

컨볼루션 신경망 모델을 이용한 분류에서 입력 영상의 종류가 정확도에 미치는 영향

김민정^{1†} · 김정훈² · 박지은³ · 정우연¹ · 이종민^{4*}

¹경북대학교대학원 의용생체공학과, ²경북대학교병원 생명의학연구원
³경북대학교 비선형동역학연구소, ⁴경북대학교 의과대학 영상의학교실

The Effect of Type of Input Image on Accuracy in Classification Using Convolutional Neural Network Model

Min Jeong Kim^{1†}, Jung Hun Kim², Ji Eun Park³, Woo Yeon Jeong¹ and Jong Min Lee^{4*}

¹Department of Biomedical Engineering Kyungpook National University

²Bio-Medical Research institute, Kyungpook National University Hospital

³Nonlinear Dynamics Research Center, Kyungpook National University

⁴Department of Radiology, School of Medicine, Kyungpook National University

(Manuscript received 31 May 2021 ; revised 8 August 2021 ; accepted 12 August 2021)

Abstract: The purpose of this study is to classify TIFF images, PNG images, and JPEG images using deep learning, and to compare the accuracy by verifying the classification performance. The TIFF, PNG, and JPEG images converted from chest X-ray DICOM images were applied to five deep neural network models performed in image recognition and classification to compare classification performance. The data consisted of a total of 4,000 X-ray images, which were converted from DICOM images into 16-bit TIFF images and 8-bit PNG and JPEG images. The learning models are CNN models - VGG16, ResNet50, InceptionV3, DenseNet121, and EfficientNetB0. The accuracy of the five convolutional neural network models of TIFF images is 99.86%, 99.86%, 99.99%, 100%, and 99.89%. The accuracy of PNG images is 99.88%, 100%, 99.97%, 99.87%, and 100%. The accuracy of JPEG images is 100%, 100%, 99.96%, 99.89%, and 100%. Validation of classification performance using test data showed 100% in accuracy, precision, recall and F1 score. Our classification results show that when DICOM images are converted to TIFF, PNG, and JPEG images and learned through preprocessing, the learning works well in all formats. In medical imaging research using deep learning, the classification performance is not affected by converting DICOM images into any format.

Key words: X-ray, Convolutional neural network, Classification, Deep learning

I. 서 론

딥러닝(Deep Learning)은 인간의 뇌에 있는 뉴런의 동작을 모방한 인공 신경망(Artificial Neural Networks)을 여러 층 쌓은 심층 신경망(Deep Neural Network)을 이용하여 입력 데이터로부터 다양한 특징을 추출하고 계층적

로 학습하여 뛰어난 성능을 이끌어내는 연구 분야이다. 과거의 딥러닝은 계산 성능의 한계와 느린 학습 시간, 학습 데이터 수집의 어려움 등의 문제로 사용하기에 어려움이 있었다. 그러나 수 년 전부터 새로운 학습 알고리즘이 개발되고 빅데이터의 증가와 하드웨어 기술의 발전 등 여러 문제가 해결되면서 그 응용 분야를 넓혔다. 딥러닝을 이용한 영상 인식은 2012년 ILSVRC(Large Scale Visual Recognition Challenge)에서 컨볼루션 신경망(Convolutional Neural Network)을 여러 층으로 쌓은 AlexNet이 사람보다 높은 성능을 보이면서 컴퓨터 비전 분야에서 각광을 받기 시작했다[1-2].

*Corresponding Author : Jong Min Lee

Department of Radiology, School of Medicine, Kyungpook National University, Dongin-dong, Jung-gu, Daegu, South Korea
Tel: +82-53-950-7956
E-mail: jonglee@knu.ac.kr

영상 인식과 분류 등 여러 분야에서 뛰어난 성능을 보이는 컨볼루션 신경망 모델은 의료 영상의 다양한 문제를 해결하기 위한 연구에 활용되고 있다. 의료 영상 분야에서 컨볼루션 신경망 모델은 의료 영상을 특정 질환에 따라 분류하거나 병변의 위치를 찾고, 장기를 분할하는 등 다양한 형태로 활용되고 있다[3-5]. 또한, 의사들이 질병의 진단과 치료 방법과 같은 임상적 의사결정을 내리는 데에도 중요한 도움을 주고 있다[6].

심장비대는 심장이 흉곽의 내경보다 50% 이상 커진 상태를 의미하며, 심전도검사(Electrocardiogram, ECG), 단순 흉부 방사선 검사(X-ray), 심장 초음파 검사(Echocardiography) 등으로 진단이 가능하다. 심장비대는 심혈관계의 합병증을 유발하여 임상적으로 문제가 되고 있으며 선별진단이 매우 중요한 질환이기 때문에 자동화된 판독이 도움될 것이다[7-10].

