

XAI 기반의 임상 의사결정시스템에 관한 연구

A Study on XAI-based Clinical Decision Support System

안윤애*, 조한진**

한국교통대학교 컴퓨터공학전공*, 극동대학교 에너지IT공학과**

Yoon-Ae Ahn(yeahn@ut.ac.kr)*, Han-Jin Cho(hanjincho@kdu.ac.kr)**

요약

임상 의사결정시스템은 누적된 의료 데이터를 활용하여 머신러닝으로 학습된 AI 모델을 환자의 진단 및 진료 예측에 적용한다. 그러나 기존의 블랙박스 기반의 AI 응용은 시스템이 예측한 결과에 대해 타당한 이유를 제시하지 못하여 설명성이 부족한 한계점이 존재한다. 이와 같은 문제점을 보완하기 위해 이 논문에서는 임상 의사결정시스템의 개발 단계에서 설명이 가능한 XAI를 적용하는 시스템 모델을 제안한다. 제안 모델은 기존의 AI 모델에 설명성이 가능한 특정 XAI 기술을 추가로 적용시켜 블랙박스의 한계점을 보완할 수 있다. 제안 모델의 적용을 보이기 위해 LIME과 SHAP을 활용한 XAI 적용 사례를 제시한다. 테스트를 통해 데이터들이 모델의 예측 결과에 어떤 영향을 미치는지 다양한 관점에서 설명할 수 있다. 제안된 모델은 사용자에게 구체적인 이유를 제시함으로써 사용자의 신뢰를 높일 수 있는 장점을 가진다. 아울러 XAI의 적극적인 활용을 통해 기존 임상 의사결정시스템의 한계를 극복하고 더 나은 진단 및 의사결정 지원을 가능하게 할 것으로 기대한다.

■ 중심어 : | 임상 의사결정시스템 | 설명 가능한 AI | 머신러닝 | 피쳐 기반 모델 | 진단 예측 |

Abstract

The clinical decision support system uses accumulated medical data to apply an AI model learned by machine learning to patient diagnosis and treatment prediction. However, the existing black box-based AI application does not provide a valid reason for the result predicted by the system, so there is a limitation in that it lacks explanation. To compensate for these problems, this paper proposes a system model that applies XAI that can be explained in the development stage of the clinical decision support system. The proposed model can supplement the limitations of the black box by additionally applying a specific XAI technology that can be explained to the existing AI model. To show the application of the proposed model, we present an example of XAI application using LIME and SHAP. Through testing, it is possible to explain how data affects the prediction results of the model from various perspectives. The proposed model has the advantage of increasing the user's trust by presenting a specific reason to the user. In addition, it is expected that the active use of XAI will overcome the limitations of the existing clinical decision support system and enable better diagnosis and decision support.

■ keyword : | CDSS | Explainable AI | Machine Learning | Feature-based Model | Diagnostic Prediction |

* 본 연구는 2020년 한국교통대학교의 지원을 받아 수행하였음.

접수일자 : 2021년 09월 30일

수정일자 : 2021년 10월 14일

심사완료일 : 2021년 10월 14일

교신저자 : 조한진, e-mail : hanjincho@kdu.ac.kr

I. 서론

병원 내의 전자의무기록시스템(EMR/EHR)을 활용하여 환자의 진단 및 치료에 보조적으로 활용하기 위해 사용되는 시스템을 임상 의사결정시스템(Clinical Decision Support System : CDSS)이라고 한다[1]. 최근에는 AI 기술을 접목시켜 좀더 우수한 임상 진단 및 치료 시스템을 만들고자 하는 연구가 활발하다.

이와 같이 의료분야에서 AI 기술을 도입하여 진료지원을 도와주는 보조 시스템에 활용하는 이유는 의료진이 좀 더 신속하고 정확하게 진료 의사결정을 내릴 수 있도록 지원하는 것이며, 여러 의료기관의 의료진이 표준화된 의료 서비스를 제공하도록 하는 것이다[2]. 의료분야에서 AI 기술이 활용되는 유형은 임상 의사결정 시스템 분야, 의료영상의 판독 분야, 환자의 질병 예측 분야, 빅데이터 기반의 환자 맞춤형 정밀의료 분야가 대표적이다[3].

지금까지 AI 응용시스템은 의사결정을 수행할 때 결과만 알려주고 어떤 이유와 근거로 결과를 예측했는지 논리적으로 설명하지 못하는 점이 한계로 지적되었다[4]. 이러한 머신러닝의 블랙박스 측면은 AI를 다양한 분야의 응용에 적용하기 위해 반드시 보완해야 할 문제이다. 특히 의료 진단 및 예측과 같이 높은 정확도가 요구되는 분야에서는 시스템의 결과가 애매모호한 경우 전문가가 재검토할 수 있도록 하는 보완적인 기능이 필요하다[4].

따라서 AI를 활용한 질병 진단 및 의료분야의 다양한 의사결정에서 시스템이 제시하는 결과를 사용자에게 설명해야 하는 필요성이 더욱 증가하고 있다. AI 임상 의사결정시스템이 환자의 질병 유형 및 진단 결과의 이유를 자세하게 설명할 수 있는 설명성을 가진다면 AI를 사용하는 의료 종사자의 신뢰 수준을 높이는 장점을 가질 수 있다[5]. 여기에서 설명성(Explainability)이란 특정 AI 시스템이 예측한 결과에 대해 타당한 이유 및 근거를 제시할 수 있는 시스템의 특성을 의미한다[5].

