Deep Learning-Based Bone Age Growth Disease Detection (BAGDD) Using RSNA Radiographs

Muhammad Ali[1], Muhammad Faheem Mushtaq[2], Saima Noreen Khosa[3],
Naila Kiran[4], Humaira Arshad[5], Urooj Akram[6]
Faculty of Computing, The Islamia University of Bahawalpur, Bahawalpur 63100, Pakistan[1,2,4,5,6]
Institute of Computer Science, Khwaja Fareed University of Engineering and Information Technology,
Rahim Yar Khan 64200, Pakistan[3]

***Abstract*—Radiological bone age assessment is essential for diagnosing pediatric growth and developmental disorders. The conventional Greulich-Pyle Atlas, though widely used, is manual, time-intensive, and prone to inter-observer variability. While deep learning methods such as Convolutional Neural Networks (CNNs) offer automation potential, most existing models rely on transfer learning from natural image datasets and lack specialization for medical radiographs. This study aims to address the gap by developing a domain-specific, custom CNN for pediatric bone age prediction. This research proposes a customized CNN architecture trained on the RSNA pediatric bone age dataset, which includes over 12,000 annotated hand X-ray images labeled with age and gender. The pipeline incorporates pre-processing techniques such as image resizing, normalization, and Contrast Limited Adaptive Histogram Equalization (CLAHE) to enhance input quality. A YOLOv3 object detector is utilized to localize the hand region prior to model training, focusing on the most relevant anatomical structures. Unlike traditional transfer learning models such as ResNet50, VGG19, and InceptionV3, the proposed CNN is tailored for radiographic features using optimized convolutional blocks and domain-aware augmentations. This design improves generalization and reduces overfitting on small or imbalanced subsets. The proposed model achieved a Mean Absolute Error (MAE) of 3.27 months on the test set and 3.08 months on the validation set, outperforming state-of-the-art transfer learning approaches. These results demonstrate the model's potential for accurate and consistent bone age estimation and highlight its suitability for integration into clinical decision-support systems in pediatric radiology.**

Keywords*—Bone age estimation; pediatric healthcare; convolutional neural networks; transfer learning; YOLOv3; medical imaging

I. Introduction

Pediatric endocrinologists mostly rely on Bone Age Growth Disease Detection (BAGDD) to detect growth disorders and evaluate a child's bone development. For conditions such as growth hormone deficiency, precocious puberty and delayed growth, it matters a lot [1]. BAGDD is used by clinicians to understand if a child's bones are growing at the appropriate pace for their age and to predict what might happen in their growth later on. Typically, an X-ray image of the left hand and wrist is used to view the bones and changes such as ossification, that reveal the growth status of a child.

Bone age was initially measured using two traditional methods: the Greulich-Pyle (GP) Atlas [2] and the Tanner-Whitehouse (TW) approach [3]. A GP Atlas allows a trained radiologist to use the worksheet to compare a patient's bone density to those of reference images based on both gender and age. Rather than TW, the TW method assigns a score to every bone depending on its stage of development and the total gives an estimate for bone age. Despite being around for decades, these methods are held back by expert interpretations, slow procedures and the influence of personal opinions [4].

Automating bone age assessment is becoming possible thanks to the growth in AI and its specialty, Deep Learning. CNNs have helped smart systems reach the same performance in image analysis as the human eye in recent times. CNNs are particularly useful for image recognition since they can automatically pick out patterns such as edges, textures and shapes at different parts of an image [5]. Tickmarks depends on how images are used; they can be used for labeling or predicting values in both types of tasks. CNNs have been used successfully in many medical image applications, including finding tumors, analyzing retinas and, not long ago, evaluating bone age [6]. In 2017, the Radiological Society of North America made available the RSNA Bone Age Dataset which consists of over 12,000 X-rays of children's hands each labeled with bone age information and evenly divided by gender. The information in this dataset aids in the development and assessment of algorithms designed for automatic bone age estimation [7].

Such models are pre-set with knowledge from millions of pictures found in ImageNet of everyday objects like cats, cars and buildings. These models are mainly used for object classification because their learned features highlight important differences in objects' shapes, colors and textures. Still, medical images, including those of hands for assessing bone age, have very different details regarding bone density, shape and growth as compared to regular X-rays [8]. The results of further studies have suggested that development of bone age is not the same in males and females and males generally mature slower than females do [9]. Because of this, transfer learning models may not fit the unique gender differences in how bones develop. Thus, these models do not always capture the important information related to each domain and usually perform poorly on these imaging operations [10].

This work introduces a new CNN model tailored for computing the bone age from radiographs of hands. Everything is learned directly by the model from scratch, so no information is transferred from different bone-related picture datasets and the model is trained explicitly for predicting bone age. The goal is to improve the certainty and precision of bone age determination by using the CNN design to meet the requirements of this medical imaging task. The proposed custom CNN model performed well, with the MAE being 3.27 months in the training set and 3.08 months on the validation set. These results significantly outperform the transfer learning models evaluated in this work, including ResNet50, VGG19, and InceptionV3. The custom CNN proves to be effective in capturing fine-grained features in bone development, allowing for a more accurate age estimation. These findings support the conclusion that domain-specific deep learning models can substantially outperform generic transfer learning approaches when applied to specialized medical imaging tasks such as bone age assessment [9].

Whereas deep learning algorithms, especially the ones that apply the methodology of transfer learning, such as ResNet50, VGG19, and InceptionV3 models have demonstrated good potential when it comes to assessing the age of the bones, the utilized models have been predominantly trained on the datasets of natural images (e.g., ImageNet) and have not been optimized to work with the peculiarities of medical radiographs. They do not routinely produce good imaging of the subtle anatomical markings characteristic of X-ray pediatric hands required to make precise assessment of the bone-age. In addition to that, none of the current models often provide the combination of the domain-specific architecture design alongside specific preprocessing (e.g., the CLAHE-enhanced contrast and the ROI extraction using the YOLOv3) to form a unified framework applicable to the setting of the pediatric bone age growth disease detection.

The novelty of this work is the modification of the CNN model into a custom design, which trained on the scratch available RSNA Bone Age Dataset that consists of data on bone development provided specifically to be used in predicting an age. In contrast to transfer learning methods devised over generically, the proposed CNN includes the task-specific architectural components, including depth of layers and dropout regularization, along with the paradigm-specific preprocessing of medical images. This is more accurate, more robust and better aligned to the domain compared to current solutions in bone age estimation.

This research focuses on the architectural design, training, and performance evaluation of a custom CNN model specifically developed to predict bone age from hand radiographs. The contributions of this research are as follows:

- A custom Convolutional Neural Network (CNN) is developed and trained from scratch, incorporating task oriented design choices such as optimal layer depth, kernel sizes, ReLU activations, and dropout regularization. Unlike generic models, custom CNN was tailored to capture fine-grained anatomical features (e.g. bone edges, growth plates) specific to pediatric hand radiographs, resulting in improved precision and robustness.
- The performance of proposed model is significantly boosted by incorporating Contrast Limited Adaptive Histogram Equalization (CLAHE) to enhance image contrast and Region of Interest (ROI) detection using YOLOv3 to focus the learning on relevant hand structures.
- The proposed custom CNN was extensively evaluated using Mean Absolute Error (MAE) and Mean Squared Error (MSE), and benchmarked against popular pre-trained models including ResNet50, VGG19, and InceptionV3.
- The proposed model was optimized through manual adjustment of the hyperparameters, including the learning rate, batch size, and number of training epochs. In addition, the architecture was refined by adjusting the depth and number of filters that incorporate dropout layers to mitigate overfitting.

The organization of the paper is as follows. Section 2 reviews existing work relevant to the proposed study. Section 3 discusses materials and methods and outlines the development procedures of the proposed BAGDD system. Section 4 presents the results and the discussion, while Section 5 concludes the study.

II. Related Work

Deep learning has played a revolutionary role in medical imaging, particularly in the diagnosis of musculoskeletal disorders. The effectiveness of deep learning applications in bone age prediction has been demonstrated in various contexts, including fracture identification, osteoarthritis severity evaluation, and pediatric assessments. Recent research highlights its ability to detect pathological abnormalities in CT and MRI scans, such as infections, fractures, joint degeneration, internal derangements, and metastases. However, challenges remain due to the complexity of CT and MRI datasets, which often involve varying tissue contrasts. For widespread clinical adoption, further technological advancements and extensive, multi-institutional prospective studies are required to ensure the generalizability, consistency, and reliability of these techniques [11].

New methods for finding and handling different medical problems have been made possible by using CNNs. Related to Bone Age Assessment using hand and wrist radiographs, CNNs have consistently performed better than other common approaches. This section deals with applying deep learning principles for medical image analysis, paying special attention to CNN designs and transfer learning created for BAA [12].

CNNs are very important in medical image analysis, with a key application in bone age estimation. CNNs built from the ground up are created to handle the issues that come with using medical images. During Bone Age Assessment (BAA), one must make models that are skilled in finding growth plates and different bone parts so the process is accurate. Many studies have shown that custom CNN models usually outperform methods such as the Greulich-Pyle atlas or Tanner-Whitehouse standards.

