International Journal of Electrical and Computer Engineering (IJECE)
Vol. 13, No. 2, April 2023, pp. 2078~2085
ISSN: 2088-8708, DOI: 10.11591/ijece.v13i2.pp2078-2085

Bone age assessment based on deep learning architecture

Alaa Jamal Jabbar, Ashwan A. Abdulmunem
Department of Computer Science, Computer Science and Information Technology, University of Kerbala, Kerbala, Iraq

Article Info

Article history:

Received May 26, 2022
Revised Sep 22, 2022
Accepted Oct 16, 2022

Keywords:

Bone age assessment
Classification
Convolutional neuron network
Deep learning
X-ray images

ABSTRACT

The fast advancement of technology has prompted the creation of automated systems in a variety of sectors, including medicine. One application is an automated bone age evaluation from left-hand X-ray pictures, which assists radiologists and pediatricians in making decisions about the growth status of youngsters. However, one of the most difficult aspects of establishing an automated system is selecting the best approach for producing effective and dependable predictions, especially when working with large amounts of data. As part of this work, we investigate the use of the convolutional neural networks (CNNs) model to classify the age of the bone. The work's dataset is based on the radiological society of North America (RSNA) dataset. To address this issue, we developed and tested deep learning architecture for autonomous bone assessment, we design a new deep convolution network (DCNN) model. The assessment measures that use in this work are accuracy, recall, precision, and F-score. The proposed model achieves 97% test accuracy for bone age classification.

Corresponding Author:

Alaa Jamal Jabbar
Department of Computer Science, College of Computer Science and Information Technology, University of Kerbala
Kerbala, Iraq
Email: alaa.jamal@s.uokerbala.edu.iq

1. INTRODUCTION

The identification of bone aging is an important problem in the medical industry. Determining bone age is a standardized process in which doctors scan a child's hands to determine a child's skeletal maturity [1]. Due to the nature of bone growth, the test is only accurate between the ages of 0 and 19 years old [2], [3]. It is often used as an indicator of developmental problems in children compared to chronological age. It can also be used to determine the age when a birth certificate cannot be obtained [4]. Due to the discriminatory stage of ossification in the non-dominant hand, it is normally done by a radiological examination of the left hand., Following that, a comparison with chronological age is made: a disparity between the two figures is found to indicate abnormality [5], [6]. Left-hand radiograph analysis is widely used to assess bone maturity because of its ease of use, low radiation exposure, and availability of different ossification centers [7]. Due to the task's similarities to deep learning's normal object recognition and classification problems, bone age assessments have grown to be a prominent focus of the machine learning community [8].

Bone age assessment (BAA) can be performed according to Greulich and Pyle (GP) or according to Tanner-Whitehouse (TW2) [9]. Advances in machine learning, image processing, statistical learning, and many other domains have given rise to breakthrough technologies with new and novel solutions [10]. The machine learning community has paid close attention to medical imaging in particular, resulting in new approaches to old problems [11]. The emergence and spread of convolutional neural networks (CNNs), a deep learning technology, has recently occurred and has attracted interest in medical imaging analysis. Many of these restrictions are addressed by deep-learning techniques, which enable an algorithm to autonomously acquire the properties without any direct human participation throughout the training phase, picture interpretation is critical [12]. Early research has shown promise in a variety of medical imaging applications, including lung nodules, interstitial lung disease, breast cancer, cerebral microbleeds, brain malignancies, and spinal metastases [13]. Traditional machine learning and image processing approaches were used in automated BAA procedures. Approaches based on CNNs have been only recently introduced to BAA [14]. Instead of extracting information from specific locations based on clinical expertise, these approaches frequently encode visual features directly [15]. Typically, a bone age category reflects the bone age of hand x-ray in these approaches. The label is usually for certain years. Nonetheless, bone age markers that are discrete, on the other hand, cannot adequately capture the complex and continual growth of bone. It invariably results in a semantic mismatch between the real scenario and the labels, limiting CNN's ability to learn better [16].

There are two types of recent popular CNNs techniques for BAA; methods of regression and categorization. These procedures almost often make use of specific bone age designations. CNNs produce continuous results as a result of regression algorithms, which are then utilized to anticipate bone age [17]. Lee *et al.* [18] in this work the range of bone ages from 5 to 18 years were discovered and developed to isolate a BAA-interested zone, a fully automated deep learning approach was used. Larson *et al.* [19] resnet-50 was used to measure bone age, with the classifier's output being a probability distribution for bone ages ranging from 0 to 19 years in 1-month increments.

