의료 분야에 활용되는 설명 가능한 인공지능(XAI) 활용 동향

박 진 환*

Abstract

설명 가능한 의사 결정 체계라는 용어로 첫 등장을 알린 XAI는 인공지능의 의사결정 과정을 이해하기 위해 사용되었다. 컴퓨터의 연산 모델과정과 결과 도출 과정을 투명하게 만들기 위해 발전한 설명 가능한 인공지능(XAI)는 인공지능의 의사결정의 근거를 확인할 수 있고, 오류가 발생했다면 해당 오류의 원인을 찾을 수 있다는 점에서 장점이 있다.

본 고에서는 설명 가능한 인공지능(XAI)의 발전 과정과 다양한 종류의 XAI에 대해 설명하고 있다. 또한 의료 분야에서 설명 가능한 인공지능(XAI)이 적용된 사례와 최근까지 발표된 연구 사례에 대해서 논의해 보고자 한다.

I. 서 론

설명 가능한 인공지능(XAI)는 2004년에 반 렌트(Michel van Lent), 피셔(William Fisher)와 만쿠소(Michael Mancuso)에 의해 명명되었으며 explainable artificial intelligence의 약자이다. 꽤나 오래된 역사를 가지고 있는 인공지능은 컴퓨팅 파워가 발전함에 따라 훌륭한 예측력을 보여주면서 다양한 분야에 적용되고 있지만 인공지능이 의사결정을 하는 과정에서 나오는 결과물이 정확히 어떠한 근거로 형성되는 것인지에 대해 알 수 없다는 것이 단점이었다. 이를 우리는 블랙박스(black box)라고 부른다. 이유를 알 수 없는 의사결정을 단순히 인공지능의 결정이라고 하여 믿는 것은 어려운 일이었기 때문에 인공지능에 해석 가능성을 부여할 수 있는 XAI의 발전은 필연적이었다고 볼 수 있다. 그럼에도 불구하고 과

* 한국과학기술원(KAIST) 예측연구실, 박사, krapynot@gmail.com

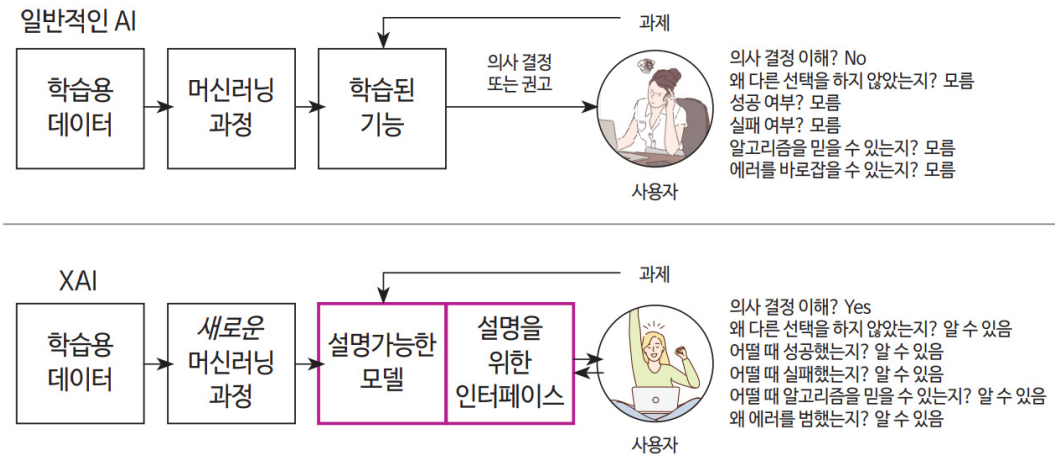
거 2000년대에는 XAI가 제한된 범위에서만 사용되었다. 이는 계산력을 비롯한 물리적 제약 조건 등의 이유가 있었는데, 시간이 지남에 따라 하드웨어의 기능이 급속도로 발전하게 되면서 활용도가 올라가게 된 것이 주요 원인이었다. 간단히 말하자면, XAI의 발전은 알고리즘 개발자와 의사 결정자를 설득하는데 도움을 준다. 어떠한 요인이 결정에 가장 유의미한 영향을 미쳤는지 등을 계량화하여 의사 결정자에게 제시하기 때문이다. 즉, 최종적인 의사결정의 주체가 되는 인간이 인공지능의 결정 과정을 더 잘 이해할 수 있게 도움을 제공함으로써 오류의 원인을 상대적으로 쉽게 찾을 수 있고 결과의 신뢰도가 올라감과 동시에 때에 따라 인공지능 알고리즘을 개선하는데도 도움을 줄 수 있다는 말이다. 또한, 인공지능의 오류 원인을 개선할 수 있다는 점에서 인공지능에 대한 대중의 신뢰도를 향상시킬 수 있다는 장점이 있다.

