International Journal of Computer Trends and Technology
ISSN: 2231 – 2803 / https://doi.org/10.14445/22312803/IJCTT-V71I2P104
Volume 71 Issue 2, 19-24, February 2023

Original Article

Bone Age Prediction with AI Models

Chi-Chang Chen[1], Yu-Xian Chou[2]

[1,2]*Information Engineering Department, I-Shou University, Taiwan.*

Received: 21 December 2022 Revised: 22 January 2023 Accepted: 03 February 2023 Published: 14 February 2023

Abstract *- Artificial intelligence (AI) models have been developed to assist in the process of bone age prediction by automating the assessment of radiographic images. These models use machine learning algorithms to learn from a dataset of previously assessed images and can then make predictions about the bone age of new images with high accuracy. In this paper, we use four AI models, namely, VGG16, ResNet50, ResNet152, and Xception, to automatically predict the bone ages of X-ray images from the Radiological Society of North America (RSNA). According to our experiments, Xception got better results than the others three models. Both the mean absolute error(mae) and median absolute error of Xception was 7.21 months. These AI models have the potential to improve the accuracy, consistency, and efficiency of bone age prediction. However, there are also limitations and challenges to using AI models for bone age prediction, such as the need for large and diverse training sets and robust validation and testing. Further research and development are needed to address these challenges and limitations to ensure that the AI models for bone age prediction are reliable and accurate in real-world settings.*

Keywords *- Bone Age Prediction, Machine Learning, Deep Learning, Medical image processing.*

1. Introduction

Bone age prediction is an important aspect of pediatric and endocrinologic care, as it is used to diagnose and monitor growth disorders in children. The traditional method of assessing a child's bone maturity involves radiographic images of the hand and wrist, which are then interpreted by a radiologist or paediatrician[1,2,3]. This process involves the interpretation of the images by a specialist, who then assigns a bone age based on their own experience and expertise. This method is time-consuming and subjective, as different radiologists or pediatricians may have different experience levels and expertise. As a result, there is a need for a more accurate, consistent, and efficient method of assessing bone age.

In recent years, Artificial intelligence (AI) models have emerged as a potential solution to these issues by automating the assessment of radiographic images. These models use machine learning algorithms to learn from a dataset of previously assessed images and can then make predictions about the bone age of new images with high accuracy. These models are designed to mimic a radiologist's or paediatrician's decision-making process, but with the added benefit of being less time-consuming and less subjective.

AI models for bone age prediction have several potential advantages over traditional methods. They can be less time-consuming and less subjective, as they do not require a radiologist's interpretation. They can also be more consistent, as they do not rely on the experience or expertise of the person assessing the images. Additionally, these models can be used in remote or low-resource settings where radiologists are unavailable. However, there are also limitations and challenges to using AI models for bone age prediction, such as the need for large and diverse training sets and robust validation and testing.

In this paper, we study four convolutional neural networks: VGG16, ResNet50, ResNet152, and Xception, and use them for bone age prediction from the dataset of the Radiological Society of North America (RSNA) [4]. We compare their accuracy and efficiency of bone age predictions and conclude that Xception has high accuracy and that ResNet50 takes less execution time for training the model.

2. Background

The traditional method of assessing bone age is based on the Greulich and Pyle (G&P) method, which was first published in 1959 [5]. The G&P method involves using radiographic images of the left hand and wrist, which a radiologist or pediatrician then assesses. The radiologist or pediatrician uses their experience and expertise to assign a bone age based on the images, using the G&P atlas as a reference. The G&P atlas includes a set of radiographic images of the hand and wrist at different stages of bone maturity, along with corresponding bone ages.

This traditional method of assessing bone age is widely used but has several limitations. The process is time-consuming, requiring a radiologist or paediatrician's interpretation of radiographic images. It can also be subjective, as different radiologists or pediatricians may have different levels of experience and expertise, leading to variations in the assigned bone age. Furthermore, the G&P

atlas is based on a sample of American white children, which may not be representative of other populations.

