Pediatric Bone Age Assessment: A Comparative Analysis of Regression and Classification Deep Learning Models

PRML Project Report

By
Akshith Pagadala

Abstract

This study explores the application of deep learning for automated pediatric bone age assessment using the RSNA dataset. Two distinct modeling approaches were implemented using a shared custom Convolutional Neural Network (CNN) backbone: a regression model to predict continuous age, and a classification model to categorize developmental stages. The regression approach yielded a Mean Absolute Error (MAE) of **1.235 years** with an R^2 of **0.795**. The classification approach achieved a test accuracy of **78%**, highlighting challenges with class imbalance. This report details the architectural design, preprocessing decisions, and visual analysis of learned representations via Grad-CAM.

1 Preprocessing and Feature Choices

1.1 Data Preparation Strategy

The dataset consists of 12,600 hand radiographs. Specific preprocessing choices were made to optimize for GPU training and model convergence:

- **Label Transformation (Regression):** The original labels (months) were converted to years ($Age_y = Age_m/12$). This scales the target variable to a smaller range (0–20), stabilizing the Mean Squared Error (MSE) loss gradients during training.
- **Label Engineering (Classification):** For the comparative approach, ages were binned into three developmental stages: *Child* ($< 10y$), *Adolescent* (10–18y), and *Adult* ($> 18y$).
- **Input Standardization:** Images were resized to $128 \times 128 \times 3$ to fit larger batch sizes while retaining anatomical details. Pixel values were normalized to $[0, 1]$.
- **Dynamic Augmentation:** Random horizontal flips, brightness shifts, and contrast adjustments were applied through the `tf.data` pipeline.

2 Model Architecture Summary

Both approaches utilize a shared convolutional backbone for feature extraction but diverge at the fully connected heads depending on the task.

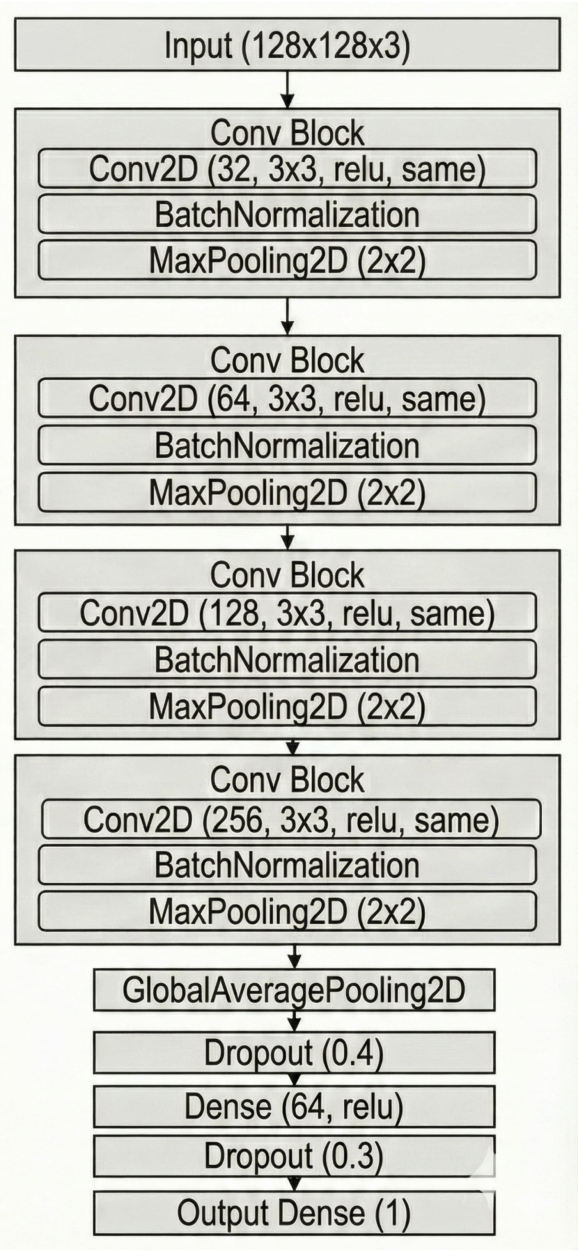
2.1 Shared Backbone (Feature Extractor)

The backbone processes the $128 \times 128 \times 3$ input through four sequential convolutional blocks with increasing depth ($32 \rightarrow 64 \rightarrow 128 \rightarrow 256$):