Specifically, a custom CNN for estimating bone age from hand radiographs designed by [13] makes use of different layers to grab both general and detailed hand and wrist features, improving the mean error over methods used before. In a similar way, [14] recommended an automatic deep learning process to make the process more efficient and less dependant on human opinion. In their method, they chose key points in images of 3,000 X-rays, working on the carpal, metacarpal and phalangeal bones. Intensity was adjusted using preprocessing and different models were made for each gender since each follows a different pattern of growth. Unlike various other similar studies, this one looked at infants as part of the full pediatric population that was considered. Using deep learning, it was possible to create a model with an MAE of 8.890 months.

An improved CNN was built in [15] to be used with the RSNA Bone Age dataset. Thanks to deep residual learning, the model had optimized gradient flow and avoided the problem of overfitting in the larger layer. Using this kind of architecture, models successfully pointed out main anatomical landmarks for an MAE of 4.2 months. The team discovered that CNNs designed to identify certain anatomical features did better than general architectures in predicting CT results.

Still, building a custom CNN architecture is complicated and takes a long time. As described in [16], making custom CNNs for medical imaging tasks draws on expertise in both radiology and deep learning which presents a problem for both small healthcare centers and seasoned developers. All the same, the advantages of custom CNNs in medical areas make them a good alternative because they can accomplish important tasks more successfully than standard approaches, for example bone age estimation [17].

Researchers have recently shown that transfer learning can successfully estimate bone age. The results presented in [18] indicate that a fine-tuned EfficientNet model had an MAE of 4.5 months when used on the RSNA Bone Age dataset. Training the model with this approach was much faster than using a custom CNN model with similar results. The study says that EfficientNet's design means it functions well for medical imaging, especially in places where resources are limited.

The authors of [19] used transfer learning by training a DenseNet on images for recognizing bone age. The authors managed to train the model quickly and accurately by freezing the early layers and adding extra training to the last ones. By using transfer learning on the RSNA Bone Age dataset, the model showed an MAE of 14.9 months, showing that transfer learning can be equally successful, while having a more variable precision [20].

In another study, [21] evaluated a commercial deep learning based bone age assessment tool (BoneAge, Vuno, Korea) using left hand X-ray images of 371 healthy Korean children and adolescents aged 4–17 years. While the estimated bone age was highly correlated with the actual chronological age ($r = 0.96$), the overall agreement rate was only 58.8%, and the model tended to overpredict bone age in younger children, particularly those under 8.3 years of age.

The work presented by [22] focuses on evaluating age and sex from computed tomography (CT) images of vertebrae using computer aided diagnosis. The study utilized a dataset comprising CT scans from 166 patients of varying genders. All images were rescaled before feature extraction using the qMaZda software. The researchers integrated conventional machine learning algorithms—both regression and classification with deep convolutional networks. Using the BART algorithm, the bone age regression model achieved a Mean Absolute Error (MAE) of 3.14 years. Gender classification accuracy was 69% with machine learning models and 59% with deep learning models [23].

In a related study, [24] proposed a regression based, multi-model deep learning framework for bone age assessment using hand radiographic images and clinical data (e.g., gender and chronological age). The approach combined EfficientNetV2S CNNs for image analysis with a simple deep neural network (DNN) for processing clinical data, resulting in robust diagnostic performance. This study underscored the advantages of multi-modal learning in improving the precision of bone age analysis [25].

To address the limitations of traditional methods such as Greulich–Pyle and Tanner–Whitehouse, which heavily depend on domain expertise, [26] developed a deep learning model for osteoporosis classification and bone density prediction using opportunistic CT scans. The model was validated across datasets from multiple hospitals and CT scanners. It employed the VB-Net module for segmentation and DenseNet for classification and regression. The model demonstrated high accuracy and strong correlation with Quantitative Computed Tomography (QCT) reference standards, achieving AUC scores of 0.999 (training set), 0.970 (test set), and 0.933 (independent set). The bone density prediction model also performed well, showing potential as a cost effective, low-radiation solution for osteoporosis screening [27], [28].

In [29] and [30] review existing Bone Age Assessment (BAA) methodologies, emphasizing the use of deep learning algorithms to address the inconsistencies of traditional approaches. Although deep neural networks have demonstrated high effectiveness in classifying bone age within specific age intervals, their complexity and reliance on numerous parameters for each region of interest (ROI) present challenges.

In [31] introduced a novel approach known as the Dual Attention Dual Path Network (DADPN). Experimental validation was performed using the RSNA Pediatric Bone Age Challenge dataset, comparing DADPN with nine other popular BAA methods. The results showed that DADPN achieved the highest accuracy, with a Mean Absolute Error (MAE) of 4.76 months.

In a separate study, [32] and [33] proposed a continuous radiological age assessment method based on clavicle ossification in CT scans. Their deep learning model was trained on 4,400 scans from 1,935 patients with a mean age of 24.2 years and evaluated on a test set of 300 scans from 300 patients. The model achieved an MAE of 1.65 years. Absolute error values ranged from 6.40 years for females and 7.32 years for males, with comparison to human reader estimates yielding errors of 3.40 years for females and 3.78 years for males, resulting in a comparative MAE of 1.84 years [34].

Although these improvements exist, additional scientist work is necessary to improve the accuracy and dependability of BAA methods. A number of CNN and transfer learning frameworks struggle to perform the same on diverse data which lowers their stability. Further work is needed to decrease the errors in bone age predictions and maintain good performance in every healthcare setting.

III. Materials and Methods

The structure of this study's research framework emphasizes building an effective auto bone age prediction method using deep learning. There are five main phases in this framework: collecting data, checking the data, preprocessing data, making features, training the model and assessing the model. All the data used in this work is taken from the RSNA Bone Age Dataset which consists of about 12,000 hand X-rays. The variety of bone growth patterns in this dataset gives it a good fit for using and assessing deep learning models.

The aim of the research is to create a new CNN model that can identify and classify bones primarily based on age. To review performance, the proposed model is compared with three commonly used transfer learning models: ResNet50, VGG19 and InceptionV3. The framework is designed to highlight how well the custom CNN predicts bone age, while transfer learning models help understand the comparison.

First, several stage were taken to boost the quality of the images before supplying them to the CNN models. First, all the images were sized to 224 × 224 pixels which is a typical input measurement for most deep learning models, including the ones employed in this study. Stretching each image to the same size reduces the amount of processing required and assures even results during training. Each value was then normalized by putting it through (255/2) and rescaled to values between 0 and 1 to support more efficient training. Besides, contrast in the images was improved by applying Contrast Limited Adaptive Histogram Equalization (CLAHE) only to the bone structures. Using this method, important anatomical features are made more visible, so the CNN can recognize and understand them better during fixing.

In order to enhance the performance of the model in more situations, this study used various data enhancement techniques. Examples of applied image transformations were rotation, flipping and shifting which introduced little changes in this work. Data augmentation boosts the size of the dataset and has the side benefit of preventing overfitting due to the extra input data. So, the model can withstand variations in real situations better. In addition, ROI detection was completed with a YOLOv3 model. At this point, the developer removed information behind the hand, helping the model concentrate on what it should focus on and therefore improving the training.

Following the preprocessing and feature extraction stages, the next phase involved model development. The CNN architecture was customized specifically for the bone age prediction task. The model consists of multiple convolutional layers that progressively learn features from low-level edges and textures to more abstract patterns crucial for bone age estimation. Since the model was trained from scratch, it was able to focus exclusively on learning features relevant to bone development, as opposed to transfer learning models that rely on representations learned from general purpose natural image datasets. This task specific design allowed the custom CNN to achieve better performance by concentrating on features uniquely important for pediatric hand X-ray analysis.

In addition to the custom CNN, this study also employed pre-trained models using transfer learning, specifically ResNet50, VGG19, and InceptionV3. These models were originally trained on the large scale ImageNet dataset and feature architectures optimized for general image recognition tasks. For the purpose of this research, they were adapted to perform bone age estimation, which is framed as a regression problem. Fine-tuning allowed these models to leverage previously learned features while adjusting to the specific domain of pediatric hand X-ray images, ensuring better performance without complete retraining.

The final component of the research framework involves model evaluation. Both the custom CNN and the transfer learning models were assessed using two key performance metrics: Mean Absolute Error (MAE) and Mean Squared Error (MSE). These metrics are well suited for regression tasks such as bone age prediction, as they quantify the deviation between predicted and actual bone ages. MAE, expressed in months, was chosen as the primary evaluation metric due to its straightforward interpretability in clinical contexts lower values indicate higher model accuracy. MSE was also considered, as it squares the error terms and therefore assigns greater penalty to larger prediction errors.

The complete research framework including preprocessing steps, the custom CNN architecture, transfer learning models, and evaluation metrics ensures a systematic and robust approach to designing and testing the bone age prediction system. The overall methodology of the proposed BAGDD system is illustrated in Figure 1.

