

머신러닝을 활용한 응급실 내원 환자 퇴실 조치 결정 조기 예측

황하은¹ · 강현구¹ · 이의선² · 김정윤² · 윤영훈² · 김성범^{1*}

¹고려대학교 산업경영공학과 / ²고려대학교 구로병원 응급의학과

Early Prediction of Patient Disposition for Emergency Department Visits Using Machine Learning

Ha Eun Hwang¹ · Hyungu Kahng¹ · Eu Sun Lee² · Jung Youn Kim² · Young Hoon Yoon² · Seoung Bum Kim¹

¹Department of Industrial and Management Engineering, Korea University

²Department of Emergency Medicine, Korea University Guro Hospital, Seoul, Republic of Korea

Overcrowding within emergency departments (ED) affects patient satisfaction and quality of care. The leading causes of ED overcrowding are systematic delays between procedures and patient disposition after ED treatment. Early prediction of patient disposition can improve patient flow and optimize allocation of hospital resources. While studies for predicting disposition using machine learning methods have been actively conducted abroad, few have been conducted in South Korea in spite of the lagging emergency medical environment. Previous studies are limited to binary predictions; either hospital admission or discharge. In this study, we attempted to predict disposition (discharge, general ward admission, ICU admission) of patients using initial information of ED patients from the Korean national emergency department information system (NEDIS). We used five machine learning methods including logistic regression, decision tree, random forest, CatBoost, and TabNet. The results showed that CatBoost yielded the best performance. This result can aid in decision making by providing standard indicators for hospital admission.

Keywords: Emergency Department, Prediction of ICU Admission, Discharge, Disposition Decision Prediction, Machine Learning

1. 서론

2017년 12월부터 중증 응급환자의 빠른 진료/검사 및 응급실 과밀화 해소를 위하여, 국내 상급종합병원을 대상으로 응급실 24시간 체류 제한 법률이 제정되었다. 더불어 최근 여러 응급 의료센터에서 응급실 과밀화 해소를 위한 다양한 해결방안을 제시하고 있다. 구급차의 분산 정책과 진료 인력의 증대 및 호출 시스템의 개발, 병상 수 증가 등이 그 예가 될 수 있다. 또한, 비 응급환자들을 2차 병원으로 분산하는 한편 환자들이 적절한 병원을 찾을 수 있도록 안내를 강화해야 한다는 지적이 나오고 있다. 그럼에도 불구하고 응급실 과밀화 현상은 지속적으로 발생하고 있다.

응급실 과밀화의 주요 원인은 응급실 내의 절차 지연, 응급실 진료 후 입원, 외래 추적, 전원의 지연 발생으로 볼 수 있다 (Kim *et al.*, 2014). 특히, 입원에 해당하는 환자의 퇴실 조치 지연은 15개 국가의 응급실 과밀화의 일반적인 주요 원인임을 확인할 수 있다(Pines *et al.*, 2011; Pines and Bernstein, 2015). 입원 병실이 부족할 경우 의료진의 결정에 따라 귀가 또는 타원으로 전원시키는 것을 원칙으로 하는데, 전원의 경우 이송 의료기관의 수용 가능 여부의 확인과 적절한 이송 수단의 확보가 선행되어야 한다. 따라서 입원 여부에 대한 조기 예측은 응급실 흐름에 매우 중요한 요소라고 하겠다.

입원이 필요한 환자들에 대한 조기 예측을 통해 입원 기준 확인 및 승인 요청 등의 조기 작업과 병상 준비 등의 병원 자원에

* 연락처: 김성범 교수, 02841 서울특별시 성북구 안암로 145 고려대학교 산업경영공학과, Tel : 02-3290-3397, Fax : 02-929-5888, E-mail : sbkim1@korea.ac.kr

2020년 12월 21일 접수; 2021년 2월 28일 수정본 접수; 2021년 3월 9일 게재 확정.

대한 최적화를 도모할 수 있다(Lee *et al.*, 2020). 특히 응급실 진료 과정에서 조기 예측은 퇴실 절차를 신속하게 개선할 수 있고 의료진의 의사결정에 도움을 줄 수 있는 표준 지표로 활용될 수 있으며, 이를 통해 환자의 안정성이 확보됨과 동시에 응급실 혼잡도 문제도 완화할 수 있다. 더불어 귀가 조치에 대한 조기 예측은 경증 환자에게 2차 병원으로의 효율적인 분산을 가능하게 하여 중증 치료가 주목적인 상급 병원에 경증 환자가 물리는 현상을 완화할 수 있다. 머신러닝을 활용한 조기 예측은 신속성과 환자의 만족을 모두 충족할 수 있으며, 비용적인 측면에서도 효율적인 영향을 가져올 수 있다. 추가적으로 응급실 퇴실 결정은 119와 같은 타 기관의 구급차 수급과도 밀접한 관련성을 가지고 있으므로 활용 관점에서 매우 유의미한 연구라고 하겠다.

머신러닝 모델을 활용한 응급실 퇴실 조치 조기 예측에 대한 다양한 시도들이 있었다. 주로 응급실 진료 관련 정보와 환자의 과거 병력 정보를 머신러닝 기법에 적용하여 진행되었다(Hong *et al.*, 2018). 하지만 대부분의 연구는 응급실 진료 후 퇴실 조치를 단순히 입원, 귀가 2가지로 구분하여 이진 분류를 진행하는 한계점을 가지고 있다(Barak-Corren *et al.*, 2017; Graham *et al.*, 2018; Raita *et al.*, 2019; Sun *et al.*, 2011; Zhang *et al.*, 2017). 이전 연구에서 간과한 중환자의 경우 고위험 환자로 일반 병실에 입원하는 환자에 비해 신속한 입원 절차가 필요한 환자군이다. 아울러 타 국가의 데이터 분석 결과는 응급의료환경이 국가 간 상이하기 때문에 국내 응급의료환경에 타 국가 결과를 직접적으로 적용하기 어렵다는 한계점을 가지고 있다(Shin *et al.*, 2012).

