



The Latest Trends in the Use of Deep Learning in Radiology Illustrated Through the Stages of Deep Learning Algorithm Development

딥러닝 알고리즘 개발과정을 통해 본 영상의학분야에서 딥러닝의 최신 경향

Kyoung Doo Song, MD^{1,2} , Myeongchan Kim, MD¹, Synho Do, PhD^{1*} 

¹Department of Radiology, Massachusetts General Hospital, Boston, MA, USA

²Department of Radiology, Samsung Medical Center, Seoul, Korea

Received January 15, 2019
Accepted March 22, 2019

***Corresponding author**

Synho Do, PhD
Department of Radiology,
Massachusetts General Hospital,
25 New Chardon Street Suite
400B, Boston, MA 02114, USA.


Tel 1-617-643-0247

Fax 1-617-724-5597

E-mail sdo@mgh.harvard.edu

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<https://creativecommons.org/licenses/by-nc/4.0>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

ORCID iDs

Synho Do 

<https://>

orcid.org/0000-0001-6211-7050

Kyoung Doo Song 

<https://>

orcid.org/0000-0002-2767-3622

Recently, considerable progress has been made in interpreting perceptual information through artificial intelligence, allowing better interpretation of highly complex data by machines. Furthermore, the applications of artificial intelligence, represented by deep learning technology, to the fields of medical and biomedical research are increasing exponentially. In this article, we will explain the stages of deep learning algorithm development in the field of medical imaging, namely topic selection, data collection, data exploration and refinement, algorithm development, algorithm evaluation, and clinical application; we will also discuss the latest trends for each stage.

Index terms Artificial Intelligence; Deep Learning; Algorithms

서론

최근 인공지능(artificial intelligence)은 지각정보를 해석하는데 있어 상당한 진보를 이루었으며, 이를 통해 기계는 매우 복잡한 데이터를 보다 잘 해석할 수 있게 되었다. 실제로 인공지능은 웹 검색, 자동차의 자율주행, 자연어 처리 및 컴퓨터 비전(computer vision)과 같은 분야의 상당한 발전을 이끌고 있다. 최근 인공지능의 여러 기술 중 가장 각광을 받고 있

는 것은 딥러닝(deep learning)으로 이는 인간의 뇌에서 영감을 받은 신경 네트워크 구조를 기반으로 하는 기계학습(machine learning)의 한 분야이다. 최근 몇 년간 이러한 딥러닝 기술의 의료 및 생물의학(biomedical) 연구분야로의 적용이 기하급수적으로 증가하고 있다. 의료분야의 논문을 검색할 수 있는 PubMed를 이용하여 최근 몇 년간의 딥러닝 관련 연구를 대략적으로 조사해보면 이러한 증가 추세를 확인할 수 있다(Fig. 1). 이러한 의료분야 전반으로의 인공지능의 확산은 영상의학 분야에서의 성공이 계기가 되었으며, 영상의학 분야는 여전히 기술적인 부분 뿐 아니라 이와 관련된 가이드라인의 제정에 있어서도 주도적인 역할을 하고 있다.

흔히 딥러닝에 대해서 설명할 때 알고리즘(algorithm)에 초점을 맞추게 된다. 어떠한 문제를 딥러닝을 이용하여 해결했을 때 결국 결과물은 개발된 알고리즘이기 때문일 것이다. 하지만 딥러닝을 보다 명확하게 이해하고 이를 활용하기 위해서는 단순히 데이터를 입력하고 변수를 조정하여 결과를 산출하는 과정이 아닌 딥러닝에 적합한 주제를 선정하여 데이터를 모으는 것부터 만들어진 알고리즘을 검증하고 실제 임상에 적용하는 여러 단계를 전반적으로 이해하는 것이 필수적이다. 이번 기고문에서는 저자들이 시행했던 딥러닝을 이용한 두개 내 출혈(intracranial hemorrhage)을 진단하는 연구를(1) 통해 딥러닝 알고리즘을 개발하는 단계에 대해서 살펴보고 각각의

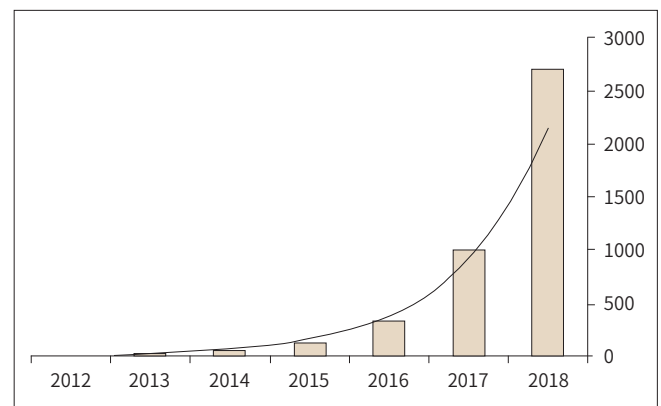
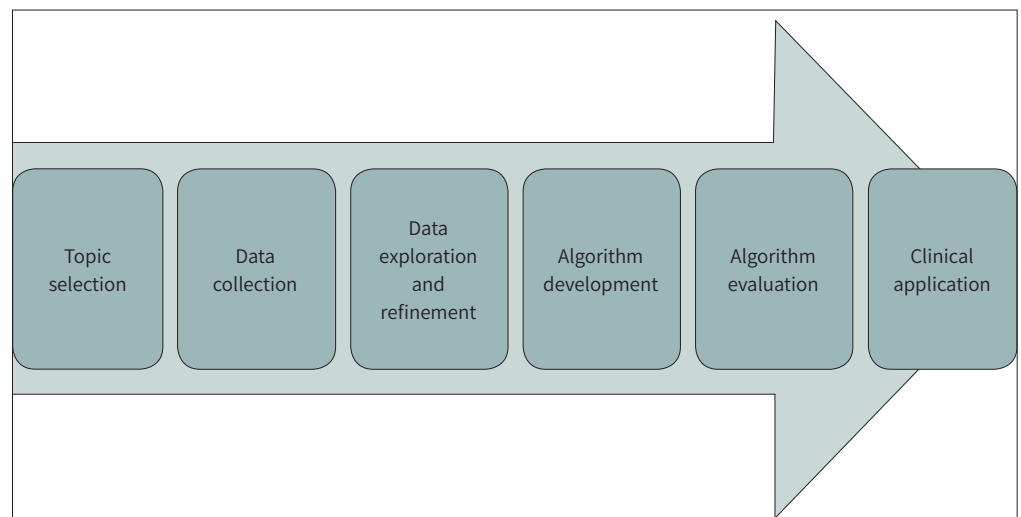