Wu *et al.* [20] developed an integrated network for hand classification and determining bone maturity at the same time. The ages of the bones were determined using their classification model the residual attention. Souza and Oliveira [21] provided residual learning as a method and inter the sex as input in the full section of their neural network. Wagner [22] to determine bone ages, researchers utilized both classification and regression approaches. Both approaches classified bone age by month (1 to 228). The discrete labels may not be a major issue in classification algorithms since these methods regard bone age labels as separate groups. However, the issue arises when deciding on a crucial portion between bone age groups. Furthermore, since classification algorithms treat each class individually as a distinct entity with no connection to the two bone age groups, they are unsuitable for a long-term problem like BAA. Also, label distribution optimization is a significant factor connected project. For apparent age estimation, CNNs with distribution-based loss functions were presented, which employed distributions as well as the training assignments to leverage the uncertainty given by hand class [23]. This work approach is based on a multiclass classification problem.

Researchers categorized the bone age by months (0 to 228) in this work, albeit we did not explicitly employ these discrete bone age classifications. Rather, create bone age ranges and replace original labels with the classification component of these ranges. Then, at the same time, a CNN is taught to produce five distinct bone age ranges. Finally, bone age is determined based on the five age range outputs. The suggested technique offers two key benefits over traditional regression and classification methods. i) it can bridge the semantic gap by indicating not just a precise bone age, as typical bone age labels do, but also the continuity of bone growth; and ii) it is resistant to incorrect labeling. Mistakes are inherent when radiologists manually identify radiographs for bone age, although these errors always fall within a certain range. As a result, it is recommended to use many bones age ranges rather than a single bone age designation.

In this section, we described the introduction to bone age assessment, as well as the numerous methods used to identify age based on X-ray images, and it was said that the survey on this issue, as well as several approaches and performance factors, had been explored. Section 2 should elaborate on the new approach using a flow chart, and section 3 should discuss the outcomes and compare them to existing procedures. According to the BAA, section 4 specified the conclusion and future scope.

2. METHOD

CNNs are more widely utilized in classification jobs. An evaluation of the bone age technique based on hand radiograph images is proposed in this section. Using hand X-ray scans, the proposed system tries to establish a person's age group proposed method's three basic operations are image preprocessing, extraction of features, and classification. The hand X-ray pictures are grayscale color modeled and scaled to a certain size during the image preprocessing procedure. A data augmentation procedure is also used to increase the dataset's size. The concept of the design CNN model was used to execute this work in the feature extraction and classification stage.

2.1. Dataset description

The dataset used in this investigation was collected from the radiological society of North America (RSNA) pediatric bone age machine learning challenge, and it was uploaded from Kaggle. This dataset is nine GB in volume and contains 12,611 x-ray photos with a comma-separated values (CSV) file containing id, bone age, and gender (ages from 1 to 228 months for males and females) see in Figure 1. The gender distribution was 5,778 for females and 6,833 for males. Experts carefully designated the bone age in years for each photograph. Some of these pictures were discovered to be warped and misshapen, making them less suitable for training the models. As a result, these data must be deleted before being fed into the training model. Along with eliminating the distorted pictures, the dataset must be standardized in order to retain the balance between radiograph classes. So, it is broken into 8,829 training images, 1,891 testing, and 1,891 validation images. This is divided based on the 70:15:15 training-testing split principle. Because of the varying number of data for ages from one to 18 years, where the ages less than 5 years were very few, as well as the ages from 15 years and above, so they divided the data into 5 groups to make a balance between the number of samples in each category in the proposed bone age evaluation method. Table 1 shows the specifics of the various age groups as well as the numbers of photos featured in each group after convert's to years.

