

Priscila Satomi Acamine

Estimação de medidas do centro de pressão utilizando Inteligência Artificial

São Bernardo do Campo

Dezembro de 2020

Priscila Satomi Acamine

Estimação de medidas do centro de pressão utilizando Inteligência Artificial

Relatório Final de Trabalho de Graduação
apresentado à Universidade Federal do ABC
como parte dos requisitos para a conclusão
da graduação em Engenharia Biomédica

Universidade Federal do ABC – UFABC

Engenharia Biomédica

Orientador: Prof. Dr. Renato Naville Watanabe

São Bernardo do Campo

Dezembro de 2020

Sumário

	Resumo	3
1	Introdução	4
2	Objetivos	8
2.1	Objetivo Geral	8
2.2	Objetivos Específicos	8
3	Metodologia	9
3.1	Base de dados	9
3.2	Estimativa do espectro do Centro de Pressão	10
3.3	Extract, Transform and Load (ETL) dos dados	12
3.4	Estrutura do modelo de classificação	13
3.5	Treino, Validação e Teste das Redes Neurais Artificiais	14
3.6	Análise e Seleção das Redes Neurais Artificiais	16
4	Resultados	18
4.1	Tratamento dos dados	18
4.2	Análise das Redes Neurais Artificiais	21
5	Discussão e Conclusão	24
	Referências	26

Resumo

O controle postural é requisito importante para a execução de diversas atividades no dia-a-dia e o estudo desta tarefa envolve uma plataforma de força de custo elevado para a análise de características como o centro de pressão. Deste modo, o objetivo deste trabalho é desenvolver um método para estimar medidas do centro de pressão de uma maneira menos custosa, utilizando-se uma abordagem de inteligência artificial. Com uma base de dados de 207 indivíduos, diversas Redes Neurais Artificiais para classificação multiclasse de estrutura simples foram treinadas e testadas a partir de 24 características dos sujeitos. As redes classificaram a frequência mediana do COP (PF50) de indivíduos em tarefa de postura ereta quieta, em superfície rígida com os olhos fechados, em 4 classes. Um método de votação foi implementado, combinando-se os resultados das redes neurais obtidas de melhor desempenho. Ao final, o método resultou em uma acurácia de 62,5% no teste. O método de votação proposto neste trabalho mostrou conseguir melhorar a predição final do modelo em comparação às acurácias individuais das Redes Neurais Artificiais e representa uma inovação de aplicação do *machine learning* em dados de controle postural.

Palavras-chave: rede neural artificial, centro de pressão, PF50, controle postural, inteligência artificial.

1 Introdução

O controle da postura ereta bípede, apesar de aparentar ser uma tarefa bastante simples, é um grande desafio para o corpo humano, evidenciado principalmente nos idosos e em patologias motoras e do sistema vestibular. Para manter a postura ereta estática, o sistema de controle postural procura manter a projeção horizontal do centro de gravidade (center of gravity - COG) do indivíduo dentro da base de suporte definida pela área dos pés, regulando o equilíbrio através das informações sensoriais recebidas. Este controle é afetado pela natureza da tarefa, condições ambientais, informações sensoriais disponíveis e condições intrínsecas da pessoa [3].

Durante a postura ereta, o corpo não permanece sem movimento, existe uma oscilação de baixa frequência, que é normalmente investigada utilizando-se uma plataforma de força. Uma das variáveis mais importantes neste tipo de análise é a posição do centro de pressão (center of pressure - COP). O COP é definido como sendo o ponto de aplicação da resultante das forças verticais que atuam na superfície de suporte e seu deslocamento representa a somatória das ações do sistema de controle postural e da força da gravidade [3]. A Figura 1, ilustra um sujeito realizando a tarefa de postura ereta bípede, em que permanece parado em posição ereta durante um certo período de tempo sobre a plataforma de força, que quantifica essa oscilação, medindo-se os níveis de estabilidade do sujeito [3].

A medida e o registro da contínua oscilação do corpo humano durante a postura ereta é chamada de estabilometria - estabilografia ou estatocinesiografia - e tem sido por décadas a principal ferramenta biomecânica para o entendimento do equilíbrio corporal [16]. Um estabilograma é o mapeamento do COP ântero-posterior (a-p) versus o COP

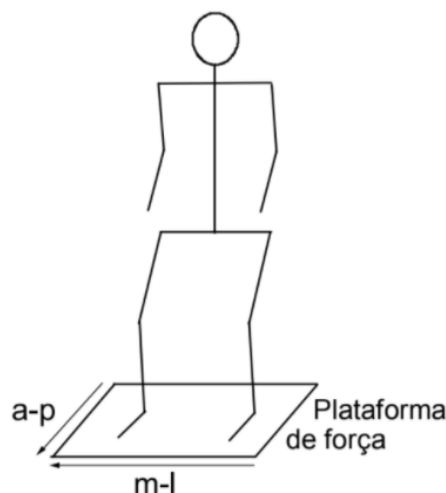


Figura 1: Tarefa de postura ereta quieta (Adaptado de Duarte,2000)

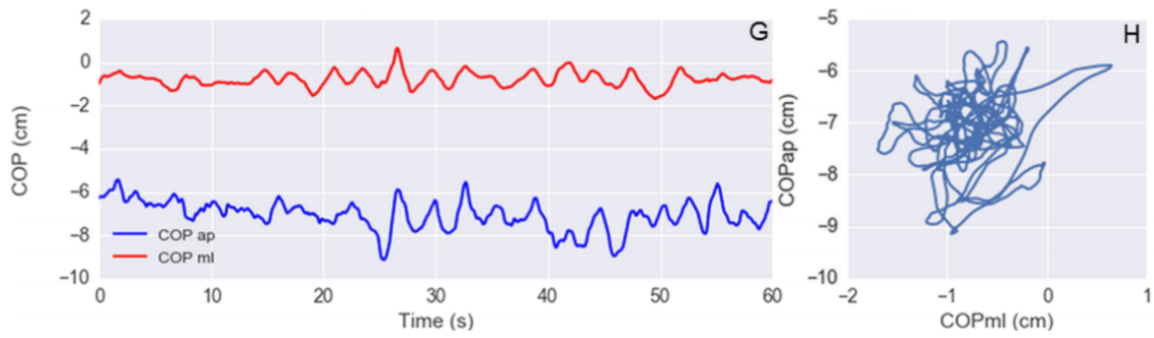


Figura 2: Gráficos de deslocamento do COP (Adaptado de Santos e Duarte, 2016).

médio-lateral (m-l) representativo durante a postura ereta quieta. A Figura 2 mostra o deslocamento do COP nas direções a-p e m-l versus o tempo e o gráfico do COP a-p versus o COP m-l [13].

Dentre várias medidas existentes para caracterizar a estabilidade postural está o PF50. O PF50 representa a frequência mediana, ou seja, a frequência em que, abaixo dela, 50% da potência total é encontrada. Ele tem a habilidade antecipada de estimar a distribuição do espectro de potência do COP, sendo uma boa medida para caracterização dos indivíduos e para investigação [18, 11].

A tarefa da postura ereta quieta é de importância para diversas atividades no dia-a-dia. Por isso, entendê-la vem a ser importante para possibilitar possíveis intervenções e análises clínicas. Uma barreira para o desenvolvimento deste tipo de estudo seria o alto custo e baixa acessibilidade de plataformas de força e laboratórios bem equipados para estudo do movimento humano. Há estudos propondo a utilização de alternativas mais baratas, como o Nintendo Wii Balance Board, à essa plataforma de força [7]. Outra alternativa para este problema seria a utilização de uma abordagem de aprendizado de máquina (*machine learning*), utilizando Redes Neurais Artificiais.

