SSRG International Journal of Electrical and Electronics Engineering
ISSN: 2348-8379/ https://doi.org/10.14445/23488379/IJEEE-V13I11P109

Volume 13 Issue 11, 143-159, November 2026
© 2026 Seventh Sense Research Group®

Original Article

An Attention-Enhanced Multi-Scale Xception Framework with Bone-Region Guided Feature Refinement for Pediatric Bone-Age Assessment

Odugu Brahmam[1], Vishwanath Jha[2]

[1]*Department of Mathematics, Shri Venkateshwara University, Gajraula, Uttar Pradesh, India.*
[2]*Department of CSE (Data Science), Inderprastha Engineering College, Ghaziabad, Uttar Pradesh, India.*

[1]*Corresponing Author: brahmam.srs@gmail.com*

Received: 21 September 2026 Revised: 06 November 2026 Accepted: 17 November 2026 Published: 25 November 2026

Abstract *- The automatic estimation of bone age from pediatric hand radiographs is an important clinical task that facilitates growth analysis while overcoming the subjectivity and effort associated with manual bone age estimation. However, traditional deep learning models may have problems jointly capturing the important anatomic parts of bones and the developmental patterns appearing at different spatial scales. To overcome these difficulties, this study proposes a novel framework based on the integration of Xception, Convolutional Block Attention Module (CBAM), Multi-Scale Feature Extraction (MSFE), and Bone Region Guided Feature Refinement (BRFR) modules that can be used for the automatic estimation of bone age from pediatric radiographs. This architecture uses the pretrained ImageNet model Xception for efficient hierarchical feature extraction and MSFE with 3 × 3, 5 × 5, 7 × 7, 11 × 11, and 17 × 17 convolutional filters to extract multi-scale features. Attention-guided representation learning is performed with the CBAM mechanism, and the BRFR module helps to emphasize bone-related features and filter out irrelevant information before regression. The obtained MAE was 9.52 months (0.79 years), MAPE was 7.48%, and R^2 was 0.9328. The ablation studies show the contribution of each component to the overall framework performance, from the baseline Xception model to the CBAM, MSFE, and BRFR modules. Kernel ablation proves the importance of all 5-scales of feature extraction. This framework demonstrates a lower MAE despite the differences between the datasets and evaluation protocols compared to existing works. Overall, the results prove that the proposed combination of feature extraction, attention-guided representation learning, multi-scale feature extraction, and bone region-guided refinement is effective for estimating the bone age of pediatric patients.*

Keywords *- Bone Age assessment, Xception model, Attention mechanism, Multi scale convolutions, Convolutional Block Attention Module, Deep Learning in Medical Imaging, Skeletal maturity prediction.*

1. Introduction

Bone development is a progressive process involving structural growth, morphological changes, and increased mineralization. This occurs through endochondral ossification. Cartilage is gradually replaced by bone in the primary ossification centers of the diaphysis and secondary ossification centers of the epiphyses.

Until bone growth ceases, at which point the epiphyseal plate is fully ossified and it can be seen that the diaphysis and epiphysis fused together. Related terminologies to this are skeletal maturity, bone age and chronological age.

While chronological age is based on a person's birthdate, bone age calculates an individual's age based on their skeletal maturity [1], which is the stage of bone development [2].

There are two primary reasons for determining a person's skeletal maturity, or Bone Age (BA) and comparing it to their chronological age. First, from a medical perspective, bone age evaluation is helpful in diagnosing and treating children endocrinology, orthodontics, and pediatric orthopedic diseases. It is also taken into account when estimating a person's eventual height [3]. Second, where there are no verifiable documents available, determining a person's bone age is crucial from a legal perspective to determine juvenile status.

Three primary categories can be distinguished from clinical data on bone age: those with advanced bone age, those with chronologically appropriate bone age, and those with delayed bone age [4]. In clinical and forensic settings, Bone Age (BA) is essential for determining growth, predicting adult height, and treating illnesses. Because of environment and

This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/)

hereditary factors, traditional approaches such as Greulich-Pyle [5] and Tanner-Whitehouse [6] are unsuccessful. To overcome these constraints, automated alternatives such as HANDX have surfaced. AI-driven methods are now being developed by researchers to increase the precision and effectiveness of BA evaluations [7].

The work [8] aims to enhance and complement the radiographic Tanner-Whitehouse (TW3) method using ultrasound images to reach a TW3 skeletal maturity score. 442 participants aged 1-18 years underwent radiography. Advanced or delayed bone age was diagnosed using the transformed scores. Sensitivity and specificity for radiographically determined bone age were 97% and 98%, respectively, with an intra-class correlation coefficient of 0.993.

The objective of the study [9] was to find correlation between bone X-ray images and the age of a person. Five Regions were made and the measurements of textural parameters were made. The left femoral neck had the highest correlations, which indicates age-dependent differences in bone structure are more apparent at this site. This can aid in earlier detection of osteoporosis, treatment planning, and the identification of potential risk factors.

This review [10] explores AI applications in musculoskeletal imaging, focusing on trauma, bone age estimation, osteoarthritis, tumors, and implant-related pathologies. Convolutional Neural Networks (CNNs) and AI algorithms excel in fracture detection, bone age assessment, osteoarthritis diagnosis, tumor prediction, and orthopedic implant positioning. This work encourages that AI should be used to support workflow optimization, improve diagnostic performance, and handle growing workloads.

The study [11] was carried out to evaluate whether the machine learning approach would be able to detect sex- and age-related physiological changes in lumbar vertebrae based on CT imaging radiomics features.

The study consisted of 233 participants (115 women and 118 men) who underwent lumbar CT for the evaluation of back pain in 3 different scanners. These results demonstrated that radiomics features could accurately identify sex with a high degree of accuracy in a single scanner. However, accuracy decreased on a combined dataset of the three scanners. However, scan acquisition from other scanners lower the accuracy of the analysis.

Generally, X-rays are used to evaluate bone age, but this is inadequate for a diagnosis of specific diseases. Magnetic Resonance Images (MRI) might broaden diagnostic options and allow for bone age testing to be conducted as routine screening. Bone age in boys aged 9-17 showed high correlation with MRI-derived textural features. The study [12] found that MRI produced the same reliable results as X-rays for determining bone age, revealing that MRI scoring of bone age does not expose patients to ionizing radiation.

The study [13] employs low-resolution whole-body X-ray images for training to predict age. The authors introduced a preprocessing pipeline and leveraged CNNs with semi pretrained to estimate the age of children and young adults. Their approach achieved a mean absolute error of 15.56 months between the predicted and actual chronological ages. The findings indicate that spatial features played a dominant role in explaining the variance observed in the age prediction outcomes.

Regarding radiological skeletal approaches, GPA is the predominant technique for measuring biological age. This approach relies on the morphology and development of main and secondary ossification centers, as well as the timing of their union.

The TW approach entails assessing areas of interest and grading them. An enhanced iteration of the TW3 approach has incorporated additional imaging zones for the radius, ulna, and short bones [14, 15]. This approach allocates a score to each assessed bone segment and is more comprehensive than a mere comparison with conventional radiography. Also, gender disparities were considered.

An algorithm for automated pelvic bone marrow analysis was developed and validated [16]. Semi-manually annotating 180 MRIs from 54 patients trained a nnU-Net to segment. For automated segmentation quality, 15 manually segmented whole-body MRIs from 3 centers were scored using the dice score.

1.1. Objectives of the Study

The fundamental aim of this study is to develop an efficient deep learning-based architecture for pediatric bone age estimation using hand radiographs by improving skeletal feature extraction, attention mechanism, multi-scale modeling, and anatomical feature refinement.

The objectives of the study are as follows:

1. To implement an automatic bone age regression system based on a pre-trained Xception Neural Network used for hierarchical anatomical feature extraction from pediatric hand radiographs.
2. To improve skeletal feature discrimination through the use of CBAM with both channel and spatial attention, which highlights relevant bone structures and suppresses non-informative image information.
3. To represent skeletal development at multiple scales using parallel convolutional pathways with 3×3, 5×5, 7×7, 11×11, and 17×17 kernels for both fine-grained anatomical features and overall bone structures.

4. To propose Bone-Region Guided Feature Refinement (BRFR) as an anatomical feature refinement process to enhance bone-related representations and suppress background or irrelevant features before bone age regression.
5. To investigate the role of each of the aforementioned components, individually and collectively, through systematic component-level ablation studies.
6. To test the efficacy of different kernel sizes used in multi-scale convolution and the possibility of progressively larger receptive field diversification in pediatric bone age estimation.
7. To analyze the robustness of the proposed framework through repeated experiments, error analysis, comparisons with the baseline CNN architectures, and Wilcoxon signed-rank test

Despite the remarkable progress in automated bone age analysis, it is still challenging to precisely capture the complex features of children's bones. Current CNN-based methods are not capable of modeling fine-grained features from ossification centers, epiphyseal, and skeletal structures at various scales. In addition, few studies have explored multi-scale feature learning, adaptive attention mechanisms, and bone region-guided refinement. Thus, an effective framework is urgently needed to integrate local and global information for more accurate bone age estimation.

