

UDC (004.032.26:612.65-053.2)

DOI: <https://doi.org/10.17721/3041-1491/2024.11-42>

Ivan RIABKO^{1,2}, Master's Student

ORCID ID: 0009-0000-6748-8686

e-mail: vanoryabko@gmail.com

Andrii TERETS^{1,2}, Master's Student

ORCID ID: 0000-0003-3668-6131

e-mail: terets.andrii@gmail.com

Stanislav RIEBIENKOV², Head of the Department

ORCID ID: 0000-0001-8116-5277

e-mail: rebenkov@gmail.com

Sergiy RADCHENKO¹, PhD (Phys. & Math.), Assoc. Prof.

ORCID ID: 0000-0001-9153-5470

e-mail: sprad@knu.ua

¹Taras Shevchenko National University of Kyiv, Kyiv, Ukraine,

²National Children's Specialized Hospital "OHMATDYT", Kyiv, Ukraine

ASSESSING A CHILD'S BONE AGE USING DEEP LEARNING ON HAND X-RAY IMAGES

Introduction. *Assessing children's bone age is essential for detecting growth and developmental disorders. By comparing bone age to chronological age, physicians can identify deviations and diagnose issues.*

Methods. *In this study, we used a neural network model to predict the bone age of children based on X-ray images of their hands.*

Results. *The effectiveness of such application of the model is demonstrated in comparison with the average error results on the validation dataset.*

Conclusions. *Our study has demonstrated that the DenseNet169 architecture, when integrated with patient gender information, can effectively predict bone age from X-ray images, yielding competitive results compared to established models like Milvue. We hope this approach will significantly improve prediction accuracy and provide deeper insights into bone development. By focusing on critical anatomical regions and integrating relevant clinical data, we aim to set new standards in automated bone age assessment, ultimately benefiting pediatric healthcare.*

Keywords: *DenseNet169, Bone Age Assessment, X-ray Image Analysis, Statistical Estimation.*

Introduction

Assessing children's bone age is essential for detecting growth and developmental disorders (Greulich, & Pyle, 1959). By comparing bone age to chronological age, physicians can identify deviations and diagnose issues. Hand X-ray analysis is a key tool in this process, offering a clear view of bone development. Automating this assessment using neural networks improves accuracy and objectivity, reducing human error (Lindner et al., 2020). In this study, we propose and evaluate a neural network model trained on X-ray images, using mean absolute error as the metric.

Recent advances in deep learning for predicting bone age have gained attention due to their potential to automate and enhance accuracy in pediatric diagnostics (Lindner et al., 2020; Jabbar, & Abdulmunem, 2023). Studies have shown the effectiveness of convolutional neural networks (CNNs) in this domain (Tan, & Le, 2020; Huang et al., 2017). The RSNA Bone Age Challenge provided a large dataset, benchmarking models like ResNet and DenseNet for bone age prediction (RSNA Bone Age Competition Dataset: <https://www.kaggle.com/kmader/rsna-bone-age>). DenseNet169, in particular, improves accuracy by incorporating patient data like gender (Huang et al., 2017).

Data augmentation techniques, such as scaling, rotation, and flipping, have been crucial in addressing image variability and preventing overfitting (Shorten, & Khoshgoftaar, 2019). Including clinical features like gender further refines predictions by accounting for biological differences in bone development (Jabbar, & Abdulmunem, 2023). Fine-tuning pre-trained models on medical datasets, such as ImageNet, enhances performance and reduces computational demands (Tan, & Le, 2020).

Overall, the combination of deep learning architectures with large datasets positions automated bone age estimation as a reliable tool in pediatric healthcare. These advancements have the potential to surpass traditional methods in speed and accuracy, making them valuable in clinical applications.

Methods

Model Architecture and Methodology. In this study, a transfer learning model based on DenseNet169 was used for bone age assessment (Huang et al., 2017). The model was adapted to work with X-ray images of hands and patient gender information. In addition to the image, to enhance the accuracy of age prediction, additional input—patient gender information—was included. This information is first converted into numerical format and passed through a separate fully connected layer with 32 neurons and a ReLU activation function.

The extracted features from DenseNet169 and the gender information are then combined and passed through two fully connected layers with 512 and 256 neurons respectively, before reaching the output layer with a single neuron and a linear activation function, which predicts the age in months.