As redes neurais artificiais representam uma tentativa de imitar conjuntos formados pelas células neurais, suas características básicas e organização. Podemos dizer que estes modelos possuem uma inteligência artificial, sendo possível aprender, errar e aprender com o erro. O conhecimento é adquirido pela rede a partir de seu ambiente através de um processo de aprendizagem e é armazenado nas forças de conexão entre neurônios (pesos sinápticos) [4].

Um modelo simples não-linear de um neurônio artificial, que forma a base para o projeto de Redes Neurais Artificiais, é apresentado na Figura 3. Neste esquema podem ser identificados três elementos básicos do modelo neuronal:

1. Um conjunto de sinais de entrada (x_j), cada um caracterizado por um peso sináptico (w_{kj}) próprio, que se conectam ao neurônio k . O peso sináptico de um neurônio

artificial pode estar em um intervalo que inclui valores negativos e positivos.

2. Um somador para os sinais de entrada, ponderados pelos respectivos pesos sinápticos.

$$v_k = \sum_{j=0}^i w_{kj} x_j \quad (1.1)$$

3. Uma função de ativação não-linear para restringir a amplitude da saída de um neurônio.

$$y_k = \varphi(v_k) \quad (1.2)$$

Há, ainda, a ilustração do cálculo do erro, em que compara-se, ao final, a saída do neurônio (y_k) com o valor alvo (t_k).

$$\epsilon_k = t_k - y_k \quad (1.3)$$

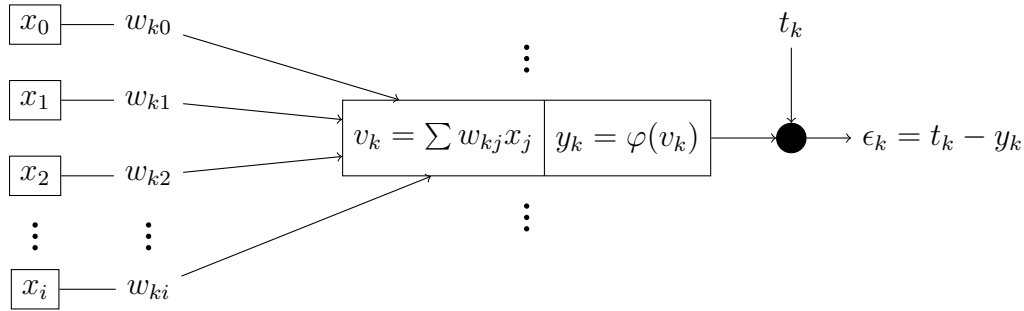
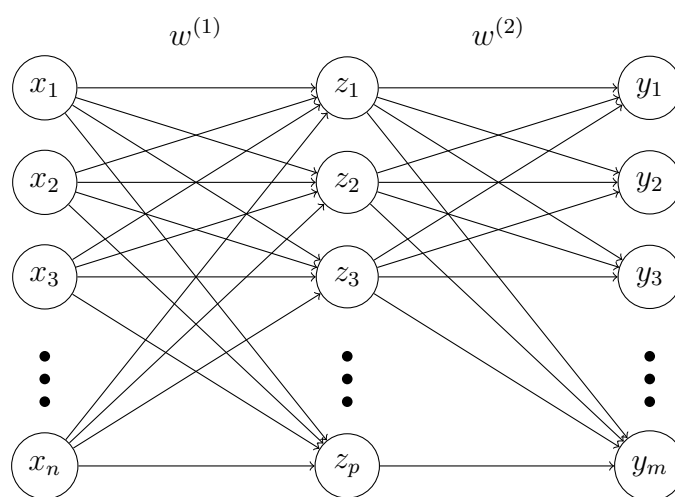


Figura 3: Modelo de neurônio artificial.

Um tipo comum de arquitetura de rede são as redes alimentadas diretamente com múltiplas camadas, também chamadas de perceptrons de múltiplas camadas (multi-layer perceptron - MLP). Neste tipo de rede, há a presença de uma ou mais camadas ocultas de neurônios, que ficam entre a camada de entrada e a de saída [4]. A Figura 4 mostra um exemplo de Rede Neural Multicamada totalmente conectada. Isso significa dizer que cada um dos nós de uma camada da rede está conectado a todos os nós da camada adjacente seguinte.

O Teorema da Aproximação Universal afirma que um perceptron de múltiplas camadas com uma única camada intermediária é capaz de aproximar qualquer mapeamento dado por uma função contínua, dado um conjunto de treinamento suficientemente significativo para representar a função. Na teoria, não existiriam limitações em redes com uma camada oculta, entretanto, na prática, podemos obter sucesso limitado com essas arquiteturas [5].



Camada de Entrada Camada Oculta Camada de Saída

Figura 4: Ilustração de uma Rede Neural de Múltiplas Camadas.

2 Objetivos

2.1 Objetivo Geral

- Desenvolver um método baseado em Inteligência Artificial para estimar a frequência mediana do centro de pressão (PF50) a partir de dados individuais dos sujeitos.

2.2 Objetivos Específicos

- Calcular o espectro e a frequência mediana do centro de pressão (COP) de indivíduos em uma tarefa de controle postural;
- Implementar uma arquitetura de Rede Neural Artificial para classificação multiclasse;
- Implementar um método de votação do resultado a partir da combinação de várias redes simples;
- Disponibilizar um código aberto do cálculo das estimativas das medidas do COP a partir de características individuais dos sujeitos analisados.

3 Metodologia

3.1 Base de dados

Foram analisados os registros provenientes de dois repositórios públicos de dados com avaliações relacionadas ao equilíbrio humano, elaborados por Santos, Duarte (2016) e Dos Santos et al. (2017). Os dados encontram-se disponíveis na PhysioNet (DOI: 10.13026/C2WW2W) e ambos na Figshare (DOI: 10.6084/m9.figshare.3394432 e DOI: 10.6084/m9.figshare.4525082).

O primeiro conjunto de dados é composto por uma amostra de 163 sujeitos (116 mulheres e 47 homens), que passaram por uma avaliação estabilográfica [13]. O segundo repositório é composto de dados da tarefa de postura ereta quieta de 49 indivíduos (23 mulheres e 26 homens) [2]. Ao todo, tem-se dados de 212 indivíduos. As idades variam entre 18 e 85 anos, a massa corporal, de 44 a 126,3 kg, alturas de 140,0 a 192,0 cm. Ao total, em ambos os estudos, 23 dos participantes apresentavam pelo menos uma ou mais deficiência severa dentre déficits auditivos, vestibulares, visuais, musculoesqueléticos e intelectuais.

A aquisição dos dados em ambos repositórios seguiu o mesmo protocolo e as mesmas condições de tarefas. A tarefa consistiu em manter-se parado durante 60s em quatro condições distintas: em uma superfície rígida com os olhos abertos, em uma superfície rígida com os olhos fechados, em uma superfície instável (bloco de espuma de 6 cm) com os olhos abertos e na superfície instável com os olhos fechados. Cada condição foi realizada três vezes em ordem randomizada [13].

Em todas as condições, os sujeitos foram instruídos a permanecer de pé, descalços e o mais imóvel possível, com os braços ao lado do corpo, e olhar direcionado a um alvo redondo de 5 cm colocado na altura dos olhos dos sujeitos em uma parede 3 m à frente. Nas condições de olhos fechados, após o indivíduo encontrar uma postura estável e confortável, dadas as exigências, ele era orientado a fechar os olhos [13].