본 연구에서는 환자의 퇴실 결정을 귀가, 일반 병실 입원, 중환자실 입원으로 구분하여 조기 예측의 효율성을 제고하고자 한다. 더불어, 전국 응급의료 기관을 통해 수집된 데이터로 국내 의료 환경에 적합한 연구를 진행하였다. 아울러 응급실 퇴실 조치 결정에 영향을 주는 변수를 추출하여 병원 내 유동성 있는 스케줄링과 환자의 안전을 도모하고자 한다.

2. 실험 방법

2.1 데이터 수집 및 전처리

본 연구는 2016년 1월 1일부터 2018년 12월 31일까지 36개월 동안 전국 응급의료기관을 방문한 환자를 대상으로 수행되었다. 국가응급진료정보망(National Emergency Department Information System, NEDIS)에 입력된 권역 응급의료센터와 지역 응급의료센터의 응급환자 진료 관련 정보를 수집하여 분석하였다.

데이터셋을 구축하는 과정에서 연령이 15세 이하인 경우에는 환자의 초기 평가 정보에 해당하는 분당 맥박수, 호흡수 등의 활력 징후가 성인에 비해 변동성이 크고 정상 수치 기준과 다르기 때문에 제외하였다. 또한, 환자의 주요 호소증상과

KTAS(Korean Triage and Acuity Scale) 등급 분류가 결측 값인 경우, 내원 시 수축기 혈압, 이완기 혈압, 분당 맥박수, 호흡수, 체온, 산소 포화도가 0 이하의 값을 가지는 경우는 제외하였다. 또한 퇴실 조치 결정이 전원과 사망인 경우에도 제외하였다. 전처리 과정을 거쳐 구축한 데이터셋은 총 5,818,005개로 구성되었다.

본 연구는 응급실 진료 후 퇴실 조치에 대한 조기 예측을 목표로 하고 있으므로 환자가 응급실에 도착한 시간 기준으로 약 10분 이내로 수집 가능한 변수만 사용하고 진료 과정 및 이후에 기입된 변수는 제외하였다. 요일 및 시간 변수에 따른 입원 여부의 차이가 있는지 알아보기 위해 응급실 내원 년도, 월, 요일, 시간대 변수를 응급실 내원 시간을 통해 새롭게 생성하였다. 또한 증상 발현 및 중증도에 따른 영향을 보기 위해 증상 발현 시간과 응급실 내원 시간 간의 차이, 응급실 내원 시간과 최초 중증도 분류 시간의 차이 변수를 추가적으로 생성하였다. 변수 사용의 타당성을 확보하기 위해 모든 범주형 설명변수에 대하여 반응변수와 카이 제곱 독립성 검정을 실시하였고, p-value 값이 각각 0.01 이하로 모든 범주형 설명변수와 반응변수는 통계적으로 독립이 아님을 확인하였다. 최종적으로 연속형 변수 9개와 범주형 변수 19개로 구성된 설명변수를 분석에 활용하였다. 반응변수인 퇴실 조치는 귀가, 일반 병실 입원, 중환자실 입원 세 가지의 범주로 구성하였다.

<Table 1>는 모든 범주형 설명변수에 따른 귀가, 일반 병실 입원, 중환자실 입원 관측치 개수(%)를 나타낸 표이다. 응급실 내원 월, 요일과 시간대에 따른 비율은 크게 차이가 없음을 알 수 있었다. 여성의 경우 남성보다 귀가의 비율이 높지만, 유의미한 차이를 보이지는 않았다. 보험유형 중 의료급여 1종 환자의 경우 일반 병실 입원 비율이 35.5%로 상대적으로 높은 비율을 보이고 있었다. 최초KTAS 중증도 결과에 따라서는 1등급(소생) 환자의 경우는 과반수 이상이 중환자실에 입원하였고, 2등급(긴급) 환자의 경우는 각각 35%, 30%로 일반 병실 입원과 중환자실 입원에서 비슷한 비율을 보이고 있었다. 이를 통해 최초KTAS 중증도 분류에 따라 퇴실 결정 비율이 상이함을 알 수 있었다. 응급실 내원 시 이동수단의 경우 중환자실에 입원한 환자 중 119 구급차를 이용한 환자 수가 176,876명으로 가장 많았지만, 119구급차를 이용한 환자의 과반수 이상은 최종 귀가 조치로 결정되었음을 알 수 있었다. 내원 경로가 외부에서 전원인 경우 전원을 보낸 의료기관이 상급기관일수록 입원의 비율이 높음을 알 수 있었다.

2.2 분석 방법

전처리가 완료된 데이터를 학습:평가=8:2의 비율로 분할하였다. 5-겹 교차검증(5-fold cross validation)을 수행하여 실험 시 학습 데이터와 검증 데이터를 분할하였다. 이때 머신러닝 알고리즘 마다 가장 우수한 분류 성능을 보이는 매개변수를 선정하여 최종 분류 모델을 구축하였다.

