OPEN ACCESS
Journal of Mechanics in Medicine and Biology
Vol. 22, No. 8 (2022) 2240020 (14 pages)

DOI: 10.1142/S0219519422400206

DEVELOPMENT OF CONVOLUTIONAL NEURAL NETWORK MODEL FOR CLASSIFICATION OF CARDIOMEGALY X-RAY IMAGES

MIN-JEONG KIM

Department of Biomedical Engineering
Kyungpook National University
Sangyeok-dong, Buk-gu, Daegu, South Korea
kmj5832@naver.com

JUNG-HUN KIM*

Bio-Medical Research Institute
Kyungpook National University Hospital
Samdeok-dong, Jung-gu, Daegu, South Korea
fainal2@naver.com

Received 14 September 2021
Accepted 9 June 2022
Published 16 September 2022

In this study, the effect of the input image format on the accuracy in deep learning (DL) was evaluated using tagged image file format (TIFF), PNG, and JPEG images converted from DICOM images. Based on the evaluation results, a convolutional neural network (CNN)-based DL model was proposed. The training and test data used in this study were chest X-ray images of patients diagnosed with normal heart or cardiomegaly conditions. CNN models, namely VGGNet, ResNet, InceptionNet (GoogleNet), DenseNet, and EfficientNet, were used to compare the results according to the types of input images, that is, images converted from DICOM images into TIFF, PNG, and JPEG images. The classification performance was validated by proposing a CNN model that can classify normal hearts and cardiomegaly in chest X-ray images. In this study, through medical imaging research using deep learning, it was demonstrated that the classification performance was unaffected even when the DICOM image was converted into any format and used as an input image. In addition, the proposed CNN model exhibited excellent performance in classifying normal hearts and cardiomegaly on X-ray images. This can be used in various studies that aim to apply DL to medical images by providing information according to the input image type and is also considered to aid in the selection of image type and learning parameters. The proposed model is expected to yield useful results for the classification of diseases in chest X-ray images.

*Corresponding author.

Keywords: Chest X-ray; cardiomegaly; convolutional neural network; classification; deep learning.

1. Introduction

Cardiomegaly refers to a condition in which the heart is more than 50% larger than the inner diameter of the thorax owing to the enlargement of the inner space of the heart, or ventricular hypertrophy.[1–4] Cardiomegaly is a clinically problematic disease that can cause cardiovascular system complications. It is critical to detect cardiomegaly early via electrocardiogram (ECG), simple chest radiograph (X-ray), echocardiography, etc., to prevent further complications.[5–8] However, because the competence level of doctors who interpret medical images differs, there is a risk associated with the differences in the results depending on the degree of fatigue at the time of examination, even if it is the same doctor interpreting the same image.[9,10]

Deep learning (DL) can be used to extract features from input data and perform hierarchical learning using a deep neural network, in which several layers of artificial neural networks that mimic the behavior of neurons in the human brain are used to achieve excellent performance.[11,12] Image recognition using DL rose to prominence in the field of computer vision after AlexNet, which was built with multiple layers of convolutional neural networks (CNNs), outperformed humans in the ImageNet large-scale visual recognition challenge (ILSVRC) in 2012.[13] In particular, CNNs, which exhibit excellent performance in application fields such as image recognition and classification, exhibit expert-level performance in the medical imaging field and are used for disease diagnosis and classification.[14–16]

In contrast to a typical image, a DICOM file expresses a large amount of information by storing unique information, such as patient and examination information, in addition to the image, allowing the file to be used effectively as a medical image. Most open-source datasets used in medical imaging research are in PNG or JPEG file format. Even when using DICOM files, most of them are converted into image file formats such as PNG or JPEG for research because of hardware limitations.[17–19] Recently, a tagged image file format (TIFF) image file format has been reported as being the most similar to a DICOM video file format.[20] A lossless compression method that preserves the original data is used when converting DICOM images into PNG, whereas a lossy compression method that loses some image data is used when converting DICOM images into JPEG.[21] It is necessary to determine whether using an image converted from a 16-bit DICOM file containing a large amount of information into an 8-bit PNG or JPEG file is optimal.

