JOIV : Int. J. Inform. Visualization, 7(1) - March 2023 105-114

INTERNATIONAL JOURNAL ON INFORMATICS VISUALIZATION

journal homepage : www.joiv.org/index.php/joiv

Human Bone Age Estimation of Carpal Bone X-Ray Using Residual Network with Batch Normalization Classification

Anisah Nabilah [a], Riyanto Sigit [a,*], Arna Fariza [a], Madyono Madyono [b]

[a] *Information and Computer Engineering Departement, Politeknik Elektronika Negeri Surabaya, Jl. Raya ITS, Surabaya, 60111, Indonesia*
[b] *Electrical Engineering Departement, Politeknik Elektronika Negeri Surabaya, Jl. Raya ITS, Surabaya , 60111, Indonesia*
*Corresponding author: *riyanto@pens.ac.id*

***Abstract*—Bone age is an index used by pediatric radiology and endocrinology departments worldwide to define skeletal maturity for medical and non-medical purposes. In general, the clinical method for bone age assessment (BAA) is based on examining the visual ossification of individual bones in the left hand and then comparing it with a standard radiographic atlas of the hand. However, this method is highly dependent on the experience and conditions of the forensic expert. This paper proposes a new approach to age estimation of human bone based on the carpal bones in the hand and using a residual network architecture. The classification layer was modified with batch normalization to optimize the training process. Before carrying out the training process, we performed an image augmentation technique to make the dataset more varied. The following augmentation techniques were used: resizing; random affine transformation; horizontal flipping; adjusting brightness, contrast, saturation, and hue; and image inversion. The output is the classification of bone age in the range of 1 to 19 years. The results obtained when using a VGG16 model were an MAE value of 5.19 and an R^2 value of 0.56 while using the newly developed ResNeXt50(32x4d) model produced an MAE value of 4.75 and an R^2 value of 0.63. The research results indicate that the proposed modification of the residual training model improved classification compared to using the VGG16 model, as indicated by an MAE value of 4.75 and an R^2 value of 0.63.**

***Keywords*— Forensics; Carpal bone; convolution neural network; bone age; batch normalization.**

Manuscript received 13 Jun. 2022; revised 15 Aug. 2022; accepted 29 Sep. 2022. Date of publication 31 Mar. 2023.

I. INTRODUCTION

In the case of an unknown victim of a crime or disaster, examining the body is important to determine the victim's identity [1]. Forensic doctors commonly use several parts of the hand bones to assess the age of a person's bones. Forensic doctors' most widely used parameters to analyze the age of hand bones are the epiphysis and the metaphysis of the bone. Evaluation of bone age is usually carried out by radiological examination of the skeletal development of the left hand and then comparing the result with chronological age [2].

Changes in the human skeleton are similar for all bones because the bone development process is continuous and passes through the same stages, where bones have certain characteristics at each stage. Therefore, compared to chronological age (CA), bone age (SA) is a more accurate way of reflecting an individual's growth and maturation rate [3]. Carpal bones are formed after birth by the ossification of carpal cartilage. The capitate and uncinatus bones are the first to show a center of ossification (2nd to 4th months), while the pisiform bones are the last (9 to 12 years) [4]. The dominant method used by experts in clinical practice are the methods of Greulich and Pyle (GP) [5], and Tanner and Whitehouse (TW2/3) [6].

However, there are issues with manual bone age estimation, for example, as a result of insufficient bone X-ray image quality, the difficulty of assessing the level of ossification in the carpal bone phalanges, and factors such as the experience, health condition, concentration level, and biases of the radiologist. These issues influence the bone age estimation result, as considerable intra-observer and inter-observer variability always exist in clinical practice [7], [8].

Deep learning has recently received a great deal of attention as an approach to realizing artificial intelligence. Deep learning is a machine learning method often used in fields such as image recognition or categorization, speech recognition, and natural language recognition. Convolution neural networks are known to be very efficient for bone age estimation from X-ray images of the hand [9]. Several bone age recognition techniques have been developed based on the

hand carpal bone using CNNs. In 2018, Van Steenkiste et al. [10] used a VGG16 module with architectural modifications in the direction of regression, resulting in an accuracy of 94.45%.

Furthermore, in 2020, a study was conducted to classify bone age using a VGG16 training model, resulting in a mean absolute difference (MAD) value – also known as the mean absolute error (MAE) – of 0.50 [9]. Bulò et al. [11] added a batch normalization process to enhance the training process. Batch normalization makes the training process more optimal, faster, and more stable because it can reduce internal covariate shifts in the network. Bulò et al. focused on improving the memory efficiency of modern architecture training processes such as ResNet, ResNeXt, Inception- ResNet, wideResNet, etc., on making the neural network performance more optimal for semantic image classification or segmentation.

This paper proposes a new approach to estimating human bone age using a residual network architecture based on the carpal bones in hand. The classification layer was modified by batch normalization to optimize the training process. Before carrying out the training process, image augmentation was done to enhance and add variation to the hand X-ray images used in the training process to maximize the accuracy of the results. The augmentation techniques used were resizing; random affine transformation; horizontal flipping; adjusting brightness, contrast, saturation, and hue; and image inversion. The model used for the training process was ResNeXt50(32x4d). The output of the proposed system is an estimation of human bone age in the range of 1 year to 19 years.

The remainder of this paper is organized as follows. Section 2 extensively surveys the latest papers on bone age estimation based on deep learning. Section 3 describes the proposed bone age estimation method using deep learning. Section 4 compares the experimental results of the commonly used VGG16 method and the proposed residual network method. Section 5 contains the conclusion of this paper.

II. Materials and Method

The stages of research carried out in this study start from identifying the problem and objectives, then conducting a literature study, collecting data used as input, performing system design, system testing, analyzing results, and finally, concluding.

A. Research Analysis

Based on Cavallo et al. [12] observations of bone maturation, it is best to define bone age based on hand and wrist radiographs because the hand and wrist correctly reflect the maturity of various types of skeletal bones. Carpal bones are very suitable to be used because they harden gradually throughout the ossification process, starting from the primary centers. Al-Khater et al. [13] conducted a study to collect data on the carpal bone ossification centers (the lower end of the radius and ulna).

In addition, they examined the order of appearance of the carpal bones and the relationship of this sequence to the appearance of the distal epiphyses of the radius and ulna. The research consisted of a retrospective radiological study from 2012 to 2020 at King Fahad University Hospital, Al- Khobar, Saudi Arabia. The dataset contained 279 X-ray images of Saudi children's wrists. It was revealed that the first bones in the wrist area that appear are the capitate, hamate, and distal radius epiphyses. These bones appear in the first year of life, after which other bones develop at annual intervals. The last to appear is the piriformis, late in the first decade of life or late childhood (before puberty) [14], [15]. Skeletal Age Assessment (SAA) is a clinical procedure used to determine the skeletal age of children and adolescents. Several factors, including nutrition, hormonal secretions, and genetics, influence bone development.

