

A Proficient Convolutional Neural Network for Classification of Bone Age from X-Ray Images

Sajid Faysal Fahim, Md Sakib Morshed, Shodorson Nath,
Nafisa Tasnim, Zareen Tasnim Nishat, Anik Lal Dey, Golam Kibria,
Sumayea Bintey Azad, Mir Ariyan Shuddho
Department of Computer Science and Engineering East West University
Dhaka, Bangladesh {sajidfaysalfahim, morshedsakib41, shodorsomannath,
nafisatasnim063, zarin.nishu99, aniklal2020, golamkibria11265, sumayea14,
mirariyanshuddho}@gmail.com

Nishat Tasnim Niloy
Department of Computer Science
and Engineering East West
University Dhaka, Bangladesh
nishat.niloy@ewubd.edu

Abstract—Bone age evaluation is crucial for identifying and planning interventions for numerous disorders. Estimating bone age is distinct from assessing physical development based on an individual's birth date. This evaluation of bone age reveals growth and progression, facilitating the identification and management of pediatric diseases. Significant obstacles in bone age evaluation often stem from low-quality X-ray images, obscured bone formations, and the intricacies of feature extraction due to compromised image quality, which greatly affects the performance of models. This research introduces VGG19, a groundbreaking Convolutional Neural Network (CNN) method, to classify bone age utilizing the RSNA dataset and its associated images. This tailored model is adept at recognizing patterns with a newly assembled dataset of regionspecific images, excelling in categorizing diverse bone types. The efficacy of ResNet50 is affirmed through extensive 5-fold crossvalidation, where it outperforms sophisticated models like VGG16 and Xception, attaining outstanding performance metrics with an accuracy of 96.46%, precision of 96.408%, recall of 96.450%, F-score of 96.475%, and specificity of 96.726%. The results of this research carry substantial implications for improving the precise classification of bone age.

Index Terms—component, formatting, style, styling, insert.

I. INTRODUCTION

THE AGE of bones indicates an individual's skeletal and biological progression, whereas chronological age refers to the time elapsed since one's birth. Pediatricians and endocrinologists utilize bone age evaluations (BAE) alongside chronological age to identify conditions that lead to growth disorders in children, whether through excessive or insufficient growth. Bone age evaluations can serve as a valuable tool in diagnosing various endocrine abnormalities, including precocious puberty and idiopathic dwarfism [1]. This facilitates timely and appropriate treatment for children exhibiting atypical growth patterns. BAE often plays a crucial role in assessing athletes' eligibility and in legal investigations, guaranteeing precision and dependability in all these contexts [2]. The key contributions of this manuscript are outlined as follows:

- A novel method that delivers environmental advantages while also saving manpower and time has been proposed.

- To address the challenge, an innovative CNN-powered system known as ResNet50 has been developed, which leverages this specific set of data.
- ResNet50 surpasses other cutting-edge models such as VGG16 and Xception when it comes to assessment criteria [3].

This article is divided into several sections. The second one brings the information on bone age techniques. Section 3 then explains the experiment setup, covering data processing and algorithm evolution. Sections 4 and 5 then present the results obtained due to the observations. Ultimately, the findings derived from the investigation are detailed in Section 6.

II. LITERATURE REVIEW

Historically, research carried out by BAA was oriented towards traditional methods like the Greulich-Pyle and the Tanner-Whitehouse [4] approaches. These approaches rely on radiographic atlases and involve the comparison of radiographs to evaluate the maturation of bones. The GilsanzRatib [5] digital atlas improves this accuracy by providing categorized images for different age groups and sexes. Under the auspices of CAD, the initial focus was on the correct segmentation of the X-ray so that skeletal structures could be isolated. This pursuit had issues distinguishing bone from soft tissue and backgrounds, prompting research into numerous various methods.

Wibisono et al. (2020) designed a decision support system based on ML and DL, utilizing RB-FCL for certain regions in hand images and DL models: DenseNet121, InceptionV3, and InceptionResNetV2 to extract bone-related features, obtaining an MAE of 6.97 months on RSNA; this approach outperformed the traditional DL models and represents a better score compared to the conventional DNN with a score of 9.41 months bone age prediction from X-ray images [6].