Os sinais foram coletados utilizando uma plataforma de força comercial de 40 cm x 60 cm (OPT400600-1000; AMTI, Watertown, MA, EUA) e um amplificador (Optima Signal Conditioner; AMTI, Watertown, MA, EUA) a uma frequência de amostragem de 100 Hz. A placa de força da marca Optima foi calibrada na fábrica e apresentou uma precisão média de COP de 0,02 cm [13].

3.2 Estimativa do espectro do Centro de Pressão

Em certas situações, o conteúdo de frequência de um sinal fornece informações mais úteis do que a representação no domínio do tempo. Muitos sinais biológicos demonstram propriedades interessantes ou úteis para o diagnóstico quando visualizadas no domínio da frequência. Essa determinação do conteúdo de frequência de uma forma de onda é chamada de análise espectral [15].

A análise espectral decompõe uma forma de onda no domínio do tempo em suas frequências constituintes. Os métodos clássicos são baseados na Transformada de Fourier que decompõe um sinal em suas componentes elementares senoidais equivalentes [15].

A determinação precisa do espectro da forma de onda exige que o sinal seja periódico, de comprimento finito e ruído livre. Normalmente, apenas uma parte dos bio-sinais está disponível para análise e estes são frequentemente corrompidos por ruídos ou artefatos. Nestes casos, as técnicas de análise espectral devem ser necessariamente aproximadas, isto é, são estimativas do verdadeiro espectro [15].

A Transformada Discreta de Fourier (discrete fourier transform - DFT) utiliza um número finito de pontos no domínio do tempo e define uma representação discreta do sinal no domínio da frequência. A discretização da função de interesse é bastante desejável principalmente em aplicações que se baseiam em técnicas digitais. Para a estimação de espectro a partir da DFT, assume-se que o dado representa uma função aperiódica, ou seja, que a série temporal original é sempre nula exceto durante o período de tempo capturado no computador. A partir desta premissa, a DFT faz uma boa aproximação para os coeficientes da Série de Fourier de funções periódicas [15].

Na DFT, a resolução é igual a frequência de amostragem dividida pelo número total de pontos do sinal digitalizado. Assim, para uma determinada frequência de amostragem, quanto maior o número de amostras no sinal, maior será a resolução espectral [15].

O periodograma é o método em que a DFT é calculada em segmentos isolados do dado seguido pelo cálculo da média desses espectros, que produz do espectro global representante do dado [1]. Essa abordagem é mais utilizada quando a forma de onda disponível é apenas uma amostra de um sinal maior. Nesse caso, a análise espectral é necessariamente um processo de estimação e o cálculo de uma média aumenta a propriedade estatística do resultado, ou seja, aumenta a confiabilidade da estimativa do espectro. Uma desvantagem do periodograma em relação a simples aplicação da DFT é que perde-se resolução espectral, já que a segmentação do dado reduz o número de amostras analisadas ao número em cada segmento [15].

O espectro de potência é usualmente definido como a Transformada de Fourier da função de autocorrelação. A densidade espectral de potência representa a função de densidade da energia através da frequência, ou seja, descreve como a energia de um sinal

ou de uma série temporal será distribuída na frequência. Ela pode ser calculada como sendo o quadrado da magnitude da transformada de Fourier da função de onda de interesse. Um dos métodos de média espectral mais utilizados para a análise do espectro de potência é o periodograma de Welch [15].

Tradicionalmente nos periodogramas utilizam-se segmentos com sobreposição de 50% em janelas não retangulares. A função de janelamento produz efeitos no espectro do sinal até mesmo quando janelas retangulares são utilizadas. No janelamento, o espectro original é convoluído com o espectro da janela. Como o espectro de uma senóide é um impulso na frequência da mesma, a convolução de um espectro qualquer com esse impulso desloca-o para a posição do impulso. Os efeitos observados são o truncamento da senóide, necessário para o cálculo da DFT e o deslocamento do espectro da janela para a posição da frequência da senóide. As funções de janelamento mais utilizadas são: retangular, de Bartlett, de Hamming, de Hanning e de Blackman-Harris [15, 8]. Na biblioteca `scipy.signal` do Python, a função do periodograma de Welch utiliza como padrão a janela de Hanning [14].

A partir da estimativa do espectro do COP, poderá ser obtido o valor do PF50 de cada indivíduo. O índice PF50 representa a frequência mediana ou a frequência que divide ao meio a potência total do COP [18]. Para esse fim, o menor valor de frequência discretizada (u) que satisfaça a seguinte equação:

$$\sum_{m=1}^u m \Delta f G[m] \geq 0,5 \text{Potência}, \quad (3.1)$$

é calculado. Então o PF50 é calculado como

$$\text{PF50} = u \Delta f \quad (3.2)$$

onde Δf é a frequência de amostragem.

Quanto maior a potência em baixas frequências, menores serão os valores de PF50. Em geral, características universais dos padrões de COP parecem estar associadas com os componentes lentos de baixa frequência do balanço [18]. Prieto et al. (1996) relataram que esses índices eram independente da idade e condições oculares [11].

A estimativa do espectro do COP e o índice PF50 de cada trial foram calculados como descrito. Calculou-se, então, a média do PF50 de cada indivíduo em cada uma das diferentes condições de visão e superfície realizados nos experimentos. Cada condição da tarefa realizada foi separada em conjunto de dados distintos sendo eles: `open_rigid` (superfície rígida com os olhos abertos), `closed_rigid` (superfície rígida com os olhos fechados), `open_foam` (superfície de espuma com os olhos abertos) e `closed_foam` (superfície de espuma com os olhos fechados).

3.3 Extract, Transform and Load (ETL) dos dados

Os dois conjuntos de dados disponíveis foram unificados, selecionando apenas as características que ambos tinham em comum. Desta forma, informações que só haviam sido descritas em uma das bases de dados tiveram de ser excluídas da análise. Divergências relacionadas a nomenclaturas (de rótulos das características e de identificação dos sujeitos) e unidades de medidas foram resolvidas entre as duas bases de dados.

A unificação das duas bases de dados totalizou informações de 2518 tarefas de postura ereta realizadas por um total de 212 sujeitos, sendo que cada tarefa possuía 25 características descritoras, relacionadas tanto a parâmetros da tarefa quanto ao sujeito que a realizou ('Trial', 'Subject', 'Vision', 'Surface', 'Age', 'AgeGroup', 'Gender', 'Height', 'Weight', 'BMI', 'FootLen', 'Nationality', 'SkinColor', 'Ystudy', 'Footwear', 'Illness', 'Illness2', 'Nmedication', 'Medication', 'Ortho-Prosthesis', 'Ortho-Prosthesis2', 'Disability', 'Disability2', 'Falls12m', 'Date').

Para que as características possam ser utilizadas como entradas da Rede Neural é necessário que as informações sejam dadas de forma numérica, uma vez que a rede não consegue manipular palavras. Desta forma, as características qualitativas tiveram de ser ajustadas para se adequarem ao propósito de auxiliarem na predição da Rede.

Cada uma dessas categorias foi tratada, listando-se todas as variações descritas e transformando-as em colunas individuais de uma tabela preenchida por zeros e uns, em que o 0 representa a ausência e 1, a presença, de determinada variação para cada sujeito. Para as características quantitativas (Idade, Altura, Massa Corporal, índice IMC, etc.) não foi necessário fazer alterações.

