



# DETECÇÃO AUTOMÁTICA DE MATURIDADE ÓSSEA UTILIZANDO REDES NEURAIIS CONVOLUCIONAIS

Bianca Bertoldo de Oliveira\*<sup>1</sup> e Milena Bueno Pereira Carneiro<sup>1</sup>

<sup>1</sup>FEELT - Universidade Federal de Uberlândia

**Resumo** - A avaliação da idade óssea de pacientes pediátricos é uma prática clínica padrão para determinar sua maturidade biológica, sendo um exame de grande importância na identificação das condições de crescimento e desenvolvimento em crianças e na previsão de sua altura futura. Normalmente, esse processo envolve a revisão manual de imagens de radiografia da mão, no entanto, este método é demorado e propenso a erros no cálculo da idade esquelética. Portanto, um sistema inteligente de avaliação da idade óssea usando inteligência artificial poderia automatizar a tarefa repetitiva de análise de imagens. O algoritmo proposto, desenvolvido em Python com o auxílio de suas bibliotecas, utiliza uma arquitetura de regressão de rede neural convolucional com múltiplas entradas. Este método tem se mostrado promissor não apenas para demonstrações teóricas de teor acadêmico na área de Processamento Digital de Imagens, como também para uso clínico prático, com uma taxa de erro médio de 7,31 meses, possibilitando auxiliar profissionais médicos a avaliar a idade óssea de forma objetiva.

**Palavras-Chave** - Idade Óssea, Inteligência Artificial, Processamento Digital de Imagens, Python, Redes Neurais Convolucionais.

## AUTOMATIC DETECTION OF BONE MATURITY USING CONVOLUTIONAL NEURAL NETWORKS

**Abstract** - The evaluation of the bone age of pediatric patients is a standard clinical practice to determine their biological maturity, being an exam of great importance in identifying the conditions of growth and development in children and in predicting their future height. Typically, this process involves manual review of radiographic images of the hand, however, this method is time consuming and prone to errors in skeletal age calculation. Therefore, an intelligent bone age assessment system using artificial intelligence could automate the repetitive task of image analysis. The proposed algorithm, developed in Python with the aid of its libraries, uses a multi-input convolutional neural network regression architecture. This method has shown to be promising not only for theoretical demonstrations of academic content in the field of Digital Image

Processing, but also for practical clinical use, with an average error rate of 7,31 months, making it possible to help medical professionals to assess bone age in an accurate way.

**Keywords** - Boneage, Artificial Intelligence, Digital Image Processing, Python, Convolutional Neural Networks.

## I. INTRODUÇÃO

A estimativa da idade óssea é uma técnica amplamente utilizada para avaliar distúrbios do crescimento em pacientes pediátricos. Ele fornece informações valiosas sobre o desenvolvimento da maturidade óssea em relação à idade cronológica. No entanto, esse processo só é preciso para indivíduos entre 0 e 19 anos, onde o crescimento ósseo para completamente. Na prática clínica, uma idade óssea 20% abaixo ou acima da idade cronológica é considerada anormal [1]. Resultados tardios ou avançados na avaliação podem ser indicativos de distúrbios pediátricos mais graves, como problemas relacionados ao estado nutricional, precocidade puberal e hipertireoidismo [2]. A técnica padrão para análise da maturação óssea é baseada em uma varredura radiológica da mão não dominante e é comparada às referências dos métodos disponíveis de forma manual por um radiologista. Fazer estimativas precisas da idade óssea é uma tarefa complexa que requer uma compreensão completa de vários fatores relacionados ao desenvolvimento ósseo. Portanto, esse método tem uma carga de trabalho alta e consome recursos significativos, resultando em um longo tempo para realizar a análise. Além disso, a precisão da estimativa da idade óssea é vulnerável ao julgamento humano, o que pode levar a incertezas na avaliação.

Para uma breve revisão bibliográfica do trabalho de alguns outros pesquisadores nas áreas de pré-processamento e classificação da maturação óssea em radiografias de mão, os trabalhos escolhidos foram desenvolvidos de 2018 a 2022, e avaliados com relação a metodologia e resultados.