Li et al., 2021 proposed a DL-based computer-assisted evaluation for BAA based on MobileNet and MLP with one hidden layer using unsupervised learning to identify informative regions, which achieved an MAE of 5.1 months on the Clinical dataset by inputting sex information into the

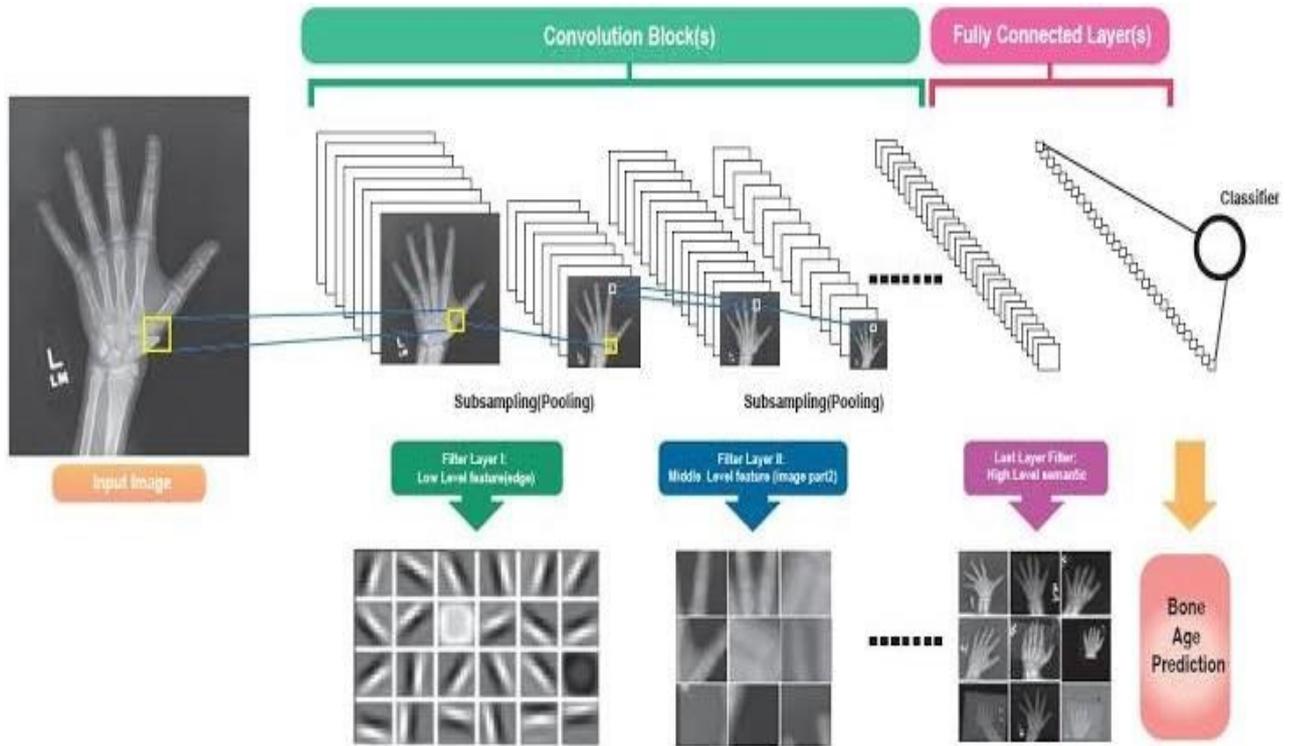


Figure 1: Proposed Method of the Solution

prediction process to perform better in clinical research and 6.2 months on the RSNA dataset [7].

Xu et al. [2022] proposed a hierarchical CNN, YOLOv5, for BAA using ROI detection and bone score classification on a dataset from Xuzhou Central Hospital (2158 X-ray images), and achieved an MAE of 6.53 months on the public RSNA dataset and 7.68 months on the clinical dataset, showing competitive performance and beating current fine-grained image classification approaches in BAA [8].

