



저작자표시-비영리-변경금지 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.

다음과 같은 조건을 따라야 합니다:



저작자표시. 귀하는 원저작자를 표시하여야 합니다.



비영리. 귀하는 이 저작물을 영리 목적으로 이용할 수 없습니다.



변경금지. 귀하는 이 저작물을 개작, 변형 또는 가공할 수 없습니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.

이것은 [이용허락규약\(Legal Code\)](#)을 이해하기 쉽게 요약한 것입니다.

[Disclaimer](#)

전자간호기록을 이용한 한국형
응급환자 분류도구(KTAS) 3, 4 단계
환자의 임상결과 예측 모델 개발

연세대학교 대학원

간 호 학 과

신 현 아

전자간호기록을 이용한 한국형
응급환자 분류도구(KTAS) 3, 4 단계
환자의 임상결과 예측 모델 개발

지도교수 최 모 나

이 논문을 박사 학위논문으로 제출함

2024 년 12 월

연세대학교 대학원

간 호 학 과

신 현 아

신현아의 박사 학위논문을 인준함

심사위원 최 모 나 (인)

심사위원 장 연 수 (인)

심사위원 이 승 은 (인)

심사위원 박 창 기 (인)

심사위원 차 원 철 (인)

연세대학교 대학원

2024 년 12 월

감사의 글

2014 년, 학문에 대한 열정과 즐거움을 가지고 호기롭게 시작했던 박사 과정이 우여곡절 끝에 10 여년이 지난 이 순간, 학위 논문을 마무리하며 끝을 맺었습니다. 드디어 고대하던 이 시간이 다가오니 지난 시간들이 주마등처럼 스치며, 박사학위 논문을 완성하기까지 도움과 격려, 응원을 해 주신 많은 분들께 감사의 마음을 전합니다.

우선, 지도 교수님이신 최모나 교수님께 진심으로 깊이 감사의 말씀을 드립니다. 연구의 방향 설정부터 최종 논문 작성에 이르기까지 아낌없는 지도와 조언을 해주셨기에 이 논문이 완성될 수 있었습니다. 특히 제가 연구 과정에서 여러 차례 어려움을 겪고 방향할 때마다 할 수 있다는 믿음을 주시고 따뜻한 마음으로 격려해 주신 덕분에 포기하지 않고 오늘에 이를 수 있었습니다. 연구 진행이 원활하지 않아 여러 차례 다시 찾아 봐어도 언제나 밝은 미소로 반겨 주시고 응원해 주시던 장연수 교수님과 이승은 교수님, 세심한 피드백과 조언을 아끼지 않으시며 제 연구가 더욱 견고해질 수 있도록 지도해 주신 점에 깊이 감사드립니다. 석사 논문 지도부터 박사 논문에 이르기까지 간호학 연구가 더 빛날 수 있도록 늘 새로운 가르침을 주시는 박창기 교수님, 제게 항상 용기를 주시고 논문을 마무리할 수 있도록 응원해 주신 점에 깊은 감사를 드립니다. 또한, 연구 방향에 대해 함께 고민해주시고, 응급실 데이터를 통찰력 있게 분석할 수 있도록 도움을 주신 차원철 교수님께도 감사 인사를 드립니다. 그리고 제가 간호사로서, 간호학자로서 앞으로 나아가야 할 방향과 간호학문의 발전을 위해 제가 무엇을 할 수 있을지 고민할 기회를 주신 자랑스러운 연세대학교 간호학과 모든 교수님들께 진심으로 감사의 말씀을 드립니다.

논문이 진행되는 동안 어려움이 없도록 늘 배려해 주시고, 논문을 잘 마무리할 수 있도록 격려해 주신 김나리 파트장님 감사드립니다. 또한 박사과정동안 함께 배움의 즐거움뿐만 아니라 고된 시간과 고민을 나누며 힘들 때 큰 위로가 되어 주신 장미라 선생님, 연구 진행이 원활하지 않아 포기하고 싶을 때 함께 결실을 맺자고 손 내밀어

주시며 격려와 응원을 해 주신 나선경 선생님께 깊이 감사드립니다. 간호 정보학과 관련된 연구에 대해 고민을 나누며, 논문의 데이터 분석 과정에서 도움을 주신 서준혁 선생님께도 감사드립니다.

마지막으로 가족들의 변함없는 사랑과 응원이 없었다면, 이 모든 과정을 끝까지 이어 나가기 어려웠을 것입니다. 여러 가지로 부족한 딸을 자랑스럽게 여기며, 항상 딸의 일과 학업을 응원하시면서 어린 손녀들을 부족함 없이 돌보아 주시는 사랑하는 부모님, 자식으로서의 도리를 다하지 못해도 늘 이해해 주시고 항상 격려해주시는 시부모님께 진심으로 감사드립니다. 부모님들의 사랑과 헌신 덕분에 이 긴 여정을 잘 마무리할 수 있었습니다. 지치고 힘들 때 늘 위로가 되고 제가 하고자 하는 일에 적극적인 지지를 보내주는 남편 지한씨, 일과 공부로 바쁜 엄마의 부재에도 늘 씩씩하고 건강하고 밝게 성장하고 있는 우리 딸 연진, 연서, 연재에게 진심으로 고맙고 사랑한다고 전합니다.

이 논문이 끝이 아니라 새로운 시작이 될 것이라 믿으며, 앞으로도 더 나은 연구로 학문적 발전에 기여할 수 있도록 노력하겠습니다.

감사합니다.

2024 년 12 월
연구자 신 현 아 올림

차 례

차례	i
표 차례	iv
그림 차례	v
부록 차례	v
국문 요약	vi

I. 서 론

1. 연구의 필요성	1
2. 연구의 목적	4
3. 용어 정의	4

II. 문헌고찰

1. KTAS 3, 4 단계 응급환자의 임상적 중요성	6
2. 응급실 환자의 임상 결과 예측 연구 동향	10
3. 간호 데이터를 활용한 연구 동향	22

III. 개념적 기틀

1. HPM-ExpertSignals 개념틀	25
2. 연구의 개념적 기틀	27

IV. 연구방법

1. 연구설계	30
2. 자료원	30
3. 연구대상	31
4. 분석 변수	32
5. 표본 크기 산출	34
6. 자료 구축	34
7. 예측 모델 개발 및 타당도 평가	40
8. 자료 분석	43
9. 윤리적 고려	45

V. 연구결과

1. 코호트 선정 및 데이터 특성	46
2. KTAS 3, 4 단계 환자의 임상결과 및 특성	50
3. KTAS 3, 4 단계 환자의 임상결과에 따른 간호기록 패턴 분석	55
4. 예측 모델 개발 및 성능 평가	58
5. 최종 모델 선정 및 변수 중요도	66
6. 보정 곡선	75

VI. 논의

1. KTAS 3, 4 단계 응급 환자의 임상결과 예측 모델 개발	77
2. 연구의 제한점	82
3. 연구의 간호학적 의의	84

VII. 결론

1. 결론	87
2. 제언	88

참고 문헌	89
부 록	99
Abstract	106
영문요약	108

표 차례

Table 1. Summary of clinical outcome prediction models	17
Table 2. Eligibility criteria	31
Table 3. Summary of candidate predictors	33
Table 4. Summary of selected features	36
Table 5. Comparison of patient characteristics between training and test sets	48
Table 6. Characteristics and clinical outcomes of the included cohort	53
Table 7. Nursing documentation patterns in the included cohort	57
Table 8. Confusion matrix of clinical outcome predictions	60
Table 9. Model performance of clinical outcome prediction models	63

그림 차례

Figure 1. Healthcare process modeling framework to phenotype clinician behaviors for exploiting the signal gain of clinical expertise (HPM-ExpertSignals)	27
Figure 2. Conceptual framework for this study	29
Figure 3. Model development and validation process	42
Figure 4. Flowchart illustrating the cohort selection process for the study	47
Figure 5. Receiver operating characteristic curves of prediction models	64
Figure 6. Precision-recall curves of prediction models	65
Figure 7. Feature importance in the random forest model 2	67
Figure 8. SHAP summary plots for feature contributions in random forest model 2	70
Figure 9. Calibration plot of the random forest model 2	76

부록 차례

Appendix 1. TRIPOD+AI checklist: Model development and validation	99
Appendix 2. Data definition document	102
Appendix 3. Approval from the institutional review board	105

국 문 요 약

전자간호기록을 이용한 한국형 응급환자 분류도구(KTAS)

3, 4 단계 환자의 임상결과 예측 모델 개발

신 현 아

연세대학교 대학원 간호학과

응급실 환자의 임상 결과를 예측하는 것은 의료 자원의 효율적 배분과 신속한 의사결정을 지원하는 데 필수적이다. 특히, KTAS 3, 4 단계 환자는 임상 상태가 다변적이고 예측이 어려워 더욱 정교한 예측 도구가 필요하다. 본 연구는 HPM-ExpertSignals를 개념적 기틀로 하여 KTAS 3, 4 단계 성인 환자의 triage 정보와 간호 데이터를 통합하여 임상 결과 예측 모델을 개발하고 평가하였다.

본 연구는 2023년 1월 1일부터 12월 31일까지 서울의 2,000 병상 규모 상급종합병원 응급실을 방문한 환자의 임상 데이터 웨어하우스에서 추출한 익명화된 데이터를 활용한 후향적 서술 조사 연구이다. 연구 대상은 KTAS 3, 4 단계 성인 환자이며, ICU 입원, 일반병동 입원, 전원, 퇴원, 사망 등 다섯 가지 임상 결과를 예측하는 모델을 개발하였다. 일반 triage 정보를 활용한 Model 1과, triage 정보에 간호 데이터를 추가한 Model 2를 각각 구축하였으며, 다항 로지스틱 회귀, 랜덤 포레스트, 그래디언트 부스트 알고리즘을 적용하였다.

모델 성능은 AUROC, AUPRC, 민감도, 특이도, F1 점수, 보정 곡선 등 다양한 평가 지표를 통해 비교하였다. 클래스 불균형 문제를 해결하기 위해 SMOTE와 비용 민감 학습을 적용하였고, 주요 예측 변수를 식별하고 모델의 해석 가능성을 높이기 위해 셰플리 가산 설명 기법(Shapley Additive Explanations, SHAP)을 사용하였다.

연구 결과, Model 2는 모든 성능 지표에서 Model 1보다 우수한 결과를 보였다. 특히, 랜덤 포레스트 알고리즘을 적용한 Model 2는 AUROC 0.964를 기록하며 가장 높은 예측 성능을 보였고, 최종 모델로 선정되었다. 주요 예측 변수로는 산소포화도 측정 빈도, 응급실 방문 경로, 과거 입원 이력이 도출되었으며, SHAP 분석을 통해 각 변수의 기여도를 시각화하여 모델의 해석 가능성을 강화하였다. SMOTE와 비용 민감 학습은 소수 클래스인 ICU 입원과 사망의 예측 성능을 일부 개선하였으나, 클래스 불균형 문제는 여전히 한계로 남아 있었다.

본 연구는 triage 정보와 간호 데이터를 통합하여 KTAS 3, 4 단계 응급 환자의 임상 결과를 예측하는 모델의 가능성과 한계를 탐구하였다. 개발된 랜덤 포레스트 기반 모델은 높은 예측 성능과 해석 가능성을 바탕으로 임상 의사결정 지원 도구로 활용될 수 있는 잠재력을 보여주었다. 향후 연구에서는 다기관 데이터를 활용해 모델의 일반화 가능성을 검증하고, 클래스 불균형 문제를 해결하기 위한 새로운 접근법을 모색할 필요가 있다.

주요어: KTAS, 응급실, 예측 모델, 간호 데이터, SHAP, 머신러닝, 클래스 불균형, 임상 의사결정 지원

I. 서론

1. 연구의 필요성

국내 응급실에 내원하는 환자 수는 코로나19 팬데믹 이전 연간 1,000만 명을 넘었으나, 2020년 800만 건 미만으로 감소한 뒤 점차 증가하여 2022년에는 약 887만 건으로 보고되었다(응급의료통계연보, 2023). 특히 중증 응급환자의 증가로 응급실 과밀화 문제는 여전히 심각한 상황이다. 과밀화된 응급실에서 모든 환자에게 최적의 의료서비스를 신속하게 제공하는 것이 매우 어렵기 때문에 생명을 위협하는 질환을 가진 환자를 조기에 인지하고 적절한 조치를 취하는 것이 중요하다(Lee et al., 2019).

이러한 예측되지 않는 패턴으로 내원하는 다양한 중증도의 급성 질환을 가진 환자들에게 골든 타임 내에 신속하고 적절한 응급 의료서비스를 제공하기 위해 환자의 중증도 분류 및 한정된 의료자원을 배정하는 의사결정 과정을 triage라고 한다(Patel et al., 2008). Triage는 국내외 대부분 간호사에 의해 수행되며 (Göransson et al., 2005; Park et al., 2014), 대표적인 분류 도구로 Canadian Emergency Department Triage and Acuity Scale(CTAS), Australian Triage Scale(ATS), Emergency Severity Index(ESI), Manchester Triage Scale(MTS)가 응급 환자를 5단계로 분류하는 체계로써 사용되고 있다(Zachariasse et al., 2019).

국내에서는 2016년부터 캐나다 응급환자 분류도구(CTAS)를 기반으로 개발된 한국형 응급환자 분류도구인 Korean Triage and Acuity Scale(KTAS)가 표준 도구로 사용되고 있다. KTAS는 환자의 중증도와 긴급도를 평가하여 5단계로 분류하며, 각 단계의 환자 상태에 따라 진료 권고 시간을 정한다. KTAS 1단계 “소생”은 즉각적인 처치를, 2단계 “긴급”은 15분 이내, 3단계 “응급”은 30분 이내, 4단계 “준응급”은

1시간 이내, 5단계 “비응급”은 2시간 이내 처치 또는 대기 중 재평가를 시행하도록 권고되고 있다(대한응급의학회 KTAS 위원회, 2021). 이에 따라 KTAS 1, 2단계 환자는 즉각적인 처치가 이루어지지만, 3단계 이하로 분류된 환자들은 과소평가 될 경우 신속한 진료 및 처치가 지연됨에 따라 환자 상태가 악화되는 결과가 초래될 수 있고, 과대평가 될 경우 인적, 물적 자원 낭비 문제가 발생할 수 있다(Ekins & Morphet, 2015). 국내 KTAS 재평가 분석 결과, 환자가 과소평가되는 경향이 있으며, 특히 2단계 이상 과소평가된 경우 중환자실 입원 및 사망 위험이 높아 초기 KTAS 분류 시 더 신중한 평가가 필요하다고 언급하였다(이은실, 2019).

2022년 통계에 따르면, 응급실 환자 중 KTAS 1단계는 1.3%, 2단계 5.8%, 3단계 43.4%, 4단계 39.4%, 5단계 10.0% 로 보고되어(응급의료통계연보, 2023), 3단계 “응급”과 4단계 “준응급” 환자가 전체의 80% 이상 대다수를 차지한다. 이들 환자들은 고위험 환자(KTAS 1, 2단계)와 비교할 때 임상 결과 예측이 상대적으로 어렵지만 적절한 응급 치료와 자원의 효율적 운용을 위해 매우 중요한 그룹이다. 선행연구에 따르면 중증도 1, 2단계로 분류된 환자의 임상결과(병원 입원, 중환자실/수술실 입원 또는 사망) 예측력은 높은 반면, 3단계 환자군은 가장 많은 비중을 차지함에도 불구하고 예측력이 낮아 이 그룹에서 위험도가 높은 환자를 식별하기 위한 연구의 필요성이 제기되었다(Kwon et al., 2018; Lee et al., 2021; Ruger et al., 2007).

특히, 다중범주 결과변수를 사용하는 접근법의 유용성은 선행 연구에서 확인된 바 있다. Riordan et al.(2017)의 연구는 ESI 3단계 환자를 대상으로 퇴원을 예측하는 이분 변수 모델을 개발하였으나, AUROC(area under the receiver operating characteristic)가 0.730으로 보고되어 중간 단계 환자의 다양한 임상 양상을 충분히 반영하지 못하였다. 이는 단순한 이분 변수 모델이 복잡한 환자군의 특성을 설명하는

데 한계가 있음을 보여준다. 반면, Lee et al.(2020b)의 연구는 귀가, 중환자실 입원, 텔레모니터 병동 입원, 일반 병실 입원, 관찰 병동 입원의 다중범주 결과변수를 활용하여, 환자 흐름을 세부적으로 예측하고 자원 배치를 최적화하려는 접근법을 제시하였다. 이 연구는 주요 입원 유형을 예측하며 AUROC 0.84에서 0.97의 수준을 보고하였고, 특히 중환자실 환자 예측에서 가장 우수한 민감도와 정확도를 나타냈다.

