

# XGBoost 와 SHAP 을 활용한 병원전 이송단계 중증응급질환 예측 모델

<sup>1</sup> 전예림, <sup>2</sup> 권은정, <sup>2</sup> 박현호, <sup>2</sup> 이민정\*

<sup>1</sup> 순천향대학교 AI 빅데이터학과, <sup>2</sup> 한국전자통신연구원 과학치안공공 ICT 연구센터

9261190@sch.ac.kr, ejkwon@etri.re.kr, hyunhopark@etri.re.kr, \*minjunglee@etri.re.kr

## Pre-hospital diagnostic model of emergency patients using XGBoost and SHAP

<sup>1</sup>Ye Lim Jeon, <sup>2</sup>Eun Jung Kwon, <sup>2</sup>Hyunho Park, <sup>2</sup>Minjung Lee\*

<sup>1</sup> Soonchunhyang Univ., <sup>2</sup>ETRI

### 요약

본 논문은 중증응급질환 환자임을 병원전 이송단계에서 확보된 정보를 사용하여 예측하는 모델을 제안한다. 제안모델은 환자 이송 중에 구급대원이 작성하는 구급활동일지와 응급처치 세부상황표를 사용해 중증외상과 뇌정지를 비롯한 24 종의 중증응급질환을 이진 분류한다. 우리는 앙상블 모델의 한 종류인 XGBoost (eXtreme Gradient Boosting)와 XAI (eXplainable AI) 기법 중 하나인 SHAP (SHapley Additive exPlanations)을 사용하여 우수한 예측 정확도를 확보하고 각 특성이 예측에 미치는 영향을 확인하고자 했다. 해당 결과를 통해 병원 의료진은 어떤 구급처치와 환자 상태 정보가 중증응급질환자로 예측하는데 중요했는지 확인할 수 있고, 중증응급질환자에게 신속한 의료 서비스를 제공할 수 있게 될 것으로 기대한다.

### I. 서론

119 구급대원의 의사결정은 환자의 예후에 결정적인 역할을 한다. 특히 중증 응급 질환에 해당하는 환자인 것을 신속하고 정확하게 판단하는 것은 환자의 생명과 직결되는 중요한 문제이다[1]. 그러나 현재 중증응급질환 예측은 구급대원의 주관적일 수 있는 판단과 불명확한 기억에 의존해 이루어진다[2]. 이 때문에 생명을 구하는데 적절한 초기 응급 조치가 이루어지지 않을 우려가 있다. 이와 더불어, 병원 의료진에게 환자가 인계될 때도 환자의 상태와 이송 중 발생한 이벤트, 처치 내역 등이 구두로 간략히 전달되는 등 정보공유가 제한된 수준으로 이루어지고 있다. 본 연구에서는 이러한 일련의 문제를 해결하는 기술을 개발하여 병원전 이송단계에서 환자의 상태, 특히 중증응급질환자 여부를 객관적으로 예측하고 해당 정보를 병원에 전달되 신속한 의료 처치가 이루어지기를 기대했다.

과거 연구[2]에서는 머신러닝 모델이 응급상황에서 구급대원을 위한 의사결정지원도구로서 중요한 역할을 할 수 있다는 것을 보여줬다. 해당 연구는 구급신고 전화에서 환자의 음성으로부터 키워드를 분석하여 심장마비를 감지하는 머신러닝 모델이 연구되었다. 모델은 구급대원보다 10 초 더 빠르게 심정지를 인식하였다. [3]의 연구에서는 머신러닝과 딥러닝 모델을 활용하여 응급 심각도를 예측했다. 그 결과 XGBoost 와 Deep Neural Networks 모델이 가장 높은 수준의 예측 정확도를 보였다. 따라서, 본 연구에서도 앙상블 기법 중 하나인 XGBoost 모델을 활용하여 중증응급질환을 예측하는 모델을 개발하고자 한다. 더불어 구급대원, 의료진들에게 모델이 어떤 근거로 예측을 수행하였는지 설명하기 위해 SHAP 을 사용하여 모델의 예측 결과를

해석 가능한 형태로 제시하였다. 이를 통해 의료진이 중증응급질환자에게 신속하게 의료 처리를 제공할 수 있는 기술에 대한 방향성을 제시하고자 하였다.