Liu et al. (2019) introduced a novel BAA method by combining NSCT and CNNs, enhancing BAA on DHA using VGGNet-16 and achieved MAE of 8.28 months with multi-scale data fusion, outperforming the traditional spatial domain methods [9].

III. STEP OF METHOD

This segment will shed light on the research approach, providing an insight into the techniques for gathering and analyzing data. It will also showcase Xception, VGG16 with the innovative ResNet50 architecture, all brought to life through Python with Tensorflow and Keras on a powerful Ubuntu machine [10]. Figure 1 illustrates the method of the proposed solution. In this illustration, the innovative approach for violation detection powered by deep neural networks is detailed. It showcases the entire journey of the project. The procedure will commence with data acquisition, progressing through training and processing phases, while also encountering various conditions.

A. Dataset Description

This research is grounded in the comprehensive RSNA



Figure 2: Sample Images of the Bones

Paediatric Bone Age Challenge dataset, established in 2017, comprising 12,611 X-ray images for Bone Age Assessment (BAA), with an age range from 0 to 217 months, and including 6,833 male and 5,778 female records to ensure accurate estimation [11]. In the Figure 2, sample images of the dataset has been provided.

B. Data Preprocessing

Image pre-processing encompasses sophisticated techniques that enhance image fidelity by correcting distortions and enriching data content, with operations such as batch manipulation, rescaling, labeling, and range exploration yielding optimal outcomes.

C. Model Training and Evaluation

The voyage of the Training Set begins as it navigates through the intricate layers of the Convolutional Neural Network, where each layer plays a vital role in shaping the final outcome. From engaging in convolution with multiple filters

TABLE 1: A CONCISE OVERVIEW OF THE FEATURE MAPS WITHIN THE SUGGESTED ResNet50 FRAMEWORK.

Layer	Filter Sets	Dimension of Filter	Step Size	Feature Map Dimensions	Function of Activation
Image				227227 3*	
Convolution	50	11 11	3	7373 50	ReLU
Normalization of Batches				7373 50	
Maximum Pooling	-	2 2	2	3636 50	
Convolution	100	11 11	1	3636 100	ReLU
Normalization of Batches				3636 100	
Max Pool		2 2	2	1818 100	
Convolution	150	5 5	1	1818 150	ReLU
Normalization of Batches				1818 150	
Convolution	100	5 5	1	1818 100	ReLU
Normalization of Batches				1818 100	
Maximum Pooling		2 2	2	99 100	
Convolution	90	3 3	1	99 90	ReLU
Normalization of Batches				99 90	
Maximum Pooling		2 2	2	44 90	
Flatten				1440	
FC	800			800	ReLU
Dropout	rat e=0.5				
FC	800			800	ReLU
Dropout	rat e=0.5				
FC	8				Softmax

to selecting maximum values and transforming outputs, each layer contributes uniquely to the network's progression. Evaluating a model is crucial in model development, guiding towards the most accurate representation of data through methods like cross-validation and hold-out, ensuring the model's true potential is revealed while guarding against over-fitting [12].

D. Cutting-Edge Algorithms

This section will discuss the architectures of two advanced algorithms, VGG16 and Xception along with the proposed ResNet50, for classifying imbalanced waste.

1) VGG16 Architecture

The VGG16 architecture depicted in Figure 3 delineates its layers, feature maps, activation functions, and parameters, featuring an initial increase in channels followed by a gradual reduction across five convolutional blocks and two fully connected layers, with essential feature maps highlighted while most max pooling layers are omitted, processing a three-channel RGB input to classify eight labels through deep learning methodologies [13].

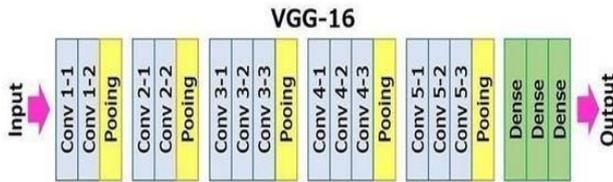


Figure 3: VGG16 Architecture.