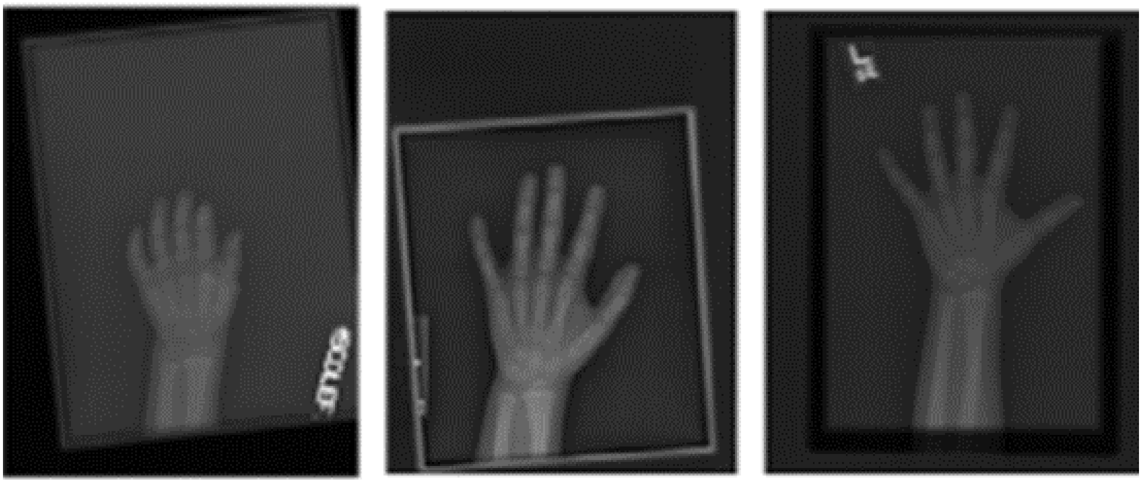


Figure 1. A sample radiograph images from the RSNA dataset

Table 1. The Information of the RSNA dataset

Classes	Age group (year)	Number of images
A	[1 to 5]	1083
B	[6 to 8]	2179
C	[9 to 11]	3322
D	[12 to 14]	4832
E	>=15	1201

2.2. Proposed CNN model

Automatic bone evaluation is essentially a classification problem in which the models are meant to predict the age values of left-hand X-ray pictures. Our primary goal is a development and tests several techniques for automated bone assessment: approaches to deep learning. Deep learning can automatically extract visual attributes from raw images. In this experiment, we utilize the design new deep learning-based convolution neural network. Figure 2 shows the overall step process employed in this study. The following are the specifics for each step:

2.2.1. Pre-processing

As mentioned in subsection 2.1, the dataset comprises thousands of photos of varying sizes. We use pre-processing to standardize picture sizes. We resize all photos in the dataset to 224×224 pixels for the deep learning-based CNN technique. To preserve the aspect ratio of the original images, horizontal or vertical padding has been added to all images. Furthermore, despite the fact that the original photos were greyscale, the input images had to be rendered as colored images owing to design architectural assumptions.

Data augmentation is also employed during several training cycles in an effort to reduce overfitting specification and increase generalization. With varying degrees of effectiveness, 15° rotation, vertical and horizontal flipping, height and width adjustments were all used. Overall, data augmentation was ineffective in increasing accuracy but reducing overfitting. After data augmentation, we applied groping for it.