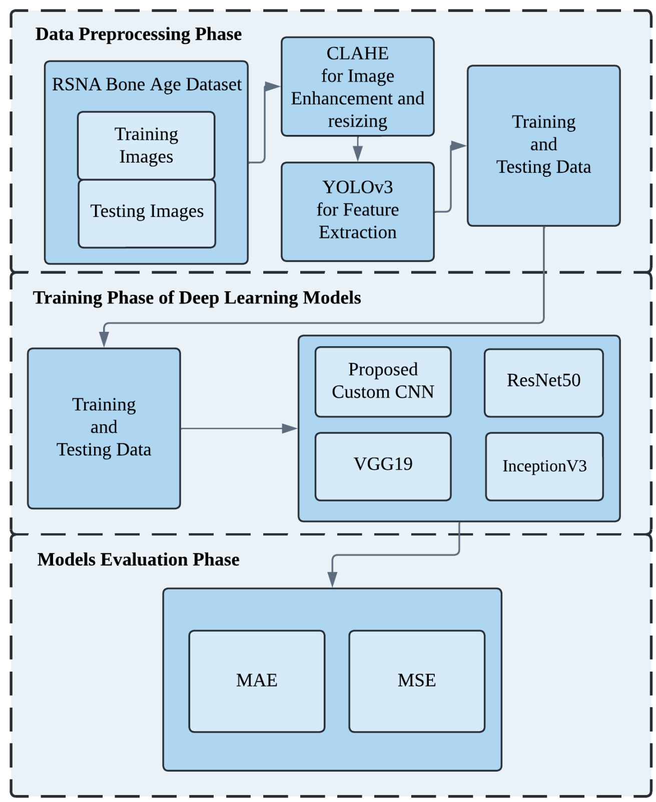

Fig. 1. The Proposed methodology of the presented BAGDD system based on deep learning.

A. RSNA Bone Age Dataset Presentation

The RSNA Bone Age Dataset is one of the most significant resources for pediatric radiology and artificial intelligence research. Provided by the Radiological Society of North America for research purposes, the dataset includes 12,611 right hand X-ray images of children, each labeled with bone age using the Greulich-Pyle (GP) method a widely accepted standard in clinical practice for bone age assessment. The dataset spans a broad age range, covering children from 0 to 18 years old, and offers comprehensive variability in growth patterns and bone development.

Because there are people of all ages, researchers can analyze both common and unusual bone growth which is essential for spotting diseases and growth issues in children. How the RSNA Bone Age Dataset is divided by age and gender is shown in Figure 2.

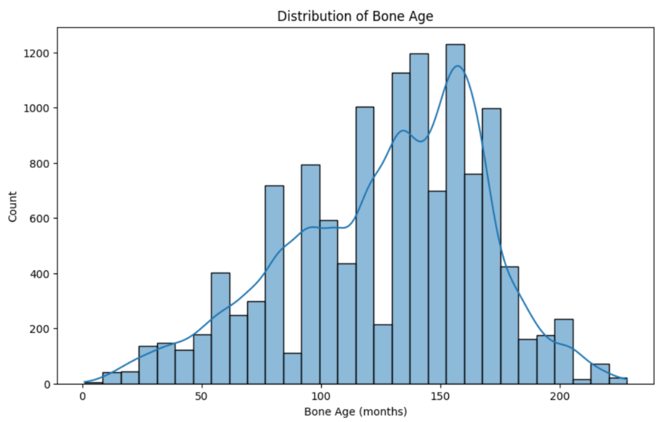

Fig. 2. Representation of the RSNA bone age dataset.

The important aspect of the dataset comes from the quality, its size and how it is relevant to medical care. All of the subjects have their pictures labeled with the bone age used to represent when the X-ray was done. They use the GP technique, a system that compares bone development with pictures in an atlas. With this relabeling, the data is truly representative of what doctors do, ideal for helping machines determine bone age.

Additionally, the dataset includes Many cases, showing how various health issues can slow bone development in kids. Having a diverse set of examples in the data ensures it's useful for training AI, by providing enough material for in-depth learning. Figure 3 shows the makeup of the dataset divided by gender and Figure 4 by age group. You can see the RSNA Bone Age dataset in Figure 5.

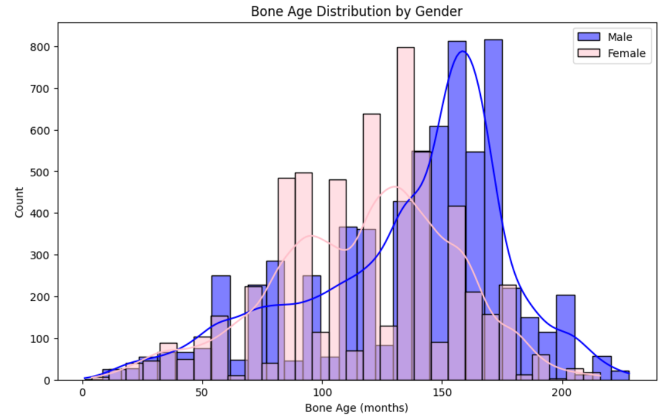

Fig. 3. Representation of the RSNA bone age distribution by gender.

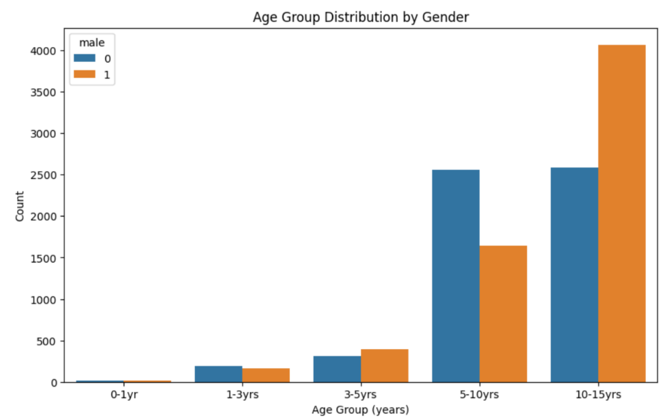

Fig. 4. Representation of the RSNA bone age distribution by age group.

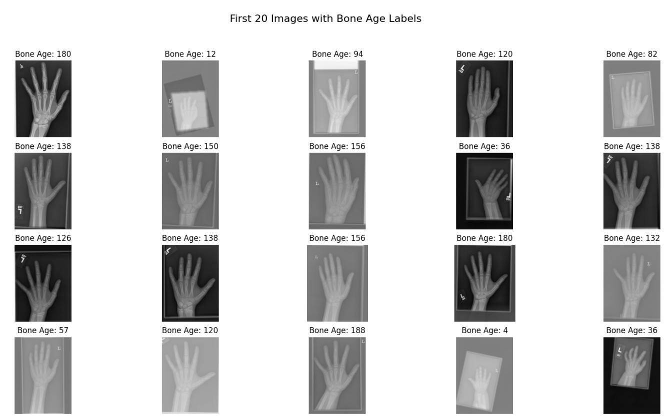

Fig. 5. Pictorial representation of the RSNA bone age dataset.

B. Preprocessing

Boosting the efficiency of predictive outcomes from deep learning models depends largely on careful data preprocessing. Certain improvement techniques were added to the X-ray images to enhance their usability:

1) Resizing: All X-ray images were adjusted to be 224 pixels wide and 224 pixels high. Following this standard lets Convolutional Neural Networks (CNNs) accept exactly these input images, since they require images of the same size. Being resized, the images fit better into the training process and avoid dialing them back in size because they come from different sources. It allows use of the same feature extraction process for all the data.

2) Normalization: All pixel intensity values in the images were scaled so that they fell between 0 and 1. To learn better, the model needs input data that is standard across points. When lighting changes and contrast in the original images are reduced, normalization allows the training process to be both stable and faster. This also helps avoid problems like gradient explosion or disappearance which both damage the model's performance.

3) CLAHE (Contrast Limited Adaptive Histogram Equalization): Since the regions of interest in X-ray images such as bones and growth plates often exhibit low contrast, CLAHE was applied to enhance their visibility. This technique enhances image contrast locally, allowing critical anatomical structures to stand out more clearly and be more easily identified by the model. Enhancing these features is essential for accurate bone age prediction, as it highlights subtle differences associated with bone growth and development. The use of CLAHE significantly improves the model's ability to detect important features that might otherwise be overlooked due to low contrast in the raw X-ray images. Figure 6 illustrates the visual improvement in an X-ray image after applying CLAHE.

Fig. 6. Before and After applying CLAHE.

4) Data augmentation: To increase the effective size of the dataset and reduce the risk of model overfitting, several data augmentation techniques were employed. These included random image rotations (0–15 degrees), horizontal and vertical flipping, shifting, and adjustments to image brightness and contrast. These transformations introduce variability into the training data, enabling the model to generalize better to unseen samples. This approach supports the third objective of the project by enhancing the model's robustness and adaptability—qualities that are particularly crucial in medical imaging, where anatomical structures can vary widely across patients. Data augmentation helps the model retain relevant features despite spatial or visual alterations, improving its performance in real world clinical scenarios.

5) ROI Detection: To further enhance model performance, a pre-trained YOLOv3 model was used for Region of Interest (ROI) detection in the X-ray images. In this context, the ROI specifically refers to the hand, which is the key anatomical region for bone age prediction. YOLOv3 enables semi automatic segmentation of the hand, allowing the CNN to focus solely on the most relevant structures. This targeted approach eliminates irrelevant background noise, helping the model avoid distractions and improving prediction accuracy. By reducing the input to only the essential region, ROI detection also contributes to computational efficiency and model precision. Figure 7 illustrates the application of YOLOv3 on an image enhanced with CLAHE.

