

# Latest Update in Bone Age Assessment Model with Deep Learning

Chadaporn Keatmanee, PhD<sup>1</sup>; Songphon Klabwong, MSc<sup>2</sup>; Chittiwat Suprasongsin, MD<sup>3</sup>; Kamolphong Osatavanichvong, MD<sup>4</sup>; Chirotchana Suchato, MD<sup>5</sup>



Chadaporn Keatmanee, PhD

## Abstract

Deep convolutional neural network (CNN) have performed remarkably well on medical image classification, including bone age assessment. The model is designed for evaluating the maturity of a child's skeletal system using x-ray images. The outstanding CNN model can accurately estimate bone age with a mean average error at 4.49 months. This survey presents existing CNN models for bone age assessment, promising developments, and high-level technical descriptions for implementing the CNN model. It aims to give background knowledge for non-computer scientists interested in deep learning application of bone age assessment. Readers will understand how CNN can improve the performance of bone age assessment models as well as its challenges for future research.

**Keywords :** deep learning, convolutional neural network, bone age, growth disorder, maturity estimation, transfer learning

Deep learning is a recent advancement and emerges from Computer Science. It is considered to be one of the most impactful inventions after electricity. The term Deep Learning is referring to a characteristic of computation structure in such a way that it contains many layers. Medical Imaging<sup>1</sup>, among many other fields in medicine, has the potential in applications to perform preventive and diagnostic measures. A systematic review<sup>2</sup> shows that deep learning consistently outperforms health-care professional's counterparts in various tasks in both sensitivity and specificity. In recent years, several deep learning approaches for Bone Age Assessment (BAA) have been purposed. Many of them provided promising results compared to health-care professionals and this creates the need for a comprehensive review of them. There is a recent literature review on BAA by Dallora et al.<sup>3</sup> However, it puts a limited focus on the technical aspects. Hence, this article will pay particular attention to automated BAA using deep learning and its latest developments in model architectures and training techniques as well as examining its strengths and weaknesses.

The structure of this article will be divided into three major components: model architectures, training techniques, and challenges. Different deep learning models will be introduced in the model architectures section, which will address both the basic and the latest state-of-the-art models. Training techniques will investigate two main techniques for training deep learning models, namely transfer learning and data augmentations. Lastly, the challenges section will discuss various types of challenges associated with BAA based on deep learning approaches and suggestions for possible solutions.

## Model architectures

### *Convolutional Neural Network (CNN) and the problem formulation*

An x-ray image of the hand and wrist are used to determine the discrepancy between skeletal bone age and chronological age. Two primary methods are TW3<sup>4</sup> and GP<sup>5</sup>, which are commonly performed by

<sup>1</sup> Department of Computer Science, Faculty of Science, Ramkhamhaeng University, Bangkok, Thailand

<sup>2</sup> Divertise Asia Co. Ltd., Bangkok, Thailand

<sup>3</sup> Research Center, Ramathibodi Hospital, Bangkok, Thailand

<sup>4</sup> Urupong Medical Center, Bangkok, Thailand

<sup>5</sup> Radiology Department, Bangkok Hospital, Bangkok, Thailand

\* Address Correspondence to author:  
Department of Computer Science,  
Faculty of Science, Ramkhamhaeng University,  
2086 Ramkhamhaeng 24, Hua Mak,  
Bangkok, 10240, Thailand.  
email: chadaporn@ru.ac.th

Received: March 25, 2020

Revision received: March 25, 2020

Accepted after revision: April 27, 2020

BKK Med J 2020;16(2): 231-236.

