sEMG, um modelo $AR(p)$ foi aplicado para a obtenção dos coeficientes $\Phi = \{\phi_1, \dots, \phi_P\}$ que representam as séries temporais de eletromiografia de superfície. Dessa forma, através da biblioteca *AutoReg* do Python, o modelo de autoregressão foi aplicado e ajustado com uma ordem de $P = 12$, isto é, a autoregressão que constrói o sinal observado possui 12 coeficientes: $\phi_1, \dots, \phi_{12}$. A Tabela 3.3 mostra os coeficientes obtidos pelo modelo, onde padrões podem ser notados para cada grupo em relação a magnitude dos coeficientes de cada sinal e sua orientação ($\pm$). A partir desse resultado, o grupo controle demonstrou menor variação de magnitudes entre os coeficientes, podendo estar relacionado com a estabilidade do sinal. Em contraste, o grupo DCF mostrou coeficientes mais altos, maiores variações de coeficientes conforme o aumento da ordem P , possivelmente pela intensa dinâmica do sinal, sua instabilidade e maior atividade muscular [V]. Assim, através desse modelo, pode-se quantificar padrões relacionados com a instabilidade do movimento e o aumento do estresse muscular para compensação motora. Além disso, esses resultados demonstram agrupamentos nos valores

dos coeficientes obtidos pelo modelo $AR(p)$, permitindo que outras análises possam ser realizadas para o diagnóstico mais preciso da DCF através dos sensores de sEMG.

Tabela 3.3: Coeficientes AR de cada série temporal, onde DS_i representa o grupo de sinais displásicos e HS_i o grupo controle. Os coeficientes do sinal de referência também são apresentados abaixo.

	DS_1	DS_2	DS_3	DS_4	HS_1	HS_2	HS_3	HS_4	HS_{Ref}
ϕ_0	2.938239e-16	6.733320e-16	5.638260e-16	7.540440e-16	-7.633989e-16	3.222375e-17	7.693169e-17	9.685663e-18	2.166842e-17
ϕ_1	5.314582	5.319570	5.317071	5.321861	4.097503	4.077429	4.076252	4.082881	4.077925
ϕ_2	-9.865076	-9.887012	-9.875157	-9.896480	-4.872470	-4.810228	-4.806899	-4.827689	-4.811893
ϕ_3	5.250869	5.279560	5.262528	5.290916	-0.341271	-0.376873	-0.378411	-0.366266	-0.375778
ϕ_4	4.468023	4.469335	4.469956	4.470818	2.631754	2.589182	2.587044	2.601428	2.590386
ϕ_5	-3.099851	-3.120772	-3.107919	-3.129080	1.473223	1.483191	1.483275	1.479595	1.482731
ϕ_6	-4.207771	-4.214697	-4.211599	-4.217712	-1.095317	-1.055466	-1.053854	-1.067653	-1.056758
ϕ_7	1.186862	1.201962	1.191889	1.207634	-1.982222	-1.963472	-1.962330	-1.968464	-1.963891
ϕ_8	4.160063	4.173863	4.165720	4.178501	-0.401631	-0.422772	-0.423200	-0.415469	-0.421881
ϕ_9	-0.237891	-0.247527	-0.240358	-0.250974	1.669185	1.638244	1.636862	1.647374	1.639149
ϕ_{10}	-4.411523	-4.429870	-4.418442	-4.434893	1.211658	1.213633	1.213115	1.211767	1.213245
ϕ_{11}	3.096057	3.114795	3.102060	3.119858	-1.984319	-1.954124	-1.952597	-1.962539	-1.954865
ϕ_{12}	-0.654333	-0.659201	-0.655746	-0.660453	0.593908	0.581256	0.580744	0.585035	0.581629

A partir dessa tabela, pode-se utilizar técnicas de visualização para averiguar a evolução dos coeficientes ao longo do aumento da ordem P . A Figura 3.5 mostra a variação da magnitude dos coeficientes ϕ_i conforme i aumenta até $P = 12$. Além disso, o coeficiente ϕ_0 é o próprio ruído inicial ε , dessa forma sendo um dos únicos que diferem da dinâmica de alternância observada abaixo.

