UNIVERSIDADE FEDERAL DO RIO GRANDE DO NORTE
CENTRO DE TECNOLOGIA
DEPARTAMENTO DE ENGENHARIA BIOMÉDICA

Trabalho de Conclusão de Curso

UTILIZAÇÃO DA TEORIA DE RESPOSTA AO ITEM COMO MÉTRICA DE AVALIAÇÃO DE DECODIFICADORES PARA INTERFACES HOMEM-MÁQUINA

Jade Marina Dias Gomes

Natal/RN
2022

UNIVERSIDADE FEDERAL DO RIO GRANDE DO NORTE
CENTRO DE TECNOLOGIA
DEPARTAMENTO DE ENGENHARIA BIOMÉDICA

UTILIZAÇÃO DA TEORIA DE RESPOSTA AO ITEM COMO MÉTRICA DE AVALIAÇÃO DE DECODIFICADORES PARA INTERFACES HOMEM-MÁQUINA

Trabalho de conclusão de curso apresentado ao Departamento de Engenharia Biomédica da Universidade Federal do Rio Grande do Norte para obtenção do título de Graduada em Engenharia Biomédica.

Graduando: Me. Jade Marina Dias Gomes
Orientador: Dr. Ernano Arrais Junior
Coorientador: Dr. Abner Cardoso Rodrigues Neto

Natal/RN
2022

Universidade Federal do Rio Grande do Norte - UFRN
Sistema de Bibliotecas - SISBI
Catalogação de Publicação na Fonte. UFRN - Biblioteca Central Zila Mamede

Gomes, Jade Marina Dias.
Utilização da teoria de resposta ao item como métrica de avaliação de decodificadores para interfaces homem-máquina / Jade Marina Dias Gomes. - Natal, 2022.
54f.: il.

Monografia (graduação) - Universidade Federal do Rio Grande do Norte, Centro de Tecnologia, Curso de Engenharia Biomédica. Natal, RN, 2022.
Orientador: Dr. Ernano Arrais Junior.
Coorientador: Dr. Abner Cardoso Rodrigues Neto.

1. Aprendizagem de Máquina - Monografia. 2. Interface Homem-Máquina - Monografia. 3. Eletromiografia - Monografia. 4. Teoria de Resposta ao Item - Monografia. 5. Calibração - Monografia. I. Junior, Ernano Arrais. II. Neto, Abner Cardoso Rodrigues. III. Título.

RN/UF/BCZM CDU 004.8

Elaborado por MARJORIE ROSIELLE SILVA DO AMARAL - CRB-15/351

UNIVERSIDADE FEDERAL DO RIO GRANDE DO NORTE
CENTRO DE TECNOLOGIA
DEPARTAMENTO DE ENGENHARIA BIOMÉDICA

UTILIZAÇÃO DA TEORIA DE RESPOSTA AO ITEM COMO MÉTRICA DE AVALIAÇÃO DE DECODIFICADORES PARA INTERFACES HOMEM-MÁQUINA

Banca Examinadora do Trabalho de Conclusão de Curso:

Orientador(a): Prof. Dr. Ernano Arrais Junior
Universidade Federal do Rio Grande do Norte - UFRN

Coorientador(a): Prof. Dr. Abner Cardoso Rodrigues Neto
Instituto Santos Dumont - ISD

Prof. Dr. Hertz Wilton de Castro Lins
Universidade Federal do Rio Grande do Norte - UFRN

Prof. Me. Amauri Marcos Costa de Morais Júnior
Universidade Federal do Rio Grande do Norte - UFRN

Prof. Karolayne Santos de Azevedo
Instituto Federal de Educação, Ciência e Tecnologia do Rio Grande do Norte - IFRN

Dedico este trabalho às crianças adultas que,
quando pequenas, sonharam ser cientistas.

Agradecimentos

Agradeço primeiramente às instituições que me forneceram o conhecimento e as ferramentas fundamentais para este trabalho: a Universidade Federal do Rio Grande do Norte (UFRN) e o Instituto Internacional de Neurociências Edmond e Lilly Safra (IINELS).

Agradeço ao meu orientador Ernano Arrais pela orientação acadêmica e conselhos pessoais ao longo da minha trajetória no TCC e no curso de maneira geral, e também ao meu coorientador Abner Cardoso por me propor esse desafio como temática do trabalho, pela orientação ao longo dos experimentos e análises dos resultados.

Agradeço também aos meus colegas de turma que dividiram comigo todas as maiores angústias e alegrias durante o curso, em especial ao grupo "DEB que lute"formado por Jackson, Epitácio e João Pedro, junto a mim.

Agradeço a Core Engenharia Biomédica por ter proporcionado uma vivência do mercado durante a graduação e por ter ampliado meu horizonte para as inúmeras áreas do mundo corporativo, em especial à melhor gestão formada por Lucas, Gabriela, Jackson, Alicia e Lizandra, junto a mim.

Agradeço às organizações que me deram a oportunidade de experenciar uma das áreas da engenharia Biomédica, a engenharia Clínica durante meu período de estágio: a Clinservice e a SESAP. Em especial as pessoas que fizeram parte dessa jornada: Antônio Moliterno, Victor, Nando, Pinheiro, Jadymilla, Alicia, Dalyne, Antônio Nunes, Wellington, Fransueldo, João Pedro, Márcia, Camila e Laura.

Agradeço a minha família e amigos que estiveram comigo ajudando a suportar todos os desafios da graduação de uma maneira mais leve. Meu muito obrigada a Sílvia, Moacir, Ana Júlia, Yuri, Tâmara, Ábida, Marcela, Marcellus, João Henrique, Victor, Ana Luiza, Bruna, Jesus e Morgana.

E por fim, um agradecimento especial ao meu companheiro Igor de Oliveira por todo o apoio, fosse emocional ou de suporte técnico audiovisual ao longo dos projetos remotos da graduação.

Resumo

As Interfaces Homem-Máquina (IHM) baseadas em sistemas biônicos podem ser alternativas para fornecer comunicação, controle e reabilitação para pessoas com deficiências motoras graves utilizando para isso sinais provenientes da contração muscular para execução de tarefas ou comandos de máquina. Para que o sinal de Eletromiografia (EMG) seja compreendido como sinal de máquina, modelos decodificadores de Aprendizagem de Máquina (AM) fazem o reconhecimento dos padrões, os decodificam e os utilizam para o controle de dispositivos eletrônicos. O ajuste do modelo de decodificação baseado na diferença entre o sinal captado do músculo e a intenção da tarefa é necessário para a calibração da IHM. Entretanto, a calibração dos decodificadores sofre uma grande variabilidade, pois pode haver alteração de pessoa para pessoa e até mesmo para uma mesma pessoa em momentos diferentes de treinamento. Logo, existe um interesse em reconhecer essa variabilidade na calibração e avaliar os decodificadores, permitindo quantificar a variação da habilidade do algoritmo ou da dificuldade da tarefa necessárias para ajustar o decodificador. Nesse contexto, a Teoria de Resposta ao Item (TRI), amplamente aplicada em testes educacionais, e recentemente utilizada para avaliação de modelos de AM, pode ser uma alternativa para reconhecer a variabilidade dos decodificadores e avaliá-los identificando as ações (itens) difíceis, que podem se beneficiar de mais exemplos, isto é, diferenciar os treinamentos IHM entre suficientes e insuficientes na modulação da atividade muscular. A TRI consiste em um conjunto de modelos logísticos que relacionam as respostas dadas aos itens às habilidades de seus respondentes e aos parâmetros do item, como: dificuldade, discriminação e adivinhação. Essa abordagem é interessante para avaliação de modelos decodificadores pois a TRI se adapta muito bem às tarefas de AM, cujos dados (exemplos de treinamento ou teste) correspondem aos itens e os modelos correspondem aos alunos. Portanto, o objetivo deste trabalho é utilizar TRI como método para avaliar algoritmos de decodificação em IHM por níveis de habilidade, e tarefas por níveis de dificuldade. Em seguida, pretende-se realizar a aumentação de instâncias difíceis até que se obtenha um resultado suficiente de calibração em IHM. As vantagens desse método incluem: diminuição do tempo de coleta necessário, quando a decodificação for bem sucedida e a instância for classificada como “fácil”; alerta sobre a necessidade de coletar mais dados quando o item for classificado como “difícil”; e melhor avaliação de decodificadores por níveis de habilidade e proficiência. Para as análises, os dados partem de um banco de dados de sinais de EMG de gestos das mãos.

Palavras-chave: Interface Homem-Máquina; Eletromiografia; Calibração; Teoria de Resposta ao Item; Aprendizagem de Máquina.

Abstract

*otherlanguage*english Human-Machine Interfaces (HMI) based on bionic systems can be alternatives to provide communication, control and rehabilitation for people with severe motor disabilities, using signals from muscle contraction to perform tasks or machine commands. In order for the Electromyography (EMG) signal to be understood as a machine signal, Machine Learning (ML) decoder models recognize the patterns, decode them and use them to control electronic devices. The adjustment of the decoding model based on the difference between the signal captured from the muscle and the task intent is necessary for the HMI calibration. However, the decoders calibration suffers a great variability, as it can change from person to person and even to the same person at different times of training. Therefore, there is an interest in recognizing this variability in the calibration and evaluating the decoders, allowing to quantify the variation of the algorithm's skill or the difficulty of the task necessary to adjust the decoder. In this context, the Item Response Theory (IRT), widely applied in educational tests, and recently used to evaluate ML models, can be an alternative to recognize the decoders variability and evaluate them by identifying the difficult actions (items), which can benefit from more examples, i.e., to differentiate between sufficient and insufficient HMI training in modulating muscle activity. The IRT consists of a set of logistic models that relate the answers given to the items to the abilities of their respondents and to the item parameters, such as: difficulty, discrimination and guessing. This approach is interesting for the evaluation of decoding models because the IRT adapts very well to ML tasks, whose data (training or test examples) correspond to the items and the models correspond to the students. Therefore, the objective of this work is to use TRI as a method to evaluate decoding algorithms in HMI by skill levels, and tasks by difficulty levels. Then, it is intended to perform the augmentation of difficult instances until a sufficient result of calibration in HMI is obtained. The advantages of this method include: decreased collection time required, when decoding is successful and the instance is classified as "easy"; alert about the need to collect more data when the item is classified as "difficult"; and better evaluation of decoders by skill and proficiency levels. For the analyses, the data are obtained from the hand gesture EMG signal database.*

***Keywords**: Human-Machine Interface; calibration ; Item Response Theory; Machine Learning*

Lista de ilustrações

Figura 1 – Sinal de EMG 19
Figura 2 – Duas técnicas de eletromiografia 20
Figura 3 – Unidade motora e a superposição de sinais 21
Figura 4 – Diagrama de blocos para o desenvolvimento de dispositivos assistivos baseados na classificação de sinais EMG 22
Figura 5 – Variáveis de análise de EMG no domínio do Tempo 23
Figura 6 – Variáveis de análise de EMG no domínio da Frequência 24
Figura 7 – Tabela resumo dos trabalhos analisados sobre geração artificial de sinais EMG. 26
Figura 8 – Tabela resumo dos trabalhos analisados sobre TRI em AM. 31
Figura 9 – Tabela resumo dos trabalhos analisados sobre TRI em IHM. 34
Figura 10 – Função Logística Sigmoide Padrão. 38
Figura 11 – Exemplo de uma Curva Característica do Item. Os coeficientes indicam: (a) discriminação do item; (b) dificuldade do item; (c) adivinhação 38
Figura 12 – Gráfico do nível de habilidade pela probabilidade da resposta correta. A Curva Característica do Item é representada em azul e Curva da Informação do Item, em vermelho 40
Figura 13 – Características do sinal 42
Figura 14 – Fluxograma do Experimento 1 43
Figura 15 – Fluxograma do Experimento 2 44
Figura 16 – Dificuldade por Taxa de Aumentação 47
Figura 17 – Dificuldade média por taxa de aumentação 48

Lista de tabelas

Tabela 1 – Coeficientes de TRI para classificação com e sem aumentação. 45
Tabela 2 – Acurácia dos modelos para classificação do conjunto de dados original, conjunto aumentado por discriminação, por dificuldade e por chute. 45

Lista de abreviaturas e siglas

AC	Autocorrelação
AM	Aprendizagem de Máquina
ARV	Average Rectified Value (Valor Retificado Médio)
BC	Classificador Bayesiano
CC	correlação cruzada
CCI	Curva Característica do item
DCGAN	*Deep Convolutional Generative Adversarial Network*
EL	Envelope Linear
EMG	Eletromiografia
ENEM	Exame Nacional do Ensino Médio
EQM	Erro Quadrático Médio
FL	Lógica Fuzzy
FM	Filtro Mediano
FRI	Função de Resposta do item
HMM	Modelos Ocultos de Markov
IA	Inteligência Artificial
ICM	Interface Cérebro-Máquina
IHM	Interfaces Homem-Máquina
IN	Integração Numérica
INEP	Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira
KNN	K-vizinho mais próximo
LDA	Análise Discriminante Linear
MAV	Mean Absolute Value (valor médio absoluto)
MDF	Median Frequency (Frequência Mediana)

MEC	Ministério da Educação
ML1-4	Modelo Logístico de 1-4 parâmetros
MLP	Perceptron Multicamada
MM	Média Móvel
MNF	Mean Frequency (Frequência Média)
NFB	Neurofeedback
QDA	Análise Quadrática Discriminante
RMS	Root mean square (raiz quadrada média)
RMS	Root Mean Square (Raiz Quadrática Média)
RNA	Redes Neurais Artificiais
SD	Standard Deviation (desvio padrão)
SHAP	SHapley Additive exPlanations
SSC	Slope Sign Changes (alterações do sinal de inclinação)
SVM	Máquinas de vetor de Suporte
TRI	Teoria de Resposta ao Item
VAR	variância
WL	Wavelength (comprimento de onda)
ZC	Zero Crossing (cruzamento de zero)
ZDP	Zona de Desenvolvimento Proximal

Sumário

1 INTRODUÇÃO 14
1.1 Relevância do Tema 14
1.2 Motivação 15
1.3 Objetivos 16
1.3.1 Geral 16
1.3.2 Específicos 16
1.4 Contribuições 17
1.5 Metodologia 17
1.6 Organização do Trabalho 18

2 REVISÃO BIBLIOGRÁFICA 19
2.1 Eletromiografia 19
2.1.1 Processamento do sinal eletromiográfico 21
2.2 Classificação de EMG 22
2.3 Geração de EMG Artificiais 24
2.4 Teoria de Resposta ao Item 26
2.4.1 Teoria de Resposta ao Item em Aprendizagem de Máquina 27
2.4.2 Teoria de Resposta ao Item em Interface Homem-Máquina 30

3 TRI 35
3.0.1 Definição 35
3.0.1.1 Função Logística 35
3.0.1.2 Modelo TRI de 3 Parâmetros 36
3.0.1.3 Modelo TRI de 4 Parâmetros 36
3.0.2 Curva Característica do Item 37
3.0.3 Função de Informação do Item 39
3.0.4 Função de Informação do Teste 39

4 MÉTODO PROPOSTO 41
4.1 Dados 41
4.2 *Design* dos experimentos 41
4.2.1 Experimento 1 42
4.2.2 Experimento 2 43
4.3 Métricas de Desempenho 44
4.3.1 Erro Médio Quadrático 44
4.3.2 Habilidade pela TRI 44

5 RESULTADOS E DISCUSSÃO 45
5.1 Experimento 1 45
5.2 Experimento 2 46

6 CONCLUSÕES 49
6.1 Trabalhos Futuros 49

REFERÊNCIAS 51

1 Introdução

Neste capítulo será feita uma abordagem geral acerca do trabalho proposto apresentando a sua relevância, motivação, objetivos, contribuições, metodologia e organização do trabalho.