Fig. 1. Annual trend in the number of papers related to deep learning in the medical field. Results from PubMed in December 2018 using 'deep learning' and 'convolutional' search terms.

Fig. 2. Stages of deep learning algorithm development.



단계에서 최신의 동향을 알아보고자 한다.

딥러닝 알고리즘 개발과정

딥러닝 알고리즘 개발은 주제 선정, 데이터 수집, 데이터 탐색 및 정제, 알고리즘 개발, 알고리즘 평가 그리고 임상적용의 단계를 거치게 된다(Fig. 2).

1단계: 주제 선정

주제 선정이란 해결하고자 하는 문제를 찾는 것이다. 의료분야에는 수없이 많은 해결해야 할 문제가 있으나 당연히 모든 문제를 딥러닝으로 해결할 수는 없기에 이러한 해결과제 중에서 딥러닝에 적합한 과제를 찾는 것이 중요하다. 먼저 의료영상분야에서 최근에 발표된 논문들을 통해서 딥러닝에 적합한 과제는 어떠한 것인지에 대해서 알아보자. 2017년도에 Litjens 등(2)은 의료영상분야에서 딥러닝을 이용한 연구에 대해서 조사한 결과를 발표하였다. 이중 영상의학분야에 한정하

Table 1. Latest important papers on deep learning in the field of medical imaging

Specialty	Images	Publication
Radiology/Neurology	CT head, acute neuro events	Titano, Nature Medicine, 2018
	CT head for brain hemorrhage	Arbabshirani, NPJ (Nature) Digital Medicine, 2018
	CT head for trauma	Chilamkurthy, Lancet, 2018
	CXR for metastatic lung nodules	Nam, Radiology, 2018
	CXR for multiple findings	Singh, PLOS One, 2018
Pathology	Breast cancer	Bejnordi, JAMA, 2017
	Lung cancer (+ driver mutation)	Coudray, Nature Medicine, 2018
	Brain tumors (+ methylation)	Capper, Nature, 2018
	Breast cancer metastases*	Steiner, Am J Surgical Pathology, 2018
	Breast cancer metastases	Liu, Arch Path Lab Med, 2018
Dermatology	Skin cancers	Esteva, Nature, 2017
	Melanoma	Haenssle, Annals of Oncology, 2018
	Skin lesions	Han, Journal of Investigative Dermatology, 2018
Ophthalmology	Diabetic retinopathy	Gulshan, JAMA, 2016
	Diabetic retinopathy*	Abramoff, NPJ (Nature) Digital Medicine, 2018
	Diabetic retinopathy*	Kanagasingam, JAMA Open 2018
	Congenital cataracts	Long, Nature Biomedical Engineering, 2017
	Retinal diseases (OCT)	De Fauw, Nature Medicine, 2018
	Macular degeneration	Burlina, JAMA Ophthalmology, 2018
	Retinopathy of prematurity	Brown, JAMA Ophthalmology, 2018
AMD and diabetic retinopathy	Keremany, Cell, 2018	
Gastroenterology	Polyps at colonoscopy*	Mori et al, Annals Internal Medicine, 2018
Cardiology	Echocardiography	Madani, NPJ (Nature) Digital Medicine, 2018
	Echocardiography	Zhang, Circulation, 2018

*Prospective assessment.

Adapted from Eric Topol. Available from: URL: <https://twitter.com/EricTopol/status/1051174567882907648>, with permission of author (3).

