



MODELAGEM E CLASSIFICAÇÃO DA MARCHA HUMANA USANDO SENSORES INERCIAIS E DADOS CLÍNICOS PARA PREDIÇÃO DE SESSÕES FISIOTERAPÊUTICAS

Trabalho de Conclusão de Curso

Engenharia da Computação

Mariana de Albuquerque Campos
Orientador: Prof. Fernando Buarque de Lima Neto
Co-orientador: M. Sc. Rafael Caldas



UNIVERSIDADE
DE PERNAMBUCO

**Universidade de Pernambuco
Escola Politécnica de Pernambuco
Graduação em Engenharia de Computação**

MARIANA DE ALBUQUERQUE CAMPOS

**MODELAGEM E CLASSIFICAÇÃO DA
MARCHA HUMANA USANDO SENSORES
INERCIAIS E DADOS CLÍNICOS PARA
PREDIÇÃO DE SESSÕES
FISIOTERAPÊUTICAS**

Monografia apresentada como requisito parcial para obtenção do diploma de Bacharel em Engenharia de Computação pela Escola Politécnica de Pernambuco – Universidade de Pernambuco.

Recife, Dezembro 2018.

MONOGRAFIA DE FINAL DE CURSO

Avaliação Final (para o presidente da banca)*

No dia 20 de dezembro de 2018, às 11:00 horas, reuniu-se para deliberar a defesa da monografia de conclusão de curso do discente **MARIANA DE ALBUQUERQUE CAMPOS**, orientado pelo professor **Fernando Buarque de Lima Neto**, sob título **MODELAGEM E CLASSIFICAÇÃO DA MARCHA HUMANA USANDO SENSORES INERCIAIS E DADOS CLÍNICOS PARA PREDIÇÃO DE SESSÕES FISIOTERAPÊUTICAS**, a banca composta pelos professores:

Diego José Rátiva Millan

Fernando Buarque de Lima Neto

Após a apresentação da monografia e discussão entre os membros da Banca, a mesma foi considerada:

Aprovada Aprovada com Restrições* Reprovada

e foi-lhe atribuída nota: 9,0 (Nove)

*(Obrigatório o preenchimento do campo abaixo com comentários para o autor)

O discente terá 30 dias para entrega da versão final da monografia a contar da data deste documento.

DIEGO JOSÉ RÁTIVA MILLAN

FERNANDO BUARQUE DE LIMA NETO

* Este documento deverá ser encadernado juntamente com a monografia em versão final.

Dedico essa monografia a minha família, especialmente meus pais, cujo apoio foi essencial para que chegasse onde estou.

Agradecimentos

Primeiramente, agradeço a Deus pela oportunidade e capacidade de fazer uma graduação, permitindo realizar um sonho de me tornar engenheira da computação.

A meus pais que deram todo apoio não apenas no período da graduação, como também em todos os momentos de decisão. Nunca mediram esforços de proporcionar o melhor para mim e meu irmão. Eu amo vocês. Agradeço também a meu irmão Guilherme que compreendeu minha ausência em assistir os jogos do Sport Clube do Recife por que precisava estudar.

A meus amigos que estiveram nos momentos bons e ruins durante esta trajetória. O curso me proporcionou conhecer pessoas fantásticas que levarei sempre comigo, especialmente o grupo Ecomp da Deprê. As pessoas desse grupo me apoiaram e ajudaram em vários momentos, até naqueles que pensava em desistir. Obrigada, galera.

Aos meus orientadores, Fernando Buarque e Rafael Caldas, pelo suporte e apoio durante um ano de projeto. Agradeço por todo suporte, correções, incentivos e dedicação a este projeto.

Ao meu namorado, Felipe, por todo apoio, carinho e compreensão. Principalmente nesses últimos meses que foram mais difíceis. Obrigada por tudo.

A Carlos Santos e Rodrigo de Paula que auxiliaram com seus conhecimentos para que pudesse implementar os algoritmos.

E a todos, que diretamente ou indiretamente contribuíram para minha formação acadêmica. Muito obrigada.

*“We’ve all got both light and dark inside us.
What matters is the part we choose to act on. That’s who we really are.”*
(Harry Potter and Prisoner of Azkaban)

Resumo

A marcha humana é um padrão de movimento fisiológico que corresponde ao maior nível de independência funcional motora. Vários tipos de disfunções, como trauma ortopédico, distúrbios neurológicos, podem comprometer a marcha. Os profissionais de fisioterapia enfrentam algumas dificuldades no tratamento de pacientes pois métodos convencionais apresentam limitações. Além disso, determinar o número de sessões fisioterapêuticas com precisão, para cada deficiência torna-se difícil principalmente para profissionais mais jovens. Uma alternativa de baixo custo é o uso de sensores inerciais que associados a dados clínicos e fisioterapêuticos compõem um sistema de predição para sugerir o número de sessões de fisioterapia. Sensores são capazes de capturar a aceleração e velocidade angular e a partir desses dados obter informações que determinam características da marcha. Ademais, os dados clínicos e fisioterapêuticos provenientes de uma pesquisa realizada em conjunto com o setor de traumatologia do IMIP foram usados como entrada do sistema preditivo. Floresta de árvores de decisão foram usadas para construção do sistema preditivo, onde foram aplicados 4 tipos de algoritmos e observou-se que o GOSS (*Gradient-based One-Side Sampling*) ou Amostragem um lado Gradiente produzem melhor resultado para predição com cerca de 17% de erro na predição de sessões fisioterapêuticas. Como o algoritmo é altamente parametrizável, melhores resultados podem ser obtidos com a realização de mais testes e aumentando-se a base de dados dos pacientes.

Palavras-chave: Predição da marcha humana, Raios de simetria, IMU, Floresta de árvores de decisão.

Abstract

Human gait is a pattern of physiological movement that corresponds to the highest level of functional motor independence. Several types of disabilities, such as orthopedic trauma, neurological and rheumatic disorders can jeopardize the gait. Physiotherapy professionals face some difficulties in treating patients because conventional methods have restrictions. Furthermore, determining the number of physiotherapeutic sessions accurately for each disability is difficult, especially for younger professionals. A low-cost alternative is to use inertial sensors to capture the acceleration and angular velocity to obtain information that determines gait characteristics. This monography uses the gait characteristics, plus clinical and physiotherapeutic data collected from a research conducted in conjunction with the IMIP traumatology sector as input to a predictive system that suggests the number of physiotherapy sessions. Decision trees forest were used to construct a predictive system, where four types of algorithms were applied. The results showed that GOSS (Gradient-based One-Side Sampling) produce the best result for the prediction of physiotherapeutic sessions with about 17%. As the algorithm is highly parameterizable, better results may be achieved by performing more tests and increasing the subjects database.

Keywords: Gait human prediction, Symmetry ratio, IMU, Decision forest

Sumário

Capítulo 1 Introdução	1
1.1 Motivação.....	1
1.2 Objetivos	3
1.2.1 Objetivo Geral.....	3
1.2.2 Objetivos Específicos	3
1.3 Estrutura da Monografia	4
Capítulo 2 Fundamentação Teórica	5
2.1 Fases da marcha humana	5
2.2 Unidades de Sensores Inerciais	6
2.4 Processamento Digital de Sinais	10
2.4.1 Representação de Sinais Digitais	11
2.4.2 Filtros Digitais	11
2.5 Padronização.....	13
2.6 Árvores de Decisão	13
2.6.1 Floresta de árvores de decisão	16
Capítulo 3 Materiais e Métodos	18
3.1 Aquisição de dados	18
3.1.2 Dados clínicos e Fisioterapêuticos.....	19
3.2 Pré-processamento	20
3.3 Processamento.....	22

3.4 Avaliação do sistema	26
3.4.1 Erro Quadrático Médio (EMQ).....	26
3.4.2 Variação explicada.....	26
Capítulo 4 Testes e Resultados	28
4.1 Extração de dados temporais da marcha.....	28
4.2 Regressão	33
Capítulo 5 Considerações Finais	36
5.1 Conclusão.....	36
5.2 Trabalhos Futuros.....	37
Bibliografia	38
Apêndice A - Termo de consentimento livre e esclarecido	41
Apêndice B - Ficha de coleta de dados - IMIP	42

Índice de Figuras

Figura 1.	Diferentes fases da marcha humana.....	6
Figura 2.	Sensor inercial.....	8
Figura 3.	Visão sistêmica de um sensor inercial.....	8
Figura 4.	Unidade Inercial	10
Equação 1.	Equação gaussiana para duas dimensões	12
Figura 5.	Representação de um problema em forma de árvore	14
Figura 6.	Exemplo de flor de íris - exemplo	15
Figura 7.	Gráfico da distribuição de tipos de íris - exemplo.....	15
Figura 8.	Partição dos exemplos de íris e a árvore de decisão gerada.....	16
Figura 9.	Gráfico do valor da aceleração X tempo	20
Figura 10.	Gráfico da aceleração com filtro gaussiano X tempo	21
Figura 11.	IC, EC e MS no sinal da marcha humana	21
Figura 12.	LightGBM – Crescimento das folhas nas árvores de decisão	23
Figura 13.	Exemplo de variância explicada aplicada a uma função	27
Figura 14.	Gráfico dos sensores dos pés, no eixo z e velocidade de 1,0 m/s	29
Figura 15.	Gráfico dos sensores das pernas, no eixo z na velocidade de 1,0 m/s	30
Figura 16.	Gráfico dos sensores das mãos, no eixo z e na velocidade de 1,0 m/s.....	30

Índice de Tabelas

Tabela 1 - Valor médio dos 6 parâmetros temporais em segundos.....	31
Tabela 2 - Valor médio dos raios de simetria.....	32
Tabela 3 - Diagnósticos dos pacientes em atendimento.....	33
Tabela 4 - Avaliação dos algoritmos	35

Índice de Equações

Equação 1. Equação gaussiana para duas dimensões	12
Equação 2. Erro Quadrático Médio	26

Tabela de Símbolos e Siglas

DART – *Dropouts Additive Regression Trees* (Descarte aditivo de árvores de regressão)

EC – *End contact*(Fim do contato)

EMQ – Erro Quadrático Médio

GBDT – *Gradient Boosting Decision Tree* (Árvore de decisão com aumento de gradiente)

GOSS – *Gradient-based One-Side Sampling* (Amostragem um lado gradiente)

GPU – *Graphic Process Unit* (Unidade de processamento gráfico)

IC – *Initial Contact* (Contato Inicial)

IMIP – Instituto de Medicina Integral Professor Fernando Figueira

IMU – *Inertial Measurement Units* (Unidades de sensors inerciais)

MEMS – *Micro Electro Mechanical Systems* (Sistemas microeletromecânicos)

MS – *Mid-swing* (Balanço médio)

USB – *Universal Serial Bus*

RF – *Random Forest* (Floresta randômica)

Capítulo 1

Introdução

Este capítulo encontra-se dividido em três seções. Em seu início é descrita a motivação para o desenvolvimento deste trabalho. Em seguida, são discutidos os seus principais objetivos. Por fim, é apresentada a estruturação adotada para os demais capítulos desta monografia.