2) Xception Architecture

The structure of Xception is illustrated in Figure 4 to clarify its parameters and information flow. As part of the generic VGG architectures, it employs multiple convolu-

tional layers followed by max pooling and fully connected layers to predict 8 classes.

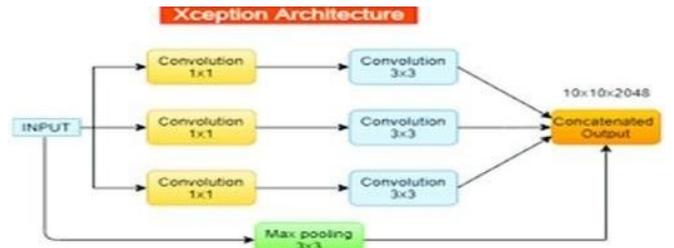


Figure 4: Xception Architecture

3) Suggested Structure of the Convolutional Neural Network (ResNet50)

The suggested design follows the VGG16 methodology of first amplifying and then reducing the quantity of filters or channels during the extraction of feature maps. Each convolutional segment, barring one, comprises a convolutional (CONV) layer paired with a max pooling layer (Max Pool), reminiscent of VGG16, yet it is more streamlined with a reduced number of channels. The layout also includes two fully connected dense layers (FC) alongside a softmax layer for producing predictions, featuring a markedly lower count of neurons. A concise overview of the proposed design is illustrated in Table 1, while Figure 5 presents a graphical depiction of ResNet50 [14].

IV. PARAMETERS OF INFLUENCE, INSTRUCTIONAL APPROACHES, AND EVALUATION TECHNIQUES

Cross-validation is utilized to assess each fold without the necessity of distinct testing instances, employing a 5-fold method with a random seed that allocates 80% of the data to

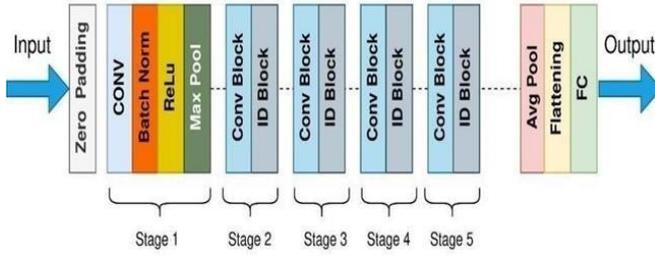


Figure 5: ResNet50 Architecture.

training, 10% to validation, and 10% to testing, as detailed in Table 2 regarding hyper-parameters and training considerations, while Table 3 illustrates the varying training durations for each model [15].

TABLE 2: INFORMATION REGARDING HYPER-PARAMETERS.

Cost metric	Multi-class cross-entropy
Optimizer	Stochastic Gradient Descent (SGD)
Learning Rate	0.001
Early stopping	60
Size of the batch Maximum	15
Total epochs for execution	230

TABLE 3: ANALYSIS OF THE MEAN TRAINING DURATION MEASURED.

Average Training Time				
	Per Batch (CPU)	Per Batch (GPU)	Per Epoch (CPU)	Per Epoch (GPU)
ResNet50	2000 ms/step	7 ms/step	607000 ms	980 ms
VGG16	2034 ms/step	14 ms/step	625000 ms	2225 ms
Xception	22500 ms/step	55 ms/step	6569000 ms	15300 ms

V. FINDINGS AND INSIGHTS

The subsequent section elucidates the study's findings, encompassing training loss and accuracy metrics per fold, parameter count comparison, and testing dataset accuracy.

