

UNIVERSIDADE FEDERAL DE OURO PRETO
INSTITUTO DE CIÊNCIAS EXATAS E APLICADAS
DEPARTAMENTO DE COMPUTAÇÃO E SISTEMAS

FABIO DUARTE GUIEIRO LOPES

**ANÁLISE DA CINEMÁTICA DA BIOMECÂNICA POR MEIO DE TÉCNICAS DE
PROCESSAMENTO DE SINAIS E APRENDIZAGEM DE MÁQUINA**

João Monlevade

2021

FABIO DUARTE GUIEIRO LOPES

**Análise da Cinemática da Biomecânica por meio de Técnicas de
Processamento de Sinais e Aprendizagem de Máquina**

Monografia apresentada ao curso de Sistemas de Informação do Instituto de Ciências Exatas e Aplicadas, da Universidade Federal de Ouro Preto, como requisito parcial para aprovação na Disciplina “Trabalho de Conclusão de Curso II”.

Orientador: Talles Henrique de Medeiros

João Monlevade

2021

SISBIN - SISTEMA DE BIBLIOTECAS E INFORMAÇÃO

L864a Lopes, Fabio Duarte Guieiro .
Análise da cinemática da biomecânica por meio de técnicas de processamento de sinais e aprendizagem de máquina. [manuscrito] / Fabio Duarte Guieiro Lopes. - 2021.
58 f.: il.: color., gráf., tab..

Orientador: Prof. Dr. Talles Henrique de Medeiros.
Monografia (Bacharelado). Universidade Federal de Ouro Preto.
Instituto de Ciências Exatas e Aplicadas. Graduação em Sistemas de Informação .

1. Biomecânica . 2. Cinemática. 3. Fourier, Transformadas de. 4. Inteligência artificial. 5. Processamento de sinais. I. Medeiros, Talles Henrique de. II. Universidade Federal de Ouro Preto. III. Título.

CDU 531.1:004.8

Bibliotecário(a) Responsável: Flavia Reis - CRB6-2431



FOLHA DE APROVAÇÃO

Fábio Duarte Guieiro Lopes

**Análise da Cinemática da Biomecânica por meio de Técnicas de
Processamento de Sinais e Aprendizagem de Máquina**

Monografia apresentada ao Curso de Sistemas de Informação da Universidade Federal de Ouro Preto como requisito parcial para obtenção do título de Bacharel em Sistemas de Informação

Aprovada em 30 de Abril de 2021

Membros da banca

Doutor - Prof. Talles Henrique de Medeiros - Orientador(a) - Universidade Federal de Ouro Preto
Doutor - Prof. Luiz Carlos Bambirra Torres - Universidade Federal de Ouro Preto
Doutor - Prof. Rafael Frederico Alexandre - Universidade Federal de Ouro Preto

Prof. Talles Henrique de Medeiros, orientador do trabalho, aprovou a versão final e autorizou seu depósito na Biblioteca Digital de Trabalhos de Conclusão de Curso da UFOP em 14/06/2021.



Documento assinado eletronicamente por **Talles Henrique de Medeiros, PROFESSOR DE MAGISTERIO SUPERIOR**, em 15/06/2021, às 13:50, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



A autenticidade deste documento pode ser conferida no site http://sei.ufop.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **0182553** e o código CRC **98EECF73**.

AGRADECIMENTOS

A todos que contribuem à democratização do saber, aos que viabilizam o acesso de pessoas de origem pobre à educação superior, meio mais efetivo de ascensão social. Aos profissionais que se dedicam a ensinar, pois só por meio do conhecimento o homem se transforma e, se homens se transformam, a realidade se transforma. Por fim, agradeço aos que se dedicam à arte, que aplaca a dureza do viver.

“Qualquer tecnologia suficientemente
avançada é indistinguível de magia”

Sir Arthur C. Clarke

RESUMO

A busca pelas práticas de atividades físicas como forma de lazer e de manutenção de uma vida saudável tem crescido significativamente nos últimos anos. Ao mesmo tempo, lesões musculoesqueléticas têm se tornado mais comuns, como reflexo da prática de exercícios físicos sem o devido acompanhamento e preparo. Este trabalho aplica conceitos matemáticos e de inteligência artificial com o objetivo de identificar os padrões de cinemática dos membros inferiores usando como insumo uma base de dados de valores de movimentações articulares (cinemática) de 28 atletas em 3 diferentes velocidades, aplicando algoritmos como a transformada de Fourier, K-Means e KNN para buscar fazer inferências sobre o que os dados de movimentação articular podem dizer sobre o atleta. Após a análise dos dados não foi possível identificar padrões suficientes para que fossem feitas tais inferências, o que motivou um estudo mais profundo sobre essas características revelando grande homogeneidade dos indivíduos.

Palavras-chave: Biomecânica de corrida, Cinemática, Transformada de Fourier, K-Means, KNN

ABSTRACT

The search for physical activities as a way of leisure on healthy life maintenance has grown significantly in the past few years. In the same rhythm, musculoskeletal injuries have become more common, as a reflex of the practice of physical activities without supervision and preparation. This work applies mathematical and artificial intelligence concepts with the objective of identifying kinematics patterns using as input a database about articulation movements (kinematics) of 28 athletes at 3 different speeds, applying algorithms such as Fourier transform, K-Means and KNN to make inferences about what the articulation movements data have to say about the athlete. After the data analysis it was not possible to identify enough patterns to make such inferences, which motivated a deeper study about the individuals features, revealing big homogeneity in it.

Palavras-chave: Running Biomechanics, Kinematic, Fourier Transform, K-Means, KNN

LISTA DE FIGURAS

Figura 1 - Marcadores utilizados na coleta de dados de movimentos cinemáticos.	18
Figura 2 - Visão ampliada do Laboratório de Biomecânica e Controle Motor.	20
Figura 3 – Histograma da Altura dos Atletas.	21
Figura 4 – Histograma da Experiência dos Atletas.	22
Figura 5 – Histograma da Idade dos Atletas.	22
Figura 6 – Histograma de IMC dos Atletas.	23
Figura 7 – Histograma do ritmo de corrida (min/km) dos atletas.	23
Figura 8 – Histograma de Peso dos Atletas.	24
Figura 9 – Histograma de volume de Volume Semanal de corrida dos atletas.	24
Figura 10 – Histograma de Sessões por Semana.	25
Figura 11 – Exemplo da aplicação da transformada de Fourier em processamento de imagens.	26
Figura 12 – Demonstração da aplicação da transformada de Fourier sobre sinal resultante de um dos marcadores articulares.	27
Figura 13 – Demonstração do funcionamento do algoritmo K-Means.	29
Figura 14 – Ilustração do funcionamento do algoritmo KNN.	32
Figura 15 – Diagrama de fluxo do algoritmo.	33
Figura 16 – Exemplo de arquivo de dados.	34
Figura 17 – Gráfico da crista íliaca direita (dimensão X) do arquivo RBDS001runT25markers.	35
Figura 18 – Gráfico da crista íliaca direita (dimensão Y) do arquivo RBDS001runT25markers.	35
Figura 19 – Gráfico da crista íliaca direita (dimensão Z) do arquivo RBDS001runT25markers.	36
Figura 20 – Gráfico da espinha íliaca esquerda (dimensão X) do arquivo RBDS001runT25markers e o gráfico da sua transformada de Fourier.	37
Figura 21 – Gráfico da espinha íliaca esquerda (dimensão Y) do arquivo RBDS001runT25markers e o gráfico da sua transformada de Fourier.	37

Figura 22 – Gráfico da espinha ilíaca esquerda (dimensão Z) do arquivo RBDS001runT25markers e o gráfico da sua transformada de Fourier.....	37
Figura 23 – Gráfico de transformada de Fourier com destaque em alguns picos...	39
Figura 24 – Gráfico resultante dos testes de número de componentes periódicos da transformada de Fourier.....	40
Figura 25 – Estrutura do vetor de dados.....	41
Figura 26 – Gráficos originais e transformados para uma mesma articulação nas dimensões X, Y e Z, respectivamente.....	45
Figura 27 – Gráfico da Transformada de Fourier para a coluna ilíaca ântero-superior direita de um atleta a 2,5 m/s.....	46
Figura 28 – Gráfico da transformada de Fourier para a coluna ilíaca ântero-superior direita de um atleta a 3,5 m/s.....	46
Figura 29 – Gráfico da Transformada de Fourier para a coluna ilíaca ântero-superior direita de um atleta a 4,5 m/s.....	47
Figura 30 – Agrupamento dos dados retornado pelo K-Means.....	48
Figura 31 – Gráfico comparativo da altura média (em centímetros) dos grupos.....	48
Figura 32 – Gráfico comparativo da idade média dos grupos.....	49
Figura 33 – Gráfico comparativo da experiência média (em meses) dos grupos...	49
Figura 34 – Gráfico comparativo do peso médio (em quilos) dos grupos.....	49
Figura 35 – Gráficos dos perfis dos corredores dentro dos grupos.....;	50
Figura 36 – Gráficos da quantidade de sessões de corrida que os corredores fazem por semana em cada grupo.....	50
Figura 37 – Gráficos da distância percorrida pelos corredores por semana em cada grupo.....	50
Figura 38 – Gráfico da acurácia alcançada pelo KNN conforme a variação do valor de K.....	51

LISTA DE TABELAS

Tabela 1 - Marcadores articulares utilizados na coleta de dados.....	18
Tabela 2 - Testes de tempo de execução do algoritmo com as articulações do quadril isoladas X todas as articulações.....	38
Tabela 3 - Testes com métricas do K-Means.....	42
Tabela 4 - Testes do melhor valor de K no KNN.....	43
Tabela 5 - Matriz de confusão resultante do KNN.....	52

SUMÁRIO

1 Introdução	13
1.1 Objetivos	13
1.2 Justificativa	14
1.3 Estrutura do Trabalho	15
2 Conceitos Gerais e Revisão da Literatura	16
2.1 Biomecânica de Corrida	16
2.2 Base de Dados	17
2.2.1 O Sistema	19
2.2.2 Análise do Perfil dos Indivíduos	21
2.3 Trabalho sobre Cinética de Corrida	25
2.4 Transformada de Fourier	26
2.5 Agrupamento de Dados	27
2.5.1 K-Means	28
2.5.2 Métricas de Validação de Cluster	29
2.5.2.1 Coeficiente de Silhueta	29
2.5.2.2 Davies-Bouldin	30
2.5.2.3 Calinski-Harabasz	30
2.6 Classificação de Dados	31
2.6.1 K Nearest Neighbors	31
3 Metodologia	33
3.1 Pré-processamento dos Dados	33
3.1.1 Manipulação da Base de Dados	34
3.1.2 Transformada de Fourier	36
3.1.3 Seleção das Componentes da Transformada de Fourier	38
3.1.4 Preparação do vetor de dados	40
3.2 Aplicação do K-Means	41
3.3 Aplicação do KNN	42
4 Apresentação e Análise dos Resultados	44
4.1 Análise da Transformada de Fourier	44
4.2 K-Means	47
4.3 Análise do KNN	51
5 Conclusões	53
Referências	55

1 Introdução

Nos dias de hoje, as discussões sobre a qualidade de vida e a saúde têm ganhado cada vez mais espaço na sociedade. Neste cenário, a longevidade e bem-estar tornaram-se, além de uma preocupação de saúde pública. Como consequência, o número de indivíduos que buscam a prática de algum tipo de atividade física, vem se tornando cada vez mais expressivo.

Passamos a almejar não só a saúde física, mas também um equilíbrio entre saúde física, mental e emocional. Dentro desse âmbito, destacam-se as atividades físicas ao ar livre, por exemplo, caminhadas e corridas, também conhecidas como pedestrianismo (SALGADO, J.V. 2006).

Segundo Williams et al (1996) os benefícios da prática de corrida estão relacionados às funções metabólicas, prevenção de doenças ósseas (osteoporose), doenças crônico-degenerativas (hipertensão arterial e diabetes), diminuição dos riscos de doenças cardiovasculares, além de melhoras nos níveis de colesterol, aumentando o HDL e diminuindo o LDL.

Um estudo de Weineck (2003) aponta ainda o sentimento de prazer, a sensação de bem-estar, realização própria, o controle do estresse e a superação de limites, como fatores motivacionais para a procura da corrida de rua. Por outro lado, tem sido observado que muitos dos novos praticantes não contam com o acompanhamento de um profissional, e isto tem culminado num aumento do número, e da gravidade das lesões musculoesqueléticas.

Apesar de a taxa de lesões reportada em diferentes estudos variar de 15 a 85%, devido a um desencontro nos critérios para categorizar uma lesão (Heiderscheit, 2014), nos últimos 40 anos, ela tem se mostrado estar próximo de 50%, a despeito da evolução tecnológica significativa no desenvolvimento do calçado de corrida (Nigg et al, 2015). Um estudo de Borel, Wyngrid Porfirio et al.(2019) analisou diversos artigos e ponderando os critérios de lesão que cada autor utilizou, chegou ao resultado que a prevalência de lesões no Brasil é de 36,5%.

A fim de melhorar a qualidade de vida dos praticantes e, conseqüentemente, diminuir o índice de lesões nestes indivíduos, é cada vez mais frequente o uso de algoritmos no auxílio ao processo de diagnóstico (Paniagua et. al 2015) e (Steiner MTA et al 2006), à predição de riscos (HORNG S et al 2011) e à biomedicina (OBENSHAIN M. K. 2014).

1.1 Objetivos

Este estudo busca analisar e compreender os dados de biomecânica de indivíduos praticantes de corrida, sobre uma base de dados pública catalogada em 2017 (disponível no endereço: <https://peerj.com/articles/3298/>). Este estudo captou movimentações de corredores, dentre eles amadores e profissionais, com o intuito de analisar a incidência de padrões que permitam fazer inferências sobre a relação dos dados de cinemática, isto é, movimentações dos corredores, e resultados, por exemplo, da incidência de lesões.

O estudo vem dar continuidade ao trabalho de Mônica Oliveira (2018), que estudou padrões de cinética, peso resultante da passada, e sua influência no histórico de lesão dos atletas. O presente estudo, somado aos esforços de Mônica, pode representar grande avanço no estudo da biomecânica e sua relação com lesões musculoesqueléticas.

A base de dados utilizada foi gerada por sensores que captaram, por um período de tempo, a movimentação da corrida, que pode ser interpretado como um sinal. Para que seja possível manipular e interpretar as informações contidas na base de dados, foi estudada a aplicação da Transformada de Fourier (BARON DE FOURIER, J. B. J.1822), a fim de dividir o sinal em componentes periódicos. Desta forma, foram obtidas as principais características do sinal, isto é, das movimentações dos atletas.

A Transformada de Fourier (TF) foi usada para extrair as características dos sinais coletados, armazenados e disponibilizados pela equipe (referência). Estes dados brutos foram processados pela TF para gerar um conjunto de características relevantes para um estudo mais profundo da base, considerando análise de *clustering* e classificação de possíveis padrões.

Portanto, as hipóteses levantadas no início deste estudo questionavam a existência de similaridades entre grupos de indivíduos e, ao mesmo tempo, se haveriam padrões que permitissem uma análise preditiva de lesões dentro do grupo de indivíduos estudados. Caso algumas das hipóteses se confirmassem, quais as explicações poderiam ser oferecidas.

1.2 Justificativa

De acordo com Hatze (1974) e Hall (2016), a biomecânica é a ciência que estuda o corpo humano produzindo movimento sob conceitos da mecânica. A mecânica encontra-se dividida em, estática e dinâmica. Sendo a estática o corpo parado, ou seja, sem deslocamento e a dinâmica com deslocamento. A mecânica dinâmica pode ser analisada utilizando os conceitos da cinética e cinemática.

A cinemática descreve os aspectos dos movimentos lineares e angulares, como deslocamento, velocidade, aceleração e tempo e foi utilizada como objeto do presente trabalho.

Já a análise biomecânica quantitativa consiste em utilizar a tecnologia para capturar informações acerca do movimento (dinâmica) e transcrever os aspectos do movimento em formas numéricas. Compõem a análise biomecânica quantitativa os seguintes métodos: Antropometria, Dinamometria, Cinemetria e Eletromiografia (EMG).

Neste trabalho foram utilizados os métodos de antropometria, dinamometria e cinemetria. O primeiro, de acordo com Melo e Santos. (2000), serve para criar modelos antropométricos para auxiliar nos treinos otimizando a individualidade de cada corredor. Sua utilização foi feita nas duas primeiras etapas do estudo da base de dados. Já a dinamometria é definida por Suzuki et al (2016) como método utilizado para mensurar as forças atuantes no sistema e pode ser observado na sexta etapa, onde se analisa o desempenho de cada um dos participantes através de velocidades de 2,5 m/s, 3,5 m/s e 4,5 m/s. Por fim, a cinemetria é, de

acordo com Chung (2000), a análise de movimento a partir de imagens, e isso pode ser visto na seção que aborda o sistema, onde são explicados os 12 pontos de câmera utilizados no laboratório.

1.3 Estrutura do Trabalho

Este texto está dividido em cinco capítulos. No primeiro capítulo, há uma breve introdução sobre o objeto de estudo, bem como os objetivos a serem alcançados e a justificativa, também dita como relevância desse presente.

