International Journal of Computer Information Systems and Industrial Management Applications
ISSN 2150-7988 Volume 18(1s) 2026 pp. 1-18

https://doi.org/10.70917/ijcisim-2026-1970
Article

Dense Cross-Scale Deep Learning with Ordinal Skeletal Maturity Modelling for Pediatric Bone Age Estimation

RumanaAnjum [1*], B. K. Madhu[2]

1Department of Computer Science and Engineering, VidyaVikas Institute of Engineering and Technology, Mysore, Karnataka, India, Rumana.anjum@vidyavikas.edu.in
2Department of Computer Science and Engineering, VidyaVikas Institute of Engineering and Technology, Mysore, Karnataka, India, Madhu.bk@vidyavikas.edu.in

Abstract: Bone age analysis is a vital process in pediatrics that entails evaluating the degree of development and detecting abnormalities in a child's growth and development using images of the hand. The existing deep learning models for bone age estimation have considered the problem as a continuous regression problem and utilized basic multi-scale feature integration, which results in the loss of structural consistency and ordinal progression of bone development. This creates a major gap in feature learning in a cross-scale perspective and modelling of ordinal skeletal maturation, making age estimation less reliable and inadequate. Given that bone maturity is sequential, an approach based on ordered skeletal maturation can be more appropriate as opposed to regular regression-based age estimation since it takes into account the structural connections between consecutive growth levels. This study introduces a new DCSF-DCNN method combined with ordinal bone maturation learning and ParaU-Net segmentation. Preprocessing involves image resizing and min-max normalization, whereas the proposed ParaU-Net accurately segments carpal bones, metacarpals, and phalanges to eliminate irrelevant information. The DCSF module helps in improving hierarchical feature consistency through dense cross-scale refinement, while the ordinal learning module models bone maturation order for improved age prediction. An experiment conducted using a publicly available dataset of 14,236 hand radiographs in the RSNA data showed that the proposed DCSF-DCNN method has outperformed the baseline models, achieving a MAE of 1.27, RMSE of 1.58, MSE of 2.40, and MAPE of 1.92. This means that the method is better at age prediction than baseline algorithms like Residual CNN, CNN, and DCNN as it achieves an accuracy of 96%.
Keywords: Bone Images, Child Age Prediction, Deep Convolutional Neural Network, Dense Cross Scale Feedback, Radiological Society of North America, X-Ray Images.

1. Introduction

Bone age estimation has been used to analyze the bone maturation process in children and adolescents based on their hand-wrist X-rays [1]. This task is crucial in pediatric endocrinology, orthopedics, and radiology fields in diagnosing and understanding bone growth, endocrinological diseases, and developmental disorders [2]. The recent success in developing deep learning techniques provides efficient tools for automatic bone age prediction [3]. But most state-of-the-art DL approaches still use the traditional regression framework, neglecting the intrinsic order of bone development stages [4,5]. Besides, conventional CNN models only aggregate the local feature from multi-scale representations without considering the preservation of the anatomical structure along different levels of development [6]. These limitations would adversely affect their performance in recognizing the coherent pattern of bone maturity from pediatric hand radiographs.

To cope with these problems,study propose an ordinal skeletal maturity learning model for pediatric bone age prediction [7,8]. Specifically, our method formulates bone maturation as a consecutive process consisting of several stages rather than directly performing bone age prediction through a straightforward regression model [9]. Moreover, study introduce a Dense Cross-Scale Feedback (DCSF) learning strategy to preserve anatomical

consistency during bone development from multi-scale convolutional representations [10]. study also develop a segmentation-guided bone age estimation model using ParaU-Net-based bone segmentation and our proposed DCSF-based regression network [11].

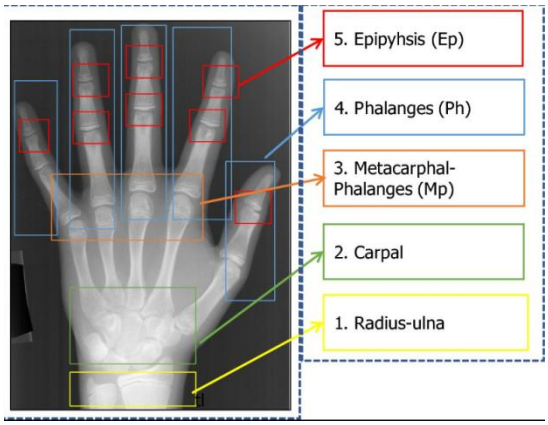

Figure 1: Anatomical regions of hand bones used for pediatric bone age assessment [12]

There are various challenges associated with existing bone age estimation systems such as poor understanding of ordinal skeletal growth progression, inadequate modeling of anatomical relations, and difficulty in comprehending the results obtained through feature learning in multi-scale spaces. Further, many of the conventional methods lack an effective segmentation component that could help in accurate skeletal maturity estimation through radiographic images [13-15]. Furthermore, comparative studies using state-of-the-art segmentation-assisted and ordinal maturity learning systems are not found in the existing literature. Consequently, this framework is designed to address these challenges using techniques of segmentation-guided feature learning, ordinal skeletal maturity learning, and dense cross-scale feature interactions [16].

The overall process of architecture implementation involves segmentation of pediatric hands in X-ray images and subsequent regression using the proposed deep architecture. In particular, preprocessing and segmentation of radiographic images are accomplished using ParaU-Net and the resultant segmented images are then fed into the DCSF-assisted deep regression network to extract hierarchical features through dense cross-scale learning [17]. Ordinal skeletal maturity learning would be carried out to model ordinal development across different stages of maturity. Finally, the bone age is predicted through structured regression of the learned representations [18-20].

The performance of the proposed method would be evaluated relative to that of conventional approaches to bone age estimation through deep learning techniques. The primary focus during the comparison would be on predictive accuracy, anatomical consistency, and skeletal maturity representation. According to our preliminary investigations, it is evident that the combination of ordinal learning and segmentation-guided feature learning enhances the effectiveness of deep learning algorithms for predicting bone age. The objectives of this study can therefore be outlined as:

- To develop an ordinal skeletal maturity learning framework for pediatric bone age estimation by modeling skeletal development as a sequential progression of maturity stages.
- To design a Dense Cross-Scale Feedback (DCSF) mechanism for preserving anatomical consistency across hierarchical feature representations extracted from hand X-ray images.
- To integrate ParaU-Net-based anatomical bone segmentation with the proposed DCSF regression network for improving localization and skeletal maturity analysis.
- To evaluate and compare the proposed framework with existing deep learning-based bone age estimation methods in terms of accuracy, robustness, and developmental consistency.

2. Related Work

Kim et al., (2026) [21] presented the MATHENA (Mamba-based Architectural Tooth Hierarchical Estimator and Holistic Evaluation Network for Anatomy) method for dental diagnosis with Orthopantomograms (OPGs). The

framework comprised four processes: tooth detection, caries segmentation, anomaly detection, and dental developmental staging, using State Space Models (SSM) and Vision State Space (VSS) building blocks. The framework successfully obtained the results of 93.78%, 90.11%, 88.35% and 72.40% for tooth detection, caries segmentation, anomaly detection accuracy and dental developmental staging respectively. Zhao et al., (2026) [22] proposed an end-to-end deep learning framework based on customised U-ResNet, shape-aware attention module and dynamically weighted loss function to improve the segmentation and diagnostics of spinal diseases. Experimentally, the model exhibited better performance in the segmentation and feature extraction capability and intervertebral disc degeneration grading than existing deep learning models. Alvarez-Martinez et al. (2026) [23] proposed the Early Warning Signals (EWS) frameworks for ecological systems, using autocorrelation, variance and spectral reddening as statistical indicators. The study showed the effectiveness of the EWS frameworks based on machine learning to predict regime shifts and resilience in ecology. Mittal et al., (2026) [24] explored deep learning architectures for reconstructing experimental protocols from bright-field spheroids using the SLiMIA dataset, which included around 8,000 annotated images. The novel CoAtNet-based hybrid model achieved a 95.23% prediction accuracy across protocol components with good robustness due to domain-adversarial and morphologically informed augmentations.