A. Evaluation of parameter quantities

ResNet50, VGG16, and Xception present several benefits, including accelerated training durations and enhanced capability to generalize to novel datasets based on varying parameters [16]. While ResNet50 necessitates a smaller number of parameters in comparison to VGG16 and Xception, it is imperative to consider both the architectural design and the training methodology to ensure the integrity of the model. VGG16 is primarily oriented towards image classification and is characterized by its numerous convolutional layers; it possesses a reduced number of filters yet features a more profound network. Conversely, ResNet50 integrates both manually designed and learned features, thus rendering it particularly suitable for smaller datasets [17]. A compari-

son among the parameters of different architecture has been provided in Table 4.

TABLE 4: ASSESSMENT OF AGGREGATE PARAMETERS VERSUS COUNT OF ADJUSTABLE PARAMETERS

Model Designation	Aggregate Parameter Count	Count of adj. Parameters
ResNet50	3,568,709	3,195,627
VGG16	56,322,676	56,366,652
Xception	125,280,820	128,283,450

B. Dimensions of the preserved weights for every design

Table 5 displays the average size of weight files for various architectures post-training; the proposed architecture is notably the lightest, offering decent accuracy over Xception despite being significantly lighter, which may be acceptable given the practical nature of the problem [18].

TABLE 5: DIMENSIONS OF THE PRESERVED WEIGHT FILES FOR ALL THE DESIGNS AVAILABLE IN HDF5 FORMAT.

Design Title	Dimensions in Megabytes (MB) of the preserved weights
ResNet50	12.8
VGG16	450.6
Xception	520.4

C. Evaluation matrices

In every iteration, each framework was executed multiple times, and the accuracy attained in each attempt was calculated and subsequently averaged, as presented in Table 6 [19]. Emphasizing average performance and employing ResNet50 can significantly improve architectural efficacy by mitigating biases, as indicated in Table 6, where ResNet50 surpasses VGG16 despite being a comparatively lighter model, whereas Xception consistently demonstrates inferior performance due to constraints in sample size and variations within the dataset.

TABLE 6: THE MEAN OUTCOMES DERIVED FROM NUMEROUS TRIALS WITHIN EACH CROSS-VALIDATION FOLD ACROSS VARIOUS ARCHITECTURAL TESTING DATASETS.

Architecture Name	Accuracy	Precision	Recall	F-Score	Specificity
VGG16	92.50	92.230	92.080	92.044	92.820
Xception	93.04	93.840	93.674	93.706	93.768
ResNet50	96.46	96.408	96.650	96.775	96.726

VI. CONCLUSION

After detailed analysis for BAA, it was remarked that the pretrained models, On the other side, the SGD optimizer was the worst among the optimizers tried on the pretrained models. Adam is usually the first choice in most CNN architectures. This work flags the importance of selection and optimization methods in BAA tasks by showing the subtle influence these decisions could have on the final performances obtained from the deep learning models. A much deeper fine-tuning strategy and architectural adjustments can be performed in further researches to improve the BAA performance. Further increasing the dataset size and using good-quality images will also increase the accuracy of the BAA.

The error can also be reduced by accurately finding the ROI to enhance the performance of the pre-trained models.