Todos os métodos propostos utilizaram a base de dados da RSNA - Radiological Society of North America (Sociedade de Radiologia da América do Norte) [3], que possui 12611 imagens de treino e validação. Em 4 dos 7 trabalhos citados foi empregado Data Augmentation – geração de novas imagens

\*bianca.bertoldo@ufu.br

utilizando cópias alteradas por exemplo em rotação, contraste, redimensionamento e orientação. Para avaliação dos resultados de predição, a métrica utilizada é o MAE (Mean absolute error - erro médio absoluto). Em [4], o MAE atingido foi de 6,44 meses. [5] encontrou MAE de 6,10 meses, porém, este possui algumas desvantagens. Ao todo, foram gerados 15 modelos diferentes combinados, diferenciados em região segmentada, sexo, classificação e regressão. Primeiro, foi necessário a rotulagem manual de 100 imagens para realizar a segmentação da mão inteira, e a rotulagem de mais 800 imagens para segmentação de regiões de interesse menores, tais como centros de ossificação. Ainda, as idades entre 0 e 3 anos foram excluídas do treinamento. [6] obteve MAE de 9 meses. [7] alcançou MAE de 8,59, onde foi preciso rotular manualmente 300 imagens para a etapa de segmentação, e também foi agrupado 2 modelos treinados. No método proposto por [8], o MAE atingido foi de 12 meses. Em [9], o MAE foi de 7,69 meses. Para [10], o MAE foi de 6,07 meses, onde foram agrupados 6 modelos diferentes obtidos treinando cada centro de ossificação separado. A Tabela 1 apresenta todas as informações observadas.

Tabela 1: Comparação de resultados para avaliação por RNAs de regressão

| Método | Segmentação | Data Augmentation | Resultados (MAE) |
|--------|-------------|-------------------|------------------|
| [4]    | Não         | Sim               | 6,44 meses       |
| [5]    | Sim         | Sim               | 6,10 meses       |
| [6]    | Não         | Sim               | 9 meses          |
| [7]    | Sim         | Não               | 8,59 meses       |
| [8]    | Não         | Não               | 12 meses         |
| [9]    | Sim         | Sim               | 7,69 meses       |
| [10]   | Sim         | Não               | 6,07 meses       |

O objetivo final deste trabalho é realizar testes que garantam a eficácia das técnicas propostas. A intenção é desenvolver uma ferramenta que possa ajudar radiologistas e outros profissionais médicos a fazerem diagnósticos mais precisos, aumentando assim a eficiência no tratamento de distúrbios relacionadas ao crescimento.

## II. FUNDAMENTAÇÃO TEÓRICA

As estruturas de Redes Neurais Artificiais (RNAs) ganharam destaque como a principal metodologia para análise de imagens, devido às suas grandes capacidades de aprendizado e vantagens em lidar com padrões complicados. A maioria dos algoritmos modernos de aprendizado de máquina inclui um estágio de pré e/ou pós-processamento que é integrado a uma rede neural profunda. Essas fases, baseadas em abordagens clássicas de processamento de imagem, são usadas para ajudar a resolver desafios de classificação, detecção ou segmentação resultantes.

RNA é uma estrutura de rede de predição e processamento de recursos não linear de autoaprendizagem. Com base em pesquisas modernas de neurociência, a RNA processa informações imitando a rede neural do cérebro. Geralmente, é organizada em três camadas: uma camada de entrada, camadas ocultas e uma camada de saída. As camadas ocultas são compostas por um determinado número de unidades de nós conhecidas como neurônios. Cada neurônio está ligado a cada unidade de nó na camada anterior. O papel dos neurônios é fazer uma transformação linear e uma transformação não linear

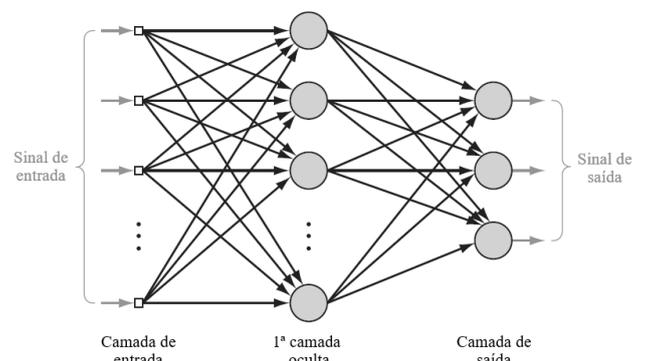
nos dados de entrada da camada anterior [11]. A topologia dos neurônios de uma rede neural está intimamente ligada ao método de aprendizado usado para treinar a rede.