As colunas "*Medication*", "*Ortho-Prothesis2*", "*Footwear*", "*Illness2*" e "*Disability2*" contém informações qualitativas dos variados tipos de medicamentos, orto-próteses e calçados que os indivíduos utilizam e enfermidades e deficiências que possuem, respectivamente. A coluna "*Medication*" descreve todos os medicamentos que o sujeito utiliza de forma contínua. Foram listados todos os medicamentos descritos para todos os indivíduos, resolvendo-se casos de diferenças de grafia para o mesmo medicamento, o que totalizou em 67 variações. O rótulo "*Medication*" foi substituído, então, por uma tabela de 67 tipos diferentes de medicamentos em que, para cada sujeito, o índice 1 representa o uso do determinado medicamento, enquanto o 0 representa a não utilização do mesmo. Para os demais rótulos qualitativos obteve-se 17 variações de orto-próteses, 5 tipos de calçados, 77 enfermidades diferentes e 4 tipos de deficiência.

Outros rótulos qualitativos, como "*Gender*", que especifica o sexo do sujeito, foi definido como 0 para homens e 1 para mulheres. O rótulo "*AgeGroup*", que especifica o grupo de idade do indivíduo, foi definido como 0 para "Jovem" e 1 para "Idoso". Além disso, "*Illness*", "*Ortho-Prosthesis*" e "*Disability*", que indicam se o indivíduo possui doenças,

orto-próteses e deficiências, respectivamente, foram definidos como 0 no caso de ausência, ou 1, no caso de presença.

Após o tratamento da base de dados, foram excluídas as características que apareciam em menos de 10% dos indivíduos, pois estas teriam pouca contribuição no treinamento da rede. As características de “*Nationality*” e “*SkinColor*” foram desconsideradas da análise, pois julgou-se que tais características não estariam relacionadas à medida PF50.

Foram excluídos de cada conjunto de dados os sujeitos que haviam realizado a condição em questão de forma incompleta (menos de 3 trials). Além disso, foram retirados dados de sujeitos que apresentavam um valor médio de PF50 inconsistente (*outliers*) para cada condição. O critério utilizado para definir os outliers foi o do z-score, adotando um limiar de 3. Essa medida foi tomada pois valores atípicos atrapalham no treinamento das Redes Neurais Artificiais.

A média do PF50 de cada indivíduo foi classificada em 4 grupos. A determinação de cada classe foi feita baseada nos quartis (Q1, Q2 e Q3) da base de treinamento e validação de cada condição para Visão e Superfície. Os quartis são valores dados a partir do conjunto de observações ordenado em ordem crescente, que dividem a distribuição em quatro partes iguais. Esse método de divisão visa que as classes ficassem balanceadas.

3.4 Estrutura do modelo de classificação

Os dados foram analisados através de uma abordagem de Inteligência Artificial e Aprendizagem de Máquina. Implementou-se uma estrutura de Rede Neural Artificial (RNA) de classificação multiclasse com arquitetura de Perceptron de Múltiplas Camadas e método de aprendizagem supervisionada. Inúmeras redes foram treinadas nessa configuração de rede simples e posteriormente os resultados de cada uma foram analisados a verificar as que obtiveram maior desempenho. O algoritmo foi desenvolvido em linguagem de programação Python utilizando-se da biblioteca PyTorch (funções de ativação, de perda, de otimização e modelagem de redes neurais) e scikit-learn (escalamento de dados, divisão de grupos e matriz de confusão), que contém ferramentas próprias para o desenvolvimento de arquiteturas de *machine learning* e *deep learning*.

Foi desenvolvida uma estrutura base de Rede Neural Artificial com tamanho de entrada referente ao número de características utilizadas para a classificação (24 neurônios), 2 camadas ocultas com número aleatório de neurônios (variando de 1 a 24) e a camada de saída com 4 neurônios, correspondentes ao número de classes em que deseja-se classificar os valores de PF50 dos indivíduos (Figura 5). Todas as camadas da RNA foram definidas como sendo lineares, com função de ativação linear retificada (ReLU) nas camadas intermediárias e função *softmax* na última camada.

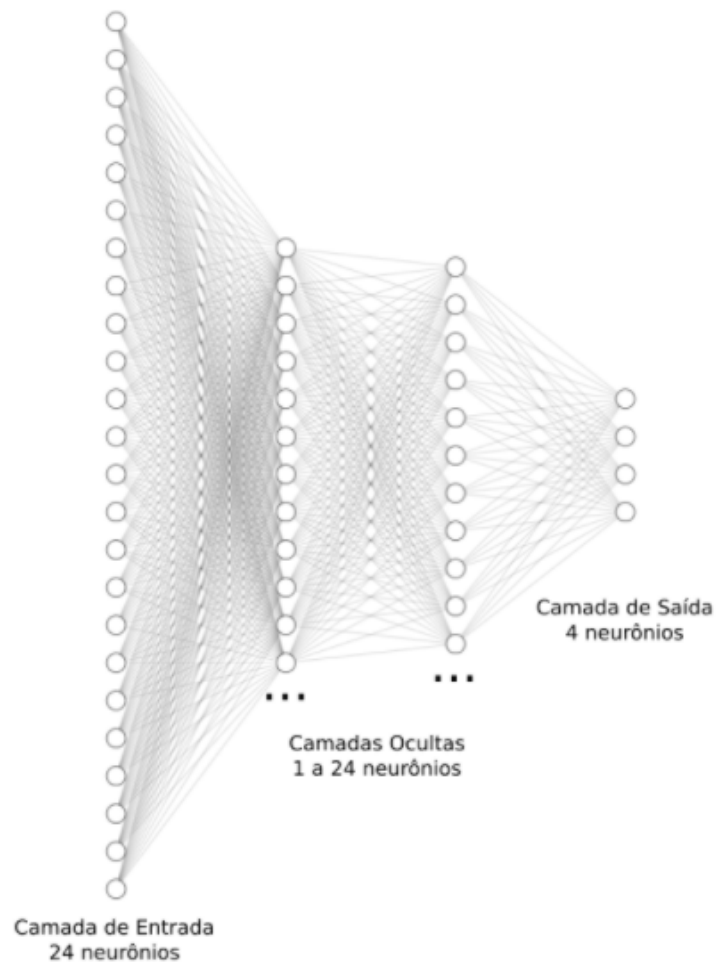


Figura 5: Estrutura da Rede Neural Artificial.

Foi utilizado como entrada da RNA as características individuais dos sujeitos, tais como idade, massa corporal, sexo, altura, e o parâmetro de saída é a classe a qual a frequência média do COP pertence.

3.5 Treino, Validação e Teste das Redes Neurais Artificiais

Foram utilizados apenas os dados referentes à condição em que o sujeito se manteve parado na superfície rígida com os olhos fechados (*closed_rigid*). Este conjunto de dados é composto de 207 sujeitos, cada um com as 24 características descritas.

O conjunto de dados foi dividido em três grupos: treino, validação e teste. O grupo de teste consistiu em 30% da base, os outros 70% compõem o grupo de desenvolvimento. Para cada RNA o grupo de teste se manteve o mesmo, mas a divisão do grupo de desenvolvimento foi aleatorizada, sendo que 80% deste foi utilizado para o treino e 20% para a validação. Ao final utilizou-se 56% dos dados para o treino, 14% para a validação e 20% para o teste.