Data set's structure and properties. The RSNA Bone Age dataset (RSNA Bone Age Competition Dataset: <https://www.kaggle.com/kmader/rsna-bone-age>) contained 12,606 X-ray images of hands with images distribution by children age (Fig. 1). The dataset was divided into two independent groups: 90 % for model training and 10 % for testing. The convolutional neural network DenseNet169 was employed for predicting bone age as this model type is designed to perform image classification (Densely Connected Convolutional Networks: <https://arxiv.org/abs/1608.06993>), Skeletal Development of the Hand and Wrist (Greulich, & Pyle, 1959).

To further enhance the model's quality and address a wide range of potential image imperfections, various data augmentation techniques were meticulously applied (Shorten, & Khoshgoftaar, 2019; Jabbar, & Abdulmunem, 2023). These imperfections, which could include issues such as dark or underexposed images, misaligned or inverted images, and varying lighting conditions, have the potential to

negatively impact the accuracy of the model. By applying data augmentation, we aimed to simulate a broader variety of scenarios that the model might encounter in real-world applications, thus improving its robustness and adaptability.

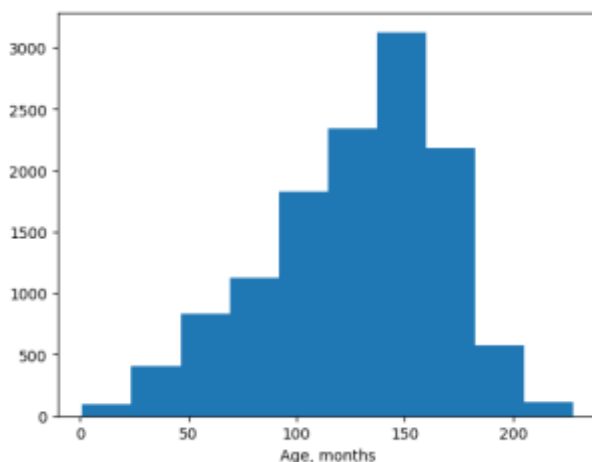


Fig. 1. Age category distribution

The data augmentation process involved a series of carefully selected image transformations, such as scaling, shifting, rotation, zooming, and horizontal flipping. Each of these transformations plays a crucial role in increasing the diversity of the training data, which in turn helps to prevent overfitting – a common issue in deep learning models where the model performs well on the training data but poorly on unseen data. By exposing the model to a wider variety of augmented images, we aimed to enhance its ability to generalize across different imaging conditions and patient populations.

For the purpose of maintaining consistency in the input data, all X-ray images were uniformly scaled to a resolution of 500×500 pixels. This standardization ensures that the model processes each image with the same level of detail, which is critical for extracting consistent features across the entire dataset. Additionally, to further diversify the training data, transformations such as random rotation and horizontal flipping were repeatedly applied, creating a more comprehensive set of input images for the model to learn from (Alhassan, & Mumuni, 2022).

In addition to image augmentation, the age data was normalized to ensure that the model could efficiently learn from it without being biased by the absolute scale of the values. Gender information, another crucial input feature, was carefully converted into a binary format, allowing the model to seamlessly integrate this categorical data alongside the numerical image data during training. This combination of data augmentation and careful preprocessing contributed significantly to the model's ability to learn effectively from the available data, ultimately improving its performance in predicting bone age (Fig. 2).

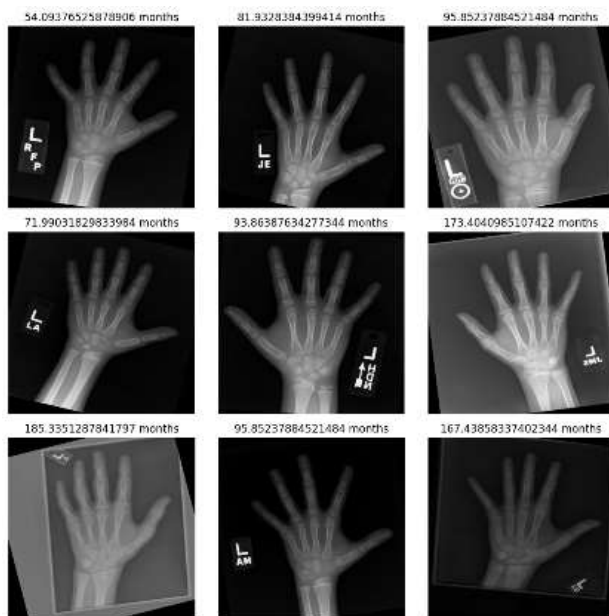


Fig. 2. Example of a training batch

Two CSV tables were prepared for the RSNA Bone Age dataset: one for the training set and one for the test set. These tables contain the following information:

- **ID:** A unique identifier for the X-ray image.
- **BoneAge:** The bone age in months, as determined by radiologists.
- **Gender:** The gender of the child ("1" denotes boys, "0" – girls).