There are several factors to assess the maturity of the framework. These include inter- method variability, degree of variability in skeletal maturation estimates, low sources of accuracy, and dispersion of skeletal maturation values. Currently, the main clinical methods for skeletal age assessment are the Greulich and Pyle (GP) and Tanner and Whitehouse (TW) methods [15]. The GP method is based on a hand atlas, which consists of a series of X-ray templates of children's growth stages with varying degrees of bone maturity. The patient's X-ray images were then compared with the samples in the template series, and the most suitable template was selected as the patient's equivalent bone age. This approach is straightforward and can be done quickly.

However, the GP Method is characterized by inter- and intra-observer variability. Moreover, it is difficult to accurately assess bone with large size variations, and the resulting bone age is very rough because the template series is arranged in intervals ranging from six months to one year [16]. TW2 increases intra- and inter-rater variability by proposing discrete stages [17]. In conclusion, several standard methods have been developed to evaluate bone maturity from hand and wrist radiographs [8], [9]. However, there are still weaknesses in manual bone age estimation, starting from factors such as the radiographer's knowledge, experience, and conditions, which may affect the estimation results. So there is a need for a system that can analyze in detail and automatically in bone age observations that can help forensic experts.

Deep learning has been proven to be very successful in image feature recognition. Van Steenkiste et al. [10] estimated bone age by applying deep learning to X-ray images of the bones of the hands of children whose ages ranged from 0 to 228 months. The dataset contained 12,661 X-ray images. The convolution neural network VGG16 was used as the learning method. The VGG16 architecture consists of 16 layers. Since it was originally designed for classification, the architecture was modified in the direction of regression. This study succeeded in achieving a classification accuracy of 94.45%.

Marouf et al. [9] also developed an automated method for age estimation and worked on classifying bone images to identify gender, age, race, and target status in forensic identification. The method used to predict gender was a deep convolution neural network (DCNN), and the gender classification results had an accuracy of 79.6%. The VGG16 model was used to estimate age, for which a mean absolute deviation (MAD) value of 0.50 years and a root mean squared (RMS) value of 0.67 years were obtained.

Automated bone classification methods are carried out with a large number of images, and the complexity level of the CNN architecture is high, so computation takes a long time. Batch normalization is commonly used in CNNs to increase

training speed and stability. Rota, Lorenzo, and Peter [11] modified a deep learning method by adding batch normalization to the plugin activation layer to reduce the training time. The results of the experiments they carried out proved that their method could save up to 50% of memory usage for image classification. Chai, Pilanci, and Murmann [18] used concepts from the traditional adaptive filter domain to provide insight into the dynamics and workings of batch norms. Their experiments showed that the method provides several benefits in terms of speed and stability. If the learning speed is low, then batch normalization of the smallest eigenvalue increases the convergence speed, whereas if the learning speed is high, batch normalization of the largest eigenvalue ensures stability. In the training process, classification with batch normalization achieved the same level of optimization as the normalized least mean squares (NLMS) method.

In several studies [11], [16], [17], attempts have been made with deep learning approaches to reduce complexity and produce better age estimates. After developing several VGG16 models, Xie et al. recently introduced a completely new model [19], i.e., ResNeXt50. This architecture was a winner in ILSVRC 2016. The network is built by repeating ResNet architecture blocks. This simple design results in a homogeneous, multi-pronged architecture that has only a small number of hyperparameters to set. It combines a set of transformations with the same topology. This method could increase the cardinality when the capacity of the dataset was increased and could also increase the classification accuracy on the ImageNet 1,000 dataset. In the present study, we used a combination of the latest methods and models for the training process to produce higher accuracy of bone age estimation. We used a CNN for the model to predict bone age, i.e., ResNeXt50(32x4d) [19]. Architectural modifications were made to the model, namely adding a batch normalization layer, activating ReLu, and fully connecting the layer to make the data learning process more stable and save memory.

B. Proposed Methodology

The proposed method for bone age estimation is based on X-rays of the carpal bones in hand. Fig. 1 is a flowchart of the research reported in this paper, and the input consisted of hand radiographic images.

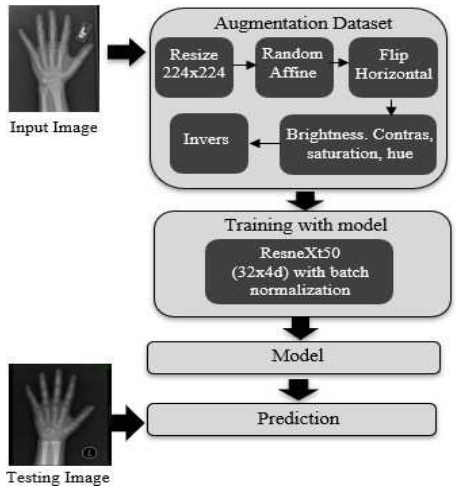

Fig. 1 General proposed method

The first step was image augmentation, carried out with five techniques, namely resizing (the image is resized to 224 x 224 pixels); random affine transformation; horizontal flipping; adjusting brightness to 0.4, a contrast to 0.1, saturation to 9, hue to 0.5; and finally, inversion of the image. After the augmentation process, dataset training was carried out using a convolution neural network with a residual model, namely ResNeXt50(32x4d), which was modified by adding batch normalization to make the training process more optimal, faster, and more stable.

C. Data Collection

The dataset consisted of X-ray images of male and female carpal bones of human hands in the bone age range of 1 to 19 years obtained from the Radiological Society of North America (RSNA) [20]. Fig. 2 shows the images used were photos of the left hand, file type .png with a size of 1514 x 2044 pixels and a bit depth of 8. In this study, 12,732 left-hand X-ray images of males and females were used for the training process, and 134 left-hand X-ray images of males and females were used for testing.

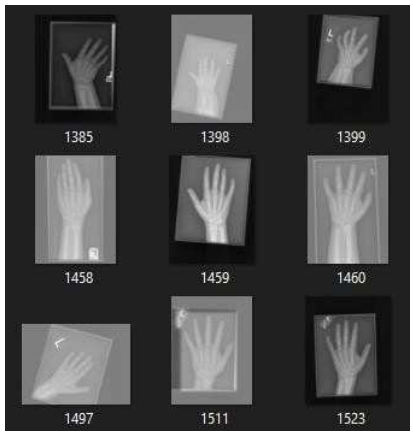

Fig. 2 X-ray hand images

Because of the uneven distribution between classes, as can be seen in Fig. 3, less than optimal accuracy will be achieved when using this dataset in the training process. Thus, the dataset needed to be balanced first. This was done using the subset random sampler function, which samples elements according to a given list of indexes without replacement. The resulting balanced dataset contained 11,339 X-ray images of carpal bones in each class.