1.1 Motivação

Nos últimos anos, a procura pelos serviços de traumatologia vem crescendo em decorrência do aumento de acidentes automobilísticos e traumas, associados muitas vezes às características sociodemográficas são as causas mais frequentes das principais patologias atendidas nos setores de traumatologia. Além disso, pode-se apontar também lesões esportivas, as mais comuns são: entorses, estiramentos e distensões.

A incapacidade funcional decorrente do trauma ortopédico configura um problema de saúde pública, uma vez que causa transtornos tanto para o indivíduo quanto para seus familiares, além do custo social agregado a tal limitação. Kfuri *et. al.* (2011) comenta que o trauma ortopédico é uma das condições mais mórbidas na sociedade, comprometendo a função do indivíduo, sua participação econômica na comunidade e sua integração familiar.

Segundo Kfuri, a unidade de emergência do Hospital das Clínicas de Ribeirão Preto da Universidade de São Paulo, que é um hospital de urgências terciário, o trauma ortopédico responde em média por 45 a 65% do movimento cirúrgico semanal, com uma ocupação acima da capacidade do hospital em alguns momentos. Apesar do cenário retratado ser no interior de São Paulo, o mesmo se repete nos hospitais do Brasil. A falta de infraestrutura mínima de trabalho faz com que pacientes sejam atendidos tardiamente, muito deles após a consolidação de suas fraturas originais.

Outro problema associado à infraestrutura é a falta de bons equipamentos para avaliação do paciente, principalmente para o diagnóstico do estado da marcha, dependendo assim clinicamente da experiência do profissional. De acordo com Ma & Liao (2017) para análises mais detalhadas, alguns centros especializados contam com o padrão ouro que corresponde a um sistema óptico e placas de força que geram dados dos pacientes o qual é bastante caro e necessita de profissionais de fisioterapia qualificados para análise dos dados gerados. Esse cenário também é reforçado por George *et al.* (2017).

Diante desse contexto, novas pesquisas e equipamentos de avaliação em conjunto com metodologias de Inteligência Artificial e algoritmos de aprendizado de máquinas vem sendo amplamente estudados e difundidos em hospitais e clínicas fisioterapêuticas.

A marcha humana possui características do indivíduo capazes de identificar anormalidades com base em padrões pré-estabelecidos. Com a identificação desses padrões, a aplicação dessas tecnologias e algoritmos de aprendizado torna-se fácil e de grande ajuda para este ramo da fisioterapia, Caldas & Lima Neto (2018).

Ma & Liao (2017), por exemplo, tendo em vista esses obstáculos propõem uma abordagem mais simples e menos custosa com o uso de sensores inerciais fixados em palmilhas, que permite uma precisão dos dados, leveza e baixo custo. Com os dados provenientes do hardware, é possível fazer a modelagem da marcha humana usando o algoritmo do Semi Processo de Markov (algoritmo de aprendizado de máquina) e com bons resultados.

Observando todo esse contexto, compreende-se a importância e contribuição desse projeto, uma vez que auxiliará os profissionais de fisioterapia em determinar a quantidade de sessões necessárias para sua recuperação. Principalmente para os profissionais mais jovens que não possuem tanta experiência, poderão ter em mãos uma ferramenta inovadora e precisa.

1.2 Objetivos

1.2.1 Objetivo Geral

O foco deste projeto tem como objetivo principal o desenvolvimento de um sistema preditivo para a quantidade de sessões fisioterapêuticas usando informações obtidas dos sensores, dados clínicos e fisioterapêuticos coletados de pacientes em parceria com profissionais do IMIP. Além disso, a pesquisa visa classificar a marcha humana e extrair informações relacionadas a simetria da marcha com base em dados provenientes de sensores inerciais. O modelo proposto pode auxiliar no diagnóstico do paciente.

1.2.2 Objetivos Específicos

Como objetivos específicos, esse projeto se dispôs a:

- Extrair informações dos sensores inerciais;
- Construir uma base de dados com dados fisioterapêuticos e clínicos de pacientes com algum tipo de traumato-ortopedia;
- Desenvolver um sistema de predição para o número de sessões de fisioterapia;
- Treinar, testar e validar a base de dados para avaliar a eficiência do algoritmo;

1.3 Estrutura da Monografia

Esta monografia está dividida em 5 capítulos, sendo esse capítulo 1, destinado à introdução do tema central do trabalho, sua motivação, seus objetivos e a estrutura da monografia.

No capítulo 2 é discutida a fundamentação teórica acerca dos algoritmos utilizados, com a explicação dos conceitos como padrões da marcha humana, sensores inerciais e floresta de árvore de decisão.

Por sua vez, no capítulo 3, são apresentados os materiais e métodos desta monografia, definindo e caracterizando o objetivo de estudo. Neste capítulo está detalhado a metodologia aplicada, desde a coleta e processamento dos dados até a construção do sistema de predição.

O capítulo 4 relata os resultados obtidos frente a comparação dos algoritmos de floresta de árvore de decisão que foram empregados. Os resultados serão discutidos e avaliados sobre a implicação dos mesmos.

Por fim, o capítulo 5 está reservado com uma conclusão do trabalho, assim como listar possíveis melhorias e passos a serem executados para evolução dos resultados obtidos.

Capítulo 2

Fundamentação Teórica

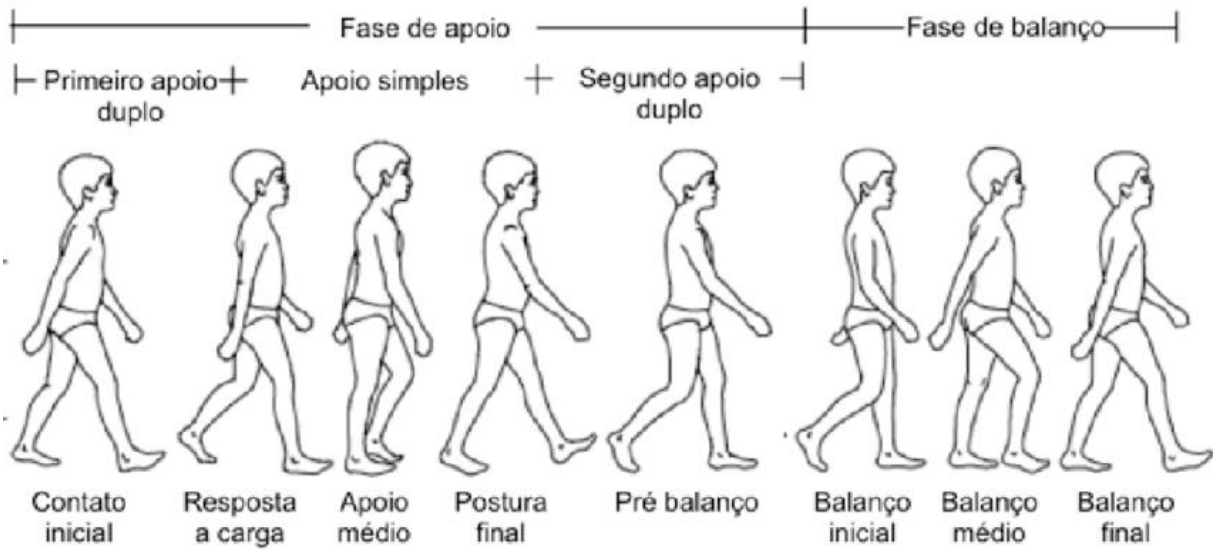
Este capítulo visa a apresentação de conceitos básicos de dispositivos e metodologias usados na pesquisa. Nas próximas páginas será apresentado o conceito de sensores inerciais, uma explicação sobre marcha humana, os parâmetros fisioterapêuticos avaliados e sobre floresta de árvore de decisão. Esses conceitos embasam o desenvolvimento deste trabalho, portanto são necessários para melhor entendimento.

2.1 Fases da marcha humana

Para uma boa compreensão deste projeto é necessário o conhecimento de nomenclaturas usadas por profissionais de fisioterapia. A marcha humana possui algumas nomenclaturas específicas que sem uma brévia explicação podem confundir o leitor. A seguir será descrito alguns dos termos que serão mencionados durante a monografia.

A marcha humana corresponde ao modo fisiológico de locomoção, que pode ser afetado por diversas disfunções. Assim, a análise da marcha humana desempenha um papel importante para fisioterapeutas, por que tal análise envolve a avaliação de várias características biomecânicas. Isso permite que o profissional possa identificar causas de várias lesões, possibilitando maior efetividade no tratamento do paciente.

Figura 1. Diferentes fases da marcha humana



Fonte: Leite *et al.* (2017)

A figura 5 mostra a análise da marcha humana. Existem diferentes nomenclaturas para cada evento em que a marcha é subdividida. Como mostrado na parte inferior da figura 5, a marcha é constituída por duas fases principais: *Stance* e *swing*, que podem ser traduzidas respectivamente por apoio e balanço. A nomenclatura em português será adotada ao longo deste projeto.

Essas duas principais fases podem ser subdivididas em outras fases. A fase de apoio começa com o contato inicial (IC), seguida por resposta à carga (LR), apoio médio (MS), apoio terminal (TS), pré-balanço (PS) e finaliza com contato final (EC). A fase de balanço começa com balanço inicial (IS), seguido por balanço médio (MS) e conclui com balanço terminal (TS).

2.2 Unidades de Sensores Inerciais

Sensores inerciais constituem sensores que calculam forças de um objeto partindo do princípio da primeira lei de *Newton*, a inercia, dentre eles se destacam os acelerômetros e giroscópios. Os acelerômetros são capazes de medir a aceleração própria do objeto, assim calculando as forças exercidas sobre ele.

Os acelerômetros podem ser criados como Microeletromecânicos, ou seja, ser implementados em microdispositivos, que possibilita seu uso em pequenos sensores. Nos sensores inerciais, Almeida *et al.* (2014) explica o tipo de acelerômetro implementado, chamado capacitivo. Nesta implementação o sensor possui uma pequena massa que é deslocada quando submetida a algum movimento. Essa massa, que é uma espécie de placa, aproximam ou distanciam-se dos capacitores, modificando assim a capacitância deles. Através dessa diferença é possível calcular a força aplicada no objeto.