### II. 본론

본 연구에서는 병원 도착하기 전의 이송 단계에서 수집되는 정보를 사용하여 중증응급질환을 조기 예측하고자 한다. 이를 위해, 구급대원이 작성한 구급활동일지와 심뇌혈관질환자, 심폐정지환자, 중증외상환자의 상세 응급처치 내용이 기록되어 있는 대한 세부상황표 데이터를 활용했다. 해당 데이터에 치명적 부정맥과 심정지를 비롯한 24 종에 대한 중증응급질환 여부를 라벨링 하기 위해 의료 전문가들이 전자의무기록 EMR (Electronic Medical Record) 내 병원에서 실제로 환자에게 진단된 질환명을 맵핑했다.

개별 환자를 식별하는 'KEY'로 병합된 네 가지 데이터에서 구급활동일지에서는 '긴급구조종결부분' 변수에서 정상 값만 추출하였으며, '등록완료여부' 변수에서는 등록 완료된 관측치만을 사용하여 데이터를 정제하였다. 모델 학습 전 범주형 변수는 원-핫 인코딩 (One-Hot Encoding) 방법을 사용해 데이터를 적절히 변환시켰다. 변수의 결측 정보도 중요한 데이터로 활용될 수 있다고 판단하여, float 변수의 결측치들은 -1 로, object 변수의 결측치는 '결측'으로 대체하여 사용했다. 위의 방식으로 간단하게 전처리가 완료된 데이터는 총 703 개의 변수로 구성된 134,172 건의 데이터이고, 변수에는 환자 연령, 환자 성별, 환자 증상 등의 입력변수와 24 가지 중증응급질환 해당 여부(해당 1, 미해당 0)의 출력변수를 포함한다. 우리를 개별

출력변수를 이진 분류하는 24 개의 XGBoost 모델을 학습했다.

이 연구에서는 모델 평가를 위해 정확도(Accuracy)를 평가지표로 사용하였다. 정확도는 직관적으로 분류 모델 예측 성능을 나타내는 평가 지표로 실제값과 예측값이 얼마나 일치하는지 측정한다. 중증응급질환별 차이가 나긴 하지만, 24 개 모델의 평균 정확도는 97%로, 이는 병원진 이송단계에서 확보되는 정보로 상당히 높은 정확도로 중증응급질환이 예측 가능함을 시사한다.

본 연구에서는 SHAP의 Shapley Value를 활용하여 각 변수가 모델의 예측에 어떤 영향을 미치는지 분석하고 모델을 해석하고자 하였다. 아래 그림 1은 치명적 부정맥 질환에 대한 SHAP의 Summary plot을 예시로 보여준다. x축은 Shapley value의 크기를, y축은 변수의 중요도를 의미한다. Shapley Value가 음수일 경우 예측값을 감소(중증응급질환 비해당)시킨다고 해석하는 반면, 양수일 경우 예측값을 증가(중증응급질환 해당)시킨다고 해석할 수 있다. 모델은 환자연령이 높으면서 기존 병력에 심장질환이 있었고 구급활동에서 ECG 측정이 수행되었으면서 의식이 없던 경우 해당 환자를 치명적 부정맥일 확률이 높다고 예측했음을 해석할 수 있었다. 또한 치명적 부정맥의 예측 영향을 주는 주요 변수를 찾아보면 심장질환, 환자발생유형, 환자연령, 맥박수임을 확인할 수 있었다. 이를 통해 이송 단계에서 병원에 중요한 정보만 신속하게 전달할 수 있도록 도와줄 수 있다.

이러한 결과를 토대로 병원 의료진에게 중증응급질환에 대한 예측에서 중요한 역할을 하는 변수들을 제공할 수 있다. 이를 통해 이송 완료된 환자에 대해 신속하고 명확한 인사이트를 얻을 수 있을 것으로 기대한다.

