University of Jordan

King Abdullah || School of Information Technology

Graduation Project

Supervisor: Dr. Tamam Alsarhan

Dr.Ali Alrodan

Project subject: Bone Age Estimation

using ResNet34

Name	ID
Mohammad Haidar Mohammad Sammour	0202735
SAIF ALDEN Ahmad Rislan AlMoghear	0201248
Leen Hasan Suleiman Al Adam	0201310

A PROJECT SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENT

FOR THE BACHELOR OF Artificial Intelligence

Winter Semester 2024/2025

Table of Contents

Acknowledgments

We would like to express our sincere gratitude to our supervisors, Dr. Tamam Alsarhan and Dr. Ali Alrodan, for their continuous support and invaluable guidance throughout the course of this project. Their expert advice and thoughtful feedback were instrumental in helping us navigate the various stages of our work. Our heartfelt thanks also go to our families and friends for their unwavering emotional support and encouragement. Their constant belief in our abilities provided us with the motivation needed to persevere. Finally, we extend our appreciation to everyone who contributed to this project, and we will forever cherish the knowledge and experience gained during this transformative journey.

Abstract

Bone age estimation is an essential diagnostic tool in pediatric medicine, aiding in the assessment of growth disorders and developmental anomalies. Traditional methods like the Greulich-Pyle and Tanner-Whitehouse standards rely on manual interpretation, often leading to subjectivity, variability, and inefficiencies. In this study, we present an automated approach using ResNet34, a convolutional neural network, trained on the RSNA Pediatric Bone Age Estimation dataset. The model leverages preprocessing techniques such as cropping, normalization, and augmentation to extract critical skeletal features and deliver accurate and consistent predictions. Our results demonstrate that ResNet34 effectively overcomes the limitations of traditional methods, offering a scalable, efficient, and objective solution for clinical and research applications, thereby showcasing the transformative potential of deep learning in medical imaging.

1. Introduction

1.1 Preamble

Bone age estimation is a critical process in pediatric medicine, used to assess the development and growth of children. It is particularly important for diagnosing growth-related disorders and monitoring treatment effectiveness in various clinical settings [1].

Traditionally, bone age has been assessed manually by comparing X-ray images of the hand and wrist to reference atlases, see *Figure 1*. It is usually performed by radiological examination of the left hand. Currently, bone age assessment is primarily conducted by trained radiologists, who manually assess hand bones in X-ray images in accordance with the Greulich-Pyle (GP) method. However, this approach is subjective, time-consuming, and prone to inter-observer variability.

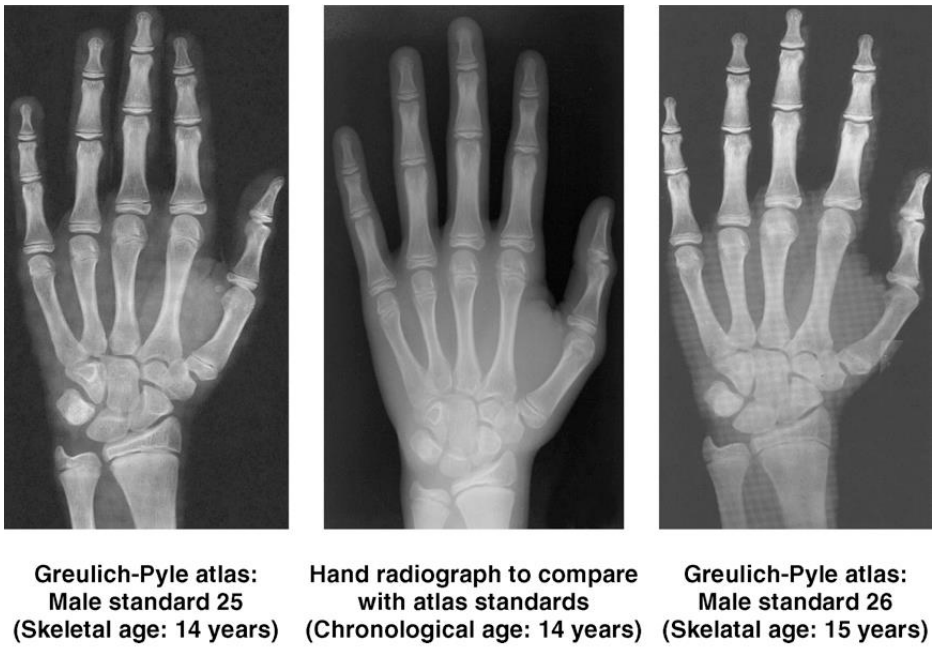

Figure 1:Approach to estimation of skeletal age using Greulich–Pyle method. Middle X-ray of an investigated subject definitely meets criteria of Greulich–Pyle standard 25 (skeletal age: 14 years; left)

Recent advancements in artificial intelligence, particularly in the field of deep learning, have opened new avenues for automating and improving the accuracy of bone age estimation [2]. One such innovation is the ResNet34 architecture, a powerful deep learning model that has demonstrated excellent performance in various image analysis tasks.

This project aims to explore the potential of ResNet34 for automated bone age estimation, offering a more accurate, reliable, and efficient alternative to traditional methods. By applying state-of-the-art deep learning techniques, we seek to contribute to the growing field of medical image analysis and provide an improved tool for clinical practice [3].

1.2 Project Motivation

1.2.1 Importance of Bone Age Estimation

Bone age estimation is a process used to evaluate the maturity of a child's skeletal system [4]. It is crucial in a variety of medical scenarios, such as:

1. Growth disorders: To identify whether a child's growth is delayed or accelerated, which can help diagnose conditions like growth hormone deficiencies or puberty-related abnormalities.

2. Hormonal imbalances: Conditions such as precocious puberty or delayed puberty can be detected by assessing the rate of bone development.

3. Monitoring treatments: For example, in children receiving hormone therapy or other growth-related treatments, bone age estimation helps track progress and adjust treatments accordingly.

The traditional methods of bone age assessment involve comparing an X-ray image of a child's hand and wrist with a reference atlas (e.g., Greulich and Pyle or Tanner-Whitehouse methods) [5]. This process is subjective, as it requires a skilled clinician to interpret the radiographs and make a judgment about the bone age based on visual comparison, see *Figure 2*. Human error and inter-observer variability (differences in opinion between clinicians) can lead to inconsistent results [6].

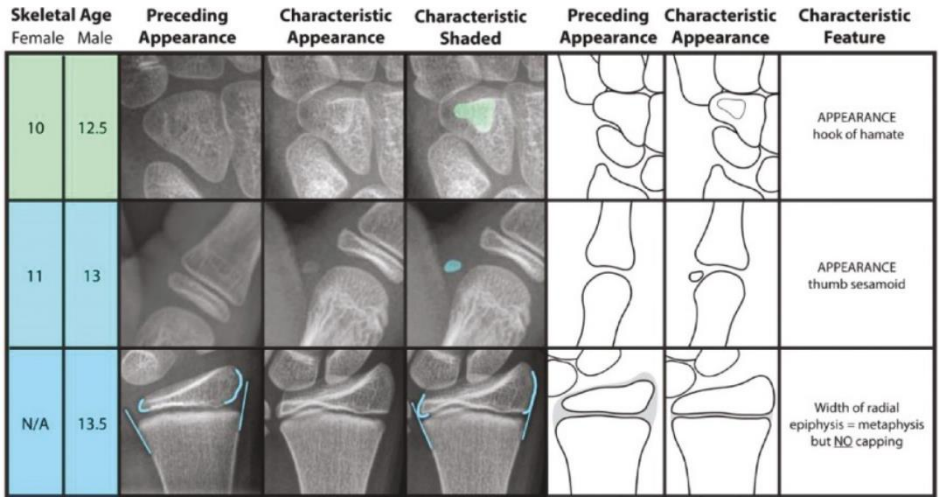

Figure 2:skeletal age based on hand ossification features

1.2.2 The Need for Automation and AI in Bone Age Estimation

In light of the limitations of manual bone age estimation, there is an increasing demand for automation in the healthcare sector, particularly in the field of pediatric diagnostics. Automating this process can have several benefits:

1. Efficiency: Automation can significantly reduce the time needed to evaluate bone age, which is especially beneficial in busy clinical settings where time is critical.

2. Consistency and Accuracy: By eliminating human error and subjective judgment, an automated system can provide more consistent and reliable results.

3. Scalability: Automated systems can handle large volumes of radiographs quickly, making it easier to scale the assessment process in hospitals or health systems with many patients.

One of the most promising approaches to automation in image analysis is the use of deep learning, which includes techniques such as Convolutional Neural Networks (CNNs), see *Figure 3*. These models are capable of learning complex patterns from large amounts of data and have shown impressive performance in fields like medical imaging, facial recognition, and object detection [7].

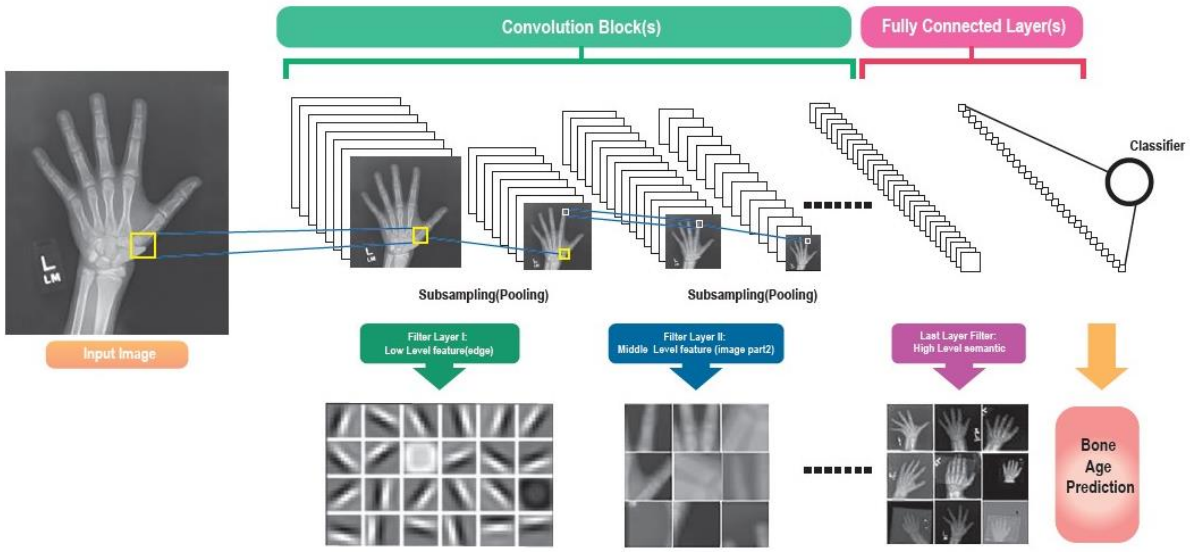

Figure 3:The architecture of CNNs model for bone age estimation

1.2.3 Why ResNet34?

ResNet34 is a powerful deep learning architecture that has been shown to be highly effective for image analysis tasks. It employs residual connections, which allow the model to learn deeper representations without suffering from vanishing gradient problems. This makes it well-suited for tasks like bone age estimation, where complex patterns need to be extracted from medical images.

1.2.4 Motivation Behind the Project

The main motivation for this project is to explore how ResNet34 can be applied to automate bone age estimation with high accuracy and efficiency [8]. By training a ResNet34 model on a large dataset of radiographs, the project seeks to achieve the following goals:

1. Enhance accuracy: By leveraging ResNet34, the model can potentially make more accurate bone age predictions, surpassing the performance of traditional methods and other deep learning models like CNNs.

2. Reduce human error: Automation minimizes the subjectivity of manual bone age assessment, offering a more consistent and objective solution.