Os sensores inerciais são dispositivos capazes de monitorar variações de velocidade e aceleração, linear ou angular, através da conversão de forças inerciais em alguma mudança física conhecida que possa ser capturada por um transdutor correspondente e convertida em um sinal elétrico, segundo Torres *et al.*(2014). Este sinal elétrico é submetido a processos de filtragem a fim de se criar uma estimativa do sinal de entrada. A saída final representará um valor calibrado da aceleração.

Torres *et al.* (2014) ainda comenta que estes tipos de sensores tem por objetivo perceber os efeitos da ação de forças que provoquem uma mudança do estado inercial de sistemas sobre os quais estas forças são exercidas, sensores estes baseados na tecnologia MEMS (Sistemas microeletromecânicos). Devido a um grande número de aplicações, estes sensores se tornaram um dos sistemas mais populares.

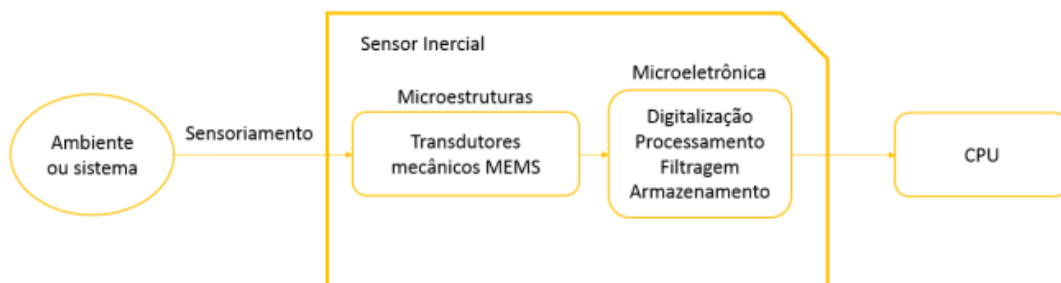
Figura 2. Sensor inercial



Fonte: <www.noraxon.com>. Acesso em 28 nov. 2018

Em um único circuito integrado encontramos estruturas mecânicas microusinadas formando transdutores mecânicos, responsáveis por realizar a tarefa de sensoriamento, e a microeletrônica, competente o suficiente para nos fornecer sinais elétricos analógicos correspondentes às forças as quais o sensor foi submetido. A figura a seguir mostra um resumo do funcionamento de um sensor inercial:

Figura 3. Visão sistêmica de um sensor inercial



Fonte: <www.embarcados.com.br>. Acesso em 28 nov. 2018

As microestruturas transformam os fenômenos físicos aferidos em um sinal elétrico proporcional. Esse sinal é convertido e capturado por um sistema computacional, que trata e armazena os dados.

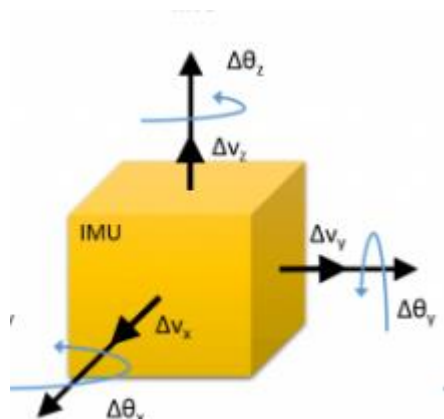
A categoria de sensores inerciais baseados em MEMS é representada por dois tipos de dispositivos:

- **Acelerômetros:** Capazes de aferir a aceleração linear na direção de um eixo referencial. A aceleração é a taxa de variação da velocidade no tempo, representada em m/s^2 (metro por segundo ao quadrado) no Sistema Internacional;
- **Giroscópios:** Mede a velocidade angular em torno de um eixo de referência. A velocidade angular é uma grandeza que representa a taxa de variação da posição angular no tempo, cuja unidade de medida no Sistema Internacional é o rad/s (radianos por segundo).

Acelerações e velocidades angulares são sinais vetoriais, possuindo módulo, direção e sentido. Caso apenas um componente do sensor seja medido, o sensor é caracterizado como de 1 eixo. Para este projeto foram utilizados sensores que capturam três eixos e que possuem os módulos de acelerômetros e giroscópios.

A combinação dos módulos de giroscópio e acelerômetro permite a criação do IMU (*Inertial Measurement Unit*) ou Unidade de medida inercial. Esse sistema é composto por três giroscópios, capaz de medir a velocidade angular em três eixos ortogonais. Associado a eles, três acelerômetros para medir a aceleração também em cada eixo ortogonal.

Conectando o IMU a um computador pode-se processar o sinal dos 6 sensores. Neste caso, não trata-se apenas de um circuito integrado, mas de um sistema embarcado completo, capaz de realizar a fusão dos sensores através dos algoritmos que agregam os dados de cada sensor individualmente. Isso proporciona os dados de velocidade e posicionamento para controle de atitude e direção.

Figura 4. Unidade Inercial

Fonte: <www.embarcados.com.br>. Acesso em 28 nov. 2018

A figura 4 mostra os eixos capturados, e as variações das velocidades angular e aceleração captadas pelo IMU. ΔV_x é a representação geral para a variação de velocidade reconhecido pelo sensor. $\Delta \Theta_x$ é a representação geral para variação da angulação do sensor em relação ao eixo correspondente.

Esses sensores são bastante atrativos não só pela quantidade de informação que pode ser obtida, mas também pela simplicidade e baixo custo. Uma vez que esse dispositivo tem seus componentes internos feitos de silício, deixa o IMU mais viável economicamente. Além disso, pode ser aplicado em várias áreas de conhecimento, principalmente para fins acadêmicos.

2.4 Processamento Digital de Sinais

O processamento de sinais tem por objetivo extrair informações de sinais e/ou torná-los mais apropriados a uma determinada aplicação, como sugere Proakis & Manolakis (1996). O processamento digital de sinais é um resumo de uma série de conceitos que pode ser traduzida na aplicação de algoritmos computacionais ou em um dispositivo eletrônico para realizar operações específicas sobre esses dados, segundo Puhlmann *et. al.* (2014).

Diversas são as aplicações possíveis no processamento digital de sinais, tais como reconhecimento de voz, extração e reconhecimento de padrões de imagens, sensores, transmissão de rádio e TV, dentre outras.

2.4.1 Representação de Sinais Digitais

Um sinal emitido por um sensor é um sinal contínuo, ou seja, ele está definido para qualquer instante do tempo. São várias as formas de representação do sinal proveniente de um sensor, como cita Lathi *et. al.*(2004), no domínio da frequência, a representação do sinal é feita pela magnitude do sinal.

No processo de tratamento de sinais, uma ferramenta bastante útil e de grande importância são os filtros. Tanto os analógicos que utilizam componentes analógicos físicos como resistores, capacitores e indutores, para produzir efeito de filtragem, como os digitais. Os filtros tem como vantagens: Programação via *software* e não sofrem variação com o tempo.

2.4.2 Filtros Digitais

O processo de filtragem permite suavizar a frequência de um sinal. Os filtros digitais são geralmente aplicados com dois propósitos: Separação ou restauração de um sinal distorcido.

Ao filtrar um sinal, presume-se que haja uma parte das informações que queremos separar de outra parte. Seja separar o sinal do ruído, ou um sinal de outros sinais. Para isso, várias estratégias podem ser utilizadas, iremos focar em duas: Filtros passa-baixas e Filtros Gaussianos.

Filtros passa-baixas

O filtro passa-baixa é um dos filtros mais utilizados, com ele pode dar origem a diversos outros tipos de filtros. Este filtro tem por finalidade permitir a passagem de sinais com frequências abaixo de um limiar estabelecido, denominado de frequência de corte, e atenuar (na física, é a diminuição da intensidade do sinal) sinais com

frequência acima desse limiar, fazendo com que seja gerado um sinal que evidencia as frequências abaixo da frequência de corte.

Filtro Gaussiano

A distribuição normal, também conhecida como curva de Gauss ou gaussiana é uma das mais importantes curvas utilizadas na matemática e na estatística. Segundo Jesus & Costa Jr.(2015), essa curva que originou-se nos estudos de *Abraham de Moivre* em um artigo no ano de 1733, gerou vários estudos a posteriori e hoje possui várias aplicações, principalmente no processamento digital de imagens.

Outra aplicação bastante comum é o filtro gaussiano. Ele é utilizado para borrar ou desfocar a imagem na qual ele é aplicado, com o objetivo de reduzir os ruídos presentes na imagem. Da mesma forma, pode ser aplicado ao sinal de sensores suavizando o sinal e retirando o ruído.

Matematicamente, a aplicação do filtro gaussiano é realizado da mesma forma que a convolução da imagem com uma função gaussiana, como um filtro passa-baixa ou como um filtro de média simples. A função que define o filtro gaussiano é:

Equação 1. Equação gaussiana para duas dimensões

$$G(x, y) = G(x) \cdot G(y) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}}$$

Fonte: Jesus & Costa Jr.(2015)

A equação acima representa a calculo do filtro gaussiano para uma matriz, onde o σ é o desvio padrão. O resultado será tão mais suave quanto maior for o valor de sigma, σ . A vantagem do filtro gaussiano é que ele permite uma suavização mais delicada, preservando mais os contornos que outros tipos de filtros.

2.5 Padronização

Este é um conceito importante para abordagens com aprendizado de máquina e que precisa de um bom resultado para o algoritmo, principalmente se estiver trabalhando com uma grande quantidade de dados. Existem algumas técnicas que auxiliam essa normalização dos dados, como: Normalização, Padronização, Redimensionamento, dentre outros.

A padronização é um processo comum para muitos estimadores de aprendizado de máquina. Quando esse requisito não é feito, o algoritmo pode ter uma taxa de aprendizagem pequena, uma vez que os dados não estão distribuídos normalmente. Isso significa que antes de começar a treinar ou prever em seu conjunto de dados, primeiro deve-se eliminar os extremos, ou seja, valores que não estão centralizados em torno do zero.

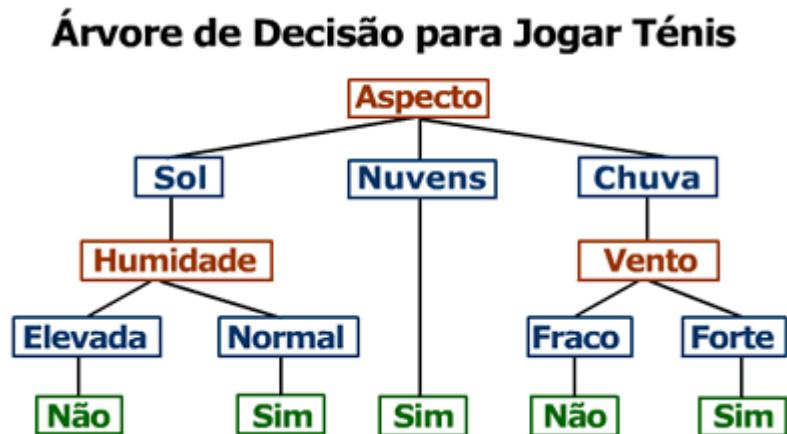
Este processo deve ser empregado no início do algoritmo em cada coluna da base de dados do problema. São calculados a média e o desvio padrão dos valores dessa tabela. Após isso, cada valor é avaliado e dimensionado conforme a média da coluna de dados que está sendo avaliada. Este processo é considerado uma boa prática em algoritmos de aprendizado de máquina.

