

Classificação de Escoliose Utilizando Um Baropodômetro e Aprendizado de Máquina

1st Éric Fadul Cunha Yoshida - RA 148143

Universidade Federal de São Paulo

São José dos Campos - São Paulo

eric.fadul@unifesp.br

2nd Luiz Eduardo Casella - RA 156702

Universidade Federal de São Paulo

São José dos Campos - São Paulo

casella@unifesp.br

I. INTRODUÇÃO E MOTIVAÇÃO

Segundo a OMS (Organização Mundial da Saúde) cerca de 3% da população mundial tem escoliose. Essa patologia recorrente consiste em uma curvatura anormal da coluna para um dos lados do tronco, determinada pela rotação das vértebras. Para diagnosticar e mensurar a gravidade do quadro é necessário fazer exames de raio-x, porém para avaliação de melhora ou piora do quadro torna-se inviável realizar esse exame com muita frequência, ao buscar uma alternativa temos o estudo do baropodômetro, junto a inteligência artificial, para diagnosticar o paciente classificando-o de acordo com o ângulo de Cobb. Este trabalho explicará conceitos fundamentais para o entendimento da escoliose e utilizando dados coletados por um baropodômetro, que consistem em 149 amostras de 61 voluntários, com informações dos sensores e altura, peso e grau de escoliose de cada paciente, com isso será criado um modelo de aprendizado de máquina que classificará em C1 ou C2 de acordo com o ângulo de Cobb.

II. CONCEITOS FUNDAMENTAIS

A escoliose é uma condição médica caracterizada por uma curvatura lateral anormal da coluna vertebral. A coluna vertebral, também conhecida como espinha dorsal, é uma estrutura complexa e essencial que proporciona suporte e flexibilidade ao corpo humano, sendo ela composta por 33 vértebras individuais que se subdividem em cinco regiões:

- **Cervical** : Composta por 7 vértebras (C1 a C7), é responsável pelo suporte da cabeça e pelos movimentos do pescoço. Sendo que as duas primeiras vértebras, atlas (C1) e axis (C2), permitem a rotação e a flexão da cabeça;
- **Torácica**: Compostas por 12 vértebras (T1 a T12), cada uma conectada a um par de costelas, responsável pela estabilidade e proteção dos órgãos vitais do tórax, como o coração e os pulmões.
- **Lombar**: Composta por 5 vértebras (L1 a L5), cuja função é suporta a maior parte do peso do corpo e é responsável pela flexibilidade e movimento da parte inferior das costas;
- **Sacro**: Composta por 5 vértebras fusionadas (S1 a S5), as quais se conectam aos ossos do quadril, formando a pelve;
- **Coccígea** :Composta por 4 vértebras fusionadas, o cóccix é o vestígio de uma cauda e tem pouca função na

mobilidade, entretanto serve como ponto de ancoragem para alguns músculos e ligamentos.

Assim, percebe-se que a coluna vertebral é de suma importância para o suporte estrutural, uma vez que sustenta o peso do corpo e mantém a postura ereta, e para a proteção, por proteger a a medula espinhal, um feixe de nervos que transporta sinais entre o cérebro e o corpo, também a flexibilidade e mobilidade pois permite uma ampla gama de movimentos, incluindo flexão, extensão, rotação e inclinação lateral, e por fim, não menos importante, a ancoragem muscular que serve como ponto de fixação para músculos e ligamentos, facilitando o movimento e a estabilidade do tronco e dos membros superiores e inferiores.

Outrossim, compreender a anatomia e a fisiologia da coluna vertebral é essencial para apreciar a complexidade da escoliose e seu impacto no corpo humano. Uma vez que, a escoliose perturba essa delicada estrutura, exigindo uma abordagem multidisciplinar para seu diagnóstico e tratamento eficaz.

A condição de escoliose, pode se dividir em três tipos:

- **Escoliose Idiopática**: representa a maioria dos casos, ocorre sem uma causa conhecida;
- **Escoliose Congênita**: é causada por anormalidades na formação das vértebras antes do nascimento;
- **escoliose adquirida**: pode ser causada por lesões, infecções, distúrbios neuromusculares ou condições médicas subjacentes.

Essa condição pode gerar diversos problemas para saúde, dependendo da sua gravidade, como dores nas costas e no pescoço, problemas respiratórios e cardíacos, alteração na postura e no caminhar.

A gravidade da escoliose é definida pelo ângulo da curvatura, esse ângulo é chamado de ângulo de Cobb, e é uma medida extraída a partir do método de Cobb que, por meio do desenho de duas retas sob uma radiografia, é possível identificar o grau de escoliose de um paciente.

- **Possui-se escoliose**: o paciente que tem o ângulo de Cobb maior que 10°;
- **escoliose moderada**: entre 25° e 45°;
- **escoliose grave**: maior que 45°.

A baropodometria é um método de análise que avalia a distribuição da pressão plantar em posição estática. Os baropodômetros utilizados são dispositivos equipados com 120

sensores, 60 para cada pé, que medem a pressão exercida em diferentes pontos. Através desses dados, é possível identificar anomalias na postura. No contexto da escoliose, a análise baropodométrica pode fornecer percepções sobre como a escoliose afeta a distribuição da pressão plantar, o equilíbrio e a postura dos indivíduos.

