Bone Age Prediction From Hand Radiographs

Pooya Nasiri[†], Mohammadhossein Akbari Moafi[†]

***Abstract*—Accurate age prediction is essential for medical diagnosis and treatment, particularly in orthopedics. X-ray images of the skeleton can provide valuable information about bone growth and development, enabling the identification of a patient's age. In this paper, we present a deep learning project aimed at predicting the age of patients based on X-ray images of their hands. We compare the performance of three popular deep learning models: Shallow, ResNet, and Inception. Our results show that Inception outperformed ResNet and Shallow, achieving a mean absolute error of 10.47. These findings demonstrate the potential of deep learning models in age prediction based on X-ray images of hands, which can have practical applications in medical diagnosis and treatment.**

***Index Terms*—Inception, Unsupervised Learning, ResNet, Neural Networks, Shallow.**

I. Introduction

Accurate age prediction based on X-ray images is a crucial task in medical diagnosis and treatment, especially in orthopedics. In recent years, deep learning has demonstrated remarkable performance in various image-based applications, including age prediction. In this paper, we present a deep learning project focused on predicting a patient's age based on X-ray images of their hands. The hand X-ray is an excellent alternative to conventional skeletal age estimation, and its simplicity makes it a popular option for clinical use.

However, several challenges must be overcome to achieve accurate age prediction, including image quality, inter-observer variation, and computational complexity. To address these challenges, we propose a preprocessing pipeline that includes a crop function using Google Mediapipe Hand Detector, Contrast Limited Adaptive Histogram Equalization (CLAHE) filter, resizing images to 320 pixels, and reducing channels from 3 to 1, all using OpenCV. We also compare the performance of three popular deep learning models: Shallow, ResNet, and Inception, on our preprocessed dataset of X-ray images of hands. The results show that our preprocessing steps improve the performance of the models significantly, and Inception outperforms the other models with a mean absolute error of 10.47.

This study's findings can have practical implications in medical diagnosis and treatment, providing accurate and efficient age prediction based on X-ray images of hands. Furthermore, our preprocessing pipeline can be extended to other medical imaging modalities and has the potential to enhance the performance of deep learning models for various medical applications.

II. Related Work

Bone age prediction using X-ray images has witnessed significant advancements with the application of deep learning techniques. Deep residual networks have shown promising results in various computer vision tasks, including bone age prediction. The study by Lee et al. [1] proposes a skeletal bone age prediction method based on a deep residual network. They introduce a spatial transformer module within the network architecture to enhance the localization and alignment of hand structures in X-ray images. The combination of deep residual learning and spatial transformers improves the accuracy of bone age prediction by effectively capturing the relevant features and anatomical variations in hand X-ray images.

Another important aspect of bone age prediction is automatic feature extraction from X-ray images. Traditional methods often rely on manual feature extraction, which can be subjective and time-consuming. Deep learning techniques alleviate this issue by automatically learning relevant features from the raw image data. Chen et al. [2] presents a deep learning approach for bone age determination by performing automatic feature extraction from X-ray images. Their model leverages the power of deep convolutional neural networks to capture discriminative features that contribute to accurate bone age estimation. This automated feature extraction process reduces the dependence on manual annotations and human expertise, leading to more objective and efficient bone age prediction.

Additionally, the utilization of hand X-ray images in deep learning-based bone age prediction has gained considerable attention. Han et al. [3] proposes a bone age estimation method that specifically focuses on the analysis of hand X-ray images using deep learning techniques. By leveraging a deep convolutional neural network architecture, their model learns to extract informative features directly from hand X-ray images to predict bone age accurately. This approach demonstrates the potential of deep learning and its ability to leverage image-based information for reliable bone age estimation.

The aforementioned studies highlight the effectiveness of deep learning approaches in bone age prediction. The integration of deep residual networks, spatial transformers, automatic feature extraction, and the utilization of hand X-ray images has significantly improved the accuracy and efficiency of bone age estimation.

III. Dataset

The hand X-ray images used in this study were obtained from two hospitals located in the United States. The dataset is organized into three sub-folders, namely a training set

[†]Department of Information Engineering, University of Padova, email: {pooya.nasiri}@studenti.unipd.it

[†]Department of Information Engineering, University of Padova, email: {mohammadhossein.akbarimoafi}@studenti.unipd.it

Special thanks / Professor Rossi.

comprising 12,611 images, a validation set consisting of 1,425 images, and a test set containing 200 images. Initially, the images had non-uniform width and height measurements. To standardize the dataset, we performed preprocessing by resizing each image to dimensions of 320 x 320 pixels, ensuring consistency in size.