의료 영상 연구에 사용되는 대부분의 Open-source dataset은 PNG(Portable Network Graphics) 또는 JPEG(Joint Photographic Experts Group) 파일 형태로 이루어져 있다. DICOM(Digital Imaging and Communications) 영상을 사용할 때에도, PNG, JPEG와 같은 이미지 파일 형식으로 변환하여 연구를 수행하는 것이 대부분이다[11-13]. 현재, DICOM 영상과 가장 유사한 파일 형식으로 TIFF(Tagged Image File Format) 이미지 파일 형식이 보고되고 있다. DICOM 파일을 지원하지 않는 디지털 의료 영상 및 평가 프로그램을 이용한 기초 연구 분야에서도 DICOM 영상과 유사한 결과를 나타내는 TIFF 이미지 파일을 기준으로 제시하고 있다[14]. DICOM 파일은 일반적인 영상과 달리 환자 정보, 검사 정보 등 의료 영상으로서 효과적으로 사용하기 위해 영상 자체 정보 외에 고유한 정보들을 함께 저장하여 많은 양의 정보를 표현하는 파일이다. 현재 의료 영상 연구는 이러한 DICOM 파일을 PNG나 JPEG로 변환시킨 데이터를 사용하여 알고리즘의 분류 결과를 발표한 것이 대부분이다. DICOM 영상을 변환시킬 때, PNG는 원본 데이터

를 유지하는 무손실 압축 방식을 사용하는 반면, JPEG는 일부 이미지 데이터를 손실시키는 손실 압축 방식을 사용한다[15]. 많은 양의 정보를 가지고 있는 16비트의 DICOM 파일을 8비트의 PNG나 JPEG 파일로 변환시킨 영상을 사용하는 방법이 최적인지 확인해볼 필요가 있다고 생각한다.

본 연구의 목적은 대부분의 Open-source dataset를 구성하고 있는 8비트의 PNG, JPEG 파일과 16비트의 파일을 입력 영상으로 사용하였을 때, 비슷한 분류 결과를 나타내는지 확인해보는 것이다. 이를 위해 의료 영상의 표준인 DICOM 파일을 가장 유사한 파일 형식으로 보고되고 있는 16비트의 TIFF 파일로 변환시킨 영상과 8비트의 PNG, JPEG 파일로 변환시킨 영상을 입력 영상으로 사용하여 분류 결과를 확인해보고자 하였다. 본 연구에서는 영상 인식 및 분류에서 좋은 성과를 냈던 5개의 심층 신경망 모델들에 흉부 X-ray DICOM 영상을 변환시킨 TIFF, PNG, JPEG 영상을 적용하여 학습 성능을 비교하고자 한다.

II. 연구 방법

1. 데이터 수집

본 연구에서는 경북대학교병원의 임상시험심사위원회(IRB)의 승인을 받은 후 수집을 진행하였다(KNUH 2020-07-018). 경북대학교병원에서 심장비대증 진단을 받은 환자의 흉부 X-ray 영상을 DICOM 영상으로 4,000장 획득하였다. 획득한 DICOM 영상들을 ImageJ(LOCI, University of Wisconsin)를 사용하여 원본 영상과 같은 비트인 16비트의 TIFF 파일과 원본 영상과 다른 비트인 8비트의 PNG, JPEG 파일로 변환하였다. 변환된 TIFF, PNG, JPEG 영상은 다음 그림 1과 같다.