### 1) MLP - Multilayer Perceptron

A primeira rede neural descrita algorítmicamente foi o perceptron, o tipo mais básico de rede neural utilizada para categorização de padrões, composto por um único neurônio, podendo realizar classificação de padrões com duas ou mais classes. Para resolver as restrições práticas do perceptron, existe uma estrutura de rede neural conhecida como Multilayer Perceptron (Perceptron de Multicamada) (MLP).

Um método popular para o treinamento de MLPs é o algoritmo de retropropagação, e decorre em duas fases: Na primeira, os pesos sinápticos da rede são fixados na fase de avanço, e o sinal de entrada é propagado pela rede, camada por camada, até atingir a saída. Na fase reversa, um sinal de erro é gerado comparando a saída da rede com uma resposta desejada. O sinal de erro resultante é enviado pela rede novamente, camada por camada, mas desta vez de maneira inversa. Durante a segunda fase, os pesos sinápticos da rede são gradualmente ajustados. A Figura 1 mostra o gráfico arquitetônico de um MLP com uma camada oculta e uma camada de saída.

Figura 1: Gráfico arquitetônico de um MLP com uma camada oculta.



### 2) CNN - Convolutional Neural Network

Convolutional Neural Networks (Redes Neural Convolucionais) (CNNs) são baseadas em MLPs projetado especificamente para reconhecer e avaliar formas bidimensionais. Essas redes empregam uma arquitetura única que é adequada para a classificação de imagens. As CNNs utilizam três conceitos fundamentais: convolução, pesos e bias compartilhados e pooling.

**Convolução:** Os pixels de entrada são conectados a uma camada oculta normalmente, mas apenas em partes pequenas e restritas da imagem de entrada. Cada neurônio na primeira camada oculta será vinculado a uma pequena região de neurônios de entrada, chamada de campo receptivo local [12]. O campo receptivo local move-se por toda a imagem de entrada.

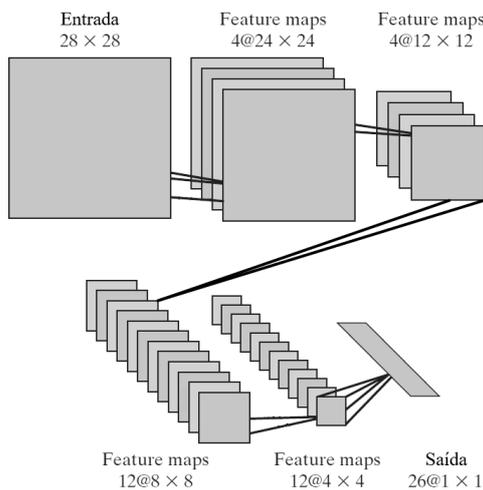
**Pesos e bias compartilhados:** Pesos e bias idênticos são utilizados para todos os neurônios ocultos, ou seja, todos os

neurônios da primeira camada oculta identificam a mesma característica. Como resultado, o mapa da camada de entrada para a camada oculta é conhecido como feature map (ou mapa de características). Para fazer o reconhecimento de imagem é necessário mais de um feature map, e assim uma camada convolucional completa consiste em vários feature maps diferentes.

**Camadas de pooling:** Seleciona e filtra as informações adquiridas, diminuindo o tamanho do modelo, acelerando o cálculo e aumentando a robustez dos recursos extraídos [13]. Uma camada de pooling prepara um feature map condensado. Cada unidade na camada de pooling pode resumir uma região de neurônios da camada anterior. A camada de saída executa um estágio adicional de convolução, remodelando o feature map recebido em um vetor para executar a tarefa de classificação final [13].

A Figura 2 representa a arquitetura de uma CNN, que consiste em uma camada de entrada, quatro camadas ocultas e uma camada de saída.

Figura 2: Rede convolucional para processamento de imagem.