이 연구에서는 지금까지 개발된 블랙박스 형태의 AI 임상 의사결정시스템이 가지는 설명성 부족의 문제점을 보완하기 위해, 임상 의사결정시스템의 개발 단계에서 설명이 가능한 AI(Explainable AI)를 적용하는 시스템

모델을 제안한다. 아울러 임상료 전문 사용자가 임상 의사결정시스템이 제시하는 결과에 대한 최소한의 설명자료를 제공받을 수 있는 XAI 기술의 구현을 통해 설명 유형의 사례를 제시한다.

제안하는 XAI 기반의 CDSS 모델에서는 분류기법의 모델 훈련 알고리즘인 XGBoost를 활용한다. XGBoost 알고리즘은 성능과 속도를 위해 개발된 Gradient Boosted Decision Tree의 구현으로, 부스팅 프레임워크를 기반으로 하는 의사결정 트리 기반 기계학습 알고리즘이다. 이것은 XAI의 필수인 기능 중요성을 도출할 수 있는 유일한 알고리즘의 하나로, XGBoost는 모든 종류의 분류 문제에 매우 적합하다[6]. 이를 통해 제안한 XAI 기반 임상 의사결정시스템 모델은 기존 시스템 모델에 비해 투명성 및 설명성이 보완된 예측 결과를 제공하는 장점을 가질 수 있다. 또한 임상 전문가들의 판단을 좀 더 명확하게 지원해주는 시스템으로 발전할 것이다.

이 논문의 구성 및 내용은 다음과 같다. 먼저 2장의 관련연구에서는 임상 의사결정시스템의 개요, 설명 가능한 AI의 개요 및 관련 기술의 특징, 기존의 AI 기반 임상 의사결정시스템의 한계점을 제시한다. 3장에서는 설명 가능한 AI를 결합한 임상 의사결정시스템 모델을 제안한다. 4장에서는 제안 모델의 적용을 위한 XAI 구현 사례를 통해 설명성을 제시하는 형태 및 이해 방법에 대해 기술하고, 제안 모델의 장점 및 한계점을 검토한다. 마지막으로 5장에서는 결론 및 향후 연구방향을 제시한다.

II. 관련 연구

1. 임상 의사결정시스템

CDSS는 의료 지식과 환자 데이터를 활용하여 환자에게 제공되는 서비스 품질을 개선하기 위한 권장 사항을 제공하고, 의사결정 지원 요청에 응답하는 컴퓨터 기반 시스템이다[7]. 의료를 위해 전자의무(건강)기록을 분석하고 활용하는 CDSS는 환자 중심의 의료를 추구하며, 의료를 개선하고 환자의 안전을 높이면서 진료 비용을 절감할 수 있는 가능성을 보여주었다[8]. 그러나

부적절하거나 결함이 있는 CDSS는 잠재적으로 의료의 질을 저하시키고 환자를 위험에 빠뜨릴 수 있다. 또한, 의도된 사용자가 신뢰, 관련성 또는 실행 가능성의 부족으로 인해 시스템의 결과를 무시할 수 있으므로, CDSS의 채택이 실패할 수 있는 문제점도 발생한다[9].

CDSS의 궁극적인 목적은 방대한 양의 정보 처리를 시스템이 실행함으로써, 임상전문가가 보다 신속하고 명확하게 의사결정을 할 수 있도록 지원하는 것이다[9]. 시스템 목적의 효과를 높이기 위해서 최근 AI 기반의 CDSS에 관한 연구가 활발하다. 대표적인 사례로는 진료지원을 위해 머신러닝을 사용하는 시스템, 환자의 위험을 예측하는 시스템, 헬스 시스템의 학습을 위한 시스템 등이 있다[10].

그러나 다른 분야와 마찬가지로 의료분야에서도 매우 대중적으로 사용되는 AI 알고리즘은 진단 및 예측 결과에 대한 직접적인 설명성을 구체적으로 제시하지 못하는 한계점으로 인해 임상 전문가의 신뢰를 얻지 못하는 제약사항을 가진다[7]. 따라서 최근에는 설명성의 제공이 가능한 XAI를 CDSS에 적용하려는 연구가 대두되고 있다. 의학 및 보건의료 분야에 적용된 AI 시스템의 결과는 높은 정확성 및 정당성을 보유했어야 하며, 의료 전문가가 시스템을 활용하여 명확한 임상 의사결정을 내릴 수 있도록 도와주는 XAI의 역할이 중요하다 [11].

2. XAI(Explainable AI)

XAI는 AI 시스템 결과를 사용자가 쉽게 이해하고 시스템이 제공하는 근거를 토대로 명확한 의사결정을 내릴 수 있도록 하는 연구 분야이다. XAI라는 용어는 2004년에 Van Lent에 의해 처음 만들어졌다. 1970년대 중반 연구자들이 전문가 시스템에 대한 설명을 연구한 이후로 설명성 문제가 존재했으며, XAI는 2016년부터 급격하게 관심이 높아진 연구분야로 AI의 신뢰도 및 투명성을 높일 수 있는 하나의 해결책으로 대두되고 있다[12]. XAI 기술이 도입된 시스템의 목표는 머신러닝 결과에 대해 인간의 가독성을 증진시키고, 머신러닝이 내린 결정이 납득 가능한지 설명하고, 결정에 대한 책임을 명확히 하고, 차별을 피하고, 사회적 편향을 줄이는 것이다[13]. 보건의료 분야에서 XAI는 현재 신경과

학 및 행동 과학, 유전체학, 유방암, 만성질환, 생의학 및 공중보건의 여러 분야에서 광범위하게 응용되고 있다[11].