2. 모델

본 연구에서는 ImageNet과 같은 영상 데이터 세트를 이

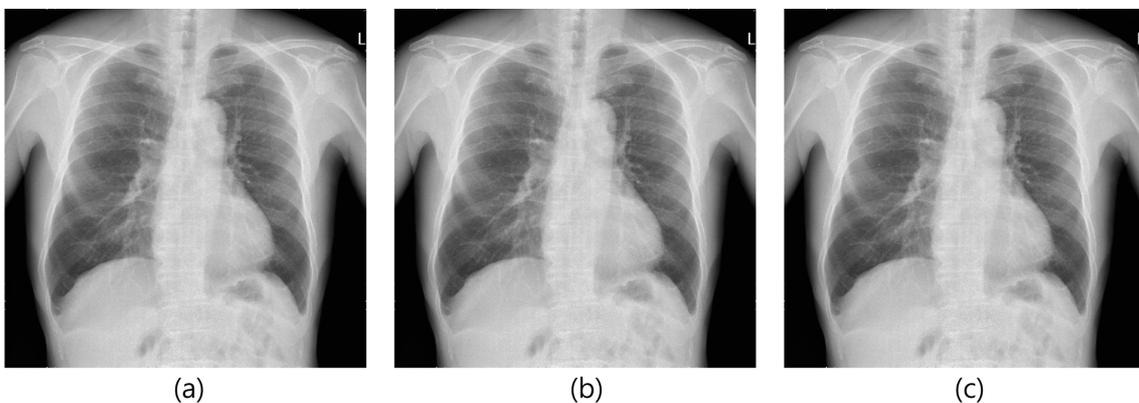


그림 1. DICOM 영상을 변환시킨 TIFF, PNG, JPEG 영상의 예, (a) TIFF 영상, (b) PNG 영상, (c) JPEG 영상
Fig. 1. The Examples of TIFF, PNG, JPEG converted from DICOM, (a) TIFF, (b) PNG, (c) JPEG

표 1. ImageNet을 사용하였을 때 각 모델들의 성능
Table 1. Results of all CNN models using ImageNet

Model	Year	Top-1 Accuracy	Top-5 Accuracy	Parameters
VGG16	2014	74.4	91.9	138,357,544
ResNet50	2015	77.15	93.29	25,636,712
InceptionV3	2015	78.8	93.7	23,851,784
DenseNet121	2016	74.98	92.29	8,062,504
EfficientNetB0	2019	78.3	93.2	5,330,571

용한 학습에서 높은 정확도를 보여주었던 5개의 컨볼루션 신경망 모델을 사용하여 입력 영상의 종류에 따른 학습 성능을 비교하였다. 연구에 사용된 CNN 모델은 많이 활용되는 VGGNet, ResNet, InceptionNet(GoogleNet), DenseNet, EfficientNet 모델이다[16-19,22]. 위의 표 1에 본 연구에서 사용된 모델들을 ImageNet으로 학습하였을 때의 특징을 정리하였다.

(1) VGGNet

VGGNet은 옥스퍼드 대학의 Visual Geometry Group 연구팀에서 개발한 모델로 이전의 신경망 모델들과 달리 3×3의 작은 필터를 사용하고 컨볼루션 네트워크의 깊이를 증가시켜 영상 인식에서 더 좋은 성능을 보여주었다[16]. 네트워크 전반에 걸쳐 여러 개의 3×3의 작은 필터만을 사용하여 이전의 5×5 필터나 7×7 필터를 사용할 때보다 학습 파라미터(parameter)의 수를 감소시켜 학습 속도를 더 높이고, 네트워크의 깊이를 16층과 19층까지 깊게 하여 비선형성을 증가시켜 증가된 특징의 식별성으로 높은 정확도를 보여주었다.

(2) ResNet

ResNet은 보다 좋은 성능을 보여주기 위해 CNN 모델의 계층을 깊게 만들수록 나타나는 Vanishing과 Exploding gradients 등의 문제를 해결한 모델이다[17]. 계층이 깊어지며 누적되는 문제를 초기화 하는 방법으로 정규화를 하거나 중간 정규화 계층을 두어 해결하였고, 수십 층의 계층들이 확률적 경사 하강법(Stochastic Gradient Descent, SGD)를 통해 수렴하도록 도와주었다. 또한, 깊은 네트워크가 수렴하는 과정에서 과적합(overfitting)의 원인이 아니라 계층이 쌓일수록 복잡해지는 최적화 알고리즘(optimize)때문에 발생하는 정확도 저하 문제를 해결하기 위해 잔차(residual)에 맵핑하는 잔차 함수(residual function)를 만들었다. 기존의 학습된 네트워크에 입력(identity)을 더하는 간단한 구조의 Shortcut Connection은 추가적인 파라미터와 복잡한 연산이 필요하지 않아 효율과 정확도를 높일 수 있다. ResNet은

ImageNet에서 VGGNet보다 8배나 깊어 졌지만, 비교적 덜 복잡한 연산으로 더 좋은 성능을 보여주었다.

(3) InceptionNet

GoogLeNet이라고도 불리는 InceptionNet은 한정적인 컴퓨팅 자원에서 네트워크의 크기를 늘리는 것보다 컴퓨팅 자원을 효율적으로 분배하는 것이 중요하다는 관점에서 개발된 모델이다[18]. 정교한 구조 설계로 네트워크의 깊이와 너비를 늘려도 연산량이 증가하지 않고 유지되어 소모되는 자원의 사용 효율이 개선되었다. InceptionNet은 AlexNet 보다 12배나 적은 파라미터에도 불구하고 훨씬 정확한 성능을 보여주었다.

