

<원저>

핵의학 감마카메라 정도관리의 딥러닝 적용

정의환¹⁾·오주영¹⁾·이주영²⁾·박훈희³⁾¹⁾(주)키즈밋·²⁾송호대학교 방사선과·³⁾신구대학교 방사선과

Deep Learning Application of Gamma Camera Quality Control in Nuclear Medicine

Euihwan Jeong¹⁾·Joo-Young Oh¹⁾·Joo-Young Lee²⁾·Hoon-Hee Park³⁾¹⁾KISMITS²⁾Department of Radiological Technology, Songho University³⁾Department of Radiological Technology, Shingu College

Abstract In the field of nuclear medicine, errors are sometimes generated because the assessment of the uniformity of gamma cameras relies on the naked eye of the evaluator. To minimize these errors, we created an artificial intelligence model based on CNN algorithm and wanted to assess its usefulness. We produced 20,000 normal images and partial cold region images using Python, and conducted artificial intelligence training with Resnet18 models. The training results showed that accuracy, specificity and sensitivity were 95.01%, 92.30%, and 97.73%, respectively. According to the results of the evaluation of the confusion matrix of artificial intelligence and expert groups, artificial intelligence was accuracy, specificity and sensitivity of 94.00%, 91.50%, and 96.80%, respectively, and expert groups was accuracy, specificity and sensitivity of 69.00%, 64.00%, and 74.00%, respectively. The results showed that artificial intelligence was better than expert groups. In addition, by checking together with the radiological technologist and AI, errors that may occur during the quality control process can be reduced, providing a better examination environment for patients, providing convenience to radiologists, and improving work efficiency.

Key Words: Nuclear medicine, Quality Control, AI, CNN, Deep Learning

중심 단어: 핵의학, 정도관리, 인공지능, 콘볼루션 신경망, 딥러닝

I. 서 론

인공지능(Artificial Intelligence; AI)이 4차 산업 혁명의 견인차 역할을 할 것으로 기대되면서 최근 많은 영역에서 관련 연구가 활발히 진행되고 있다. 의료영상 분야에서도 인공지능을 활용한 진단 시스템에 대한 연구가 영상의학, 핵의학 및 종양학을 중심으로 활발히 이루어지고 있다 [1]. 특히 인공지능을 이용한 의료영상 분석 기술은 특징을 추출하는 방법으로 최근 딥러닝(deep learning)에 기반을 둔 학습모델이 주목받고 있다[2].

딥러닝은 인간의 신경계를 모방한 인공신경망(Artificial

Neural Network; ANN)의 일종으로 여러 층으로 쌓인 신경망 내부에서 각종 매개변수(parameter)를 갱신하며 학습을 수행하는 기법이며[3], 인간의 시각 처리 과정을 모방하기 위해 개발된 컨볼루션신경망(Convolution Neural Network; CNN)은 영상 인식 분야에 다양하게 적용하는 기술이다[4].

특히 의료영상 분야에서는 방사선 장비의 발전으로 영상의 화질, 정확도 등이 상승하였고, 수요 또한 증가하고 있으며, 인체의 형태학적인 정보와 생물학적, 기능적 정보를 얻을 수 있는 핵의학의 임상적 중요성이 대두됨에 따라 많은 관심이 집중되고 있다[5, 6]. 이러한 핵의학 검사 시 장비의 관리가 면밀하게 이루어지지 않는다면 장비문제 뿐만 아니

라 이로 인하여 불필요한 피폭선량을 증가시킬 가능성이 높아진다[7, 8]. 따라서 장비를 최적의 성능으로 유지 시킬 수 있는 정확한 정도 관리는 핵의학 의료영상 분야에서 중요하게 다루어진다[9].

대한핵의학회는 영상의 질관리를 위하여 감마카메라 정도관리 지침서를 발표하여 장시야 균일성(uniformity), 공간분해능(spatial resolution), 에너지 분해능(energy resolution), 직선성(linearity), 계수율 손실(count rate loss) 등의 항목을 통해 평가하고 있으며, 항목 중 가장 범용적으로 사용되는 균일도 평가항목은 적분균일성(Integral Uniformity; IU)와 미분균일성 (Differential Uniformity; DU)를 이용한다. 이 두 값은 전체적인 영상에 대한 평가로서 국소적으로 비정상적인 열소(hot region)나 냉소(cold region)를 특정하지 못한다는 단점이 있다[10].