2.6 Árvores de Decisão

Uma árvore tem muitas analogias na vida real, e descobriu-se que ela influenciou uma ampla área de aprendizagem de máquina, cobrindo tanto regressão quanto classificação. Como o próprio nome diz, ela usa um modelo de decisões em forma de árvore. Apesar de ser uma ferramenta mais utilizada para mineração de dados, ela pode ser amplamente aplicada em aprendizagem de máquina.

Árvores de um modo geral, em computação, são estruturas de dados formadas por um conjunto de elementos que armazenam informações chamadas folhas ou nós. Estes modelos utilizam a estratégia de dividir para conquistar: um problema complexo é decomposto em sub-problemas mais simples e recursivamente, esta técnica é aplicada a cada sub-problema.

Figura 5. Representação de um problema em forma de árvore



Fonte: < <http://web.tecnico.ulisboa.pt>>. Acesso em 29 nov. 2018.

De acordo com o exemplo da figura 6, cada quadro representa um nó. O quadro denominado “Aspecto” representa o nó chamado raiz. Esse nó possui o maior nível hierárquico (ponto de partida) e ligações para outros elementos, denominados filhos. Esses filhos podem ter seus próprios filhos que por sua vez também possui os seus filhos. O nó que não possui filhos são conhecidos como nó folha ou terminal, representados na figura anterior como os quadros em verde.

Em uma árvore de decisão, uma decisão é tomada através do caminho a partir de um nó raiz até o nó folha. Cada nó que não representa o nó folha tem em si uma regra de decisão. E os nós folhas representam uma decisão a ser tomada. A ideia geral é particionar o espaço recursivamente, nos quais um modelo simples é aprendido.

Tomando-se como exemplo um conjunto de dados que contém exemplos de três espécies de flor de íris. Cada um dos quais possui quatro atributos: largura e comprimento da sépala e da pétala da flor. Para esta demonstração, usaremos os atributos de largura e comprimento da pétala.

Figura 6. Exemplo de flor de iris - exemplo

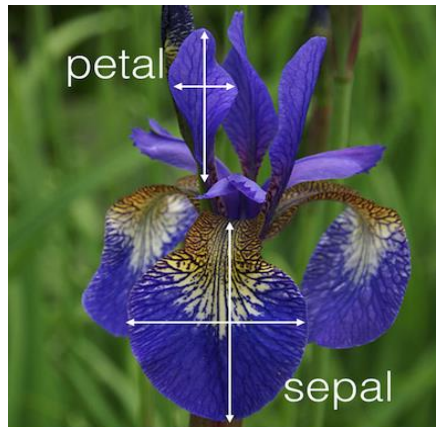
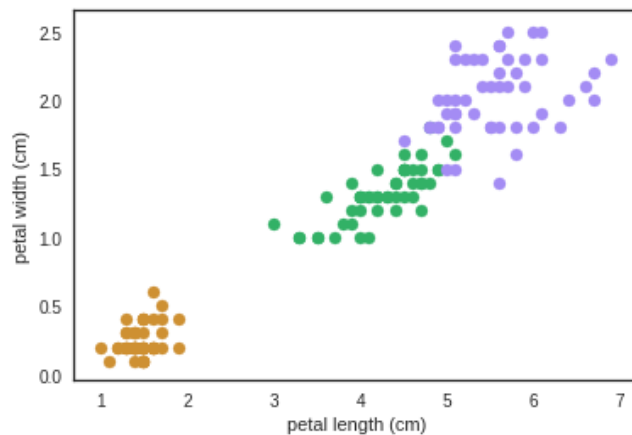


Figura 7. Gráfico da distribuição de tipos de íris - exemplo



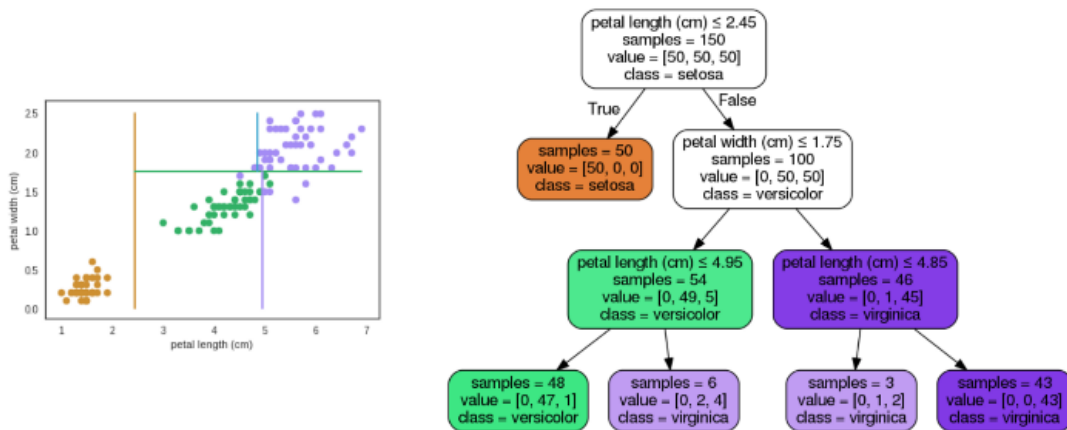
Fonte: < <https://medium.com/machine-learning-beyond-deep-learning/árvores-de-decisão> >. Acesso em 29 nov. 2018

A figura 7 mostra um exemplo da flor de iris com suas medidas. A figura 8 mostra a distribuição dos exemplos para as três espécies de iris, onde cada cor representa uma espécie diferente, são elas: setosa (amarela), versicolor (verde) e virginica (roxa).

Analisando o gráfico da figura 7, percebemos que cada tipo de íris possui um intervalo de medidas de pétala característicos. Isso permite criar uma árvore de decisão onde seja possível classificar, analisando em cada nó os limites de cada tipo. Aplicando-se todos os exemplos da base, a árvore de decisão ganha forma e

permite classificar qualquer exemplo pelas medidas de largura e comprimento da pétala. A figura a seguir mostra o formato da árvore após treinamento.

Figura 8. Partição dos exemplos de íris e a árvore de decisão gerada



Fonte: < <https://medium.com/machine-learning-beyond-deep-learning/árvores-de-decisão> >. Acesso em 01 dez. 2018

2.6.1 Floresta de árvores de decisão

As árvores de decisão são excelentes preditores, no entanto podem apresentar uma certa dificuldade, pois nem sempre generalizam bem, segundo Lucas *et. al.* (2011). As florestas de árvores de decisão utilizam técnicas capazes de cobrir problemas maiores, de forma a evitar o aprendizado em apenas uma direção.

Alternativas para otimizar esse processo são dadas através do uso de técnicas como:

- **Bagging:** Que pode ser traduzido como alargamento, usa uma pluralidade de amostras como base para uma boa solução que seja mais generalizável.

- **Boosting:** Traduzido como impulsionar, utiliza um conjunto de técnicas menos precisas para se obter um resultado de consenso que tenha boa qualidade, já que uma solução completa outra.
- **Randomizing:** Traduzido como aleatorização, em cada método/técnica empregada, utiliza-se um conjunto de diferentes variáveis.

Uma floresta de árvores de decisão emprega técnicas na criação das árvores a fim aumentar o aprendizado. Durante o projeto foram utilizadas algumas metodologias que diferem quanto ao *boosting*, ou seja, técnicas mais fracas que se complementam. Essas técnicas tem como objetivo priorizar alguns ramos com dados mais relevantes e que ofereçam um maior aprendizado no processo em detrimento de outros. Logo, as sub-regiões em árvores identificadas com baixa taxa de ganho são descartados pelo algoritmo.

Capítulo 3

Materiais e Métodos

Neste capítulo é apresentado o modelo proposto para o desenvolvimento de um sistema de predição para sessões fisioterapêuticas. São mostrados também desde a coleta de dados dos sensores inerciais, as metodologias usadas para tratamento dos sinais, assim como a coleta de dados clínicos dos pacientes até a utilização de todas essas entradas para o sistema preditivo. Finalmente são detalhadas todas as ferramentas que foram utilizadas em cada fase e também os critérios de avaliação.

3.1 Aquisição de dados

Como proposta, o sistema deve ter como entrada dados coletados de sensores inerciais e dados clínicos, provenientes de consulta de pacientes do IMIP com profissionais de fisioterapia. Esta pesquisa tem autorização condicionada ao cumprimento dos requisitos da Resolução 466/12 e suas complementares, como mostrado no Apêndice A, comprometendo-se a utilizar os dados pessoais dos participantes da pesquisa, exclusivamente para os fins científicos, mantendo o sigilo e garantindo a não utilização das informações em prejuízo das pessoas e/ou das comunidades.

3.1.1 Sensores Inerciais

Para este projeto foram utilizados oito sensores inerciais do tipo IMU onde cada sensor projetou velocidade angular e aceleração em três eixos. Para adquirir esses valores da marcha, os sensores foram posicionados da seguinte forma: um sensor para cada pé, um sensor para cada mão, um sensor na parte superior da coluna, um sensor na parte inferior da coluna, um sensor em cada tíbia. Foram escolhidas esse posicionamento dos sensores a fim de obter a maior variabilidade de movimentos dos pacientes.

Os dados cinemáticos de cada sensor são coletados por um software próprio do fabricante de sensores. O *software* deve estar instalado em um computador, e este deve ter uma câmera integrada para realizar a captura dos movimentos. O programa responsável utiliza as imagens com um rápido algoritmo de processamento que permite calcular valores de velocidade angular e acelerômetro.

A partir de cada paciente foram realizadas cinco capturas de movimentos em cinco velocidades diferentes: 0.8, 1.1, 1.4, 1.7 e 2.0 m/s, gerando um total de 25 planilhas que serviram para compor uma das entradas do sistema de predição.

3.1.2 Dados clínicos e Fisioterapêuticos

Uma parceria com profissionais do Centro de Reabilitação e Medicina Física Prof. Ruy Neves Baptista, do IMIP permitiu a coleta de dados sociodemográficos e clínicos dos pacientes com alguma disfunção que afete sua deambulação, atendidos no ambulatório de traumatologia-ortopedia.