DOI: 10.31524/bkkmedj.2020.23.001

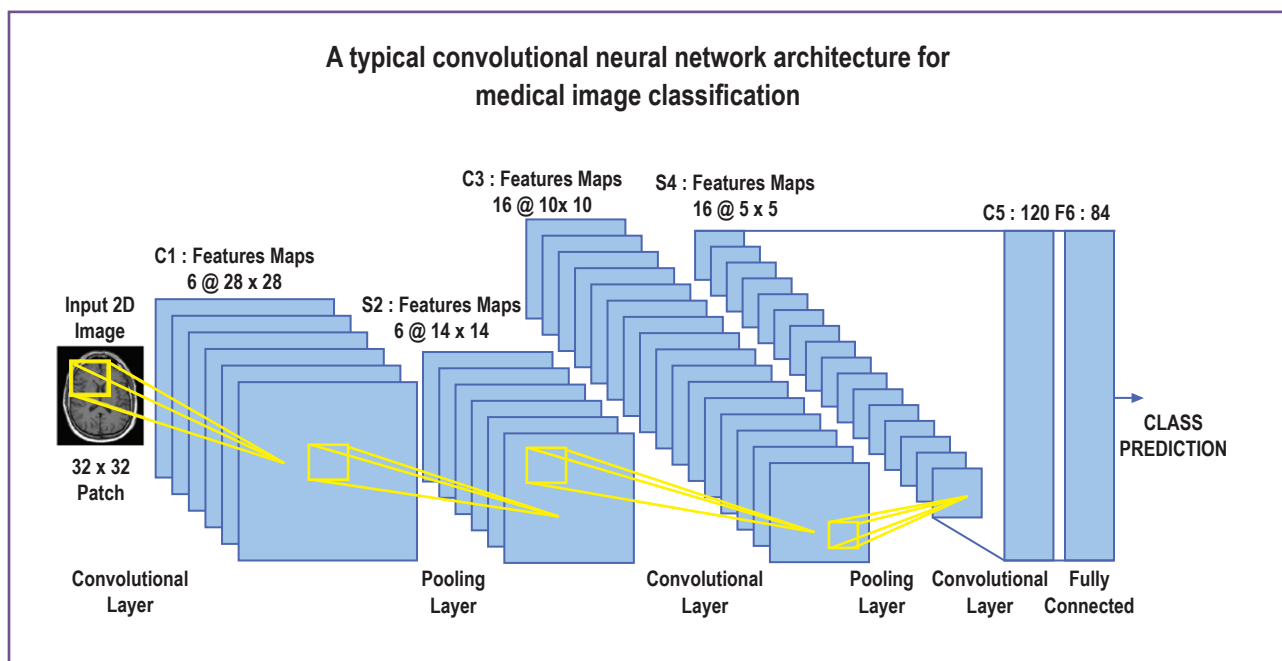
www.bangkokmedjournal.com

radiologists. In contrast, the state-of-the-art result of bone age predictions is using a technique called Convolutional Neural Network (CNN). Convolution refers to an intuitive mathematical operation, where the operation takes multiple input values to create one output value. It is a variation of Artificial Neural Network (ANN), a class of algorithms inspired by biological neural networks. It is one of the staple techniques used for medical image analysis.<sup>6</sup>

CNN is not necessarily interpreting an image in the same way as conventional methods as it does not look for a specific region of bones. Rather, it takes in all pixels in the input image into consideration and assigns different weights to each of them. These weights are key to the learning process, where they are adjusted through a training process with the *Backpropagation* algorithm. The algorithm calculates the disagreement values between predictions and ground truths given from a *training dataset*. These values are called loss

values and will be propagated back through the entire network. Thus, the training dataset is an essential part that shapes how well the model performs. The fact that the model statistically captures characteristics of training data implies a large amount of data will lead to better generalization.

By default, a CNN comprised repetitive groups of convolutional and pooling layers, followed by a fully connected layer, as depicted in Figure 1. Each layer consists of a specific number of the computational node so-called neurons. The neurons are merely mathematical functions that will be determined by the training process. This function is called the activation function. The most common *activation functions* in use are Rectified Linear Unit (ReLU) and Logistic Sigmoid (Logsig) functions. These are nonlinear functions, so that it can reflect the nature of non-linearity in real-world problems.