6) Image quality assessment: In addition to the preprocessing steps outlined above, image quality assessment was also considered. This involves evaluating the clarity, contrast, and overall quality of each X-ray image to ensure that only the highest quality images are used for training and validation. Since image quality directly affects model performance, filtering for optimal image clarity enhances the effectiveness of feature extraction and improves the overall learning process.

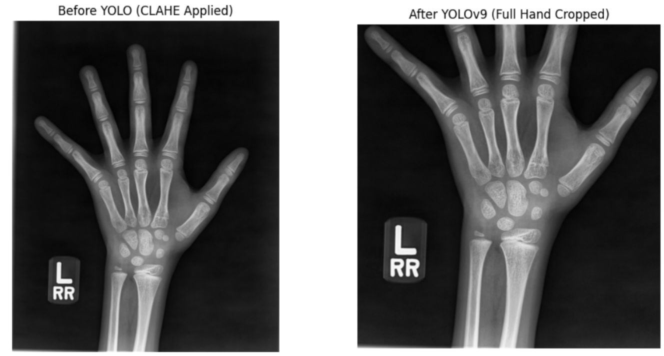

Fig. 7. Before and After Applying YOLOv3 on CLAHE Output.

Application of these methods allowed for better data preparation in the RSNA Bone Age Dataset for model training. Following these steps strengthens the model's understanding of what is required for correct prediction of bone age. As well as getting the data ready for machine learning, the preprocessing pipeline helps to make the model steady and more trustworthy which helps increase its use in pediatric radiology. By following this approach, future AI studies intended for bone age estimation are easily supported.

C. Convolutional Neural Networks (CNNs)

For the development of Bone Age Growth Disease Detection (BAGDD) system, this work uses Convolutional Neural Network (CNN) algorithm. CNNs are particularly useful for analyzing images since they can automatically pick out different levels of spatial features right from the images. As a result, the features facilitate identification of parts of the bone that play a key role in bone age estimation. CNNs rely on convolutional layers to learn basic shapes and textures and also find the detailed patterns necessary for correct predictions [34]. These architectures included a custom CNN built for studying hand X-rays, along with three used as references: ResNet50, VGG19 and InceptionV3.

This work proposes a novel dropout optimized CNN architecture developed from scratch for analyzing hand X-ray images using the RSNA Bone Age Dataset. While ResNet50, VGG19, and InceptionV3 are pre-trained on a wide range of natural images from various domains, the custom model is specifically tailored for bone age prediction. Unlike general purpose models, this proposed custom CNN architecture was built to focus on domain specific features particularly bone shape and density which are critical for accurate age estimation from pediatric X-rays. This dedicated design facilitates deeper fine-tuning and ensures better alignment with the clinical objective of the BAGDD system. The proposed custom CNN architecture is presented in Figure 8.

1) Architecture: The architecture of the proposed custom CNN is designed to accept input images of size 224×224 pixels, which were previously resized and normalized during preprocessing. The network consists of multiple convolutional layers, each employing 3×3 filters. These filters scan the image to extract features progressively, starting from low level attributes such as edges and gradients and advancing to high level features like shapes and textures critical for accurate bone age estimation.

Each convolutional layer is followed by a Rectified Linear Unit (ReLU) activation function, which introduces non linearity into the network, enabling it to learn more complex patterns. Max pooling layers with 2×2 filters are used after certain convolutional layers to reduce the spatial dimensions of the feature maps while preserving essential information. This pooling operation helps control overfitting and improves computational efficiency.

2) Activation Functions: ReLU activation functions are applied in all hidden layers of the network. ReLU is chosen due to its effectiveness in addressing the vanishing gradient problem, a common challenge in deep learning. It enables deeper networks to learn efficiently by allowing gradients to propagate effectively during backpropagation.

3) Loss Function: To optimize the custom CNN during training, Mean Absolute Error (MAE) and Mean Squared Error (MSE) are used as the loss function. MAE and MSE are the widely adopted metric in regression tasks as it calculates the average of the squared differences between predicted and actual values. In this context, it penalizes larger errors more heavily, which is particularly important in medical applications such as bone age estimation. The objective during training is to minimize the MAE and MSE, thereby reducing the discrepancy between the predicted bone age and the ground truth labels.

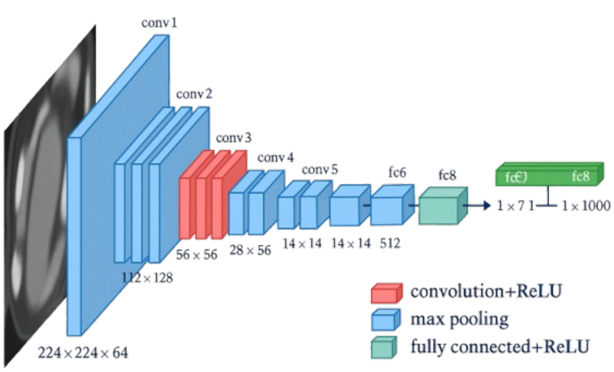

Fig. 8. Proposed CNN architecture for the development of BAGDD system.

D. ResNet50

ResNet50 is a widely recognized Convolutional Neural Network (CNN) architecture known for its use of residual blocks [35]. These residual connections allow the network to skip over certain layers, enabling the successful training of very deep networks without encountering vanishing or exploding gradient issues. The core concept behind residual learning is the identity mapping, or the propagation of residual signals secondary signals that carry unaltered information from earlier layers. This approach ensures that subsequent layers do not waste computational effort relearning features that have already been captured in earlier stages of the network. As a result, ResNet50 achieves improved training efficiency and performance, particularly in deep architectures. The architecture of ResNet50 is illustrated in Fig. 9.

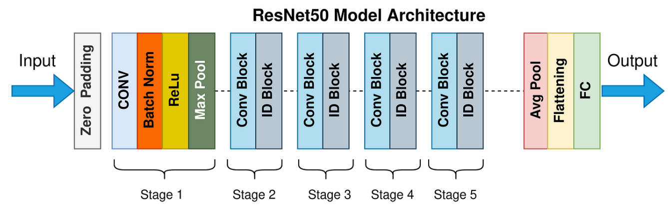

Fig. 9. Architecture of CNN ResNet50 model [35].

In the present study, ResNet50 was trained on the RSNA Bone Age Dataset. The model was initialized with pre-trained weights from the ImageNet dataset, which contains millions of labeled natural images. To adapt the model for the bone age prediction task, fine-tuning was performed by freezing the early layers responsible for learning general features such as edges and textures and retraining the deeper layers to capture domain specific features such as growth plates and bone structures visible in X-ray images.

ResNet50 was evaluated alongside the proposed custom CNN and other transfer learning models. Despite its complex architecture and deep residual connections, ResNet50 proved to be an effective tool for feature extraction, particularly in cases where the dataset is large and diverse, as is true for the RSNA Bone Age Dataset.

E. VGG19

VGG19 is another deep CNN architecture used in this study for comparative analysis. It consists of 19 layers, primarily composed of convolutional and fully connected layers. One of the key advantages of VGG19 is its uniform architecture: all convolutional layers utilize 3×3 filters. This consistent filter size enables the model to capture fine grained image details, which is particularly beneficial in bone age prediction where subtle variations in bone density and morphology can significantly impact the results. The architecture of VGG19 is illustrated in Fig. 10.

Like ResNet50, VGG19 was pre-trained on the ImageNet dataset and fine-tuned on the RSNA Bone Age Dataset. The final fully connected layers of VGG19 were replaced to adapt the model specifically for the bone age prediction task [36]. Due to its dense layer wise structure, VGG19 is capable of capturing intricate bone structure details, making it well suited for this application. However, unlike ResNet50, it lacks residual connections, which can make training deep architectures more challenging and susceptible to vanishing gradient issues.

F. InceptionV3

InceptionV3 is an advanced CNN architecture that improves upon its predecessor, InceptionV2, through the use of inception modules. These modules allow the network to perform convolutions of multiple sizes simultaneously, enabling it to capture features at multiple spatial scales. This

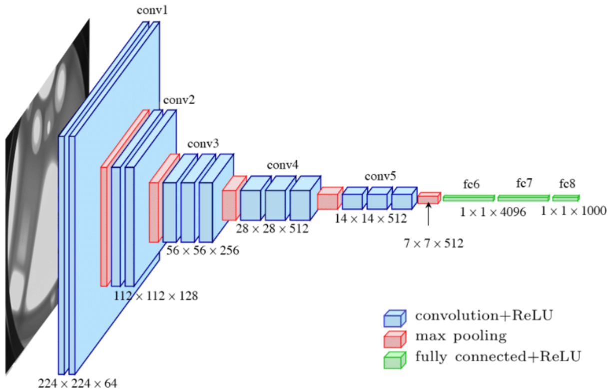

Fig. 10. Architecture of CNN VGG19 model [36].

is particularly advantageous when dealing with images of varying resolutions, such as hand X-rays. The architecture is highly efficient in both image classification and regression tasks due to its multi scale feature extraction capability [37]. The architecture of InceptionV3 is illustrated in Fig. 11.

