

# Improved Skeletal Maturity Estimation through Combined Tanner-Whitehouse and Greulich and Pyle Approaches with Artificial Intelligence

\*Note: Subtitles are not captured in Xplore and should not be used

1<sup>st</sup> André Brasiliano da Silva

*Dept. of Electrical and Computer Engineering*

*Mackenzie Presbyterian University (UPM)*

São Paulo, Brazil

ORCID: 0000-0001-5531-0792

2<sup>nd</sup> Cristiane

*dept. name of organization (of Aff.)*

*name of organization (of Aff.)*

City, Country

email address or ORCID

3<sup>rd</sup> Carlos Longui

*dept. name of organization (of Aff.)*

*name of organization (of Aff.)*

City, Country

email address or ORCID

4<sup>th</sup> Rogério de Oliveira

*dept. name of organization (of Aff.)*

*name of organization (of Aff.)*

City, Country

email address or ORCID

5<sup>th</sup> Gustavo Scalabrin

*dept. name of organization (of Aff.)*

*name of organization (of Aff.)*

City, Country

email address or ORCID

6<sup>th</sup> Leandro A. Silva

*dept. name of organization (of Aff.)*

*name of organization (of Aff.)*

City, Country

email address or ORCID

**Abstract**—Accurate estimation of skeletal maturity is crucial to the diagnosis and treatment of growth-interfering diseases, such as hypothyroidism, celiac disease, or growth hormone deficiency. Traditionally, this evaluation has relied on manual processes using hand radiography, which are time consuming and heavily reliant on the evaluator's expertise. There are two main diagnostic approaches, namely Greulich and Pyle (GP) and Tanner-Whitehouse (TW), each with its limitations. In TW, an image is compared to an atlas to identify the closest match and estimate bone age, while GP focuses on region-based hand analysis, producing better results. Despite significant advancements, Artificial Intelligence (AI)-based automation methods for skeletal maturity estimation still fall short of the accuracy achieved by experienced professionals. AI approaches have mainly focused on TW, which tends to yield lower precision estimates. This paper proposes a novel combination of the TW and GP approaches using Convolutional Neural Networks (CNN) and YOLO (You Only Look Once). Our experiments were carried out using radiographs from Brazilian patients, which included data from associated diseases - a contribution of this study, as age estimation often ignores the influence of diseases. By integrating the TW and GP methods, we achieved a notable 35% percent improvement compared to the better-isolated approaches. We could identify the regions of the hand that showed superior precision. Overall, this research presents a significant advancement in skeletal maturity estimation by leveraging the strengths of both the TW and GP approach through AI. The results demonstrate the potential for improved accuracy in diagnosing and treating growth-involving diseases, thus reducing the reliance on manual processes and improving the efficiency of healthcare professionals.

**Index Terms**—radiology, pediatrics, bone age, machine learning, deep neural networks.

## I. INTRODUCTION

Estimating bone age (BAA - Bone Age Assessment) [1] is crucial for endocrinology and genetic evaluation of patients

with growth disorders and endocrine abnormalities [2]. As described by [3], it is a standard clinical examination used to determine a child's skeletal maturity and to find discrepancies between skeletal age (the stage of bone development) and chronological age (the age obtained from the date of birth). In addition, the skeletal maturity of a child from a radiological bone image of a hand, compared with chronological age, allows for estimating the future adult height of the child.

The methods used most traditionally to examine radiographs to estimate skeletal maturity are the Greulich and Pyle method [4] preferred by 76% pediatricians and radiologists and the different versions of the Tanner and Whitehouse method [5] used by about 20%, which are preferentially used in Brazil. Other less used methods are the Fels method [6] and [7] used by about 4% of professionals.

About 30 years ago, a bone age behavior system called HANDX [8] proposed replacing manual estimation with a computer-aided diagnosis system. Among the few commercial systems available is BoneXpert [9], [10], which estimates the TW and GP methods. Since then, some studies have been carried out for the safety of bone age; for example, the works presented by [9], [11]–[15], focused on studies on bone age trail from X-ray images.

Technologies based on artificial intelligence (AI) and machine learning (ML) have been transforming the health area, particularly in the last two decades, with examples of applications in disease detection, more accurate diagnosis, personalized treatment, and drug development, among others. So many others. ML systems can learn and improve their performance by adding new data. [16] briefly review the main applications and methods used in the medical field. They include algorithms

for prediction, image processing, classification, clustering, dimensionality reduction, etc., with applications in assisted diagnosis, medical robotics, and several medical areas [17].

However, the estimation of bone age is performed manually most of the time by visually comparing the radiograph of the patient's hand based on a standard atlas. [18] presented a study in which clinical practice evaluators generated an error in the estimation process of up to 0.50 years for the Tanner-Whitehouse (TW) estimation method and 0.82 years for the Greulich and Pyle (GP) estimation method. [5] describes that the error in estimating bone age using the proposed method can be up to 0.25 years.

In this context, this paper proposes constructing an algorithm for estimating bone age using machine learning techniques, convolutional neural networks (CNN), and computer vision techniques for preprocessing hand X-ray images, using and providing a database of data containing the diagnoses previously performed in the methods of, applying for both estimation methods Greulich and Pyle (GP) and Tanner-Whitehouse (TW) combining the results in the same hand X-ray image.

## II. RELATED WORK

[19], [20] reviews and discusses techniques, uses, and trends. Within radiology, these techniques have also been applied in the automation of bone age estimation [21]–[23], with an emphasis more recently on works employing deep neural networks (DL, Deep Learning) [14], [24]–[28].

The work presented by [29] approaches the TW [30] method, executing the process in two phases, the first being the detection of regions of interest, six areas of the 13 indicated by TW, using the Digital Hand Atlas dataset with 1,375 samples, applying CNN Inception V4, obtain a Mean Absolute Error (MAE) result of 0.59 months.

[31] highlights the study carried out by [32] on the Greulich and Pyle (GP) and Tanner-Whitehouse (TW2) methods, highlighting that the methods are highly subjective with visual examinations, using radiological images of the left hand, depending a lot on the professional's experience. Using the Digital Hand Atlas dataset with 1,390 samples, applying the convolutional neural network VGG and some computer vision techniques, for example, histogram equalization and data augmentation, we obtain a mean absolute error (MAE) result of 0.79 months.

[33] describes some steps for using a convolutional neural network (CNN) for the benefit of images of good quality and low quality images. First, a baseline CNN model called “BoNet+” based on *DenseNet* [34] was used to evaluate sound quality X-ray images and investigate the choice of loss function. The main idea is to solve the problem of poor quality X-ray images. Using the Digital Hand Atlas dataset with 1400 samples, applying CNN VGG and the data augmentation technique, we obtain a mean absolute error (MAE) result of 0.76 months.