현재 임상에서는 장비의 균일성을 측정하기 위해 국소적으로 발생하는 열소나 냉소부위를 평가하기 위해 일반적으로 방사선사의 육안적인 평가에 의존하고 있다[11]. 따라서 본 연구에서는 인공지능을 이용할 경우, 육안으로 찾기 힘든 국소적 냉소 부분을 발견하는 성능을 확인하여 정도관리에 적용이 가능성을 평가하는데 목적을 두었다.

II. 대상 및 방법

1. 프로그램

프로그래밍 언어로 파이썬(Python)과 딥러닝 프레임워크로 파이토치(Pytorch)를 사용하였으며, 딥러닝 모델로 Resnet18모델을 적용하였다. 국소적 냉소 부위를 인위적으로 제작하기 위해서 파워포인트(Power Point)를 사용하였으며, 이미지 합성 작업을 위해 파이썬의 필로우(Pillow) 모듈을 사용하였다. 인공지능 학습을 위해 구글 플랫폼의 코랩(colaboratory)를 사용했다.

2. 정상 정도관리 영상 제작

정상 정도관리 영상은 실제 핵의학 영상이 아닌 핵의학 정도관리 영상이 생성되는 과정을 코드로 작성하여 제작하였고, 감마카메라(Siemens E.Cam single head gamma camera)에서 코발트-60(⁶⁰Co) 장선원으로 200,000 카운트(counts)의 외인성(extrinsic) 정도관리 영상을 기반으로 제작하였다(Fig. 1). 이 중 50개의 영상을 참고하였으며, 영상 크기(size)는 238 × 175, 영상의 픽셀(pixel)의 평균값은 98이었다. 또한, 핵의학 영상에서 방사선이 계수되는 부분이 각

픽셀값이 255가 되도록 한 후 1/2 확률로 포아송(poisson) 분포를 따라 각 픽셀값이 감소하도록 진행하였다(Fig. 1). 이와 같은 방식으로 총 20,000개의 정도관리 영상을 제작하고, 영상을 실제 정도관리 영상과 히스토그램을 비교한 결과 유사한 그래프를 나타내어 유사한 영상임을 확인하였다(Fig. 2).

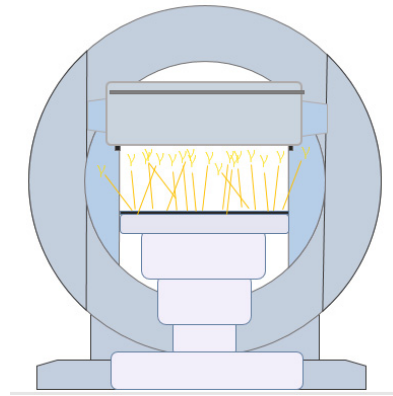


Fig. 1. An image was acquired using Siemens E.cam single head gamma camera.

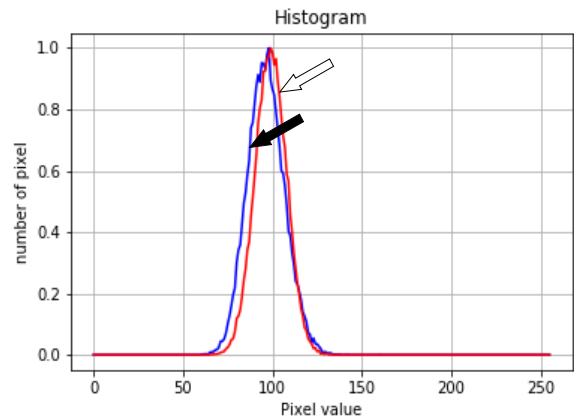


Fig. 2. Histogram of nuclear medicine Quality Control(QC) image(blue line graph) and created image(red line graph)

3. 국소적 냉소 영상 제작

자체 제작한 20,000개의 정상 정도관리 영상 중, 10,000개의 영상에 파워포인트로 제작한 원형의 영상을 필로우(Pillow)로 합성하여 국소적 냉소 영상을 제작하였다. 서로 다른 라인 프로파일(line profile)을 가진 1, 2, 3, 4단계의 원형 영상을 제작하였고(Fig. 3), 4단계의 원형 영상의 투명도와 이미지 크기를 다양하게 설정하여 중앙에 합성하였다. 영상을 확인하여 전문가의 육안으로 판단하기 힘든 지점을 선택하여 따라 투명도(α 값)를 0.02, 0.04, 0.06, 0.08로