Table 1. Characteristics of Variables

변수	상위 빈도 범주	입원		
		일반 병실 (n = 1,582,727)	중환자실 (n = 430,554)	귀가 (n = 3,804,724)
응급의료 기관 구분	권역	636,376 (28.82)	193,365 (8.76)	1,378,704 (62.43)
	지역	946,351 (26.22)	237,189 (6.57)	2,426,020 (67.21)
응급실 내원 년도	2016	473,687 (27.68)	132,032 (7.72)	1,105,438 (64.6)
	2017	530,068 (27.47)	145,029 (7.51)	1,254,835 (65.02)
	2018	578,972 (26.6)	153,493 (7.05)	1,444,451 (66.35)
응급실 내원 월	12월	137,366 (25.98)	38,152 (7.21)	353,292 (66.81)
	8월	137,803 (26.35)	36,638 (7.01)	348,437 (66.64)
	9월	131,814 (25.39)	36,000 (6.93)	351,317 (67.67)
	10월	136,108 (26.7)	37,718 (7.4)	335,884 (65.9)
	7월	136,117 (26.82)	36,196 (7.13)	335,179 (66.05)
응급실 내원 요일	월요일	266,737 (30.63)	67,546 (7.76)	536,531 (61.61)
	화요일	231,749 (29.55)	62,473 (7.96)	490,159 (62.49)
	수요일	224,901 (29.43)	61,611 (8.06)	477,600 (62.5)
	목요일	225,722 (29.58)	62,348 (8.17)	475,026 (62.25)
	금요일	226,425 (28.97)	64,136 (8.21)	491,036 (62.82)
	토요일	207,256 (23.89)	58,567 (6.75)	601,607 (69.36)
	일요일	199,937 (20.27)	53,873 (5.46)	732,765 (74.27)
응급실 내원 시간대	11:00 ~12:00	123,491 (36.02)	28,616 (8.35)	190,688 (55.63)
	10:00~11:00	117,042 (34.4)	27,227 (8.0)	195,951 (57.6)
	20:00~21:00	65,776 (21.02)	19,703 (6.3)	227,497 (72.69)
	19:00~20:00	73,871 (23.91)	21,633 (7.0)	213,473 (69.09)
	15:00~16:00	105,912 (34.54)	25,873 (8.44)	174,841 (57.02)
성별	남성	850,918 (29.04)	258,883 (8.84)	1,820,035 (62.12)
	여성	731,809 (25.34)	171,671 (5.94)	1,984,689 (68.72)
보험유형	건강보험	1,371,923 (27.03)	354,474 (6.98)	3,348,797 (65.98)
	의료급여 1종	127,074 (35.5)	43,669 (12.2)	187,182 (52.3)
	자동차보험	47,497 (23.19)	15,397 (7.52)	141,933 (69.29)
	일반	13,594 (15.83)	9,236 (10.76)	63,020 (73.41)
	의료급여 2종	11,023 (21.97)	3,960 (7.89)	35,200 (70.14)
증상 발현 월	12월	137,867 (25.86)	38,025 (7.13)	357,306 (67.01)
	9월	132,529 (25.29)	36,042 (6.88)	355,428 (67.83)
	8월	137,283 (26.26)	36,663 (7.01)	348,805 (66.72)
	7월	136,282 (26.88)	36,238 (7.15)	334,403 (65.97)
	10월	136,110 (27.02)	37,813 (7.51)	329,840 (65.48)
증상 발현 요일	월요일	506,314 (29.52)	239,099 (7.96)	64,479 (62.52)
	화요일	484,242 (28.96)	222,262 (7.96)	61,061 (63.09)
	수요일	487,256 (28.75)	221,234 (7.92)	60,925 (63.33)
	목요일	500,105 (28.31)	221,632 (7.82)	61,205 (63.88)
	금요일	546,150 (27.77)	234,432 (7.52)	63,497 (64.7)
	토요일	663,787 (23.92)	227,907 (6.42)	61,203 (69.66)
	일요일	616,870 (24.25)	216,161 (6.53)	58,184 (69.22)
증상 발현 시간대	12:00~13:00	121,666 (33.37)	28,400 (7.79)	214,524 (58.84)
	10:00~11:00	121,509 (33.4)	28,878 (7.94)	213,425 (58.66)
	9:00~10:00	105,996 (31.4)	27,485 (8.14)	204,115 (60.46)
	8:00~9:00	101,198 (30.27)	27,413 (8.2)	205,756 (61.54)
	20:00~21:00	71,276 (22.05)	20,198 (6.25)	231,766 (71.7)
질병 여부	질병	1,372,255 (29.69)	360,421 (7.8)	2,889,316 (62.51)
	질병 외	209,450 (17.6)	69,783 (5.86)	910,676 (76.53)
	미상	1,022 (16.74)	350 (5.73)	4,732 (77.52)
의도성 여부	미상	1,374,605 (29.66)	362,996 (7.83)	2,897,692 (62.51)
	비의도적 사고	193,884 (17.9)	51,891 (4.79)	837,638 (77.31)
	자해/자살	9,066 (26.38)	13,330 (38.78)	11,974 (34.84)
	폭력/타살	4,096 (8.7)	1,012 (2.15)	41,988 (89.15)
	기타	1,076 (6.03)	1,325 (7.43)	15,432 (86.54)
손상기전	미상	1,374,688 (29.64)	362,444 (7.82)	2,900,068 (62.54)
	미끄러짐	59,399 (26.06)	9,922 (4.35)	158,595 (69.58)
	부딪힘	19,020 (9.88)	3,932 (2.04)	169,488 (88.07)
	기타-성폭력, 감전 등	13,350 (7.25)	2,001 (1.09)	168,662 (91.66)
	베임, 찔림	10,542 (7.07)	1,824 (1.22)	136,693 (91.7)

Table 1. Characteristics of Variables(Continue)