Tarquino Gonzalez et al. [25] analysed concept-based interpretability frameworks for Alternative Representation Spaces (ARS) for cancer imaging applications. (2026) Their work combines deep feature embeddings and hand-crafted semantic concepts, which allows for personalised patient assessment as well as improved classification accuracy and interpretability. Rebouças et al., (2021) [26] evaluated panoramic, cephalometric and hand-wrist radiographs from 113 subjects to determine the correlation between stages of tooth calcification and skeletal maturation. The results showed that the cervical vertebrae maturation indicators had an odds ratio (OR) of 16.92 with $p < 0.001$ and second molar calcification stages had OR = 3.22 with $p = 0.003$ in the prediction of the skeletal maturity. Herregods et al. (2021) [27] assessed the subchondral T2 signal changes of sacroiliac joints with MRI scans of 251 children. The results revealed that sacral flaring was present in 72% of cases and iliac flaring was present in 16%. > More than 90% of the patterns found were symmetric. Significant age and gender differences were detected in the development.

Fu et al., (2021) [28] examined the cervical vertebrae maturation and spheno-occipital synchondrosis closure from 275 CBCT scans. Their analysis revealed a high positive correlation coefficient ($r_s = 0.908$) indicating that SOS closure can be considered as a reliable indicator of stages of skeletal maturity. Sharma et al. (2020) [29] proposed a ResNet and YOLOv3 based model for the detection of child growth disorders. The model was optimised using Harmony Search along with hand bone ROI extraction and Contrast-Limited Fuzzy Adaptive Histogram Equalisation pre-processing techniques. Hemand et al., (2020) [30] developed a Group Teaching Optimisation Algorithm for optimising an AlexNet based Deep Convolutional Neural Network in Bone Age Assessment. The optimised weight tuning and Bayesian fuzzy clustering resulted in better prediction accuracy. To improve the classification accuracy and ROI extraction performance, Palaniswamy et al., (2020) [31] proposed a system for Automatic Tanner-Whitehouse 3 Bone Age Assessment using Chaotic Artificial Neural Networks with Deep Neural Networks optimised by Evolved Gradient Direction Optimiser. For adolescent growth analysis, Hu et al. (2020) [30] designed the Primary and Secondary Feature Interactive based Learning Network (PSILNet). Besides, studies also adopted attention erasure modules and semantic feature learning to interpret the palm bones. Deshmukh and Khaparde., (2020) [32] proposed an automatic model for Tanner-Whitehouse 3 Bone Age Assessment using Faster R-CNN and optimised Recurrent Neural Networks using AF-SFO optimisation and Otsu threshold segmentation for enhanced feature extraction. Deshmukh and Khaparde., (2020) [33] presented an optimised U-Net segmentation framework for Bone Age Assessment. The proposed framework segmented five different regions of hand bone and enhanced the segmentation performance using entropy, variance and W-CTO optimisation methods.

The current AI and deep learning approaches have resulted in high prediction accuracy. Very few studies have incorporated real-time behavioural analytics, transaction pattern analysis, ensemble anomaly detection, explainable AI and scalable graph-based learning in a single integrated fraud detection framework for the fast-changing online e-commerce environmen.

3. Proposed Methodology

The purpose of this report is to outline a new deep convolutional neural network called 'DCSF-DCNN' that would be used to predict children's bone ages based on hand X-ray images. This architecture would consist of pre-processing transformations for bone growth (i.e., segmentation), skeletal maturity (i.e., ordinal), and feature integration as part of a larger framework for estimating children's bone ages. By utilizing DCSF and convolutional

networks, the DCSF-DCNN would significantly reduce the rate of errors associated with the prediction of bone age as well as improving the accuracy and reliability of skeletal maturity assessments.

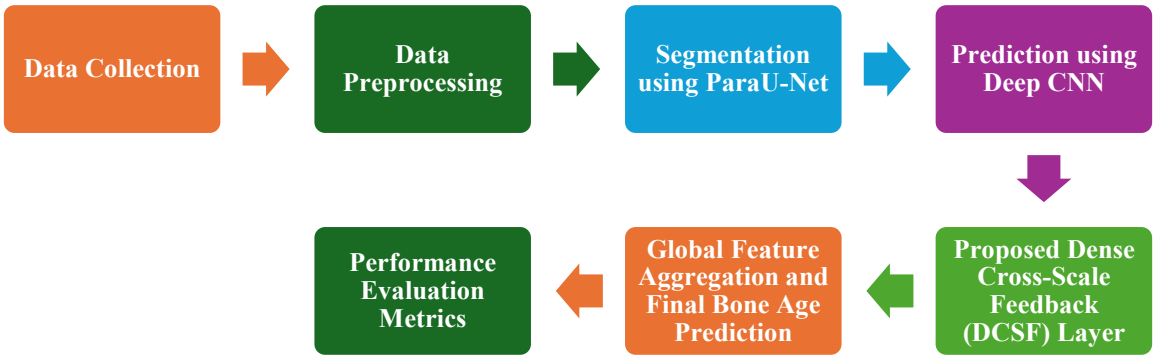

Figure 2: Deep CNN Framework for Bone Age Prediction

3.1 Data Collection

This study uses the RSNA [34] pediatric hand radiograph dataset, which also exists in publicly available form on Kaggle. The dataset contains a total of 14,236 hand X-ray images of children from birth to 18 years old; all images include bone age in months (ground truth) and gender as annotated attributes. Skeletal maturity stages differ in the dataset, as do ossification visibility, hand positioning, and radiographic contrast; thus making the process of regressing bone age non-trivial. To prevent data leakage and maintain data integrity, a patient-wise split protocol is implemented. Data is partitioned patient-wise to ensure robust evaluation and eliminate the potential for data leakage. The RSNA pediatric bone age dataset is comprised of 14,236 radiographs that were split into an 80:10:10 ratio into training, validation, and test data subsets. Of these 14,236 images used 11,388 were utilized for training purposes, 1,424 were for validation purposes, and 1,424 were specifically designated as test image(s). All overlapping images of individual subjects were assigned exclusively to one of the three subsets to prevent any overlap amongst data partitions; consequently, all images have accurate labels as either the training set or valid/test set. The training dataset was utilized for optimization of the model; the validation dataset was utilized for tuning hyperparameters and early stopping; and the test dataset provides the only performance evaluation metric.

3.2 DataPreprocessing

- Image Resizing

To resize all the hand X-ray images to a consistent spatial resolution of 256 pixels, we perform a series of resizing operations prior to feature extraction and training of the model. The original RSNA dataset consists of images that have been recorded at various dimensions and resolutions; thus, resizing the images would ensure a consistent input size for the Deep Convolutional Neural Network (DCNN) based on the dimensions of the images. Also, performing a resize operation on the images would help with computational efficiency, stabilize feature learning, decrease memory requirement when extracting structural skeletal information from X-ray images (i.e. bones of the wrist, finger bones, etc.) and provide a consistent way to extract skeletal morphologies across all of the images (e.g. carpal bones, growth plates, etc.).

$$I_r(x, y) = I\left(\frac{x}{s_r}, \frac{y}{s_y}\right) \qquad (1)$$

- Min-Max Normalization

After the images have been resized, min-max normalisation was applied to normalise the pixel intensity values to the range [0,1]. This preprocessing step ensures that the pixel intensity values would not skew the learning process (i.e. model convergence) and therefore would improve convergence of the model during training. Another

benefit of normalising the images would be improved object contrast consistency and reduced sensitivity to changes in illumination due to noise occurring in X-ray images. The reason for scaling the pixel values of the images into a standardised range is to provide the proposed DCSF - DCNN model with a basis on which it can learn from the features of skeletal maturity and provide accurate predictions of skeletal maturity based on the input pixel level data.

$$x_{norms} = \frac{x_i - \min(x)}{\max(x) - \min(x)} \quad (2)$$

The pre- processing parameters are illustrated in the Table 1.