여 결과를 살펴보면, 연구 대상은 뇌, 폐, 복부, 심장, 유방, 뼈 등 많은 장기에 걸쳐 있다. 뇌의 경우 자기공명영상, 폐는 흉부 x-선 촬영과 컴퓨터단층촬영, 복부는 컴퓨터단층촬영, 유방은 유방촬영 영상이 주로 이용되었다. 대상 주제는 주로 장기 및 병변의 분할(segmentation), 검출(detection), 그리고 분류(classification) 였다. 이러한 결과를 딥러닝 알고리즘 개발에 적합한 주제 선정과 관련하여 해석해보면, 많은 양의 비교적 균질한 영상데이터를 수집할 수 있는 분야에서 장기 및 병변을 분할, 검출 혹은 분류함으로써 해결할 수 있는 문제가 딥러닝 알고리즘 개발에 적합한 주제라는 것을 알 수 있다. 다시 말해, 희귀한 질환이나 영상데이터를 확보하기 어려운 질환은 딥러닝 알고리즘 개발에 적합하지 않은 주제들이다. 알고리즘의 활용 방향의 측면에서는 사람이 할 수 있으나 반복적인 작업, 혹은 사람이 할 수 있으나 실수할 경우 치명적일 수 있는 문제가 적절한 주제가 될 수 있다. “The Patient Will See You Now (청진기가 사라진 이후)”의 저자인 Eric Topol(3)이 선정한 의료영상 딥러닝 분야에서 최신의 중요한 논문들의 주제를 살펴봐도 이러한 경향을 볼 수 있다(Table 1).

2단계: 데이터 수집

전통적인 기계학습에서는 검출하고자 하는 피처(feature)를 전문가가 미리 정의하였지만 딥러닝은 피처의 정의 없이 데이터로부터 자동적으로 피처를 학습한다. 이러한 이유로 딥러닝에 있어서 가장 중요한 부분 중에 하나가 바로 데이터라 할 수 있다. 데이터를 수집하는데 있어 고려해야 할 사항들이 있다. 첫째는 모든 데이터의 익명화가 필요하다. 대부분의 국가에서 의료정보는 중요한 개인정보로 간주되어 법으로 엄격하게 규제되고 있다. 딥러닝의 경우 일반적으로 많은 양의 데이터를 다루게 되므로 정보가 유출될 경우 매우 심각한 상황이 발생할 수 있다. “기관생명윤리위원회 정보포털 인체유래물은행관련 개인정보보호 및 익명화” 혹은 “빅데이터 활용을 위한 개인정보 비식별화 기술 활용 안내서” 가이드라인을 참조할 수 있다(4). 비슷한 해외의 가이드라인으로는 미국의 Health Insurance Portability and Accountability Act와 유럽의 General Data Protection Regulation이 있으니, 이를 활용한 프로그램들을 이용하는 것이 바람직하다. 둘째는 충분한 데이터의 확보이다. 해결하고자 하는 문제에 따라서 필요한 데이터의 크기는 달라지겠지만 딥러닝의 작동원리를 고려한다면 일반적으로 많은 양의 데이터가 필요하게 된다. 또한 데이터 크기가 작으면 성능에 비하여 오버피팅(overfitting)이 쉽게 발생하기 때문에 충분한 데이터의 확보가 필요하다. 셋째로 가능한 알고리즘이 사용될 실제환경을 대변할 수 있는 데이터를 확보하는 것이 중요하다. 그렇지 않을 경우 알고리즘의 개발 후 검증 및 적용 단계에서 일반화의 저해를 초래할 수 있다. 이러한 점을 고려할 때 데이터를 수집하는데 있어 기관간의 협력이 중요할 수 있다. 규모가 큰 기관에서는 기관 안에서 충분한 데이터를 확보할 수도 있겠지만 아무리 규모가 큰 기관이라 할지라도 해결과제에 따라서 그렇지 못할 수도 있다. 또한 기관간에 데이터의 특성이 다를 수 있다. 만약 여러 기관에서 데이터를 수집한다면 데이터 크기의 측면뿐 아니라 개발된 알고리즘의 일반화의 측면에서도 도움이 될 수 있다. 하지만 기관간에 데이터가 이동할 경우 정보유출의 위험도는 증가할 수 있고 데이터 소유권의 문제가 발생할 수도 있다. 이러한 문제를 해결하기 위해 기관간에 입력 데이터의 이동이나 공유는 없이 학습 모델(training model)만을 공유하여 알고리즘을

개발하는 새로운 개념이 제안되었다(5). 또한 최근 발표된 논문에서는 시뮬레이션을 통해 데이터를 한 곳에 모아 훈련한 모델과 데이터를 각각의 소속기관에 보관한 상태에서 모델을 순환하면서 훈련한 경우를 비교하였을 때 비슷한 정확도를 보였다고 발표하였다(6). 이러한 결과는 의료분야의 딥러닝 연구에 있어서 기관들간의 협업이 제한된 데이터의 문제를 극복할 수 있는 방법이 될 수 있음을 보여주고 있다. 이는 아직 개념 확립단계이기는 하지만 개인정보보호의 훼손 없이 지속 가능한 ‘데이터로부터 인공지능개발’의 생태계를 만들기 위해서는 필요한 과정으로 생각된다.

