IJARSCT

International Journal of Advanced Research in Science, Communication and Technology

International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal

ISSN: 2581-9429 Volume 6, Issue 3, July 2026 Impact Factor: 8.2

Automated Bone Age Classification from Pediatric Hand Radiographs Using a Convolutional Neural Network

A. H. N. V. S. Amani[1], Swathi Terli[2]

[1] PG Scholar, [2] Assistant Professor, Department of Master of Computer Applications

Sir C. R. Reddy College (Autonomous), PG Courses, Eluru, Andhra Pradesh, India

Affiliated to Adikavi Nannaya University, Rajamahendravaram

Abstract: *Bone age assessment is a common clinical procedure for judging the skeletal maturity of children and for detecting growth, endocrine, and metabolic disorders. In routine practice a radiologist reads a radiograph of the hand and wrist and compares it against a reference standard such as the Greulich–Pyle atlas or the Tanner–Whitehouse method. That process is manual and slow, it depends on the reader's experience, and two radiologists do not always agree on the same film. This paper describes an automated system that uses a Convolutional Neural Network (CNN) to predict skeletal maturity directly from a pediatric hand radiograph. The pipeline covers image acquisition, preprocessing (resizing, grayscale conversion, normalization, and augmentation), feature extraction by the network, and prediction. We implemented the model in Python with TensorFlow and Keras, used OpenCV for preprocessing, and delivered it through a Django web application that handles user authentication, image upload, prediction, and record keeping. The complete workflow was checked with a set of functional cases covering registration, login, upload, prediction, result display, and history, and every case passed. The paper also sets out the protocol we use to report performance, based on accuracy, mean absolute error, root mean square error, precision, recall, and F1-score. The results suggest that a CNN can give fast and repeatable predictions that support a radiologist's reading. We close by discussing how transfer learning, regression toward an exact age, larger datasets, and the addition of clinical variables could push accuracy further.*

Keywords: Bone age assessment; Convolutional Neural Network; deep learning; medical image analysis; skeletal maturity; pediatric radiography; computer-aided diagnosis.

I. INTRODUCTION

Skeletal maturity, usually reported as "bone age," is one of the standard indicators of how a child is developing. Chronological age and skeletal development do not always move in step, so a measure of bone age gives paediatricians, endocrinologists, and orthopaedic surgeons something concrete to work from when they monitor growth or look for problems such as growth-hormone deficiency, early or delayed puberty, hormonal imbalance, malnutrition, and a range of genetic and endocrine conditions. Getting that estimate right, and getting it quickly, matters in everyday paediatric care.

In the clinic, bone age is read from a radiograph of the left hand and wrist. The radiologist compares the film against a reference, most often the Greulich–Pyle (G&P) atlas, where the whole hand is matched to the closest plate, or the Tanner–Whitehouse (TW) method, where individual bones are scored and the scores are added up. Both are well established, and both carry the same drawbacks. Reading a film by hand takes time, particularly in a busy department, and it calls for an experienced radiologist. The larger issue is agreement: the same film can be scored differently by two readers, or by the same reader on a different day, which makes the result harder to reproduce.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 3, July 2026
Impact Factor: 8.2

Deep learning has changed how medical images are analyzed, because a model can learn useful features straight from the pixels instead of relying on features a person designed in advance. For image tasks the usual choice is a Convolutional Neural Network (CNN). A CNN passes the image through repeated convolution, activation, and pooling steps and builds up a layered description of it, moving from simple edges to larger structures. That kind of description fits a hand radiograph well, since the shape and density of the developing bones and the growth plates carry most of the information a reader is after.