1.1 Relevância do Tema

No Brasil, mais de 45 milhões de pessoas possuem algum tipo de deficiência, segundo o Censo de 2010, realizado pelo Instituto Brasileiro de Geografia e Estatística (IBGE). As deficiências motoras, severas ou não, podem advir de diferentes etiologias. As principais causas dos quadros de maior incidência são: malformações congênitas, como luxações; afecções congênitas sistêmicas ou generalizadas do esqueleto; osteocondrose; afecções articulares, como artrose; afecções neuromusculares, como paralisia cerebral; ou outras afecções neurológicas (Nunes *et al.* 1998).

Pensando numa inclusão cada vez maior das pessoas acometidas por algum tipo de deficiência, a tecnologia se insere com o objetivo de trazer maior autonomia para essas pessoas. Denomina-se tecnologia assistiva a área de conhecimento interdisciplinar, que apresenta produtos, recursos, métodos, estratégias, práticas e serviços que promovem a funcionalidade de pessoas com deficiência, incapacidades ou mobilidade reduzida, visando sua autonomia, independência, qualidade de vida e inclusão social (Santos *et al.* 2017).

Para alguns casos, a reabilitação fisioterapêutica aliada a tecnologias de órtese passiva podem ser suficientes para devolver à pessoa com deficiência formas de incluir-se à sociedade. Entretanto, para acometimentos de graus mais elevados, as tecnologias agem ativamente como facilitadoras nesse processo, como por exemplo com o desenvolvimento de dispositivos controlados por sinais biológicos que permitem a mobilidade ou maximizam a capacidade comunicativa desse indivíduo. A capacidade de movimento do indivíduo pode ser adquirida através do controle por outros meios, tais como: oculares, vocais, musculares, visuais e neurais. A tecnologia assistiva que utiliza sinais biônicos (provenientes do corpo humano) para o controle de dispositivos eletrônicos ou sistemas computacionais é chamada de Interface Homem-Máquina (IHM). A IHM visa reconhecer os padrões de sinais biônicos, decodificá-los e usá-los para controle de dispositivos eletrônicos (Cannan e Hu 2011).

A decodificação em IHM é um processo que consiste na utilização de modelos de classificação para reconhecimento de padrões de sinais musculares e associação à um movimento específico. Os atuais sistemas de controle de próteses mais avançados baseiam-se em Aprendizagem de Máquina (AM) para o reconhecimento desses padrões. Esse método exige que os usuários "ensinem" o dispositivo a reconhecer padrões específicos de atividade muscular e os traduz em

comandos – como abrir ou fechar uma mão protética (Huang *et al.* 2008).

O ajuste do modelo de decodificação baseado na diferença entre o sinal captado do músculo e a ação motora do indivíduo é usualmente chamado de calibração da IHM. Entretanto, a calibração dos decodificadores não é um processo fácil, pois sofre grande variabilidade, podendo variar de pessoa para pessoa e até mesmo para uma mesma pessoa em momentos diferentes de registro (Tsinganos *et al.* 2020). Logo, existe um grande interesse científico em reconhecer essa variabilidade na calibração, avaliar decodificadores, e quantificar essa variabilidade da habilidade do algoritmo ou da dificuldade da tarefa necessárias para ajustar o decodificador.

Além disso, a grande variabilidade das calibrações em IHM faz com que sejam necessárias várias coletas de dados e que o tempo de cada coleta seja prolongado. Essa morosidade é desgastante tanto para o profissional responsável pelo treinamento da IHM, quanto principalmente para o usuário da tecnologia (Zanini e Colombini 2020).

Nesse contexto, a Teoria de Resposta ao Item (TRI), amplamente aplicada principalmente em testes educacionais e de avaliação psicométrica (Prudêncio *et al.* 2015), é uma alternativa que já vem sendo usada para compreensão de modelos de Aprendizagem de Máquina (AM) (Martínez-Plumed *et al.* 2016), e que pode ser estendida para identificação de exemplos cuja calibração de IHM é prejudicada. A TRI considera um conjunto de modelos que associam as respostas dadas aos itens com habilidades latentes dos respondentes (Embretson e Reise 2013). Nesse modelo, a probabilidade de uma resposta certa é uma função logística da habilidade do respondente e da dificuldade do item (Chen e Ahn 2020).

Uma maneira de compreender a ideia de TRI para a avaliação de decodificadores e tarefas de IHM é considerar cada dado de exemplo do conjunto de treinamento (uma leitura do potencial mioelétrico) como um item de um questionário como o ENEM (Exame Nacional do Ensino Médio); e cada modelo decodificador como um respondente, isto é, um aluno. Essa abordagem traz vantagens para avaliação e interpretação de modelos, tais como: melhor avaliação das habilidades dos modelos; identificação das instâncias mais difíceis; determinação do nível de proficiência do classificador; e determinação do parâmetro de adivinhação, ou seja, a probabilidade de que um classificador possa classificar corretamente a situação sem realmente aprender as características dos dados de treinamento (Dias *et al.* 2022).

1.2 Motivação

Alguns trabalhos na literatura partem da mesma problemática de variabilidade na calibração de decodificadores (Zanini e Colombini 2020). Uma das estratégias mais utilizadas é a aumentação de dados de Eletromiografia (EMG).

O trabalho de Zanini e Colombini (2020) realiza a aumentação de sinais de EMG para validar abordagens de tratamento da doença de Parkinson em diferentes cenários de movimento.

A metodologia utilizada é aplicação de redes neurais artificiais generativas (DCGAN - Deep Convolutional Generative Adversarial Network) que aprendem as principais características do sinal e conseguem gerar novos com características semelhantes (Zanini e Colombini 2020). Apesar de trazer bons resultados, o DCGAN realiza aumentação considerando todos os dados do sinal, mas para o contexto de calibração em IHM, e pensando também na redução do custo computacional, convém concentrar a aumentação apenas nos itens problemáticos, isto é, nos itens que confundiram os classificadores na calibração. Dessa forma, é possível utilizar TRI como critério para identificar esses itens problemáticos (ações de movimento no treinamento) que podem se beneficiar de mais exemplos.

A probabilidade de acerto em TRI pode ser modelada em função da dificuldade de um item e da capacidade do sujeito (Bauer e Gharabaghi 2015). Supondo que esse último seja invariável, segundo Bauer e Gharabaqui (2015), a dificuldade do item depende quase que exclusivamente do design da tarefa, que inclui: algoritmo de classificação, a estrutura do ensaio, as pistas e quaisquer instruções ou aspectos da coleta, provocando alterações na calibração.

O trabalho de Bauer e Gharabaqui (2015) fornece pistas importantes da aplicação de TRI no campo de IHM, mais especificamente a Interface Cérebro-Máquina (ICM), demonstrando que existe uma relação natural entre a classificação das funções de ICM com as funções de TRI, bem como, que os modelos fornecem informações sobre a capacidade do cérebro em realizar a auto-regulação em uma tarefa de treinamento.

No entanto, até onde se sabe, nenhum trabalho desenvolveu um algoritmo de aumentação de exemplos baseados em TRI para aplicação em IHM. Logo, este trabalho tem por objetivo identificar os casos em que a calibração IHM é difícil, e aumentar através da repetição de dados até que os modelos TRI declarem a calibração como “item fácil”. O trabalho tem relevância tanto na identificação de sucesso das calibrações IHM, quanto na redução de tempo no número de coletas para os itens já considerados suficientes ou fáceis.

1.3 Objetivos

1.3.1 Geral

O objetivo geral deste trabalho é promover a aumentação de dados de treinamento para decodificadores de Interfaces Homem-Máquina baseados na Teoria de Resposta ao Item (TRI) com a finalidade de reduzir o tempo de coleta necessário para calibração em IHM.

1.3.2 Específicos

- Criar alguns modelos decodificadores baseados em AM, usando diferentes parâmetros para um conjunto de dados de IHM;

- Avaliar os decodificadores e os exemplos de treinamento utilizando TRI;
- Interpretar os exemplos de treinamento pela TRI;
- Realizar aumentação pelos critérios da TRI: itens mais discriminantes, mais difíceis e de maior chute;
- Comparar o desempenho de sistemas aumentados com os não aumentados;
- Comparar o desempenho por modelos decodificadores com base em TRI e na acurácia;
- Realizar aumentação progressiva e avaliar seu desempenho.

1.4 Contribuições

Este trabalho traz potenciais contribuições para o campo de estudo em Interface Homem-Máquina, de interpretabilidade de modelos de Aprendizagem de Máquina, e de aplicações da TRI. O trabalho em si trata-se de ensaios experimentais para validação de uma metodologia de aumentação de dados, mas sua aplicabilidade é diversa, como construção de sistemas de aumentação de dados em IHM automatizados, sistemas de alerta para calibrações mal sucedidas com base nos critérios da TRI, ou sistemas de ranqueamento de modelos que mais se aplicam ao conjunto de dados específicos de um paciente ou grupo muscular, entre outros.

Assim, as principais contribuições deste trabalho são:

- Construção de um protocolo de aumentação de dados de sinais de EMG;
- Proposição de uma métrica de avaliação de decodificadores IHM; e
- Discussão de possíveis interpretações de dados de EMG do ponto de vista da Teoria de Resposta ao Item.

1.5 Metodologia

Inicialmente, como primeira etapa do estudo, foi realizada uma revisão bibliográfica acerca dos sinais musculares baseados em Eletromiografia (EMG) e suas características. Em seguida, foi efetuada uma revisão bibliográfica acerca das técnicas de classificação e geração de sinais de EMG artificiais. Uma vez compreendido o cenário da pesquisa, o trabalho foi direcionado a um embasamento teórico a respeito da metodologia da análise proposta, a Teoria de Resposta ao Item (TRI), de forma a contextualizar e estabelecer paralelos com outras aplicações. Uma vez contextualizado o tema, foram desenvolvidas soluções numéricas de forma a reduzir as arestas das limitações encontradas nas pesquisas bibliográficas, melhorando a interpretação comportamental dos sinais de EMG. Desenvolvida e validada a análise baseada em TRI, os

resultados foram utilizados como base para implementação de algoritmos de geração de sinais artificiais de EMG (aumentação de dados). Finalmente, a pesquisa foi documentada, sendo elaborado este Trabalho de Conclusão de Curso (TCC).

1.6 Organização do Trabalho

O trabalho foi organizado e distribuído de acordo com os capítulos abaixo:

- Introdução: neste capítulo o trabalho é contextualizado, são apresentadas suas principais motivações, relevância da pesquisa e do tema, objetivos, contribuições para a sociedade, metodologia e organização do trabalho.
- Revisão Bibliográfica: este capítulo apresentará a revisão bibliográfica acerca das tecnologias e ferramentas voltadas para classificação e aumentação de sinais EMG, com o estado da arte dessa área de pesquisa;
- TRI: neste capítulo serão descritas as aplicações e os modelos matemáticos;
- Método Proposto: neste capítulo serão descritos os métodos dos experimentos realizados para validação das hipóteses, cujos algoritmos serão apresentados em forma de fluxograma para melhor compreensão;
- Resultados e Discussões: neste capítulo os resultados obtidos serão discutidos, com possíveis interpretações;
- Conclusões e Trabalhos Futuros: neste capítulo serão apresentadas as conclusões e possíveis trabalhos que podem derivar deste.

O projeto completo pode ser acessado através do link na nota de rodapé[1]

[1] <https://github.com/jademdias/TCC>.

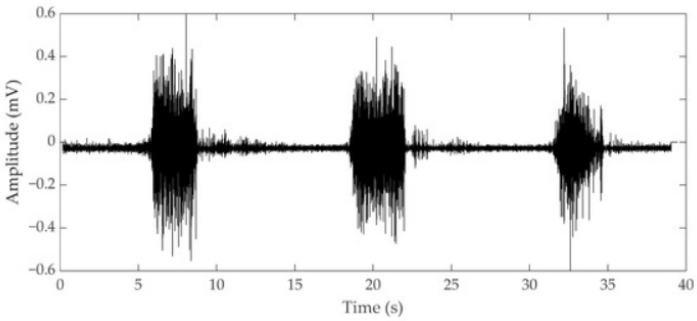
2 Revisão Bibliográfica

Neste Capítulo será apresentado alguns conceitos importantes acerca do sinal biomédico estudado, classificação do sinal, aumentação de dados, e TRI. E em seguida, será apresentado um levantamento de estado da arte acerca das aplicações de TRI em AM e TRI em IHM.

2.1 Eletromiografia

A Eletromiografia (EMG) é uma técnica que permite o registro da atividade elétrica no músculo (Mills 2005), isto é, dos sinais bioelétricos que emanam de um músculo em contração (Luca 2006) gerados pela despolarização das membranas das células musculares (Aparicio 2005). Desde 1940, a EMG tem sido amplamente utilizada para compreender as funções e disfunções do sistema muscular durante o movimento humano (Soderberg e Knutson 2000). Este recurso tem possibilitado pesquisas em diversas áreas de interesse de fisioterapeutas e de outros profissionais que tenham como foco de interesse o movimento humano (Ocarino *et al.*). O uso da EMG nas ciências da reabilitação tem contribuído para a compreensão dos padrões neuromusculares utilizados na execução de diferentes gestos motores. Ajuda também a compreender as possíveis causas e consequências de uma lesão e/ou disfunção no comportamento eletrofisiológico do músculo (Guzmán-Muñoz e Méndez-Rebolledo 2018), pois possibilita estimar a magnitude de uma contração muscular (ativação), o tempo em que o músculo é ativado (latência) e os níveis de fadiga que podem ocorrer durante uma contração (Guzmán-Muñoz e Méndez-Rebolledo 2018). O sensor EMG é um dispositivo que mede os sinais elétricos dos músculos durante esse processo. A Figura 1 apresenta um exemplo de sinal obtido por um sensor EMG representando a contração do músculo bíceps braquial nos intervalos cuja amplitude é maior.

Figura 1 – Sinal de EMG

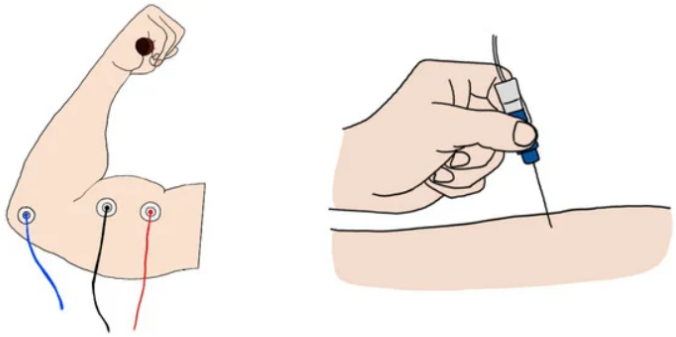


Fonte: Ebashi *et al.* 1969

O sinal eletromiográfico permite a construção do eletromiograma, que corresponde à soma temporal e espacial dos potenciais de ação das unidades motoras (MUAPs - *Motor*

Unit Action Potential) durante a contração, permitindo a quantificação precisa e objetiva do comportamento bioelétrico muscular (Guzmán-Muñoz e Méndez-Rebolledo 2018). Existem duas técnicas de EMG: EMG invasiva e EMG de superfície (Figura: 2). A EMG invasiva consegue obter o registro do potencial de ação gerado por uma determinada unidade motora. Para isso, um eletrodo é inserido por via intramuscular através de agulhas (Guzmán-Muñoz e Méndez-Rebolledo 2018). Já a EMG de superfície permite o estudo da atividade bioelétrica do músculo registrando as diferenças de potencial registradas na superfície da pele. Baseia-se na utilização de eletrodos de superfície com fio ou sem fio que são colocados diretamente na pele do músculo a ser avaliado (Guzmán-Muñoz e Méndez-Rebolledo 2018).