In recent years, there has been an increased interest in using artificial intelligence (AI) for bone age prediction. AI-based methods have the potential to improve the accuracy and consistency of bone age assessment, as well as reduce the time and costs associated with manual assessment.

One of the most promising applications of AI in bone age prediction is the use of deep learning algorithms. These algorithms, such as convolutional neural networks (CNNs) [6,7,8], can be trained on large datasets of radiographic images and automatically learn features associated with bone age. Once trained, the algorithms can then be used to predict the bone age of new radiographic images with high accuracy.

In this paper, we study four convolutional neural networks: VGG16, ResNet50, ResNet152, and Xception, and use them for bone age prediction from the dataset of RSNA. RSNA is a non-profit organization and an international society of radiologists, medical physicists and other medical imaging professionals representing 31 radiologic subspecialties from 145 countries around the world. In 2017 RSNA conducted a challenge to assess bone age from pediatric hand radiographs, a routine task that determines an important developmental indicator [9,10,22].

Several studies have shown that AI-based methods can achieve high levels of accuracy in bone age prediction [12,13,14]. For example, a study by Lee et al. used a CNN trained on a dataset of over 14,000 hand radiographs and achieved an accuracy of 95% in bone age prediction [3]. Another study by Kooi et al. used a deep learning algorithm trained on a dataset of over 7,000 hand radiographs and achieved an accuracy of 96.7% [16].

Despite the promising results of these studies, there are still some challenges to be addressed before AI-based methods can be widely adopted in clinical practice. One of the main challenges is the lack of a large, diverse, and well-annotated dataset of radiographic images for training the algorithms. Another challenge is the lack of standardization in assessing bone age, which can lead to variations in the radiographic images and make it difficult to develop a generalizable algorithm.

In this paper, we study four CNN models for bone age prediction. They are Vgg16, ResNet50, ResNet152, and Xception. In the following, we make a brief survey of these models.

Vgg16 [17] is a convolutional neural network model trained on the ImageNet dataset. The "VGG" in the name refers to the architecture of the network, which is made up of multiple layers of small convolutional filters arranged in a stack of "blocks", where each block is composed of two or more convolutional layers. The "16" in the name refers to the number of weight layers in the network. VGG16 is widely used as a benchmark for image classification and object recognition tasks, and its architecture has been used as the basis for many other models.

ResNet50 [18] is a convolutional neural network model trained on the ImageNet dataset. The "ResNet" in the name stands for "Residual Network," which refers to the model's architecture. This architecture utilizes "residual connections," which are additional connections between the layers of the network that help to alleviate the problem of vanishing gradients. The "50" in the name refers to the number of weight layers in the network. ResNet50 is a deeper version of the original ResNet architecture. It is known for its ability to achieve high accuracy in image classification and object recognition tasks while being relatively easy to train. It is also a popular architecture in computer vision and image processing tasks.

ResNet152 [18] is a convolutional neural network model trained on the ImageNet dataset. Like ResNet50, the "ResNet" in the name stands for "Residual Network," which refers to the architecture of the model that utilizes "residual connections" to alleviate the problem of vanishing gradients. The "152" in the name refers to the number of weight layers in the network. ResNet152 is even deeper than ResNet50 and is known for its ability to achieve high accuracy in image classification and object recognition tasks. However, it requires much more computational power and memory than ResNet50. It is also commonly used in computer vision and image processing tasks.

Xception [19,20] is a convolutional neural network model trained on the ImageNet dataset. The name "Xception" stands for "Extreme Inception," which refers to the model's architecture. Xception is an extension of the Inception architecture, known for its ability to process images at multiple scales efficiently. Xception utilizes depthwise separable convolutions, which can reduce the number of parameters and computational costs compared to traditional convolutional layers. The Xception model was designed to be more efficient than Inception while maintaining similar or even better performance on image classification and object recognition tasks. It is also widely used in computer vision, image processing, and other related fields.

In summary, VGG16 is a simple and strong model, ResNet models are deeper models with high accuracy and are easy to train, and Xception is an efficient model. The best model for a specific task depends on the characteristics of the task, computational resources, and the desired trade-off between accuracy and efficiency.