(4) DenseNet

DenseNet은 네트워크 계층이 진행됨에 따라 계층 간의 정보 흐름을 최대화하기 위해 모든 계층들이 서로 순방향으로 연결되는 구조를 가진 모델이다[19]. 각각의 계층은 이전의 모든 계층으로부터 추가적인 정보를 얻고, 고유의 특징 맵(feature map)을 그 다음의 모든 계층으로 전달한다. 이러한 방식은 이전의 모든 계층으로부터 정보를 받기에, 특징 맵이 더 견고 해져 더 적은 파라미터 수를 요구한다. 각각의 계층들은 손실 함수 및 입력 신호의 기술기에 직접 접근할 수 있기 때문에 네트워크의 구조가 깊어도 학습이 쉬워진다는 장점이 있다. 연산량을 줄인 bottleneck layer를 사용하고 Transition layer를 통해 특징 맵의 수를 줄여 적은 수의 연산량으로도 좋은 성능을 나타냈다.

(5) EfficientNet

컨볼루션 신경망은 한정된 자원 내에서 개발되어왔으며, 추가적으로 지원되는 자원의 한도 내에서 더 높은 정확도를 위해서 계층 또는 필터의 수를 늘리고 입력 이미지의 해상도를 높여 그 크기를 키워가는 방향으로 발전되어왔다[20-21]. 이러한 scaling-up 방법에서 네트워크를 깊고 넓게 하며 입력 이미지의 해상도를 조절하는 3가지 요소의 균형을 맞추는 compound scaling method이 제안되었다. EfficientNet은

표 2. 각 데이터 세트의 학습, 검증, 테스트 데이터 영상 개수
Table 2. Number of images in Train, Validation, Test Data

Data set	Train (Validation) Data		Test Data	
	Image Type	Image Number	Image Type	Image Number
Set1		2400 (800)	TIFF	800
Set2	TIFF	2400 (800)	PNG	800
Set3		2400 (800)	JPEG	800
Set4		2400 (800)	PNG	800
Set5	PNG	2400 (800)	TIFF	800
Set6		2400 (800)	JPEG	800
Set7		2400 (800)	JPEG	800
Set8	JPEG	2400 (800)	TIFF	800
Set9		2400 (800)	PNG	800

강화학습을 기반으로 최적의 네트워크를 찾는 방식인 Neural Architecture Search(NAS)를 사용하여 기존 네트워크를 설계하였고, compound model scaling을 적용하였다. EfficientNet는 기존의 컨볼루션 신경망에 비해 훨씬 적은 파라미터 수를 사용하여 가장 높은 Top-1 정확도를 나타내었다[22].

3. 데이터 세트

본 연구에서 DIOCM 파일을 변환시킨 TIFF, PNG, JPEG 파일을 이용하여 총 9개의 데이터 세트를 구성하였다. 모든 데이터 세트는 학습, 검증, 테스트 데이터를 6:2:2의 비율로 나눴고, 표 2에서 각 데이터 세트에 대한 구체적인 학습, 검증, 테스트 데이터의 영상 개수를 정리하였다.

첫 번째 데이터 세트는 학습, 검증 및 테스트 데이터를 TIFF 파일로 구성하였다. 두 번째 데이터 세트는 TIFF 파일로 학습과 검증 데이터를 구성하고, PNG 파일로 테스트 데이터를 구성하였다. 세 번째 데이터 세트는 TIFF 파일로 학습과 검증 데이터를 구성하고, JPEG 파일로 테스트 데이터를 구성하였다. 네 번째 데이터 세트는 학습, 검증, 테스트 데이터를 PNG 파일로 구성하였다. 다섯 번째 및 여섯 번째 데이터 세트는 PNG 파일로 학습과 검증 데이터를 구성하고, 각각 TIFF와 JPEG 파일로 테스트 데이터를 구성하였다. 일곱 번째 데이터 세트는 학습, 검증, 테스트 데이터를 JPEG 파일로 구성하였다. 여덟 번째와 아홉 번째 데이터 세트는 JPEG 파일 학습 및 검증 데이터를 구성하고, 각각 TIFF 및 PNG 파일로 테스트 데이터를 구성하였다.

4. 모델 학습

컨볼루션 신경망을 이용하여 딥 러닝 학습을 진행하기 위해서는 입력 및 출력 영상의 가로 및 세로 크기가 동일해야 한다. 따라서 본 연구에서는 모든 입력 및 출력 영상의 크

기를 224×224 크기로 변환하였고, InceptionV3 모델의 경우에는 299×299 크기로 변환하였다. 그 후, 픽셀 값을 정규화하여 전처리를 진행하였다. 모든 데이터 셋에서 학습에 사용된 최적화 알고리즘(optimizer)은 Adam을 사용하였고, 손실 함수로는 Cross Entropy를 사용하였다. 배치 사이즈(batch size)는 32로 지정하였고, 에폭(epoch)은 10으로 설정하여 학습을 진행하였다.