인공지능 중 딥러닝을 비롯한 머신러닝 알고리즘은 불투명한 특성을 가지고 있다. 이러한 경우 XAI 기법 중 피처 중요도(Feature Importance), 필터 시각화(Filter Visualization), LRP(Layerwise Relevance Propagation)등을 활용하여 인공지능의 의사결정과정을 분해할 수 있다. 본 고에서는 이 중 대표적인 기법들과 이를 활용한 연구들에 대해 다룰 예정이다.

인공지능은 다양한 분야에 적용되고 있고, 의료 분야에서 또한 활발하게 적용되고 있다. 의료 분야에서는 데이터의 양과 질이 풍부하게 제공되고 있고, 특히 대한민국은 전국민 의료보험이 적용되는 국가적 특성에 의해 환자들의 의료 비용 지출이나 병력을 데이터화하여 관리하고 있다. 의료 이용 기록 뿐 아니라, 영상, 음성, 이미지 등의 데이터를 처리하기에 인공지능이 적합한 경우가 많기 때문에 의료분야에서는 활발히 연구가 진행되고 있고 실전에서도 활용되는 중이다. 하지만 인공지능이 본연으로 가지고 있는 블랙박스의 특성은 의학 전문가 및 환자들 불신으로 이어지기도 한다. 인공지능에서의 오류 발생이 치명적인 결과로 이어질 수 있는 분야인만큼, 심층 신경망의 불투명성은 사용자로 하여금 의사결정을 신뢰할 수 있는지를 판단하는 데에 어려움을 준다. 이에 설명 가능한 인공지능(XAI)은 의료 분야에서 빠르게 발전하고 있다. [그림 1]은 설명 가능한 인공지능(XAI)의 개념을 도식화하여 나타낸 것이다.

기존의 인공지능 모델에서 해석성을 부여한 설명 가능한 인공지능(XAI)은 다양한 기술적 접근을 통해 탄생하였다. 설명 가능한 인공지능 적용 모델은 크게 세 가지로 나뉘는데, 역산 과정을 넣는 등의 기존 학습 모델을 변형하는 방법, 설명을 위한 알고리즘을 추가하는 새로운 학습 모델을 개발하는 방법, 타 학습 모델과 비교하여 예측 결과에 대한 근거를 제공하는 방법이 있다.

그림 1 일반적인 AI와 설명가능한 인공지능(XAI)의 차이



자료: 미 국방연구원 '설명가능 인공지능(XAI)', 한국정보화진흥원(2018)

의료분야에 설명가능한 인공지능(XAI)이 적용된 사례를 서두에 간단히 소개하면 영국 NHS와 캠브리지 대학에서는 조기 유방암 치료를 받은 여성 환자가 수술 후 어떻게 생존율을 향상시킬 수 있는지 설명 가능한 인공지능(XAI)을 활용하여 분석하였다. 또한, 흉부 X-ray 분석을 통해 위험 정도를 표시하는 것을 딥러닝을 활용하여 연구하였는데, 기존의 모델보다 위험 정도를 표시하게 되는 원리를 설명 가능한 인공지능(XAI)을 통하여 임상적으로 정확하게 설명하였다는 결과도 있다. 이처럼 임상 현장에서 의료진의 의사결정을 도우는 것이 필요하기 때문에 설명성은 매우 중요하다. 투명성과 신뢰성이 미달인 경우 법적인 문제 등이 발생할 수 있는 문제가 있기 때문이다. 즉, 설명성의 부여는 일반 대중들의 인공지능을 향한 불안감을 해소할 수 있고 규정을 준수하면서 인공지능 서비스의 가치 창출에 또한 긍정적인 영향을 미친다.