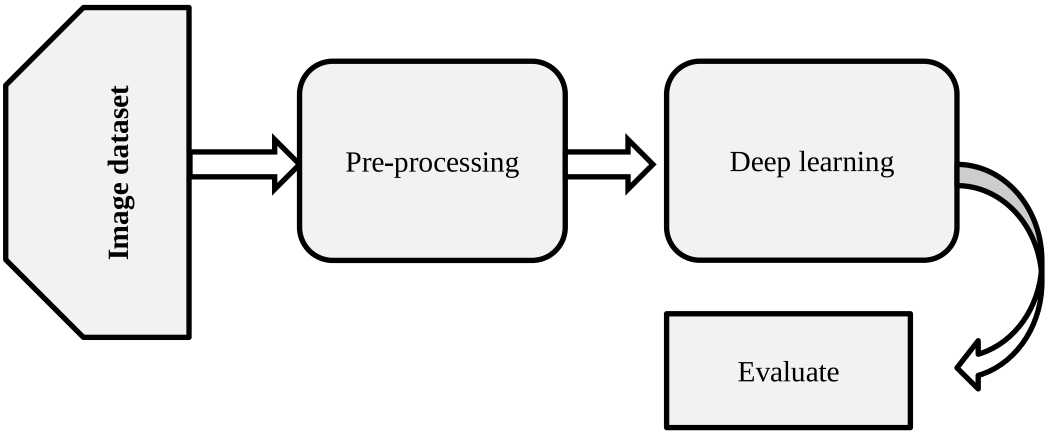


Figure 2. General steps for the proposed method

2.2.2. **Architecture of the proposed deep learning**

Deep learning has grown in popularity in recent years due to the ability to learn features automatically. convolutional neural networks (CNNs) typically have three-layer types: convolutional layer, pooling layer, and fully connected layer see in Figure 3 [24]. The convolutional layer is responsible for computing the weighted sum, including a bias value into the weighted sum, and then using a function of activation known as the rectifier linear unit (ReLu) to the addition result., which is described using (1). The goal of pooling layers, on the other hand, is to prevent over-fitting by lowering the number of convolutional layer features acquired. Finally, utilizing the final layer, the entire connected layers attempt to collect all of the descriptor features to be classified [25].

$$ReLu(x) = max(0, x) \tag{1}$$

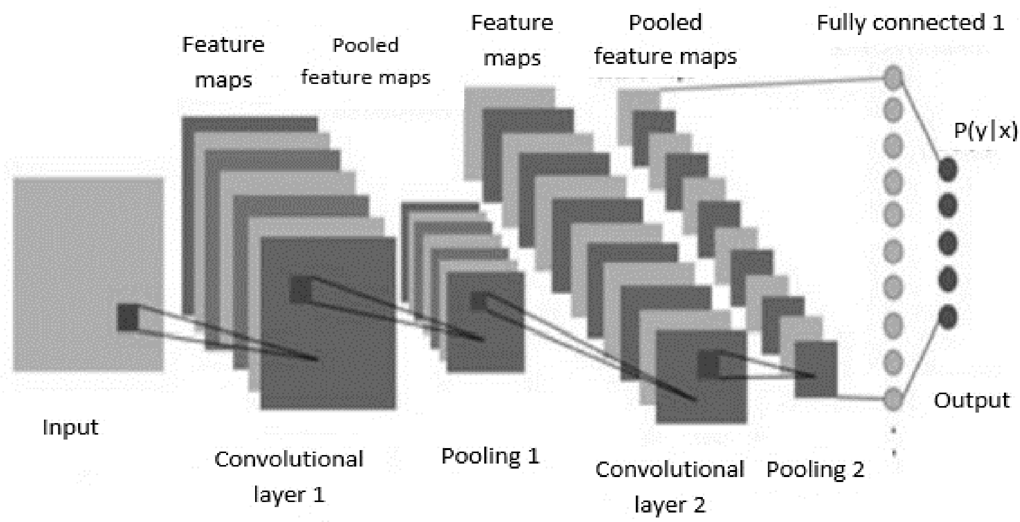


Figure 3. Convolutional network architecture (CNN) in general [24]

In this work, we evaluate the performance of several designs and compare the results. This includes a design model comprised of 10 learnable weighted layers: 7 convolutional blocks and 3 fully-connected layers. Every convolution employs a 3x3 kernel with a stride of 2 and a padding of 1, with stride 2 and no padding, a 2×2 max-pooling is done. Batch normalization is used after each convolutional layer and the first with second block use 64 filter, third and four block use 128 filters, five and six block use 256 filters and seven block use 512 filter also dropout layer used after each Dense layer see in Figure 4. The activation function is ReLu and replace with SoftMax in the output layer. The summary model as show in Table 2.