두개 내 출혈에 관한 저자의 연구에서는 데이터의 수집을 위해 Massachusetts General Hospital (이하 MGH)의 Research Patient Data Registry (이하 RPDR)를 이용하였다. RPDR은 일종의 Clinical Data Warehouse로 병원의 의무기록 시스템으로부터 환자의 정보를 취합하여 보관하는 시스템이다. RPDR에는 기본적인 환자정보와 각종 검사결과, 영상검사 보고서, 병리검사 보고서, 상병코드, 입퇴원 요약지 등 연구에 필요한 대부분의 정보를 포함하고 있다. 이러한 RPDR의 장점은 연구자들이 연구 주제에 맞는 대상군을 연구 시작 전 쉽게 탐색해봄으로써 연구의 실현가능성 여부를 평가할 수 있다. 또한 연구자 스스로 연구 대상자를 선정하고 선정된 대상자의 자료를 수집할 수 있기에 연구의 진행속도 측면에서도 장점을 가진다. 한가지 아쉬운 점은 현재의 RPDR에는 영상데이터 자체가 보관되어 있지는 않다. 따라서 RPDR을 통해 대상환자군을 어느 정도 선정한 후 영상데이터 서버에 접속하여 데이터를 수집하는 별도의 과정을 거치게 된다. 연구 개발을 장려하는 MGH는 이러한 연구 플랫폼을 오래전부터 개발하여 사용하고 있는데, 이는 MGH가 인공지능 연구를 선도할 수 있게 하는 일종의 비밀 병기라 할 수 있다.

3단계: 데이터 탐색 및 정제

데이터가 수집된 이후에는 가장 먼저 데이터를 탐색하는 과정이 필요하다. 수집된 데이터의 질이 알고리즘 개발에 적합한지, 데이터의 분포가 왜곡되어 있지는 않은지, 데이터의 특성은 어떠한지 등에 대한 탐색이 필요하다. 이 과정을 탐색적 자료 분석(exploratory data analysis)이라고 하며, 이러한 과정을 통해 데이터 수집과정에 문제는 없었는지를 파악할 수 있고, 데이터 정제 방법

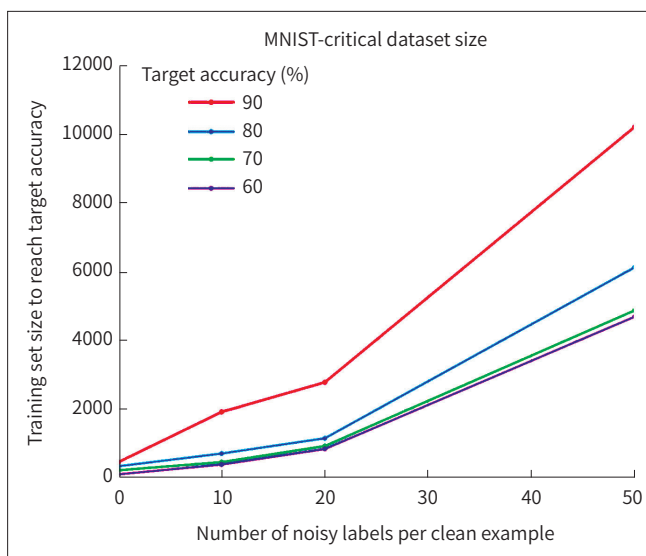


Fig. 3. Relationship between the amount of noise in the dataset and the critical number of clean training examples needed to achieve high test accuracy.

MNIST = Modified National Institute of Standards and Technology
Adapted from Rolnick et al. arXiv preprint 2017;arXiv:1705.10694, with permission of author (7).

에 대한 아이디어를 얻을 수 있다.

다음은 딥러닝 알고리즘 개발에 있어서 가장 많은 시간과 인력을 필요로 하는 데이터 정제 단계이다. 앞서 언급하였듯이 딥러닝 알고리즘 개발에는 많은 양의 데이터가 필요하다. 하지만 일반적으로 잡음이 많은 데이터의 비율이 높아질수록 필요한 데이터의 크기도 증가하게 되므로(Fig. 3) (7) 데이터 정제를 통해 필요한 데이터의 크기를 감소시키고 알고리즘의 정확도를 향상할 수 있다.

데이터 정제 과정을 세부적으로 살펴보면 데이터를 알고리즘 개발에 적합하게 변환하는 과정과 데이터 레이블링(labeling) 과정이 있다. 처음 수집된 데이터는 각각의 데이터마다 특성이 다르며 노이즈로 작용할 수 있는 부분들을 포함하고 있다. 이에 각각의 영상데이터의 크기를 동일하게 맞추고, 실제 알고리즘 개발에 입력할 영상 이외의 부분을 잘라내는 등의 과정이 필요하다. 또한 의료영상은 대부분 제한된 수의 데이터만을 사용할 수 있는 경우가 많아 데이터 수 자체를 증폭(augmentation) 시키기 위해 영상의 왜곡(skewing), 대조도나 해상도의 조정, 반전, 회전, 확대, 영상 내의 병변의 위치 변화 등의 방법을 이용할 수 있다. 다음 과정은 데이터 레이블링이다. 현재 대부분의 영상데이터를 활용한 딥러닝 알고리즘 개발은 지도학습(supervised learning) 방법을 이용하고 있다. 이는 비지도학습(unsupervised learning) 방법이 일반적으로 지도학습방법에 비해서 필요로 하는 데이터의 크기가 크기 때문에 비교적 제한된 양의 데이터를 이용할 수 있는 의료영상분야에서는 아직까지 널리 이용되고 있지는 않다. 지도학습은 입력 데이터 값에 상응하는 결과값을 함께 주는데 여기서 결과값을 만들어내는 과정이 레이블링과정이다. 레이블링을 자세하게 할수록 필요한 데이터의 크기를 감소시키고 개발된 알고리즘의 정확도를 향상시킬 수 있다. 예를 들어, 흉부 x-선 촬영영상에서 폐렴 유무를 판단하는 딥러닝 알고리즘을 개발하고자 할 때, 단순히 폐렴 유무를 결과값으로 레이블링하는 것보다 영상에서 폐렴 부위를 레이블링하는 것이 알고리즘 개발에 있어 장점이 있다. 하지만 이와 같이 레이블링을 자세하게 하기 위해서는 많은 시간과 인력이 필요하게 된다. 한가지 예로, 2018 Radiological Society of North America (RSNA)에서 시행된 Pneumonia Detection Challenge에 이용된 데이터셋을 만들기 위해 약 20명의 영상의학과 전문의가 몇 달에 걸쳐 레이블링을 시행하였다. 영상에서 장기나 병변을 분할하여 레이블링하는 데는 MRICron이나 3Dslicer와 같은 상용화된 프로그램을 사용할 수 있다. Pneumonia Detection Challenge에서는 여러 명의 전문가가 함께 작업하는 것을 수월하게 하기 위해서는 MD.ai가 개발한 프로그램을 사용하였다. 이 밖에도 여러 가지 영상 분할 혹은 정보 추출을 위한 프로그램들이 개발되어 있으며 이러한 프로그램을 이용한다면 레이블링에 투입되는 시간과 인력을 감소시킬 수 있을 것이다.