In addition, because it is difficult to collect medical images for learning because of the low accessibility of medical data and complicated procedures such as institutional review board (IRB) collection approval, the amount of training data is insufficient, in contrast to data used for classification or recognition of natural images. Hence, most CNN models used in medical imaging research are trained using general natural image data. In contrast to general natural images, medical images are composed of

grayscale rather than RGB values and represent the inside of the human body, therefore it is necessary to develop a model that can reflect image characteristics in special fields such as grayscale medical images.

In this study, we aimed to determine whether the 8-bit PNG and JPEG files and the 16-bit files used in most studies show similar classification results when used as input images. To this end, we evaluated the classification results using an image converted from a DICOM file, the standard of medical imaging, into a 16-bit TIFF file, which is reported as the most similar file format, and an image converted into an 8-bit PNG or JPEG file as an input image. Moreover, based on the classification results, we proposed a CNN model that can classify a normal heart and a cardiomegaly based on chest X-ray images. We attempted to classify normal hearts and cardiomegaly using the proposed CNN model and verified how close a classified image was to the correct answer.

2. Materials and Methods

2.1. *Development and implementation environment*

In this study, the Jupyter Notebook, an integrated development environment based on the Python (version 3.8.8) language, was used to implement DL modeling. The system used was an Intel(R) Core (TM) i7-8700k (Intel, Santa Clara, CA, USA), NVIDIA GeForce GTX 1080 Ti (NVIDIA, Santa Clara, CA, USA). Tensorflow 2.3.0 and Keras 2.4.3 were used for the DL modeling implementation.

2.2. *Data collection*

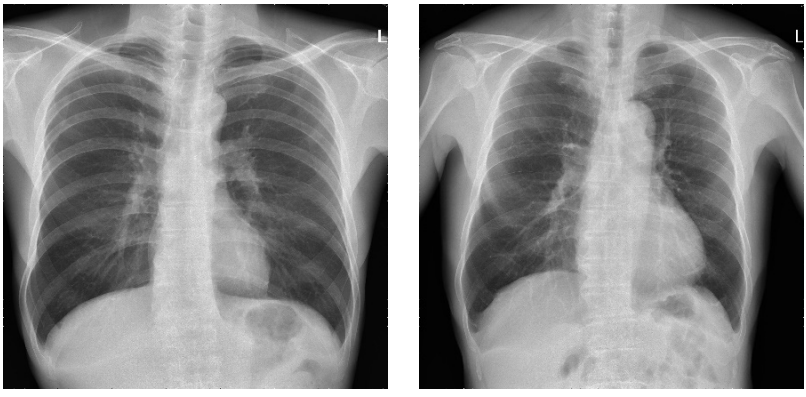
The data used in this study were approved according to the review procedure of IRB of Kyungpook National University Hospital (IRB No.: KNUH 2020-07-018-002). From January 2010 to December 2020, chest X-ray images were obtained from patients aged between 20 and 95 years who visited the Kyungpook National University Hospital. The collected image data consisted of 526 normal hearts and 500 cardiomegaly. For the composition of the data for classification, 80% of the total data were used as the training data, and the remaining 20% were used as the test dataset to validate the final model.

2.3. *Data annotation*

The data used in this study were obtained by a chest radiologist and categorized into normal hearts and cardiomegaly, as shown in Fig. 1. A specialist annotated each data point on the chest X-ray image using AVIEW (Coreline Europe GmbH, Eschborn, Germany) to draw a rectangular outline according to the heart size.

2.4. *Model*

In this study, five CNN models with high-learning accuracy using image datasets such as ImageNet were used to compare the learning results according to the types of



(a) Normal heart (b) Cardiomegaly

Fig. 1. Chest X-ray images.

input images. The CNN models used in this study were VGGNet, ResNet, InceptionNet (GoogleNet), DenseNet, and EfficientNet.[22–25,28]

2.4.1. *VGGNet*

VGGNet is a model developed by a research team at Oxford University that has shown good performance in image recognition using a small 3×3 filter and increasing the depth of the convolutional network, contrasting previous neural network models.[22] Using only several small 3×3 filters throughout the network, more learning parameters were reduced compared to using the previous 5×5 or 7×7 filter, which improved the learning speed and increased the depth of the network to 16 layers. High accuracy with increased feature discrimination was achieved by increasing the nonlinearity by deepening up to 19 layers.