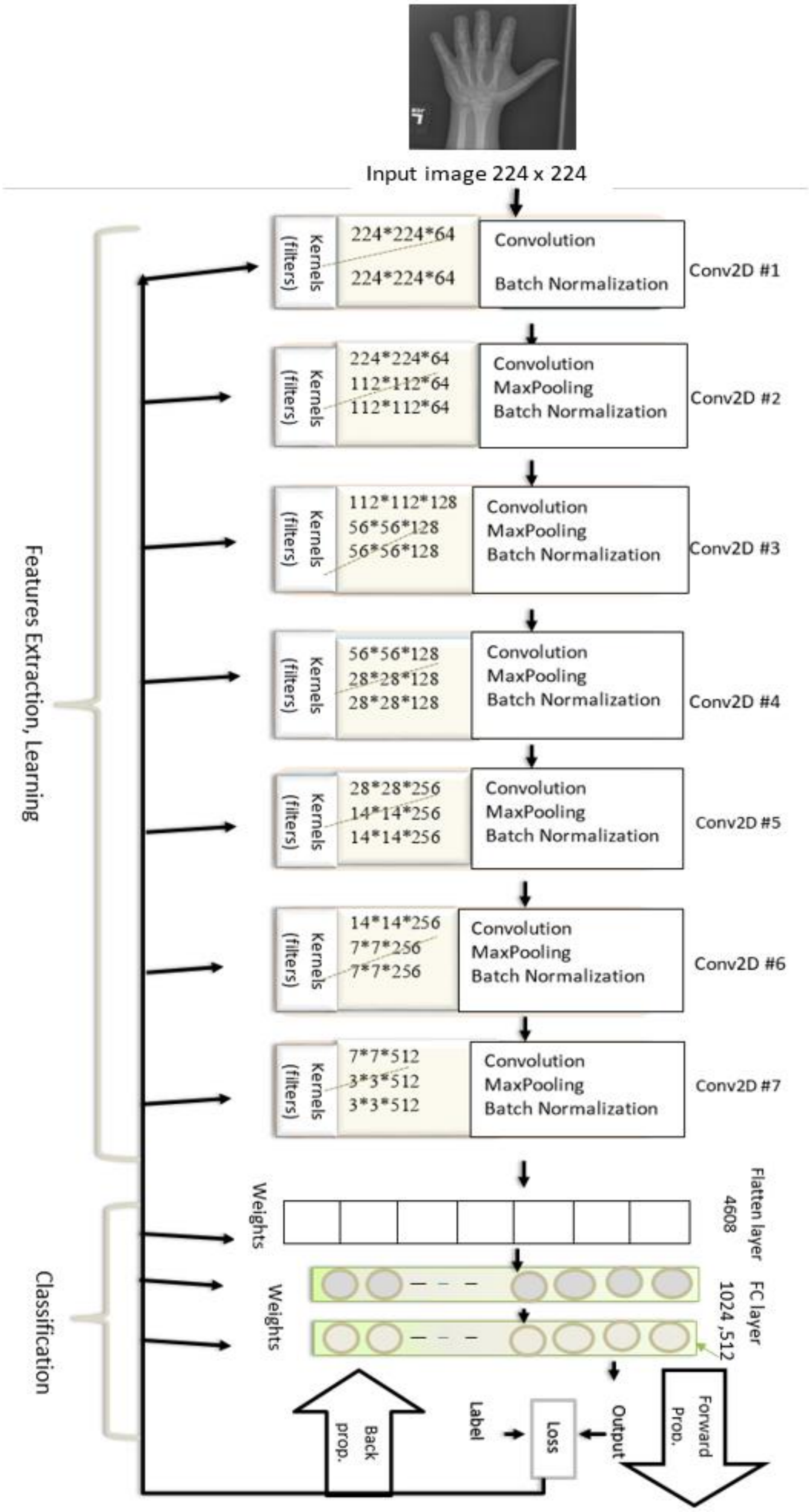


Figure 4. The architecture of the proposed CNN model with the training process

Table 2. Summary of the proposed model

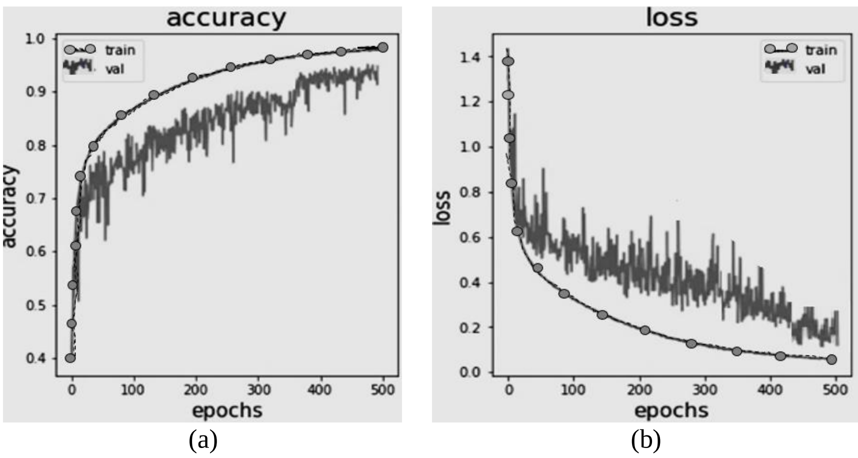
Layer (Type)	Output shape	Parameters
conv2d (Conv2D)	(None, 224, 224, 64)	640
Batch_ normalization	(None, 224, 224, 64)	256
conv2d_1	(None, 224, 224, 64)	36928
max_pooling2d	(None, 112, 112, 64)	0
batch_normalization_1	(None, 112, 112, 64)	256
conv2d_2	(None, 112, 112, 128)	73856
max_pooling2d_1	(None, 56, 56, 128)	0
batch_normalization_2	(None, 56, 56, 128)	512
conv2d_3	(None, 56, 56, 128)	147584
max_pooling2d_2	(None, 28, 28, 128)	0
batch_normalization_3	(None, 28, 28, 128)	512
conv2d_4	(None, 28, 28, 256)	295168
max_pooling2d_3)	(None, 14, 14, 256)	0
batch_normalization_4	(None, 14, 14, 256)	1024
conv2d_5	(None, 14, 14, 256)	590080
max_pooling2d_4	(None, 7, 7, 256)	0
batch_normalization_5	(None, 7, 7, 256)	1024
conv2d_6	(None, 7, 7, 512)	1180160
max_pooling2d_5	(None, 3, 3, 512)	0
batch_normalization_6	(None, 3, 3, 512)	2048
flatten	(None, 4608)	0
dense	(None, 1024)	4719616
dropout	(None, 1024)	0
dense_1	(None, 512)	524800
dropout_1	(None, 512)	0
dense_2	(None, 5)	2565