Após a divisão das classes, as características de treino foram normalizadas utilizando uma estatísticaa robusta a *outliers*, chamada *Robust Scaler*. O *Robust Scaler* remove a mediana e dimensiona os dados de acordo com o intervalo interquartil (Intervalo entre o Q1 e Q3) [19]. Como o intervalo de valores dos dados brutos varia muito, em alguns algoritmos de aprendizado de máquina, as funções não funcionarão corretamente sem essa normalização. Além disso, com os dados normalizados, o gradiente descendente converge mais rápido, otimizando o modelo. O escalamento foi ajustado, então, ao dado de treinamento, realizando em seguida a transformação desses dados. Os dados de validação e os de teste também foram normalizados utilizando a mesma escala do dado de treino.

Finalizado o pré-processamento dos dados de entrada, definiu-se então, os parâmetros da RNA. Foi definido 4 camadas e a quantidade de neurônios em cada uma dessas camadas, sendo que o número de neurônios da primeira camada deve ser igual ao número de características escolhidas como entrada da rede e a última camada deve ter 4 neurônios, uma vez que desejamos como saída quatro classes, o maior valor de saída corresponde à classe predita do PF50 para o indivíduo.

A fase de treinamento do modelo consiste em ajustes dos pesos de cada camada da Rede Neural Artificial através do método de *error backpropagation*, que faz uma análise do erro baseado na diferença entre o valor real da medida e o valor predito pelo modelo.

Definiu-se a entropia cruzada (*Cross Entropy Loss*) como função de perda. Também chamada de perda logarítmica, ela mede o desempenho de um modelo de classificação cuja saída é um valor de probabilidade entre 0 e 1. A perda de entropia cruzada aumenta à medida que a probabilidade prevista diverge do rótulo real. A perda pode ser descrita pela função abaixo [17]:

$$loss(x, classe) = -\log \left(\frac{\exp(x[classe])}{\sum_j \exp(x[j])} \right) = -x[classe] + \log \left(\sum_j \exp(x[j]) \right) \quad (3.3)$$

Como otimizador, utilizou-se o algoritmo de Adam. Diferente do clássico *Stochastic Gradient Descent* (SGD), que mantém uma única taxa de aprendizado para todas as atualizações de pesos e não a muda durante o treinamento, este método calcula taxas de aprendizagem adaptativas individuais para diferentes parâmetros a partir de estimativas do primeiro e segundo momento dos gradientes.

A taxa de aprendizado inicial foi escolhida de forma aleatória entre 0,001 a 0,1. O treinamento da rede foi realizado ao longo de 50000 passos, sendo que a cada 100 passos, é feita uma validação da rede.

A partir dos dados de validação normalizados, calcula-se a acurácia, perda e correlação do modelo. O melhor modelo é salvo, utilizando-se o critério de melhor acurácia obtida na validação. Para este modelo, foram salvas as métricas de avaliação obtidas e

gravou-se a matriz de confusão, uma tabela que mostra as frequências de classificação para cada classe do modelo, apresentando os rótulos preditos em relação aos rótulos reais de cada indivíduo.

Além disso, calculou-se as mesmas métricas de avaliação para um grupo de dados (*surrogate*). O *surrogate* consiste nos dados de validação embaralhados de maneira aleatória, de forma que o desempenho da rede nesse grupo aleatório foi analisado para verificar se a rede estava de fato aprendendo.

Então, o melhor modelo salvo durante o treinamento é carregado e realiza-se o teste com os dados normalizados. Avaliou-se a acurácia, perda, correlação e matriz de confusão com dados totalmente novos, para verificar a capacidade de generalização do modelo. Para a função de perda, quanto menor o valor, menor é o erro da predição. Já para o coeficiente de correlação R, quanto mais próximo de 1, melhor é a performance.

Um laço foi criado para testar diferentes arquiteturas de redes. O algoritmo foi executado no parque computacional de alta performance da UFABC (HPC/UFABC) no Cluster Titânio. A Titânio possui 2620 *cores*, 10,4 Tb de RAM e 180 Tb de storage, com limite de 100Gb de uso por usuário [12].

Alguns parâmetros foram randomizados: tamanho da primeira e segunda camada intermediária (variando de 1 a 24 neurônios) e a taxa de aprendizado da rede (variando de 0,001 a 0,1). O grupo de desenvolvimento foi dividido em grupos de treino e validação utilizando uma função da biblioteca sklearn (`sklearn.model_selection.train_test_split`), que tem como um dos parâmetros o *random_state*, que é uma semente utilizada para randomizar tal divisão. Este parâmetro também foi randomizado no laço de RNA's, variando de 1 a 10000.

Todos os parâmetros necessários para a reprodutibilidade de cada rede foram gravados em um documento *xlsx*, assim como as métricas de avaliação correspondentes. Os próprios modelos PyTorch, com os parâmetros aprendíveis, foram salvos para possibilitar seu restauramento posteriormente. Um arquivo com o escalamento aplicado aos dados de treino, teste e validação também foi salvo para cada modelo treinado.

3.6 Análise e Seleção das Redes Neurais Artificiais

Para a análise das redes salvas, foi retirada uma amostra aleatória de 8 sujeitos do grupo de teste (*leave-8-out*). Calculou-se a acurácia do teste retirando-se essa amostra. O critério de seleção das redes com melhor desempenho baseou-se nessa métrica de avaliação. Foram selecionadas as redes com os maiores valores de acurácia.

O resultado das redes selecionadas passou, então, por um método de votação em que calcula-se a moda de todos os resultados combinados para cada indivíduo. Assumiu-se

o resultado da votação como predição final do modelo. O modelo final de votação foi avaliado e comparado aos resultados individuais de cada rede simples para verificar se houve uma melhora no desempenho da predição.

Ao final, a amostra separada que não foi utilizada na seleção de redes é testada no modelo de votação com as redes selecionadas e avaliada em relação às métricas de acurácia e correlação, assim como sua matriz de confusão.

4 Resultados

4.1 Tratamento dos dados

Após o tratamento dos dados, a base resultou em 212 sujeitos com 24 características descritoras mais o valor do PF50:

- ‘Age’: Idade do indivíduo (anos);
- ‘AgeGroup’: Grupo de Idade (jovem ou idoso);
- ‘Gender’: Sexo (masculino ou feminino);
- ‘Height’: Altura (cm);
- ‘Weight’: Peso (kg);
- ‘BMI’: Índice de massa corporal (IMC);
- ‘FootLen’: Tamanho do pé (cm);
- ‘Ystudy’: Anos de estudo;
- ‘Illness’: Portador de doenças;
- ‘Nmedication’: Número de medicações;
- ‘Ortho-Prosthesis’: Portador de orto-próteses;
- ‘Disability’: Portador de deficiência;
- ‘Falls12m’: Número de quedas nos últimos 12 meses;
- ‘Angiotensin II receptor antagonist’: Uso de antagonistas de receptores de angiotensina II;
- ‘HMG-CoA reductase inhibitor’: Uso de inibidor da HMG-CoA redutase;
- ‘Synthetic thyroid hormone’: Uso de hormônio sintético da tireóide;
- ‘Hypercholesterolemia_total’: Portador de Hipercolesterolemia;
- ‘Hypertension_total’: Portador de Hipertensão;
- ‘Hypothyroidism_total’: Portador de Hipotireoidismo;
- ‘Normal Shoes_total’: Uso de sapatos normais;
- ‘Sandal_FlipFlop_total’: Uso de sandália e chinelo;
- ‘Dental_total’: Uso de prótese dentária;
- ‘Corrective_lens’: Uso de lentes corretivas;