본 고는 앞서 서술한 설명 가능한 인공지능(XAI) 모델 종류에 대해 알아보고 해당 모델들을 활용한 최신 연구 동향을 소개하려 한다. 또한 최신 연구 동향의 시사점과 앞으로의 비전을 제시한다.

Ⅱ. 의료 분야에서 활용되는 설명 가능한 인공지능(XAI) 모델의 종류

1. 의사 결정 트리

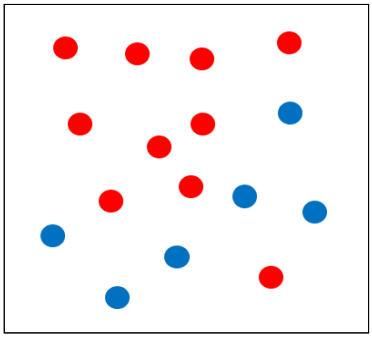
의사 결정 트리(Decision Tree)는 일련의 분류 규칙을 통해 데이터를 분류, 회귀하는 지도 학습 모델 중 하나이다. 해당 모델은 질문에 답을 하는 과정을 지속적으로 반복하면서 예측 혹은 분류를 진행하며, 결과가 트리 구조를 가지고 있기 때문에 의사 결정 트리라고 불린다. 해당 모형은 해답을 찾는 과정을 시각화하여 표현할 수 있다. 이는 도식화되어 표현되기 때문에 인간이 이해하기 쉽다는 장점이 있다.

의사 결정 트리는 정보 이득 수치를 계산한 뒤 최적의 트리를 완성하는 방식으로 진행이 되는데, 정보 이득이라는 것은 엔트로피의 변화량을 활용하여 계산된다. 엔트로피는 다음과 같이 계산할 수 있다.

$$Entropy(A) = -\sum_{k=1}^{n} p_k \log_2(p_k)$$

A는 의사 결정을 수행하는 전체 영역, n은 범주 개수, p_k는 A영역 속 데이터들 중 k 범주에 속하는 데이터의 비율이다. 이해를 돕기위해 [그림 2]의 엔트로피를 계산해보면 다음과 같다.

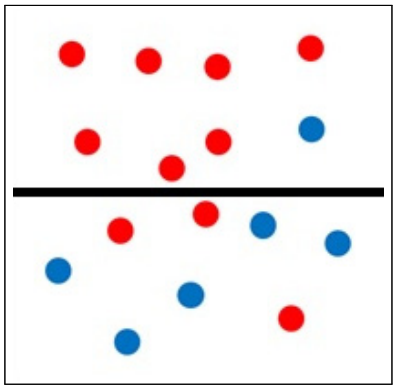
그림 2 엔트로피 계산을 위한 예시 1



해당 그림 영역에서의 엔트로피를 계산해보면 전체 16개의 점 중 적색 동그라미는 10개, 청색 동그라미는 6개인 것을 알 수 있다. 그렇다면 엔트로피는 다음과 같이 계산된다.

$$Entropy(A) = -\frac{10}{16} log_2(\frac{10}{16}) - \frac{6}{16} log_2(\frac{6}{16}) \approx 0.95$$

그림 3 엔트로피 계산을 위한 예시 2



[그림 2]의 점들을 분류하기 위해서 [그림 3]처럼 직선을 그렸을 때, 엔트로피는 다음과 같이 변화한다.

$$Entropy(A) = 0.5 \times (-\frac{7}{8} log_2(\frac{7}{8}) - \frac{1}{8} log_2(\frac{1}{8})) + 0.5 \times (-\frac{3}{8} log_2(\frac{3}{8}) - \frac{5}{8} log_2(\frac{5}{8}))$$
$$\approx 0.75$$

이는 [그림 3]에서 그은 직선에 의해 0.2만큼의 엔트로피가 감소되었음을 의미한다. 이처럼 어떠한 방식으로 선을 그어서 점들을 분류하느냐에 따라서 엔트로피의 변화량은 차이가 있다. 의사 결정 트리는 다양한 방식의 분류 방식을 시도하고, 정보 이득량(엔트로피의 변화량)이 가장 커지는 방향으로 학습을 진행한다. 단, 해당 과정에서 모든 경우에 대하여 분류를 모두 수행하는 경우 과적합의 문제가 발생할 수 있다는 것을 유의해야 한다.