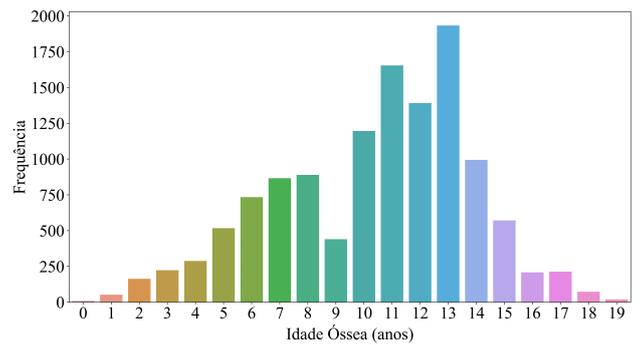
### III. METODOLOGIA DO TRABALHO

#### 1) Banco de Dados

Para fins de pesquisa, a RSNA publicou um conjunto de dados de fotografias da mão direita de jovens entre 0 e 20 anos. O acervo contém 12611 imagens de radiografia. Junto à base de imagens, é fornecido também uma tabela informando a idade óssea aferida por radiologistas e o gênero.

A aparência das radiografias no banco de dados varia muito devido aos diferentes métodos de aquisição utilizados na captura. O brilho, o contraste, a resolução e até a proporção das imagens podem diferir umas das outras. Os dados de imagem de raios-X usados para apoiar as conclusões deste artigo foram depositados no repositório RSNA em doi:10.1148/radiol.2018180736. Na Figura 3 é apresentado um gráfico de frequência da idade óssea em anos para melhor visualização.

Figura 3: Distribuição da base considerando a idade óssea.



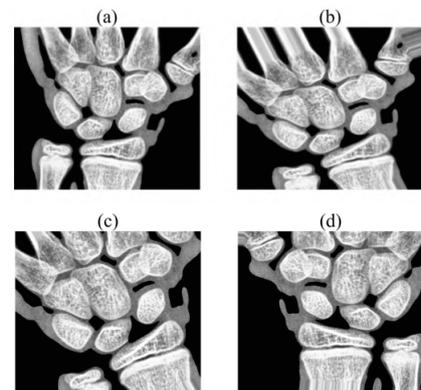
#### 2) Softwares e Bibliotecas

A linguagem de programação utilizada neste trabalho foi Python, desenvolvida na plataforma Anaconda com o ambiente de desenvolvimento do Jupyter Notebook. As bibliotecas Python utilizadas foram: NumPy e Matplotlib para manipulação de dados numéricos e criação de gráficos, OpenCV para manipulação de imagens, Pandas para manipulação da base de dados contendo informações de sexo e idade, e TensorFlow, Keras e Scikit-learn para desenvolvimento da RNA.

#### 3) Data Augmentation

A técnica de Data Augmentation foi empregada para expandir o conjunto de dados de imagem, e ao mesmo tempo para tornar a rede mais resistente à variâncias, duplicando coleções de imagens originais invertendo, girando, ampliando e alterando o brilho. Os procedimentos de Data Augmentation são realizado apenas para amostras de idade óssea em meses que possuem menos de 80 exemplares, dessa forma, produzindo cópias variadas até que se atinja o valor de 80 imagens. As distorções aleatórias aumentaram significativamente o tamanho da base para treinamento, de  $\approx 12600$  para  $\approx 21600$ . A Figura 4 exibe uma imagem original seguida de três imagens obtidas artificialmente.

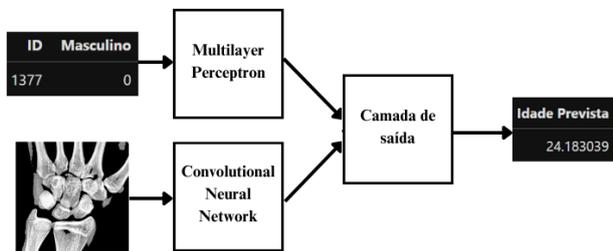
Figura 4: Resultados do uso de Data Augmentation. a) Imagem original. b), c) e d) Imagens geradas automaticamente.



#### 4) Arquitetura da Rede Neural

Dados mistos, que incluem dados numéricos, categóricos e visuais, podem ser processados usando uma variedade de modelos de aprendizado de máquina. Uma RNA complexa é necessária para lidar com o modelo de múltiplas entradas. O banco de dados contém imagens segmentadas de radiografias de mãos, e a característica categórica de sexo é reunido em um formato de arquivo de dados tabulares, que é posteriormente associado junto com as imagens. O processo de análise é triplo: imagem, características e concatenação. A CNN manipula a seção de processamento de imagens, enquanto a MLP lida com a seção de categorização. No estágio final, os resultados são integrados usando uma rede neural para fornecer a saída final prevista. A Figura 5 mostra em resumo o fluxo da rede completa.