한편, 간호기록은 환자의 상태 악화 및 사망률을 예측하는 중요한 도구가 될 수 있다(Collins et al., 2013). 간호사들은 환자의 상태 변화에 대한 우려 및 비정상 범주의 임상 데이터 모니터링을 강화하기 위해 전자의무기록(EMR) 시스템을 통해 활력 징후를 더 자주 기록하고, 임상 측정과 관련된 비정형화된 기록을 입력하는 경향이 있다(Collins & Vawdrey, 2012). 따라서 이러한 전자간호기록의 패턴 및 특징을 분석하면 환자의 임상 결과를 예측하는 데 중요한 통찰을 얻을 수 있다. 하지만 그동안 응급 환자의 입원, 사망 등 예후를 예측하기 위한 모델 개발 연구가 국내외 다수 시행되었음에도 지금까지 대부분의 연구들은 응급실 triage 정보나 검사 결과에 초점을 맞추고, 전자간호기록 등의 간호 관련 정보를 예측 변수로 하여 응급실 환자의 예후와의 관련성을 보고한 연구는 드물다(Brink et al., 2022; Larburu et al., 2023).

따라서 본 연구는 전자간호기록의 패턴 및 특징을 포함한 예측 변수를 기반으로 KTAS 3, 4단계 환자의 임상 결과 예측 모델을 개발하고 그 효과를 평가하고자 한다. 이를 통해 응급실 내원 환자의 임상 결과를 보다 정확하게 예측하여, 의료 자원의 효율적 배분과 신속한 의사결정을 가능하게 하고, 궁극적으로 응급의료 서비스의 질을 향상시키는데 기여하고자 한다. 특히, 본 연구는 간호기록을 주요 예측 변수로 활용함으로써 기존 모델의 한계를 보완하고, 응급실 내에서 간호기록의 임상적 가치와 과학적 근거를 강화하고자 한다.

2. 연구의 목적

본 연구의 목적은 KTAS 3, 4단계 응급실 환자를 대상으로, 전자간호기록의 패턴 및 특징을 포함한 다양한 예측 변수를 활용한 머신러닝 기반 임상 결과 예측 모델을 개발하고 검증하는 것이다. 구체적인 연구 목적은 다음과 같다.

- 1) KTAS 3, 4단계 응급 환자의 임상 결과 예측 모델을 개발하기 위하여, 일반 triage 데이터를 활용한 모델 1과 triage 데이터에 간호 데이터를 통합한 모델 2를 각각 구축한다.
- 2) 머신러닝 알고리즘을 활용하여 예측 모델을 개발한다.
- 3) 내부 검증 코호트를 활용하여 개발된 예측 모델의 성능을 평가 지표를 통해 비교 분석하고, 응급실 환경에 적합한 최적 모델을 선정한다.
- 4) 최적 모델에서 중요한 예측 변수를 확인하고, 신뢰성을 평가한다.

3. 용어의 정의

1) KTAS 3, 4단계 응급 환자

한국형 응급환자 분류도구(Korean Triage and Acuity Scale; KTAS)에 따르면 KTAS 3단계는 “응급 처치가 필요하며 심각한 문제로 진행할 수 있는 잠재적 위험이 있는 상태”를 의미하며, “응급”으로 분류한다. 4단계는 “환자의 나이, 불편한 정도, 악화될 가능성 등을 고려할 때 한 두 시간 안에 치료 혹은 재평가하면 되는 상태”를 의미하며, “준응급”으로 분류한다(대한응급의학회 KTAS 위원회, 2021).

본 연구에서는 한국형 응급환자 분류도구에 따라 서울시 소재 일개 상급종합병원 응급실 triage 간호사에 의해 최초 평가된 KTAS 분류 결과가 3단계와 4단계인 환자로

정의한다.

2) 전자간호기록

전자간호기록이란 전자의무기록 중 간호사에 의해서 작성된 기록으로 환자 사정, 간호 문제 및 목표 확인, 간호 중재 계획 및 수행 결과, 평가 등의 내용 등을 포함한다. 전자간호기록의 주요한 목적은 효율적이고 양질의 간호를 제공하는 것이며 과거, 현재, 예상되는 미래의 기록을 포함하며 인증된 사용자들이 접근할 수 있으며 쉽게 저장되고 교환 가능한 특징을 가진다(Häyrinen et al., 2008).

본 연구에서는 연구기관의 전자의무기록 시스템을 통해 응급실 간호사가 작성한 간호초기평가, 간호기록, 임상관찰기록지(flowsheet)를 전자간호기록으로 정의한다.

3) 응급실 환자의 임상 결과

본 연구에서 응급실 환자의 임상 결과란 응급실 최종 진료 결과를 의미하며, “일반병동으로 입원”, “중환자실로 입원”, “귀가”, “전원”, “사망”의 5가지로 구분된다.

II. 문헌 고찰

본 장에서는 한국형 응급환자 분류도구의 개념과 관련 선행 문헌에 대해 살펴보고, 응급실 환자의 임상 결과를 예측하기 위한 목적으로 시행된 선행 문헌을 고찰하여 주요 결과 변수 및 관련 예측 요인을 파악하고자 한다. 또한 간호기록을 활용한 선행 연구의 동향을 확인함으로써 응급실 환자의 임상 결과 예측에 있어 간호기록의 중요성을 논의하고자 한다.

1. KTAS 3, 4단계 응급환자의 임상적 중요성

한국형 응급환자 분류도구(Korean Triage and Acuity Scale, KTAS)는 캐나다 응급환자 분류도구(CTAS)를 기반으로 개발되었으며, 환자의 상태를 신속하고 체계적으로 평가하여 적절한 의료 자원의 배분과 신속한 응급 처치를 가능하게 한다. KTAS 는 2016 년부터 국내 전역의 응급의료기관에서 표준 도구로 사용되며, 응급 환자의 중증도와 긴급도를 평가하여 5 단계로 분류하는 체계이다(대한응급의학회 KTAS 위원회, 2021). KTAS 1 단계는 “생명이나 사지를 소실할 정도의 위협이 있어 적극적인 처치를 필요로 하는 상황”을 의미하며, 심정지나 쇼크, 심각한 호흡곤란, 무의식 상태 등 환자들은 아주 명백한 고통을 호소하며 불안정한 활력 징후를 보인다. 2 단계는 “생명 혹은 사지를 소실할 정도의 잠재적인 위협이 있어 의사 혹은 의료 지시에 따라 빠른 처치가 필요한 상황”을 의미한다. 3 단계는 “응급 처치가 필요하며 심각한 문제로 진행할 수 있는 잠재적 위험이 있는 상태”를 의미하며, 4 단계는 “환자의 나이, 불편한 정도, 악화될 가능성 등을 고려할 때 한 두 시간 안에 치료 혹은 재평가하면 되는 상태”를 의미한다. 5 단계는 “급성 발병이지만

긴급하지 않은 상황이며 악화된 또는 악화되지 않더라도 만성적인 문제의 일부분일 수 있는 상태”를 의미한다(대한응급의학회 KTAS 위원회, 2021).

따라서 KTAS 1, 2 단계로 분류된 환자는 즉각적인 처치가 이루어지고, KTAS 5 단계 환자는 비응급 환자이기 때문에 이들에 대한 임상결과는 쉽게 예측이 가능한 반면, KTAS 3 단계와 4 단계 환자는 중간 위험군으로 분류되며, 임상적 결과 예측하는 것은 더 복잡하고 어려운 과제로 여겨진다(Ekins & Morphet, 2015). KTAS 3 단계는 전체 응급실 환자 중 가장 큰 비중(약 43.4%)을 차지하며, 심각한 문제로 진행할 가능성이 높은 환자와 그렇지 않은 환자가 혼재되어 있어, 적절한 치료 우선순위를 결정하기 어려운 경우가 많다. KTAS 4 단계는 상대적으로 긴급성이 낮은 환자로 간주되지만, 과소평가된 환자의 경우 상태 악화 가능성이 존재한다. 실제로 국내 연구에서 KTAS 3, 4 단계 환자가 과소평가될 가능성이 높고, 이로 인해 중환자실 입원, 사망 또는 타 병원으로 전원되는 사례가 많아 초기 평가의 신중함이 필요하다고 보고되었다(이은실, 2019).

따라서 KTAS 3, 4 단계 환자에 대한 중증도 분류를 더욱 정교화하고, 임상적 결과를 정확히 예측하기 위한 도구와 방법이 필요하다. 최근 KTAS 분류의 정확성과 신뢰성을 향상시키고, 분류자의 주관적 판단에서 기인하는 비일관성과 예측 오류의 한계를 극복하기 위해 데이터 기반의 머신러닝 및 딥러닝 기법을 활용한 다양한 연구가 수행되었다. Kwon et al.(2018)과 Yu et al.(2020)의 연구는 딥러닝 모델을 활용하여 KTAS의 임상적 결과 예측을 개선하는데 중점을 두었다. Kwon et al.(2018)의 연구는 국가응급진료정보망(NEDIS, national emergency department information system) 데이터를 활용하여 딥러닝 기반의 분류 도구(DTAS, deep learning based triage and acuity score)를 개발하였다. 병원 내 사망률, 중환자 치료 필요성, 입원 여부 등 다양한 결과를 예측하였으며, AUROC 0.804 ~ 0.935 로 기존의 KTAS 보다(AUROC:

0.681~0.797) 더 높은 정확도를 나타냈다. 특히 외부 검증에서 높은 성능을 유지하여 일반화 가능성을 입증하였으며, 응급실에서 신속하고 객관적인 분류 도구로 활용될 가능성을 제시하였다. 또한 기존 KTAS 및 SOFA(Sequential Organ Failure Assessment) 모델의 한계를 보완하기 위해 초기 간호평가((Initial Nursing Assessment) 데이터를 활용하여 딥러닝 기반 분류 모델을 개발한 연구(Yu et al., 2020)에서는 중환자실 입원 또는 사망 결과를 예측하였는데, 초기 간호평가 데이터를 활용한 딥러닝 기반 모델이 AUROC 0.876 의 높은 성능을 기록하며 기존 KTAS(0.768) 및 SOFA(0.740)보다 우수한 임상결과(중환자실 입원 또는 사망) 예측력을 보고 하였다. 특히 주증상, 나이, 심박수, 산소포화도와 같은 변수를 활용한 간소화된 모델(AUROC 0.807)도 기존 KTAS, SOFA 보다 성능이 높은 것으로 보고하였다.

한편, Kim et al.(2021)과 Chang et al.(2023)의 연구는 KTAS 의 적용성과 신뢰성을 확장하는 데 초점을 맞추었다. Kim et al.(2021)은 음성 인식과 자연어 처리 기술을 결합하여 환자와 의료진 간의 대화를 분석함으로써 KTAS 분류를 자동화하는 시스템을 개발하였다. KTAS 분류 예측에서 AUROC 0.890 의 높은 성능을 나타낸 이 연구는 응급의료 환경에서 시간과 자원이 제한된 상황에서도 신속하고 정확한 triage 를 지원할 수 있는 자동화된 의사결정 도구로 활용될 가능성을 제시하였다. Chang et al.(2023)은 다기관 데이터의 품질 차이를 보완하고 개인정보를 보호하며 학습이 가능하도록 설계된 연합 학습(Federated Learning)을 활용하여 KTAS 단계 수정 필요성을 예측하는 모델을 개발하였다. 이 모델은 외부 검증에서 AUROC 0.750 의 예측 성능을 나타냈으며, KTAS 단계 수정이 필요한 환자는 입원률과 응급실 사망률이 높은 경향을 보였음을 보고하였다. 이는 초기 평가에서 과소평가된 환자를 조기에 식별하고 적절한 치료를 제공할 수 있는 가능성을 시사한다.

이처럼 머신러닝 및 딥러닝 기반의 연구들은 KTAS 분류의 정확성과 효율성을 향상시키기 위한 좋은 전략으로 평가되고 있다. 이러한 기법들은 분류자 간의 주관적 편차를 줄이고, 신속하고 정확한 환자 분류를 가능하게 함으로써 응급실 내 자원 배분과 환자 치료 우선순위 결정에 기여할 수 있다. 하지만 대부분의 연구는 전체 응급실 환자 또는 KTAS 1, 2 단계 환자에 집중되어 있으며, KTAS 3, 4 단계 환자만을 대상으로 한 연구는 상대적으로 부족하다.

KTAS 3,4 단계 환자와 같은 중증도 분류 중간 단계에 해당하는 환자들은 임상적으로 중요한 그룹으로, 이들에 대한 정확한 중증도 분류와 임상 결과 예측은 응급실 운영의 효율성과 환자 안전과 관련된다(Ketseridou et al., 2021; Levin et al., 2018). 따라서 이들을 대상으로 예측 모델을 정교화하고, 임상적 검증을 통해 응급의료 현장에서 실질적인 적용 가능성을 탐구할 필요가 있다.

2. 응급실 환자의 임상 결과 예측 연구 동향

응급실에서 환자의 임상 결과를 예측하는 다양한 연구들이 진행되어 왔으며, 특히 입원 여부, 사망 위험, 퇴원 가능성 등과 같은 주요 결과 변수를 예측하기 위해 다양한 예측 모델이 사용되었다. 이러한 국내외 응급실 환자의 임상 결과 예측 모델 개발 연구 현황 및 주요 결과 변수, 관련 예측 요인을 살펴보고자 문헌고찰을 시행하였다.

본 연구의 문헌고찰을 위하여 사용한 검색엔진은 PubMed, CINAHL, Cochrane CENTRAL, EMBASE, IEEE Xplore Library, Web of Science, and Scopus 이다. 검색어는 MeSH 용어를 기반으로 구성되었으며, 각 데이터베이스의 특성에 맞게 조정되었다. 검색어는 OR 및 AND 연산자를 사용하여 결합하였으며, 주요 MeSH 용어로는 "emergencies", "triage", "prognosis"를 사용하였다. 검색 결과 제목 및 초록을 검토하여 응급실 환자를 대상으로 한 예측 모델 개발 연구를 포함하였으며, 타당도 검증만 시행한 연구와 대상자가 특정 질환이나 노인 또는 소아만을 대상으로 한 연구는 제외하였다.

위와 같은 선정 기준으로 총 26 개의 문헌을 검토하였는데 각 문헌의 저자, 년도, 국가, 자료원, 표본 수, 결과 변수, 예측 변수, 모델 개발 분석 기법, AUROC를 정리하였다(Table 1). 문헌고찰에 포함된 26 개의 연구 모두 수집된 데이터를 후향적으로 분석하여 조사한 연구였으며, 미국(9 편), 한국(4 편), 호주(3 편), 싱가포르(3 편), 대만(1 편), 유럽 국가 등 다양한 국가에서 연구가 진행되었다. 자료원은 국가응급의료정보를 활용한 연구가 6 편 (Kwon et al., 2018; Lee et al., 2020a; Parker et al., 2019; Raita et al., 2019; Yun et al., 2021; Zhang et al., 2017), 다기관 자료를 활용하여 분석한 연구가 6 편이고, 나머지 14 편의 연구는 모두 단일 기관의 자료를 사용하여 모델을 개발하였다. 표본 크기는 Lucke et al.(2018)의

연구에서 활용한 소규모 데이터셋(21,287 건)부터 Kwon et al.(2018)의 연구에서 활용한 대규모 데이터셋(10,967,518 건)까지 연구마다 큰 차이를 보였다.

예측하고자 하는 결과변수는 “입원(hospital admission)”이 가장 많았고(n=15), 중환자실 입원과 사망을 모두 포함하여 정의한 “중증치료(critical care)”, “사망(mortality)”을 예측한 연구가 대다수였으며 “귀가(discharge)”를 예측하는 연구가 1 편(Riordan et al., 2017) 있었다. 대부분의 연구가 단일 결과변수를 대상으로 예측 모델을 개발했지만, 2 개 이상의 결과변수를 다룬 연구는 3 편(Kwon et al., 2018; Levin et al., 2018; Raita et al., 2019)이었다. 예를 들어, Kwon et al.(2018)은 주요 결과변수인 사망 이외에 2 차 결과변수로 critical care, 입원에 대한 예측 결과를 보고하였다. 또한 결과변수는 대부분 이분변수로 정의하여 분석한 연구가 대부분이었으며, 다중범주 변수로 분석한 연구가 1 편(Lee et al., 2020b)이 있었다. Lee et al.(2020b)의 연구는 중환자실, 텔레모니터 병동, 일반병동, 관찰병동 입원과 귀가 등 5 개 범주로 결과변수를 정의하여 다중범주 모델링을 시도했으며, 이는 이분변수 모델링과 비교해 임상적 의사결정에 보다 구체적인 정보를 제공한다는 점에서 의의가 있다고 강조하였다.

응급실 환자의 입원 예측 모델을 개발한 연구는 총 15 편으로 이 중 3 편에서는 이차 결과변수로 제시되었다. 15 편의 연구에서 제시한 입원율은 13.4%에서 40.7%까지 다양하게 보고되었는데, 이는 국가마다 의료체계가 상이하기 때문인 것으로 생각된다. 미국과 스페인 병원의 응급실 자료를 이용한 연구(Raita et al., 2019; Zhang et al., 2017; Zlotnik et al., 2016)에서는 약 13~16%로 낮은 입원율을 보고하고 있는 반면 호주나 싱가포르 병원의 자료를 이용한 연구(Dinh et al., 2016; Kim et al., 2014; Parker, 2019; Sun, 2011)에서는 약 30~40%의 높은 입원율을 보고하였다. 국내 연구는

국가응급의료정보망(NEDIS)을 통해 수집된 자료를 분석하여 보고하였는데 입원율은 27.0%로 보고하였다(Kwon et al., 2018).