Figura 2 – Duas técnicas de eletromiografia

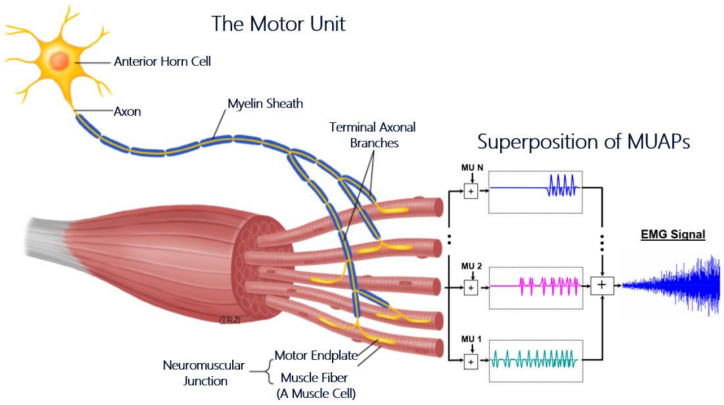


Fonte: Ebashi *et al.* 1969

O sinal eletromiográfico é baseado em potenciais de ação na membrana da fibra muscular que resultam dos processos de despolarização e repolarização na fibra muscular ao ser excitada por um neurônio motor. Após a excitação inicial desta área, o potencial percorre a fibra muscular e em seu caminho passa pela região onde os eletrodos estão localizados. Na EMG de superfície, a separação dos eletrodos é geralmente de 10 mm. Portanto, quando um eletrodo está captando o sinal de despolarização, o outro está registrando um potencial de repouso. Dessa forma, o sinal registrado pela EMG corresponde à diferença de potenciais de ação existentes entre os dois eletrodos (Guzmán-Muñoz e Méndez-Rebolledo 2018) (Figura: 3).

O EMG tem componentes em frequência desde 20Hz até 10kHz e apresenta amplitudes de 100mV até 90mV, dependendo do sinal estudado e do tipo do eletrodo. Em virtude das características do sinal de EMG e da dependência da sua morfologia com a tarefa motora em estudo, o eletromiograma bruto quase não apresenta utilidade no estudo eletrofisiológico. Assim, pesquisas têm sido realizadas na busca por métodos para extração de informações associadas a processos fisiológicos (Ferreira *et al.* 2010).

Figura 3 – Unidade motora e a superposição de sinais



Fonte: Ebashi *et al.* 1969

2.1.1 Processamento do sinal eletromiográfico

Antes de serem analisados, os dados eletromiográficos passam por uma série de processamentos. Pesquisadores desenvolveram métodos de processamento com o objetivo de minimizar interferências de outras fontes que não o campo eletromagnético externo, e de permitir a quantificação do sinal eletromiográfico (Ocarino *et al.*). O processamento de sinais consiste em um grupo de técnicas matemáticas que podem ser aplicadas para extrair informações de sinais, inclusive os de origem biomédica. As ferramentas para processamento podem ser divididas de acordo com o domínio: tempo, frequência ou tempo-frequência (idem) (Ferreira *et al.* 2010). Dentre as diversas formas de análise do sinal eletromiográfico, destacam-se dois parâmetros principais: a amplitude e a frequência do sinal. A partir da amplitude é possível determinar o nível de ativação do músculo ou, especificamente, o percentual de ativação. Este pode ser calculado analisando a média, o pico, a área sob a curva, a raiz quadrada média do sinal (Guzmán-Muñoz e Méndez-Rebolledo 2018); e as variáveis de frequência podem ser obtidas analisando a frequência, média e mediana, cruzamento zero, número de voltas (Guzmán-Muñoz e Méndez-Rebolledo 2018).

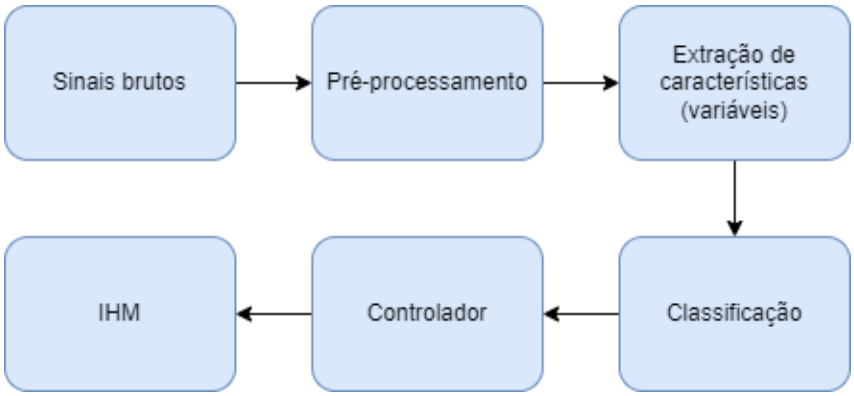
As principais técnicas de processamento descritas na literatura incluem: Raiz Quadrática Média (*Root Mean Square* – RMS), Valor Retificado Médio (*Average Rectified Value* – ARV); Retificação de meia onda e de onda completa; Média Móvel (MM), Filtro Mediano (FM); Envelope Linear (EL); Integração Numérica (IN); Autocorrelação (AC) e correlação cruzada (CC); Normalização em amplitude; Frequência Média (*Mean Frequency* – MNF), Frequência Mediana (*Median Frequency* – MDF) e Índice MNF/MDF; Adição de zeros ao sinal temporal amostrado; Filtração digital (Guzmán-Muñoz e Méndez-Rebolledo 2018)

2.2 Classificação de EMG

A classificação de sinais mioelétricos baseada em reconhecimento de padrões consiste em registrar um intervalo de tempo específico de sinais EMG vindos dos músculos e realizar muitas repetições com diferentes movimentos, para posteriormente segmentá-los. A classificação é realizada extraindo as características em cada intervalo do sinal para reconhecer as informações características de cada movimento (Toledo-Pérez *et al.* 2019). Em 1993, cinco características foram propostas por Hudgins et al.: valor absoluto médio (MAV - *Mean Absolute Value*), inclinação do valor absoluto médio, alterações do sinal de inclinação (SSC - *Slope Sign Changes*), cruzamento de zero (ZC - *Zero Crossing*) e comprimento de onda (WL - *Wavelength*). E em especial para a detecção de movimentos da mão, Ahsan et al. realizaram a extração de características usando MAV, ZC, SSC, raiz quadrada média (RMS - *Root mean square*), variância (VAR) e desvio padrão (SD - *Standard Deviation*) (Ahsan *et al.* 2009).

Geralmente, o processo de desenvolvimento de dispositivos assistivos baseados na classificação de sinais EMG é como o apresentado na Figura 4. Os três módulos principais em cascata consistem no pré-processamento, na extração de características e na classificação. Inúmeras abordagens têm sido propostas para alcançar um melhor desempenho de classificação, porém a fim de obter maior precisão de classificação, a seleção das características selecionadas são o núcleo principal usado na análise de sinais EMG (Nazmi *et al.* 2016).

Figura 4 – Diagrama de blocos para o desenvolvimento de dispositivos assistivos baseados na classificação de sinais EMG



Fonte: Adaptado de Nazmi et al. 2016.

De acordo com a figura 4, inicialmente os dados são segmentados a partir dos sinais EMG brutos. Para cada segmento dividido, o sinal passa pelo estágio de pré-processamento, isto é, filtragem e retificação para melhorar a precisão e o tempo de resposta do controlador de dados. Em seguida, um conjunto de características são extraídas desses dados para alimentar o classificador (Nazmi *et al.* 2016).

Com o objetivo de realizar o reconhecimento de padrões para aplicações mioelétricas, uma série de características são extraídas do sinal mioelétrico para fins de classificação (Toledo-Pérez

et al. 2019). A extração de características desempenha um papel crítico na obtenção de melhor desempenho de classificação para reconhecimento de padrões de movimento. Este processo envolve a transformação de sinais EMG brutos em um vetor de características, cujas características podem ser divididas em três categorias: características no domínio do tempo (TD) (Figura 5), características no domínio da frequência (FD) (Figura 6) também conhecido como domínio espectral e características no domínio do tempo-frequência (TFD) (Nazmi *et al.* 2016). A partir dessas características, é realizado o treinamento – que varia de acordo com o método – e, dessa forma, é possível classificar o tipo de movimento, ou o tipo de região muscular ativada (Toledo-Pérez *et al.* 2019).

Figura 5 – Variáveis de análise de EMG no domínio do Tempo

Features	Abbreviation
Integrated EMG	IEMG
Mean Absolute Value	MAV
Modified mean absolute value 1	MAV1
Modified mean absolute value 2	MAV2
Root Mean Square	RMS
Variance	VAR
Waveform length	WL
Zero crossing	ZC
Slope sign change	SSC
Willison amplitude or Wilson amplitude	WAMP
Kurtosis	KURT
Skewness	SKEW
Moving Approximate Entropy	moving ApEn
Fuzzy approximate entropy	fApEn
Simple square integral	SSI
v-Order	V
Log detector	LOG
Average amplitude change	AAC
Difference absolute standard deviation value	DASDV
Mean absolute value slope	MAVSLP
Multiple hamming windows	MHW
Multiple trapezoidal windows	MTW
Histogram of EMG	HIST
Auto-regressive coefficients	AR
Cepstral coefficients	
Standard deviation	SD
Cepstral coefficients	CC
Sample entropy	SampEn
Integral absolute value	IAV
Variance	VAR
Maximum amplitude	MAX

Fonte: (Nazmi *et al.* 2016)

As informações extraídas dos sinais EMG serão então alimentadas no classificador para mapear diferentes padrões e combiná-los adequadamente. Os classificadores devem ser implantados para distinguir diferentes categorias de características extraídas. Várias técnicas são implementadas para classificar dados EMG, como redes neurais artificiais (ANN- *Artificial neural networks*), classificador Bayesiano (BC- *Bayesian classifier*), lógica fuzzy (FL - *Fuzzy Logic*), perceptron multicamada (MLP- *Multilayer Perceptron*), máquinas de vetor de suporte (SVM- *Support Vector Machine*), análise discriminante linear (LDA - *Linear Discriminant*

Figura 6 – Variáveis de análise de EMG no domínio da Frequência

Features	Abbreviation
Mean frequency	MNF
Median frequency	MDF
Mean power frequency	MNP
Peak frequency	PKF
Total power	TTP
Frequency ratio	FR
Power spectrum ratio	PSR
The power spectrum deformation	Ω
Variance of central frequency	VCF
Signal-to-motion artifact ratio	SMR
Signal-to-noise ratio	SNR
Spectral moment	SM
Energy	EN
Wavelet decomposition	WDC
Wavelet decomposition difference	WDCDIF
Modified mean frequency	MMNF
Modified median frequencies	MMDF
Short Time Fourier transform	STFT

Fonte: (Nazmi *et al.* 2016)

Analysis), oculta Modelos de Markov (HMM - *Hidden Markov model*) e K-vizinho mais próximo (KNN - *K-nearest neighbor*) (Nazmi *et al.* 2016).

A escolha da metodologia de classificação depende do campo de aplicação. Na área de IHM, os estudos constataram que a maioria dos classificadores são baseados em redes neurais. Isso porque tem sido usado por muitos pesquisadores no passado, bem como tem inúmeras vantagens no processamento e classificação de biossinais (Ahsan *et al.* 2009). Algumas técnicas de inteligência artificial (IA) baseadas principalmente em redes neurais têm sido propostas para processamento e discriminação de sinais EMG. Rede neural é uma técnica de computação que evoluiu de modelos matemáticos de neurônios e sistemas de neurônios. E durante os últimos anos, as redes neurais tornaram-se uma ferramenta útil para categorização de dados multivariados (Ahsan *et al.* 2009).

2.3 Geração de EMG Artificiais

O EMG tem sido usado para diversas aplicações de interface homem-máquina, como controle de interfaces de computador, navegação em ambientes de realidade virtual, controle de robôs, drones e outras aplicações interessantes. No entanto, adquirir esses conjuntos de dados de pacientes é uma tarefa complicada e às vezes dolorosa. Existe uma alta variabilidade do sujeito (ou seja, a mesma pessoa pode realizar o mesmo gesto de muitas maneiras diferentes, enquanto fatores como fadiga e suor podem afetar as propriedades do sinal gravado) (Tsinganos *et al.* 2020). Além disso, a maioria dos pacientes experimentam efeitos desagradáveis durante tais experimentos, como cansaço, fadiga, e uma grande variedade de movimentos geralmente não são possíveis devido à limitação de movimento do paciente e ao comprometimento devido à

doença (Zanini e Colombini 2020). Em contrapartida, as análises de sinais de EMG requerem uma grande quantidade e grande variabilidade de dados para ajuste da interface, logo o aumento de dados é uma abordagem alternativa promissora para estender os conjuntos de dados existentes, permitindo mais pesquisas e análises.

O trabalho de Zanini e Colombini (2020) propõe duas abordagens de aumento de dados baseadas em *Deep Convolutional Generative Adversarial Networks* (DCGANs) e *Style Transfer* para aumentar os sinais de EMG da doença de Parkinson com o uso de dois bancos de dados EMG distintos. Na metodologia do trabalho as Redes Neurais são treinadas para aprender os padrões específicos de tremor do sinal EMG, sendo capazes de reproduzir tal tremor para cada paciente e assim avaliar a condição do tremor para diferentes cenários (diferentes pessoas). Este trabalho fornece resultados interessantes que dão suporte ao desenvolvimento de novos tratamentos para redução de tremores em pacientes com a doença de Parkinson ao estender os conjuntos de dados existentes possibilitando a simulação do tratamento em diferentes casos artificiais e reduzindo o tempo necessário para experimentos de pacientes reais para a captura de dados.

O trabalho de Tsinganos et al. (2020) investiga a aplicação de diferentes técnicas de aumentação de dados ao problema específico de reconhecimento de gestos das mãos com base em sinais EMG de superfície. Ele fornece uma avaliação comparativa dos métodos mais comuns de aumentação como ruído aditivo e janelas sobrepostas; e também dos métodos mais novos como distorção de magnitude, decomposição wavelet e modelos EMG sintéticos. Os resultados do trabalho mostram que a acurácia de reconhecimento pouco varia de um método para outro de aumentação, no entanto são mais favorecidos os métodos cujos dados são inflados com sinais de maior diversidade, para aplicações em simulações e dispositivos que testam inúmeros casos. Ao comparar métodos comuns com métodos novos, os modelos mais modernos tiveram a precisão significativamente melhorada em 1% usando os métodos de aumentação investigados (Tsinganos *et al.* 2020). Além disso, os pesquisadores advertem que a seleção apropriada da quantidade de aumentação é importante, pois o tamanho de dados necessário depende do conjunto de dados e da arquitetura do modelo.

O trabalho de Yang et al. (2018) considerou as operações para aumentação de dados como inversão e rotação que já vinham se mostrado eficientes para reconhecimento de imagens e se inspirou nessa metodologia para projetar abordagens semelhantes para a aumentação do EMG para decodificação de movimentos no punho. Dentre as abordagens, a disposição errada dos eletrodos, alternância de canais, eletrodos colocados inversamente foram estratégias utilizadas para aumentar a quantidade de dados. De acordo com os resultados, os autores recomendam a abordagem de aumentação de dados colocando todos os eletrodos na direção oposta e alternando aleatoriamente dois canais no processo de treinamento (Yang *et al.* 2018).

O trabalho de Mendez; Lhoste e Micera (2022) reforça que o registro de grandes quantidades de sinais EMG enfrenta vários problemas, pois o registro de dados dos pacientes

Figura 7 – Tabela resumo dos trabalhos analisados sobre geração artificial de sinais EMG.