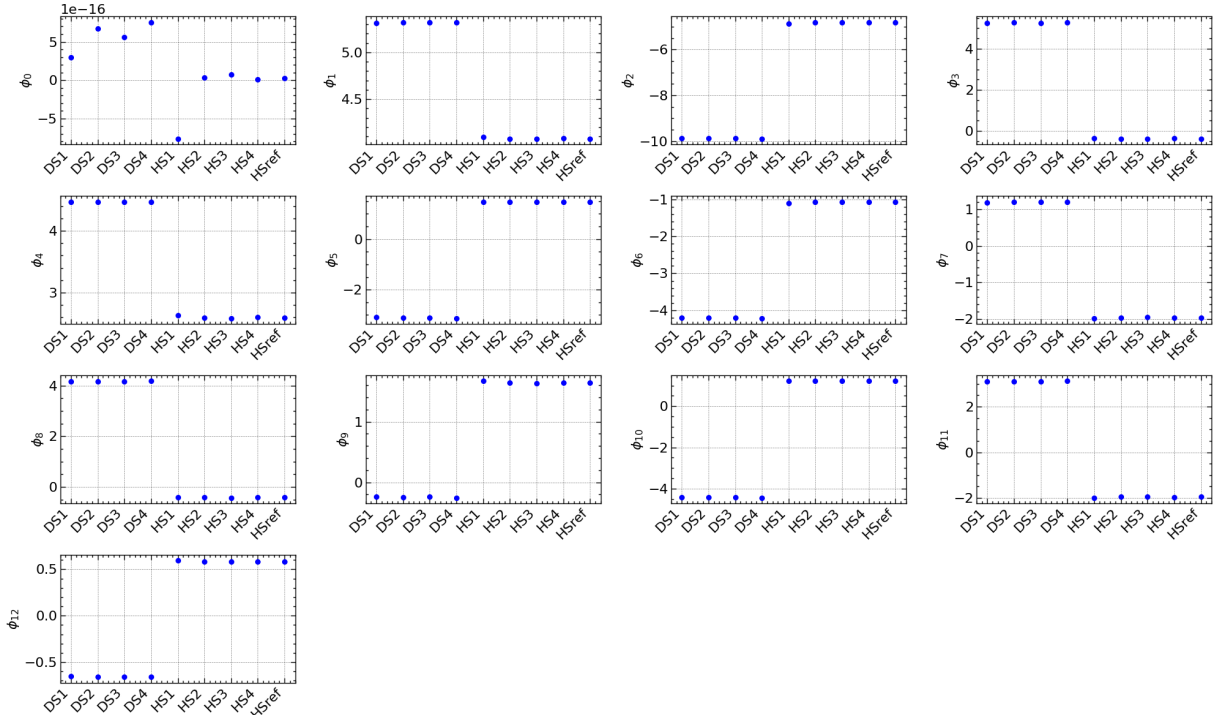


Figura 3.5: Avaliando a dispersão entre as magnitudes dos coeficientes conforme $i \rightarrow P$.
Fonte: Elaborado pelo próprio autor.

Conclusão

4.1 Conclusões Preliminares

Em suma, esse estudo demonstra o potencial de combinar sinais de sensores de sEMG e ACC com o algoritmo DTW e o método $AR(p)$ para auxiliar no diagnóstico da displasia coxofemoral em cães, como um método alternativo e não-invasivo. A análise dos sinais de ACC pelo algoritmo DTW revelou a existência de uma sólida separação entre os grupos de cães controle e displásico, contendo uma incerteza de apenas 1,3%, indicando potenciais de classificação através deste método com certa acurácia. Ademais, o modelo AR aplicado aos sinais de sEMG quantitativamente encontrou padrões entre os grupos de cães, providenciando parâmetros numéricos para possíveis diagnósticos através do agrupamento dos coeficientes ϕ_i da autoregressão. Essas descobertas sugerem que essa abordagem combinada oferece uma alternativa promissora aos métodos diagnósticos tradicionais e mais invasivos, aproveitando equipamentos amplamente disponíveis e de baixo custo, como acelerômetros. Além disso, esses métodos podem auxiliar estudos futuros com um número maior de cães e múltiplas raças. Embora os resultados sejam encorajadores, o estudo é limitado pelo pequeno tamanho da amostra e pelo uso exclusivo de cães da raça Labrador Retriever. Ou seja, mais pesquisas devem ser feitas em diferentes raças e portes, visando analisar os padrões das métricas computadas pelos modelos, a fim de estabelecer diferenças nos padrões capturados pelos modelos e validar o estudo acima. Além disso, a aplicação de algoritmos de aprendizado de máquina à grande quantidade de dados gerados por esses métodos apresenta um potencial significativo para o desenvolvimento de uma ferramenta diagnóstica automatizada e objetiva para displasia coxofemoral, podendo reduzir o tempo de diagnóstico e melhorar a precisão. Portanto, com base nessas conclusões, futuros algoritmos de aprendizado de máquina poderão ser projetados para diagnosticar automaticamente essa condição motora a partir da extração de características dos sinais biológicos pré-processados.