This paper describes an automated bone age classifier built on a CNN. The aim is to read skeletal maturity from a hand radiograph without the manual atlas comparison, and to return that reading quickly and consistently. The model is written in Python using TensorFlow and Keras, with OpenCV handling the image preparation, and it runs behind a web application so a clinician can upload a film and get a result in a few seconds. The rest of the paper is organized as follows. Section 2 reviews earlier work on automated bone age assessment. Section 3 sets out the research design, the dataset, the preprocessing, and the network. Section 4 describes how the system is built. Section 5 gives the evaluation protocol and the validation results. Sections 6 and 7 close the paper and consider what could come next.

1.1 Objectives

The work set out to do the following:

- build a CNN that classifies skeletal maturity from hand radiographs and removes the manual atlas step;
- put together a preprocessing pipeline (resizing, grayscale conversion, normalization, and augmentation) that gives the network consistent input;
- wrap the trained model in a web application a clinician can actually use, covering upload, prediction, results, and history;
- measure the system with the usual metrics (accuracy, MAE, RMSE, precision, recall, and F1) and check the full workflow with functional tests.

These objectives point to one question: can a CNN read pediatric bone age from a hand radiograph about as accurately and consistently as the manual atlas method, while taking far less time? Our expectation is that the model returns a repeatable prediction within a few seconds of upload and gives the radiologist a quick second reading to check their own assessment against.

II. RELATED WORK

Automated bone age assessment has a well-developed literature, and over the past decade CNN methods have taken over from the older image-processing pipelines. Spampinato et al. (2017) built BoNet, a CNN that reads bone age from a hand radiograph and copes with the hand being placed differently from one film to the next. It found the skeletal features on its own and outperformed the earlier image-processing methods, which is why CNNs became the default choice for this problem.

Larson et al. (2017) trained a network on more than 14,000 pediatric hand radiographs and found that it matched experienced radiologists at estimating skeletal maturity; they argued that deep learning could cut reporting time without giving up accuracy. Iglovikov et al. (2017) took a two-step route, segmenting the hand with a U-Net and then feeding both the whole hand and specific bone regions into ResNet-50 and VGG-style networks. Their system did well on the RSNA Pediatric Bone Age Challenge data and showed that pairing segmentation with a CNN was worth the extra effort.

The RSNA Pediatric Bone Age Machine Learning Challenge, written up by Halabi et al. (2019), gave the field a large public benchmark and drew in many teams, most of them using CNNs. The leading entries came within roughly four to five months of the reference age, close to what a radiologist achieves. Pan et al. (2019) went back to the 2017 entries and showed that averaging several of them into an ensemble reduced the error further and made the predictions steadier than any single network.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
IJARSCT International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 3, July 2026
Impact Factor: 8.2

A systematic review by Dallora et al. (2019) pulled these threads together. Comparing machine-learning and deep-learning approaches, it found CNNs consistently ahead of the traditional methods, mainly because they learn the image features themselves. The review also flagged the problems that keep recurring: datasets that are not diverse enough, uneven image quality, and models that do not always transfer to a new population. Those same concerns shape the system described here, which applies a CNN to classify skeletal maturity from hand radiographs with less manual work and steadier output.

III. MATERIALS AND METHODS

3.1 Research Design

This is an applied, build-and-test study: we build the artifact (the CNN and the application around it) and then measure it against data it has not seen. The work is quantitative. Labeled hand radiographs go in, a predicted skeletal maturity comes out, and gender is kept alongside as an extra input. The study runs in five steps: collecting and labeling the images; preparing and augmenting them; building and training the network; checking it on a held-out test set; and deploying the finished application and testing it end to end. Splitting the data into training, validation, and test sets keeps the tuning separate from the final measurement, so the reported performance reflects how the model behaves on new films.

3.2 Dataset

The system is trained on pediatric hand radiographs, each one labeled with a bone age in months and with the patient's gender. The images cover a range of ages so that the network sees the full span of skeletal growth; Fig. 1 shows a few examples. Every record holds an image identifier, the radiograph itself (JPEG or PNG), the target bone age, and gender, which is kept because boys and girls mature on different timelines. The images are divided into training, validation, and test sets before any training begins.