입원 예측 모델에서 사용된 주요 예측 변수를 정리해 보면, 크게 인구학적 요인, triage 에서 수집되는 변수, 그리고 병원 정보를 통해 수집되는 변수로 구분할 수 있다. 인구학적 요인은 나이, 성별, 인종, 언어, 고용 상태, 보험 상태, 결혼 상태, 종교, 주거지 정보(postal code)등이 포함된다. triage 에서 수집되는 변수는 내원 시간, 내원 수단, 활력징후(혈압, 맥박, 호흡수, 체온, 산소포화도), 중증도 평가 결과, 주증상 등으로 보고되었는데 중증도 평가 결과는 각 나라마다 다양한 도구 사용되었다. 또한 환자의 병원 정보를 통해 수집할 수 있는 변수는 병원 이용 상태(최근 1년 동안 응급실 내원 횟수, 최근 1년 동안 입원 횟수, 이전 응급실 내원 시 퇴실 결과, 과거 시술 또는 수술력), 과거 상병 정보 등이 포함되었다. 이러한 변수 중에 나이, 내원 수단, 활력징후, 주증상은 대부분의 연구에서 공통적으로 사용된 변수로, 입원 여부를 예측하는 데 중요한 변수로 확인되었다.

응급실 환자의 중증치료(critical care) 예측 모델은 총 7 편의 연구에서 시행되었으며, 이 중 Kwon et al.(2018)의 연구는 중증치료를 이차 결과변수로 제시하였다. 중증치료 비율은 2.1%부터 13.7% 범위로 보고되었으며, 가장 높은 비율을 나타낸 Joseph et al.(2020)의 연구에서는 critical care 에 대한 정의가 구체적으로 명시되어 있다. 이 연구는 응급실 내원 24 시간 이내 사망, 중환자실 입원, 그리고 일반병동 입원 후 24 시간 이내 중환자실로 전동한 경우를 포함하였다. 반면, 다른 연구에서는 주로 중환자실 입원과 사망을 포함하여 critical care 를 정의하였는데, Joseph et al.(2020)과 같은 넓은 범위의 정의를 사용한 경우, 발생율이 더 높게 보고된 경향이 있다. 3 편의 국내 연구(Kwon et al., 2018; Yu et al., 2020; Yun et al., 2021)에서 보고한 중증치료 발생율은 3.5%~5.7% 이다.

중증치료를 예측하기 위해 개발된 모델들은 인구학적, 임상적, 그리고 상황적 정보를 포괄하는 다양한 변수를 활용하고 있다. 인구학적 변수로는 모든 연구에서 공통적으로 나이와 성별이 주요 예측 변수로 포함되었다. triage 에서 수집되는 변수로는 활력징후, 의식수준, 주증상, 내원수단, 증상 발생 후 응급실 도착까지 소요시간이 포함되었다. 또한 일부 연구에서는 환자의 동반질환, 과거 병력을 포함하였다. triage acuity scale 이 입원 예측 모델에서는 대부분 예측 변수에 포함된 반면, 중증치료 예측 모델에서는 포함되지 않았다.

중증치료 예측 모델을 개발하기 위해 로지스틱 회귀분석, 앙상블 모델인 랜덤 포레스트, 그래디언트 부스트, 딥러닝 등 다양한 알고리즘이 사용되었다. 딥러닝 알고리즘은 특히 대규모 데이터셋에서 높은 예측 성능을 보여주며, 텍스트 데이터와 같은 비정형 데이터의 활용에서도 두각을 나타내었다. 특히, Chen et al.(2023)의 연구와 Joseph et al.(2020)의 연구에서는 비구조화된 텍스트 데이터를 포함하여 신경망 모델을 이용한 예측 모델을 개발하였는데 두 연구 모두 AUROC 0.850 이상의 안정적인 모델 성능을 보고하였다. 또한 Kwon et al.(2018)은 약 1천만 건 이상의 대규모 NEDIS 데이터를 사용하여 딥러닝을 적용하였고 중증치료 예측에서 AUROC 0.894 의 높은 모델 성능을 보고하였다. 가장 높은 성능을 보고한 연구는 Levin et al.(2018)의 연구로 랜덤 포레스트를 사용해 AUROC 0.900-0.920 이라는 매우 높은 성능을 달성하였다.

응급 환자 사망(mortality) 예측 모델 개발 연구는 총 6편으로 사망률은 0.48%에서 7.99%까지 다양한 범위로 보고되었다. 이는 국가별 의료시스템의 차이와 연구마다 사망률에 대한 정의가 상이하기 때문인 것으로 보인다. 예를 들어, 일부 연구(Teubner et al., 2015)는 응급실에서 사망한 경우를 포함하지 않은 반면, 다른 연구에서는 포함하였다. 또한, 일부 연구(Klug et al., 2020; Lee et al., 2020a; Xie

et al., 2021)는 시간 범위에 따라 세분화하여 보고하였다. 가장 높은 사망률을 보고한 연구는 Lee et al.(2020a)의 연구로, 한국의 NEDIS 데이터를 기반으로 30 일 이내 사망률을 7.99%로 보고하였다. 이 연구는 사망률을 24 시간(3.04%), 48 시간(4.08%), 7 일(5.82%), 그리고 30 일(7.99%)로 기간에 따라 세분화하여 제시하였다. 마찬가지로, 싱가포르에서 단일 병원 데이터를 활용한 Xie et al.(2021) 연구에서는 30 일 이내 사망률이 5.9%로 보고되었다. 반면, 특정 기간에 대해 구체적으로 명시하지 않은 Teubner et al.(2015)의 연구에서는 사망률을 1.64%에서 3.04%로 보고하였고, 국내 대규모 데이터셋을 활용한 Kwon et al.(2018) 연구는 일차 결과변수로써 사망율을 1.4%로 보고하였다.

사망 예측 모델에서 사용된 주요 예측 변수를 정리해 보면, 인구학적 요인으로 대부분의 연구에서 나이와 성별이 기본적인 예측 변수로 사용되었다. triage에서 수집되는 변수로는 활력징후, 의식수준, 주증상, 내원수단이 포함되었다. 또한 일부 연구 (Teubner et al., 2015; Klug et al., 2020)에서 triage scale 이 예측 변수에 포함되었다.

사망 예측 모델 개발에서도 다양한 알고리즘이 사용되었으며, 모델 성능은 AUROC 0.821 에서 0.962 까지 다양하게 보고되었다. 특히, Fernandes et al.(2020) 연구의 XGBoost 모델(AUROC 0.960)과 Klug et al.(2020) 연구의 XGBoost 모델(AUROC 0.962)이 가장 높은 성능을 보고하였다. 이는 복잡한 데이터 패턴을 처리할 수 있는 앙상블 학습의 장점을 보여준다. Teubner et al.(2015)와 Lee et al.(2020a)는 로지스틱 회귀 기반 모델을 사용하여 AUROC 0.832~0.910 수준으로 안정적인 성능을 나타냈으며, 간단하고 해석 가능한 모델링 접근 방식을 수행했다. Teubner et al.(2015)는 사망률을 예측하는 데 있어 모델 개발과 검증 데이터에서 각각 AUROC 0.859 와 0.837 로 외부 검증에서도 일관된 성능을 나타냈다. Lee et al.(2020a)는 기간별로

세분화된 사망률을 예측하며, 예측 기간이 길어질수록 AUROC 가 점진적으로 감소하는 경향을 나타냈다. 이는 예측 기간이 길어질수록 추가적인 변수가 필요하며, 초기 정보만으로는 장기적인 결과를 정확히 예측하기 어렵다는 점을 반영한다.

Riordan(2017)의 연구는 응급실 자원 관리의 중요성에 초점을 맞추어, 결과변수를 귀가(discharge)로 설정하여 중증도 중간 단계인 ESI 3 단계 환자를 대상으로 예측모델을 개발하였다. 주요 예측변수로는 나이, 성별, 내원 수단, 활력징후가 포함되었으며, 특히 통증 점수를 포함한 점이 다른 모델과 차이가 있었다. 로지스틱 회귀분석을 활용하여 개발된 모델의 AUROC 는 0.730 으로 보고되었으며, 이는 예측 성능이 상대적으로 낮은 수준이었다. 이는 ESI 3 단계 환자 집단의 다양한 임상 양상이 모델 성능에 영향을 미쳤을 가능성을 시사한다. Riordan et al.(2017)의 연구는 응급실 환자가 귀가할 가능성을 예측하여 자원 분배를 효율적으로 관리하려는 시도를 하였다. 그러나 결과 변수를 이분화하여 설정한 제한점으로 인해, 환자별로 세부적인 자원 배치를 최적화하는 데에는 한계가 있었다.

한편, Lee et al.(2020b)의 연구는 응급실 환자의 입원 준비 시간(boarding time)을 감소시켜 응급실 과밀화 문제를 해결하려는 노력에 초점을 맞추었다. 이 연구는 결과변수를 귀가, 중환자실 입원, 텔레모니터 병동 입원, 일반병실 입원, 관찰병동 입원의 5 개 다중범주 변수로 설정하여, 환자 흐름을 세부적으로 예측하고 자원을 적시에 배치할 수 있도록 설계되었다. 특히, 계층적 다중 클래스 분류 모델을 도입하여 결과 변수 간 상호 연관성을 반영하였으며, 데이터를 네 가지 주요 단계(도착, 분류, 초기 평가, 첫 진단검사 결과 보고 시점)로 나누어 각 단계에서의 예측 성능을 분석하였다. Lee et al.(2020b)의 연구에서 입원 여부를 예측하는 모델의 AUROC 는 진단검사 결과 보고 시점에서 중환자실 입원인 경우 0.970 으로 가장 높은 성능을 보였다. 이 연구는 단순한 이분 변수 모델에서 벗어나 다중 클래스 모델을

통해 병동별 자원 배치를 최적화하고, 응급실 내 환자 흐름의 병목 현상을 줄이기 위한 실질적인 기여를 하고자 했다.

위와 같은 선행 문헌 고찰을 통해 각 결과변수에 대한 주요 예측 변수와 모델 개발에 효과적인 알고리즘을 확인함으로써 본 연구의 방향성을 구체화할 수 있었다. 본 연구는 기존 연구에서 주로 이분변수를 사용하여 입원 여부, 사망, 중환자실 입원 등을 예측한 한계를 극복하고, 다중범주 결과변수를 활용하여 보다 세부적이고 실질적인 임상결정을 지원하고자 한다. 특히 KTAS 3,4 단계 환자를 대상으로 전자간호기록을 포함한 다양한 예측변수를 활용하여 일반 병동 입원, 중환자실 입원, 사망, 귀가, 전원이라는 다섯 가지 주요 임상결과를 예측함으로써, 기존 모델 대비 임상적 의사결정에 구체적이고 포괄적인 정보를 제공할 수 있을 것으로 기대된다. 이러한 접근은 Lee et al.(2020b)의 연구와 같이 다중범주 변수의 적용이 응급실 자원 관리와 환자 흐름 최적화에 기여할 수 있음을 보여주는 선행 연구를 기반으로 한다. 나아가, 간호 데이터를 포함한 예측 모델 개발은 기존의 triage 정보 중심의 모델링을 보완하고, 환자 개인별 및 환경적 요인을 포괄하는 새로운 관점을 제시할 것이다.

Table 1. Summary of clinical outcome prediction model

Author (year)	Country (setting)	Data source (sample size)	Outcome rates	Predictors	Algorithm	AUROC (95% CI)
Hospital admission						
Araz (2019)	USA (SC)	ED data (118,005)	-	age, gender, arrival mode, triage acuity level, flu season, time of day, ICD-9 coded complaints	LR	0.832
					ANN	0.835
					SVM	0.790
					DT	0.810
					RF	0.770
					XGBoost	0.863
Cameron (2015)	Scotland (MC)	ED data (322,846)	-	age, NEWS, triage category, referred by GP, arrive by ambulance, admitted within 1 year	LR	0.877 (0.875-0.880)
Dinh (2016)	Australia (MC)	EDDC (1,721,294)	40.7%	Age, ambulance arrival, triage category, previous admission, presenting problem, ED arrival time	LR	0.820 (0.810-0.820)
Graham (2018)	Ireland (MC)	ED data (107,545)	24.0%	hospital site, age, arrival mode, triage category, care group, previous admission	LR	0.849
					DT	0.824
					GBM	0.859
Handly (2015)	USA (SC)	ED data (derivation; 74,056 validation; 85,144)	30.8% derivation 26.3% validation	age, sex, race, time of arrival, day of arrival, emergency severity index (ESI), coded chief complaint (CCC)	NN	0.840
					without CCC	(0.838-0.842)
					with CCC	0.860 (0.858-0.862)
Hong (2018)	USA (MC)	ED data (560,486)	29.7%	ESI level, age, gender, ethnicity, primary language, number of ED visits within 1 year, number of admissions within 1 year, disposition of the previous ED visit, total number of prior surgeries or procedures, medication counts, insurance type	LR	Full set: 0.909 (0.906-0.911)
					XGBoost	0.924 (0.922-0.927)
					DNN	0.920 (0.917-0.922)
Kim (2014)	Australia (SC)	ED data (100,123)	38.6%	age, sex, time of day, day of the week, ATS category, arrival by ambulance, referral by local medical officer, blood test results	LR	0.835 (0.833-0.837)

Abbreviations. AUROC = area under the receiver operating characteristic curve, CI = confidence interval, ED = emergency department, EDDC = emergency department data collection, ANN = artificial neural network, SVM = support vector machine, DT = decision tree, RF = random forest, XGBoost = extreme gradient boosting, NEWS = national early warning score, GP = general practitioner, LR = logistic regression, GBM = gradient boosting machine, NN = neural network, ESI = emergency severity index, CCC = coded chief complaint, ATS = Australasian triage scale, DNN = deep neural network, ICD-9 = international classification of diseases, 9th revision, MC = multicenter, SC = single center.

Table 1. Summary of clinical outcome prediction model (continued)

Author (year)	Country (setting)	Data source (sample size)	Outcome rates	Predictors	Algorithm	AUROC (95% CI)
Hospital admission						
Kwon (2018)	Korea (MC)	NEDIS data (10,967,518) EMR data (13,989)	27.0%	age, sex, chief complaint, symptom onset time, arrival mode, trauma, vital signs, mental status	deep learning	0.804 (0.803–0.804)
Levin (2018)	USA (MC)	ED data (172,726)	22.3– 26.0%	age, sex, arrival mode, vital signs (BT, PR, RR, SBP, SpO2), chief complaint, medical history	RF	0.820–0.840
Lucke (2018)	Netherlands (SC)	ED data (21,287)	① < 70 y : 23.1% ② ≥ 70 y : 43.2%	age, sex, triage category (MTS), mode of arrival, performance of blood test, chief complaint, ED revisit, type of specialist, blood sample, vital signs	multivariable LR	① 0.860 (0.850–0.870) ② 0.770 (0.750–0.790)
Parker (2019)	Singapore (SC)	eHINTS (1,232,016)	38.7%	age, race, postal code, day of week, shift time, mode of arrival, triage category (PAC), fever status	LR	0.825 (0.824–0.827)
Raita (2019)	USA (MC)	NHAMCS (135,470)	16.2%	age, sex, mode of arrival, vital signs, chief complaints, comorbidities	LR, Lasso Regression RF, GBDT, DNN	DNN 0.820 (0.820–0.830)
Sun (2011)	Singapore (SC)	ED data (317,581)	30.2%	age, PAC status, arrival mode, ethnic group, chronic conditions (DM, HTN, dyslipidemia), ED/hospital visit in last 3 months	LR	0.849 (0.847–0.851)
Zhang (2017)	USA (MC)	NHAMCS (47,200)	13.4%	age, sex, race, vital signs, triage level, comorbidities, arrival mode, chief complaint (PCA from NLP)	LR MLNN	0.846 (0.839–0.853) 0.844 (0.836–0.852)
Zlotnik (2016)	Spain (SC)	ED data (255,668)	13.6%	age, gender, visit source, arrival mode, MTS triage level, MTS chief complaint group, insurance status, visit	LR ANN	0.857 (0.851–0.858) 0.858 (0.854–0.861)

Abbreviations. EMR = electronic medical record, NEDIS = national emergency department information system, BT = body temperature, PR = pulse rate, RR = respiratory rate, SBP = systolic blood pressure, SpO2 = peripheral capillary oxygen saturation, MTS = Manchester triage system, PAC = patient acuity category, NHAMCS = national hospital ambulatory medical care survey, NLP = natural language processing, ANN = artificial neural network, RF = random forest, GBDT = gradient boosting decision tree, DNN = deep neural network, LR = logistic regression, MLNN = multilayer neural network, SC = single center, MC = multicenter, DM = diabetes mellitus, HTN = hypertension.

