

Comparative Analysis by Batch Size when Diagnosing Pneumonia on Chest X-Ray Image using Xception Modeling

Ji-Yul Kim¹, Soo-Young Ye^{2,*}

¹Daewoo general hospital

²Department of Radiology Catholic University of the Pusan

Received: August 10 2021. Revised: August 26, 2021. Accepted: August 31, 2021.

ABSTRACT

In order to quickly and accurately diagnose pneumonia on a chest X-ray image, different batch sizes of 4, 8, 16, and 32 were applied to the same Xception deep learning model, and modeling was performed 3 times, respectively. As a result of the performance evaluation of deep learning modeling, in the case of modeling to which batch size 32 was applied, the results of accuracy, loss function value, mean square error, and learning time per epoch showed the best results. And in the accuracy evaluation of the Test Metric, the modeling applied with batch size 8 showed the best results, and the precision evaluation showed excellent results in all batch sizes. In the recall evaluation, modeling applied with batch size 16 showed the best results, and for F1-score, modeling applied with batch size 16 showed the best results. And the AUC score evaluation was the same for all batch sizes. Based on these results, deep learning modeling with batch size 32 showed high accuracy, stable artificial neural network learning, and excellent speed. It is thought that accurate and rapid lesion detection will be possible if a batch size of 32 is applied in an automatic diagnosis study for feature extraction and classification of pneumonia in chest X-ray images using deep learning in the future.

Keywords: Deep learning, batch size, pneumonia, Pneumonia, Automatic Diagnosis of Pneumonia, chest X-ray imaging

I. INTRODUCTION

2019년 12월 중국의 우한시에서 원인 미상의 폐렴이 코로나바이러스라는 새로운 질병으로 확인이 된 후 전 세계적으로 코로나바이러스가 확산되었다^[1]. 그리고 세계보건기구는 코로나바이러스가 세계적 대유행(pandemic) 상황이라고 선언을 하였다^[1]. 코로나바이러스는 평균 14일간의 잠복기를 거치다가 발열, 오한, 근육통, 마른기침, 콧물, 인후통, 식욕부진, 설사 등의 증상을 수반하는 호흡기 증후군이다^[1]. 코로나바이러스는 동일한 호흡기 질환들에 비해 낮은 치사율이지만 폭발적인 전파력을 가지고 있다^[2]. 사물에 묻은 코로나바이러스는 수 시간 동안 생존이 가능하며 코로나바이러스를 만진

후 호흡기와 관련된 기관에 접촉이 된다면 감염을 일으킬 수 있다^[3].

딥러닝(deep learning) 기술은 음성인식과 영상인식을 비롯한 다양한 패턴인식 분야의 성능향상을 이끄는 중요한 인공지능기술이다^[4]. 그리고 의료영상의 다양한 분야에 적용이 되어 특정 질환별로 분류를 하거나, 병변의 위치 파악 및 장기를 분할하는 등의 다양한 형태로 이용되고 있다^[5].

이러한 딥러닝 기술의 여러 기법 중 합성곱 신경망(convolutional neural network)은 각 레이어(layer)마다 훈련 데이터(training data)로부터 최적의 필터를 설계하여 시험데이터(test data)를 분류하는 알고리즘을 의미하며 2차원 데이터의 원형을 이용할 수 있다는 특징을 가지고 있어서 이미지 분류에 널리

* Corresponding Author: Soo-Young Ye

E-mail: syye@cup.ac.kr

Tel: +82-51-510-0586

Address: Catholic University of Pusan, 57, Oryundae-ro, Geumjeong-gu, Busan, Korea 46252

사용된다^[6].

코로나바이러스의 진단을 위하여 흉부 X선 영상을 사용할 경우 많은 판독의를 비롯한 인적 및 의료 자원의 소모는 코로나바이러스 환자 증가로 인하여 더욱 가중될 수 있다^[7]. 그리고 원인 병원체를 확인하는데 2~3일 이상이 소요되며 실제 판독자 간 해석이 달라질 수도 있다^[8]. 그러므로 코로나바이러스의 감염 예방 및 환자를 적절히 치료하기 위해서는 흉부 X선 영상의 신속하고 정확한 진단이 필요하다^[9].