Training and Evaluation. The model was trained using the L1 Loss function, which measures the absolute differences between predicted and actual values. The Adam optimizer, known for its efficiency in providing adaptive learning rates, was employed during the training process. The model was trained for 100 epochs, where an epoch refers to a complete pass through the entire training dataset. The primary evaluation metric was L1 Loss, calculated on both the training and validation sets. After 100 epochs, the model achieved a Mean Absolute Distance (MAD) of 5.2 months on the validation set, indicating reliable predictions for unseen X-ray images. To monitor the optimization of the model's parameters, we tracked the loss function throughout the training process. Fig. 3 illustrates the convergence of the loss function, showing a consistent decrease over time, which indicates effective learning by the model. Additionally, testing on a holdout test set confirmed that the model generalizes well, providing similar MAD values (Fig. 5), where the actual values versus predicted values on the validation dataset are shown.

Results

Throughout our research, we conducted a comprehensive comparative analysis between our neural network model and the well-known French model, Milvue (Milvue, n. d.), using a robust set of clinical data. The analysis focused on X-ray images of the

hands, which represent one of the most reliable and widely used methods for assessing bone age in both clinical and research settings. The hands contain numerous bones at various stages of development, making them an ideal site for examining skeletal growth and maturity, which is particularly relevant in pediatric assessments.

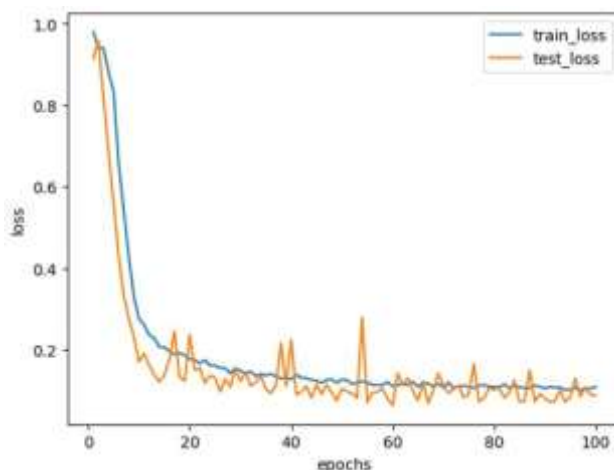


Fig. 3. Loss depending on the eras

Our findings showed that the L1 Loss for the Milvue model was 9.6 months, while our model achieved an L1 Loss of 11.2 months. Although this 1.6-month difference indicates that our model's accuracy is slightly lower, it is crucial to highlight that the difference is relatively minor. Additionally, our model achieved a Mean Absolute Distance (MAD) of 5.2 months on the validation set, indicating reliable predictions for unseen X-ray images. The performance of our model falls well within the expected range, particularly when considering results from the RSNA Bone Age test set. This demonstrates that despite the slight discrepancy in accuracy, our model is still highly capable of delivering reliable and clinically useful predictions.

The graph below provides a visual representation of the comparison between the two models. The X-axis represents the actual chronological age of the patients, while the Y-axis indicates the predicted age as calculated by the models. In the graph, predictions made by our model are illustrated in blue, while those made by the Milvue model are shown in red (Fig. 4-5).

This visual comparison shows that both models perform consistently across various age groups, reinforcing the clinical utility of our model. The higher L1 Loss in our model highlights areas for improvement, particularly in image preprocessing and focusing on growth zones, which are essential for understanding skeletal maturity. Our analysis reveals opportunities for enhancement. We plan to integrate a more advanced architecture focused on regions of interest to prioritize the most informative parts of the images for improved precision in bone age predictions. Additionally, we will implement a dual-model approach: one model will segment growth zones, while the other will predict age based on information from these

regions. We will test this new architecture on the same dataset as our initial model to ensure comparable results and validate its effectiveness with real clinical data. In conclusion, our model shows competitive performance and potential for development. By refining the architecture and targeting key anatomical regions, we aim to achieve greater accuracy in bone age assessment.