Table 1. Summary of clinical outcome prediction model (continued)

Author (year)	Country (setting)	Data source (sample size)	Outcome rates	Predictors	Algorithm	AUROC (95% CI)
Critical care						
Chen (2023)	Taiwan (SC)	EHR data (171,275)	2.5%	age, gender, arrival mode, vital signs, clinical narratives (chief complaints, present illness, medical history)	Clinical narrative-aware deep neural network	0.874 (0.873–0.882)
Joseph (2020)	USA (SC)	ED data (445,925)	13.7%	age, sex, vital signs (HR, RR, BP, SpO2, temp), chief complaint (text data)	LR	0.805 (0.801–0.808)
					2-layer NN	0.812 (0.811-0.814)
					XGBoost	0.820 (0.818-0.821)
					NN (+text data)	0.857 (0.856-0.858)
Kwon (2018)	Korea (MC)	NEDIS data (10,967,518) EMR data (13,989)	5.7%	age, sex, chief complaint, symptom onset time, arrival mode, trauma, vital signs, mental status	deep learning	0.894 (0.894–0.895)
Levin (2018)	USA (MC)	ED data (172,726)	1.6–2.0%	age, sex, arrival mode, vital signs (BT, PR, RR, SBP, SpO2), chief complaint, medical history	RF	0.900–0.920
Raita (2019)	USA (MC)	NHAMCS (135,470)	2.1%	age, sex, mode of arrival, vital signs, chief complaints, comorbidities	LR, Lasso Regression, RF, GBDT, DNN	DNN 0.860 (0.850-0.870)
Yu (2020)	Korea (SC)	CDW (86,309)	3.5%	age, gender, vital signs, chief complaint, level of consciousness, arrival mode	deep learning	0.876 (0.863–0.889)
Yun (2021)	Korea (SC)	NEDIS data (80,433)	4.6%	age, gender, mode of ED arrival, time interval between onset and ED arrival, reason for ED visit, chief complaints, vital signs, consciousness	XGBoost	0.861 (0.848–0.874)
					DNN	0.833 (0.819–0.848)

Abbreviations. EHR = electronic health record, EMR = electronic medical record, NEDIS = national emergency department information system, CDW = clinical data warehouse, BT = body temperature, PR = pulse rate, RR = respiratory rate, SBP = systolic blood pressure, SpO2 = peripheral capillary oxygen saturation, HR = heart rate, BP = blood pressure, NN = neural network, RF = random forest, GBDT = gradient boosting decision tree, DNN = deep neural network, LR = logistic regression, SC = single center, MC = multicenter.

Table 1. Summary of clinical outcome prediction model (continued)

Author (year)	Country (setting)	Data source (sample size)	Outcome rates	Predictors	Algorithm	AUROC (95% CI)
Mortality						
Fernandes (2020)	Portugal (SC)	EDIS data (235,826)	0.48%	age, GCS, pulse oximetry, arrival mode, triage vitals, chief complaints (TF-IDF vectorized)	LR	0.950 (0.940–0.960)
					RF	0.940 (0.930–0.950)
					XGBoost	0.960 (0.950–0.970)
Klug (2020)	Israel (SC)	EMR data (799,522)	① early; 0.60% ② short- term; 2.50%	age, arrival mode, structured chief complaint, vital signs (BT, SpO ₂ , HR, SBP, DBP), ESI level	XGBoost	① 0.962 (0.956–0.968) ② 0.923 (0.919–0.926)
Kwon (2018)	Korea (MC)	NEDIS data (10,967,518) EMR data (13,989)	1.40%	age, sex, chief complaint, symptom onset time, arrival mode, trauma, vital signs, mental status	deep learning	0.935 (0.935–0.936)
Lee (2020a)	Korea (MC)	NEDIS data (81,520)	① 24 hours ; 3.04% ② 48 hours ; 4.08% ③ 7 days ; 5.82% ④ 30 days ; 7.99%	age, sex, SBP, HR, RR, BT, SpO ₂ , level of consciousness, oxygen supply	multi- variable LR	① 0.910 (0.907–0.914) ② 0.899 (0.895–0.903) ③ 0.876 (0.872–0.880) ④ 0.832 (0.828–0.837)
Teubner (2015)	Australia (MC)	EDIS datasets (derivation; 424,316, internal validation; 179,082, external validation; 34,434)	① derivation ; 3.04% ② internal ; 2.27% ③ external ; 1.64%	age, gender, calendar year, ambulance arrival, ATS category, triage complaint codes	LR	① 0.859 (0.856–0.865) ② 0.848 (0.840–0.856) ③ 0.837 (0.823–0.851)
Xie (2021)	Singapore (SC)	EMR data (280,833)	① 2-days ; 0.8% ② 7-days ; 2.2% ③ 30-days ; 5.9%	age, heart rate, respiratory rate, systolic and diastolic blood pressure, comorbidities (Charlson Comorbidity Index)	AutoScore- based machine learning + logistic regression	① 0.821 (0.796–0.847) ② 0.826 (0.811–0.841) ③ 0.823 (0.814–0.832)

Abbreviations. EDIS = emergency department information system, EMR = electronic medical record, NEDIS = national emergency department information system, BT = body temperature, HR = heart rate, RR = respiratory rate, SBP = systolic blood pressure, DBP = diastolic blood pressure, SpO₂ = peripheral capillary oxygen saturation, GCS = Glasgow coma scale, TF-IDF = term frequency-inverse document frequency, ESI = emergency severity index, LR = logistic regression, RF = random forest, XGBoost = extreme gradient boosting, MC = multicenter, SC = single center, ATS = Australasian triage scale.

Table 1. Summary of clinical outcome prediction model (continued)

Author (year)	Country (setting)	Data source (sample size)	Outcome rates	Predictors	Algorithm	AUROC (95% CI)
Hospital admission						
Lee (2020b)	USA (SC)	EHR data (172,809)	admission to	age, gender, prior ED	multinomial	① 0.970
			different units	visits, arrival time,	LR	② 0.950
			① ICU (3.6%)	severity score, chief	NN	③ 0.890
			② TU (2.2%)	complaints, vital signs,	SVM	④ 0.840
			③ GPU (11.9%)	lab test results,		
④ OU (6.6%)	imaging orders					
Discharge						
Riordan (2017)	USA (SC)	EHR data (49,755)	Discharge vs. non-discharge	age, sex, arrival mode, systolic BP, diastolic BP, pulse, respiratory rate, SpO2, temperature, pain score	LR	0.730 (0.720–0.740)

Abbreviations. EHR = electronic health record, BP = blood pressure, SpO₂ = peripheral capillary oxygen saturation, ICU = intensive care unit, TU = telemetry unit, GPU = general practice unit, OU = observation unit, LR = logistic regression, NN = neural network, SVM = support vector machine, SC = single center.

3. 간호 데이터를 활용한 연구 동향

간호 데이터는 환자의 행동 및 생리적 징후를 바탕으로 한 간호사의 임상적 판단이 포함된 데이터로, 환자의 악화를 예측하는 데 중요한 역할을 할 수 있는 요소로 주목받으면서 최근 많은 연구에서 그 의미와 가치가 강조되고 있다(Kim et al., 2023; Rossetti et al., 2024). 특히, 간호 데이터는 생리적 데이터 외에도 환자의 상태에 대한 임상적 판단과 간호사의 직관을 반영한다는 점에서 임상 의사결정 지원 도구로서의 잠재력을 보여준다.

Collins와 Vawdrey(2012)는 간호사들이 환자 상태 악화에 대한 우려가 있을 때 활력징후를 더 자주 기록하고, 비정상적인 측정값과 관련된 코멘트를 추가로 기록하는 경향이 있음을 발견하였다. 이는 간호사의 문서화 행동이 환자의 상태 악화를 조기에 반영할 수 있는 신호임을 보여준다. 이후 Collins et al.(2013)은 간호사의 추가 기록 패턴과 심정지 발생 가능성 간의 유의미한 연관성을 밝히며, 간호 데이터가 환자의 상태 악화를 감지하는 데 중요한 도구로 활용될 수 있음을 확인하였다(Collins et al., 2013).

Collins 연구는 간호 데이터의 중요성을 확인하며 다양한 후속 연구의 기반이 되었다. Kang et al.(2020)은 간호사가 환자의 상태 악화를 예측할 때 사용하는 “우려 (concern)” 개념을 체계적으로 정의하고, 이를 분석 가능하도록 표준화하였다. 이는 간호 데이터의 비정형적 특성을 구조화함으로써 예측 모델에서 간호 데이터를 활용할 수 있는 가능성을 제시하였다. 또한 Rossetti et al.(2020)은 간호사의 코멘트와 문서화 패턴을 활용해 기존 조기경고시스템인 Modified Early Warning System(MEWS)의 성능을 향상시키며, 간호 데이터가 환자 상태 악화를 조기에 감지하는데 중요한 역할을 할 수 있음을 확인하였다. 이후 Rossetti et al.(2021)은 간호사의 행동 데이터를 체계적으로 분석하여 환자의 상태 악화를 감지하는 HPM-

ExpertSignals 프레임워크를 제안하였다. 최근에는 간호사의 서술적 및 구조적 데이터를 통합한 Communicating Narrative Concerns Entered by RNs Early Warning System (CONCERN EWS)을 개발하여 실제 임상에 적용한 결과, 환자의 악화를 기존 시스템보다 최대 42시간 더 조기에 감지하고, 병원 내 사망률을 35.6% 감소시키는 등의 성과를 보여주었다(Rossetti et al., 2024).

간호 데이터를 활용한 예측 모델은 주로 중환자실 환자를 대상으로 한 연구가 많았다. 중환자실 환자를 대상으로 한 체계적 문헌고찰 연구에서는 간호 데이터를 활용한 예측 모델이 생리적 데이터만을 사용한 모델보다 더 높은 예측 성능을 보였음을 보고하였다(Kim et al., 2023). 주로 사용된 간호 데이터는 간호 기록, 평가, 문서 작성 빈도, flowsheet 코멘트 등으로 구분되었다. 이와 같은 간호 데이터는 간호사의 주관적 판단과 환자의 상태를 반영하는 중요한 요소로, 이를 활용한 예측 모델은 환자의 악화를 조기에 감지할 가능성을 증가시킬 수 있다는 점을 강조하였다.

국내에서도 응급실 간호 데이터를 활용한 연구가 일부 시도되었다. 예를 들어, 백승연 등(2019)은 전자 간호 기록 데이터를 분석하여 응급 환자의 중증도가 높을수록 간호 진단과 중재 빈도가 증가하는 경향을 확인하였고, 이효진(2022)은 COVID-19 팬데믹 이후 응급실 임종 환자의 전자간호 기록을 분석하여 임종 간호의 특성과 빈도를 파악함으로써, 응급 간호 데이터의 활용 가능성을 보여주었다. Choi et al.(2019)은 응급실의 구조화된 임상 데이터와 간호사가 작성한 비정형 텍스트 데이터를 사용하여 KTAS 단계를 예측하는 머신러닝 모델을 개발하였다. 랜덤 포레스트와 그래디언트 부스트 알고리즘을 적용한 결과, AUROC 0.922로 높은 성능을 보였으며, 이는 간호사가 작성한 비정형 데이터가 KTAS 단계 예측에서 중요한 역할을 할 수 있음을 시사한다. 이처럼 일부 연구에서 간호 데이터가 예측 모델의 중요한

변수로 작용할 수 있음을 확인하였지만 국내 응급실 간호 데이터를 활용한 예측 모델 개발 연구는 아직 초기 단계에 머물러 있다.

간호데이터는 환자의 상태를 실시간으로 반영하며, 복합적이고 역동적인 상태 변화를 예측하는 데 있어 중요한 역할을 한다. 이러한 데이터는 환자의 상태를 조기에 감지하고, 적시에 중재를 시행할 수 있는 잠재력을 보여준다. 그러나 국내에서는 간호 데이터를 활용한 연구와 예측 모델 개발이 아직 초기 단계에 머물러 있으며, 데이터 표준화 및 활용 체계 구축이 중요한 과제로 남아 있다.

본 문헌고찰에서는 KTAS 3,4단계 환자의 임상적 중요성과 응급실 환자의 임상 결과 예측 연구 동향, 그리고 간호 데이터를 활용한 연구 동향을 종합적으로 살펴보았다. KTAS 3,4단계 환자는 중간 위험군으로 분류되며, 임상적 결과 예측이 복잡하고 자원 배분 및 치료 우선순위 결정에 중요한 그룹임에도 불구하고, 이들을 대상으로 한 연구는 부족한 실정이다. 또한, 입원, 사망, 중환자실 입원 등 주요 결과 변수를 예측하는 다양한 연구가 진행되었으나, 대부분 이분변수를 활용하여 세부적이고 포괄적인 임상결정을 지원하기에는 한계가 있었다. 특히, 간호 데이터는 환자 상태를 실시간으로 반영하며 예측 모델의 성능을 향상시키는 중요한 변수임에도, 국내 연구는 초기 단계에 머물러 있어 활용 가능성이 충분히 탐구되지 않았다. 본 연구는 KTAS 3,4단계 환자를 대상으로 다중범주 결과변수를 적용하고, 전자간호기록을 포함한 다양한 예측변수를 활용하여 입원, 중환자실 입원, 사망, 귀가, 전원과 같은 세부적인 임상결과를 예측함으로써 기존 연구의 한계를 극복하고 응급실 자원 관리 및 환자 안전 향상에 기여하고자 한다.

Ⅲ. 개념적 기틀

본 연구의 개념적 틀은 임상 정보 시스템(Clinical Information System, CIS)에서 임상 전문가들의 전문성과 지식을 반영한 행동 신호를 활용하기 위해, 이들의 행동을 모델링한 Healthcare Process Modeling Framework (HPM-ExpertSignals)을 적용하여 구축되었다. HPM-ExpertSignals 개념적 틀은 Donabedian의 구조-과정-결과 모델을 응용하여 개발되었으며, 특히 임상 전문가의 활동 과정에서 생성된 임상 데이터 구조에서 추출 가능한 정보와 지식 기반의 행동에 의해 주도되는 정보에 중점을 둔다(Rossetti et al., 2021). 이 장에서는 HPM-ExpertSignals 프레임워크에 대해 살펴보고, 이를 바탕으로 본 연구의 개념적 기틀에 대해 논하고자 한다.

1. HPM-ExpertSignals framework

HPM-ExpertSignals는 의료 전문가들의 행동 패턴을 모델링하여, 임상적 지식과 전문성을 예측 모델에 반영하기 위한 개념적 프레임워크이다. 이 프레임워크는 주로 의사와 간호사가 임상 정보 시스템(CIS)과 상호작용할 때 나타나는 행동을 분석하여, 그들의 결정, 지식, 전문성을 대리할 수 있는 신호를 찾아내고 이를 환자 결과 예측 모델에 적용하는 것을 목적으로 한다(Rossetti et al., 2021). 이 프레임워크는 CONCERN (Communicating Narrative Concerns Entered by Registered Nurses) 모델을 기반으로 발전한 것으로, CONCERN 모델은 주로 간호사가 환자의 상태에 대해 우려할 때 나타나는 행동패턴을 분석하여 병원 내 악화 위험을 예측하는데 사용된다. HPM-ExpertSignals는 이러한 간호사의 우려를 반영한 패턴을 더 넓은 범위의 임상적 행동으로 확장하고, 이를 통해 임상적 판단을 예측 모델에 반영하여 보다 정확한

의료 상황 파악을 가능하게 한다(Rossetti et al., 2021).

HPM-ExpertSignals는 Donabedian의 구조-과정-결과 모델을 응용하여 개발되었으며, 다음 세 단계의 모델링 기술을 활용한다(Figure 1). 첫 번째 단계에서는 임상 정보 시스템과 사용자의 상호작용에서 추출된 행동 패턴을 분석하여 의료 전문가들의 행동을 이해한다. 두 번째 단계는 이러한 행동 패턴을 전문가들의 결정, 지식, 전문성의 대리 지표로 정량화하고 체계화한다. 세 번째 단계에서는 이 패턴들을 예측 모델에 적용하여 환자 결과와의 연관성을 분석한다. 이 프레임워크의 궁극적인 목표는 행동 패턴을 통해 나타나는 임상적 전문성을 체계화하여 이를 기반으로 병원 내 환자 경로를 예측하는 것이다. 특히 간호사들의 의사결정 과정에서 나타나는 미묘한 행동적 신호들이 의료 결과 예측에서 중요한 요소로 활용될 수 있음을 강조하고 있다. 예를 들어, 간호사가 언제 데이터를 기록하고, 주기적으로 수정하거나 추가하는지 등의 패턴을 통해 환자 상태를 조기에 파악할 수 있다. 즉, 생리학적 변화가 나타나기 훨씬 전에 활성화되는 전문가의 행동 패턴에서 도출된 신호를 활용함으로써, 환자의 악화 여부를 더 신속하고 정확하게 예측할 수 있도록 돕는다고 설명하고 있다(Rossetti et al., 2021).