3. Improve clinical workflow: By automating the bone age assessment, healthcare providers can save time and effort, allowing them to focus on other aspects of patient care.

4. Support better healthcare outcomes: Accurate and timely bone age assessment can help detect growth disorders earlier, leading to more effective treatments and interventions.

1.3 Problem Statement

Bone age estimation is a critical task in pediatric healthcare, often used for diagnosing growth-related disorders and evaluating developmental progress [9]. However, the traditional methods of bone age assessment, which rely on manual comparison of hand and wrist radiographs to reference atlases, suffer from several limitations. These methods are subjective, time-consuming, and highly dependent on the expertise of the clinician, resulting in potential inter-observer variability and inconsistencies in the assessments [10]. As a result, there is a pressing need for a more accurate, efficient, and objective approach to automate the bone age estimation process [11].

While recent advancements in machine learning have introduced automated solutions, most existing models have relied on Convolutional Neural Networks (CNNs) [12], which, despite their success in many image analysis tasks, struggle to capture long-range dependencies and contextual information across an image. This limitation is particularly significant for bone age estimation, where the relationships between various bones and the overall skeletal development need to be considered in a holistic manner.

ResNet34 has demonstrated superior performance in capturing global context and relationships within images. However, there is limited research on applying ResNet34 specifically to the problem of bone age estimation. Therefore, the problem addressed by this project is the development and application of a ResNet34-based model for automated bone age estimation, aimed at improving the accuracy, consistency, and efficiency of the process compared to existing methods.

1.4 Project Aim and Objectives

1.4.1 Project Aim

The primary aim of this project is to develop and evaluate an automated system for bone age estimation using ResNet34. By leveraging the capabilities of ResNet34 in capturing long-range dependencies and contextual information in medical images, the project seeks to enhance the accuracy, consistency, and efficiency of bone age estimation, offering a reliable tool for pediatric healthcare professionals [13].

1.4.2 Project Objectives

To achieve the aim of this project, the following specific objectives will be pursued:

1. Dataset Collection and Preprocessing

- Collect a comprehensive dataset of hand and wrist radiographs that are annotated with bone age labels.

- Preprocess the dataset to ensure uniformity in image size, quality, and format, and apply techniques like data augmentation to improve model robustness.

2. Model Development

- Design and implement a ResNet34-based model for bone age estimation. The model will be trained to extract relevant features from X-ray images of the hand and wrist, which are key to determining bone age.

- Explore hyperparameter optimization techniques to enhance model performance and prevent overfitting.

3. Model Evaluation

- Evaluate the model's performance on the dataset using standard metrics such as mean absolute error (MAE), accuracy, and root mean square error (RMSE).

- Compare the performance of the ResNet34-based model with traditional machine learning models (e.g., CNNs) to assess improvements in accuracy and robustness.

4. Clinical Applicability Assessment

- Assess the model's practical applicability in a clinical setting by analyzing its scalability, consistency, and real-time performance in bone age estimation.

5. Documentation and Reporting

- Document the development process, methodology, experimental results, and insights gained from the project.

- Provide recommendations for future work and potential clinical implementation.

1.5 Project Scope

This project aims to develop a ResNet34-based model for automated bone age estimation from hand and wrist X-rays. The key aspects of the project are as follows:

1. Data and Dataset: The project will use publicly available datasets (e.g., RSNA bone age dataset) with annotated pediatric radiographs. Preprocessing will include resizing, normalization, and data augmentation to ensure data quality.

2. Model Development and Training: The core focus will be designing and implementing the ResNet34 architecture for bone age estimation using deep learning frameworks (TensorFlow/PyTorch). Model optimization and hyperparameter tuning will be conducted to improve performance [14].

3. Evaluation Metrics: The model's performance will be evaluated using mean absolute error (MAE), root mean square error (RMSE), and accuracy. Results will be compared to traditional methods like CNNs and manual assessments.

4. Clinical Relevance: The project will assess the model’s practical application in clinical settings, focusing on its potential for automating bone age estimation and integrating into healthcare workflows.

5. Exclusions: The project will focus on hand and wrist X-rays only, excluding other body parts and imaging techniques (MRI/CT scans). It will not develop new radiographic hardware but improve existing methods [15].

6. Timeframe and Resources: The project will be completed within a set timeframe, using available computing resources (GPUs), and primarily working with Python and deep learning frameworks like TensorFlow and PyTorch.

1.6 Project Limitation

While ResNet34 has shown promise in various medical image analysis tasks, including bone age estimation, several limitations and challenges persist:

1. Data Quality and Quantity

- Image Quality: The quality of input images significantly impacts the model's performance. Variations in image resolution, contrast, and noise can introduce uncertainty in the estimation process.

- Data Quantity: Large and diverse datasets are crucial for training robust models. However, acquiring a sufficient number of high-quality annotated images can be challenging, especially for specific bone types or clinical conditions.

2. Anatomical Variability

- Inter-individual Variation: Individuals exhibit significant anatomical variations in bone structure, size, and maturation rate. These variations can pose challenges for accurate estimation, especially when dealing with atypical cases [16].

- Intra-individual Variation: Even within the same individual, bone development can be asynchronous across different skeletal sites. This variability can affect the reliability of estimates based on a single image [16].

3. Interpretability and Explainability:

- Deep learning models, including ResNet34, are often considered black boxes [17], making it difficult to understand the underlying decision-making process. This lack of interpretability can hinder trust and clinical adoption.

- Explainable AI (XAI) techniques can provide insights into model predictions, but their effectiveness in medical image analysis remains an active research area.

1.7 Project Expected Output

The expected output of a bone estimation project using ResNet34 is a model capable of accurately predicting bone age or maturity level from medical images, such as X-rays, see *Figure 4*. This model should:

1. Process Input Images: The model should be able to efficiently process and extract relevant features from input images, which may include various bone types like hand and wrist and imaging modalities like X-ray and CT.

2. Predict Bone Age or Maturity: The model should output a numerical estimate of bone age or a categorical classification of maturity level like early, average and late. The prediction should be accurate and reliable, even in the presence of image variations and noise.

3. Provide Confidence Intervals: Ideally, the model should provide confidence intervals or uncertainty estimates for its predictions, allowing clinicians to assess the reliability of the estimated bone age.

4. Generalize Across Populations: The model should be able to generalize well to different populations, including individuals with varying genetic backgrounds, growth patterns, and disease conditions.

5. Integrate with Clinical Workflows: The model should be designed to seamlessly integrate into existing clinical workflows, providing a user-friendly interface and efficient processing times.

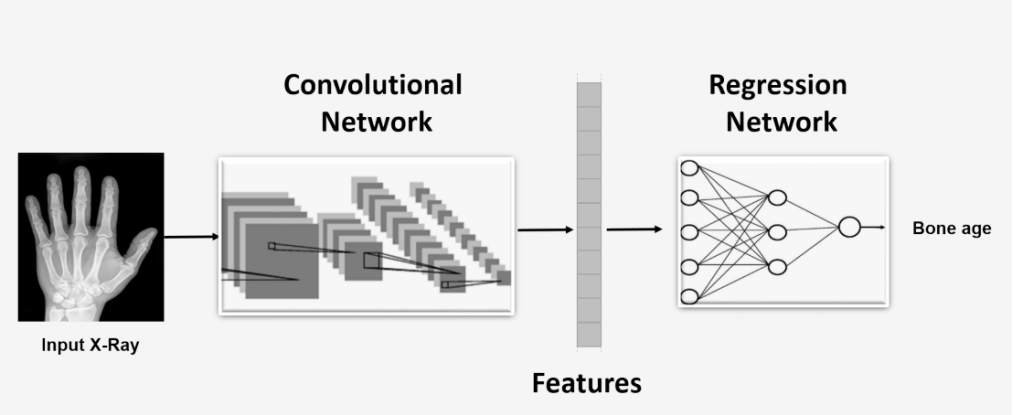

Figure 4: General architecture of deep learning methods for bone age assessment.

2. Related Work

Bone age estimation plays a critical role in pediatrics, particularly in assessing and managing growth disorders, evaluating hormonal imbalances, and estimating maturity levels in young individuals. This process involves comparing the skeletal development of an individual to a reference standard, providing valuable insights into growth potential and deviations from normal patterns. Traditionally, bone age is determined using radiographic images, particularly of the left hand and wrist, following methodologies like the Greulich-Pyle (GP) which is a well-established technique for estimating a child's bone age. It involves comparing a hand and wrist X-ray to a standardized atlas of bone development [18]. This atlas contains images representing different stages of bone maturation, allowing for a visual comparison and estimation of the child's skeletal ages, see *Figure 5*.

The GP method is based on the principle that bones develop in a predictable sequence and at a relatively consistent rate. By analyzing the degree of ossification and the appearance of specific bone centers, clinicians can assess how advanced a child's skeletal development is compared to chronological age [18].

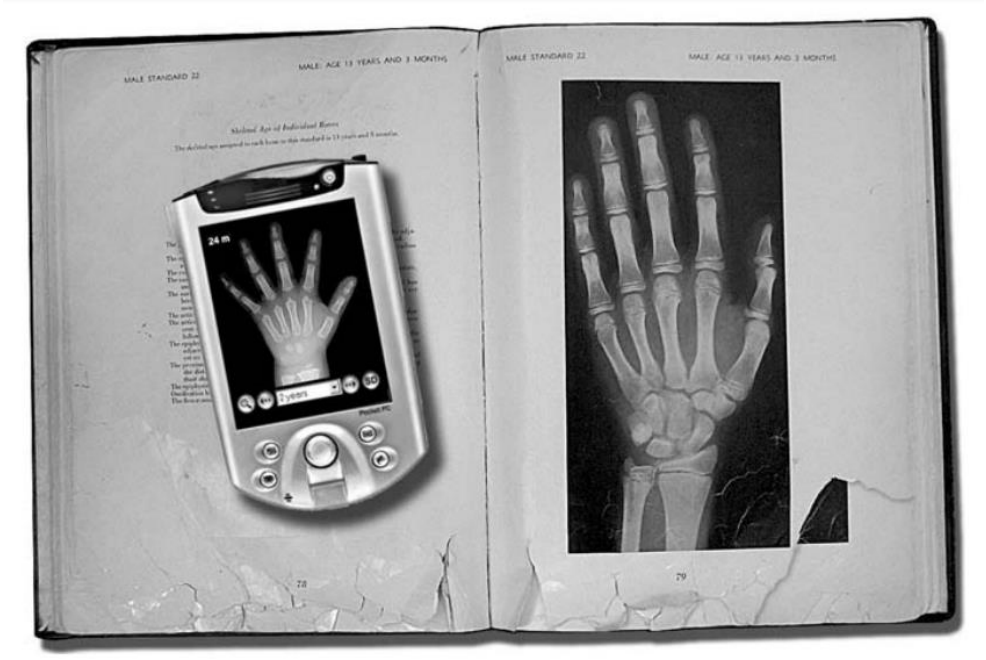

Figure 5: Comparison of the traditional Greulich and Pyle atlas used for determination of bone maturity.

While the GP method is widely used, it has some limitations. It relies on subjective interpretation by the clinician, and variations in technique and interpretation can lead to differences in bone age estimates. Additionally, the original GP atlas was based on a specific population, and its accuracy may vary in different ethnic groups [18]. Another traditional way to determine the bone age is Tanner-Whitehouse (TW) standards, it is a well-established technique for estimating a child's bone age. Unlike the Greulich-Pyle method, which relies on visual comparison to a standard atlas, the TW method involves a more quantitative approach.

A hand and wrist X-ray is taken, and specific bones are identified. Each bone is assigned a score based on its maturity level, considering factors like ossification centers, epiphyseal fusion, and bone shape. These scores are then summed to calculate a total score. This total score is compared to standardized tables or charts to estimate the child's bone age.