3. Experiments and Results

Programing Environments:

3.1. Ubuntu 18.4

Ubuntu is a popular open-source operating system based on Linux. It is commonly used for servers and desktop computers.

3.2. Anaconda3-2020.05 with python 3.9

Anaconda is a distribution of Python and R for scientific computing and data science. It includes popular packages such as Jupyter, NumPy, and Pandas and allows for easy installation of additional packages and environments. Anaconda can be installed on Ubuntu and other Linux distributions, as well as on Windows and macOS.

3.3. Nvidia RTX 6000 GPU

The NVIDIA RTX 6000 is a high-performance graphics processing unit (GPU) designed for use in professional and scientific computing applications. It is built on the NVIDIA Ampere architecture and is based on the 7nm process technology.

3.4. Tensorflow-gpu==2.2.0

TensorFlow-GPU is a version of TensorFlow that is optimized to run on NVIDIA GPUs, which allows for faster training of deep learning models. To use TensorFlow-GPU, you need to have a compatible NVIDIA GPU and the appropriate NVIDIA CUDA and cuDNN libraries installed on your system.

3.5. Pytorch-cuda=11.6

PyTorch is an open-source machine learning library developed by Facebook's AI Research group. It is based on the Torch library and is primarily used for deep learning and computer vision applications. One of the key features of PyTorch is its support for automatic differentiation, which allows for easy implementation of complex models such as neural networks. It also provides a high-level interface for building and training models, making it easy for researchers and practitioners to use.

During the programming of AI models, we use the following three important parameters: EarlyStopping, ModelCheckpoint, and ReduceLROnPlateau [23].

EarlyStopping is a callback that can be used to stop the training process before the specified number of epochs if the model's performance on a validation set does not improve after a certain number of epochs. This can help prevent overfitting.

ModelCheckpoint is a callback that can be used to save the best model during training. The best model is determined by a specified metric, such as accuracy or loss.

ReduceLROnPlateau is a callback that can be used to reduce the learning rate of the optimizer when the performance of the model on a validation set has stopped improving. This can help the model converge more quickly and prevent overfitting.

These three callbacks can be very useful during training neural networks, particularly in deep learning applications.

In summary, the EarlyStopping callback stops training when the model's performance on a validation set does not improve after a certain number of epochs. The ModelCheckpoint callback saves the best model during training. The ReduceLROnPlateau callback reduces the learning rate of the optimizer when the performance of the model on a validation set has stopped improving.

The training results of each model are displayed as figures in the following order.

- Predicted age vs. Actual age
- Predicted mean absolute errors (mae) with epochs
- Loss values with epochs