본 논문에서는 이송단계에서 병원이 어떤 구급처치와 환자정보가 중증응급질환자로 예측하는데 중요했는지 확인할 수 있는 중증응급질환에 대한 설명 가능한 예측 모델을 개발하고 성능을 확인하였다. 특히, 치명적 부정맥 예측을 중심으로 SHAP의 Summary plot을 통해 변수와 예측 값의 영향관계를 알아보고 해석함으로써 개발된 모델을 평가하였다. 본 연구의 결과는 병원에 응급환자에 대한 정확하고 신속한 정보를 제공함으로써 의료 서비스의 향상을 기대한다.

본 연구에서 개발된 모델은 97.3%라는 높은 정확도를 보여주지만, 구급환자 중 중증응급질환환자의 비율은 극히 작아 추후에는 이를 반영하는 평가지표로 변경 후 모델의 성능을 더욱 정밀하게 평가해 개선할 예정이다. 중증응급질환자를 놓치는 예측모델은 환자의 생명과 직결된 심각한 문제 야기할 수 있기에 비해당 환자를 중증응급질환자로 오분류하는 비율의 증가를 감수하더라도 모델의 성능 고도화는 필수적이라 판단한다. 또한 Tableau 데이터에 대해서 XGBoost 모델의 높은 성능을 고려하여 해당 모델을 선택하였지만, 추후에는 다양한 모델을 비교 분석하여 모델을 개선하는 것이 필요하다.

### ACKNOWLEDGMENT

이 논문은 한국산업기술평가관리원 소방구급서비스스마트 첨단기술개발 사업의 지원을 받아 수행된 연구임 (과제번호 : RS-2024-00247687).

### 참 고 문 헌

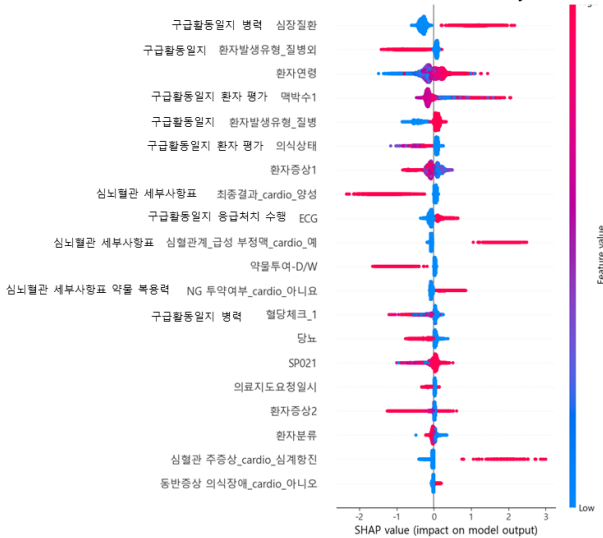
[1] Sasson, Comilla, et al. "Predictors of survival from out-of-hospital cardiac arrest: a systematic review and meta-analysis." *Circulation: Cardiovascular Quality and Outcomes* 3.1 (2010): 63-81.

[2] Tae-Young Moon, Mi-Young Choi "Model for the Structural Relationships of Behavioral Attitude, Subjective Norm and Perceived Behavioral Control to Safety Intention and Safety Behavior in 119 Emergency Medical Technicians" *Journal of the Korea Academia-Industrial cooperation Society* Vol. 19, No. 9 pp. 215-224, 2018.

[3] BLOMBERG, Stig Nikolaj, et al. Machine learning as a supportive tool to recognize cardiac arrest in emergency calls. *Resuscitation*, 2019, 138: 322-329.

[4] SÁNCHEZ-SALMERÓN, Rocío, et al. Machine learning methods applied to triage in emergency services: A systematic review. *International Emergency Nursing*, 2022, 60: 101109.

치명적 부정맥 SHAP Summary Plot



<그림 1> 치명적 부정맥 SHAP Summary Plot 시각화

### III. 결론