The TW method is considered more precise and objective than the GP method, as it relies on numerical scores rather than subjective visual comparison. However, it requires careful technique and experience in interpreting X-rays to accurately assign scores to each bone.

While the TW method is a valuable tool for assessing skeletal maturity, it has limitations. It can be time-consuming and requires specialized training to perform accurately. Additionally, variations in technique and interpretation can still influence the results [19].

Name	**Method**	**Advantages**	**Disadvantages**	**Radiation risk**
GP	Visual inspection Correspondence method.	Quick execution. Used by more than 76% of pediatricians.	Greater variability between observers compared to the TW method.	Very low
TW	Visual and scoring method: the sum of scores reflects general skeletal development.	More reliable than GP method.	Subjective evaluation of bone age. Takes time.	Very low

Table 1: Comparison between GP and TW.

While effective, these manual methods are inherently subjective, time-consuming, and reliant on expert radiologists, resulting in potential inter and intra-observer variability. These challenges have fueled the pursuit of automated solutions to enhance precision, efficiency, and accessibility.

The foundation for automation in bone age estimation was laid with the advent of computer-aided diagnosis (CAD) systems [20]. Early attempts relied on handcrafted features, such as edge detection and geometric analysis, derived from radiographs. These methods required domain expertise to engineer relevant features manually, which often limited their generalizability across diverse populations and imaging conditions [21]. With the rise of machine learning, statistical models began to replace rule-based approaches, demonstrating moderate success. However, it wasn't until the emergence of deep learning that significant breakthroughs were achieved in this domain.

Deep learning, particularly convolutional neural networks (CNNs), has revolutionized medical image analysis by enabling automatic feature extraction and end-to-end learning directly from image data. Various deep learning architectures have been explored for bone age estimation, including AlexNet, a groundbreaking convolutional neural network (CNN) architecture, has significantly impacted the field of medical image analysis, including bone age estimation. Its powerful feature extraction capabilities have been leveraged to develop sophisticated deep learning models for this task.

AlexNet's success in image classification tasks demonstrated the potential of deep learning in medical image analysis. This inspired the development of more complex and specialized CNN architectures for medical imaging, including bone age estimation.

AlexNet's architecture, with multiple convolutional and pooling layers, is effective in extracting relevant features from medical images, such as X-rays of hands and wrists. These features can be used to train models that accurately predict bone age.

Pre-trained AlexNet models, trained on large datasets like ImageNet, can be fine-tuned on smaller medical image datasets, including bone age X-rays. This transfer learning approach accelerates model training and improves performance [22].

AlexNet's design principles have influenced the development of newer architectures like Inception, ResNet, and DenseNet, which have been successfully applied to bone age estimation [23].

AlexNet, while influential, has limitations such as its relatively shallow architecture and limited capacity to capture complex features, especially for fine-grained tasks like bone age estimation.

The Inception architecture, a powerful deep learning model, has significantly contributed to the advancement of bone age estimation. Characterized by its use of multiple convolutional layers with different kernel sizes, Inception allows the network to extract features at various scales, capturing both fine-grained and coarse-grained details within the X-ray images.

In the context of bone age estimation, Inception-based models have proven to be highly effective. They can accurately identify and analyze key skeletal features, such as ossification centers and bone maturation patterns, which are crucial for determining a child's biological age [8].

By using multiple convolutional layers with different kernel sizes, Inception models can capture features at various scales, allowing for a more comprehensive understanding of the bone structures. Additionally, the Inception architecture is designed to be computationally efficient, enabling faster training and inference times.

Inception architectures can be computationally expensive, especially when dealing with large-scale medical image datasets. Additionally, their complex structure can make them challenging to train and fine-tune [24].

The ResNet family of neural networks has significantly impacted the field of computer vision, and its influence extends to medical image analysis, including bone age estimation. These networks, characterized by their "residual learning" architecture, have proven to be highly effective in tackling the challenges associated with training deep neural networks.

One of the major challenges in training deep neural networks is the vanishing gradient problem, where gradients become smaller and smaller as they propagate backward through the layers. ResNet's residual connections help alleviate this issue by allowing gradients to flow more directly, enabling the training of deeper networks [25].

By effectively addressing the vanishing gradient problem, ResNets can achieve higher accuracy in bone age estimation tasks. They can capture complex patterns and subtle variations in bone development, leading to more precise predictions. Additionally, the residual connections in ResNets contribute to faster training convergence, reducing the time required to train these models [25].

ResNet34, a member of the ResNet family, is particularly well-suited for tasks like bone age estimation. With its 34-layer architecture, ResNet34 strikes a balance between depth and computational efficiency, making it capable of capturing complex patterns in medical images while maintaining a manageable training complexity. The use of residual blocks in ResNet34 allows for the learning of intricate features without overfitting, which is especially important given the variability in bone structure across different age groups and individuals [26].

DenseNet, a powerful deep learning architecture, has significantly contributed to the advancement of bone age estimation. Its unique design, characterized by dense connections between layers, enables efficient feature propagation and reuse [27].

By connecting each layer to every subsequent layer, DenseNets encourage feature reuse, allowing the network to learn more efficiently and effectively. This dense connectivity also helps mitigate the vanishing gradient problem, enabling the training of deeper networks [28].

The dense connections in DenseNets promote feature diversity, as each layer receives input from multiple layers. This leads to richer feature representations, which are crucial for accurately identifying and analyzing key skeletal features in X-ray images [28].

DenseNets have consistently demonstrated strong performance in bone age estimation tasks, often surpassing other architectures like ResNet and Inception. They can accurately predict a child's biological age, which is essential for diagnosing growth disorders, predicting adult height, and monitoring the effectiveness of treatments [28].

DenseNets can be computationally intensive, especially for large-scale models. Additionally, their dense connectivity can make them prone to overfitting, requiring careful regularization techniques [29].

The field of bone age estimation has significantly benefited from advancements in deep learning. Architectures like AlexNet, ResNet, Inception, and DenseNet have played pivotal roles in improving the accuracy and efficiency of bone age prediction.

AlexNet, a pioneering architecture, laid the foundation for many subsequent models. While effective for general image classification, its capabilities for the intricate details of bone age assessment were limited.

Inception architectures, with their multi-scale feature extraction, can effectively capture both fine-grained and coarse-grained details within X-ray images. This makes them well-suited for bone age estimation, especially when combined with attention mechanisms.

ResNet's residual connections address the vanishing gradient problem, enabling the training of deeper networks. This architecture excels in capturing both low-level and high-level features, leading to improved performance in bone age estimation.

DenseNets, characterized by their dense connections, efficiently propagate and reuse features. This architecture has shown promising results in bone age estimation, often achieving high accuracy.

While each architecture offers unique advantages, the optimal choice depends on factors such as dataset size, computational resources, and desired performance metrics.

The success of any deep learning-based approach to bone age estimation hinges on the quality and diversity of the dataset. The RSNA Pediatric Bone Age Estimation Challenge dataset has become a cornerstone for research in this field. This dataset comprises thousands of labeled hand and wrist radiographs annotated with corresponding bone age labels. The availability of this dataset has significantly propelled the development and benchmarking of deep learning models by providing a standardized platform for experimentation. However, raw radiographs often include irrelevant information such as background noise or variations in image intensity, necessitating robust preprocessing techniques.

Preprocessing techniques typically involve normalization, resizing, and augmentation of images to enhance model robustness and generalizability. Cropping the region of interest (ROI) to focus on the hand and wrist, see *Figure 6*, adjusting for variations in brightness or contrast, and employing data augmentation strategies such as rotation, flipping, or zooming are common practices [30]. These steps not only improve the quality of input data but also mitigate overfitting by artificially expanding the dataset size.

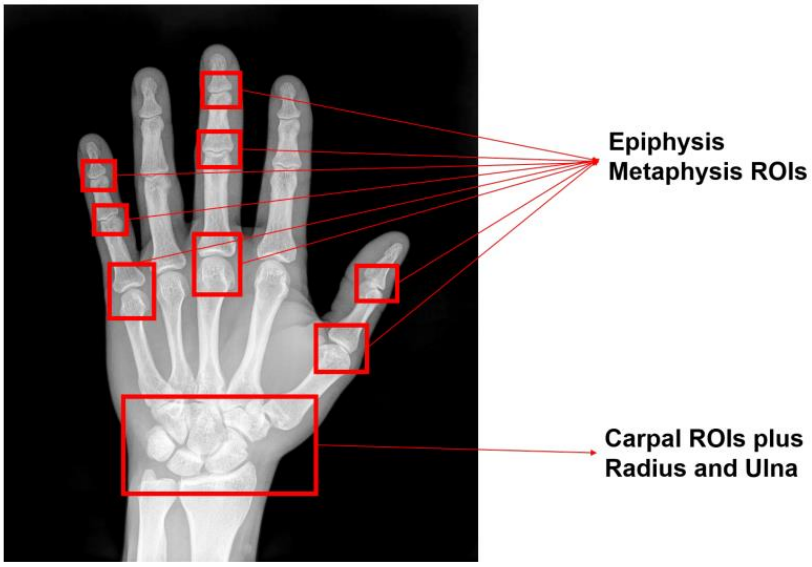

Figure 6: Regions of Interest (ROIs) used in the Tanner Whitehouse method.

Feature extraction remains a critical aspect of bone age estimation. Important key features include the structural characteristics of bones, such as the presence, size, and shape of ossification centers, as well as the width and appearance of epiphyseal plates. Additionally, the overall shape and contour of bones, like the metacarpals, phalanges, and carpal bones, provide valuable information. Textural properties of bone tissue, including bone density and texture patterns within bone structures, offer further insights into bone age. For instance, the trabecular bone pattern can vary with age and maturity, providing clues about skeletal development. Furthermore, the spatial relationships between bones play a significant role. These include the relative positions and orientations of bones, as well as the distances between specific bone landmarks [18]. By analyzing these spatial relationships, deep learning models can better understand the overall skeletal maturity of the individual. ResNet34, with its hierarchical feature extraction capabilities, excels in identifying these subtle yet vital patterns. Its ability to process intricate details of bone morphology while learning age-specific characteristics has proven invaluable in improving the accuracy and reliability of bone age predictions.

The RSNA dataset, besides its size, also stands out for its diversity, representing a wide range of ages and demographic groups, see *Figure 7*. This diversity ensures that models trained on this dataset are more likely to generalize well to unseen data [6]. Additionally, the inclusion of expert annotations provides a reliable ground truth for supervised learning, facilitating the development of robust prediction models.

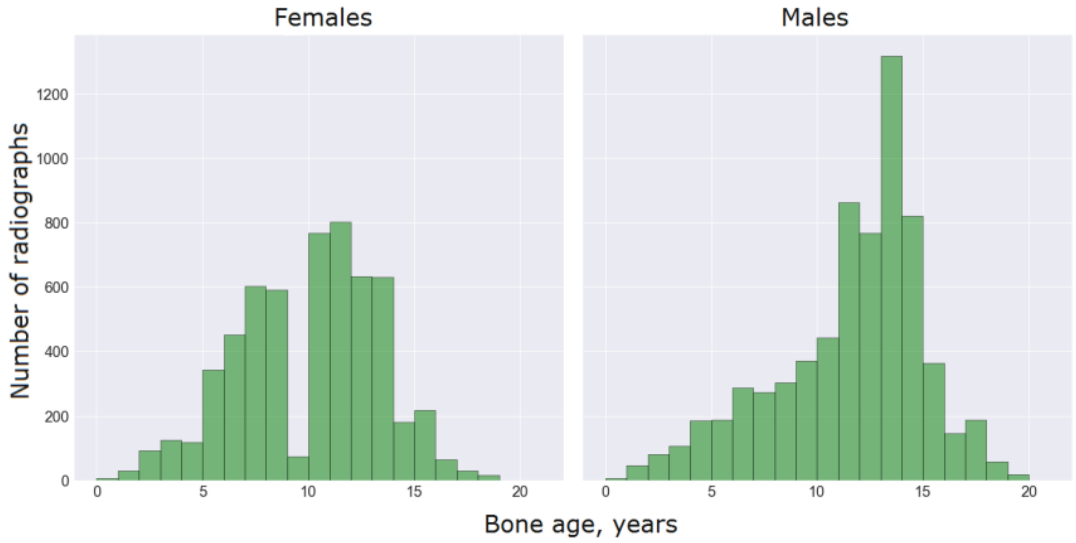

Figure 7: Bone age distribution for females and males in the training dataset.