O capítulo 2 traz o referencial teórico sobre os trabalhos que inspiraram e trouxeram fomento para que este trabalho fosse possível, bem como a introdução de conceitos que precisam ser inseridos para a contextualização do estudo.

O capítulo 3 traz a metodologia e todos os passos percorridos no desenvolvimento da pesquisa, enquanto o capítulo 4 traz os resultados que foram alcançados e algumas análises sobre eles. Por fim, no capítulo 5 são apresentadas as conclusões e sugestões para trabalhos futuros.

2 Conceitos Gerais e Revisão da Literatura

2.1 Biomecânica de Corrida

Biomecânica é uma disciplina entre as ciências derivadas das ciências naturais, que se ocupa com análises físicas de sistemas biológicos e, conseqüentemente, análises físicas de movimentos do corpo humano (AMADIO et al, 1999). Através desse ponto de vista, mais especificamente pensando em exercício físico, a biomecânica busca estudar os esforços externos para que exista uma melhor compreensão das repercussões destes nas estruturas internas.

Em geral, a biomecânica da corrida é estudada definindo-se primeiramente o ciclo da marcha, que é a unidade básica de medida de eventos em uma análise da locomoção humana, uma vez que o andar ou o correr representam movimentos que ocorrem de forma cíclica e repetitiva. O ciclo da marcha pode ser identificado com o início do contato do pé com o solo (contato inicial) e terminando no toque seguinte do mesmo pé. Esse ciclo também pode ser referido como uma passada e, por sua vez, uma passada é constituída por dois passos, cada um iniciado por uma das pernas. Na marcha, esse ciclo pode ser dividido em duas fases: apoio e balanço. A fase de apoio inicia-se no contato inicial e termina quando o mesmo pé decola do chão, o que é conhecido como *toe off* (retirada dos dedos, em uma tradução direta do inglês).

O *toe off* denota o início da fase aérea (balanço) que, por sua vez, termina quando este mesmo pé retorna a realizar um novo contato. O período das fases de apoio e balanço durante o ciclo da marcha é afetado pela velocidade da marcha. A fase de apoio nos diferentes tipos de marcha pode representar cerca de 60, 50, 30 ou 20% do ciclo à medida que a velocidade aumenta, ou seja, para o andar, marcha atlética, corrida e tiro de velocidade, respectivamente (VAUGHAN C. L., 1984).

É importante salientar também que, para o mesmo tipo de marcha, a duração da passada diminui à medida que aumenta a velocidade da marcha, e esta redução ocorre, principalmente, por causa da redução do tempo de apoio (contato com o solo). Há muitos estudos na área da saúde que observam o impacto da atividade da corrida no corpo humano, como o de Mangini (2016) que aponta o pico de concentração de lactato sanguíneo quando o corredor eleva ao máximo sua velocidade de corrida.

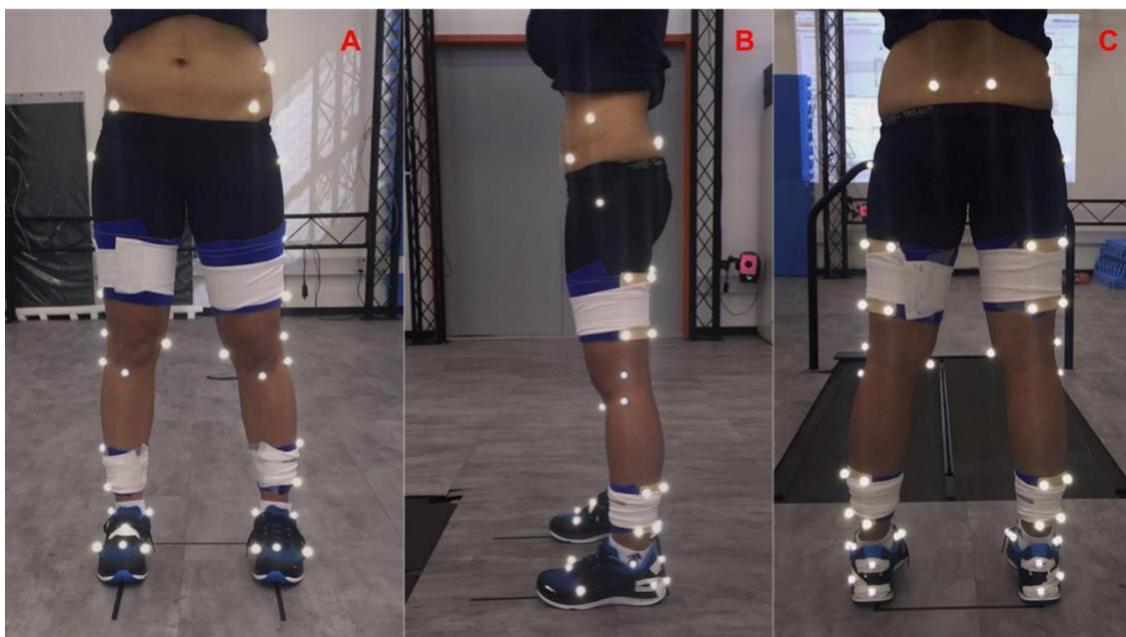
Por outro lado, não são tão volumosas as pesquisas que relacionam a biomecânica de corrida com a ocorrência de lesões musculoesqueléticas, destacando-se o estudo de Fukuchi (2017) sugere que as estratégias para prevenção de lesões em idosos corredores devem se concentrar na melhora da mobilidade articular, principalmente do joelho. Já Portela et al (2019) indica que o treinamento do atrator permite corrigir o padrão de movimento (cinemática), reeducando a corrida de atletas, o que poderia ser um preventivo de lesões.

2.2 Base de Dados

O presente trabalho se utiliza de uma base de dados sobre a movimentação articular de 28 indivíduos em 3 diferentes velocidades. Esta base de dados é o resultado de um estudo desenvolvido por Reginaldo Fukuchi, Claudiane Fukuchi e Marcos Duarte, pesquisadores da Universidade Federal do ABC.

Para o desenvolvimento do estudo, os pesquisadores recrutaram voluntários através de panfletos e de propagandas via internet. Os requisitos mínimos para participação eram que: 1) o participante fosse um corredor regular, isto é, com hábito de corrida de, pelo menos, 20 km por semana, 2) que fosse capaz de manter um ritmo de corrida de 5 minutos por quilômetro durante corridas de 10 quilômetros e 3) que tivesse familiaridade com corridas em esteiras.

O protocolo do estudo foi dividido em etapas, a primeira consistia em uma entrevista, onde o voluntário era questionado sobre a frequência com que corria, o seu nível de experiência, tipo de calçado utilizado, entre outras questões relevantes ao desempenho do atleta. A segunda etapa era a de medições preliminares, onde eram medidos altura, massa e tamanho do calçado. Todos os participantes utilizaram calçados neutros escolhidos pelo laboratório (Nike Dual Fusion X). Como ilustra a figura 1, na terceira etapa eram posicionados os marcadores articulares. O estudo utilizou 48 marcadores (Tabela 1). Na quarta etapa foi feita a calibragem dos marcadores de acordo com as medidas de cada atleta. Na quinta etapa os marcadores foram removidos, exceto pelos considerados anatômicos e técnicos. Finalmente, na sexta etapa as balanças da esteira foram zeradas, e foram feitas baterias de corrida a velocidades de 2,5 m/s, 3,5 m/s e 4,5 m/s. Cada atleta correndo em 3 velocidades, possibilitou a compilação de 3 arquivos contendo dados de movimentação por atleta, multiplicando pelo número de atletas (28) temos no total 84 arquivos contendo dados referentes às movimentações dos marcadores articulares.

Figura 1 - Marcadores utilizados na coleta de dados de movimentos cinemáticos

Fonte: FUKUCHI et al, 2015

Tabela 1 - Marcadores articulares utilizados na coleta de dados.

#	Rótulo	Tipo	Nomes das colunas
1	R.ASIS	T/A	Espinha ilíaca anterior superior direita
2	L.ASIS	T/A	Espinha ilíaca anterior superior esquerda
3	R.PSIS	T/A	Espinha ilíaca posterior direita
4	L.PSIS	T/A	Espinha ilíaca posterior esquerda
5	R.Iliac.Crest	T	Crista ilíaca direita
6	L.Iliac.Crest	T	Crista ilíaca esquerda
7	R.Thigh.Top.Lateral	T	Marcador lateral superior da coxa direita
8	R.Thigh.Bottom.Lateral	T	Marcador lateral inferior da coxa direita
9	R.Thigh.Top.Medial	T	Marcador medial superior da coxa direita
10	R.Thigh.Bottom.Medial	T	Marcador medial inferior da coxa direita
11	R.Shank.Top.Lateral	T	Marcador lateral superior da perna direita
12	R.Shank.Bottom.Lateral	T	Marcador lateral inferior da perna direita
13	R.Shank.Top.Medial	T	Marcador medial superior da perna direita
14	R.Shank.Bottom.Medial	T	Marcador medial inferior da perna direita
15	R.Heel.Top	T/A	Calcanhar direito
16	R.Heel.Bottom	T/A	Calcanhar direito inferior
17	R.Heel.Lateral	T	Lateral do calcanhar direito
18	L.Thigh.Top.Lateral	T	Marcador lateral superior da coxa esquerda
19	L.Thigh.Bottom.Lateral	T	Marcador lateral inferior da coxa esquerda

20	L.Thigh.Top.Medial	T	Marcador medial superior da coxa esquerda
21	L.Thigh.Bottom.Medial	T	Marcador medial inferior da coxa esquerda
22	L.Shank.Top.Lateral	T	Marcador lateral superior da haste esquerda
23	L.Shank.Bottom.Lateral	T	Marcador lateral inferior esquerdo
24	L.Shank.Top.Medial	T	Marcador medial superior da perna esquerda
25	L.Shank.Bottom.Medial	T	Marcador medial inferior da perna esquerda
26	L.Heel.Top	T/A	Calcanhar esquerdo superior
27	L.Heel.Bottom	T/A	Calcanhar esquerdo inferior
28	L.Heel.Lateral	T	Lateral do calcanhar esquerdo
29	R.GTR	A	Trocanter maior direito
30	R.Knee	A	Joelho direito
31	R.Knee.Medial	A	Joelho direito medial
32	R.HF	A	Cabeça direita da fíbula
33	R.TT	A	Tuberosidade tibial direita
34	R.Ankle	A	Tornozelo direito
35	R.Ankle.Medial	A	Tornozelo direito medial
36	R.MT1	A	1º metatarso direito
37	R.MT5	A	5º metatarso direito
38	R.MT2	A	2º metatarso direito
39	L.GTR	A	Trocanter maior esquerdo
40	L.Knee	A	Joelho esquerdo
41	L.Knee.Medial	A	Joelho esquerdo medial
42	L.HF	A	Cabeça esquerda da fíbula
43	L.TT	A	Tuberosidade tibial esquerda
44	L.Ankle	A	Tornozelo esquerdo
45	L.Ankle.Medial	A	Tornozelo esquerdo medial
46	L.MT1	A	1º metatarso esquerdo
47	L.MT5	A	5º metatarso esquerdo
48	L.MT2	A	2º metatarso esquerdo

Fonte: FUKUCHI et al, 2015

As demais etapas incluíam ajustes específicos no equipamento de medição para que fosse possível chegar ao resultado esperado com a pesquisa, maiores detalhes sobre essas etapas são dispensáveis a este estudo.

2.2.1 O Sistema

O método de coleta dos dados obtidos nesse estudo foi realizado a partir da captura de movimento 3D com 12 câmeras de resolução 4Mb e o software Cortex 6.0. Esses dados foram coletados em uma esteira ergométrica, enquanto os voluntários caminhavam sobre o aparelho a diferentes velocidades (2,5 m/s, 3,5 m/s e 4,5 m/s) usando calçados neutros padrão.

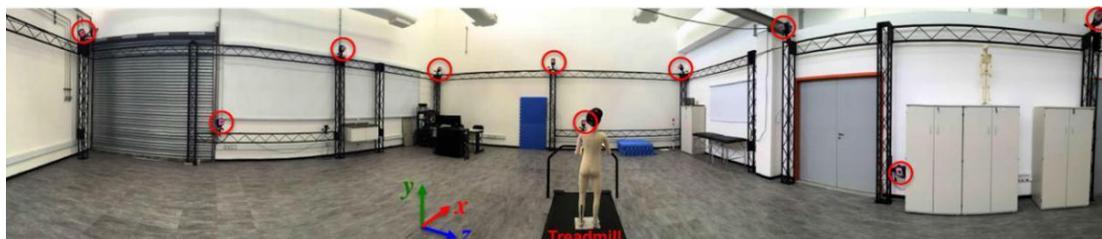
A figura 2 representa o posicionamento de doze câmeras que são distribuídas ao redor do laboratório, de modo a apontar para o volume de captura de movimento da esteira instrumentada. As câmeras foram montadas em uma estrutura de treliça metálica com 11,5 m de comprimento, 9,3 m de largura e 2,8 m de altura. Essa estrutura permitiu o posicionamento de algumas câmeras com elevações variadas.

O software Cortex 6.0 foi utilizado para: (1) calibrar o volume de captura de movimento; (2) capturar e identificar os marcadores refletivos; e (3) preparar os dados e exportá-los para o formato de arquivo c3d.

O sistema de coordenadas de laboratório utilizado para o estudo foi o mesmo que o proposto pela Sociedade Internacional de Biomecânica (Wu & Cavanagh, 1995) e, como mostrado na Figura 2, continha o seguinte:

- Eixo X na direção da progressão da marcha e positivo apontando para a frente.
- Eixo Y na direção vertical e positivo apontando para cima.
- Eixo Z na direção medial-lateral e apontando positivo para a direita.

Figura 2: Visão ampliada do Laboratório de Biomecânica e Controle Motor, mostrando 10 das 12 câmeras do sistema de captura de movimento (marcadas com círculos vermelhos), a esteira instrumentada e o sistema de coordenadas do laboratório.



Fonte: FUKUCHI et al, 2015

Como resultados do estudo, os autores conseguiram gerar 84 arquivos sobre movimentações dos marcadores (cinemática), bem como 84 arquivos contendo dados sobre as forças das passadas (cinética) dos atletas que se voluntariaram. Todos os arquivos gerados são nomeados iniciando com a sigla RBDS que significa “*Running Biomechanics Dataset*”, base de dados sobre biomecânica de corrida, em tradução livre.

Além dos arquivos de dados, foi gerado um arquivo em forma de tabela que contém informações obtidas na entrevista com os voluntários que participaram da pesquisa. Os dados contidos na tabela incluem nome, idade, altura, peso, gênero, membro inferior dominante (perna esquerda ou direita), tempo de experiência em corridas, tipos de solo com que este tem

costume de correr, tamanho e tipo do calçado, histórico de lesão, entre outras informações relevantes ao perfil do atleta.

2.2.2 Análise do Perfil dos Indivíduos

Analisando os indivíduos integrantes do estudo que deu base a este trabalho, é possível perceber uma alta similaridade entre os indivíduos. De todos os 28 corredores convidados para o estudo, apenas um deles é mulher. A diferença de altura do indivíduo mais alto e o mais baixo é de 22cm, já a diferença de peso entre a pessoa mais pesada e a mais leve é de 26kg, mas quando calculamos o IMC (índice de massa corporal) dos voluntários, temos que apenas um deles está fora da faixa de peso ideal para a sua altura.

Figura 3 – Histograma da Altura dos Atletas.

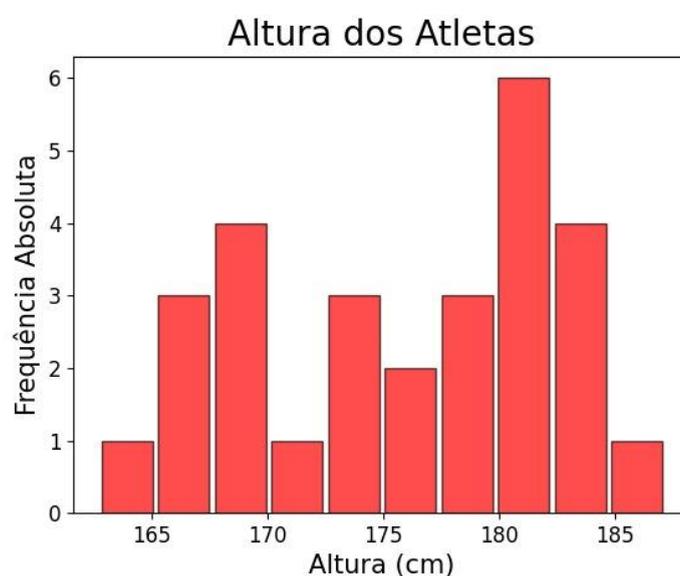


Figura 4 – Histograma da Experiência dos Atletas.