3.3 Image Preprocessing

Radiographs vary in size and quality, so each one is cleaned up before it reaches the network:

- resized to a fixed 224 × 224 so that every input has the same shape;
- converted to grayscale, since the diagnostic detail lies in the intensity rather than in color;
- scaled to the [0, 1] range to keep training stable;
- cleaned with noise removal and contrast adjustment so that the bone edges and growth plates stand out;
- augmented with rotations, flips, zooms, and shifts, which enlarges the training set, makes the model less sensitive to how the hand is positioned, and reduces overfitting.

3.4 CNN Architecture

The classifier is a CNN that learns its features from the radiograph rather than from hand-coded rules; Fig. 2 shows the whole pipeline. The image enters at 224 × 224 with a single channel. From there it passes through several convolutional layers with ReLU activations that pick out edges, textures, and bone structures, each one followed by max-pooling that shrinks the spatial size and keeps the computation manageable. A flatten step then hands the features to one or more fully connected layers, and a final layer outputs the predicted maturity. Because the network finds its own features, there is no separate feature-engineering stage, and the whole model is trained together.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 3, July 2026
Impact Factor: 8.2

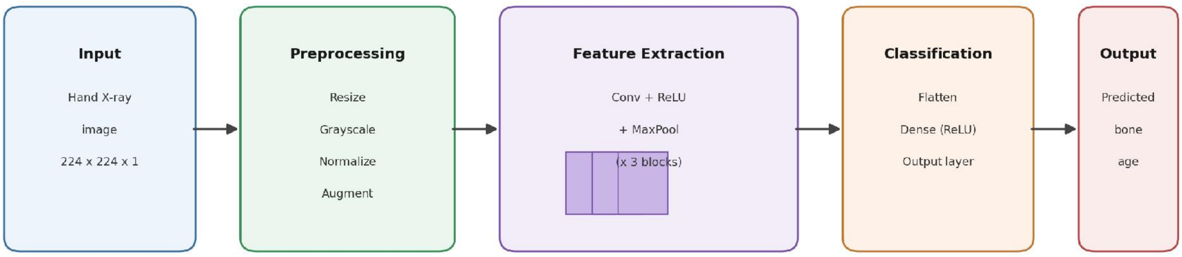

Fig. 2. Architecture of the proposed bone age classification pipeline.

3.5 Training Procedure

Training runs on the prepared training set. In each epoch the network sees the images in batches, compares its predictions with the labels, and updates its weights by backpropagation with a gradient-based optimizer. After every epoch we check accuracy and loss on the validation set to see whether the model is still improving or beginning to overfit, with augmentation applied only to the training stream. Once the validation numbers settle, the best model is saved as an .h5 file for later use.

IV. SYSTEM DESIGN AND IMPLEMENTATION

The trained model sits inside a web application, so a clinician can use it without touching any of the deep-learning code underneath. The application is split into the usual layers: a browser client, an application server, storage for the model, storage for the dataset, and a database that keeps the predictions.

4.1 Software and Technology Stack

Everything is written in Python 3. TensorFlow and Keras build and train the CNN. OpenCV, together with the Python Imaging Library and NumPy, handles the image preparation and the array work. Pandas keeps track of the dataset metadata, Matplotlib draws the training and validation curves, and Scikit-learn takes care of the data splits and the evaluation. The web side is Django, which brings its own authentication, request routing, and database layer. Development and training were done in Jupyter Notebook, Google Colab, and Visual Studio Code, with a GPU used for training where one was available.

4.2 Application Workflow

A session runs like this. The user logs in and goes to the prediction page, uploads a hand radiograph, and selects the patient's gender. The server checks the file, prepares it with OpenCV, and passes it to the trained CNN. The model returns a bone age, which the application turns into years and months and shows next to the original film along with a maturity reading. Each prediction is saved, so the user can return to a history of earlier assessments, and an admin dashboard summarizes figures such as the total number of predictions and the split of cases by gender.