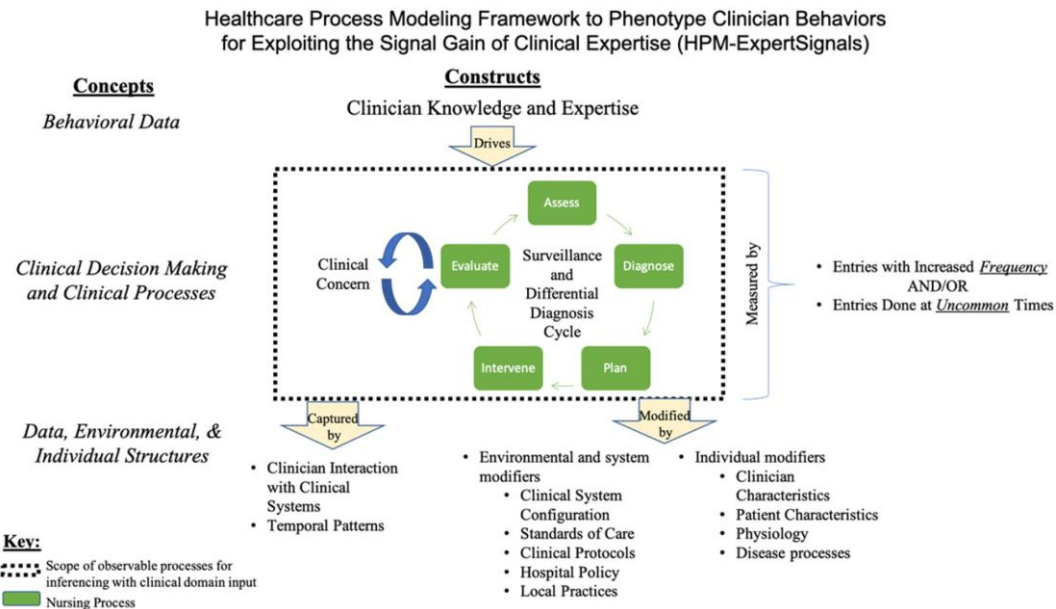


Figure 1. Healthcare Process Modeling Framework to Phenotype Clinician Behaviors for Exploiting the Signal Gain of Clinical Expertise (HPM-ExpertSignals)

2. 연구의 개념적 기틀

본 연구에서는 KTAS 3, 4단계 응급 환자의 임상 결과를 예측하기 위한 예측 모델을 개발하고 평가하기 위해 HPM-ExpertSignals 프레임워크를 수정한 개념적 기틀을 적용하였다. 이 개념적 기틀은 간호사들이 지식과 전문성에 따라 환자에 대한 임상적 우려를 바탕으로 감시(surveillance) 활동을 수행하고, 그와 관련된 정보를 전자 간호기록에 기록한다는 가정을 기반으로 한다. 간호사의 우려와 환경적 및 개별적 조정 요인에 따라 각 환자마다 기록된 정보 패턴이 다르게 나타나며, 이 패턴은 환자의 상태를 반영한다. 이 개념적 기틀은 응급실에서 수집된 triage 정보 및 간호

기록 정보를 활용하여 환자의 임상 결과를 예측하는 구조를 제시하며, 개념적 기틀에 포함된 변수는 문헌고찰을 기반으로 선정하였다(Figure 2).

간호 데이터는 간호 사정과 간호기록 패턴으로 구성된다. 간호 사정은 triage 에서 수집한 초기평가 정보로써 활력징후, 의식 상태, 주증상, KTAS 단계, 통증 평가를 포함한다. 간호기록 패턴은 기록의 빈도와 중재로 구분되는데 빈도 패턴은 환자의 활력징후, 의식상태, 통증 사정과 관련된 간호 데이터의 기록 횟수를 포함한다. 중재 패턴은 간호기록에 나타난 의료진에게 알림, 비정상 검사결과의 기록, 추가 검사의 의뢰에 대한 기록 등을 포함하며, 특히 flowsheet의 코멘트 내용과 같은 간호사가 환자 상태에 대한 우려를 문서화한 내용이 반영된다. 이러한 중재 활동은 환자의 상태 변화와 시간적 연관성을 가지고 있으며, 모델에 중요한 변수로 포함된다.

개별적 조정 요인은 환자의 나이, 성별, 1년 이내 입원 횟수, 1개월 이내 응급실 방문 횟수와 같은 개인적 특성을 의미한다. 환경적 조정 요인은 응급실의 운영 환경과 관련된 요인으로, 환자가 응급실에 도착한 시간대, 내원 경로, 수단, 진료 대기 시간, 증상 발현 후 응급실 도착까지 소요 시간을 포함한다.

본 연구는 전자간호기록에 기록된 간호 데이터를 주요 예측 변수로 활용함으로써, 기존의 예측모델들이 간과했던 간호사 관점에서의 환자 상태 변화를 반영할 수 있는 새로운 프레임워크를 제공하고자 한다. 이를 통해 KTAS 3, 4단계 환자의 임상 결과 예측 모델을 개발하고, 궁극적으로 응급실 체류 시간 단축, 의료 자원 효율화, 환자 결과 향상에 기여할 수 있는 근거 기반의 도구를 제시할 수 있을 것으로 기대한다.

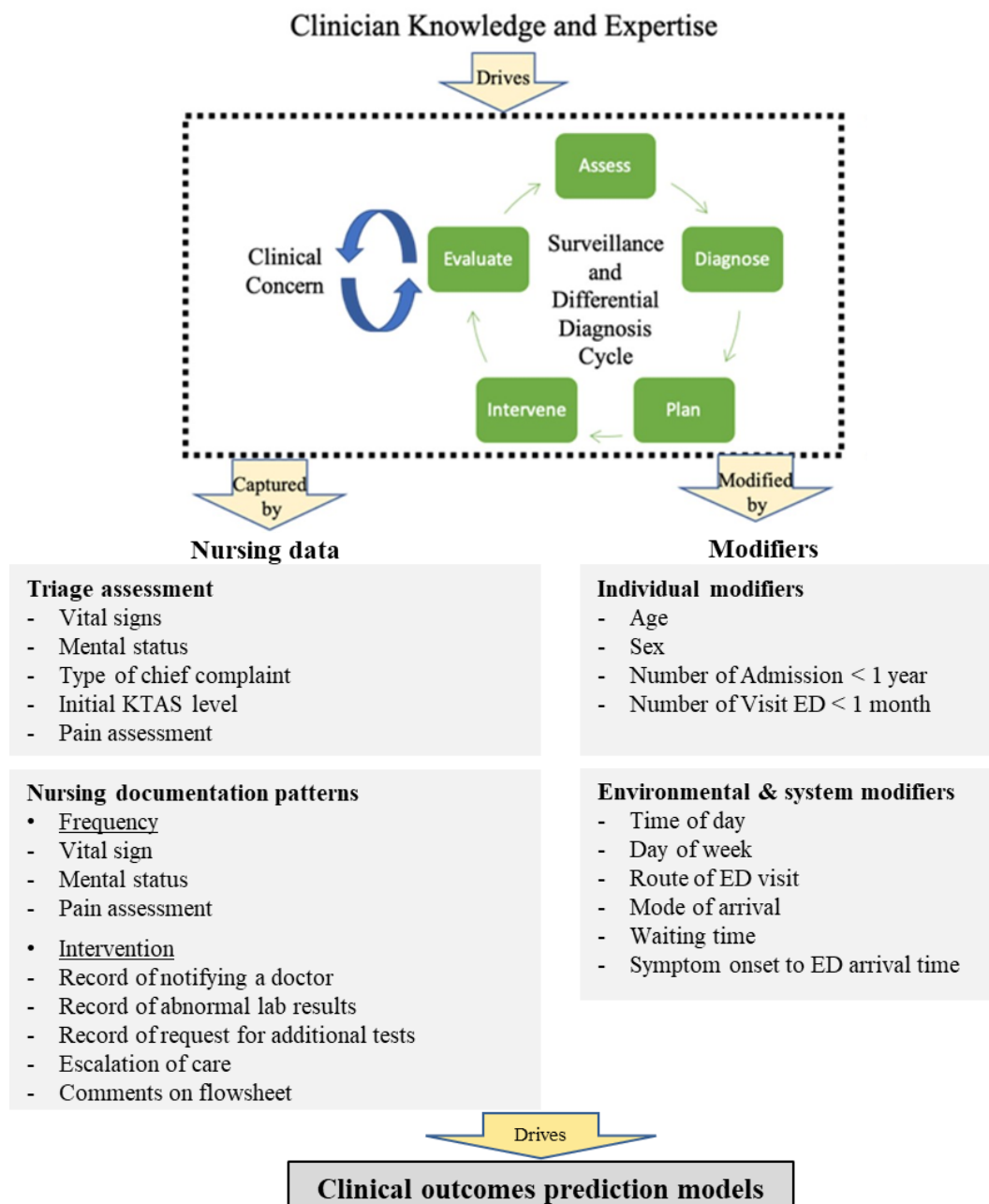


Figure 2. Conceptual framework for this study

IV. 연구 방법

1. 연구 설계

본 연구는 응급실에 내원한 KTAS 3, 4 단계 응급 환자의 특성과 간호기록의 내용 및 패턴을 분석하여 임상 결과와의 연관성을 조사하고, 이를 바탕으로 임상 결과를 예측하는 모델을 개발하고 평가하는 후향적 서술 조사 연구이다. 본 연구는 인공지능 및 머신러닝 기법을 사용한 다변수 예측 모델의 투명하고 일관된 보고를 위한 가이드라인인 "Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis + Artificial Intelligence" (TRIPOD + AI)에 따라 작성되었다(Collins et al., 2024)(Appendix 1).

2. 자료원

본 연구는 2023 년 1 월 1 일부터 12 월 31 일까지 서울시 소재 2,000 병상 규모의 상급종합병원 응급실에 내원한 환자 데이터를 기반으로 진행되었다. 이 병원 응급실은 연간 약 60,000 명 이상의 환자가 방문하며, 모든 데이터는 병원의 의료정보시스템인 Clinical Data Warehouse (CDW) DARWIN-C 에서 추출되었다. 모든 자료는 환자의 개인 식별 정보를 포함하지 않은 익명화된 형태로 수집되었다. 조사 기간은 COVID-19 로 인한 응급실의 비정상적 운영 영향을 고려하여 2023 년 1 월부터 12 월까지 1 년간의 자료를 활용하였다.

3. 연구대상

본 연구의 분석 대상은 2023 년 1 월 1 일부터 12 월 31 일까지 서울시 소재 상급종합병원 응급실에 내원한 환자 중, 다음의 선정 및 배제 기준을 충족한 환자이다(Table 2). 선정 기준은 내원일시 기준 만 18 세 이상의 성인 환자, KTAS 도구로 중증도 분류 결과 3 단계와 4 단계인 환자이며, 배제 기준은 접수취소한 환자, 진료 외 방문 환자, 자의 퇴원 환자, 탈원 환자, 도착 시 사망 환자, 결측 데이터가 있는 환자이다.

동일한 환자가 여러 차례 응급실에 내원한 경우, 각 내원일시를 기준으로 데이터를 포함하였으며, 따라서 분석 대상은 환자 수가 아닌 응급실 방문 건수이다. 연구에 사용된 자료는 응급실 내원 시 작성된 간호 초기평가 기록, 간호 기록, 임상 관찰 기록(flowsheet) 등 전자의무기록(EHR)을 기반으로 수집되었다.

Table 2. Eligibility criteria

Inclusion criteria	Exclusion criteria
<ul style="list-style-type: none"> • Age ≥ 18 • Initial KTAS level 3 or 4 	<ul style="list-style-type: none"> • Canceled registrations • Visit without treatment • Left against medical advice • Left without being seen • Dead on arrival (DOA) • Missing data

4. 분석 변수

KTAS 3, 4 단계 환자의 임상 결과를 예측할 것으로 예상하는 잠재적 예측변수는 선행문헌 고찰을 기반으로 하여 선정하였으며, 개념적 기틀에 따라 triage 정보, 간호기록 패턴을 포함한 간호 데이터와 개별적 조정요인, 환경적 조정요인으로 구분하였다(Table 3). 간호 데이터 중 triage 정보는 간호초기평가를 통해 수집된 정보로 내원 시 활력징후, 의식상태, 주증상, 초기 KTAS 분류 결과, 통증 여부 및 점수 등 총 11 개 변수를 포함하였다. 간호기록의 패턴을 파악하기 위해 임상관찰 기록지(Flow Sheet)에 작성된 활력징후 측정 빈도, 의식상태 측정 빈도, 통증사정 빈도를 포함하였으며, 간호기록의 중재 관련 요인으로는 간호기록에 작성된 의사 보고, 비정상 검사결과에 대한 기록, 추가 검사에 대한 의뢰 기록, 상위 진료 구역으로의 이동, 임상관찰 기록지의 Event 특이사항 메모 등을 포함하여 총 12 개의 변수를 선정하였다. 임상관찰 기록지와 간호기록의 빈도 측정은 내원일시를 기준으로 6 시간 이내에 작성된 기록의 횟수를 측정하였다.

개별적 조정 요인은 성별, 연령, 1 년 이내 입원이력, 1 달 이내 응급실 방문이력으로 총 4 개 변수를 포함하였으며, 환경적 조정요인은 내원 시간대, 내원 요일, 내원 경로, 내원 수단, 진료대기시간, 증상 발현부터 내원시까지 소요시간으로 총 6 개 변수를 포함하였다. 결과변수는 임상 결과이며, “일반병동 입원”, “중환자실 입원”, “전원”, “사망”, “귀가”의 5 가지로 범주로 구분된다. 임상 결과는 응급실 환자의 응급 진료가 완료되어 퇴실하는 시점에 결정되며, 간호사가 작성한 퇴실 정보에 기록된 정보이다.

Table 3. Summary of candidate predictors

Category	Candidate predictors	Type of data	CDW tables
Nursing data			
Triage assessment	Vital signs	Continuous	NIA-ED
	Mental status	Discrete	NIA-ED
	Type of chief complaint	Discrete	NIA-ED
	Initial KTAS level	Continuous	NIA-ED
	Pain	Discrete	Flowsheet
	Pain scale	Continuous	Flowsheet
Nursing documentation patterns	Frequency	Continuous	Flowsheet
	• Vital signs		
	• Mental status		
	• Pain assessment		
	Intervention		
	• Record of notifying a doctor*	Continuous	Nursing records
	• Record of abnormal lab results*		
	• Record of request for additional tests*		
	• Escalation of care	Discrete	ED location history
	• Comments on flowsheet*	Continuous	Flowsheet
Modifiers			
Individual modifiers	Age	Continuous	NIA-ED
	Sex	Discrete	NIA-ED
	Number of Admission < 1 year	Continuous	Admission information
	Number of Visit ED < 1 month	Continuous	Emergency information
Environmental & system modifiers	Time of day	Discrete	NIA-ED
	Day of week	Discrete	NIA-ED
	Route of ED visit	Discrete	NIA-ED
	Mode of arrival	Discrete	NIA-ED
	Waiting time	Continuous	ED location history
	Symptom onset to ED arrival time	Continuous	NIA-ED

Abbreviations. NIA-ED = nursing initial assessment-emergency department, KTAS = Korean triage and acuity scale, PRN = pro re nata, ED = emergency department

* Record and comments in nursing documentation patterns indicate the frequency of entries.

5. 표본 크기 산출

본 연구는 약 47,000 건의 데이터를 기반으로 하고 있어 예측 모델을 개발하는 데 충분한 표본을 제공한다. 변수당 사건 비율(Events Per Predictor, EPP)이 충분히 확보된 경우, 과적합의 위험이 줄어들고 예측 모델의 신뢰성이 높아질 수 있다. 일반적으로 EPP가 10 이상이면 안정적인 모델을 구축할 수 있다고 보고되었다(Riley et al., 2020).

본 연구의 결과 변수는 5 가지 범주로, 가장 낮은 사건 비율은 사망(0.1%)며, 이는 총 60 건의 사건에 해당한다. 본 연구에서 사용된 예측 변수는 총 33 개로, 이를 기반으로 계산하면 EPP 기준에 따라 최소 330 건의 사건이 필요하다. 따라서 사망 사건에서는 이 기준을 충족하지 못하므로 과적합의 가능성이 있다. 하지만 일반 병동 입원 등 발생 빈도가 높은 다른 결과 변수의 경우, 충분한 사건 수와 데이터 양을 확보하고 있어, 추가적인 표본 크기 산출이 반드시 필요하지 않다고 판단된다.

Riely et al.(2020)은 충분한 데이터가 이미 확보된 상황에서 중요한 것은 데이터의 질과 변수의 적합성임을 강조하며, 모델의 과적합을 방지하고 성능을 개선하기 위해 데이터 검증 절차를 통해 내부적으로 평가할 것을 권장하고 있다. 본 연구는 이러한 권고를 반영하여 데이터 질을 보장하고 모델 성능을 개선하는 데 초점을 맞추었다.

6. 자료구축

1) 분석 변수 자료원 조사

문헌고찰을 통해 수집한 분석 변수의 자료원을 파악하기 위하여 CDW에 포함되어 있는 모든 테이블을 조사하고 각 변수 별 수집 가능한 자료원을 모두 확인하였다.

간호 데이터는 간호초기평가-응급, Flowsheet 검색항목, 환자간호기록, 응급환자 위치변경이력 테이블에서 확인하였으며, 환경/개별적 조정요인은 간호초기평가-응급, 응급환자 위치변경이력, 입원정보, 응급정보 테이블을 통해 조사하였다.

2) 분석 변수별 자료 정의서 작성

각 변수별 수집 자료원을 조사한 결과를 종합하여, 변수 별 수집 자료원과 추출할 자료 항목 및 코드를 정리하였다(Table 4, Appendix 2). 간호 데이터의 triage 정보 중 주증상의 종류 변수는 질병과 외상으로 구분하였다. 간호기록패턴의 빈도 자료는 CDW 내 Flowsheet 검색항목의 하위 그룹에서 내원 일시 기준 6시간 이내 해당 항목의 빈도 분석을 통해 확인하였고, 중재에 해당하는 간호기록은 환자간호기록 테이블의 간호기록내용 검색을 통해 확인하였다. 치료 수준의 상향은 위치변경 이력 정보를 통해 경증 환자 진료 구역에서 중증 환자 진료 구역으로 이동한 환자군을 확인하였다.