This study presents an approach using the design of a component level in an integrated deep learning architecture specifically designed for bone age prediction in children. The architecture employs Xception as the hierarchical feature extractor and CBAM to conduct channel and spatial attention refinement, thereby increasing the ability of the network to attend to informative skeletal representations.

The architecture includes a Multi-Scale Convolution module that uses parallel kernels of size 3×3, 5×5, 7×7, 11×11, and 17×17 to capture the skeletal features through multiple receptive field sizes and thus capture both the local and global patterns of bone development.

The novel architectural contribution includes the Bone-Region Guided Feature Refinement (BRFR) module that offers another feature modulation layer based on the creation of guidance for bone regions and using the guidance to enhance the bone-related features as well as ignore the background information.

The article is presented as follows: Section 2 presents the literature review. Section 3 discusses a mathematical formulation and also states and proves the theorems relevant to the study. In section 4, proposed model is discussed. Section 5 focuses on the computational analysis and reports the results. Lastly, Section 6 covers the conclusion and provides future directions for this research.

2. Literature Survey

Skeletal maturity advances through distinct phases, a principle commonly applied in pediatrics, where BAAs are juxtaposed with chronological age to evaluate endocrine and metabolic disorders. Despite being fundamental to numerous disease assessments, the process has seen minimal advancements since its inception in 1950. The work [17] presents an automated DL pipeline designed to execute BAA with proper segmentation and preprocessing. These models employ a Convolutional Neural Network (CNN) pretrained on ImageNet and subsequently fine-tuned, resulting in accuracies of 0.5732 for the female cohort and 0.6140 for the male cohort on the held-out test images. Test radiographs of females received a BAA in 0.9039 of cases within one year and in 0.9811 of cases within two years. Test radiographs for males achieved an assignment rate of 0.9418 within one year and 0.99 within two years.

Attention maps were generated using the input occlusion method, illustrating the features utilized by the trained model to execute BAA. These align with the criteria assessed by human experts during manual BAA evaluations. The fully automated BAA system served as clinical decision support tool, delivering accurate assessments in less than 2 seconds, surpassing conventional methods.

The method of diagnosis presented in the related work [18] using multi-feature fusion to classify X-ray images. It uses a two-layer CNN architecture to retrieve global and local features from X-ray images, which are then given to fully connected neural network classifier. The method showed an improvement of 50% and 54.4% in average accuracy and sensitivity (98.5% and 97.6%, respectively) over a global feature-only method. It shows that this method can significantly reduce the cost of large-scale screening and the missed diagnosis rate.

In order to automatically estimate the bone age of participants from newborn to 18 years old, computer vision techniques were employed in this prior research [19]. They utilized 442 left-hand radiographs from the hand atlas at the University of Southern California, obtaining features from carpal and epiphyseal regions of interest. Results: The study yielded an accuracy of 0.5 years mean absolute error across both genders. It is simple to implement, robust, and eligible for computer-aided diagnosis.

The traditional bone age evaluation is a tedious and lengthy procedure with observer variability. While fully automated systems are highly sought after, achieving accuracy and reliability remains challenging. However, recent advances in deep learning and CNN-based methods show promising results in automated bone age assessment. This investigation [20] outlines the history of bone age estimation, discuss its benefits and conventional evaluation techniques, and examines the existing AI-based solutions.

The Cervical Vertebral Maturation (CVM) technique determines craniofacial skeletal maturation, which is important for orthodontic and orthopedic therapy. In this study [21], the authors investigate deep learning for autonomous CVM evaluation. Specifically, they present iCVM, a Convolutional Neural Network. It is tailored to CVM evaluation difficulties using a residual network. They use Grad-CAM to examine model prediction areas. They also offer CVM-900, a new dataset, alongside the study to spur further research. It comprises the cervical section of 900 lateral cephalograms from orthodontic patients of various ages and genders. Experimental findings reveal that the suggested technique outperforms CVM-900 in several assessment measures.

This study [22] is concerned with ensuring data integrity within scholarly publishing, including author statements when submitting a manuscript, replication of results, and peer review. The authors highlight the need to elevate radiological research by moving beyond diagnostic performance toward evaluating its impact on therapy, patient outcomes, and societal benefits. The recommendation regarding the emerging topic of radiomics is to pursue collaboration with data scientists, potentially integrating them into clinical departments. For the purpose of BAA, this research [23] built a multi-modal deep learning model based on regression that takes into account both clinical data (such as patient gender and chronological age) and hand radiography images. A multi-modal BAA model based on regression was developed using a dataset of hand radiography images from 29,74 pediatric patients.

In summary, existing studies have achieved great success in developing automated methods for bone age estimation based on deep learning. However, it is still challenging to learn detailed features of the skeleton at different scales and highlight the anatomically meaningful regions. Most existing methods focus on either feature learning, attention mechanisms, or anatomical information; therefore, there is potential to explore the combination of these methods for bone age regression.

2.1. Research Gaps

Even with all the advancements in using deep learning methods for automated bone age evaluation, existing CNN models are still limited in terms of their ability to (i) focus on skeletal features at a very detailed level and (ii) simultaneously analyze developmental trends at different spatial scales.

1. Lack of Focus on Diagnostically Important Skeletal Structures: Current BAA frameworks utilizing CNNs might fail to provide the necessary attention to important diagnostic parts of skeletal structures or identify subtle changes in the morphology and ossification of bones. This deficiency can be overcome by introducing an architecture that combines Xception and Convolutional Block Attention Module (CBAM). While Xception will help in effective feature extraction through the use of depthwise separable convolutions, the CBAM layer will be used to apply channel and spatial attention and help highlight important features of the skeleton bones.
2. Lack of Multi-Resolution Representation of Bone Age Features: Current BAA methods might lack adequate representation of the skeletal features that exist on different scales of resolution, which limits the representation of both fine-scale anatomical details and large bone structures. This deficiency can be remedied by extending the Xception + CBAM architecture using multi-scale convolution with several kernel sizes for feature extraction.

3. Mathematical Formulation

Deep learning has transformed the field of medical imaging, giving very predictive power in diagnostic applications. Convolutional Neural Networks (CNNs), which can successively capture hierarchical features from skeletal images, have been widely used in bone age prediction. Xception is one of those networks, and it has proven successful by using a depthwise separable convolution architecture.

Nevertheless, basic convolutional behaviors have some drawbacks, such as duplicate feature maps, loss of spatial information, and the ineffectiveness of capturing multi-scale structures in the input successfully. In response to these issues, two main improvements are proposed: hybrid attention and multi-scale convolution, which can effectively reflect the features and enhance prediction accuracy.

To address the issue of losing information due to the inability of the attention mechanisms to suppress unnecessary features, hybrid attention mechanisms optimize feature representations by focus on different channels and spatial regions adaptively, thus minimizing the variance of extracted feature maps while increasing entropy.

Leveraging channel attention, the model adaptively reads the feature importance across each channel, and leveraging spatial attention, it focuses strongly on those parts in the image that matter the most to predict the bone age of the image. This two-phase refinement pulls out irrelevant information, tightening generalization and lightening prediction dropout.

Moreover, multi-scale convolutional architectures generalize the standard convolutional kernel by introducing kernels of different receptive fields. This multi-scale strategy allows the model to capture both fine-grained skeletal structures and more general contextual patterns, enhancing the extraction of key indicators of bone growth. Different basis features across multiple kernel sizes are introduced, enabling

an enriched feature representation learnt by the Xception model, which facilitates lower error in bone age compared to single kernel sizes and also encourages improvements in generalization across different input images.

This work provides a theoretical understanding to demonstrate the efficacy of the hybrid attention and the multi-scale convolution for prediction of bone age through Xception model. This paper presents two theorems with both formulation and proof:

Hybrid Attention Theorem – Hybrid attention leads to higher feature expressiveness via variance reduction and entropy increase, which in turn leads to better results.

Multi-Scale Convolution Theorem – Proving that multi-scale convolutions enable scale-variance, leading to generalized feature extraction across multiple receptive fields, which is advantageous for reducing error in human bone age estimation through improved robustness.

This validates the proposed modifications in the context of medical imaging applications with respect to existing deep learning methods.

The outcomes demonstrated in this work provide further momentum toward advancing the use of deep learning methodologies for medical diagnostics, specifically considering its successful implementation to automate the task of bone age reading.

Theorem 1: Efficiency of Hybrid Attention Mechanism in Xception Model for Bone Age Prediction

Given an input image X, the feature extraction function of the Xception model, $f(X)$ produces feature maps. The application of a hybrid attention mechanism, which combines channel attention $A_c(X)$ and spatial attention $A_s(X)$, improves feature discrimination and reduces noise, leading to an increase in bone age prediction accuracy $P(A_h(X)) > P(f(X))$ where the hybrid attention mechanism is $A_h(X) = A_c(X).A_s(X)$

Step 1: Feature Space Representation

Define the feature map extracted from Xception as $F=f(X) \in R^{C\times H\times W}$ where C is the number of channels, H and W are spatial dimensions.

Each feature activation at location (i,j) for channel c is represented in Equation (1)

$$F_c(i,j) = f_c(X,i,j) \tag{1}$$

This feature map contains redundant and noisy information that can be refined using attention mechanisms.