4.3 Functional Modules

The application breaks into a handful of modules. Authentication and user management handle registration and login through Django's built-in system, with hashed passwords and separate access for ordinary users and administrators. The upload module accepts the standard image formats and rejects anything it cannot read, returning a clear message rather than failing outright. Preprocessing and inference prepare the film with OpenCV and run it through the CNN.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 3, July 2026
Impact Factor: 8.2

The results module shows the estimated bone age in a readable form, and the history and administration module keeps each user's past predictions and gives the administrator a dashboard for oversight.

V. RESULTS AND DISCUSSION

5.1 Evaluation Protocol

Bone age assessment can be treated as a classification problem, sorting a film into a maturity band, or as a regression problem, predicting a continuous age. The protocol reports both kinds of number. For classification we use accuracy, precision, recall, F1-score, and the confusion matrix. For continuous estimation we report the mean absolute error (MAE) and the root mean square error (RMSE) in months. MAE is the figure most papers quote, and the one the RSNA benchmark used, where the best systems came within about four to five months. Reporting the error in months makes it easy to line up against earlier work and against how closely two radiologists agree.

5.2 Functional Validation

We checked the application with a set of functional tests that walk through the whole workflow. Every case passed, from registration through prediction, results, history, and the admin view. Table 1 lists them.

ID	Test Case	Input	Expected Result	Status
TC-01	User registration	Valid user details	Account created	Pass
TC-02	User login	Valid credentials	User logged in	Pass
TC-03	Upload radiograph	Valid hand X-ray	Image uploaded	Pass
TC-04	Bone age prediction	Uploaded radiograph	Bone age displayed	Pass
TC-05	View result	Predicted image	Age shown to user	Pass
TC-06	Prediction history	Prior records	History displayed	Pass
TC-07	Admin login	Valid admin creds	Dashboard opens	Pass
TC-08	Delete record	Selected record	Record deleted	Pass

Table 1. Functional test cases and outcomes.

5.3 Quantitative Performance

The classification model is trained and then measured on the held-out test set using the protocol above. Table 2 is the template for those numbers. The value cells are left blank on purpose: they should be filled with the results from your own training run rather than with invented figures.

Metric	Formulation	Value
Accuracy	Classification	[report]
Precision (macro)	Classification	[report]
Recall (macro)	Classification	[report]
F1-score (macro)	Classification	[report]
Mean Absolute Error (months)	Regression	[report]
Root Mean Square Error (months)	Regression	[report]

Table 2. Reporting template for quantitative performance metrics.

The training behavior is worth reporting next to the final scores. Fig. 3 plots training and validation accuracy and loss across the epochs, which shows whether the model converged and whether it began to overfit. Fig. 4 gives the confusion matrix over the maturity bands, which shows where the model does well and which classes it tends to confuse.

5.4 Discussion

The functional tests show that the pipeline works from end to end and returns a prediction within a few seconds of upload. Set against the manual atlas method, the automated version has a few clear benefits. It skips the repeated comparison against reference plates. It gives the same answer every time for the same film, which the manual method cannot promise. And it handles large batches of radiographs without slowing down. None of this is meant to take the radiologist out of the loop; the system is there to provide a fast second reading.

There are real limits, too. How well the model does depends heavily on how large, how clean, and how varied the training set is. Films differ from one machine and protocol to the next, and sorting a film into a fixed band will always be coarser than predicting an exact age. Because boys and girls mature at different rates, it also makes sense to feed gender into the model directly rather than leave it as a side field. Section 7 takes up these points.