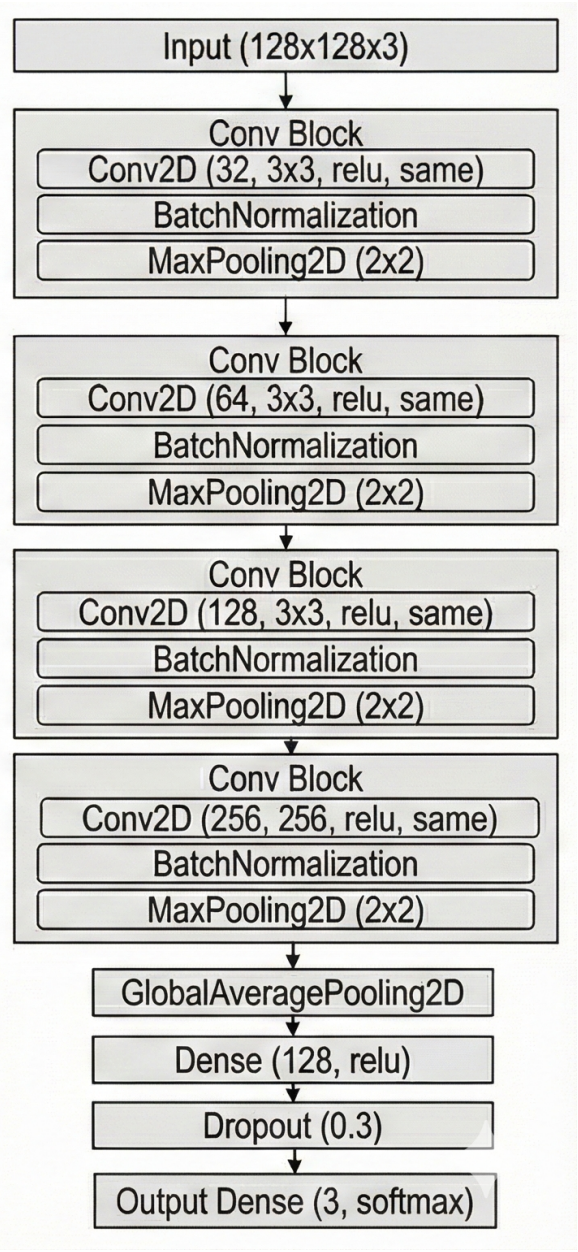
- **Conv Blocks:** Each block contains a `Conv2D` layer (ReLU, 3×3), `BatchNormalization`, and `MaxPooling2D`.
- **Bottleneck:** A `GlobalAveragePooling2D` layer reduces the 3D feature map to a 1D vector for efficient learning.

2.2 Task-Specific Heads

- **Regression Head:** Dense(256) $\rightarrow$ Dropout(0.4) $\rightarrow$ Dense(64) $\rightarrow$ Dropout(0.3) $\rightarrow$ Dense(1, linear).
- **Classification Head:** Dense(128) $\rightarrow$ Dropout(0.3) $\rightarrow$ Dense(3, softmax).



(a) Regression Model Architecture



(b) Classification Model Architecture

Figure 1: Side-by-side comparison of the Regression and Classification model architectures.

3 Comparison of Approaches

3.1 Approach 1: Regression (Continuous Prediction)

- MAE: 1.235 years
- RMSE: 1.594 years
- Strengths: Granular and clinically relevant predictions.
- Limitations: Higher variance in predictions for younger age groups (0–2 years).

3.2 Approach 2: Classification (Developmental Staging)

- Accuracy: 78%

- Issues: Severe imbalance — "Adult" class had near-zero recall (< 1% representation).
- Strengths: High recall and precision for common classes.

4 Error Analysis and Model Behavior Insights

4.1 Regression Model: Understanding the Variance

The discrepancy between the Mean Absolute Error (MAE = 1.235 years) and the Root Mean Squared Error (RMSE = 1.594 years) provides a critical insight into the model's behavior. Since RMSE penalizes large errors more heavily than MAE, this gap indicates the presence of significant outliers where the model's prediction deviates drastically from the ground truth.

Sources of Outliers:

- **The "Infant Gap" (0-2 Years):** As visualized in the scatter plot, predictions for subjects under 2 years of age show high variance. At this developmental stage, the carpal bones (a primary feature for bone age assessment) have often not yet ossified and are invisible on X-rays. The model is forced to rely on soft tissue size and general hand shape, which are less reliable predictors than distinct bony landmarks.
- **The "Maturity Ceiling" (17+ Years):** A "ceiling effect" is observed where predictions plateau for older subjects. Once the epiphyseal plates fuse completely, the radiological appearance of the hand becomes static, even as chronological age continues to advance. The model correctly identifies "maturity" but struggles to regress precise ages beyond this physiological limit.