□ 피쳐 중요도

피쳐 중요도를 구하는 것은 설명 변수로 입력되는 요소들 중 어떠한 것이 종속 변수에 가장 큰 영향을 미치는가로 이해하면 편리하다. 설명 변수 역할을 하는 데이터의 피쳐가 정확한 분류에 얼만큼의 영향을 미치는지를 계산하는 방법으로, 특정 피쳐에 변화를 주었을 때 모델의 성능에 어떠한 변화가 있는지를 관찰한다. 만약 특정 피쳐가 중요하지 않다면, 피쳐의 값을 임의로 바꾸더라도 모델의 성능에는 큰 영향을 주지 않을 것이다. 이러한 개념을 바탕으로 퍼뮤테이션 피쳐 중요도(Permutation Feature Importance)라는 것이 등장한다. 이는 정해진 숫자만큼 피쳐의 값을 변형하면서 오류의 변화를 측정하고 이를 피쳐의 중요도로 치환하는 과정이다.

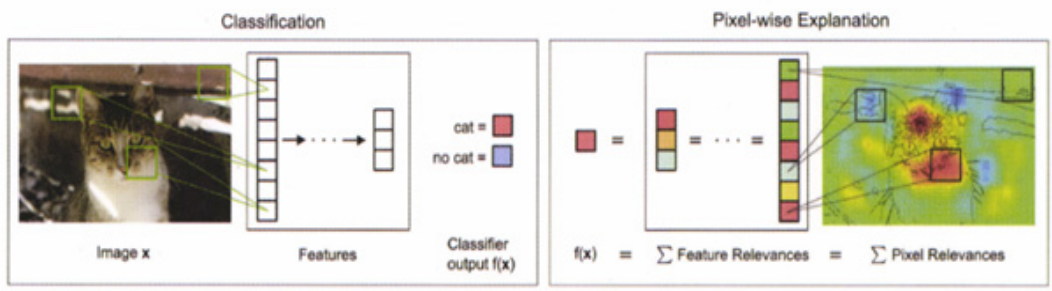
2. LRP(Layer-wise Relevance Propagation)

심층 신경망 모델은 신경망들이 복잡하게 얽힌 형태이다. 각각의 신경망은 가중치, 노드, 편향의 수열합으로 이루어져있고 수많은 파라미터로 구성되어있기 때문에 입력된 값이 결과로 이어지기까지의 과정을 온전히 이해하기란 불가능에 가깝다.

LRP 알고리즘은 어떤 데이터의 부분이 분류 결과를 도출하게 하였는지 히트맵 형식으로 알려준다. 심층 신경망 모델이 분류한 결과를 역순으로 탐지하여 분해하고 이 요소들이 원본 이미지에 도달했을 때 기여도를 표시하는 방식으로 모델을 해석하는 방식이다. LRP는 기본적으로 타당성 전파(Relevance Propagation)와 분해(Decomposition)를 활용하여 모델을 해석한다. 출력값으로부터 계산을 시작하여 기여도(Relevance Score)를 역방향으로 계산해 나가며 비중을 분배한다. 타당성 전파는 특정 결과가 나오게 된 원인의 비중을 분배하고 분해는 타당성 전파로 얻어낸 원인을 가중치로 환원하고 해부하는 과정이다.

분해(Decomposition)는 입력된 값(피쳐) 하나가 결과 도출에 얼마만큼의 영향을 미치는지 해체하는 방법으로 특정 이미지의 픽셀이 해당 결과를 도출하는데 도움이 되는지, 그렇지 않은지를 알 수 있다. [그림 4]의 경우 해당 이미지를 고양이로 인식하는 원인을 픽셀단위로 쪼개어 나타낸 것으로 붉게 나타난 부분이 고양이를 판단하는데 도움이 되는 부분이라고 볼 수 있다. 이는 해당 이미지를 고양이로 예측할 가능성을 역으로 계산하여 판단한 것이다. 분해 과정을 거친 뒤에는 타당성 전파(Relevance Propagation)를 통해 입력값을 전달 받은 은닉층이 결과값 출력에 어떠한 양의 기여를 했는지를 계산한다.