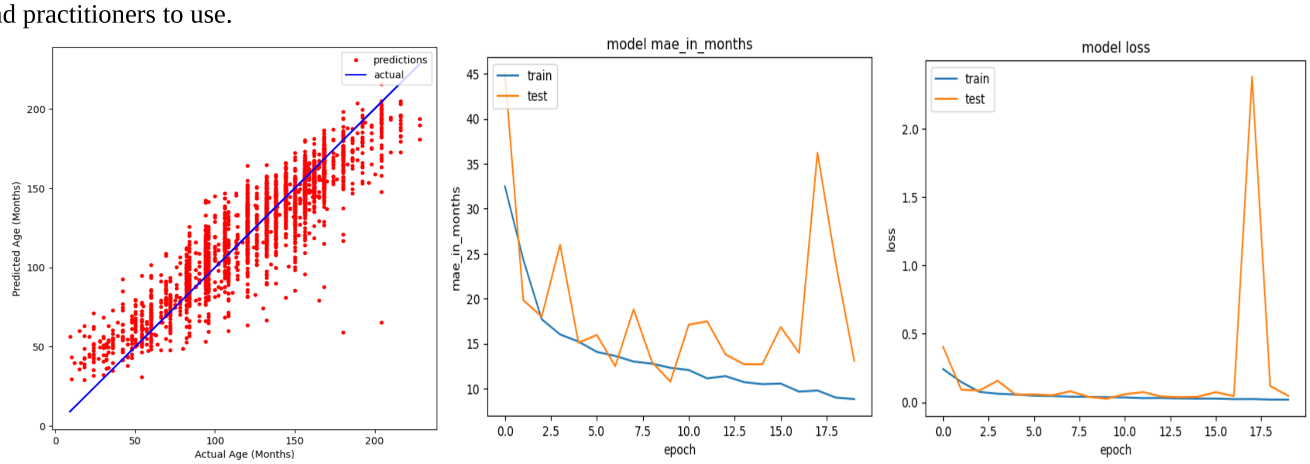

Fig. 1 VGG-16 training results

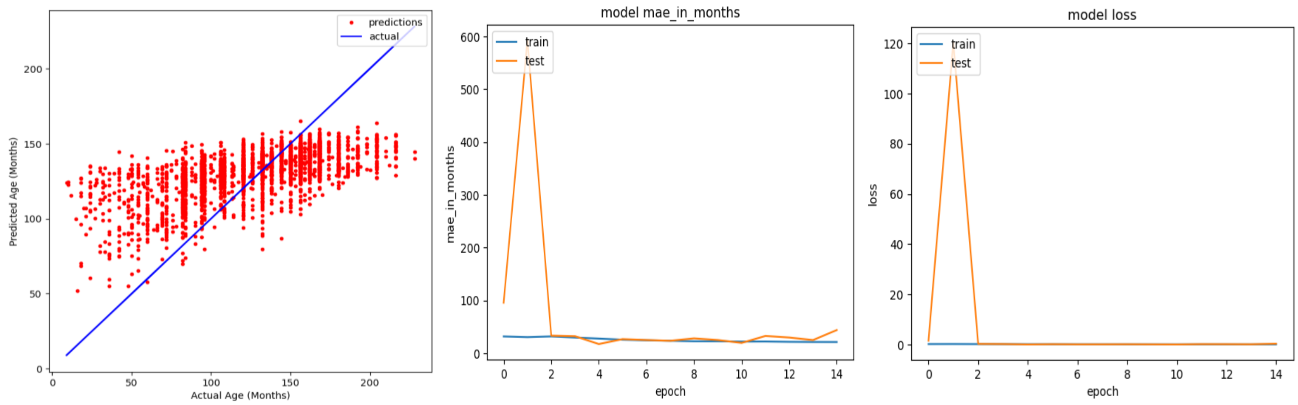

Fig. 2 ResNet50 training results

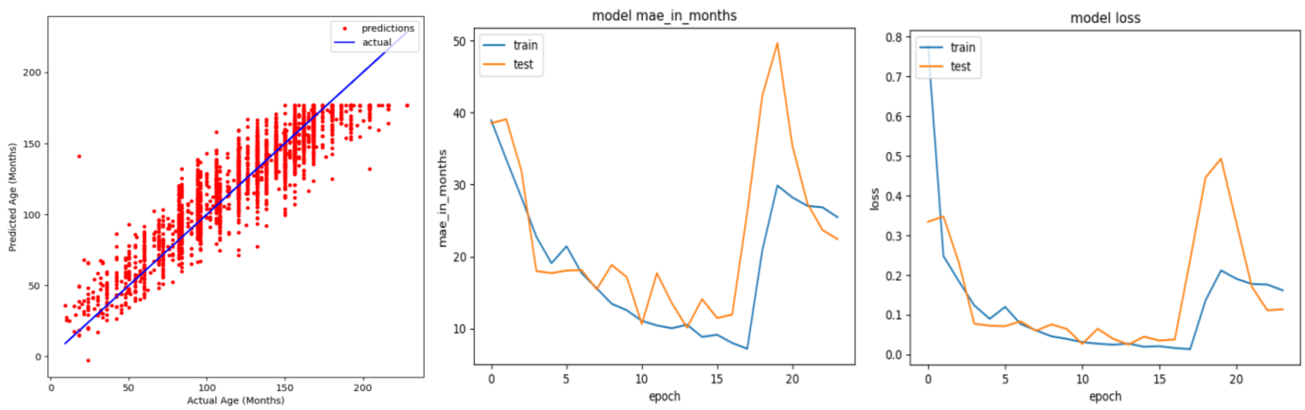

Fig. 3 Resnet152 training results

Fig. 4 Xception training results

Table 1. Training results of the four AI models

Models / Results	**Vgg16**	**ResNet50**	**ResNet152_2**	**Xception**
Training Time (RTX6000 GPU, 50 epochs)	78min39s	59.0min53s	101min55s	219min49s
Mean_absolute_error (months)	12.47	26.49	12.28	7.21
Median_absolute_error (months)	9.95	21.17	9.81	7.21

Comparison of the four AI models according to the execution time required to find the best model, the mean absolute errors of ages and the median absolute errors of ages. Some of the prediction results of the Xception model are listed in figure 5.