In summary, bone age estimation has undergone a remarkable transformation, moving from traditional, manual methods reliant on expert interpretation to highly automated, data-driven approaches enabled by the power of deep learning. This evolution has not only addressed the limitations of subjectivity and variability inherent in conventional techniques but has also expanded the possibilities for accuracy, efficiency, and scalability. ResNet34, with its advanced architecture featuring residual connections and hierarchical feature extraction capabilities, has emerged as a cornerstone in this transformation, demonstrating its ability to capture intricate skeletal features critical for precise bone age assessment. Central to this progress has been the RSNA Pediatric Bone Age Estimation dataset, which has provided a diverse and meticulously annotated collection of hand and wrist radiographs. This dataset has not only facilitated rigorous training and validation of models like ResNet34 but has also served as a benchmark for innovation in the field. Together, these advancements weave a compelling narrative of how technology is revolutionizing traditional medical workflows. By integrating cutting-edge algorithms with high-quality data, clinicians now have access to powerful, automated tools that enable faster, more consistent, and highly accurate bone age predictions, ultimately improving patient outcomes and broadening access to critical diagnostic capabilities.

3. Project Requirements

Developing an automated bone age estimation system using ResNet34 involves addressing a range of requirements that span hardware, software, and data considerations. These requirements ensure the system is capable of effectively training and deploying a deep learning model while maintaining accuracy, scalability, and reliability. The system must also be designed to handle the complexity of medical imaging tasks, adhere to regulatory standards, and operate within acceptable time constraints. Below, the system requirements are categorized into hardware, software, data, and operational needs.

3.1 Hardware Requirements

Deep learning models, such as ResNet34, require substantial computational resources for training and inference. Hardware selection is critical to ensuring the system's performance is both efficient and scalable. Training ResNet34 involves processing large volumes of high-resolution medical images, which is computationally intensive. A high-performance Graphics Processing Unit (GPU) is essential to accelerate matrix computations and optimize training times. GPUs such as NVIDIA RTX 3090, A100, or equivalent provide the necessary computational power. Additionally, medical imaging datasets, particularly the RSNA Pediatric Bone Age Estimation dataset, can be sizeable, so storage capacity is required.

3.2 Software Requirements

The software stack for bone age estimation using ResNet34 encompasses the operating system and deep learning frameworks. Popular frameworks like TensorFlow or PyTorch are essential for implementing and training the ResNet34 model. PyTorch, in particular, is favored for its flexibility, ease of debugging, and extensive community support.

3.3 Data Requirements

The quality and volume of data are pivotal in determining the success of the ResNet34-based bone age estimation system. Proper management and preprocessing of the data ensure reliable and interpretable results.

3.4 Operational Requirements

To ensure the successful deployment and usability of the automated bone age estimation system, several operational aspects must be carefully considered. The system should support an iterative model training environment, allowing for continuous improvement and adaptation to new data. Real-time monitoring of metrics like mean absolute error (MAE), accuracy, and loss is crucial to identify potential issues such as overfitting or underfitting. Visualization tools can aid in this process, providing insights into the model's learning behavior.

For practical clinical use, the system must deliver predictions in real-time or near-real-time. The inference pipeline should be optimized to minimize latency while maintaining accuracy. To accommodate future growth and evolving needs, the system must be designed with scalability in mind. A user-friendly interface is essential for radiologists and clinicians to interact with the system effectively. The UI should allow for easy upload of radiographs, viewing of predictions, and exploration of visual explanations of the model's decisions.

4. Problem Statement

Bone age estimation is a critical diagnostic tool in pediatric healthcare, aiding in the evaluation of growth disorders, hormonal imbalances, and developmental anomalies. It involves determining the skeletal maturity of a child by analyzing radiographs, typically of the left hand and wrist. Despite its importance, the process remains largely dependent on traditional methods which are manual, subjective, and time-intensive. This reliance on expert interpretation not only limits scalability but also introduces variability in results, which can have significant consequences for patient outcomes. Addressing this issue is essential to improve efficiency, accuracy, and accessibility in clinical workflows.

4.1 Limitations of Traditional Methods

The GP and TW methods, although widely used, are fraught with limitations. The GP method requires radiologists to visually compare a patient's radiograph to an atlas of standard images, selecting the closest match to estimate the bone age. While straightforward, this approach is inherently subjective, leading to inter- and intra-observer variability. A radiologist's experience, fatigue, and even cognitive biases can affect the accuracy and consistency of the results.

On the other hand, the TW method, which scores individual bones and calculates an aggregate score to determine bone age, offers a more systematic approach. However, it is significantly more time-consuming and still relies on expert judgment at multiple stages. Both methods are further constrained by the need for extensive training and specialization, making them less accessible in under-resourced settings where radiological expertise may be scarce. Additionally, these methods are not easily scalable to accommodate large volumes of patients, particularly in high-demand healthcare facilities.

4.2 Challenges in Manual Bone Age Estimation

The subjective nature of traditional bone age estimation creates several challenges. First, the lack of reproducibility in results can lead to inconsistent diagnoses and treatment plans. For instance, a slight variation in the estimated bone age could alter a clinician's decision about initiating or modifying treatment for conditions such as growth hormone deficiency or precocious puberty.

Second, the time-intensive nature of the process delays decision-making. Third, the reliance on experienced radiologists limits the availability of these services in resource-limited areas.

These challenges are compounded by the increasing demand for bone age assessments in both clinical and non-clinical contexts. For example, in sports, bone age is often used to verify the eligibility of young athletes in age-restricted competitions. The growing need for bone age assessments across diverse applications highlights the urgency of developing more efficient and reliable methods.

4.3 The Need for Automation

Automation in bone age estimation offers a promising solution to the limitations of traditional methods. By leveraging advancements in artificial intelligence (AI) and deep learning, automated systems can provide objective, reproducible, and efficient estimates of bone age. These systems have the potential to not only reduce the workload of radiologists but also extend the reach of bone age estimation to settings where expertise is lacking. Automation can improve accuracy by eliminating subjective biases and enabling consistent analysis of radiographs, regardless of operator experience or fatigue.

4.4 Deep Learning as a Transformative Approach

Deep learning has revolutionized medical imaging by enabling algorithms to learn complex patterns directly from data. CNNs in particular, have demonstrated exceptional performance in tasks involving image analysis, such as object detection, segmentation, and classification. In the context of bone age estimation, CNNs can automatically extract relevant features from radiographs, such as the size, shape, and ossification of bones, without requiring handcrafted feature engineering.

4.5 Data Challenges and Requirements

The effectiveness of an automated bone age estimation system depends heavily on the quality and diversity of the data used for training and evaluation. Medical imaging datasets, such as the RSNA Pediatric Bone Age Estimation dataset, provide a valuable resource for this purpose. However, several challenges must be addressed to ensure reliable outcomes.

First, radiographs must be of high resolution and free from artifacts such as noise, motion blur, or improper exposure. Poor-quality images can negatively impact the model's ability to learn relevant features. Second, the dataset must be representative of the population it is intended to serve, encompassing a wide range of ages, genders, and ethnic backgrounds. This diversity ensures that the model generalizes well to unseen data and avoids biases that could skew predictions. Third, accurate annotations of bone age by expert radiologists are

essential for supervised learning. Inconsistent or erroneous labels can introduce noise into the training process, undermining model performance.

4.6 Research Gaps and Opportunities

While deep learning has shown great promise in automating bone age estimation, several research gaps remain. One challenge is the explainability of deep learning models. Clinicians often prefer to understand the rationale behind a model's predictions, which can be difficult with black-box architectures like ResNet34. Developing methods to visualize and interpret the features learned by the model could enhance its acceptance in clinical settings.

Another area for improvement is the integration of clinical metadata, such as the patient's chronological age, gender, and medical history, into the prediction process. Combining radiographic features with contextual information could further improve accuracy and provide more comprehensive insights.

5.0 Methodology

This chapter outlines the methodology employed in developing the bone age estimation project using ResNet34. The methodology contains multiple stages, from image preprocessing and augmentation to the development and training of our deep learning model. We provide a detailed approach to addressing the challenges we faced in automated bone age estimation from hand radiographic x-ray images.

5.1 Preprocessing

The preprocessing workflow starts with excluding low contrast images, manual mask creation through automated segmentation, region filling, and various image enhancement steps. Each step in our pipeline plays an important role and purpose in improving image quality and highlighting relevant anatomical features for more accurate results.

The following sections provide detailed explanations of each preprocessing step, including the technical specifications, parameters used, and the reason behind using it. We also discuss how these techniques work to produce high quality preprocessed images suitable for our bone age estimation project.

5.1.1 Excluding Low Contrast Images

Because the project was requested from King Hussein Cancer Center (KHCC), and due to good resolution hand x-ray images produced by the center, the first step we did is excluding low contrast images from the RSNA dataset, and the reason behind this process is to ensure that the model receives the most relevant data, leading to better performance and more reliable results, see *Figure 8*. Moreover, removing the low-contrast images helps by reducing unnecessary noise and improving the overall efficiency of the model training process.

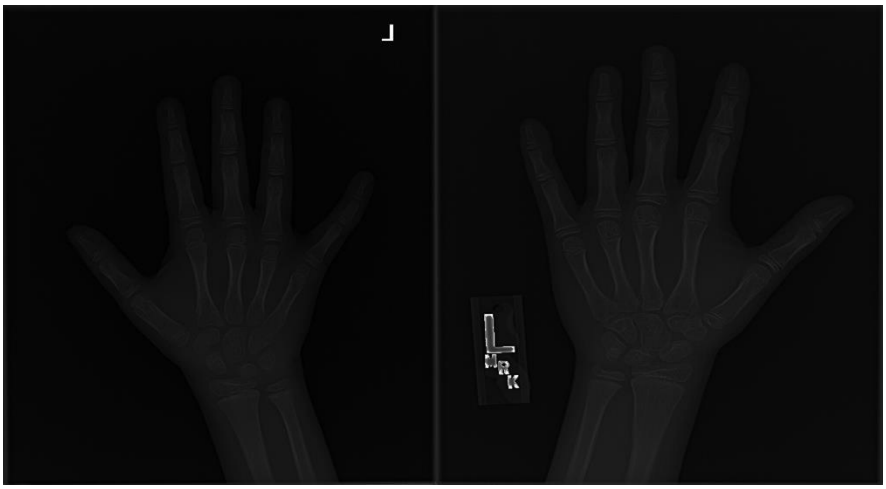

Figure 8: Examples of low contrast images

5.1.2 Draw Masks Manually

Removing the background from hand X-ray images is good for several reasons:

1. Improved Model Focus: By removing the background, the model can focus on the relevant anatomical structures, such as bones, joints, and tissues. This ensures that the model doesn't waste a lot of computational resources on irrelevant information, so this improves its efficiency.