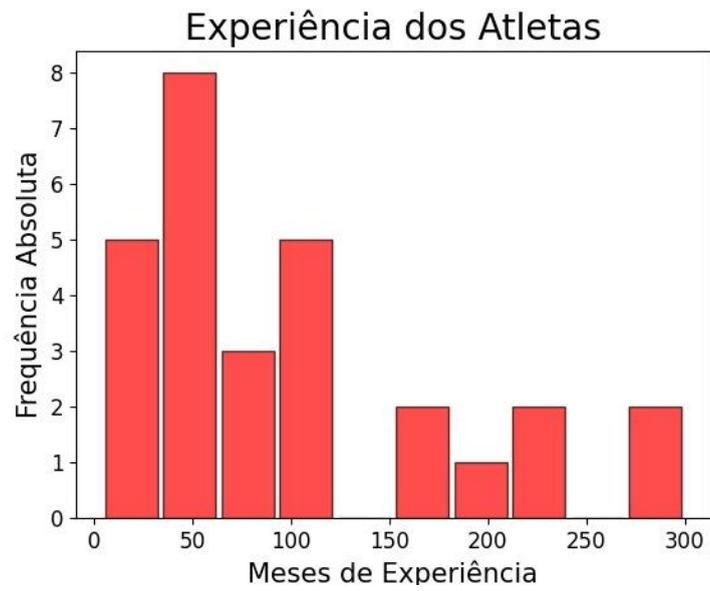


Figura 5 – Histograma da Idade dos Atletas.

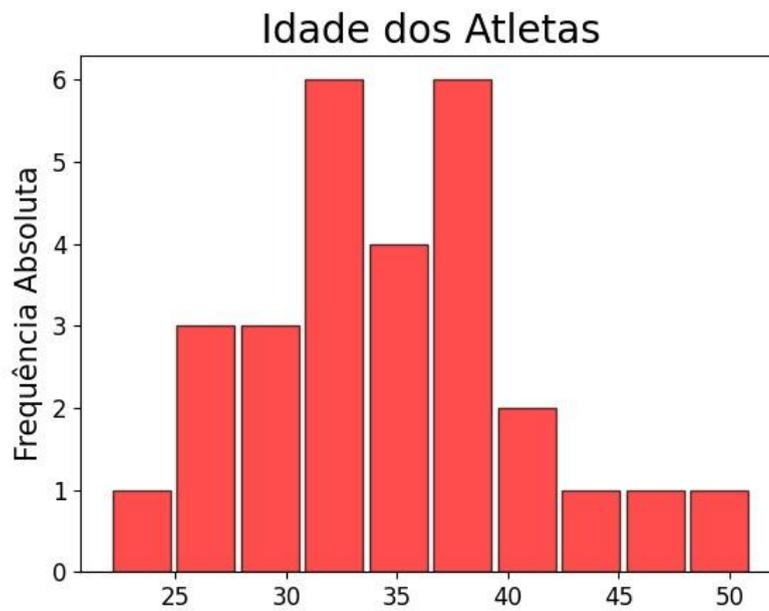


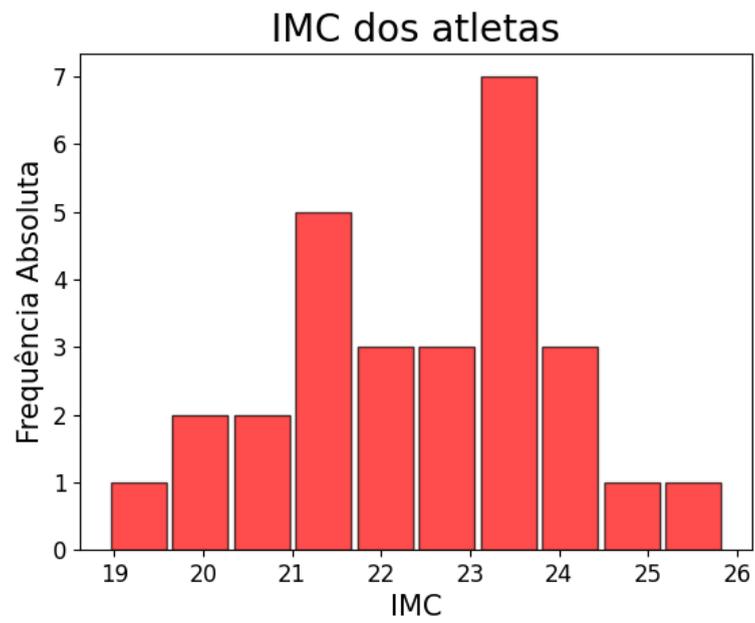
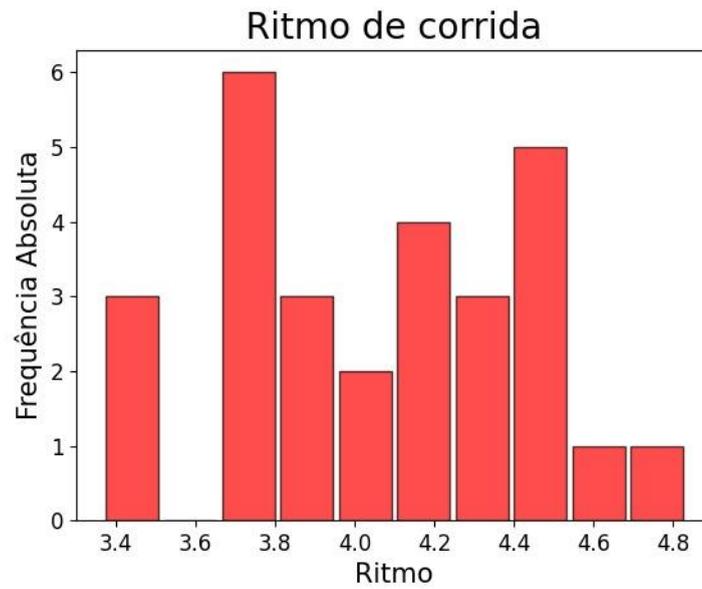
Figura 6 – Histograma de IMC dos Atletas.**Figura 7 – Histograma do ritmo de corrida (min/km) dos atletas.**

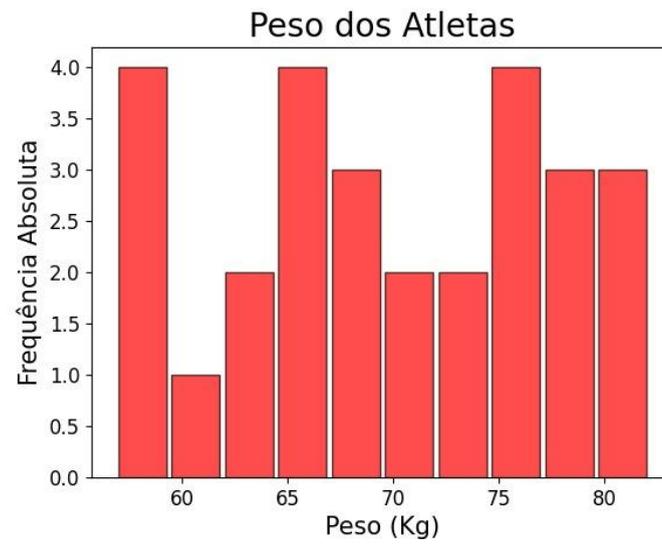
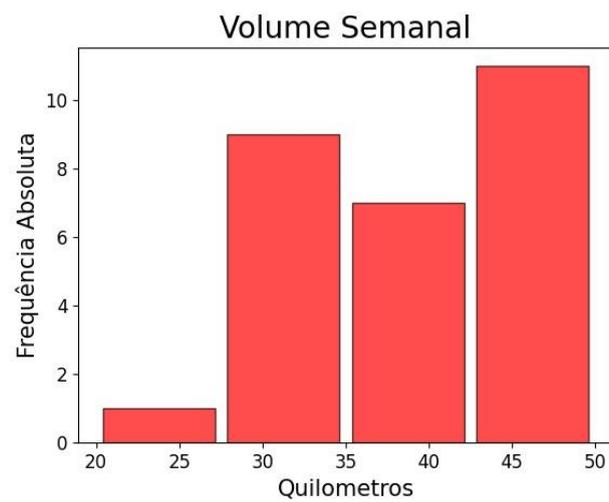
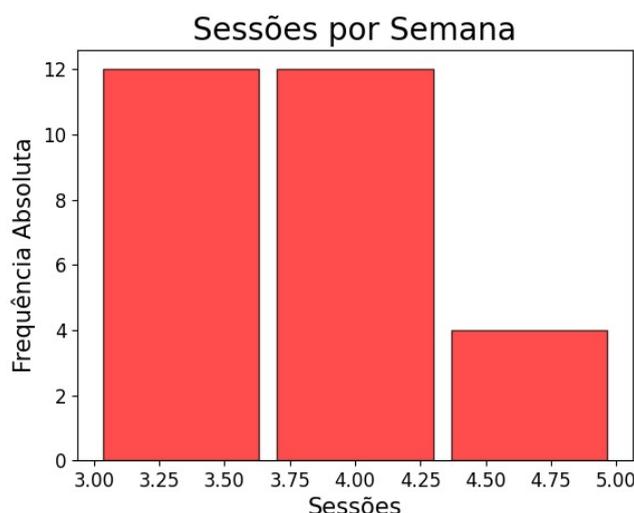
Figura 8 – Histograma de Peso dos Atletas.**Figura 9** – Histograma de volume de Volume Semanal de corrida dos atletas.

Figura 10 – Histograma de Sessões por Semana.

Tem-se, nas figuras 3 a 10, o histograma de alguns dos dados analisados dos corredores. Esses gráficos explicitam a homogeneidade dos atletas selecionados para a pesquisa. Os histogramas, em sua maioria, tem suas barras niveladas, o que comprova que os dados coletados não tem amplitude, tampouco diversidade de amostra, o que certamente influenciou o resultado indefinido no agrupamento dos dados.

2.3 Trabalho sobre Cinética de Corrida

Como dito anteriormente, este trabalho vem para dar sequência ao trabalho feito por Mônica Oliveira (2018), onde a autora utilizou dados do mesmo estudo sobre coleta de dados de corredores, porém analisou dados sobre a cinética da corrida, isto é, a força resultante da passada do atleta.

No estudo, a autora se utiliza da Transformada de Fourier, se aproveitando da propriedade cíclica dos dados recebidos, ou seja, do movimento da passada, extraindo as características chave desse sinal. Mônica aplicou a transformada de Fourier nos arquivos de entrada e aplicou algoritmos de aprendizagem de máquina como o *K Nearest Neighbor* (KNN), frequentemente aplicado em reconhecimento de padrões, e o *Support Vector machine* (SVM), capaz de separar os dados em classes distintas utilizando hiperplano de separação, sendo ambos algoritmos para classificação de dados. Foi obtido êxito na classificação dos episódios de lesão, com acurácia de até 77,7%.

Finalmente, a autora concluiu, através de sua pesquisa, que o uso da Transformada de Fourier desempenhou papel fundamental para escolher e para visualizar padrões. Além disso, o SVM encontrou um valor de classificação ótimo. Em contrapartida, pôde perceber que o kNN possui pouca relevância para o estudo, uma vez que não revelou uma alta taxa de acerto máxima, obtendo apenas 55,5%. Enfim, sua pesquisa não contrasta com a

literatura utilizada, pois ela já apontava o kernel *Radial Basis Function* (Função de Base Radial) como opção promissora em diversas áreas de pesquisa.

2.4 Transformada de Fourier

A descoberta da transformada de Fourier é atribuída ao físico e matemático francês Jean-Baptiste Joseph Fourier (1768-1830), que dedicou seus estudos à decomposição de funções periódicas em séries trigonométricas. As séries, que hoje são chamadas de séries de Fourier, são usadas para representar funções periódicas definidas na reta em termos de senos e cossenos. A ação de efetuar a decomposição de uma função periódica é chamada de transformada de Fourier.

A transformada de Fourier é amplamente utilizada em técnicas de processamento de sinais devido a sua aplicabilidade na extração de características chave do sinal. Ela tem sido amplamente utilizada em diversas áreas, por exemplo, na solução dos problemas de processamento digital de imagens (UZUN et al, 2005), como ilustra a figura 11. Na prática, quando se deseja trabalhar uma imagem no domínio da frequência, por exemplo, aplica-se a transformada de Fourier da referida imagem e a multiplica pela função de transferência de um filtro (convenientemente de acordo com a aplicação). No entanto, muitas vezes, é mais simples "zerar " os coeficientes das componentes de frequência que quer filtrar e tomar, em seguida, em ambos os casos, a transformada inversa obtendo, assim, a imagem filtrada (processada) (NETO, 1999).

Figura 11 – Exemplo da aplicação da transformada de Fourier em processamento de imagens.

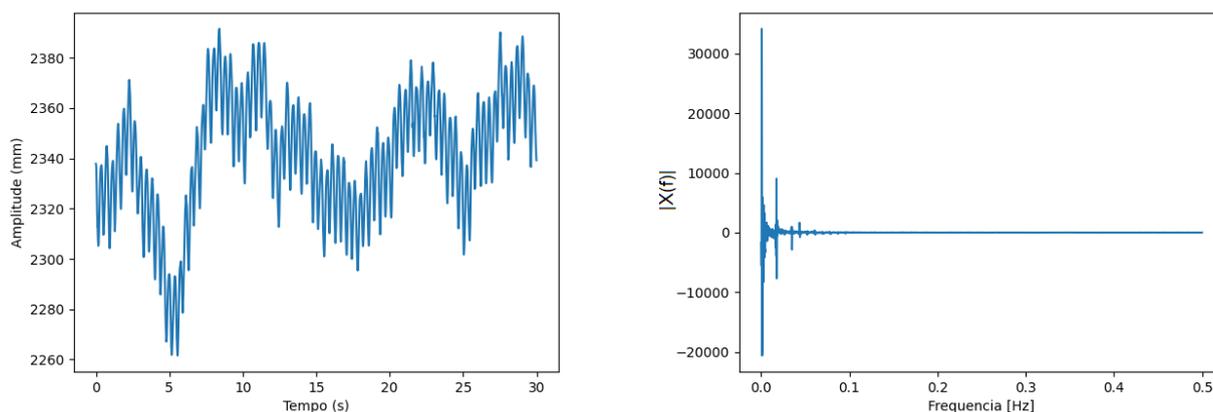


Fonte: UFSC, 2011

A transformada de Fourier é um processo matemático extensamente utilizado hoje em dia em espectrometria de Ressonância Magnética Nuclear, de Infravermelho, etc., na forma do engenhoso processo de cálculo digital simplificado, desenvolvido por vários autores (COOLEY, JW et al), denominado *Fast Fourier Transform* (FFT). A simplificação introduzida por esse algoritmo reduz as operações necessárias de N^2 multiplicações complexas para $N \log_2 N$ multiplicações (N é o número de pontos da função a ser transformada).

Neste trabalho, como foi manipulada uma grande quantidade de dados, ao tentar expressar esses dados de forma gráfica, não foram obtidos resultados que pudessem gerar implicações práticas. Assim sendo, optou-se por aplicar a análise de Fourier, já que, o movimento da passada pode ser expresso como um sinal. Portanto, se aplicada a Transformada nesses dados, serão encontrados, como resultados, a extração de características desse sinal. Permitindo, como mostra a figura 12, gerar gráficos que podem ser interpretados e que conclusões possam ser tiradas destes.

Figura 12 – Demonstração da aplicação da transformada de Fourier sobre sinal resultante de um dos marcadores articulares.



2.5 Agrupamento de Dados

Desde que existe o advento da computação, busca-se entender dados. Com o avanço das tecnologias e sistemas, o volume de dados disponíveis, de uma maneira geral, tem crescido exponencialmente, segundo o IDC (2014) chegando a dobrar a cada 2 anos.

Possuir grandes massas de dados em nada ajuda, se não os arranjá-los de forma a entendê-los e investigar padrões, por isso é necessário, de alguma forma, organizá-los. Neste contexto, surge como forte opção a clusterização, ou agrupamento dos dados, conceito que consiste em reunir os dados semelhantes entre si, formando grupos, ou *clusters*, e revelando padrões de similaridade.

Agrupamento ou *Clustering* é uma técnica de mineração de dados não supervisionada, isto é, agrupa as entradas de dados sem ter o conhecimento prévio do “*target*”, valor de saída que o algoritmo busca descobrir. Algoritmos de *clustering* tem como objetivo descobrir padrões em objetos de dados, gerando uma partição da base de dados que contenha grupos (*clusters*) de objetos com características comuns entre si (FONTANA et al, 2009).

Segundo Jain et al (1999), o agrupamento é de grande utilidade em várias aplicações que buscam análise de padrões. Como exemplos, na mineração de dados, na recuperação de documentos, na segmentação de imagens e na classificação de padrões.

2.5.1 K-Means

É um algoritmo de agrupamento cujo objetivo do algoritmo é dividir a entrada em K grupos distintos, minimizando a distância intra-cluster e maximizando a distância inter-cluster. Em outras palavras, ele particiona os dados em K grupos distintos de acordo com a distância entre os membros do mesmo grupo.