이러한 이유로 본 논문에서는 합성곱 신경망 기반의 Xception 모델을 이용하여 흉부 X선 영상의 폐렴 유무를 자동진단하고 모델링의 신속성과 정확성에 관여하는 배치 사이즈들을 비교 평가하여 향후 유사 연구 시 기초자료가 되고자 한다.

II. MATERIAL AND METHODS

1. 실험 재료

1.1 데이터셋 & 프로그래밍언어 & 프레임워크

Kaggle에서 제공하는 X선 영상 데이터셋(license: CC BY 4.0)를 사용하였다. 그리고 프로그래밍언어로 파이썬(Python, version 3.7.3 Python Software Foundation, Delaware, USA)을 이용하였으며, 프레임워크는 케라스(Keras, version 2.2.4)를 적용하였다. 케라스는 딥러닝을 쉽게 사용하기 위한 API들로 구성된 장점이 있다^[10].

1.2 가상환경 및 웹 기반의 파이썬 환경

가상환경으로 아나콘다(Anaconda, version 4.7.11, Continuum Analytics, Texas, USA)를 사용하였다. 그리고 웹 기반의 파이썬 환경으로 주피터 노트북(Jupyter notebook)을 사용하였다. 주피터 노트북은 파이썬 코드를 단계별 진행하고 데이터 분석에 쓰이며 사용 방법이 간단하다^[11].

1.3 합성곱 신경망

합성곱 신경망은 입력된 값으로부터 이미지의 고유한 특징 맵(feature map)을 만들고 신경망에 연결하여 이미지가 어떤 레이블에 속하는지를 분류

하는 역할을 한다^[12]. 입력 및 출력 부분에서 뉴런들이 학습해야 하는 가중치의 수가 작고 상대적으로 훈련 시간 대비 높은 성능을 가지며 이미지 데이터 분류 기법에 활발히 쓰이고 있다^[13].

1.4 Xception

합성곱 신경망기반의 Xception은 채널 방향의 컨볼루션만 진행하는 Pointwise convolution과 연산량을 기하급수적으로 줄여 실시간 동작이 가능한 Depthwise convolution을 완전히 분리하여 수행한다^[14]. Xception은 입력값(input)에 대해 1×1 컨볼루션 연산을 통해 채널 축소 후 3×3 컨볼루션 연산이 수행된다^[14]. 그 결과 Xception은 Inception V3와 비교하여 파라미터 및 용량의 증가 없이 성능이 크게 향상되었다^[15].

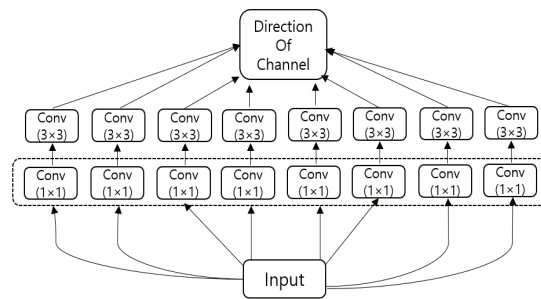


Fig. 1. Structure of Xception.

1.5 배치 사이즈(Batch size)

딥러닝 구현의 주요 파라미터 중에서 배치 사이즈는 한 번에 학습되는 데이터 개수를 의미하며, 신경망의 계산 속도와 정확성에 관여한다^[16]. 배치 사이즈는 딥러닝 모델링에 주요 설정값인 4, 8, 16, 32를 선정하였다.

2. 실험 방법

흉부 X선 영상을 이용한 폐렴 유무 자동진단을 위해 딥러닝 모델로 Xception을 선정하였으며, 동일 딥러닝 모델에 배치 사이즈를 각각 설정하여 모델링을 3회 실시하였다.

2.1 탐색적 데이터 분석

탐색적 데이터 분석은 그래프를 그려 시각화하

고 데이터를 직관화 하는 과정을 의미한다¹⁷⁾. 본 연구에서는 Kaggle에서 제공하는 흉부 X선 영상 데이터 세트 총 5,856장에 대한 탐색적 데이터 분석을 시행했다. 라벨링 1은 폐렴 흉부 X선 영상, 라벨링 0은 정상 흉부 X선 영상으로 하여 데이터 세트의 라벨링 수를 막대 그래프로 나타냈다. 학습 데이터 세트의 총 2,930장 중 라벨링 1로 분류된 폐렴 흉부 X선 영상은 2,140장, 0으로 분류된 정상 흉부 X선 영상은 790장으로 분류하였고, 검증 데이터 세트의 총 2,426장 중 폐렴 흉부 X선 영상은 1,833장, 정상 흉부 X선 영상은 593장으로 분류하였으며, 테스트 데이터 세트의 총 500장 중 폐렴 흉부 X선 영상은 300장, 정상 흉부 X선 영상은 200장으로 Table 1과 같이 분류하였다.