2. Enhanced Accuracy: The model may incorrectly associate certain features of the background with the bones or other relevant structures, and this will reduce the overall performance of bone age estimation.

3. Faster Training: With fewer irrelevant pixels, the model can process the images faster, leading to reduced computational costs and less training time. This is especially beneficial for deep learning models with large datasets.

Images were collected at different hospitals and simple background removal methods did not produce satisfactory results. Thus, there is a compelling need for a reliable hand segmentation technique. After a lot of searches, we were unable to find a good source of labeled masks for hand X-ray images. To solve this problem, we first manually label 100 masks using the online annotation service Supervisely that takes approximately 2 min to process each image, see *Figure 9*. These masks are used to train the U-Net model that then is used to perform hand segmentation on the rest of the training set.

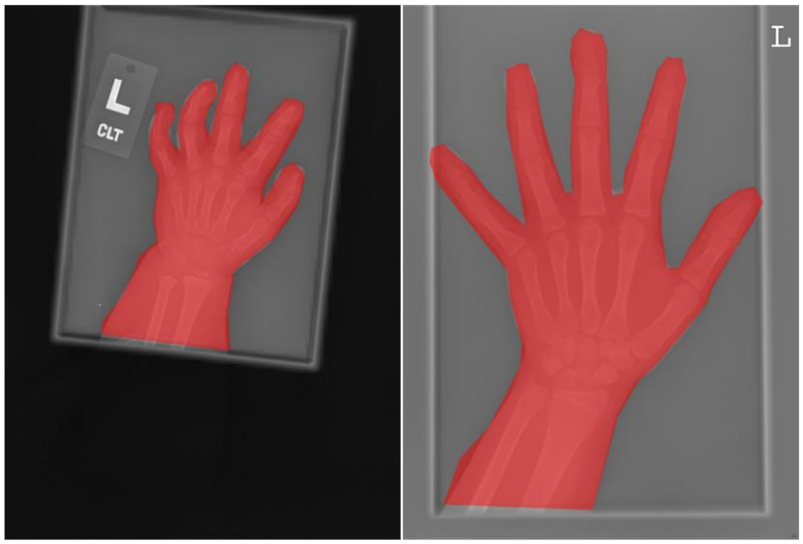

Figure 9: Drawing Masks Manually for 100 Image

After drawing 100 masks, we convert the red regions into white, see *Figure 10*, and the other regions into black using a simple code.

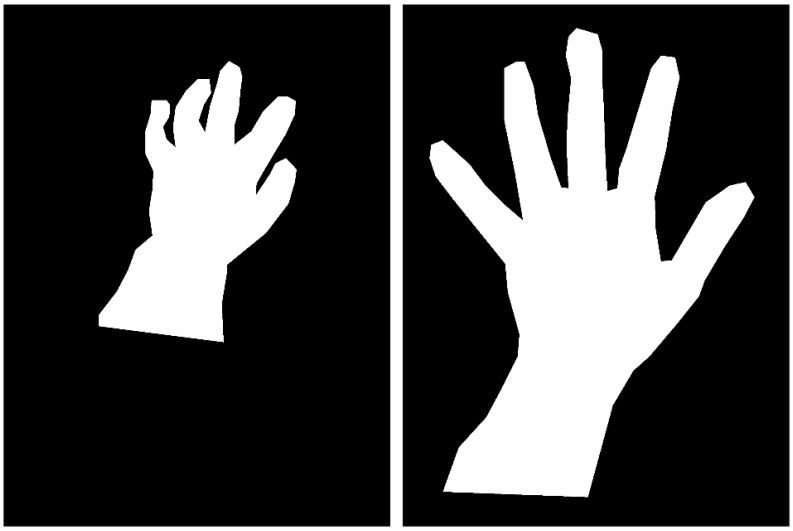

Figure 10: Binary Images

5.1.3 Create a Segmentation Model

Among the myriads of segmentation techniques and models that have emerged over the years, **U-Net** stands out as a **state-of-the-art solution** that has revolutionized the field. Developed by researchers at the Computer Science Department of the University of Freiburg in 2015, U-Net has gained widespread acclaim.

U-Net is a convolutional neural network (CNN) architecture that was specifically designed for biomedical image segmentation tasks. Developed in 2015, U-Net has become one of the go-to architectures for various segmentation tasks due to its effectiveness and efficiency.

The **U-Net architecture** is characterized by its U-shaped structure, which gives it its name, see *Figure 11*. It consists of an encoding path and a decoding path.

- **Encoding Path**: This part of the network captures the context of the input image by using a series of convolutional and max-pooling layers to downsample the spatial dimensions. It "contracs" the original images.
- **Decoding Path**: The decoding path uses upsampling and convolutional layers to produce a segmentation map that has the same spatial dimensions as the input image. It "expands" the contracted images.

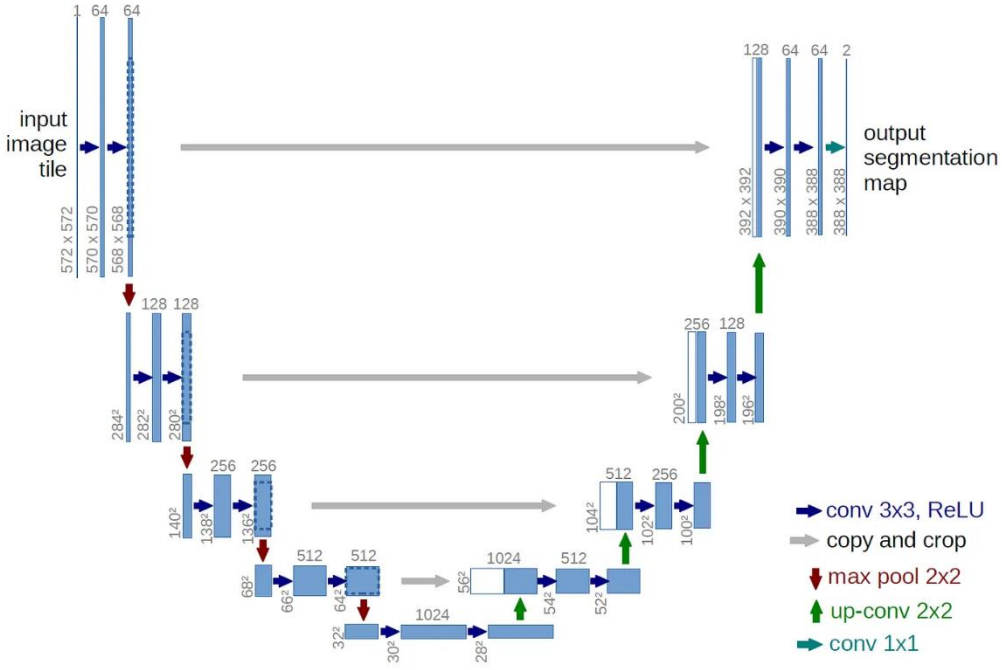

Figure 11: U-Net Architecture

U-Net's strength in segmentation comes from its use of **skip connections**, which **connect the encoding and decoding paths by merging features**. This helps retain spatial details lost during downsampling, preserving the image's local and global context. By maintaining this spatial information, U-Net achieves more accurate segmentation masks. The skip connections assist the network in grasping the relationships between image parts, leading to improved segmentation results.

Inside our code of hand X-ray segmentation, we use patch extraction because our images are big (high width and length), and we have just 100 images. Patch extraction is a crucial strategy in this scenario for several important reasons when dealing with high-resolution images and small dataset (100 image):

- Data Augmentation Through Patches: Each large image can generate multiple patches (hundreds depending on image size). Also, converting 100 large images into thousands of small patches effectively increases the training dataset and this will help in preventing overfitting and improving model generalization.

- Better Feature Learning: Patches help the model focus on learning local patterns and details. In addition, the overlapping strategy (stride = 64 or 96) ensures continuity in feature learning, and different features capture various contexts and object parts.

We use data augmentation to create modified versions of training data such as rotation, shifting, zooming and so on.

The entire training process includes:

- Loads and patches the data.
- Normalizes pixel values to [0,1].
- Augments the training data.
- Splits into training/validation sets (80/20).
- Reshapes data for model input.
- Creates and compiles the U-Net model.
- Trains for 50 epochs with binary cross-entropy loss.

We apply full image prediction function on large images by:

1. Breaking images into overlapping patches.
2. Making predictions on each patch.
3. Averaging overlapping predictions.
4. Reconstructing full-size segmentation mask.

The Key Implementation Details consist of:

- Image size: Works with 128×128 patches
- Overlapping strategy: Uses 96-pixel stride for prediction (32-pixel overlap)
- Normalization: Scales pixel values to [0,1]
- Model optimization: Uses Adam optimizer
- Loss function: Binary cross-entropy (suitable for binary segmentation)

After applying the segmentation model to the images, we obtain resulting mask images that focus on the target regions of interest. However, these masks often include unintended black regions within the segmented areas, which may compromise the accuracy and clarity of the segmentation. And because of that, we apply further refinement or post-processing techniques to ensure the masks accurately represent the desired features.

Here are some examples of the results after applying the segmentation models:

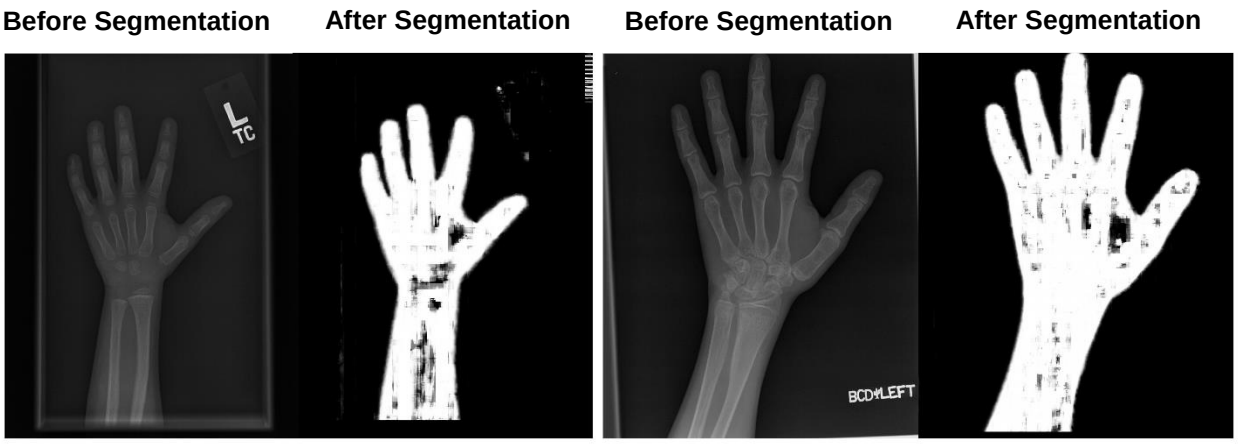

Figure 12: Resulted Masks After Applying Segmentation Model

5.1.4 Applying Contours to Resulted Masks

To resolve unintended black regions within the segmented area, we apply contour method to the images. Contouring is a fundamental technique in computer vision and image processing, playing a crucial role in identifying and extracting object boundaries within an image. OpenCV, a powerful open-source computer vision library, offers the cv2.findContours() function for contour detection.

The algorithm operates on a **binary image** where pixels have values of either 0 (background) or 255 (foreground). The binary image is typically obtained through the segmentation model we applied above. Below, we describe the process in detail:

1. Reading the Input Image: The function begins by loading the image using the cv2.imread method, which reads the file from the specified path. If the image cannot be loaded, an error message is displayed to prevent further processing.

2. Converting to Grayscale: To simplify subsequent processing, the loaded image is converted to grayscale using OpenCV's cv2.cvtcolor function. Grayscale images reduce computational complexity by eliminating color channels while retaining structural information essential for contour detection.