2.4.2. *ResNet*

ResNet is a model that solves problems such as vanishing and exploding gradients that occur as the layer of the CNN model is deepened to improve the performance.[23] The problem that accumulates as the layer deepens is initialized through regularization or the use of an intermediate regularization layer, helping tens of layers to converge via stochastic gradient descent (SGD). Moreover, the mapping of the residual function to the residual during the convergence of a deep neural network is not the cause of overfitting; rather, the mapping to the residual to solve the accuracy degradation problem caused by the optimization algorithm, which becomes more complex as more layers are added, causes overfitting. A shortcut connection, which has a simple structure that adds identity to an existing learned network, does not require additional parameters or complex calculations; thus, efficiency and accuracy can be improved. ResNet is eight times deeper than VGGNet in ImageNet; however, it shows better performance with relatively fewer complex operations.

2.4.3. *InceptionNet*

InceptionNet, also called GoogLeNet, is a model developed from the viewpoint that it is more important to distribute computing resources efficiently than to increase the size of the network when there are limited computing resources.[24] With a sophisticated structural design, the amount of computation does not increase, even if the depth and width of the network are increased, and thus, the efficiency of the use of consumed resources is improved. Despite having 12 fewer parameters than AlexNet, InceptionNet outperforms it.

2.4.4. *DenseNet*

DenseNet is a model with a structure in which all layers are connected in a forward direction to maximize the information flow between the layers as the network layer progresses.[25] Each layer obtains additional information from all previous layers and passes its feature map to all subsequent layers. Because this method receives information from all the previous layers, the feature map becomes more robust and requires fewer parameters. Because each layer can directly access the loss function and slope of the input signal, it has the advantage that it simplifies learning, even when the network structure is complex. By using a bottleneck layer with a reduced computational amount and reducing the number of feature maps through a transition layer, good performance was achieved even with a small number of computations.

2.4.5. *EfficientNet*

CNNs have been developed with limited resources and with the goal of increasing the size of the input image by increasing the number of layers or filters and the resolution of the input image for higher accuracy within the limits of additionally supported resources.[26,27] In this scaling-up method, a compound scaling method was proposed that renders the network deep and wide and balances the three factors that control the resolution of the input image. EfficientNet designed a reference network using a neural architecture search, a method to find an optimal network based on reinforcement learning, and applied compound model scaling.[28] EfficientNet showed the highest top-1 accuracy using a significantly smaller number of parameters than the conventional CNN.

2.5. *Proposed convolutional neural network structure*

In this study, we proposed a CNN consisting of 13 layers. The proposed CNN model consists of eight convolution layers and three fully connected layers. Moreover, it consists of two max-pooling layers that reduce their size by subsampling the input image. Figure 2 shows the CNN model proposed in this study.

The first four layers consist of 160 convolutional kernels of size 2×2. The fifth layer is a max-pooling layer with a size of 2×2, which reduces the size of the input

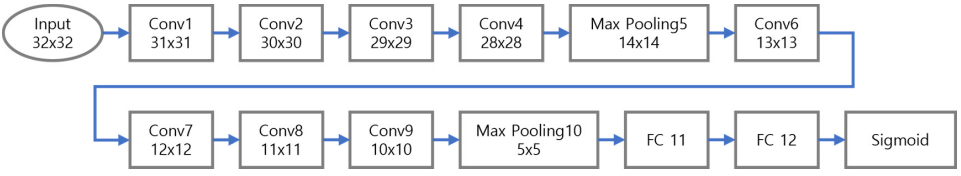


Fig. 2. A flowchart of the layer constituting the CNN model for the classification of normal heart and cardiomegaly.

image by half and outputs an image of size 14×14. The sixth to ninth layers consist of 320 convolutional kernels of size 2×2. The 10th layer consists of a max pooling layer with a size of 2×2. These convolution and max pooling layers extract features used to classify normal heart and cardiomegaly from the input chest X-ray images and select better features through a learning process.

The 11th–13th layers consist of fully connected layers that are connected to all the neurons in the previous layer. The 11th layer is composed of a flattened layer that converts two-dimensional data into one-dimensional data to transmit the selected feature to the fully connected layer. The 12th layer is a fully connected layer, consisting of 8001 nodes, and classifies a normal heart and a cardiomegaly based on the features selected through the previous learning process. The last output layer is a sigmoid layer with one node that outputs a result based on whether the input image is a normal heart or a cardiomegaly.