환경적 조정요인 중 내원 시간대는 의료진의 근무시간에 따라 3가지 시간대(0시, 8시, 16시 기준)로 구분하였고, 내원일의 주중, 주말 여부에 따른 변수도 정의하였다. 진료 대기시간은 응급환자 위치변경이력을 확인하여 내원일시로부터 진료구역으로 위치변경이 실시된 일시까지의 소요시간을 계산하였고, 증상발현 후 내원시까지 소요시간은 발병일시와 내원일시 사이의 소요시간을 계산하여 정의하였다. 개별적 조정요인 중 1년 이내 입원 횟수는 응급실 내원일시를 기준으로 입원정보의 입원일자와 기간 조건(-365일부터 0일)을 설정하여 추출하였고, 1달 이내 응급실 내원 횟수도 마찬가지로 응급실 내원일시를 기준으로 응급정보의 내원일시와 기간 조건(-31일부터 0일)을 설정하여 추출하였다.

Table 4. Summary of selected features

Category	Predictors (range)	Possible value
Nursing data		
Triage assessment (11)	Vital signs <ul style="list-style-type: none"> • Systolic blood pressure • Diastolic blood pressure • Pulse rate • Respiratory rate • Body temperature • Oxygen saturation Mental status Type of chief complaint Initial KTAS level (1-5) Pain Pain scale (0-10)	Maximum and median values AVPU Disease/Trauma 3/4 Yes/No Maximum value
Nursing documentation patterns (12)	Frequency <ul style="list-style-type: none"> • Single vital sign: systolic blood pressure, pulse rate or heart rate, respiratory rate, body temperature, oxygen saturation • Mental status (AVPU or GCS) • Pain assessment Intervention <ul style="list-style-type: none"> • Record of notifying a doctor • Record of abnormal lab results • Record of request for additional tests • Escalation of care • Comments on flowsheet 	Frequency/6hours Frequency/6hours Zone classification Frequency/6hours
Modifiers		
Individual modifiers (4)	Age, years Sex Number of Admission < 1 year Number of Visit ED < 1 month	Maximum value Female/Male Maximum value Maximum value
Environmental & system modifiers (6)	Time of day Day of week Route of ED visit Mode of arrival Waiting time Symptom onset to ED arrival time	Day/Evening/Night Weekdays/Weekends Direct/Transfer/From OPD Ambulance/Others Maximum and median values Maximum and median values

Abbreviations. KTAS = Korean triage and acuity scale; AVPU = alert, verbal, pain, unresponsive scale, GCS = Glasgow coma scale, ED = emergency department, OPD = outpatient department

3) 원시 자료(raw data) 추출

최종 작성된 자료 정의서는 전문가 2인 이상에게 검토를 받은 후, CDW 전문가의 자문을 받아 연구자가 CDW 에서 변수 자료 정의에 따라 필요한 원시자료를 직접 추출하였다.

4) 자료 전처리(data preprocessing)

추출된 자료의 신뢰도를 높이기 위하여 자료를 통합 정제하는 과정을 통해 자료 전처리를 수행하였다. 본 연구에서는 정형 데이터의 경우 이상치 검출, 결측 데이터 처리, 데이터 균형을 통해 데이터 전처리 단계를 수행하였다.

(1) 이상치 검출

이상치는 데이터세트 내에서 다른 데이터와 일치하지 않는 비정상적인 값으로 정의된다(Barnett & Lewis, 1984). 이상치는 모델 성능을 저하시키고 오류 분산을 증가시킬 수 있기 때문에, 통계 분석을 진행하기 전에 이상치를 탐지하는 것이 필수적이다. 이상치는 주로 장비 오류, 인간의 실수, 또는 환자 개체 간 자연적 변이로 인해 발생할 수 있다(Salgado et al., 2016b).

본 연구에서는 이상치를 찾아내기 위해 활력 징후의 경우 과학적으로 유효한 양의 값을 사용하였고, 이상치 식별에는 사분위 범위(IQR; 25-75 percentile)를 적용하였다(Steyerberg, 2009). 이에 따라 활력 징후의 상위 및 하위 사분위수(Q1, Q3)를 기준으로 극단적인 값들이 처리되었다. 이상치로 발견된 데이터는 해당 범주의 중위값으로 대체하였다.

(2) 결측값 처리

대부분의 머신러닝 알고리즘은 완전한 데이터를 요구하기 때문에 결측값을 처리하는 과정이 필수적이다(Salgado et al., 2016a). 변수 간의 상호작용을 고려하여 결측값을 결정해야 하며(Little & Rubin, 2019), 이는 데이터의 정확성을 높이는 중요한 단계이다.

본 연구에서는 triage 에서 측정된 활력징후의 결측값을 내원 시간과 가장 가까운 flowsheet 의 활력징후 값으로 대체하였다. triage 활력징후의 결측은 주로 긴급하거나 혼잡한 상황 또는 활력징후 측정 기기의 오류 등으로 인해 환자의 활력징후를 즉시 측정하기 어려운 경우 발생한다. 이러한 상황에서는 전산 시스템에는 ‘활력징후 측정 불가’로 기록된 후 진료가 시작되며, 이후 flowsheet 에 활력징후 값이 추가 입력되기도 한다. 따라서 flowsheet 에 기록된 초기 활력징후 값으로 대체하는 것이 초기평가에서 환자의 상태를 적절히 반영할 수 있다고 판단하였다. 총 235 건의 결측값 중에서 173 건이 flowsheet 에 기록된 초기값으로 대체 가능하였고, 나머지 62 건은 초기평가 및 flowsheet 모두에서 활력징후 값이 기록되지 않아 결측 데이터로 간주하고 본 연구의 분석 대상에서 제외하였다.

또한 통증 여부 측정불가, 수면 중, 결측값은 “없음”으로, 통증 점수 결측값은 "0"으로 대체하여 처리하였다. 통증에 대한 결측값은 대부분 소수 클래스인 사망 환자에서 발생했기 때문에 데이터를 제외하는 경우 데이터 불균형이 심화될 수 있어 해당 결측값을 제외하지 않고, 대체값으로 처리하였다. 진료대기시간과 증상 발생 후 내원 시까지 소요시간에 대한 결측값은 각 데이터가 속한 결과변수 범주의 중위값으로 대체하였다.

(3) 데이터 불균형 처리

데이터의 불균형 분포는 소수 클래스에 대한 학습 부족과 다수 클래스에 대한 편향을 초래하여 머신러닝 모델의 학습과 예측 성능을 저하시킬 수 있다(López et al., 2013; Fernández et al., 2018). 이러한 문제는 특히 의료 데이터를 활용한 연구에서 두드러지며, 중환자실 입원, 전원, 사망과 같은 소수 클래스와 귀가와 같은 다수 클래스 간의 불균형이 모델의 성능과 실질적 활용 가능성에 부정적인 영향을 미칠 수 있다. 이를 해결하기 위해 데이터 수준에서의 오버샘플링 및 언더샘플링 기법, 알고리즘 수준 전략, 비용 민감 학습(cost-sensitive learning), 그리고 앙상블 기반 전략의 네 가지 주요 접근이 제안되고 있다(Fernández et al., 2018).

본 연구는 데이터 불균형 문제를 해결하기 위해 One-vs-Rest SMOTE(synthetic minority over-sampling technique)와 비용 민감 학습을 결합하는 방식을 사용하였다. One-vs-Rest SMOTE는 소수 클래스 데이터를 증강하여 데이터 불균형을 완화하며, 각 소수 클래스를 독립적으로 처리함으로써 클래스 간 데이터 비율을 균형 있게 맞춘다(Imbalanced-learn Development Team, n.d.). 이후 증강된 데이터에 비용 민감 학습을 적용하여 소수 클래스의 오분류 비용을 증가시켜 해당 클래스의 중요도를 높였다. 이러한 접근은 모델이 소수 클래스에서 민감도를 향상시키고 다수 클래스에 치우치지 않도록 균형 잡힌 예측 성능을 확보하는데 기여하였다.

7. 예측 모델 개발 및 타당도 평가

triage 데이터는 결측치와 이상치가 적고 추출이 용이하다는 장점이 있어, 기존 연구에서는 주로 이를 기반으로 예후 및 입원 가능성을 예측하는 모델이 개발되어 왔다. 그러나 간호 데이터를 활용한 예측 모델 개발 연구는 상대적으로 드물다. 본 연구는 개념적 기틀에서 제시한 바와 같이, 간호기록의 패턴 변수가 예측 모델 성능에 미치는 영향을 평가하기 위해 간호기록 패턴 변수를 포함하지 않은 모델(model 1)과 포함한 모델(model 2)을 각각 개발하고 두 모델 간 성능 차이를 비교하였다.

분류 모델링에서 과적합을 방지하고 모델 성능을 검증하기 위해 데이터셋은 학습 세트(80%)와 테스트 세트(20%)로 무작위 분할되었다. 학습 세트는 One-vs-Rest SMOTE와 비용 민감 학습을 통해 데이터 불균형 문제를 해결하였다.

본 연구에서는 다항 로지스틱 회귀(Multinomial logistic regression), 랜덤 포레스트(Random forest), 그래디언트 부스트(Extreme Gradient Boosting, XGBoost)를 사용하여 예측 모델을 개발하였다. 다항 로지스틱 회귀는 여러 범주형 결과 변수를 예측하는 데 적합한 기법으로, 결과 변수가 3 개 이상의 범주로 나뉘는 경우 각 범주의 발생 확률을 추정할 수 있어, 본 연구의 다양한 임상 결과를 모델링하는 데 적합하다(Hosmer, Lemeshow, & Sturdivant, 2013). 랜덤 포레스트는 여러 결정 트리의 앙상블을 통해 과적합을 방지하면서도 높은 예측 정확도를 제공하는 기법으로, 특히 결측값 처리와 변수 간의 복잡한 상호작용을 잘 반영할 수 있어 임상 데이터 분석에 자주 사용된다(Breiman, 2001). 그래디언트 부스트는 이전 단계에서 발생한 오류를 보정하며 점진적으로 성능을 개선하는 기법으로, 대규모 데이터에서도 매우 높은 예측 성능을 보이며 복잡한 임상 데이터를 분석하는 데 강력한 도구로 사용된다(Friedman, 2001).

비용 민감 학습은 각 알고리즘에 추가적으로 적용되어, 클래스별 오류 비용을 조정함으로써 모델의 민감도와 특이도 간 균형을 개선하였다. 학습 데이터를 활용하여 알고리즘을 훈련하기 위해 계층적 10-겹 교차 검증(stratified 10-fold cross-validation)이 적용되었다. 학습 세트는 무작위로 10 개의 동일 비율 부분(학습 세트 1~10)으로 나누어졌으며, 학습 세트와 검증 세트의 비율은 9:1로 설정되었다. 학습 과정은 각기 다른 학습 세트에서 10 회 수행되었으며, 이 절차는 10 회 반복되었다.

개발된 예측 모델의 성능은 테스트 세트를 이용한 내부 검증을 통해 평가되었다(Figure 3). 모델 성능을 평가하여 최적 모델을 선정한 후 변수 중요도를 확인하였다.

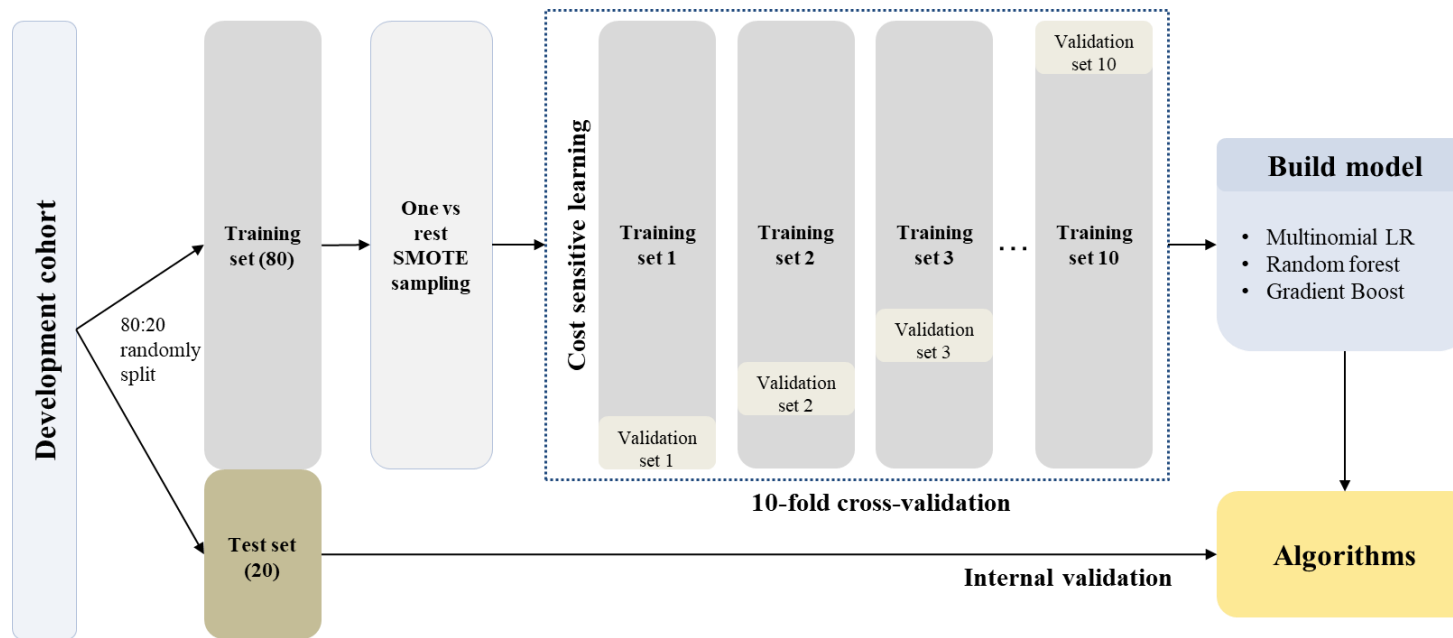


Figure 3. Model development and validation process

8. 자료분석

데이터 전처리, 모델 개발 및 타당도 평가는 Python version 3.10.12 프로그램을 이용하여 수행하였으며, 주요 분석에 사용된 라이브러리는 pandas(1.5.3), numpy(1.24.4), scipy (1.12.0), scikit_posthocs (0.11.2), matplotlib(3.8.4), scikit-learn(1.2.0), imbalanced-learn(0.12.4), xgboost(2.0.3), shap(0.46.0) 등이 포함되었다. 각 단계별 데이터 분석 방법은 다음과 같다.

첫째, 개발 및 검증 코호트의 기술 통계는 연속형 변수에 대해 평균 및 표준편차(SD), 명목형 변수에 대해 빈도 및 백분율로 제시하였다. 코호트 간 동질성을 평가하기 위해 연속형 변수는 독립표본 t-검정을, 명목형 변수는 카이제곱 검정을 사용하였다.

둘째, 예측 변수와 결과 변수 간의 관계 분석을 위해 명목형 변수는 카이제곱 검정을 사용하여 결과 변수 간의 분포 차이를 확인하였다. 연속형 변수는 정규성 검정 결과 정규성을 만족하지 않아 크루스칼-왈리스 검정(Kruskal-Wallis Test)을 통해 결과변수의 그룹 간 차이를 분석하였다. 그룹 간 유의미한 차이가 있는 경우 사후분석으로 Dunn's test 를 Bonferroni 보정을 적용하여 추가 검토하였다.

셋째, 모델 개발 과정에서는 여러 하이퍼파라미터 조합을 체계적으로 탐색하여 모델의 성능을 최적화하는 기법으로, 교차 검증을 통해 각 조합의 성능을 평가하는 GridSearchCV 를 사용하여 하이퍼파라미터 최적화를 수행하였다. 각 모델에 대해 최적화된 파라미터는 다음과 같다. 그래디언트 부스트(XGBClassifier)의 경우 learning_rate= 0.1, max_depth= 5, n_estimators= 200 으로 설정되었다. 다항 로지스틱 회귀(LogisticRegression)는 규제 계수 C= 10 으로, 패널티 방식은 'l2'로 설정하였다. 랜덤 포레스트(RandomForestClassifier)는 max_depth 를 제한하지 않았으며, min_samples_split= 2, n_estimators 는= 200 으로 설정하였다.

넷째, 모델 성능 평가는 AUROC, AUPRC (area under the precision-recall curve), 정확도(accuracy), 혼동행렬(confusion matrix), 민감도(sensitivity)/재현율(recall), 특이도 (specificity), 양성 예측 값(positive predictive value, PPV)/정밀도(precision), 음성 예측 값(negative predictive value, NPV), F1 점수를 사용하여 이루어졌다.

다섯째, 예측 변수의 중요도는 최적 모델이 선정 이후 랜덤 포레스트 모델 내의 Gini importance 기법을 사용하여 초기 분석을 수행하였다. 추가로 셰플리 가산 설명 기법(Shapley Additive Explanations, SHAP) 분석을 수행하여 예측 변수의 중요도를 심층적으로 평가하였다. SHAP 분석은 결과 변수의 각 클래스별로 수행되었으며, 각 클래스에 대한 주요 변수와 이들의 상대적 기여도를 파악하였다. 특히, SHAP summary plot 을 활용하여 각 클래스에서 변수들의 상대적 중요도를 시각적으로 제시함으로써 모델의 해석 가능성을 효과적으로 높였다.