[35] use the Radiological Society of North America (RSNA) dataset with 12,611 samples to estimate bone age

using the GP method, applying the Canny filter in the image to detect corners, and VGG CNN, obtain a Mean Absolute Error (MAE) result of 1.23 months.

[36] present work with the RSNA dataset with 12,611 samples and the Digital Hand Atlas dataset with 1425 samples, estimating bone age for the TW methods, Inception V2 CNN obtain a Mean Absolute Error (MAE) of 0.38 months.

[37] bring a case of bone age estimation for the GP and TW methods, using a Digital Hand Atlas dataset, with 1229 samples, resizing the images to  $300 \times 300$  pixels, and applying Gaussian and watershed filters, with Inception V2 CNN, obtain a Mean Absolute Error (MAE) result of 0.346 months, the best result compared to all related works.

[38] show a bone age estimation for the TW method using YOLOv5, using the RSNA Dataset with 12,611, obtain a Mean Absolute Error (MAE) of 6.53 months.

For detecting and extracting ROIs from X-ray images [29], [39] shows that most of the studies for the estimation of bone age used Faster R-CNN, and all of these studies had the phenomenon of missing ROI detection.

Table VI, see Appendix A, shows the results of all related works, demonstrating the database used in each experiment, the estimation method, the computer vision techniques used in some preprocessing steps, the neural network, the validation method and their respective performance.

In this context, the state of art shows works that bring different techniques and datasets. For this work, we intend to improve the methods presented in the related works for a better bone age estimation with a minimum MAE and show the estimation in both ways, TW and GP, in the same output image, applying machine learning techniques to increase the accuracy.

## III. METHODOLOGY

This section presents the steps to implement the method that can estimate bone age and combine the estimation method in the same result.

### A. Object Detection and Image Segmentation for Tanner-Whitehouse (TW) ROIs

In bone age estimation using the Tanner-Whitehouse (TW) method, Regions of Interest (ROIs) are identified on hand X-rays to determine bone age. ROIs are specific areas of the bones that show the most noticeable changes as the bones mature [40].

The TW with ROIs is a reliable and accurate method for estimating bone age widely used in clinical practice. However, manually identifying and measuring ROI can be time-consuming and requires a skilled radiologist or technician [41].

YOLO (You Only Look Once) algorithms [42] have a high detection speed, and it is a widely developed object detection and image segmentation model developed; YOLOv8 is used to identify regions of interest (ROIs) with object detection and image segmentation [42]–[44]. Figure 1 shows the timeline of the YOLO versions already released.

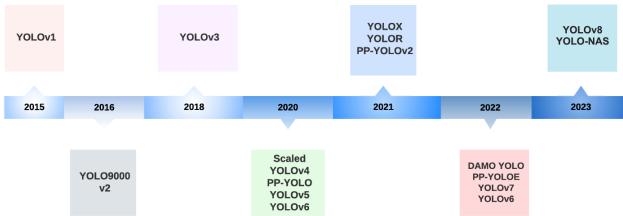


Fig. 1. A timeline of YOLO versions [45].

YOLOv8 [45] uses an anchor-free model with a decoupled head to independently process objectness, classification, and regression tasks. This design allows each branch to focus on its task and improves the overall accuracy of the model. In the YOLOv8 output layer, they used the sigmoid function as the activation function for the objectness score, representing the probability that the bounding box contains an object. It uses the softmax function for the class probabilities, representing the probabilities of objects belonging to each possible class.

For medical images, [46] shows a method for using Yolo for medical images to detect lung cancer. Figure 2 illustrates object detection using YOLOv5 in two stages, which first extract the candidate regions, then classify the candidate regions, and finally locate them accurately.

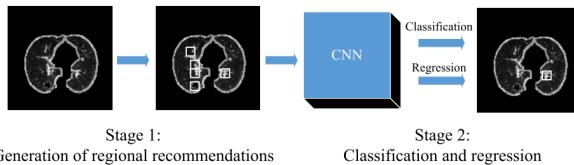


Fig. 2. Object detection using Yolo for medical images of lung cancer [46].

Figure 3 illustrates the Regions of Interest (ROIs) of the TW method [28] annotated in the x-ray image using YoloV8 for image segmentation.

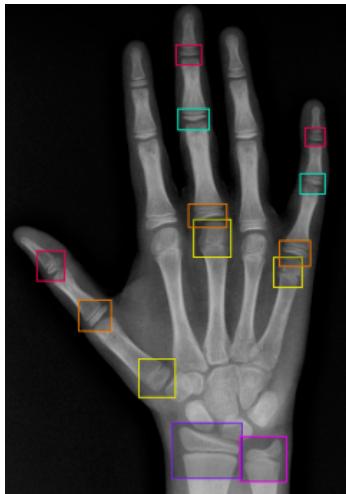


Fig. 3. TW ROIs annotating in the x-ray image for training using YoloV8 for image segmentation.

## B. Convolutional neural network (CNN) VGG16

Visual Geometry Group (VGG) is a deep convolutional neural network (CNN) architecture used in image classification tasks [47].

To apply VGG16 to bone age estimation, the pre-trained weights from the X-ray hand images can be used as a starting point, and the model can be fine-tuned on the bone age dataset by updating the last few layers to output a regression prediction of bone age. X-ray images can be pre-processed and normalized, and bone age labels can be scaled and standardized to improve training performance [48].

One potential advantage of using VGG16 for bone age estimation is its ability to learn high-level features from X-ray images, such as bone structure and shape, which are essential for accurate bone age estimation [49].

[50] refer to VGG16 as a promising approach to estimate bone age and has shown competitive performance compared to other deep learning methods such as ResNet and DenseNet. However, more research and validation are needed to establish its utility and effectiveness in clinical practice. The Figure 4 shows the VGG16 architecture [51].

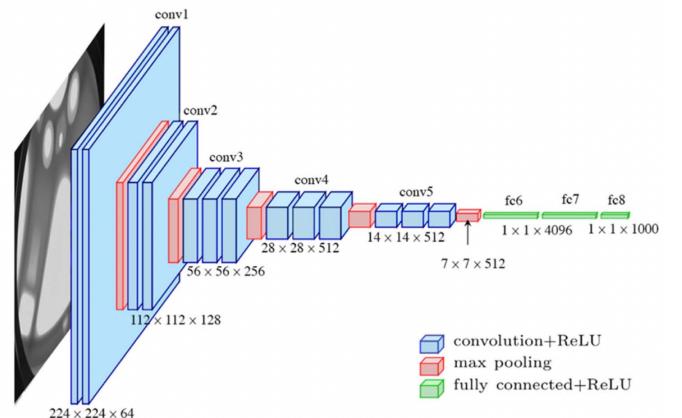


Fig. 4. Arquitetura da rede VGG16 [51].