그림 4 히트맵으로 결과를 산출하는 LRP



자료: Bach, Sebastian 외 "On pixel-wise explanations for non-linear classifier decisions by layer-wise relevance propagation", PloS one 10.7(2015): e0130140

□ LRP 등장 이후 설명 가능한 인공지능(XAI) 동향

LRP는 심층 신경망 내부의 활성화 함수를 분해하는 것부터 시작한다. 이 기법은 민감도 분석(Sensitivity Analysis)이 시작이었는데 이를 발전 시킨 것이 디컨볼루션 기법이다. 디컨볼루션 기법은 2014년쯤 등장하였고 신경망의 연결을 역재생하는 방법이다. 이를 발전시킨 것이 LRP 방식이다. LRP 내에서도 계산 방법에 따라 모델의 명칭이 달라진다.

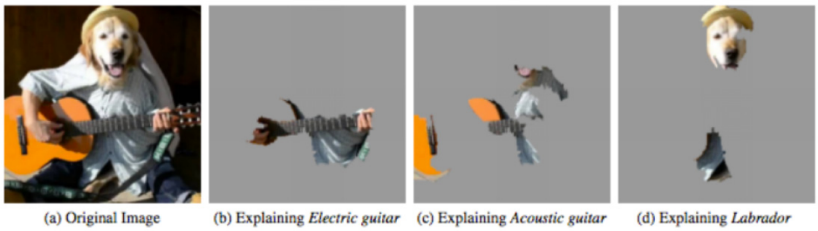
LRP는 심층 신경망 내에서 입력값이 어떠한 방식으로 해석되고 있는지를 시각적으로 보여준다. LRP의 개발 이후에는 CAM(Class Activation Map)이라는 기법도 등장하였다. CAM은 데이터의 양이 부족한 경우 문제에 취약함을 보여줬다는 단점이 있었고 이를 보완한 것이 Grad-CAM 방식이다. 그 이후에는 2018년에 TCAV(Testing with Concept Activation Vectors)가 등장했고 이는 이미 구축된 심층 신경망 모형을 변형하지 않는다는 특징이 있다. 그 이후에도 꾸준히 단점이 보완되면서 성능이 발전된 모델들이 연구되면서 활용되고 있다.

3. LIME, SHAP

LIME은 Local Interpretable Model-agnostic Explanations의 약자이며 텍스트 하이라이트 기능을 제공하는 것으로 알려져 있다. SHAP은 SHapley Additive exPlanations의 약자인데, LIME과 SHAP의 공통점은, 인공지능 모델과 상관없이 학습이 완료된 후에 사후적으로 모델에게 설명력을 부여한다는 점이다.

LIME은 입력데이터에 대해 조금씩 변화를 주는데 이를 일반적으로 변형(permutation)이나 샘플 퍼뮤테이션(sample permutation)이라고 부른다. 즉, 입력데이터가 들어오는 순간 이를 인식 단위로 쪼개고 이미지를 해석한다. 이미지를 쪼갠 뒤 해당 부분 중 원본 이미지를 가장 추정할 확률이 높은 부분을 찾는다. 해당 방법을 활용하여 최적의 설명 부분을 찾는다고 볼 수 있다.