이와 같은 데이터 정제 과정을 거쳐 완성된 데이터를 이후 사용되는 목적에 따라 훈련데이터(training dataset), 검증데이터(validation dataset), 그리고 평가데이터(test dataset)로 나누게 된다.

4단계: 알고리즘 개발

의료영상 처리 시 사용하는 알고리즘은 전처리, 모델학습, 후처리의 세 가지 단계로 나눌 수 있다. 전처리 단계에서는 각 영상데이터가 원래 저장되어 있던 상태에서, 딥러닝 모델이 가장 효율

적으로 문제를 해결할 수 있는 형태로 변환하는 과정이다. 대부분의 의료영상데이터 정보를 Digital Imaging and Communications in Medicine (DICOM) 데이터의 형태로 저장된다. 이를 일반적인 딥러닝 모델이 인식하기 쉬운 8비트 이미지로 변환하는 과정이 필요하다. 예를 들면, 컴퓨터단층촬영 영상의 경우 Hounsfield Unit 값 데이터를 이미지로 변환하는 Windowing 과정이 필요하다. Shin 등(8)의 논문에서는 영상의학과 의사가 여러 Window setting으로 병변을 관찰하듯이 Window setting 값을 다양하게 설정하여, 다채널 이미지 데이터를 학습 시켰을 때 성능의 향상을 보였다.

앞서 데이터 탐색 및 정제 단계에서 언급한 데이터 증폭(augmentation)은 전처리 단계에서 또한 시행된다. 증폭의 방법이 점점 다양해지면서, 모든 데이터를 미리 증폭해 놓을 수 없는 경우 전처리 단계에서 모델학습과 동시에 증폭을 적용하는 것을 실시간 증폭이라 한다. 저자의 두개 내 출혈 논문에서는 컴퓨터단층촬영 영상의 슬라이스와 슬라이스 사이의 보간(interpolation)을 이용한 데이터 증폭을 학습 전에 미리 시행하였고, 선형 매핑(affine) 변환에 의한 증폭은 학습 중에 시행하였다.

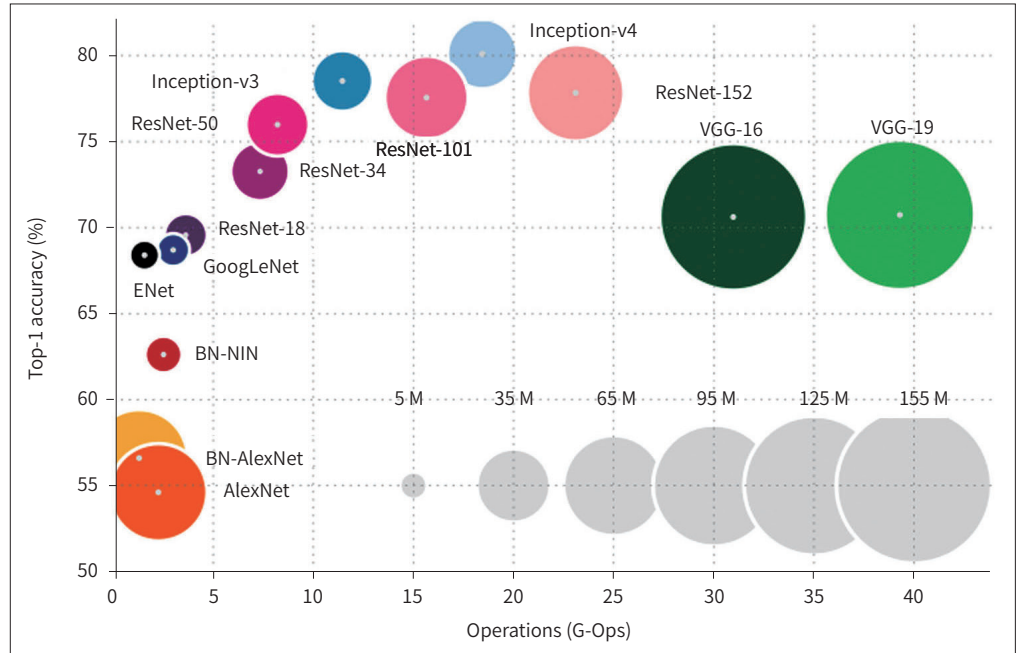
전처리 단계에서는 고전적인 기계학습 방법이나, 연구자가 직접 정의한 규칙에 의한 방법도 많이 쓰인다. 전처리 단계는 연구 주제나, 연구자에 따라서 방법이 많이 달라질 수 있다. Roth 등(9)은 장기 분할에서 장기가 있을 만한 부분을 Random forest 등의 고전 기계학습 방법을 통하여 먼저 찾고, 이후에 딥러닝 모델을 적용하여 성능을 향상시켰다. 최근에는 전처리까지 한꺼번에 딥러닝 모델에 포함시켜 한 번에 학습시키기도 하는데, 저자는 컴퓨터단층촬영 영상의 Window setting을 딥러닝 모델에 포함시켜 학습하는 방법을 제안하였다(10).