변수	상위 빈도 범주	입원		
		일반 병실 (n = 1,582,727)	중환자실 (n = 430,554)	귀가 (n = 3,804,724)
응급실 내원 경로	직접 내원	1,006,439 (21.27)	250,287 (5.29)	3,474,277 (73.44)
	외부에서 내원	475,108 (52.1)	169,421 (18.58)	267,450 (29.33)
	외래에서 내원	100,530 (58.09)	10,623 (6.14)	61,900 (35.77)
	기타	538 (34.69)	160 (10.32)	853 (55.0)
	미상	112 (26.73)	63 (15.04)	244 (58.23)
응급실 내원 시 이동수단	기타 자동차	917,665 (23.24)	112,193 (2.84)	2,918,016 (73.91)
	119 구급차	395,308 (29.74)	176,876 (13.31)	756,959 (56.95)
	기타 구급차	186,880 (54.1)	98,378 (28.48)	60,170 (17.42)
	의료기관 구급차	59,667 (50.35)	38,050 (32.11)	20,783 (17.54)
	도보	16,027 (28.79)	1,427 (2.56)	38,212 (68.65)
응급증상 해당 여부	응급	1,511,301 (31.3)	424,056 (8.78)	2,892,431 (59.91)
	비응급	71,426 (7.21)	6,498 (0.66)	912,293 (29.13)
환자 내원 시 반응	명료	1,488,359 (26.94)	303,351 (5.49)	3,732,362 (67.57)
	기면	60,825 (37.4)	54,978 (33.81)	46,823 (28.79)
	혼미 또는 반혼수	31,189 (27.33)	59,082 (51.77)	23,847 (20.9)
	혼수	2,342 (13.65)	13,143 (76.6)	1,672 (9.75)
최초 KTAS 중증도 결과	1등급(소생)	18,753 (29.36)	38,447 (60.19)	6,679 (10.46)
	2등급(긴급)	211,505 (35.26)	182,027 (30.34)	206,377 (34.4)
	3등급(응급)	919,304 (36.78)	180,626 (7.23)	1,399,366 (55.99)
	4등급(준응급)	384,662 (17.94)	26,260 (1.22)	1,732,799 (80.83)
	5등급(비응급)	48,366 (9.47)	3,162 (0.62)	459,034 (89.91)
	기타	137 (21.47)	32 (5.02)	469 (73.51)
전원 보낸 의료기관	모름	1,108,960 (22.59)	261,464 (5.33)	3,538,530 (72.08)
	병원	167,463 (53.33)	62,476 (19.9)	84,085 (26.78)
	종합병원	172,656 (54.85)	79,599 (25.29)	62,542 (19.87)
	의원	94,959 (44.43)	12,686 (5.94)	106,080 (49.63)
	상급종합병원	31,156 (61.86)	12,118 (24.06)	7,092 (14.08)

본 연구에서는 의료 분야에서는 많이 사용하고 있는 다항 로지스틱 회귀 모델, 의사결정 나무 모델과 분류 성능이 우수하다고 알려진 앙상블 계열의 랜덤 포레스트 모델, Catboost 모델을 사용하였다. 더불어 범주형 변수에 대한 원 핫 인코딩 과정이 필요 없고 관측치 별 중요 변수 해석이 가능한 딥러닝 계열의 TabNet 알고리즘을 실험에 사용하였다.

다항 로지스틱 회귀는 설명변수와 여러 범주가 있는 명목형 반응변수 간의 선형 관계를 추정하기 위한 단순 로지스틱 회귀의 확장된 모델이다(Vittinghoff *et al.*, 2011). 의사결정 나무는 규칙을 기반으로 데이터를 분석하여 패턴을 예측 가능한 규칙들의 조합으로 나타내는 알고리즘이다(Breiman *et al.*, 1984). 데이터를 부분 집합으로 분할하고 비용 함수가 최소로 얻어지는 최상의 분할을 찾는다. 본 연구에서 classification and regression tree (CART) 의사결정 알고리즘을 사용했으며, 각 분할에 속해 있는 값들의 불순도가 최소가 되는 방향으로 학습한다. 일반 의사결정 나무 모델의 경우 설명변수의 수가 많게 되면 의사결정 나무의 가치가 많아지게 되어 학습 데이터만 잘 맞추게 되는 과적합이 일어난다. 이러한 과적합을 막기 위해 랜덤 포레스트 모델은 하나의 의사결정나무 생성시 설명변수 일부만을 임의로 샘플링하여 사용하며 복원 추출을 통해 다수의 의사결정 나무를 생성한다(Breiman, 2001). 앞의 과정을 통해 만들어진 여러 개의 의사

결정나무들이 예측값 중 가장 많이 예측한 값이 최종 예측값으로 선정된다. CatBoost는 순서형 TBS(target-based statistics)와 순서형 부스팅(ordered boosting)을 통해 범주형 변수를 효과적으로 처리하는 모델이다(Prokhorenkova *et al.*, 2018). 순서형 TBS는 해당 카테고리의 레벨마다 반응 변수의 평균을 통해 범주형 변수를 변환하는 TBS기법에 이전까지의 데이터를 통해 산출하는 방식을 적용하여 데이터 누수를 막는 기법이다. 순서형 부스팅은 각 단계마다 새로운 데이터셋을 독립적으로 샘플링하고 학습 중인 현재 모델에 적용하여 잔차를 구함으로 기존 그라디언트 부스팅 계열 모델의 한계점을 극복한 기법이다. TabNet은 정형 데이터에 적합한 해석 가능한 딥러닝 모델로 성능과 해석 가능성을 향상시키기 위해 의사결정 단계에서 변수 선택이 가능하도록 어텐션 구조(attention mechanism)를 활용하였다(Ai and Tomas, 2019). TabNet은 feature transformer와 attentive transformer 단계를 추가하여 기존 뉴럴 네트워크와 차별점을 가진다. 우선, feature transformer는 전처리 과정 없이 입력 값을 받아 변수를 처리하고 의사결정 단계에서 변수가 공유될 수 있도록 유지하고 다음 의사결정 단계에 영향을 줄 수 있도록 정보를 유지하는 역할을 한다. Attentive transformer는 이전 의사결정 단계에서 변수가 선택되었는지 반영하여 현재 의사결정 단계에서 마스킹(masking)할 변수를 결정하는 변수 선택의 역할을 수행한다.