When using the VGG16 model for bone age estimation, similar to other image classification tasks, several key hyperparameters need to be considered for effective training. They control how the model learns from the data and affect its performance. For example, the learning rate is similar to the speed with which the model learns. A higher learning rate means that the model learns faster, but it may also jump around and not learn accurately. A lower learning rate makes the model learn more slowly, but it could also learn more precisely.

The batch size of the hyperparameters is analogous to the number of images the model is exposed to simultaneously. A larger batch size implies that the model is exposed to more images at once, which can accelerate the training process, but requires more memory. On the contrary, a smaller batch size implies that the model is exposed to fewer images at a time, which can be slower but may result in better learning.

The number of epochs is the number of times the model goes through the entire dataset during training. Training for more epochs can help the model learn more from the data, but too many epochs can lead to overlearning and poor performance on new data.

Other hyperparameters, like the optimizer and dropout rate, influence the model's learning process. The optimizer is like the strategy the model uses to adjust its learning based on the data it sees. The dropout rate is a way to prevent overfitting, where the model memorizes training data instead of learning general patterns.

### C. Data Augmentation

Data augmentation [52] is a common technique used in machine learning and computer vision to artificially increase the size of a dataset by creating modified versions of existing images. It is often used in image classification, object detection, and other image-based tasks.

The process involves applying various transformations to the original images, such as rotations, translations, scaling, flipping, and changing brightness, contrast, or color. These modifications are intended to simulate real-world variations in the appearance of the objects being imaged and increase the diversity of the dataset.

For example, a data augmentation process for a dataset of images might include flipping some images horizontally to create a new set of images in the opposite direction. The process might also include rotating the images by a few degrees, adding noise, or adjusting the brightness and contrast [52].

In general, data augmentation is an essential technique for improving the performance of image-based machine learning models and is widely used in computer vision tasks [28].

Image data augmentation is a technique explored in this work, with the objective of increasing the number of samples of hand X-ray images with more variations, improving the bone age learning process.

Figure 5 illustrates the result of the data augmentation process implemented by [28].

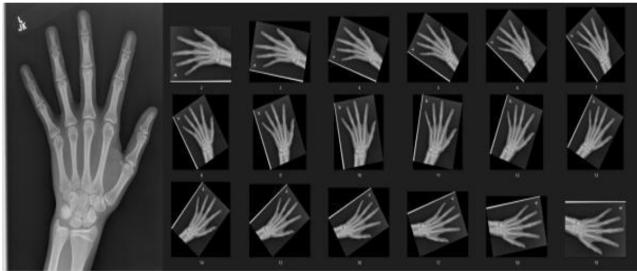


Fig. 5. Data augmentation process implemented by [28].

One commonly used technique is image rotation, in which bone radiographs are rotated at certain angles to mimic different perspectives. Flipping the images horizontally or vertically provides mirror images, enhancing the diversity of the dataset. Scaling and resizing the images help the model handle different resolutions and image scales. Image translation, involving

horizontal or vertical shifts, allows the model to recognize bones at different positions.

Shearing is another technique applied to create deformations and irregularities in bone radiographs. By adjusting image brightness and contrast, the model becomes more adaptable to varying lighting conditions. Image cropping is useful for focusing the model's attention on relevant bone regions.

### D. Dataset

Created by **Santa Casa de Misericórdia de São Paulo** [53], the database used in this work consists of a structure that does not exist in the current literature on the subject of estimating bone age. This database contains 434 samples of hands X-ray images, images that have a standard dimension of  $2704 \times 3404$ , and a data structure containing all patient diagnoses applied in Greulich and Pyle (GP) estimation methods [4] and Tanner-Whitehouse (TW) estimation methods [5].

The database is composed of some information that will be crucial for the implementation of the bone age estimation algorithm and is partially represented in Table I, where Male refers to gender in a binary type, fields Radio, Ulna, Metacarpo I, Metacarpo III, Metacarpo V, Prox phalanx I, Prox phalanx III, Prox phalanx V, Middle phalanx III, Med V phalanx, Distal phalanx I, Distal phalanx III and Distal phalanx V are the 13 regions of interest of the Tanner-Whitehouse (TW) method [5] and represents the stage of bone maturity that will be explained in the section IV-A3, the DOB field is the date of birth of the analyzed patients, Weight is the patient's Weight, Height is the patient's Height, Age is the patient's age, Greulich and Pyle is the bone age estimated using the GP method, BonExpert is the bone age estimated using the BoneXpert application [9], Score is the score assigned after classifying the regions of interest, TW2 RUS is the estimated bone age by the TW method.

The representation of the stage of bone maturity could be grouped into some parts [54]. These parts are Ulna, Epyphysis, Metaphysis, Radius, Phalanx, and Metacarpal, which are shown in Figure 6.

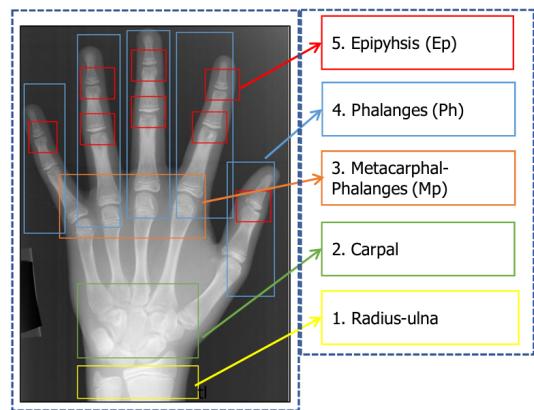


Fig. 6. ROI areas and parts of hand X-ray [54].

TABLE I  
THE TABLE CONTAINING INFORMATION FROM THE DATABASE CREATED BY THE SANTA CASA TEAM [53].

Index	Column Name	Type
1	Male	Binary
2	Radio	Categorical (A-I)
3	Ulna	Categorical (A-I)
4	Metacarpus I	Categorical (A-I)
5	Metacarpus III	Categorical (A-I)
6	Metacarpus V	Categorical (A-I)
7	Phalanx prox I	Categorical (A-I)
8	Phalanx prox III	Categorical (A-I)
9	Phalanx prox V	Categorical (A-I)
10	Phalanx med III	Categorical (A-I)
11	Phalanx med V	Categorical (A-I)
12	Distal Phalanx I	Categorical (A-I)
13	distal phalanx III	Categorical (A-I)
14	distal phalanx V	Categorical (A-I)
15	DOB	Date
16	Weight	Numeric
17	Height	Numeric
18	Age	Numeric
19	Greulich and Pyle	Numeric
20	BonExpert	Numeric
21	Score	Numeric
22	TW2 RUS	Numeric

Figure 7 represents an example of images extracted from samples of training database data and their related bone age, presented in months.