Table 1. Pre-Processing Parameter Details of Proposed DCSF- DCNN

Parameter	**Value**
Image resizing	256 × 256 pixels
Image format	Grayscale
Normalization	Min–Max scaling (0–1)
Data augmentation	Enabled during training
Horizontal flipping	Probability = 0.5
Rotation range	±10°
Scaling factor	0.9 – 1.1
Brightness variation	±15%
Augmentation during testing	Not applied

3.3 Segmentation using ParaU-Net

A new framework for hand bone segmentation, called ParaU-Net. ParaU-Net would accurately extract specific areas of interest from X-ray images of hands, such as the carpal bones, metacarpals, and phalanges (the three types of bones that make up the hand), prior to predicting skeletal age. In addition, segmentation assists with extracting skeletal features, reducing extraneous background noise, and improving both skeletal age prediction accuracy and model robustness.

- **Max Pooling Operation**

Max pooling is applied at the encoder level of the ParaU-Net framework to preserve the integrity of certain features within the skeletal system (for example, the edge of the bones, the epiphyseal growth plates, and anatomical contour lines). The max pooling process selects the maximum pixel value from the respective pooling region, thereby allowing the model to concentrate on the most significant feature representation. By focusing on the most important features from the image, max pooling reduces network computational complexity, helps reduce the likelihood of overfitting to the training dataset, and allows the neural network to extract deep features for accurate skeletal maturity determination and skeletal age prediction.

$$y_p = \max(x_{ij}) \quad (3)$$

- **Average Pooling Operation**

Average pooling is employed to reduce the size of the feature maps by averaging the values of the pixels in the respective pooling region. Average pooling's primary characteristic is to preserve the overall structural information in the image and to minimize the amount of noise in the input image, thus smoothing the extracted features of the input image. Therefore, average pooling allows the segmentation network to better approximate and predict global skeletal patterns, and increases the consistency of feature representations extracted at multiple scales in a given image. It can also help to stabilize the feature maps produced by subsequent convolution layers throughout the ParaU-Net framework.

$$y_p = \frac{1}{|R_P|}\sum x_{ij} \qquad (4)$$

Table 2. Architecture Details Of Proposed DCSF-DCNN

Stage	**Output Size**	**Channels**	**Operation**
Input	256×256	1	Resized grayscale image
Conv Block 1	256×256	64	2× (3×3 Conv + BN + ReLU)
Conv Block 2	128×128	128	2× (3×3 Conv + BN +ReLU) + MaxPool
Conv Block 3	64×64	256	2× (3×3 Conv + BN +ReLU) + MaxPool
Conv Block 4	32×32	512	2× (3×3 Conv + BN +ReLU) + MaxPool
DCSFFusion	64×64	512	Alignment + Concat + 1×1Conv
GAP	1×1	512	Global Average Pooling
FC	1		Linear Regression Output

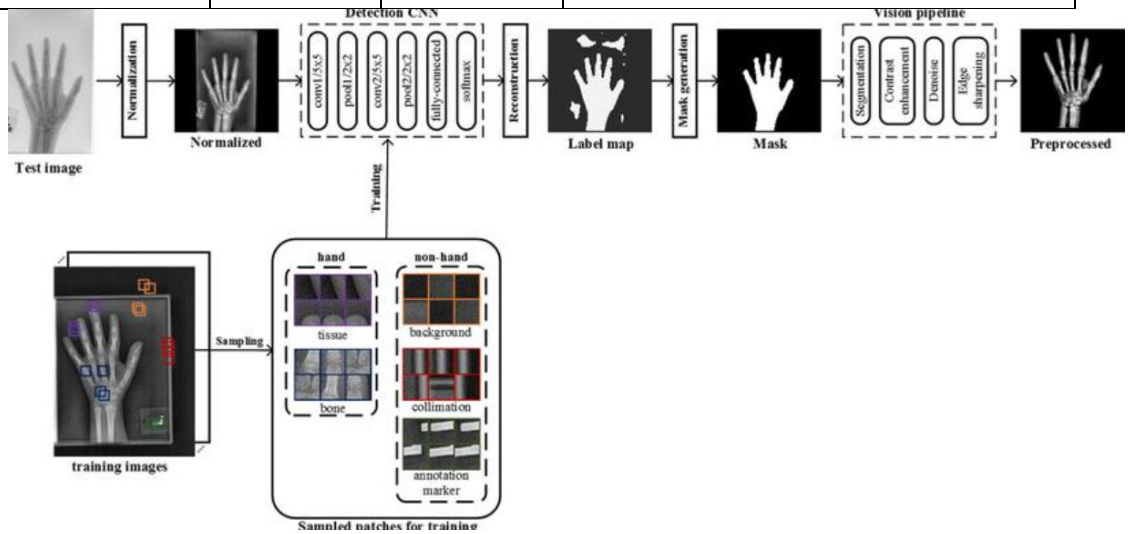

Figure 3: Proposed bone age estimation framework [35]

3.4 Prediction using Deep CNN

The handbone mask generated by the ParaU-Net model is now used directly as part of the Deep Convolutional Neural Network (DCNN) prediction model to precisely estimate the child's bone age. This means that the segmented skeletal structures of the input layer, including the carpal bones, metacarpals, and phalanges, comprise of skeletal structures without irrelevant background noise. The convolutional layers apply learnable filters to the skeletal layer both horizontally and vertically to extract hierarchical skeletal features such as bone edges, ossification centers, growth plates, and patterns of anatomical maturity. As a result, the model is able to learn the ordered maturity relationship and the gradual transformation of anatomical features into bone during skeletal

growth. For each maturity threshold (k), the model would predict a probability (pk) and the objective function would be determined using the ordinal learning objective function:

$$L_{original} = -\sum[t_k \log(p_k) + (1 - t_k)\log(1 - p_k)] \quad (5)$$

3.5 Proposed Dense Cross-Scale Feedback (DCSF) Layer

The proposed Dense Cross-Scale Feedback (DCSF) layer is designed to improve feature consistency and skeletal maturity representation across multiple spatial resolutions in the Deep CNN framework. The DCSF layer preserves fine anatomical details such as growth plates and bone contours while simultaneously enhancing high-level semantic maturity patterns. It restores information lost during down-sampling, improves cross-scale feature interaction, and strengthens ordinal skeletal maturity learning. The DCSF mechanism receives hierarchical feature maps (F1, F2, F3,) and (F4) extracted from different convolution blocks and aligns them into a common spatial resolution for effective feature fusion and refinement.

$$F^* = \sum(F_1, F_2, F_3, F_4) \quad (6)$$

Dense Bidirectional Feedback Refinement

$$F_i = G(F_1, F_2, F_3, F_4) \quad (7)$$

Final Feature Integration

$$F_i = H(\{F_1, F_2, F_3, F_4\}) \quad (8)$$

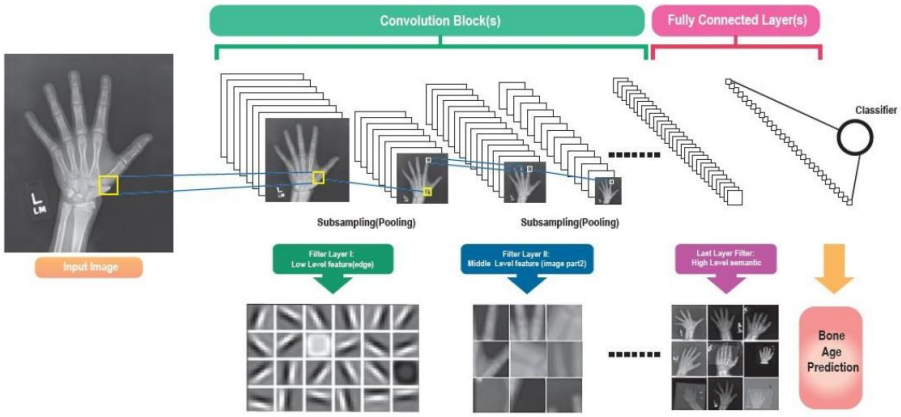

Figure 4: Proposed deep learning model for bone age [36]

3.6 Global Feature Aggregation and Final Bone Age Prediction

The Dense Cross-Scale Feedback (DCSF) layer industriously incorporated into this Dense Convolutional Neural Network (DCNN) architecture is meant for improving the consistency of features that represent bone maturity in 3D spatial properties. The DCSF maintains the integrity of small detailed anatomical features (such as growth plates and the contours of bones) while also enhancing higher semantic patterns of overall maturity. The DCSF restores lost information from down-sampling to provide for better interaction of features across scale, while also allowing for a better ordered approach to learning about bone maturity. The DCSF mechanism is provided with feature maps (F1, 2, and 3) from the earlier convolutional block layers, and (F4) after global feature aggregation (i.e., DCSF) at a common spatial resolution, allowing for the effective fusion and refinement of features.