Table 1. Classification of dataset

	Labeling : 0 (Normal)	Labeling : 1 Pneumonia	Total
Training dataset	790	2140	2,930
Validation dataset	593	1833	2,426
Test dataset	200	300	500
Total	1,583	4,273	5,856

2.2 딥러닝 모델링

흉부 X선 영상의 신속성과 정확성을 평가하기 위해 Xception의 입력 이미지 사이즈는 224×224, 모델링의 epoch(시도회수)는 200회, 활성화 함수는 Sigmoid 함수, 최적화 하이퍼 파라미터를 적용하였다. 그리고 동일한 모델링에 배치 사이즈를 4, 8, 12, 16으로 적용하여 모델링을 각각 3회 실행하였으며 출력된 결과값을 산출하였다.

3. 성능 평가

Xception의 폐렴 유무 자동진단에 대한 배치 사이즈 별 성능을 정확하게 비교 평가하기 위해 딥러닝 모델링의 정확도, 손실 함수값, 평균 제곱 오차 (Mean Square Error; MSE), 1 epoch 당 학습 소요 시간과 Test Metric의 정확도, 정밀도, 재현율, F1-score, AUC score의 3회 모델링한 평균값을 산출한 후 성능평가를 수행하였다.

III. RESULT

본 연구에서는 Xception을 이용하여 흉부 X선 영상의 폐렴 유무 분류에 대한 각각의 배치 사이즈별 모델링의 신속성과 정확성을 평가하고자 하였다.

1. 딥러닝 모델링 평가

Xception을 이용한 각각의 배치 사이즈별 딥러닝 모델링에 대한 성능 평가 결과의 3회 평균값은 Table 2와 같다.

Table 2. Modeling result

Batch size	Validation			
	Accuracy (%)	Loss (%)	MSE (%)	1 epoch per training time[s]
4	98.48	3.87	2.05	76.33
8	98.74	2.94	1.67	60.67
16	98.79	2.88	1.68	52
32	98.98	2.68	1.54	48.67

1.1 정확도 평가

딥러닝 모델링의 정확도 평가는 Table 2와 같다. 배치 사이즈 4의 결과값은 98.48%, 배치 사이즈 8의 결과값은 98.74%, 배치 사이즈 16의 결과값은 98.79%, 배치 사이즈 32의 결과값은 98.98%의 정확도를 나타내었다. 이러한 결과로 배치 사이즈 32를 적용한 딥러닝 모델링이 흉부 X선 영상의 폐렴에 대한 특징 추출 및 분류에 있어 가장 우수한 모델링이라고 평가되었다.

1.2 손실함수 값 평가

딥러닝 모델링의 손실함수 값 평가는 Table 2와 같다. 배치 사이즈 4의 결과값은 3.87%, 배치 사이즈 8의 결과값은 2.94%, 배치 사이즈 16의 결과값은 2.88%, 배치 사이즈 32의 결과값은 2.68%의 손실함수 값을 나타내었다. 이러한 결과로 배치 사이즈 32를 적용한 딥러닝 모델링이 흉부 X선 영상의 폐렴에 대한 특징 추출 및 분류에 있어 인공지능경망의 학습상태가 가장 우수한 모델링이라고 평가되었다.

1.3 MSE 평가

딥러닝 모델링의 MSE 값 평가는 Table 2와 같다. 배치 사이즈 4의 결과값은 2.05%, 배치 사이즈 8의 결과값은 1.67%, 배치 사이즈 16의 결과값은 1.68%, 배치 사이즈 32의 결과값은 1.54%의 MSE 값을 나타내었다. 이러한 결과로 배치 사이즈 32를 적용한 딥러닝 모델링이 흉부 X선 영상의 폐렴에 대한 특징 추출 및 분류에 있어 가장 편차가 적은 모델링이라고 평가되었다.