A inteligência artificial (IA) refere-se à simulação de processos de inteligência humana por sistemas computacionais. Através de algoritmos é possível criar uma inteligência artificial que aprende a realizar determinada tarefa. No contexto do trabalho, a inteligência artificial é utilizada para analisar os dados obtidos pelos sensores baropodométricos e classificar os indivíduos em dois grupos através do reconhecimento de padrões entre os dados. Para uso de um modelos de IA tem-se alguns passos como o pré processamento dos dados, que a partir dos dados brutos, molda eles para que possam ser utilizados pelo modelo, também usualmente divide-se os dados em treino e teste em uma proporção definida, para que os dados utilizados no treino do modelo não sejam utilizados no teste, considerando que isso poderia ter um impacto falso na eficácia do modelo.

Conceitos de inteligência artificial básicos são:

- **Modelo:** Conjunto de instruções lógicas que permite que um sistema de computador aprenda a partir de dados e experiências passadas para tomar decisões ou prever resultados futuros.
- **Acurácia:** Valor que mensura a eficácia do modelo e no caso de uma classificação, consiste na relação dos dados previstos e sua classificação real.
- **Parâmetro:** Entrada da função do modelo que influencia na maneira como o modelo irá trabalhar, podendo utilizar outras funções ou limitar em um número de iterações de treino do modelo.
- **Classificação:** Objetivo do modelo, classificar dados de acordo com seu comportamento ao passar pelo modelo.
- **Single Layer Perceptron (SLP):** tipo básico de rede neural artificial, composta por uma única camada de neurônios que realiza a transformação linear dos dados de entrada. Essa estrutura simples permite que o SLP seja utilizado para resolver problemas de classificação linear, onde os dados podem ser separados por uma linha reta ou hiperplano. Apesar de sua simplicidade, o SLP serve como uma introdução fundamental aos conceitos de redes neurais e é frequentemente usado como base para modelos mais complexos.
- **Multilayer Perceptron (MLP):** expande o conceito de SLP ao adicionar uma ou mais camadas ocultas entre as camadas de entrada e saída. Essas camadas adicionais permitem que o MLP aprenda representações mais complexas dos dados, tornando-o capaz de resolver problemas de classificação não linear. O MLP é amplamente utilizado em diversas aplicações de aprendizado de máquina, devido à sua capacidade de modelar relações complexas e capturar padrões ocultos nos dados.
- **Acurácia Balanceada:** uma métrica de avaliação que leva em conta o desempenho do modelo em ambas as classes,

especialmente em cenários de dados desbalanceados. Ao contrário da acurácia tradicional, que pode ser enganosa em casos onde uma classe é muito mais prevalente que a outra, a acurácia balanceada calcula a média da sensibilidade (recall) de cada classe, proporcionando uma medida mais justa do desempenho do modelo em cenários onde há uma distribuição desigual entre as classes.

- **Convolutional Neural Network (CNN):** um tipo de rede neural projetada para processar dados em forma de grade, como imagens. Ela é composta por varias camadas, e algumas delas são as camadas convolucionais, que utilizam de filtros para extrair características e padrões da imagem.
- **Kfold:** método de divisão de dados em treino e teste, onde K é o número de partes e assim se cria um modelo para cada iteração, separando 1/k para validação e o restante para treino.
- **SVM:** Support Vector Machine é um modelo de aprendizado de máquina que utiliza vetores de suporte

III. TRABALHOS RELACIONADOS

Dentre os trabalhos relacionados ao tema, temos como destaque o artigo "Evaluation of scoliosis using baropodometer and artificial neural network" publicado em 2017, o qual será utilizado como base e o trabalho "Sistema baropodométrico e classificação de escoliose utilizando técnicas de machine learning" que introduz a relação entre as distribuições de força na região plantar com o grau de desvio lateral da coluna vertebral, apesar de não garantir causalidade, abrindo possibilidade para outros fatores envolvidos.

No trabalho "Evaluation of scoliosis using baropodometer and artificial neural network", possui-se 61 voluntários, de diferentes idades e sexo, e 149 amostras, foram feitas mais de uma medição por voluntário. Para a coleta dos dados, os voluntários ficavam por 120 segundos em cima do baropodômetro, que fazia captação dos sensores em uma frequência de 20hz, possibilitando captar a pressão da região plantar dos voluntários em tempo real. Desses 120 segundos em que o voluntário ficava em cima do baropodômetro, 60 segundos eram com os olhos abertos e 60 segundos eram com os olhos fechados. Após a coleta dos dados, o trabalho desenvolve uma rede neural perceptron com input das médias dos sensores e a rede neural classificaria o voluntário em dois grupos:

- C1: ângulo de Cobb menor que 10°
- C2: ângulo de Cobb entre 10° e 19°

Foi utilizado como referencia a tese de doutorado do professor Fábio Rodrigo Castro Forero, que deu o próximo passo a comprovação da relação introduzida no artigo anterior, ao utilizar a mesma base de dados e métricas mais confiáveis, além de adicionar cálculos, conseguindo uma acurácia de $74,89 \pm 7,19$ % em seu melhor modelo.

IV. OBJETIVOS

Nesse trabalho, vai-se utilizar o dataset do artigo "Evaluation of scoliosis using baropodometer and artificial neural

network” e a tese de doutorado ”Sistema Baropodométrico e Classificação de Escoliose Utilizando Técnicas de Machine Learning” do professor Fabian Rodrigo Castro Forero tem-se como objetivo melhorar a acurácia dos trabalhos já realizados. Em que, a partir dos dados de 120 sensores medidos por um baropodômetro, cálculos de centro de massa e medidas antropométricas dos pacientes, busca-se classificar, usando inteligência artificial, os voluntários em dois grupos conforme a angulação da escoliose: C1 (angulação menor que 10°) e C2 (angulação entre 10° e 19°).