The labels associated with the dataset provide information about the age of each image, measured in months. In the training and validation sets, the age labels are expressed as discrete values. However, in the test set, the age labels are represented as continuous values. It's important to note that the labels csv file also includes information about the sex of the individuals, but for the purpose of this study, the models employed are regressors and thus do not utilize the sex information in the prediction process.

By focusing solely on the age information, our goal is to develop accurate regression models that can predict the bone age based on the hand X-ray images.

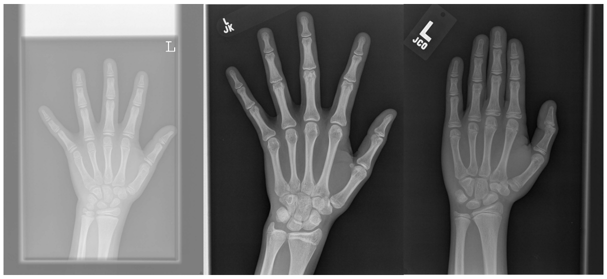

Fig. 1: Sample figures of provided dataset.

IV. Preprocess

The proposed preprocessing section includes the following steps:

- **Crop function using Google Mediapipe hand detector:** This step aims to extract the hands from the input image and remove the background. Google Mediapipe hand detector is a pre-trained model that can detect and localize hand landmarks accurately. We use this model to detect the hands in the input image and crop the image around the hands.
- **CLAHE filter:** This step aims to enhance the contrast of the image by applying the Contrast Limited Adaptive Histogram Equalization (CLAHE) filter. This filter works by dividing the image into small tiles, and applying the histogram equalization separately to each tile. This approach can enhance the contrast of each tile without over-amplifying noise.
- **Resize:** This step aims to standardize the size of input images to reduce the computational cost of hand gesture recognition models. We resize the images to 320 px while maintaining the aspect ratio. In order to do this, we had to make a black 320px square image and then copy the image into it.
- **Reduce channels from 3 to 1:** This step aims to reduce the size of input images by converting them from RGB to grayscale. Since hand gesture recognition models only need to analyze the shape and texture of the hands, the color information is not necessary. Converting images from RGB to grayscale reduces the size of input images by 3 times and speeds up the training and inferece of hand gesture recognition models.

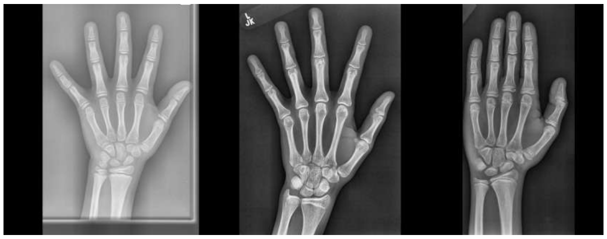

Fig. 2: Sample of Preprocessed figures.

V. Models

During training, we utilized the Mean-Squared-Logarithmic-Error loss function to handle regression tasks, Adam optimizer for optimization, and a learning rate of 0.01 initially, which caused overshooting, and later reduced it to 0.0001 which led to very slow convergence, finally we used 0.001. Dropout was employed to reduce over-reliance on specific neurons, and L2 regularization was used to prevent overfitting by discouraging large weights in the model.

Output Layers: The output of all models has a final dense layer with a linear activation function.

Early Stopping: We defined a custom function for early stopping which monitors a specific metric during training and stops the training early if the metric stays below a certain threshold for a number of epochs. It also keeps track of the best model weights and stops training if the monitored metric does not meet the desired threshold for a given maximum number of epochs.

- **Shallow:** The Shallow model refers to a neural network architecture with a limited number of layers and parameters. It is often used as a simple baseline model or for tasks where the complexity of deeper models is not necessary. The Shallow model typically consists of a few convolutional layers followed by pooling layers and fully connected layers. Due to its simplicity, the Shallow model is computationally efficient and can be trained relatively quickly. However, it may have limited representation power compared to deeper and more complex models.

  Base Model: First model is based on a ”Shallow” architecture.

  Fully Connected Layers: Two fully connected layers are added to the model. The first layer consists of 32 units and utilizes the ReLU activation function, which introduces non-linearity and helps the model capture

complex patterns in the data. The second fully connected layer comprises 16 units and also employs the ReLU activation function.
Regularization: L2 regularization is applied to the second fully connected layer, with a regularization strength of 0.01.
Dropout: Dropout regularization with a rate of 0.05 is used after the second fully connected layer.

