CBIS'24 XX Congresso Brasileiro de Informática em Saúde
08/10 a 11/10 de 2024 - Belo Horizonte/MG - Brasil

Localização e classificação de epífises em radiografias carpais utilizando modelos YOLO

Localization and classification of epiphyses in carpal radiographs using YOLO models

Localización y clasificación de epífisis en radiografías carpianas utilizando modelos YOLO

Guilherme Nique da Silva[1], Viviane Rodrigues Botelho[2] , Thatiane Alves Pianoschi[2] , Carla Diniz Lopes Becker[2]

1 Master's Student, Federal University of Health Sciences of Porto Alegre – UFCSPA, Porto Alegre (RS), Brazil.
2 Ph.D., Federal University of Health Sciences of Porto Alegre – UFCSPA, DECESA, Porto Alegre (RS), Brazil.

Corresponding author: Guilherme Nique da Silva

E-mail: guilhermend@ufcspa.edu.br

Resumo

Objetivo: Neste estudo é proposto o desenvolvimento de um modelo de detecção de epífises em imagens de raio X, utilizando modelos de aprendizado de máquina. **Metodologia:** descrevemos o processo de aquisição do *dataset* e conduzimos testes com modelos como YOLOv5, YOLOv8 e *faster* R-CNN. **Resultados:** O modelo YOLOv8 obteve erro de 1% no *dataset* DHA, enquanto o modelo YOLOv5 em torno de 5%. **Conclusão:** Após uma análise comparativa, o YOLOv8 foi selecionado como o modelo ideal para a detecção final das epífises

Descritores: Detecção de objetos, epífise, idade óssea.

Abstract

Objective: This study proposes the development of a model for detecting epiphyses in X-ray images using machine learning models. **Methods:** We describe the process of dataset acquisition and conduct tests with models such as YOLOv5, YOLOv8, and Faster R-CNN. **Results:** The YOLOv8 model achieved a 1% error rate on the DHA dataset, while the YOLOv5 model achieved around 5%. **Conclusion:** After a

comparative analysis, YOLOv8 was selected as the ideal model for final epiphyses detection.

Keywords: Object detection, epiphysis, bone age.

Resumen

Objetivo: Este estudio propone el desarrollo de un modelo para detectar epífisis en imágenes de rayos X utilizando modelos de aprendizaje automático. **Metodología:** Describimos el proceso de adquisición de datos y realizamos pruebas con modelos como YOLOv5, YOLOv8 y Faster R-CNN. **Resultados:** El modelo YOLOv8 obtuvo un error del 1% en datos DHA, mientras que el modelo YOLOv5 obtuvo cerca del 5%. **Conclusión:** Tras un análisis comparativo, YOLOv8 se seleccionó como modelo ideal para la detección final de epífisis.

Descriptores: Detección de objetos, epífisis, edad ósea.

Introduction

Chronological age, as discussed by Schneider and Irigaray[6], refers to the age measured from the date of birth. This indicator is universal and standardized, serving as a basic reference for human development in chronological, biological, psychological, and social spheres. Despite its importance, this indicator does not allow the knowledge of skeletal and biological maturity of an individual. Bone age concerns the degree of maturation of the skeletal system. The difference between bone and chronological age can indicate developmental issues such as delays or advancements in skeletal growth, this differences in skeletal growth may suggest problems like, endocrine disorders, congenital disorders and others. This discrepancy can have significant implications for childs and adolescents' health and physical development, influencing aspects such as physical capability and growth potential. Pinto *et al.*[4] highlights the relation between bone age, hormonal markers, and physical capability, underscoring the complexity and importance of accurate assessment of bone age for a comprehensive understanding of human development.