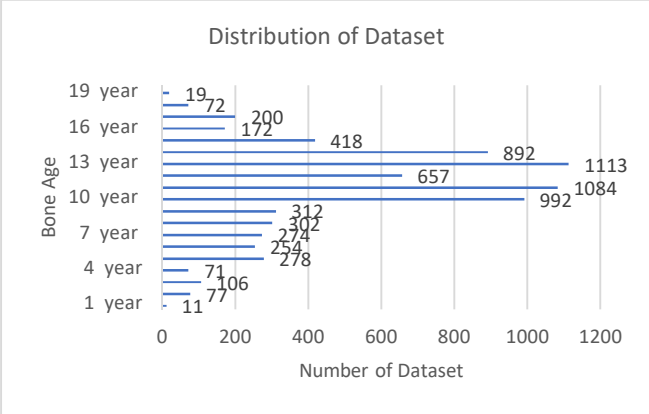

Fig. 3 Distribution of bone age in the hand X-ray images in the dataset

Fig. 4 Distribution of the images for training, validation, and test processes

Next, we divided the training and validation datasets, namely 80% for training and 20% for validation. Fig. 4 shows the distribution of the images for the training, validation, and testing processes. The training and validation dataset consisted of secondary data from the pediatric hand radiographs dataset (RSNA 2017), while the testing dataset consisted of primary data from Dr. Soetomo Hospital, Surabaya.

D. Augmentation Dataset

Data augmentation is a technique that can be used to artificially expand the size of a training dataset by creating modified versions of the images in a dataset [21]. Adding more images to the training dataset can produce a better model. The results of the augmentation step are shown in Fig. 3. The augmentation process changes the image size to 224 x 224 pixels. Then, a random affine transformation is done by simultaneously applying translation, rotation, scale enlargement, and crop.

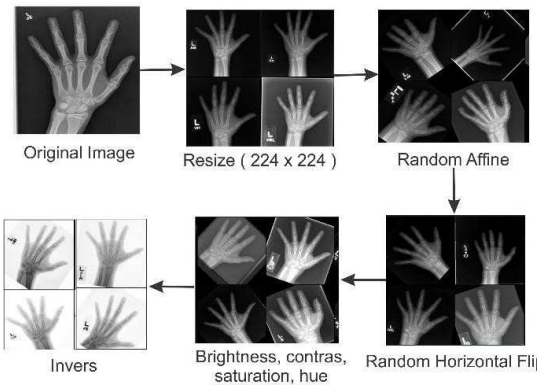

Fig. 5 X-ray images after augmentation

The affine transformation function is used to vary the images by transforming the 2D coordinate values into a new 2D coordinate system. Eq. (1) is used to calculate the affine transformation.

$$x' = a_1 x + b_1 x + c_1$$
$$y' = a_2 x + b_2 x + c_2 \quad (1)$$

where $a1$, $b1$, $c1$, $a2$, $b2$, and $c2$ are the transformation parameters, with $a1 \neq b2$, $a2 \neq b1$. This does not produce the right shape, so changes in angles and distances are done by crammer elimination to complete the transformation step. Then the affine transformation matrix is formed with Eq. (2).

$$\begin{pmatrix} Q_X \\ Q_Y \\ 1 \end{pmatrix} = \begin{pmatrix} m_{11}P_x + m_{12}P_y + m_{13} \\ m_{21}P_x + m_{22}P_y + m_{23} \\ 1 \end{pmatrix} \quad (2)$$

Eq. (2) shows that the affine transformation affects four basic transformations: translation, scale, rotation, and shear. The next augmentation process is horizontal flipping, which consists of horizontally flipping the row and column of the image pixels. The reason for using this function is to rotate the image horizontally. The next step is to adjust the brightness, saturation, and hue to improve the visibility of the target of the research, i.e., the carpal bones in hand X-ray images. Using these three augmentation techniques, the carpal portion of the hand is still not optimally visible. Therefore, the image is further improved by adjusting the brightness, contrast, and saturation to enhance the sharpness of the image and, finally, adjusting the hue to enhance the color nuances in the image. The next step is image inversion to convert a positive image into a negative one, for which Eq. (3) is used.

$$f_0(x,y) = f\ maximum - f_i\ (x,y) \quad (3)$$

The maximum value of f in Eq. (1) is the highest value in color bits. In this study, images with a grayscale value of 8 bits were used, so the maximum value of f was 255. As can be seen in Fig. 6, the augmentation process makes the X-ray image of the hand clearer, especially the carpal bone part. The image then proceeds to the next process, namely feature extraction using a convolution neural network.


Fig. 6 X-ray images after augmentation

E. ResNeXt50(32x4d) with batch normalization in the classification stage

Feature extraction is performed using the ResNeXt architecture. It is designed simply to produce a homogeneous multi-pronged architecture with only a few hyperparameters to define. This strategy exposes a new dimension, which we call "cardinality" (size of the transformation set), as an important factor in depth and width dimensions. In this paper, we use ResNeXt with a cardinality of 32 for feature extraction as shown in Fig. 7

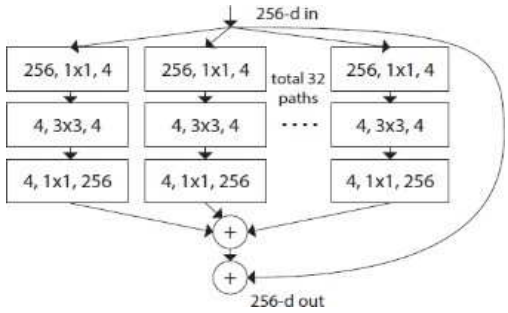

Fig. 7 ResNeXt architecture block

Fig. 8 shows a detailed architecture of ResNeXt-50 with cardinality C = 32 and bottleneck area d = 4. This is the ResNeXt50 (32x4d) architecture is preceded by an input layer for images with a size of 224 x 224 pixels. The images are grayscale, which means there is only one channel. Then, the convolution process is carried out with 64 neurons and an image size of 7 x 7. Convolution is carried out by repetition gradually, as shown in Figure 8(a),(b),(c),(d), then convolution will be lowered after convolution highest to avoid overfitting.