VI. CONCLUSION

This paper described a bone age classifier that reads skeletal maturity from a pediatric hand radiograph with a CNN. The pipeline, from image to preprocessing to feature extraction to prediction, is written in Python with TensorFlow, Keras, and OpenCV, and runs as a Django web application that handles authentication, upload, prediction, results, and record keeping. The functional tests passed across the whole workflow, and the paper lays out the protocol, using accuracy, MAE, RMSE, precision, recall, and F1-score, for reporting how well the model does. The main point is that a CNN can stand in for the slow manual atlas reading and give a faster, steadier answer, which lightens the radiologist's load and makes the result easier to reproduce. It still needs testing on larger and more varied datasets before it could be trusted at a clinical level, but it is a workable step in that direction.

VI. FUTURE WORK

A few changes would make the system more accurate and closer to something usable in a clinic. The first is simply more data, and more varied data: films from different age groups, regions, and machines would help the model hold up on patients it was not trained on. A second is a stronger backbone. Pretrained networks such as ResNet, DenseNet, EfficientNet, or a Vision Transformer, brought in through transfer learning, would likely read the films better and train faster than a network built from scratch. The task could also move from bands to an exact age by switching to a regression output. Adding clinical details such as height, weight, and growth history alongside the image would turn it into a multimodal model that lines up better with how a clinician actually reasons. Attention maps or class-activation heatmaps would show which part of the film drove each prediction, which tends to make clinicians trust the output more. Finally, integrating the model with hospital systems and offering it through a responsive web or mobile interface would let it be used in real time, including in places where a specialist radiologist is hard to reach.

REFERENCES

[1] C. Spampinato, S. Palazzo, D. Giordano, M. Aldinucci, and R. Leonardi, "Deep learning for automated skeletal bone age assessment in X-ray images," Medical Image Analysis, vol. 36, pp. 41–51, 2017.

[2] D. B. Larson, M. C. Chen, M. P. Lungren, S. S. Halabi, N. V. Stence, and C. P. Langlotz, "Performance of a deep-learning neural network model in assessing skeletal maturity on pediatric hand radiographs," Radiology, vol. 287, no. 1, pp. 313–322, 2018.

[3] V. Iglovikov, A. Rakhlin, A. Kalinin, and A. Shvets, "Pediatric bone age assessment using deep convolutional neural networks," arXiv preprint arXiv:1712.05053, 2017.

[4] S. S. Halabi, L. M. Prevedello, J. Kalpathy-Cramer, A. B. Mamonov, A. Bilbily, M. Cicero, et al., "The RSNA pediatric bone age machine learning challenge," Radiology, vol. 290, no. 2, pp. 498–503, 2019.

[5] I. Pan, H. H. Thodberg, S. S. Halabi, J. Kalpathy-Cramer, and D. B. Larson, "Improving automated pediatric bone age estimation using ensembles of models from the 2017 RSNA machine learning challenge," Radiology: Artificial Intelligence, vol. 1, no. 6, e190053, 2019.

[6] A. L. Dallora, P. Anderberg, O. Kvist, et al., "Bone age assessment with various machine learning techniques: a systematic literature review and meta-analysis," PLOS ONE, vol. 14, no. 7, e0220242, 2019.

[7] 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," Journal of Digital Imaging, vol. 30, pp. 427–441, 2017.

[8] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR), pp. 770–778, 2016.

[9] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in Advances in Neural Information Processing Systems, vol. 25, 2012.

[10] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," Nature, vol. 521, pp. 436–444, 2015.

[11] I. Goodfellow, Y. Bengio, and A. Courville, Deep Learning. Cambridge, MA: MIT Press, 2016.

[12] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, et al., "A survey on deep learning in medical image analysis," Medical Image Analysis, vol. 42, pp. 60–88, 2017.

[13] F. Chollet, Deep Learning with Python. Shelter Island, NY: Manning Publications, 2017.

[14] A. Géron, Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow, 3rd ed. Sebastopol, CA: O'Reilly Media, 2022.

[15] World Health Organization, Ethics and Governance of Artificial Intelligence for Health. Geneva: WHO Publications, 2021.