Usually, bone age is assessed from carpal radiographs. Among the avaliable methods for its determination, the Greulich-Pyle (GP) and Tanner-Whitehouse (TW) are two predominant approaches in determining bone age, both utilizing X-rays of the left

hand and wrist. The GP method involves comparing the patient carpal radiograph with a standard atlas of radiographic images. In this method, the evaluator seeks the image in the atlas that most closely resembles the patient's X-ray, considering factors such as the individual's sex and biological age. This comparison results in an estimation of bone age, based on the best match found in the atlas, evaluating based on the maturation of individual epiphyses[2]. The TW method adopts a more detailed approach, assigning points to specific bones in the hand based on criteria related to the patient's sex. After scoring the bones in the defined region, which varies according to the version of the method, an overall score is calculated. This score is then used to generate a chart that is evaluated in relation to the patient's chronological age. Comparing this chart with chronological age allows the evaluator to estimate the patient's bone age. The TW method has undergone several revisions and is currently in its third version[2].

To achieve the task of predicting bone age using algorithms, deep learning models can be used. As defined by Chollet[7]:

> Deep learning is a specific subfield of machine learning: a new take on learning representations from data that puts an emphasis on learning successive layers of increasingly meaningful representations.(...) Modern deep learning often involves tens or even hundreds of successive layers of representations, and they're all learned automatically from exposure to training data.

In the context of applying deep learning models in medicine, the study "A deep learning model to classify and detect brain abnormalities in portable microwave based imaging system"[1] is of great relevance. This work highlights the efficient use of the YOLOv5 model for automatic detection and classification of brain abnormalities in microwave magnetic resonance images. Through the analysis of its results, the model proves to be a reliable and effective tool for real-time medical diagnostics. This technological advancement in the field of deep learning enables high reliability in disease diagnosis and monitoring, reinforcing the importance of these models in improving healthcare. "Aplicação de Deep Learning para Diagnóstico de Pneumonia Causada por COVID -19 a partir de Imagens de Raio X"[8] also shows the applicabitlity of deep learning models in health, developing a model to detecting abnormalities in chest X-ray images.

The epiphysis detection has great clinical importance and represent an important role on TW and GP methods, allowing the overall assessment to produce more solid and reliable information. Developed methods involving bone age prediction using the

entire hand, while accurate, end up being a black box, not presenting the methodology for calculating the final age, which may not always be as interesting in the clinical area. On the other hand, with the epiphysis method, the process becomes clearer for healthcare professionals, allowing them to have a better understanding of the intelligent system's workflow.

For the present study, the study of Koitka[3] served as the base for comparison, evaluating the accuracy of the models. The study uses the Faster R-CNN model, and although it is a competent model, YOLO was chosen for its ease of implementation. The model in the evaluated work achieved an average precision of 99%, proving effective for the proposed objective, but it has a higher prediction time compared to YOLO models. The referenced work reports a F1 score of 97%. The F1 score is a metric that represents the harmonic mean between precision and recall, where precision measures the proportion of true positives relative to the total positive predictions made by the model, and recall measures the proportion of true positives relative to the total examples that are positive.

Therefore, this paper presents the development of an epiphysis detection model, and a comparison with a similar work. To achieve this, YOLO models where developed, which have proven effective in the rapid and accurate detection of objects in as shown in the papers of Hossain[1] and Redmon[5].