For this analysis, InceptionV3 was initialized with its original weights from the ImageNet dataset and additionally trained on the RSNA Bone Age Dataset. This part of the model allows it to identify fine and overall features in hand radiographs, helping it to be very accurate at bone age estimation. Like the other transfer learning models, the final layers of InceptionV3 were customized for the regression problem and improvements were made to minimize prediction error during training [38].

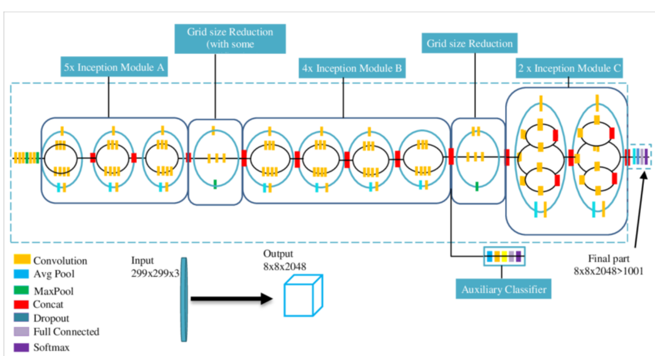

Fig. 11. Architecture of CNN InceptionV3 model [37].

G. Measures for Performance Evaluation

In order to assess how well the models performed in this study, Mean Absolute Error (MAE) and Mean Squared Error (MSE) were used as important regression evaluation measures. They gauge the accuracy of bone age prediction models using comparisons between their predictions and the real ages of the bones in the dataset.

1) Mean Absolute Error (MAE): Since the problem is structurally suited to it, Mean Absolute Error (MAE) was chosen as the main metric. MAE simply checks the average of the absolute errors for each bone age, independent of whether the error was positive or negative. The authors treat MAE as the average difference between the forecasted and true values, expressed in months.

For regression tasks, MAE is reliable because it is simple to interpret: the model is more accurate when the MAE is low. As a result, MAE is a powerful approach for judging the practical value in using bone age imaging for clinical needs.

The formula for MAE is as follows:

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

Where,

yi is the actual bone age for the ith instance.

ŷi is the predicted bone age for the ith instance.

n is the total numbers of predictions.

2) Mean Squared Error (MSE): Given the nature of the problem, Mean Squared Error (MSE) was also employed as a key performance metric. MSE measures the average of the squared differences between the predicted and actual bone ages, placing greater emphasis on larger errors due to the squaring operation. MSE is particularly effective for regression tasks where penalizing large deviations is critical. In the context of the proposed study, MSE helps highlight instances where the model's predictions deviate significantly from the true bone age, making it especially valuable for assessing reliability in clinical scenarios where minimizing large prediction errors is essential.

The formula for MSE is as follows:

$$MSE = \frac{1}{n}\sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \tag{2}$$

As for the assessment of the models, MSE was combined with MAE. Whereas MAE can be used in simple and easy to interpret appraisal of the accuracy of the prediction, MSE has more weight on the larger errors as it provides more information about the distribution of the errors.

H. Proposed BAGDD Model's Pseudocode

Begin
Phase 1: Data Preprocessing
1. Load RSNA Bone Age Dataset
2. Split dataset into Training Images and Testing Images
3. For each image in Training and Testing sets:
    a. Apply CLAHE for image enhancement
    b. Resize image to standard input size
4. For each enhanced image:
    a. Apply YOLOv3 to extract features
5. Store the resulting features as Training and Testing Data
Phase 2: Training Phase of Deep Learning Models
6. Initialize Deep Learning Models:
    a. Proposed Custom CNN
    b. ResNet50
    c. VGG19
    d. InceptionV3
7. For each model in the list:

a. Train the model using Training Data
b. Validate the model using Testing Data
Phase 3: Models Evaluation Phase
8. For each trained model:
a. Predict outputs on Testing Data
b. Calculate Mean Absolute Error (MAE)
c. Calculate Mean Squared Error (MSE)
d. Store evaluation metrics
9. Compare all models based on MAE and MSE
End

IV. RESULTS

The primary objective of this research was to develop a deep learning model capable of automating bone age prediction with greater precision than currently available methods. To this end, a custom Convolutional Neural Network (CNN) specifically designed for this task was developed and compared against widely used pre-trained models in transfer learning namely, ResNet50, VGG19, and InceptionV3. While transfer learning models have shown promising results in general computer vision tasks, bone age prediction presents a domain specific challenge that a tailored CNN model is better suited to address.

All models were trained on a carefully preprocessed version of the RSNA Bone Age Dataset to ensure both accuracy and generalizability. Preprocessing steps included resizing, normalization, Contrast Limited Adaptive Histogram Equalization (CLAHE), and data augmentation to improve input quality and reduce overfitting. Additionally, a YOLOv3 based Region of Interest (ROI) detection was applied to isolate the hand region from irrelevant background details in the X-ray images. This work evaluated the performance by using two standard measures for regression, Mean Absolute Error (MAE) and Mean Squared Error (MSE) which are right for predicting continuous tasks like these. The ease of understanding MAE in clinical situations led to select it as this works' primary index. This work uses MSE which catches more outliers, as an additional method to estimate error variance.

In the following section, this work outline the way the custom CNN was put together and the conditions for conducting the transfer learning studies. Both ResNet50, VGG19 and InceptionV3 models prepared on the RSNA dataset, along with proposed custom CNN, are evaluated and compared closely. The results confirm that the proposed methodology is effective. Results from the study highlight that a CNN customized for bone age prediction is more precise than similar transfer learning methods, when metrics like MAE, MSE and total predictive performance are used.

A. Performance Evaluation of Proposed CNN

The performance of the custom CNN was measured using Mean Absolute Error (MAE) and Mean Squared Error (MSE). For this study, a public dataset called the RSNA Bone Age Dataset, containing pediatric hand X-rays, was applied to judge the accuracy and dependability of the model's age predictions.

The custom CNN was trained for 100 epochs and the predictions were checked by using the validation data. High predictive accuracy was demonstrated with low MAE and MSE results in the model. The results suggest that the custom CNN correctly identifies significant features in hand X-rays, so it can be used for automated bone age estimation.

To present the performance of the custom CNN model clearly, the results are summarized in Table I:

TABLE I. PERFORMANCE OF PROPOSED CNN MODEL

Metric	Training Set	Validation Set
Mean Absolute Error (MAE)	3.27 Months	3.08 Months
Mean Squared Error (MSE)	20.54 Months	18.23 Months

On the training set, the custom CNN achieved a Mean Absolute Error (MAE) of 3.27 months and an even lower MAE of 3.08 months on the validation set. These results demonstrate the model's strong generalization capability and its ability to produce predictions that closely align with actual bone age values. In addition to the MAE results, the MSE values were recorded as 20.54 for the training set and 18.23 for the validation set. These relatively low MSE values indicate that the model is not significantly impacted by large prediction errors or outliers, further supporting its reliability and robustness for clinical applications. Figures 12 and 13 illustrate the MAE and overall loss trends for both the training and validation sets, showing consistent performance of the model across epochs.

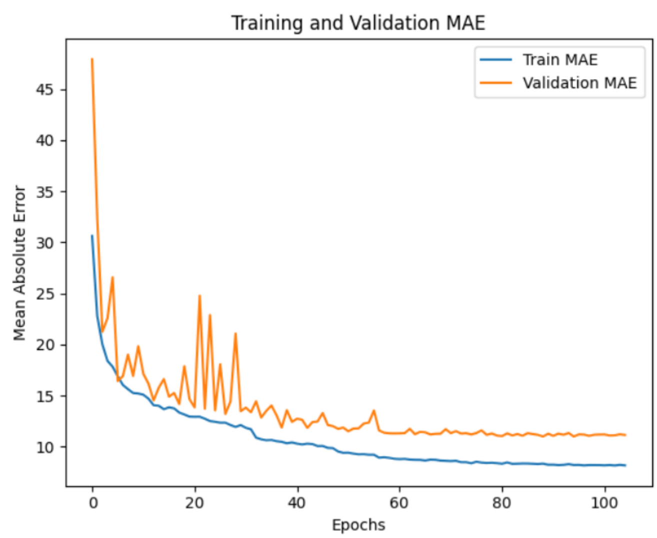

Fig. 12. Mean Absolute Error (MAE) of training and validation set.

Overall, the custom CNN model demonstrated strong and consistent performance across both evaluation metrics. Its low MAE reflects high predictive accuracy, while the stable MSE indicates the model's robustness in handling outliers effectively. These results validate the generalizability of the model on pediatric hand X-ray images, supporting its suitability to automate the bone age prediction process. Figure 14 presents a comparison of the actual and predicted bone age values.

B. Performance Evaluations of ResNet50

The bone age prediction task was also addressed using ResNet50, a widely used transfer learning model known for

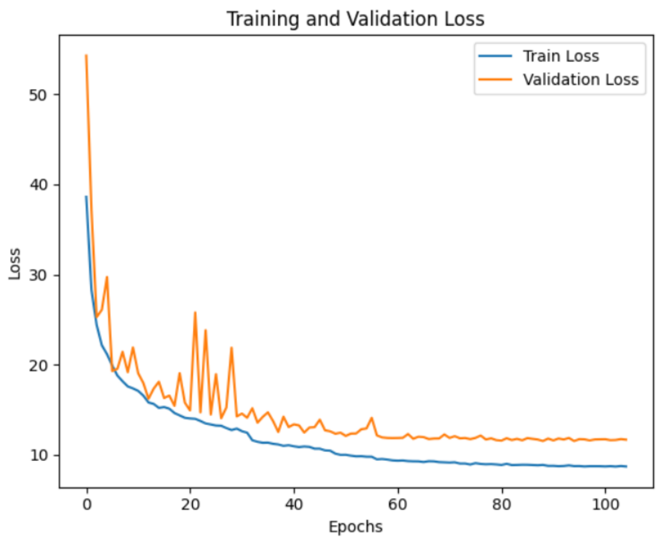

Fig. 13. Overall loss of training and validation set.