V. METODOLOGIA EXPERIMENTAL

Inicialmente, com acesso a base de dados fornecida, a mesma utilizada pelo artigo ”Evaluation of scoliosis using baropodometer and artificial neural network” e pela tese de doutorado ”Sistema Baropodométrico e Classificação de Escoliose Utilizando Técnicas de Machine Learning” do professor Fabian Rodrigo Castro Forero, foram pré-processados os dados buscando inconsistências para minimizar a influência de possíveis erros.

Sequencialmente, os dados foram divididos em 80% para treino e 20% para teste, utilizando o método Kfold, considerando que temos mais de uma amostra do mesmo paciente, assim será necessário procurar estratégias para não correr o risco de enviesar o modelo.

Com a base dividida foram definidos os principais dados para treinamento do modelo, visto que temos acesso a vários dados agrupados separadamente, sendo esses as medidas antropométricas de cada paciente, a média dos sensores e cálculos de centro de massa, assim explorou-se as possibilidade de seleção dos dados para cada treino.

Com todos esses fatores definidos definiu-se o modelo, o qual para primeiro treino foi o mesmo utilizado em um dos trabalhos relacionados, uma rede neural artificial do tipo multi layer perceptron com topologia [30-44-1]. Outra possibilidade explorada foi o uso de redes convolucionais, que utilizam imagens, assim foi necessário transformar os dados brutos dos sensores em imagens.

Em seguida foi mensurado através da acurácia balanceada, e outras métricas avaliativas para definir o melhor modelo, caso o resultado seja inconsistente.

VI. PROTOCOLO EXPERIMENTAL

Nesse trabalho foram utilizadas duas bases de dados, a primeira sendo uma planilha e por isso será mencionada como planilha, e a segundo é referente aos dados brutos dos sensores, assim será mencionada como dados brutos, ambas as bases descritas a seguir.

A. Domínio da aplicação:

A planilha consiste em informações referentes a amostras coletadas, em cada amostra temos a média dos valores em N (Newtons) de cada sensor de um baropodômetro, o qual tem 120 sensores posicionados 6x10 para cada pé, formando uma matriz 12x10. Na base também temos dados referentes às medidas antropométricas e ao juntar esses valores, calculou-se:

- **CoP_ML:** Center of Pressure - Mediolateral Refere-se ao deslocamento do centro de pressão (CoP) na direção médio lateral (lado a lado).
- **CoP_AP:** Center of Pressure - Anteroposterior Refere-se ao deslocamento do centro de pressão na direção anteroposterior (frente para trás).
- **T_Displacement:** Total Displacement O deslocamento total do centro de pressão (CoP) durante o período de teste.
- **Disp_ML:** Displacement - Mediolateral O deslocamento do CoP na direção médio lateral.
- **Disp_AP:** Displacement - Anteroposterior O deslocamento do CoP na direção anteroposterior.
- **T_avg_speed:** Total Average Speed Velocidade média total do CoP durante o teste.
- **Speed_ML:** Speed - Mediolateral Velocidade média do CoP na direção médio lateral.
- **Speed_AP:** Speed - Anteroposterior Velocidade média do CoP na direção anteroposterior.
- **Amp_ML:** Amplitude - Mediolateral Amplitude do deslocamento do CoP na direção médio lateral.
- **Amp_AP:** Amplitude - Anteroposterior Amplitude do deslocamento do CoP na direção anteroposterior.
- **RMS_ML:** Root Mean Square - Mediolateral Valor RMS do CoP na direção médio lateral. Este é um valor estatístico que mede a magnitude da oscilação na direção médio lateral.
- **RMS_AP:** Root Mean Square - Anteroposterior Valor RMS do CoP na direção anteroposterior. Este é um valor estatístico que mede a magnitude da oscilação na direção anteroposterior.

Essas métricas são usadas para avaliar o equilíbrio e a estabilidade postural, especialmente em testes que comparam as condições de olhos abertos e fechados. A análise dessas variáveis pode fornecer informações sobre a capacidade de uma pessoa de manter o equilíbrio em diferentes condições sensoriais.

Foram avaliados 61 voluntários no total. A grande maioria da população avaliada foi do sexo feminino com 47 participantes, e o restante do sexo masculino, com 14 participantes.

Agrupando esses valores, formou-se a planilha onde com um total de 149 x 151, em que 149 é o número de amostras, 150 o número de atributos para cada amostra, e os valores da última coluna são o grau de escoliose do indivíduo. Mais de uma coleta foi realizada com o mesmo indivíduo.

Os dados brutos consistem nos valores brutos dos sensores em cada amostragem, em que planejou-se monitorar o equilíbrio do paciente durante 60 segundos com os olhos abertos e 60 segundos com olhos fechados, e os sensores captam 20 valores por segundo, assim teríamos um total de 2400 momentos para cada amostragem, porém a coleta foi feita de maneira não homogênea, assim parte das amostragens tem mais valores ou menos valores.