그림 5 LIME의 입력 이미지 데이터를 인식 단위로 쪼갠 결과

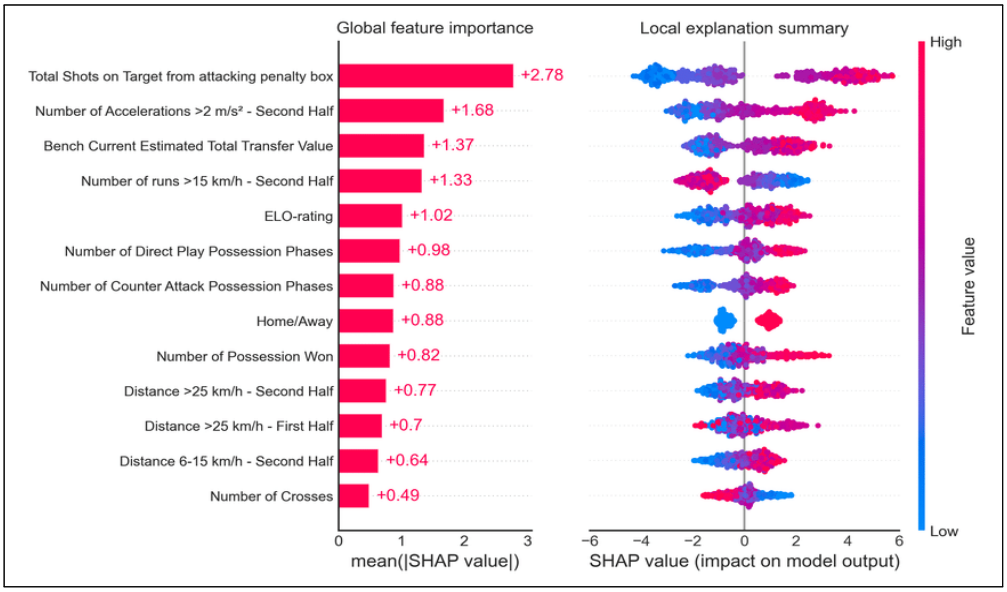


자료: cs.washington.edu/~

LIME은 실제로 의료 데이터 분석에서 상당한 빈도로 사용되는데. 이는 LIME은 인공지능 모델이 어떤 것이든 적용할 수 있다는 장점을 가지고 있기 때문이다. 또한, GPU를 사용해야하는 무거운 기법이 아니기 때문에 모델 자체가 무거워지지 않는다는 장점이 있다. 모델의 학습에 모든 컴퓨팅 파워를 쏟아야하는 정교한 결과를 요구하는 의학 분야에서 최소한의 컴퓨팅 파워로 설명력을 부여할 수 있다는 점에서 선호된다고 해석될 수 있다.

SHAP은 로이드 섀플리가 만든 이론 위에 피쳐간 영향력을 더할 수 있도록 한 것으로, 해당 값은 각 요소의 기여도는 해당 요소를 제외하고 모형의 결과값이 얼마나 영향을 받는가를 수치적으로 측정하여 해당 값을 요소의 영향력이라고 추론하는 방법이다. [그림 6]은 SHAP을 사후적으로 인공지능 모형에 적용하여 해당 피쳐 혹은 변수가 어느정도의 영향력을 예측값에 행사하는지를 시각화하여 표현한 것이다.

그림 6 SHAP 값의 결과 예시



자료: Geurkink, Y., Boone, J., Verstockt, S., & Bourgois, J. G.(2021). Machine learning-based identification of the strongest predictive variables of winning and losing in Belgian professional soccer. Applied Sciences, 11(5), 2378.

Ⅲ. 설명 가능한 인공지능(XAI)를 활용하여 연구된 최신 연구 동향 소개

□ 질병 진단에서 활용된 설명 가능한 인공지능(XAI) 연구

본 고에서는 앞서 설명 가능한 인공지능(XAI)에 대한 간략한 소개와 해당 분야에 속하는 다양한 설명 가능한 인공지능 모형들에 대해 살펴보았다. Kavya 외(2021)의 연구에서는 알러지 진단을 위해 다양한 인공지능 알고리즘을 적용하였다. Amorso 외(2021)의 경우 유방암 치료를 위한 XAI 프레임워크를 제시했다. 군집화(clustering)와 차원 축소 방법(dimension reduction method)으로 유방암 치료를 가장 효과적으로 만들어주는 피쳐를 파악할 수 있었다. El-Sappagh 외(2021)는 알츠하이머 초기와 진행정도 진단을 인공지능 모형을 활용하여 분석하였다. 여기서 해당 연구는 어떠한 요소들이 진단에 가장 주요한 역할을 하는지 알아보기위해 SHAP을 활용하였는데, 이는 앞서 설명한 퍼뮤테이션 피쳐 중요도(Permutation Feature Importance)에서 파생된 개념으로, 각각의 피쳐를 제거했을때와 제거하지 않았을 때 모델의 퍼포먼스를 비교하는 방식으로 피쳐의 중요도를 측정한다. Peng 외(2021)에서는 간염 환자의 징조를 의사가 알 수 있도록 도울 수 있는 XAI 프레임워크를 제안했다. 해당 연구에서 저자는 의사 결정 트리, 로지스틱 회귀분석, kNN과 같은 XAI 프레임워크를 상대적으로 복잡한 모형들과 비교하였고 해당 결과를 SHAP, LIME과 같은 피쳐 중요도 모형을 더하여 비교하였다.