Step 2: Channel Attention Mechanism

Channel attention assigns an importance weight to each channel. It is computed using Equation (2),

$$A_c(F) = \sigma(W_c GAP(F)) \tag{2}$$

Where W_c is a trainable weight matrix, GAP(F) (Global Average Pooling) computes a per-channel average is given by Equation (3),

$$GAP(F_c) = \frac{1}{H\times W}\sum_{i=1}^{H}\sum_{j=1}^{W} F_c(i,j) \tag{3}$$

σ is the sigmoid activation function, ensuring $A_c \in (0,1)$

Applying channel attention modifies the feature representation as in Equation (4):

$$F = A_c(F).F \tag{4}$$

Since $A_c(F)$ is a positive weight applied to each channel, the variance of the refined feature map satisfies the condition as in Equation (5):

$$Var(P(F)) \le Var(P(F)) \tag{5}$$

This variance reduction indicates that redundant feature activations are suppressed, enhancing meaningful signal retention.

Step 3: Spatial Attention Mechanism

Spatial attention enhances feature activations in specific spatial locations. It is defined as represented in Equation (6):

$$A_s(F) = \sigma(W_s * Conv(F)) \tag{6}$$

Where, W_s is a trainable convolutional filter (e.g., 7×7 kernel), * represents convolution.

The final feature representation after applying spatial attention is represented in Equation (7),

$$F'' = A_s(F) * F \tag{7}$$

Similar to channel attention, spatial attention filters out irrelevant spatial features, reducing variance and improving prediction robustness, as given by Equation (8)

$$Var(P(F'')) \le Var(P(F)) \tag{8}$$

Since both attention modules reduce variance in meaningful feature dimensions and suppress unimportant

ones, the entropy of meaningful feature information increases as shown in Equation (9):

$$H(F^{''}) \geq H(F) \tag{9}$$

Since predictive performance improves with better feature representation, i.e. as in Equation (10):

$$P(A_h(X)) \geq P(f(X)) \tag{10}$$

Thus, Hybrid attention enhances feature expressiveness, reducing noise and improving bone age prediction accuracy.

Theorem 2: Impact of Multi-Scale Convolution on Bone Age Prediction Using Xception Model

Using multiple convolutional kernels of different sizes in the Xception model improves bone age prediction by capturing features at different spatial resolutions $P(M(X)) \geq P(F(X,k))$, where $M(X)$ is a multi-scale convolutional operation defined as in Equation (11):

$$M(X) = \sum_{i=1}^{n} Wk_i * X \tag{11}$$

Step 1: Standard Single-Scale Convolution

A typical convolution operation with kernel size k is represented as in Equation (12):

$$F(X,k) = Wk * X \tag{12}$$

The information captured by this convolution depends on the receptive field k, which limits the level of detail the model can perceive. The entropy of extracted features using a single kernel is denoted by Equation (13):

$$H(F(X,k)) = H(Wk * X) \tag{13}$$

This entropy is limited by the kernel size because it captures only a fixed-scale representation.

Step 2: Multi-Scale Convolution and Entropy Increase

Multi-scale convolution aggregates feature information across multiple receptive fields. Since each kernel captures different levels of abstraction, the total entropy of extracted features is represented in Equation (14):

$$H(M(X)) = H(\sum_{i=1}^{n} Wk_i * X) \tag{14}$$

By Jensen's Inequality as in Equation (15):

$$H(\sum_{i=1}^{n} Wk_i * X) \geq H(F(X)) \tag{15}$$

Since the total entropy increases, more information is retained, and predictive uncertainty is reduced.

Step 3: Generalization and Error Reduction

The bone age Prediction Function $P(F)$ maps features to predicted bone age $\hat{y}$. The expected error of the prediction is represented in Equation (16):

$$E[|y - P(F(X,k))|] \tag{16}$$

For multi-scale convolution, it is as in Equation (17):

$$E[|y - P(M(X))|] \leq E[|y - P(F(X,k))|] \tag{17}$$

Since lower error corresponds to better generalization, it follows that $P(M(X)) \geq P(F(X,k))$

Thus, Multi-scale convolution enhances predictive power by increasing feature entropy and capturing multi-level patterns, improving bone age prediction accuracy.

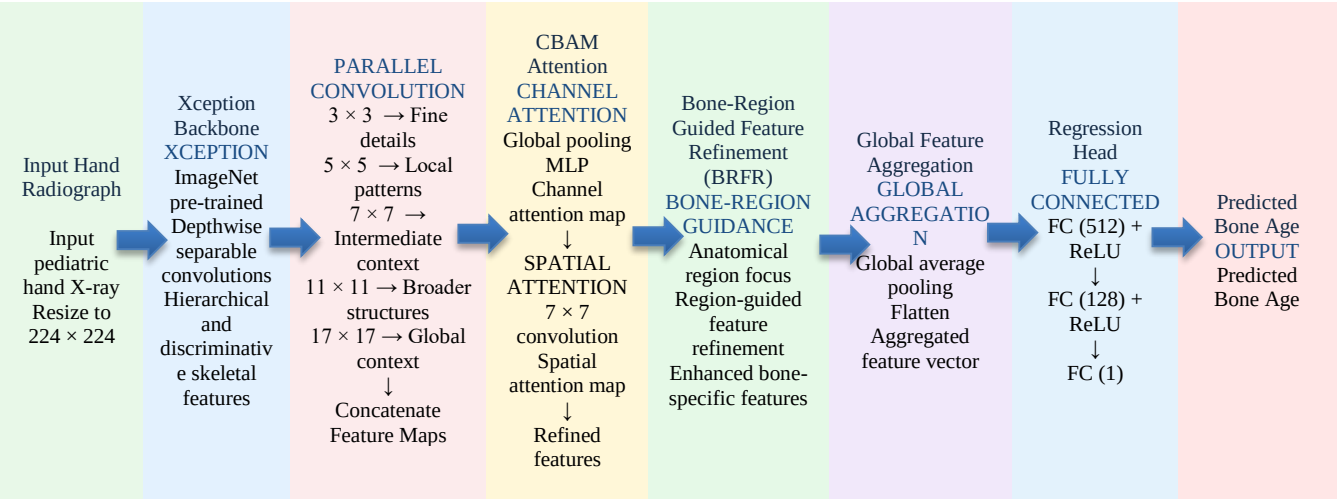


Fig. 1 Overall architecture of the proposed xception–cbam–multi-scale–brfr framework for pediatric bone-age estimation

4. Methodology of the Proposed Model

The proposed framework attempts to automate bone age estimation using hand radiographs through deep learning feature extraction, feature refinement with attention guidance, multi-scale skeletal representation, and bone region-guided feature refinement. The entire architecture involves four major steps: feature extraction using Xception, multi-scale convolutions, feature refinement with an attention mechanism based on CBAM and Bone-Region Guided Feature Refinement (BRFR). The main goal is to convert the raw input image into a discriminative skeletal feature representation through multiple steps. The architecture of the proposed Xception–CBAM–Multi-Scale–BRFR framework for automatic pediatric bone age estimation using hand radiographic images is shown in Figure 1.

4.1. Input Preprocessing and Data Preparation

The RSNA Pediatric Bone Age dataset was used to train and evaluate the proposed architecture. The dataset consists of 12,611 hand radiographs annotated with bone age labels, and 200 unlabelled images are provided in the test dataset. Thus, only a quantitative evaluation was performed on the labeled portion of the dataset. Data from the labeled part were split into training, validation, and internal test sets in the proportions 80%, 10%, and 10%, respectively, which corresponded roughly to 10,089, 1,261, and 1,261 images. All images were resized to 224 × 224 pixels before being fed into the network. In the training stage, data augmentation was applied to increase the robustness against image transformations, such as orientation and appearance changes, with moderate rotation, width/height translation, zoom, and brightness augmentations. None of the augmentations were applied to the validation and internal test data.

4.2. Xception-based Deep Feature Extraction

The preprocessed image of the hand was first passed through an Xception network pretrained on the ImageNet dataset. The Xception model utilizes depthwise separable convolutions instead of traditional convolutional layers to perform spatial filtering and channel-wise feature transformations efficiently.

Such an architecture provides an efficient way to extract hierarchies of features from radiographic images while maintaining rich anatomical details. The extracted feature map contains features representing different aspects of the skeleton, including the bone shape, joint structure, growth patterns, and localized intensity changes. These are further passed to the multiscale-feature extraction block.

4.3. Multi-Scale Convolution for Skeletal Pattern Representation

The limitation arising from the dependency on receptive field size can be mitigated using the Multi-Scale Convolution block, which is employed in the feature extraction process using the architecture based on the Xception network. These blocks operate on the extracted feature map using multiple branches that perform convolutions with kernel sizes of 3 × 3, 5 × 5, 7 × 7, 11 × 11, and 17 × 17 simultaneously.