4.2 Procedimentos Futuros

Por fim, os próximos passos dessa pesquisa são: explorar novos bancos de dados com maiores quantidades de amostras de cães, explorar todas as características e métricas que o modelo $AR(p)$ e o algoritmo DTW podem oferecer, construir um banco de dados pré-processado com dados multifatoriais, criar um modelo multimodal de aprendizado de máquina para ser treinado através dos dados pré-processados e validá-lo devidamente. Esses passos serão fundamentais, num momento em que novas métricas podem alimentar o futuro modelo de *Machine Learning* para classificação automatizada da DCF. Além disso, o caráter multifatorial do possível modelo de ML tem um grande potencial de receber diferentes tipos de dados, como os custos de alinhamento, os coeficientes ϕ dos

sinais de sEMG e as matrizes de custo acumulado das sequências alinhadas. Assim, essa abordagem engloba diferentes parâmetros, podendo aumentar a quantidade de análises do futuro modelo e na qualidade do diagnóstico. Outro fator importante são as matrizes de custo acumulada, podendo ser interpretadas como imagens, abrindo margem para o processamento do espaço de características $\mathcal{F}$ e na interpretação dos padrões capturados por cada matriz, possibilitando novas análises e observação de padrões entre os grupos de indivíduos.

Referências Bibliográficas

- [1] PAL, Kunal et al. Medical signal processing in biomedical and clinical applications. *Journal of Healthcare Engineering*, v. 2018, p. 3932471, 2018.
- [2] HYNDMAN, R. J.; ATHANASOPOULOS, G. Forecasting: principles and practice. 2. ed. Melbourne, Australia: OTexts, 2018. Disponível em: <https://OTexts.com/fpp2>. Acesso em: 26 mar. 2025.
- [3] SU, Weiwei et al. Pairwise vs Higher-order interactions: Can we identify the interaction type in coupled oscillators from time series?. *arXiv preprint arXiv:2503.13244*, 2025.
- [4] KIM, Jongseon et al. A Comprehensive Survey of Time Series Forecasting: Architectural Diversity and Open Challenges. *arXiv preprint arXiv:2411.05793*, 2024.
- [5] CRYER, Jonathan D. Time series analysis. Boston: Duxbury Press, 1986.
- [6] ZHANG, Z. Time series analysis. In: *Environmental Data Analysis: Methods and Applications*. Berlin; Boston: De Gruyter, p. 1-44, 2017.
- [7] BOX, George EP et al. Time series analysis: forecasting and control. John Wiley & Sons, 2015.
- [8] BROCKWELL, Peter J.; DAVIS, Richard A. Time series: theory and methods. Springer science & business media, 1991.
- [9] REGIS, Marta; SERRA, Paulo; VAN DEN HEUVEL, Edwin R. Random autoregressive models: A structured overview. *Econometric Reviews*, v. 41, n. 2, p. 207-230, 2022.
- [10] CORDUAS, Marcella; PICCOLO, Domenico. Time series clustering and classification by the autoregressive metric. *Computational statistics & data analysis*, v. 52, n. 4, p. 1860-1872, 2008.
- [11] MÄKI, K. et al. An indication of major genes affecting hip and elbow dysplasia in four Finnish dog populations. *Heredity*, v. 92, n. 5, p. 402-408, 2004.
- [12] ROCHA, Fábio Perón Coelho da et al. Displasia coxofemoral em cães. *Revista Científica Eletrônica de Medicina Veterinária*, v. 4, n. 11, p. 1-7, 2008.
- [13] GINJA, M. M. D. et al. Diagnosis, genetic control and preventive management of canine hip dysplasia: a review. *The Veterinary Journal*, v. 184, n. 3, p. 269-276, 2010.
- [14] BARR, A. R. S.; DENNY, H. R.; GIBBS, C. Clinical hip dysplasia in growing dogs: the long-term results of conservative management. *Journal of Small Animal Practice*, v. 28, n. 4, p. 243-252, 1987.
- [15] GENUÍNO, P. C. et al. Parâmetros radiográficos de displasia coxofemoral na raça Rottweiler. *Arquivo brasileiro de medicina veterinária e zootecnia*, v. 67, p. 1178-1182, 2015.
- [16] LOPEZ, M. J. et al. Prevalence of canine hip dysplasia in 10 breeds in France, a retrospective study of the 1997-2017 radiographic screening period. *PLOS ONE*, v. 15, n. 7, p. 1-11, 2020. DOI: 10.1371/journal.pone.0235847.
- [17] HENRIGSON, B.; NORBERG, I.; OLSSONS, S.-E. On the etiology and pathogenesis of hip dysplasia: a comparative review. *Journal of Small Animal Practice*, v. 7, n. 11, p.

673-688, 1966.