O K-Means funciona da seguinte forma:

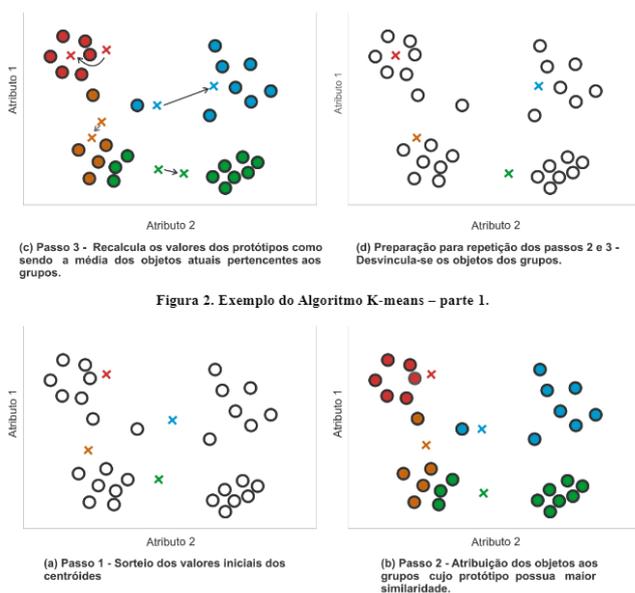
- 1) Inicialmente são selecionados os centróides $(x_1, x_2 \dots x_k)$
- 2) Associa os elementos ao centróide mais próximo através de alguma equação de cálculo de distância entre dois pontos, como a distância euclidiana:

$$d = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$$

Na formula da distancia euclidiana, mostrada acima, “d” representa a distância, x1 e x2 representam os valores dos dois pontos no eixo X, já y1 e y2 representam os valores dos pontos no eixo Y.

- 3) Recalcular posição dos centróides como sendo a média dos atuais valores do grupo.
- 4) Repetir os passos 2 e 3 até que não haja mais mudança na posição dos centróides.

Figura 13 – Demonstração do funcionamento do algoritmo K-Means.



Fonte: FONSECA; BELTRAME, 2011

2.5.2 Métricas de Validação de Cluster

O ideal, quando se trata de agrupamento de dados, é que os dados pertencentes a um mesmo grupo estejam o mais próximos possível, enquanto os dados que pertencem a grupos diferentes devem estar o mais distante possível. Assim pode-se afirmar que os grupos são heterogêneos e os dados foram divididos de forma eficaz. Existem várias métricas que auxiliam na avaliação de quão bem os dados estão agrupados, algumas delas serão apresentadas nesta seção

2.5.2.1 Coeficiente de Silhueta

Proposta pela primeira vez por Peter Rousseeuw (1987), a métrica de coeficiente de silhueta tem uma abordagem espacial para definir seu valor de saída. O valor de silhueta é a medida do quanto um objeto é similar aos demais do mesmo grupo (coesão), e quanto ele é distante dos outros grupos (separação). A saída da função é um valor entre 1 e -1 onde valores mais próximos de 1 indicam um bom agrupamento, enquanto valores próximos de -1 indicam um agrupamento ruim.

$$C_S = \frac{b_i - a_i}{\max(a_i, b_i)}$$

O coeficiente de silhueta é calculado através da equação acima, onde a_i representa a média da distância de uma entrada aleatória i e as demais entradas do grupo a , da mesma

forma, b_i representa a média das distâncias entre uma entrada aleatória i e os demais elementos do grupo b . A função $\max()$ retorna a maior das entradas a_i e b_i .

2.5.2.2 Davies-Bouldin

Proposta por David L. Davies and Donald W. Bouldin (1979) é uma função da razão da soma da dispersão dentro dos agrupamentos pela dispersão entre os agrupamentos (PAKHIRA et al, 2004). Quanto menor é o valor de saída desta função, melhor é o resultado do agrupamento.

$$\frac{1}{c} \sum_{k=1}^c \max_{c \neq l} \left\{ \frac{S_c(Q_k) + S_c(Q_l)}{d_{ce}(Q_k, Q_l)} \right\}$$

O índice Davies-Bouldin é dado pela equação acima, onde c é o número de agrupamentos, $S_c(Q_k)$ representa a distância intra-agrupamento baseado na distância para o centróide, $d_{ce}(Q_k, Q_l)$ representa a distância entre os agrupamentos Q_k e Q_l , também baseado na distância entre os centróides.

2.5.2.3 Calinski-Harabasz

É uma métrica que avalia tanto a distância intra-cluster, distância entre os elementos de um mesmo cluster, quanto a distância inter-cluster, distância entre diferentes clusters. Em um estudo de 2001, esta métrica foi classificada como uma das melhores para validação de clusters (LIPOR e BALZANO, 2020).

$$CH(K) = \frac{tr(B_K)/(K - 1)}{tr(W_K)/(N - K)}$$

O índice Calinski-Harabasz é calculado segundo a fórmula acima, onde B_K e W_K são, respectivamente, as matrizes de dispersão entre grupos e intragrupo, das quais são calculados os traços (operador $tr()$), K representa o número de *clusters* do K-means e N o número de objetos da base de dados.

2.6 Classificação de Dados

Ao contrário da clusterização, a classificação de dados utiliza aprendizagem supervisionada, isto é, o algoritmo é informado do resultado esperado para a entrada, para que assim ele possa treinar e fazer testes com as novas entradas e fazer inferências mais assertivas sobre as novas entradas que forem submetidas.

2.6.1 K Nearest Neighbors

O *K Nearest Neighbors* (KNN), ou K Vizinhos mais próximos, em tradução livre) é um algoritmo simples e ainda assim poderoso na tarefa de classificação de dados, por isso é um dos algoritmos mais utilizados em aprendizagem de máquina. Foi implementado como uma melhoria do algoritmo *Nearest Neighbor* (COVER; HART, 1967). O algoritmo de Cover e Hart consistia simplesmente em encontrar o vizinho mais próximo de uma entrada, isto é, qual entrada se assemelha mais à nova entrada, que procura-se classificar.

O KNN, por sua vez, utiliza um número K de vizinhos próximos para decidir a que classe a nova entrada pertence, aumentando assim a precisão do algoritmo e as taxas de sucesso.

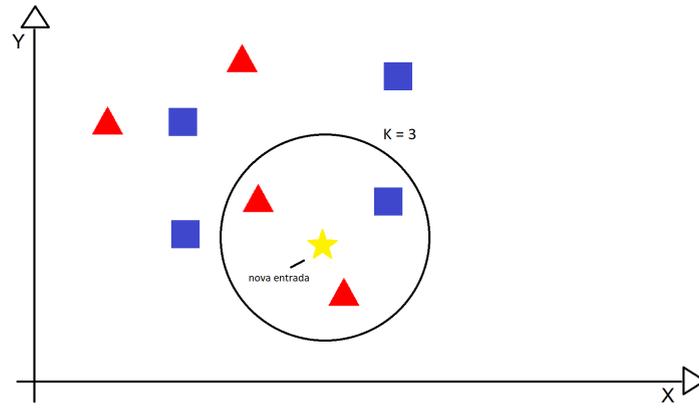
O algoritmo é amplamente utilizado nos mais diversos campos de aplicação, como agronomia, medicina, economia, entre muitas outras. Bijalwan et al (2014), por exemplo, demonstraram a eficácia do algoritmo frente a outros similares em seu trabalho de mineração de textos e documentos.

Mustafa et al. (2012), por sua vez, provaram que o KNN se mostrou superior que o ANN (*Artificial Neural Networks*, Redes Neurais Artificiais) em sua aplicação de equilíbrio mental.

O KNN Funciona da seguinte forma:

1. Recebe um dado não classificado
2. Utiliza algum método para medir a distância entre este dado e os dados já classificados. Alguns métodos para medir a distância entre as entradas são: distância euclidiana, de Manhattan, Minkowsky, entre outras.
3. Descobre os K dados mais próximos da nova entrada.
4. Verifica como estão classificados os K vizinhos encontrados
5. Classifica o novo dado com a classe que mais aparecer no passo anterior.

Na Figura 14 tem-se duas classes (triângulos e quadrados) e uma nova entrada é apresentada (estrela), inicialmente, sem classificação. O algoritmo KNN é representado pelo círculo, que delimita e seleciona os K vizinhos mais próximos, neste caso K é igual a 3. No exemplo ilustrado pela figura, a classificação do novo dado (estrela) será “triângulo” pois, dos 3 vizinhos mais próximos, 2 tem essa classificação.

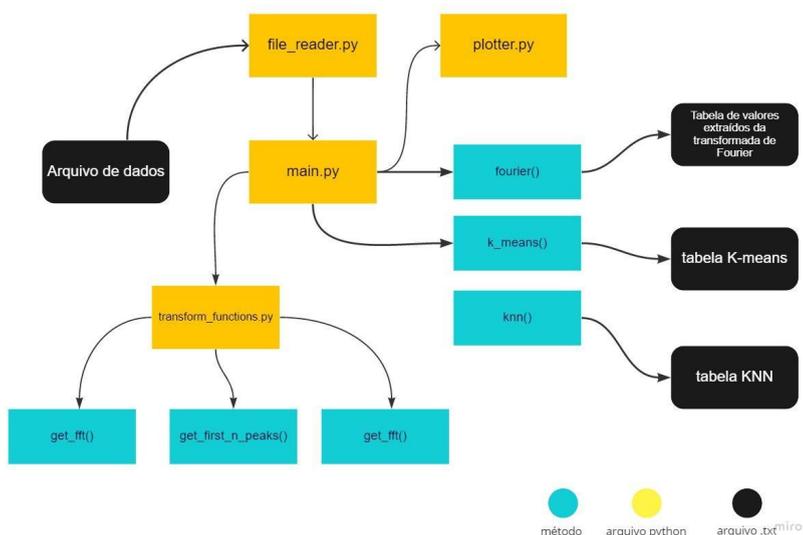
Figura 14 – Ilustração do funcionamento do algoritmo KNN.

3 Metodologia

A abordagem utilizada para o desenvolvimento do algoritmo utilizado neste trabalho pode ser ilustrada pelo diagrama da Figura 15. Nesse diagrama, são exibidas três etapas: a leitura dos arquivos de dados, o processamento dos dados por meio da transformada de Fourier e a aplicação de algoritmos de aprendizagem de máquina (K-Means e KNN).

O pré-processamento dos dados consiste em organizar a estrutura das características obtidas utilizando a Transformada de Fourier. Os arquivos de dados são lidos pelo código “file_reader.py” e envia as informações extraídas ao código “main.py”, que se comunica com o arquivo “transform_functions.py” para executar as funções da transformada de Fourier. O arquivo principal também faz chamadas ao arquivo “plotter.py”, responsável por gerar gráficos. Por fim, o arquivo “main.py” possui métodos que aplicam os algoritmos K-means e KNN e exportam as tabelas úteis na análise dos dados.

Figura 15 – Diagrama de fluxo do algoritmo.



3.1 Pré-processamento dos Dados

A base de dados descrita no capítulo anterior é composta por 84 arquivos em formato de texto. Cada um desses arquivos contém uma tabela de 4500 linhas e 97 colunas. Como ilustrado na Figura 16, a primeira coluna representa um instante de tempo, enquanto as 96 colunas seguintes representam a exata posição espacial dos marcadores articulares nesse instante de tempo.

Figura 16 – Exemplo de arquivo de dados com campos omitidos, devido ao tamanho da tabela. Foram exibidos os primeiros e últimos valores das linhas e colunas, para possibilitar a visualização dos dados.

Os valores de movimentação dos marcadores são atualizados a cada 7 milissegundos, aproximadamente.

Time	R.ASISX	R.ASISY	R.ASISZ	L.ASISX	L.ASISY	L.ASISZ	R.PSISX	R.PSISY	R.PSISZ	...	R.MTSX	R.MTSY	R.MTSZ	L.MT1X	L.MT1Y	L.MT1Z	L.MT5X	L.MT5Y	L.MT5Z
0	2420.63	1027.69	1154.7	2382.92	1036.06	948.608	2242.8	1059.59	1114.45	...	1637.27	509.596	1157.78	2717.16	119.751	1065.83	2686.71	829.453	965.566
0.007	2418.94	1021.43	1154.41	2380.65	1031.09	947.348	2241.08	1052.69	1115.67	...	1653.06	518.527	1160.69	2696.73	101.722	1066.29	2663.28	680.879	967.37
0.013	2417.13	1014.76	1154.34	2378.4	1025.77	946.213	2239.22	1045.82	1117.12	...	1670.41	524.643	1164.05	2676.39	818.196	1069.01	2639.48	532.085	969.248
0.02	2414.99	1007.88	1154.34	2376.13	1020.2	945.564	2237.09	1039.14	1118.31	...	1688.98	528.531	1167.92	2654.42	651.207	1072.31	2616.47	468.804	969.107
0.027	2412.63	1000.76	1154.19	2373.6	1014.31	945.543	2234.4	1032.61	1119.3	...	1708.61	529.972	1172.44	2627.85	554.323	1078.44	2592.22	469.196	973.332
0.033	2409.54	993.835	1154.08	2369.84	1008.01	946.856	2230.74	1026.33	1120.28	...	1729.07	529.024	1177.14	2600.56	56.067	1081.41	2562.46	486.225	976.241
0.04	2405.53	987.697	1153.89	2364.65	1001.77	949.351	2226.36	1020.34	1121.18	...	1750.01	525.863	1182.02	2571.32	543.623	1083.18	2534.09	48.18	976.535
0.047	2400.78	981.941	1153.59	2358.99	995.339	950.462	2221.73	1014.3	1121.94	...	1771.27	520.445	1187.15	2542.97	536.109	1081.26	2506.23	484.439	978.091
0.053	2395.89	976.689	1153.67	2354.11	990.259	949.562	2217.33	1008.68	1123.25	...	1793.77	511.667	1188.91	2513.59	528.624	1080.24	2476.8	479.751	977.389
0.06	2391.22	972.023	1154.09	2349.31	985.398	948.32	2213.41	1003.83	1124.59	...	1815.84	501.412	1193.15	2483.98	520.883	1080.25	2447.21	478.719	977.718
0.067	2387.25	968.055	1154.41	2345.47	981.495	947.143	2209.86	1000.73	1125.23	...	1838.22	488.956	1196.89	2453.77	513.849	1079.85	2417.09	475.735	977.569
0.073	2383.9	965.057	1154.65	2342.29	978.678	946.276	2206.77	998.781	1125.3	...	1860.83	474.488	1200.13	2423.31	506.871	1079.86	2386.57	474.394	977.777
0.08	2381.11	962.835	1154.87	2339.67	976.251	945.621	2204.29	997.852	1124.92	...	1883.56	458.34	1202.73	2392.57	501.185	1079.94	2355.84	476.649	978.152
0.087	2378.77	961.68	1154.62	2337.66	974.28	944.859	2202.38	998.129	1124.08	...	1906.46	440.768	1204.54	2361.95	495.968	1080.24	2325.08	474.961	978.847
0.093	2377.03	961.831	1154.32	2336.45	972.654	944.104	2200.74	999.228	1122.7	...	1929.64	422.111	1205.51	2331.3	494.743	1079.87	2294.22	473.381	979.297
0.1	2375.8	963.626	1153.87	2336.1	971.904	943.566	2199.41	1001.08	1121.21	...	1952.9	402.735	1205.84	2300.82	493.905	1079.5	2263.42	473.035	979.873
0.107	2375.17	967.258	1153.6	2336.91	972.168	943.162	2198.58	1003.71	1119.7	...	1977.95	382.289	1205.94	2270.29	492.792	1079.5	2232.65	473.129	980.405
0.113	2374.97	971.684	1153.1	2338.62	973.655	943.01	2198.26	1007.08	1117.98	...	2002.21	362.368	1205.06	2239.82	493.424	1079.46	2202.05	473.943	981.002
0.12	2374.96	976.734	1152.92	2341.12	976.175	942.974	2198.59	1010.9	1116.19	...	2025.8	341.803	1203.36	2209.62	496.262	1079.29	2171.25	475.193	981.104
0.127	2375.35	982.249	1152.85	2344.14	979.509	942.878	2199.27	1015.11	1114.34	...	2051.63	321.312	1201.18	2179.36	498.531	1078.93	2140.57	476.858	981.36
0.133	2376.24	988.38	1152.85	2347.42	982.835	942.76	2200.38	1019.88	1112.46	...	2078.5	301.216	1198.79	2149.42	50.114	1078.43	2109.94	479.915	981.497
0.14	2377.51	995.155	1152.95	2351.39	986.183	942.665	2201.42	1025.29	1110.51	...	2106.75	281.242	1195.97	2119.81	503.082	1077.66	2079.63	484.679	981.274
0.147	2379.19	1002.45	1152.92	2355.81	990.201	942.916	2203.56	1030.89	1108.38	...	2134.34	262.653	1194.54	2090.12	502.135	1077.15	2049.72	489.902	980.794
0.153	2381.22	1010.39	1153.22	2360.98	994.64	943.375	2206.22	1036.98	1106.11	...	2163.71	243.758	1191.86	2060.8	498.432	1077.14	2020.71	493.985	978.964
0.16	2383.51	1018.33	1153.2	2366.47	999.29	943.932	2209.21	1043.28	1103.98	...	2193.86	225.268	1188.45	2032.35	494.104	1077.33	1992.74	501.289	978.789
0.167	2385.87	1026.24	1153.27	2372.25	1003.93	944.584	2212.55	1049.98	1101.92	...	2225.02	207.831	1185.27	2004.98	48.883	1077.38	1965.96	509.754	977.535
0.173	2388.39	1033.79	1153.4	2377.78	1008.44	946.253	2216.28	1057.02	1099.85	...	2256.9	190.987	1182.29	1978.94	490.118	1077.6	1939.97	526.063	975.772
0.18	2390.94	1040.79	1153.47	2384.09	1013.07	947.475	2220.39	1064.12	1097.81	...	2290.01	175.411	1179.23	1955.4	502.554	1076.26	1915.03	553.024	973.834
...
29.947	2493.43	1046.82	1158.92	2464.51	1053.43	953.264	2319.97	1083.98	1107.26	...	1657.67	542.812	1158.14	2822.69	158.528	1052.18	2801.59	114.377	953.192
29.953	2492.74	1041.05	1158.69	2462.08	1048.9	952.382	2318.63	1077.24	1109.37	...	1668.97	557.602	1160.24	2808.85	145.744	1054.55	2787.3	101.663	955.671
29.96	2491.83	1035.11	1158.17	2459.66	1044.07	951.183	2317.38	1070.45	1111.32	...	1682.07	570.128	1162.83	2792.65	130.388	1057.23	2770.14	886.067	957.951
29.967	2490.64	1028.91	1157.4	2457.24	1039.04	949.796	2316.06	1063.74	1113.12	...	1696.97	580.363	1165.91	2773.86	113.087	1059.71	2747.73	74.742	959.986
29.973	2489.25	1022.39	1156.53	2454.79	1033.78	948.309	2314.59	1057.06	1114.96	...	1713.21	588.18	1169.69	2753.88	937.896	1060.98	2723.52	586.897	962.572
29.98	2487.83	1015.61	1155.78	2452.41	1028.38	947.058	2312.78	1050.55	1116.79	...	1730.75	593.689	1173.99	2733.64	728.362	1063.37	2699.31	446.118	962.358
29.987	2486.21	1008.71	1155.27	2449.98	1022.54	946.398	2310.55	1044.34	1118.57	...	1749.38	596.784	1178.39	2708.57	576.671	1069.95	2678.73	448.868	961.788

Todos os testes apresentados neste capítulo foram executados em uma máquina com processador Intel Core i5 x64 2.71 GHz, 8 GB de memória primária e 500 GB de memória secundária executando sobre o sistema operacional Windows 10. A linguagem utilizada foi Python versão 3.7.