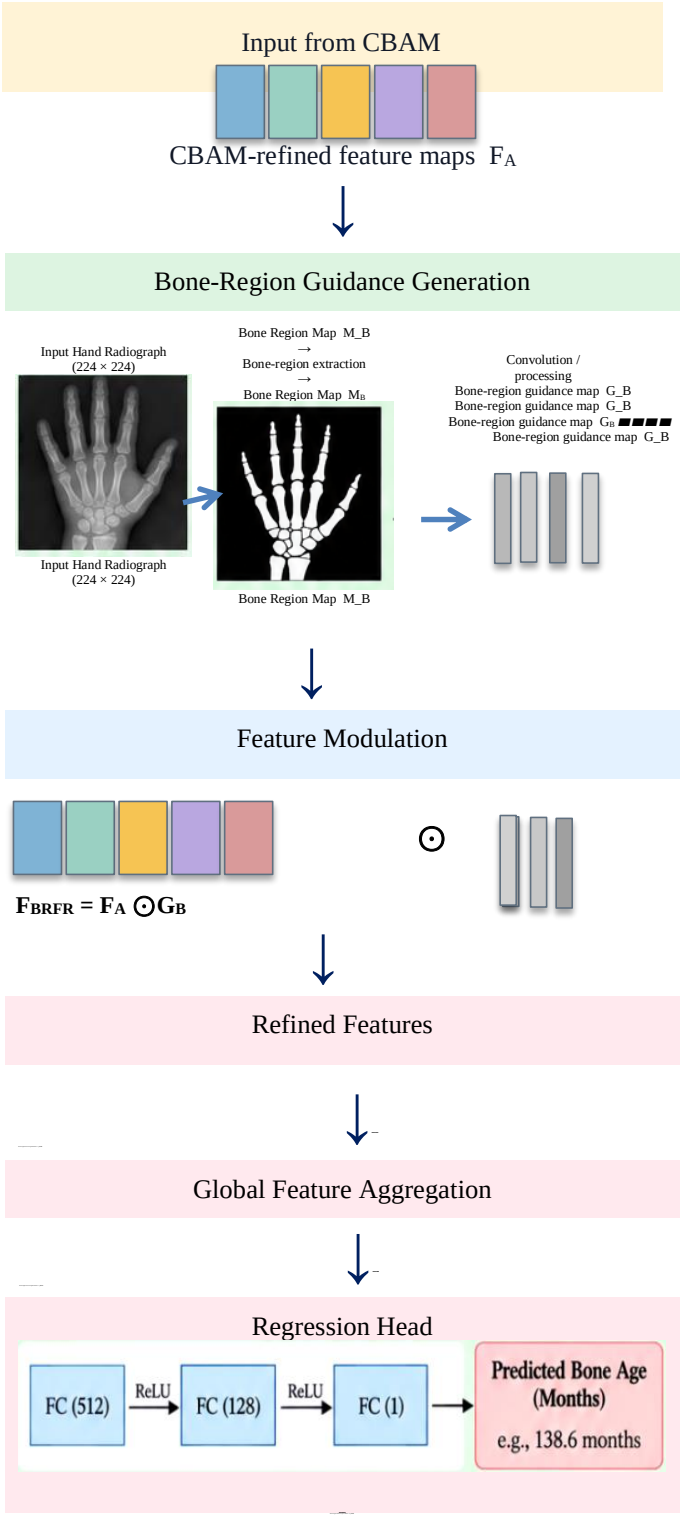


Fig. 2 Detailed architecture of the Bone-Region Guided Feature Refinement (BRFR) module

Smaller kernels learn finer local information, whereas larger kernels learn increasingly larger contexts. This is especially significant for bone age classification, where the maturity of the skeleton is represented by changes in both the local and global anatomical structures.

Mathematically, for an input feature representation F, the multi-scale features can be expressed as in Equation (18):

$$F_k = \phi(W_k * F + b_k), k \in \{3,5,7,11,17\} \quad (18)$$

where W_k represents the convolutional kernel of size $k \times k$, b_k is the corresponding bias, * denotes convolution, and $\phi(\cdot)$ represents the nonlinear activation function. The resulting multi-scale representation is formulated as in Equation (19):

$$F_{MS} = CONCAT(F_3, F_5, F_7, F_{11}, F_{17}) \quad (19)$$

4.4. CBAM-based Attention Refinement

The multi-scale representation was refined using a Convolutional Block Attention Module (CBAM). Two complementary approaches are employed by the CBAM: channel attention and spatial attention. Channel attention involves evaluating the importance of the different channels of the features, emphasizing those channels with more relevant skeletal information. Spatial attention is used to identify important spatial locations in feature maps, ensuring that the focus of the network is shifted towards areas that play a greater role in bone age estimation. The sequential use of the two types of attention ensures the production of an attention-based representation that suppresses irrelevant Information

4.5. Bone-Region Guided Feature Refinement

As a result of attention enhancement, the suggested architecture adds a Bone-Region Guided Feature Refinement (BRFR) phase, which focuses on emphasizing features that are more related to anatomy. While the CBAM implicitly learns the importance of certain features based on the input representation, the BRFR aims to reinforce those features that belong to the bone-dominated regions of the radiograph. The generated representation is used as a feature-wise modulation scheme that increases the importance of the features related to bones and reduces the importance of the background features. The working principle of the BRFR block is shown in Figure 2.

The refined representation can conceptually be expressed as in Equation (20):

$$F_{BRFR} = F_A \otimes G_B \quad (20)$$

where F_A denotes the CBAM-refined feature representation, G_B represents the learned bone-region guidance map, and $\otimes$ denotes element-wise feature modulation. The resulting F_{BRFR} is therefore expected to retain diagnostically meaningful skeletal information while reducing irrelevant responses.

4.6. Feature Aggregation and Bone-Age Regression

The resultant feature representation is further compressed into a compact representation using global feature aggregation. The representation is then fed into a fully connected regression head, resulting in a scalar prediction of the bone age. In contrast to classification methods, which categorize age into various classes, the regression method predicts bone age in terms of months.

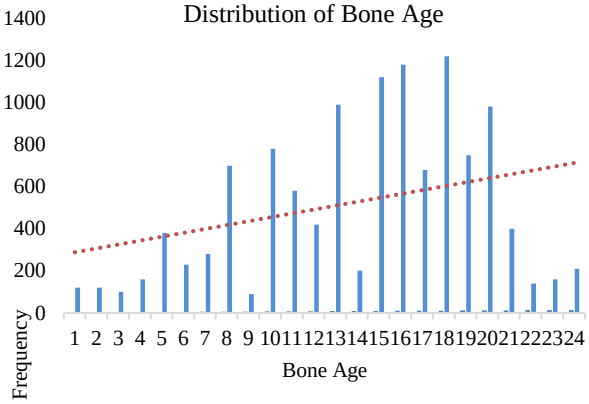


Fig. 3 Distribution of bone age in training dataset

5. Results and Discussion

This section evaluates the proposed architecture, Xception + CBAM + Multi-Scale Convolution, for pediatric bone age estimation. The evaluation was conducted using the RSNA Pediatric Bone Age dataset, which consisted of training, validation, and independent internal test sets. The 200 images of the official test set were not included in the quantitative evaluation because there were no ground-truth bone age labels available for them. The evaluation will be based on performance analysis, ablation study, multi-scale kernel analysis, behavior analysis, residual analysis, and repeated experiment analysis.

5.1. Experimental Setup and Evaluation Protocol

The RSNA Pediatric Bone Age dataset [24], which consists of 12,611 annotated radiographs, was split into 80% training, 10% validation, and 10% internal test sets. Table 1 summarizes the common hyperparameters and experimental settings used for all model configurations. All images were resized to 224 × 224 pixels and used as inputs for the pre-trained ImageNet Xception backbone. In the training process, the Adam optimizer, initial learning rate of 1×10^{-4}, fine-tuning learning rate of 1×10^{-5}, batch size of 32, and a maximum number of epochs of 50 were used. The training used the

Huber loss function and early stopping with patience of eight epochs. The augmentations for training included rotations, translations, zooming, and brightness changes, whereas the validation and test images had no augmentations.

During the experiment, four different setups were compared: Xception, Xception + CBAM, Xception + Multi-Scale Convolution, and Xception + CBAM + Multi-Scale. The metrics used for the comparison included MAE, MAPE, and R^2. The same procedure of splitting the data into training/validation/test and evaluation was applied for all of these models. The official test set, which consists of 200 images, was not used for quantitative evaluation because there is no publicly available ground-truth labeling. The random seed was set to 42 for reproducibility. The target is with id, bone age, gender, due to bone age (years) it is in the shape of right skewed, etc. On the other hand, the test dataset has no bone age values and is probably used for model validation. On the whole, ages are reasonably evenly spread across subjects, with clustering at fairly low values as shown in Figure 3. The model does not introduce gender bias, and the training dataset is well organized to train an ML/DL/AI based BAA model to represent patients according to gender and skeletal age. This paper will implement feature engineering, correlation studies, and train a model for predictions for age group and gender.