본 연구에서는 전처리 과정 및 학습을 위해 Python(version 3.8.3) 언어 기반의 통합 개발 환경인 주피터 노트북(Jupyter notebook)을 사용하였다. 딥러닝 학습을 위한 시스템은 Intel (R) Core (TM) i7-8700k(Intel, Santa Clara, CA, USA), NVIDIA GeForce GTX 1080 Ti(NVIDIA, Santa Clara, CA, USA), Tensorflow 2.3.1, Keras 2.4.3를 사용하였다. 학습에 사용된 5개의 구체적인 컨볼루션 신경망 모델은 VGG16, ResNet50, InceptionV3, DenseNet121, EfficientNetB0이다. 본 연구에서 학습과 테스트를 포함한 전체적인 진행 순서는 다음의 그림 2와 같다.

5. 평가

본 연구에서는 학습을 진행한 모델들의 성능을 검증 데이터와 테스트 데이터를 사용하여 비교하였다. 모델 성능 비교에는 정량적 지표인 일반적으로 널리 사용되는 정확도(Accuracy), 정밀도(Precision), 재현율(Recall), F1 Score을 사용하여 나타냈다[23].

(1) 정확도(Accuracy)

정확도(Accuracy)는 전체 판별된 집합에서 얼마나 많은 클래스가 올바른 라벨을 가졌는지에 대한 확률을 나타내는 지표이다. 정확도는 아래의 식 (1)로 계산할 수 있다.

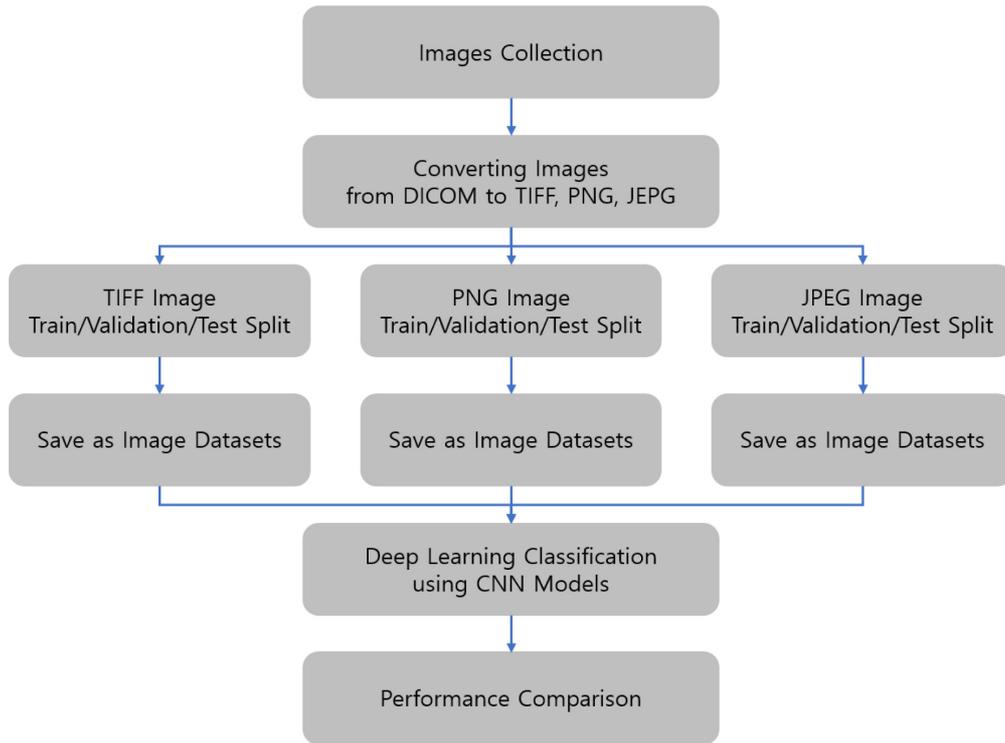


그림 2. 딥러닝 TIFF, PNG, JPEG 영상 분류 순서도

Fig. 2. Deep Learning TIFF, PNG, JPEG image Classification Flow chart

$$\text{Accuracy} = \frac{TP+TN}{TP+FP+TN+FN} \times 100 \quad (1)$$

$$\frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \times 100 \quad (4)$$

(2) 정밀도(Precision)