Figura 5: Fluxo da RNA desenvolvida.



A rede MLP lida com o dado categórico de sexo, dado por 0 ou 1, sendo 0 feminino e 1 masculino. Inicia uma camada densa totalmente conectada, seguida por uma camada oculta totalmente conectada, e uma ativação linear é usada para a camada de saída de regressão ideal final. CNN é um bom classificador preditivo, mas deve enquadrar-se em um certo intervalo de valores para ser usado para regressão. Todos os valores de intensidade da imagem são divididos por 255, o valor máximo da imagem, para dimensioná-lo entre 0 e 1. A rede aplica uma convolução em três fases e performa a função de ativação ReLU antes de entrar em cada camada de pooling.

Cada convolução tem três tipos diferentes de filtro, ou seja, 16, 32 e 64, onde cada filtro tem o mesmo tamanho de 3x3 em loop. O resultado segue pela camada de pooling máximo, aplicada para reduzir a dimensão espacial do tamanho de entrada. O tamanho de pooling é usado duas vezes do tamanho 2x2 para diminuir o tamanho de entrada das imagens e acrescentar feature maps. O vetor de saída da CNN é acoplado com o vetor de saída do MLP no último estágio para formar um vetor de estrutura regressiva. A estrutura é processada usando a arquitetura de CNN + MLP para obter uma saída linear. A última camada linear é aplicada para fornecer previsão de saída numérica.

O desempenho dos algoritmos de RNA depende especificamente da seleção de um conjunto sólido de hiper parâmetros, e automatizar a otimização desses parâmetros pode tornar a construção de algoritmos complexos mais eficiente e menos demorada. Keras inclui algumas técnicas de otimização de hiper parâmetros, das quais foi escolhida a Hyperband [14]. É um método de early-stopping baseado em princípios que aloca um recurso predefinido, como iterações ou amostras de

dados, para configurações amostradas aleatoriamente e, em seguida, treina o modelo com cada configuração, interrompendo as configurações de treinamento com baixo desempenho ao alocar recursos adicionais a configurações promissoras. Hyperband faz uso significativo de reduções pela metade sucessivas, que atribui uma provisão a um conjunto de configurações de hiper parâmetros. Isso é feito de forma consistente e, uma vez gasto o recurso, metade das configurações é descartada com base no desempenho. As 50% melhores são retidas e ensinadas com uma nova provisão, enquanto as 50% restantes são descartadas. O processo continua até que reste apenas uma configuração [15]. A configuração do otimizador é feita de acordo com a Figura 6.

Figura 6: Inicialização do algoritmo de otimização.

```
tuner = kt.Hyperband(model_builder,
                    objective='val_accuracy',
                    max_epochs=10,
                    factor=3,
                    directory='my_dir',
                    project_name='intro_to_kt')
```

As épocas de treinamento (max\_epochs) foram definidas em 50, o fator foi mantido como padrão e o objetivo, por se tratar de um modelo de regressão, foi val\_mape, o valor do MAPE - Mean absolute percentage error (Erro percentual absoluto médio) para os dados de validação de cada iteração.

## IV. RESULTADOS

Para a estimativa da idade óssea foi testado um modelo de regressão contendo um mosaico das imagens do dedo e pulso como entrada, como demonstrado na Figura 7.

Figura 7: Imagem de entrada do modelo.



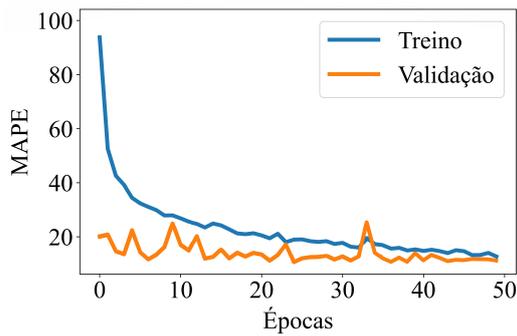
O modelo foi programado para treinar por 50 épocas, com taxa de aprendizado de 0,01 definido pelo otimizador. Early-stopping com uma paciência de 5 épocas foi empregada. A Tabela 2 mostra os valores de dimensões das camadas definidos pelo otimizador.