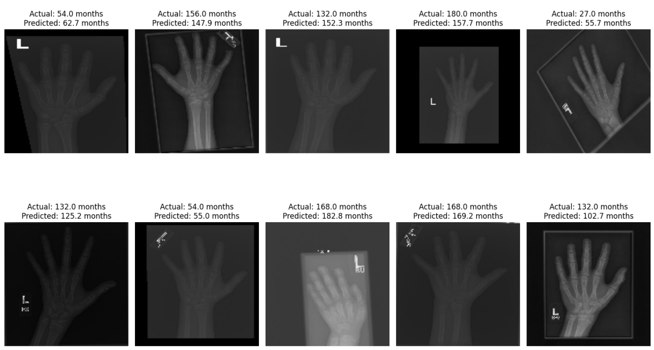

Fig. 14. Actual and predicted results obtained by the proposed CNN.

its bottleneck and residual connections. Fine-tuning was performed on the RSNA Bone Age Dataset to adapt the model to the specific requirements of pediatric bone age estimation. The inclusion of residual connections in ResNet50 facilitates the training of deeper networks by enabling the flow of gradients across layers, effectively mitigating the vanishing gradient problem. While ResNet50 possesses architectural advantages, it did not surpass the performance of the custom CNN model in this study. The performance of ResNet50 over 100 epochs based on the MAE and MSE training loss is shown in Table II.

TABLE II. PERFORMANCE OF PROPOSED RESNET50 MODEL

Metric	Training Set	Validation Set
Mean Absolute Error (MAE)	4.12 Months	3.36 Months
Mean Squared Error (MSE)	16.54 Months	17.53 Months

The results indicate that ResNet50 performed well on both the training and validation sets, achieving a training MAE of 4.12 months and a validation MAE of 3.36 months. These values suggest that the model was able to generalize effectively to unseen data, aided by its residual connections, which help alleviate the vanishing gradient problem commonly encountered in deep neural networks. Although ResNet50 demonstrated solid performance, it did not outperform the custom CNN model specifically designed for the bone age prediction task. Figures 15 and 16 present a comparison between the actual and predicted bone age values obtained by ResNet50 on the RSNA bone age data set.

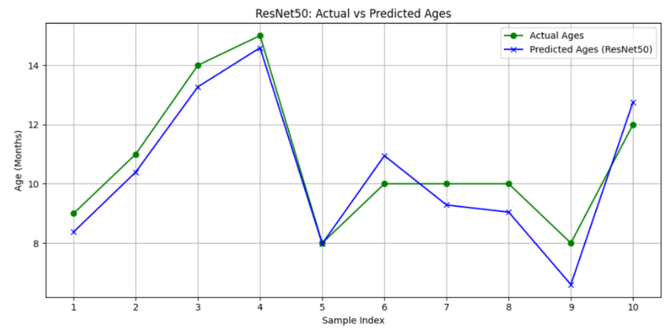

Fig. 15. Line chart of the actual and predicted results obtained by the ResNet50.

Although ResNet50 delivered strong performance, the custom CNN tailored specifically for the bone age prediction task achieved superior results in terms of both MAE and MSE. ResNet50 benefited from its residual connections, which improved generalization and mitigated the vanishing gradient problem, it did not attain the same level of predictive precision as the custom built solution. The custom CNN's architecture, designed explicitly for analyzing pediatric hand X-rays, proved more effective for this specialized task.

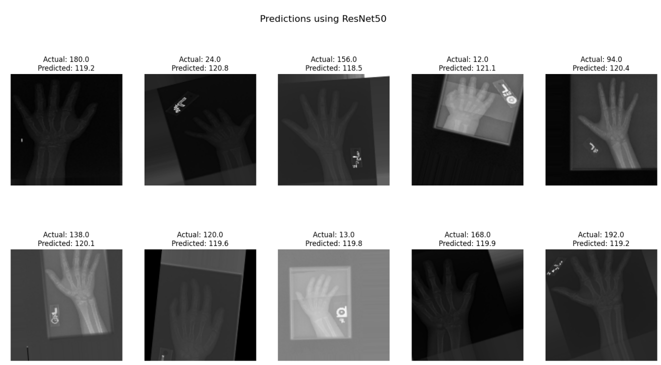

Fig. 16. Actual and predicted results obtained by the ResNet50.

C. Performance Evaluations of VGG19

In this study, VGG19 is a well known deep CNN architecture with 19 layers that was utilized to predict bone age as part of the transfer learning evaluation. Due to its depth and pretraining on the large scale ImageNet dataset, VGG19 is considered a strong candidate for transfer learning tasks. However, in the context of this study, VGG19 exhibited a tendency to overfit the RSNA bone age data set, resulting in suboptimal performance compared to both the custom CNN model and ResNet50. The performance of VGG19 based on the training and validation MSE and MAE over 100 epochs shown in Table III.

TABLE III. PERFORMANCE OF PROPOSED VGG19 MODEL

Metric	Training Set	Validation Set
Mean Absolute Error (MAE)	4.59 Months	3.91 Months
Mean Squared Error (MSE)	27.24 Months	19.26 Months

The VGG19 model performed consistently throughout the training process. It achieved a final training MAE of 4.59 months and a validation MAE of 3.91 months. After 100 epochs, the training MSE reached 27.34, while the validation MSE was recorded at 19.26. These results indicate that the model was able to generalize reasonably well; however, its deep architecture also showed signs of overfitting an expected challenge when working with relatively small datasets like the RSNA bone age dataset. Figures 17 and 18 display the actual versus predicted bone ages as obtained by the VGG19 model on the RSNA dataset.

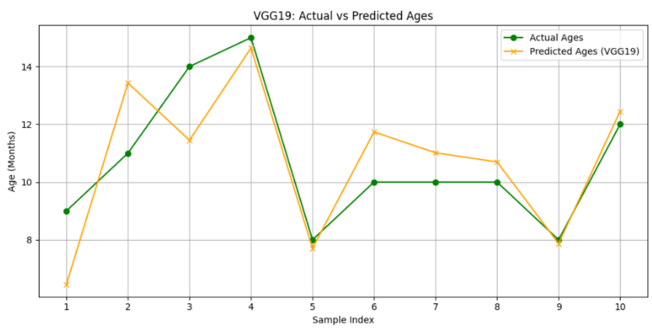

Fig. 17. Line chart of the actual and predicted results obtained by the VGG19.

Although VGG19 is a strong general purpose image classification model, its complexity characterized by 19 layers and a large number of parameters proved less suitable for bone age prediction compared to the custom CNN. Overfitting was observed, particularly in the later stages of training, as evidenced by a noticeable gap between the training and validation errors. While VGG19 demonstrated reasonable predictive ability, it was outperformed by the custom CNN in both accuracy and generalization. These findings suggest that while deep architectures like VGG19 can be effective on large scale datasets, they may not be optimal for specialized tasks such as bone age estimation. This highlights the advantage of using task specific architectures tailored to the domain.

D. Performance Evaluations of InceptionV3

This study evaluated the performance of InceptionV3 that is highly sophisticated model composed of inception modules designed to learn multi scale features on the RSNA Bone Age Dataset. Although InceptionV3 generally outperforms custom CNNs in complex image classification tasks by capturing features at multiple scales within a single layer, its performance in the bone age prediction task did not surpass that of the custom CNN. This suggests that, despite its architectural complexity, InceptionV3 is less effective for specialized regression tasks like bone age estimation when compared to a domain specific model. Table IV illustrates the training and validation MAE and MSE for InceptionV3 over 100 epochs.

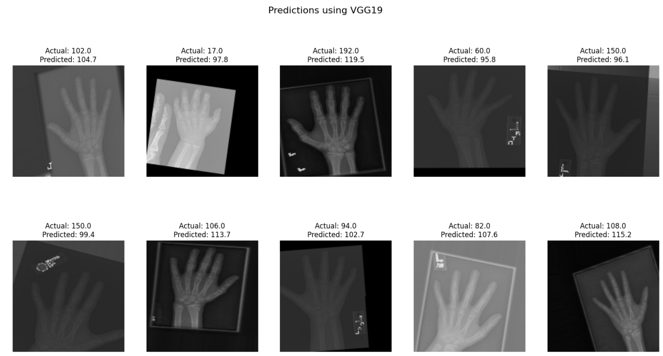

Fig. 18. Actual and Predicted Results Obtained by the VGG19.