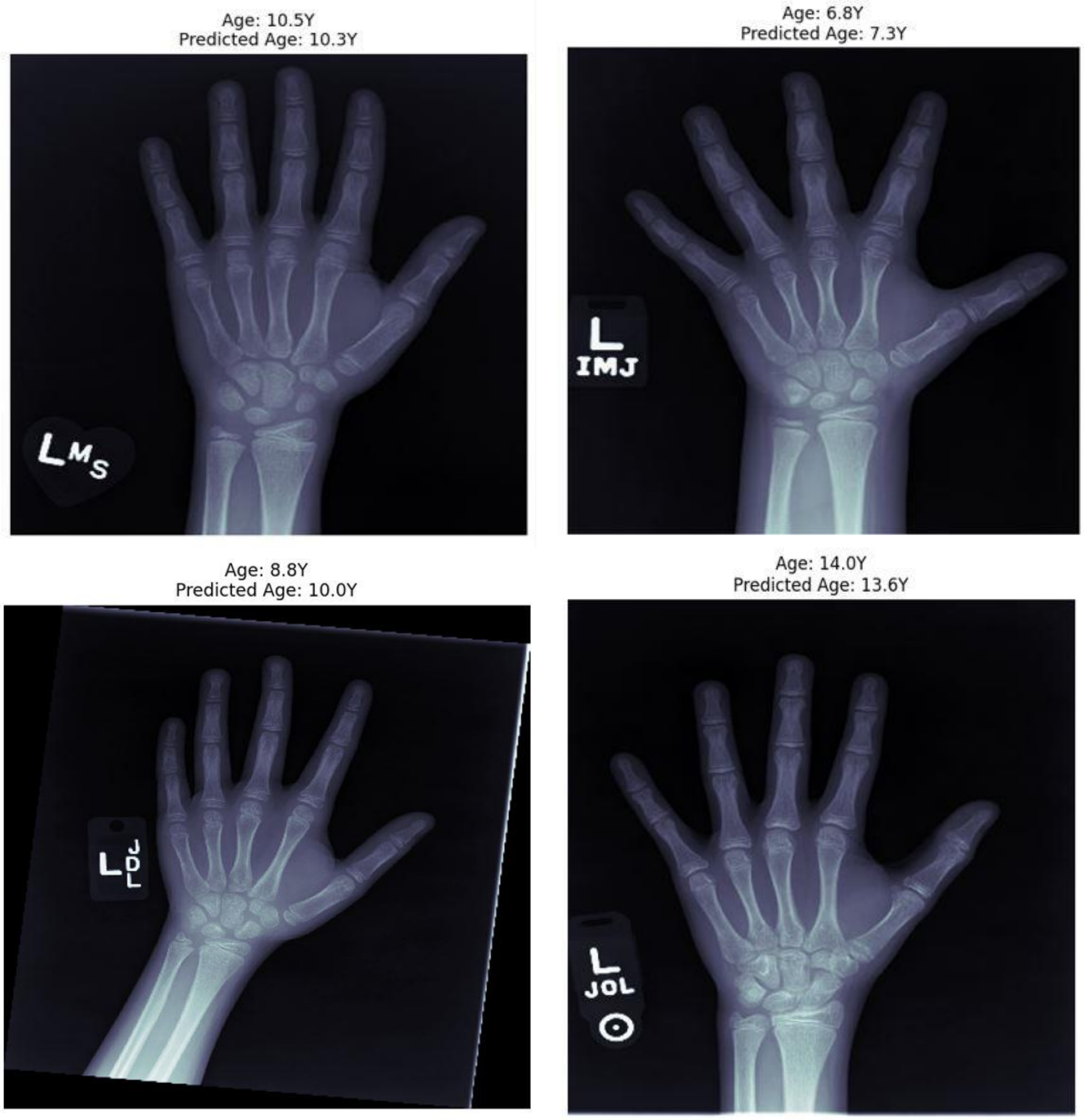

Fig. 5 Examples of prediction results of the Xception model

4. Discussion and Conclusion

From the experiments, we have the following observations.

1. We add Discriminators for both Vgg16 and ResNet50. The accuracy of Vgg16 is better than that of ResNet50, though ResNet50 has more layers.
2. Although ResNet152 use more time for training than ResNet50 dose, it still can converse and has better accuracy.
3. Xception model was designed to be more efficient than Inception3 while maintaining similar or even better performance on image classification and object recognition tasks. Xception utilizes depthwise separable convolutions, which can reduce the number of parameters and computational costs compared to traditional convolutional layers. Its accuracy is better than the other three models, although it takes a litter longer training time.

In conclusion, AI models have shown promise as a tool for bone age prediction, with the potential to improve this process's accuracy, consistency, and efficiency. However, further research and development are needed to address these models' challenges and limitations and ensure they are reliable and accurate in real-world settings. Using AI models in bone age prediction can lead to more accurate and efficient diagnosis and treatment of growth disorders in children.

References

[1] Xue-Lian Zhou et al., "Diagnostic Performance of Convolutional Neural Network-based Tanner-Whitehouse 3 Bone Age Assessment System," *Quantitative imaging in medicine and surgery*, vol. 10, no. 3, pp. 657-667, 2020. *Crossref*, https://doi.org/10.21037/qims.2020.02.20