Autor	Técnica	Resultado
Yang et al. (2018)	Aumentação de dados através de inversão e rotação.	A colocação dos eletrodos na direção oposta alternando aleatoriamente dois canais foi o que obteve o resultado mais satisfatório.
Zanini e Colombini (2020)	Deep Convolutional Generative Adversarial Networks (DCGANs) e Style Transfer	Estenderam os conjuntos de dados existentes possibilitando a simulação do tratamento em diferentes casos artificiais, reduzindo o tempo necessário para experimentos de pacientes reais para a captura de dados
Tsinganos et al. (2020)	Ruído aditivo; janelas sobrepostas; distorção de magnitude; decomposição wavelet; e modelos de EMG sintéticos.	A acurácia de reconhecimento pouco varia de um método para outro de aumentação, no entanto são mais favorecidos os métodos cujos dados são inflados com sinais de maior diversidade, para aplicações em simulações e dispositivos que testam inúmeros casos.
Mendez; Lhoste e Micera (2022)	Aprendizado profundo usando redes adversarias generativas (GANs)	A geração de dados sintéticos de alta qualidade aumentaram o desempenho da classificação.
Bird et al. (2021)	GPT (Generative Pretrained Transformer)	Os resultados mostram que os dados sintéticos melhoram a classificação dos dados reais tanto em termos de benchmarking de modelo quanto na classificação de amostras não vistas

Fonte: De autoria própria

consome muito tempo e pode resultar em fadiga muscular (Mendez *et al.* 2022). Este trabalho explorou uma estratégia de aumento de dados de aprendizado profundo usando redes adversarias generativas (GANs) para criar dados sintéticos de alta qualidade para aumentar o desempenho da classificação.

O trabalho de Bird et al. (2021) mostra que alguns modelos de GPT (*Generative Pretrained Transformer*) podem gerar sinais biológicos sintéticos (EMG e EEG) e melhorar a classificação de dados reais. Os resultados mostram que os dados sintéticos melhoram a classificação dos dados reais tanto em termos de *benchmarking* de modelo quanto na classificação de amostras não vistas (Bird *et al.* 2021). A Figura 7 descreve resumidamente os trabalhos analisados que tratam sobre geração artificial de sinais EMG.

2.4 Teoria de Resposta ao Item

A TRI consiste em um conjunto de modelos que relacionam as respostas dadas aos itens de interesse com as habilidades latentes de seus respondentes (Embretson e Reise 2000). Tais habilidades latentes são características que não podem ser medidas diretamente, mas podem ser mensuradas através de variáveis secundárias relacionadas aos traços latentes de seus respondentes (Andrade *et al.* 2000). Os métodos de TRI ficaram popularmente conhecidos através das provas do Exame Nacional do Ensino Médio (ENEM) pelo Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP) do Ministério da Educação (MEC), ao aplicar a teoria para pontuar os estudantes bonificando-os pelas suas habilidades. O método permite valorizar aqueles que acertam as questões de forma mais coerente, ou seja, aqueles que acertam mais questões fáceis do que difíceis (JUNIOR 2010).

A aplicação de TRI foi recentemente sugerida para novas perspectivas, como por exemplo: para uma melhor compreensão dos resultados de experimentos de AM (Martínez-Plumed *et al.* 2016). Aspectos da utilidade da TRI na literatura de AM, incluem: comparações de filtragem colaborativa (Bergner *et al.* 2012); avaliação de sistemas de processamento de linguagem natural (Lalor *et al.* 2016); identificação de itens para aprendizagem adaptativa (Pliakos *et al.* 2019); avaliação da utilidade de classificadores de AM (Martínez-Plumed *et al.* 2019); e seleção de atributos de AM (Kline *et al.* 2020). De forma a estabelecer um paralelo entre as aplicações, pode-se entender os itens, parte fundamental da TRI, como análogo às questões respondidas pelos estudantes do ENEM e os pontos dos dados de treinamento em AM; e os estudantes (ou respondentes), aos classificadores.

O modelo TRI básico, o modelo Rasch (Rasch 1960), é baseado na ideia de que a probabilidade de resolver corretamente um item é uma função logística da diferença entre um parâmetro de pessoa (ou de um classificador) e um parâmetro de item, que são frequentemente interpretados como o parâmetro de habilidade da pessoa, e o parâmetro de dificuldade do item (Pliakos *et al.* 2019). Por modelos logísticos, a TRI fornece vários parâmetros com a finalidade de caracterizar os itens e seus respondentes, como: dificuldade, nível de proficiência ou habilidade, discriminação e adivinhação. Atualmente, esses parâmetros são frequentemente usados em testes adaptativos computadorizados (Linden 2009), em que, após cada resposta, a estimativa de capacidade de um participante do teste é atualizada e um item é fornecido com uma dificuldade que corresponde aproximadamente a estimativa da capacidade. Desta forma, testes mais curtos são suficientes para obter uma visão precisa da habilidade do respondente (Pliakos *et al.* 2019), fazendo com que o objetivo da aprendizagem adaptativa seja não somente de avaliação de aprendizagem, mas de motivação para melhorar o progresso da aprendizagem, fornecendo novos itens mais personalizados (Zhang e Chang 2016).

2.4.1 Teoria de Resposta ao Item em Aprendizagem de Máquina

Uma área onde a análise de dificuldade foi investigada recentemente é a AM (Martínez-Plumed *et al.* 2019). Pesquisas em aprendizagem de máquina têm a tradição de avaliar diferentes técnicas com muitos problemas, numa área chamada meta-aprendizagem (Brazdil *et al.* 2008), cujo objetivo é analisar as características dos classificadores partindo de um conjunto de dados único. No entanto, o uso de parâmetros como dificuldade e proficiência não é tão comum em comparação às métricas usuais como precisão, acurácia e custo computacional (Ferri *et al.* 2009).

A definição de "dificuldade" ainda é vaga nesse contexto, mas existe uma noção intuitiva percebida através das medições comuns de avaliação que alguns problemas são mais difíceis que outros, e que algumas técnicas são mais capazes do que outras. Assim, o estudo de TRI aplicado a AM visa caracterizar a relação existente entre dificuldade e habilidade; identificar os melhores classificadores, sejam os que resolvem mais problemas fáceis ou os que resolvem poucos problemas, porém difíceis (Martínez-Plumed *et al.* 2019). A AM está sendo apresentada

na literatura de saúde com resultados de estudos frequentemente relatados usando métricas como precisão, sensibilidade, especificidade, memória ou escore (Kline *et al.* 2020). Essas métricas convencionais podem classificar os modelos com base em resultados gerais para bonificar o método A em comparação com o método B, mostrando inclusive diferenças significativas entre as respostas (Martínez-Plumed *et al.* 2019). Mas, sabendo que cada caso possui particularidades que o diferencia dos outros, as abordagens avaliativas atuais de AM não são suficientes para apresentar uma visão estratificada (Kline *et al.* 2020) dos modelos seguindo critérios de classificação, isto é, as métricas não levam em conta a distribuição desses resultados. Não é avaliado, por exemplo, se os sistemas são melhores nos problemas mais difíceis ao custo de falhar em problemas fáceis (Martínez-Plumed *et al.* 2019), ou se suas respostas têm coerência com o nível de suas habilidades e complexidades. Embora cada métrica forneça uma perspectiva diferente sobre o desempenho, elas continuam a ser medidas gerais para toda a amostra, desconsiderando a singularidade de cada problema ou caso (Kline *et al.* 2020).

Utilizando TRI, é possível estratificar os dados com base na dificuldade de classificação de cada caso, permitindo uma avaliação dos classificadores de AM em forma de uma distribuição em vez de um único valor escalar. A avaliação a partir de qualquer medida escalar única tem limitações significativas (Drummond e Japkowicz 2010), e essas medidas podem simplificar questões complexas e combinar coisas que deveriam ser mantidas separadas (Drummond e Japkowicz 2010). A estratificação é realizada, uma vez que todos os parâmetros dos itens e que todas as habilidades dos respondentes — tanto individuais como populacionais — de todos os grupos avaliados estão numa mesma métrica, ou seja, quando todos os parâmetros envolvidos são comparáveis (Andrade *et al.* 2000).

Assim, a TRI pode construir escalas de conhecimento interpretáveis com aplicação para qualquer esfera de aprendizagem. E com a estratificação, infere-se não apenas a descrição de um sistema específico de IA, como também observa-se a progressão dos métodos de IA frente a diferentes exemplos de dados.

Alguns trabalhos têm apoiado a concepção de que algoritmos de AM devem superar especialistas humanos, quanto ao nível de abstração e complexidade (Senders *et al.* 2018). Porém, outros já suportam a ideia de que abordagens excessivamente complexas usadas em alguns modelos de AM não são melhores do que modelos mais simples e intuitivos (Christodoulou *et al.* 2019). Os modelos devem fornecer aos usuários uma compreensão das razões pelas quais os algoritmos funcionam, e tornar os modelos ditos “caixa preta” interpretáveis, ao invés de simplesmente fornecer resultados de diferenças entre dois valores escalares (Langley 2011).

O trabalho de Martínez-Plumed et al. (2016) uniu uma ampla gama de classificadores AM com uma diversidade de conjunto de dados para projetar um cenário experimental realista e reproduzível, semelhante a sua aplicação em testes educacionais, onde os exemplos correspondem aos itens e classificadores correspondem aos alunos. O trabalho reuniu uma grande população de modelos de AM, 128 classificadores provenientes de 15 famílias distintas, com modificação de

parâmetros para aumentar a diversidade da população, dentre elas: árvores de decisão, métodos baseados em regras, análise discriminatória, Bayesiana, redes neurais, máquina de vetor de suporte, boosting, ensacamento, empilhamento, florestas aleatórias, vizinho mais próximo, mínimos quadrados parciais, regressão de componente principal, regressão global e multinomial. Nesse trabalho foi utilizado o pacote R chamado ltm para implementação do modelo TRI, sendo responsável por esclarecer aspectos importantes para interpretação dos parâmetros de TRI, como discriminação, adivinhação e habilidade. O estudo mostra que, se um classificador está prevendo todas as instâncias corretamente, ou conjunto de dados é muito simples, ou o classificador está superdimensionando os dados ou, como usual no TRI, o classificador está "trapaceando". Segundo o trabalho, existem três potenciais áreas de aplicação da TRI a partir do que foi observado em seus estudos: (1) Aprimorar os métodos de classificação, pois observaram que o parâmetro de discriminação pode ser usado para identificar instâncias com ruído, com características particulares, ou com sobreajuste. (2) Seleção do classificador, pois com uma antecipação da dificuldade de uma instância, pode-se decidir qual classificador é preferível para uma dada instância particular, olhando para as curvas características do classificador. (3) Avaliação, pois uma direção possível de estudo é o uso de TRI para produzir mais conjuntos de dados discriminativos, que é o foco deste trabalho.

O trabalho de Kline et al. (2020) usou dois conjuntos de dados disponíveis gratuitamente, que abriga dados de pacientes de unidades de terapia intensiva eletrônica para estimar a previsão de mortalidade. Um modelo logístico de dois parâmetros foi usado para fornecer pontuações para cada caso usando o programa IRTPRO (Teoria da resposta ao item para resultados relatados pelo paciente), e alguns algoritmos de AM foram usados na demonstração para classificar os casos com base em suas características relacionadas à saúde, dentre eles: regressão logística, análise discriminante linear, K-vizinhos mais próximos, árvore de decisão, Bayes e rede neural. O trabalho fornece uma maneira de avaliar os classificadores de AM em um contexto de saúde, mostrando que os classificadores podem ser avaliados com base em seu desempenho em casos de dificuldade variada. Segundo os autores, um classificador dito bom é o que apresenta consistência em todos os níveis de dificuldade. O método do trabalho apresenta uma interseção entre a medicina personalizada e AM, mantendo explicabilidade e transparência na seleção de atributos e precisão modelada, os quais são fundamentais para sua compreensão na área da saúde.

O trabalho de Pliakos et al. (2019) utiliza a mesma ideia de implementar um algoritmo de aprendizagem de máquina que incorpora TRI em uma estrutura unificada. Foram utilizados dados de experimentos educacionais para esse estudo, cuja finalidade é minimizar o efeito problemático no campo da aprendizagem adaptativa conhecido como “problema de inicialização a frio”. Os sistemas de aprendizagem adaptativa visam fornecer itens de aprendizagem adaptados ao comportamento e às necessidades dos alunos de forma individual. No entanto, o problema surge pois esses sistemas não têm informações sobre os níveis de habilidade inicial de alunos novatos, logo a qualidade da recomendação de itens personalizados durante a fase inicial do ambiente de aprendizagem fica prejudicada. O trabalho utiliza alguns modelos de AM para comparação, como:

árvore de regressão; máquina de vetor de suporte; floresta aleatória; rede neural convolucional; e regressão linear. A contribuição do trabalho foi que aplicar TRI combinado a florestas aleatórias fornece menor erro para estimativa de habilidade e maior precisão na previsão da resposta em relação a outros modelos, diminuindo o efeito da inicialização a frio.

Com uma abordagem diferente, Chen e Ahn (2020) propõem uma nova estrutura probabilística a fim de melhorar a precisão de um algoritmo de votação ponderada. Eles tratam os classificadores como concorrentes, assim, utilizam TRI como forma de avaliar simultaneamente a dificuldade e a habilidade dos classificadores das amostras atribuindo pesos para eles com base nas suas habilidades. Os limites de classificação são construídos pelos pontos que frequentemente são classificados incorretamente e pesos mais altos são atribuídos aos classificadores com habilidades superiores. Na pesquisa, foram utilizados quinze conjuntos de dados reais e dois conjuntos de dados artificiais, e os modelos AM para classificação foram: floresta aleatória; ensacamento; árvore de decisão; análise discriminante linear; e máquina de vetores de suporte. Foi observado que quando a variação dos parâmetros é grande o suficiente, pesos mais altos são atribuídos aos classificadores mais fortes e assim, eles conseguem superar outros métodos.

O trabalho de Prudêncio et al. (2015) propõe a avaliação de classificadores pelos exemplos mais difíceis que ele é capaz de resolver através do modelo de TRI, que fornece uma série de ferramentas para caracterizar instâncias, incluindo dificuldade e poder discriminante. Com essa caracterização, a avaliação do modelo em AM pode ser repensada, de forma a escolher instâncias que são mais discriminativas durante a avaliação e seleção do modelo. O estudo de caso do trabalho utilizou um conjunto de classificadores produzidos por Florestas Aleatórias, treinados com diferentes números de árvores, e aplicou uma função logística dicotômica (de dois parâmetros). A quantidade de parâmetros se baseia no resultado preliminar experimental ao estudo de que não foi observado nenhum ganho significativo no ajuste do modelo ao adicionar o terceiro parâmetro (de adivinhação). Os autores reforçam que para um conjunto de classificadores mais heterogêneos, o parâmetro de adivinhação pode ser importante e provavelmente relaciona-se à classe de distribuição.

A Figura 8 descreve resumidamente os trabalhos analisados que utilizam TRI em aprendizagem de máquina.

2.4.2 Teoria de Resposta ao Item em Interface Homem-Máquina

Os parâmetros das funções dos itens são geralmente estimados em conjuntos de dados de vários respondentes e itens (Bauer e Gharabaghi 2015). Foi observada uma aplicação da TRI na caracterização de itens em uma bateria de testes de aptidão física. Essa aplicação exibe uma problemática de interesse nacional, uma vez que as versões para baterias de teste têm variado expressivamente ao longo dos anos, e uma base para modificação das versões do teste é, apropriadamente, em função da evidência clínica e dos resultados dos pacientes (Safrit *et al.* 1989). Por exemplo, existem várias versões do teste de abdominais, no entanto, informações

Figura 8 – Tabela resumo dos trabalhos analisados sobre TRI em AM.