설명가능 AI 분야에서 연구된 XAI 설명 기술은 크게 피쳐기반 모델과 예제기반 모델로 구분된다. 피쳐기반 모델의 설명 가능성 기술(Feature-based Techniques)은 입력 피쳐가 모델의 출력에 얼마나 기여하는지를 나타내는 것을 의미한다. LIME(Local Interpretable Model-Agnostic Explanations)은 전체 모델을 설명하지 않는 방법으로 로컬 모델 즉, 특정 인스턴스 하나에 대한 해석 및 설명이 가능하다[7]. LIME을 활용하면 AI 응용시스템이 제공한 예측 결과값에 대해, 시스템의 입력에 활용된 개별 인스턴스의 피쳐 값들을 토대로 가중치 값과 예측에 대한 기여도를 상세하게 설명할 수 있는 장점을 가진다[6].

SHAP(SHapley Additive exPlanations)은 모델의 투명도를 높이는 방법으로 로컬 설명 방법으로 작동하는 방식을 보여주기 위해 개별 인스턴스에서 실행하거나, 글로벌 설명을 위해 수집된 SHAP 값이 최종 예측에 긍정/부정적으로 기여하는 방식을 보여준다[7]. SHAP을 활용하면 AI 응용시스템이 제공한 예측 결과값에 대해, 개별 인스턴스에 대한 해석, 특정 피쳐 간의 관계에 대한 해석, 모든 인스턴스에 대한 종합적인 해석, 모든 피쳐들의 모델 영향력에 대한 해석 등의 다양한 방법으로 예측 결과에 대해 상세하게 설명하고 분석할 수 있는 장점을 가진다[6].

예제기반 모델의 설명 가능성 기술(Example-based Techniques)은 인스턴스가 모델의 출력에 기여하는 정도를 나타내는 것이다. 여기에는 Anchors, Counterfactuals, Counterfactuals with Prototypes, CEM(Contrastive Explanation Methods), Kernel Shapley, Integrated Gradients와 같은 기술들이 대표적이다[7].

3. 기존 임상 의사결정시스템의 한계점

임상의학에서 AI 기반 CDSS는 의사의 진단 및 치료의 결정 과정을 돕는 역할을 한다. 기존의 CDSS는 지식베이스에 저장된 데이터들을 활용하여 개별 환자의 특징과 유사하거나 일치하는 경우를 예측한다. 이에 반

해 AI 기반 CDSS는 누적된 의료 데이터를 활용하여 머신러닝으로 학습된 AI 모델을 환자의 진료 예측에 적용하게 된다. 그러나 AI 모델은 아직 해결 및 보완할 점이 많이 존재하고 있는 실정이다. 그 이유 중 대표적인 한 가지는 예측 결과에 대한 설명의 기능이 부족하다는 점이다. 설명성은 AI 시스템이 예측한 결과에 대해 타당한 이유 및 근거를 제시할 수 있는 특성을 의미한다[5].

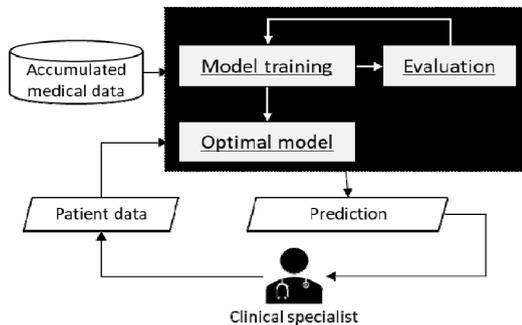


그림 1. 블랙박스 형태의 CDSS

[그림 1]은 블랙박스 형태의 간단한 CDSS 구조를 나타낸 것이다. 이러한 블랙박스 형태의 모델에서는 최적 모델(Optimal Model)의 정확도는 최대한으로 높일 수 있을지도 모른다. 그러나 어떤 부분에서도 모델이 제공하는 처리결과에 대한 합리적인 이유를 설명하는 부분이 보이지 않는다. 따라서 기존의 블랙박스 기반의 AI 응용은 이와 같은 설명성의 중요도 및 가치를 완벽하게 지원하기에는 아직 한계점이 존재한다[10]. 이와 같은 한계점을 보완하기 위한 솔루션 중 하나가 설명 가능한 AI(XAI) 기술을 모델에 적용하는 것이며, XAI를 기반으로 하는 CDSS의 연구가 필요한 계기가 되었다. 이 논문에서는 XAI를 적용한 CDSS 모델을 제안하고, 피쳐기반 설명 가능성의 결과를 제공하는 LIME과 SHAP 기술을 당뇨병 데이터 세트에 적용하고 그 결과를 해석하여 설명성의 특성을 검토한다.

III. XAI 기반의 임상 의사결정시스템 모델

설명성이 추가된 CDSS 모델은 기존의 AI 모델을 적용한 시스템 구조에 설명성이 가능한 특정 XAI 기술을

추가로 적용시켜 블랙박스의 한계점을 보완하는 구조이다. 이때 모델 훈련에 사용 가능한 AI 모델은 일반적인 머신러닝 알고리즘을 다양하게 적용할 수 있다. 평가 분석을 통해 최적의 모델을 찾은 후 개별 환자의 예측 결과에 대한 상세 설명이 필요할 때 최적 모델을 XAI 알고리즘에 적용하여 분석 결과를 얻는 것이다.