정밀도(Precision)는 양성이라고 예측된 집합에서 얼마나 많은 클래스가 진짜 양성인지 예측 능력을 평가한 지표이다. 정밀도에 대한 계산식은 아래의 식 (2)로 표현할 수 있다.

$$\text{Precision} = \frac{TP}{TP+FP} \times 100 \quad (2)$$

(3) 재현율(Recall)

재현율(Recall)은 sensitivity로도 불리며, 전체 양성 집합에서 얼마나 많은 클래스가 양성 클래스로 분류되는지에 대한 확률을 나타내는 지표이며, 아래의 식 (3)으로 계산식을 나타낼 수 있다.

$$\text{Recall} = \frac{TP}{TP+FN} \times 100 \quad (3)$$

(4) F1 Score

F1 Score는 Precision과 Recall의 조화평균으로 데이터 라벨이 불균형을 이루는 경우, 모델의 성능을 하나의 숫자로 표현하여 정확하게 평가할 수 있으며, 계산식은 아래의 식 (4)로 나타낼 수 있다.

각 평가지표에서 진양성(true positive, TP)는 실제로 양성이며 양성으로 판별된 경우를 의미하고, 위양성(false positive, FP)은 실제로 음성이지만 양성으로 판별된 경우를 의미한다. 진음성(true negative, TN)은 실제로 음성이며 음성으로 판별된 경우, 위 음성(false negative, FN)은 실제로 양성이지만 음성으로 판별된 경우를 의미한다.

III. 연구 결과

본 논문에서는 5개의 컨볼루션 신경망(Convolution Neural Network)을 기반으로 DICOM 영상을 변환시킨 TIFF, PNG, JPEG 영상을 적용하여 성능 평가 지표인 정확도, 정밀도, 재현율, F1 Score를 비교하였다.

1. 학습 데이터에 대한 성능 평가

(1) 정확도 평가

입력 영상으로 흉부 X-ray 영상을 변환시킨 TIFF 영상, PNG 영상, JPEG 영상을 사용하였을 때의 분류 성능을 비교하기 위하여 분류 정확도를 구하고 비교하였다. 학습 데이터에 대한 분류 정확도는 식 (1)을 통하여 계산하였다. 흉부

표 3. 5개의 컨볼루션 신경망 모델에서의 TIFF, PNG, JPEG 영상의 분류 정확도 (%)

Table 3. Classification accuracy (%) of five CNN models using TIFF, PNG, JPEG data

	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

X-ray 영상을 변환시킨 TIFF, PNG, JPEG 영상을 5개의 컨볼루션 신경망 모델에 학습시킨 분류 정확도를 VGG16, ResNet50, InceptionV3, DenseNet121, EfficientNetB0 모델 순서로 살펴본다. 5개의 컨볼루션 신경망 모델에서 TIFF 영상의 정확도는 차례로 99.86%, 99.86%, 99.99%, 100%, 99.89%이다. PNG 영상의 정확도는 차례로 99.88%, 100%, 99.97%, 99.87%, 100%이다. JPEG 영상의 정확도는 차례로 100%, 100%, 99.96%, 99.89%, 100%이다. 위의 표 3에서 5개의 컨볼루션 신경망 모델에서의 TIFF, PNG, JPEG 영상의 분류 정확도를 정리하여 나타냈다.

2. 테스트 데이터에 대한 성능 평가

(1) 정확도, 정밀도, 재현율, F1 Score 평가

입력 영상으로 흉부 X-ray 영상을 변환시킨 TIFF, PNG, JPEG 영상을 사용하였을 때의 분류 성능을 비교하기 위하여 테스트 데이터를 사용하여 정확도, 정밀도, 재현율, F1 Score를 구하고 평가하였다. 학습 영상과 동일한 영상 포맷을 가진 테스트 데이터를 사용하였을 때와 다른 영상 포맷을 가진 테스트 데이터를 사용하였을 때의 분류 성능도 확인하였다. 테스트 데이터에 대한 성능 평가 지표들은 식 (1)~(4)를 통하여 계산하였다. 아래의 표 4에 TIFF, PNG, JPEG 파일을 테스트 데이터로 사용하였을 때의 정확도, 정밀도, 재현율, F1 Score의 결과를 정리하여 나타냈다.