3.1.1 Manipulação da Base de Dados

Lendo o arquivo de dados, e armazenando os valores de marcadores articulares isoladamente, foi possível gerar gráficos (figuras 17, 18 e 19) que permitiram visualizar o comportamento dos valores de movimentação ao longo do tempo. É possível perceber a diferença do formato da onda de um mesmo marcador articular quando são produzidos seus gráficos de dimensões X, Y e Z.

Os nomes dos arquivos da base de dados por padrão iniciam com a sigla RBDS (*Running Biomechanics Dataset*, base de dados sobre biomecânica de corrida, em tradução livre), seguido do número do indivíduo a quem pertence os dados, um indicativo da velocidade que foi impressa (runT25 para 2,5m/s, runT35 para 3,5m/s e runT45 para 4,5m/s), por fim uma palavra chave que indica se este é um arquivo de forças ou de marcadores articulares. Sendo assim, o arquivo denominado RBDS001runT25markers é o arquivo que guarda os dados de marcadores articulares do atleta de número 001 correndo a uma velocidade de 2,5m/s.

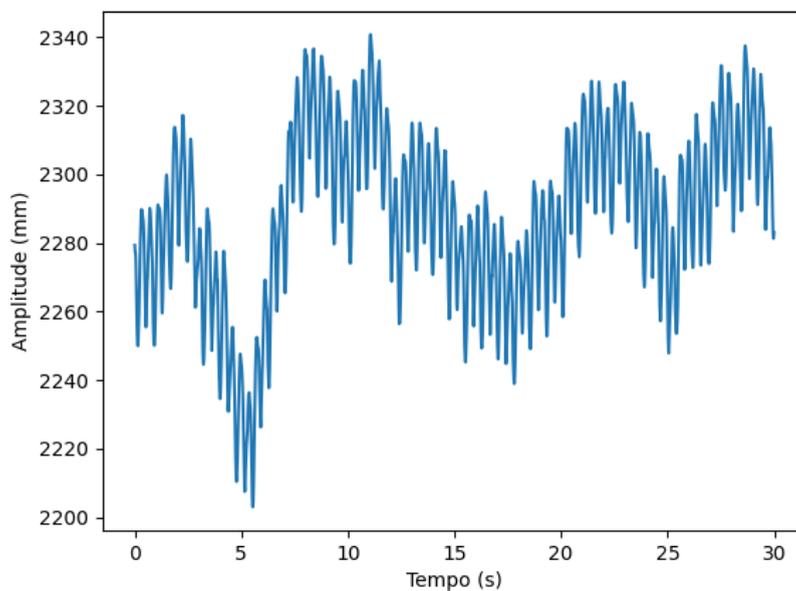
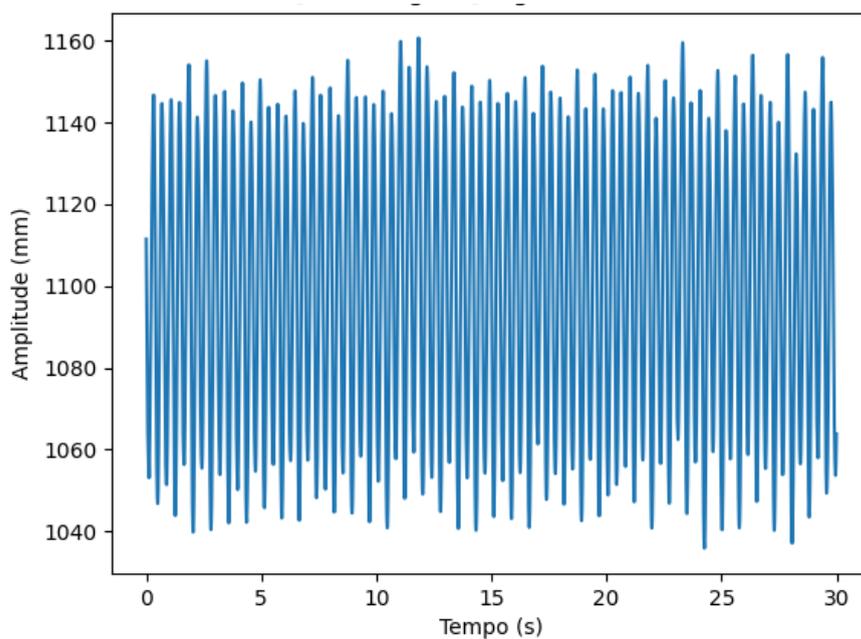
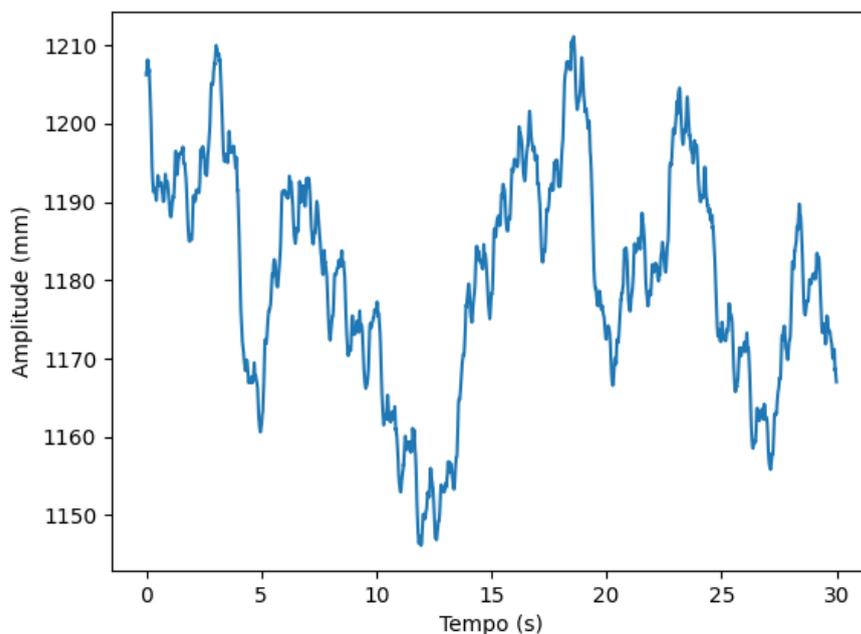
Figura 17 – Gráfico da crista ílfaca direita (dimensão X) do arquivo RBDS001runT25markers.**Figura 18** – Gráfico da crista ílfaca direita (dimensão Y) do arquivo RBDS001runT25markers.

Figura 19 – Gráfico da crista íliaca direita (dimensão Z) do arquivo RBDS001runT25markers.



3.1.2 Transformada de Fourier

Uma vez que os dados foram corretamente extraídos dos arquivos, percebe-se que esses dados formam um sinal contínuo em relação ao tempo. O sinal extraído dos dados precisa ser transformado a fim de analisar os componentes periódicos que melhor expressam esse sinal. Para isso, a solução escolhida foi aplicar a Transformada de Fourier nos dados.

A escolha da ferramenta para esta etapa se deve ao comportamento que sugere que os sinais possuem periodicidade nos padrões. A Transformada de Fourier, como anteriormente abordado, é utilizada para decompor o sinal em seus componentes periódicos, componentes estes que serão posteriormente selecionados e alimentarão o algoritmo de agrupamento e de classificação de dados.

As Figuras 20, 21 e 22 ilustram o efeito da Transformada de Fourier sobre os dados. São representados 3 marcadores articulares referentes à articulação denominada espinha íliaca esquerda, localizada no quadril, em suas 3 dimensões (X, Y e Z). Na parte esquerda das figuras estão os gráficos originais, referentes às movimentações dos marcadores e, à direita, a Transformada de Fourier do sinal original indicando as principais componentes identificadas.

Figura 20 – Gráfico da espinha ilíaca esquerda (dimensão X) do arquivo RBDS001runT25markers e o gráfico da sua transformada de Fourier.

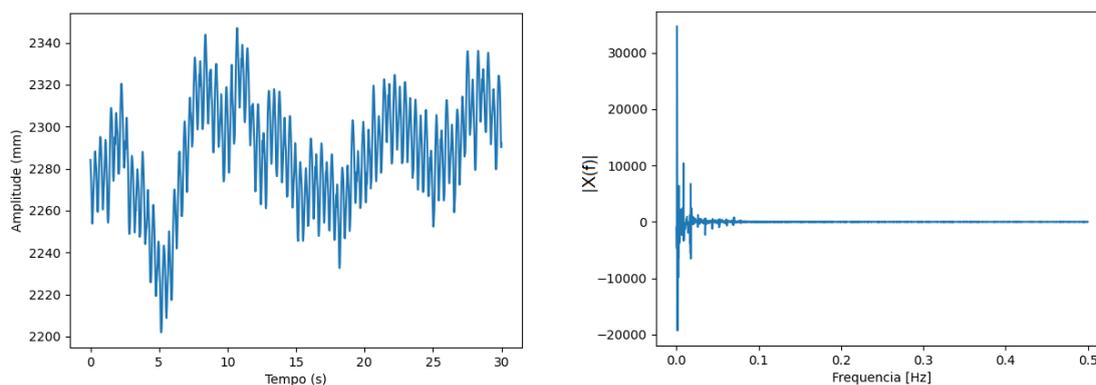


Figura 21 – Gráfico da espinha ilíaca esquerda (dimensão Y) do arquivo RBDS001runT25markers e o gráfico da sua transformada de Fourier.

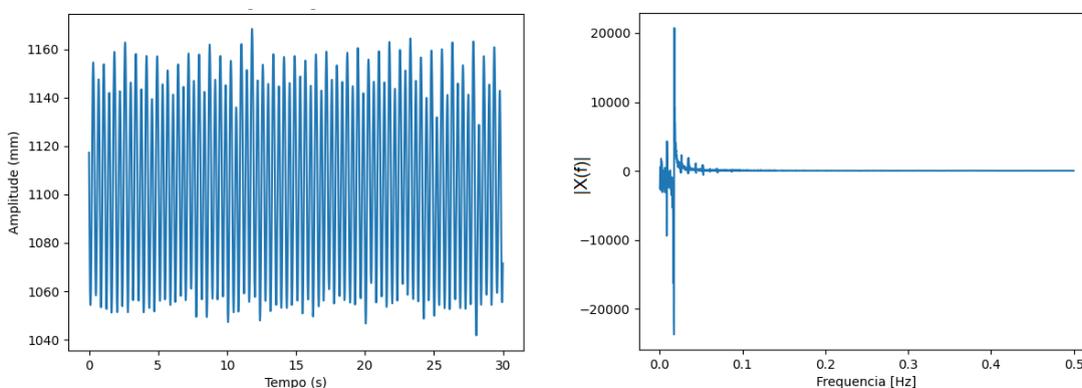
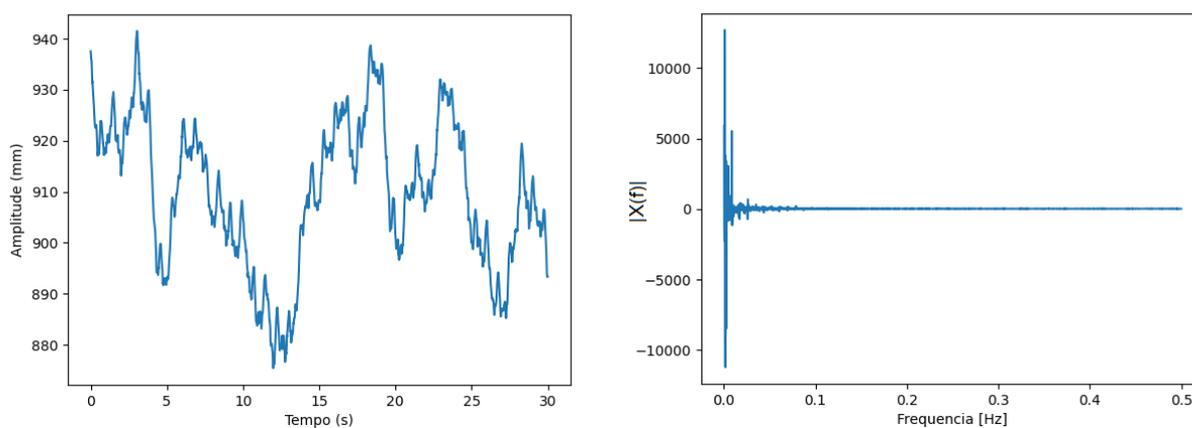


Figura 22 – Gráfico da espinha ilíaca esquerda (dimensão Z) do arquivo RBDS001runT25markers e o gráfico da sua transformada de Fourier.



Como dito anteriormente, os sinais extraídos sobre movimentação dos corredores apresentaram comportamento indicativo de periodicidade, por isso, a fim de obter seus

componentes periódicos, é possível aplicar a Transformada de Fourier sobre esses sinais. Assim, permitiu interpretar melhor as informações disponibilizadas, o que não seria possível a partir dos sinais originais.

Inicialmente, foram isolados apenas os marcadores articulares da região do quadril devido ao grande volume da base de dados. O isolamento dessas articulações foi um teste para entender o desempenho do algoritmo.

Posteriormente, as demais articulações foram acrescentadas, lendo os arquivos de dados de forma completa. Foram gerados testes comparativos de tempo de execução, o resultado desses testes se encontra na Tabela 2. Ao realizar os testes foi possível perceber que o aumento do tempo de execução não inviabilizava o trabalho com todas as articulações.

Tabela 2 – Testes de tempo de execução do algoritmo com as articulações do quadril isoladas X todas as articulações.