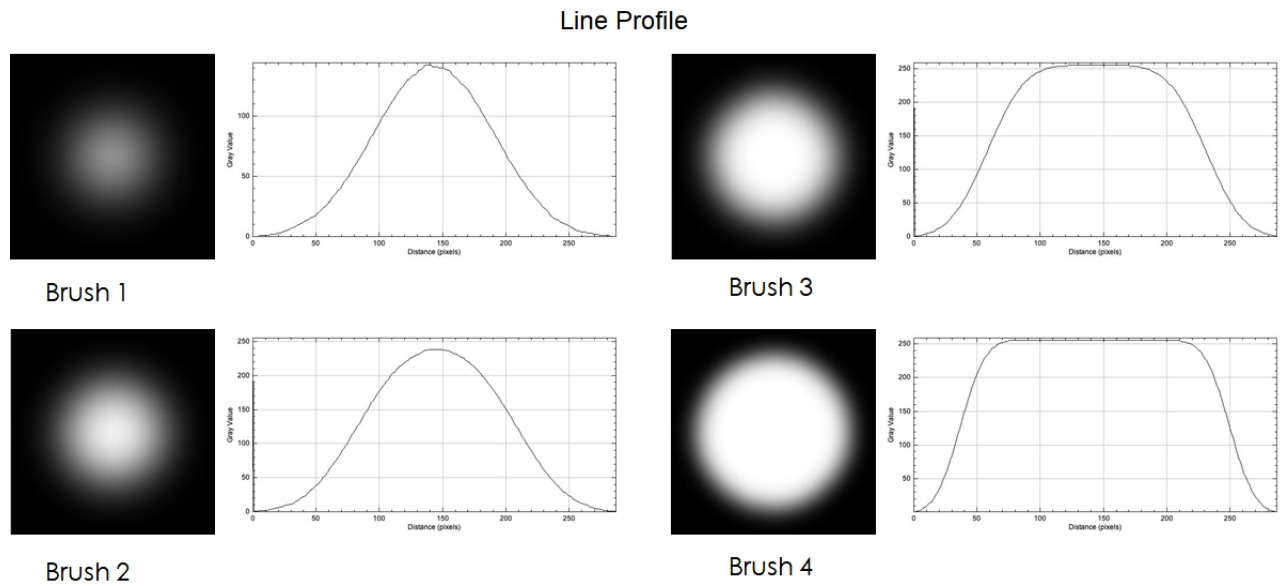


Fig. 3. Line profiles of 4 brushes-

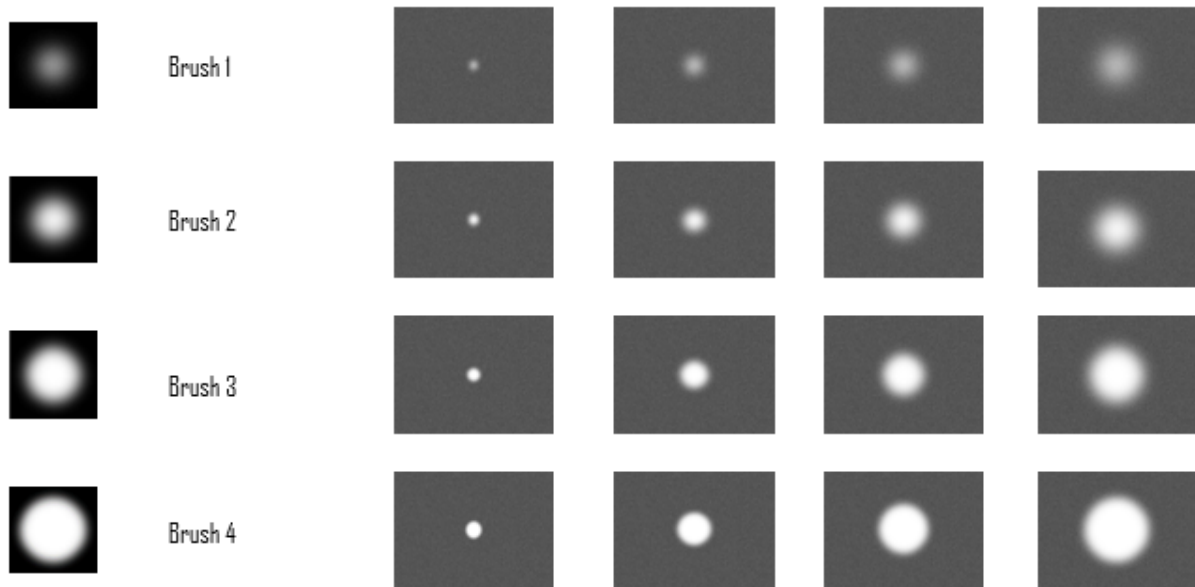


Fig. 4. Images are 4 brushes with different transparency and size.