Total parameters: 7,577,029; Trainable parameters: 7,574,213; Non-trainable parameters: 2,816

3. RESULTS AND DISCUSSION

For training, we use accuracy to evaluate our model. Furthermore, we use the Stochastic gradient descent (SGD) optimizer and an initial learning rate of 0.01 and implemented learning rate reduction when validation accuracy plateaued, for sequential architecture, we explored using batch sizes of 32 and 64, and we found the batch size of 32 allowed us to achieve better results. Additionally, we found that using random horizontal and vertical flips improved our overall performance. We trained our models for 500 epochs (about 40 hours for model architecture). Here is the plot for train/Val error vs. epoch for the model and accuracy in Figure 5. Figure 5(a) for accuracy, and Figure 5(b) for loss. We use an Anaconda and Jupyter environment with CPU Intel(R) 2.60 GHz, RAM 16 GB, and an NVIDIA GetForce GTX 1660Ti GPU. All the codes execute in Python 3.8 with Keras and Tensorflow used in the experiment.

After tuning hyperparameters, we found the accuracy of the model architecture is chivied 98% with a 0.05 loss function for the training model and 97.22% with 0.09 val loss for the testing model after make data augmentation. Thus, for our final model, we selected the model that includes the following hyper-parameter in Table 3. The obtained result of the confusion matrix for test evaluation is shown in Figure 6 and the classification report of the proposed system using the design new deep convolution network is shown in Table 4.



(a) (b)

Figure 5. The plot for train/val error vs. epoch for the proposed model: (a) model train accuracy and validation accuracy, (b) train loss and validation loss

Table 3. Hyperparameters for the proposed model

Parameters	Hyperparameters
Optimizer	SGD
Learning rate	0.01
Batch size	32
Epochs	500

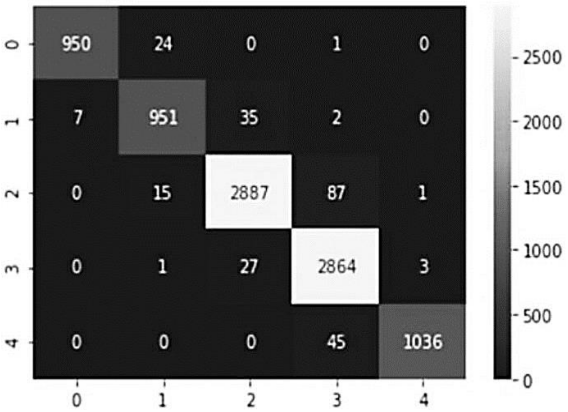


Figure 6. The *confusion_matrix* for proposed CNN

Table 4. The classification report of the proposed DCNN and metrics values

Performance Metrics (%)	Age Classes				
	E= 4	D=3	C=2	B =1	A = 0
Accuracy	95.8%	98.9%	96.5%	95.5%	97%
precision	100%	95%	98%	96%	99%
Recall	96%	99%	97%	96%	97%
F-Measure	98%	97%	97%	96%	98%