3. Thresholding: A binary threshold is applied to the grayscale image using cv2.threshold. Pixels with intensity values above a threshold (127 in this case) are set to white (255), while those below are set to black (0). This step creates a binary image where structures of interest stand out, making them easier to detect.

4. Contour Detection: Using OpenCV's cv2.findContours function, contours are extracted from the binary image. The function retrieves only the external contours (cv2.RETR_EXTERNAL) and applies a simplified chain approximation (cv2.CHAIN_APPROX_SIMPLE) to reduce the number of points representing each contour, optimizing memory usage.

5. Drawing Contours: The identified contours are superimposed onto the original image using cv2.drawContours. Each contour is drawn in green (0, 255, 0) with a thickness of 2 pixels, highlighting the boundaries of skeletal features or other regions of interest.

This step enhances the visual distinction of skeletal boundaries and other key features, facilitating accurate analysis and feature extraction. Contour detection is particularly useful for isolating regions of interest, such as growth plates or ossification centers, in subsequent preprocessing and modeling stages, see *Figure 13*.

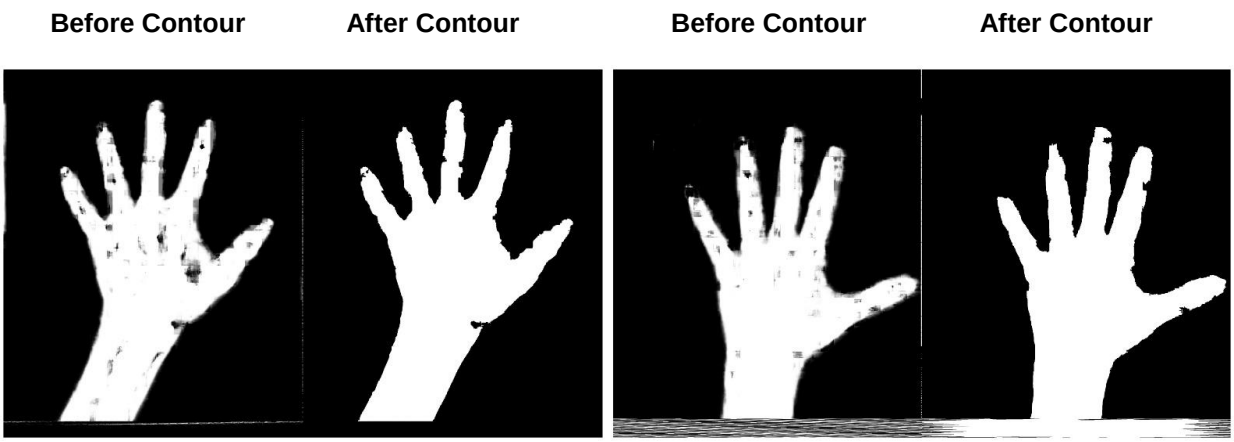

Figure 13: Before and after applying Contours to Resulted Masks

5.1.5 Apply Dilation to Images

Dilation is a critical morphological operation used in image preprocessing, particularly for medical imaging tasks like bone age estimation. The operation expands the boundaries of white regions (foreground) in a binary image, which is particularly useful for enhancing features such as skeletal structures, making regions like bones or growth plates more pronounced. It also helps bridge gaps by closing small holes or imperfections caused by noise or faulty segmentation. Additionally, dilation improves contour detection by thickening boundaries, ensuring better performance in subsequent detection steps. Below, we describe the process in detail:

1. Input and Output Folder Management: The function takes the path to the input folder containing the images and creates an output folder for storing the processed results. If the output folder does not exist, it is automatically created using os.makedirs.

2. Filtering Image Files: To ensure the function processes only valid image files, it filters the files in the folder by their extensions. Common image formats.

3. Loading and Validating Images: Each image is loaded using OpenCV's cv2.imread. If an image fails to load, the function skips it and prints an error message, ensuring the process continues for other files.

4. Grayscale Conversion: The loaded image is converted to grayscale using cv2.cvtColor, simplifying the image by removing color information. Grayscale images are essential for operations like thresholding and dilation, as they focus solely on intensity values.

5. Binary Thresholding: A binary threshold is applied to convert the grayscale image into a binary format, where pixel intensities above a threshold (127) are set to white (255), and those below are set to black (0). This step isolates the regions of interest for dilation.

6. Defining the Structuring Element: A kernel (structuring element) is created using NumPy's np.ones. The kernel size, specified as (11, 11) by default, determines the extent of dilation. Larger kernels result in greater expansion of white regions.

7. Applying Dilation: The cv2.dilate function performs the dilation operation. The iterations parameter, set to 1 by default, controls how many times the dilation is applied. Repeated iterations lead to progressively thicker boundaries.

Dilation plays a crucial role in preparing images for feature extraction and analysis. By enhancing key structures like growth plates, dilation makes them more distinguishable for machine learning models, improving the accuracy of the estimation process, see *Figure 14*. Additionally, dilation helps reduce noise by expanding connected components and closing gaps, which minimizes the impact of segmentation errors on downstream tasks. The thickened contours resulting from dilation also facilitate more accurate measurements of skeletal structures, making it easier to compute geometric and morphological features necessary for bone age analysis.

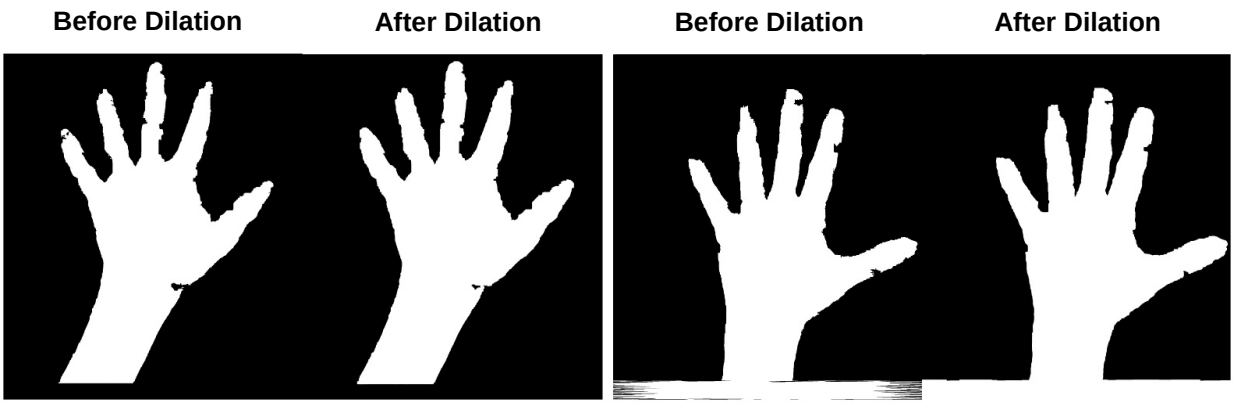

Figure 14: Before and after applying Dilation to Images

5.1.6 Apply overlay to images

This step in the preprocessing pipeline applies binary masks to isolate regions of interest in images, which is essential for tasks like feature extraction and model training. In medical imaging tasks, such as bone age estimation, masks help focus the analysis on specific areas, like bones, while ignoring irrelevant regions such as background or soft tissues. This targeted approach improves the signal-to-noise ratio by eliminating extraneous details, enhancing the quality of the data and the performance of machine learning models. Additionally, uniformly applying masks ensures consistency across the dataset, reducing variability that could impact the model's learning process. Below, we describe the process in detail:

1. Input and Output Folder Management: The function manages three folders, the images folder which contains the original images, the masks folder which holds the corresponding binary masks and the output folder, where the resultant masked images are saved. If the output folder does not already exist, it is created using os.makedirs.

2. Image and Mask File Matching: The function assumes that each image in the images folder has a corresponding mask with the same filename in the masks folder. It retrieves the list of image files using os.listdir and filters by common image extensions.

3. Loading Images and Masks: Each image is loaded using OpenCV's cv2.imread. The corresponding mask is loaded in grayscale using cv2.IMREAD_GRAYSCALE. This ensures the mask is a single-channel image suitable for binary thresholding.

4. Ensuring Binary Mask: To confirm the mask is binary, a thresholding operation is applied using cv2.threshold. This converts the mask into a binary format where pixel values are either 0 (black) or 255 (white).

5. Applying the Mask: The binary mask is applied to the original image using cv2.bitwise_and. This operation retains only the regions where the mask is white, effectively isolating the regions of interest. For pixels where the mask is black, the corresponding image pixels are set to zero (black).

Overlaying binary masks on wrist X-ray images is a crucial preprocessing step in bone age estimation. This process isolates skeletal structures, ensuring that only the relevant features are considered during model training and evaluation, see *Figure 15*.

By removing irrelevant regions, it reduces noise, allowing the model to learn more effectively. Masked images also have a consistent structure and format, which is essential for training robust and reliable deep learning models. The function is designed to handle batches of images and masks, making it scalable and suitable for large datasets, which is particularly important for medical imaging projects with thousands of samples. Automation ensures efficiency and reduces the potential for human error, making the process repeatable and reliable. In conclusion, by automating the application of binary masks, this function ensures high-quality, consistent preprocessed data, laying the foundation for effective feature extraction and classification in medical imaging tasks.

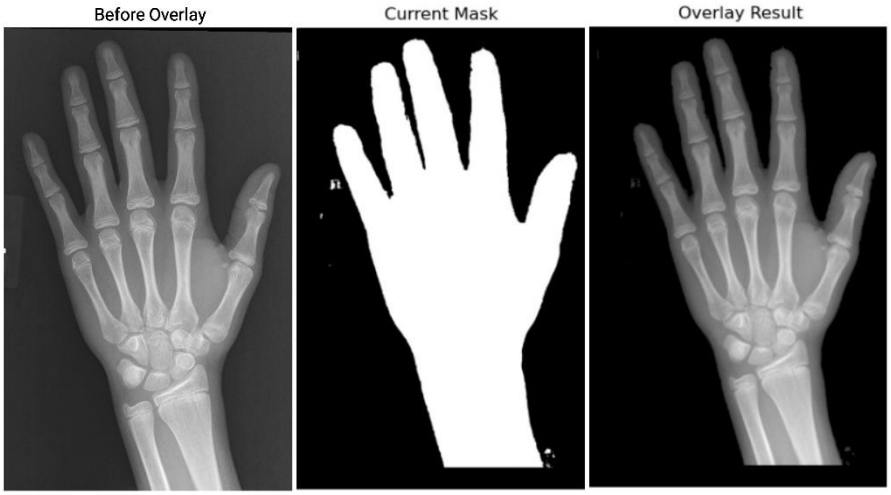

Figure 15: Images after applying overlay

5.1.7 Apply CLAHE to Images (Histogram Equalization)

Contrast Limited Adaptive Histogram Equalization (CLAHE) is a preprocessing technique used to enhance the quality of X-ray images by improving the visibility of finer details in localized regions. This step is crucial for tasks like bone age estimation, as it enhances contrast and makes subtle features more distinguishable, ultimately improving the performance of machine learning models. Unlike global histogram equalization, which adjusts contrast uniformly across the entire image, CLAHE operates on small, localized regions called tiles.This localized approach prevents over-amplification of noise, a common issue in global methods, particularly in low-quality X-ray images. By applying a clipping limit to the histogram, CLAHE ensures no specific intensity value becomes overly dominant, resulting in enhanced image contrast without introducing excessive noise. Below, we describe the process in detail:

1. Input and Output Folder Management: The source folder contains the original overlayed images that require enhancement and the destination folder is created dynamically if it doesn't exist, ensuring a separate location to store the processed images.

2. Iterating Over Images: The code iterates through all files in the source folder, filtering only valid image ignoring non-image files.