흉부 X-ray 영상을 변환시킨 TIFF 영상을 학습시켜 TIFF, PNG, JPEG 파일을 테스트 데이터로 사용하였을 때의 정확도를 VGG16, ResNet50, InceptionV3, DenseNet121, EfficientNetB0 모델 순서로 살펴본다. TIFF 영상의 경우 차례로 100%, 100%, 100%, 100%, 100% 이고 PNG 영상의 경우 차례로 100%, 100%, 100%, 100%, 100% 이고 JPEG 영상은 차례로 100%, 100%, 100%, 100%, 100% 이다. 흉부 X-ray 영상을 변환시킨 PNG 영상을 학습시켜 TIFF, PNG, JPEG 파일을 테스트 데이터로 사용하였을 때, 모든

모델에서 정확도가 100%로 동일한 결과를 나타냈다. 흉부 X-ray 영상을 변환시킨 JPEG 영상을 학습시켜 TIFF, PNG, JPEG 파일을 테스트 데이터로 사용하였을 때에도 모든 모델에서 정확도가 100%로 동일한 결과를 나타냈다.

흉부 X-ray 영상을 변환시킨 TIFF 영상을 학습시켜 TIFF, PNG, JPEG 파일을 테스트 데이터로 사용하였을 때의 정밀도, 재현율, F1 Score를 살펴본다. TIFF 영상의 경우 모든 모델에서 차례로 100%, 100%, 100% 이고 PNG 영상의 경우 차례로 100%, 100%, 100% 이고 JPEG 영상은 차례로 100%, 100%, 100% 이다. 흉부 X-ray 영상을 변환시킨 PNG 영상을 학습시켜 TIFF, PNG, JPEG 파일을 테스트 데이터로 사용하였을 때, 모든 모델에서 정밀도, 재현율, F1 Score가 100%로 동일한 결과를 나타냈다. 흉부 X-ray 영상을 변환시킨 JPEG 영상을 학습시켜 TIFF, PNG, JPEG 파일을 테스트 데이터로 사용하였을 때에도 모든 모델에서 정밀도, 재현율, F1 Score가 100%로 동일한 결과를 나타냈다.

IV. 고찰 및 결론

컨볼루션 신경망 모델을 이용한 의료 영상 분류 연구는 활발히 진행되고 있지만, 많은 양의 정보를 가지고 있는 16비트의 DICOM 영상을 사용하기에는 하드웨어와 같은 제약이 존재하여 PNG나 JPEG와 같은 8비트의 일반적인 이미지 포맷으로 변환시킨 영상을 이용하여 연구를 진행하는 것이 대부분인 상황이다. 따라서 8비트의 일반적인 이미지 포맷을 이용하였을 때의 연구 결과가 16비트의 의료 영상을 이용한 영상 분류에서 동일하게 적용되지 못할 수도 있기 때문에 이러한 의료 영상 분석에 초점을 맞춘 연구들이 필요하다고 생각된다.

본 논문에서는 의료 영상을 이용한 연구에서 최근 많이 활용되고 있는 컨볼루션 신경망 모델들을 사용하여 입력 영상의 종류가 분류 정확도에 영향을 미치는 지 확인하고자 하였

표 4. 테스트 데이터에서의 TIFF, PNG, JPEG 영상 분류 성능 검증 결과

Table 4. Validation result of classification of TIFF, PNG, JPEG image using test data

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

다. 이를 위해 DICOM 영상을 변환시킨 TIFF, PNG, JPEG 영상을 5개의 컨볼루션 신경망 모델인 VGG16, ResNet50, InceptionV3, Densnet121, EfficientNetB0에 적용하여 학습 및 분류 성능을 비교하였다.

그 결과, 입력 영상의 포맷에 관계없이 모두 정확도 99% 이상의 좋은 성능을 보여주었다. 또한, 입력 영상과 같은 포맷을 가진 테스트 데이터와 다른 포맷을 가진 테스트 데이터를 사용하여 분류 성능을 검증하였을 때도 모두 100%의 성능을 보여주었다.

원본 영상에서 관심영역의 크기는 심장의 크기에 따라 제각기 다르기 때문에 관심영역을 잘라내는 영상의 크기도 다르다. 본 연구에서는 입력 및 출력 영상의 가로 및 세로 크기가 동일해야 학습이 진행되는 딥러닝의 특성에 맞춰 영상의 크기를 모두 일정한 크기로 변환하여 연구에 사용하였다. 이때, 영상의 크기는 알고리즘에 맞춰 선정하였는데 영상의 크기에 따라 학습 결과가 달라질 수도 있을 것이라는 생각이 들어 최적의 영상의 크기를 찾는 연구를 추가로 진행하고자 한다.