4. CONCLUSION

In this part, we provided a computerized approach for BAA based on x-ray scans of human wrist bones. The approach was created employing cutting-edge deep learning techniques. The suggested BAA algorithm improves radiologists' ability to diagnose bone age by reducing human observation variances, resulting in cost savings in hospitals or clinics. Furthermore, the provided solution speeds up the BAA process compared to traditional methods. The technique was assessed using model accuracy, recall, precision, and F-score metrics derived from the weights of design models are 97.22%, 97%, 97.2%, and 97.26%. Among the suggested deep learning models, the model architecture includes two stages: preprocessing stage and design new model stage this achieves a more accurate result as an automated technique for detecting bone age, the accuracy of training is 98% and the test is 97%. Data augmentation is used to reduce the overfitting and also convert original x-ray images into a standardized form to improve the training procedure for the classification network.

ACKNOWLEDGEMENTS

This work support by Kerbala University. Authors thanks the University of Kerbala. Iraq.

REFERENCES

[1] A. Wibisono *et al.*, "Deep learning and classic machine learning approach for automatic bone age assessment," in *2019 4th Asia-Pacific Conference on Intelligent Robot Systems (ACIRS)*, Jul. 2019, pp. 235–240, doi: 10.1109/ACIRS.2019.8935965.

[2] B. Liu, Y. Zhang, M. Chu, X. Bai, and F. Zhou, "Bone age assessment based on rank-monotonicity enhanced ranking CNN," *IEEE Access*, vol. 7, pp. 120976–120983, 2019, doi: 10.1109/ACCESS.2019.2937341.

[3] P. Hao, S. Chokuwa, X. Xie, F. Wu, J. Wu, and C. Bai, "Skeletal bone age assessments for young children based on regression convolutional neural networks," *Mathematical Biosciences and Engineering*, vol. 16, no. 6, pp. 6454–6466, 2019, doi: 10.3934/mbe.2019323.

[4] M. C. Chen, "Automated bone age classification with deep neural networks," *Stanford Univ. USA, Technical Report*. pp. 1–7, 2016.