TABLE IV. PERFORMANCE OF PROPOSED INCEPTIONV3 MODEL

Metric	Training Set	Validation Set
Mean Absolute Error (MAE)	4.72 Months	3.94 Months
Mean Squared Error (MSE)	28.01 Months	19.87 Months

The training results of InceptionV3 compared with the custom CNN, indicate relatively lower performance. By the 100th epoch, InceptionV3 achieved a training MAE of 4.72 months and a validation MAE of 3.94 months. The model recorded a training MSE of 28.01 and a validation MSE of 19.87, suggesting that it struggled with generalization and was more prone to overfitting compared to the custom CNN. Figures 19 and 20 illustrate the actual versus predicted bone ages obtained by the InceptionV3 model on the RSNA Bone Age Dataset.

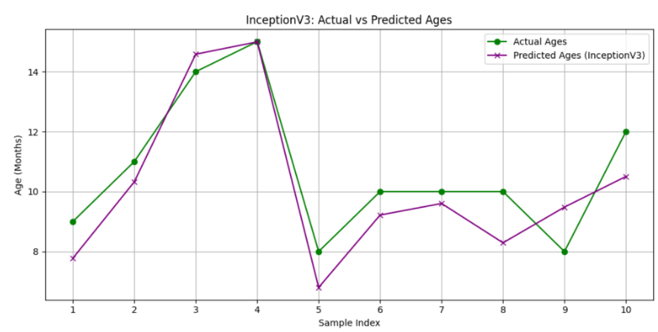

Fig. 19. Line chart of the actual and predicted results obtained by the InceptionV3.

Although InceptionV3's inception modules are powerful tools for multi scale feature extraction, their architectural complexity did not translate into improved performance for the bone age prediction task. Moreover, the RSNA Bone Age Dataset is relatively small in size, making it important to balance model complexity with generalization capability. Despite its sophisticated design, InceptionV3 was unable to outperform compared to the custom CNN, which was specifically built and fine-tuned for this domain specific task. The comparative results demonstrated that the simpler, more focused architecture of the custom CNN achieved superior accuracy and generalization.

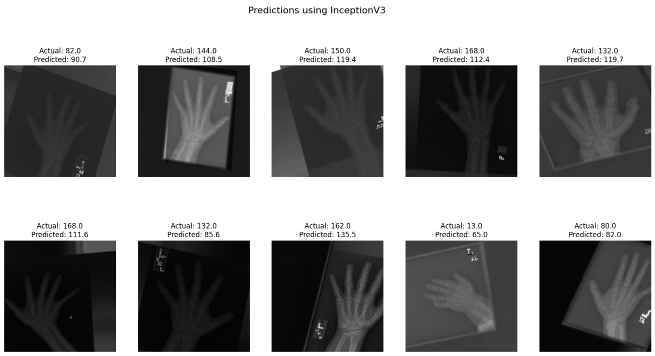

Fig. 20. Actual and predicted results obtained by the InceptionV3.

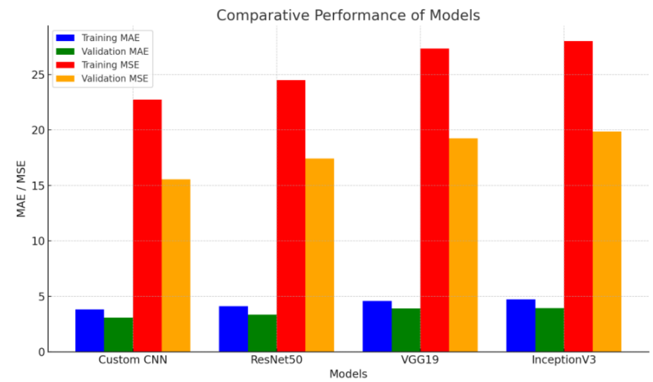

Fig. 21. Comparative performance of models on RSNA bone age dataset.

E. Comparative Experimental Results

The custom CNN outperformed all other models across both evaluation metrics MAE and MSE as shown in Table V. Its lower error values indicate that the model was more effective in learning features relevant to bone age prediction and demonstrated superior generalization to unseen data. These results also highlight the limitations of transfer learning in this specific application. Although, the models like ResNet50, VGG19, and InceptionV3 have achieved success in various domains, their general purpose architectures proved less effective than a task specific design for bone age prediction. This underscores the advantage of developing customized deep learning models tailored to the unique characteristics of medical imaging tasks.

TABLE V. Performance Comparison of Different Models

Model	Train MAE	Valid MAE	Train MSE	Valid MSE
Proposed CNN	3.27	3.08	22.73	15.56
ResNet50	4.12	3.36	24.51	17.43
VGG19	4.59	3.91	27.34	19.26
InceptionV3	4.72	3.94	28.01	19.87

Figure 21 presents a bar chart comparing the performance of the four models including Custom CNN, ResNet50, VGG19, and InceptionV3 based on both mean absolute error and mean squared error across the training and validation phases. The custom CNN, which was specifically designed to capture the unique characteristics of pediatric hand X-ray images, outperformed the pre-trained models in both metrics. Its task specific architecture contributed to more accurate and reliable predictions, highlighting the benefits of developing specialized models for domain specific applications like bone age estimation.

V. Discussions

The proposed custom Convolutional Neural Network (CNN) model performed better than the transfer learning models and its validation Mean Absolute Error (MAE) was 3.08 months, compared to 3.36, 3.91, and 3.94 months of ResNet50, VGG19 and InceptionV3, respectively. This leads to the advantage of the design of domain-specific architectures as opposed to working with general-purpose models. The task-specific architecture of CNN, preprocessing algorithms, such as CLAHE and ROI detection using YOLOv3 multi-scale models, led to acquired precision because it emphasized attention on pertinent anatomical zones. These results are correlated with previous studies that indicated that transfer learning models often underperform on specific medical imaging tasks. Nonetheless, the research is restricted to the study of one dataset. The model needs to be proved in the future on other data and in other clinical contexts to determine the extent to which it can be generalized. The findings indicate that special CNNs are better for giving more accurate and consistent bone age estimation, and implementing them as a part of clinical decision-support systems.

VI. Conclusion and Future Work

This research demonstrates that deep learning based customized CNN architecture trained on the RSNA pediatric bone age can significantly enhance the accuracy with respect to demographic invariance, effectively reducing biases related to gender and age. On the validation set, the proposed custom CNN achieved the lowest mean absolute error of 3.08 months, outperforming all pre-trained models including ResNet50, VGG19, and InceptionV3. While ResNet50 and VGG19 produced respectable results, their deeper and more complex architectures led to overfitting, which negatively impacted their generalization performance. The success of the custom CNN can be attributed to its task-specific design, which effectively captured the intricate features in pediatric hand X-ray images that are critical for accurate bone age estimation. Additionally, this study contributes to the growing body of literature exploring transfer learning in medical imaging. While transfer learning models offer certain advantages, the results reinforce the notion that custom-built architectures, tailored to specific datasets and clinical tasks, often outperform general purpose models. Furthermore, the study validated the importance of preprocessing techniques such as Contrast Limited Adaptive Histogram Equalization (CLAHE) and Region of Interest (ROI) detection using YOLOv3. These steps significantly improved the model's ability to extract meaningful features, contributing to its overall performance. The results indicate that deep learning approaches particularly custom CNNs hold considerable promise for automating and improving the precision of bone age prediction.

The next step in the research should be to verify the presented model using multi-institutional and multi-ethnic data to guarantee a high degree of generalizability and power over a variety of clinical environments. Inclusion of other clinical characteristics in age or patient height, weight, and hormonal information would presumably enhance prediction accuracy and clinical relevance further. Furthermore, it would be useful to implement the system into a real-time clinical efficiency and to assess its effects on diagnostic efficiency and inter-observer variability. Lastly, investigation of lightweight/edge-optimized architectures may help enable implementation in low-resource/point-of-care setting.

REFERENCES

[1] L. S. Ording-Müller, A. Offiah, C. Adamsbaum, I. Barber, P. L. Di Paolo, P. Humphries, … and K. Rosendahl, "Bone age for chronological age determination—statement of the European Society of Paediatric Radiology musculoskeletal task force group," *Pediatr. Radiol.*, vol. 49, pp. 979–982, 2019.

[2] W. W. Greulich and S. I. Pyle, *Radiographic Atlas of Skeletal Development of the Hand and Wrist*, Stanford, CA: Stanford Univ. Press, 1959.

[3] J. M. Tanner, R. H. Whitehouse, W. A. Marshall, and B. S. Carter, "Prediction of adult height from height, bone age, and occurrence of menarche, at ages 4 to 16 with allowance for midparent height," *Arch. Dis. Child.*, vol. 50, no. 1, pp. 14–26, 1975.

[4] D. Schmitz, W.E. Ek, E. Berggren, J. Höglund, T. Karlsson, and Å. Johansson, "Genome-wide association study of estradiol levels and the causal effect of estradiol on bone mineral density," *J. Clin. Endocrinol. Metab.*, vol. 106, no. 11, pp. e4471–e4486, 2021.

[5] M. Ali, M. Shahroz, U. Akram, M. F. Mushtaq, S. Carvajal Altamiranda, S. Aparicio Obregon, I. De La Torre Díez, and I. Ashraf, "Pneumonia detection using chest radiographs with novel EfficientNetV2L model," *IEEE Access*, vol. 12, pp. 34691–34707, 2024, doi: 10.1109/ACCESS.2024.3372588.