**Figure 1:** Repetitive structure of CNN<sup>6</sup> (Adaptation from J med Syst 2018;42:226)

The problem of BAA can be directly translated into the classification problem of image processing. Specifically, given an image  $X$ , what is the corresponding number  $y$  to  $X$ . In this case,  $X$  is an image of hand bone and  $y$  is a predicted age in months. Several studies view BAA as a regression problem such as Lee et al.,<sup>7</sup> in which a prediction result is a continuous number instead of an integer. Nonetheless, this difference will not affect how CNN operates except only the last layer of the model.

### Residual Network

In theory, CNN will perform better as we keep adding more layers. However, in the training process of a deep

network. Backpropagation tends to decrease the propagated values (so-called Gradient) exponentially. The results will become very close to zero on a deep network. This problem is called Gradient vanishing. Likewise, Gradient exploding is the exponential increase in gradients, resulting in too big a number to handle. These problems limited the number of layers in a model.

Deep Residual Network<sup>8</sup> is arguably one of the most crucial architectures that enable deep networks to be trained efficiently. The network uses a skip connection between layers, as represented in Figure 2. This setting preserves values flow through each layer by avoiding activation functions. As a result, the network can be trained up to a thousand layers in depth.

Automatic Whole-body Bone Age Assessment System suggested by Nguyen et al.<sup>9</sup> employed a modified version of VGGNet<sup>10</sup> by adding Residual connections between layers and has shown substantial improvement over the original VGGNet.

### Attention mechanisms

While CNN takes all pixels into consideration, the group of pixels that contribute to predictions may be limited to a subset of the pixels. Jetley et al.,<sup>11</sup> proposed trainable attentions in which the model will pay particular attention to certain areas of images. The attention mechanism also contributes to model interpretability which will be discussed in the challenges section. Figure 3 shows the bone age assessment model which was designed based on the CNN model architectures as described in this section.