[5] J. E. Pinsker, M. S. L. Ching, D. P. Chiles, M. Lustik, C. B. Mahnke, and V. J. Rooks, “Automated bone age analysis with lossy image files,” *Military Medicine*, vol. 182, no. 9, pp. e1769--e1772, Sep. 2017, doi: 10.7205/MILMED-D-17-00011.
[6] C. Liu, H. Xie, Y. Liu, Z. Zha, F. Lin, and Y. Zhang, “Extract bone parts without human prior: End-to-end convolutional neural network for pediatric bone age assessment,” in *Medical Image Computing and Computer Assisted Intervention – MICCAI 201*, 2019, pp. 667–675.
[7] C. Spampinato, S. Palazzo, D. Giordano, M. Aldinucci, and R. Leonardi, “Deep learning for automated skeletal bone age assessment in X-ray images,” *Medical Image Analysis*, vol. 36, pp. 41–51, Feb. 2017, doi: 10.1016/j.media.2016.10.010.
[8] B.-D. Lee and M. S. Lee, “Automated bone age assessment using artificial intelligence: The future of bone age assessment,” *Korean Journal of Radiology*, vol. 22, no. 5, 2021, doi: 10.3348/kjr.2020.0941.
[9] M. W. Nadeem, H. G. Goh, A. Ali, M. Hussain, M. A. Khan, and V. Ponnusamy, “Bone age assessment empowered with deep learning: A survey, open research challenges and future directions,” *Diagnostics*, vol. 10, no. 10, Oct. 2020, doi: 10.3390/diagnostics10100781.
[10] J. C. Castillo, Y. Tong, J. Zhao, and F. Zhu, “RSNA bone-age detection using transfer learning and attention mapping.” *ECE228 and SIO209 Machine learning for physical applications*, Spring 2018, pp. 1–5, 2018.
[11] B. J. Erickson, P. Korfiatis, T. L. Kline, Z. Akkus, K. Philbrick, and A. D. Weston, “Deep learning in radiology: Does one size fit all?,” *Journal of the American College of Radiology*, vol. 15, no. 3, pp. 521–526, Mar. 2018, doi: 10.1016/j.jacr.2017.12.027.
[12] B.-U. Bae, W. Bae, and K.-H. Jung, “Improved deep learning model for bone age assessment using triplet ranking loss,” *MIDL 2018*, pp. 2–4, 2018.
[13] S. Mutasa, P. D. Chang, C. Ruzal-Shapiro, and R. Ayyala, “MABAL: A novel deep-learning architecture for machine-assisted bone age labeling,” *Journal of Digital Imaging*, vol. 31, no. 4, pp. 513–519, Aug. 2018, doi: 10.1007/s10278-018-0053-3.
[14] B. S. Reddy, D. Sreenivasarao, and S. K. Saheb, “An automated system for identification of skeletal maturity using convolutional neural networks based mechanism,” *International Journal of Innovative Technology and Exploring Engineering*, vol. 8, no. 11, pp. 2221–2227, Sep. 2019, doi: 10.35940/ijitee.K2049.0981119.
[15] M. Prokop-Piotrkowska, K. Marszałek-Dziuba, E. Moszczyńska, M. Szalecki, and E. Jurkiewicz, “Traditional and new methods of bone age assessment-an overview,” *Journal of Clinical Research in Pediatric Endocrinology*, vol. 13, no. 3, pp. 251–262, Aug. 2021, doi: 10.4274/jcrpe.galenos.2020.2020.0091.
[16] T. Langner, “Deep regression and segmentation for medical inference,” Diss. Acta Universitatis Upsaliensis, 2021.
[17] A. Z. Alsinan, V. M. Patel, and I. Hacihaliloglu, “Automatic segmentation of bone surfaces from ultrasound using a filter-layer-guided CNN,” *International Journal of Computer Assisted Radiology and Surgery*, vol. 14, no. 5, pp. 775–783, May 2019, doi: 10.1007/s11548-019-01934-0.
[18] H. Lee *et al.*, “Fully automated deep learning system for bone age assessment,” *Journal of Digital Imaging*, vol. 30, no. 4, pp. 427–441, Aug. 2017, doi: 10.1007/s10278-017-9955-8.
[19] D. B. Larson, M. C. Chen, M. P. Lungren, S. S. Halabi, N. V Stence, and C. P. Langlotz, “Performance of a deep-learning neural network model in assessing skeletal maturity on pediatric hand radiographs,” *Radiology*, vol. 287, no. 1, pp. 313–322, Apr. 2018, doi: 10.1148/radiol.2017170236.
[20] E. Wu *et al.*, “Residual attention based network for hand bone age assessment,” in *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*, Apr. 2019, pp. 1158–1161, doi: 10.1109/ISBI.2019.8759332.
[21] D. Souza and M. M. Oliveira, “End-to-end bone age assessment with residual learning,” in *2018 31st SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI)*, Oct. 2018, pp. 197–203, doi: 10.1109/SIBGRAPI.2018.00032.
[22] D. Wagner, *Lecture notes in computer science: Preface*. Ljubljana, Slovenia: Springer Berlin Heidelberg, 2008.
[23] W. Ketwongsa, S. Boonlue, and U. Kokaew, “A new deep learning model for the classification of poisonous and edible mushrooms based on Improved AlexNet convolutional neural network,” *Applied Sciences*, vol. 12, no. 7, Mar. 2022, doi: 10.3390/app12073409.
[24] J. Chen, B. Yu, B. Lei, R. Feng, D. Z. Chen, and J. Wu, “Doctor imitator: A graph-based bone age assessment framework using hand radiographs,” in *Medical Image Computing and Computer Assisted Intervention – MICCAI 2020*, 2020, pp. 764–774.
[25] K. Li *et al.*, “Automatic bone age assessment of adolescents based on weakly-supervised deep convolutional neural networks,” *IEEE Access*, vol. 9, pp. 120078–120087, 2021, doi: 10.1109/ACCESS.2021.3108219.

BIOGRAPHIES OF AUTHORS

Alaa Jamal Jabbar received her B.Sc. degree in computer science from Thiqar University, Iraq. Recently, she is a Master’s Student at College of Computer Science and Information Technology, University of Kerbala, Iraq. She can be contacted at email: alaa.jamal@s.uokerbala.edu.iq.

Ashwan A. Abdulmunem received her Ph.D. degree in computer vision, Artificial Intelligence from Cardiff University, UK. Recently, she is an Asst. Prof. at College of Computer Science and Information Technology, University of Kerbala. Her research interests include computer vision and graphics, computational imaging, pattern recognition, artificial intelligence, machine and deep learning. She can be contacted at email: ashwan.a@uokerbala.edu.iq.