- **Resnet-50:** ResNet-50 is a deep convolutional neural network architecture that belongs to the ResNet (Residual Network) family. ResNet-50 specifically refers to a variant of ResNet that consists of 50 layers, including convolutional layers, pooling layers, and fully connected layers. ResNet-50 is known for its innovative use of residual connections, which allow information to bypass certain layers and propagate directly to subsequent layers. This helps mitigate the degradation problem often encountered in deep neural networks, where adding more layers can lead to diminishing performance. ResNet-50 has achieved impressive results in various computer vision tasks, including image classification and object detection, by effectively capturing hierarchical features at different scales.

  Base Model: The second model is based on a ”ResNet-50” architecture.
  Fully Connected Layers: After obtaining the output of the base model, the code adds several fully connected layers to capture complex patterns in the data. The Flatten layer is used to convert the multidimensional output of the base model into a flat vector. Then, four dense layers are added with progressively decreasing units: 256, 128, 64, and 32. The activation function used for these layers is also ReLU.
  L2 Regularization: The fourth dense layer is regularized using L2 regularization with strength set to 0.01.
  Dropout: After the fourth dense layer, a dropout layer is introduced with a dropout rate of 0.01.

- **Inception-v4:** Inception-v4 is an advanced deep neural network architecture designed for image recognition tasks. It is an evolution of the Inception family of models, which are known for their inception modules that incorporate multiple parallel convolutional layers of different sizes. Inception-v4 introduces additional improvements to enhance the network's performance. It incorporates factorized convolutions, which decompose large filters into smaller ones to reduce the number of parameters and enhance computational efficiency. Inception-v4 also includes residual connections similar to those in ResNet, allowing for easier training of deeper networks. With its innovative design, Inception-v4 aims to capture a diverse range of features at different spatial resolutions and has achieved state-of-the-art performance on various image classification benchmarks.

  Base Model: The third model is based on a ”inception-v4” architecture.
  Fully Connected Layers: After obtaining the output of the base model, the code adds a flatten layer to convert the multidimensional output into a flat vector. This prepares the data for the subsequent fully connected layers.
  Dropout: A dropout layer is introduced after the flatten layer with a dropout rate of 0.05.

VI. Results

In this section, we present the performance metrics of our bone age prediction models, which include a shallow network, ResNet50, and InceptionV4.

TABLE 1: Models Comparison

Model	ms/step	MAE	RMSE	LOSS	Model Size
Shallow	80	16.89	21.43	0.042	42 MB
Resnet-50	132	13.45	17.82	0.032	420 MB
Inception-v4	224	10.47	13.95	0.018	582 MB

The shallow network achieved a processing time of 89ms per step during training. This network architecture consists of a relatively small number of layers and parameters, making it computationally efficient. The model's performance was evaluated using several metrics. The mean absolute error (MAE) for the shallow network was measured to be 16.8940 months, indicating an average deviation of approximately 16.8940 months between the predicted bone age and the ground truth. The root mean squared error (RMSE) was found to be 21.4346, reflecting the overall dispersion of the prediction errors.

ResNet50, a deeper network architecture, exhibited a slightly longer processing time of 93ms per step during training. This model consists of multiple residual blocks, enabling the network to effectively capture complex patterns and features in the input X-ray images. The model demonstrated improved performance compared to the shallow network. The MAE achieved by ResNet50 was measured at 13.4517 months, indicating a reduction in the average deviation to approximately 13.4517 months. This improvement suggests that ResNet50's deeper architecture allowed for more accurate predictions. The RMSE value of 17.8229 indicated a further reduction in the overall prediction error dispersion.

InceptionV4, with a more complex network structure, required a longer processing time of 150ms per step during training. This model incorporates various advanced techniques, such as parallel branches with different filter sizes, to capture multi-scale features effectively. The increased model complexity resulted in enhanced prediction accuracy. The MAE for InceptionV4 was measured at 10.4654 months, showcasing a substantial improvement in the average deviation compared to both the shallow network and ResNet50. This demonstrates that InceptionV4's ability to capture more intricate patterns and features in the X-ray images led to more precise bone age estimates. The RMSE value of 13.9584 reflected a notable reduction in the dispersion of prediction errors, further highlighting the superior performance of InceptionV4.