A visão geral dos dados biométricos mostra que:

Linha/Coluna	Homem	Mulher
Quantidade	14	47
Média de Idade	26.78	41.7
Média de Peso	67.07kg	64.57kg
Média de Altura	1.68m	1.59m

TABLE I
TABELA DE MEDIDAS ANTROPOMÉTRICAS

B. Pré-processamento:

A etapa de pré-processamento é crucial para qualquer projeto de inteligência artificial por diversas razões fundamentais, tais como garantir a integridade dos dados, buscando ruídos, erros, valores ausentes ou inconsistências que podem prejudicar o desempenho de modelos, essa etapa permite estudar mais a fundo a base de dados e visualizar e corrigir esses problemas. Os processos mais comuns são a limpeza de dados, onde se procuram inconsistências e as corrigem, a transformação de dados, que podem envolver normalização de dados, seleção de variáveis, ou até redução de dimensionalidade como PCA, e a etapa de criação de dados, onde se tem como possibilidade estratégias como o aumento de dados a partir de variáveis existentes.

1) *Limpeza dos dados*:: Na primeira base de dados (planilha) as alterações iniciais consistiram da exclusão de duas colunas com informações redundantes, a primeira foi de lado dominante, onde tínhamos duas colunas nomeadas de “R_Handed” e “L_Handed”, e os atributos eram contrários, ou seja, quando um indivíduo era “R_Handed” implicava que ele não era “L_Handed”, assim retiramos a coluna “L_Handed” e renomeamos a coluna “R_Handed” para “Handedness”, o mesmo aconteceu com as colunas de “Female” e “Male”, onde excluímos a male e a renomeamos para “Sex”.

Para a segunda base de dados (valores brutos) que consiste em diversos arquivos de texto, cada um contendo uma matriz 12x10 referente ao valor dos sensores naquele momento, o primeiro passo foi a verificação dos arquivos. Os arquivos estavam organizados sobre a seguinte estrutura, primeiro uma pasta geral, a qual continha 149 pastas referentes às 149 amostras, e cada pasta estava nomeada com o nome do voluntário seguido do número daquela amostra, “nome do voluntário 1”, e dentro de cada pasta haviam arquivos de texto nomeados como “Olhos Abertos_nome do voluntário_número”.

Para manipulação dos dados utilizamos as bibliotecas pandas e numpy.

2) *Transformação de dados*:: Para os dados brutos a primeira verificação realizada foi buscar erros de escrita no nome das pastas, alguns nomes estavam errados, e foram corrigidos manualmente. Com isso, através de scripts utilizando a biblioteca OS, foi realizada a anonimização dos dados. O primeiro script renomear as pastas referentes a cada amostra, considerando que o mesmo indivíduo pudesse ter mais que uma amostra, e o novo formato foi “individual_número do indivíduo_número da amostra”, em seguida, o segundo script renomeou os arquivos utilizando o nome do diretório para o seguinte formato “Estado_individual_numero

do indivíduo_número da amostra_número do arquivo”, sendo o estado ou CE (Closed Eyes) para olhos fechados ou OE (Opened Eyes) para olhos abertos.

Para ambas as bases essa etapa buscou estratégias de normalização, estudou-se as seguintes possibilidades:

- MinMax: Transforma os dados para que todos os valores estejam em um intervalo específico, geralmente [0, 1].
- Standard: Normaliza os dados para que tenham média 0 e desvio padrão 1.
- MaxAbs: Escala os dados de modo que os valores absolutos máximos em cada feature sejam 1.
- L2: Escala os dados de modo que a soma dos quadrados dos elementos em cada vetor seja igual a 1

Utilizando matplotlib, visualizamos os resultados da normalização de valores de um momento coletado pelos sensores para identificar qual normalização utilizar. Considerando

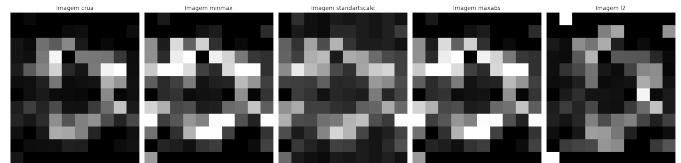


Fig. 1. Teste de normalização 1

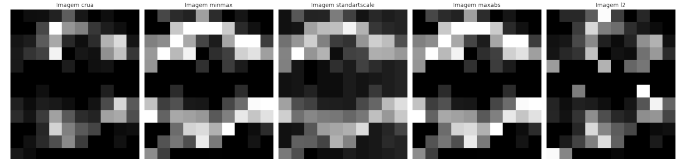


Fig. 2. Teste de normalização 2

o mínimo de informações perdidas, optamos por utilizar a normalização standard. Para aplicação de normalização nos utilizou-se a biblioteca sklearn.

Em seguida foi preciso transformar os dados brutos de arquivos de texto em imagens, para isso foi utilizado a biblioteca opencv junto ao pandas e os para salvar os arquivos. Esse processo foi realizado através de um script que lia todos os dados, normalizava cada arquivo e o salvava no formato png. Cada arquivo de texto deu origem a uma respectiva imagem e planejou-se utilizar modelos de CNN, assim a próxima etapa foi agrupar essas imagens em grupos de 20, assim cada imagem agrupada teria informações referentes a um segundo. Para realizar esse agrupamento utilizou-se um script que lia todos os arquivos png de uma pasta e juntava 5 imagens por linha e 4 linhas para formar uma imagem agrupada, formando imagens 48x50. Para realização do treino foi necessário em seguida redimensionar as imagens para 224x224.