3. Loading Images in Grayscale: Each image is loaded in grayscale using cv2.IMREAD_GRAYSCALE. Grayscale images reduce computational complexity and focus on structural details.

4. Creating and Configuring the CLAHE Object: The CLAHE object is initialized with two key parameters, clipLimit=3.0, which controls contrast enhancement by limiting the amplification of histogram peaks, and tileGridSize=(8, 8), which defines the grid size for histogram equalization. Smaller tiles enhance finer details, while larger tiles provide broader contrast adjustments.

5. Applying CLAHE: The apply method processes each grayscale image, enhancing contrast in localized regions. This step is particularly effective for highlighting the bone structures in X-rays.

This CLAHE application step enhances the contrast of X-ray images, making key anatomical features more prominent for subsequent analysis, see *Figure 16*. By improving image quality, this preprocessing step directly contributes to the accuracy and robustness of the bone age estimation model. The systematic application of CLAHE ensures that the dataset is prepared for feature extraction and model training, laying the foundation for reliable predictions.

Figure 16: Images after applying CLAHE

5.1.8 Apply Sharpening to Images

This section describes the implementation of Laplacian sharpening on CLAHE-processed images to further enhance their visual clarity and accentuate edges. This technique is particularly useful for improving the visibility of fine structural details in X-ray images, such as bone edges and growth plates, which are critical for bone age estimation. Below, we describe the process in detail:

1. Input and Output Folder Management: The script processes all images in a source folder (CLAHE) and saves the sharpened versions to a destination folder (SHARPENING). If the destination folder does not exist, it is created dynamically to store the output images.

2. Grayscale Conversion: Images are converted to grayscale using OpenCV's cv2.cvtColor function. This step is crucial because the Laplacian operator is typically applied to single-channel grayscale images for edge detection.

3. Laplacian Filter Application: The Laplacian filter is computed using cv2.Laplacian, which calculates the second-order derivatives of the grayscale image. This operation highlights regions with rapid intensity changes, such as edges.

4. Sharpening Process: The sharpened image is obtained by subtracting the Laplacian result from the original grayscale image. This operation enhances edges while preserving the overall structure of the image.

Laplacian sharpening is a powerful preprocessing technique that enhances the clarity of medical images by emphasizing structural details, see *Figure 17*. By applying this method to CLAHE-processed X-ray images, the preprocessing pipeline prepares the dataset for more accurate and reliable feature extraction, ultimately contributing to the success of the bone age estimation task.

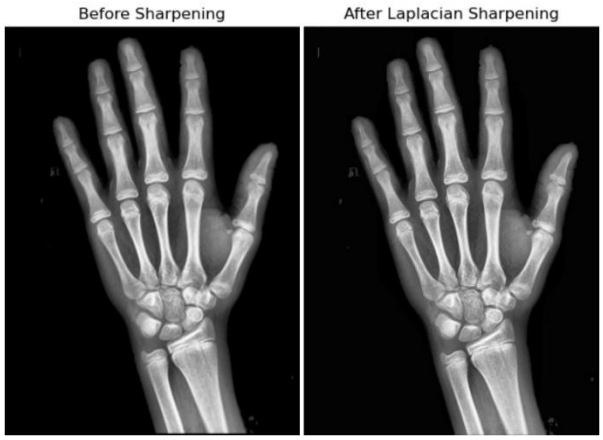

Figure 17: Images after applying Sharpening

5.1.9 Image Resizing

All images were resized to a consistent size of (384 × 384) pixels. The resizing ensures that the input dimensions are uniform across all images, which is essential for the neural network.

5.1.10 Standardize Pixel Values

To improve training stability, the pixel values were normalized to the range [1,0] by dividing by 255.

5.1.11 Data Encoding

For the gender data, the gender of each subject was encoded into binary numerical values. Male subjects were assigned a value of 1, and female subjects were assigned a value of 0. This encoding transformed the categorical gender feature into a format suitable for input into the model.

5.1.12 Weighted Augmentation

To improve the model's generalization ability and prevent overfitting, data augmentation was applied. Augmentation helps to artificially increase the size of the training dataset by introducing variations in the images. The augmentation techniques included random rotations, random translations, random zooms, and contrast adjustments. These transformations were designed to mimic real-world variations that might occur during image acquisition, such as different angles, positions, and lighting conditions.

A custom augmentation generator was also developed to address class imbalances in the dataset. Some age categories are underrepresented, which can lead to model bias. To counter this, we calculate category weights inversely proportional to the frequency of each bone age category. Categories with fewer samples receive higher weights, which are then normalized so that the sum of all weights equals 1.

The idea is that underrepresented categories will have more augmentations applied, while overrepresented categories will receive fewer. This ensures the model doesn’t become biased toward more frequent categories. Augmentations include random transformations like rotation, translation, zooming, and contrast adjustments to introduce variability into the dataset.

To decide how many augmentations to generate for each image, the weight of the image’s category is used. For instance, a category with a higher weight gets more augmentations. This weighted approach helps the model learn more from underrepresented categories and improves its generalization ability.

A data generator is used to handle the augmentation process efficiently. It loads and augments images in batches, applying the correct number of augmentations based on the category's weight. The generator then returns both the augmented images and their corresponding bone age values for training.

After augmentation, the original and augmented data are combined, resulting in a more balanced dataset. This increased diversity helps the model generalize better and reduces the risk of overfitting. The effectiveness of the augmentation is evaluated by comparing the distribution of bone age categories in the original and augmented datasets, see *Figure 18.*

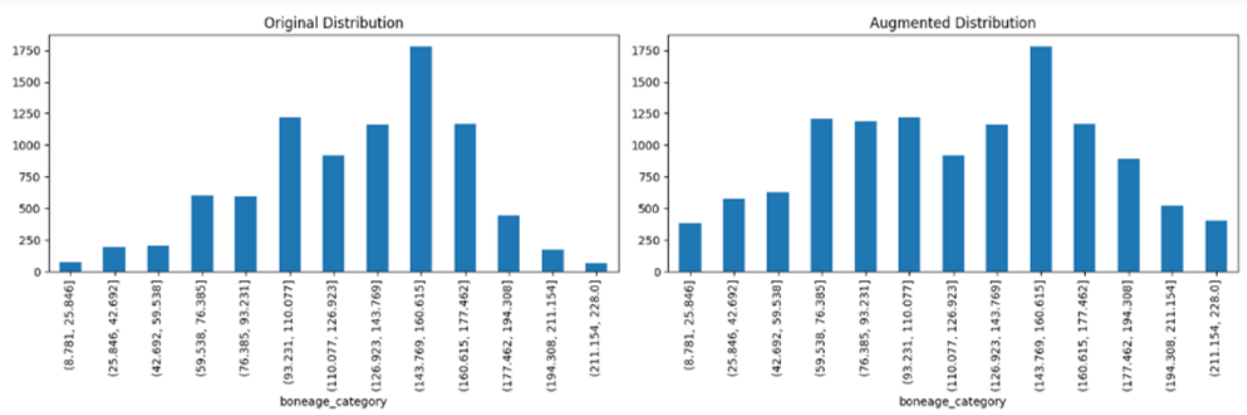

Figure 18:The distribution of bone age categories in the original and augmented datasets

5.2 Model Development and Training

This section covers the architecture design and training strategies used to optimize the model's performance. It also provides a detailed overview of the steps taken to ensure the model learns from the data efficiently and generalizes well to unseen examples.

5.2.1 Model Architecture

The core of the model is based on the ResNet-34 architecture, which is designed to capture features from both the images and the gender data. The architecture starts with an initial convolutional layer followed by a max-pooling layer, which reduces the spatial dimensions of the image while simultaneously extracting important features. The ResNet-34 model uses residual blocks, which consist of skip connections to mitigate the vanishing gradient problem and allow for the training of deeper networks, see *Figure 19.*

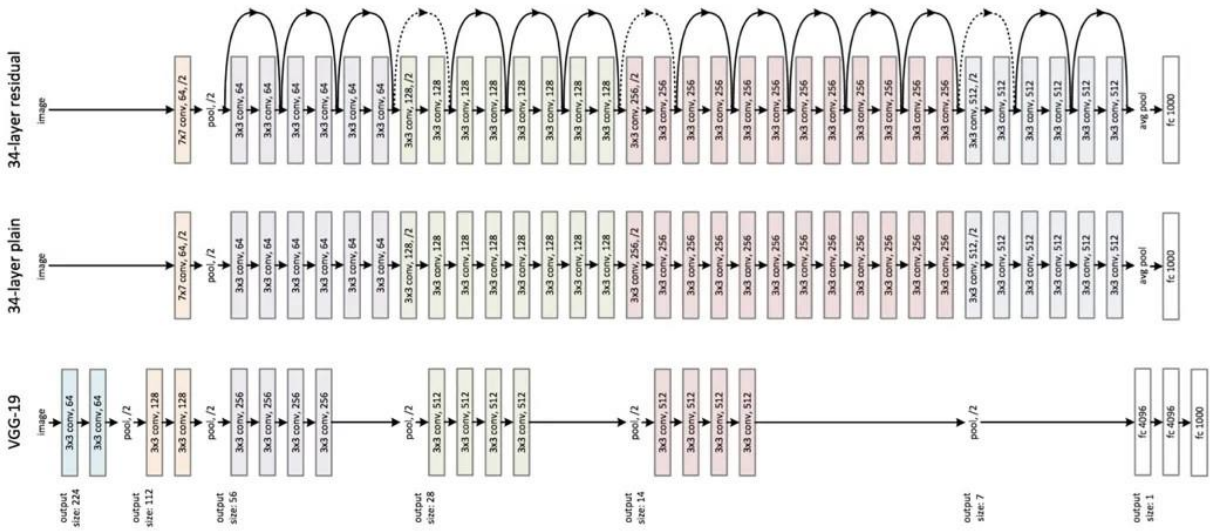

Figure 19:ResNet-34 architecture compared to othe architectures

ResNet-34 architecture is hierarchical, with increasing filter sizes as we move deeper into the network. The first block uses 64 filters, the second block uses 128 filters, followed by 256 filters in the third block, and 512 filters in the final block. These blocks are responsible for progressively extracting higher-level features from the input images.

A second input for gender is incorporated into the model to account for physiological differences between males and females. This input is concatenated with the feature representation extracted from the image data after a global average pooling operation. The model then passes this concatenated information through two fully connected layers, each with 2048 neurons. Dropout regularization is applied in these layers to prevent overfitting. Finally, the output layer consists of a single neuron with a linear activation function, as the task is a regression problem predicting bone age.

5.2.2 Training Pipeline

The model was trained using Mean Absolute Error (MAE) as the loss function. MAE was chosen because it directly measures the average magnitude of the errors, without considering their direction. This is a suitable metric for regression tasks, especially when interpretability is desired. The model was optimized using the Adam optimizer with a learning rate of 1×10^{-3}. Adam is a widely-used optimization algorithm known for adapting learning rates during training, which helps achieve better convergence.

Several callbacks were integrated to monitor training and improve the model's efficiency. The ReduceLROnPlateau callback was used to reduce the learning rate by a factor of 0.1 if the validation loss plateaued for 10 epochs. This helps the model fine-tune its parameters.

The EarlyStopping callback was employed to halt training if the validation loss did not improve for 30 epochs, preventing unnecessary computation and overfitting. The ModelCheckpoint callback saved the best model based on the validation loss during training. Finally, a custom ModelSaver callback was created to save the model at regular intervals (every 7 epochs) to allow for recovery in case of interruptions.

In addition to the initial training, the model was fine-tuned starting from epoch 7, leveraging a previously saved checkpoint. This approach allows the model to build upon its existing knowledge and continue improving without starting from scratch.