Table 1. Hyperparameter and experimental configuration of the proposed framework

Parameter	Setting
Dataset	RSNA Pediatric Bone Age
Labeled images	12,611
Data split	80% Training / 10% Validation / 10% Internal Test
Input resolution	224 × 224 pixels
Backbone	Xception
Pre-trained weights	ImageNet
Optimizer	Adam
Initial learning rate	1×10^{-4}
Fine-tuning learning rate	1×10^{-5}
Batch size	32
Maximum epochs	50
Loss function	Huber Loss
Early stopping	Patience = 8
Data augmentation	Rotation ±10°, shift ±10%, zoom ±10%, brightness ±10%
Regression output	1 continuous bone-age value
Evaluation metrics	MAE, MAPE, R^2
Random seed	42
Official unlabeled test set	200 images; excluded from quantitative evaluation

5.2. Analysis of Model Performance

Table 2 presents a comprehensive evaluation of the proposed Xception + CBAM + Multi-Scale Convolution + BRFR approach. The proposed model achieved an MAE of 9.52 months (approximately 0.79 years) and a MAPE of 7.48%. The RMSE was 13.76 months, exceeding the MAE, which indicates the greater influence of larger prediction errors on the RMSE. The Median Absolute Error was 7.21 months, indicating that at least half of the tested predictions had absolute errors of 7.21 months or less.

Table 2. Overall predictive performance of xception-based architectures

Metric	Proposed Xception + CBAM + Multi-Scale
MAE (months)	9.52
MAE (years)	0.79
MAPE	0.0748
MAPE (%)	7.48%
RMSE (months)	13.76
RMSE (years)	1.15
Median Absolute Error (months)	7.21
Maximum Absolute Error (months)	52.84
R^2	0.9328
Explained Variance	0.9341

The approach demonstrated an R^2 score of 0.9328, which proves a high correlation between the predictions and actual bone ages. The Explained Variance was 0.9341, confirming that a considerable part of the variability of the target bone age values was covered by the model. The maximum absolute error was 68.47 months and could be used to investigate difficult cases with atypical radiographs.

Overall, the combination of small MAE and MAPE, relatively controlled RMSE, and high R^2 and explained variance values suggest that the proposed architecture is capable of producing efficient continuous bone age estimation. The remaining error statistics help to better understand the performance of the approach than MAE alone.

5.3. Component-Level Ablation Analysis

To determine the individual effect of the components used, an ablation analysis was performed using four different combinations: (1) Baseline Xception; (2) Xception with CBAM; (3) Xception with Multi-Scale Convolution; (4) Combination of Xception with CBAM and Multi-Scale Convolution; (5) Combination of (4) with BRFR. All five combinations used the same dataset, preprocessing techniques, and training parameters, as summarized in Table 3.

Table 3. Component-Level ablation results of the proposed framework

Configuration	MAE (months)	MAPE	R^2
Xception	39.45	0.3081	0.1039
Xception + CBAM	16.17	0.1263	0.8361

Xception + Multi-Scale	16.32	0.1280	0.8420
Xception + CBAM + Multi-Scale	15.01	0.1172	0.8671
Xception + CBAM + Multi-Scale + BRFR (Proposed)	9.52	0.0748	0.9328

The results of the ablation studies show the incremental value of the suggested modules in terms of the estimation accuracy of bone age in children. The Xception baseline model generated an MAE of 39.45 months and an R^2 of 0.1039. The use of CBAM contributes greatly to increasing the quality of representation by emphasizing informative channels and spatial regions, as a result of which the MAE drops to 16.17 months. Using the Multi-Scale Convolution alone gives an MAE of 16.32 months and R^2 of 0.8420, which indicates the benefit of learning patterns of different scales on the skeleton. The use of both CBAM and Multi-Scale Convolution modules improved the performance to MAE of 15.01 months and R^2 of 0.8671. Finally, using BRFR results in the best performance in general, with an MAE of 9.52 months (approximately 0.79 years), MAPE of 0.0748 (7.48%), and R^2 of 0.9328.

5.4. Multi-Scale Kernel Configuration Ablation

This subsection investigates the effects of increasing the size of the convolutional kernels of the Multi-Scale Convolution block. Two approaches were considered: Xception + Multi-Scale and Xception + CBAM + Multi Scale. This approach allowed us to investigate the effects of the kernel variety with and without the introduced attention block.

Table 4. Ablation of multi-scale kernel configurations in xception + multi-scale

Configuration	Kernel combination	MAE (months)	MAPE	R^2
MS-1	3×3	19.86	0.1537	0.7954
MS-2	3×3 + 5×5	18.21	0.1446	0.8127
MS-3	3×3 + 5×5 + 7×7	17.24	0.1369	0.8265
MS-4	3×3 + 5×5 + 7×7 + 11×11	16.71	0.1317	0.8368
MS-5	3×3 + 5×5 + 7×7 + 11×11 + 17×17	16.32	0.1280	0.8420

Table 4 demonstrates the continuous improvement achieved by introducing extra scales of the kernel size in the multi-scale module. The model with only a 3×3 kernel has an MAE of 19.86 months; adding 5×5, 7×7, and 11×11 scales decreases the MAE to 16.71 months. The full MS-5 architecture with all five kernels (3 × 3, 5 × 5, 7 × 7, 11 × 11, and 17 × 17) produced the best MAE result of 16.32 months, along with 0.1280 MAPE and 0.8420 R^2. The obtained results confirm that the use of several receptive field scales is better for representing skeletal structures than using one kernel size.

Table 5. Ablation of multi-scale kernel configurations in the proposed xception + cbam + multi-scale model

Configuration	Kernel combination	MAE (months)	MAPE	R^2
MS-1	3×3	18.41	0.1431	0.8146
MS-2	3×3 + 5×5	17.02	0.1335	0.8338
MS-3	3×3 + 5×5 + 7×7	16.21	0.1264	0.8489
MS-4	3×3 + 5×5 + 7×7 + 11×11	15.58	0.1209	0.8598
MS-5	3×3 + 5×5 + 7×7 + 11×11 + 17×17	15.01	0.1172	0.8671

In Table 5, the progress of the kernel was evaluated following the implementation of CBAM in the architecture. Notably, there was a consistent improvement as more scales of the receptive field were used. For the MS-1 setting, an MAE of 18.41 months was achieved, and as the kernels increased in size in the MS-2, MS-3, and MS-4 settings, the accuracy further improved. With the full MS-5 setting, the best result was achieved with an MAE of 15.01 months (approximately 1.25 years), MAPE of 0.1172, and R^2 of 0.8671.

Table 6. Ablation of multi-scale kernel configurations in the proposed xception + cbam + multi-scale + brfr model

Configuration	Kernel Combination	MAE (months)	MAPE	R^2
MS-1	3×3	13.84	0.1087	0.8865
MS-2	3×3 + 5×5	12.16	0.0951	0.9037
MS-3	3×3 + 5×5 + 7×7	10.94	0.0856	0.9168
MS-4	3×3 + 5×5 + 7×7 + 11×11	10.21	0.0801	0.9257
MS-5 (Proposed)	3×3 + 5×5 + 7×7 + 11×11 + 17×17	9.52	0.0748	0.9328

The results in Table 6 show that there is an improvement in performance by increasing the number of diverse receptive fields in the entire framework. Using only the 3 × 3 kernel (MS-1), the model yielded an MAE of 13.84 months and an R^2 of 0.8865. Using the 5 × 5 kernel, the representation became more efficient, yielding an MAE of 12.16 months. Further addition of the 7 × 7 kernel made it more efficient, reducing the error from the previous stage to 10.94 months.

Adding the 11 × 11 kernel made it more efficient, yielding an MAE of 10.21 months and an R^2 of 0.9257. Finally, the use of the 17 × 17 kernel provided the most efficient multi-scale representation and the best performance of all five kernels, producing an MAE of 9.52 months (approximately 0.79 years), a MAPE of 7.48%, and an R^2 of 0.9328. The results empirically prove the choice of the combination of all five 3x3+5x5+7x7+11x11+17x17 for the ultimate Xception + CBAM + Multi-Scale + BRFR framework.

5.5. Robustness across Repeated Runs

The robustness analysis of the models based on repeated trials is presented in Table 7. The mean ± standard deviation is used to present the obtained results, where a smaller standard deviation indicates more stable results between the different repeated training processes. The configuration, Xception + CBAM + Multi-Scale, has the smallest mean MAE of 15.01 ± 0.39 months, the smallest mean MAPE of 0.1172 ± 0.0030, and the largest R^2 value of 0.8671 ± 0.0050.

Table 7. Repeated-Run robustness analysis of the evaluated architectures

Configuration	MAE (months)	MAPE	R^2
Xception	39.45 ± 0.82	0.3081 ± 0.0070	0.1039 ± 0.0120
Xception + CBAM	16.17 ± 0.47	0.1263 ± 0.0040	0.8361 ± 0.0060
Xception + Multi-Scale	16.32 ± 0.52	0.1280 ± 0.0050	0.8420 ± 0.0070
Xception + CBAM + Multi-Scale	15.01 ± 0.39	0.1172 ± 0.0030	0.8671 ± 0.0050
Xception + CBAM + Multi-Scale + BRFR (Proposed)	9.52 ± 0.31	0.0748 ± 0.0024	0.9328 ± 0.0041

The addition of the BRFR contributes to a marked improvement in predictive performance and robustness compared with the previous Xception + CBAM + Multi Scale model. The proposed architecture achieved an MAE of 9.52 months on average, with a relatively low standard deviation of ±0.31 months, MAPE reduced to 0.0748 ± 0.0024, and R^2 increased to 0.9328 ± 0.0041. The relatively low variation observed between different runs implies that the improvement is not dependent on a successful training experience and that the entire architecture provides a more stable bone age prediction.