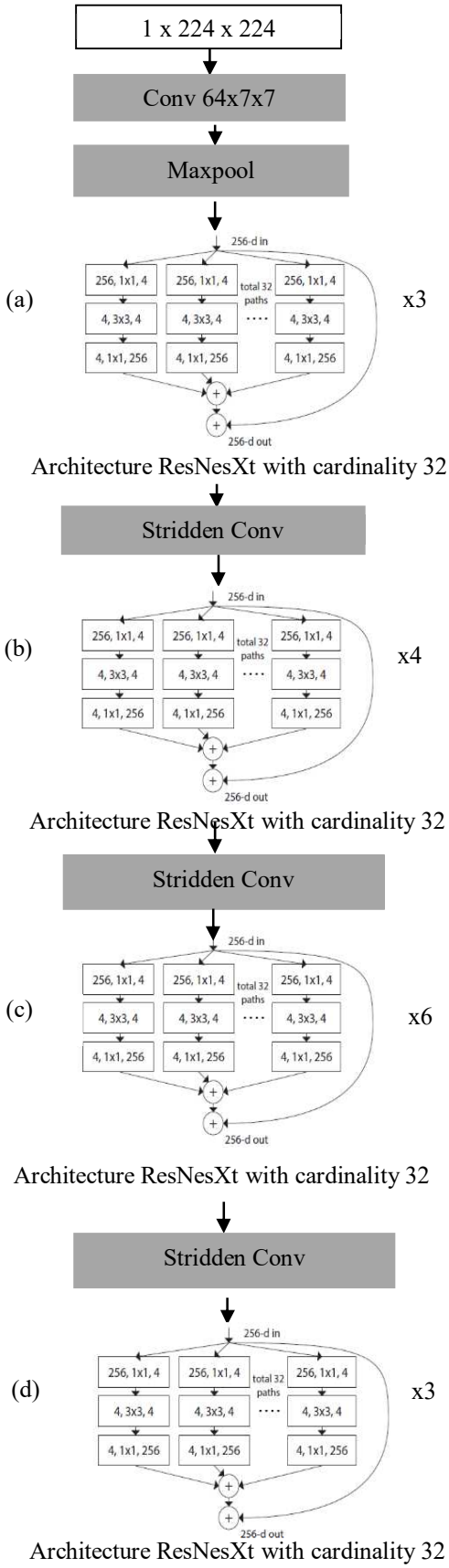

Fig. 8 ResNeXt50 (32x4d) convolution network architecture (a) convolution three times (b) convolution four times (c) convolution six times (d) convolution three times

Next is the max pooling layer, which aims to reduce the input spatially by taking its highest value. The next layer consists of the ResNext50 architecture carried out in parallel three times. Then follows stride convolution and ResNext50 parallelized four times. This process is repeated until the last layer of the ResNext50 architecture has been parallelized three times.

Classification is done to estimate bone age in the range of 1 to 19 years, so 19 classes must be classified. Classification is carried out through a training process with a convolution neural network, but layer modifications were made to optimize the training process and produce a higher accuracy level. Fig. 9 is the modified architecture of the CNN model for classification.

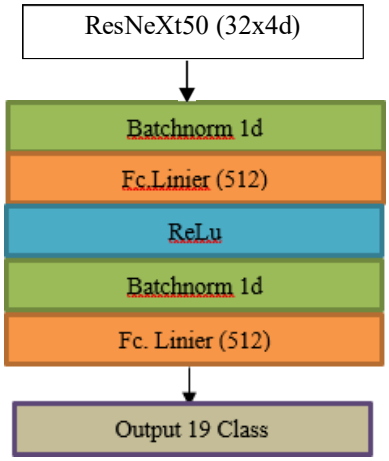

Fig. 9 Modification of the 19-class classification architecture using batch normalization.

The batch normalization layer already has a dimension of 1D because it has undergone a flattening process. After that, it enters a fully connected layer, which uses 512 neurons, aiming for more detailed features in the classification. After that, a ReLu activation function was added, followed by two more layers: a batch normalization 1D layer and a fully connected layer, like in the previous step. The output of this classification process is 19 classes.

F. Optimized Hyperparameters

To control the training process, we must set optimized hyperparameters before the training process creates the model to be used. The model needs different limits, weights, and learning speeds to generalize different data patterns. This process aims to make the machine learning process run optimally. In this study, the following hyperparameters were set: the number of epochs was set at 200; the learning rate was set at 0.0001 using the Adam optimizer; the batch size was set at 50; the loss was set using loss entropy, as given in Table 1.

TABLE I
OPTIMIZED HYPERPARAMETERS

Hyperparameter	Value
Batch size	50
Epoch	200
Optimizer	Adam
Loss	Cross entropy
Learning rate	0.0001

Eq. (4) was used to calculate the loss using cross entropy [22].

$$loss = \sum_i \log\left[\frac{\exp(W_{x_i})}{\sum_j \exp(W_{x_j})}\right] \quad (4)$$

Likewise, the derived gradient was used for updating the weights, following the Adam optimizer, asexpressed in Eq. (5) [22].

$$m_t = \beta_1 m_{t-1} + (1-\beta_1) g_1$$
$$v_t = \beta_2 v_{t-1} + (1-\beta_2) g_t^2 \quad (5)$$

III. Result and Discussion

An augmentation experiment was carried on the dataset training process using a residual convolution neural network (CNN) to estimate bone age in the range from 1 to 19 years, using Spyder Notebook 4.1.5 and the Python 3.8 programming language on an Intel Core i7 system with 16 GB RAM and 2 TB ROM storage and running the Windows 10 operating system. The model's performance was tested by entering a primary dataset of 134 hand X-ray images obtained from Dr. Soetomo Hospital Surabaya. The model's performance is presented in terms of the mean absolute error (MAE), and R squared.

$$MAE = \sum \frac{|Y'-Y|}{n} \quad (6)$$

Formula 6 was used to calculate the mean absolute error (MAE), where Y' is the predicted value, Y is the actual value, and n is the number of images.

$$SSR = (\sum Y_{Prediction} - Y_{average})^2 \quad (7)$$

$$SST = (\sum Y_{Actual} - Y_{average})^2 \quad (8)$$

$$R^2 = \frac{SSR}{SST} \quad (9)$$

Formula (9) was used to calculate R squared. Sum Square Regression (SSR) is the square of the difference between the predicted Y value and the average Y value, as shown in Eq. (7). Total Sum of Squared (SST) is the square of the difference between the actual Y value and the average Y value, as shown in Eq. (8).

TABLE II
Accuracy and Loss for Training with CNN Models

No	Model	Result	
		Accuracy	Loss
1	VGG16	87.6%	3.25
2	VGG16 with Batch Norm	87.5%	2.99
3	ResNeXt50(32x4d)	86.6%	3.67
4	ResNeXt50(32x4d) with Batch Norm	95.9%	3.25

The next stage was feature extraction, which was carried out with a CNN-based training process, with ResNeXt50(32x4d) and VGG16 [23] as comparison models. After that, the training process was added with the proposed layer modifications for classification. Table 2 presents the accuracy and loss values that resulted from the training process that was carried out with the two CNN models.