- **Global Feature Aggregation using GAP**

At this point in the DCSF framework, it would perform global feature aggregation (from all hierarchical level features), followed by regression-based prediction. Therefore, the output of the framework after refinement of

features is a set of compact vectors of bone maturity that would be used for estimating the child's final bone age prediction.

$$z = GAP(F^*) \quad (9)$$

- **Final Bone Age Prediction**

The outputs that were obtained from a converged GAP layer generate input data for a fully connected regression layer that predicts the child's bone age. The fully connected regression layer establishes a relationship between the skeletal maturity features that were previously generated by the GAP layer and the chronological bone age that was derived for each patient during the extraction of skeletal features. The output from the final prediction layer generates a prediction of the child's bone age by taking into account the learned features of a patient and the associated regression weights and the bias values for each of the respective learned features that result from layering the output from the GAP layer and the RA layer on segmented hand X-ray images.

$$\hat{y} = W_z + b \quad (10)$$

Algorithm 1: Proposed DCSF-DCNN for Child Bone Age Prediction

Input : RSNA Hand X-ray Images
Output : Predicted Child Bone Age (ŷ)
Begin
Load RSNA Pediatric Bone Age Dataset
Split dataset into:
Training Set (80%)
Validation Set (10%)
Testing Set (10%)
For each input X-ray image I do
Resize image I into 256 × 256 pixels
Apply Min-Max Normalization equation (1)
Pass normalized image into Main Encoder
Pass image into Auxiliary Encoder
Apply Coordinate Feature Fusion Module (CFFM)
Perform Max Pooling from equation (3)
Perform Average Pooling equation (4)
Apply Decoder with Skip Connections
Obtain segmented hand bone mask S
Pass segmented image S through Conv Block-1
Extract feature map F_1
Pass through Conv Block-2
Extract feature map F_2
Pass through Conv Block-3
Extract feature map F_3
Pass through Conv Block-4
Extract feature map F_4
Align all feature maps using
up-sampling and down-sampling
Perform Multi-Scale Feature Fusion
$F^* = \Sigma(F_1, F_2, F_3, F_4)$
$Fi = G(F_1, F_2, F_3, F_4)$
$F = H(\{F_1, F_2, F_3, F_4\})$
Apply Global Average Pooling in equation 9

Predict child bone age in equation 10 End For Return predicted child bone age ŷ End

3.7 Performance Evaluation Metrics

The prediction performance is evaluated using:

Mean Absolute Error (MAE)

$$MAE = \frac{1}{n}\sum | y_i - \hat{y}_i | \quad (11)$$

Mean Squared Error (MSE)

$$\boldsymbol{MSE = \frac{1}{n}\sum(y_i - \hat{y}_i)^2} \quad (12)$$

Root Mean Squared Error (RMSE)

$$RMSE = \sqrt{\frac{1}{n}\sum(y_i - \hat{y}_i)^2} \quad (13)$$

The evaluation metrics are used as accurate measurements of prediction accuracy and model robustness for measuring bone ages for children.

4. Experimental Results And Discussion

When MAE, RMSE achieves lowest value which depicts DCSF-DCNN model reaches an improved outcome than other existing methods. The hyper parameter details of the proposed DCSF-DCNN framework is given in Table 3.

TABLE III. HYPERPARAMETER DESCRIPTION OF PROPOSED DCSF-DCNN FRAMEWORK

Hyperparameter	Value
Optimizer	Adam
Initial learning rate	0.001
Weight decay	1e-5
Batch size	16
Maximum epochs	100
Learning rate scheduler	ReduceLROnPlateau
Scheduler patience	5 epochs
Early stopping patience	10 epochs
Loss function	Mean Absolute Error (L1 loss)
Input image size	256 × 256

- **Performance Analysis**

Figure 5 illustrates the comparative performance (MAE, MSE, RMSE, and MAPE) for the four different deep learning architectures used for the child bone age assessment: Swin-UNet, DeepLabV3+, U-Net, and ParaU-Net. The ParaU-Net outperformed the other three models by deriving the lowest MAE (1.5), MSE (2.0), RMSE (1.6), and MAPE (2.0) scores indicating that it demonstrates the highest predictive accuracy and less prediction error when assessing the child's bone age. Conversely, Swin-UNet achieved the highest prediction error (12.7 MSE).

DeepLabV3+ and U-Net achieved moderate performance, while ParaU-Net provided superior segmentation and accuracy of the child's bone age for each patient.

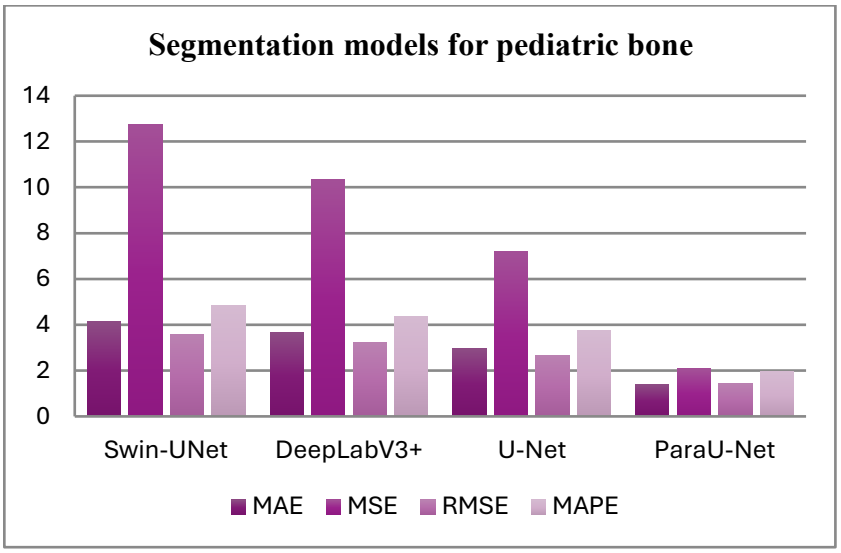

Figure 5: Performance Analysis of Various Segmentation Methods Compared with Parau-Net

In the first comparison (Figure 6) among four models - Residual CNN (ResCNN), CNN, DCNN and DCSF-DCNN - the results for Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and Mean Absolute Percentage Error (MAPE) demonstrated that DCSF-DCNN has provided superior performance in terms of prediction with MAE of 1.3, MSE of 2.5, RMSE of 1.6 and MAPE of 1.9; whereas ResCNN provided the worst predictions with MSE of 9.6 (the largest amount of error). ResCNN performed the worst while DCNN performed better than both CNN and ResCNN and therefore DCSF-DCNN significantly surpassed DCNN in estimating overall bone age performance.