2.6. *Data conversion*

In this study, DICOM images acquired using ImageJ (LOCI, University of Wisconsin) were converted into 16-bit TIFF files with the same bit as the original image and into 8-bit PNG and JPEG files with different bits from the original image. Using the converted TIFF, PNG, and JEPG files, we compared the learning and testing accuracy according to the types of input images.

2.7. *Data augmentation*

Data augmentation is performed when sufficient learning is required with limited data in the image-recognition field. It is accomplished by applying operations such as translation to the input image.

In this study, the converted JEPG image was rotated using data augmentation within the range of 0–90° clockwise and counterclockwise, as well as enlarging and reducing the width and height of the entire image by 60–140% to randomly increase the number of images. We used this method to augment only the training datasets by 30 and 60 times, excluding the test dataset. Using augmented data, the learning and test accuracies were compared according to the number of training data points.

2.8. *Model optimization*

In this study, we compared the learning results based on the type of input image. To proceed with DL using a CNN, the horizontal and vertical sizes of the input and output images must be identical. Therefore, in this study, the size of the Dicom-converted input image was converted to 224×224, and in the case of the InceptionV3 model, it was converted to 299×299. Preprocessing was performed by normalizing pixel values. Adaptive moment estimation (Adam) was used as the optimizer for training in all datasets for learning according to the type of input image, and cross-entropy was used as the loss function.[29] The batch size was set to 32, and the number of epochs was set to 10 to proceed with learning.

In this study, to optimize the proposed CNN model, we attempted to find parameters that can achieve optimal performance by comparing the accuracy of four optimizers and six learning rates. The four specific optimizers used in the training to compare the classification accuracy according to the optimizer are SGD, root mean square propagation (RMSprop), adaptive gradient (Adagrad), and Adam.[29–31] BinaryCrossentropy was used as the loss function for learning, and because the training requires a long time owing to the large amount of data, the batch size was set to 64 for an effective learning speed. The epoch was initially set to 50; however, after a specific time, the accuracy did not increase further and was either maintained or decreased, thus it was set to 20. The six specific learning rates used to compare the classification accuracy according to the learning rate were 0.005, 0.003, 0.001, 0.0005, 0.0003, and 0.0001.

2.9. *Model evaluation*

In this study, the performances of the models trained according to the type of input image and the proposed CNN model were verified on the test dataset. To validate the model's performance, quantitative indicators such as accuracy, precision, recall, and F1 score were used.[32]

Accuracy represents the probability of the number of classes with the correct label in the class classified as normal or cardiomegaly in the entire chest X-ray image. Precision is an index that evaluates the ability to predict the number of actual normal heart data points in a classified dataset. Recall, also called sensitivity, is an index that indicates the probability of how many data points are classified as normal heart data in actual normal heart data. The F1 score is a harmonic average of precision and recall. The calculation formulas for the performance indicators are expressed in the following equations:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{TN} + \text{FN}} \times 100, \tag{1}$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \times 100, \tag{2}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \times 100, \tag{3}$$

$$\text{F1 Score} = \frac{2 * \text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} \times 100. \tag{4}$$

3. Results

3.1. *Accuracy according to the type of input image*

To compare the classification performance when using TIFF, PNG, and JPEG images converted from DICOM images as input images, the classification accuracy was obtained and compared. The classification accuracy for the training data was calculated using Eq. (1). The classification accuracy obtained by training the five CNN models on TIFF, PNG, and JPEG images converted from DICOM images was examined in the following order: VGG16, ResNet50, InceptionV3, DenseNet121, and EfficientNetB0 models. Table 1 shows the training accuracy of TIFF, PNG, and JPEG images on the VGG16, ResNet50, InceptionV3, DenseNet121, and EfficientNetB0 models. The training accuracy of the TIFF images for the VGG16, ResNet50, InceptionV3, DenseNet121, and EfficientNetB0 models were 99.86%, 99.86%, 99.99%, 100, and 99.89%, respectively; the learning accuracies of the PNG images were 99.88%, 100%, 99.97%, 99.87%, and 100%, respectively; and the training accuracies of the JPEG images were 100%, 100%, 99.96%, 99.89%, and 100%, respectively.