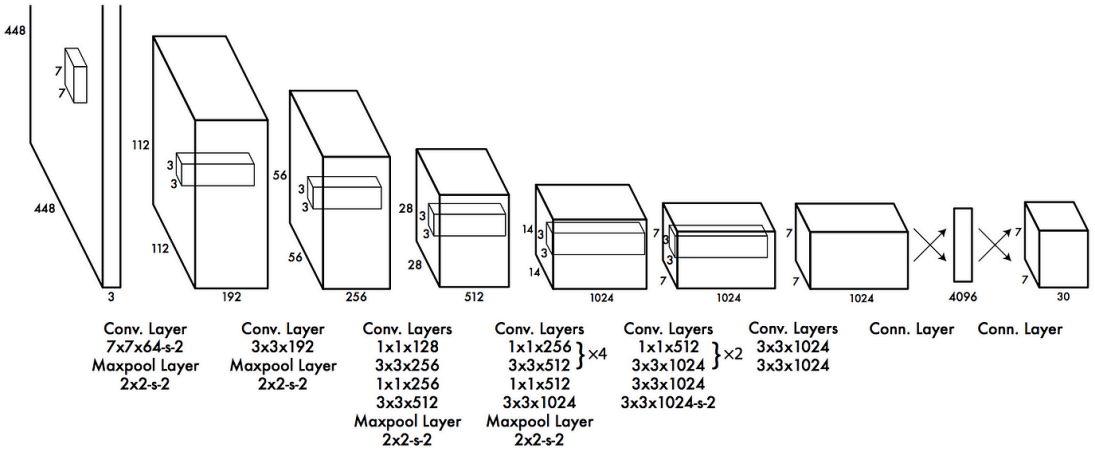
Materials and Methods

In this section, the YOLO model is presented and the steps that led to the development of the final model, including obtaining the dataset, training the model, and validating the model.

YOLO Model

Convolutional neural networks (CNN), especially those specialized in object detection like YOLO (You Only Look Once) represent a significant advancement in the field of computer vision. An illustrative depiction of this architecture introduced by Redmon[5] is presented in Figure 1.

Figure 1 – YOLO network detection has 24 convolutional layers followed by 2 fully connected layers. Alternating 1 x 1 convolutional layers reduce the features space from preceding layers [5].



YOLO architecture is close to GoogleNet, and it has overall 24 convolutional layers, four max-pooling layers, and two fully connected layers. Unlike previous methods that performed detection in two stages, first selecting regions of interest and then classifying these regions, YOLO unifies these processes, standing out for its ability to perform real-time detections by processing images in a single step. It divides the image into a grid and, for each grid cell, predicts bounding boxes and class probabilities, identifying objects and their locations simultaneously. This method not only increases detection speed, but also maintains considerable accuracy, making it ideal for real-time applications.

Dataset

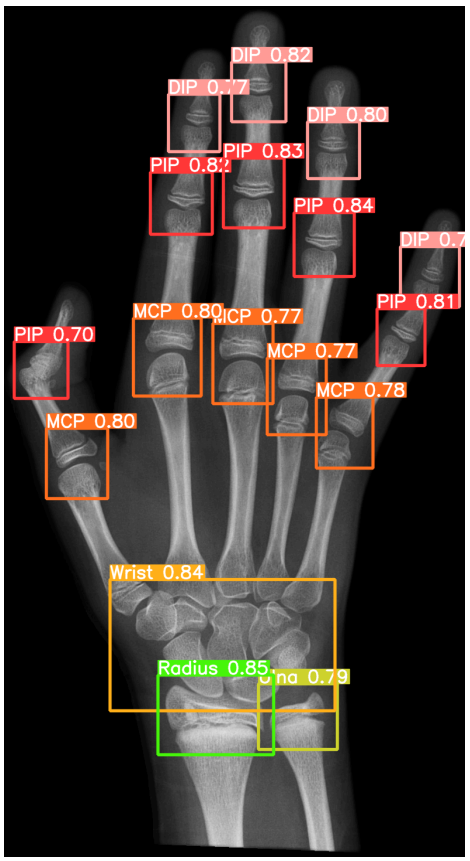
The dataset from the Koitka's paper[10] was used. This dataset consists of annotations of epiphysis from the Radiological Society of North America bone age dataset (RSNA)[9]. The RSNA bone age dataset[9] is publicly available and contain approximately twelve thousand hand X-ray images, each one accompanied by its respective bone age labeled by two pediatric radiologists.

For validation was used the DHA[11] dataset, which is also a bone age dataset containing 1390 images of hands X-rays and its respective labels, at the moment of the acquisition of the dataset, 954 images were able to be downloaded, resulting in a smaller dataset, but still useful for validation.

For epiphysis detection, a dataset with these epiphyses pre-annotated is necessary. The dataset from the referenced work[10] includes epiphysis annotations of 329 images from RSNA dataset.

Thus, the final dataset consisted of 240 training images and 89 test images. The classes in the dataset are Wrist, Metacarpophalangeal (MCP), Distal Interphalangeal (DIP), Proximal Interphalangeal (PIP), Ulna, and Radius. These classes correspond to the epiphyses present in the hand and contain information such as the maturation state, allowing them to be used for the analysis of the individual's bone age. These classes can be seen in Figure 2.