C. Replicação da tese do professor Fabian

Como primeiro passo do trabalho, realizamos a replicação de um trabalho já existente na bibliografia, que utiliza o mesmo banco de dados, a tese do professor Fabian Rodrigo Castro Forero. Nosso objetivo era replicar os resultados

obtidos na tese. Para alcançar esse objetivo, adotamos uma abordagem rigorosa, buscando reproduzir todos os parâmetros e métodos utilizados no trabalho original. Ao replicar esses procedimentos, buscamos não apenas validar os resultados, mas também fornecer um resultado base para nosso trabalho final da disciplina de Inteligência Artificial

Iniciamos o experimento com a etapa de pré-processamento dos dados, a fim de garantir a qualidade e consistência das informações utilizadas. Como os dados utilizados neste experimento foram derivados de um trabalho acadêmico já existente, partimos de um conjunto de informações que já havia passado por um processo inicial de limpeza. Isso significou que a necessidade de realizar uma limpeza profunda dos dados foi reduzida, já que grande parte das inconsistências, valores ausentes e outliers já havia sido tratada pelos autores do estudo original. No entanto, realizamos algumas transformações nos dados que julgamos necessários, como: exclusão de colunas que não eram relevantes para as análises, a junção de colunas, e o renomeamento de colunas para padronizar e facilitar a interpretação dos dados. Assim como no trabalho original, selecionamos apenas as primeiras 70 variáveis. Após todo processo de transformação da base de dados, realizamos a normalização utilizando a função `StandardScaler()`, da biblioteca `sklearn.preprocessing`.

Neste experimento, nosso objetivo principal é desenvolver e treinar um modelo de inteligência artificial capaz de classificar as amostras em dois grupos distintos: C1, que corresponde às pessoas com ângulo de escoliose inferior a 10 graus, e C2, para aquelas com ângulo superior a 10 graus. Assim como foi feito no trabalho original que serviu de base para o nosso estudo, optamos por utilizar uma Rede Neural Artificial (RNA) do tipo Perceptron Multicamadas (MLP), utilizamos a biblioteca `sklearn`, com uma topologia [30-44-1]. A rede original foi treinada utilizando o algoritmo `trainsec` como solucionador, porém no `sklearn` não existe essa função, assim utilizou-se uma semelhante `sgd`.

Para garantir a robustez do modelo e a generalização dos resultados, aplicamos o método de validação cruzada *K-fold*, com K igual a 5, para a separação dos dados em conjuntos de treino e teste. Essa abordagem permitiu que o modelo fosse treinado e validado em diferentes subconjuntos dos dados, reduzindo o risco de *overfitting* e assegurando que o desempenho do modelo fosse avaliado de maneira equilibrada.

Os resultados obtidos com o modelo MLP [30-44-1] foram satisfatórios dentro do contexto do experimento. Utilizando a métrica de acurácia balanceada, que leva em consideração a distribuição desigual das classes, o modelo alcançou uma acurácia de 65,85%. Esse resultado demonstra a capacidade do modelo em diferenciar entre os dois grupos de amostras, C1 e C2, mesmo diante de possíveis desequilíbrios nos dados. Embora haja espaço para melhorias, o desempenho atingido indica que a abordagem utilizada é promissora para a classificação baseada nos ângulos de escoliose.