To compare the classification performance when using TIFF, PNG, and JPEG images converted from DICOM images as input images, the accuracy, precision, recall, and F1 score were obtained and evaluated using test data. The performance evaluation indicators for the test data were calculated using Eqs. (1)–(4). After learning the TIFF, PNG, and JPEG images converted from DICOM images, the classification performance evaluated using the test data was analyzed in order of accuracy, precision, recall, and F1 score. Table 2 shows the accuracy, precision, recall, and F1 scores of TIFF, PNG, and JPEG test images. For accuracy, precision, recall, and F1 score, the classification performance of the TIFF images was 100%, 100%, 100%, and 100%, respectively for all models; the classification performance of the PNG images was 100%, 100%, 100%, and 100%, respectively, for all models; and the classification performance of the JPEG images for all models was 100%, 100%, 100%, and 100%, respectively.

Table 1. The training accuracy (%) of five CNN models using TIFF, PNG, and JPEG images.

	VGG16	ResNet50	InceptionV3	DenseNet121	EfficientNetB0
TIFF	99.86	99.86	99.99	100.0	99.89
PNG	99.88	100.0	99.97	99.87	100.0
JPEG	100.0	100.0	99.96	99.89	100.0

Table 2. Validation results of the classification of TIFF, PNG, and JPEG images using test data.

	Accuracy (%)	Precision (%)	Recall (%)	F1 Score (%)
Test data	100.0	100.0	100.0	100.0

3.2. *Accuracy according to data augmentation and optimizer change*

The classification accuracy of normal hearts and cardiomegaly was obtained and compared to determine the effect of data augmentation of the input data on accuracy. Moreover, the effect of using an optimizer on learning accuracy was validated. The classification accuracy based on multiple augmented input data and different optimizers was calculated using Eq. (1). Table 3 shows the classification accuracy when using input data without augmented and input data augmented 30 times and 60 times. The classification accuracy with and without augmented input data was examined in terms of optimizers and the order of SGD, RMSprop, Adagrad, and Adam, without augmented input data, the classification accuracies were 52.52%, 52.52%, 52.52%, and 52.52%, respectively; with input data augmented 30 times, the classification accuracies were 91.54%, 93.93%, 84.47%, and 53.32%, respectively; and with input data augmented 60 times, the classification accuracies were 95.91%, 96.44%, 86.64%, and 52.52%, respectively. All optimizers except Adam showed the highest accuracy.

3.3. *Accuracy according to learning rate change*

To compare the accuracy according to the change in the learning rate, six learning rates were set, and the classification accuracies of normal hearts and cardiomegaly were obtained and compared. A 60-fold augmented dataset that showed the highest accuracy amongst the datasets was augmented. Table 4 shows the learning accuracy according to the learning rate in the input data augmented 60 times. We examined the learning accuracy according to the learning rate in the following order: 0.005, 0.003, 0.001, 0.0005, 0.0003, and 0.0001.

A comparison was conducted using SGD and RMSprop, which showed more than 90% accuracy in the previous experiment. For 0.005, 0.003, 0.001, 0.0005, 0.0003, and 0.0001, the learning accuracies when using SGD were 93.75%, 90.62%, 71.75%,

Table 3. Accuracy (%) according to data augmentation times and optimizer.

	Data	30-times augmentation	60-times augmentation
SGD	52.52	91.54	95.91
RMSprop	52.52	93.93	96.44
Adagrad	52.52	84.47	86.64
Adam	52.52	53.32	52.52

Table 4. Accuracy with changing the learning rate of SGD and RMSprop.

	0.005	0.003	0.001	0.0005	0.0003	0.0001
SGD	93.75	90.62	71.75	52.52	52.52	52.52
RMSprop	52.52	52.52	96.43	97.92	98.21	99.44

52.52%, 52.52%, and 52.52%, respectively; and the learning accuracies when using RMSprop were 52.52%, 52.52%, 96.43%, 97.92%, 98.21%, and 99.44%, respectively. The highest accuracy was obtained when RMSprop was used and the learning rate was 0.0001.

3.4. *Accuracy of the proposed model*

In this study, the accuracy of the classification of normal hearts and cardiomegaly in chest X-ray images was evaluated, and the classification performance was evaluated using the proposed CNN model. A 60-times augmented dataset that showed the highest accuracy in accuracy comparison according to data increase was used. RMSprop, which showed the highest accuracy in accuracy comparison according to optimization and learning rate, was set to 0.0001, and training was performed. Figure 3 shows the learning accuracy and learning loss for the classification of normal hearts and cardiomegaly of the CNN model proposed in this study. The training accuracy of the proposed CNN model for the classification of normal hearts and cardiomegaly was 99.44% and the learning loss value was 0.0284.