3. 결 과

3.1 실험 평가 지표

본 연구에서 퇴실 조치 분류 모델의 성능을 평가하는 척도로 One vs. Rest AUROC score, 정확도(accuracy), 정밀도(precision), 재현율(recall), F1 점수(F1 score)을 사용하였다. <Table 1>을 통해 확인한 범주 간 불균형 문제를 모델 성능 평가시 완화하기 위해 모든 평가 척도는 각 범주의 관측치 수에 상관없이 평균을 취하는 방식으로 계산하였다. 각 평가 척도는 식 (1)~식 (5)을 통해 산출하였다. 수식에서 y, \hat{y} 는 각각 실제 값과 예측값을 의미한다. 또한, C는 반응변수의 모든 범주를 의미하고 n은 총 관측치 수를 의미한다.

$$AUC_{ovr,macro} = \frac{1}{C} \sum_{i=1}^C AUC(i, rest_i) \tag{1}$$

$$Accuracy(y, \hat{y}) = \frac{1}{n} \sum_{j=1}^n I(\hat{y}_j = y_j) \tag{2}$$

where, $I(\hat{y}_i = y_i) = \begin{cases} 1 & \text{if } \hat{y}_i = y_i \\ 0 & \text{if } \hat{y}_i \neq y_i \end{cases}$

$$Precision_{macro} = \frac{1}{C} \sum_{i=1}^C \frac{TP_i}{TP_i + FP_i} \tag{3}$$

$$Recall_{macro} = \frac{1}{C} \sum_{i=1}^C \frac{TP_i}{TP_i + FN_i} \tag{4}$$

$$F1\ Score_{macro} = 2 \times \frac{Precision_{macro} \times Recall_{macro}}{Precision_{macro} + Recall_{macro}} \tag{5}$$

3.2 실험 결과

<Table 2>는 훈련 및 검증 데이터를 통해 선정된 모델의 평가 데이터에 대한 성능 결과를 보여주고 있다. AUROC Score 기준으로 모든 모델이 0.77 이상의 성능을 보이고 있으며 CatBoost 모델이 가장 높은 성능을 보이고 있음을 알 수 있다. CatBoost와 TabNet이 모든 평가 지표에서 거의 유사하게 높은 성능을 도출했음을 확인할 수 있다.

<Figure 1>은 데이터의 불균형으로 인한 편향성이 발생했는지 확인할 수 있는 혼동행렬을 보여주고 있다. 혼동행렬의 대각선 부분은 실제값과 예측값이 일치하는 비율을 의미하며 해당 비율이 높은 분류 결과가 이상적이라고 해석할 수 있다. 일반병실 입원의 경우 귀가와 중환자실 입원으로 다소 혼동하는 양상을 띠지만 불균형의 정도가 심한 데이터 특징을 고려하였을 때 어느정도 올바른 분류 결과를 보임을 확인할 수 있다.

Table 2. Model Performance for Classification

	One vs. Rest AUROC	Accuracy	Precision	Recall	F1 score
Catboost	0.847	0.691	0.58	0.671	0.602
TabNet	0.84	0.695	0.578	0.664	0.599
Logistic regression	0.804	0.639	0.531	0.621	0.547
Random forest	0.790	0.626	0.522	0.612	0.534
Decision tree	0.773	0.645	0.516	0.594	0.529