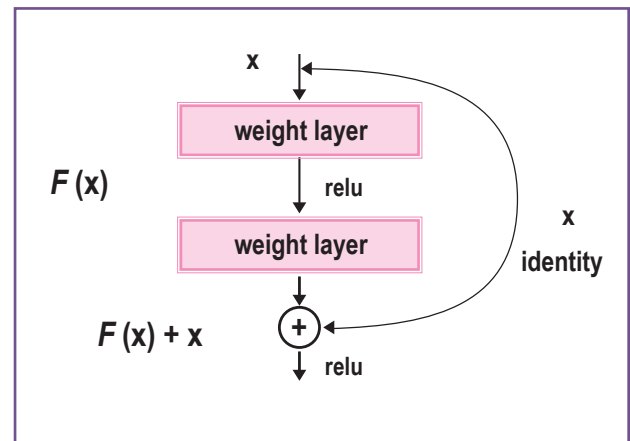


Figure 2: A building box of Residual learning<sup>8</sup>

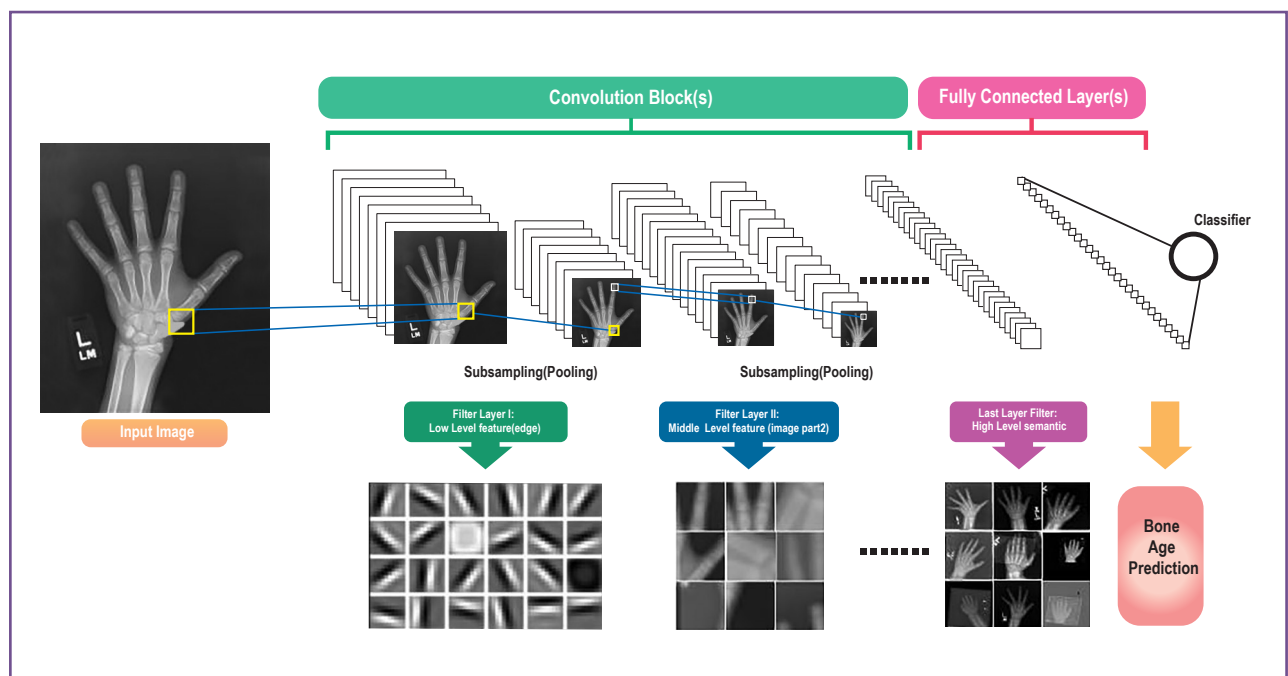


Figure 3: The model architecture of CNNs model for bone age assessment<sup>12</sup>

### The latest state-of-the-art BAA using CNN

PRNet<sup>13</sup> is the latest development in CNN-based BAA. The main idea of the paper is to capture the relationship between parts of the hand bone and the selection of discriminative parts as opposed to earlier proposed models, which consider the whole image or a part of it. The network consists of three main components, namely, the part relation module, part selection module and feature concatenation module. Firstly, the part relation module will connect the strong-correlated parts of the hand bones. The model does not require manual labelling for parts connection and is trained in an end-to-end manner. Secondly, the part selection module

selects useful parts to create different anchors, as shown in Figure 4. In the end, the selected anchors will be passed to the ranking module to compute anchor scores and confidences.

This approach shows a significant improvement of a result prediction of mean average error at 4.49 months, compared to baseline at 6.52 months and human reviewers at 7.32 months. The chart in Figure 5 depicts a comparison of different approaches. Although the network is currently the latest state-of-the-art model for BAA, there are still challenges present in its application as we will discuss in a later section.