Autor	Técnica	Resultado
Prudêncio et al. (2015)	o trabalho utilizou um conjunto de classificadores produzidos por Florestas Aleatórias, treinados com diferentes números de árvores, e aplicou uma função logística dicotômica (de dois parâmetros).	A quantidade de parâmetros se baseia no resultado preliminar experimental ao estudo de que não foi observado nenhum ganho significativo no ajuste do modelo ao adicionar o terceiro parâmetro (de adivinhação).
Martínez-Plumed et al. (2016)	Utilizaram 128 classificadores provenientes de 15 famílias distintas: árvores de decisão, métodos baseados em regras, análise discriminatória, Bayesiana, redes neurais, máquina de vetor de suporte, boosting, ensacamento, empilhamento, florestas aleatórias, vizinho mais próximo, mínimos quadrados parciais, regressão de componente principal, regressão global e multinomial. E o pacote R chamado ltm para implemetação do modelo TR.	Segundo o trabalho, existem três potenciais áreas de aplicação da TRI a partir do que foi observado em seus estudos: (1) Aprimorar os métodos de classificação (2) Seleção do classificador (3) Avaliação.
Pliakos et al. (2019)	O trabalho utiliza alguns modelos de AM para comparação, como: árvore de regressão; máquina de vetor de suporte; floresta aleatória; rede neural convolucional; e regressão linear.	A contribuição do trabalho foi que aplicar TRI combinado a florestas aleatórias fornece menor erro para estimativa de habilidade e maior precisão na previsão da resposta em relação a outros modelos, diminuindo o efeito da inicialização a frio.
Kline et al. (2020)	Um modelo logístico de dois parâmetros foi usado para fornecer pontuações para cada caso usando o programa IRTPRO, e alguns algoritmos de AM foram usados para classificação, são eles: regressão logística, análise discriminante linear, K-vizinhos mais próximos, árvore de decisão, Bayes e rede neural.	Segundo os autores, um classificador dito bom é o que apresenta consistência em todos os níveis de dificuldade. O método do trabalho apresenta uma interseção entre a medicina personalizada e AM, mantendo explicabilidade e transparência na seleção de atributos e precisão modelada, os quais são fundamentais para sua compreensão na área da saúde.
Chen e Ahn (2020)	modelos AM para classificação foram: floresta aleatória; ensacamento; árvore de decisão; análise discriminante linear; e máquina de vetores de suporte.	Foi observado que quando a variação dos parâmetros é grande o suficiente, pesos mais altos são atribuídos aos classificadores mais fortes e assim, eles conseguem superar outros métodos.

Fonte: De autoria própria

adicionais seriam desejáveis para a seleção da versão mais apropriada para um determinado propósito clínico (Safrit *et al.* 1989) e isso pode se aplicar aos treinamentos de calibração de uma IHM como estamos propondo. Usando TRI, os níveis de dificuldade de uma variedade de versões podem ser estimados e expressos na mesma escala da habilidade (Safrit *et al.* 1989). Isso nos permite quantificar a respectiva variabilidade da habilidade do sujeito e dificuldade da tarefa necessária para algoritmos de ajuste, cujas taxas marginais de sucesso nos permitem estimar a dificuldade de um teste motor específico e a habilidade dos pacientes (Bauer e Gharabaghi 2015). De modo geral, os sujeitos com maior habilidade motora abdominal alcançam uma taxa de sucesso mais alta, assim como os testes mais fáceis devem resultar em taxas médias de sucesso mais altas. Essa informação pode permitir ao especialista obter uma melhor compreensão do funcionamento de vários testes abdominais em diferentes níveis de habilidade (Safrit *et al.* 1989). Embora a TRI não possa fornecer as respostas a todas as perguntas relacionadas à seleção e ao desenvolvimento de testes, ela pode gerar informações que permitem aos desenvolvedores de testes a tomada de decisões mais inteligentes sobre essas questões (Safrit *et al.* 1989).

O trabalho de Bauer e Gharabagui (2015) utiliza os conceitos de TRI em Aprendizagem de máquina para uma aplicação muito importante no ramos de IHM, as interfaces cérebro-máquina.

É, até então, o único trabalho que traz a abordagem do TRI para ICMs no contexto da decodificação. O trabalho estuda o treinamento de *neurofeedback* (NFB) com objetivo de reduzir sintomas de transtornos neurológicos e neuropsiquiátricos através de ICMs.

Para esse propósito, eles utilizam uma gama de algoritmos de classificação para identificar diferentes estados cerebrais e aplicam um método analítico para determinar a Zona de Desenvolvimento Proximal (ZDP) como uma medida dos recursos cognitivos de um sujeito e da eficácia instrucional do NFB. Define-se ZDP como "aquelas funções da aprendizagem que ainda não amadureceram, mas que estão em processo de maturação, ou funções que amadurecerão, mas que estão, presentemente, em estado embrionário" (VYGOTSKY 1994).

No contexto de ICM, ZDP pode ser definida como a faixa entre as tarefas mais difíceis que podem ser realizadas pelo sujeito com e sem ajuda. Essa abordagem é baseada em uma reconsideração da teoria de resposta ao item e da teoria de carga cognitiva para design instrucional, combinada à curva de classificação de acurácia máxima para fornecer uma medida de desempenho para ICM.

A teoria da carga cognitiva argumenta que a carga mental pode ser dividida em três categorias: carga intrínseca, carga extrínseca e carga pertinente (Jong 2010). Os autores trazem essa teoria para o contexto da aprendizagem da ICM, interpretando-as como tipos distintos de dificuldade do item. Segundo o trabalho, a carga intrínseca se assemelha à dificuldade da tarefa e é causada principalmente pelo elemento de interatividade da tarefa; a carga externa é causada principalmente por informações irrelevantes; e a carga pertinente é causada pela construção e posterior automação de esquemas, ou seja, aprendizagem.

Os dados para elaboração do trabalho foram colhidos a partir da atividade cerebral de dois indivíduos destros e saudáveis, uma mulher (19 anos) e um homem (31 anos), que apresentavam diferentes habilidades de autorregulação cerebral. Cada um deles realizou 75 testes de imagens motoras com indicação. A estrutura do ensaio consistia em fases consecutivas de preparação (2 s), imagiologia motora (6 s) e repouso (6 s), cada uma das quais iniciada por uma pista auditiva específica. A eletroencefalografia (EEG) foi registrada em 64 canais com foco nas regiões sensório-motoras (FC3, C3 e CP3).

O trabalho demonstrou que existe uma relação natural entre a classificação das funções de taxa e as funções de resposta ao item, revelando a conexão entre a curva de classificação de acurácia máxima e a capacidade do cérebro de realizar a autorregulação cerebral em uma tarefa de NFB. O método se mostrou satisfatório, servindo como uma medida para avaliar se uma tarefa NFB é ou não instrucionalmente eficaz; para adaptar o algoritmo de classificação e corrigir medições com ruído; para interpretar dificuldades individuais, permitindo aplicação em abordagens personalizadas, cujas necessidades de ensino para alunos e ambientes específicos sejam respeitadas.

Um outro trabalho envolvendo TRI em aplicações com sinais biomédicos foi realizado

por Sturma et. al. (2015). Ele utilizou a TRI para estabelecer uma escala no treinamento de controle protético a fim de determinar se um paciente tem potencial (habilidade) para controlar determinada configuração protética. Semelhante a este presente trabalho, um dos objetivos dele era de determinar se é necessário mais treinamento ou se outras estratégias seriam mais apropriadas para a adaptação de um dispositivo protético. Nesse caso, entretanto, a aplicação do trabalho é de avaliar a habilidade do indivíduo na realização de atividades motoras pré-configuração protética para facilitar reabilitação, diferente da proposta deste presente trabalho que é de avaliação de decodificadores na realização da calibração da atividade motora, com o que de fato é a intenção do movimento.

Para estabelecer um esquema de classificação clinicamente útil para avaliação primária, apoio à decisão para adaptação protética e monitoramento do treinamento de habilidades, a pontuação do TRI foi dividida em cinco classes (Sturma *et al.* 2015): Incapaz (nenhuma função muscular); Pobre (muito treinamento é requerido para uma função protética); Médio (não é capaz de controlar a prótese ainda, mas provavelmente aprenderá com treinamento); Bom (alguns refinamentos precisam ser feitos, mas não são obrigatórios) e Excelente (nenhum treinamento é requerido). A principal capacidade de um sujeito controlar repetidamente a intensidade EMG de músculos e grupos musculares únicos, conforme refletido por esta classificação, pode ser um dos critérios decisivos para o investimento e escolha de uma solução protética (Sturma *et al.* 2015).

Para validar os resultados da classificação de habilidade conforme a TRI, os autores compararam com uma métrica muito utilizada para interpretabilidade na AM (o SHAP - *SHapley Additive exPlanations*), estabelecendo um paralelo entre as duas ferramentas. Segundo os autores, a escolha da prótese mais adequada deve ser acordada pela equipe médica em conjunto com o paciente e esta decisão é baseada em seus objetivos, estilo de vida, nível de amputação e condição física geral diagnosticada pelo médico. No entanto, a capacidade do paciente de gerar sinais EMG para controle protético pode ser avaliada quantitativamente através da utilização de TRI como métrica, para decidir se a adaptação mioelétrica é, de fato, a escolha mais eficaz.

Além disso, a ferramenta também apoia a tomada de decisão sobre a quantidade de treinamento necessária para alcançar o controle proporcional adequado e mostra ao terapeuta quais sinais ou músculos precisam ser treinados, e é esse o ponto que o presente trabalho utilizará como motivador para calibração em IHM. Para o trabalho de Sturma et al. (2015), se um grupo muscular funcional for identificado com uma pontuação de 3 na TRI, enquanto para outro, uma pontuação na TRI próxima de 1 pode ser alcançada, o terapeuta deve se concentrar em treinar os grupos/funções musculares com as pontuações mais altas. E para este trabalho de TCC, o decodificador deve se concentrar em treinar mais vezes itens de pontuação mais baixa (através da aumentação).

A figura 9 descreve resumidamente os trabalhos analisados que utilizam TRI em IHM.

Figura 9 – Tabela resumo dos trabalhos analisados sobre TRI em IHM.

Autor	Técnica	Resultado
Bauer e Gharabagui (2015)	Utilizam vários algoritmos de classificação para identificar diferentes estados cerebrais e aplicam um método analítico para determinar a Zona de Desenvolvimento Proximal (ZDP) como uma medida dos recursos cognitivos de um sujeito e da eficácia instrucional da tarefa.	O trabalho demonstrou que existe uma relação natural entre a classificação das funções de taxa e as funções de resposta ao item, revelando a conexão entre a curva de classificação de acurácia máxima e a capacidade do cérebro de realizar a autorregulação cerebral em uma tarefa de NeuroFeedback (NFB). O método se mostrou satisfatório, servindo como uma medida para avaliar se uma tarefa NFB é ou não instrucionalmente eficaz; para adaptar o algoritmo de classificação e corrigir medições com ruído; para interpretar dificuldades individuais, permitindo aplicação em abordagens personalizadas, cujas necessidades de ensino para alunos e ambientes específicos sejam respeitadas.
Sturma et. al. (2015)	Utilizaram TRI para estabelecer uma escala no treinamento de controle protético a fim de determinar se um paciente tem potencial (habilidade) para controlar determinada configuração protética. Para validar o resultado da TRI, o método SHAP foi utilizado.	A ferramenta apoia a tomada de decisão sobre a quantidade de treinamento necessária para alcançar o controle proporcional adequado e mostra ao terapeuta quais sinais ou músculos precisam ser treinados.

Fonte: De autoria própria

3 TRI

A TRI é um conjunto de modelos matemáticos que procuram representar a probabilidade de um indivíduo dar uma certa resposta a um item. Trata-se de uma função matemática logística que depende dos parâmetros do item e da habilidade (ou habilidades) do respondente. Os modelos propostos na literatura dependem fundamentalmente de três fatores principais: (1) da natureza do item — sejam dicotômicos ou não dicotômicos; (2) do número de populações envolvidas — apenas uma ou mais de uma; e (3) da quantidade de traços latentes que está sendo medida — apenas um ou mais de um.

Os modelos apresentados neste trabalho são formulações válidas apenas para análise de itens de natureza dicotômica, isto é, passíveis de apenas duas respostas (verdadeiro ou falso; certo ou errado; 1 ou 0). Na prática, a análise é feita a partir de itens de múltipla escolha dicotomizados ou a partir de itens abertos (de resposta livre), quando avaliados de forma dicotomizada. Os modelos logísticos para itens dicotômicos são os modelos de resposta ao item mais utilizados, diferenciando-se apenas quanto ao número de parâmetros utilizados para descrever um item. Esses podem ser conhecidos como os modelos logísticos de 1, 2, 3 e 4 parâmetros, que consideram, respectivamente: (1) somente a dificuldade do item; (2) a dificuldade e a discriminação; (3) a dificuldade, a discriminação e a adivinhação, ou seja, a probabilidade de resposta correta dada por indivíduos de baixa habilidade; (4) os três primeiros parâmetros, além do limite superior e inferior, que representam a probabilidade máxima ou mínima de acertar determinado item.

3.0.1 Definição

3.0.1.1 Função Logística

A TRI consiste em modelos dados por funções logísticas. A função logística foi nomeada em 1844–1845 por Pierre François Verhulst, que estudou a função relacionando-a ao crescimento populacional. Essa função tem aplicações em grande diversidade de áreas, incluindo rede neural artificial, ecologia, biomatemática, química, demografia, ciências políticas e estatísticas. A função ou curva logística tem um formato de S, chamada curva sigmoide (Fig. 10) e é dada por:

$$f(x) = \frac{L}{1 + e^{-k(x-x_0)}}$$

onde e é a base dos logaritmos naturais, também conhecido como número de Euler; x_0 é o valor de x no ponto médio da curva sigmoide; L é o valor máximo da curva e k é a declividade da curva.

3.0.1.2 Modelo TRI de 3 Parâmetros

Dos modelos propostos pela TRI, o modelo logístico unidimensional de 3 parâmetros (ML3) é atualmente o mais utilizado e é dado por:

$$P(U_{ij} = 1|\theta_j) = c_i + (1 - c_i)\frac{1}{1 + e^{-Da_i(\theta_j - b_i)}}$$

com i = 1, 2, $\cdots$, I, e j = 1, 2, $\cdots$, n, onde:

U_{ij} é uma variável dicotômica que assume os valores 1, quando o indivíduo j responde corretamente o item i, ou 0 quando o indivíduo j não responde corretamente ao item i.

θ_i representa a habilidade (traço latente) do j-ésimo indivíduo.

$P(U_{ij} = 1|\theta_j)$ é a probabilidade de um indivíduo j com habilidade θ_j responder corretamente o item i e é chamada de Função de Resposta do Item (FRI).

b_i é o parâmetro de dificuldade (ou de posição) do item i, medido na mesma escala da habilidade.

a_i é o parâmetro de discriminação (ou de inclinação) do item i, com valor proporcional à inclinação da Curva Característica do Item (CCI) no ponto b_i.

c_i é o parâmetro do item que representa a probabilidade de indivíduos com baixa habilidade responderem corretamente o item i (muitas vezes referido como a probabilidade de acerto casual, ou adivinhação).

D é um fator de escala, constante é igual a 1. Utiliza-se o valor $1,7$ quando deseja-se que a função logística forneça resultados semelhantes ao da função ogiva normal.

As funções que descrevem os modelos de 1 ou 2 parâmetros podem ser facilmente obtidas a partir do modelo de 3 parâmetros. Por exemplo, quando não existe a possibilidade de acerto ao acaso, faz-se c_i = 0 do modelo anterior, e obtém-se o chamado modelo logístico unidimensional de 2 parâmetros (ML2). E se além de não existir resposta ao acaso, todos os itens possuam o mesmo poder de discriminação, tem-se o chamado modelo logístico unidimensional de 1 parâmetro (ML1), também conhecido como modelo de Rasch (Rasch, 1960).

3.0.1.3 Modelo TRI de 4 Parâmetros

Embora o modelo de 3 parâmetros seja amplamente utilizado para fornecer informações que caracterizam os itens, existem informações essencialmente finitas que não são identificadas: como a possibilidade de "deslizar" super ou inferiormente, que reduz a informação contida nas respostas, pois os parâmetros de assíntota no modelo 3 são parcialmente identificados.