조합하였으며, 원의 크기는 5×5 , 10×10 , 15×15 , 20×20 , 25×25 , 30×30 , 35×35 , 40×40 , 45×45 , 50×50 , 55×55 , 60×60 , 65×65 , 70×70 , 75×75 의 순차적으로 다양한 값들을 조합하였다(Fig. 4). 10,000개의 정상영상에 무작위 위치로 합성하여 10,000개의 국소적 냉소 영상을 제작하였으며, 실제 훈련데이터로 제작된 영상은 알파값을 조정하여 육안으로 인식하기 어려운 난이도로 진행하였다(Fig. 5).

4. Resnet18 모델 적용

CNN 모델은 Resnet18을 사용하였으며(Fig. 6), 18개의 계층으로 된 컨볼루션 신경망 모델을 적용하였다[12]. Resnet18 모델의 특징은 최소 7×7 의 컨볼루션 레이어가 있고, 이후 3×3 의 컨볼루션레이어를 쌓으며, 2개의 층마다 해당 출력값에서 2개 층의 전 출력값을 제거함으로써 나머지(residual)를 학습하지만, 본 연구에서는 Resnet은 1,000개의 완전연결계층(Fully-Connected Layer)로 되어 해당 부분을 2개

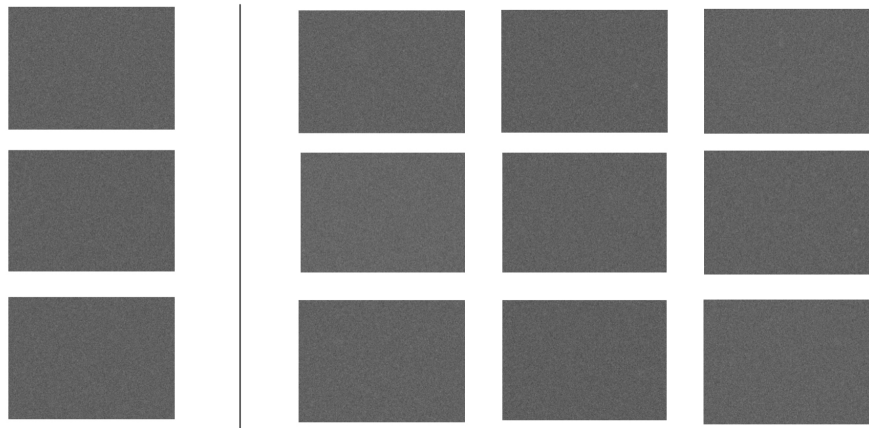


Fig. 5. Normal image(left) versus Cold region image used as AI data set(right)

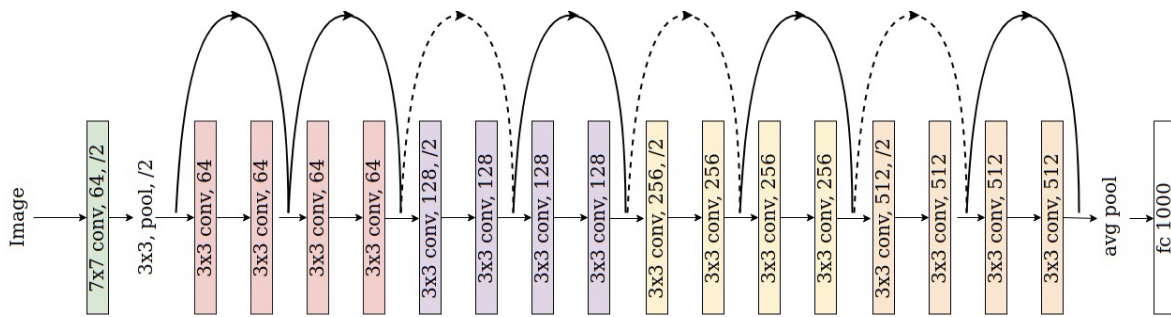


Fig. 6. Resnet18 Architecture

로 분류하는 층으로 변경하여 소프트맥스(softmax)층에서 값이 0.5 이상일 경우는 국소적 냉소 영상, 0.5 미만인 경우는 정상 영상으로 분류하는 모델로 변경하도록 하였다.