Desenvolvemos também uma segunda versão da MLP, utilizando a função `RandomUnderSampler()` com o intuito de resolver a disparidade entre a quantidade de dados nas classes.

```

Training fold: 100% [██████████] 3000/3000 [00:28:00:00, 104.44it/s]
Fold Train Accuracy: 0.9748, Fold Test Accuracy: 0.6667
Fold Balanced Train Accuracy: 0.9888, Fold Balanced Test Accuracy: 0.4762
Confusion Matrix for Fold 1:
[[ 0  9]
 [ 1 20]]

Training fold: 100% [██████████] 3000/3000 [00:34:00:00, 86.00it/s]
Fold Train Accuracy: 0.9832, Fold Test Accuracy: 0.7333
Fold Balanced Train Accuracy: 0.9878, Fold Balanced Test Accuracy: 0.7014
Confusion Matrix for Fold 2:
[[ 0  7]
 [ 1 16]]

Training fold: 100% [██████████] 3000/3000 [00:30:00:00, 98.48it/s]
Fold Train Accuracy: 0.9664, Fold Test Accuracy: 0.7667
Fold Balanced Train Accuracy: 0.9759, Fold Balanced Test Accuracy: 0.7545
Confusion Matrix for Fold 3:
[[ 8  6]
 [ 1 15]]

Training fold: 100% [██████████] 3000/3000 [00:32:00:00, 92.83it/s]
Fold Train Accuracy: 0.6387, Fold Test Accuracy: 0.7667
Fold Balanced Train Accuracy: 0.5000, Fold Balanced Test Accuracy: 0.5000
Confusion Matrix for Fold 4:
[[ 0  7]
 [ 0 13]]

Training fold: 100% [██████████] 3000/3000 [00:27:00:00, 107.69it/s]
Fold Train Accuracy: 0.9583, Fold Test Accuracy: 0.8021
Fold Balanced Train Accuracy: 0.9675, Fold Balanced Test Accuracy: 0.8004
Confusion Matrix for Fold 5:
[[ 6  1]
 [ 3 19]]

Mean Train Accuracy: 0.9043 ± 0.1331
Mean Test Accuracy: 0.7591 ± 0.0631
Mean Balanced Train Accuracy: 0.8824 ± 0.1913
Mean Balanced Test Accuracy: 0.6585 ± 0.1484
Aggregated Confusion Matrix:
[[ 4.  0. ]
 [ 1.2 18.6]]

Training fold: 100% [██████████] 3000/3000 [00:22:00:00, 133.68it/s]
Fold Train Accuracy: 0.8659, Fold Test Accuracy: 0.5333
Fold Balanced Train Accuracy: 0.8659, Fold Balanced Test Accuracy: 0.4762
Confusion Matrix for Fold 1:
[[ 3  0]
 [ 0 13]]

Training fold: 100% [██████████] 3000/3000 [00:19:00:00, 151.17it/s]
Fold Train Accuracy: 0.9324, Fold Test Accuracy: 0.6667
Fold Balanced Train Accuracy: 0.9324, Fold Balanced Test Accuracy: 0.6516
Confusion Matrix for Fold 2:
[[ 7  6]
 [ 4 13]]

Training fold: 100% [██████████] 3000/3000 [00:21:00:00, 137.45it/s]
Fold Train Accuracy: 0.9061, Fold Test Accuracy: 0.7333
Fold Balanced Train Accuracy: 0.9061, Fold Balanced Test Accuracy: 0.7366
Confusion Matrix for Fold 3:
[[11  3]
 [ 5 11]]

Training fold: 100% [██████████] 3000/3000 [00:19:00:00, 157.08it/s]
Fold Train Accuracy: 0.7093, Fold Test Accuracy: 0.4000
Fold Balanced Train Accuracy: 0.7093, Fold Balanced Test Accuracy: 0.4099
Confusion Matrix for Fold 4:
[[ 3  4]
 [14  9]]

Training fold: 100% [██████████] 3000/3000 [00:18:00:00, 159.93it/s]
Fold Train Accuracy: 0.5000, Fold Test Accuracy: 0.7586
Fold Balanced Train Accuracy: 0.5000, Fold Balanced Test Accuracy: 0.5000
Confusion Matrix for Fold 5:
[[ 0  7]
 [ 0 22]]

Mean Train Accuracy: 0.7987 ± 0.1759
Mean Test Accuracy: 0.6184 ± 0.1343
Mean Balanced Train Accuracy: 0.7987 ± 0.1759
Mean Balanced Test Accuracy: 0.5549 ± 0.1205
Aggregated Confusion Matrix:
[[ 4.8  5.2]
 [ 6.2 15.6]]

```

Fig. 3. Resultado - Folds MLP

Fig. 4. Resultado - Folds MLP com `RandomUnderSampler`

No entanto, os resultados obtidos com essa abordagem foram inferiores, alcançando uma acurácia balanceada de 55,49%. Apesar da tentativa de equilibrar a distribuição das classes, a performance do modelo piorou, indicando que essa estratégia de subamostragem pode não ter sido a mais eficaz para melhorar a classificação no contexto deste experimento. Acreditamos que, pela baixa quantidade de amostras da menor classe, o treinamento foi comprometido ao tentar balanceá-lo utilizando C1 como parâmetro.

Além disso, desenvolvi uma outra versão da MLP, mantendo os mesmos parâmetros utilizados nas versões anteriores, mas utilizando um conjunto de dados diferente. Nesta versão, utilizei exclusivamente os cálculos derivados dos dados obtidos pelos sensores. Essa abordagem foi adotada para avaliar se o uso de dados mais específicos e diretamente relacionados ao problema poderia melhorar o desempenho do modelo. Esse modelo obteve 52.14% ± 7.47% de acurácia balanceada.

```

using device: cpu
Training fold: 100% [██████████] 5000/5000 [00:37:00:00, 131.82it/s]
Fold Balanced Train Accuracy: 0.9653, Fold Balanced Test Accuracy: 0.4630
Confusion Matrix for Fold 1:
[[ 1  2]
 [11 16]]

using device: cpu
Training fold: 100% [██████████] 5000/5000 [00:20:00:00, 173.51it/s]
Fold Balanced Train Accuracy: 0.5000, Fold Balanced Test Accuracy: 0.5000
Confusion Matrix for Fold 2:
[[ 0 11]
 [ 0 19]]

using device: cpu
Training fold: 100% [██████████] 5000/5000 [00:28:00:00, 174.85it/s]
Fold Balanced Train Accuracy: 0.5000, Fold Balanced Test Accuracy: 0.5000
Confusion Matrix for Fold 3:
[[ 0 10]
 [ 0 20]]

using device: cpu
Training fold: 100% [██████████] 5000/5000 [00:29:00:00, 171.81it/s]
Fold Balanced Train Accuracy: 0.5000, Fold Balanced Test Accuracy: 0.5000
Confusion Matrix for Fold 4:
[[ 0 15]
 [ 0 15]]

using device: cpu
Training fold: 100% [██████████] 5000/5000 [00:29:00:00, 169.69it/s] Fold Balanced Train Accuracy: 0.9321, Fold Balanced Test Accuracy: 0.6439
Confusion Matrix for Fold 5:
[[ 5  6]
 [ 3 15]]

Mean Balanced Train Accuracy: 0.6795 ± 0.2201
Mean Balanced Test Accuracy: 0.5214 ± 0.0629
Aggregated Confusion Matrix:
[[ 1.2  9.6]
 [ 2.8 17. ]]

```

Fig. 5. Resultado - Folds MLP apenas com calculos.