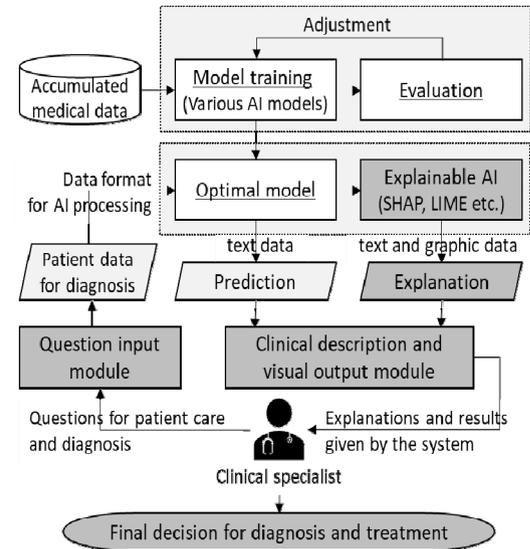


그림 2. XAI와 결합된 CDSS 모델

[그림 2]는 기존의 AI 최적 모델에 XAI 기술을 한 번 더 적용하는 형태의 시스템 구성도이다. 최적 모델에서 예측한 결과 값을 그대로 임상 전문가에게 제시할 수도 있고, XAI 알고리즘에 입력하여 그래픽 형태의 설명 및 예측 결과를 동시에 제공할 수도 있다. 임상 전문가는 텍스트 및 그래픽 결과를 통해 시스템의 예측 결과에 대한 상세 설명을 제공받으며, 본인의 임상 경험과 시스템의 다양한 분석결과를 종합적으로 판단하여 진단 및 치료 방법을 결정하면 된다. 이러한 과정을 통해 임상 전문가는 시스템에서 제공되는 예측 결과 및 설명과 함께 임상 기록을 사용하여 더 수준 높은 진단 예측 통찰력과 진료 권장 사항을 생성할 수 있다.

IV. 제안 모델 기반의 XAI 적용 및 해석

1. 모델 훈련

이 논문에서는 Kaggle[14]에서 제공하는 피마 인디언 당뇨병 데이터 세트를 사용하여 적용한 사례를 기술한다. 테스트의 목적은 당뇨병 환자를 예측하고 이에 대한 적합한 설명을 제시하는 것이다. 당뇨병은 만성질환의 하나로서 당뇨병 진단을 위한 예측은 CDSS 분야에서 많이 응용되는 영역 중 하나이다. 모델 훈련을 위해 XGBoost를 사용하였으며 데이터 세트는 다음과 같다.

	Pregna	Glucose	BloodPr	SkinTh	Insulin	BMI	DiaPediFun	Age	Out
0	6	148	72	35	0	33.6	0.627	50	1
1	1	85	66	29	0	26.6	0.351	31	0
2	8	183	64	0	0	23.3	0.672	32	1
3	1	89	66	23	94	28.1	0.167	21	0
4	0	137	40	35	168	43.1	2.288	33	1
5	5	116	74	0	0	25.6	0.201	30	0
6	3	78	50	32	88	31.0	0.248	26	1
7	10	115	0	0	0	35.3	0.134	29	0
8	2	197	70	45	543	30.5	0.158	53	1
9	8	125	96	0	0	0.0	0.232	54	1

(a) 실제 데이터 값

	Pregna	Glucose	BloodPr	SkinTh	Insulin	BMI	DiaPediFun	Age	Out
0	6	148.0	72.0	35.0	80.0	34.0	0.627	50	1
1	1	85.0	66.0	29.0	80.0	27.0	0.351	31	0
2	8	183.0	64.0	21.0	80.0	23.0	0.672	32	1
3	1	89.0	66.0	23.0	94.0	28.0	0.167	21	0
4	0	137.0	40.0	35.0	168.0	43.0	2.288	33	1
5	5	116.0	74.0	21.0	80.0	26.0	0.201	30	0
6	3	78.0	50.0	32.0	88.0	31.0	0.248	26	1
7	10	115.0	69.0	21.0	80.0	35.0	0.134	29	0
8	2	197.0	70.0	45.0	543.0	30.0	0.158	53	1
9	8	125.0	96.0	21.0	80.0	32.0	0.232	54	1

(b) 누락된 값을 평균값으로 대체

그림 3. 데이터 세트 예제

[그림 3]과 같이 피마 인디언 당뇨병 데이터 세트는 모두 9개의 필드로 구성되며, 이 중 8개는 당뇨병 유무를 예측하는 피쳐로 사용되고 마지막 1개의 필드는 예측 결과를 판별하는 값으로 사용한다. 8가지 피쳐는 Pregna(임신히수), Glucose(포도당농도), BloodPr(최소혈압), SkinTh(삼두근피부두께), Insulin(혈청인슐린), BMI(체질량지수), DiaPediFun(당뇨내력 가중치

값), Age(나이) 이고, Out(질병유무)는 1 또는 0으로 구분된다. 전체 데이터 세트에 대한 describe 함수 결과를 통해서 Pregna, Glucose, BloodPr, SkinTh, Insulin, BMI 값에 결측치가 '0'으로 되어있는 것을 확인하였으며, 이 중에서 Pregna(임신히수)를 제외한 나머지 피쳐에 대해 평균값으로 대체하여 테스트에 사용하였다.

2. 최적 모델 생성

모델 훈련에 사용된 XGBoost(eXtreme Gradient Boosting)는 XAI의 필수인 기능 중요성을 도출할 수 있는 유일한 알고리즘 중의 하나로 모든 종류의 분류 문제에 매우 적합한 특징을 가진다[6]. XGBoost 알고리즘의 모델 훈련을 위해서는 XGBClassifier 함수에 사용될 최적의 파라미터를 찾아야 한다. 여기에서는 [15]에서 GridSearchCV를 통해 찾아낸 파라미터 값을 사용하여 최적화 모델을 생성하였다.