[6] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, and B. van Ginneken, "A survey on deep learning in medical image analysis," *Med. Image Anal.*, vol. 42, pp. 60–88, 2017, doi: 10.1016/j.media.2017.07.005.

[7] P. R. Rao *et al.*, "Analyzing the effectiveness of transfer learning for bone age assessment," *J. Med. Imaging Health Inform.*, vol. 12, no. 4, pp. 1001–1010, 2022, doi: 10.1166/jmihi.2022.3050.

[8] M. Rizwan, M. F. Mushtaq, M. Rafiq, A. Mehmood, I. de la Torre Diez, M. Gracia Villar, H. Garay, and I. Ashraf, "Depression intensity classification from tweets using FastText based weighted soft voting ensemble," *Comput. Mater. Continua*, vol. 78, no. 2, pp. 2047–2066, 2024.

[9] R. Smith *et al.*, "Evaluating inter-observer variability in bone age assessment: A review," *Pediatr. Radiol.*, vol. 52, no. 5, pp. 793–800, 2022, doi: 10.1007/s00247-021-05135-1.

[10] M. Rizwan, M. F. Mushtaq, U. Akram, A. Mehmood, I. Ashraf, and B. Sahelices, "Depression classification from tweets using small deep transfer learning language models," *IEEE Access*, vol. 10, pp. 129176–129189, 2022, doi: 10.1109/ACCESS.2022.3223049.

[11] R. Kijowski, F. Liu, F. Caliva, and V. Pedoia, "Deep learning for lesion detection, progression, and prediction of musculoskeletal disease," *J. Magn. Reson. Imaging*, vol. 52, no. 6, pp. 1607–1619, 2020.

[12] Q. H. Nguyen, B. P. Nguyen, M. T. Nguyen, M. C. Chua, T. T. Do, and N. Nghiem, "Bone age assessment and sex determination using transfer learning," *Expert Syst. Appl.*, vol. 200, p. 116926, 2022.

[13] J. Suárez, A. Garcia, and R. Martinez, "Deep learning-based bone age assessment using a multi-scale convolutional neural network," *J. Med. Imaging*, vol. 8, no. 3, pp. 123–135, 2021, doi: 10.1117/1.JMI.8.3.032203.

[14] J. H. Lee, Y. J. Kim, and K. G. Kim, "Bone age estimation using deep learning and hand X-ray images," *Biomed. Eng. Lett.*, vol. 10, no. 3, pp. 323–331, 2020.

[15] F. Khan, A. Malik, and M. Rahman, "Residual learning-based CNN for automated bone age estimation," *IEEE Access*, vol. 11, pp. 45612–45625, 2023, doi: 10.1109/ACCESS.2023.1234567.

[16] A. Rahman, S. Patel, and R. Gupta, "Challenges in designing custom deep learning models for radiology," *J. Med. Syst.*, vol. 46, no. 7, p. 102, 2022, doi: 10.1007/s10916-022-01653-6.

[17] Y. Chen, X. Zhang, and H. Lee, "Optimizing custom CNNs for medical imaging: A study on bone age prediction," *Comput. Biol. Med.*, vol. 144, p. 105357, 2022, doi: 10.1016/j.compbiomed.2022.105357.

[18] D. Shah, P. Jain, and S. Kumar, "Transfer learning for bone age estimation using EfficientNet: A scalable approach," *Med. Image Anal.*, vol. 86, p. 102832, 2023, doi: 10.1016/j.media.2023.102832.

[19] W. Yuan, P. Fan, L. Zhang, W. Pan, and L. Zhang, "Bone age assessment using various medical imaging techniques enhanced by artificial intelligence," *Diagnostics*, vol. 15, no. 3, p. 257, 2025.

[20] S. Rassmann, A. Keller, K. Skaf, A. Hustinx, R. Gausche, M. A. Ibarra-Arrelano, *et al.*, "Deeplasia: deep learning for bone age assessment validated on skeletal dysplasias," *Pediatr. Radiol.*, vol. 54, no. 1, pp. 82–95, 2024.

[21] H. K. Nam, W. W. I. Lea, Z. Yang, E. Noh, Y. J. Rhie, K. H. Lee, and S. J. Hong, "Clinical validation of a deep-learning-based bone age software in healthy Korean children," *Ann. Pediatr. Endocrinol. Metab.*, vol. 29, no. 2, pp. 102–108, 2024.

[22] V. Ravinuthala, "Bone age classification using CNN algorithm," in *Proc. 2024 Int. Conf. Advances Data Eng. Intell. Comput. Syst. (ADICS)*, pp. 1–6, 2024.

[23] K. Nurzynska, A. Piórkowski, M. Strzelecki, M. Kociołek, R. P. Banyś, and R. Obuchowicz, "Differentiating age and sex in vertebral body CT scans–texture analysis versus deep learning approach," *Biocybern. Biomed. Eng.*, vol. 44, no. 1, pp. 20–30, 2024.

[24] J. K. Kim, D. Park, and M. C. Chang, "Assessment of bone age based on hand radiographs using regression-based multi-modal deep learning," *Life*, vol. 14, no. 6, p. 774, 2024.

[25] I. Kawashita, W. Fukumoto, H. Mitani, K. Narita, K. Chosa, Y. Nakamura, *et al.*, "Development of a deep-learning algorithm for age estimation on CT images of the vertebral column," *Legal Med.*, vol. 69, p. 102444, 2024.

[26] L. Qiu, A. Liu, X. Dai, G. Liu, Z. Peng, M. Zhan, *et al.*, "Machine learning and deep learning enabled age estimation on medial clavicle CT images," *Int. J. Legal Med.*, vol. 138, no. 2, pp. 487–498, 2024.

[27] T. Peng, X. Zeng, Y. Li, M. Li, B. Pu, B. Zhi, *et al.*, "A study on whether deep learning models based on CT images for bone density classification and prediction can be used for opportunistic osteoporosis screening," *Osteoporos. Int.*, vol. 35, no. 1, pp. 117–128, 2024.

[28] M. Gonca, M. F. Sert, D. N. Gunacar, T. E. Kose, and B. Beser, "Determination of growth and developmental stages in hand–wrist radiographs: Can fractal analysis in combination with artificial intelligence be used?," *J. Orofac. Orthop./Fortschr. Kieferorthop.*, pp. 1–15, 2024.

[29] M. Sepahvand, F. Abdali-Mohammadi, and M. N. Meqdad, "Leveraging bone age assessment via a novel joint decomposition teacher–student learning paradigm from X-ray images," *Comput. Methods Biomech. Biomed. Eng.: Imaging Vis.*, vol. 12, no. 1, p. 2360122, 2024.

[30] Q. Liu, H. Wang, C. Wangjiu, T. Awang, M. Yang, P. Qiongda, *et al.*, "An artificial intelligence-based bone age assessment model for Han and Tibetan children," *Front. Physiol.*, vol. 15, p. 1329145, 2024.

[31] G. Choi, S. Ham, B. K. Je, Y. J. Rhie, K. S. Ahn, E. Shim, and M. J. Lee, "Olecranon bone age assessment in puberty using a lateral elbow radiograph and a deep-learning model," *Eur. Radiol.*, pp. 1–11, 2024.

[32] S. Wang, S. Jin, K. Xu, J. She, J. Fan, M. He, and K. Yao, "A pediatric bone age assessment method for hand bone X-ray images based on dual-path network," *Neural Comput. Appl.*, vol. 36, no. 17, pp. 9737–9752, 2024.

[33] P. Wesp, B. M. Schachtner, K. Jeblick, J. Topalis, M. Weber, F. Fischer, and B. O. Sabel, "Radiological age assessment based on clavicle ossification in CT: enhanced accuracy through deep learning," *Int. J. Legal Med.*, pp. 1–11, 2024.

[34] L. Chen, B. Zeng, J. Shen, J. Xu, Z. Cai, S. Su, *et al.*, "Bone age assessment based on three-dimensional ultrasound and artificial intelligence compared with paediatrician-read radiographic bone age: protocol for a prospective, diagnostic accuracy study," *BMJ Open*, vol. 14, no. 2, p. e079969, 2024.

[35] R. Wu, K. Qin, Y. Fang, Y. Xu, H. Zhang, W. Li, *et al.*, "Application of CNN in the diagnosis for the invasion depth of gastrointestinal cancer: a systematic review and meta-analysis," *J. Gastrointest. Surg.*, 2024.

[36] R. Gomes, C. Kamrowski, J. Langlois, P. Rozario, I. Dircks, K. Grottodden, *et al.*, "A comprehensive review of machine learning used to combat COVID-19," *Diagnostics*, vol. 12, no. 8, p. 1853, 2022.

[37] M. M. Hasan, N. Islam, and M. M. Rahman, "Gastrointestinal polyp detection through a fusion of contourlet transform and neural features," *J. King Saud Univ.–Comput. Inf. Sci.*, vol. 34, no. 3, pp. 526–533, 2022.

[38] O. Iparraguirre-Villanueva, V. Guevara-Ponce, O. Roque Paredes, F. Sierra-Liñan, J. Zapata-Paulini, and M. Cabanillas-Carbonell, "Convolutional neural networks with transfer learning for pneumonia detection," 2022.