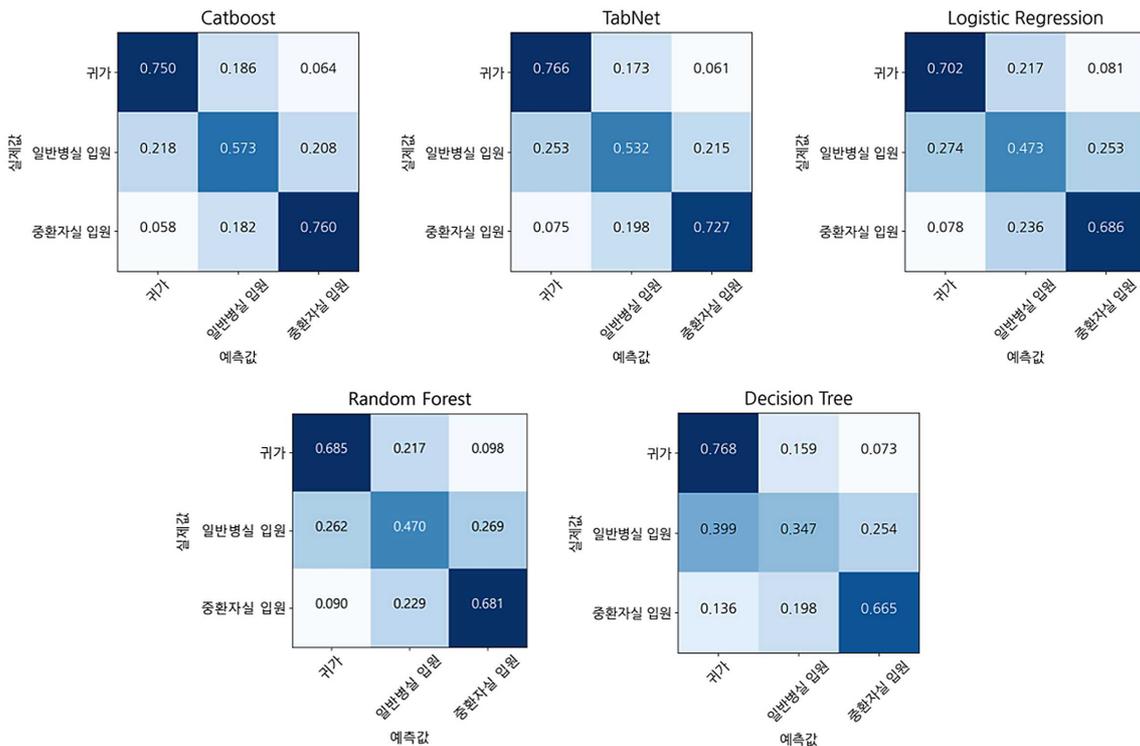


Figure 1. Confusion Matrix for Classification

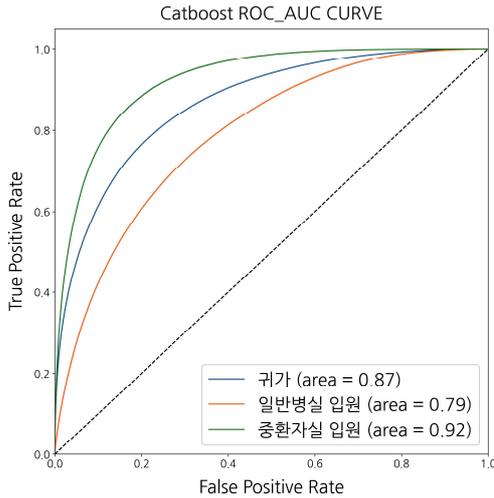


Figure 2. ROC curve of Catboost Model for Each Class

<Figure 2>를 통해 CatBoost 모델의 각 범주별 receiver operating characteristics(ROC) 그래프를 확인할 수 있다. ROC 그래프는 모든 임계 값에서 분류 모델의 성능을 보여주는 그래프이다. ROC 그래프 밑에 해당하는 면적 값을 나타내는 area under the receiver operating characteristics(AUROC)값은 실제 참일 경우에 참으로 예측한 비율을 나타내는 지표로 의학 분야에서 대표적으로 사용되는 성능 평가 지표이다(Provost and Domingos, 2000). 우수한 분류 모델일수록 AUROC값이 1에 가깝고, 범주에 대한 분류 성능이 뛰어난 것을 의미한다. 중환자실 입원의 경우, AUROC 값이 0.92로 가장 우수한 분류 정확도를 보였으며, 왼쪽 상단으로 굽어지는 형태를 확인할 수 있다. 다음으로 귀가 범주의 0.87로 높은 성능을 보였고, 일반 병실 입원은 0.79의 AUROC값을 보여줌으로 신뢰할 만한 성능을 보여주고 있음을 확인하였다.

3.3 변수 중요도

부스팅 계열인 CatBoost로부터 도출된 변수 중요도를 <Figure 3>을 통해 확인할 수 있다. 각각 변수 중요도 기준으로 내림차순

정렬하여 상위 15개 변수를 그래프로 시각화 하였다.

CatBoost는 관측치가 변경될 경우에 예측이 평균적으로 얼마나 변하는지에 대한 척도인 prediction value change를 기준으로 변수 중요도를 산출한다. 최초 KTAS 중증도 결과, 연령과 응급실 내원 시 이동수단 등이 높은 변수 중요도를 보이며 퇴실 조치 결정에 영향을 많이 주는 변수로 선정되었다. 새벽에 내원하는 환자들 중에는 응급 환자가 많기 때문에 응급실 내원 시간대가 퇴실 조치 결정에 유의미한 변수로 선정되었다고 해석할 수 있다.

아울러 주 증상 발현 시간과 응급실 내원 시간 간의 차이의 경우 유의미한 변수로 선정되어 적절한 응급실 방문 시점에 대한 흥미로운 결과가 도출되었다. 또한, 환자의 건강 상태를 나타내는 변수들에 비해 응급실 내원 년도, 월, 요일 등 외부 변수는 크게 중요하지 않음을 확인할 수 있었다. 내원 년도에 따른 퇴실 조치 결정의 차이는 보험체계의 변경 또는 감염병 여부에 의해서만 변동하는 경향을 가진다. 또한, 중환자의 경우에는 요일과 같은 외부 변수에 영향을 받지 않고 입원하는 경향을 가지고 있기 때문에 의학적으로 합리적인 결과라고 하겠다.

해석에 어려움이 있는 기존 덤러닝 모델의 한계점을 보완한 TabNet을 통해 각 관측치마다 지역 변수 중요도와 글로벌 변수 중요도를 산출할 수 있다. 현 단계의 의사결정에 기여한 정도로 지역 변수 중요도가 계산되고 각 단계마다 가중치를 두어 지역 변수 중요도를 선형 결합하여 글로벌 변수 중요도를 계산할 수 있다. 글로벌 변수 중요도가 높다는 것은 모델 학습 과정 전반에서 중요한 변수임을 의미한다. <Figure 4>는 샘플링을 통해 선별한 임의의 30개의 관측치 마다 산출된 글로벌 변수 중요도의 정도를 표현한 그래프이다. 진한 파란색을 뿔수록 높은 변수 중요도를 의미하며 연한 파란색을 뿔수록 낮은 변수 중요도를 의미한다. 모든 관측치에 대하여 변수마다 중요도를 확인할 수 있기 때문에 개별 환자에 대한 해석이 가능하다는 장점을 가지고 있다.

<Figure 5>는 반응변수의 범주에 대해 각각 글로벌 변수 중요도의 평균을 시각화한 그래프이다.