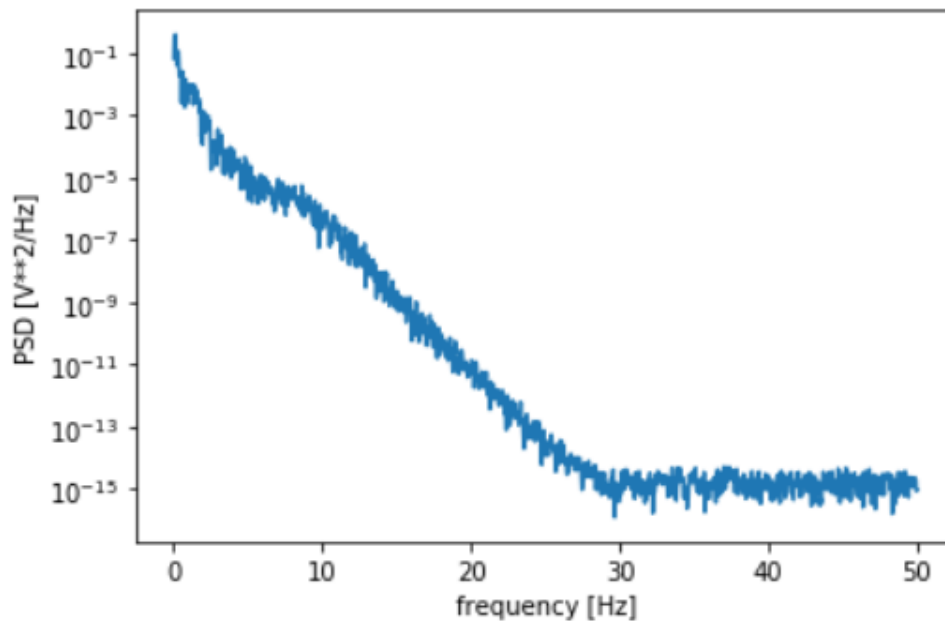


Figura 6: Gráficos de Densidade Espectral do COP (log) versus Frequência.

- ‘Hearing_dis_ortho_total’: Portador de deficiência ou prótese auditiva.

Para cada condição de realização da tarefa foram identificados e excluídos os sujeitos que não haviam concluído todos os trials sendo: 5 de closed_foam, 1 de closed_rigid e 3 de open_foam. Todos os sujeitos executaram de forma completa a condição de open_rigid. Além disso, foram excluídos os outliers da cada base.

A Figura 6 mostra um exemplo do gráfico do espectro calculado do COP a partir do qual o PF50 foi identificado.

Para o dataset de closed_rigid, que foi utilizado para o desenvolvimento das redes, os quartis foram calculados referente à base de dados separada para treinamento e validação. A Figura 7 apresenta o boxplot desses dados. Foi classificado como classe 0 os valores de PF50 abaixo de Q1, classe 1 entre Q1 e Q2, classe 2 entre Q2 e Q3, e classe 3 para os valores acima de Q3, conforme valores abaixo:

Classe 0: $PF50 < 0,13$

Classe 1: $0,13 \leq PF50 < 0,175$

Classe 2: $0,175 \leq PF50 < 0,2125$

Classe 3: $0,2125 \leq PF50$

A Figura 8 mostra o histograma dos dados após a classificação do PF50. Pode-se observar que a partir deste método de divisão, as classes ficam balanceadas, com número de observações parecido para cada classe.

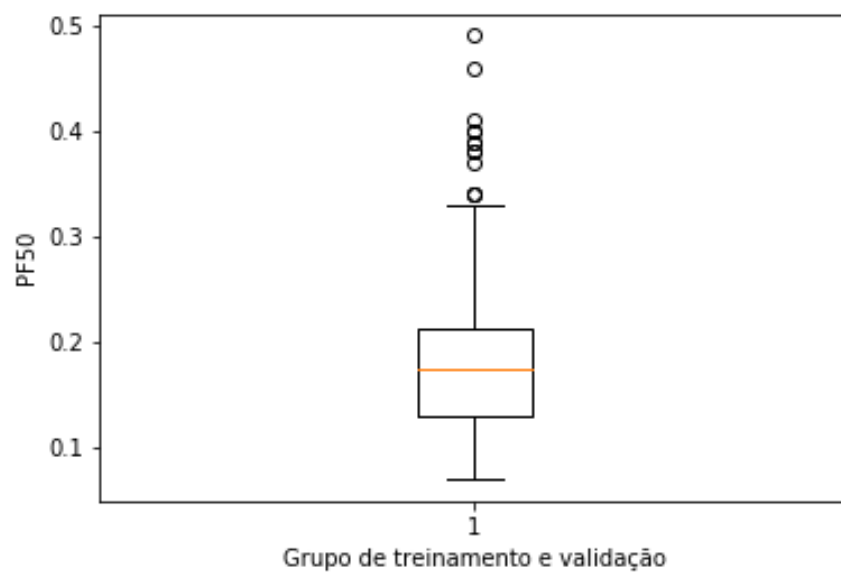


Figura 7: Boxplot dos dados de treinamento e validação.

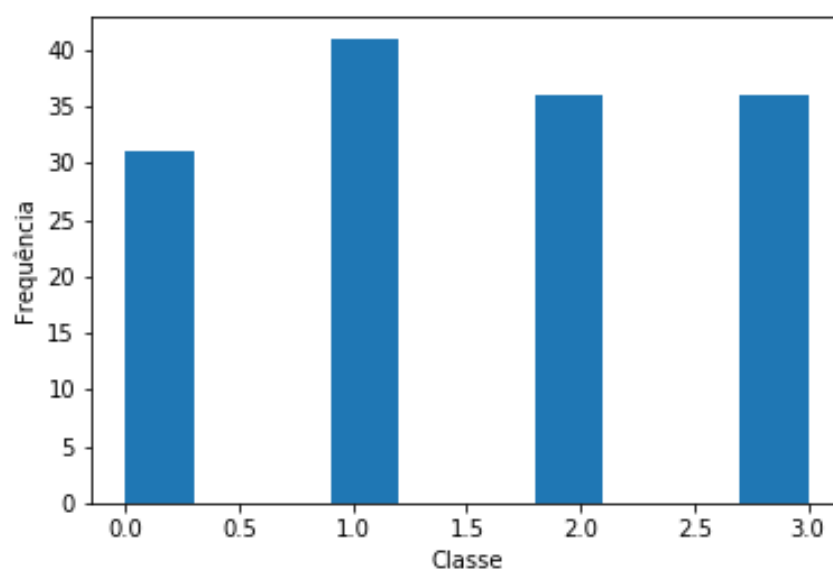


Figura 8: Histograma dos dados de treinamento e validação.

Rede	Tamanho das camadas ocultas	Taxa de aprendizado	Passo melhor modelo	Validação			Teste			Acurácia leave-8-out (%)
				Acurácia (%)	Perda	Correlação R	Acurácia (%)	Perda	Correlação R	
1	[13, 2]	0,079	8100	48	19,34	0,40	46	25,50	0,34	45,45
2	[4, 8]	0,022	7900	51	21,94	0,64	44	39,95	0,46	45,45
3	[3, 7]	0,002	10800	27	65,71	0,11	44	52,06	0,31	45,45
4	[11, 23]	0,031	1400	31	14,95	0,26	46	7,94	0,44	45,45
5	[12, 11]	0,068	1700	34	108,14	0,48	46	53,84	0,33	45,45
6	[12, 18]	0,041	5800	34	24,20	0,25	49	41,71	0,29	47,27
7	[21, 4]	0,034	500	27	4,61	0,36	49	8,42	0,48	45,45
8	[8, 6]	0,061	1400	48	8,40	0,43	46	14,73	0,25	45,45
9	[3, 7]	0,080	400	37	18,29	0,08	46	12,08	0,35	47,27
10	[7, 12]	0,080	100	37	22,60	0,12	46	11,52	0,35	47,27

Figura 9: Características das redes selecionadas.