최근 영상의학분야에서 쓰이는 딥러닝 모델들은 컴퓨터 공학에서 뛰어난 성능을 보인 모델, 특히 컨볼루션 신경망 모델(Convolutional Neural Network Model)을 가져와 그대로 쓰거나 조금 변형하는 경우가 많다. 최근 모델들의 가장 주목할만한 특징으로는 많은 컨볼루션 레이어(layer) 개수와 레이어 간의 우회 연결을 들 수 있다.

여러 딥러닝 모델 중 분류 과제 해결에는 ResNet (11), DensNet (12), Inception v3 (13)가 많이 쓰이는데, 이 세 모델 모두 최대 100층 이상의 컨볼루션 레이어를 사용하고, 레이어를 우회하는 연결이 사용된다. 분할 과제 해결에 있어서는 U-Net (14)이 주로 쓰인다. U-Net 또한 23개의 컨볼루션 레이어가 쓰였고, 4개의 우회 연결이 존재한다. 이 밖에도 다양한 크기의 딥러닝 모델들이 인공지능 분야에서 개발되었는데, 일반적으로는 계산속도와 정확도 사이에는 trade-off (상충)가 존재한다. Canizi 등(15)은 여러 딥러닝 모델의 ImageNet-Top1 정확도와 함께 모델 파라미터의 크기와 연산량을 비교하였다(Fig. 4). 컴퓨터단층촬영이나 자기공명영상의 경우에는 수초에서 수분 안에 정확한 예측을 목표로 하기 때문에 큰 연산량의 모델을 주로 선택하고, 수분의 1초 단위로 빠른 예측이 필요한 경우에는 상대적으로 작은 모델을 고려해 볼 수 있다.

모델이 예측한 이후에 의사나 사용자가 인식할 수 있는 형태로 변환하는 과정이 반드시 필요하다. 이를 후처리(post-processing)라 하는데, 분류 문제에서 역치(threshold)를 정하는 기본적인 역할부터 또 다른 딥러닝 모델을 이용해 성능을 향상시키는 방법까지 다양하다. 여러 개의 딥러닝 모델의 결과를 이용하는 것을 앙상블(ensemble) 과정이라 하는데 각 예측 모델의 값의 평균을 구

Fig. 4. Correlation between ImageNet Top-1 one-crop accuracy and amount of operations. Adapted from Canziani et al. arXiv preprint 2016;arXiv:1605.07678, with permission of author (15).



하는 방법이 흔하게 이용된다. 이 밖에 최대값을 선택하거나, 가장 다수가 선택한 결과를 사용하는 방법 등도 쓰인다. 타깃 데이터의 불균형(imbalance)이 심한 경우에는, 여러 단계에 걸쳐 위의 방법을 사용하기도 하는데, 이를 캐스케이드(Cascade) 모델이라고 부른다. 후처리는 연구 목적이나 연구자에 따라 다양하게 나타나며, 연구 목적에 맞는 적절한 후처리 방식을 사용하는 것이 중요하다. 전처리와 후처리는 연구 목적을 잘 이해한 개발자와 그렇지 못한 개발자의 역량이 구별되는 중요한 과정이다.

5단계: 알고리즘 평가

훈련데이터와 검증데이터를 이용하여 알고리즘이 개발되면 이를 평가하는 단계가 필요하다. 가장 일반적인 방법은 개발된 알고리즘을 평가데이터에 적용하여 그 정확도를 평가하는 방법이다. 과제에 따라서 기존의 알고리즘 혹은 해당 분야의 전문가와 정확도 비교를 통해서 개발된 알고리즘을 평가하기도 한다. 만약 이러한 평가과정에서 알고리즘의 성능이 기대치에 미치지 못한다면 앞서 설명한 알고리즘 개발 단계들에 문제가 없었는지를 확인하고 이를 개선하는 과정을 반복하게 된다.

저자는 두개 내 출혈 예측 알고리즘을 실제 임상에 적용하기 전 단계로 개발된 알고리즘을 일정 기간 동안 MGH의 응급실에서 촬영한 비조영 뇌 컴퓨터단층촬영 영상에 적용하여 평가하였다. 그런데 알고리즘의 정확도가 기존의 평가데이터를 이용한 평가에서의 정확도에 비해서 낮아지는 현상을 발견하였다. 이에 알고리즘 개발단계를 되짚어 보았으며 훈련데이터의 수집에 문제가 있었음을 발견하였다. 즉, 훈련데이터에서는 특정 제조사의 컴퓨터단층촬영 기계에서 촬영된 영상이 많이 포함되었는데 공공롭게도 이러한 영상들에서 두개 내 출혈의 빈도가 높았다. 반면 응급실

에서 촬영한 뇌 컴퓨터단층촬영 영상에는 앞서 언급한 제조사 이외의 다른 제조사의 기계들에서 촬영한 영상들이 많이 포함되어 있었다. 다시 말해 훈련데이터가 실제 임상환경을 정확하게 반영하지 못했기에 이러한 결과가 발생한 것을 알 수 있었다. 이에 여러 제조사의 영상을 포함하는 훈련데이터를 재구성하여 이러한 문제를 해결할 수 있었다.