To verify the classification performance of the proposed CNN model, the accuracy, precision, recall, and F1 scores were obtained and evaluated using the test dataset. The performance evaluation indicators for the test data were calculated using Eqs. (1)–(4). Table 5 shows the classification performance of the proposed CNN model. The classification performance of the proposed CNN model when using

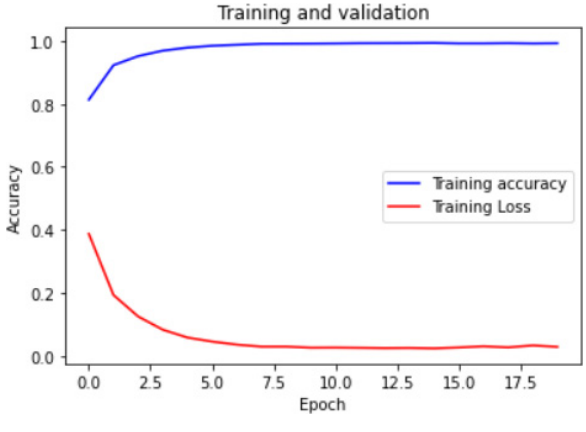


Fig. 3. Classification result of the proposed model.

Table 5. Classification results using test data.

	Accuracy	Precision	Recall	F1 score
Normal	91	85	90	88
Cardiomegaly	91	90	85	88

normal hearts and cardiomegaly data was reviewed in the order of accuracy, precision, recall, and F1 score; in the case of a normal heart, it was 91.0%, 85.0%, 90.0%, and 88.0%, respectively; whereas in the case of an cardiomegaly, it was 91.0%, 90.0%, 85.0%, and 88.0%, respectively.

4. Discussion

Although medical image classification research using a CNN model is being actively conducted, there is a hardware limitation when using a 16-bit DICOM image with a large amount of information. In most cases, medical image classification studies have been conducted using converted 8-bit PNG or JPEG images. Therefore, research results using the 8-bit general image format may not be equally applicable to image classification tasks involving 16-bit medical images; therefore, a study focusing on such medical image analysis is required.

In this study, we attempted to determine whether the input image type affects the classification accuracy using CNN models, which are widely used in medical image analysis. To this end, TIFF, PNG, and JPEG images converted from DICOM images were applied to five CNN models, namely, VGG16, ResNet50, InceptionV3, Densnet121, and EfficientNetB0, to compare the learning and classification performance. Regardless of the format of the input data, all the models showed good performance, each with an accuracy of 99% or more. Moreover, when the classification performance was verified using the test data in the same format as the input image, all the models showed 100% performance. Although the TIFF, PNG, and JPEG images differed in terms of the number of pixels and image compression methods, the models showed similar classification performance. When converting DICOM images to JEPG and PNG images, the number of bits representing pixels was reduced to eight bits; however, the characteristics of the image were not changed; therefore, it was presumed that a similar classification performance was achieved. In addition, it is estimated that similar classification performance was obtained by creating the ranges of TIFF, PNG, and JPEG images with different pixel ranges through the preprocessing process for normalizing pixel values during model training.

In this study, we proposed a DL model based on CNNs, which are widely used in the medical imaging field, to diagnose chest X-ray images. For this purpose, a model for classifying normal hearts and cardiomegaly was presented. It achieved a high learning accuracy of 99.44%, and a low loss value, converging to zero at 0.0284 in the case of the learning loss value. In the performance evaluation using test data, in terms of the accuracy, precision, recall, and F1 score, the normal heart rates were

91.0%, 85.0%, 90.0%, and 88.0%, respectively, and the cardiomegaly rates were 91.0%, 90.0%, 85.0%, and 88.0%, respectively, indicating high classification performance. In terms of the classification performance using these training and test data, the proposed CNN model performed efficiently in the classification of normal heart and cardiomegaly chest X-ray images.