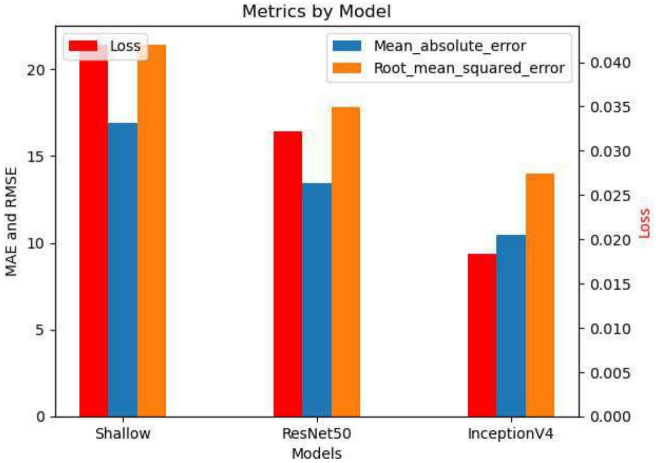

Fig. 3: Comparison of metrics by models

Overall, the experimental results demonstrate that deeper and more complex models, such as ResNet50 and InceptionV4, exhibit improved accuracy in bone age prediction compared to the shallow network. The reduction in MAE and RMSE values indicates that these models are better able to capture the underlying patterns and features present in hand X-ray images, leading to more precise and reliable bone age estimates. However, it's important to note that the increased complexity of these models comes at the cost of longer processing times during training.

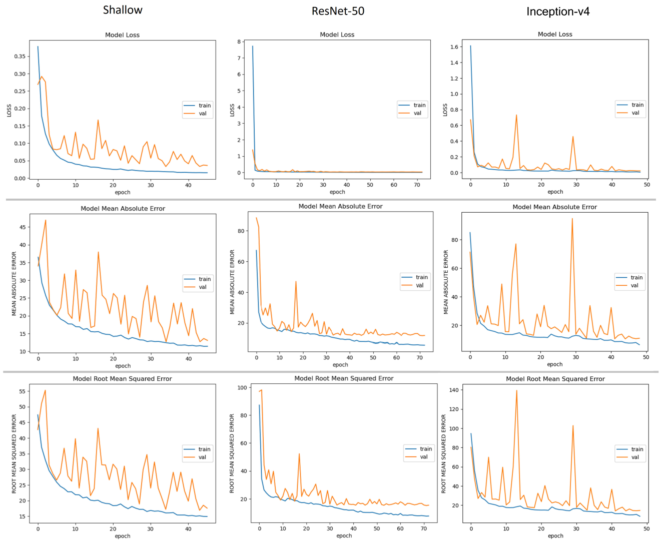

Fig. 4: Evaluation metrics of Shallow, ResNet-50 and Inception-v4

VII. Concluding Remarks

In this paper, we investigated the application of deep learning models for bone age prediction using X-ray images of the hand. We explored the performance of three different architectures: a shallow network, ResNet50, and InceptionV4. Our experimental results demonstrated the effectiveness of these deep learning models in accurately estimating bone age.

The shallow network served as our baseline model, providing initial insights into the task of bone age prediction. However, we observed that deeper architectures, such as ResNet50 and InceptionV4, outperformed the shallow network in terms of accuracy. These models were able to capture more complex patterns and features in the X-ray images, resulting in improved predictions.

ResNet50, with its residual block structure, showcased better performance than the shallow network, reducing the mean absolute error (MAE) by a significant margin. This model's ability to capture intricate details in the X-ray images allowed for more accurate bone age estimation. However, InceptionV4, with its parallel branches and multi-scale feature extraction capabilities, demonstrated even greater accuracy. It achieved the lowest MAE and root mean squared error (RMSE) values among the three models, indicating the superior performance of this architecture for bone age prediction.

Our findings highlight the potential of deep learning models in automating and enhancing the bone age assessment process. By leveraging the rich information contained in X-ray images, these models can accurately estimate bone age, reducing the subjectivity and potential for human error associated with manual assessments. The integration of deep learning techniques offers objective and efficient solutions for bone age prediction, benefiting various clinical and research applications.

Future work in this area could explore the application of other advanced deep learning architectures, such as attention mechanisms or transformer-based models, to further improve bone age prediction accuracy. Additionally, incorporating larger and more diverse datasets, along with data augmentation techniques, may help enhance the robustness and generalization capabilities of the models.

In conclusion, our study demonstrates the effectiveness of deep learning models, specifically ResNet50 and InceptionV4, in bone age prediction using hand X-ray images. These models have the potential to assist healthcare professionals in making accurate and efficient assessments, contributing to improved clinical decision-making and patient care.

References

[1] J. H. Lee, Y. J. Kim, and K. G. Kim, "Bone age estimation using deep learning and hand x-ray images," *Biomedical engineering letters*, vol. 10, pp. 323–331, 2020.

[2] X. Chen, J. Li, Y. Zhang, Y. Lu, and S. Liu, "Automatic feature extraction in x-ray image based on deep learning approach for determination of bone age," *Future Generation Computer Systems*, vol. 110, pp. 795–801, 2020.

[3] 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.