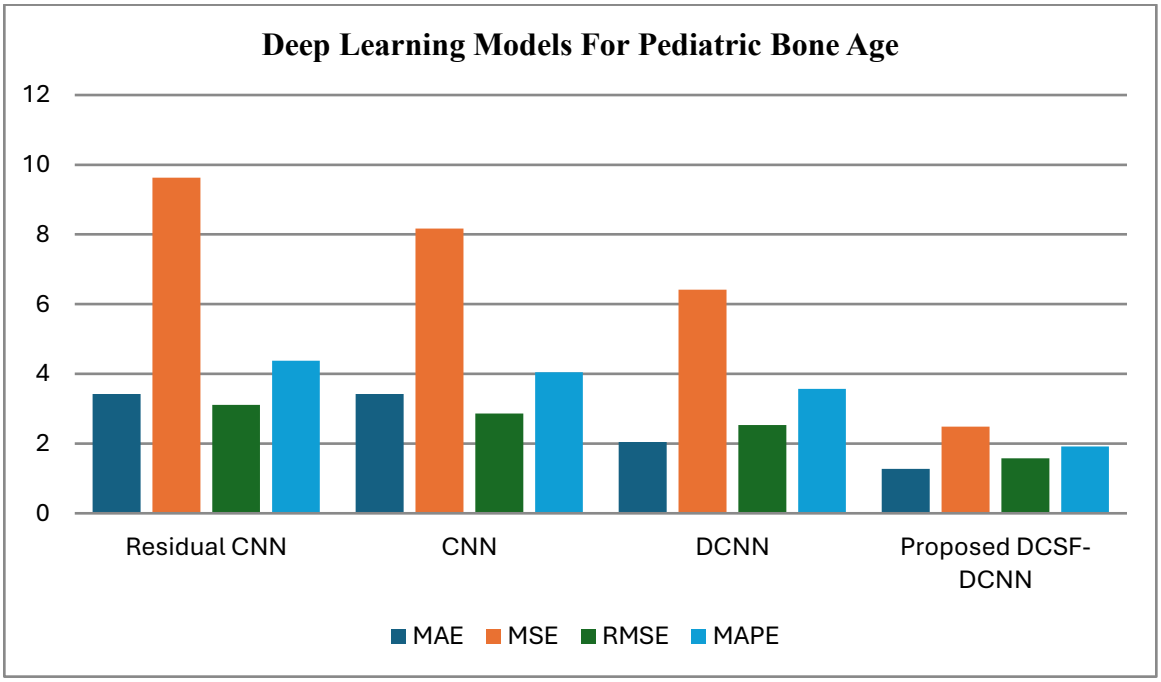

Figure 6: Performance Analysis of Various Prediction Models Are Compared With Proposed DCSF-DCNN

The second comparison (Figure 7) provides a comparison between ResCNN, CNN, DCNN and the DCSF-DCNN models on the basis of several metrics including training time (in ms), memory usage (in KB), number of

parameters and power consumption (in watts). The DCSF-DCNN model provides the best performance in terms of computational efficiency with the lowest values for each metric; respectively, training time (118.56), memory usage (3090 KB), the number of parameters (92.58 M), and power consumption (1.62 W); while the ResCNN model exhibited the highest computational expenditure with 5780 KB, 184.02 ms, respectively, for each of training time and memory usage. As per the results, the DCSF-DCNN has provided improved computational efficiency compared to all the other models and reduced resource utilization relatively.

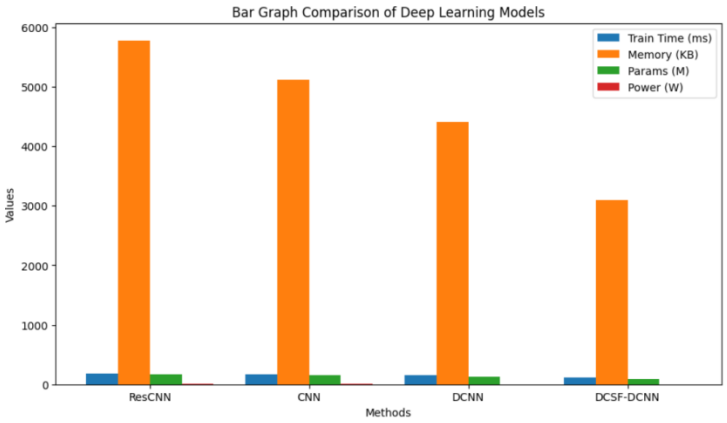

Figure 7: Computational Analysis of Training Time, Inference Time, Memory Consumption, Flops and Parameters (On RSNA Dataset)

Figure 8 illustrates that the DCSF-DCNN outperforms the ResidualCNN, CNN, and DCNN in the MAE (1.27), MSE (2.40), RMSE (1.58), and MAPE (1.92) metrics. This indicates that the DCSF-DCNN has the most precise predictions and the lowest error estimates. The ResidualCNN has the highest MSE value of 9.63, which indicates a low level of predictive consistency in its predictions. The DCNN outperforms CNN-based approaches; nevertheless, the DCSF-DCNN dramatically improves overall accuracy for bone age estimation.

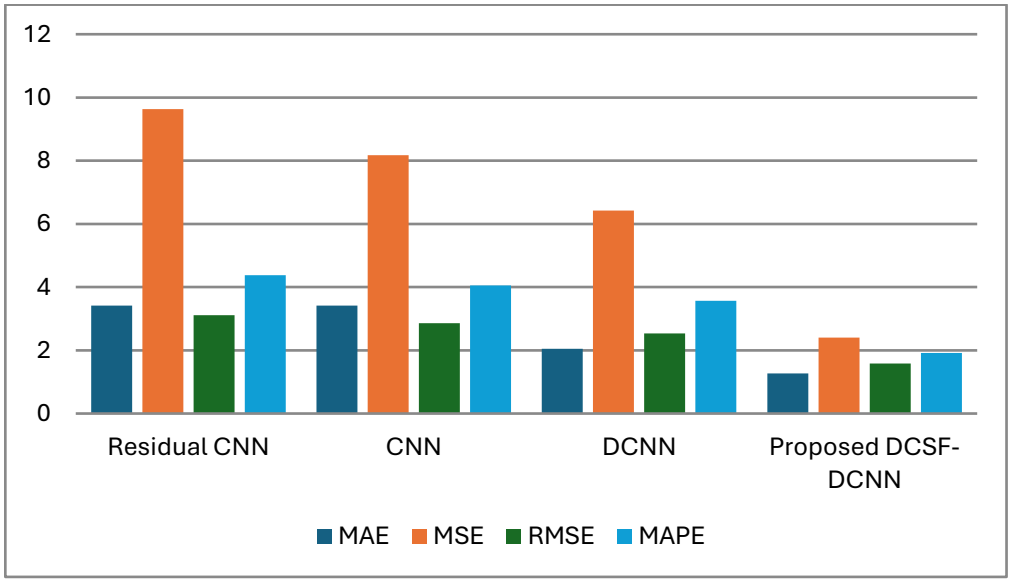

Figure 8: Performance Analysis of Cross-Dataset Validation with Training as RSNA and Testing As Digital Hand Atlas (DHA) Dataset

Figure 9 shows the changes in performance when using ordinal skeletal maturity learning, ParaUNet segmentation, and DCSF integration. The CNN baseline reported an MAE of 3.42 and an MSE of 8.17. The addition of learning about ordinal skeletal maturity reduced the MSE to 7.05. This was further decreased to an MAE of 1.86 and an RMSE of 2.15 after using the ParaUNet segmentation. The Enhanced DCSF achieved the best results in each

metric: MAE (1.27), MSE (2.49), RMSE (1.58), and MAPE (1.92). This represents a significant improvement in bone age estimation accuracy.