Ou seja, dada uma quantidade infinita de dados, as estimativas dos parâmetros de interesse em geral não convergem para um valor verdadeiro. Logo, o modelo de 4 parâmetros (ML4) traz uma generalização do modelo de três parâmetros incluindo uma assíntota superior para o probabilidade de uma resposta correta, e é dado por:

$$P(U_{ij} = 1|\theta_j) = c_i + \frac{d_i - c_i}{1 + e^{-Da_i(\theta_j - b_i)}}$$

onde a, b, c e d são os parâmetros de discriminação de item, dificuldade, assíntota inferior e assíntota superior, respectivamente. Para o ML4, d representa a probabilidade máxima de $U_{ij} = 1$ em um determinado item.

3.0.2 Curva Característica do Item

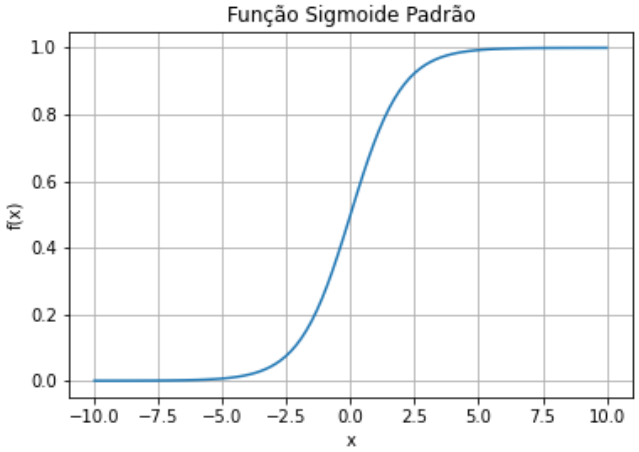
A relação existente entre $P(U_{ij} = 1|\theta_j)$ e os parâmetros do modelo pode ser observada na chamada Curva Característica do Item (CCI) (Fig. 11). A curva baseia-se no fato de que indivíduos com maior habilidade possuem maior probabilidade de acertar o item e essa relação, portanto, não é linear. A CCI tem forma de "S" com inclinação e deslocamento na escala de habilidade definidos pelos parâmetros do item. Na curva, o eixo da abscissa corresponde à habilidade do respondente e o eixo da ordenada corresponde a probabilidade de resposta correta.

Na figura 11, o parâmetro a é proporcional à derivada da tangente da curva no ponto de inflexão e representa o poder de discriminação de um item. Assim, baixos valores de a indicam que o item tem pouco poder de discriminação, isto é, alunos com habilidades bastante diferentes têm aproximadamente a mesma probabilidade de responder corretamente o item; da mesma maneira, valores muito altos de a indicam itens com curvas características muito "íngremes", logo discriminam os alunos em dois grupos: os que possuem habilidades abaixo do valor do parâmetro b e os que possuem habilidades acima do valor do parâmetro b.

O parâmetro b representa a habilidade necessária para uma dada probabilidade de acerto. Esse parâmetro é denominado parâmetro de dificuldade (ou de posição) do item. Assim, quanto maior o valor de b, mais difícil é o item, e vice-versa. Através da figura 11 é possível observar que o intervalo que compreende o valor b, de dificuldade do item, está associado diretamente ao intervalo que compreende o poder discriminante. Por isso, é correto afirmar que a discriminação entre bons respondentes é feita a partir de itens difíceis e não a partir de itens considerados fáceis.

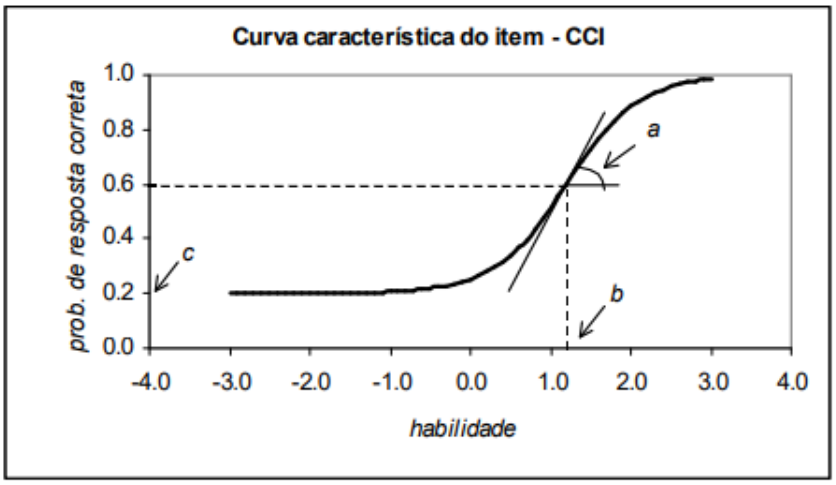
O parâmetro c representa a probabilidade de um aluno com baixa habilidade responder corretamente o item, também conhecido como a probabilidade de acerto ao acaso, ou "chute". Se, no entanto, em uma aplicação não há possibilidade de adivinhação, c seria igual a 0 e b representaria o ponto na escala da habilidade onde a probabilidade de acertar o item é 0,5.

Figura 10 – Função Logística Sigmoide Padrão.



Fonte: Produzida pela autora

Figura 11 – Exemplo de uma Curva Característica do Item. Os coeficientes indicam: (a) discriminação do item; (b) dificuldade do item; (c) adivinhação



Fonte: (Andrade *et al.* 2000)

3.0.3 Função de Informação do Item

A Função de Informação do Item é uma medida utilizada em conjunto com a CCI, e permite analisar quanto um item (ou teste) contém de informação que contribui para a medida de habilidade. A função de informação de um item é dada por:

$$I_i(\theta) = D^2 a_i^2 \frac{Q_i(\theta)}{P_i(\theta)} \left[\frac{P_i(\theta) - c_i}{1 - c_i} \right]^2$$

onde $I_i(\theta)$ é a informação fornecida pelo item i no nível de habilidade θ; $P_i(\theta) = P(X_{ij} = 1|\theta)$; e $Q_i(\theta) = 1 - P_i(\theta)$. Essa função mostra a contribuição dos três parâmetros por sobre a informação do item. Pela equação, observa-se que a informação é maior: (1) quando b_i se aproxima de θ; (2) quanto maior for a_i; (3) e quanto mais c_i se aproximar de 0.

3.0.4 Função de Informação do Teste

A função de Informação do Teste consiste na soma das informações fornecidas por cada item que compõe o teste completo, e é dada por:

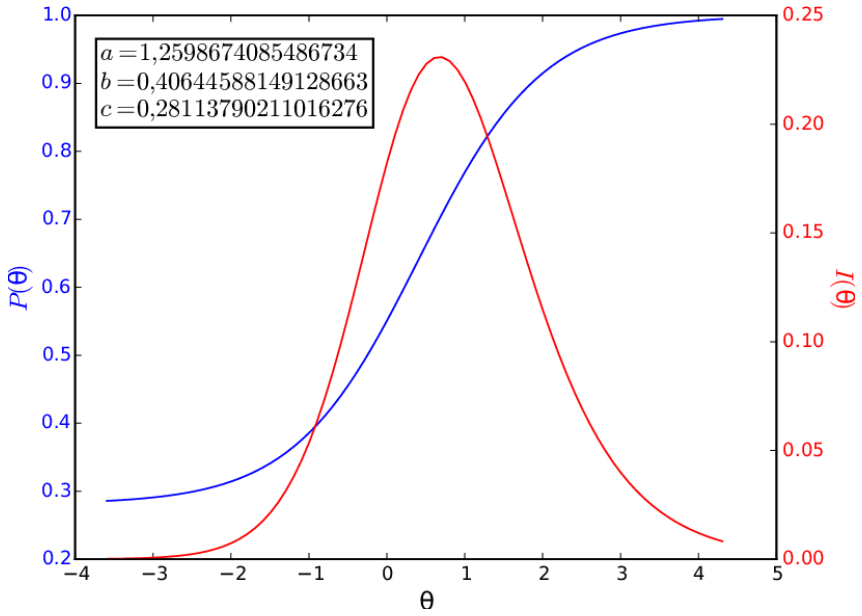
$$I(\theta) = \sum_{i=1}^{I} I_i(\theta)$$

A função de informação do teste também pode ser representada através do erro-padrão de medida. É chamado na TRI de erro-padrão de estimação, e é dado por:

$$EP(\theta) = \frac{1}{\sqrt{I(\theta)}}$$

Nota-se que as medidas de informação, tanto do item quanto do teste, dependem do valor de θ, isto é, do nível de habilidade. Logo, a amplitude do intervalo de confiança para θ depende também do seu valor. Na figura 12 pode-se observar a relação entre a curva característica do item (em azul) e a curva de informação do item (em vermelho).

Figura 12 – Gráfico do nível de habilidade pela probabilidade da resposta correta. A Curva Característica do Item é representada em azul e Curva da Informação do Item, em vermelho



Fonte: (Meneghetti 2015)

4 Método Proposto

O método consiste na aplicação do modelo de TRI de três parâmetros em uma sequência de experimentos controlados, como ferramenta para avaliar o desempenho de alguns modelos de aprendizagem de máquina, utilizando para isso uma base de dados de sinais de EMG de gestos das mãos.

4.1 Dados

O conjunto de dados consiste em uma série de registros eletromiográficos provenientes da coleta através de oito eletrodos posicionados em torno dos cinco dedos da mão, tais quais: indicador; médio; anelar; mínimo; e polegar. O banco disponibiliza os sinais de EMG brutos e também processados, com 10 características extraídas de cada sinal: desvio padrão; raiz quadrada média; mínimo; máximo; cruzamentos zero; variação de amplitude média; amplitude de primeira explosão; valor absoluto médio; comprimento de onda; e amplitude de Willison (fig: 13. O banco é composto por sinais de EMG contendo 6822 linhas de amostras e 81 colunas (sendo 80 de características e 1 para a discriminação do rótulo de 7 classes, dentre elas: 1 - dedo indicador; 2 - dedo médio; 3- dedo anelar; 4- dedo mínimo; 5- polegar; 6- mão em descanso e 7- mão em gesto).

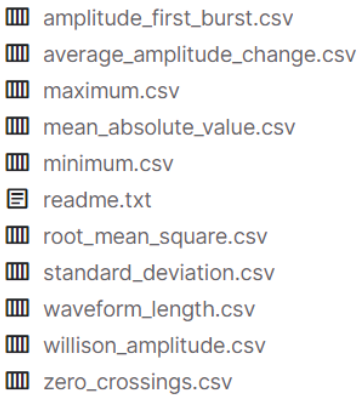
Os dados foram retirados da plataforma Kaggle, comumente conhecida por permitir aos usuários participação em competições de Aprendizagem de Máquina e pela ampla disponibilização de conjunto de dados com foco em AM. Os dados deste trabalho podem ser encontrados no endereço eletrônico: <https://www.kaggle.com/datasets/nccvector/electromyography-emg-dataset>

Informações específicas sobre a aquisição, como posicionamento exato dos eletrodos, perfil do voluntário avaliado não foram disponibilizadas no banco de dados.

4.2 *Design* dos experimentos

Com a finalidade de verificar as hipóteses iniciais, de que a aumentação de exemplos pode favorecer o treinamento de decodificadores para IHM, e de que pode reduzir a necessidade de recoletas para calibração, foram elaborados dois experimentos: o primeiro experimento busca verificar se a aumentação de exemplos baseados na dificuldade, discriminação ou chute de fato reduzem esses três parâmetros, e se o impacto da aumentação prejudica o desempenho dos modelos de decodificação; o segundo experimento visa determinar o efeito nos índices de TRI a partir da aumentação proporcional, isto é, existe um interesse em verificar se a dificuldade, discriminação e chute exibem um comportamento proporcional à aumentação de exemplos.

Figura 13 – Características do sinal



Fonte: Kaggle

4.2.1 Experimento 1

O experimento consiste em carregar todo o conjunto de dados de sinais e realizar treinamento de classificação com seis modelos AM: Árvores de Decisão; Florestas Aleatórias; Redes neurais; AdaBoost; Naive Bayes; e Análise Quadrática Discriminante (QDA). Os modelos foram selecionados por possuírem níveis diferentes de habilidade e permitir a visualização dos resultados para diferentes cenários de proficiência do classificador.

Na segunda etapa do experimento, são selecionados 100 exemplos de cada classe do conjunto de teste. Com os resultados de previsão dicotômica para os 100 itens (zero para uma previsão errada e 1 para uma previsão correta), monta-se uma tabela de erros e acertos por item e por modelo e aplica-se o modelo TRI para identificar características do item como dificuldade, discriminação e adivinhação.

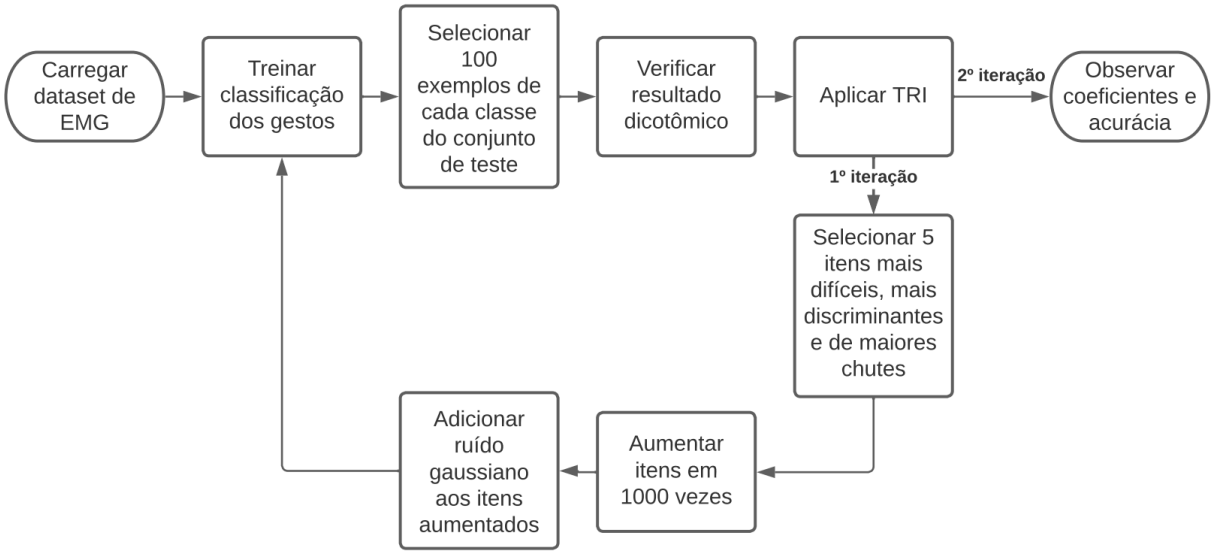
Em seguida, (1ª iteração) os itens são ranqueados por dificuldade, discriminação e chute, e os cinco maiores exemplos de cada parâmetro são aumentados. Assim, obtém-se o item mais difícil, e ele é aumentado em 1000 vezes adicionando a ele ruídos aleatórios gaussianos (média: 0; desvio-padrão: 0,1). Após a aumentação, a base de dados de treinamento aumentada passa por um novo treinamento com os 6 classificadores, obtém-se a resposta dicotômica da predição e novamente a TRI é aplicada para verificação dos coeficientes e da acurácia (2ª iteração). O mesmo é feito para o item mais discriminante e de maior chute. Para as três aumentações, os resultados são dispostos em uma tabela comparativa.

Além disso, a cada novo treinamento dos modelos de decodificação, as acurácias são armazenadas para análise de desempenho. Ver fluxograma 14.