O estudo que foi realizado no período de Agosto a Novembro de 2018 com cerca de 39 pacientes com o objetivo de montar uma base de dados para identificação e descrição do perfil do paciente em atendimento, além de identificar as condutas terapêuticas adotadas de acordo com os objetivos terapêuticos nas subsequentes etapas de tratamento.

Dentre os dados fisioterapêuticos foram usados como base em Peccin, Ciconelli & Cohen (2006) e Marx *et al.* (2006) que determinam alguns tipos de dados específicos que devem ser coletados no momento da consulta com o objetivo de reconhecer patologias dos pacientes e determinar a quantidade de sessões para tratamento.

Através dos dados clínicos obtem-se informações do histórico familiar, como diabetes, hipertensão, histórico de acidentes, histórico de cirurgias e outros. Ou seja, dados que ajudem nas metodologias de tratamento do paciente. Além destes, foram coletados dados fisioterapêuticos que constituem dados mais específicos dos

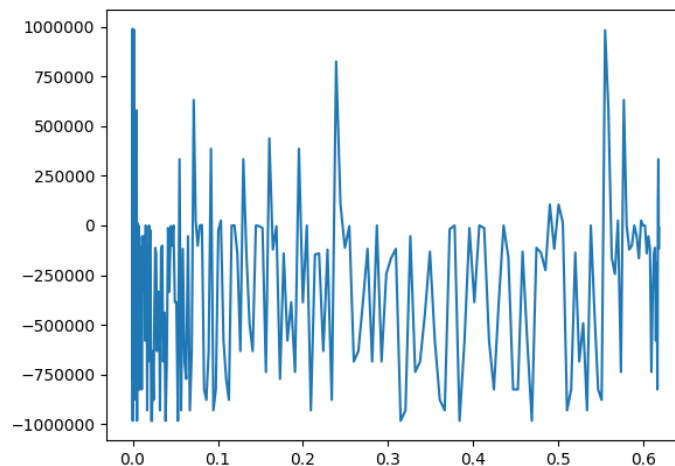
membros inferiores. Essas informações auxiliam o profissional no acompanhamento dos pacientes.

O formulário de coleta de dados dos pacientes está no Apêndice B. Apenas os dados clínicos e fisioterapêuticos foram utilizados no experimento deste projeto.

3.2 Pré-processamento

Dados provenientes de sensores possuem uma quantidade enorme de ruído que não permite a visualização dos dados de forma clara para a obtenção de informações. O sinal proveniente do sensor localizado no pé esquerdo em um dos eixos possui a seguinte característica:

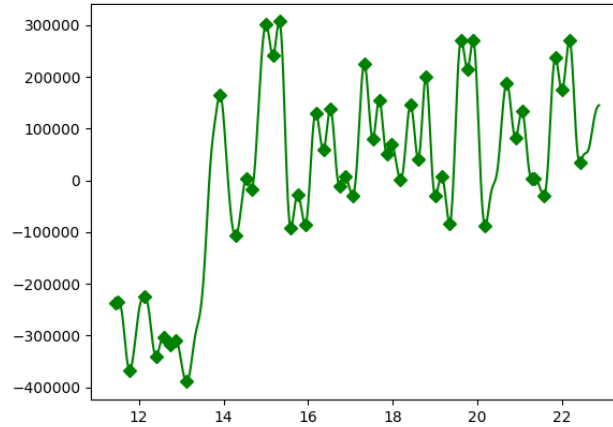
Figura 9. Gráfico acelerômetro - frequência (Hertz) X tempo (segundos)



Fonte: elaboração própria

Usou-se um filtro Gaussiano que é um operador de convolução que é usado para suavizar imagens e remover ruídos. Além do sinal o filtro gaussiano utiliza como entrada um operador de desvio padrão sigma. Quanto maior o sigma, mais a imagem será suavizada. A imagem da Figura 10 mostra o mesmo sinal com o filtro gaussiano.

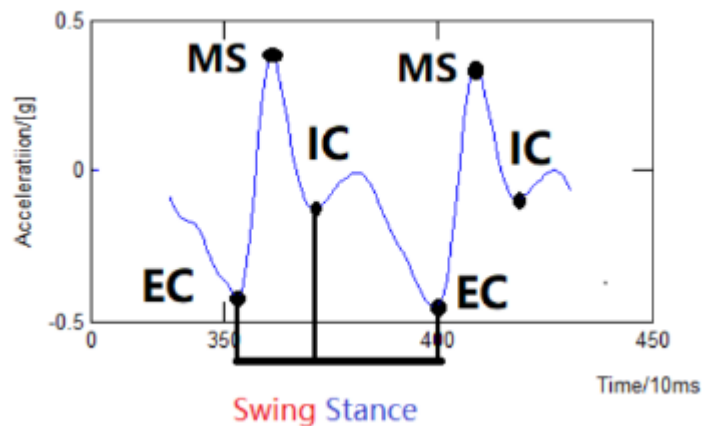
Figura 10. Gráfico acelerômetro com filtro gaussiano - frequência (Hertz) X tempo (segundos)



Fonte: elaboração própria

O pré-processamento do sinal dos sensores permitiu prepará-los para detecção de picos e vales relacionados a diferentes eventos da marcha. Os vales representam o contato inicial (IC) e contato final (EC), os picos representam o balanço médio (MS). A figura 11 mostra a distribuição desses pontos.

Figura 11. IC, EC e MS no sinal da marcha humana



Fonte: Wang *et al.* (2017)

Cada ciclo de marcha inclui uma fase de apoio e uma fase de balanço. Como mostrado na figura 11 (o eixo y representa a aceleração e o eixo x representa o

tempo em milissegundo), os pontos de maior aceleração representam apoio médio(MS). Os pontos medianos determinam o contato inicial (IC). Pode separar também a fase de balanço e apoio pelos pontos inferiores que apresentam o fim do contato (EC). Com o uso do filtro gaussiano possibilitou uma visualização mais apurada do EC, já que eles encontram-se muito próximos.

Houve também pré-processamento dos dados clínicos e fisioterapêuticos antes de serem inseridos no sistema de predição. Os dados que foram coletados de um arquivo do tipo “.csv” passaram por uma padronização de escala. A centralização e o dimensionamento acontecem independentemente em cada recurso, calculando as estatísticas relevantes das amostras no conjunto de treinamento.

A padronização de um conjunto de dados é um requisito comum para muitos estimadores de aprendizado de máquina: eles podem se comportar mal se os recursos individuais não se pareçam mais ou menos como dados padrão normalmente distribuídos.

3.3 Processamento

Após a coleta de dados e o pré-processamento dos mesmos, os dados estavam prontos para serem inseridos no sistema preditivo. Esse sistema consistiu no uso de florestas de árvores de decisão. Esta técnica que pode também ser usada para problemas de classificação.

Antes de inseri-los na predição, os dados pré-processados foram divididos em dados para treino e dados para teste. Sendo 75% deles para uso de treino, onde o sistema principal aprende com essas informações, definindo os parâmetros do algoritmo. E 25% para teste, que permite verificar se o aprendizado aconteceu de fato e se o sistema alcançou o objetivo proposto.

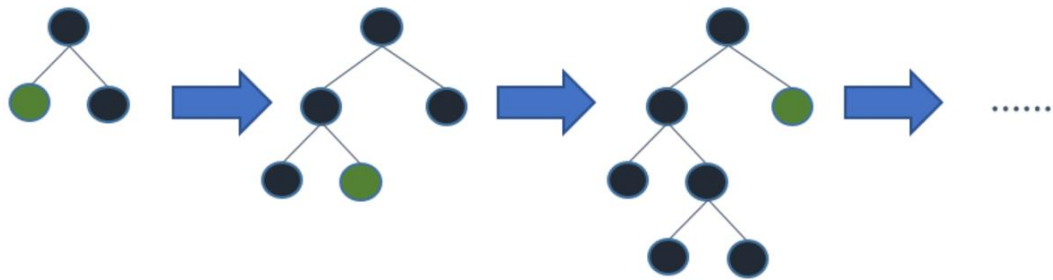
Para o sistema preditivo foi usado o lightGBM que consiste em um *framework* de aprimoramento de gradiente que usa algoritmos de aprendizado baseados em árvore. É um software que possui vantagens como:

- Diminuição no uso de memória;

- Melhor precisão;
- Suporte de aprendizagem paralela e GPU(Unidade de Processamento Gráfico);
- Velocidade de treino mais rápida e mais eficiente;
- Capaz de lidar com grandes bases de dados;

Este *framework* permitiu o uso de 4 algoritmos de florestas de árvores de decisão e diferente de outros algoritmos faz com que as árvores cresçam em termos de folha de árvore. Esta biblioteca, que está implementada em *Python*, oferece vários parâmetros que o usuário pode alterar, a fim de adequação ao problema.

Figura 12. LightGBM – Crescimento das folhas nas árvores de decisão



Fonte: extraído de Mandot *et. al.* (2018)

A configuração dos parâmetros foi necessária pois esse *framework* é indicado para bases de dados com mais de 10.000 linhas. Como nossa base é pequena, os parâmetros devem ser escolhidos cuidadosamente a fim de evitar *overfitting* (quando um modelo estatístico se ajusta muito bem ao conjunto de dados anteriormente observado, mas se mostra ineficaz para novos resultados).

Alguns parâmetros foram comuns para todos os tipos de algoritmos que o *framework* implementa:

- *Learning-rate*: 0.01 – Fator que determina o impacto de cada árvore no resultado final. O LightGBM funciona começando com uma estiva inicial que é atualizada usando a saída de cada árvore, segundo Mandot *et. al.* (2017).

- *Application: regression* – Este é um dos parâmetros mais importante pois indica o tipo de modelo do problema. Neste caso, foi usado o *regression*, pois para o problema de predição deve ser empregado este tipo de aplicação.
- *Task: predict* – Especifica a tarefa que você deseja executar nos dados.
- *Max_depth: 10* – Este parâmetro determina o máximo de profundidade da árvore. É um parâmetro importante para evitar *overfitting*.
- *Num_leaves: 31* – Número de folhas na árvore completa.
- *Bagging_fraction: 1* – Determina a fração de dados a ser usada em cada interação.
- *Min_data: 1* – Número mínimo de dados em uma folha. Para base dados pequenas, o valor deve ser igual a 1 para evitar *overfitting*.

O primeiro tipo de floresta de árvores de decisão foi o GBDT (Árvore de decisão com aumento de gradiente), esse tipo de algoritmo é considerado padrão. Tem a característica de convergir em um tempo razoável.