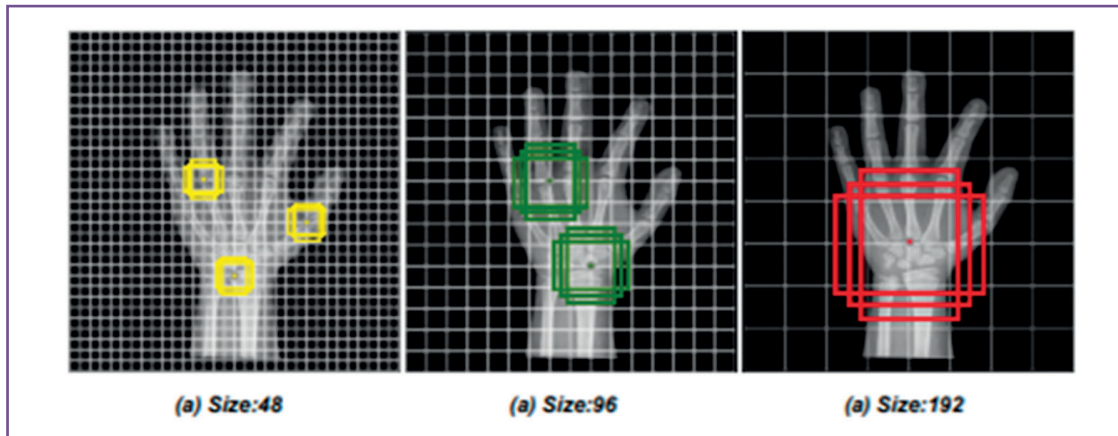


Figure 4: Example of displayed anchors in various sizes and scales<sup>13</sup>

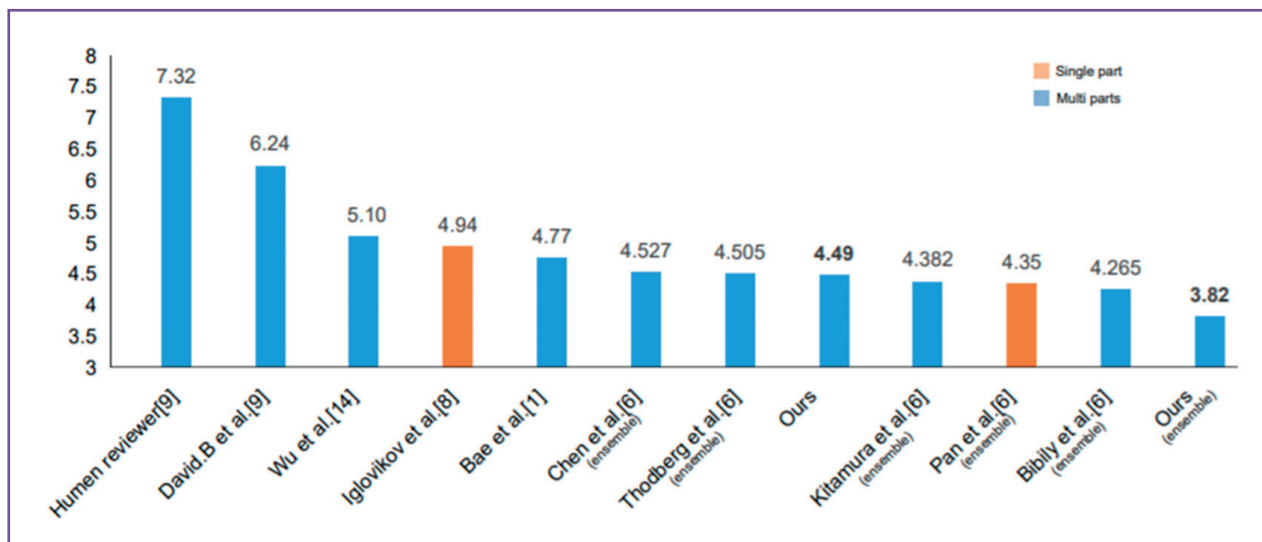


Figure 5: Comparison of the different approaches for BBA<sup>13</sup>

### Training technique

Neural networks generally require a training process. It is a method to adjust parameters in the network so that the network can perform accurate and reliable predictions. Two major techniques that are in prevalent usage for the training process are transfer learning and image augmentation. They can be used alone or more often in combinations.

#### Transfer learning

In general, CNNs require a large number of images to capture the internal structure of pixels and to associate with the corresponding target label. This becomes an issue in a particular domain where annotated training images are scarce.

Transfer learning is a technique to train a model on one data distribution and apply it to a different distribution. For example, a bone age CNN model can be trained using a large image dataset such as ImageNet<sup>14</sup> and subsequently, train with a smaller dataset of hand bone images. The works of Tang W

et al.<sup>15</sup> and Zhou J et al.<sup>16</sup> demonstrated the usage of transfer learning in BAA.

One caveat of transfer learning is that the target data distribution must not be too different from the pre-trained data distribution. Otherwise, the model is likely to suffer reduced performance from a negative transfer issue. A survey on Transfer Learning<sup>17</sup> has pointed out that few studies have been conducted on how to avoid a negative transfer.