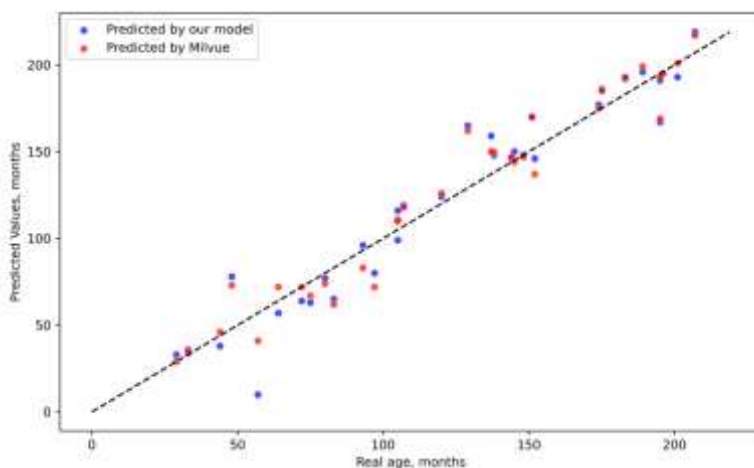


Fig. 4. Comparison our model with Milvie model.
Blue dots denotes our model prediction, and the red milvie prediction

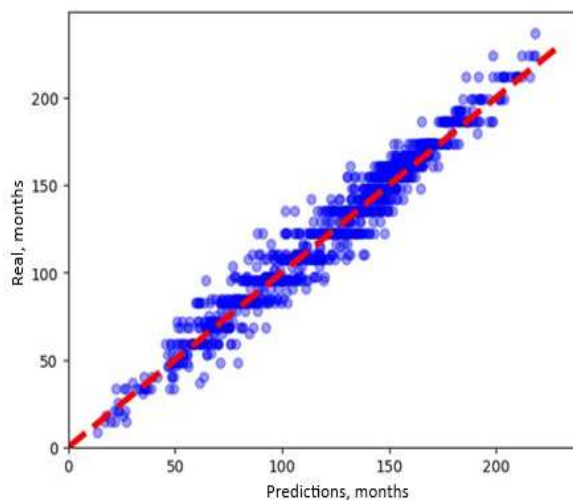


Fig. 5. Actual values versus predicted on validation dataset

Discussion and conclusions

Our study has demonstrated that the DenseNet169 architecture, when integrated with patient gender information, can effectively predict bone age from X-ray images, yielding competitive results compared to established models like Milvie (Milvie, n.d.).

While our model had a slightly higher L1 Loss compared to Milvue, the difference was minimal, underscoring the robustness of our model for clinical use. On the validation set of the RSNA dataset, we achieved a MAD of 5.2 months. To further improve prediction accuracy, we plan to implement a more advanced architecture that focuses on regions of interest, particularly growth zones. This approach will involve using a combination of models: one for segmenting growth zones and another for predicting age based on these segments. We believe this will enhance the precision of our analysis and improve the accuracy of bone age prediction. The new architecture will be tested on the same dataset as our original model to ensure comparability and will also be validated on real clinical data for practical application. We hope this approach will significantly improve prediction accuracy and provide deeper insights into bone development. In conclusion, our research highlights the importance of continually refining deep learning models in medical imaging. By focusing on critical anatomical regions and integrating relevant clinical data, we aim to set new standards in automated bone age assessment, ultimately benefiting pediatric healthcare.

Authors' contribution: Ivan Riabko – software, writing, original draft, methodology; Andrii Terets – software, methodology, validation; Stanislav Riebienkov – conceptualization, methodology, formal analysis; Sergiy Radchenko – writing, review & editing.

References

- Alhassan, M., & Mumuni, F. (2022). Data augmentation: A comprehensive survey of modern approaches. *Array*, 16, 100258. <https://doi.org/10.1016/j.array.2022.100258>
- Greulich, W.W., & Pyle, S.I. (1959). *Radiographic Atlas of Skeletal Development of the Hand and Wrist* (2nd ed.). Stanford University Press. <https://www.cambridge.org/core/journals/amg-acta-geneticae-medicae-et-gemellologiae-twin-research/article/w-w-greulich-and-s-i-pyle-radiographic-atlas-of-skeletal-development-of-the-hand-and-wrist-2nd-edition-i-volumeatlante-di-256-pagine-stanford-university-press-stanford-california-1959/40E97D3A609810F793A4228BE960A7CA>
- Huang, G., Liu, Z., Van Der Maaten, L., & Weinberger, K. Q. (2017). Densely connected convolutional networks. *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 4700–4708). <https://doi.org/10.1109/CVPR.2017.243>, https://www.researchgate.net/publication/306885833_Densely_Connected_Convolutional_Networks
- Jabbar, A.J., & Abdulmunem, A. (2023). Bone age assessment based on deep learning architecture. *International Journal of Electrical and Computer Engineering (IJECE)*, 13(2), 2078–2085. <https://doi.org/10.11591/ijece.v13i2.pp2078-2085>
- Lindner, C., Thiagarajah, S., Wang, C., Wilkinson, J. M., & Bowes, M. (2020). Bone age estimation using a deep learning model with anatomical domain knowledge. In *Digital Image Computing: Techniques and Applications (DICTA), 2020 International Conference* (pp. 610–617). Springer, Cham. https://doi.org/10.1007/978-3-030-59725-2_61
- Milvue. (n. d.). *AI-powered medical imaging solutions*. <https://www.milvue.com/>
- Shorten, C., & Khoshgoftaar, T.M. (2019). A survey on image data augmentation for deep learning. *Journal of Big Data*, 6(1), 1–48. <https://doi.org/10.1186/s40537-019-0197-0>
- Tan, M., & Le, Q.V. (2020). EfficientNet: Rethinking model scaling for convolutional neural networks. In *International Conference on Machine Learning* (pp. 6105–6114). PMLR. <https://doi.org/10.48550/arXiv.1905.11946>, <https://proceedings.mlr.press/v97/tan19a.html>