In this study, the data used were augmented, an optimizer was used, and the learning rate varied. Depending on whether the data were augmented, the learning and test accuracies increased with the increase in the number of times the data were augmented. It was concluded that the greater the amount of learning data used, the more the learning proceeded and the better were the results that were achieved. Further, with more data augmentation, the learning and test accuracies increased for all the optimizers, except for the case of the Adam optimizer. SGD and RMSprop achieved an accuracy of more than 90%. The accuracy of SGD tended to decrease with the decrease in the learning rate. RMSprop tended to improve the accuracy and perform stable learning with the decrease in the learning rate.

The classification results of the CNN model proposed in this study were compared with those of the InceptionV3 model, which is currently widely used in medical image classification. The same dataset, optimizer, and learning rate were used to compare the classification results of the two models. As a result of classifying normal heart and cardiomegaly, the learning accuracy was found to be 99.18%. In the order of accuracy, precision, recall, and F1 score in the performance evaluation using test data, the normal heart was 87.0%, 100.0%, 78.0%, and 88.0%, respectively, and the cardiomegaly was 87.0%, 78.0%, 100.0%, and 88.0%, respectively. When the classification results of the proposed CNN model and InceptionV3 model were compared, the difference in the training and test accuracy, precision, and recall was up to 7%. From a methodological point of view, the CNN model proposed in this study was developed to reflect the characteristics of the chest X-ray image, which is a medical image; therefore, it is concluded that higher classification performance results could be obtained compared with the InceptionV3 model. In addition, by using a U-Net-based CNN model, the Singapore research team showed higher accuracy than in the case of classifying normal hearts and cardiomegaly, with an accuracy of 93.75%.[33] Considering these results, it can be concluded that this study performed relatively sufficient learning using more data through data augmentation than the Singapore research team. In summary, the proposed CNN model was found to perform efficiently on chest X-ray images in the classification of normal and abnormal hearts suspected of cardiomegaly.

In the future, we believe that more research with a broader range of data should be conducted to determine whether the results of this study are the characteristics of other medical images. Moreover, based on the results, it will be necessary to investigate the characteristics of useful models in special images, such as chest X-ray images, and continuously develop models that are suitable for various fields.

5. Conclusions

In this study, the effect of the input image format on accuracy in DL was evaluated using TIFF, PNG, and JPEG images that were converted from DICOM images. Based on the obtained results, a CNN-based DL model was proposed. In this study, through medical imaging research using DL, it was validated that the classification performance was unaffected even if the DICOM image was converted into any format and was used as an input image. Moreover, the proposed CNN model exhibited excellent performance in the classification of normal hearts and cardiomegaly on chest X-ray images. It also showed excellent performance in the performance evaluation using test data. This will be helpful in providing information according to the input image type for various studies that aim to apply DL to medical images, and it is also believed to be helpful in the selection of image type and learning parameters. The proposed model is expected to achieve superior results in the classification of diseases based on chest X-ray images.

References

1. Danzer CS, The cardiothoracic ratio: An index of cardiac enlargement, *Amer J Med Sci* **157**:1827–1924, 1919.
2. Paek SH, Lee JM, Han SJ, Bahk YW, Evaluation of cardiac measurements in healthy Korean adults, *J Korean Radiol Soc* **14**:57–62, 1978.
3. Kim YS, Park HJ, Park SH, Chun HJ, Choi BG, A CT criteria of cardiomegaly, *J Korean Radiol Soc* **57**:235–238, 2007.
4. Alghamdi SS, Abdelaziz I, Albadri M, Alyanbaawi S, Aljondi R, Tajaldeen A, Study of cardiomegaly using chest X-ray, *J Radiation Res Appl Sci* **13**:460–467, 2020.
5. Yang S, Lee JS, Kim C, The Accuracy of echocardiography and ECG in the left ventricular hypertrophy, *J Korea Contents Assoc* **16**:666–672, 2016.
6. Frohlich ED, Left ventricular hypertrophy as a risk factor, *Cardiol Clin* **4**:137–144, 1986.
7. Kannel WB, Left ventricular hypertrophy as a risk factor: the Framingham experience, *J Hypertens Suppl* **9**:S3–S8, 1991.
8. Amin H, Siddiqui WJ, Cardiomegaly, *StatPearls*, StatPearls Publishing, 2021.
9. Anderson JC, Baltaxe HA, Wolf GL, Inability to show clot: One limitation of ultrasonography of the abdominal aorta, *Radiology* **132**:693–696, 1979.
10. Moon HJ, Kim EK, Park JS, Kwak JY, Thyroid ultrasound: Change of inter-observer variability and diagnostic performance after training, *J Korean Soc Ultrasound Med* **30**:23–28, 2011.
11. LeCun Y, Bengio Y, Hinton G, Deep learning, *Nature* **521**:436–444, 2015.
12. Goodfellow I, Bengio Y, Courville A, *Deep Learning*, MIT Press, 2016.
13. Krizhevsky A, Sutskever I, Hinton GE, Imagenet classification with deep convolutional neural networks, *Adv Neural Inf Process Syst* **25**:1097–1105, 2012.
14. Yap MH, Goyal M, Osman F, Ahmad E, Martí R, Denton E, Zwiggelaar R, End-to-end breast ultrasound lesions recognition with a deep learning approach, *Proc SPIE* **10578**:1057819, 2018.
15. Han S, Kang HK, Jeong JY, Park MH, Kim W, Bang WC, Seong YK, A deep learning framework for supporting the classification of breast lesions in ultrasound images, *Phys Med Biol* **62**:7714, 2017.

