Bone Age Prediction

Machine learning in medicine project

Nguyen Duc Bao Minh, Nguyen Dinh Tung, Nguyen Kien Trung, Le Quoc Trung, Ng
Tung, Do Bao Phuc

Context

Bone age assessment using X-ray images is a standard clinical procedure to detect anomaly in bone growth in human.
This is our group's presentation about the application for bone age prediction automatically using VGG19 transfer learning enhanced with attention memchanism.

Introduction

Motivation:

Bone age refers to the actual level of bone growth development, which can be assessed continuously as the skeleton bone grows, as such, it will change in shape and size over time.

The assessment is done using a non-dominant hand due to the nature of bone classification. It is vital to predict a child's age through bone X-ray images to understand if there are any health issues (such as: genetic disorders, hormonal problems, and endocrine disorders).

5 years old

15 years old

Workflow

Input

- X-ray images

1

1

Workflow

Input

- X-ray images

1

Data Preprocessing Stage

- Label Processing
- Data Split
- Class Balancing
- Image Augmentation

2

1

Workflow

Input

- X-ray images

1

Data Preprocessing Stage

- Label Processing
- Data Split
- Class Balancing
- Image Augmentation

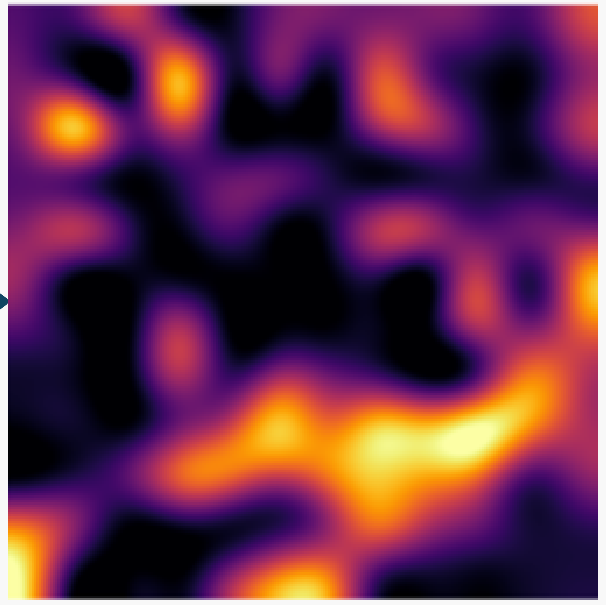
2

Prediction stage

- Spatial attention mask
- Bone Age Regression

3

1



3

Workflow

Input
- X-ray images

1

Data Preprocessing Stage
- Label Processing
- Data Split
- Class Balancing
- Image Augmentation

2

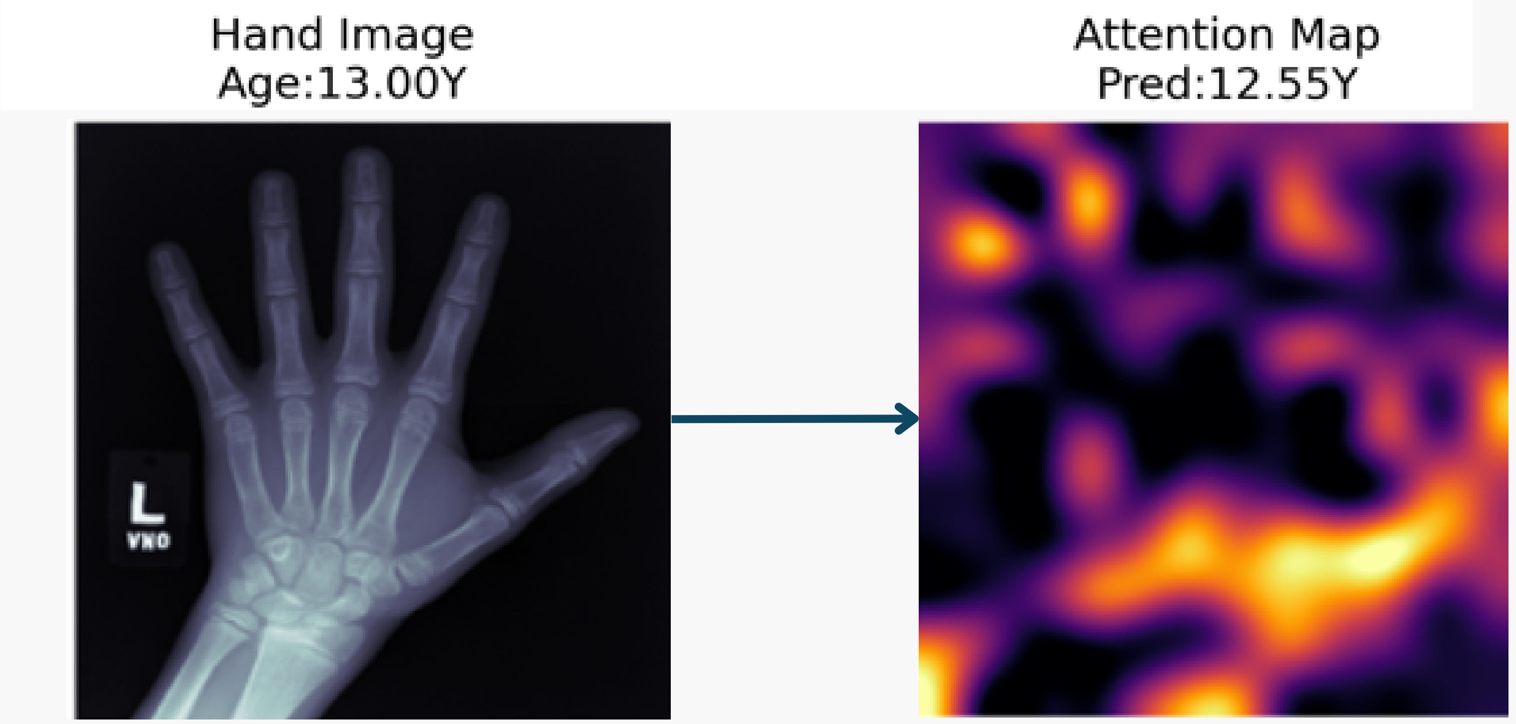
Prediction stage
- Spatial attention mask
- Bone Age Regression

3

Output
- Prediction of Bone age
- Regression performance of the model

4

4



1

3

Input

Input

Dataset:

- Obtained from the 2017 RSNA Pediatric Bone Age Challenge.
- Contained 12.611 X-ray images.
- Ground truth bone age and gender are provided.