1.4 1 epoch 당 학습 소요 시간

딥러닝 모델링의 1 epoch 당 학습 소요 시간 평가는 Table 2와 같다. 배치 사이즈 4의 결과값은 76.33sec, 배치 사이즈 8의 결과값은 60.67sec, 배치 사이즈 16의 결과값은 52sec, 배치 사이즈 32의 결과값은 48.67sec의 1 epoch 당 학습 소요 시간을 나타내었다. 이러한 결과로 배치 사이즈 32를 적용한 딥러닝 모델링이 흉부 X선 영상의 폐렴에 대한 특징 추출 및 분류에 있어 학습 속도가 가장 신속한 모델링이라고 평가되었다.

2. Test Metric 평가

Xception을 이용한 각각의 배치 사이즈별 테스트 모델링에 대한 Metric 평가 결과의 3회 평균값은 Table 3과 같다.

Table 3. Test Metric result

Batch size	Validation				
	Accuracy (%)	Precision (%)	Recall (%)	F1 -score	AUC -score
4	96.13	100.0	93.55	96.64	0.999
8	98.13	100.0	96.88	98.41	0.999
16	98.00	99.78	97.00	98.53	0.999
32	97.33	100.0	95.55	97.72	0.999

2.1 정확도(Accuracy) 평가

Metric 평가에서의 정확도란 예측값과 실제값과의 비교 시 정확한 정도를 나타내는 것으로, 가장 직관적으로 모델의 성능을 나타낼 수 있는 평가지표를 의미한다^[18].

Eq. (1)은 정확도를 구하는 수식^[19]이다.

테스트 metric의 정확도 평가는 Table 3과 같다. 배치 사이즈 4의 결과값은 96.13%, 배치 사이즈 8의 결과값은 98.13%, 배치 사이즈 16의 결과값은 98.00%, 배치 사이즈 32의 결과값은 97.33%의 정확도를 나타내었다. 이러한 결과로 배치 사이즈 8을 적용한 딥러닝 모델링이 흉부 X선 영상의 폐렴에 대한 특징 추출 및 분류에 있어 실측값과 예측값이 가장 정확한 모델링이라고 평가되었다.

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

TP : True Positive (진양성)

TN : True Negative (진음성)

FP : False Positive (위양성)

FN : False Negative (위음성)

2.2 정밀도(precision) 평가

Metric 평가에서의 정밀도란 예측 값과 실제 값과의 비교 시 Positive라고 예측한 결과값 중에서 실제 Positive의 비율을 의미한다^[18].

Eq. (2)는 정밀도를 구하는 수식^[19]이다.

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

테스트 metric의 정밀도 평가는 Table 3과 같다. 배치 사이즈 4의 결과값은 100%, 배치 사이즈 8의 결과값은 100%, 배치 사이즈 16의 결과값은 99.78%, 배치 사이즈 32의 결과값은 100.0%의 정밀도를 나타내었다. 이러한 결과로 모든 배치 사이즈를 적용한 딥러닝 모델링은 흉부 X선 영상의 폐렴에 대한 특징 추출 및 분류에 있어 실측값과 예측값의 정밀도가 매우 우수한 모델링이라고 평가되었다.

2.3 재현율(Recall) 평가

Metric 평가에서의 재현율이란 예측 값과 실제 값과의 비교 시 True인 것 중에서 모델이 Positive라고 예측한 것의 비율을 의미한다^[18].

Eq. (3)은 재현율을 구하는 수식^[19]이다.

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

테스트 metric의 재현을 평가는 Table 3과 같다. 배치 사이즈 4의 결과값은 93.55%, 배치 사이즈 8의 결과값은 96.88%, 배치 사이즈 16의 결과값은 97.00%, 배치 사이즈 32의 결과값은 95.55%의 재현율을 나타내었다. 이러한 결과로 배치 사이즈 16을 적용한 딥러닝 모델링이 흉부 X선 영상의 폐렴에 대한 특징 추출 및 분류에 있어 실측값과 예측값의 재현율이 가장 우수한 모델링이라고 평가되었다.

2.4 F1-score 평가

Metric 평가에서의 F1-score란 정밀도와 재현율의 평균을 의미한다^[18].