O segundo tipo é denominado como *random forest* ou floresta randômica é um tipo de algoritmo que adiciona aleatoriedade extra ao model, quando está criando árvores. Ao invés de procurar pela melhor característica ao fazer a partição de nodos, ele busca a melhor característica em um subconjunto aleatório das características. Este processo cria uma grande diversidade, o que geralmente leva a melhores modelos.

Outro tipo de algoritmo implementado pelo *framework* é o DART (Descarte aditivo de árvores de regressão) que usa a filosofia de descarte de árvores. Ao invés de desenvolver a próxima árvore a partir de resíduos de árvores anteriores, desenvolve a próxima árvore a partir do resíduo de uma amostra de árvores anteriores. O modelo ganha mais pois os componentes individuais são forçados a

serem mais auto-suficientes. Alguns parâmetros a mais precisaram ser adicionados para esse tipo de algoritmo, são eles:

- *Drop_rate*: 0.5 – É uma taxa que indica a fração de árvores anteriores que deverão ser abandonadas.
- *Max_drop*: 10 – Número máximo de árvores derrubadas durante uma iteração.
- *Skip_drop*: 0.4 – Probabilidade de pular o procedimento de abandono durante uma iteração de reforço.
- *Drop_seed*: 1 – Semente aleatória escolhida para abandono.
- *Uniform_drop*: *False* : Parâmetro que determina se pode ter descarte de árvores de forma uniforme. Nesse caso, selecionamos para que o descarte não fosse feito de maneira uniforme.
- *Boosting_type*: *dart* – Indicando o tipo de algoritmo que desejamos implementar.

O último algoritmo a ser testado foi o GOSS (*Gradient-based One-Side Sampling*). Esta metodologia pode alcançar um bom equilíbrio entre reduzir o número de instâncias de dados e manter a precisão das árvores de decisão aprendidas. Alguns parâmetros que são próprios desse algoritmo e foram aplicados no modelo:

- *Boosting_type*: *goss* – Indicando o tipo de algoritmo a ser testado.
- *Top_rate*: 0.2 – Taxa de retenção dos gradientes
- *Other_rate*: 0.1 – Taxa de retenção de dados para gradientes pequenos.

3.4 Avaliação do sistema

Após o desenvolvimento do sistema de predição é necessário definir métricas de avaliação. Para regressão há algumas métricas que auxiliam nos resultados comparativos. Esses cálculos avaliam os resultados da amostra de teste com os resultados do preditor. A seguir, são descritos os métodos de avaliação.

3.4.1 Erro Quadrático Médio (EMQ)

Representa uma forma de avaliação usado em predição, o erro quadrático médio é definido como sendo a média da diferença entre o valor estimador e do parâmetro ao quadrado, segundo Takacs *et al.* (1985). O EMQ é determinado somando os erros de previsão ao quadrado e dividindo pelo número de erros usados no cálculo. Pode ser expresso pela equação 2.

Equação 2. Erro Quadrático Médio

$$EQM = \frac{\sum_{t=1}^n e_t^2}{n}$$

Onde:

e_t = Erro no período t . Representa a diferença entre o valor real e a previsão.

n = Número de períodos usados

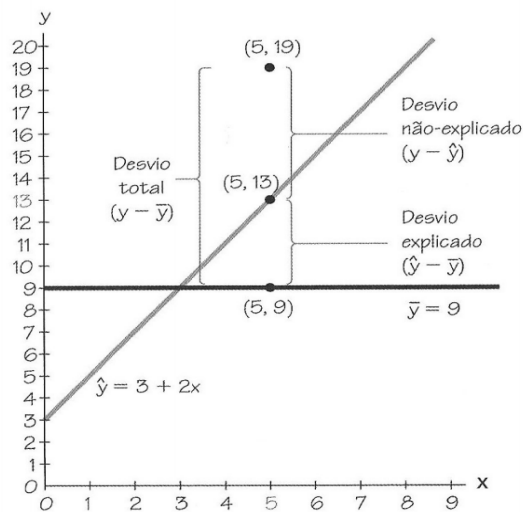
3.4.2 Variação explicada

Outro método de avaliação para predição é a variação explicada. Para entendê-la, alguns conceitos são necessários. Suponha que tenhamos um conjunto de pares de dados com o ponto amostral (x,y) , que \hat{y} seja o valor previsto de y , obtido pela regressão. E que a média dos valores da amostras de y seja \bar{y} .

- Desvio total de (x,y) : É a distância entre um ponto (x,y) e a reta da função que caracteriza o problema.

- Desvio explicado de (x,y) : É a distância $(\hat{y} - \bar{y})$ entre o valor previsto de y e a reta horizontal que passa pela média amostral.
- Desvio não-explicado de (x,y) : É a distância $(y - \hat{y})$ entre o ponto e a reta de regressão.

Figura 13. Exemplo de variância explicada aplicada a uma função



Fonte: < <http://www.ernestoamaral.com/docs/est-122/Aula09.pdf>>. Acesso em: 1 dez. 2018

Observando a figura anterior, podemos observar que os pontos $(5,9)$ e $(5,19)$ estão fora da reta. O ponto $(5,9)$ está dentro da média amostral. A variância explicada é a soma dos quadrados dos valores do desvio explicado.

Capítulo 4

Testes e Resultados

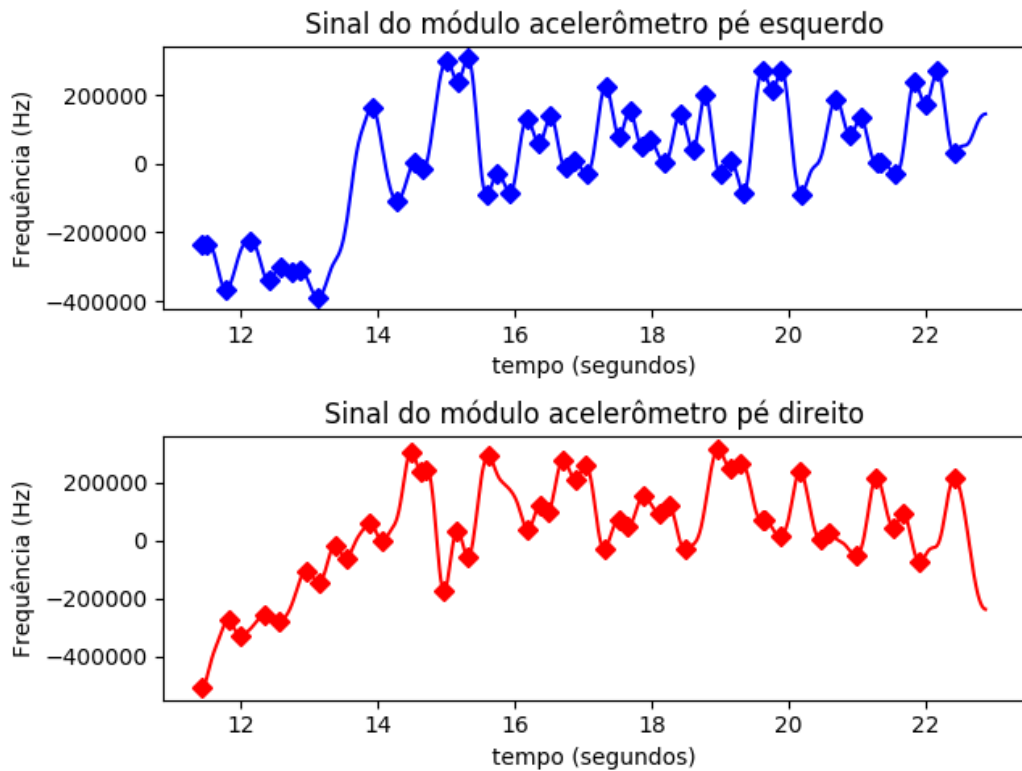
Este capítulo apresenta o processo de realização dos testes e dos resultados obtidos. Também mostra desde o tratamento dos dados, tanto para os dados clínicos quanto os dados coletados por sensores, mostrando também a aplicação do sistema de predição. Por fim, serão apresentados os resultados obtidos.

4.1 Extração de dados temporais da marcha

Os dados coletados dos sensores permitem obter informações de raios de simetria da marcha humana. Esses dados passaram antes por um pré-processamento como mencionado no capítulo anterior, permitindo a visualização de um ciclo completo de marcha e após isso foram extraídos informações. Os valores encontrados constituem entrada para o sistema preditivo.

Durante a análise do sinal, observou-se algumas características que auxiliaram na identificação dos pontos de EC, IC e MS. Constatou-se que o eixo Zs dos sensores localizados nos pés oferecem melhor visibilidade do sinal. Além disso, foram sinalizados os picos e vales para identificar as fases da marcha, como na figura 14.

Figura 14. Gráfico dos sensores dos pés, no eixo z e velocidade de 1,0 m/s



Fonte: elaboração própria

Com a identificação de um ciclo de marcha pela observação dos pontos que se assemelham ao contorno descrito na figura 7, foi possível encontrar os ciclos de marcha dos outros sensores. Os sensores foram analisados aos pares a fim de se obter os valores de raios de simetria que serão comentados mais a frente.