#### Image Augmentations

The bone images can be altered by various image operations as explored in a survey paper on augmentations<sup>18</sup>, which is an inexpensive way to increase the number of training images. In principle, the image augmentation increases prediction performance by preventing the model from remembering the answer. Even the same image with a slight undetectable change by human eyes will be interpreted differently by CNN. Several studies<sup>19-21</sup> showed an improvement in the accuracy of predictions.

Although, the image augmentations helped to reduce the overfitting problem, it cannot become the sole answer to the problem of a small dataset. In a BAA task, if there is a missing bone image of a certain age, there is no known algorithm to augment images to compensate for missing data effectively and reliably.

## Challenges

In this section, we will discuss key challenges in the latest BAA using the deep learning model. Some of the challenges persist among different problems in image analysis, while others appear only in BAA and similar classification problems.

### Data availability

Generally, a CNN model will not perform any better than the given data. Lack of enough annotated data will severely affect model performance. Although the problem can be eased by using transfer learning and image augmentations, real annotated data is still the most important thing to focus on when it comes to data availability. Another problem is data correctness. The implication of model training from data is that if the given data is skewed or is misrepresenting the true distribution of the population, the BAA model will inevitably make inaccurate predictions if it is trained on the bone images with different growth rates from different periods and locations. These problems suggest that systematic data gathering and a review process is required before the images can be used in a training process.

### Model Generalization

Model generalization is an ability to infer the larger population from its subset. When a BAA model is trained using the same set of bone images too many times, it is likely to make a prediction specific to the training dataset, and hence, does not generalize well. We called this model overfitted with the given data.

The other problem that requires a model to generalize well is the population shift. The accuracy of BAA may be reduced over time as younger generations reach puberty

earlier. For this reason, the challenge of model adaptation to different population requires further study.

Recently, several active research efforts have been focused on creating a model with high generality using Reinforcement Learning (RL) and generative adversarial network (GAN). Despite demonstrating an encouraging result on various tasks, further study of BAA and medical image analysis is required.

### Interpretability

Transparency is an essential property in any medical system. The black-box nature of CNN presents a challenge to its applications. In BAA, we need to be able to explain the particular areas that allow the model to make a conclusion. This can be achieved through the means of visual attention such as Saliency Map<sup>22</sup> or Class Activation Map (CAM) proposed by Zhou et al.<sup>23</sup> to show focused areas. CAM enables visualization of the area where the model pays particular attention. Class-selective Relevance Mapping (CRM) by Kim et al.<sup>24</sup> displays substantial improvement in multi-modality. Ossification ROI Detection<sup>25</sup> is another way to improve the model interpretability by a dedicated model structure to actively look for joints and connecting bones.

## Conclusion

In this article, we firstly reviewed deep learning, especially focused on CNN and variants of the architecture. The latest state-of-the-art PRSNet model was summarized and its advantages discussed over other architectures. Secondly, we gave an overview of the main training techniques for BAA, their advantages, and shortcomings. Lastly, we discussed challenges in the application of CNN to BAA problems and medical image analysis in general. We have discussed the potential solutions for handling different challenges. We hope this paper may help researchers to grasp a basic understanding of BAA using Deep Learning and to suggest the model architectures to be used, as well as raising awareness of possible challenges and their solutions. Deep learning approaches will play a vital role in BAA and Medical Image Analysis in the foreseeable future.