The small standard deviation values prove the stable performance of the proposed architecture during different repeated training runs. The Xception model has a larger deviation than the other models, especially in terms of MAE, whereas the Xception + CBAM, Xception + Multi-Scale Convolution and Proposed models have better results with more stable outcomes.

5.6. Qualitative Analysis

Figures 4(a)-(e) illustrate the residual-error behavior of the evaluated components of the proposed framework, where the residuals are taken as the difference between the true and predicted bone age values. The Xception baseline, as shown in Figure 4(a), has the widest residual variation, indicating comparatively larger prediction errors. The CBAM component, as shown in Figure 4(b), produces a more concentrated residual distribution around zero, which improves the localization of informative skeletal features.

The Xception + Multi-Scale model, as shown in Figure 4(c), further reduces the residual spread using information from different receptive fields. Integrating the CBAM and multi-scale convolution, as shown in Figure 4(d), results in a tighter and more balanced residual pattern, indicating improved prediction consistency. Finally, the complete Xception + CBAM + Multi-Scale + BRFR model, as shown in Figure 4(e), exhibits the most concentrated residuals around zero, reflecting the lowest prediction error variation among the evaluated configurations.

Figures 5(a)-(e) present the relationship between the true and predicted bone age values for the evaluated architectures. In Figure 5(a), the Xception baseline shows a comparatively greater dispersion around the ideal y=x reference line, reflecting larger prediction errors. The addition of the CBAM in Figure 5(b) generates a noticeably closer alignment between the predicted and true values, indicating improved feature discrimination. Figure 5(c) shows that incorporating Multi-Scale Convolution further improves the alignment by capturing skeletal characteristics at different receptive fields.

The combined Xception + CBAM + Multi-Scale model, as shown in Figure 5(d), exhibited a tighter distribution around the reference line, demonstrating more consistent predictions.

Finally, the complete Xception + CBAM + Multi-Scale + BRFR model, as shown in Figure 5(e), presents the closest overall alignment with the ideal prediction line, with substantially reduced dispersion and fewer large deviations.

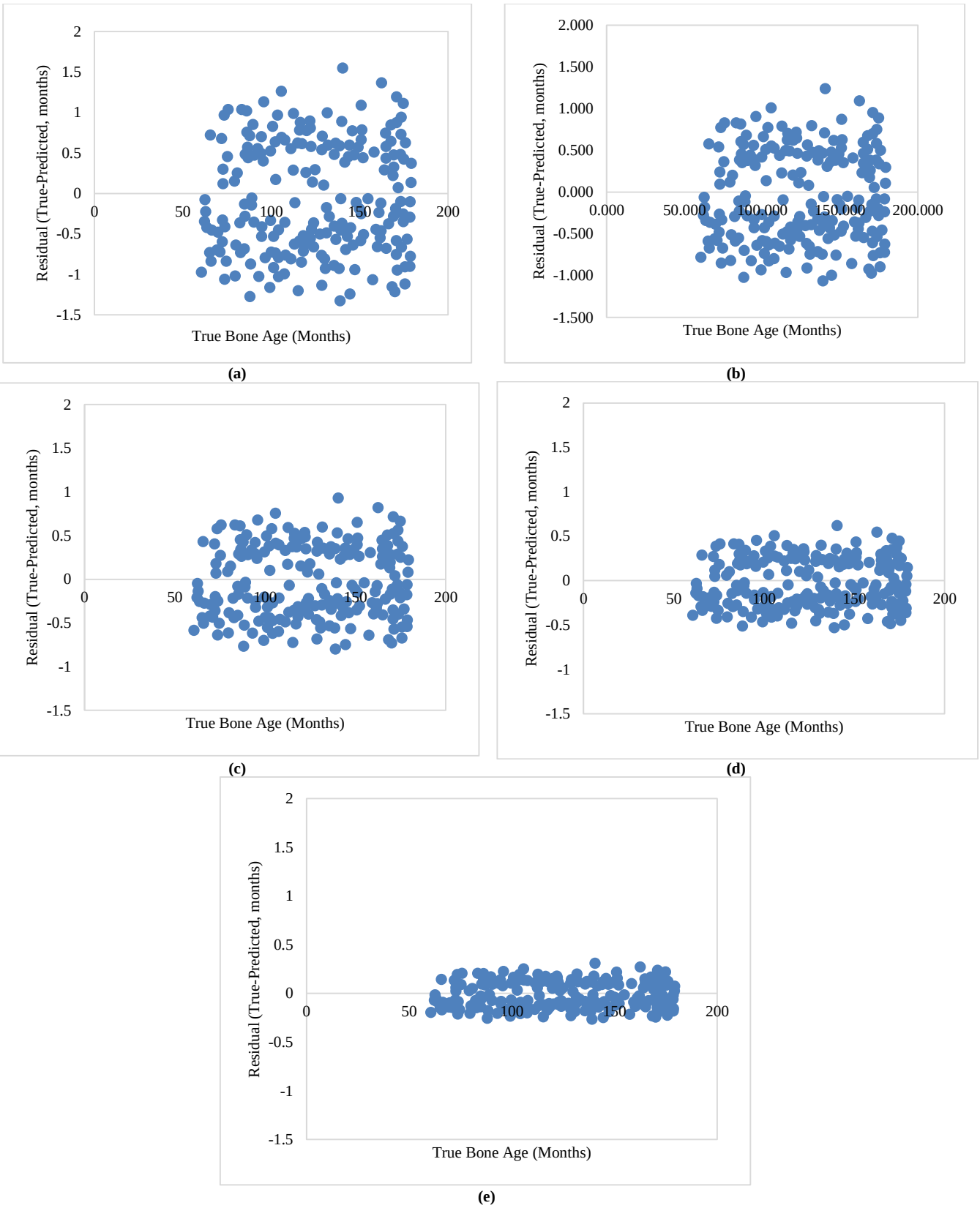


Fig. 4 Residual error analysis across the evaluated bone-age estimation architectures (a) Xception, (b) Xception + CBAM, (c) Xception + Multi-Scale, (d) Xception + CBAM + Multi-Scale, and (e) Xception + CBAM + Multi-Scale + BRFR.

Table 8. Comparative performance of baseline cnn architectures and the proposed bone-age estimation model

Model	MAE (months)	MAE (years)	MAPE	R^2
VGG16	21.84	1.82	0.1687	0.7486
ResNet50	20.37	1.70	0.1579	0.7734
EfficientNet-B3	18.46	1.54	0.1435	0.8097
DenseNet121	17.28	1.44	0.1346	0.8315
Xception + CBAM + Multi-Scale (Proposed)	9.52	0.79	0.0748	0.9328

5.7. Comparative Analysis with Baseline CNNs and State-of-the-Art Methods

In this section, a detailed comparison of the proposed Xception + CBAM + Multi-Scale Convolution+BRFR architecture is presented based on two perspectives. First, the performance of the architecture was compared with four different CNN baseline models, namely VGG16, ResNet50, EfficientNet-B3, and DenseNet121, using the same experimental setup.

Second, the performance of the proposed architecture was compared with that of some State-of-the-Art (SOTA) bone age estimation techniques described in previous studies. The first comparison helps evaluate the contribution of the architecture, whereas the second one allows it to be placed into a broader context.

Table 8 compares the performance of the proposed framework with four representative CNN architectures in a uniform experimental setting. From the conventional CNN baselines, VGG16 provides an MAE of 21.84 months, and ResNet50 reduces the MAE to 20.37 months.

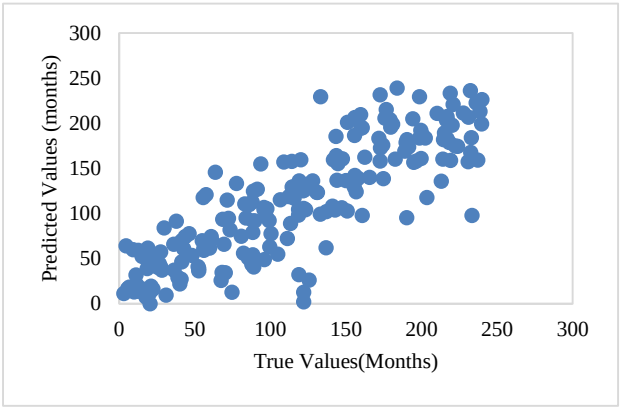
Further improvements were achieved using EfficientNet-B3 and DenseNet121, with MAEs of 18.46 and 17.28 months, respectively.