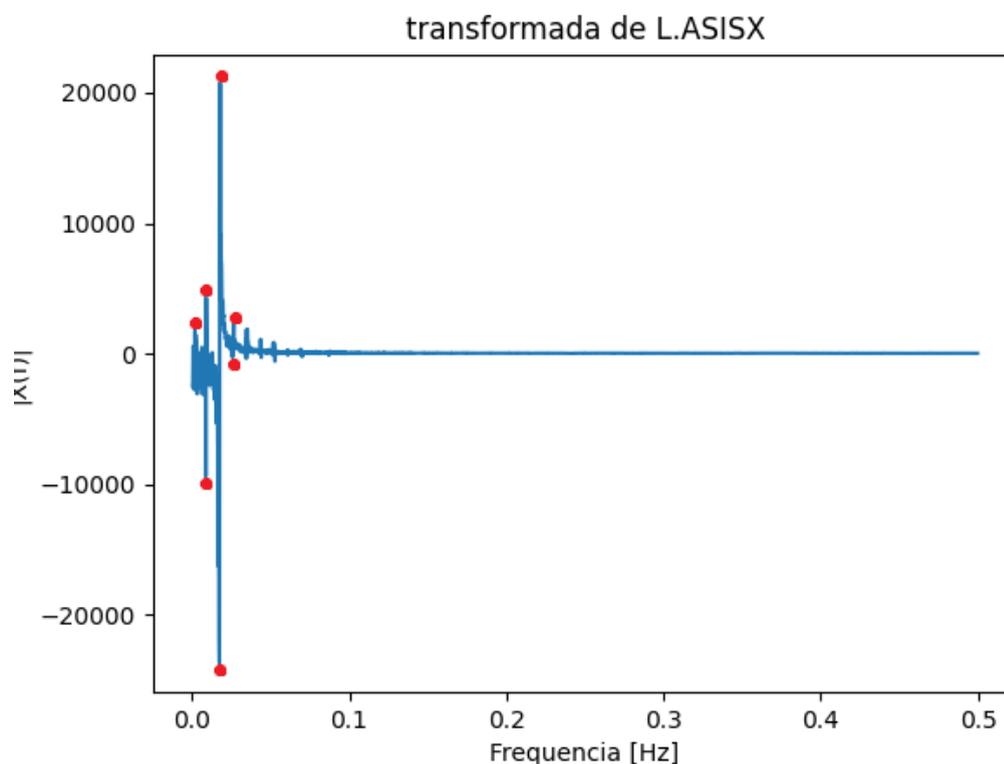
Teste de tempo de execução		
	6 articulações (s)	Todas articulações (s)
Teste 1	10,23	33,18
Teste 2	10,11	34,15
Teste 3	10,28	32,79
Teste 4	10,68	33,22
Teste 5	10,92	33,11
MÉDIA	10,44	33,29

A transformada de Fourier foi aplicada para todas as entradas de dados utilizando as funções do pacote “NumPy”, que fornece uma série de funções matemáticas para a linguagem de programação Python.

3.1.3 Seleção das Componentes da Transformada de Fourier

A execução da transformada de Fourier retorna um vetor de tamanho N de valores de magnitude e de frequência, também chamados de picos, sendo N um valor não constante. Cada um dos pontos vermelhos na Figura 23 representa um componente periódico do sinal submetido à transformada de Fourier.

Figura 23 – Gráfico de transformada de Fourier com destaque em alguns picos.



A escolha de componentes do sinal transformado se faz necessária por 2 motivos: controlar o volume de dados enviados aos algoritmos de aprendizagem de máquina e nivelar a quantidade de dados passado aos algoritmos, dada a imprevisibilidade da quantidade de componentes extraídos do sinal.

A obtenção dos picos é feita através de chamadas a funções `get_features` e `detect_peaks`, que transformam o vetor de retorno da Transformada de Fourier em uma lista de “features”, ou “características”, em tradução livre, e posteriormente a função `get_first_n_peaks` é chamada passando como parâmetro o número de componentes que se deseja extrair.

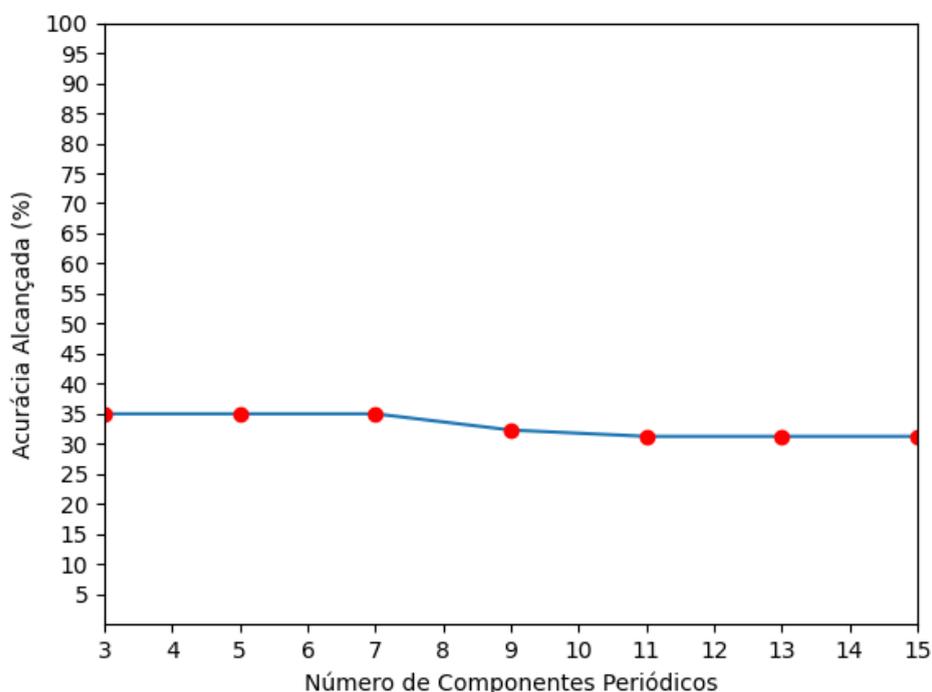
Para determinar a quantidade de picos a serem selecionados foram experimentados diferentes valores crescentes, até encontrar um valor ótimo. Os testes consistiram em selecionar um número X de picos e observar o resultado do agrupamento e classificação.

Foram testados os valores 3, 5, 7, 9, 11, 13 e 15 para definir o melhor número de componentes a serem extraídos da transformada. Ao final da execução dos testes, o valor ideal encontrado foi o de 7 componentes periódicos do sinal transformado, como ilustra a Figura 24. A acurácia alcançada com 3, 5 e 7 picos foi a mesma (35%), caindo para 32,25% quando captados 9 componentes, se mantendo estável conforme aumentou o número de componentes.

Após um trabalho de pesquisa, chegou-se a uma biblioteca denominada `detect_peaks` (disponível gratuitamente em: https://github.com/BMClab/BMC/blob/master/functions/detect_peaks.py). A biblioteca foi criada

para seleções de picos em funções de transformadas de Fourier, a fim de detectar os pontos mais altos do sinal transformado.

Figura 24 – Gráfico resultante dos testes de número de componentes periódicos da transformada de Fourier.

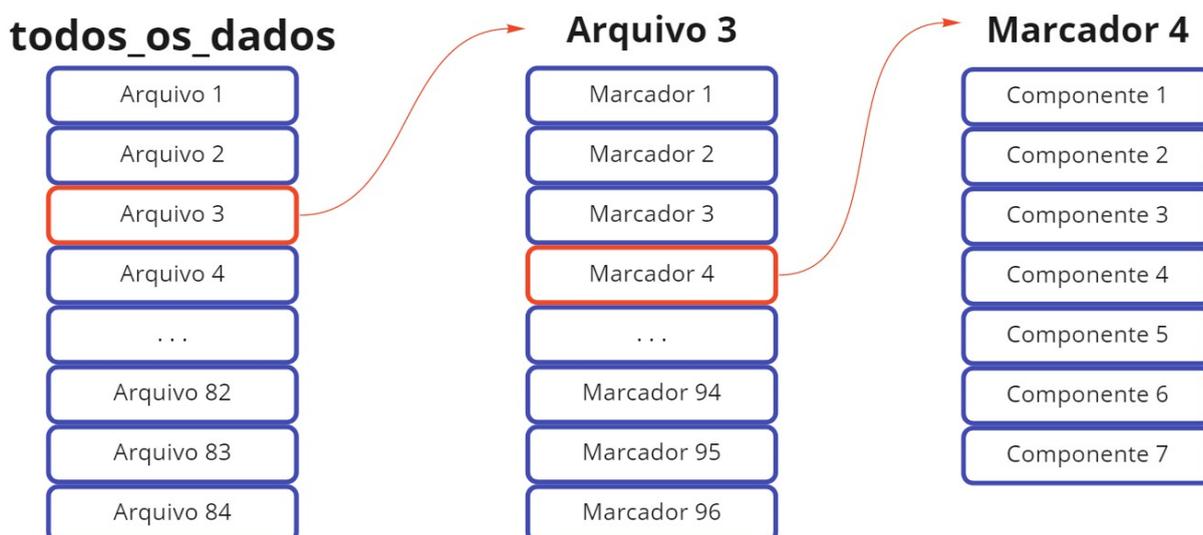


O trecho de código presente no anexo A deste estudo traz os métodos utilizados na escolha dos picos. O primeiro para selecionar os primeiros n picos, onde n é passado por parâmetro, e o segundo usado para obter as características dos sinais (`get_features`) fazendo chamada à biblioteca `detect_peaks`, citada no parágrafo anterior.

3.1.4 Preparação do vetor de dados

Após a seleção dos picos, a estrutura de dados está como representado na Figura 25: o vetor “`todos_os_dados`” contém dados extraídos de cada um dos 84 arquivos de dados, cada um dos arquivos contém informações de cada um dos 96 marcadores articulares e estes, por sua vez, contém 7 componentes de maior frequência, extraídos da Transformada de Fourier.

Figura 25 – Estrutura do vetor de dados.



No vetor de dados “todos_os_dados”, encontram-se dados extraídos dos 84 arquivos de dados. Dentro de cada um dos arquivos de dados estão os 96 marcadores articulares. Por fim, em cada um dos marcadores, localizam-se 7 componentes periódicos extraídos pela transformada de Fourier, e que representam a movimentação deste marcador.

3.2 Aplicação do K-Means

Após a etapa extração das características e pré-processamento dos dados, foi iniciada a etapa onde foram submetidos ao algoritmo de clusterização K-Means. Esta fase do estudo tem o objetivo de agrupar os dados que têm mais similaridades entre si, para que a partir disso, seja possível estudar os grupos formados, a fim de encontrar padrões que descrevam esses dados agrupados.

Para identificar o número ideal de clusters para o K-Means, novos testes foram executados, a fim de encontrar um número K ótimo para o problema. O pacote público “scikit-learn” fornece, além das funções para execução do K-means, métricas de avaliação do K escolhido.

A Tabela 3 demonstra o resultado obtido nas métricas aplicadas. Foram utilizadas as métricas: coeficiente de silhueta, Calinski-Harabasz e Davies-Bouldin. As setas no cabeçalho da tabela indicam o sentido dos valores, seta para cima significa que quanto maior o valor, melhor avaliado é o valor de K, já a seta para baixo indica que, quanto menor o valor, melhor avaliado estará o número de K. Após a execução dos testes, considerando qualquer uma das técnicas, o valor ideal para K é 3.

Tabela 3 – Testes com métricas do K-Means.

Métricas K-Means			
	Silhouette Coefficient ↑	Calinski-Harabasz ↑	Davies-Bouldin ↓
k = 2	0,209238495	23,46235924	1,723454097
k = 3	0,218102936	23,50563589	1,537917355
k = 4	0,205363685	20,1598248	1,666626053
k = 5	0,158503765	17,85519539	1,741296392
k = 6	0,174299177	16,22384081	1,635051552
k = 7	0,176545869	15,49249176	1,569326256

3.3 Aplicação do KNN

A fim de ampliar os resultados da pesquisa, foi aplicado além do K-means, o KNN. O agrupamento de dados é uma técnica de aprendizagem não supervisionada, já a classificação de dados, pelo contrário, é uma estratégia supervisionada, portanto é interessante estudar os resultados dessas abordagens em paralelo.

A aplicação da aprendizagem supervisionada aqui objetiva classificar entre lesionados e não lesionados, na busca de investigar se, com base apenas nos dados cinemáticos, seria possível fazer uma predição da lesão de um atleta.

Novamente, a exemplo do capítulo anterior, é necessário descobrir o valor ideal para K, desta vez para o KNN. Desta forma, para identificar o número ideal, foi aplicado o algoritmo com diferentes valores de K e, os resultados observados, podem ser vistos na Tabela 4.

Tabela 4 – Testes do melhor valor de K no KNN.

KNN					
valor de K	proporção para testes	Acurácia alcançada	matriz de confusão		
1	30%	0.5769230769230769		verdadeiro	falso
			verdadeiro	1	4
			falso	7	14
3	30%	0.7307692307692307		verdadeiro	falso
			verdadeiro	0	5
			falso	2	19
5	30%	0.6923076923076923		verdadeiro	falso
			verdadeiro	1	4
			falso	4	17
7	30%	0.7307692307692307		verdadeiro	falso
			verdadeiro	0	5
			falso	2	19
9	30%	0.8076923076923077		verdadeiro	falso
			verdadeiro	0	5
			falso	0	21
11	30%	0.8076923076923077		verdadeiro	falso
			verdadeiro	0	5
			falso	0	21

Na primeira coluna da tabela 4 está o valor de K, seguido da proporção com a qual os dados foram divididos entre dados de treinamento e de teste, o que é requisito do algoritmo KNN. Na terceira coluna, a acurácia que foi alcançada com o teste, seguido, por fim, da matriz de confusão resultante.

É possível perceber que o somatório do valor de qualquer uma das matrizes de confusão é 26. Isto se deve ao fato que o vetor de entrada tem tamanho 84 e 30% desses dados foram utilizados para os testes do algoritmo KNN, sendo essa porcentagem um valor bem próximo de 26.

Com os testes referidos acima, foi possível entender que o melhor valor para K seria 9 ou 11, o valor 9 foi escolhido arbitrariamente. Vale observar que, apesar da acurácia alcançada pelo KNN, a predição do algoritmo tende sempre a classificar a entrada como 0 (não lesionado), sendo assim, a aplicação do KNN não consegue atender a expectativa de predição buscada, como será explicado no próximo capítulo.

4 Apresentação e Análise dos Resultados

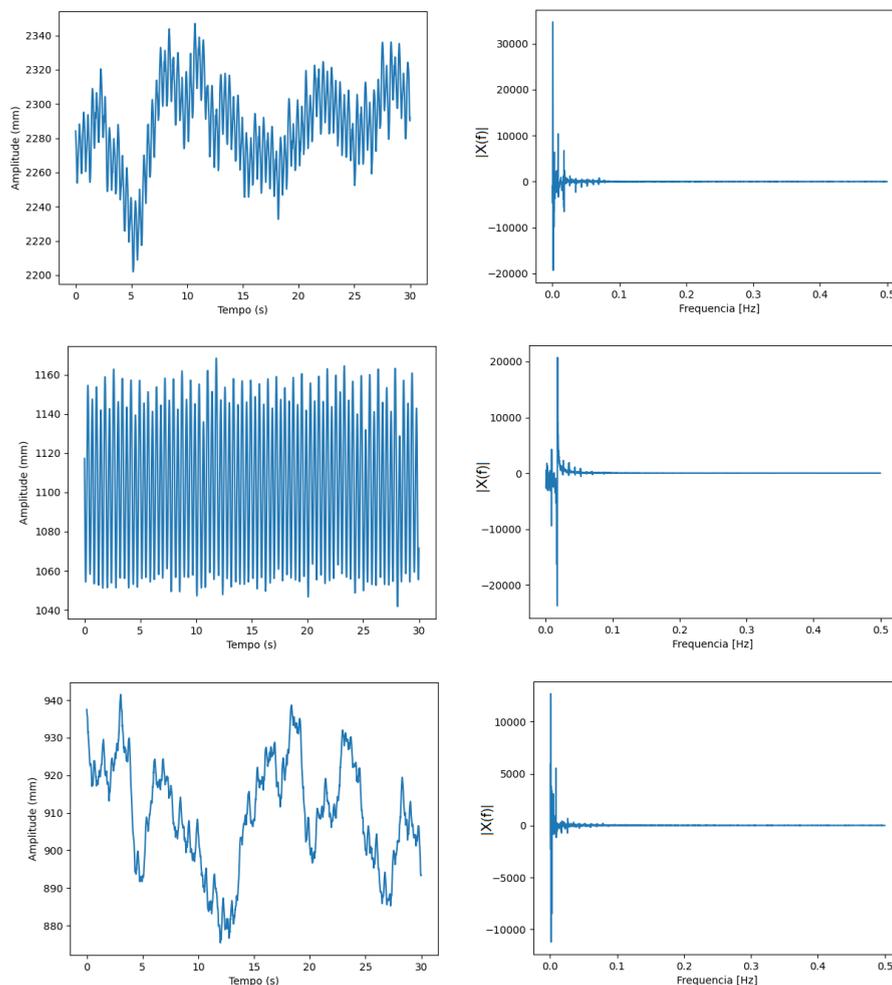
Neste capítulo serão apresentadas as análises que puderam ser feitas sobre os resultados dos esforços deste trabalho. Será abordada a transformada de Fourier e no que ela auxiliou no estudo, K-Means, estudo sobre os atletas e KNN.

4.1 Análise da Transformada de Fourier

Ao aplicar a transformada de Fourier sobre os dados deste estudo, o objetivo foi decompor o sinal das funções trigonométricas que o compõem, pois só dessa forma seria possível obter as características mais relevantes do sinal, podendo também visualizar através de gráficos a implicação do ritmo de corrida na movimentação do corredor.

Além da vantagem de retornar os fatores chave dos sinais, a transformada de Fourier também auxiliou como um fator de uniformização dos dados, com isso os sinais de articulações diferentes, em dimensões diferentes estarão no mesmo formato: senóides.

Figura 26 – Gráficos originais e transformados para uma mesma articulação nas dimensões X, Y e Z, respectivamente.