The training process results presented in Table 1 show that the highest accuracy value of 95.9% was achieved using the ResNesXt50(32x4d) training model with modified classification. The training was carried out with the epoch hyperparameter set to 200, a batch size of 50 using the Adam optimizer, loss calculated using cross-entropy, and a learning rate of 0.0001. Fig. 10 shows a graph of the accuracy and loss of the training results.

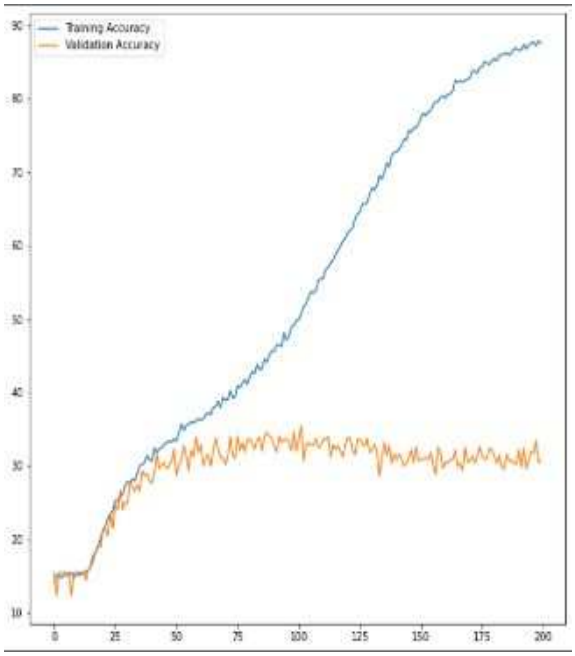

Fig. 10 Accuracy graph of VGG16 training process without batch normalization

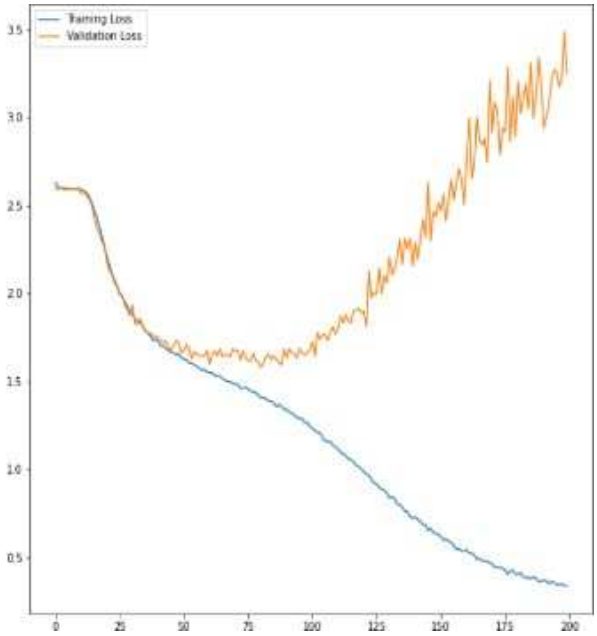

Fig. 11 Loss graph of VGG16 training process without batch normalization

The graph in Fig. 10 shows the accuracy from epochs 1 to 200 when the training process was carried out without classification layer modification. The training accuracy, represented by the blue line, increased, while the validation accuracy, represented by the orange line, was stable at a value of 30 to 40. The resulting accuracy value was 87.6%. Fig. 12

shows the accuracy graph after adding the normalization batch process to the classification layer.

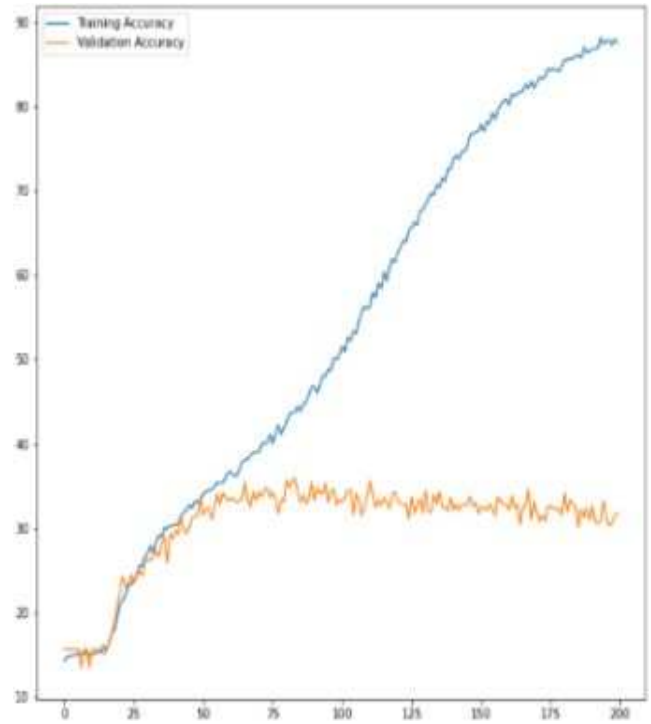

Fig. 12 Accuracy graph of VGG16 training process with batch normalization

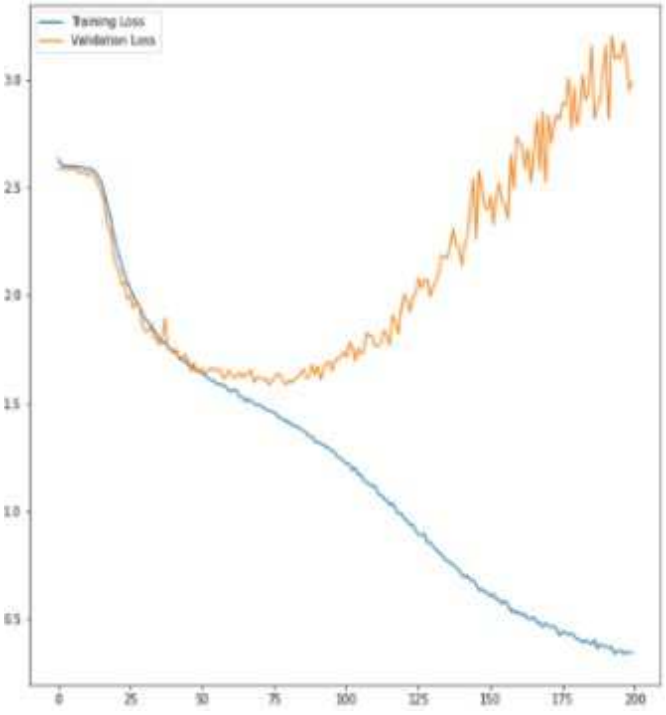

Fig. 13 Loss graph of VGG16 training process with batch normalization

The training accuracy increased, while the validation accuracy was still stable at 30 to 40, producing an accuracy value of 87.5%. Comparing the loss graphs in Fig. 11 and 13, it can be seen that the lowest loss value was 2.99, which was achieved when using the VGG16 model with batch normalization.