Image information:

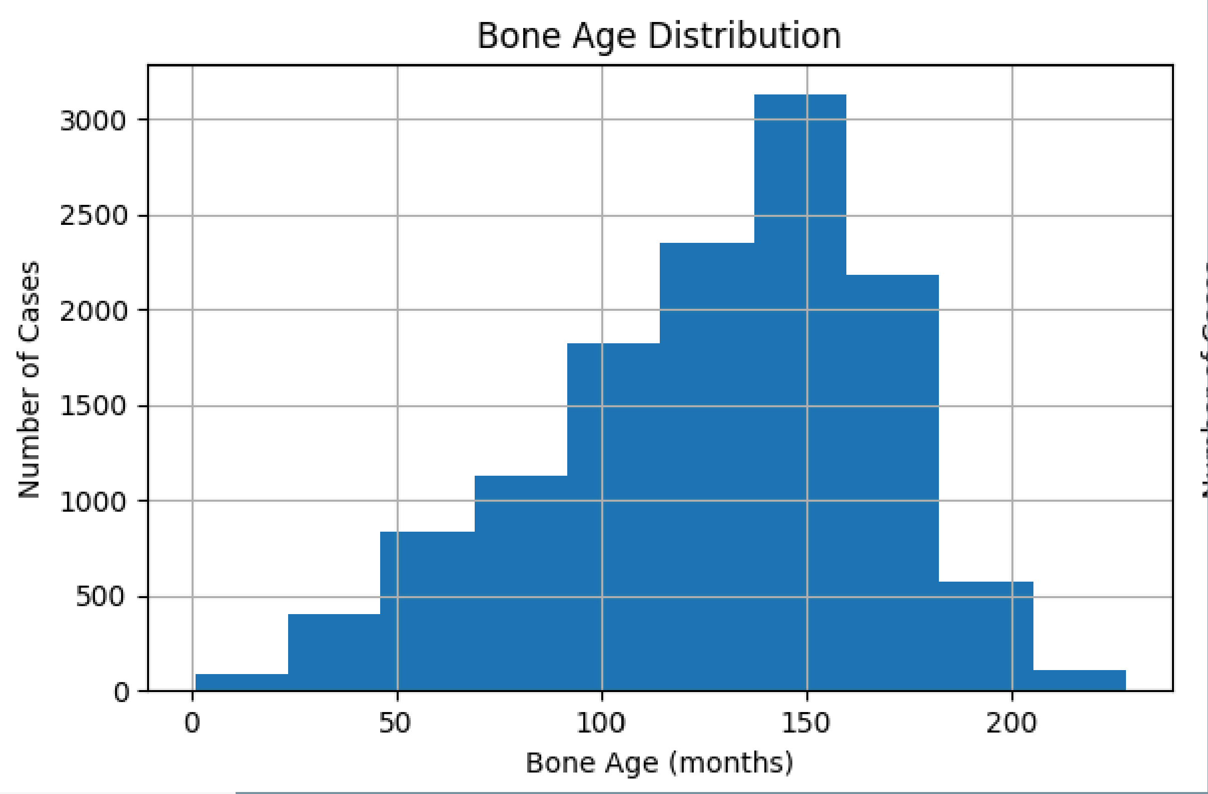
- All images are in different size.
- Channel: RGB

Data Preprocessing Stage

Data Preprocessing

Label Processing:

- We label each images with its corresponding gender.
- The bone age distribution is right-skewed, with most subject clustered around 80 months to 180 months
- There are fewer samples in the very early age and very older age.