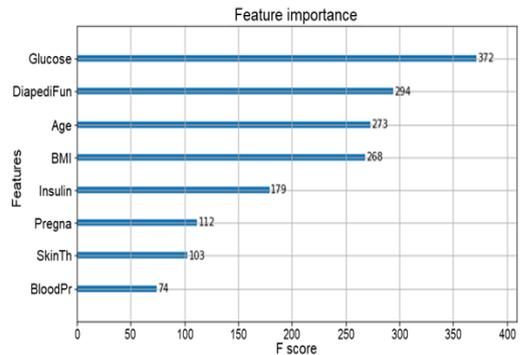


그림 4. XGBoost 모델의 피쳐 중요도

[그림 4]는 XAI를 적용하기 전에 XGBoost로 실행한 피쳐 중요도를 그래프로 나타낸 것이다. 학습 모델 자체에서 판단하는 피쳐 중요도와 XAI 기술을 적용한 이후의 결과는 서로 상이할 수 있다. [그림 4]와 같이 모델 학습에 사용된 XGBoost 알고리즘이 제시하는 결과와 4.3.2절의 [그림 11]에서 제시된 SHAP 알고리즘 적용 후의 결과가 정확하게 일치하지는 않으므로, 임상 전문가가 비교 분석하여 진단 예측의 근거 및 설명에 대한 특성을 분석하는데 활용할 수 있다.

3. XAI를 적용한 설명 가능성의 해석

최적 모델이 완성되면 이제 시스템은 개별 인스턴스의 피쳐들을 토대로 당뇨병 유무를 예측하고, XAI 기술을 적용하여 예측 결과에 대한 설명을 그래픽 및 텍스트 형태로 제시할 수 있다. 테스트용 인스턴스 데이터는 도출된 최적 모델과 XAI 모듈에 제공된다. 여기에서는 XAI 적용 기술로 LIME과 SHAP 2가지 방법으로 처리하여 예측 결과와 함께 제시된 설명성을 해석한다.

3.1 LIME을 활용한 설명

LIME은 개별 인스턴스의 피쳐 값들을 토대로 XGBoost 모델이 예측한 결과 값에 대해 로컬 설명(Local Explanation)을 제공하는 XAI 기술이다. [그림 5]와 [그림 6]은 LIME의 LimeTabularExplainer, explain_instance 함수를 사용하여 당뇨병 예측에 대해 설명한 것이다.

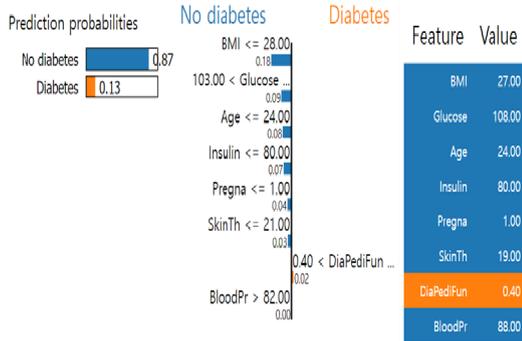


그림 5. 당뇨병이 아닌 것으로 예측된 사례

[그림 5]의 좌측에서는 87%의 확률로 당뇨병이 아니라는 예측 결과를 제시하고 있으며, 이 확률은 XGBoost 모델이 제시한 결과이다. 그림의 우측에서는 BMI(27.0), Insulin(80.0), Glucose(108.0), Age(24), Pregna(1.0), SkinTh(19.0), DiaPediFun(0.4), BloodPr(88.0)의 피쳐 값을 보여주고 있다. 그림의 중앙에서는 각 피쳐 값이 최종 예측에 영향을 미친 가중치와 예측에 영향을 미치는 값의 범위를 보여준다. [그림 5]의 결과 8개의 피쳐 중 유일하게 DiaPediFun이 0.4이고, 당뇨병으로 예측 가능한 가중치가 0.02임을 알 수 있다. 따라서 나머지 7개의 피쳐 값은 모두 당뇨

병이 아니라고 예측하는 데 기여하였으며, 이 중에서 BMI 값이 가중치 0.18로 결과에 가장 영향력이 높은 피쳐라고 해석할 수 있다.

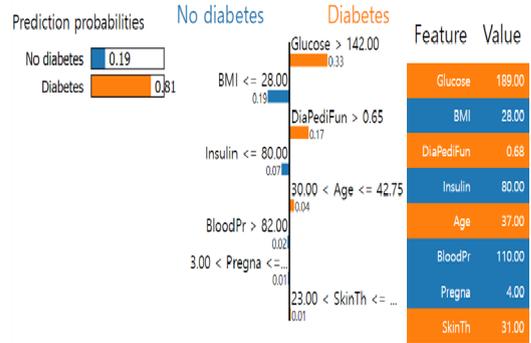


그림 6. 당뇨병으로 예측된 사례

[그림 6]에서는 최종적으로 당뇨병일 확률이 81%라고 예측하였다. 4개의 피쳐 Glucose, DiaPediFun, Age, SkinTh가 당뇨병으로 예측한 결과에 기여했음을 설명하고 있다. 특히 환자의 Glucose 수치가 189.0이므로 기준값 142.0보다 크고 이는 0.33의 가중치로 결과에 가장 크게 기여한 피쳐라고 해석할 수 있다.