[2] Byoung-Dai Lee, and Mu Sook Lee, "Automated Bone Age Assessment using Artificial Intelligence: The Future of Bone Age Assessment," *Korean Journal of Radiology*, vol. 22, no. 5, pp. 792-800, 2021. *Crossref*, https://doi.org/10.3348%2Fkjr.2020.0941

[3] Kyu-Chong Lee et al., "Clinical Validation of a Deep Learning-based Hybrid (Greulich-Pyle and Modified Tanner-Whitehouse) Method for Bone Age Assessment," *Korean Journal of Radiology*, vol. 22, no. 12, pp. 2017-2025, 2021. *Crossref*, https://doi.org/10.3348/kjr.2020.1468

[4] RSNA Pediatric Bone Age Challenge, 2017. [Online]. Available: https://www.rsna.org/education/ai-resources-and-training/ai-image-challenge/rsna-pediatric-bone-age-challenge-2017

[5] William Walter Greulich, and S. Idell Pyle, *Radiographic Atlas of Skeletal Development of the Hand and Wrist*, Stanford: Stanford University Press, 1959.

[6] Keiron O'Shea, and Ryan Nash, "An Introduction to Convolutional Neural Networks," *Neural and Evolutionary Computing*, 2015. *Crossref*, https://doi.org/10.48550/arXiv.1511.08458

[7] Jiuxiang Gu et al., "Recent Advances in Convolutional Neural Networks," *Pattern Recognition*, vol. 77, pp. 354-377, 2018. *Crossref*, https://doi.org/10.1016/j.patcog.2017.10.013

[8] Zewen Li et al., "A Survey of Convolutional Neural Networks: Analysis, Applications, and Prospects," *IEEE transactions on neural networks and learning systems*, vol. 33, no. 12, pp. 6999-7019, 2022. *Crossref*, https://doi.org/10.1109/TNNLS.2021.3084827