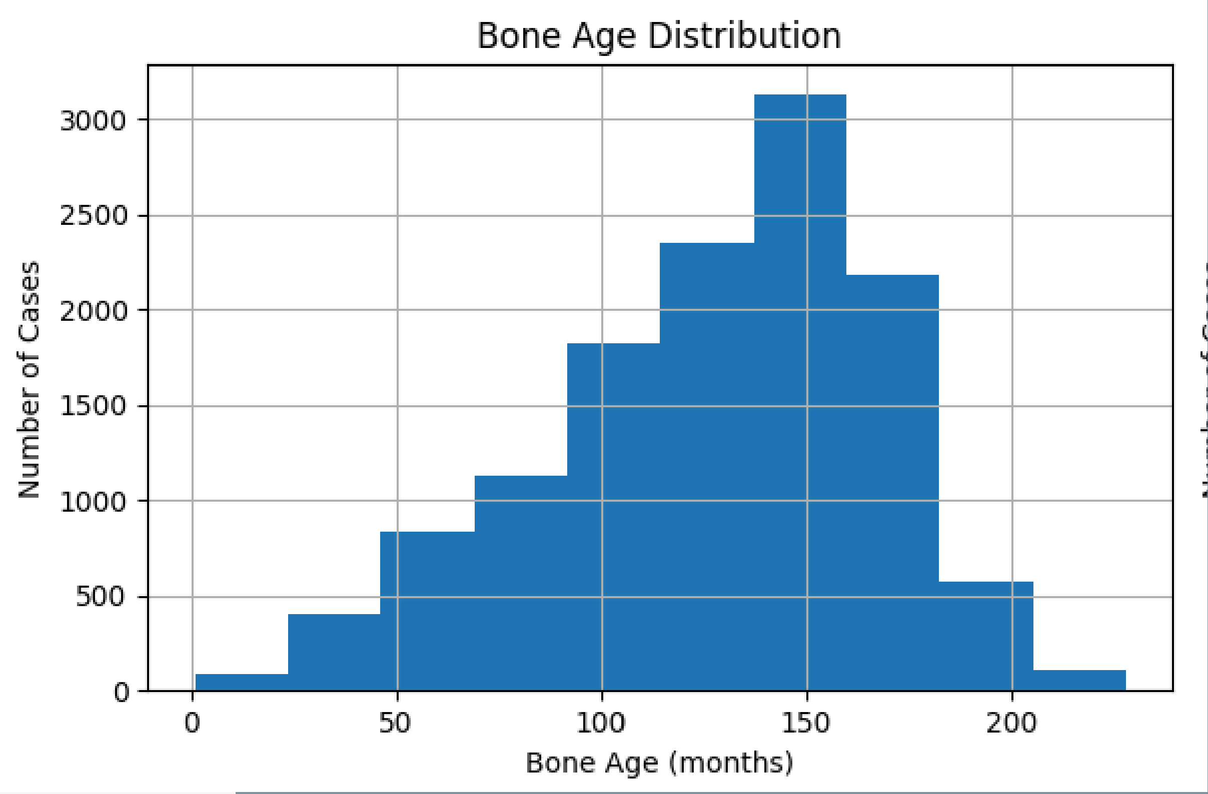
Data Preprocessing

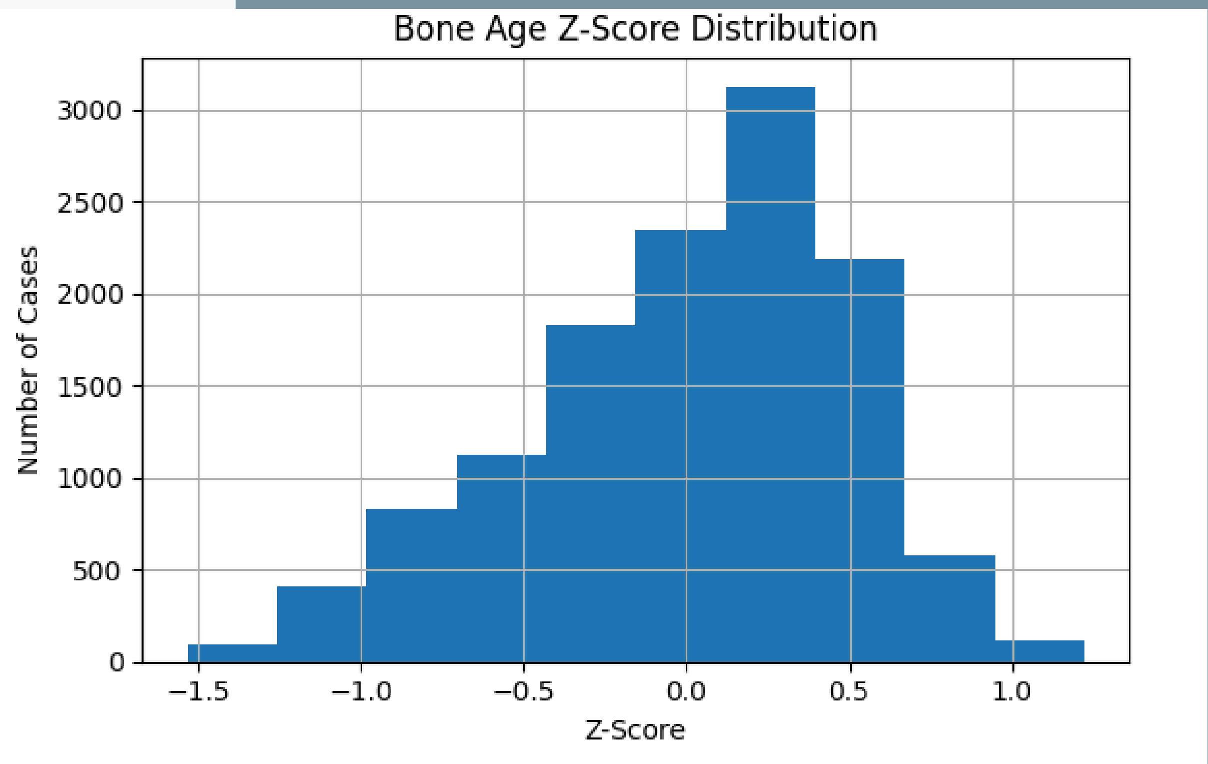
Label Processing:

- We label each images with its corresponding gender.
- The bone age distribution is right-skewed, with most subject clustered around 80 months to 180 months
- There are fewer samples in the very early age and very older age.

EDA:

- The bone age z-score distribution has symmetrics around mean (z = 0)