3.2 SHAP을 활용한 설명 방법

(1) 개별 인스턴스에 대한 해석

SHAP의 force_plot 함수를 사용하면 개별 인스턴스에 대한 로컬 설명이 가능하다. 당뇨병 예측의 결과를 실행한 사례를 2가지로 비교하여 설명한다. 먼저 [그림 7]은 당뇨병이 아닌 것으로 예측된 사례이다.

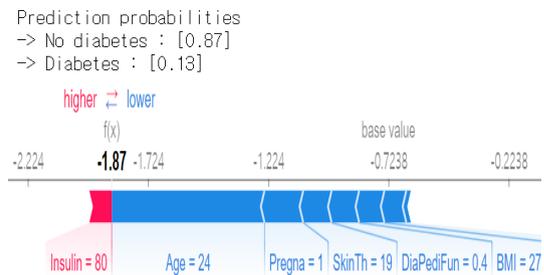


그림 7. 당뇨병이 아닌 것으로 예측된 사례

[그림 7]과 같이 SHAP의 실행 결과는 각 인스턴스의

피쳐 값에 따라 계산된 샐플리 값(Shapely Value)을 기준으로 제공된다. 표기된 값 중에서 가장 진하게 출력된 샐플리 값이 -1.87이다. 이는 환자가 당뇨병이 아님을 예측한 것이다. 이 예측 결과에 영향을 미친 가장 중요한 피쳐는 Age, Pregna, SkinTh, DiaPediFun 순으로 나타났음을 확인할 수 있다.

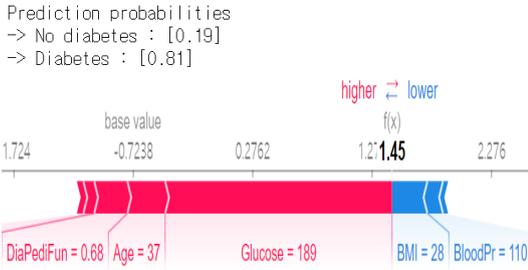


그림 8. 당뇨병으로 예측된 사례

[그림 8]에서는 가장 진하게 출력된 샐플리 값이 1.45이다. 이는 환자가 당뇨병이라고 예측한 결과이다. 이 예측의 결과에 영향을 미친 가장 중요한 피쳐는 Glucose, Age, DiaPediFun 순으로 나타났음을 확인할 수 있다.

(2) 특징 피쳐 간의 관계에 대한 해석

다음은 여러 가지 피쳐들 중에서 2가지 항목 간의 관계와 샐플리 값을 보여주는 것이다.

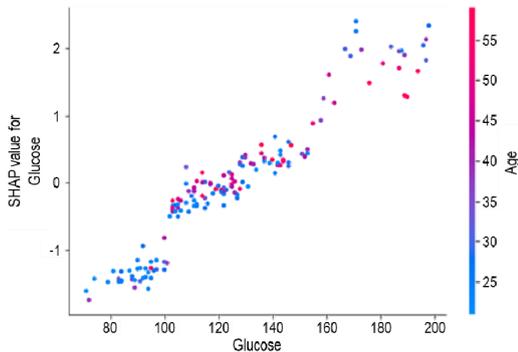


그림 9. 피쳐 Glucose와 Age의 관계

[그림 9]에서 x축은 Glucose 수치이고, y축은 샐플리 값을 나타낸다. Age와 Glucose 수치가 적을수록

샐플리 값이 낮게 나타나고, Age와 Glucose 값이 커질수록 샐플리 값도 커지는 특징을 보인다. 결국 샐플리 값이 커질수록 당뇨병 환자로 예측되고 있음을 이해할 수 있다.

(3) 모든 인스턴스에 대한 종합적인 해석

SHAP의 글로벌 설명(Global Explanation)은 테스트에 활용되는 모든 인스턴스 집합에 대한 예측 결과를 한번에 확인할 수 있는 기능이다. 여기에서는 모두 154개의 인스턴스가 사용되었다.

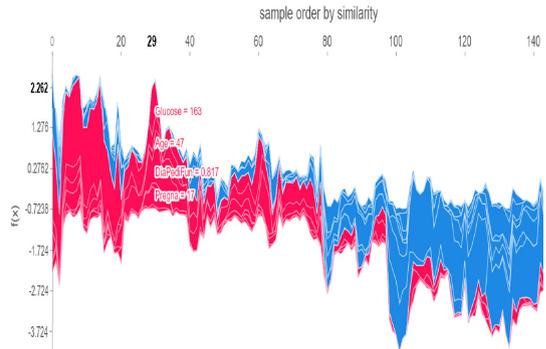
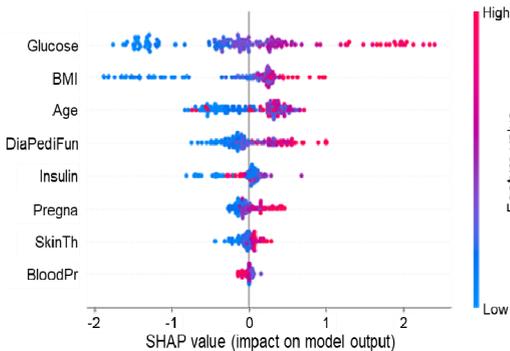