## References

1. Hosny A, Parmar C, Quackenbush J, et al. Artificial intelligence in radiology. *Nat Rev Cancer* 2018;18(8):500-10.
2. Liu X, Faes L, Kale AU, et al. A comparison of deep learning performance against health-care professionals in detecting diseases from medical imaging: a systematic review and meta-analysis. *Lancet Digital Health*. 2019;1(6):271-97.
3. Dallora AL, Anderberg P, Kvist O, et al. Bone age assessment with various machine learning techniques: A systematic literature review and meta-analysis. *PloS one* 2019;14(7):220-42
4. Greulich WW, Pyle SI, Todd TW. Radiographic atlas of skeletal development of the hand and wrist. *Stanford: Stanford University Press*, 1959;2:150-9.
5. Tanner JM, Whitehouse RH, Cameron N, et al. Assessment of skeletal maturity and prediction of adult height (TW2 method). *London: Academic press* 1975;16.
6. Anwar SM, Majid M, Qayyum A, et al. Medical image analysis using convolutional neural networks: a review. *J Med Syst* 2018;42(11):226.
7. Jang L, Kwanggi K. Applying Deep Learning in Medical Images: The Case of Bone Age Estimation. *Health Inform Res* 2018;24(1):86-92.
8. He K, Zhang X, Ren S, et al. Deep Residual Learning for Image Recognition. *IEEE CVPR* 2016; DOI: 10.1109/CVPR.2016.90

9. Nguyen HD, Kim SH. Automatic whole-body bone age assessment using deep hierarchical features, 2019. (Accessed January 10, 2020, at <https://arxiv.org/pdf/1901.10237v1.pdf>)
10. Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition, 2014. (Accessed January 10, 2020, at <https://arxiv.org/pdf/1409.1556.pdf>)
11. Jitley S, Lord NA, Lee N, et al. Learn to pay attention, 2018. (Accessed January 10, 2020, at <https://arxiv.org/pdf/1804.02391v2.pdf>).
12. Keatmanee C, Klabwong S, Osatavanichvong K, et al. Performance of convolutional neural networks and transfer learning for skeletal bone age assessment. *BKK Med J* 2019;15(1).
13. Yuanfeng Ji, Hao C, Dan L, et al. PRSNet: Part relation and selection network for bone age assessment, 2019. (Accessed January 10, 2020, at <https://arxiv.org/pdf/1909.05651.pdf>).
14. Deng J, Dong W, Socher R, et al. Imagenet: A large-scale hierarchical image database. (Accessed January 10, 2020, at [http://www.image-net.org/papers/imagenet\\_cvpr09.pdf](http://www.image-net.org/papers/imagenet_cvpr09.pdf)).
15. Tang W, Wu G, Shen G. Improved automatic radiographic bone age prediction with deep transfer learning. *12<sup>th</sup> International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI)*, Suzhou, China, 2019:1-6. DOI: 10.1109/CISP-BMEI48845.2019.8965906
16. Zhou J, Li Z, Zhi W, et al. Dawes. Using convolutional neural networks and transfer learning for bone age classification. 2017 International Conference on Digital Image Computing: Techniques and Applications (DICTA), Sydney, NSW, 2017:1-6. DOI: 10.1109/DICTA.2017.8227503
17. Pan SJ, Yang Q. A survey on transfer learning. *IEEE Transactions on knowledge and data engineering* 2009;22(10):1345-59.
18. Shorten C, Khoshgoftaar TM. A survey on image data augmentation for deep learning. *J Big Data* 2019;6(1): 60.
19. Wang J, Perez L. The effectiveness of data augmentation in image classification using deep learning. *Convolutional Neural Networks Vis. Recognit*, 2017:11.
20. Mikołajczyk A, Grochowski M. Data augmentation for improving deep learning in image classification problem. *International interdisciplinary PhD workshop (IIPhDW)*, 2018:117-22.
21. Inoue H. Data augmentation by pairing samples for images classification, 2018. *arXiv preprint arXiv:1801.02929*.
22. Simonyan K, Vedaldi A, Zisserman A. Deep inside convolutional networks: Visualising image classification models and saliency maps, 2013. *arXiv preprint arXiv:1312.6034*.
23. Zhou B, Khosla A, Lapedriza A, et al. Learning deep features for discriminative localization. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016:2921-9.
24. Kim I, Rajaraman S, Antani S. Visual interpretation of convolutional neural network predictions in Classifying Medical Image Modalities. *Diagnostics* 2019; 9(2):38.
25. Koitka S, Demircioglu A, Kim MS, et al. Ossification area localization in pediatric hand radiographs using deep neural networks for object detection. *PloS one* 2018;13(11):0207496