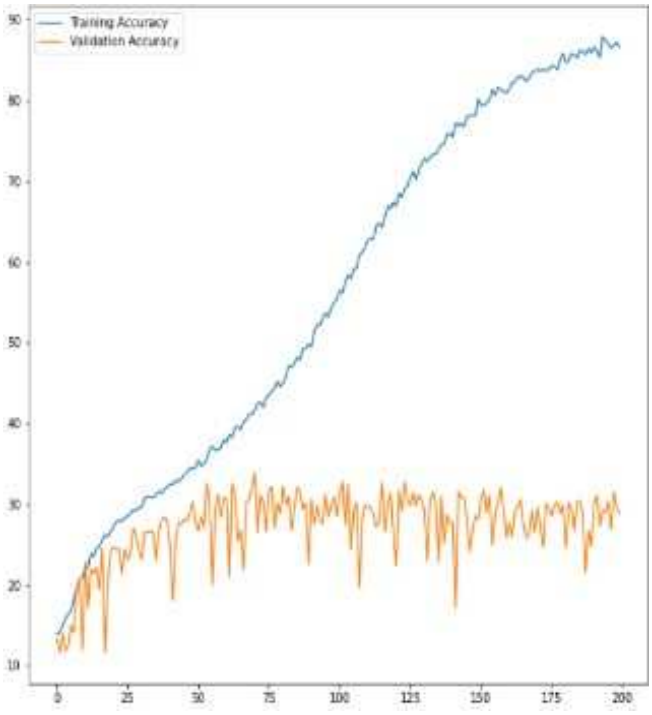

Fig. 14 Accuracy graph of the ResNeXt50(32x4d) training process without batch normalization

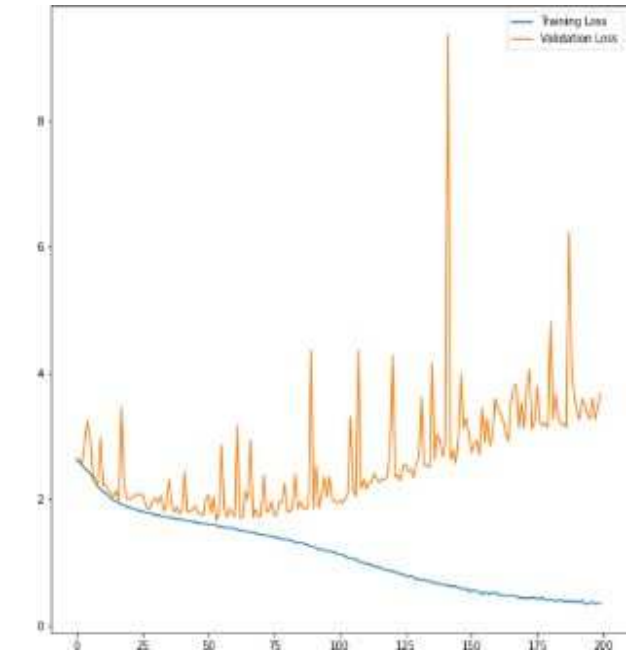

Fig. 15 Loss graph of the ResNeXt50(32x4d) training process without batch normalization

Fig. 14 shows the accuracy graph for the ResNeXt50(32x4d) training model from epochs 1 to 200. The training accuracy, indicated by the blue line), increased, while the validation accuracy, indicated by the orange line, increased to a value of 30. The resulting accuracy value was 86.6%. Furthermore, in Figure 16, the training accuracy increased, while the validation accuracy also increased to a value over 30, producing an accuracy of 95.9%.

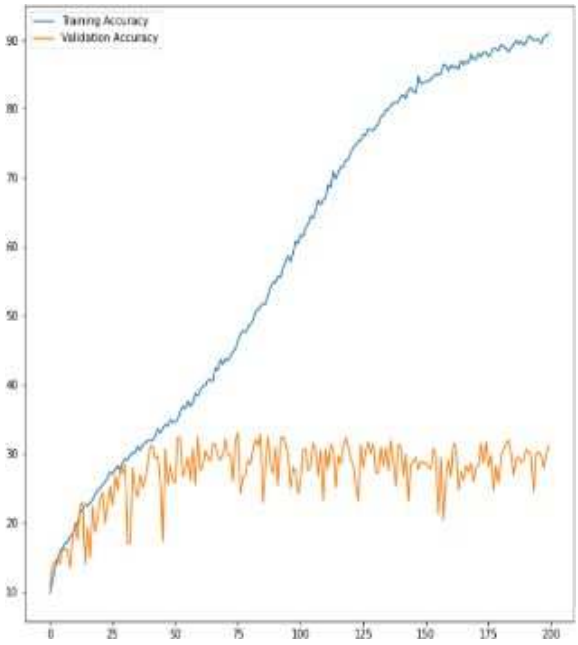

Fig. 16 Accuracy graph of the ResNeXt50(32x4d) training process without batch normalization

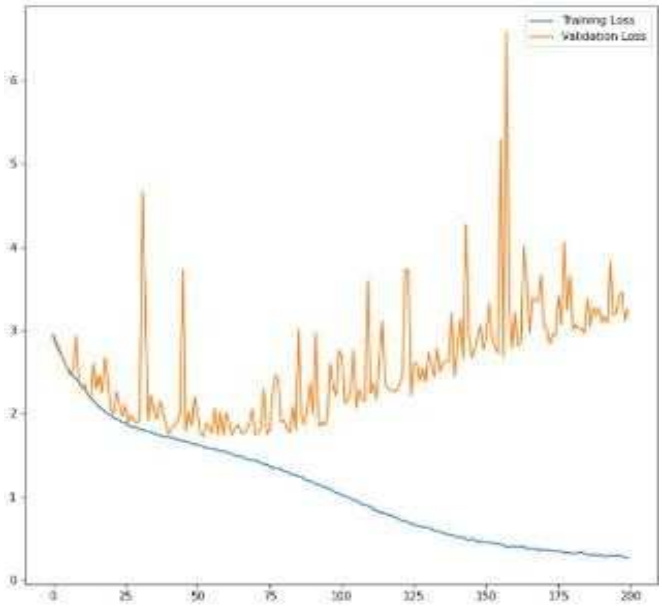

Fig. 17 Loss graph of the ResNeXt50(32x4d) training process with batch normalization

Fig.13 and 17 show the loss graphs for the ResNeXt50(32x4d) training model. Comparing both graphs, we see that the validation loss in Fig. 16 (without batch normalization) is 3.67, while in Fig. 17 (with batch normalization, it is 3.25. Thus, it can be concluded that training with the ResNeXt50(32x4d) model performed better when the batch normalization process was added to the classification architecture.