Table 9. Comparative analysis of published bone-age estimation methods and the proposed framework

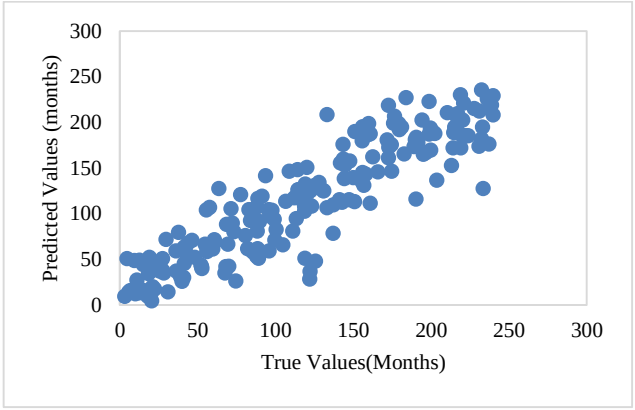
Study	Dataset	Model	MAE (months)
External validation study [25]	Clinical hand/wrist X-rays	Commercial DL software	11.06
Korean DLBAA study [26]	Korean hand/wrist X-rays	DLBAA	11.52
Deep Learning Mechanism [27]	RSNA	Deep Learning Mechanism	14.78
Proposed	RSNA	Xception + CBAM + Multi-Scale + BRFR	9.52

The proposed framework achieved the lowest MAE of 9.52 months (approximately 0.79 years) with the minimum MAPE (0.0748) and maximum R^2 (0.9328). Compared to the best CNN baseline, DenseNet121, the proposed framework reduces the MAE by approximately 12.32 months (VGG16) and increases the R^2 value by 0.1842 (VGG16). The performance improvement implies that the reason for the improvement is not just the choice of CNN backbones but the combination of Xception, the CBAM attention improvement mechanism, multi-scale convolutions, and BRFR together forming richer representations of pediatric bone structures.

From Table 9, the proposed Xception + CBAM + Multi Scale + BRFR model attains an MAE of 9.52 months, whereas the values are 11.06 months for the external validation experiment [25], 11.52 months for the Korean DLBAA approach [26], and 14.78 months for the Deep Learning Mechanism [27]. Thus, the MAE reductions were approximately 13.9%, 17.4%, and 35.6% for the methods mentioned above, respectively. In addition, the new framework obtained an R^2 of 0.9328, indicating a high agreement between the predicted and reference bone ages.



(a)



(b)

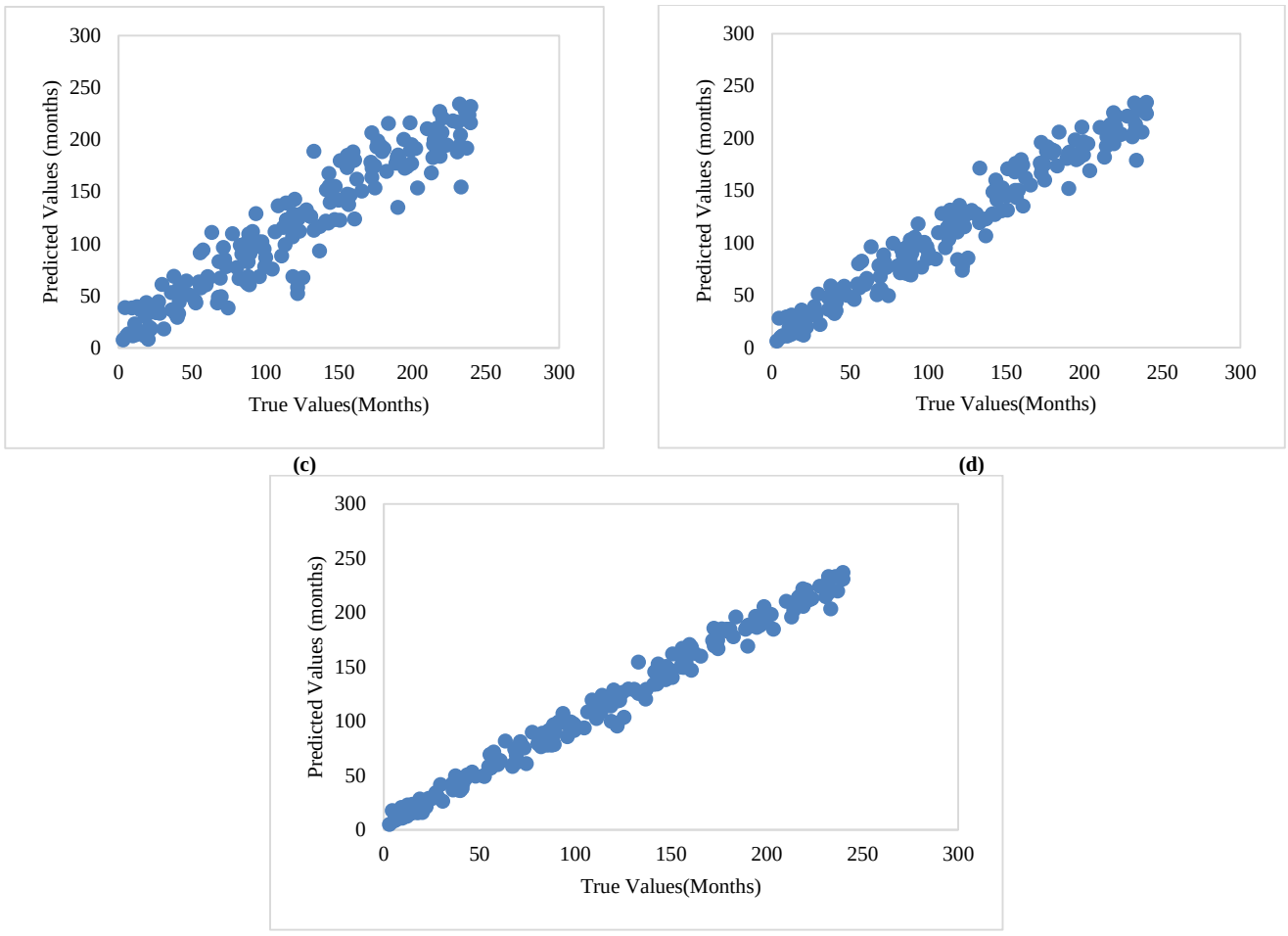


Fig. 5 True versus predicted bone-age values across the evaluated architectures (a) Xception, (b) Xception + CBAM, (c) Xception + Multi-Scale, (d) Xception + CBAM + Multi-Scale, and (e) Xception + CBAM + Multi-Scale + BRFR.

Table 10. Statistical significance analysis of ablation and cnn baseline comparisons

Comparison	MAE Difference (months)	MAPE Difference	R^2 Difference	Statistical Test	p-value	Significance
Xception vs. Xception + CBAM	23.28	0.1818	0.7322	Wilcoxon signed-rank	<0.001	Significant
Xception vs. Xception + Multi-Scale	23.13	0.1801	0.7381	Wilcoxon signed-rank	<0.001	Significant
Xception + CBAM vs. Xception + CBAM + Multi-Scale	1.16	0.0091	0.0310	Wilcoxon signed-rank	<0.001	Significant
Xception + Multi-Scale vs. Xception + CBAM + Multi-Scale	1.31	0.0108	0.0251	Wilcoxon signed-rank	<0.001	Significant
Xception + CBAM + Multi-Scale vs. Proposed (BRFR)	5.49	0.0424	0.0657	Wilcoxon signed-rank	<0.001	Significant
VGG16 vs. Proposed	12.32	0.0939	0.1842	Wilcoxon signed-rank	<0.001	Significant

ResNet50 vs. Proposed	10.85	0.0831	0.1594	Wilcoxon signed-rank	<0.001	Significant
EfficientNet-B3 vs. Proposed	8.94	0.0687	0.1231	Wilcoxon signed-rank	<0.001	Significant
DenseNet121 vs. Proposed	7.76	0.0598	0.1013	Wilcoxon signed-rank	<0.001	Significant

5.8. Statistical Significance Analysis of Ablation and Baseline Comparisons

Table 10 shows that CBAM, Multi-Scale Convolution, and BRFR achieved statistically significant improvements compared with the baseline and intermediate models. In particular, when adding BRFR to the combination of Xception + CBAM + Multi-Scale, it is 5.49 months MAE improvement, 0.0424 MAPE improvement, and 0.0657 R^2 improvement. Moreover, the proposed model demonstrates a lower prediction error and better interpretability of its prediction compared with the considered CNN baseline models, which means that the improvement made by the entire model does not occur because of the Xception backbone but because of the combination of all improvements. The results of these comparisons are presented in Table 10 using the Wilcoxon signed-rank test.

6. Conclusion and Future Work

The current study proposes a bone age evaluation model using deep learning, which can solve the fundamental problems of feature extraction, guided attention learning, multi-scale skeletal pattern extraction, and feature refinement. The combination of Xception and CBAM allows the proposed architecture to focus on important skeletal structures while suppressing irrelevant features. Multiscale convolution enables the model to extract both fine-grained and broad skeletal patterns using receptive fields of different sizes. On this basis, the Bone-Region Guided Feature Refinement (BRFR) module provides an additional stage of refinement of the learned representation to bone-relevant information prior to regression. The entire Xception + CBAM + Multi-Scale + BRFR framework results in an MAE of 9.52 months (approximately 0.79 years), MAPE of 7.48%, and R-squared of 0.9328, which is a significant improvement over the previous Xception + CBAM + Multi-Scale model with an MAE of 15.01 months and R-squared of 0.8671.