Figure 2 – Hand X-ray example with labeled classes.



Models

As mentioned, the goal is a fast and efficient epiphysis location model Therefore, although the model used in the reference work[3] was the Faster R-CNN, tests were conducted with YOLO models because they are easy to implement, test and have good

results. They come with a comprehensive library, allowing for quick testing and validation.

The YOLOv5 model tested was the intermediate model YOLOv5s, which takes an input image of 640 pixels. Its performance on the COCO dataset, widely used for model validation, is mAP@50 64.1. For training, 240 annotated images were used, while the remaining 89 were used for validation. The model was trained for approximately 100 epochs, with other hyperparameters set to their default values.

The YOLOv8 model was trained using the same parameters as mentioned above, configured with the same input image size, epochs, etc. The main difference is the model itself, as the YOLOv8s model was used. Both models utilized their respective pretrained weights. One of the advantages of YOLO models is their size. The models were trained using a relatively old GPU, a RTX 2060 (NVIDIA), for example, training the YOLOv5s model took 13 minutes.

Interface

In addition to the models, an interface for prediction and visualization of epiphyses in X-ray images was also developed. This application allows users to upload an image for visualization of located epiphyses and to make modifications to the original image. The purpose of this interface is to simulate the workflow of a professional when performing an automatic prediction on an X-ray image, thus enabling the identification of possible needs of the professional when conducting their assessment.

Results

Initially, the YOLOv5 model was tested, achieving an F1 score of 99%, which presented quite good results for the proposed objective but still did not match the model presented in the reference work.

However, the YOLOv8 model ended up yielding significantly better results in the tests conducted, achieving an F1 score of 98% and curves with very positive outcomes. As can be seen in the curves in Figures 3 and 4, the models achieved 0.98 at 0.66 and 0.99 F1 at 0.47 of confidence, meaning they were able to learn the detection task, and having only a slightly difference, where the YOLOv8 model is more conservative having

a higher confidence on a lower F1 score and the YOLOv5 the opposite, but still needing an external validation to understand better the differences.

In Figure 5, additional predictions of the model on validation images can be observed, displaying some of the obtained results.

In addition to the validation conducted above, the model was also tested on another dataset, the Digital Hand Atlas (DHA). This dataset consists of X-ray images divided by ethnicity, age, and sex, with ethnicities including Asians, Africans, Caucasians, and Hispanics, totaling 954 images. For predictions on the DHA images, only predictions above 0.5 confidence were considered valid. The results were as follows: out of 954 images.