REFERENCES

- [1] K. N. Sami, Z. M. A. Amin, and R. Hassan, "Waste management using machine learning and deep learning algorithms," *Int. J. Perceptive Cogn. Comput.*, vol. 6, no. 2, pp. 97–106, Dec. 2020, doi: 10.31436/ijpcc.v6i2.165.
- [2] S. Shahab, M. Anjum, and M. S. Umar, "Deep learning applications in solid waste management: A deep literature review," *Int. J. Adv. Comput. Sci. Appl.*, vol. 13, no. 3, pp. 381–395, 2022, doi: 10.14569/IJACSA.2022.0130347.
- [3] M. Triassi, R. Alfano, M. Illario, A. Nardone, O. Caporale, and P. Montuori, "Environmental pollution from illegal waste disposal and health effects: A review on the 'triangle of death,'" *Int. J. Environ. Res. Public Health*, vol. 12, no. 2, pp. 1216–1236, Jan. 2015, doi: 10.3390/ijerph120201216.
- [4] Md. T. and S. Mst. S. R. and H. S. A. and C. N. R. Tusher Abdur Nur and Islam, "Automatic Recognition of Plant Leaf Diseases Using Deep Learning (Multilayer CNN) and Image Processing," in *Third International Conference on Image Processing and Capsule Networks*, 2022, pp. 130–142.
- [5] J. Wu, "Introduction to Convolutional Neural Networks," National Key Lab for Novel Software Technology, 2017. Accessed: Feb. 08, 2023. [Online]. Available: Introduction to Convolutional Neural Networks <https://cs.nju.edu.cn> >> paper > CNN
- [6] R. C. Ploetz, "The Major Diseases of Mango: Strategies and Potential for Sustainable Management," *Acta Horti*, vol. 645, pp. 137–150, 2004, doi: 10.17660/ActaHortic.2004.645.10.
- [7] P. Kumar, S. Ashtekar, S. S. Jayakrishna, K. P. Bharath, P. T. Vanathi, and M. Rajesh Kumar, "Classification of Mango Leaves Infected by Fungal Disease Anthracnose Using Deep Learning," in *2021 5th International Conference on Computing Methodologies and Communication (ICCMC)*, 2021, pp. 1723–1729. doi: 10.1109/ICCM-C51019.2021.9418383.
- [8] Bogdan Alexe, Thomas Deselaers, and Vittorio Ferrari, "What is an object?", *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2010.
- [9] Bogdan Alexe, Thomas Deselaers, and Vittorio Ferrari, "Measuring the objectness of image windows", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34(11), 2189-2202, 2012.
- [10] Sean Bell, Paul Upchurch, Noah Snaveley, and Kavita Bala, "Material recognition in the wild with the materials in context database." *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 3479-3487, 2015.
- [11] Alhamdan, W. S., & Howe, J. M. (2021). Classification of date fruits in a controlled environment using convolutional neural networks. *Advances in Intelligent Systems and Computing*, 154–163. https://doi.org/10.1007/978-3-030-69717-4_16
- [12] Alzu'bi, R., Anushya, A., Hamed, E., Al Sha'ar, Eng. A., & Vincy, B. S. (2018). Dates fruits classification using SVM. *AIP Conference Proceedings*. <https://doi.org/10.1063/1.503204>
- [13] Kandel, I., Castelli, M., Popovič, A.: Comparative Study of First Order Optimizers for Image Classification Using Convolutional Neural Networks on Histopathology Images. *J. Imaging*
- [14] Sharma, P., Anand, R.S.: A comprehensive evaluation of deep models and optimizers for Indian sign language recognition. *Graphics and Visual Computing*. 5, 200032 (2021). <https://doi.org/10.1016/j.gvc.2021.200032>.
- [15] Maggio, A., Flavel, A., Hart, R., Franklin, D.: Assessment of the accuracy of the Greulich and Pyle hand-wrist atlas for age estimation in a contemporary Australian population. *Australian Journal of Forensic*
- [16] Adler, B.H.: Vicente Gilsanz, Osman Ratib: Bone age atlas. *Pediatr Radiol*. 35, 1035– 1035 (2005). <https://doi.org/10.1007/s00247-0051527-2>.
- [17] Spampinato, C., Palazzo, S., Giordano, D., Aldinucci, M., Leonardi, R.: Deep learning for automated skeletal bone age assessment in X-ray images. *Medical Image Analysis*. 36, 41–51 (2017). <https://doi.org/10.1016/j.media.2016.10.010>.
- [18] Liu, Y., Zhang, C., Cheng, J., Chen, X., Wang, Z.J.: A multiscale data fusion framework for bone age assessment with convolutional neural networks. *Computers in Biology and Medicine*. 108, 161–173 (2019). <https://doi.org/10.1016/j.compbiomed.2019.03.015>.
- [19] Bui, T.D., Lee, J.-J., Shin, J.: Incorporated region detection and classification using deep convolutional networks for bone age assessment. *Artificial Intelligence in Medicine*. 97, 1–8 (2019). <https://doi.org/10.1016/j.artmed.2019.04.005>.