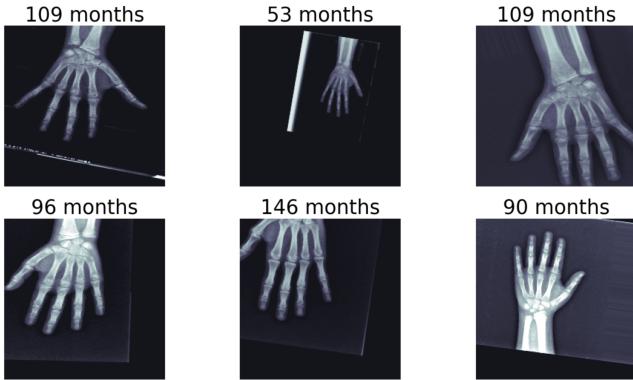


Fig. 7. Samples of images used in the training database, after separating between training and testing data, with their representation in months.

The dataset samples are selected from patients with no kind of pathology or disease because they could affect the model [55].

#### E. Cross Validation

Cross-validation is a common technique used in machine learning to assess the performance and generalizability of a model [56]. In the context of bone age estimation, cross-validation can be used to evaluate the accuracy and reliability of a bone age estimation model using reference data estimating bone ages and to evaluate its generalizability to unseen data. Helps mitigate issues such as overfitting. Using cross-validation, researchers and practitioners can obtain a more robust estimate of the model's performance and reliability in

estimating bone ages, which is essential in academic research and clinical practice in endocrinology.

It involves splitting the dataset into K subsets or folds. The model is trained K times, each time using K-1 folds as training data and the remaining fold as validation data. This process is repeated K times, each fold serving as the validation set exactly once. It involves partitioning the dataset into multiple folds, each serving as a validation/test set, while the remaining folds are used for training the model. The model is trained on a subset of the folds and evaluated on the remaining fold, and this process is repeated multiple times with different fold combinations. The model's performance is then assessed using appropriate evaluation metrics, such as mean absolute error (MAE), by comparing the model's predicted bone ages with the reference bone ages in the validation/test set.

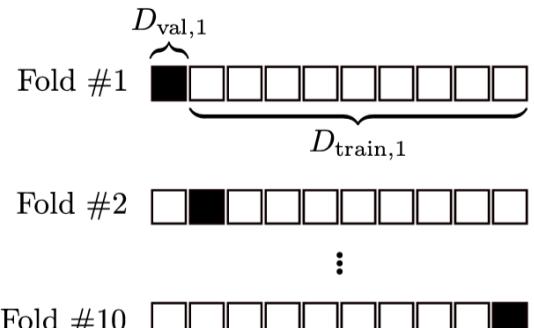


Fig. 8. 10-fold cross-validation [56]. The data set is randomly split into ten disjoint subsets, each containing (approximately) 10% of the data. The model is trained on the training set and then applied to the validation set. Illustrates this process for k=10, i.e., 10-fold cross-validation. In the first fold, the first subset serves as validation set  $D_{val,1}$ , and the remaining nine subsets serve as training set  $D_{train,1}$ .

For this work, the cross-validation can be applied to bone age estimation with references:

- Number of folds (k): Cross-validation involves dividing the dataset into k folds of equal size, where k is a hyperparameter that determines the number of folds. Typical values for k are 5 or 10, but other values such as 3, 7, or even higher can also be used depending on the size of the dataset and the complexity of the model [57]. In this case, the value of 5 folds is used because of the size of the dataset.
- Randomization: Randomization was used to shuffle the dataset before partitioning it into folds. This helps reduce potential bias in the data and ensures that the folds represent the overall dataset. Randomization can be particularly useful when dealing with intrinsic ordering or pattern datasets.

#### F. Mean Absolute Error (MAE)

The MAE of the bone age assessment technique is being evaluated. MAE [35] can then be calculated with the equation 1, where  $n$  is the sample value,  $y_i$  is the measured value, and  $\hat{y}_i$  is the actual value.

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (1)$$

The mean absolute error(MAE) is a useful measure widely used in model evaluations, originates from a measure of average error [58] and is often used in the evaluation of vector-to-vector regression models [59].

#### IV. EXPERIMENTAL METODOLOGY

The development of the algorithm consists of two implementation steps to divide the scope of the bone age estimation algorithm. The first step aims to estimate bone age using the Greulich and Pyle (GP) method.

The sequence of the implementation process is represented in Figure 9, which highlights each step of the dataset, pre-processing, data augmentation, cross-validation, image segmentation, and algorithm execution, resulting in a pre-trained model based on the methods of Greulich and Pyle (GP) and Tanner-Whitehouse (TW).

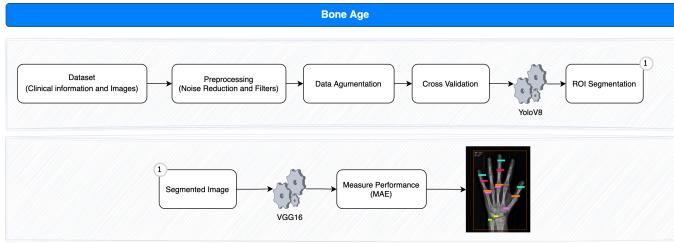


Fig. 9. The sequential flow of an implementation of the first stage of the estimation algorithm implemented in this work, based on the method of Greulich and Pyle (GP) [4].

The first step is the data entry step that uses the dataset presented in Section III-D, which contains the parameters necessary to train the CNN to estimate bone age, with patient characteristics such as age, gender, and bone age estimation, bone age previously performed by specialists, and hand X-ray images. Information is used to estimate bone age and data analysis after the execution of the algorithm.

The preprocessing step is used to apply filters and noise reduction techniques in the image dataset, for input parameters in the proposed neural network.

The data augmentation step applies the process to increase the image samples in the dataset, presented in the section III-C.

The cross-validation step prepares the database for the execution of CNN YoloV8 for image segmentation and extraction of the ROIs of the images.

After the segmentation image process, CNN VGG16 is used for bone age estimation based on Greulich and Pyle (GP) and Tanner-Whitehouse (TW) methods.

The division of the algorithm implementation process into steps allows for the verification of the results for each of the stages in isolation, also guaranteeing the gradual progress of the research work.