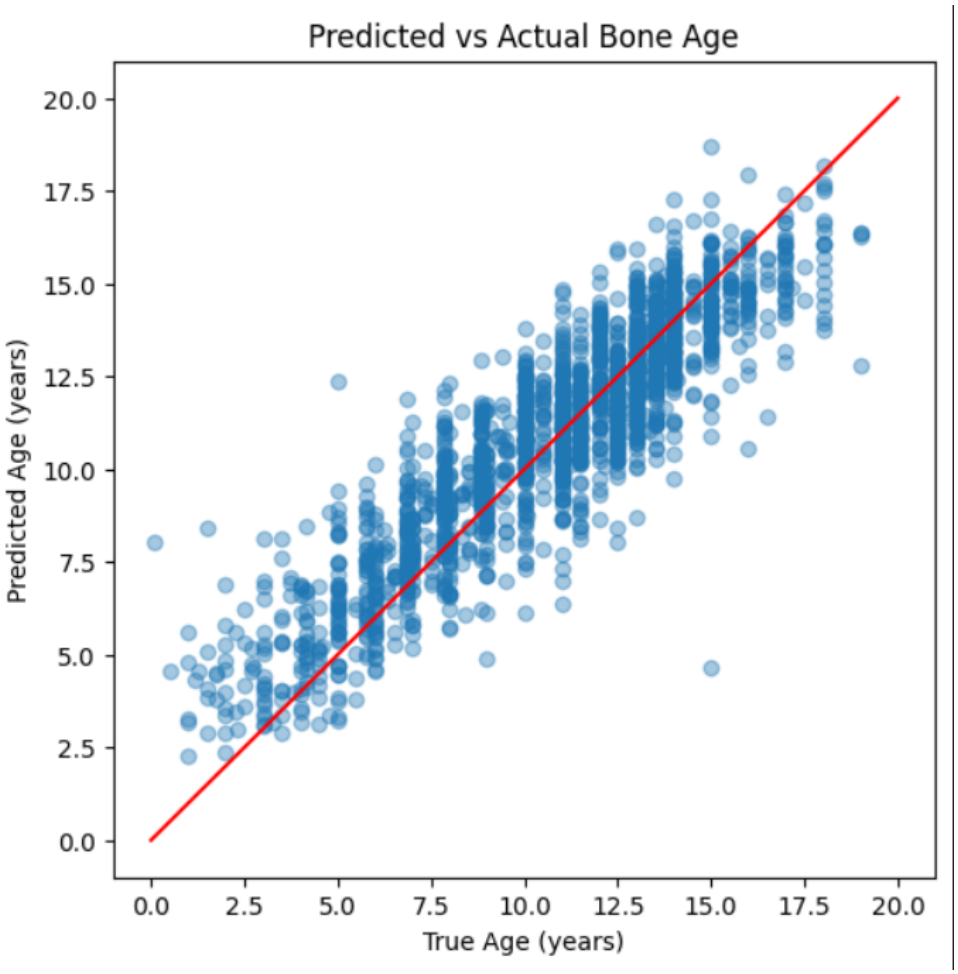


Figure 2: Placeholder for Predicted vs. True Age Scatter Plot.

4.2 Classification Model: Confusion Matrix Analysis

The classification model (Test Accuracy: 78%) revealed specific weaknesses in distinguishing adjacent developmental stages.

Key Findings from Confusion Matrix:

- **Child vs. Adolescent Ambiguity:** The model achieved a Recall of only 0.40 for the "Child" class (¡10 years), frequently misclassifying them as "Adolescents". This suggests that the transition period around age 10—where puberty onset varies widely between individuals—creates a visually ambiguous decision boundary that the current 128×128 resolution cannot resolve.
- **The Minority Class Failure:** The "Adult" class had a precision and recall of 0.00. This is a direct result of extreme class imbalance (only 93 adults vs. 8,230 adolescents in the dataset). The model learned to optimize global accuracy by effectively ignoring this rare class, highlighting the need for oversampling techniques or class-weighted loss functions in future iterations.

4.3 Visual Validation via Grad-CAM

To ensure the model wasn't relying on spurious correlations (e.g., background artifacts or tags), Gradient-weighted Class Activation Mapping (Grad-CAM) was employed.

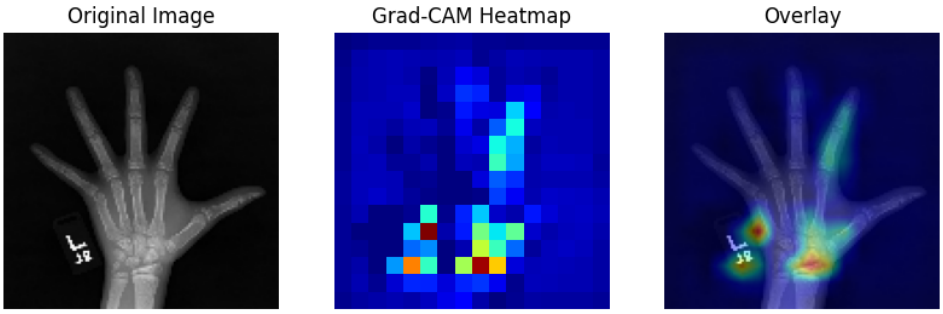


Figure 3: Placeholder for Grad-CAM Visualization.

The activation heatmaps confirm that the model correctly focuses on anatomically significant regions:

- **Primary Focus:** The carpal bones (wrist area), which provide the strongest indicators of maturation in younger subjects.
- **Secondary Focus:** The metacarpal-phalangeal joints, where epiphyseal fusion is most visible in adolescents.

This visual evidence validates that the custom CNN is learning robust, medically relevant features despite the noisy nature of X-ray data.

5 Conclusion

This study demonstrates that a custom CNN can effectively estimate bone age using hand radiographs. The regression model provides more clinically useful continuous predictions, while the classification model performs well only for majority categories. Improvements such as balanced sampling, gender integration, and higher-resolution inputs could boost performance further.