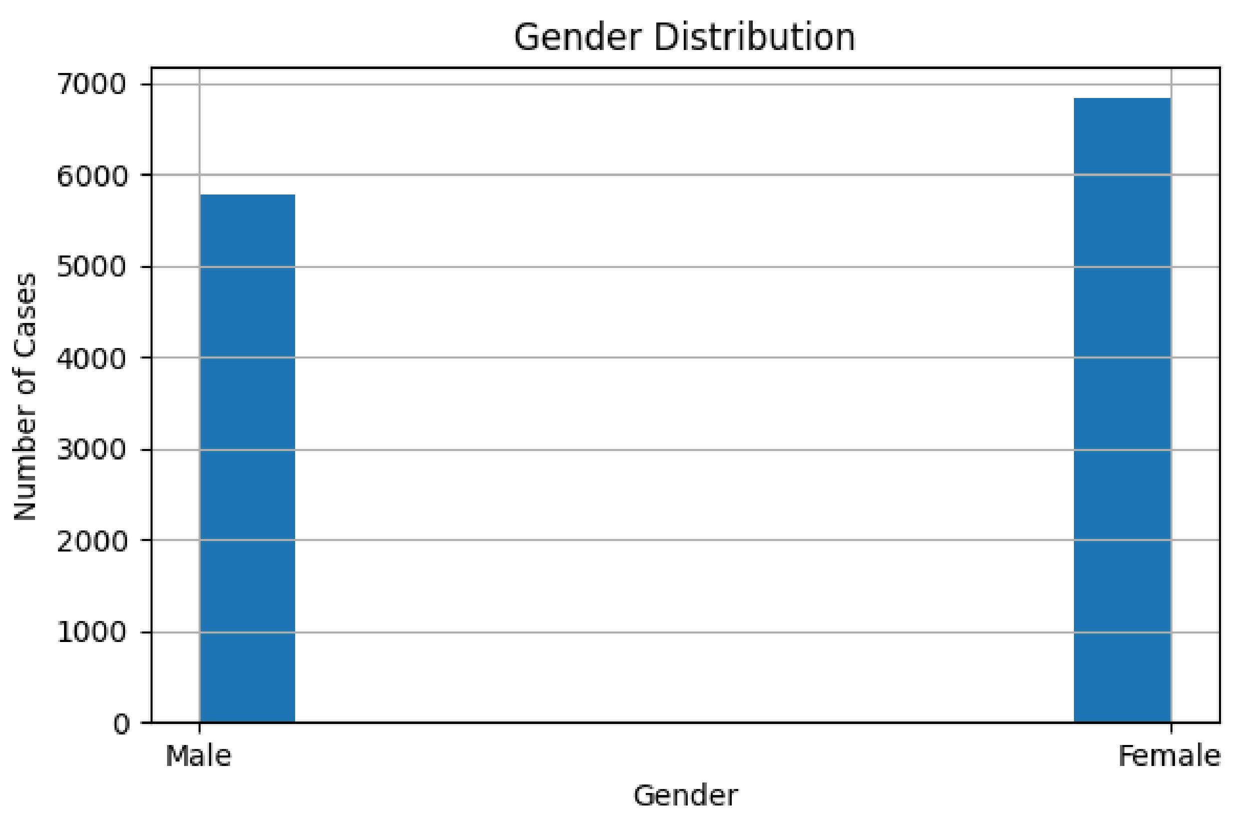
Data Preprocessing

Label Processing:

- We label each images with its corresponding gender.
- The bone age distribution is right-skewed, with most subject clustered around 80 months to 180 months
- There are fewer samples in the very early age and very older age.

EDA:

- The bone age z-score distribution has symmetrics around mean (z = 0)
- There are slightly more female subjects than male subjects.



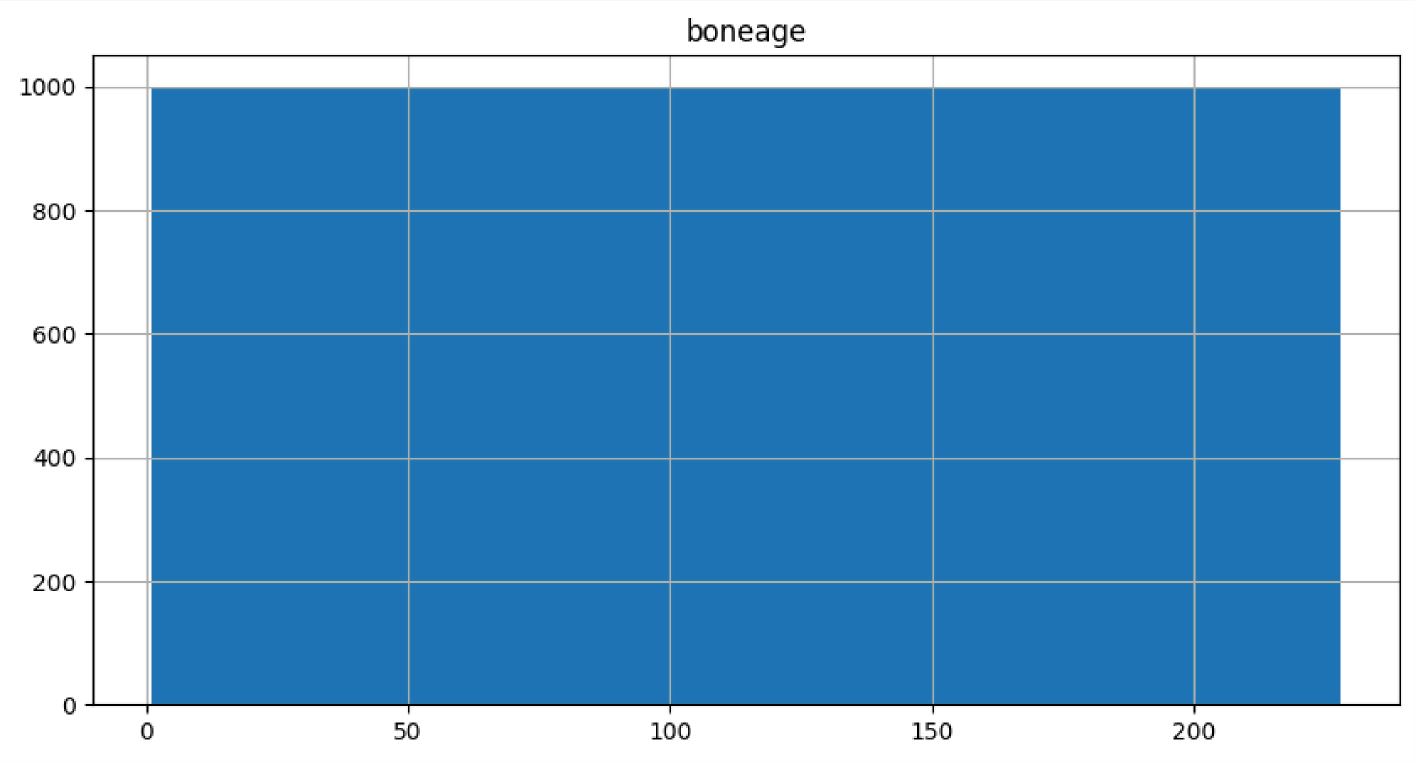
Data Preprocessing

Data split:

- The Bone age dataset is then split into 2 subset: train set and validation set.
- From the validation set, there will be 1024 random images picked to be test set.