TIFF, PNG, JPEG 영상은 픽셀의 수와 이미지 압축 방식에서 뚜렷한 차이점을 가지고 있지만, 본 연구의 결과에 따르면 비슷한 분류 성능을 나타냈다. 이는 모델 학습에서 픽셀 값을 정규화 시키는 전처리 과정을 진행하여 다른 픽셀 범위를 가지고 있는 TIFF, PNG, JPEG 영상의 범위를 동일하게 만들어주었기 때문에 비슷한 분류 성능이 나타난 것이라고 추측된다. 향후 본 논문의 결과가 8비트의 일반적인 영상 포맷 또는 16비트의 의료 영상 포맷에서 일반적으로 확인되는 특징인지 알아보기 위해서 더욱 다양하고 많은 데이터를 이용한 연구가 진행될 필요가 있다.

최종적으로, 본 연구를 통해 DICOM 영상을 TIFF, PNG, JPEG 영상으로 변환시켜 전처리 과정을 통해 학습을 진행하면 어떠한 포맷을 사용하여도 학습이 잘 된다는 것을 확인할 수 있었다. 딥러닝을 이용한 의료 영상 연구에서, DICOM 영상을 어떠한 포맷으로 변환시켜 입력 영상으로 사용하여도 분류 성능에 영향을 미치지 않는다는 것을 의미한다. 이는 의료 영상에 딥러닝을 적용하려는 여러 연구들에게 입력 영상 종류에 따른 정보 제공에 도움이 될 것이며 영상 종류를 선택하는 데에도 도움이 될 것이라고 생각된다.

References

- [1] LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature*. 2015; 521(7553):436-444.
- [2] Krizhevsky A, Sutskever I, Hinton GE. Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*. 2012;25:1097-1105.
- [3] Yap MH, Goyal M, Osman F, Ahmad E, Marti R, Denton E, Zwiggelaar R. End-to-end breast ultrasound lesions recognition with a deep learning approach. In *Medical Imaging 2018: Biomedical Applications in Molecular, Structural, and Functional Imaging*. International Society for Optics and Photonics. 2018;10578:1057819.
- [4] 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. *Physics in Medicine & Biology*. 2017;62(19):7714.
- [5] Sahiner B, Pezeshk A, Hadjiiski LM, Wang X, Drukker K, Cha KH, Giger ML. Deep learning in medical imaging and radiation therapy. *Medical physics*. 2019;46(1):e1-e36.
- [6] 송하연, 김지은, 김태년. 인공지능과 헬스 커뮤니케이션. *인공지능연구*, 2020;57.3:196-238.
- [7] Yang SH, Lee J-S, Kim CS. The Accuracy of Echocardiography and ECG in the Left Ventricular Hypertrophy. *The Journal of the Korea Contents Association*, 2016;16.2:666-672.
- [8] Frohlich, ED. Left ventricular hypertrophy as a risk factor. *Cardiology clinics*, 1986;4.1:137-144.
- [9] Amin H, Siddiqui WJ. Cardiomegaly. *StatPearls* [internet]. 2020.
- [10] Alghamdi SS, et al. Study of cardiomegaly using chest x-ray. *Journal of Radiation Research and Applied Sciences*, 2020; 13.1:460-467.
- [11] 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. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017;2097-2106.
- [12] Tschandl P, Rosendahl C, Kittler H. The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions. *Scientific data*. 2018;5(1):1-9.
- [13] Kwon SM. Change of Image Quality within Compression of AAPM CT Performance Phantom Image Using JPEG2000 in PACS. *Journal of The Korean Society of Radiology*. 2012; 6(3):217-226.
- [14] Kim JY, Ko SJ. Comparison of DICOM images and various types of images. *Journal of the Institute of Convergence Signal Processing*. 2017;18(2):76-83.
- [15] Kabachinski J. TIFF, GIF, and PNG: get the picture?. *Bio-medical instrumentation & technology*. 2007;41(4):297-300.
- [16] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv*. 2014;1409.1556.
- [17] He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016;770-778.
- [18] Szegedy C, Vanhoucke V, Ioffe S, Shlens J, Wojna Z. Rethinking the inception architecture for computer vision. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016;2818-2826.
- [19] Huang G, Liu Z, Van Der Maaten L, Weinberger KQ. Densely connected convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017; 4700-4708.
- [20] Zagoruyko S, Komodakis N. Wide residual networks. *arXiv*

- preprint arXiv. 2016;1605.07146.
- [21] Huang Y, Cheng Y, Bapna A, Firat O, Chen D, Chen M, Wu Y. Gpipe: Efficient training of giant neural networks using pipeline parallelism. *Advances in neural information processing systems*. 2019;32:103-112.
- [22] Tan M, Le Q. Efficientnet: Rethinking model scaling for convolutional neural networks. In *International Conference on Machine Learning*. PMLR. 2019;6105-6114.
- [23] Japkowicz N, Shah M. *Evaluating learning algorithms: a classification perspective*. Cambridge University Press, 2011.