6단계: 임상적용

알고리즘 개발의 목적이 단순한 연구가 아닌 이상 최종 단계는 의료현장으로의 적용일 것이다. 개발된 알고리즘을 임상에 적용하기 위해서는 미국의 Food and Drug Administration (이하 FDA)이나 한국의 식품의약품안전처와 같은 정부기관의 허가를 받아야 한다. 이와 관련하여 2018년 미국 FDA는 “Medical Device Safety Action Plan: Protecting Patients, Promoting Public Health”라는 지침서를 발표하였다. 이전에는 새롭게 개발된 저위험 혹은 중등도 위험 장치의 경우 기존 시장에 있던 장치와 직접 비교를 통해 비슷하거나 우월한 성능을 보일 때 승인을 해주었으나, 이번에 발표된 지침서에 따르면 기존의 장치와 직접 비교 대신 성능이 비슷하거나 우월하다는 것을 판단할 수 있는 데이터를 바탕으로 상당한 동일성(substantial equivalence)이 인정되면 승인을 해줄 수 있게 되었다. 이러한 미국 FDA의 정책 변화는 최근 매우 빠르게 변화하고 있는 의료환경의 특성을 고려하여, 새롭게 개발된 장치나 소프트웨어를 좀 더 빨리 의료환경에 적용할 수 있게 하고자 하는데 그 목표가 있다. 실제로 2018년 말에 의료영상을 대상으로 하는 많은 딥러닝 알고리즘과 플랫폼이 미국 FDA의 승인을 받았다.

딥러닝을 이용한 알고리즘과 플랫폼이 정부기관의 승인을 받았다고 해서 이들이 의료환경에 곧바로 널리 사용되는 것은 아니다. 이전에도 Computer Aided Detection or Diagnosis라는 이름으로 많은 프로그램들이 시장에 소개되었지만 이들의 대부분이 현재 의료환경에서 실제 사용되고 있지 않다. 이는 개발된 알고리즘이 실제 의료환경에서 살아남기 위해서는 단순히 연구 개발 단계에서 제시하는 단순한 수치로 나타나는 “성능의 향상” 이외의 많은 변수들이 작용함을 시사하는 것이다. 아무리 어떠한 알고리즘이 진단 능력에 있어서 탁월하다고 하더라도 실제 의료환경에서 사용하기 너무 복잡하거나 기존의 워크플로(workflow)를 급격하게 변화시켜야 한다면 이는 사용자들에게 받아들여지기 어렵게 된다. 이러한 측면에서 최근 딥러닝 알고리즘 개발도 “이음새 없는 통합(seamless integration)”의 중요성을 강조하기 시작하였다. 이음새 없는 통합이란 사용자가 특정한 일을 함에 있어서 새로운 어떠한 프로그램이나 플랫폼을 사용할 때 기존과 다른 무언가를 사용하고 별도로 무언가를 인식해야 할 필요가 없도록 하는 것을 말한다. 말 그대로 이전 환경과 새로운 환경간에 이음새가 필요치 않도록 하는 것이다.

이러한 이음새 없는 통합의 의미를 실제 예를 통해 알아보자. 미국의 의료영상 인공지능 회사인 “AIdoc”은 뇌 컴퓨터단층촬영에서 두개 내 출혈, 척추 컴퓨터단층촬영에서 골절, 폐 컴퓨터단층촬영에서 폐색전, 기흉, 늑골 골절, 폐결절, 그리고 복부 컴퓨터단층촬영에서 복강 내 공기, 출혈, 혈관의 박리와 동맥류 등 영상을 판독함에 있어서 중요한 소견들을 발견하는 딥러닝 알고리즘을 개발하였다. 개발된 알고리즘은 의료기관에서 촬영된 컴퓨터단층촬영들에 적용되어 위에서 언급한 중요한 소견들의 가능성이 있는 경우 Picture Archiving and Communication System (PACS)

의 판독 폴더에서 판독의 우선순위를 최상위로 이동시켜 담당의사로 하여금 이러한 환자들의 영상을 우선 확인하도록 하였다. 이를 통해 중요한 소견을 놓치는 가능성을 낮추고 담당의사로 하여금 응급한 환자를 좀 더 빨리 찾을 수 있도록 하였다. 전체적인 적용과정을 생각해보면 사용자, 즉 담당의사의 입장에서는 일의 우선순위의 변화 이외에는 워크플로 자체에 큰 변화가 없는 것을 알 수 있다. 즉 앞서 설명한 이음새 없는 통합의 개념이 잘 적용된 예라 할 수 있을 것이다.