Tabela 2: Dimensões das redes CNN e MLP.

| Camada 1 CNN   | Camada 2 CNN         | Camada 3 CNN |
|----------------|----------------------|--------------|
| (100, 150, 16) | (50, 75, 32)         | (25, 37, 64) |
| Pooling Final  | Saída CNN Vetorizada | MLP          |
| (12, 18, 64)   | 13824                | 416          |

A curva de convergência da validação da rede neural aproximou-se do MAPE do conjunto de treinamento de 6,58%. A relação entre o erro no conjunto de treino e de validação e as épocas treinadas é mostrada na Figura 8.

Figura 8: Erros de treino e validação no treinamento do modelo.



Os cálculos de erro mais comuns são: Mean absolute percentage error (Erro percentual absoluto médio) (MAPE) e Mean absolute error (Erro médio absoluto) (MAE), definido pelas equações (1) e (2).

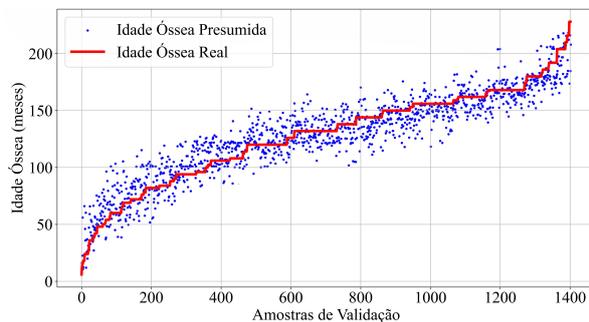
$$MAPE = \frac{1}{n} \sum_{i=1}^n \frac{|\text{Valor real} - \text{Valor previsto}|}{\text{Valor real}} \quad (1)$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |\text{Valor real} - \text{Valor previsto}| \quad (2)$$

O valor de MAE e MAPE do método de avaliação da idade óssea proposto para a base de teste foi de 7,31 meses e 6,75% respectivamente. O resultado do MAE ficou comparável aos resultados de referência de estado da arte. Quando analisado os resultados de MAPE em comparação com os valores de referência na literatura [16], o valor encontrado indica um modelo com previsão de alta precisão.

Para a construção do gráfico de dispersão, todas as amostras de idade óssea real na base de teste foram rearranjadas de forma crescente, representadas pela linha vermelha do gráfico. Os pontos azuis representam a idade óssea presumida para uma determinada amostra. O resultado é uma relação linear entre o valor real e o valor encontrado pelo modelo, indicando uma forte correlação entre os dois valores. A Figura 9 mostra o gráfico de dispersão do modelo aplicado na base de teste.

Figura 9: Gráfico de dispersão das previsões do modelo.



## V. CONCLUSÕES

Com o rápido desenvolvimento da tecnologia, o processamento de imagens digitais passou por uma rápida evolução e é amplamente utilizada em diversas áreas. É extremamente útil em sensoriamento remoto, medicina, reconhecimento e detecção de objetos. A profissão médica pode se beneficiar tre-

mentadamente do emprego de RNAs em imagens para colaborar com análises e diagnóstico.

A precisão da rede proposta foi obtida realizando a avaliação das radiografias do conjunto de dados de teste. O modelo proposto alcançou um MAE de 7,31 meses, comparável ao desempenho do estado da arte. Isso indica um nível substancial de concordância entre as revisões obtidas e a avaliação absoluta. A precisão da solução proposta é semelhante à obtida por radiologistas especializados e semelhante aos sistemas automatizados anteriores.

Para comparação direta, foram selecionados os métodos utilizando RNAs de regressão com maior e menor erro. Para estes trabalhos, o método com maior MAE foi o elaborado por [8] de 12 meses. O mesmo não utilizou nenhuma segmentação para a entrada do modelo e não fez uso de técnicas de Data Augmentation. [10] alcançou o melhor MAE de 6,07 meses sem empregar também Data Augmentation, contudo foi necessário segmentar 6 centros de ossificação para cada mão e dispor de técnicas de combinações de 6 modelos diferentes. A diferença desse método com o proposto neste trabalho foi de 1,24 meses, mostrando-se satisfatório, onde tal diferença não seria clinicamente significativa.