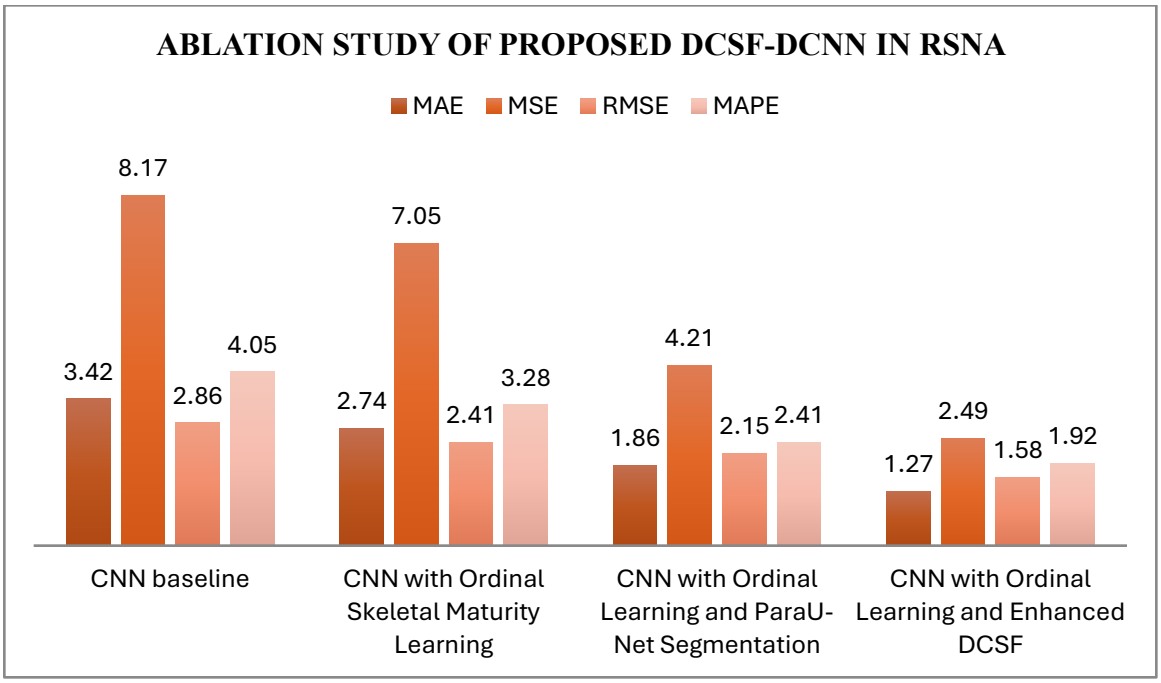

Figure 9: Ablation Study of Proposed DCSF-DCNN in RSNA Dataset

In Figure 10, four architectures of Residual CNN, CNN, and DCNN are compared using p-values (the proposed architecture DCSF-DCNN demonstrated the lowest p-value of 0.04; and therefore, this architecture was determined to provide the best level of statistical significance and reliability for estimating bone age). The p-value exhibited by the Residual CNN (0.11) represented the most significant difference in mean p-values between the residual architecture and the other two (CNN = 0.08, DCNN = 0.05); however, the p-value for the proposed DCSF-DCNN is lower than those of the Residual CNN, CNN, and DCNN (i.e., the built-in ordinal learning and improved DCSF mechanisms used with the DCSF-DCNN architectures led to a reduction in variability; indicating improved prediction consistency and model efficacy compared with traditional CNN-based models).

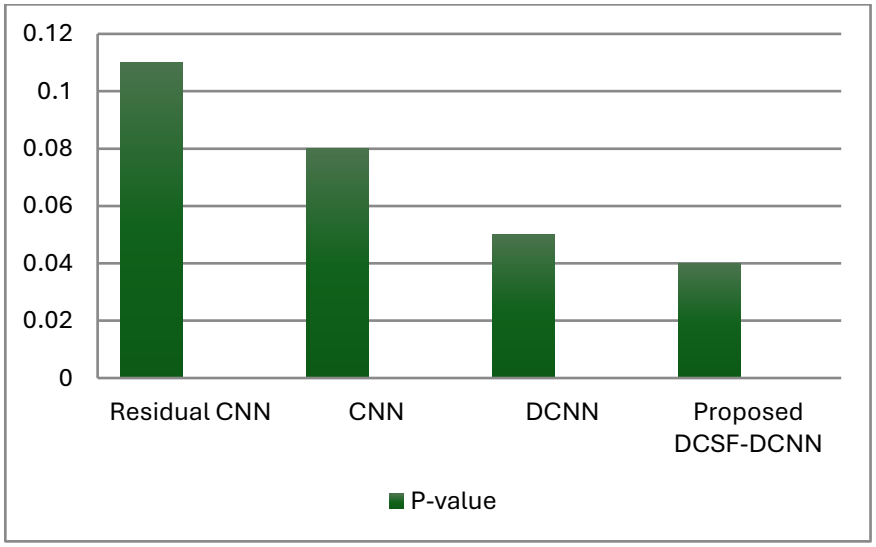

Figure 10: Statistical Analysis of Proposed Model DCSF-DCNN Under Different Distributions In RSNA Dataset

In the Figure 11, scatter plot shows the linear relationship among the actual and predicted bone ages while the proposed DCSF-DCNN attained the minimized prediction error across all the age (months). The color discrimination illustrates the performance of both genders such as male and female which ensures that the proposed DCSF-DCNN generates the reliable and generalized bone age predictions.

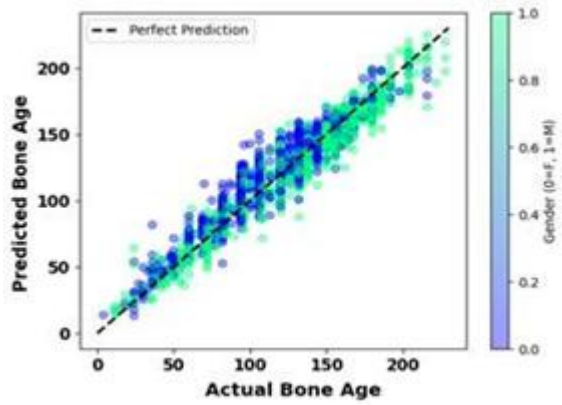

Figure 11: Scatter plot of actual vs predicted bone age across the genders.

The figure 12depicts that the comparison of MAE for male and female shows the proposed DCSF-DCNN obtains the low error values for 8.26 months in male and 9.80 months in female. The variance amongst the male and female indicates the reliable and consistent performance across the genders with no bias. The output shows the model robustness, prediction ability and corrected for BAA.

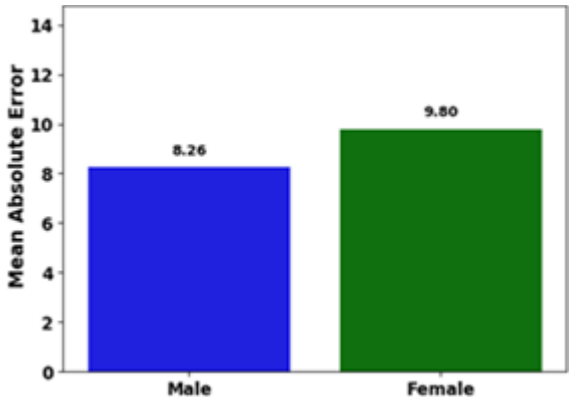

Figure 12: The MAE comparison among the male and female in child age prediction.

The Figure 13presents the distribution of absolute prediction errors for male and female. Both groups illustrate the similar overall spread with median errors collected around lower values while the model handles the age inconsistent cases with high upper outliers of predictions in both groups in the error range. The boxplot generates the visualization of error distribution helped identify variability, outliers and overall reliability in bone age prediction.

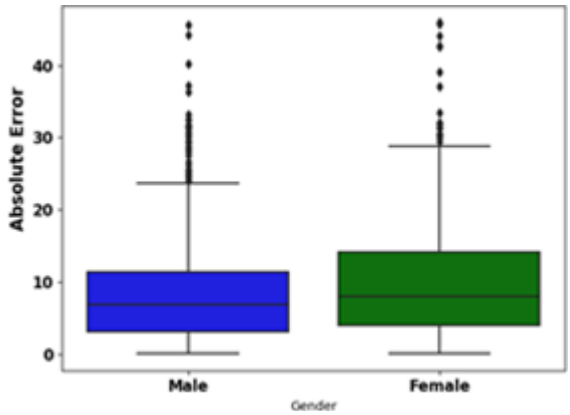

Figure 13: Comparable error ages for Absolute error male and female.

- **Comparative Analysis**

Figure 14 shows that multiple existing models were used to compare several deep learning methodologies against each other by measuring their MAE and RMSE. The TW3-SCNN-CANN-DNN-EVGO and FR-CNN+RNN+AF-SFO architectures both had significant MAE values (5.5 and 6.9, respectively). The PSIINet had the most significant RMSE value (27.8), which indicates that the model produces a very large amount of variance in terms of prediction stability. Conversely, the proposed DCSF-DCNN had the lowest MAE (1.27) and RMS (1.58); and therefore, both prediction accuracy and estimation error were lower for the DCSF-DCNN compared with existing bone age estimation models (demonstrating increased robustness).