- [18] PINNA, Stefania et al. Prevalence of primary radiographic signs of hip dysplasia in dogs. *Animals*, v. 12, n. 20, p. 2788, 2022.
- [19] GASPAR, Ana R. et al. The Norberg angle is not an accurate predictor of canine hip conformation based on the distraction index and the dorsolateral subluxation score. *Preventive Veterinary Medicine*, v. 135, p. 47-52, 2016.
- [20] BOCKSTAHLER, Barbara B. et al. Correlation of surface electromyography of the vastus lateralis muscle in dogs at a walk with joint kinematics and ground reaction forces. *Veterinary Surgery*, v. 38, n. 6, p. 754-761, 2009.
- [21] NEGRÃO, Roberta Rocha et al. Analysis of time series of surface electromyography and accelerometry in dogs. *Biomedical Signal Processing and Control*, v. 74, p. 103518, 2022.
- [22] BOCKSTAHLER, Barbara et al. Pelvic limb kinematics and surface electromyography of the vastus lateralis, biceps femoris, and gluteus medius muscle in dogs with hip osteoarthritis. *Veterinary Surgery*, v. 41, n. 1, p. 54-62, 2012.
- [23] MARSCHOLLEK, Michael et al. A performance comparison of accelerometry-based step detection algorithms on a large, non-laboratory sample of healthy and mobility-impaired persons. In: 2008 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE, 2008. p. 1319-1322.
- [24] MILLS, Kerry R. The basics of electromyography. *Journal of Neurology, Neurosurgery & Psychiatry*, v. 76, n. suppl 2, p. ii32-ii35, 2005.
- [25] GRAȚIELA-FLAVIA, Deak; FLAVIA, Rusu; EMILIA, Grosu. SURFACE ELECTROMYOGRAPHY IN BIOMECHANICS: APPLICATIONS AND SIGNAL ANALYSIS ASPECTS. *Journal of Physical Education & Sport/Citius Altius Fortius*, v. 25, n. 4, 2009.
- [26] DE LUCA, Carlo J. The use of surface electromyography in biomechanics. *Journal of applied biomechanics*, v. 13, n. 2, p. 135-163, 1997.
- [27] FAISAL, I. Arun; PURBOYO, T. Waluyo; ANSORI, A. Siswo Raharjo. A review of accelerometer sensor and gyroscope sensor in IMU sensors on motion capture. *J. Eng. Appl. Sci*, v. 15, n. 3, p. 826-829, 2019.
- [28] FILIPPESCHI, Alessandro et al. Survey of motion tracking methods based on inertial sensors: A focus on upper limb human motion. *Sensors*, v. 17, n. 6, p. 1257, 2017.
- [29] SENIN, Pavel. Dynamic time warping algorithm review. *Information and Computer Science Department University of Hawaii at Manoa Honolulu, USA*, v. 855, n. 1-23, p. 40, 2008.
- [30] GIORGINO, Toni. Computing and visualizing dynamic time warping alignments in R: the dtw package. *Journal of statistical Software*, v. 31, p. 1-24, 2009.
- [31] VELICHKO, V. M.; ZAGORUYKO, N. G. Automatic recognition of 200 words. *International Journal of Man-Machine Studies*, v. 2, n. 3, p. 223-234, 1970.
- [32] KAUR, Jatinder; PARMAR, Kulwinder Singh; SINGH, Sarbjit. Autoregressive models in environmental forecasting time series: a theoretical and application review. *Environmental Science and Pollution Research*, v. 30, n. 8, p. 19617-19641, 2023.
- [33] ULLRICH, Torsten. On the autoregressive time series model using real and complex analysis. *Forecasting*, v. 3, n. 4, p. 716-728, 2021.
- [34] KLEIN, Judy L. *Statistical visions in time: a history of time series analysis, 1662-1938*. Cambridge University Press, 1997.
- [35] PAISS, Omry; INBAR, Gideon F. Autoregressive modeling of surface EMG and its spectrum with application to fatigue. *IEEE transactions on biomedical engineering*, n.

10, p. 761-770, 1987.

[36] CHEN, Weitian et al. Solutions of Yule-Walker equations for singular AR processes. *Journal of Time Series Analysis*, v. 32, n. 5, p. 531-538, 2011.

[37] ZHENG, Wei Xing. A least-squares based method for autoregressive signals in the presence of noise. *IEEE Transactions on Circuits and Systems II: Analog and Digital Signal Processing*, v. 46, n. 1, p. 81-85, 1999.

[38] ABDI, Hervé; WILLIAMS, Lynne J. Tukey's honestly significant difference (HSD) test. *Encyclopedia of research design*, v. 3, n. 1, p. 1-5, 2010.

[39] MIDWAY, Stephen et al. Comparing multiple comparisons: practical guidance for choosing the best multiple comparisons test. *PeerJ*, v. 8, p. e10387, 2020.

[40] KAMPER, Herman. Código disponível em: https://github.com/kamperh/lecture_dtw_notebook. 2021. Disponível sob a licença Creative Commons Attribution-ShareAlike 4.0 (CC BY-SA 4.0).

[41] MÜLLER, Meinard. *Information retrieval for music and motion*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2007.