Figure 3 – YOLOv8 F1-Confidence curve

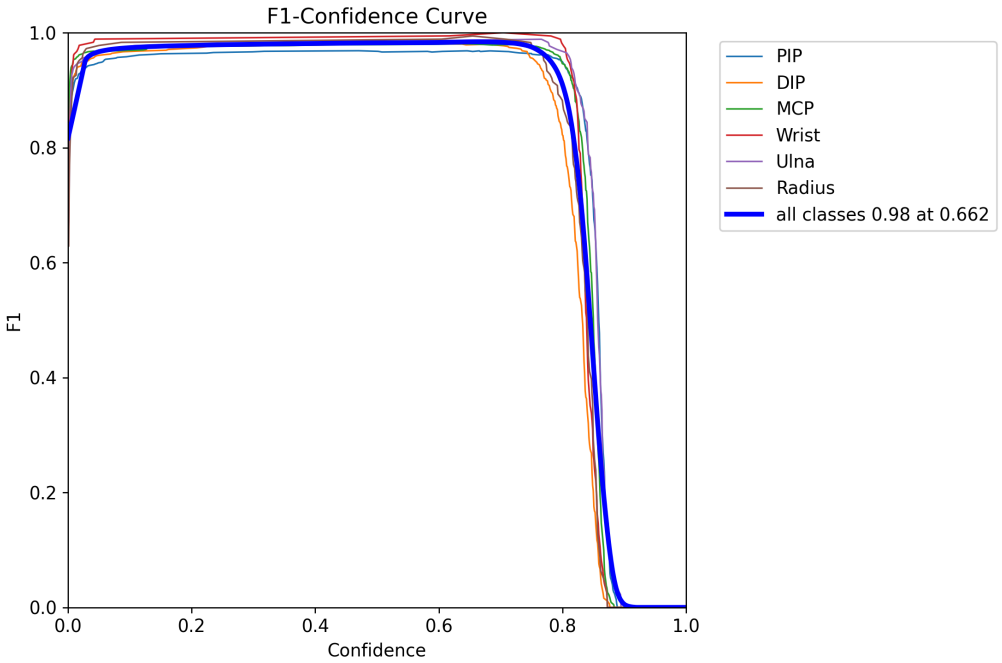


Figure 4 – YOLOv5 F1-Confidence curve.

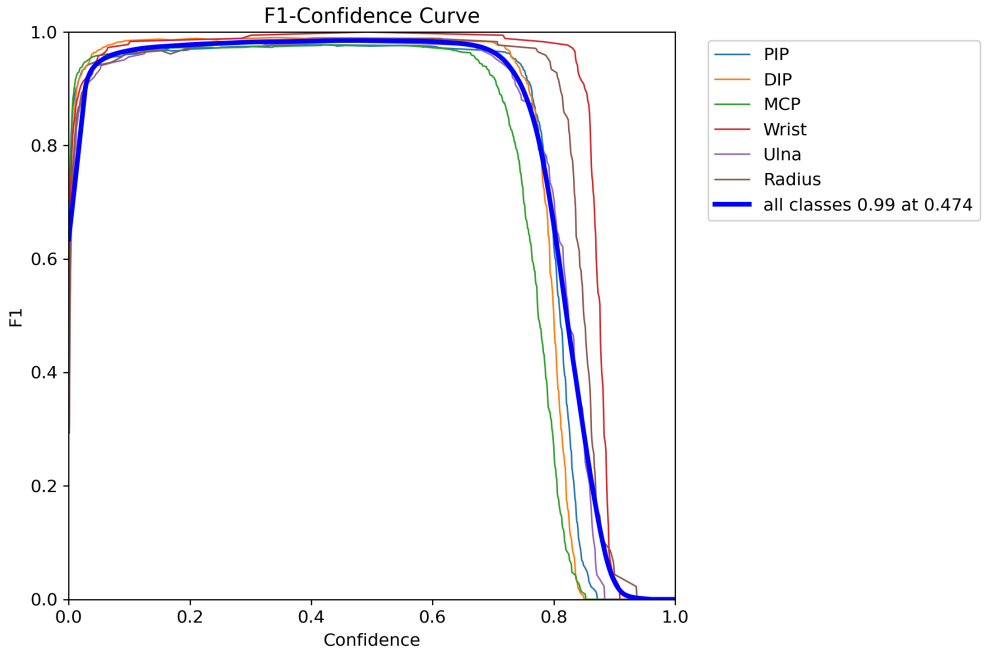
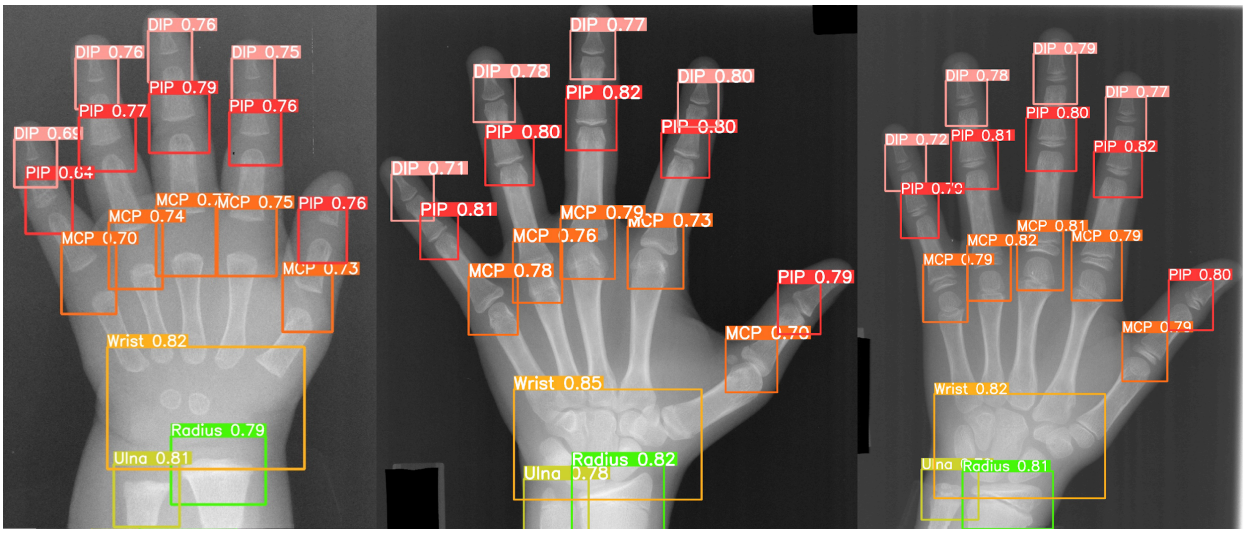