Figura 15. Gráfico dos sensores das pernas, no eixo z na velocidade de 1,0 m/s

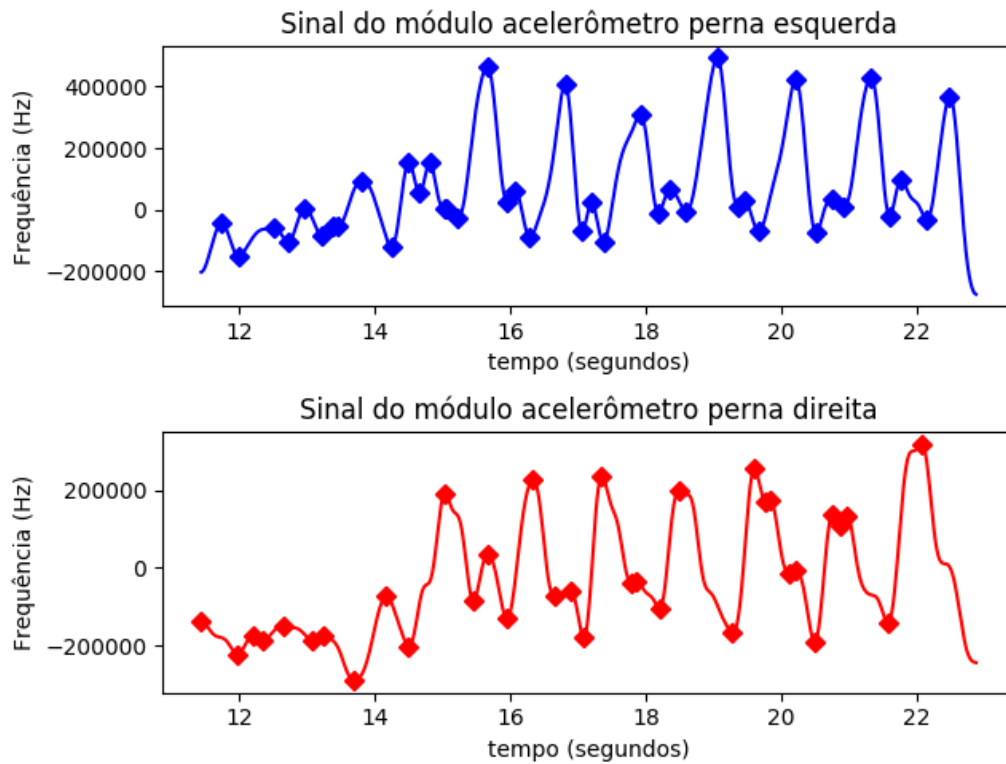
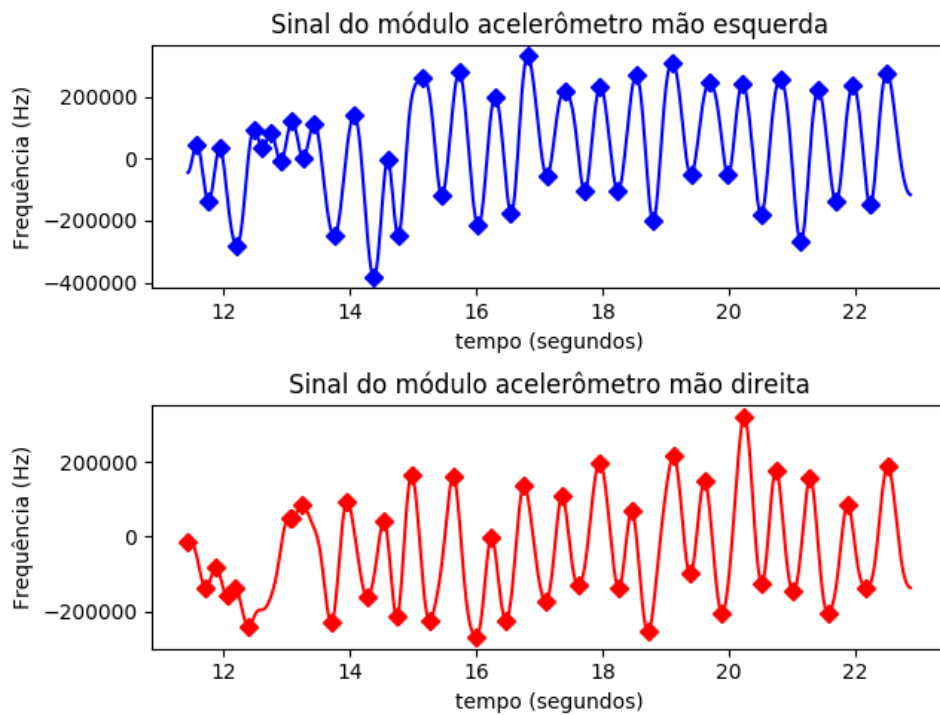


Figura 16. Gráfico dos sensores das mãos, no eixo z e na velocidade de 1,0 m/s



Fonte: elaboração própria

Como mostrado nas figuras 14, 15 e 16, o vale inferior determina o EC e o vale superior aponta o IC. A distância de um ponto de início de contato para um ponto de fim de contato obtem-se o tempo de apoio e a distância de um ponto de fim de contato para um ponto de início de contato corresponde ao tempo de balanço. Desde que o movimento é alternado entre as pernas, capa pico corresponde a um membro inferior. A partir dessas constatações pode-se obter valores mais específicos da marcha.

Os parâmetros temporais como são chamados representam algumas das variáveis que podem ser calculadas a partir do sinal. São eles:

- Tempo de apoio médio (T_{st}): Valor médio de todo o tempo de apoio de uma amostra. T_{stL} e T_{stR} (siglas cujas últimas letras representam as palavras em inglês *Left* e *Right*) representam o valor médio das pernas esquerda e direita, respectivamente.
- Tempo médio de balanço (T_{sw}): Valor médio de todo o tempo de balanço de uma amostra. T_{swL} e T_{swR} representam o valor médio das pernas esquerda e direita, respectivamente.

O valor médio de cada parâmetro temporal encontra-se na tabela 1, a seguir:

Tabela 1 - Valor médio dos 6 parâmetros temporais em segundos

	Média (segundos)	Desvio Padrão (segundos)
T _{st}	0.63	0.051
T _{sw}	0.52	0.040
T _{stL}	0.45	0.047
T _{stR}	0.52	0.051
T _{swL}	0.51	0.049
T _{swR}	0.52	0.050

A partir dos valores obtidos, as 5 variáveis de entrada para o sistema preditivo puderam ser calculadas na tabela 2, são elas:

- Razão do tempo de apoio (R_{st}):

$$R_{st} = T_{stR} / T_{stL} ;$$

- Razão de balanço temporal (R_{sw}):

$$R_{sw} = T_{swR} / T_{swL} ;$$

- Razão entre as fases de apoio e balanço (R_{ss}) para cada membro:

$$R_{ssR} = T_{stR} / T_{swR};$$

$$R_{ssL} = T_{stL} / T_{swL};$$

- Razão geral de simetria (R_s):

$$R_s = R_{ss} / R_{ssL};$$

Tabela 2 - Valor médio dos raios de simetria

	Média
R_{st}	1.15
R_{sw}	1.01
R_{ssR}	1.00
R_{ssL}	0.88
R_s	0.98

4.2 Regressão

Finalizados a extração e o pré-processamento dos dados e em posse dos resultados obtidos, seguimos para a análise de relação entre as variáveis de saída calculadas pelo sistema preditivo e os valores do conjunto de testes. Essa análise tem como objetivo avaliar o aprendizado do algoritmo.

O sistema de predição desenvolvido foi baseado no conceito de floresta de árvores de decisão. Para construção desse sistema foram utilizados quatro algoritmos com metodologias diferentes. Foram eles: GBDT, RF, DART e GOSS. Como mostrado no capítulo dois, esses algoritmos diferem entre si, com relação ao modo da remoção de árvores, do aprendizado de cada árvore e de como as árvores são distribuídas.

As análises foram feitas coletando dados de 39 pacientes atendidos no setor de traumatologia e ortopedia de um centro de reabilitação público na cidade de Recife. Desses pacientes que estavam em atendimento, 23 do sexo masculino e 16 do sexo feminino, e com idades idades entre 22 e 80 anos, com uma média de idade de 49,36 anos e desvio padrão de 13,16.

Foram coletados os seguintes diagnósticos :

Tabela 3 - Diagnósticos dos pacientes em atendimento

Diagnóstico	Número de pacientes	Média do número de sessões	Desvio padrão
Reconstrução pós-operatória do LCA	4	14,75	4,50
Gonartrose	3	8,30	1,52
Torção de tornozelo	4	9,75	1,25
Artroplastia total do joelho	3	26	8,71
Fascite plantar	3	12,70	4,16

Reconstrução pós-operatória do LCA + menisco	3	14,70	5,51
Artrodese pós-operatório do tornozelo	2	15	7,07
Fratura tibial pós-operatório	2	15	7,07
Pós-operatório de tumor de osso no joelho	2	26,50	28,99
Artrofibrose do joelho	1	10	0
Artroplastia total do quadril	1	17	0
Esporão do calcâneo	1	20	0
Fratura fibular + Ruptura do ligamento lateral	1	16	0
Menisco medial – Tratamento conservador	1	8	0
Pós-operatório do menisco	3	22,34	6,42
Reconstrução parcial pós-operatório do tendão patelar	1	20	0
Tendinopatia fibular pós-operatório	1	40	0
Tenoplastia do tendão de Aquiles	1	36	0

Os testes foram realizados com a inserção dos dados de cada paciente. Esses dados foram extraídos do formulário do apêndice B, 75% deles foram usados para treinos dos algoritmos e 25% para realizar os testes de predição. Aplicando-se o erro quadrático médio e a variação explicada para avaliação dos algoritmos de previsão, resultaram nos seguintes valores:

Tabela 4 - Avaliação dos algoritmos

Algoritmos	Erro quadrado médio	Variação explicada
GBDT	18,80%	0.66
RF	20,40%	0.58
DART	65,83%	0.120
GOSS	17,47%	0.69

Os valores de predição foram encontrados de acordo com a melhor combinação de parâmetros para cada algoritmo. As informações da tabela 4 indicam o grau de relevância para o problema proposto.

Diante dos algoritmos testados, o GOSS, apresentou menor taxa de erro médio quadrático e melhor acurácia da predição. Isso indica que o algoritmo tem uma taxa de acerto com mais de 85% na predição de sessões fisioterapêuticas. Enquanto o RF apresentou menor taxa de variância explicada.

Não foram encontrados projetos que possuem o mesmo objetivo de prever o número de sessões fisioterapêuticas. Logo não há como fazer uma comparação direta das metodologias usadas. Contudo, a facilidade de acesso aos sensores inerciais, pode trazer mais pesquisas nesta área.

Alguns trabalhos como o de Ma & Liao (2017), que buscam modelar a marcha humana através de sensores usando o algoritmo do Semi Processo de Markov, podem servir de base para sistemas preditivos. Uma vez que os autores desse projeto visam obter o estado da marcha com a probabilidade do estado atual para identificar patologias no padrão de marcha, acredita-se ser possível usar a mesma metodologia para criação de um sistema preditivo.

Capítulo 5

Considerações Finais

Este capítulo descreve as conclusões acerca do desenvolvimento desta monografia na seção 5.1 e as sugestões para trabalhos futuros para dar continuidade ao projeto seguem na seção 5.2.

5.1 Conclusão

Este trabalho foi elaborado com o intuito de montar um sistema preditivo para a quantidade de sessões fisioterapêuticas, a partir de dados coletados de sensores inerciais e por fichas técnicas de pacientes atendidos no hospital do IMIP, do setor de traumatologia ortopedia. Tendo como objetivo auxiliar o profissional de fisioterapia na determinação do número de sessões, de acordo com o diagnóstico.

A extração dos dados a partir dos sensores inerciais permitiram adquirir informações específicas da marcha que poderão ser integradas com os dados clínicos e fisioterapêuticos. O profissional da área poderá ter uma ferramenta de baixo custo e que mostra detalhes, as diferenças entre a simetria de uma marcha padrão e a simetria de uma marcha deficiente.

O algoritmo de floresta de árvores de decisão escolhido é altamente parametrizável, isso fez com que fosse possível o teste com várias configurações de floresta de árvore de decisão. Além disso, o grande número de combinações possíveis para a entrada do sistema com a análise de correlação feita com os dados extraídos e calculados fez com que o modelo pudesse ser testado de várias formas possíveis, auxiliando assim na sua otimização.