Class balancing:

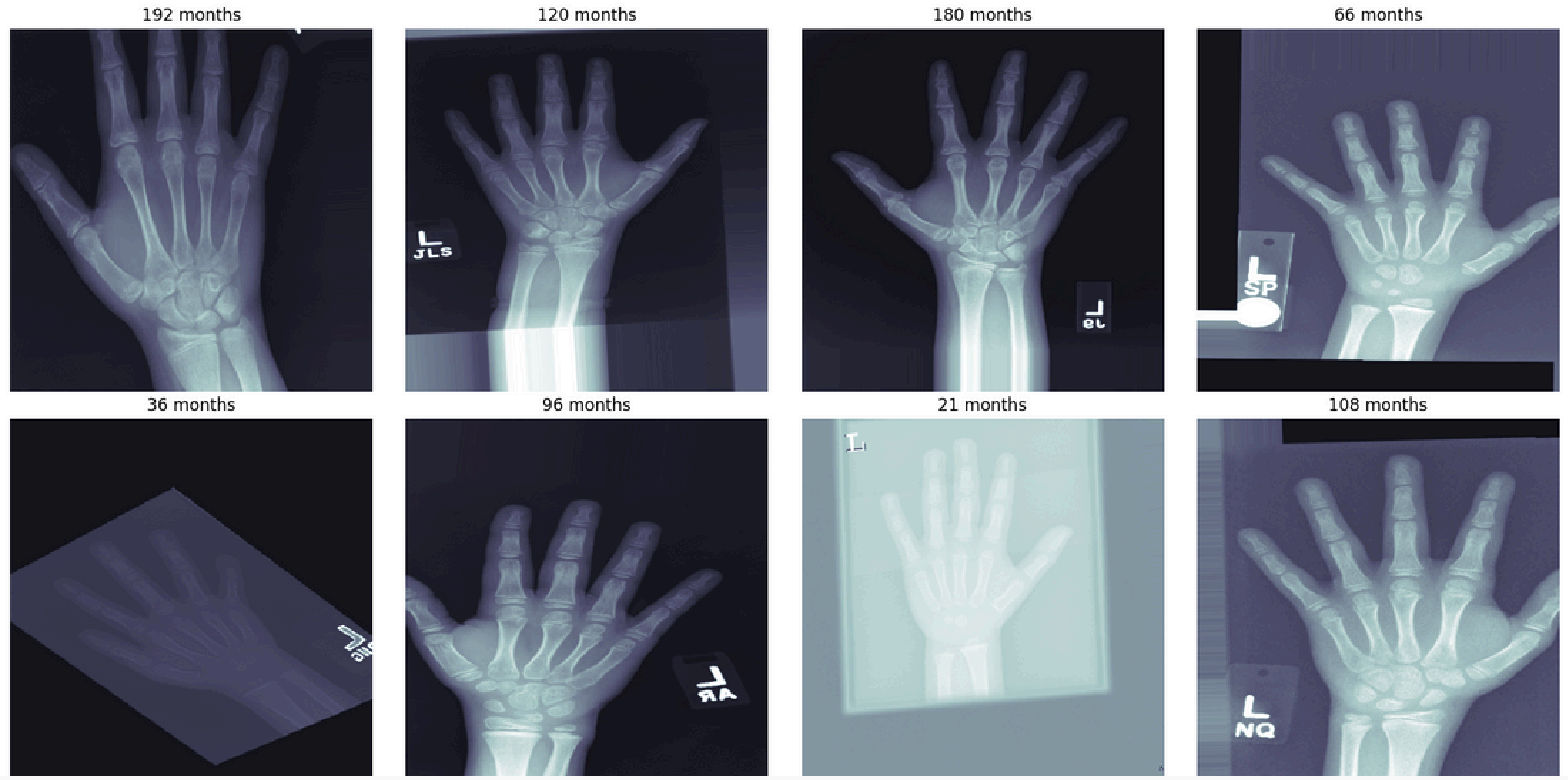
- On train set only.
- A stratified sampling strategy is applied:
  - Group the data by bone age category
  - Sampling to 500 samples each.
  - Results in the new train set, which make the male and female sample equally.



Data Preprocessing

Image Augmentation:

Augmentation Processing Step	Train	Valid	Test
Input image size	384×384		
Color channels	3 (RGB)		
Horizontal flip	√		
Vertical flip	–	–	–
Height shift (15%)	√	–	–
Width shift (15%)	√	–	–
Rotation (±5°)	√	–	–
Shear (1%)	√	–	–
Zoom (±25%)	√	–	–
Fill mode	Nearest		
Batch size	32	256	1024



Prediction Stage

Prediction

Layer (Type)	Output Shape	Param #	Connected to
input 2 (InputLayer)	(None, 384, 384, 3)	0	[]
vgg19 (Functional)	(None, 12, 12, 512)	20,024,384	input 2[0][0]
batch normalization	(None, 12, 12, 512)	2,048	vgg19[0][0]
conv2d (Conv2D)	(None, 12, 12, 64)	32,832	batch normalization[0][0]
conv2d 1 (Conv2D)	(None, 12, 12, 16)	1,040	conv2d[0][0]
locally connected2d	(None, 12, 12, 1)	2,448	conv2d 1[0][0]
conv2d 2 (Conv2D)	(None, 12, 12, 512)	512	locally connected2d[0][0]
multiply (Multiply)	(None, 12, 12, 512)	0	conv2d 2[0][0], batch_normalization[0][0]
global_average_pooling2d	(None, 512)	0	multiply[0][0]
global_average_pooling2d_1	(None, 512)	0	conv2d_2[0][0]
RescaleGAP (Lambda)	(None, 512)	0	global_average_pooling2d[0][0], global_average_pooling2d_1[0][0]
dropout (Dropout)	(None, 512)	0	RescaleGAP[0][0]
dense (Dense)	(None, 1024)	525,312	dropout[0][0]
dropout_1 (Dropout)	(None, 1024)	0	dense[0][0]
dense_1 (Dense)	(None, 1)	1,025	dropout_1[0][0]
Total params:			20,589,601
Trainable params:			563,681
Non-trainable params:			20,025,920

Model:

- We transfer learning the VGG19 model from pre-trained ImageNet weights.
- Batch normalization is applied.

Attention mechanism:

- A locally connected 2D layer is employed to activate the sigmoid function.
- It dynamically hightlight bone-relevant region in the image.