Os gráficos da Figura 26 representam movimentos de um mesmo marcador de movimentação articular, de um mesmo atleta, na mesma velocidade, mas em dimensões (X, Y e Z) diferentes. É perceptível a diferença dos gráficos originais, mas ao aplicar a Transformada de Fourier, nota-se que o formato do gráfico que representa os dados é bem similar, já que está lidando, após a transformada, com uma soma de componentes senoidais.

Apenas com os gráficos de transformadas foi possível fazer uma observação do impacto da relação ritmo de corrida com a amplitude dos valores impressos no gráfico.

Figura 27 – Gráfico da Transformada de Fourier para a coluna ilíaca ântero-superior direita de um atleta a 2,5 m/s.

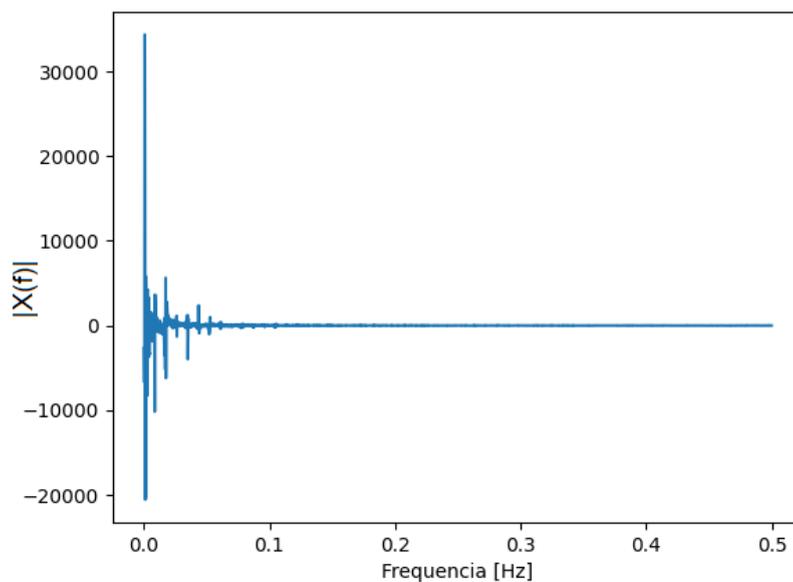


Figura 28 – Gráfico da transformada de Fourier para a coluna ilíaca ântero-superior direita de um atleta a 3,5 m/s.

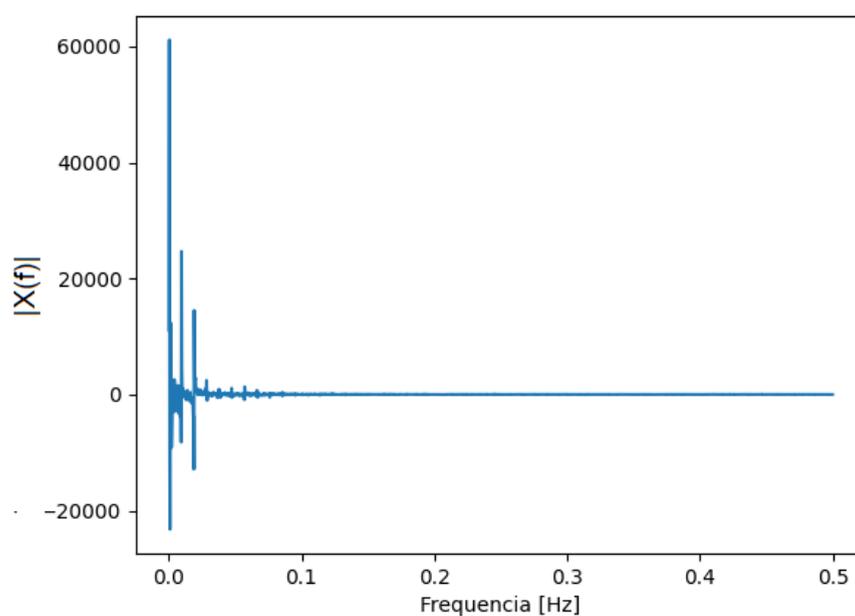
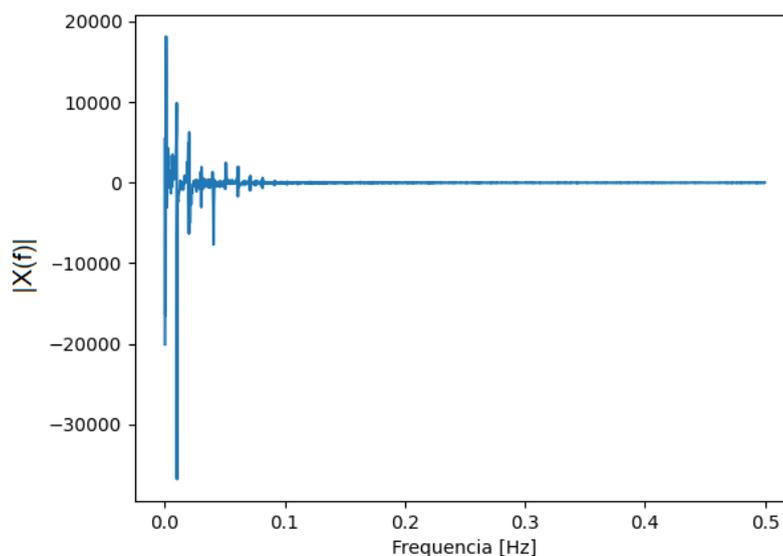


Figura 29 – Gráfico da Transformada de Fourier para a coluna ilíaca ântero-superior direita de um atleta a 4,5 m/s.



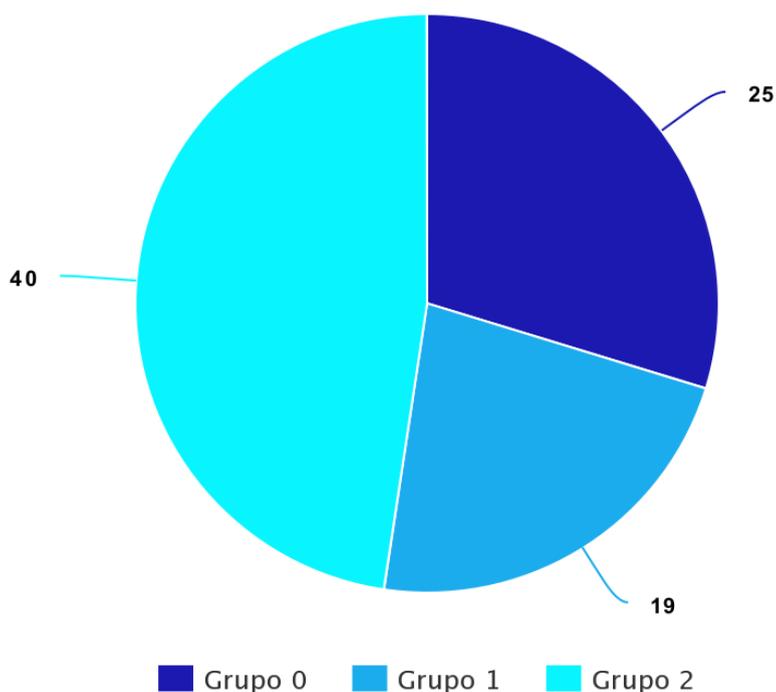
Como é possível observar nas figuras 27 a 29, há um aumento significativo na magnitude do sinal, mesmo se tratando de um gráfico referente ao mesmo corredor, nas mesmas condições de corrida, com exceção da velocidade.

O aumento da amplitude dos movimentos do corredor quando a velocidade aumenta de 2,5 m/s para 3,5 m/s é de aproximadamente 90% e, se for comparado com a velocidade máxima, 4,5 m/s, terá um salto de quase 500%. O padrão de aumento de amplitude dos movimentos se repete para todas as articulações, de todos os indivíduos. Isso permite inferir o impacto desse aumento drástico na fadiga muscular de um competidor em relação a um corredor ocasional.

4.2 K-Means

Ao implementar o algoritmo K-Means e submeter os dados pré-processados, se obtém como retorno, entre outras informações, um vetor de “labels”, ou rótulos, em tradução livre, do mesmo tamanho do vetor de entrada. Isto é, para cada entrada, o K-Means gera uma classificação de 0 a $(K - 1)$, sendo K o número de clusters que se espera dividir os dados, nesse caso, K tem o valor 3, portanto o algoritmo retornará rótulos de 0 a 2.

Ao aplicar o K-Means, o retorno obtido foi o de 3 grupos divididos como mostra a Figura 30, o grupo com mais elementos foi o grupo 2, com 40, seguido do grupo 0, com 25, e por último está o grupo 1, com apenas 19 elementos.

Figura 30 – Agrupamento dos dados retornado pelo K-Means.

A partir do agrupamento, o próximo passo foi analisar os padrões que poderiam existir nestes grupos. Para isto, foi utilizado o arquivo de características dos corredores citado na seção 2.2. Cruzando os dados contidos da tabela, com os grupos formados foi possível gerar gráficos, que auxiliaram na análise dos padrões. Estes gráficos serão apresentados a seguir nas figuras de 31 a 37.

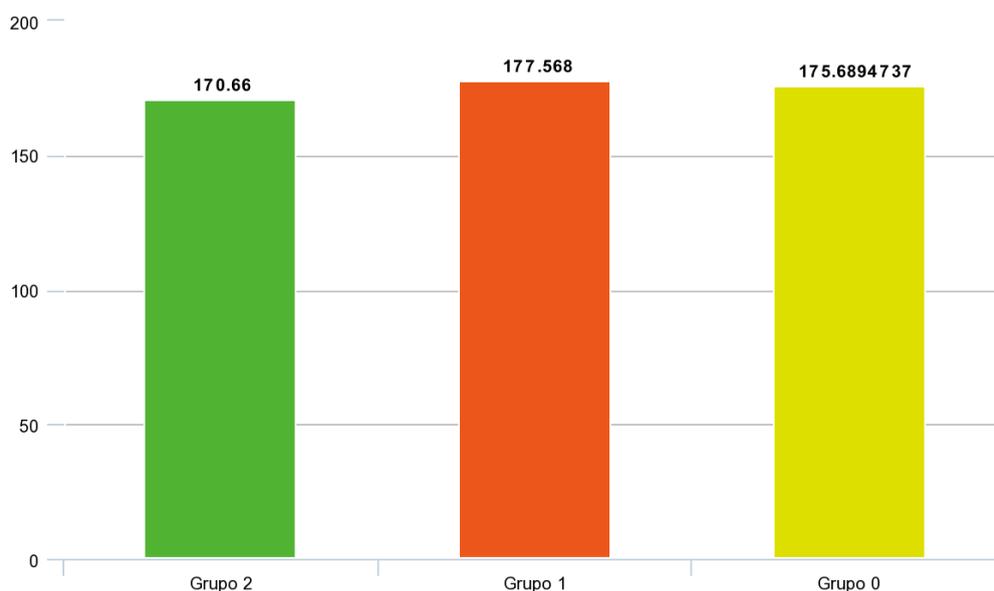
Figura 31 – Gráfico comparativo da altura média (em centímetros) dos grupos.

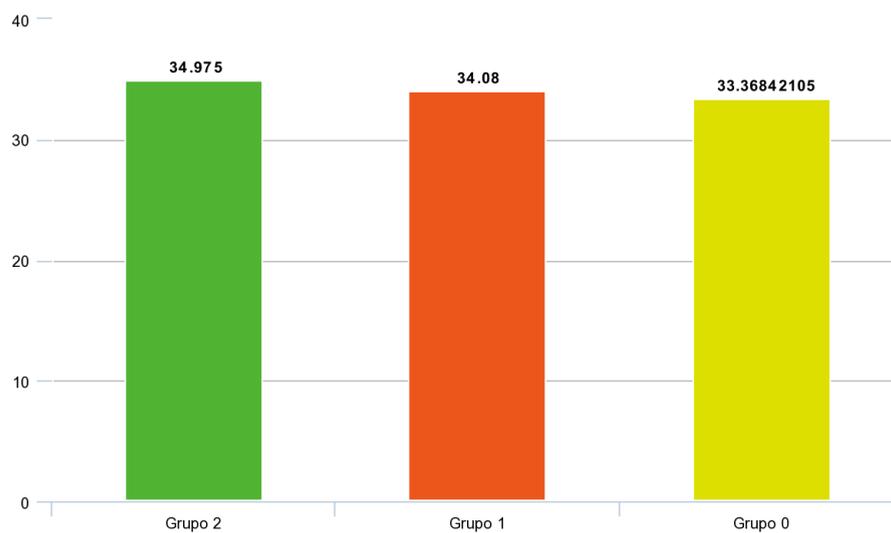
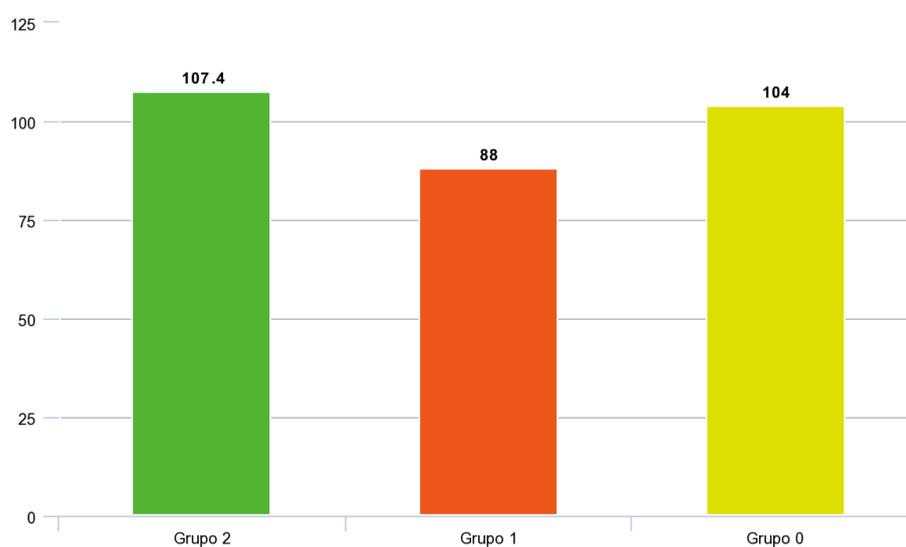
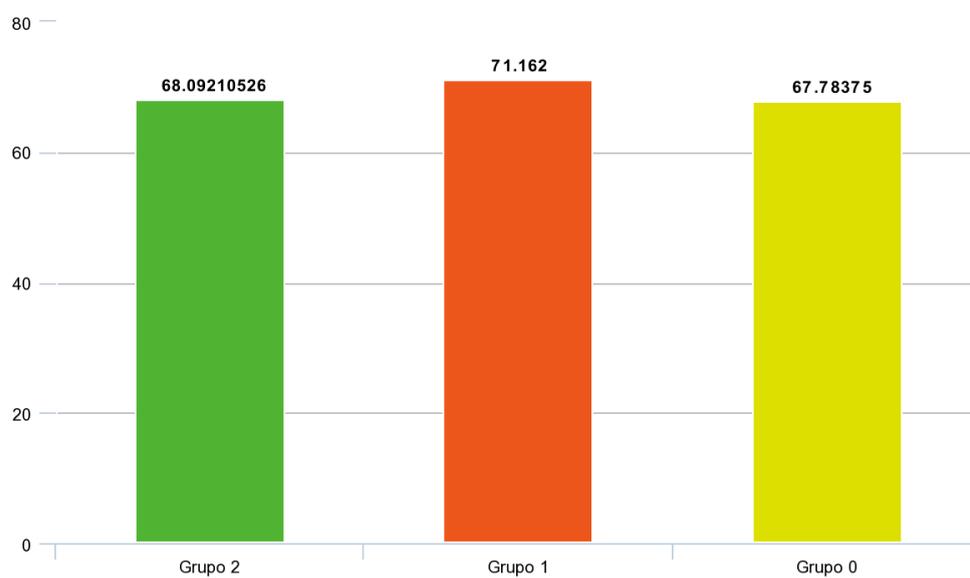
Figura 32 – Gráfico comparativo da idade média dos grupos**Figura 33 – Gráfico comparativo da experiência média (em meses) dos grupos.****Figura 34 – Gráfico comparativo do peso médio (em quilos) dos grupos.**

Figura 35 – Gráficos dos perfis dos corredores dentro dos grupos.

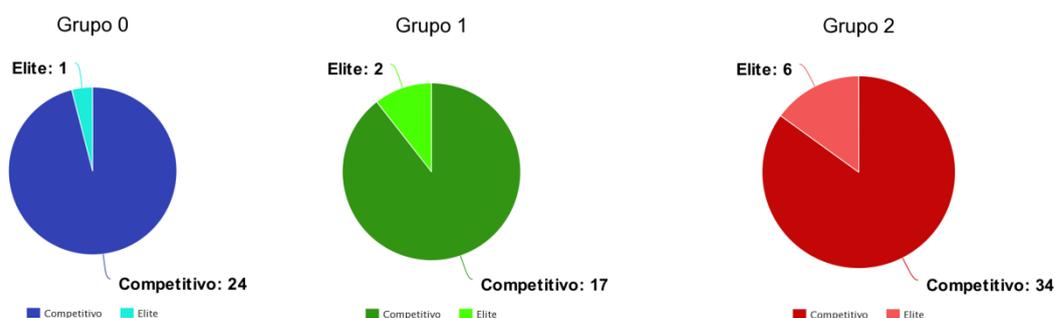


Figura 36 – Gráficos da quantidade de sessões de corrida que os corredores fazem por semana em cada grupo.