#### A. Experiments and parameters

After the preparation of the dataset with pre-processing and image segmentation using YOLOv8, the VGG16 CNN is used for the estimation of bone age based on Greulich and Pyle (GP) and Tanner and Whitehouse (TW) methods.

1) *Data augmentation parameters:* In this work, the data augmentation parameters used for increase the samples are:

- **Flipping:** Mirroring the images horizontally and vertically.
- Shear:  $\pm 20^\circ$  Horizontal,  $\pm 20^\circ$  Vertical.
- Crop: 0% Minimum Zoom, 30% Maximum Zoom.
- Hue: Between  $-30^\circ$  and  $+30^\circ$ .
- Brightness: Between -20% and +20%.

The dataset increment went from 434 samples to 1039 samples, an increase of more than 41% of the original dataset.

2) *Estimation model training parameters:* The hyperparameters for training a VGG16 model for bone age estimation are set following:

- Learning rate: 0.0001.
- Batch size: 64 for the training samples and 256 for the valid samples.
- Number of epochs: 250.
- Dropout rate: 0.5.
- Weight decay: l2(0.01).

3) *Estimating by the stage of maturity of the bone:* The dataset contains the maturity stage of each bone in the regions of interest presented by Tanner-Whitehouse (TW). [60] details each maturity bone stage and is shown in Figure 10 and each description is show in Table II.

TABLE II  
MATURITY STAGE OF BONE [61].

Maturity Stage	Description
A	Absent
B	Single calcium deposit
C	Center is distinct in appearance
D	The maximum diameter is half or more the width of the metaphysis
E	The border of the epiphysis is concave
F	The epiphysis is as wide as the metaphysis.
G	Epiphysis caps the metaphysis
H	The fusion of epiphysis and metaphysis has begun
I	Completed epiphyseal fusion

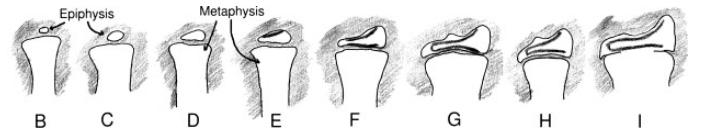


Fig. 10. Stage of Bones of TW ROIs [60].

The calculation of the score with estimated ROIs is used to determine the bone age estimate and the MAE is calculated with the difference between the measured estimate and the actual estimate.

## V. EXPERIMENTAL RESULTS

The result of the algorithm execution process is the combination of both GP and TW estimation methods, shown in the output image.

Figure 11 represents the image classification using YoloV8 for each ROI, performed from the training dataset for the TW method. The YoloV8 network first takes the X-ray image as input, for recognition, and obtains the prediction result of bone maturity.

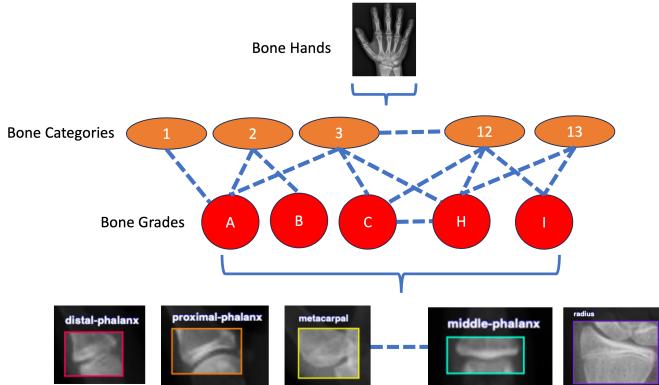


Fig. 11. The tree structure of bone categories and grades for image segmentation of each ROI of the TW method using YoloV8.

The prediction result for the bone category of the input image is obtained through the detection model. The corresponding branch structure is selected on the basis of the result. Finally, bone age is estimated according to the TW method based on the 13 regions of interest, and the result is the output.

Figures 12 and 13 show bone age estimation, which leads to actual age estimation with GP and TW methods, and the annotation of the regions of interest using YoloV8. Validation was carried out with the mass of test data, obtaining a result of a difference in months of up to 0.21 months, a satisfactory result compared to the related work, showing the estimation of both methods GP and TW.

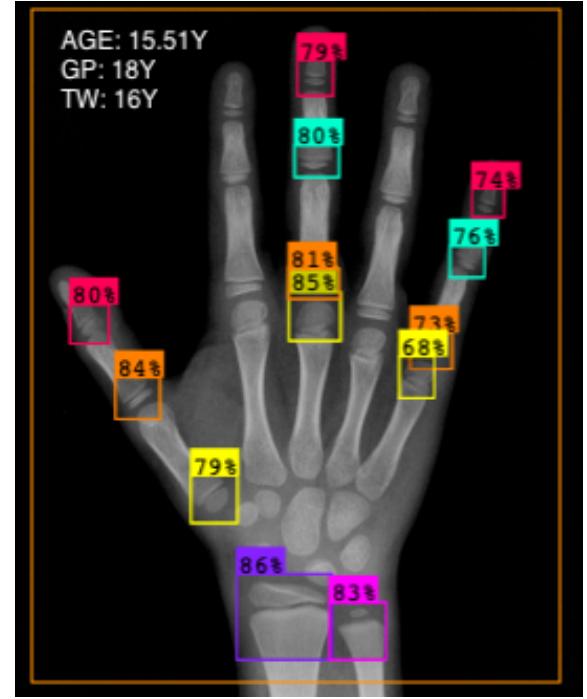


Fig. 12. Bone age estimation result with GP and TW methods, actual age 15.51Y, estimated GP 18Y, and TW 16Y.

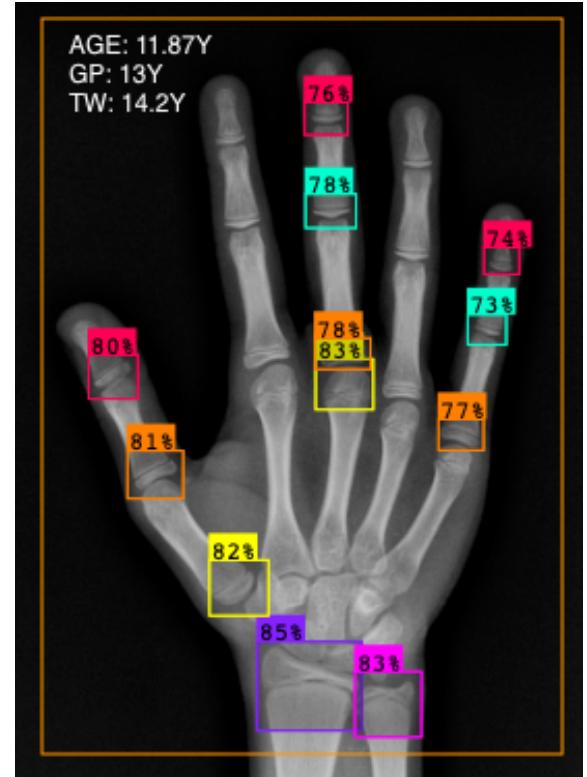


Fig. 13. Bone age estimation result with GP and TW methods, actual age 11.87Y, estimated GP 13Y, and TW 14.2Y.