여섯째, 예측 모델의 신뢰성은 적합도 보정 곡선(calibration plot)을 통해 평가하였다. 보정 곡선은 모델의 예측 확률과 실제 결과 간의 일치도를 시각화 하여 예측된 확률의 신뢰성을 평가하는 도구이다. 이를 통해 모델이 예측에서 과대평가하거나 과소평가하는 부분을 파악할 수 있다(Van Hoorde et al., 2015).

9. 윤리적 고려

본 연구는 삼성서울병원 기관생명윤리위원회(IRB)의 심의 면제 승인을 받아 진행되었다(IRB No. SMC 2024-07-063) (Appendix 3). 자료 수집은 연구 대상자 중에서 선정 기준을 만족하는 응급실 내원 환자의 개인식별정보를 제외한 익명화된 데이터 형태로 수집되었으며, 전자간호기록 및 응급간호초기평가 자료는 임상 데이터 웨어하우스(CDW)에서 추출하여 사용하였다.

데이터의 기밀성을 유지하기 위해, 수집된 자료는 비밀번호로 보호된 파일에 저장되었으며, 연구자 이외의 접근은 철저히 제한되었다. 또한 데이터의 보안 유지를 위하여, 연구는 연구 대상 기관의 원내망에서만 이루어졌으며, 외부로의 데이터 유출이 원천적으로 차단되었다. 수집된 자료 및 분석 파일은 연구자 이외의 타인이 열람할 수 없도록 철저히 관리되며, 연구 종료 후 3년간 보관한 뒤 모든 관련 파일을 영구히 삭제할 예정이다. 이러한 절차를 통해 연구 자료의 기밀성과 데이터 보안이 유지될 수 있도록 하였다.

V. 연구 결과

1. 코호트 선정 결과 및 특성

본 연구의 코호트 선정 과정은 Figure 4 와 같다. 2023 년 1 월 1 일부터 12 월 31 일까지 총 50,251 명의 환자가 응급실에 내원하여 총 71,000 건의 응급실 내원 기록이 도출되었다. 이 중 선정 및 배제 기준에 따라 24,100 건(33.9%)의 데이터는 연구에서 제외되었고, 총 33,885 명의 응급 환자에 대한 46,900 건(66.1%)의 내원환자 데이터가 최종 연구에 포함되었다. 선정 기준에 따라 18 세 미만의 환자($n=10,395$), KTAS 레벨 1, 2, 5 환자($n=5,373$)가 제외되었고, 배제 기준에 따라 접수 취소($n=6,132$), 진료 외 방문($n=1,706$), 자의 퇴원($n=408$), 탈원($n=24$), 결측값이 있는 자료($n=62$)가 제외되었다. 본 연구에서 결측 데이터로 인해 제외된 데이터 비율은 0.1% 미만으로 나타났다.

선정 기준을 충족한 데이터를 무작위로 할당하여 80%는 훈련 세트($n=37,520$), 20%는 테스트 세트($n=9,380$)로 분할하였다. 이때 결과 변수가 유사한 분포를 유지할 수 있도록 데이터 분포를 조정하였다. 훈련 세트는 예측 모델 개발에 사용되었으며, 테스트 세트는 모델의 내부 검증을 위해 활용되었다. 훈련 세트와 테스트 세트에 포함된 환자 특성은 Table 5 에 제시된 바와 같이 모든 변수에서 통계적으로 유의미한 차이가 없어 두 코호트 간 동질성이 확인되었다.

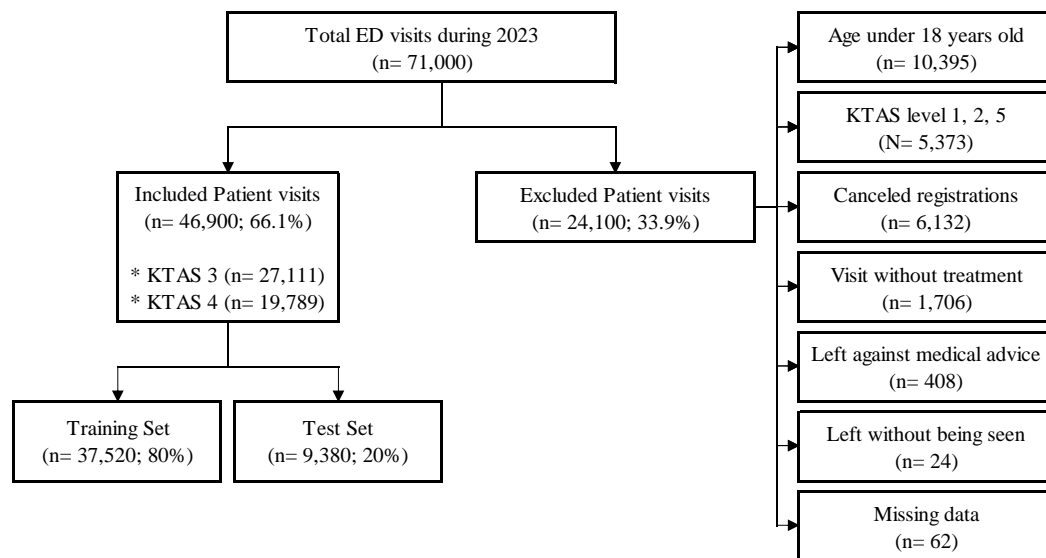


Figure 4. Flowchart illustrating the cohort selection process for the study

Table 5. Comparison of patient characteristics between training and test sets (n= 46,900)

Variables	Training set (80%) (n= 37,520)	Test set (20%) (n= 9,380)	t/ χ^2	p
Individual modifiers				
Age	58.69 \pm 17.46	58.74 \pm 17.49	-0.214	0.830
Sex			0.641	0.423
Female	18,912 (50.4)	4,727 (50.4)		
Male	18,608 (49.6)	4,653 (49.6)		
Number of admission < 1 year	1.09 \pm 2.12	1.08 \pm 2.11	0.246	0.805
Number of visit ED < 1 month	0.25 \pm 0.71	0.25 \pm 0.66	0.509	0.611
Environmental & system modifiers				
Time of day			3.214	0.523
08:00~16:00	18,776 (50.0)	4,703 (50.1)		
16:00~24:00	12,454 (33.2)	3,103 (33.1)		
00:00~08:00	6,290 (16.8)	1,574 (16.8)		
Day of week			0.251	0.616
Weekday	27,488 (73.3)	6,851 (73.0)		
Weekend	10,032 (26.7)	2,529 (27.0)		
Route of ED visit			1.899	0.754
Direct visit	26,954 (71.8)	6,703 (71.5)		
Refer from OPD	3,568 (9.5)	907 (9.7)		
Transfer from other hospital	6,998 (18.7)	1,770 (18.8)		
Mode of arrival			0.558	0.455
Ambulance	6,079 (16.2)	1,542 (16.4)		
Others	31,441 (83.8)	7,838 (83.6)		
Waiting time	55.73 \pm 85.05	56.26 \pm 85.20	-0.541	0.588
Symptom onset to ED arrival time (hour)	50.78 \pm 137.2	49.45 \pm 124.08	0.859	0.391
Outcomes			8.738	0.924
ICU admission	635 (1.7)	159 (1.7)		
GW admission	10,077 (26.9)	2,519 (26.9)		
Transfer	1,416 (3.8)	354 (3.8)		
Death	48 (0.1)	12 (0.1)		
Discharge	25,344 (67.5)	6,336 (67.5)		

Abbreviations. ICU = intensive care unit, GW = general ward, ED = emergency department, OPD = outpatient department

Table 5. Comparison of patient characteristics between training and test sets (n= 46,900) (continue)

Variables	Training set (80%) (n= 37,520)	Test set (20%) (n= 9,380)	t/ χ^2	p
Triage assessment				
Vital signs				
SBP (mmHg)	134.33 \pm 25.40	134.13 \pm 25.25	0.677	0.499
DBP (mmHg)	81.63 \pm 16.15	81.46 \pm 15.96	0.904	0.366
PR (bpm)	92.02 \pm 19.26	92.01 \pm 19.44	0.030	0.976
RR (per min)	18.38 \pm 2.28	18.38 \pm 2.22	-0.063	0.950
BT ($^{\circ}$ C)	36.88 \pm 0.80	36.87 \pm 0.81	1.084	0.278
SpO2 (%)	97.45 \pm 2.25	97.45 \pm 2.24	0.175	0.861
Mental statue			7.607	0.574
Alert	37,139 (99.0)	9,301 (99.1)		
Verbal response	269 (0.7)	63 (0.7)		
Pain response	88 (0.2)	14 (0.1)		
Unconscious	24 (0.1)	2 (<0.1)		
Type of chief complaint			1.080	0.299
Disease	33,437 (89.1)	8,314 (88.6)		
Trauma	4,083 (10.9)	1,066 (11.4)		
Initial KTAS level			0.007	0.935
3	21,693 (57.8)	5,418 (57.8)		
4	15,827 (42.2)	3,962 (42.2)		
Pain location			0.322	0.570
Yes	10,657 (28.4)	2,629 (28.0)		
No	26,863 (71.6)	6,751 (72.0)		
Pain scale	1.40 \pm 2.42	1.40 \pm 2.44	-0.093	0.926
Nursing documentation patterns				
Frequency (within 6hrs)				
SBP/DBP	2.44 \pm 1.77	2.44 \pm 1.82	-0.130	0.897
PR or HR	2.53 \pm 1.85	2.55 \pm 1.98	-0.777	0.437
RR	2.36 \pm 1.59	2.37 \pm 1.68	-0.768	0.442
BT	2.14 \pm 1.13	2.14 \pm 1.15	0.012	0.990
SpO2	4.01 \pm 4.04	4.04 \pm 4.12	-0.664	0.507
Mental status (AVPU or GCS)	1.25 \pm 1.26	1.26 \pm 1.30	-0.328	0.743
Pain scale	0.32 \pm 0.77	0.31 \pm 0.76	1.175	0.240
Intervention (within 6hrs)				
Notifying a doctor	0.16 \pm 0.45	0.17 \pm 0.45	-0.676	0.499
Abnormal lab results	0.32 \pm 0.83	0.32 \pm 0.83	-0.613	0.540

Request for additional tests	0.14 ± 0.47	0.15 ± 0.50	-1.852	0.064
Escalation of care			39.959	0.106
A zone to B zone	6,838 (18.2)	1,732 (18.5)		
A zone to C zone	1,046 (2.8)	267 (2.8)		
A zone to R zone	923 (2.5)	226 (2.4)		
B zone to C zone	12 (<0.1)	4 (<0.1)		
B zone to R zone	11 (<0.1)	3 (<0.1)		
C zone to R zone	9 (<0.1)	3 (<0.1)		
None	28,681 (76.4)	7,145 (76.2)		
Comments on flowsheet	0.31 ± 1.12	0.32 ± 1.15	-0.722	0.470

Abbreviations. SBP = systolic blood pressure, DBP = diastolic blood pressure, PR = pulse rate, RR = respiratory rate, BT = body temperature, SpO₂ = peripheral oxygen saturation, KTAS = Korean triage and acuity scale, A zone = ambulatory zone, B zone = bed zone, C zone = critical care zone, R zone = resuscitation zone

2. KTAS 3, 4 단계 환자의 특성 및 임상결과

총 46,900 건의 응급실 내원 환자 데이터를 분석한 결과, 임상 결과에 따른 환자 분포는 다음과 같이 나타났다(Table 6). 중환자실로 입원한 환자는 794 건 (1.7%)이었으며, 일반 병동 입원 환자는 12,596 건(26.9%)으로 확인되었다. 또한, 다른 병원으로 전원된 환자는 1,770 건(3.8%), 사망 환자는 60 건(0.1%)이었고, 퇴원하여 귀가한 환자는 31,680 건(67.5%)으로 나타났다. 본 연구에서 KTAS 3, 4 단계 응급실 환자의 총 입원율은 28.6%로 확인되었다.

개인적 조정 요인을 분석한 결과는 다음과 같다. 전체 환자의 평균 연령은 58.70 세(±17.47)였으며, 결과 변수에 따라 유의미한 차이를 보였다(H=1055.69, $p<.001$). 사망 환자군(71.18±13.55)과 중환자실 입원 환자군(65.29±14.85)이 다른 군에 비해 나이가 많았으며, 귀가 환자군(56.83±18.01)이 가장 낮았다. 사후 분석 결과 사망 환자군과 중환자실 입원환자군은 일반 병동 입원 환자 및 전원 환자보다 나이가 많았고, 귀가 환자와는 더 큰 차이를 보였다. 성별 분석에서는 남성이 중환자실 입원 환자군(60.2%)과 전원 환자군(56.9%)에서 더 높은 비율을

차지했다($p<.001$). 입원 횟수와 응급실 방문 횟수도 임상 결과와 유의미한 차이를 보였다. 1년 내 평균 입원 횟수는 일반 병동 입원 환자(1.97 ± 2.62)와 중환자실 입원 환자(1.66 ± 1.82)가 다른 그룹보다 유의하게 많았다($p<.001$). 1개월 내 응급실 방문 횟수는 사망 환자군에서 가장 많았다(0.58 ± 0.85 , $p<.001$).

환경적 조정 요인에 대한 분석 결과 일반 병동 입원 환자군의 경우 주중에 내원한 비율과 주간 시간대 내원한 비율이 다른 결과군에 비해 많았다. 응급실 내원 경로와 수단에서도 차이가 나타났는데, 타 병원에서 전원 온 환자가 사망 환자(43.3%)와 중환자실 입원 환자(34.3%)에서 더 높은 비율을 차지했다($p<.001$). 또한, 구급차를 이용하여 도착한 환자도 중환자실 입원 환자(35.9%)와 사망 환자(66.7%)에서 높은 비율을 보였다. 대기 시간에 대한 사후분석 결과 일반 병동 입원 환자(67.44 ± 90.74)와 전원 환자(64.70 ± 93.37)가 귀가 및 중환자실 입원 환자보다 대기 시간이 유의하게 길게 나타났다($p<.001$). 증상 발현에서 응급실 도착까지의 시간은 전체 환자에서 평균 50.52 시간으로 나타났으며, 일반 병동 입원 환자(59.30 ± 133.28)가 다른 결과군에 비해 유의하게 길었다($p<.001$).

응급실 환자 내원 시 triage 간호사의 초기평가 분석 결과는 다음과 같다. 활력징후 중에서 수축기 혈압(SBP)의 평균은 귀가 환자군(136.20 ± 24.88 mmHg)이 가장 높았으며, 사망군(118.65 ± 32.92 mmHg)이 가장 낮았다($H=593.79$, $p<.001$). 사후 분석 결과, 귀가 환자군은 모든 다른 군보다 수축기 혈압이 유의하게 높았다. 이완기 혈압(DBP) 또한 귀가 환자(83.24 ± 15.68 mmHg)가 가장 높았고, 사망 환자(70.53 ± 23.13 mmHg)가 가장 낮았다($p<.001$). 귀가 환자군은 모든 다른 군보다 유의하게 이완기 혈압이 높았다. 맥박수(PR)는 사망군(106.28 ± 25.74 bpm)이 가장 높았으며, 귀가 환자군(90.28 ± 18.38 bpm)이 가장 낮았다($p<.001$). 사후 분석 결과, 사망 환자군은 다른 군보다 유의하게 맥박수가 높았고, 귀가 환자군은 다른 군보다

유의하게 맥박수가 낮았다. 호흡수(RR)는 사망군(21.47 ± 4.36 회/분)이 가장 높았으며, 귀가 환자군(18.18 ± 1.97 회/분)이 가장 낮았다($p < .001$). 중환자실 입원군과 사망군은 다른 군보다 유의하게 호흡수가 높았다. 체온(BT)의 경우, 일반 병동 입원군($37.02 \pm 0.87^\circ\text{C}$)과 전원군($36.99 \pm 0.85^\circ\text{C}$)이 귀가 환자군($36.82 \pm 0.76^\circ\text{C}$)보다 유의하게 높았다($p < .001$). 산소포화도(SpO_2)는 귀가 환자군($97.70 \pm 1.83\%$)이 가장 높았으며, 사망군($95.27 \pm 5.02\%$)이 가장 낮았다($p < .001$). 사후 분석 결과, 귀가 환자군은 모든 다른 군보다 산소포화도가 높았고, 일반 병동 입원군은 중환자실 입원군보다 유의하게 높았다.

환자의 의식 수준 역시 임상 결과에 따라 유의미한 차이를 보였다. 전체 환자의 99.0%는 명료 상태였으나, 사망 환자군에서는 의식이 명료한 상태가 83.3%로 나타나 가장 낮았다($p < .001$). 사망 환자군에서 8.3%는 언어 반응을 보였고, 6.7%는 통증 반응으로 나타나 중증 환자일수록 비정상적인 반응을 보이는 비율이 높았다. 주증상에 대한 분석에서는, 전체 환자의 89.0%가 질병을 주 증상으로 보고하였고, 외상 환자는 11.0%로 나타났다. 사망 환자의 98.3%는 질병을 주 증상으로 보고하였으며, 외상 환자는 1.7%로 상대적으로 적었다($p < .001$).