Para melhor compreensão, considere o seguinte exemplo com números: a base de dados carregada é composta por 6822 amostras. Destas, dividem-se 5116 para o conjunto de treino e 1706 para o conjunto de testes. Os 6 classificadores são treinados com essas 5116 amostras do conjunto de treino. O conjunto de teste é aplicado aos modelos treinados e 1706 respostas de

predição são dadas por cada modelo classificador. Destas 1706 seleciona-se de forma aleatória 100 exemplos de cada classe (como são 7 classes, 700 exemplos) e gera-se uma tabela composta por 700 linhas (700 itens) e 6 colunas (6 classificadores), atribuindo o valor 0 caso a predição tenha sido incorreta e 1 caso tenha sido correta. O modelo logístico TRI é aplicado nessa tabela de respostas dos classificadores e os valores de dificuldade, discriminação e chute são fornecidos para cada um dos 700 itens. Então, seleciona-se os 5 itens melhores ranqueados para a etapa de aumentação. Cada item é aumentado em 100 vezes, logo: 5 mais difíceis x 1000 = 5000 novos exemplos que serão acrescidos ao conjunto de treino (assim: 5116 + 5000 = 10116 amostras no conjunto de treino após a aumentação).

Figura 14 – Fluxograma do Experimento 1



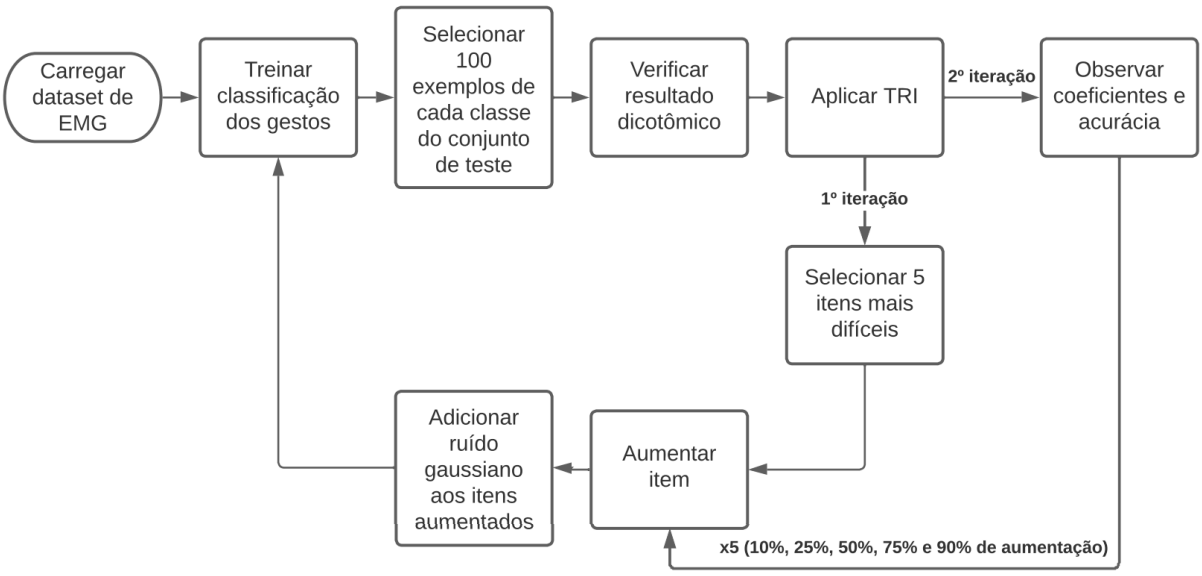
Fonte: De autoria própria.

4.2.2 Experimento 2

O experimento consiste em carregar todo o conjunto de dados de sinais e realizar o treino de classificação com os mesmos seis modelos AM. Da mesma forma, o modelo TRI é aplicado ao resultado dicotômico das predições e os cinco maiores índices de dificuldade são aumentados.

Nesse experimento, ao invés de realizar a aumentação em 1000 vezes, a aumentação é feita de maneira gradual a cada iteração, isto é, 10%, 25%, 50%, 75%. O objetivo da aumentação gradual é verificar se a taxa de aumentação é diretamente proporcional a diminuição dos coeficientes de TRI, bem como o impacto da taxa de aumentação no resultado dos modelos. O experimento visa verificar também uma porcentagem ideal de aumentação que promova a redução dos coeficientes sem prejuízos para o desempenho dos modelos. Ver fluxograma 15.

Figura 15 – Fluxograma do Experimento 2



Fonte: De autoria própria.

4.3 Métricas de Desempenho

As métricas de desempenho são utilizadas para analisar e quantificar a qualidade dos sistemas AM com o intuito de identificar o quão bom é o modelo. Existem muitas métricas de desempenho para AM, mas neste trabalho, foram adotadas duas: O Erro Médio Quadrático, e o nível de habilidade pela TRI.

4.3.1 Erro Médio Quadrático

O Erro Quadrático Médio (EQM) consiste na variância dos erros de predição. É calculado através pela média dos quadrados da diferença entre os valores reais e os valores estimados, dado por:

$$EQM = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x}_i)^2$$

onde n é número de pontos, x_i é o valor real na dimensão i, e $\overline{x}_i$ é o valor estimado na dimensão i.

4.3.2 Habilidade pela TRI

A habilidade ou proficiência de um modelo de AM pode ser calculada pela TRI através de uma função logística que depende das três características do item (dificuldade, discriminação e chute). Os três parâmetros da TRI revelam sobre o conjunto de dados, mas também pode explicar muito a respeito do modelo, sobre quão habilidoso ele é em relação aos demais, e sobre como ele toma suas decisões, e isso é o que se entende por explicabilidade na AM.

5 Resultados e Discussão

Neste capítulo, será apresentada a consolidação dos resultados de cada experimento através de tabelas e gráficos.

5.1 Experimento 1

Ao aplicar TRI às predições do conjunto de testes, ranquearam-se os índices cujos coeficientes foram os mais elevados, provavelmente caracterizando itens problemáticos. Realizou-se uma aumentação em 1.000 vezes desses itens e observou-se novos coeficientes. Este experimento foi realizado 10 vezes e obteve-se a média dos resultados por coeficiente. A Tabela 1 apresenta a média dos coeficientes do item original (sem aumentação) e os coeficientes após o aumentação. A Tabela 2 apresenta a média das acurácias originais (sem aumentação) para cada classificador, e acurácias após aumento por discriminação, por dificuldade e por adivinhação.

Tabela 1 – Coeficientes de TRI para classificação com e sem aumentação.

Coeficientes	**Original**	**Aumentados**
Discriminação (a)	3.337	1.396
Dificuldade (b)	4.529	1.038
Chute(c)	0.45	0.009

Tabela 2 – Acurácia dos modelos para classificação do conjunto de dados original, conjunto aumentado por discriminação, por dificuldade e por chute.

Classificadores	**Original**	**Aumentados por discriminação**	**Aumentados por dificuldade**	**Aumentados por adivinhação**
Decision Tree	0.726	0.727	0.664	0.733
Random Forest	0.739	0.723	0.743	0.784
Neural Net	0.894	0.872	0.878	0.864
Ada Boost	0.396	0.413	0.378	0.477
Naive Bayes	0.583	0.582	0.552	0.572
QDA	0.869	0.564	0.708	0.691

Através da Tabela 1 nota-se que os três coeficientes foram, de fato, reduzidos quando avaliamos a média do coeficiente original em relação ao coeficiente cujos itens foram aumentados (o parâmetro de discriminação teve 41,83% de redução; o parâmetro de dificuldade teve 77,08% de redução; e o parâmetro de adivinhação teve 98% de redução).

A redução do parâmetro de dificuldade observada na tabela sugere que a aumentação pode fazer com que os exemplos mais difíceis se tornem mais fáceis para o classificador. Um exemplo de dificuldade no sinal EMG é quando ocorre a contração de músculos muito próximos.

Isso faz com que o modelo decodificador confunda o grupo muscular ou até mesmo o gesto realizado, provocando erros de classificação, e consequentemente prejudicando a execução de uma tarefa para a IHM. Através da aumentação desses exemplos localizados numa zona de fronteira de decisão, conseguimos separá-los mais facilmente.

A redução do parâmetro de discriminação observado no experimento está intimamente relacionado a redução do parâmetro de dificuldade. A discriminação alta de um exemplo de contração muscular indica um exemplo que consegue diferenciar bem os bons decodificadores dos maus decodificadores, isto é, é um exemplo que discrimina bem pelo nível de habilidade dos modelos classificadores. É muito natural que os exemplos discriminantes sejam também exemplos difíceis, pois são nos exemplos difíceis onde fica claro quais os modelos que conseguem decodificar corretamente e quais não conseguem. Desta forma, realizar aumentação baseada na discriminação pode ser ótimo se o objetivo for a redução da dificuldade de calibração em IHM, no entanto, a aumentação baseada em discriminação pode não ser tão interessante caso o objetivo seja de avaliar modelos decodificadores em IHM.

O parâmetro de adivinhação ou chute refere-se àqueles exemplos cujos modelos não conseguem classificar ao certo e acabavam por chutar uma resposta que pode ou não estar certa. Exemplos de alto chute geralmente são exemplos de fronteira de decisão ou exemplos *outliers* gerados por ruídos ou interferências no momento da coleta do sinal. Logo, os modelos de decodificação podem se beneficiar da aumentação baseada em chute para reduzir a porcentagem de adivinhações, no entanto, fazer isto pode estar criando mais ilhas no mapa de fronteiras, o que prejudica o desempenho da classificação.

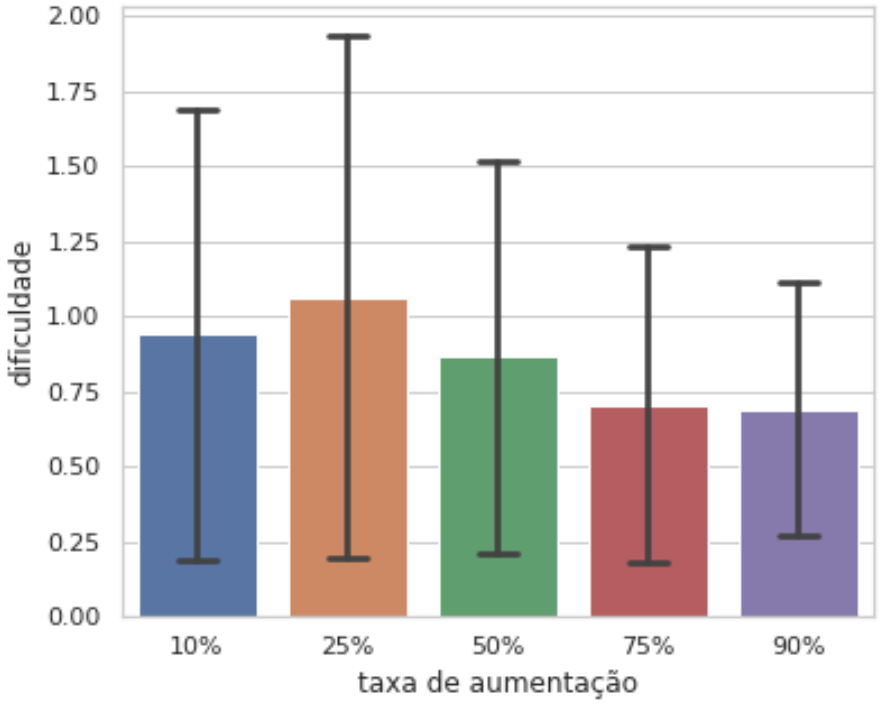
Apesar das grandes reduções nos parâmetros da TRI, o interesse é verificar também se o resultado da aumentação compromete o desempenho do modelo, isto é, se há impacto na acurácia. Através da Tabela 2 verifica-se que a acurácia dos classificadores permaneceram ou sofreram baixa redução, e até aumentaram em alguns casos. Isso quer dizer que a aumentação promoveu uma redução nos parâmetros da TRI, sem alterar a eficiência dos classificadores.

5.2 Experimento 2

Com a intenção de verificar o comportamento do coeficiente de dificuldade conforme taxa de aumentação, foram realizadas uma série de aumentações sucessivas modificando apenas a porcentagem de aumentação para os 5 exemplos de maior dificuldade. A Figura 16 mostra a distribuição do resultado dos coeficientes dos itens que sofreram aumentação através de um gráfico boxplot.

Através da Figura 16 nota-se uma redução da dificuldade à medida que a taxa de aumentação aumenta. A diferença é estatisticamente significante entre a taxa de 10% e 75% ($p < .001$); 10% e 90% ($p < .001$); 25% e 50% ($p < .001$); 25% e 75% ($p < .001$); 25% e 90% ($p < .001$); 50% e 75% ($p < .001$); 50% e 90% ($p < .001$). Até a taxa de 75% observa-se uma clara

Figura 16 – Dificuldade por Taxa de Aumentação



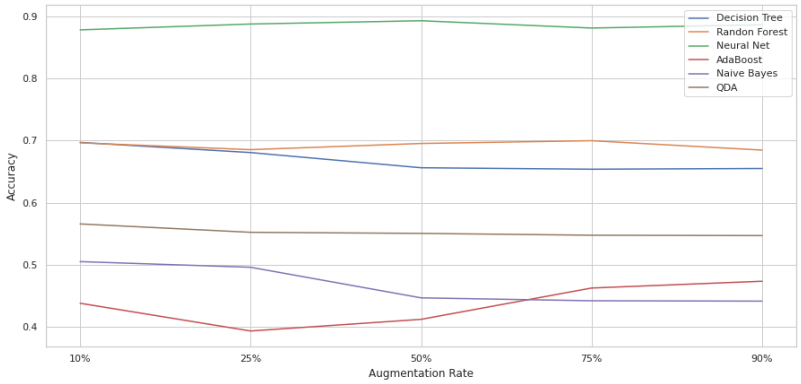
Fonte: De autoria própria

diminuição na dificuldade dos itens, entretanto a 5° taxa de aumentação não provocou diferença estatisticamente significativa em relação a anterior (p = .61). Essa constância na dificuldade pode sugerir uma saturação na aumentação, isto é, os resultados só são percebidos até a taxa de 75%, mais que isso os modelos não se beneficiam de mais exemplos, e pode inclusive provocar sobreajuste (“*overfitting*").

Assim como o Experimento 1, verificamos a acurácia dos modelos classificadores também a cada iteração. A Figura 17 apresenta um gráfico cujas curvas representam a acurácia de cada modelo a cada taxa de aumentação.

É possível notar que o desempenho dos modelos não foi comprometido pelas aumentações, ficando praticamente constante para quase todos os modelos, com exceção da árvore de decisão e do Naive Bayes, que tiveram uma queda na acurácia com aumentação a partir de 25%. Outra observação é que o desempenho permaneceu constante entre a aumentação de 75% e 90% corroborando o resultado do experimento anterior de que os modelos saturam a partir da taxa de 75% de aumentação.

Figura 17 – Dificuldade média por taxa de aumentação



Fonte: De autoria própria

6 Conclusões

O trabalho trouxe potenciais contribuições para o campo de interface Homem-Máquina, bem como para o campo de aumentação de dados em sinais de uma maneira geral. Trabalhos anteriores utilizaram a metodologia para outras aplicações e conseguiram demonstrar uma relação confiável entre a aumentação de instâncias e mudanças nos parâmetros da TRI, mas havia dúvida se essa metodologia seria adequada para aplicações em sinais biomédicos, uma vez que esses sofrem de grande variabilidade e são muito susceptíveis a interferências. Para confirmar a hipótese, este trabalho foi dividido em duas etapas.

A primeira etapa mostrou que, de fato, ao aumentar os exemplos de maior dificuldade, percebe-se uma redução no parâmetro de dificuldade para esse item, e o mesmo acontece para os outros dois parâmetros da TRI. E a segunda etapa mostrou que essa relação entre aumentação e redução dos coeficientes da TRI é proporcional. Isso significa que cada vez que aumentamos os exemplos, melhor é o coeficiente do item aumentado, até que chegue ao limite de saturação de aumentação do modelo. Além disso, o trabalho teve uma preocupação em observar não apenas a redução dos coeficientes como critério de desempenho, mas também a acurácia do modelo ao longo de todos os experimentos. Foi observado que além da redução dos coeficientes da TRI, a acurácia dos classificadores não foi prejudicada.