5. 영상학습

제작한 정상 영상 10,000개와 국소적 냉소 영상 10,000개를 데이터셋(data set)으로 선정하였고, 총 20,000개의 영상 중 70%에 해당하는 14,000개를 인공지능 훈련(training set)으로 사용했다. 나머지 6,000개는 인공지능 학습검증(validation set)으로 사용하였고, 50 에폭(epoch)으로 학습을 진행했다.

6. 인공지능 평가

인공지능 평가를 위하여 14,000개의 인공지능 훈련으로 학습한 영상을 6,000개의 학습검정으로 평가했다. 6,000장을 훈련된 인공지능에 입력하여 정확도(accuracy), 특이도(specificity), 민감도(sensitivity)의 값을 확인하였다.

7. 전문가 집단 평가

인공지능과 전문가 집단의 영상 평가 결과를 비교하기 위

해서 제작한 정상 영상 100개와 국소적 냉소 영상 100개, 총 200개로 분석하였다. 전문가 집단으로는 방사선 영상전문가 5명을 대상으로 하여 오차 행렬(confusion matrix)값을 구하고, 인공지능과 전문가 집단의 정확도, 특이도, 민감도를 비교하였다. 또한, 인공지능과 전문가 집단의 ROC(Receiver Operating Characteristic)곡선을 활용하여 AUC(Area Under Curve)값으로 인공지능과 전문가 집단의 판단력 차이를 비교하였다.

III. 결 과

1. 인공지능 평가

구글의 코랩(collaboratory)을 이용하여 Resnet18의 모델로 14,000장의 인공지능 훈련과 학습검정으로 학습한 결과, 훈련 시간은 1시간 소요되었다. 인공지능 학습 결과, 6,000개의 학습검정 기준으로 정확도는 95.01%를 나타냈다. 특이도와 민감도는 각각 92.30%와 97.73%다[Table 1]. 3,000개의 국소적 냉소 영상 중 231개를 정상 영상으로 판단했지만, 나머지 3,000개의 정상 영상 중에서는 68개를 국소적 냉소

Table 1. AI's confusion matrix evaluated with 6,000 verification images

	TP	FP	FN	TN	Accuracy	Specificity	Sensitivity
AI (6,000 Images)	2,769	68	231	2,932	95.01%	92.30%	97.73%

* TP(True Positive), FP(False Positive), FN(False Negative), TN(True Negative)

영상으로 판단했다. 정상영상의 경우 포아송분포에 의해 생긴 노이즈의 발생으로 미세하게 균질하지 못한 부분을 국소적 냉소영상으로 판단하였다. 국소적 냉소 영상 중에서 알파값이 높고, 브러쉬 사이즈가 5 × 5로 작은 경우 정답부분을 알아도 판단하기가 어려웠으며, 데이터가 데이터셋에 포함이 되어 위음성(false negative)이 낮았다.

2. 전문가 집단 평가

인공지능과 전문가 집단의 비교를 위해 200개의 영상으로 비교한 결과, 인공지능의 정확도, 특이도, 민감도는 94.00%, 91.00%, 97.00%였으며, 전문가 집단 5명의 평균은 각각

69.00%, 64.00%, 74.00%로 인공지능보다 낮은 수치였다 [Table 2]. 결과적으로 인공지능이 육안보다 세밀한 부분까지 판단하여 정상과 비정상을 분별하였다.

200개의 영상을 판단하는데 전문가 집단의 경우 30분 소요되었고, 인공지능의 경우 1초가 소요되었다. 많은 양의 데이터를 판단하는 경우, 인공지능은 같은 기준에서 지속적으로 판단할 수 있지만, 전문가 집단은 판단의 기준에서 개인적인 편차가 발생하였다.

ROC 곡선 확인 결과, 인공지능의 AUC는 0.9863으로 인공지능이 정확히 판단하였고, 정확도와 판단력이 전문가 집단 분포가 인공지능보다 낮았다[Fig. 7].