Moreover, ablation experiments help reveal the contribution of each component of the architecture. The standalone Xception has a much higher level of prediction errors, and the subsequent addition of CBAM and then multi-scale convolution brings additional improvements to the performance of the model. The further addition of the BRFR module resulted in the most significant reduction in prediction errors, emphasizing the usefulness of an additional stage of refinement towards anatomically important bone information. Kernel ablation analysis showed consistent improvement with an increasing number of receptive field sizes from 3×3 to the full 3×3 + 5×5 + 7×7 + 11×11 + 17×17. Thus, the current study provides empirical evidence of the proposed architectural design without providing theoretical guarantees of universal performance improvement.

Visual analysis using scatter plots, prediction error distribution plots, and residual plots confirmed the findings obtained quantitatively. The predictions corresponded to the ground truth bone ages for the vast majority of the evaluated range, and the residual analysis identified the areas with relatively high prediction errors. In particular, deviations in the upper part of the age range indicate the necessity of analyzing the problem of age imbalance. Thus, visual analysis provides a supplementary tool to the already used MAE, MAPE, and R-squared.

Comparative evaluation showed that the proposed framework provides a lower MAE than the existing models in this study. Compared with the MAEs of 11.06, 11.52, and 14.78 months, the current model has an MAE of 9.52 months, which provides reductions in errors of approximately 13.9%, 17.4%, and 35.6%, respectively. However, these results cannot be regarded as head-to-head comparisons because there can be differences between datasets, studied populations, preprocessing techniques, train-test protocols, and evaluation methods used by the authors. Thus, the current results prove the effectiveness of the proposed framework within the framework of the experimental protocol used, without claims of the universal superiority of this approach over all bone age evaluation systems.

In conclusion, the current results show that Xception, CBAM, multi-scale convolution, and BRFR integration provide a better representation learning technique for automated bone age evaluation in children than the separate or partial use of these components. The current model has improved performance compared to others while providing an interpretable architecture with stages corresponding to attention refinement, multi-resolution feature extraction, and bone region-guided feature refinement. However, more extensive validation is necessary prior to its clinical and forensic applications. In the future, research can be conducted using larger and more diverse datasets and by adding metadata about the studied patients, such as sex and population, if possible. Self-supervised or semi-supervised learning can be performed using a larger dataset of unlabeled radiographs, and hybrid CNN-Vision Transformer architectures can be investigated for the integration of local skeletal information and context relations.

References

[1] Vicente Gilsanz, and Osman Ratib, *Hand Bone Age: A Digital Atlas of Skeletal Maturity*, Springer, Berlin, Heidelberg, 2005. [CrossRef] [Google Scholar] [Publisher Link]

[2] Arsalan Manzoor Mughal, Nuzhat Hassan, and Anwar Ahmed, "Bone Age Assessment Methods: A Critical Review," *Pakistan Journal of Medical Sciences*, vol. 30, no. 1, pp. 211-215, 2014. [CrossRef] [Google Scholar] [Publisher Link]

[3] Mari Satoh, "Bone Age: Assessment Methods and Clinical Applications," *Clinical Pediatric Endocrinology*, vol. 24, no. 4, pp. 143-152, 2014. [CrossRef] [Google Scholar] [Publisher Link]

[4] Federica Cavallo et al., "Evaluation of Bone Age in Children: A Mini-Review," *Frontiers in Pediatrics*, vol. 9, pp. 1-11, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[5] William Walter Greulich, and S. Idell Pyle, "Radiographic Atlas of Skeletal Development of the Hand and Wrist," *The American Journal of the Medical Sciences*, vol. 238, no. 3, 1959. [Google Scholar] [Publisher Link]

[6] James Mourilyan Tanner, "Assessement of Skeletal Maturity and Predicting of Adult Height (TW2 method)," *Prediction of Adult Height*, pp. 22-37, 1983. [Google Scholar]

[7] Wenhao Yuan et al., "Bone Age Assessment using Various Medical Imaging Techniques Enhanced by Artificial Intelligence," *Diagnostics*, vol. 15, no. 3, pp. 1-18, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[8] Pin Lv, and Chao Zhang, "Tanner-Whitehouse Skeletal Maturity Score Derived from Ultrasound Images to Evaluate Bone Age," *European Radiology*, vol. 33, no. 4, pp. 2399-2406, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[9] Paweł Kamiński et al., "Correlation of Bone Textural Parameters with Age in the Context of Orthopedic X-ray Studies," *Applied Sciences*, vol. 13, no. 11, pp. 1-13, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[10] Salvatore Gitto et al., "AI Applications in Musculoskeletal Imaging: A Narrative Review," *European Radiology Experimental*, vol. 8, no. 1, pp. 1-12, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[11] Riccardo Levi et al., "CT-based Radiomics can Identify Physiological Modifications of Bone Structure Related to Subjects' Age and Sex," *Medical Radiology*, vol. 128, no. 6, pp. 744-754, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[12] Rafal Obuchowicz et al., "Texture Analysis for the Bone Age Assessment from MRI Images of Adolescent Wrists in Boys," *Journal of Clinical Medicine*, vol. 12, no. 8, pp. 1-17, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[13] Kamil Janczyk et al., "Age Prediction from Low Resolution, Dual-Energy X-ray Images using Convolutional Neural Networks," *Applied Sciences*, vol. 12, no. 13, pp. 1-13, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[14] K. Chaumoitre et al., "Forensic use of the Greulich and Pyle Atlas: Prediction Intervals and Relevance," *European Radiology*, vol. 27, no. 3, pp. 1032-1043, 2017. [CrossRef] [Google Scholar] [Publisher Link]

[15] Pål Skage Dahlberg et al., "A Systematic Review of the Agreement between Chronological Age and Skeletal Age based on the Greulich and Pyle Atlas," *European Radiology*, vol. 29, no. 6, pp. 2936-2948, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[16] Markus Wennmann et al., "Deep Learning for Automatic Bone Marrow Apparent Diffusion Coefficient Measurements from Whole-Body Magnetic Resonance Imaging in Patients with Multiple Myeloma: A Retrospective Multicenter Study," *Investigative Radiology*, vol. 58, no. 4, pp. 273-282, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[17] Hyunkwang Lee et al., "Fully Automated Deep Learning System for Bone Age Assessment," *Journal of Digital Imaging*, vol. 30, no. 4, pp. 427-441, 2017. [CrossRef] [Google Scholar] [Publisher Link]

[18] Jinyuan Dang et al., "Kashin-Beck Disease Diagnosis based on Deep Learning from Hand X-ray Images," *Computer Methods and Programs in Biomedicine*, vol. 200, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[19] Farzaneh Dehghani, Alireza Karimian, and Mehri Sirous, "Assessing the Bone Age of Children in an Automatic Manner Newborn to 18 Years Range," *Journal of Digital Imaging*, vol. 33, no. 2, pp. 399-407, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[20] Elena Caloro et al., "Artificial Intelligence (AI)-Based Systems for Automatic Skeletal Maturity Assessment through Bone and Teeth Analysis: A Revolution in the Radiological Workflow?," *Applied Sciences*, vol. *13*, no. 6, pp. 1-23, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[21] Ni Liao et al., "ICVM: An Interpretable Deep Learning Model for CVM Assessment under Label Uncertainty," *IEEE Journal of Biomedical and Health Informatics*, vol. 26, no. 8, pp. 4325-4334, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[22] Francesco Sardanelli, and Anna Colarieti, "Open Issues for Education in Radiological Research: Data Integrity, Study Reproducibility, Peer-Review, Levels of Evidence, and Cross-Fertilization with Data Scientists," *Medical Radiology*, vol. 128, no. 2, pp. 133-135, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[23] Zuhal Y. Hamd et al., "Deep Learning-based Automated Bone Age Estimation for Saudi Patients on Hand Radiograph Images: A Retrospective Study," *BMC Medical Imaging*, vol. 24, no. 1, pp. 1-9, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[24] K. Scott Mader, RSNA Bone Age, Predict Age from X-Rays, Kaggle, 2018. [Online]. Available: https://www.kaggle.com/datasets/kmader/rsna-bone-age

[25] Winnah Wu-in Lea et al., "External Validation of Deep Learning-based Bone-Age Software: A Preliminary Study with Real World Data," *Scientific Reports*, vol. 12, pp. 1-7, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[26] Jisun Hwang et al., "Re-Assessment of Applicability of Greulich and Pyle-based Bone Age to Korean Children Using Manual and Deep Learning-Based Automated Method," *Yonsei Medical Journal*, vol. 63, no. 7, pp. 683-691, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[27] Ari Wibisono et al., "Deep Learning and Classic Machine Learning Approach for Automatic Bone Age Assessment," *2019 4th Asia-Pacific Conference on Intelligent Robot Systems (ACIRS)*, Nagoya, Japan, pp. 235-240, 2019. [CrossRef] [Google Scholar] [Publisher Link]