Eq. (4)는 F1-score를 구하는 수식^[19]이다.

$$F1 - score = 2 \times \frac{1}{\frac{1}{Precision} + \frac{1}{Recall}} \quad (4)$$

$$= 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

Precision : 정밀도

Recall : 재현율

테스트 metric의 F1-score 평가는 Table 3과 같다. 배치 사이즈 4의 결과값은 96.64, 배치 사이즈 8의 결과값은 98.41, 배치 사이즈 16의 결과값은 98.53, 배치 사이즈 32의 결과값은 97.72의 F1-score를 나타내었다. 이러한 결과로 배치 사이즈 16을 적용한 딥러닝 모델링이 흉부 X선 영상의 폐렴에 대한 특징 추출 및 분류에 있어 정밀도와 재현율의 평균에 대한 F1-score 값이 가장 우수한 모델링이라고 평가되었다.

2.5 AUC score 평가

AUC score란 ROC(receiver operating characteristic) curve의 아래 면적으로 데이터의 결과값이 우수할수록 면적은 넓어지며 1에 가까운 값을 가진다^[17]. 테스트 metric의 AUC score 평가는 Table 3과 같다. 배치 사이즈 4의 결과값은 0.99, 배치 사이즈 8의

결과값은 0.99, 배치 사이즈 16의 결과값은 0.99, 배치 사이즈 32의 결과값은 0.99의 AUC score를 나타내었다.

이러한 결과로 모든 배치 사이즈를 적용한 딥러닝 모델링은 흉부 X선 영상의 폐렴에 대한 특징 추출 및 분류에 있어 데이터의 결과값이 매우 우수한 모델링이라고 평가되었다.

IV. DISCUSSION

코로나바이러스의 전염을 막고 감염된 환자를 적절히 치료하기 위해서 흉부 X선 영상이 진단검사로서 중요한 역할을 담당하며 신속한 진단 및 정확한 평가가 필요하다^[20]. 그러나 코로나바이러스 환자의 증가로 인하여 판독 의를 비롯한 인적 및 의료 자원 소모가 더욱 가중될 수 있으며^[21], 판독자 간의 해석이 달라져 정확한 판독에 어려움이 있을수도 있다^[7]. 최근의 딥러닝은 기관이나 암 부위 등의 추출, 분할, 영상 정합 및 영상 검색 등의 의료영상 분석 분야에 널리 활용되고 있다^[22]. 그리고 딥러닝의 주요 파라미터 중 배치 사이즈는 의료영상의 특징 추출에 대한 정확성과 신속성의 주요 파라미터로 적용이 된다^[16].

Rachna Jain 등(2021)의 연구에서는 흉부 X선 영상의 폐렴에 대한 자동진단에 대하여 Inception V3, Xception, Inception-ResNet의 성능을 비교 평가하였으며, 연구 결과 Xception이 가장 우수한 성능을 나타내었다고 보고 하였다^[23]. 이러한 보고를 바탕으로 본 논문에서는 딥러닝 모델로 Xception을 선정하였으며, 동일한 모델링에 4, 8, 16, 32의 배치 사이즈를 각각 다르게 적용한 모델링을 3회 실시 한 후 결과의 평균값을 비교 평가하여 적합한 배치 사이즈를 제시하고자 하였다.

최보혜 등(2018)의 연구에서는 AlexNet과 VggNet을 이용하여 간암의 병변 분류에 대한 배치 사이즈와 epoch를 다르게 적용하여 성능을 평가하고자 하였으며, 연구 결과 배치 사이즈와 epoch가 증가할수록 성능이 우수하다고 평가되었다^[23]. 본 논문에서는 Xception을 이용한 폐렴의 특징 추출에 대한 배치 사이즈 별 성능을 비교 평가하였다. 그 결과 배치 사이즈 32를 적용한 모델이 가장 우수한 결과

를 나타내었으며 이러한 결과는 선행연구와 유사한 경향을 나타낸다고 판단이 된다.

배치 사이즈 32를 적용한 딥러닝 모델링이 다른 배치 사이즈를 적용한 딥러닝 모델링과의 비교 평가에서 높은 정확도, 안정적인 인공신경망 학습, 우수한 신속성을 나타내어 흉부 X선 영상의 폐렴 유무를 자동진단하고 모델링의 신속성과 정확성을 함께 수행할 수 있는 모델링이라고 판단하였다.