De acordo com os resultados relatados, pode-se afirmar que este trabalho trouxe novas contribuições para a área de processamento de imagens médicas, sendo possível auxiliar profissionais de ortopedia, pediatria, endocrinologia, entre outros, no diagnóstico de pacientes. Foi demonstrada a viabilidade de implementar um método totalmente automatizado para determinar a idade óssea. A solução proposta foi desenvolvida do início ao fim, pode ser facilmente treinada com diferentes conjuntos de dados e requer um tempo de treinamento razoável. Na prática, o sistema proposto pode ser facilmente implantado no ambiente clínico em um computador comum. O método de concatenar MLP e CNN pode ser implementado em sistemas que consistem em formatos de dados mistos.

As propostas para melhorias neste trabalho e sugestões para trabalhos futuros são: 1. Desenvolver uma CNN com ajuste fino de Transfer Learning, utilizando pesos de soluções relacionadas à tarefa específica de imagens médicas, por exemplo, outro sistema envolvendo classificação de imagens de raios-x; 2. Aumentar a quantidade de dados no conjunto de dados original uma vez que existe um desbalanceamento do banco de dados usado para treinamento em relação às idades; 3. Realizar testes com mais camadas, aumentar as dimensões das imagens de entrada e executar o processamento em computadores mais potentes.

## REFERÊNCIAS

- [1] MELMED, S. et al. *Williams textbook of endocrinology E-Book*. [S.l.]: Elsevier Health Sciences, 2015.
- [2] ZEFERINO, A. et al. Acompanhamento do crescimento. *Jornal de pediatria*, SciELO Brasil, v. 79, p. S23–S32, 2003.
- [3] RSNA. 2017. Disponível em: <<https://www.rsna.org/education/ai-resources-and->

- training/ai-image-challenge/rsna-pediatric-bone-age-challenge-2017>.
- [4] SOUZA, D.; OLIVEIRA, M. M. End-to-end bone age assessment with residual learning. In: IEEE. *2018 31st SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI)*. [S.l.], 2018. p. 197–203.
- [5] IGLOVIKOV, V. I. et al. Paediatric bone age assessment using deep convolutional neural networks. In: SPRINGER. *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support* [S.l.], 2018. p. 300–308.
- [6] MARROCOS, M. L. L. d. L. et al. *Detecção automática de idade óssea através da radiografia de mão e punho utilizando redes neurais convolucionais*. Universidade Federal de Campina Grande, 2019.
- [7] PAN, X. et al. Fully automated bone age assessment on large-scale hand x-ray dataset. *International journal of biomedical imaging*, Hindawi, v. 2020, 2020.
- [8] WESTERBERG, E. *AI-based age estimation using X-ray hand images: A comparison of object detection and deep learning models*. 2020.
- [9] ZULKIFLEY, M. A. et al. Intelligent bone age assessment: an automated system to detect a bone growth problem using convolutional neural networks with attention mechanism. *Diagnostics*, MDPI, v. 11, n. 5, p. 765, 2021.
- [10] GUO, L. et al. Bone age assessment based on deep convolutional features and fast extreme learning machine algorithm. *Frontiers in Energy Research*, Frontiers, v. 9, p. 888, 2022.
- [11] HOU, Y. et al. *The state-of-the-art review on applications of intrusive sensing, image processing techniques, and machine learning methods in pavement monitoring and analysis*. Engineering, Elsevier, v. 7, n. 6, p. 845–856, 2021.
- [12] NIELSEN, M. A. *Neural networks and deep learning*. [S.l.]: Determination press San Francisco, CA, USA, 2015. v. 25.
- [13] RIPLEY, B. D. *Pattern recognition and neural networks*. [S.l.]: Cambridge university press, 2007.
- [14] TENSORFLOW. 2022. Disponível em: <[https://www.tensorflow.org/tutorials/keras/keras\\_tuner](https://www.tensorflow.org/tutorials/keras/keras_tuner)>.
- [15] LI, L. et al. Hyperband: Bandit-based configuration evaluation for hyperparameter optimization. In: *International Conference on Learning Representations*. [S.l.: s.n.], 2017.
- [16] LEWIS, C. *International and business forecasting methods*. Butterworths, London. 1982.