Table III presents the accuracy of YOLOv8 of the regions of interest of the TW method by gender and shows a mean accuracy of 86% for the male gender and 89% for the females.

TABLE III  
ACCURACY OF TW ROIs BY GENDER USING YOLOv8.

ROI	Male	Female
Distal Phalanx	79%	83%
Metacarpal	94%	94%
Middle Phalanx	74%	83%
Proximal Phalanx	80%	91%
Radius	98%	94%
Ulna	94%	86%
<b>All</b>	<b>86%</b>	<b>89%</b>

The Table IV presents the experimental results of MAE by gender. The tests show a better result for the female gender, with an MAE of **0.16** months for the GP + TW methods, versus an MAE of **0.26** months for the male gender, and an MEA for both genders of **0.21** months.

TABLE IV  
MAE IN MONTHS OF EXPERIMENTAL BONE AGE ESTIMATION RESULT BY GENDER.

Gender	GP	TW	GP+TW
Female	0.21	0.18	0.16
Male	0.75	0.55	0.26
<b>MAE</b>	<b>0.48</b>	<b>0.36</b>	<b>0.21</b>

Table V shows the mean absolute error (MAE) of bone age of males and females based on different age ranges.

TABLE V  
MEAN ABSOLUTE ERROR (MAE) IN MONTHS OF THE BONE AGE ABOUT MALES AND FEMALES BASED ON DIFFERENT AGE RANGES.

Age (range)	Female GP	Female TW	Female GP+TW	Male GP	Male TW	Male GP+TW
6-7	0.38	0.02	0.11	1.71	0.37	0.22
7-8	1.14	0.72	0.17	0.70	1.22	0.24
8-9	0.92	0.99	0.23	1.45	1.20	0.29
9-10	0.17	0.26	0.15	0.03	0.24	0.28
10-11	0.40	0.38	0.21	0.94	1.01	0.23
11-12	0.15	0.02	0.12	0.14	1.21	0.28
12-13	0.54	0.21	0.14	0.59	0.78	0.25
13-14	0.23	0.41	0.20	1.31	0.62	0.25
14-15	0.40	0.81	0.13	0.74	0.81	0.27
15-16	0.36	0.45	0.15	0.03	1.28	0.29
16-17	1.41	0.09	0.24	0.23	0.49	0.28
<b>6-18</b>	<b>0.22</b>	<b>0.18</b>	<b>0.16</b>	<b>0.75</b>	<b>0.55</b>	<b>0.26</b>

The comparative tables present the accuracy of a variety of details taken from the model, allowing experts to discuss the outcomes and gain a deeper understanding of the patient's data, for example, the age that has more precision or accuracy for the bone estimation.

## VI. CONCLUSIONS

This work was evaluated using a Convolutional Neural Network to estimate bone age from hand X-ray images, implementing the Greulich and Pyle (GP) and Tanner-Whitehouse (TW) estimation methods, with a dataset that includes the estimation of all the methods studied, which contain actual data from patients.

The training process consists of preprocessing images with computer vision techniques, data augmentation, cross-validation, object detection, and image segmentation using YoloV8, with the intention of recognizing the degrees of bone maturity defined by Greulich and Pyle (GP) and Tanner-Whitehouse (TW), combining the results in the same image.

A convolutional neural network based on the VGG16 architecture was used to implement the estimation model [47], modifying it with the addition of new convolutional layers and segmentation of the image to extract the regions of interest, resulting in an absolute mean error (MAE) of up to 0.21 months for the Greulich and Pyle (GP) and Tanner-Whitehouse (TW) methods, obtaining the lowest mean absolute error among related works.

This work shows the results for the gender and age group for each estimation method and the combination of both, detailing the accuracy of each region of interest and providing a piece of complete information for analysis. Implementing an automation model to estimate bone age can improve and optimize a professional analysis, showing the regions of interest, and the bone age estimation of both methods in the same image.

The contributions of this work consist of a different model for estimating bone age with hand X-ray images, combining the essential estimation techniques, using a real-world dataset and images.

In future work, a more extensive database of images can be captured to increase the accuracy of the algorithm, estimating a more significant number of scenarios with different types of patients with more age groups and more gender distributions.

The tests with other IA models for classification and image segmentation identify more image details for the regions of interest and could improve the precision in the classification of the ROIs.

A mobile application that uses this algorithm to calculate bone age could be developed to enhance the scalability of this technique with a reliable practical application that could help improve the model with the capture of the professional's feedback.

## REFERENCES

- [1] A. L. Creo and W. F. Schwenk, "Bone age: a handy tool for pediatric providers," *Pediatrics*, vol. 140, no. 6, 2017.
- [2] J. Zerin and R. Hernandez, "Approach to skeletal maturation." *Hand clinics*, vol. 7, no. 1, pp. 53–62, 1991.
- [3] V. Gilsanz and O. Ratib, *Hand bone age: a digital atlas of skeletal maturity*. Springer Science & Business Media, 2005.
- [4] W. W. Greulich and S. I. Pyle, *Radiographic atlas of skeletal development of the hand and wrist: Based on the Brush Foundation study of human growth and development initiated by T. Wingate Todd, MB, Ch. B., FRCS*. Stanford University Press, 1950.
- [5] J. Tanner, M. Healy, H. Goldstein, and N. Cameron, "Assessment of skeletal maturity and prediction of adult height: Tw3 method saunders," *London, UK*, pp. 1–110, 2001.
- [6] M. M. Hoffer, "Assessing the skeletal maturity of the hand-wrist: Fels method." *JBJS*, vol. 71, no. 1, pp. 156–157, 1989.
- [7] O. Eklöf and H. Ringertz, "A method for assessment of skeletal maturity," in *Annales de radiologie*, vol. 10, no. 3, 1967, pp. 330–336.
- [8] D. J. Michael and A. C. Nelson, "Handx: a model-based system for automatic segmentation of bones from digital hand radiographs." *IEEE transactions on medical imaging*, vol. 8, no. 1, pp. 64–69, 1989.