4.2 Análise das Redes Neurais Artificiais

Ao todo, foram treinadas e salvas 47310 diferentes Redes Neurais Artificiais. Todos os parâmetros das redes e métricas de avaliação foram salvos em arquivos xlsx para análise.

Todas as redes utilizaram 24 características de entrada e 4 neurônios de saída, com 2 camadas ocultas com número variável de neurônios. Para cada rede, variou-se a divisão dos grupo de treinamento e validação a partir de um mesmo grupo de desenvolvimento. Já o grupo de teste foi o mesmo para todos. Utilizou-se a entropia cruzada como critério de perda, o otimizador Adam e uma taxa de aprendizado inicial aleatória. Foi salvo o modelo com melhor desempenho na validação dentre 50000 passos de treinamento.

Uma amostra de 8 indivíduos aleatórios foi retirada do grupo de teste e a acurácia foi recalculada (acurácia-leave-8-out). As redes com as melhores performances foram filtradas utilizando como critério valores de acurácia-leave-8-out acima de 45%. Foram selecionadas 10 modelos. As principais características de cada rede são descritas na Figura 9.

As redes selecionadas apresentaram na validação uma acurácia média de 37,40% (DP = 8,76), perda média de 30,82 (DP = 31,82) e correlação média de 0,31 (DP = 0,18). Observou-se que a validação sobre os dados *surrogate* teve desempenho pior que os dados originais, com menor média de acurácia (22,30%, DP = 7,09) e de correlação (0,03, DP = 0,14) e maior média de perda (52,88, DP = 49,82). Isso é um indicativo de que as redes conseguiram aprender e ter um desempenho pelo menos melhor que o aleatório.

As métricas de avaliação utilizando os dados integrais do grupo de teste obtiveram média de 46,20% (DP = 1,69) de acurácia, 26,77 (DP = 18,43) de perda e 0,36 (DP = 0,08) de correlação. A acurácia-leave-8-out média foi de 46,00% (DP = 0,88).

O método de votação combinou os resultados de cada uma dessas 10 redes obtendo como predição final a moda da classificação para cada sujeito. A acurácia-leave-8-out para esse método foi de 56,36%, com 24 erros dentre 55 sujeitos testados. O coeficiente R de

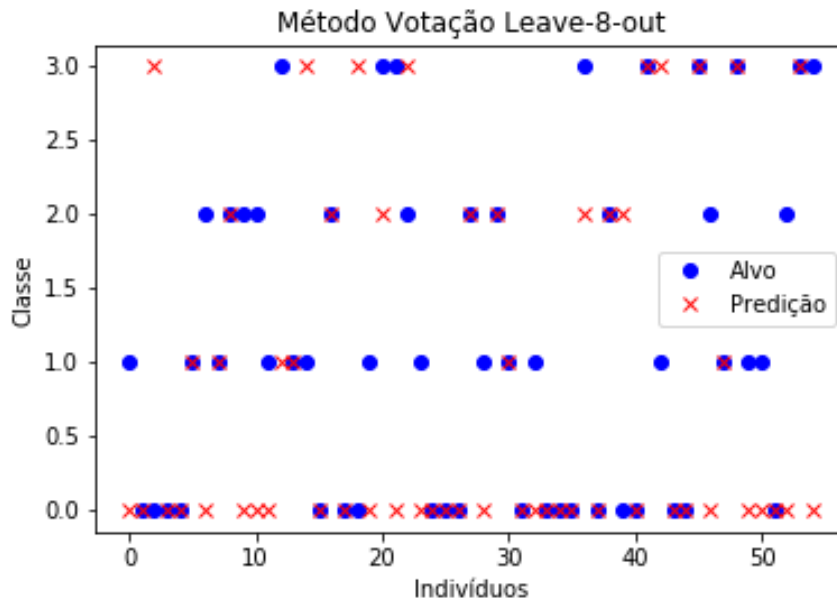


Figura 10: Resultados da Votação Leave-8-out.

		Classe Predita			
		Classe 0	Classe 1	Classe 2	Classe 3
Classe Real	Classe 0	17	0	1	2
	Classe 1	8	5	0	2
	Classe 2	5	0	5	1
	Classe 3	2	1	2	4

Figura 11: Matriz de Confusão da Votação Leave-8-out.

correlação foi de 0,4409. As Figuras 10 e 11 apresentam o gráfico de dispersão e a matriz de confusão, repectivamente, com as classes preditas e a classes alvo para cada sujeito testado.

Ao testar o método de classificação por votação nos 8 sujeitos que haviam sido retirados do grupo de teste, obteve-se uma acurácia-8-out de 62,5%. O modelo errou a predição de 3 sujeitos dentro o total de 8. A correlação obtida foi de 0,6393. A Figura 12 apresenta os resultados do teste na amostra de sujeitos 8-out e a Figura 13 mostra a sua matriz de confusão.

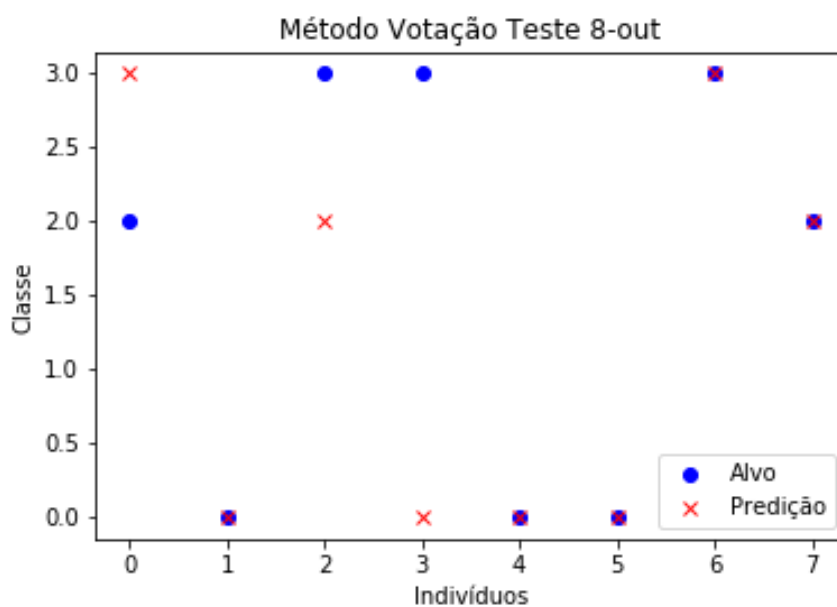


Figura 12: Resultados da Votação Teste em 8-out.

		Classe Predita			
		Classe 0	Classe 1	Classe 2	Classe 3
Classe Real	Classe 0	3	0	0	0
	Classe 1	0	0	0	0
	Classe 2	0	0	1	1
	Classe 3	1	0	1	1

Figura 13: Matriz de Confusão da Votação Teste em 8-out.

5 Discussão e Conclusão

A arquitetura de Rede Neural Artificial utilizada é bastante simples e, por isso, limitada. Pode-se perceber a partir do processo de seleção das redes que uma pequena parcela obteve um desempenho minimamente satisfatório, sendo que apenas 10 redes foram selecionadas, o que representa aproximadamente 0,02% do total de 47310 modelos treinados.