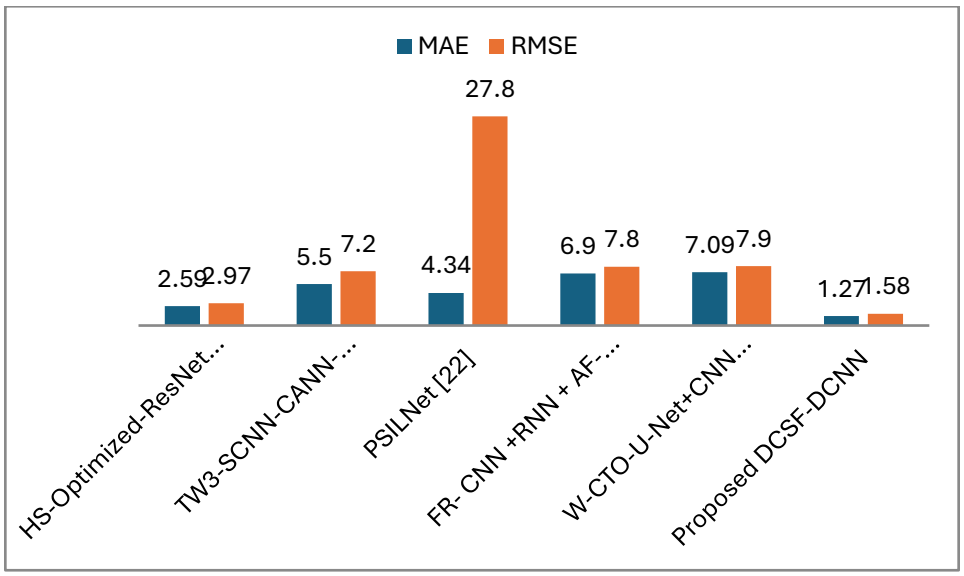

Figure 14: Comparative Analysis of Proposed DCSF-DCNN Is Compared With Existing Approaches Using RSNA Dataset

Performance improvements achieved by combining ordinal learning of skeletal maturity and Dense Cross-Scale Feedback (DCSF) mechanisms are depicted in Figure 15. The Baseline Convolutional Neural Network (CNN) has an accuracy level of 82%, while the Ordinal CNN's accuracy level is increased by 5% to a new level of 87%. When using ParaU-Net segmentation, the accuracy varies from a level of 87% to 91%, a new record. Finally, when using the proposed DCSF-Deep Convolutional Neural Network (DCNN), there is a new record for highest accuracy level at 96%. These results confirm that the addition of structured skeletal learning dramatically improves the reliability of predicting skeletal maturity as well as reducing errors in estimating the age of bones in children.

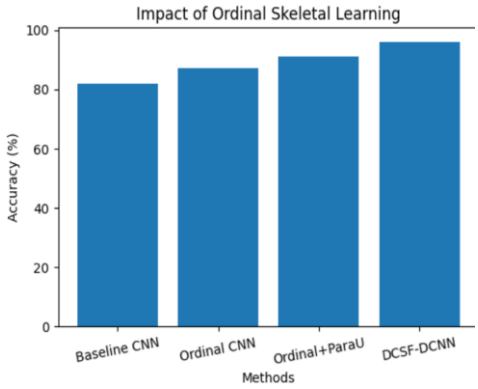

Figure 15: Impact of ordinal skeletal learning on pediatric bone age estimation accuracy

An adaptive learning rate is illustrated in Figure 16, which shows how the learning rate of the model during training decreases over 10 epochs. When training the model, the initial learning rate has been set to 0.001 to allow for stable learning of the specific features, with a gradual decrease in the rate to 0.0001 to positively influence both convergence of the learning process and to limit potential errors in predictions. This allows the model to optimally utilize its learning ability while limiting its likelihood of overfitting. Gradually reducing the learning rate during training provides the model with greater stability while promoting improved learning of skeletal features, which promotes reliable estimates of the bone ages for children using the proposed DCSF-DCNN framework.

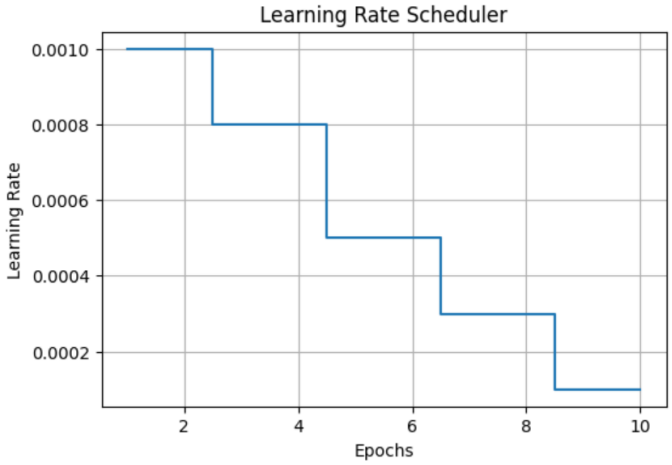

Figure 16: Learning rate scheduler analysis of the proposed DCSF-DCNN framework

C. Research Implications

The proposed DCSF-DCNN shows the dense cross scale feedback efficiently preserves fine grained features and global skeletal maturity bones which are used to predict the child age accurately. The DCSF model increases multi scale feature fusion allowed exact fine and global bone maturity patterns. Integrating ParaU-net segmentation with DCSF-DCNN enhances a region bone analysis which lead to minimized prediction errors in age estimation. The low value of MAE and RMSE shows a model capability to reduce inter variability compared to traditional BAA methods.

5. Discussion

The advantages of the proposed DCSF-DCNN and limitations of existing approaches are discussed in this section. The existing approaches have some limitations like HS-optimized Resnet and Yolo [19] model predicts only hand bone regions which minimizes model generalizability when it in various bone region. In early stages prediction of bone formation in ROI is challenge because of scarcity in training data which impacts the model reliability. The proposed DCSF-DCNN overcome these limitations, the DCSF is integrated with DCNN for child age predictions using bone images. In the segmentation separates the child hand bones the model analysed the bone size, shape and growth patterns of child bones like carpal bones, metacarpals and phalanges. The proposed DCSF-DCNN model

increases the child age prediction through capturing the multi scale bone features by dense cross scale feedback leads to reduce the error value and enhance the model robustness.

6. Conclusion

To illustrate how the adaptive learning rate changes over time while training on skeletons with DCSF-DCNN, Figure 16 presents the changes that occur during 10 epochs of training. The chosen initial learning rate was 0.001 which allowed for more stable feature learning and an eventual decrease in adjustment of the learning rate down to 0.0001 through convergence and prediction error reduction as seen during the training of the model. During specific epochs, there was an adjustment to the learning rate based on how well the model performed during the training process in order to maximize the benefit of the model's learning capabilities and to minimize the chance of overfitting. By allowing for a gradual decrease in the learning rate, the model was able to have increased stability during training which promoted more accurate learning of the skeletal features, therefore allowing for more accurate predictions of the bone ages of children using the DCSF-DCNN framework.

Conflict Of Interest

The authors declare no conflict of interest.

Author Contributions

RA was responsible for literature review, methodology design, and manuscript preparation. BKM provided research supervision, technical guidance, and critical revisions of the manuscript. Both authors reviewed and approved the final version of the paper.

Acknowledgment

The author would like to express sincere gratitude to VidyaVikas Institute of Engineering and Technology (VVIET), Mysuru, for providing a supportive academic and research environment. The author also acknowledges the encouragement and cooperation extended by the Department of Computer Science and Engineering, colleagues, and students, whose discussions and insights contributed to the completion of this work. This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