- [9] H. H. Thodberg, S. Kreiborg, A. Juul, and K. D. Pedersen, "The boneexpert method for automated determination of skeletal maturity," *IEEE transactions on medical imaging*, vol. 28, no. 1, pp. 52–66, 2008.
- [10] <https://bonexpert.com/>, accessed: 2023-04-16.
- [11] C.-W. Hsieh, T.-L. Jong, Y.-H. Chou, and C.-M. Tiu, "Computerized geometric features of carpal bone for bone age estimation," *Chinese medical journal*, vol. 120, no. 9, pp. 767–770, 2007.
- [12] A. Gertych, A. Zhang, J. Sayre, S. Pospiech-Kurkowska, and H. Huang, "Bone age assessment of children using a digital hand atlas," *Computerized medical imaging and graphics*, vol. 31, no. 4-5, pp. 322–331, 2007.
- [13] D. Giordano, I. Kavasidis, and C. Spampinato, "Modeling skeletal bone development with hidden markov models," *Computer methods and programs in biomedicine*, vol. 124, pp. 138–147, 2016.
- [14] C. Spampinato, S. Palazzo, D. Giordano, M. Aldinucci, and R. Leonardi, "Deep learning for automated skeletal bone age assessment in x-ray images," *Medical image analysis*, vol. 36, pp. 41–51, 2017.
- [15] M. Mansourvar, "Bone age assessment using hand and clavicle x-ray images," Ph.D. dissertation, University Malaya, 2014.
- [16] M. Shehab, L. Abualigah, Q. Shambour, M. A. Abu-Hashem, M. K. Y. Shambour, A. I. Alsalibi, and A. H. Gandomi, "Machine learning in medical applications: A review of state-of-the-art methods," *Computers in Biology and Medicine*, vol. 145, p. 105458, 2022.
- [17] B. Lorica and P. Nathan, <https://gradientflow.com/wp-content/uploads/2022/03/GradientFlow-2022-AI-in-Healthcare-report-v3.pdf>, 2022, accessed: 2023-04-16.
- [18] R. Bull, P. Edwards, P. Kemp, S. Fry, and I. Hughes, "Bone age assessment: a large scale comparison of the greulich and pyle, and tanner and whitehouse (tw2) methods," *Archives of disease in childhood*, vol. 81, no. 2, pp. 172–173, 1999.
- [19] G. Choy, O. Khalilzadeh, M. Michalski, S. Do, A. E. Samir, O. S. Pianykh, J. R. Geis, P. V. Pandharipande, J. A. Brink, and K. J. Dreyer, "Current applications and future impact of machine learning in radiology," *Radiology*, vol. 288, no. 2, pp. 318–328, 2018.
- [20] Z. Zhang and E. Sejdić, "Radiological images and machine learning: trends, perspectives, and prospects," *Computers in biology and medicine*, vol. 108, pp. 354–370, 2019.
- [21] R. v. Rijn and H. Thodberg, "Bone age assessment: automated techniques coming of age?" *Acta Radiologica*, vol. 54, no. 9, pp. 1024–1029, 2013.
- [22] M. Mansourvar, R. G. Raj, M. A. Ismail, S. A. Kareem, S. Shanmugam, S. Wahid, R. Mahmud, R. H. Abdullah, F. H. F. Nasaruddin, and N. Idris, "Automated web based system for bone age assessment using histogram technique," *Malaysian Journal of Computer Science*, vol. 25, no. 3, pp. 107–121, 2012.
- [23] F. C. Castro, E. L. Flôres, G. A. Carrijo, A. C. P. Veiga, and M. P. B. Carneiro, "Localização, segmentação e classificação automática de regiões de interesse para a avaliação da maturidade esquelética utilizando o método de tanner-whitehouse," *Research on Biomedical Engineering*, vol. 28, no. 1, pp. 53–67, 2012.
- [24] Y. A. Ding, F. Mutz, K. F. Côco, L. A. Pinto, and K. S. Komati, "Bone age estimation from carpal radiography images using deep learning," *Expert Systems*, vol. 37, no. 6, p. e12584, 2020.
- [25] J. R. Kim, W. H. Shim, H. M. Yoon, S. H. Hong, J. S. Lee, Y. A. Cho, and S. Kim, "Computerized bone age estimation using deep learning based program: evaluation of the accuracy and efficiency," *American Journal of Roentgenology*, vol. 209, no. 6, pp. 1374–1380, 2017.
- [26] H. Lee, S. Tajmir, J. Lee, M. Zissen, B. A. Yeshivas, T. K. Alkasab, G. Choy, and S. Do, "Fully automated deep learning system for bone age assessment," *Journal of digital imaging*, vol. 30, no. 4, pp. 427–441, 2017.
- [27] S. Li, B. Liu, S. Li, X. Zhu, Y. Yan, and D. Zhang, "A deep learning-based computer-aided diagnosis method of x-ray images for bone age assessment," *Complex & Intelligent Systems*, pp. 1–11, 2021.
- [28] N. Poojary, P. Pokhare, P. Poojary, C. Raghavani, and J. Khanapuri, "A novel approach for bone age assessment using deep learning," *International Journal of Scientific Research in Computer Science, Engineering and Information Technology*, pp. 67–75, 05 2021.
- [29] T. D. Bui, J.-J. Lee, and J. Shin, "Incorporated region detection and classification using deep convolutional networks for bone age assessment," *Artificial intelligence in medicine*, vol. 97, pp. 1–8, 2019.
- [30] H. Carty, "Assessment of skeletal maturity and prediction of adult height (tw3 method). edited by jm tanner, mjr healy, h. goldstein and n. cameron. pp 110. london, etc: Wb saunders, 2001. isbn: 0-7020-2511-9.£ 69.95." 2002.
- [31] J. Zhou, Z. Li, W. Zhi, B. Liang, D. Moses, and L. Dawes, "Using convolutional neural networks and transfer learning for bone age classification," in *2017 International Conference on Digital Image Computing: Techniques and Applications (DICTA)*. IEEE, 2017, pp. 1–6.
- [32] A. M. Mughal, N. Hassan, and A. Ahmed, "Bone age assessment methods: A critical review," *Pakistan journal of medical sciences*, vol. 30, no. 1, p. 211, 2014.
- [33] J. Guo, J. Zhu, H. Du, and B. Qiu, "A bone age assessment system for real-world x-ray images based on convolutional neural networks," *Computers & Electrical Engineering*, vol. 81, p. 106529, 2020.
- [34] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 4700–4708.
- [35] A. Wibisono, M. S. Saputri, P. Mursanto, J. Rachmad, A. T. W. Yudasubrata, F. Rizki, E. Anderson *et al.*, "Deep learning and classic machine learning approach for automatic bone age assessment," in *2019 4th Asia-Pacific Conference on Intelligent Robot Systems (ACIRS)*. IEEE, 2019, pp. 235–240.
- [36] S. Koitka, M. S. Kim, M. Qu, A. Fischer, C. M. Friedrich, and F. Nensa, "Mimicking the radiologists' workflow: Estimating pediatric hand bone age with stacked deep neural networks," *Medical Image Analysis*, vol. 64, p. 101743, 2020.
- [37] Y. Han and G. Wang, "Skeletal bone age prediction based on a deep residual network with spatial transformer," *Computer Methods and Programs in Biomedicine*, vol. 197, p. 105754, 2020.
- [38] X. Xu, H. Xu, and Z. Li, "Automated bone age assessment: A new three-stage assessment method from coarse to fine," in *Healthcare*, vol. 10, no. 11. MDPI, 2022, p. 2170.
- [39] S. Koitka, A. Demircioglu, M. S. Kim, C. M. Friedrich, and F. Nensa, "Ossification area localization in pediatric hand radiographs using deep neural networks for object detection," *PloS one*, vol. 13, no. 11, p. e0207496, 2018.
- [40] Q. Hui, C. Wang, J. Weng, M. Chen, and D. Kong, "A global-local feature fusion convolutional neural network for bone age assessment of hand x-ray images," *Applied Sciences*, vol. 12, no. 14, p. 7218, 2022.
- [41] X. Mao, Q. Hui, S. Zhu, W. Du, C. Qiu, X. Ouyang, and D. Kong, "Automated skeletal bone age assessment with two-stage convolutional transformer network based on x-ray images," *Diagnostics*, vol. 13, no. 11, p. 1837, 2023.
- [42] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 779–788.
- [43] J. Redmon and A. Farhadi, "Yolo9000: better, faster, stronger, in-proceedings—30th ieee conference on computer vision and pattern recognition(2017)," 2017.
- [44] Y. Y. O. L. Once), <https://github.com/ultralytics/ultralytics>, accessed: 2023-02-19.
- [45] J. Terven and D. Cordova-Esparza, "A comprehensive review of yolo: From yolov1 to yolov8 and beyond," *arXiv preprint arXiv:2304.00501*, 2023.
- [46] Z. Ji, Y. Wu, X. Zeng, Y. An, L. Zhao, Z. Wang, and I. Ganchev, "Lung nodule detection in medical images based on improved yolov5s," *IEEE Access*, 2023.
- [47] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.
- [48] V. I. Iglovikov, A. Rakhlina, A. A. Kalinin, and A. A. Shvets, "Paediatric bone age assessment using deep convolutional neural networks," in *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support: 4th International Workshop, DLMIA 2018, and 8th International Workshop, ML-CDS 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 20, 2018, Proceedings*. Springer, 2018, pp. 300–308.
- [49] X. Pan, Y. Zhao, H. Chen, D. Wei, C. Zhao, and Z. Wei, "Fully automated bone age assessment on large-scale hand x-ray dataset," *International journal of biomedical imaging*, vol. 2020, 2020.
- [50] Y. Gao, T. Zhu, and X. Xu, "Bone age assessment based on deep convolution neural network incorporated with segmentation," *International journal of computer assisted radiology and surgery*, vol. 15, pp. 1951–1962, 2020.