결론

의료영상분야에서 딥러닝은 기술적인 부분에서 매우 빠르게 발전할 뿐 아니라 적용 분야도 매우 다양해지고 있다. 하지만 알고리즘 및 플랫폼을 개발하는 과정은 앞서 설명한 단계에서 크게 달라지지는 않을 것이다. 따라서 개발과정을 전체적으로 이해하는 것은 보다 유용하고 실제 사용 가능한 알고리즘 및 플랫폼을 개발하는데 도움이 될 것으로 생각된다. 현재까지 의료영상분야에서 개발된 여러 딥러닝 알고리즘은 진단기구라기보다는 의사결정지원시스템(clinical decision supporting system)에 가까운 단계이다. 즉, 아직까지는 사용자의 감독과 책임하에서만 사용할 수 있는 단계이다. 하지만 의료환경에서 딥러닝의 사용이 점차 확대되고 있으며 그 변화 속도도 점차 빨라지고 있다. 한편으로는, 세계적으로 의료서비스의 불균형이 매우 심하다. 미국이나 한국 처럼 높은 질의 의료서비스를 제공하는 나라가 있는 반면 의료인력이나 시설이 절대적으로 부족하여 기본적인 의료서비스조차 제공하지 못하는 나라들도 많다. 인공지능 알고리즘의 개발은 이러한 의료취약지역의 환자도 제대로 된 의료서비스를 제공받을 수 있는 기회를 줄 수 있을 것이다. 지금은 딥러닝으로 대표되는 인공지능이 의사를 대체할지 그렇지 못할지에 대한 의미 없는 논쟁을 하기보다는 새로운 기술을 어떻게 활용하여 의료인 및 환자에게 도움이 되게 할지를 고민할 시기로 생각된다.

Conflicts of Interest

The authors have no potential conflicts of interest to disclose.

REFERENCES

1. Lee H, Yune S, Mansouri M, Kim M, Tajmir SH, Guerrier CE, et al. An explainable deep-learning algorithm for the detection of acute intracranial haemorrhage from small datasets. *Nat Biomed Eng* 2018;3:173-182
2. Litjens G, Kooi T, Bejnordi BE, Setio AAA, Ciompi F, Ghafoorian M, et al. A survey on deep learning in medical image analysis. *Med Image Anal* 2017;42:60-88
3. Topol E. There's been a recent burst of peer-reviewed deep neural network #AI publications in medicine. Available at: <https://twitter.com/EricTopol/status/1051174567882907648>. Published Oct 13, 2018. Accessed Jan 1, 2019
4. Shin SY, Lyu Y, Shin Y, Choi HJ, Park J, Kim WS, et al. Lessons learned from development of de-identification system for biomedical research in a Korean Tertiary Hospital. *Healthc Inform Res* 2013;19:102-109
5. Batten L, Kim DS, Zhang X, Li G. *Applications and Techniques in Information Security: 8th International Conference, ATIS 2017*. Auckland: Springer 2017
6. Chang K, Balachandar N, Lam C, Yi D, Brown J, Beers A, et al. Distributed deep learning networks among institutions for medical imaging. *J Am Med Inform Assoc* 2018;25:945-954
7. Rolnick D, Veit A, Belongie S, Shavit N. Deep learning is robust to massive label noise. *arXiv preprint* 2017;arX-

iv:1705.10694

8. Shin HC, Roth HR, Gao M, Lu L, Xu Z, Nogues I, et al. Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning. *IEEE Trans Med Imaging* 2016; 35:1285-1298
9. Roth HR, Lu L, Farag A, Shin HC, Liu J, Turkbey EB, et al. Deeporgan: multi-level deep convolutional networks for automated pancreas segmentation. *arXiv preprint* 2015;arXiv:1506.06448
10. Lee H, Kim M, Do S. Practical window setting optimization for medical image deep learning. *arXiv preprint* 2018;arXiv:1812.00572
11. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. *arXiv preprint* 2015;arXiv:1512.03385
12. Huang G, Liu Z, Van der Maaten L, Weinberger KQ.ensely connected convolutional networks. *arXiv preprint* 2016;arXiv:1608.06993
13. Szegedy C, Vanhoucke V, Ioffe S, Shlens J, Wojna Z. Rethinking the inception architecture for computer vision. *arXiv preprint* 2015;arXiv:1512.00567
14. Ronneberger O, Fischer P, Brox T. U-Net: convolutional networks for biomedical image segmentation. *arXiv preprint* 2015;arXiv:1812.00572
15. Canziani A, Paszke A, Culurciello E. An analysis of deep neural network models for practical applications. *arXiv preprint* 2016;arXiv:1605.07678

딥러닝 알고리즘 개발과정을 통해 본 영상의학분야에서 딥러닝의 최신 경향

송경두^{1,2} · 김명찬¹ · 도신호^{1*}

최근 인공지능은 지각정보를 해석하는데 있어 상당한 진보를 이루었으며, 이를 통해 기계는 매우 복잡한 데이터를 보다 잘 해석할 수 있게 되었다. 최근 몇 년간 딥러닝 기술로 대표되는 인공지능의 의료 및 생물의학 연구분야로의 적용이 기하급수적으로 증가하고 있다. 이번 기고문에서는 의료영상분야의 딥러닝 알고리즘 개발 단계를 주제선정, 데이터 수집, 데이터 탐색 및 정제, 알고리즘 개발, 알고리즘 평가 그리고 임상적용의 단계로 나누어 설명하고, 각각의 단계에서 최신의 동향을 소개하고자 한다.

¹메사추세츠병원 영상의학과, ²삼성서울병원 영상의학과