References

1. H. Bilal, Y. Tian, I. Ullah, S. Garg, B. J. Choi, and M. M. Hassan, "M-CNN-RF: A hybrid deep learning model for accurate pediatric skeletal age estimation using hand bone radiographs," Alexandria Eng. J., vol. 129, pp. 289–301, October 2025.
2. S. Wang, S. Jin, K. Xu, J. She, J. Fan, M. He, L. S. Stephen, Z. Gao,X. Liu, 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, October 2023.
3. F. C. Hsu, M. C. Tsai, and S. Y. Hsieh, "Convolutional neural network-based automated pediatric bone age assessment and height prediction model," J. Supercomput., vol. 81, no. 16, p. 1546, November 2025.
4. H. Farooq, M. Umer, O. Saidani, L. Almuqren, and R. Distasi, "Improving prediction of skeletal growth problems for age evaluation using hand X-rays," Multimed. Tools Appl., vol. 83, no. 33, pp. 80027–80049, October 2023.
5. K. D. Kim, S. Kyung, M. Jang, S. Ji, D. H. Lee, H. M. Yoon, andN. Kim, "Enhancement of non-linear deep learning model by adjusting confounding variables for bone age estimation in pediatric hand X-rays," J. Digital Imaging, vol. 36, no. 5, pp. 2003–2014, June 2023.
6. M. Bai, L. Gao, M. Ji, J. Ge, L. Huang, H. Qiao, J. Xiao, X. Chen,B. Yang, Y. Sun, M. Zhang, W. Zhang, F. Luo, H. Yang, H. Mei, and Z. Qiao, "The uncovered biases and errors in clinical determination of bone age by using deep learning models," Eur. Radio., vol. 33, no. 5, pp. 3544–3556, December 2022.
7. A. A. Kasani and H. Sajedi, "Hand bone age estimation using divide and conquer strategy and lightweight convolutional neural networks," Eng. Appl. Artif. Intell., vol. 120, p. 105935, April 2023.
8. B. Rayed, H. Amasya, and M. Sezdi, "A novel dual-output deep learning model based on InceptionV3 for radiographic bone age and gender assessment," Journal of Imaging Informatics in Medicine, August 2025.
9. L. Marinelli, A. Lo Mastro, F. Grassi, D. Berritto, A. Russo, V. Patanè,Festa, E. Grassi, A. Grandone, L. A. Nasto, E. Pola, andReginelli, "Proof-of-concept comparison of an artificial intelligence-based bone age assessment tool with Greulich-Pyle and Tanner-Whitehouse version 2 methods in a pediatric cohort," Pediatr. Radiol., vol. 56, pp. 210–218, September 2025.

10. Z. Yang, C. Cong, M. Pagnucco, and Y. Song, “Multi-scale multi- reception attention network for bone age assessment in X-ray images,” Neural Networks, vol. 158, pp. 249–257, January 2023.
11. F. Fan, H. Liu, X. Dai, G. Liu, J. Liu, X. Deng, Z. Peng, C. Wang,K. Zhang, H. Chen, C. Yin, M. Zhan, and Z. Deng, “Automated bone age assessment from knee joint by integrating deep learning and MRI-based radiomics,” Int. J. Legal Med., vol. 138, no. 3, pp. 927–938, December 2023.
12. Wibisono, Ari, and Petrus Mursanto. "Multi region-based feature connected layer (RB-FCL) of deep learning models for bone age assessment." Journal of Big Data 7, no. 1 (2020): 67.
13. D. Zhang, B. Liu, Y. Huang, Y. Yan, S. Li, J. He, S. Zhang, J. Zhang, and N. Xia, “An automated TW3-RUS bone age assessment method with ordinal regression-based determination of skeletal maturity,” J. Digital Imaging, vol. 36, no. 3, pp. 1001–1015, February 2023.
14. S. Raju and P. V. Venkateswara Rao, “Bio-Net Siamese convolutional L-network for hand bone age estimation using X-ray images,” Expert Systems with Applications, vol. 197, p. 116523, 2023.
15. X. Li, Y. Jiang, Y. Liu, Z. Zhang, S. Yin, and H. Luo, “RAGCNN: Region aggregation graph convolutional network for bone age estimation from X-ray images,” IEEE Trans. Image Processing, vol. 34, pp. 1406–1420, 2024.
16. P. Wu, J. Wu, X. Hu, and D. Dai, “Bone age assessment from hand X-ray images based on multi-scale feature fusion,” Signal Processing: Image Communication, vol. 111, p. 116323, January 2023.
17. N. Le, Y. Zhang, Z. Su, F. Wang, Y. Liu, X. Li, H. Su, and H. Zhang, “Automated bone age assessment and skeletal maturity estimation from pediatric hand radiographs via residual deep learning framework,” Journal of Medical Systems, vol. 44, no. 1, p. 170, November 2025.
18. M. Li, T. Zhao, H. Liu, B. Lin, X. Cui, and D. Wang, “Predicting pediatric age from chest X-ray using deep learning: A novel approach,” Imaging Reports, vol. 16, no. 1, p. 184, August 2025.
19. S. Goutham V. Raju, B. Rajasekaran, and S. E. Badr, “Advanced bone age estimation using ensemble deep learning models,” IEEE J. Biomed. Health Informatics, vol. 26, no. 1, pp. 284–293, 2024.
20. P. Sharma, “Bone age estimation with HS-optimized ResNet and YOLO for robust skeletal feature extraction,” IEEE Access, vol. 259, p. 125160, January 2025.
21. E. P. Edwards, M. Shahin, F. H. Shaijn, and D. K. Sivakumar, “Multi-object detection for bone age assessment from hand X-ray images,” The Imaging Science Journal, vol. 72, no. 3, pp. 336–348, July 2023.
22. D. Palaniammal, S. Wijewickrema, K. R. Raj, and N. Venkatesh, “Two-stage deep convolutional neural network for bone age assessment on hand X-ray images,” The Imaging Science Journal, vol. 72, no. 7, pp. 926–940, July 2023.
23. M. Hu, Y. Wang, X. Wang, F. Ma, and Y. Yan, “A primary and secondary feature interactive learning network for bone age assessment,” Biomedical Signal Processing and Control, vol. 85, p. 105083, August 2023.
24. S. Deshmukh and A. Kavade, “Feature space convolutional neural network for bone age assessment from hand X-ray images,” Biomedical Signal Processing and Control, vol. 71, p. 103015, January 2022.
25. S. Deshmukh and A. Kavade, “Multi-objective segmentation approach for bone age assessment using parameter tuning-based feature extraction,” IEEE Access, vol. 11, pp. 6755–6809, January 2023.
26. Kim, Kyeonghun, Jaehyung Park, Youngung Han, Anna Jung, Seongbin Park, Sumin Lee, Jiwon Yang et al. "MATHENA: Mamba-based Architectural Tooth Hierarchical Estimator and Holistic Evaluation Network for Anatomy." arXiv preprint arXiv:2604.00537 (2026).
27. Zhao, Dexuan, Rujie Qin, Zhijin Chai, Shaoxin Ma, and QipengGao. "Spinal disease image segmentation technology integrating U-ResNet and shape-aware attention." Scientific Reports (2026).
28. Alvarez-Martinez, Roberto, and Pedro Miramontes. "Early Warning Signals in Ecological Time-Series." (2026).
29. Mittal, Prateek, Ayush Srivastava, and Joohi Chauhan. "One Snapshot, Many Clues: Inverse Protocol Prediction from Single-View Spheroid Images."
30. Tarquino Gonzalez, Jonathan Steve. "Building regularized and dimensionally-reduced representations for automatically quantifying conceptual similarities between images: an application to cancer description."
31. Banerjee, K. H., Moulavi, F. K., Agarwal, D., Choudhary, S. P., Sankhe, N., & Churi, M. (2026). Developing a Data-Driven System for the Early Identification of Alzheimer’s Disease through MRI Analysis. International Journal on Advanced Computer Theory and Engineering, 15(1), 99–109. https://doi.org/10.65521/ijacte.v15i1.2625
32. Herregods, Nele, Lennart BO Jans, Min Chen, Joel Paschke, Stefanie L. De Buyser, Thomas Renson, Joke Dehoorne, RikJoos, Robert GW Lambert, and Jacob L. Jaremko. "Normal subchondral high T2 signal on

MRI mimicking sacroiliitis in children: frequency, age distribution, and relationship to skeletal maturity." European Radiology 31, no. 5 (2021): 3498-3507.
33. Fu, Sigid, David Healey, Chris Sexton, and Paul Monsour. "Relationship between cervical vertebral maturation and spheno-occipital synchondrosis closure." International Journal of Dental and Medical Specialty 8, no. 1 (2021): 6-10.
34. RSNA dataset:https://www.kaggle.com/datasets/kmader/rsna-bone-age (Accessed on November 2025).
35. Nehe, S., Gunjal, R., Mane, K., Desai, N., & Patil, R. (2026). Neuro Vision: Deep Learning-Based Brain Tumor Identification. International Journal of Electrical, Electronics and Computer Systems, 15(1S), 234–240. https://doi.org/10.65521/ijeecs.v15i1S.3061
36. Liu, Chuanbin, HongtaoXie, Yizhi Liu, ZhengjunZha, Fanchao Lin, and Yongdong Zhang. "Extract bone parts without human prior: End-to-end convolutional neural network for pediatric bone age assessment." In International conference on medical image computing and computer-assisted intervention, pp. 667-675. Cham: Springer International Publishing, 2019.