- [51] T. Sugata and C. Yang, "Leaf app: Leaf recognition with deep convolutional neural networks," in *IOP Conference Series: Materials Science and Engineering*, vol. 273, no. 1. IOP Publishing, 2017, p. 012004.
- [52] A. Zhang, Z. C. Lipton, M. Li, and A. J. Smola, "Dive into deep learning," 2021. [Online]. Available: <https://arxiv.org/abs/2106.11342>
- [53] "Santa casa de misericórdia de são paulo," <https://www.santacasasp.org.br/>, accessed: 2022-05-01.
- [54] A. Wibisono and P. Mursanto, "Multi region-based feature connected layer (rb-fcl) of deep learning models for bone age assessment," *Journal of Big Data*, vol. 7, no. 1, pp. 1–17, 2020.
- [55] E. Ippolito, F. Postacchini, and E. Scola, "Skeletal growth in normal and pathological conditions." *Italian Journal of Orthopaedics and Traumatology*, vol. 9, no. 1, pp. 115–127, 1983.
- [56] D. Berrar, "Cross-validation." 2019.
- [57] R. Kohavi *et al.*, "A study of cross-validation and bootstrap for accuracy estimation and model selection," in *Ijcai*, vol. 14, no. 2. Montreal, Canada, 1995, pp. 1137–1145.
- [58] C. Willmott, S. Ackleson, R. Davis, J. Feddema, K. Klink, D. Legates, J. O'donnell, and C. Rowe, "Statistics for the evaluation of model performance," *Journal of Geophysical Research*, vol. 90, no. C5, pp. 8995–9005, 1985.
- [59] J. Qi, J. Du, S. M. Siniscalchi, X. Ma, and C.-H. Lee, "On mean absolute error for deep neural network based vector-to-vector regression," *IEEE Signal Processing Letters*, vol. 27, pp. 1485–1489, 2020.
- [60] S. Aja-Fernández, R. de Luis-García, M. A. Martín-Fernández, and C. Alberola-López, "A computational tw3 classifier for skeletal maturity assessment. a computing with words approach," *Journal of Biomedical Informatics*, vol. 37, no. 2, pp. 99–107, 2004.
- [61] P. Thangam and K. Thanushkodi, "Skeletal bone age assessment from epiphysis/metaphysis of phalanges using hausdorff distance," *Scientific Research and Essays*, vol. 7, no. 28, pp. 2495–2503, 2012.

## APPENDIX A

TABLE VI: Comparative table between the referenced works, with the bases, computer vision techniques, neural networks used, and the results obtained.

<b>Author</b>	<b>Dataset</b>	<b>Samples</b>	<b>Estimation Method</b>	<b>Techniques</b>	<b>CNN</b>	<b>Validation</b>	<b>Results</b>
[29]	Digital Hand Atlas	1375	TW	N/A	Inception V4	MAE	0.59
[31]	Digital Hand Atlas	1390	GP	Histogram equalization Data Augmentation	VGG	MAE	0.79
[33]	Digital Hand Atlas	1400	GP	Data Augmentation Resizing	VGG	MAE	0.76
[37]	Digital Hand Atlas	1229	GP+TW	Gaussian filter Watershed	Inception V2	MAE	0.346
[35]	RSNA Dataset	12.611	GP	Canny	VGG	MAE	1.23
[36]	RSNA Dataset	12.611	TW	N/A	Inception V2	MAE	0.38
[38]	Digital Hand Atlas	1425	TW	N/A	Yolo V5	MAE	6.53
	RSNA Dataset	12.611	TW	N/A			