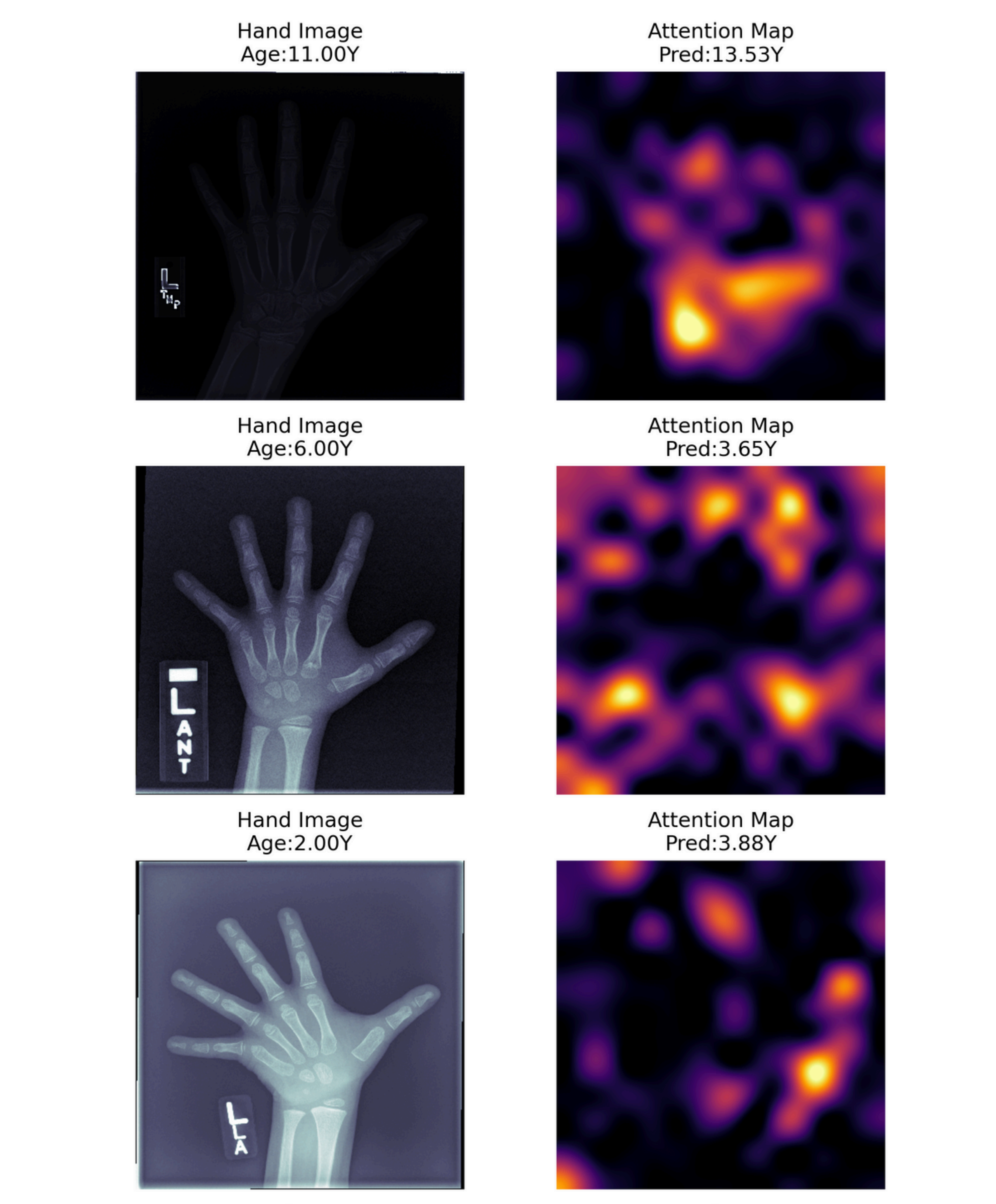
Prediction

Layer (Type)	Output Shape	Param #	Connected to
input 2 (InputLayer)	(None, 384, 384, 3)	0	[]
vgg19 (Functional)	(None, 12, 12, 512)	20,024,384	input 2[0][0]
batch normalization	(None, 12, 12, 512)	2,048	vgg19[0][0]
conv2d (Conv2D)	(None, 12, 12, 64)	32,832	batch normalization[0][0]
conv2d 1 (Conv2D)	(None, 12, 12, 16)	1,040	conv2d[0][0]
locally connected2d	(None, 12, 12, 1)	2,448	conv2d 1[0][0]
conv2d 2 (Conv2D)	(None, 12, 12, 512)	512	locally connected2d[0][0]
multiply (Multiply)	(None, 12, 12, 512)	0	conv2d 2[0][0], batch_normalization[0][0]
global_average_pooling2d	(None, 512)	0	multiply[0][0]
global_average_pooling2d_1	(None, 512)	0	conv2d_2[0][0]
RescaleGAP (Lambda)	(None, 512)	0	global_average_pooling2d[0][0], global_average_pooling2d_1[0][0]
dropout (Dropout)	(None, 512)	0	RescaleGAP[0][0]
dense (Dense)	(None, 1024)	525,312	dropout[0][0]
dropout_1 (Dropout)	(None, 1024)	0	dense[0][0]
dense_1 (Dense)	(None, 1)	1,025	dropout_1[0][0]
Total params:			20,589,601
Trainable params:			563,681
Non-trainable params:			20,025,920

Model:

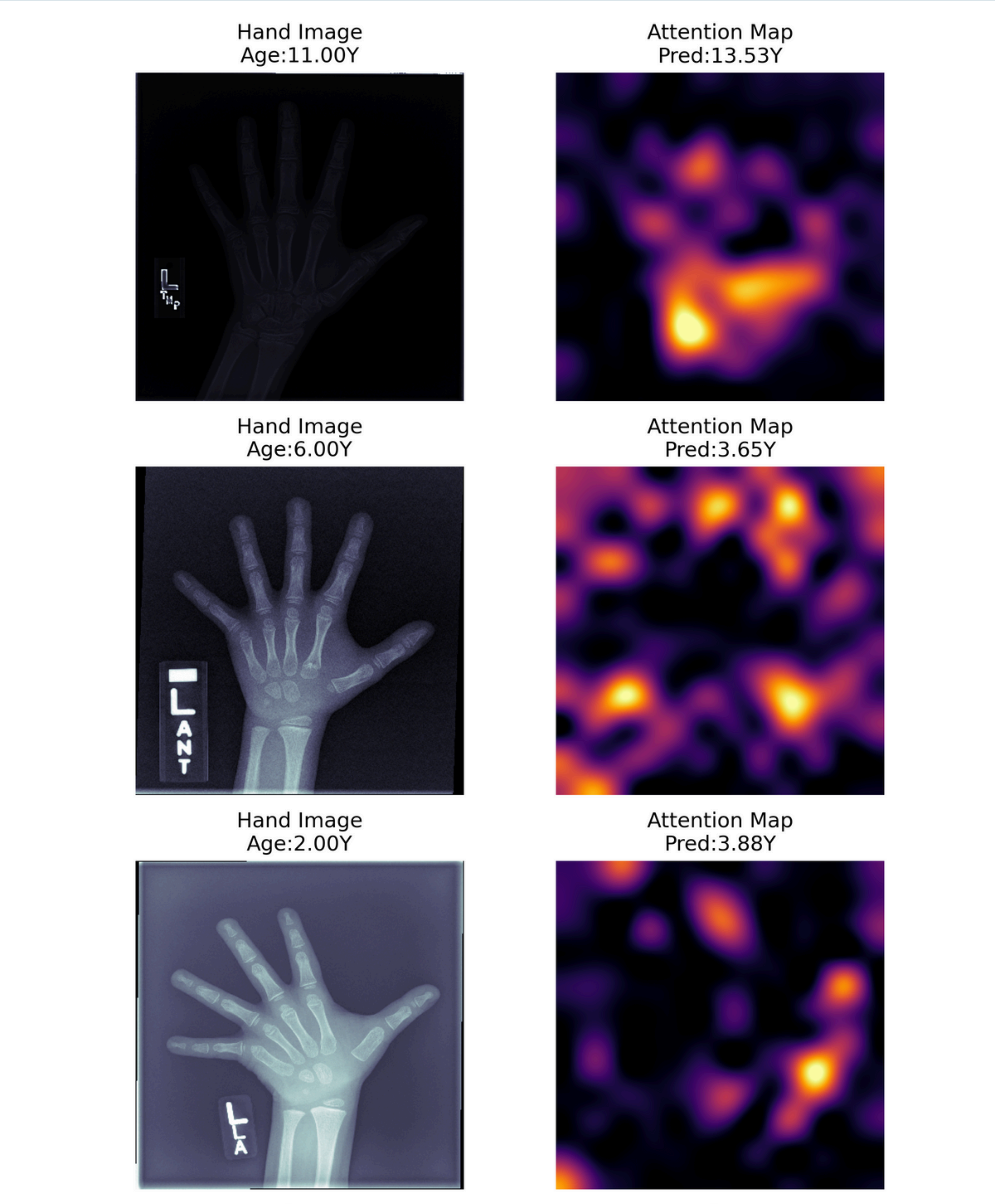
- Global average pooling (GAP) is used to condensed the features and mask .
- A regression head is used instead of default classification head of the VGG19 architecture.
- The attention and regression layers are trained from scratch.