Отримано редакцією збірника / Received: 31.08.24

Прорецензовано / Revised: 16.09.24

Схвалено до друку / Accepted: 23.10.24

Іван РЯБКО^{1,2}, магістр
ORCID ID: 0009-0000-6748-8686
e-mail: vanoryabko@gmail.com

Андрій ТЕРЕЦЬ^{1,2}, магістр
ORCID ID: 0000-0003-3668-6131
e-mail: terets.andrii@gmail.com

Станіслав РЕБЕНКОВ²
ORCID ID: 0000-0001-8116-5277
e-mail: rebenkov@gmail.com

Сергій РАДЧЕНКО¹, канд. фіз.-мат. наук, доц.
ORCID ID: 0000-0001-9153-5470
e-mail: sprad@knu.ua

¹Київський національний університет імені Тараса Шевченка, Київ, Україна

²Національна дитяча спеціалізована лікарня "ОХМАТДИТ", Київ, Україна

ПРОГНОЗУВАННЯ КІСТКОВОГО ВІКУ ДІТЕЙ ЗА ДОПОМОГОЮ ГЛИБОКОГО НАВЧАННЯ НА ОСНОВІ РЕНТГЕНІВСЬКИХ ЗНІМКІВ КИСТЕЙ РУК

Вступ. У цьому дослідженні розроблено модель на основі нейронної мережі для прогнозування віку кісток у дітей, використовуючи рентгенівські знімки їхніх рук. Мета роботи полягала в розробці та оцінюванні ефективності такої моделі для підвищення точності діагностики в педіатрії.

Методи. Для досягнення цієї мети було використано підхід перенесення навчання на основі архітектури DenseNet169, який дозволяє обробляти зображення рук та додаткову інформацію про статть пацієнта. Методика включала масштабування зображень до розміру 500x500 пікселів, нормалізацію даних віку та перетворення інформації про статть у бінарний формат. Для підвищення якості моделі було застосовано методи розширення даних, такі як масштабування, поворот, зміщення та горизонтальне віддзеркалення зображень. Крім того, модель була доопрацьована шляхом включення додаткових повністю зв'язаних шарів для покращення прогнозування віку.

Результати. Отримані результати показали, що середня абсолютна похибка (MAE) моделі становила 6,506 місяців, що свідчить про високу здатність до узагальнення та адекватне пристосування до нових даних.

Висновки. Практична цінність цієї роботи полягає в тому, що модель може бути ефективно використана у педіатричній діагностиці для автоматизованого визначення віку кісток, що значно спрощує та прискорює цей процес.

Ключові слова: DenseNet169, прогнозування віку кісток, рентгенівські знімки, статистичне оцінювання.

Автори І. Рябко, А. Тереть та С. Ребенков заявляють про відсутність конфлікту інтересів. Відповідальний редактор С. Радченко заявляє про відсутність конфлікту інтересів та невтручання в процес оцінювання цієї статті відповідно до міжнародних стандартів публікаційної етики. Спонсори не брали участі в розробленні дослідження; у зборі, аналізі чи інтерпретації даних; у написанні рукопису; в рішенні про публікацію результатів.

The authors I. Riabko, A. Terets and Stanislav Riebienkov declare no conflicts of interest. The Executive Editor S. Radchenko declares that there is no conflict of interest and has not interfered in the evaluation process of this article, in accordance with international publication ethics standards. The funders had no role in the design of the study; in the collection, analyses or interpretation of data; in the writing of the manuscript; in the decision to publish the results.