16. Sahiner B, Pezeshk A, Hadjiiski LM, Wang X, Drukker K, Cha KH, Giger ML, Deep learning in medical imaging and radiation therapy, *Med Phys* **46**:e1–e36, 2019.
17. Wang X, Peng Y, Lu L, Lu Z, Bagheri M, Summers RM, Chestx-ray8: Hospital-scale chest x-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases, *Proc IEEE Conf Computer Vision and Pattern Recognition*, pp. 2097–2106, 2017.
18. Tschandl P, Rosendahl C, Kittler H, The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions, *Sci Data* **5**:1–9, 2018.
19. Kwon SM, Change of image quality within compression of AAPM CT performance phantom image using JPEG2000 in PACS, *J Korean Soc Radiol* **6**:217–226, 2012.
20. Kim JY, Ko SJ, Comparison of DICOM images and various types of images, *J Inst Convergence Signal Process* **18**:76–83, 2017.
21. Kabachinski J, TIFF, GIF, and PNG: Get the picture?, *Biomed Instrum Technol* **41**:297–300, 2007.
22. Simonyan K, Zisserman A, Very deep convolutional networks for large-scale image recognition, arXiv:1409:1556, 2014.
23. He K, Zhang X, Ren S, Sun J, Deep residual learning for image recognition, *Proc IEEE Conf Computer Vision and Pattern Recognition*, pp. 770–778, 2016.
24. Szegedy C, Vanhoucke V, Ioffe S, Shlens J, Wojna Z, Rethinking the inception architecture for computer vision, *Proc IEEE Conf Computer Vision and Pattern Recognition*, pp. 2818–2826, 2016.
25. Huang G, Liu Z, Van Der Maaten L, Weinberger KQ, Densely connected convolutional networks, *Proc IEEE Conf Computer Vision and Pattern Recognition*, pp. 4700–4708, 2017.
26. Zagoruyko S, Komodakis N, Wide residual networks, arXiv:1605:07146, 2016.
27. Huang Y, Cheng Y, Bapna A, Firat O, Chen D, Chen M, Wu Y, Gpipe: Efficient training of giant neural networks using pipeline parallelism, *Adv Neural Inf Process Syst* **32**:103–112, 2019.
28. Tan M, Le Q, Efficientnet: Rethinking model scaling for convolutional neural networks, *Int Conf Machine Learning*, PMLR, pp. 6105–6114, 2019.
29. Kingma DP, Ba J, Adam: A method for stochastic optimization, arXiv:1412:6980, 2014.
30. Ruder S, An overview of gradient descent optimization algorithms, arXiv:1609:04747, 2016.
31. Duchi J, Hazan E, Singer Y, Adaptive subgradient methods for online learning and stochastic optimization, *J Mach Learn Res* **12**:2121–2159, 2011.
32. Japkowicz N, Shah M, *Evaluating Learning Algorithms: A Classification Perspective*, Cambridge University Press, 2011.
33. Ronneberger O, Fischer P, Brox T, U-net: Convolutional networks for biomedical image segmentation, *Int Conf Medical Image Computing and Computer-Assisted Intervention*, Springer, Cham, pp. 234–241, 2015.