[9] Safwan S. Halabi et al., "The RSNA Pediatric Bone Age Machine Learning Challenge," *Radiology*, vol. 290, no. 2, pp. 498-503, 2018. *Crossref*, https://doi.org/10.1148/radiol.2018180736

[10] Ian Pan et al., "Improving Automated Pediatric Bone Age Estimation Using Ensembles of Models from the 2017 RSNA Machine Learning Challenge," *Radiology: Artificial Intelligence*, vol. 1, no. 6, 2019. *Crossref*, https://doi.org/10.1148/ryai.2019190053

[11] Priscilla Whitin, and V. Jayasankar, "A Novel Deep Learning-Based System for Real-Time Temperature Monitoring of Bone Hyperthermia," *SSRG International Journal of Electrical and Electronics Engineering*, vol. 10, no. 1, pp. 187-196, 2023. *Crossref*, https://doi.org/10.14445/23488379/IJEEE-V10I1P118

[12] Chi Fung Cheng et al., "Report of Clinical Bone Age Assessment Using Deep Learning for An Asian Population in Taiwan," *Biomedicine*, vol. 11, no. 3, pp. 50-58, 2021. *Crossref*, https://doi.org/10.37796%2F2211-8039.1256

[13] Shahein H. Tajmir et al., "Artificial Intelligence-assisted Interpretation of Bone Age Radiographs Improves Accuracy and Decreases Variability," *Skeletal radiology*, vol. 48, no. 2, pp. 275-283, 2019. *Crossref*, https://doi.org/10.1007/s00256-018-3033-2

[14] Luciano M. Prevedello et al., "Challenges Related to Artificial Intelligence Research in Medical Imaging and the Importance of Image Analysis Competitions," *Radiology: Artificial Intelligence*, vol. 1, no. 1, 2019. *Crossref*, https://doi.org/10.1148/ryai.2019180031

[15] Sikender Mohsienuddin Mohammad, "AI Automation and Application in Diverse Sectors," *International Journal of Computer Trends and Technology*, vol. 68, no. 1, 2020.

[16] Geert Litjens et al., "A Survey on Deep Learning in Medical Image Analysis," *Medical Image Analysis*, vol. 42, pp. 60-88, 2017. *Crossref*, https://doi.org/10.1016/j.media.2017.07.005

[17] Karen Simonyan, and Andrew Zisserman, "Very Deep Convolutional Networks for Large-scale Image Recognition," *Computer Vision and Pattern Recognition*, 2014. *Crossref*, https://doi.org/10.48550/arXiv.1409.1556

[18] Kaiming He et al., "Deep Residual Learning for Image Recognition," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 770-778, 2016. *Crossref*, https://doi.org/10.1109/CVPR.2016.90

[19] François Chollet, "Xception: Deep Learning with Depthwise Separable Convolutions," *Proceedings of The IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1800-1807, 2017. *Crossref*, https://doi.ieeecomputersociety.org/10.1109/CVPR.2017.195

[20] J. Carreira, H. Madeira, and J.G. Silva, "Xception: A Technique for the Experimental Evaluation of Dependability in Modern Computers," *IEEE Transactions on Software Engineering*, vol. 24, no. 2, pp. 125-136, 1998. *Crossref*, https://doi.org/10.1109/32.666826

[21] P. Nageswari, S. Rajan, and K. Manivel, "Medical Image Segmentation Approaches: A Survey," *SSRG International Journal of Electronics and Communication Engineering*, vol. 7, no. 7, pp. 1-3, 2020. *Crossref*, https://doi.org/10.14445/23488549/IJECE-V7I7P101

[22] Elham Beheshtian et al., "Generalizability and Bias in a Deep Learning Pediatric Bone Age Prediction Model using Hand Radiographs," *Radiology*, vol. 306, no. 2, 2022. *Crossref*, https://doi.org/10.1148/radiol.220505

[23] Module: tf.keras.callbacks. [Online]. Available: https://www.tensorflow.org/api_docs/python/tf/keras/callbacks