Mesmo com o número reduzido de pacientes para treino do sistema de predição, o método se mostrou bastante promissor. A base de dados possuía alguns dados incompletos e mesmo assim o sistema conseguiu suprir essa falta. O

LightGBM trata os dados ausentes ignorando-os durante as divisões e alocando-os posteriormente nos ramos com maiores perdas.

Fica comprovado que o sistema de predição pode ser uma ferramenta eficaz para o profissional de fisioterapia, que não dispõe de muitos recursos. Portanto, sensores inerciais, filtragem de sinal e métodos de Inteligência artificial já conhecidos podem ser de grande ajuda, na prática cotidiana do fisioterapeuta que é a recomendação do número de sessões.

5.2 Trabalhos Futuros

Como trabalhos futuros recomenda-se:

- Utilização dos sensores IMU na coleta de dados, junto com os dados clínicos e fisioterapêuticos. A junção dessas informações pode trazer refinamento para o sistema.
- Comparação do sistema preditivo com o uso de dados dos sensores inerciais e com os dados clínicos separadamente a fim de observar quais dados são mais relevantes.
- Aumento no número de exemplos da base de dados dos pacientes.
- Análise comparativa do sistema proposto neste trabalho com outros sistemas preditivos de aprendizagem de máquina. Isso permitirá encontrar um sistema de predição mais preciso.

Bibliografia

- ALMEIDA, Vinicius Mauricio *et al.* **Sensores Inerciais**. Disponível em: <<http://www.decom.ufop.br/imobilis/sensores-inerciais/>> (2014). Acesso em: 22 dez. 2018.
- CALDAS, Rafael *et al.* A systematic review of gait analysis methods based on inertial sensors and adaptive algorithms. **Gait & posture**, v. 57, p. 204-210, 2017.
- CALDAS, R. LIMA NETO, F. B. An application of self-organizing maps clustering to support kinematic gait analysis and evaluation of gait symmetry. **J. Biomed. Health Inf.(Submitted for publication)**, 2018.
- DAS, Sandhitsu R. *et al.* Two-stage PCA extracts spatiotemporal features for gait recognition. **Journal of multimedia**, v. 1, n. 5, p. 9-17, 2006.
- GEORGE, Anju Susan *et al.* An Efficient Gait Recognition System for Human Identification using Neural Networks. 2017.
- GONG, Rong *et al.* Acoustic Scene Classification by Fusing LightGBM and VGG-net Multichannel Predictions. **IEEE AASP Challenge on Detection and Classification of Acoustic Scenes and Events**, 2017.
- JESUS, Edison O.; COSTA JR, Roberto. A utilização de filtros gaussianos na análise de imagens digitais. **Proceeding Series of the Brazilian Society of Computational and Applied Mathematics**, v. 3, n. 1, 2015.
- KE, Guolin *et al.* Lightgbm: A highly efficient gradient boosting decision tree. In: **Advances in Neural Information Processing Systems**. 2017. p. 3146-3154.
- KFURI JUNIOR, Mauricio. O trauma ortopédico no Brasil. **Rev. bras. ortop.**, São Paulo , v. 46, supl. 1, 2011 . Disponível em: <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0102-36162011000700003&lng=en&nrm=iso>. Acesso em 25 nov. 2018.
- LATHI, B. P. Linear Systems and Signals. 2. ed. [S.I.]: Oxford University Press, 2004.
- LEITE, Ana Karla Oliveira *et al.* Desenvolvimento de um sistema de simulação de marcha em hipogravidade. 2017.
- LUCAS, LC de S. Árvores, Florestas e Sua Função Como Preditores: Uma Aplicação na Avaliação do Grau de Maturidade de Empresas. **Revista Brasileira de Pesquisas de Marketing, Opinião e Mídia. ISSN**, v. 9456, 2011.

- MA, Hao; LIAO, Wei-Hsin. Human gait modeling and analysis using a semi-Markov process with ground reaction forces. **IEEE transactions on neural systems and rehabilitation engineering**, v. 25, n. 6, p. 597-607, 2017.
- MAKINO, Koji *et al.* Evaluation method of the gait motion based on self-organizing map using the gravity center fluctuation on the sole. **International Journal of Automation and Computing**, v. 14, n. 5, p. 603-614, 2017.
- MANDOT, Pushkar *et al.* **What is lightGBM, how to implement it ? How to fine tune the parameters?** . Medium. Disponível em: <<https://medium.com/@pushkarmandot/https-medium-com-pushkarmandot-what-is-lightgbm-how-to-implement-it-how-to-fine-tune-the-parameters-60347819b7fc>>. Acesso em: 30 out. 2018.
- MARX, Felipe C. *et al.* Tradução e validação cultural do questionário algofuncional de Lequesne para osteoartrite de joelhos e quadris para a língua portuguesa. **Revista Brasileira de Reumatologia**, 2006.
- O'GRADY, Kevin E. Measures of explained variance: Cautions and limitations. **Psychological Bulletin**, v. 92, n. 3, p. 766, 1982.
- PECCIN, Maria Stella; CICONELLI, Rozana; COHEN, Moisés. Specific questionnaire for knee symptoms-the" Lysholm Knee Scoring Scale": translation and validation into Portuguese. **Acta Ortopédica Brasileira**, v. 14, n. 5, p. 268-272, 2006.
- PRASAD, K. Venkatesh. Fundamentals of statistical signal processing: Estimation theory: by Steven M. KAY; Prentice Hall signal processing series; Prentice Hall; Englewood Cliffs, NJ, USA; 1993; xii+ 595 pp.; ISBN: 0-13-345711-7. 1994.
- PROAKIS, J. G.; MANOLAKIS, D. G. Digital signal processing: principles, algorithms, and applications. [S.l.]: Prentice Hall, 1996.
- PUHLMANN, H. Processamento de Sinais - DSP - Parte 1. 2014. Disponível em: <<https://www.embarcados.com.br/introducao-ao-processamento-digital-de-sinais-dsp-parte-1/>>.
- SABATINI, Angelo Maria; MANNINI, Andrea. **Ambulatory Assessment of Instantaneous Velocity during Walking Using Inertial Sensor Measurements**. **Sensors**, v. 16, n. 12, p. 2206, 2016.
- SANTOS, Caroline P.; VIEIRA, Mário EM; JR, Sérgio L. Stevan. **Sensores inerciais aplicados à marcha humana no esporte**, 2016.

- SHAIKHINA, Torgyn *et al.* Decision tree and random forest models for outcome prediction in antibody incompatible kidney transplantation. **Biomedical Signal Processing and Control**, 2017.
- TAKACS, L. A Two-Step Scheme for the Advection Equation with Minimized Dissipation and Dispersion Errors. **Monthly Weather Review**, v. 113, p. 1050-1065, 1985.
- TORRES, Henrique. Sensores Inerciais – Parte 1: A complexidade por trás do desenvolvimento de sensores inerciais. 2014. Disponível em <<http://www.embarcados.com/sensores-inerciais-parte-1>>. Acesso em: 29 nov. 2018.
- WANG, Zhao. Manual gait events detection and adaptative algorithms on gait analysis based in inertial-sensors data. **RWTH Aachen University Germany**. 2018.
- WILLMOTT, Cort J.; MATSUURA, Kenji. Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. **Climate research**, v. 30, n. 1, p. 79-82, 2005.
- YANG, Cheng *et al.* Autonomous gait event detection with portable single-camera gait kinematics analysis system. **Journal of Sensors**, v. 2016, 2016.
- YAZDI, Navid; AYZAZI, Farrokh; NAJAFI, Khalil. Micromachined inertial sensors. **Proceedings of the IEEE**, v. 86, n. 8, p. 1640-1659, 1998.

Apêndice A - Termo de consentimento livre e esclarecido

TERMO DE DISPENSA DO TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO

Ao Comitê de Ética em Pesquisa em Seres Humanos da FPS

Projeto Intitulado:

Perfil sociodemográfico e clínico dos usuários atendidos no serviço de Traumatologia e Ortopedia de um Centro de Reabilitação público na cidade de Recife-PE

Senhor Coordenador,

Em atendimento a este Comitê de Ética em Pesquisa referente ao projeto em epígrafe, solicitamos a dispensa do Termo de Consentimento Livre e Esclarecido, considerando que a coleta em banco de dados já existentes (prontuários) não apresentando risco direto aos seus participantes. Ademais, as informações colhidas nos devidos bancos de dados serão mantidas em sigilo pelos pesquisadores responsáveis pela pesquisa. Colocamo-nos à disposição para esclarecimentos, caso seja necessário.

Nestes termos, nos comprometemos a cumprir todas as diretrizes e normas reguladoras descritas na Resolução nº 466 de 12 de dezembro de 2012, referentes às informações obtidas com Projeto.

Recife, 02 de Julho de 2018

Atenciosamente,

Pesquisador Responsável

Apêndice B - Ficha de coleta de dados - IMIP

FICHA DE COLETA DE DADOS

Data: ____/____/____

DADOS SOCIODEMOGRÁFICOS

1. Dados Pessoais

Paciente n o : _____

Data de Nascimento: ____/____/____ Idade: _____ Estado Civil: _____

Sexo: () M () F

Escolaridade: _____

Profissão: _____

DADOS CLÍNICOS

2. Antecedentes Pessoais

Hipertensão: () sim () não Asma: () sim () não Diabetes: () sim () não

Cardiopatía: () sim () não

Cirurgia: _____

Prática de atividade física: () sim () Não

- Etiologia: _____

- Diagnóstico: _____

- História de Trauma: () Queda _____

() Acidente com veículo _____

() Outro _____

- Tratamento anterior: () Não () Sim

16

() Clínico: _____ () Fisioterapia: _____

() Cirurgia: _____ () Outros: _____

- Tratamento Cirúrgico: () Não () Sim Qual: _____

- Tratamento conservador: () Não () Sim

DADOS FISIOTERAPÊUTICOS

- EVA: _____

- ADM: Quadril: _____

Joelho: _____

Tornozelo: _____

- Perimetria: Coxa: _____

Panturrilha: _____

- Força Muscular: _____

- Questionários/Escalas Funcionais: _____

- Duração do tratamento: _____

- Número de sessões: _____

- Abandono: () Não () Sim Motivo: _____

- Tempo de alta: _____

- Objetivos terapêuticos:

Fase Inicial: _____

Fase Intermediária: _____

Fase Final: _____

- Conduta terapêutica:

1ª- 8ª sessões: _____

9ª - 16ª sessões: _____

17ª - Fim: _____

- Evolução: _____