D. Replicação artigo Scielo

Realizamos também a replicação do artigo: “Evaluation of scoliosis using baropodometer and artificial neural network”. Neste trabalho, realizamos a replicação de um estudo que utiliza a mesma base de dados composta pela média da captação de 120 sensores obtidos de dois baropodômetros para analisar ângulos de escoliose em voluntários. No artigo original, os autores desenvolveram uma inteligência artificial baseada em um Perceptron de Camada Simples (SLP, do inglês *Single-Layer Perceptron*) para classificar os dados em duas classes distintas, com base nos ângulos de escoliose dos voluntários. Nosso objetivo foi reproduzir a metodologia descrita, utilizando o mesmo conjunto de dados e implementando o modelo de SLP. Normalizamos os dados utilizando a função `StandardScaler()` e os separamos os dados utilizando o método K-Fold, com $k=5$. Nesse modelo, obtivemos como resultado de 55.40% de acurácia balanceada.

E. SVM

Outro modelo criado, foi usando o classificador SVM, usamos todos os dados, cálculos e sensores, para treiná-lo. Normalizamos os dados utilizando a função `StandardScaler()`, da biblioteca `sklearn.preprocessing`, utilizamos `kfold`, com $k=5$. O modelo obteve um acurácia balanceada de $60.34\% \pm 11.25\%$.

F. Treino CNN

O treino da CNN buscou classificar o indivíduo em C1 ou C2 partindo das imagens de sua amostra e foi realizado com estratégias diferentes, na primeira tentativa foi treinado o modelo de resnet50 com todas as imagens, onde para acessar a label respectiva de cada indivíduo foi utilizada uma planilha com todos os indivíduos e sua classe real, e para saber de qual indivíduo era a imagem utilizou-se o nome do arquivo. Porém ao alimentar todos os arquivos aleatoriamente para o modelo, percebeu-se que a diferença entre cada imagem de um mesmo indivíduo era muito pequena, assim se distribuídas aleatoriamente, parte da amostra era utilizada para treinar e outra parte para validar, e assim a acurácia deu um valor

muito alto, após a realização desse treino percebeu-se e corrigiu-se a falha através de uma organização de vetores de indexação para cada indivíduo, assim ao dividir os folds com o k fold seriam divididos os vetores que continham o index das imagens daquela amostra, com isso, consegui treinar o modelo porém os resultados foram muito ruins, piores do que uma escolha aleatória. No experimento foram utilizados os modelos de resnet50 e efficientnet, junto a um dataloader para leitura dos dados, todos da biblioteca Pytorch. O tamanho da “batch” inicial definido foi de 620, para uso do máximo das capacidades da máquina, porem alterou-se para 128 para melhorar o treino.

```

Resultados dos folds:
Média e desvio padrão da Acurácia: 58.01% ± 8.83
Média e desvio padrão da Acurácia Balanceada: 0.4787 ± 0.0605
Média e desvio padrão do F1-Score: 0.7086 ± 0.0826
Média e desvio padrão do ROC-AUC: 0.4855 ± 0.1274

```

Fig. 6. Resultado - Resnet50 sem undersampling

Como protocolo de validação foram utilizadas, além da acurácia balanceada, os valores de F1-Score e ROC-AUC. F1-Score é a média harmônica da precisão e da sensibilidade, sendo precisão a proporção de predições positivas corretas em relação a todas as predições positivas, e sensibilidade a proporção de predições positivas corretas em relação a todos os verdadeiros positivos. A média do F1 Score ao longo dos folds fornece uma visão geral de como o modelo equilibra precisão e recall em diferentes divisões dos dados. Um F1 Score consistentemente alto indica um bom desempenho geral em todas as partições. ROC-AUC é uma curva que representa a relação entre a taxa de verdadeiros positivos e a taxa de falsos positivos para diferentes limiares de classificação. A média de ROC AUC ao longo dos folds reflete a habilidade média do modelo de discriminar entre as classes em diferentes divisões dos dados. Uma alta média de ROC AUC indica que o modelo é consistentemente bom em separar as classes. Assim, analisando os resultados desse primeiro treino, optou-se por realizar um undersampling e equilibrar os dados para cada classe para ver se isso influenciaria no resultado e os resultados foram semelhantes. Com relação a acurácia balanceada, o

```

Resultados dos folds:
Média e desvio padrão da Acurácia: 57.07% ± 7.01
Média e desvio padrão da Acurácia Balanceada: 0.4884 ± 0.0695
Média e desvio padrão do F1-Score: 0.6884 ± 0.0752
Média e desvio padrão do ROC-AUC: 0.4941 ± 0.1359

```

Fig. 7. Resultado - Resnet50 com undersampling

modelo se saiu um pouco maior porem ainda pior que uma classificação de uma classe só. Com esses resultados buscou-se outra treinar outro modelo, a EfficientNet, e os resultados foram semelhantes.