Figura 37 – Gráficos da distância percorrida pelos corredores por semana em cada grupo.



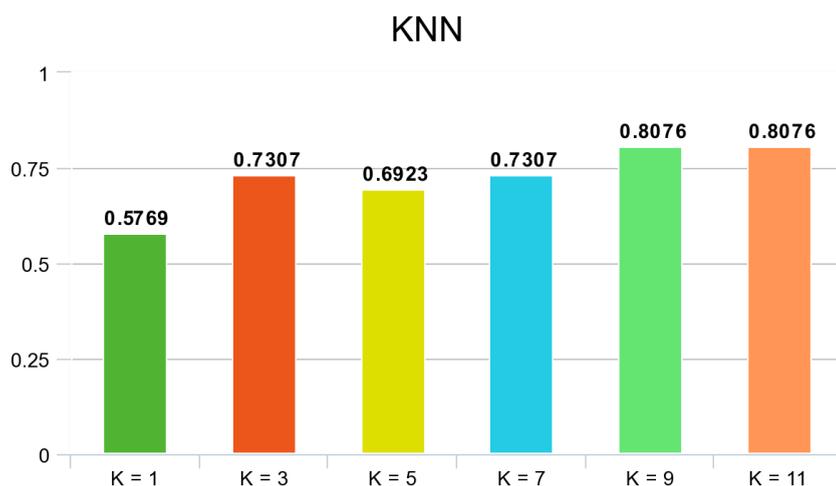
Após a análise dos gráficos apresentados, não foi possível notar heterogeneidade que sugerisse a presença de um padrão. Os resultados obtidos não foram conclusivos em relação às características predominantes nos grupos de indivíduos. Portanto, não é possível comprovar, apenas com informações de cinemática disponibilizadas pela base de dados, a classificação destes corredores.

Era esperado, com o agrupamento de dados, que algumas características se concentrassem em um dos grupos. Por exemplo: caso a maioria dos atletas de elite estivessem agrupados em um só grupo, isto comprovaria que os dados de cinemática destes atletas são consideravelmente diferentes dos demais. Tal comportamento não foi observado em nenhum dos grupos, em nenhuma das características estudadas.

4.3 Análise do KNN

Em complemento ao agrupamento de dados foi implementado, também, a classificação destes utilizando o algoritmo KNN. Procurou-se avaliar se os dados tinham uma boa distribuição espacial, com isso o KNN conseguiria classificar adequadamente os dados em classes de indivíduos com e sem histórico de lesões. Os testes executados revelaram que o valor ideal para K seria 9 ou 11. Com esses parâmetros a acurácia alcançada foi de 80,76% (Figura 38).

Figura 38 – Gráfico da acurácia alcançada pelo KNN conforme a variação do valor de K.



A acurácia alcançada pode parecer promissora num primeiro momento, mas, quando é analisada a matriz de confusão gerada após a execução do KNN (Tabela 5), é notável que este apenas apontou valores 0, isto é, a previsão que o algoritmo gerou é de que nenhum dos atletas se lesionou, estando certo em 21 dos casos teste e errado em 5. Portanto, apesar da acurácia relativamente alta, o algoritmo não é eficiente nesta base de dados por tender a uma só resposta.

Estes resultados levam a crer que o padrão de biomecânica, isoladamente analisado, não é suficiente para indicar a lesão de um atleta, ao menos se utilizando de uma base de dados com as limitações apresentadas.

Tabela 5 – Matriz de confusão resultante do KNN..

valor de K	matriz de confusão		
1		positivo	negativo
	verdadeiro	1	14
	falso	7	4
3		positivo	negativo
	verdadeiro	0	19
	falso	2	5
5		positivo	negativo
	verdadeiro	1	17
	falso	4	4
7		positivo	negativo
	verdadeiro	0	19
	falso	2	5
9		positivo	negativo
	verdadeiro	0	21
	falso	0	5
11		positivo	negativo
	verdadeiro	0	21
	falso	0	5

5 Conclusões

O presente trabalho de conclusão de curso destinou-se a estudar e a analisar a biomecânica do movimento humano em corridas estacionárias controladas. Para este estudo, foram utilizados os dados da cinemática coletados por meio de sensores inerciais.

A análise da cinemática examina o movimento sob uma perspectiva espacial e temporal, sem referência com as forças que causam o movimento. Ao realizar um estudo sob a ótica exclusiva da cinemática, estávamos interessados em analisar a descrição do movimento humano na realização de uma tarefa, neste caso uma corrida.

Uma hipótese comumente levantada na literatura é se há, ou não, padrões característicos de determinada patologia ortopédica, ou de efeitos do envelhecimento no padrão biomecânico do indivíduo.

Neste trabalho, a base de dados fornecida foi caracterizada por apresentar pouca diversidade entre os indivíduos. Isso pode ser observado pelas análises estatísticas preliminares levantadas através do questionário que foi preenchido por todos.

No entanto, ao analisar a cinemática da corrida de indivíduos em diferentes velocidades, procurou-se entender a influência do ritmo no padrão de movimento. Para isso, a análise de clustering serviu como método para identificar similaridades e agrupá-las. Nesta análise, os clusters identificados pelo algoritmo K-Means, foram validados por métricas diferentes e todas apontaram para um mesmo número de clusters, indicando fortemente semelhanças dentro dos grupos e as diferenças entre os grupos. Porém, ao se analisar grupo por grupo, viu-se a heterogeneidade das variáveis de interesse, como: velocidade de corrida do teste, episódios de lesões, volume semanal, sexo, IMC, etc. Diante desses resultados, sob a perspectiva da cinemática e, com base neste grupo de indivíduos, foi possível perceber que a análise do movimento não pode ser característica de padronização de grupos de perfis.

Ainda assim, em outra abordagem para descartar a capacidade de se identificar grupos, adotou-se um algoritmo de aprendizado supervisionado, a fim de avaliar a capacidade de se classificar indivíduos com episódios de lesões musculoesqueléticas em sua história de treinamento, com base na cinemática observada. O algoritmo K-NN usado neste procedimento registrou resultados que, analisados com base na matriz de confusão, apontaram para a falta de capacidade em obter uma separação razoável entre as classes. Apesar disso, este já era um resultado esperado de ser observado, uma vez que o padrão de movimento analisado isoladamente não é garantia de previsão de lesões.

O que foi testado neste experimento era se, mesmo não havendo uma coerência espacial entre os dados da cinemática entre indivíduos já lesionados e que ainda não apresentam nenhum episódio, o algoritmo de classificação conseguiria gerar uma diferenciação entre eles.

Além disso, o problema de previsão de lesões com base na cinemática é pouco plausível de ser solucionado, dada a enorme complexidade que envolve a origem de uma lesão musculoesquelética. Ainda, mesmo que um indivíduo pudesse ter apresentado um padrão

“indesejado” para esta atividade física e não houvesse sofrido nenhuma lesão em sua história, ainda sim haveria a possibilidade de apresentar uma lesão futura por fatores que fogem ao escopo deste estudo.

Por fim, o que fica evidente neste estudo é que a cinemática, apesar de ter sua importância no estudo da biomecânica humana, pode não possuir informação suficiente para caracterizar perfil preditivo de lesões de natureza musculoesquelética, quando analisada isoladamente. Além, é claro, da base de dados disponível possuir pouca diversidade entre os indivíduos observados.

A observação por si só, neste tipo de estudo pode, indiretamente, modificar o comportamento do observado. Mesmo que não existisse interferência física, o simples conhecimento que se está sendo observado já pode afetar a forma como se comporta durante a atividade.

Por fim, o universo amostral para um estudo dessa característica deverá ser maior e mais diversificado que o utilizado. Esta base serviu de experiência para a continuidade de um estudo anterior, baseado na cinética, e também serviu para aprimorar os métodos de pré-processamento e transformação para outras bases de dados similares que podem ser coletadas futuramente. Inclusive, uma das opções para continuidade deste estudo é realizar uma observação de população maior e devidamente estratificada.

Referências

AMADIO, A. C. et al. **Introdução à biomecânica para análise do movimento humano: descrição e aplicação dos métodos de medição**. Revista Brasileira de Fisioterapia, v. 3, n. 2, p. 41-54, 1999. <https://www.scielo.br/pdf/rbefe/v25nspe/03.pdf>

BARON DE FOURIER, Jean Baptiste Joseph. **Théorie analytique de la chaleur**. Firmin Didot, 1822. <https://doi.org/10.1017/CBO9780511693229>

BIJALWAN, V.; KUMAR, V.; KUMARI, P.; PASCUAL, J. **KNN based Machine Learning Approach for Text and Document Mining**. International Journal of Database Theory and Application, vol.7, no.1 (2014), pp.61-70. <http://dx.doi.org/10.14257/ijdta.2014.7.1.06>

BOREL, Wyngrid Porfirio et al. **PREVALENCE OF INJURIES IN BRAZILIAN RECREATIONAL STREET RUNNERS: META-ANALYSIS**. Rev Bras Med Esporte, São Paulo, v. 25, n. 2, p. 161-167, Apr. 2019. Available from <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S1517-86922019000200161&lng=en&rm=iso>. access on 06 May 2021. <https://doi.org/10.1590/1517-869220192502214466>.

CHUNG, T. M. **Avaliação cinética e cinemática da marcha de adultos do sexo masculino**. Acta Fisiátrica, v. 7, n. 2, p. 61-67. 2000. <https://doi.org/10.11606/issn.2317-0190.v7i2a102258>

COOLEY, J.; TUKEY, J. **An Algorithm for the Machine Calculation of Complex Fourier Series**. *Mathematics of Computation*, 19(90), 297-301, 1965. doi:10.2307/2003354

COVER, T.; HART, P. "Nearest neighbor pattern classification," in **IEEE Transactions on Information Theory**, vol. 13, no. 1, pp. 21-27, January 1967, doi: 10.1109/TIT.1967.1053964.

FUKUCHI, R. K.; FUKUCHI, C. A.; DUARTE, M. **A public dataset of running biomechanics and the effects of running speed on lower extremity kinematics and kinetics**. PeerJ, PeerJ Inc., v. 5, p. e3298, 2017. 10.7717/peerj.3298

HATZE, H. **A model of skeletal muscle suitable for optimal motion problems**. In: Nelson R.C., Morehouse C.A. (eds) Biomechanics IV. International Series on Sport Sciences. Palgrave, London, 1974. https://doi.org/10.1007/978-1-349-02612-8_62

HORNG, S.J. et al. **A novel intrusion detection system based on hierarchical clustering and support vector machines**. Expert systems with Applications, v. 38, n. 1, p. 306-313, 2011. <https://doi.org/10.1016/j.eswa.2010.06.066>

JAIN, A.K.; MURTY, M. N.; FLYNN, P. J. **Data clustering: a review**. ACM Comput. Surv.31, 3 (Sept. 1999), 264–323. DOI:<https://doi.org/10.1145/331499.331504>

LIPOR, John; BALZANO, Laura. **Clustering quality metrics for subspace clustering**. *Pattern Recognition*, v. 104, p. 107328, 2020. <https://doi.org/10.1016/j.patcog.2020.107328>

MANGINI, F.L.; FABRICA, G. **Mechanical stiffness: a global parameter associated to elite sprinters performance**. Rev. Bras. Ciênc. Esporte, Porto Alegre , v. 38, n. 3, p. 303-309, Sept. 2016. <https://doi.org/10.1016/j.rbce.2016.02.004>

MELO, S.IL; SANTOS, SG. **Antropometria em biomecânica: características, princípios e modelos antropométricos**. Revista Brasileira de Cineantropometria e Desempenho Humano, v.2, n.1, p.97-105. Janeiro de 2000.<https://doi.org/10.1590/%25x>

MUSTAFA, M. et al. **Comparison between KNN and ANN classification in brain balancing application via spectrogram image**. Journal of Computer Science & Computational Mathematics, v. 2, n. 4, p. 17-22, 2012.[10.20967/jcscm.2012.04.004](https://doi.org/10.20967/jcscm.2012.04.004)

NETO, A. L. **A note on projective Levi flats and minimal sets of algebraic foliations**. Annales de l'Institut Fourier, Tome 49 (1999) no. 4, pp. 1369-1385. doi : 10.5802/aif.1721. http://www.numdam.org/item/AIF_1999__49_4_1369_0/

NIGG, B.; BALTICH, J.; HOERZER, S.; ENDERS, H. **Running shoes and running injuries: mythbusting and a proposal for two new paradigms: 'preferred movement path' and 'comfort filter'**. British Journal of Sports Medicine, V. 49, p 1290-1294, 2015.[10.1136/bjsports-2015-095054](https://doi.org/10.1136/bjsports-2015-095054)

OBENSHAIN, M. K. **Application of data mining techniques to healthcare data**. Infection Control & Hospital Epidemiology, v. 25, n. 8, p. 690-695, 2004. [10.1086/502460](https://doi.org/10.1086/502460)

OLIVEIRA, M. **Reconhecimento de padrões em Biomecânica de Corrida usando Aprendizado de Máquina**. 68 f. TCC (Graduação) - Curso de Engenharia de Computação, Decsi, Universidade Federal de Ouro Preto, João Monlevade, 2018

PAKHIRA, M.K.; BANDYOPADHYAY, S.; MAULIK, U. **Validity index for crisp and fuzzy clusters**. Pattern Recognition, v. 37, n.3, p. 487-501, 2004. <https://doi.org/10.1016/j.patcog.2003.06.005>

PORTELA, M.M.; LIMA, M.O.; SILVA, P.L.; BARJA, P.R.; TEMOTEO, R.F.; ARISAWA, E.A. **Análise Biomecânica Comparativa Do Padrão De Movimento Entre Corredores Após Dois Protocolos De Treinamento De Atrator**. Revista Univap. 25. 1. [10.18066/revistaunivap.v25i48.2196](https://doi.org/10.18066/revistaunivap.v25i48.2196). 2019.

ROUSSEUW, Peter J. **Silhouettes: a graphical aid to the interpretation and validation of cluster analysis**. Journal of computational and applied mathematics, v. 20, p. 53-65, 1987. [https://doi.org/10.1016/0377-0427\(87\)90125-7](https://doi.org/10.1016/0377-0427(87)90125-7)

SALGADO, J.V. **CORRIDA DE RUA: análise do crescimento do número de provas e de praticantes**. Trabalho de conclusão de curso - Faculdade de Educação Física, UNICAMP, Campinas, 2006. <https://doi.org/10.20396/conex.v4i1.8637965>

STEINER, M.T.A. et al . **Abordagem de um problema médico por meio do processo de KDD com ênfase à análise exploratória dos dados**. Gest. Prod., São Carlos, v.13, n.2, p.325-337, May 2006 . <https://doi.org/10.1590/S0104-530X2006000200013>

SUZUKI F.S.; KALYTCZAK, M.M.; CASARIN, C.A.S.; SERRA, A.J.; EVANGELISTA, A.L.; MARCHETTI, P.H. et al. **Physical activity level does not influence the neuromuscular fatigue in adults**. Rev Bras Med Esporte, v. 2, n.22, p. 97–101. 2016. <https://doi.org/10.1590/1517-869220162202150282>

UZUN, Isa Servan; AMIRA, Abbas; BOURIDANE, Ahmed. **FPGA implementations of fast Fourier transforms for real-time signal and image processing**. IEE Proceedings-Vision, Image and Signal Processing, v. 152, n. 3, p. 283-296, 2005. 10.1109/FPT.2003.1275737

VAUGHAN, C.L. Biomechanics of running gait. **Critical Reviews in Biomedical Engineering**. 1984 ;12(1):1-48.

WILLIAMS et al. **Motivational predictors of weight loss and weight loss maintenance**. Journal of Personality and Social Psychology, Washington DC, v.70, n.1, p.115-126, 1996. 10.1037//0022-3514.70.1.115

WU, G.; CAVANAGH, P.R. **ISB Recommendations for standardization in the reporting of kinematic data**. The Center for Locomotion Studies, Penn State University, University Park, PA 16802, U.S.A. 1995. 10.1016/0021-9290(95)00017-c

APÊNDICE A – Sub Rotinas utilizadas na transformada de Fourier

```
def get_first_n_peaks(x, y, no_peaks=7):  
  
    x_, y_ = list(x), list(y)  
  
    if len(x_) >= no_peaks:
```

```
        return x_[:no_peaks], y_[:no_peaks]

    else:

        missing_no_peaks = no_peaks - len(x_)

        return x_ + [0] * missing_no_peaks, y_ + [0] * missing_no_peaks

def get_features(x_values, y_values):

    mph = 0.1 * np.nanmax(y_values)

    indices_peaks = detect_peaks(y_values, mph)

    peaks_x, peaks_y = get_first_n_peaks(x_values[indices_peaks],
y_values[indices_peaks])

    return peaks_x + peaks_y
```