Table 2. AI and 5 experts' confusion matrix evaluated with 200 images

	TP	FP	FN	TN	Accuracy	Specificity	Sensitivity
AI's results	91	3	9	97	94.00%	91.00%	97.00%
5 expert's average	64	26	36	74	69.00%	64.00%	74.00%
Expert 1	29	4	71	96	63.00%	29.00%	96.00%
Expert 2	73	37	27	63	68.00%	73.00%	63.00%
Expert 3	53	7	47	93	73.00%	53.00%	93.00%
Expert 4	84	48	16	52	68.00%	84.00%	52.00%
Expert 5	81	34	19	66	74.00%	81.00%	66.00%

* TP(True Positive), FP(False Positive), FN(False Negative), TN(True Negative)

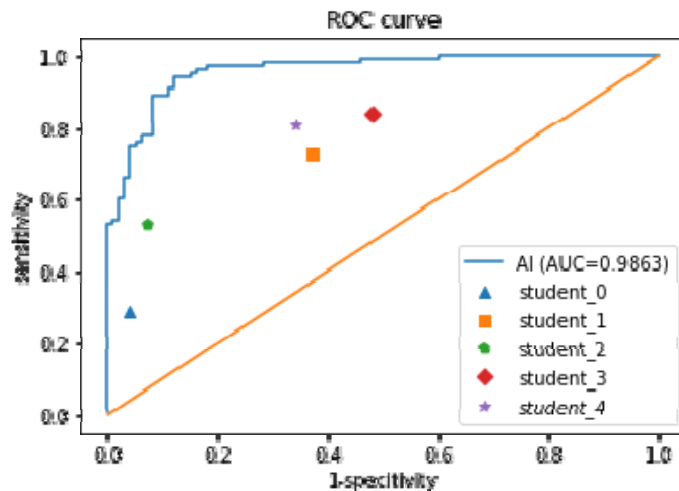


Fig. 7. The ROC curve of AI and experts. It can be seen that the accuracy of the AI is higher because the experts' distribution is located below the AI.

IV. 고 찰

의료 분야에서는 골연령 판단, 흉부 X-ray 영상 판단, 뇌 질환 판단 등 의료영상의 세부적인 질병에 대해 찾아내는 진단 보조 소프트웨어들이 상용화되고 있다[13].

본 연구는 인공지능을 이용할 경우, 육안으로 찾기 힘든 국소적 냉소 부분을 발견하는 성능을 확인하여 정도관리에 적용이 가능성을 평가하는 목적을 두었으나, 제한점으로 프로그램 학습 및 검증 시 사용되었던 영상은 실제 임상 영상이 아닌 파이썬 코드와 브러쉬로 자체 재현한 인위적인 영상을 사용하여 영상의 농도(contrast)만을 기준으로 영상의 적합 또는 부적합 판정을 확인하였다.

인공지능의 훈련용 영상의 다양성이 부족할 경우 활성화 되는 뉴런 셋(Neuron set)이 부족해져 실제 환경에 적용할 시, 일반화의 오류가 발생할 수 있다[14]. 본 연구에서 브러쉬의 종류는 4가지만 사용되어 이에 따른 결과의 다양성이 부족하였고, 다양한 형태의 냉소 부위 적용이 어려워 실제 영상 및 다른 모양의 합성 영상에 대해서 다양성의 확보가 필요할 것으로 여겨진다.

제작된 프로그램을 실제 임상 환경마다 사용되는 장비 혹은 정도관리 기준의 차이에 따른 오차가 발생할 가능성이 있다[15]. 따라서 추후 유사 연구가 진행될 경우, 균일도 정도관리에서 국소적 냉소영상 이외에 발생할 수 있는 크리스탈(crystal)의 파손, 국소적 열소를 추가하여 영상을 다양하게 적용한다면 본 연구의 인공지능보다 신뢰성과 판단능력이 향상될 것이다. 핵의학 이외의 다른 진단 영상기기에서도 다양한 정도관리 영상이 학습되어 미세한 이상 부위를 발견하고 사전에 해결할 수 있다면, 임상에서도 인공지능을 이용한 정도관리가 유용하게 사용될 것이라 기대한다.