Com esses resultados, resolvemos fazer mais um teste, utilizei o modelo que deu melhor acurácia balanceada para tentar um treino com mais épocas e obtive um resultado similar. Com esses resultados insatisfatórios buscamos outras estratégias, e assim mudamos o parâmetro de learning rate

Resultados dos folds:
 Média e desvio padrão da Acurácia: 57.07% \pm 7.01
 Média e desvio padrão da Acurácia Balanceada: 0.4884 \pm 0.0695
 Média e desvio padrão do F1-Score: 0.6884 \pm 0.0752
 Média e desvio padrão do ROC-AUC: 0.4941 \pm 0.1359

Fig. 8. Resultado - EfficientNet sem undersampling

Resultados dos folds:
 Média e desvio padrão da Acurácia: 48.60% \pm 5.61
 Média e desvio padrão da Acurácia Balanceada: 0.4313 \pm 0.0416
 Média e desvio padrão do F1-Score: 0.6262 \pm 0.0602
 Média e desvio padrão do ROC-AUC: 0.4324 \pm 0.0380

Fig. 9. Resultado - EfficientNet com undersampling

para 0.00001 e passamos a utilizar um tensorboard para visualização do treino, também foi efetuada uma seleção de imagens, e definido um padrão de quantidade de imagens para 500 para CE e 500 para OE, totalizando 1000 imagens por amostra, salvo a exceção de a amostragem não estar completa. Com esses dados separadas, prosseguiu-se para o treino ainda utilizando a validação cruzada, a quantidade de épocas foi aumentada para 50 por fold. Nesse treino, os resultados foram melhores que uma escolha aleatória e algumas distribuições de dados resultaram em uma acurácia de mais de 60% e assim resolveu-se explorar o melhor fold e fazer um treinamento mais extensivo, com 250 épocas.

VII. CONCLUSÃO

Para desenvolvimento de alternativas de acompanhamento e classificação da escoliose, estudou-se a relação comportamental do equilíbrio dos pacientes em um baropodômetro, e com esses dados treinou-se uma inteligência artificial para validar essa relação. Este estudo avaliou e replicou os experimentos realizados já existentes na literatura, e apesar de não conseguir acurácia melhor em treinos dos modelos, comprovou-se que a relação existe, uma vez que a acurácia alcançada pelo modelo de redes convolucionais foi de 62,27% para classificação de c1 e c2. Nas tentativas de replicação dos experimentos da tese do professor Fabian, os resultados se aproximaram do alcançado, utilizando SVM alcançou-se a acurácia balanceada de 73.81%, porém não foram superados os 74,89%, e foi descoberto um fator que poderia causar enviesamento dos modelos, nos treinos realizados na tese não se considerou que existem mais amostragens de um mesmo indivíduo, e essas amostragens tem semelhança e valores de medidas antropométricas idênticos, o que arriscam o enviesamento do modelo. Realizou-se um experimento com as mesmos hiperparâmetros porém considerando a diferença de indivíduos e não se conseguiu resultados consistentes. Assim, conclui-se que pode existir relação entre o desequilíbrio medido pelos baropodômetros e a gravidade da escoliose, porém não foi considerado o fator de mais de uma amostra para o mesmo indivíduo, e somados às medidas antropométricas e cálculos baseado nessas medidas e na média dos sensores, é possível alcançar um valor de acurácia que indica que essa relação existe. No caso do treino utilizando imagens, concluiu-se que pode existir relação, porém devido a variação de padrões não

Resultados dos folds:
 Média e desvio padrão da Acurácia: 55.85% \pm 6.55
 Média e desvio padrão da Acurácia Balanceada: 0.4673 \pm 0.0708
 Média e desvio padrão do F1-Score: 0.6794 \pm 0.0753
 Média e desvio padrão do ROC-AUC: 0.4670 \pm 0.1277

Fig. 10. Resultado - ResNet50 com undersampling - 30 épocas

Resultados dos folds:
 Média e desvio padrão da Acurácia: 56.54% \pm 8.11
 Média e desvio padrão da Acurácia Balanceada: 0.5223 \pm 0.0638
 Média e desvio padrão do F1-Score: 0.6778 \pm 0.0752
 Média e desvio padrão do ROC-AUC: 0.5695 \pm 0.0988

Matriz de Confusão Média:
 [[2407.4 6332.8]
 [5774.6 13562.2]]

Fig. 11. ResNet50 com undersampling - 1000 imagens por amostra

ser notável nos dados utilizados, é necessário mais amostragens e em um equipamento mais preciso para realizar esse estudo e comprovar a viabilidade do experimento.

VIII. REFERÊNCIAS BIBLIOGRÁFICAS

- FANFONI, Caroline Meireles; FORERO, Fabian Castro; SANCHES, Marcelo Augusto Assunção; MACHADO, Érica Regina Marani Daruichi; URBAN, Mateus Fernandes Réu; CARVALHO, Aparecido Augusto de. Evaluation of scoliosis using baropodometer and artificial neural network. 2017.
- CASTRO FORERO, Fabian Rodrigo. Sistema baropodométrico e classificação de escoliose utilizando técnicas de machine learning. 2019.
- CRD. Escoliose: Doença acomete mais de 50 milhões de crianças e adolescentes no mundo. Disponível em: <https://crd.med.br/sem-categoria/escoliose-doenca-acomete-mais-de-50-milhoes-de-criancas-e-adolescentes-no-mundo/>. Acesso em: 11 jul. 2024

Validation Loss: 0.8056, Accuracy: 63.41%
Balanced Accuracy: 0.6227
F1-Score: 0.7065
ROC-AUC: 0.6976
Matriz de Confusão do Fold 3:
[[3536 2464]
[4210 8031]]

Fig. 12. Resultados Fold 3 - 250 épocas