그림 10. 모든 인스턴스에 대한 포괄적인 결과

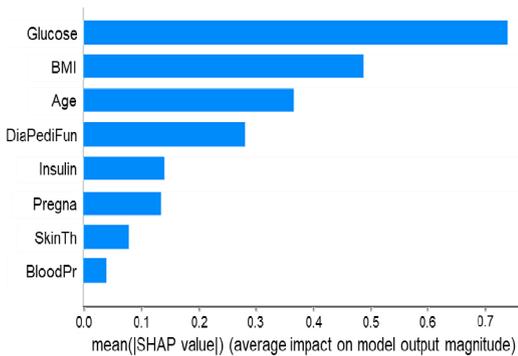
[그림 10]은 154개의 인스턴스에 대한 설명을 종합적으로 하나의 플롯으로 나타낸 것이다. 상단 x축의 숫자 중에서 진하게 출력된 29는 해당 위치에 마우스를 올렸을 경우에 보여주는 화면이고, 29번째 인스턴스의 샐플리값은 y축에서 2.262로 나타난다. 따라서 29번째 환자에 대한 예측 결과는 '당뇨병이다'로 해석되며, 이 결과에 영향을 준 피쳐는 Glucose, Age, DiaPediFun, Pregna 순으로 설명되고 있다.

(4) 모든 피쳐의 모델 영향력에 대한 해석

SHAP의 summary_plot 함수를 사용하면 모든 피쳐들이 모델에 미치는 영향력을 샐플리 값의 발생 범위에 따라 산점도 그래프 또는 막대 그래프로 한눈에 파악할 수 있다.



(a) 산점도 그래프 형식의 종합적인 결과



(b) 막대 그래프 형태의 종합적인 결과

그림 11. 모든 기능의 전반적인 영향에 대한 설명

[그림 11]의 (a)는 산점도 그래프를 통해 8개의 피쳐 값의 변화가 새플리 값에 미치는 영향력을 설명하고 있다. 위쪽에 위치하는 피쳐들 즉, Glucose, BMI, Age는 아래쪽에 위치하는 BloodPr, SkinTH, Pregna 보다 모델에 더 많이 기여함을 알 수 있다. 특히 Glucose는 새플리 값의 전 범위(-2에서 +2)에 영향을 미치고 있으므로, 가장 영향력이 큰 피쳐임을 확인할 수 있는 것이다. (b)의 막대 그래프에서도 위쪽에 존재하는 피쳐가 모델의 새플리 값에 미치는 영향이 크다는 것을 알 수 있다.

4. 제안한 XAI 기반 CDSS 모델의 고찰

4.1 XAI 기술 적용 시 고려사항

지금까지 XAI 기술 중에서 LIME과 SHAP을 사용한 예시를 통해 기존의 AI 모델보다 좀 더 구체적으로 예측 결과에 대한 설명이 제공되는 방법을 살펴보았다.

제안한 CDSS 모델이 실제 현장에서 사용되기 위해 가장 먼저 고려할 사항은 시스템 개발에 사용할 XAI 기술의 선정이다. 이때 AI 모델과 연계되는 XAI 기술이 가지는 특성과 예측의 근거 및 설명성의 제시 방법을 고려해야 한다. 구축하려는 CDSS와 응용 데이터의 특성, 시스템 사용자의 전문성, 결과의 활용이 임상 현장에서 미치는 영향력의 중요도 등을 세밀하게 검토해야 한다. 아울러 시스템 사용자의 편의성 및 설명의 이해도를 최대한으로 높일 수 있는 인터페이스의 개발을 고려해야 한다. 그래야만 XAI 기술을 접목한 시스템의 효율성이 극대화된다고 할 것이다.

4.2 제안 모델의 장점

제안한 XAI 기반의 CDSS 모델은 기존의 시스템이 갖지 못하는 예측 결과의 투명성과 신뢰성을 제공할 수 있는 장점을 가진다. 제안 모델에 적용된 SHAP 기술은 예측을 위해 시스템에 입력된 피쳐들에 대해 계산된 새플리 값을 제공한다. 새플리 값은 각 피쳐들의 최종 예측 결과에 대한 기여도 및 긍정 또는 부정적으로 기여하는지를 알려주는 값이다. 이를 통해서 제안 모델에서 제시하는 예측 결과에 대한 투명성을 확보할 수 있으며, 사용자의 신뢰성을 높일 수 있는 것이다. 본 연구에서 제안한 모델은 CDSS 분야에 XAI를 적용하는 방법론의 하나로서 학술적인 의미를 가질 것이며, 향후 더 다양한 연구 방법의 기초 모델로서 참조될 수 있을 것이다.

4.3 제안 모델의 평가 방안

일반적으로 AI 시스템의 성능을 평가하는 것은 모델이 제시하는 예측 결과의 정확성을 가지고 판단할 수 있다. 그러나 제안한 XAI 기반의 CDSS 모델은 예측 결과의 정확성으로 평가받는 것이 아니라, 시스템이 사용자에게 제공한 예측 결과의 설명성에 대한 만족도를 바탕으로 성능을 평가해야 한다. 특히 다양한 임상전문가의 만족도 평가를 위해서는 2가지 이상의 XAI 기술을 적용한 시스템을 개발하여 비교 평가할 필요가 있을 것이다. 따라서 향후 제안 모델의 프로토타입을 개발한 후 설명성에 대한 사용자의 만족도를 조사하여 제안 모델의 우수성을 평가할 계획이다.