텍스트 기반의 의료 기록 데이터를 인공지능 모형을 활용하여 분석한 경우도 있지만 이미지 데이터를 처리하는 과정을 XAI 프레임워크를 적용한 연구 사례 또한 존재한다. Sarp 외(2021)의 연구는 처음 이미지를 처리하는 심층 신경망 모델인 CNN을 기반으로 한 모델을 만성 상처를 분류하는 모형을 개발한 뒤 LIME(Local Interpretable Model-agnostic Explanation)을 적용하여 결과를 설명하였고 평균 정밀도는 95%를 달성하였다. Gu 외(2020)에서는 VINet이라는 인공지능 질병 진단 지원 시스템을 제안했는데 저자는 해당 모델을 CAM, LRP와 같은 기존의 설명 가능한 인공지능(XAI)과 성능을 비교하였다.

이처럼 해가 지날수록 이미지를 비롯한 의료 데이터를 활용하여 단순히 질병을 예측할 뿐 아니라, 해당 예측치가 나온 원인까지 설명 가능한 인공지능(XAI)을 활용하여 분석하는 연구는 발전하고 있다. 해당 정교함은 연구가 지속될수록 더해지고 있고, 의료 데이터의 크기와 복잡성을 고려할 때 적합한 모델과 함께 발전이 지속된다면 의료 전문가의 판단에 준하는 정확도의 성능을 보여주는 결과를 볼 수 있을 것으로 기대된다. 또한 설명 가능한 인공지능(XAI)를 활용한다면 해당 의사결정에 대한 원인 또한 분석이 가능하므로 의학 학문

의 발전에도 기여할 수 있을 것으로 기대된다.

□ 의료 분야에서 활용된 설명 가능한 인공지능(XAI) 연구의 비전

질병 진단 이외의 분야에서도 설명 가능한 인공지능(XAI)이 적용될 수 있는 연구와 적용은 지속되고 있다. CT, X-ray 이미지를 CNN 기반 모형에 설명 가능한 인공지능(XAI)을 적용하여 환자의 상태를 진단하고 환자의 질병 발병 확률 계산한 뒤, 어떠한 요인들이 해당 결과에 유의미한 영향을 끼쳤는지를 분석하는 등의 연구를 예시로 들 수 있다. Yoo 외(2020)의 연구의 경우 멀티클래스 XGBoost 모형을 활용하여 전문가 수준에서 레이저 수술 옵션을 선택할 수 있음을 보였다. 해당 연구에서는 이해를 돕기위해 SHAP 값을 활용하여 피쳐 중요도를 조명하고 있다.

이처럼 심층 신경망과 같은 인공지능 모형의 발달은 최근으로 갈수록 의료 분야에서 큰 역할을 하고 있다. 일부 심층 신경망 기반 질병 진단 작업의 정확성은 인간 의료 전문가의 능력치를 상회하기도 한다. 인공지능이 발달함에 따라 연구자들은 결국 모델의 정확성이 아닌 설명 가능성이 의료 분야에서 중요함을 깨달은 것이다. 근래 연구들의 비중을 관찰하였을 때, Zhang 외(2022)의 현황 조사에 의하면 설명 가능한 인공지능(XAI)를 의료분야에 적용한 연구들에서 사용한 알고리즘 중 37%는 CNN 기반의 이미지 처리 인공지능 모형을 활용했으며 LIME모형을 연구에서 25.9% 활용하였고 해당 모형의 사용 빈도가 가장 높았다. 즉, CNN 기반의 인공지능 모형에 LIME, SHAP을 첨가한 인공지능 모형이 가장 빈도 높게 쓰인 것이다.