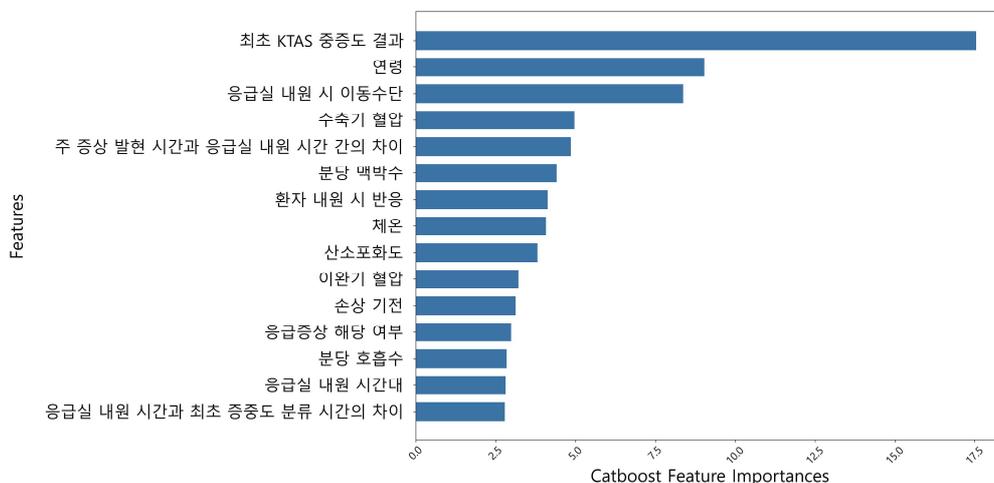


Figure 3. Feature Importance from CatBoost

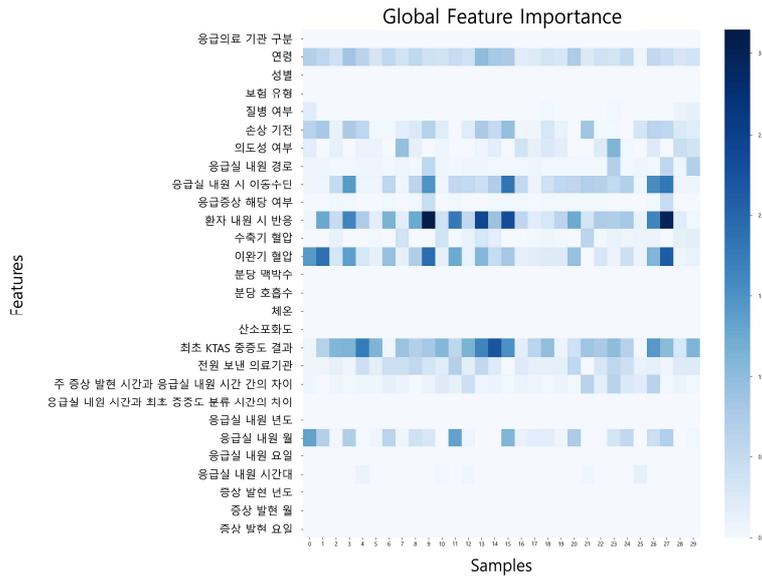


Figure 4. Global Feature Importance from TabNet for 30 Samples

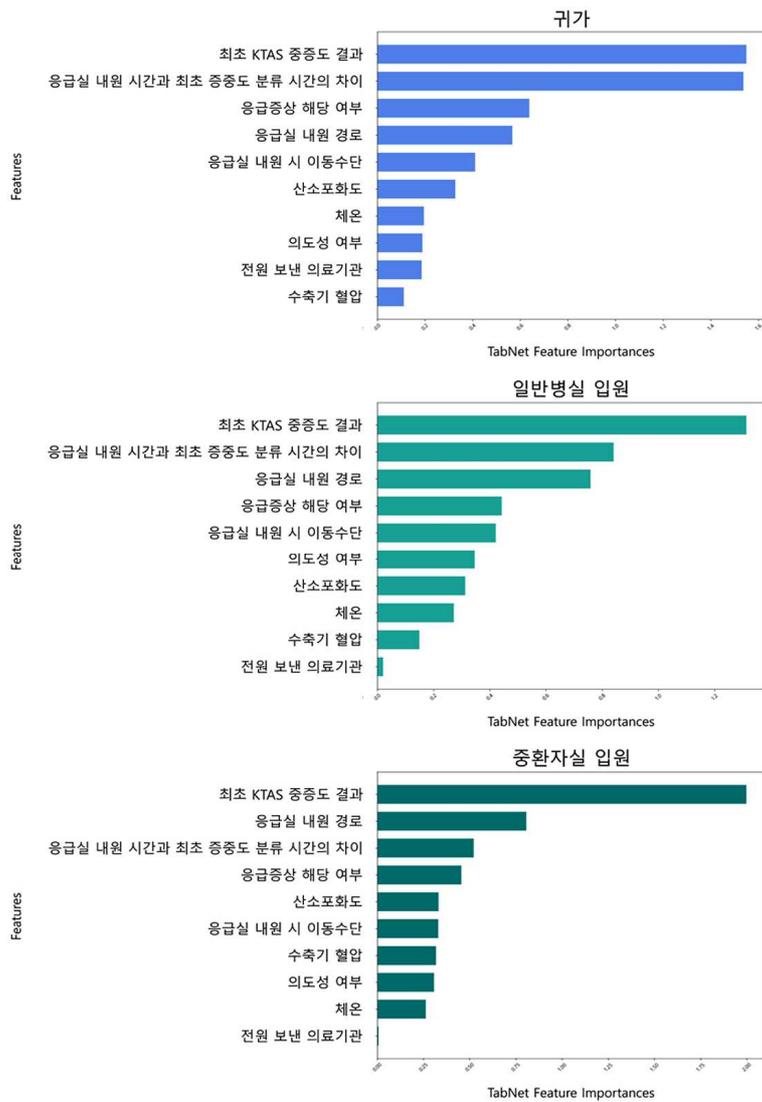


Figure 5. Global Feature Importance from TabNet for Each Class

개별 범주에 해당되는 환자들의 일반적인 특징을 파악하기 위해 평균 변수 중요도를 표현하였다. 범주 구분 없이 최초 KTAS 중증도 결과가 가장 높은 변수 중요도를 보임을 알 수 있었다. 아울러, 모든 범주에서 산소 포화도가 높은 변수 중요도를 보이고 있었다. 의학적인 관점에서 정상 산소 포화도의 경우 95이상의 값을 가지며 90 미만은 매우 낮은 수치로 해석된다. 이를 통해 산소 포화도가 퇴실 결정에 영향을 주는 유의한 변수임을 알 수 있었다. 앞선 변수 중요도에 따른 상위 변수 중 <Figure 3>의 CatBoost 모델과 공통 변수들이 다수 존재함을 알 수 있었다. <Table 1>에서 범주별 차이가 두드러진 변수들이 상위 10개 변수에 많이 해당된다는 점도 확인할 수 있다. 종합적으로 환자의 중증도 및 응급실 내원 시간과 최초 중증도 분류 시간의 차이, 응급증상 해당 여부에 따라 퇴실 조치 결정이 영향을 받는다고 해석할 수 있다.

4. 결론

본 연구의 목적은 응급실 퇴실 조치 결정을 조기 예측하고 이에 영향을 끼치는 요인들을 파악하여 해석 가능한 예측 모형을 개발하는 것이다. 이러한 목적을 달성하기 위해 국가응급진료정보망에 입력된 전국 단위의 응급환자 진료 관련 정보를 수집하여 분석하였다. 응급실 퇴실 조치 결정 분류 모형에는 총 5가지의 머신러닝 모델(다항 로지스틱 회귀, 의사결정 나무, 랜덤 포레스트, CatBoost, TabNet)을 적용하여 실험을 진행하였다. 머신러닝 모델 간 큰 성능 차이는 보이지 않았지만 CatBoost 모델이 AUROC 기준으로 0.847의 가장 높은 분류 성능을 산출하였다. 아울러 각 범주별 성능을 확인하였을 시, 중환자실 입원에 대한 CatBoost 분류 성능이 0.92로 가장 높은 성능을 보임을 알 수 있었다.