O trabalho trouxe avanços por uma perspectiva mais prática de aplicação da TRI, o utilizando como método para avaliar algoritmos de decodificação em IHM por níveis de habilidade e tarefas de treinamento da IHM por níveis de dificuldade. Além de utilizar a TRI como critério de aumentação, a TRI também consegue nos fornecer importantes interpretações acerca do sinal avaliado e da tarefa a ser executada através dos seus coeficientes, sendo também considerado um método que auxilia a interpretabilidade de modelos AM. As vantagens desse método na IHM incluem: diminuição do tempo de coleta necessário, quando a decodificação for bem sucedida e a instância for classificada como “fácil”; indicativo sobre a necessidade de coletar mais dados (ou aumentá-los quando o item for classificado como “difícil”); e melhor avaliação de decodificadores por níveis de habilidade e proficiência.

6.1 Trabalhos Futuros

Este trabalho introduziu uma potencial metodologia de aumentação de sinais e mostrou bons resultados iniciais, fazendo com que possa ser expandido para outras esferas do conhecimento ou para o desenvolvimento de plataformas, sistemas ou dispositivos.

Além da aplicação para o sinal de EMG, um possível trabalho futuro seria o teste com outros sinais biomédicos, como o Eletrocardiograma (ECG), Eletroculograma (EOG) e sinais

mais variáveis e aleatórios como o Eletroencefalograma (EEG). De maneira análoga, a aplicação da metodologia apresentada para o EEG traria benefícios no campo de ICM, cuja calibração da decodificação é ainda mais desafiadora.

Além disso, o trabalho utiliza-se de algoritmos para aumentação de maneira manual, e seria conveniente automatizar essa aumentação de instâncias. Nesse sentido, uma possível direção para trabalhos subsequentes seria a aplicação de sistemas AutoML, ou seja, utilização de algoritmos de Aprendizagem de Máquina Automatizados cujo critério de avaliação seriam os coeficientes da TRI.

Outra potencial aplicação seria no desenvolvimento de sistema para suporte aos treinamentos para a IHM, isto é, um sistema que forneça alertas sobre a qualidade da calibração, e indique se são necessárias mais coletas da atividade muscular do paciente ou se a quantidade de dados foi suficiente para uma boa calibração.

Referências

AHSAN, M. R.; IBRAHIMY, M. I.; KHALIFA, O. O. *et al.* Emg signal classification for human computer interaction: a review. *European Journal of Scientific Research*, EuroJournals Publishing, v. 33, n. 3, p. 480–501, 2009. Citado 2 vezes nas páginas 22 e 24.

ANDRADE, D. F. de; TAVARES, H. R.; VALLE, R. da C. Teoria da resposta ao item: conceitos e aplicações. *ABE, Sao Paulo*, 2000. Citado 3 vezes nas páginas 26, 28 e 38.

APARICIO, M. V. Electromiografía cinesiológica. *Rehabilitación*, Elsevier, v. 39, n. 6, p. 255–264, 2005. Citado na página 19.

BAUER, R.; GHARABAGHI, A. Estimating cognitive load during self-regulation of brain activity and neurofeedback with therapeutic brain-computer interfaces. *Frontiers in behavioral neuroscience*, Frontiers, v. 9, p. 21, 2015. Citado 3 vezes nas páginas 16, 30 e 31.

BERGNER, Y.; DROSCHLER, S.; KORTEMEYER, G.; RAYYAN, S.; SEATON, D.; PRITCHARD, D. E. Model-based collaborative filtering analysis of student response data: Machine-learning item response theory. *International Educational Data Mining Society*, ERIC, 2012. Citado na página 27.

BIRD, J. J.; PRITCHARD, M.; FRATINI, A.; EKÁRT, A.; FARIA, D. R. Synthetic biological signals machine-generated by gpt-2 improve the classification of eeg and emg through data augmentation. *IEEE Robotics and Automation Letters*, IEEE, v. 6, n. 2, p. 3498–3504, 2021. Citado na página 26.

BRAZDIL, P.; CARRIER, C. G.; SOARES, C.; VILALTA, R. *Metalearning: Applications to data mining*. [S.l.]: Springer Science & Business Media, 2008. Citado na página 27.

CANNAN, J.; HU, H. Human-machine interaction (hmi): A survey. *University of Essex*, Citeseer, 2011. Citado na página 14.

CHEN, Z.; AHN, H. Item response theory based ensemble in machine learning. *International Journal of Automation and Computing*, Springer, v. 17, n. 5, p. 621–636, 2020. Citado na página 15.

CHRISTODOULOU, E.; MA, J.; COLLINS, G. S.; STEYERBERG, E. W.; VERBAKEL, J. Y.; CALSTER, B. V. A systematic review shows no performance benefit of machine learning over logistic regression for clinical prediction models. *Journal of clinical epidemiology*, Elsevier, v. 110, p. 12–22, 2019. Citado na página 28.

DIAS, J.; RODRIGUES, C. M.; RODRIGUES, A. C. Use and interpretation of item response theory applied to machine learning. In: RIBEIRO, P. R. d. A.; COTA, V. R.; BARONE, D. A. C.; OLIVEIRA, A. C. M. de (Ed.). *Computational Neuroscience*. Cham: Springer International Publishing, 2022. p. 15–24. ISBN 978-3-031-08443-0. Citado na página 15.

DRUMMOND, C.; JAPKOWICZ, N. Warning: statistical benchmarking is addictive. kicking the habit in machine learning. *Journal of Experimental & Theoretical Artificial Intelligence*, Taylor & Francis, v. 22, n. 1, p. 67–80, 2010. Citado na página 28.

EBASHI, S.; ENDO, M.; OHTSUKI, I. Control of muscle contraction. *Quarterly reviews of biophysics*, Cambridge University Press, v. 2, n. 4, p. 351–384, 1969. Citado 3 vezes nas páginas 19, 20 e 21.

EMBRETSON, S. E.; REISE, S. P. *Item response theory for psychologists. Maheah*. [S.l.]: New Jersey: Lawrence Erlbaum Associates, Publishers, 2000. Citado na página 26.

EMBRETSON, S. E.; REISE, S. P. *Item response theory*. [S.l.]: Psychology Press, 2013. Citado na página 15.

FERREIRA, A. de S.; GUIMARÃES, F. S.; SILVA, J. G. Aspectos metodológicos da eletromiografia de superfície: considerações sobre os sinais e processamentos para estudo da função neuromuscular. *Revista Brasileira de Ciências do Esporte*, v. 31, n. 2, 2010. Citado 2 vezes nas páginas 20 e 21.

FERRI, C.; HERNÁNDEZ-ORALLO, J.; MODROIU, R. An experimental comparison of performance measures for classification. *Pattern Recognition Letters*, Elsevier, v. 30, n. 1, p. 27–38, 2009. Citado na página 27.

GUZMÁN-MUÑOZ, E.; MÉNDEZ-REBOLLEDO, G. Electromiografía en las ciencias de la rehabilitación. *Revista Salud Uninorte*, Fundación Universidad del Norte, v. 34, n. 3, p. 753–765, 2018. Citado 3 vezes nas páginas 19, 20 e 21.

HUANG, H.; ZHOU, P.; LI, G.; KUIKEN, T. A. An analysis of emg electrode configuration for targeted muscle reinnervation based neural machine interface. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, IEEE, v. 16, n. 1, p. 37–45, 2008. Citado na página 15.

JONG, T. D. Cognitive load theory, educational research, and instructional design: some food for thought. *Instructional science*, Springer, v. 38, n. 2, p. 105–134, 2010. Citado na página 32.

JUNIOR, F. d. J. M. Aplicações da teoria da resposta ao item (tri) no brasil. *Revista Brasileira de Biometria*, v. 28, n. 4, p. 137–70, 2010. Citado na página 26.

KLINE, A.; KLINE, T.; ABAD, Z. S. H.; LEE, J. Novel feature selection for artificial intelligence using item response theory for mortality prediction. In: IEEE. *2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*. [S.l.], 2020. p. 5729–5732. Citado na página 27.

KLINE, A.; KLINE, T.; ABAD, Z. S. H.; LEE, J. Using item response theory for explainable machine learning in predicting mortality in the intensive care unit: Case-based approach. *Journal of Medical Internet Research*, JMIR Publications Inc., Toronto, Canada, v. 22, n. 9, p. e20268, 2020. Citado na página 28.

LALOR, J. P.; WU, H.; YU, H. Building an evaluation scale using item response theory. In: NIH PUBLIC ACCESS. *Proceedings of the Conference on Empirical Methods in Natural Language Processing. Conference on Empirical Methods in Natural Language Processing*. [S.l.], 2016. v. 2016, p. 648. Citado na página 27.

LANGLEY, P. *The changing science of machine learning*. [S.l.]: Springer, 2011. Citado na página 28.

LINDEN, W. J. van der. Constrained adaptive testing with shadow tests. In: *Elements of adaptive testing*. [S.l.]: Springer, 2009. p. 31–55. Citado na página 27.

LUCA, C. D. Electromyography. *Encyclopedia of medical devices and instrumentation*, Wiley Online Library, 2006. Citado na página 19.

MARTÍNEZ-PLUMED, F.; PRUDÊNCIO, R. B.; MARTÍNEZ-USÓ, A.; HERNÁNDEZ-ORALLO, J. Making sense of item response theory in machine learning. In: *Proceedings of the Twenty-second European Conference on Artificial Intelligence*. [S.l.: s.n.], 2016. p. 1140–1148. Citado 2 vezes nas páginas 15 e 27.

MARTÍNEZ-PLUMED, F.; PRUDÊNCIO, R. B.; MARTÍNEZ-USÓ, A.; HERNÁNDEZ-ORALLO, J. Item response theory in ai: Analysing machine learning classifiers at the instance level. *Artificial Intelligence*, Elsevier, v. 271, p. 18–42, 2019. Citado 2 vezes nas páginas 27 e 28.

MENDEZ, V.; LHOSTE, C.; MICERA, S. Emg data augmentation for grasp classification using generative adversarial networks. In: IEEE. *2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*. [S.l.], 2022. p. 3619–3622. Citado na página 26.

MENEGHETTI, D. D. R. *Metodologia de seleção de itens em Testes Adaptativos Informatizados baseada em Agrupamento por Similaridade*. Tese (Doutorado), 08 2015. Citado na página 40.

MILLS, K. R. The basics of electromyography. *Journal of Neurology, Neurosurgery & Psychiatry*, BMJ Publishing Group Ltd, v. 76, n. suppl 2, p. ii32–ii35, 2005. Citado na página 19.

NAZMI, N.; RAHMAN, M. A. A.; YAMAMOTO, S.-I.; AHMAD, S. A.; ZAMZURI, H.; MAZLAN, S. A. A review of classification techniques of emg signals during isotonic and isometric contractions. *Sensors*, MDPI, v. 16, n. 8, p. 1304, 2016. Citado 3 vezes nas páginas 22, 23 e 24.

NUNES, L.; MAGALHÃES, A.; MADEIRA, S.; NUNES, D.; NOGUEIRA, D.; PASSOS, M.; MACEDO, E. Sistemas pictográficos de comunicação alternativa para portadores de paralisia cerebral. In: *IV Congresso RIBIE, Brasília*. [S.l.: s.n.], 1998. Citado na página 14.

OCARINO, J. de M.; SILVA, P. L. P. da; VAZ, D. V.; AQUINO, C. F. de; BRÍCIO, R. S.; FONSECA, S. T. da. Eletromiografia: interpretação e aplicações nas ciências da reabilitação. Citado 2 vezes nas páginas 19 e 21.

PLIAKOS, K.; JOO, S.-H.; PARK, J. Y.; CORNILLIE, F.; VENS, C.; NOORTGATE, W. Van den. Integrating machine learning into item response theory for addressing the cold start problem in adaptive learning systems. *Computers & Education*, Elsevier, v. 137, p. 91–103, 2019. Citado na página 27.

PRUDÊNCIO, R. B.; HERNÁNDEZ-ORALLO, J.; MARTINEZ-USÓ, A. Analysis of instance hardness in machine learning using item response theory. In: *Second International Workshop on Learning over Multiple Contexts in ECML*. [S.l.: s.n.], 2015. Citado na página 15.

RASCH, G. Studies in mathematical psychology: I. probabilistic models for some intelligence and attainment tests. Nielsen & Lydiche, 1960. Citado na página 27.

SAFRIT, M. J.; COHEN, A. S.; COSTA, M. G. Item response theory and the measurement of motor behavior. *Research Quarterly for Exercise and Sport*, Taylor & Francis, v. 60, n. 4, p. 325–335, 1989. Citado 2 vezes nas páginas 30 e 31.

SANTOS, R. F. dos; SAMPAIO, P. Y. S.; SAMPAIO, R. A. C.; GUTIERREZ, G. L.; ALMEIDA, M. A. B. de. Tecnologia assistiva e suas relações com a qualidade de vida de pessoas com deficiência. *Revista de Terapia Ocupacional da Universidade de São Paulo*, v. 28, n. 1, p. 54–62, 2017. Citado na página 14.

SENDERS, J. T.; STAPLES, P. C.; KARHADE, A. V.; ZAKI, M. M.; GORMLEY, W. B.; BROEKMAN, M. L.; SMITH, T. R.; ARNAOUT, O. Machine learning and neurosurgical outcome prediction: a systematic review. *World neurosurgery*, Elsevier, v. 109, p. 476–486, 2018. Citado na página 28.

SODERBERG, G. L.; KNUTSON, L. M. A guide for use and interpretation of kinesiologic electromyographic data. *Physical therapy*, Oxford University Press, v. 80, n. 5, p. 485–498, 2000. Citado na página 19.

STURMA, A.; ROCHE, A. D.; GÖBEL, P.; HERCEG, M.; GE, N.; FIALKA-MOSER, V.; ASZMANN, O. A surface emg test tool to measure proportional prosthetic control. *Biomedical Engineering/Biomedizinische Technik*, De Gruyter, v. 60, n. 3, p. 207–213, 2015. Citado na página 33.

TOLEDO-PÉREZ, D. C.; RODRÍGUEZ-RESÉNDIZ, J.; GÓMEZ-LOENZO, R. A.; JAUREGUI-CORREA, J. Support vector machine-based emg signal classification techniques: A review. *Applied Sciences*, MDPI, v. 9, n. 20, p. 4402, 2019. Citado 2 vezes nas páginas 22 e 23.

TSINGANOS, P.; CORNELIS, B.; CORNELIS, J.; JANSEN, B.; SKODRAS, A. Data augmentation of surface electromyography for hand gesture recognition. *Sensors*, MDPI, v. 20, n. 17, p. 4892, 2020. Citado 3 vezes nas páginas 15, 24 e 25.

VYGOTSKY, L. S. *A formação social da mente. Tradução de José Cipolla Neto et al.* [S.l.]: São Paulo: Martins Fontes, 1994. Citado na página 32.

YANG, W.; YANG, D.; LI, J.; LIU, Y.; LIU, H. Emg dataset augmentation approaches for improving the multi-dof wrist movement regression accuracy and robustness. In: IEEE. *2018 IEEE International Conference on Robotics and Biomimetics (ROBIO)*. [S.l.], 2018. p. 1268–1273. Citado na página 25.

ZANINI, R. A.; COLOMBINI, E. L. Parkinson's disease emg data augmentation and simulation with dcgans and style transfer. *Sensors (Basel, Switzerland)*, Multidisciplinary Digital Publishing Institute (MDPI), v. 20, n. 9, 2020. Citado 2 vezes nas páginas 15 e 16.

ZANINI, R. A.; COLOMBINI, E. L. Parkinson's disease emg data augmentation and simulation with dcgans and style transfer. *Sensors*, MDPI, v. 20, n. 9, p. 2605, 2020. Citado na página 25.

ZHANG, S.; CHANG, H.-H. From smart testing to smart learning: How testing technology can assist the new generation of education. *International Journal of Smart Technology and Learning*, Inderscience Publishers (IEL), v. 1, n. 1, p. 67–92, 2016. Citado na página 27.