Ⅳ. 시사점 및 앞으로의 비전 제시

의료 분야에서의 심층 신경망 모델을 결정하고 이를 설명하는 것은 근래 상당한 진전을 보이고 있다. 인공 지능 모형의 의사결정을 이해하는 것은 모델 설계자가 최종 사용자의 신뢰를 얻는다는 점에서 굉장히 중요한 이슈이다. 심층 신경망 실무자는 모델 특징을 분석하고 모델과 데이터 사이의 상호 작용을 이해함으로서 더 나은 시스템을 설계하는데 사용할 수 있다. 최종 사용자는 심층 신경망의 의사결정에 대한 설명을 제공받음으로서 모형의 결정에 대한 신뢰를 쌓고 잠재적으로 의심스러운 결정을 식별하는데 도움을 받을 수 있다.

심층 신경망은 점점 더 정교해지고 유용해지는 동시에 인간의 직관이 따라갈 수 없는 결

정을 내리는 경우가 있다. 의학적 분야에서 인간이 직관이 이해할 수 없는 결정의 경우 보수적으로 대응할 수 밖에 없는 실정이다. 그러므로 설명 가능한 인공지능(XAI)의 발전이 꼭 수반되어야하며 발전은 지금도 진행중에 있다.

참고문헌

한국정보화진흥원(2018), "설명 가능한 인공지능(eXplainable AI, XAI) 동향".

Bach, S., Binder, A., Montavon, G., Klauschen, F., M ller, K. R., & Samek, W.(2015). On pixel-wise explanations for non-linear classifier decisions by layer-wise relevance propagation. PloS one, 10(7), e0130140.

Geurkink, Y., Boone, J., Verstockt, S., & Bourgois, J. G.(2021). Machine learning-based identification of the strongest predictive variables of winning and losing in Belgian professional soccer. Applied Sciences, 11(5), 2378.

Kavya, R., Christopher, J., Panda, S., & Lazarus, Y. B.(2021). Machine learning and XAI approaches for allergy diagnosis. Biomedical Signal Processing and Control, 69, 102681.

Amoroso, N., Pomarico, D., Fanizzi, A., Didonna, V., Giotta, F., La Forgia, D., ... & Massafra, R.(2021). A roadmap towards breast cancer therapies supported by explainable artificial intelligence. Applied Sciences, 11(11), 4881.

El-Sappagh, S., Alonso, J. M., Islam, S. M., Sultan, A. M., & Kwak, K. S.(2021). A multilayer multimodal detection and prediction model based on explainable artificial intelligence for Alzheimer's disease. Scientific reports, 11(1), 1-26.

Peng, J., Zou, K., Zhou, M., Teng, Y., Zhu, X., Zhang, F., & Xu, J.(2021). An explainable artificial intelligence framework for the deterioration risk prediction of hepatitis patients. Journal of Medical Systems, 45(5), 1-9.

Sarp, S., Kuzlu, M., Wilson, E., Cali, U., & Guler, O.(2021). The enlightening role of explainable artificial intelligence in chronic wound classification. Electronics, 10(12), 1406.

Gu, D., Li, Y., Jiang, F., Wen, Z., Liu, S., Shi, W., ... & Zhou, C.(2020). VINet: A visually interpretable image diagnosis network. IEEE Transactions on Multimedia, 22(7), 1720-1729.

Yoo, T. K., Ryu, I. H., Choi, H., Kim, J. K., Lee, I. S., Kim, J. S., ... & Rim, T. H.(2020). Explainable machine learning approach as a tool to understand factors used

to select the refractive surgery technique on the expert level. Translational vision science & technology, 9(2), 8-8.

Zhang, Y., Weng, Y., & Lund, J.(2022). Applications of Explainable Artificial Intelligence in Diagnosis and Surgery. Diagnostics, 12(2), 237.