퇴실 조치 분류에 영향을 끼치는 중요 변수로는 최초 KTAS 중증도 결과, 응급실 내원 시 이동수단 등이 도출되었다. 본 연구를 통해 환자마다 응급실 퇴실 조치를 예측하여 실제 응급의료 환경에서 활용할 경우 응급실 과밀화 문제를 완화할 수 있을 것으로 기대된다. 또한 병원 내 자원의 최적화와 병원 간 이송에 필수적인 구급차 스케줄링에도 활용될 수 있을 것으로 본다. 본 연구에서는 환자의 응급실 방문 시 수집된 정보만을 고려한 모형을 개발하였지만 환자의 이전 병력 정보들을 추가적으로 활용한다면 보다 더 정교한 분류 모델을 개발할 수 있을 것으로 판단된다.

참고문헌

Ai, G. C. and Tomas, P. (2019), TabNet : Attentive Interpretable Tabular Learning, *arXiv preprint arXiv:1908.07442*.

- Barak-Corren, Y., Fine, A. M., and Reis, B. Y. (2017), Early Prediction Model of Patient Hospitalization from the Pediatric Emergency department, *Pediatrics*, **139**(5), <https://doi.org/10.1542/peds.2016-2785>.
- Breiman, L. (2001), Random Forests, *Machine Learning*, **45**(1), 5-32.
- Breiman, L., Friedman, J., Stone, C. J., and Olshen, R. A. (1984), *Classification and Regression Trees*, CRC press.
- Graham, B., Bond, R., Quinn, M., and Mulvenna, M. (2018), Using Data Mining to Predict Hospital Admissions from the Emergency Department, *IEEE Access*, **6**, 10458-10469.
- Hong, W. S., Haimovich, A. D., and Taylor, R. A. (2018), Predicting Hospital Admission at Emergency Department Triage Using Machine Learning, *PLoS One*, **13**, 1-13.
- Kim et al. (2014), Mid-Term Effects of Tertiary Hospital Department Overcrowding, *Journal of The Korean Society of Emergency Medicine*, **25**(6), 722-729.
- Lee, S.-Y., Chinnam, R. B., Dalkiran, E., Krupp, S., and Nauss, M. (2020), Prediction of Emergency Department Patient Disposition Decision for Proactive Resource Allocation for Admission, *Health Care Management Science*, **23**(3), 339-359.
- Pines, J. M. and Bernstein, S. L. (2015), Solving the Worldwide Emergency Department Crowding Problem-what can we Learn from an Israeli ED?, *Israel Journal of Health Policy Research*, **4**(1), 1-4.
- Pines et al. (2011), International Perspectives on Emergency Department Crowding, *Academic Emergency Medicine*, **18**(12), 1358-1370.
- Prokhorenkova, L., Gusev, G., Vorobev, A., Dorogush, A. V., and Gulin, A. (2018), Catboost : Unbiased Boosting with Categorical Features, *Advances in Neural Information Processing Systems*, 6638-6648.
- Provost, F. and Domingos, P. (2000), Well-trained PETs : Improving Probability Estimation Trees, Raport Instytutowy IS-00-04, Stern School of Business, New York University.
- Raita, Y., Goto, T., Faridi, M. K., Brown, D. F. M., and Jr, C. A. C. (2019), Emergency Department Triage Prediction of Clinical Outcomes Using Machine Learning Models, *Critical Care*, 1-13.
- Shin et al. (2012), Comparison of Emergency Medical Services Systems Across Pan-Asian Countries : A Web-Based Survey, *Prehospital Emergency Care*, **16**(4), 477-496.
- Sun, Y., Heng, B. H., Tay, S. Y., and Seow, E. (2011), Predicting Hospital Admissions at Emergency Department Triage Using Routine Administrative Data, *Academic Emergency Medicine*, **18**(8), 844-850.
- Vittinghoff, E., Glidden, D. V., Shiboski, S. C., and McCulloch, C. E. (2011), *Regression Methods in Biostatistics : Linear, Logistic, Survival, and Repeated Measures Models*, Springer Science & Business Media.
- Zhang, X., Kim, J., Patzer, R. E., Pitts, S. R., Patzer, A., and Schrage, J. D. (2017), Prediction of Emergency Department Hospital Admission based on Natural Language Processing and Neural Networks, *Methods of Information in Medicine*, **56**(5), 377-389.

저자소개

황하은 : 건국대학교 응용통계학과에서 2019년 학사 학위를 취득하고 고려대학교 산업경영공학과에서 석사과정에 재학 중이다. 연구분야는 해석 가능한 인공지능, 머신러닝이다.

강현구 : 고려대학교 산업경영공학과에서 2015년 학사 학위를 취득하고 고려대학교 산업경영공학과에서 석박사통합과정에 재학 중이다. 연구 분야는 Unsupervised Representation Learning, Neural Networks for Electronic Health Records, Multimodal Representation Learning이다.

이의선 : 고대구로병원 응급의학과 임상조교수로 재직 중이다. 전문진료분야는 응급의학과이다.

김정윤 : 고대구로병원 응급의학과 교수로 재직 중이다. 전문진

료분야는 응급의학과 소생학이다.

윤영훈 : 고대구로병원 응급의학과 교수로 재직 중이다. 전문진료분야는 응급의학과 외상학이다.

김성범 : 한양대학교 산업공학과에서 1999년 학사를 취득하고 2001년과 2005년 미국 Georgia Institute of Technology에서 산업공학 석사학위, 박사학위를 취득하였다. 미국 텍사스주립대학교 교수를 역임하고 2009년부터 고려대학교 산업경영공학부 교수로 재직하고 있다. 연구 분야는 인공지능, 머신러닝, 최적화이다.