6.0 Results

The model's performance was evaluated by comparing its predictions to the actual values. A scatter plot was generated to visualize the relationship between the predicted and actual values. As shown in *Figure 20*, the model's predictions generally follow the trend of the actual values, indicating a positive correlation and that it has effectively captured the underlying patterns in the data.

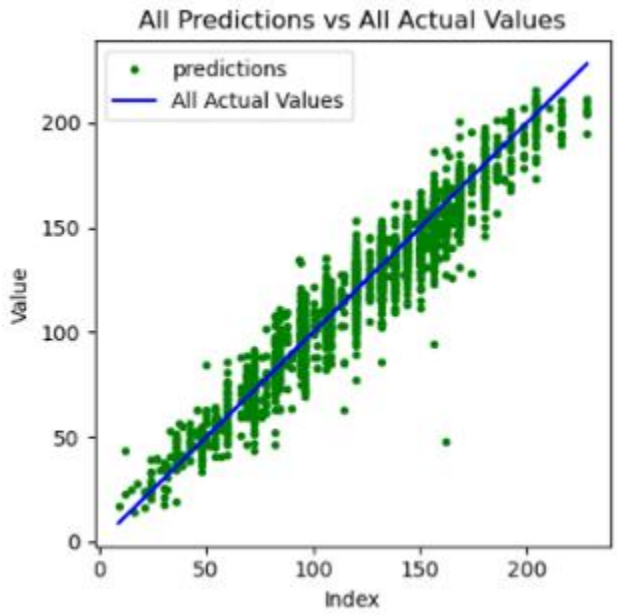

Figure 20: Scatter plot of predictions vs actual value

However, the presence of scatter around the diagonal line suggests that the model's predictions are not always perfectly accurate, leaving room for improvement in its precision.

Here are some other results from different models trained on the same dataset:

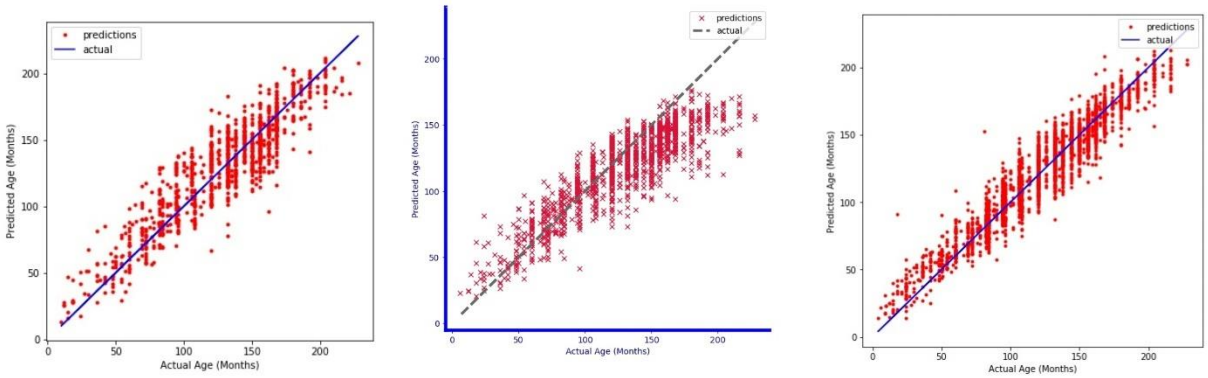

Figure 21: Results of other different models trained on RSNA dataset

In the first plot, the predictions are closely aligned with the actual values, indicating a strong correlation between them. While the predictions generally match the actual values, there are some slight deviations, particularly at higher ages. These deviations suggest the model is performing well overall but may have minor inaccuracies in specific cases, especially for older age groups. Despite these minor outliers, the clustering of points near the diagonal line demonstrates that the model achieves a high level of predictive accuracy.

In the second plot, the predictions still follow the general trend of the actual values, indicating a similar level of performance as the first plot. However, upon closer inspection, the scatter appears slightly more dispersed compared to the first plot. This wider spread implies that the prediction errors might be slightly larger in this scenario, especially for certain age ranges. Nevertheless, the majority of the points remain reasonably close to the actual values, demonstrating that the model's predictions are still accurate, albeit with some increase in variability.

The third plot resembles the first plot in terms of its results. The predictions are tightly clustered around the actual values, signifying a strong agreement between predictions and true ages. This indicates that the model consistently performs well across a wide range of ages. The scatter around the diagonal appears minimal, which further reinforces the notion that the model exhibits robust predictive accuracy. As in the first plot, there may be occasional outliers, particularly for higher age groups, but these deviations are relatively minor and do not significantly impact the overall performance.

When comparing these three plots to our result, our model clearly demonstrates the best predictive performance.

The first plot shows good accuracy, with predictions closely following the actual values, but there is some scatter, especially at higher ages. The second plot has more variability, with a wider spread of predictions around the actual values, indicating larger errors. The third plot improves upon the first and second, with tighter clustering of predictions around the actual values, but still shows slight deviations in the upper age ranges.

Our results, however, stands out with minimal scatter and the tightest alignment of predictions to the actual values across all ages. It demonstrates superior precision and reliability compared to these three plots, making it the best performing model overall.

7.0 Conclusion

In summary, we developed a deep learning model using the ResNet-34 architecture for bone age prediction from radiographic images. Through careful data preprocessing, feature extraction, and the application of weighted augmentation techniques, we improved the model's robustness and predictive accuracy. The model was trained on RSNA which is a comprehensive dataset of bone age images, with results indicating promising accuracy and generalization capabilities. By incorporating weighted augmentation, we ensured that the model could effectively handle diverse bone age categories, contributing to its ability to generalize well across different age groups. The model also underwent rigorous evaluation to minimize loss and fine-tune hyperparameters, leading to improved prediction performance. This work highlights the potential of deep learning for automating bone age estimation, offering a reliable tool for clinical applications in pediatric radiology and diagnostics. Moving forward, further optimization and the inclusion of additional data sources could enhance the model's accuracy, solidifying its potential as a powerful tool in the medical field for more precise bone age assessment.

Table of figures

Table of tables

References

[1] D. D. Martin, J. M. Wit, Z. Hochberg, L. Sävendahl, R. R. v. Rijn, O. Fricke, N. Cameron, J. Caliebe, T. Hertel, D. Kiepe, K. Albertsson-Wikland, H. H. Thodberg, G. Binder and M. B. Ranke, "The Use of Bone Age in Clinical Practice – Part 1".

[2] J. H. Lee, Y. J. Kim and K. G. Kim, "Bone age estimation using deep learning and hand X-ray images".

[3] Z. Wang , "Probing an AI regression model for hand bone age determination using gradient-based saliency mapping".

[4] F. H. Tang, J. L. Chan and B. K. Chan, "Accurate Age Determination for Adolescents Using Magnetic Resonance Imaging of the Hand and Wrist with an Artificial Neural Network-Based Approach".

[5] P. Sharma, "Bone age estimation with HS-optimized Resnet and Yolo for child growth disorder".

[6] V. Iglovikov, A. Rakhlin, A. A.Kalinin and A. Shvets, "Pediatric Bone Age Assessment Using Deep Convolutional Neural Networks".

[7] L.-S. O. Müller, A. Offiah, C. Adamsbaum, I. Barber, P. L. D. Paolo, P. Humphries, S. Shelmerdine, L. T. D. Horatio, P. Toma, CatherineTreguier and K. Rosendahl, "Bone age for chronological age determination: statement of the European Society of Pediatric Radiology musculoskeletal task force group".

[8] M. Zulkifley, S. Abdani and N. Zulkifley, "Automated Bone Age Assessment with Image Registration Using Hand X-ray Images".

[9] J. He and D. Jiang, "Fully Automatic Model Based on SE-ResNet for Bone Age Assessment".

[10] B. LIU, Y. ZHANG, M. CHU, X. BAI and F. ZHOU, "Bone Age Assessment Based on Rank-Monotonicity Enhanced Ranking CNN".

[11] E. Siegel, "What Can We Learn from the RSNA Pediatric Bone Age Machine Learning Challenge?".

[12] N. Reddy, J. Rayan, A. Annapragada, N. Mahmood, A. Scheslinger, W. Zhang and J. H. Kan, "Bone age determination using only the index finger: a novel approach using a convolutional neural network compared with human radiologists".

[13] A. L. Dallora, P. Anderberg, O. Kvist, E. Mendes, S. D. Ruiz and J. S. Berglund, "Bone age assessment with various machine learning techniques: A systematic literature review and meta-analysis".

[14] J. Zhou, Z. Li, W. Zhi, B. Liang, D. Moses and L. Dawes, "Using Convolutional Neural Networks and Transfer Learning for Bone Age Classification. In Proceedings of the 2017 International Conference on Digital Image Computing: Techniques and Applications".

[15] F. H. Tang, J. L. C. Chan and B. K. L. Chan, "Accurate age determination for adolescents using magnetic resonance imaging of the hand and wrist with an artificial neural network-based approach".

[16] F. Cavallo, A. Mohn, F. Chiarelli and C. Giannini, "Evaluation of Bone Age in Children: A Mini-Review".

[17] S. Anwar, M. Majid, A. Qayyum, M. Awais, M. Alnowami and M. Khan, "Medical Image Analysis using Convolutional Neural Networks: A Review".

[18] V. Gilsanz and O. Ratib, Hand Bone Age: A Digital Atlas of Skeletal Maturity.

[19] F. Cavallo, A. Mohn, F. Chiarelli and C. Giannini, "Evaluation of Bone Age in Children: A Mini-Review".

[20] K. Doi, "Computer-Aided Diagnosis in Medical Imaging: Historical Review,Current Status and Future Potential".

[21] S. Li, B. Liu, S. Li, X. Zhu, Y. Yan and D. Zhang, "A deep learning-based computer-aided diagnosis method of X-ray images for bone age assessment".

[22] H. Lee, S. Tajmir, J. Lee, M. Zissen, B. A. Yeshiwas, T. K. Alkasab, G. Choy and S. Do, "Fully Automated Deep Learning System for Bone Age Assessment".

[23] M. Z. Alom, T. M. Taha, C. Yakopcic, S. Westberg, M. Hasan, B. C. V. Esesnm, A. A. S. Awwal and V. K. Asari, "The History Began from AlexNet: A Comprehensive Survey on Deep Learning Approaches".

[24] C. Szegedy, V. Vanhoucke, S. Ioffe and J. Shlens, "Rethinking the Inception Architecture for Computer Vision".

[25] Z. Wu, C. Shen and A. Hengel, "Wider or Deeper: Revisiting the ResNet Model for Visual Recognition".

[26] W. Sheng, J. Shen, Q. Huang, Z. Liu, J. Lin, Q. Zhu and L. Zhou, "Symmetry-Based Fusion Algorithm for Bone Age Detection with YOLOv5 and ResNet34".

[27] Y. Deng, Y. Chen, Q. He, X. Wang, Y. Liao, J. Liu, Z. Liu, J. Huang and T. Song, "Bone age assessment from articular surface and epiphysis using deep neural networks".

[28] S. Wang, X. Wang, Y. Shen, B. He, X. Zhao, P. Cheung, J. Cheung, K. Luk and Y. Hu, "An Ensemble based Densely-Connected Deep Learning System for Assessment of Skeletal Maturity".

[29] K. Wan, B. Feng, L. Xie and Y. Ding, "Reconciling Feature-Reuse and Overfitting in DenseNet with Specialized Dropout".

[30] S. Halabi, L. Prevedello, J. Kalpathy-Cramer, A. Bilbily, M. Cicero, F. Kitamura, K. Andriole and A. Flanders, "The RSNA Pediatric Bone Age Machine Learning Challenge".