Figure 5 – Validation data results.



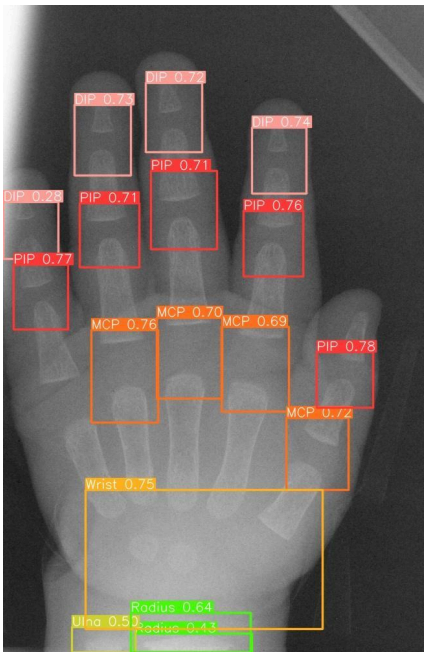
Manual evaluation of the YOLOv5 model revealed errors in 53 images (5.5%) and 91 epiphyses (0.7%). The errors followed mainly three patterns:

1. Images with low contrast between the background and the hand/bones resulted in problematic outcomes.
2. Some images exhibited a pattern of false positives in the lower right corner of the image, consistently identified as the Wrist class, possibly due to image noise or dataset balancing.

3. Hands of young children, which may be due to an imbalance in the dataset, particularly in wrist bones (Radius, Ulna, Wrist).

In the YOLOv8 model, errors were observed in 9 images (1%). Specifically evaluating the epiphyses, 13 errors were observed out of a total of 13356 epiphyses (0.1%). In general, the errors have a noticeable pattern, occurring more frequently in X-ray images of infants in their first months, as shown in Figure 6.

Figure 6 – Example of bad result.