V. CONCLUSION

본 논문에서는 합성곱 신경망 기반의 Xception을 이용하여 흉부 X선 영상의 폐렴 유무를 자동진단하고 모델링의 신속성과 정확성에 관여하는 배치 사이즈를 비교 평가하고자 하였다. 실험결과 배치 사이즈를 다르게 적용한 각각의 전체 딥러닝 모델링은 성능평가 및 metric 평가를 통하여 폐렴의 특징 추출 및 분류에 대하여 우수한 성능을 나타내었다. 향후 딥러닝을 이용한 흉부 X선 영상의 폐렴에 대한 특징 추출 및 분류의 자동진단 연구 시 배치 사이즈를 32로 적용한다면 정확하면서도 신속한 병변 검출이 가능할 것이라고 사료된다. 그리고 이러한 결과는 흉부 X선 검사 시 폐렴의 자동진단을 위한 보조적인 방법으로 활용할 수 있을 것으로 기대된다.

Acknowledgement

본 연구는 2021년 부산가톨릭대학교 교내 학술연구과제로 수행되었다.

Reference

- [1] S. E. Park, "Coronavirus Infectious Disease-19 virus (SARS-CoV-2) characteristics, transmission and clinical picture", *Pediatric Infection & Vaccine*, pp. 27, 2019.
- [2] N. Petrosillo, G. Viceconte, O. Ergonul, G. Ippolito, E. Petersen, "COVID-19 SARS and MERS: are they closely related?", *Clinical Microbiology and Infection*, Vol. 26, No. 6, pp. 729-734, 2020.
<http://dx.doi.org/10.1016/j.cmi.2020.03.026>
- [3] B. H. Kang, Y. K. Choi, C. Y. Jeon, "The Effects of Qingfei Paidu Decoction on Coronavirus Disease-19 : A Narrative Review", *The Journal of Internal Korean Medicine*, Vol. 41, No. 3, pp. 424-433, 2020.
- [4] M. G. Kim, J. H. Yun, Y. W. Cho, "Deep learning in medical imaging", *Neurospine*, Vol. 16, pp. 657-668, 2019.
- [5] J. H. Kim "Imaging informatics : a new horizon for radiology in the era of artificial intelligence, big data, and data science", *Korean Journal of Radiology*, Vol. 80, pp. 176-201, 2019.
<http://dx.doi.org/10.3348/jksr.2019.80.2.176>
- [6] K. W. Yang, J. S. Kang, H. N. Lee, "Digital Walkie-Talkie Identification scheme based on Convolutional Neural Network", *The Journal of Korean Institute of Communications and Information Sciences*, pp. 1210-1211, 2018.
- [7] S. M. Lee, "Short-term Power Consumption Forecasting Based on IoT Power Meter with LSTM and GRU Deep Learning", *The Journal of The Institute of Internet, Broadcasting and Communication*, Vol. 19, No. 3, pp. 79-85, 2019.
- [8] F. Chollet, "Xception: Deep learning with depthwise separable convolutions", *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1-8, 2017.
- [9] J. Almirall, I. Bolibar, J. Vidal, "Epidemiology of community-acquired pneumonia in adults : a population-based study", *European respiratory journal*, Vol. 15, No. 4, pp. 757-763, 2000.
- [10] <https://keras.io/ko/>
- [11] A. Krizhevsky, G. Hinton, I. Sutskever, "ImageNet classification with deep convolutional neural networks," *Advances in Neural Information Processing (NIPS)*, pp. 1-9, 2012.
- [12] H. S. Kim, "Classification and Combination of Fashion Items Using CNN-Based Deep Learning", *Journal of Digital Contents Society*, Vol. 22, No. 3, pp. 475-482, 2021.
- [13] S. J. Kim, J. C. Yu, "COVID-19 Chest X-ray reading Technique based on Deep Learning", *Korean society of computer and information*, Vol. 29, No. 3, pp. 31-32, 2021.
- [14] <https://hwiyong.tistory.com/290>