V. 결론

이 논문에서는 CDSS에서 머신러닝 기반의 AI를 사용하면서 신뢰성을 창출하기 위해 XAI 기술을 적용한 모델을 제시하였다. 또한 피마 인디언 당뇨병 데이터 세트를 사용하여 제안 모델의 적용 가능성을 테스트 사례를 통해 설명하였다. 이러한 XAI 기술의 적용 연구를 통해 얻은 가장 중요한 장점은 데이터들이 모델의 예측 결과에 어떤 영향을 미치는지 다양한 관점에서 설명할 수 있다는 것이다. 아울러 CDSS 사용자에게 좀 더 구체적인 근거를 제시함으로써 시스템 사용자에게 신뢰를 얻을 수 있다는 것이다.

향후 AI는 일상적인 임상 치료의 일부가 될 것이며, AI 모델의 임상적 사용은 현재 의료 서비스 제공 모델을 변화시킬 것으로 기대하고 있다[16]. 실제로 이들의 범위는 임상 설정을 넘어서 확장될 것이며, AI 및 XAI의 적극적인 활용을 통해 기존의 규칙기반 임상 의사결정시스템의 한계를 극복하고 더 나은 진단 및 의사결정 지원을 가능하게 할 것으로 기대한다.

현재 본 연구에서는 LIME과 SHAP이 도출한 결과를 그대로 사용하여 설명성을 해석하였다. 이로 인해 일반 사용자들은 결과를 해석하기에 다소 어려움을 겪을 수 있다. 따라서 향후에는 사용자에게 좀 더 쉬운 이해를 돕기 위한 별도의 HCI 기반의 설명 모듈 개발에 대한 연구를 구체적으로 수행할 계획이다.

본 연구의 한계점은 제한한 CDSS 모델을 아직 실제 시스템으로 구현하지 않은 점이다. 따라서 향후 연구과제로서 제안 모델의 프로토타입을 개발한 후에 사용자로부터 설명성에 대한 만족도를 평가받는 형태로 제안 모델의 객관적 타당성을 평가할 계획이다.

참고 문헌

- [1] M. Khalifa, "Clinical Decision Support: Strategies for Success," J. of Procedia Computer Science, Vol.37, pp.422-427, 2014.
- [2] 김용균, "디지털 헬스케어의 최근 동향과 시사점," 정보통신기술진흥센터, 주간기술동향, 제1846호, pp.12-23, 2018.
- [3] <https://blog.naver.com/neozensoft/221124735041>
- [4] M. Turek, *Explainable Artificial Intelligence(XAI)*, DARPA-BAA-16-53, pp.5-15, 2016.
- [5] U. Pawar, D. O'Shea, S. Rea, and R. O'Reilly, "Explainable ai in healthcare," 2020 International Conference on Cyber Situational Awareness, Data Analytics and Assessment (CyberSA), pp.1-2, 2020.
- [6] D. Dave, H. Naik, S. Singhal, and P. Patel, "Explainable AI meets Healthcare : A Study on Heart Disease Dataset," Cornell University, ArXiv: 2011.03195, pp.1-23, 2020.
- [7] U. Arioz, B. Yildiz, R. Kut, I. Agim, and K. Üğüdücü, "The Future of Applications For Clinical Decision Support Systems in Healthcare. Case Study: H2020 PERSIST Project," Proceedings of IES'20 International Engineering Symposium, Engineering Applications in Industry (Virtual), p.157, 2020.
- [8] H. Sung, B. Jung, K. Kim, S. Sung, A. Sung, and J. Park, "Trends and Future Direction of the Clinical Decision Support System in Traditional Korean Medicine," J. of Pharmacopuncture, Vol.22, No.4, pp.260-268, 2019.
- [9] G. Mahadevaiah, P. RV, I. Bermejo, D. Jaffray, A. Dekker, and L. Wee, "Artificial intelligence-based clinical decision support in modern medical physics: Selection, acceptance, commissioning, and quality assurance," J. of Medical Physics, Vol.47, No.5, pp.e228-e235, 2020.
- [10] T. Lysaght, H. Lim, V. Xafis, and K. Ngiam, "AI-Assisted Decision-making in Healthcare : The Application of an Ethics Framework for Big Data in Healthand Research," J. of Asian Bioethics Review, Vol.11, pp.299-314, 2019.
- [11] 한형진, "의료/헬스케어 분야에서의 설명 가능 인공지능(Explainable AI) 연구 동향," BRIC View 동향 리포트, BRIC View 2021-T13, pp.1-13, 2021.
- [12] A. Adadi and M. Berrada, "Peeking Inside the Black-Box: A Survey on Explainable Artificial

Intelligence (XAI),” J. of IEEE Access, Vol.6, pp.52138-52160, 2018.

[13] <https://towardsdatascience.com/googles-new-explainable-ai-xai-service-83a7bc823773>

[14] <https://www.kaggle.com/uciml/pima-indians-diabetes-database>

[15] 안재현, *XAI 설명 가능한 인공지능, 인공지능을 해부하다*, 위키북스, pp.79-90, 2020.

[16] S. Reddy, S. Allan, S. Coghlan, and P. Cooper, “A governance model for the application of AI in health care,” J. of the American Medical Informatics Association, Vol.27, No.3, pp.491-497, 2020.

저자 소개

안 윤 애(Yoon-Ae Ahn)

중신회원



- 2003년 2월 : 충북대학교 전자계산학과(이학박사)
- 2003년 3월 ~ 현재 : 한국교통대학교 컴퓨터공학전공 교수

〈관심분야〉 : 인공지능, 병원정보시스템, 빅데이터

조 한 진(Han-Jin Cho)

중신회원



- 2002년 8월 : 한남대학교 컴퓨터공학과(공학박사)
- 2002년 8월 ~ 현재 : 극동대학교 에너지IT공학과 교수

〈관심분야〉 : 인공지능, 정보보호, 클라우드