정도관리에서 영상만을 판단하여 수행하기에는 한계가 있기 때문에, 측정, 모니터링, 피드백 및 최적화 제어 알고리즘 등이 수반되어야 체계적인 정도관리를 대체할 수 있는 인공지능으로 발전 할 수 있다[16]. 정도관리 인공지능이 이상 부위만을 찾아주는 한정적인 인공지능에서 해결방안을 제시 가능한 인공지능으로 발전하면 보다 다양한 분야에서 인공지능을 통해 정도관리를 진행할 수 있을 것으로 판단된다.

V. 결 론

본 연구에서는 핵의학 정도관리 평가항목 중 장시야 균일성 평가에서 발생할 수 있는 냉소 부위를 보다 빠르고 쉽게

발견하고 해결하고자 개발한 코드를 인공지능을 통해 학습시켜 연구를 진행하였다. 그 결과 인공지능은 정확도 94.0%, 전문가 집단은 정확도 69.0%로 나타남으로써 인공지능이 더 높은 정확도를 보였다. 200장의 영상을 판단하는데 전문가 집단에서는 30분 소요되었지만, 인공지능은 1초에 판단하였고, 전문가 집단보다 정확한 결과를 도출해냈다.

본 연구는 도출된 결과를 통해 정도관리에 있어 인공지능에 대한 적용 가능성을 제시하였다. 추후 다양한 연구가 활발하게 진행된다면 정도관리에서 발생할 수 있는 오류를 감소시켜 진단 가치가 높은 영상을 제공할 수 있을 것이라 사료된다.

REFERENCES

- [1] Choi KT. Real-time artificial neural network for high-dimensional medical image. *Journal of the Korean Society of Radiology*. 2016;10(8):637-43.
- [2] Ravi D, Wong C, Deligianni F, et al. Deep learning for health informatics. *IEEE Journal of Biomedical and Health Informatics*. 2017;21(1):4-21.
- [3] LeCun Y, Boser B, Denker JS, et al. Backpropagation applied to handwritten zip code recognition. *Neural Computation*. 1989;1(4):541-51.
- [4] Li J, Mi Y, Li G, Ju Z. CNN-based facial expression recognition from annotated RGB-D images for human-robot interaction. *International Journal of Humanoid Robotics*. 2019;16(4):1941002.
- [5] Choi JG, Lee BI. Research for development of standardized system for quality control of nuclear medicine imaging equipments. *The Annual Report of Korea Food & Drug administration (ABSTRACT)*. 2008;12(11-14700000-000071-10):859-60.
- [6] Choe JG, Joh CW. Survey of current status of quality control of gamma cameras in republic of Korea. *Nuclear Medicine and Molecular Imaging*. 2008;42(4):314-22.
- [7] Choi WC. Actual condition of computerized tomography system in use in Seoul and projection evaluation using AAPM CT Phantom [master's thesis]. Korea University, Korea; 2009.
- [8] Lim JJ, Kim HK, Kim JP, Jo SW, Kim JE. Evaluation of radiation exposure to medical staff except nuclear

- medicine department. The Korean Journal of Nuclear Medicine Technology. 2016;20(2):32-5.
- [9] Noh SS, Um HS, Kim HC. Development of automatized quantitative analysis method in CT images evaluation using AAPM phantom. The Institute of Electronics and Information Engineers. 2014;51(12):163-73.
- [10] The International Atomic Energy Agency. Quality control of nuclear medicine instruments. IAEA-TECDOC-317; 1984.
- [11] National Physical Laboratory. Protocol for establishing and maintaining the calibration of medical radionuclide calibrators and their quality control. A National Measurement Good Practice Guide; 2006:93.
- [12] He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. arXiv(Computer Vision and Pattern Recognition). 2015;1512.03385.
- [13] Park HK. Artificial intelligence (AI) health care industry status and trends. Convergence Focus, 2019;148:2-9.
- [14] Lee HH. CNN Generalization Error Evaluation Method [master's thesis]. Pusan University, Korea; 2020.
- [15] Park JK. Medical clinics' quality management of X-ray units in Gyeongbuk area. The Journal of the Korea Contents Association. 2010;10(9):267-75.
- [16] Larson DB, Boland GW. Imaging quality control in the era of artificial intelligence. Journal of the American College of Radiology. 2019;16(9):1259-66.

구분	성명	소속	직위
제1저자	정의환	KISMITS	방사선사 / 연구원
공동저자	오주영	KISMITS	방사선사 / 대표
교신저자	이주영	송호대학교	조교수(박사)
단독	박훈희	신구대학교	부교수(박사)