Looking at the training process results using CNN, namely the VGG16 and ResNeXt50 (32x4d) models, the best training results were obtained with the ResNeXt50(32x4d) model with a modified classification layer, which had an accuracy value of 95.9% and a loss value of 3.25. In comparison, the learning process outcomes for the VGG16 model with a modified classification layer were an accuracy value of 87.5% and a loss value of 2.99.

After training the dataset, the model was tested. Testing was done by entering the results from the training process and the test data. The results from the testing process are the MAE and $R2$ values. The purpose of knowing the MAE value is to find out how effective the model is in predicting the dataset to solve the problem, where the lower the MAE value, the better its performance. If a good model is generated from the training results, then the MSE value can be used to find wrong estimates or errors that occur in the model.

TABLE III
MAE, MSE, AND R SQUARED VALUES FOR TRAINING WITHOUT BATCH NORMALIZATION

Predict Value	**VGG16**	**ResNeXt50(32x4d)**
MAE	5.31	5.59
R^2	0.55	0.54

TABLE IV
MAE, MSE, AND R SQUARED VALUES FOR TRAINING WITH BATCH NORMALIZATION

Predict Value	**VGG16**	**ResNeXt50(32x4d)**
MAE	5.19	4.75
R^2	0.56	0.63

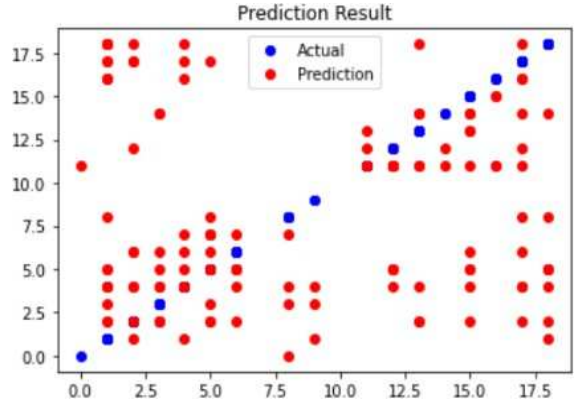

Fig. 18 Predictive graphs of the training model using ResNeXt50(32x4d) without back normalization

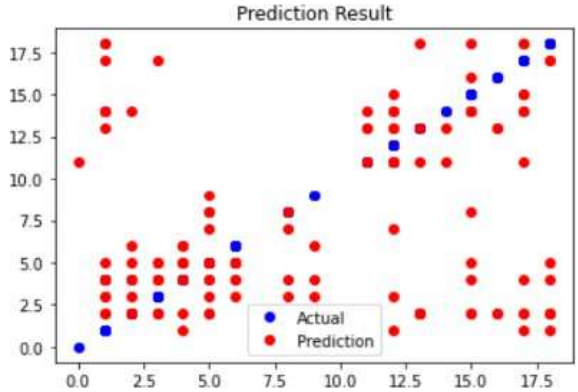

Fig. 19 Prediction graph for the training model using ResNeXt50(32x4d) with batch normalization

The squared value is used to determine the suitability ratio between bone age (BA) and chronological age (CA). Table 3 presents the predicted values for 200 epochs without batch normalization, and Table 4 presents the predicted values for

200 epochs with batch normalization. Fig. 18 shows the prediction result using the ResNeXt50(23x4d) model without batch normalization. Fig. 19 shows the prediction graph for training with batch normalization added to the classification layer.

TABLE V
COMPARISON OF PERFORMANCE BETWEEN THE PROPOSED METHOD AND PREVIOUS WORK

Model	Performance
[24]	MAD: 0.50 years (for age classification) RMS : 0.63 years (for age classification)
[25]	For age classification: MAE (male) : 0.43 years MAE (female): 0.53 year
[20]	Accuracy: 36% (age classification)
[9]	Accuracy: 79,6% (for gender) MAD : 0.50 years (for age classification) RMS : 0.67 years (for age classification)
Proposed method	MAE : 4.75 years (for age classification) $R2$: 0.63 (for age classification)

Comparing the prediction graphs for ResNext50(32x4d) in Fig. 18 and 19, we can see that the best results were achieved with modified classification because the red dots indicating the predicted data are closer to the blue dots, which are the actual data. From the experimental results in Tables 3 and 4, it can be concluded that the highest training accuracy was achieved using the ResNeXt50(32x4d) model with batch normalization, which had an accuracy of 95.9%. The errors made by neurons can be seen from the loss value. The lowest loss value in this study among the four methods was 2.99, using the VGG16 module with batch normalization, as can be seen from Table 2. To evaluate the prediction results, we used the MAE and $R2$ values. The lowest MAE value was achieved by using the ResNeXt50(32x4d) model with batch normalization at 4.75. The best $R2$ value (closest to 1) was achieved by using the ResNeXt50(32x4d) model with batch normalization at 0.63.

Table 5 shows the proposed performance systems and other previous work. As shown in Table 5, each research conducted has different performance parameters, namely the average MAE value, the proposed method has a better MAE value that can detect bone age in all genders, therefore the proposed method can outperform another method in the MAE value.

IV. CONCLUSION

This paper proposed a novel method for the estimation of bone age using a convolution neural network and a residual training model with a classification architecture modified by adding batch normalization. Most previous studies related to bone age estimation were carried out using a VGG16 model. Recently, a residual model has been introduced, i.e., ResNeXt50(32x4d). Techniques have been developed to optimize the training process because modern models have a large computational load and use a large amount of memory. Optimization can be done using batch normalization in order to stabilize and speed up the data processing and the training process, thereby increasing accuracy. After running experiments with two CNN models, namely VGG16 and ResNeXt50(32x4d), satisfactory MAE and R^2 values were produced, namely, 4.75 and 0.63, respectively, where in a previous study related to carpal bone age estimation, the MAD (MAE) value was 0.50 [9]. The experiments conducted in this study showed that the ResNeXt50(32x4d) model, modified by adding batch normalization to the classification architecture, could reduce the MAE value considerably and produce an R^2 value close to one. For further work, it is hoped that it will be possible to develop a carpal bone age estimation using the latest method to increase the value of training accuracy and prediction.

ACKNOWLEDGMENT

The authors thank Politeknik Elektronika Negeri Surabaya for supporting this research by providing equipment for research, as well as thanks to the pediatric hand radiographs dataset (RSNA 2017) and dr. Soetomo Surabaya Hospital for providing datasets for research.

REFERENCES

[1] I. Alfredo, "Bab 1 pendahuluan," Pelayanan Kesehat., no. 2014, pp. 1–6, 2010, [Online]. Available: http://library.oum.edu.my/repository/725/2/Chapter_1.pdf

[2] C. Spampinato, S. Palazzo, D. Giordano, M. Aldinucci, and R. Leonardi, "Deep learning for automated skeletal bone age assessment in X-ray images," Med. Image Anal., vol. 36, pp. 41–51, 2017, doi: 10.1016/j.media.2016.10.010.