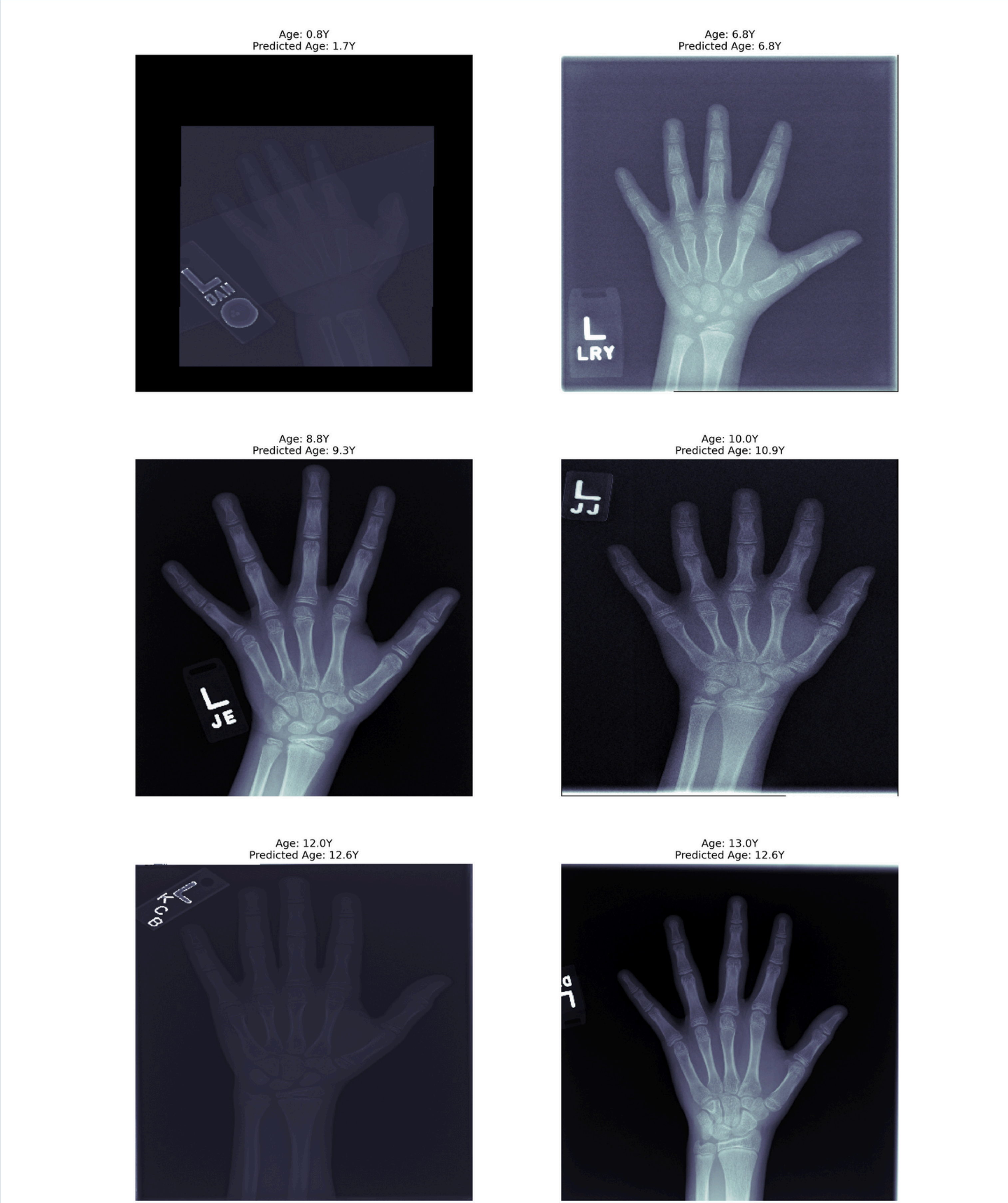
Output and Results



- The attention value for these 6 images are very low; varies from the lowest: 2.21432e-5 to the hightest: 0.04. Which is why we have to apply normalization for the attention map, adjust vmin and vmax to fit the range
- Overall, the attention mechanism sucessfully identifies relevant bone structures in older children.

Output and Results





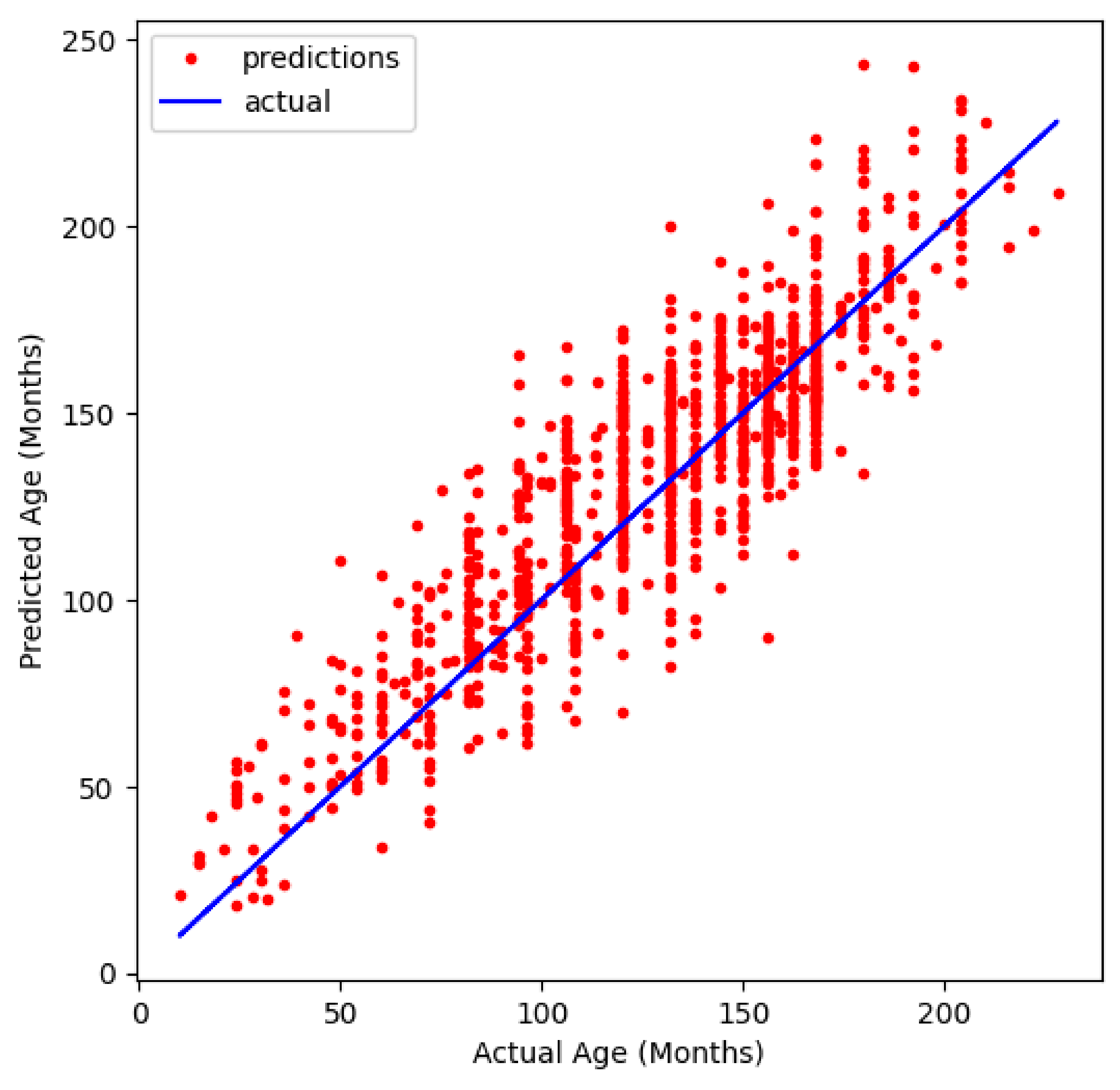
Output and Results

3 evaluation metrics used in the project are: Mean Absolute Error (MAE), Mean Absolute Deviation (MAD), and Root Mean Squared Error (R<SE). Below is the results of the output prediction on test set:

Metric	**Value**	**Units**
Mean Absolute Error (MAE)	16.2751	months
Mean Absolute Deviation (MAD)	13.5901	months
Root Mean Squared Error (RMSE)	20.6213	months

Output and Results

- As the actual age increases, the predicted age tends to increase as well.
- There are fewer data points on the lower age and older age
- The model perform equally well across the entire age range.



Output and Results

Compare to other state-of-the-art algorithm:

Method	MAE (months)	MSE (months)	MAD (months)	RMSE (months)
Proposed VGG19+Attention	16.28	425.24	13.59	20.62
AXNet	7.70	108.87	–	–
Deeplasia (SOTA)	–	–	3.87	7.67

Conclusion

We have presented our data preprocessing and prediction method for bone age prediction and compare with other studies in the same field.

Thank you