- [15] <https://sotudy.tistory.com/14>
- [16] I. Kandel, C. Mauro, "The effect of batch size on the generalizability of the convolutional neural networks on a histopathology dataset", *ICT express*, Vol. 6, No. 4, pp. 312-315, 2020.
<http://dx.doi.org/10.1016/j.ict.2020.04.010>
- [17] <https://eda-ai-lab.tistory.com/13>
- [18] K. D. Song, M. C. Kim, S. H. Do, "The Latest Trends in the Use of Deep Learning in Radiology Illustrated Through the Stages of Deep Learning Algorithm Development", *Korean Journal of Radiology*, Vol. 80, No. 2, pp. 202-212, 2019.
<http://dx.doi.org/10.3348/jksr.2019.80.2.202>
- [19] J. Warrens, "Cohen's kappa can always be increased and decreased by combining categories", *Statistical Methodology*, Vol. 7, No. 6, pp. 673-677, 2010.
<http://dx.doi.org/10.1016/j.stamet.2010.05.003>
- [20] P. Cedric, P. Matthew, E. Nicole, "acute respiratory syndrome-related coronavirus 2 : a narrative review", *Annals of internal medicine*, Vol. 172, No. 11, pp. 726-734, 2020.
- [21] G. D. Rubin, C. J. Ryerson, L. B. Haramati, "The role of chest imaging in patient management during the COVID-19 pandemic", *Radiology*, Vol. 158, pp. 106-118 2020.
- [22] H. S. Lee, M. S. Park, J. M. Kim, "Deep Learning in Medical Imaging", *Korean society of imaging information in medicine*, Vol. 20, No. 1, pp. 13-18, 2014.
- [23] R. Jain, M. Gupta, S. Taneja, "Deep learning based detection and analysis of COVID-19 on chest X-ray images", *Applied Intelligence*, Vol. 51, No. 3, pp. 1690-1700, 2021.
- [24] B. H. Choi, Y. J. Kim, S. J. Choi, K. G. Kim, "Malignant and Benign Classification of Liver Tumor in CT according to Data pre-processing and Deep- running model", *Journal of Biomedical Engineering Research*, Vol. 39, pp. 229-236, 2018.
<https://doi.org/10.9718/JBER.2018.39.6.229>

Xception 모델링을 이용한 흉부 X선 영상 폐렴(pneumonia) 진단 시 배치 사이즈별 비교 분석

김지율¹, 예수영^{2,*}

¹대우병원 영상의학과

²부산가톨릭대학교 방사선학과

요 약

흉부 X선 영상의 폐렴을 신속하고 정확하게 진단하기 위하여 동일한 Xception 딥러닝 모델에 배치 사이즈를 4, 8, 16, 32로 다르게 적용하여 각각 3회의 모델링을 실시하였다. 그리고 성능평가 및 metric 평가에 대한 결과값을 3회 평균값으로 산출하여 배치 사이즈별 흉부 X선 영상의 폐렴 특징 추출과 분류의 정확도 및 신속성을 비교 평가하였다.

딥러닝 모델링의 성능평가 결과 배치 사이즈 32를 적용한 모델링의 경우 정확도, 손실함수 값, 평균제곱 오차, 1 epoch 당 학습 소요 시간의 결과가 가장 우수한 결과를 나타내었다. 그리고 Test Metric의 정확도 평가는 배치 사이즈 8을 적용한 모델링이 가장 우수한 결과를 나타내었으며, 정밀도 평가는 모든 배치 사이즈에서 우수한 결과를 나타내었다. 재현율 평가는 배치 사이즈 16을 적용한 모델링이 가장 우수한 결과를 나타내었으며, F1-score는 배치 사이즈 16을 적용한 모델링이 가장 우수한 결과를 나타내었다. 그리고 AUC score 평가는 모든 배치 사이즈의 결과가 동일하였다. 이러한 결과를 바탕으로 배치 사이즈 32를 적용한 딥러닝 모델링이 높은 정확도, 안정적인 인공신경망 학습 및 우수한 신속성의 결과를 나타내었다. 향후 딥러닝을 이용한 흉부 X선 영상의 폐렴에 대한 특징 추출 및 분류에 관하여 자동진단 연구 시 배치 사이즈를 32로 적용한다면 정확하면서도 신속한 병변 검출이 가능할 것이라고 사료된다.

중심단어: 딥러닝, 배치 사이즈, 폐렴, 폐렴 유무 자동진단, 흉부 X선 영상

연구자 정보 이력

	성명	소속	직위
(제1저자)	김지율	대우병원 영상의학과	방사선사
(교신저자)	예수영	부산가톨릭대학교 방사선학과	부교수