[3] V. Gilsanz and O. Ratib, Hand Bone Age Bone Development. 2012. [Online]. Available: http://link.springer.com/10.1007/978-3-642-23762-1

[4] D. S. D. C, "Skeletal Bone Age Analysis Using Emroi Technique," IOSR J. Comput. Eng., vol. 12, no. 3, pp. 06–13, 2013, doi: 10.9790/0661-1230613.

[5] E. Reynolds, "Radiographic atlas of skeletal development of the hand and wrist. By W. W. Greulich and S. I. Pyle. Stanford University Press, 1950, xiii + 190 pp., ($10.00)," Am. J. Phys. Anthropol., vol. 8, no. 4, pp. 518–520, 1950, doi: 10.1002/ajpa.1330080429.

[6] G. Beunen, J. Lefevre, M. Ostyn, R. Renson, J. Simons, and D. Van Gerven, "Skeletal maturity in Belgian youths assessed by the Tanner-Whitehouse method (TW2)," Ann. Hum. Biol., vol. 17, no. 5, pp. 355–376, 1990, doi: 10.1080/03014469000001142.

[7] A. A. Aung and Z. M. Win, "Computer Assisted Bone Age Estimation of Children Using Middle Finger and Carpal Bones," Int. J. Intell. Eng. Syst., vol. 14, no. 3, pp. 119–127, 2021, doi: 10.22266/ijies2021.0630.11.

[8] D. Giordano, C. Spampinato, G. Scarciofalo, and R. Leonardi, "An automatic system for skeletal bone age measurement by robust processing of carpal and epiphysial/metaphysial bones," IEEE Trans. Instrum. Meas., vol. 59, no. 10, pp. 2539–2553, 2010, doi: 10.1109/TIM.2010.2058210.

[9] M. Marouf, R. Siddiqi, F. Bashir, and B. Vohra, "Automated Hand X-Ray Based Gender Classification and Bone Age Assessment Using Convolutional Neural Network," 2020 3rd Int. Conf. Comput. Math. Eng. Technol. Idea to Innov. Build. Knowl. Econ. iCoMET 2020, 2020, doi: 10.1109/iCoMET48670.2020.9073878.

[10] T. Van Steenkiste et al., "Automated Assessment of Bone Age Using Deep Learning and Gaussian Process Regression," pp. 674–677, 2018.

[11] S. R. Bulo, L. Porzi, and P. Kontschieder, "In-place Activated BatchNorm for Memory-Optimized Training of DNNs," Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit., pp. 5639–5647, 2018, doi: 10.1109/CVPR.2018.00591.

[12] F. Cavallo, A. Mohn, F. Chiarelli, and C. Giannini, "Evaluation of Bone Age in Children: A Mini-Review," Front. Pediatr., vol. 9, no. March, pp. 5–8, 2021, doi: 10.3389/fped.2021.580314.

[13] K. Al-Khater et al., "Time of appearance of ossification centers in carpal bones. A radiological retrospective study on Saudi children," Saudi Med. J., vol. 41, pp. 938–946, 2020, doi: 10.15537/smj.2020.9.25348.

[14] A. M. Richard Drake, A. Wayne Vogl, Gray's Anatomy for Students. Elsevier, 2019. [Online]. Available: https://www.elsevier.com/books/grays-anatomy-for-students/drake/978-0-323-39304-1

[15] V. De Sanctis, S. Di Maio, A. Soliman, G. Raiola, R. Elalaily, and G. Millimaggi, "Hand X-ray in pediatric endocrinology: Skeletal age assessment and beyond," Indian J. Endocrinol. Metab., vol. 18, no. November, pp. S63–S71, 2014, doi: 10.4103/2230-8210.145076.

[16] S. J. Son et al., "TW3-Based Fully Automated Bone Age Assessment System Using Deep Neural Networks," IEEE Access, vol. 7, no. c, pp. 33346–33358, 2019, doi: 10.1109/ACCESS.2019.2903131.

[17] D. Štern, C. Payer, N. Giuliani, and M. Urschler, "Automatic Age Estimation and Majority Age Classification from Multi-Factorial MRI Data," IEEE J. Biomed. Heal. Informatics, vol. 23, no. 4, pp. 1392–1403, 2019, doi: 10.1109/JBHI.2018.2869606.

[18] E. Chai, M. Pilanci, and B. Murmann, "Separating the Effects of Batch Normalization on CNN Training Speed and Stability Using Classical Adaptive Filter Theory," in 2020 54th Asilomar Conference on Signals, Systems, and Computers, 2020, pp. 1214–1221. doi: 10.1109/IEEECONF51394.2020.9443275.

[19] S. Xie, R. Girshick, P. Dollár, Z. Tu, and K. He, "Aggregated residual transformations for deep neural networks," Proc. - 30th IEEE Conf. Comput. Vis. Pattern Recognition, CVPR 2017, vol. 2017-Janua, pp. 5987–5995, 2017, doi: 10.1109/CVPR.2017.634.

[20] S. S. Halabi et al., "The RSNA Pediatric Bone Age Machine Learning Challenge.," Radiology, vol. 290, no. 2, pp. 498–503, Feb. 2019, doi: 10.1148/radiol.2018180736.

[21] J. Zhou, Z. Li, W. Zhi, B. Liang, D. Moses, and L. Dawes, "Using Convolutional Neural Networks and Transfer Learning for Bone Age Classification," pp. 0–5, 2017.

[22] S. Nadeemhashmi, H. Gupta, D. Mittal, K. Kumar, A. Nanda, and S. Gupta, "A Lip Reading Model Using CNN with Batch Normalization," 2018 11th Int. Conf. Contemp. Comput. IC3 2018, pp. 2–4, 2018, doi: 10.1109/IC3.2018.8530509.

[23] I. Fibriani, Widjonarko, A. Prasetyo, A. M. Raharjo, and D. E. Irawan, "Multi Deep Learning to Diagnose COVID-19 in Lung X-Ray Images with Majority Vote Technique," Int. J. Intell. Eng. Syst., vol. 13, no. 6, pp. 560–568, 2020, doi: 10.22266/ijies2020.1231.49.

[24] P. H. Radiographs, D. B. Larson, M. C. Chen, M. P. Lungren, N. V Stence, and C. P. Langlotz, "Performance of a Deep-Learning Neural Network Model in Assessing Skeletal Maturity on," vol. 287, no. 1, pp. 1–10, 2018.

[25] S. A. Adeshina, C. Lindner, and T. F. Cootes, "Automatic segmentation of carpal area bones with random forest regression voting for estimating skeletal maturity in infants," in 2014 11th International Conference on Electronics, Computer and Computation (ICECCO), 2014, pp. 1–4. doi: 10.1109/ICECCO.2014.6997559.