O método de votação proposto neste trabalho mostrou que a combinação do resultado de várias redes simples conseguiu melhorar a predição final do modelo (acurácia de 56,36%) para a classificação do PF50 de sujeitos em superfície rígida com os olhos fechados em comparação às acurácias individuais das redes (média de 46,00%, $DP = 0,88$). Este método assume que o valor que ocorre com maior frequência no conjunto de predições para cada sujeito tem maior probabilidade de ser o valor real. Assim, ao identificar as modas dos conjuntos de predições, o método diminui o erro individual de cada rede e gera uma predição com maior acurácia.

O teste feito com a amostra de 8 indivíduos excluídos do processo de seleção mostrou uma acurácia de 62,5%. Apesar de considerado um desempenho mediano, podemos verificar que houve um aprendizado da rede, com performance acima do acaso. Vale ressaltar que o uso de inteligência artificial para a estimação de medidas do centro de pressão é uma inovação e ainda não havia sido explorado no âmbito de Redes Neurais Artificiais e *Machine Learning*. Uma vez que, a conhecimento da autora, não existem trabalhos semelhantes publicados e não há possibilidade de comparações. Consideramos, então, que para uma primeira implementação, o resultado foi satisfatório e o poder de predição pode ser refinado para uma classificação mais acurada.

Uma das limitações deste trabalho é a pouca quantidade de dados disponíveis para o desenvolvimento do modelo. Em abordagens de *machine learning* sabe-se que a quantidade de dados disponíveis para treino, assim como o desbalanceamento destes em relação às classes, tem grande impacto na aprendizagem do modelo. Algoritmos de *machine learning* encontram problemas em aprender quando uma classe domina sobre outra e quando a quantidade de dados é pequena.

A base de dados utilizada continha em torno de 200 dados. Idealmente, em aplicações de algoritmos preditivos é desejável uma quantidade de dados em torno de alguns milhares. A disponibilidade de mais dados permitiria o desenvolvimento de uma rede de regressão que teria como resultado o valor de PF50. Neste trabalho, a estratificação dos valores contínuos do PF50 em classes teve como objetivo superar esta limitação de escassez de dados e permitir, ainda assim, uma boa acurácia de predição. Além disso, a escolha dos

limites de classificação a partir dos quartis também teve como objetivo garantir uma divisão de classes balanceadas de forma a não prejudicar o processo de aprendizagem das redes.

Como perspectivas futuras, deseja-se estudar a influência de cada característica para a predição; adaptar parâmetros da rede e da escolha de características para melhorar o desempenho do modelo de classificação com o método de votação; e implementar uma estrutura de Rede Neural Artificial para regressão, a depender da maior disponibilidade de dados.

O código foi disponibilizado em plataforma digital online no GitHub (<https://github.com/priscilaacamine/TG-Biomed>) de forma que qualquer pessoa que se interessar pelo método proposto possa acessar o código e acompanhar atualizações, assim como colaborar com propostas de melhoria.

O centro de pressão é um dos parâmetros mais importantes para a pesquisa sobre o equilíbrio e a postura e poderia ser útil no diagnóstico clínico [16]. Inúmeras alterações neurais ou fisiológicas modificam significativamente a estabilidade postural, que é refletida no deslocamento do COP. Como, por exemplo: a doença de Parkinson [10], osteoartrite de joelho [6], distúrbios mandibulares e na cervical [9] e presença de tumores [16].

Este trabalho teve como objetivo implementar um método baseado em Inteligência Artificial para estimar medidas do centro de pressão a partir de dados individuais dos sujeitos. Espera-se que o método de predição da classe da frequência média do COP por votação possa contribuir com o entendimento dessa característica e possivelmente ser aplicado, no futuro, como auxílio diagnóstico de distúrbios associadas a ela ou em sugestões de intervenções e tratamentos.

Referências

- [1] *Babadi Behtash, Brown Emery N.* A review of multitaper spectral analysis // IEEE Transactions on Biomedical Engineering. 2014. 61, 5. 1555–1564.
- [2] *Dos Santos Damiana A, Fukuchi Claudiane A, Fukuchi Reginaldo K, Duarte Marcos.* A data set with kinematic and ground reaction forces of human balance // PeerJ. 2017. 5. e3626.
- [3] *Duarte Marcos.* Análise estabilográfica da postura ereta humana quasi-estática. 2000.
- [4] *Haykin Simon.* Redes neurais: Principios e prática. 2005. 2nd.
- [5] *Ivan Vasilev Gianmario Spacagna Peter Roelants Valentino Zocca Daniel Slater.* Python Deep Learning: Exploring deep learning techniques, neural network architectures and GANs with PyTorch, Keras and TensorFlow. 2019. 2.
- [6] *Nagymate G, Pethes A, Szabo G, Bejek Z, Paroczaí R, Kiss RM.* The impact of the severity of knee osteoarthritis on the postural stability // Materials Today: Proceedings. 2016. 3, 4. 1069–1073.
- [7] *Nagymate Gergely, Vamos Bence, Kiss Rita M.* Validation of the Nintendo Wii Balance Board for stabilometry measurements // 2016 International Symposium on Small-scale Intelligent Manufacturing Systems (SIMS). 2016. 111–114.
- [8] *Oppenheim Alan V, Schafer Ronald W.* Discrete-time signal processing. 2014.
- [9] *Palano D, Molinari G, Cappelletto M, Guidetti G, Vernole B.* Computerized stabilometry in the diagnosis of craniomandibular disorders // Bulletin du Groupement International pour la Recherche Scientifique en Stomatologie et Odontologie. 1994. 37, 1-2. 19–22.
- [10] *Park Jang-Ho, Youm Sekyoung, Jeon Yongwoong, Park Seung-Hun.* Development of a balance analysis system for early diagnosis of Parkinson’s disease // International Journal of Industrial Ergonomics. 2015. 48. 139–148.
- [11] *Prieto Thomas E, Myklebust Joel B, Hoffmann Raymond G, Lovett Eric G, Myklebust Barbara M.* Measures of postural steadiness: differences between healthy young and elderly adults // IEEE Transactions on biomedical engineering. 1996. 43, 9. 956–966.
- [12] *Pró-Reitoria de Pesquisa - UFABC .* Computação Científica (CC). 2020. <https://propes.ufabc.edu.br/divisoess/cc>.

-
- [13] *Santos Damiana A, Duarte Marcos*. A public data set of human balance evaluations // PeerJ. 2016. 4. e2648.
 - [14] *Scipy.org* . Scipy Documentation: signal.welch. 2014. <https://docs.scipy.org/doc/scipy-0.14.0/reference/generated/scipy.signal.welch.html>.
 - [15] *Semmlow J.L., Griffel B*. Biosignal and Medical Image Processing. 2014.
 - [16] *Terekhov Yuriy*. Stabilometry as a diagnostic tool in clinical medicine. // Canadian Medical Association Journal. 1976. 115, 7. 631.
 - [17] *Torch Contributors* . PyTorch: CROSSENTROPYLOSS. 2019. <https://pytorch.org/docs/stable/generated/torch.nn.CrossEntropyLoss.html>.
 - [18] *Yamamoto Tomohisa, Smith Charles E, Suzuki Yasuyuki, Kiyono Ken, Tanahashi Takao, Sakoda Saburo, Morasso Pietro, Nomura Taishin*. Universal and individual characteristics of postural sway during quiet standing in healthy young adults // Physiological reports. 2015. 3, 3. e12329.
 - [19] *scikit-learn developers (BSD License)* . sklearn.preprocessing.RobustScaler. 2007-2020. <https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.RobustScaler.html>.