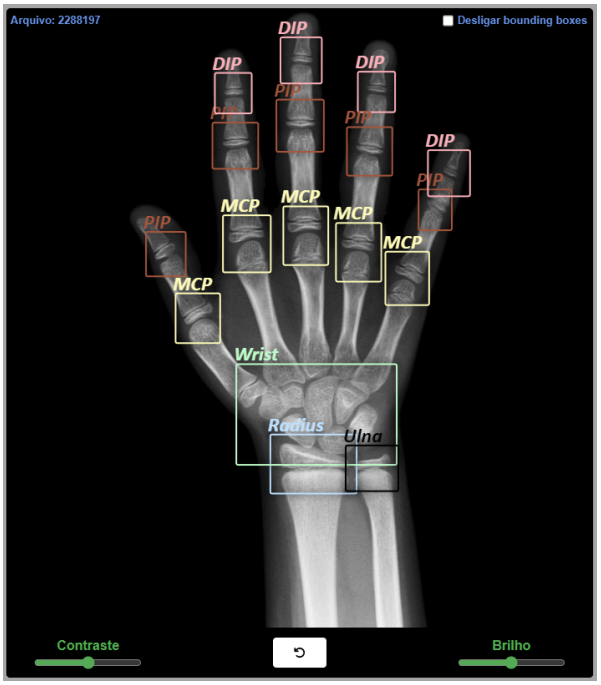


It can be observed that even though during training the models yielded comparable results, when exposed to a different dataset, differences in the models' generalization capabilities become apparent, with the YOLOv8 model proving superior. Nevertheless, the tests were conducted on an untreated dataset, where various image transformations could have been applied to assist the models in their predictions. One such modification is the application of the CLAHE filter, which increases contrast in the image and is widely used to enhance the quality of radiographic images.

Figure 7 presents the prediction result screen of the developed interface, where various information are displayed. On the side of the screen, some of these include the detected epiphyses, options for modifying the image visualization such as contrast and

brightness adjustments, and a button to toggle the visualization of the epiphyses on and off.

Figure 7 – Results of detected epiphysis shown on developed tool for visualization.



Conclusion

The proposal to develop a model for locating and classify epiphyses in carpal radiographs images arose as a preliminary step to the development of a bone age prediction model, which requires individual identification of epiphyses for the precise prediction. Therefore, there was a need to swiftly develop a detection model capable of creating a dataset for the subsequent model. The use of the YOLOv8 model for this dataset creation task appears quite feasible given the achieved results. It might even be possible to employ this model on the RSNA images themselves to create a more comprehensive dataset for epiphysis detection in X-ray images and make it available for future research.

It is believed that errors in X-ray images of younger individuals occur more frequently because the quantity of such images is proportionally smaller compared to the rest of the dataset. Additionally, structural differences tend to be more pronounced in these images, which can potentially mislead the model when encountering such cases.

As demonstrated by the developed interface, this work can serve as a foundation for a future tool that not only provides epiphysis detection but also bone age prediction, offering a complete pipeline for epiphysis detection and bone age prediction. Allowing it to be integrated into the medical environment, it can be used as a diagnostic tool, improving the process.

Referências

1. Hossain A, Islam MT, Almutairi AF. A deep learning model to classify and detect brain abnormalities in portable microwave based imaging system. 2022. Scientific Reports.

2. Jones J, Greulich and Pyle method. 2020.

3. Koitka S, Kim MS. Qu M. Fischer A, Friedrich CM, Nensa F. Mimicking the radiologists' workflow: Estimating pediatric hand bone age with stacked deep neural networks. 2020. Medical Image Analysis 64.

4. Pinto, Vanessa C. Relação da idade óssea e marcadores hormonais com a capacidade física de adolescentes. 2017. J Hum Growth Dev.

5. Redmon J, Divvala S, Girshick R, Farhadi A. You Only Look Once: Unified, Real-Time Object Detection. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016. (CVPR).

6. Schneider, R, Irigaray, T. Envelhecimento na atualidade: aspectos cronológicos, biológicos, psicológicos e sociais. Estudos de psicologia. 2008.

7. Chollet, F. Deep Learning with Python. 2017.

8. Trombetta GBW, Fröhlich W da R, Rigo SJ, Rodrigues CA. Aplicação de Deep Learning para Diagnóstico de Pneumonia Causada por COVID -19 a partir de Imagens de Raio X. J Health Inform [Internet]. 15º de março de 2021 [citado 21º de maio de 2024];12. Disponível em: https://jhi.sbis.org.br/index.php/jhi-sbis/article/view/828

9. https://www.rsna.org/rsnai/ai-image-challenge/rsna-pediatric-bone-age-challenge-2017.

10. Koitka S, Demircioglu A, Kim MS, Friedrich CM, Nensa F.Ossification area localization in pediatric hand radiographs using deep neural networks for object detection. 2018.

11. https://ipilab.usc.edu/research/baaweb/.