KTAS 결과는 중환자실 입원 환자의 88.8%가 3 단계로 나타났으며, 사망 환자의 93.3%가 3 단계로 분류되었다($p < .001$). 반면, 귀가 환자의 경우 3 단계와 4 단계가 각각 50%로 차이가 없었다. 통증 관련 분석에서는 전체 환자의 28.3%가 통증을 호소한 것으로 나타났다. 특히, 전원 환자군에서는 통증 호소 비율이 36.2%로 다른 환자군에 비해 높았다. 통증 점수의 경우, 전원군(1.86 ± 2.70)이 가장 높은 값을 보였으며, 귀가 환자군(1.44 ± 2.43)이 상대적으로 낮았다($p < .001$).

Table 6. Characteristics and clinical outcomes of the included cohort (n= 46,900)

Variables	Total set (n (%), mean ± SD)						H/ χ^2	p	post hoc
	Total (n= 46,900)	ICU admission ^a (n= 794)	GW admission ^b (n= 12,596)	Transfer out ^c (n= 1,770)	Death ^d (n= 60)	Discharge ^e (n= 31,680)			
<u>Individual modifiers</u>									
Age	58.70±17.47	65.29±14.85	62.38±15.40	62.68±16.98	71.18±13.55	56.83±18.01	1055.69	<.001	(a,d)>(b,c)>e
Sex							414.03	<.001	
Female	23,639 (50.4)	316 (39.8)	5,541 (44.0)	762 (43.1)	28 (46.7)	16,992 (53.6)			
Male	23,261 (49.6)	478 (60.2)	7,055 (56.0)	1,008 (56.9)	32 (53.3)	14,688 (46.4)			
Number of admission < 1 year	1.08±2.12	1.64±1.82	1.97±2.62	0.93±1.96	1.55±1.98	0.73±1.78	7373.03	<.001	(a,b,d)>c>e
Number of visit ED < 1 month	0.25±0.70	0.19±0.49	0.27±0.59	0.32±0.75	0.58±0.85	0.24±0.74	200.76	<.001	d>(b,c)>(a,e)
<u>Environmental & system modifiers</u>									
Time of day							438.31	<.001	
08:00~16:00	23,479 (50.1)	392 (49.4)	7,081 (56.2)	864 (48.8)	25 (41.7)	15,117 (47.7)			
16:00~24:00	15,557 (33.2)	270 (34.0)	4,083 (32.4)	595 (33.6)	22 (36.7)	10,587 (33.4)			
00:00~08:00	7,864 (16.8)	132 (16.6)	1,432 (11.4)	311 (17.6)	13 (21.7)	5,976 (18.9)			
Day of week							471.03	<.001	
Weekdays	34,339 (73.2)	608 (76.6)	10,098 (80.2)	1,348 (76.2)	47 (78.3)	22,238 (70.2)			
Weekend	12,561 (26.8)	186 (23.4)	2,498 (19.8)	422 (23.8)	13 (21.7)	9,442 (29.8)			
Route of ED visit							6300.03	<.001	
Direct visit	33,657 (71.8)	423 (53.3)	6,215 (49.3)	964 (54.5)	33 (55.0)	26,022 (82.1)			
Refer from OPD	4,475 (9.5)	99 (12.5)	2,888 (22.9)	108 (6.1)	1 (1.7)	1,379 (4.4)			
Transfer from other hospital	8,768 (18.7)	272 (34.3)	3,493 (27.7)	698 (39.4)	26 (43.3)	4,279 (13.5)			
Mode of arrival							1017.20	<.001	
Ambulance	7,621 (16.2)	285 (35.9)	2,419 (19.2)	606 (34.2)	40 (66.7)	4,271 (13.5)			
Others	39,279 (83.8)	509 (64.1)	10,177 (80.8)	1,164 (65.8)	20 (33.3)	27,409 (86.5)			
Waiting time (min)	55.83±85.08	42.22±69.28	67.44±90.74	64.70±93.37	30.34±59.99	51.11±82.11	863.82	<.001	(b,c)>(a,e)
Symptom onset to ED arrival time (hour)	50.52±134.68	45.62±218.06	59.30±133.28	45.91±91.59	25.48±40.27	47.45±134.49	262.14	<.001	b>(a,e) e>c

Table 6. Characteristics and clinical outcomes of the included cohort (n= 46,900) (continue)

Variables	Total set (n (%), mean ± SD)						H/χ2	p	post hoc
	Total (n= 46,900)	ICU admission ^a (n= 794)	GW admission ^b (n= 12,596)	Transfer out ^c (n= 1,770)	Death ^d (n= 60)	Discharge ^e (n= 31,680)			
Triage assessment									
Vital signs									
SBP (mmHg)	134.29±25.37	129.05±31.09	130.33±25.56	131.11±25.70	118.65±32.92	136.20±24.88	593.79	<.001	e>(a,b,c,d)
DBP (mmHg)	81.59±16.11	76.25±18.80	78.33±16.31	78.02±15.95	70.53±23.13	83.24±15.68	1130.87	<.001	e>(a,b,c,d)
PR (bpm)	92.02±19.29	95.90±23.00	95.60±20.45	95.50±20.28	106.28±25.74	90.28±18.38	748.03	<.001	d>(a,b,c)>e
RR (per min)	18.38±2.27	19.58±3.46	18.74±2.67	18.81±2.74	21.47±4.36	18.18±1.97	570.42	<.001	(a,d)>(b,c)>e
BT (°C)	36.88±0.81	36.90±0.92	37.02±0.87	36.99±0.85	36.79±1.09	36.82±0.76	483.78	<.001	(b,c)>e b>a
SpO2 (%)	97.45±2.25	96.42±3.34	96.98±2.83	96.87±2.75	95.27±5.02	97.70±1.83	772.10	<.001	e>(a,b,c,d) b>a
Mental statue									
Alert	46,440 (99.0)	747 (94.1)	12,413 (98.5)	1,702 (96.2)	50 (83.3)	31,528 (99.5)	715.44	<.001	
Verbal response	332 (0.7)	32 (4.0)	134 (1.1)	42 (2.4)	5 (8.3)	119 (0.4)			
Pain response	102 (0.2)	15 (1.9)	40 (0.3)	24 (1.4)	4 (6.7)	19 (0.1)			
Unconscious	26 (0.1)	0 (0)	9 (0.1)	2 (0.1)	1 (1.7)	14 (0)			
Chief complaint									
Disease	41,751 (89.0)	771 (97.1)	12,169 (96.6)	1,498 (84.6)	59 (98.3)	27,254 (86.0)	1125.71	<.001	
Trauma	5,149 (11.0)	23 (2.9)	427 (3.4)	272 (15.4)	1 (1.7)	4,426 (14.0)			
Initial KTAS level									
3	27,111 (57.8)	705 (88.8)	9,329 (74.1)	1,167 (65.9)	56 (93.3)	15,854 (50.0)	2538.91	<.001	
4	19,789 (42.2)	89 (11.2)	3,267 (25.9)	603 (34.1)	4 (6.7)	15,826 (50.0)			
Pain									
Yes	13,286 (28.3)	209 (26.3)	3,082 (24.5)	641 (36.2)	15 (25.0)	9,339 (29.5)	169.23	<.001	
No	33,614 (71.7)	585 (73.7)	9,514 (75.5)	1,129 (63.8)	45 (75.0)	22,341 (70.5)			
Pain scale									
	1.40±2.42	1.35±2.43	1.23±2.34	1.86±2.70	1.22±2.31	1.44±2.43	156.63	<.001	c>e>b c>a

Note. Mean ± SD are presented for simplicity, but group differences were analyzed using the Kruskal-Wallis test due to non-normal data distribution. Dunn's test with Bonferroni correction was used for post hoc analysis, and adjusted p-values were considered significant at $p < 0.01$.

Abbreviations. ICU = intensive care unit, GW = general ward, ED = emergency department, OPD = outpatient department, SD = standard deviation, SBP = systolic blood pressure, DBP = diastolic blood pressure, PR = pulse rate, RR = respiratory rate, BT = body temperature, SpO2 = peripheral oxygen saturation, AVPU = alert, verbal, pain, unresponsive, GCS = Glasgow coma scale

3. KTAS 3, 4 단계 환자의 임상결과에 따른 간호기록 패턴 분석

응급실 환자들의 간호 기록 패턴을 분석한 결과, 임상 결과에 따라 간호사들이 기록한 활력징후와 중재 빈도에 유의미한 차이가 있는 것으로 나타났다(Table 7). 우선 활력징후의 측정 빈도를 살펴보면, 모든 활력징후에서 사망 환자군의 평균 측정 빈도가 가장 높았으나, 사후 분석 결과에서 중환자실 환자군과의 비교에서는 유의한 차이가 없는 것으로 나타났다. 혈압 측정 빈도는 사망군(7.50 ± 6.069)이 가장 높았고, 귀가 환자군(2.10 ± 1.401)이 가장 낮았다($p < .001$). 사후 분석에서 중환자실 입원군과 사망군은 일반 병동 입원군, 전원군보다 높았고, 귀가 환자군은 가장 낮았다. 맥박수(PR)와 호흡수(RR) 측정 빈도 역시 혈압 측정 빈도 결과와 유사한 결과를 나타냈다. 체온(BT) 측정 빈도는 중환자실 입원군(3.13 ± 1.655)이 일반 병동 입원군, 전원군보다 높았고, 귀가 환자군(1.88 ± 0.990)이 가장 낮았다($p < .001$). 산소포화도(SpO₂)의 측정 빈도는 사망군(14.73 ± 10.447)이 가장 높았고, 귀가 환자군(2.52 ± 2.219)이 가장 낮았다($p < .001$). 사후 분석 결과 귀가 환자군은 모든 군보다 빈도가 유의하게 낮았다.

의식 수준의 기록 빈도는 사망군(2.92 ± 2.181)이 가장 높은 빈도를 보였고, 귀가 환자군(0.97 ± 1.037)이 가장 낮았다($p < .001$). 사후 분석에서 중환자실 입원군과 사망군은 다른 군에 비해 빈도가 유의미하게 높았다. 통증 사정 빈도는 사망군(0.87 ± 1.282)이 가장 높았으며, 귀가 환자군(0.25 ± 0.682)이 가장 낮았다($p < .001$). 사후 분석에서 중환자실 입원군과 전원군은 일반 병동 입원군보다 유의미하게 높은 빈도를 보였다.

간호 중재 항목에서, 의사에게 보고한 기록의 빈도는 사망군(0.57 ± 0.927)이 가장 높았고, 귀가 환자군(0.13 ± 0.380)이 가장 낮았다($p < .001$). 사후 분석 결과 중환자실 입원군과 사망군은 일반 병동 입원군 및 전원군보다 빈도가 유의하게 높았다.

비정상적인 검사 결과를 기록한 빈도도 사망군(1.55 ± 1.682)이 가장 높고, 귀가 환자군(0.17 ± 0.530)이 가장 낮았다($p < .001$). 사후 분석에서 사망군은 모든 군보다 빈도가 유의미하게 높았고, 중환자실 입원군은 일반 병동 입원군과 귀가 환자군보다 높았다. 추가적인 검사를 요청한 기록 역시 사망군(0.30 ± 0.788)이 가장 높았으며, 귀가 환자군(0.13 ± 0.471)이 가장 낮았다($p < .001$). 사후 분석에서 중환자실 입원군과 일반 병동 입원군은 귀가 환자군보다 유의미하게 높았다.

치료 수준의 상향은 응급실 내 진료 구역의 변경으로 정의되었으며, 분석 결과 사망 환자의 48.3%에서 진료 구역이 상향 조정된 것으로 확인되었다. 이들 모두 중증 환자 치료 구역 이상의 상위 진료 구역으로 이동한 사례였다. 반면, 일반 병동으로 입원한 환자군에서는 39.9%가 진료 구역이 상향 조정되었으며, 이 중 대부분(30.2%)이 일반 침상 구역으로 이동하였다. 귀가 환자군의 경우, 진료 구역이 상향 조정된 비율은 16.3%로 상대적으로 낮았으며, 이들 대부분은 경증 환자 구역에서 침상 구역으로 이동한 것으로 나타났다.

마지막으로, 간호사가 flowsheet 에 기록한 코멘트의 빈도는 사망군(2.45 ± 3.306)이 가장 높은 빈도를 보였고, 귀가 환자군(0.24 ± 0.967)이 가장 낮았으며($p < .001$), 사망군은 모든 군보다 유의미하게 높은 빈도를 보였다.

Table 7. Nursing documentation patterns in the included cohort (n= 46,900)

Variables	Total set (n (%), mean ± SD)							H/χ2	p	post hoc
	Total (n= 46,900)	ICU admission ^a (n= 794)	GW admission ^b (n= 12,596)	Transfer out ^c (n= 1,770)	Death ^d (n= 60)	Discharge ^e (n= 31,680)				
Nursing documentation patterns										
Frequency (within 6hrs)	range									
SBP/DBP	2.44±1.782	(0-30)	5.12±3.982	3.04±1.993	2.87±1.865	7.50±6.069	2.10±1.401	5400.27	<.001	(a,d)>b>c>e
PR	2.53±1.881	(0-38)	5.27±3.834	3.20±2.123	3.07±2.036	7.60±5.552	2.16±1.490	5522.91	<.001	(a,d)>b>c>e
RR	2.36±1.608	(0-30)	4.68±3.234	3.05±1.854	2.93±1.770	6.28±4.514	1.99±1.215	6486.29	<.001	(a,d)>b>c>e
BT	2.14±1.135	(0-22)	3.13±1.655	2.66±1.204	2.50±1.087	3.25±1.910	1.88±0.990	5528.14	<.001	a>b>c>e d>e
SpO2	4.02±4.056	(0-89)	7.50±5.617	7.11±4.940	6.91±5.533	14.73±10.447	2.52±2.219	18468.79	<.001	d>c>e b>c (a,b)>e
Mental status (AVPU, GCS)	1.26±1.268	(0-20)	2.71±1.950	1.82±1.432	1.72±1.389	2.92±2.181	0.97±1.037	6007.01	<.001	(a,d)>b>c>e
Pain scale	0.32±0.768	(0-8)	0.55±0.982	0.44±0.869	0.59±1.063	0.87±1.282	0.25±0.682	1060.97	<.001	(a,c)>b>e d>e
Intervention										
Recording of notifying a doctor	0.16±0.448	(0-6)	0.37±0.720	0.23±0.543	0.23±0.521	0.57±0.927	0.13±0.380	653.09	<.001	(a,d)>(b,c)>e
Recording of abnormal lab results	0.32±0.829	(0-15)	0.67±1.169	0.60±1.150	0.69±1.359	1.55±1.682	0.17±0.530	3155.88	<.001	d>a>b>e d>c>e
Recording of request for additional tests	0.14±0.476	(0-10)	0.19±0.496	0.15±0.482	0.16±0.512	0.30±0.788	0.13±0.471	61.53	<.001	(a,b)>e
Escalation of care								7835.63	<.001	
A zone to B zone	8,570 (18.3)	44 (5.5)	3,810 (30.2)	383 (21.6)	0 (0.0)	4,333 (13.7)				
A zone to C zone	1,313 (2.8)	96 (12.1)	576 (4.6)	71 (4.0)	3 (5.0)	567 (1.8)				
A zone to R zone	12 (<0.1)	4 (0.5)	1 (<0.1)	0 (0.0)	2 (3.3)	5 (<0.1)				
B zone to C zone	1,149 (2.4)	124 (15.6)	636 (5.0)	125 (7.1)	13 (21.7)	251 (0.8)				
B zone to R zone	16 (<0.1)	4 (0.5)	2 (<0.1)	0 (0.0)	6 (10.0)	4 (<0.1)				
C zone to R zone	14 (<0.1)	6 (0.8)	0 (0.0)	1 (0.1)	5 (8.3)	2 (<0.1)				
None	35,826 (76.4)	516 (65.0)	7,571 (60.1)	1,190 (67.2)	31 (51.7)	26,518 (83.7)				
Comments on flowsheet	0.32±1.129	(0-23)	1.3±2.46	0.42±1.288	0.37±1.185	2.45±3.306	0.24±0.967	1121.60	<.001	d>a>(b,c)>e

Note. Mean ± SD are presented for simplicity, but group differences were analyzed using the Kruskal-Wallis test due to non-normal data distribution. Dunn's test with Bonferroni correction was used for post hoc analysis, and adjusted p-values were considered significant at $p < 0.01$.

Abbreviations: SBP = systolic blood pressure, DBP = diastolic blood pressure, PR = pulse rate, RR = respiratory rate, BT = body temperature, SpO2 = peripheral oxygen saturation, AVPU = alert, verbal, pain, unresponsive, GCS = Glasgow coma scale, A zone = ambulatory zone, B zone = bed zone, C zone = critical care zone, R zone = resuscitation zone

4. 예측 모델 개발 및 성능 평가

본 연구에서는 21 개의 예측 변수를 포함한 모델 1 과 33 개의 예측 변수를 포함한 모델 2 의 성능을 비교하여 간호 기록 데이터의 추가가 임상 결과 예측 능력에 미치는 영향을 분석하였다. 두 모델은 다항 로지스틱 회귀, 랜덤 포레스트, 그래디언트 부스트의 세 가지 머신러닝 알고리즘을 활용하여 총 6 개의 예측 모델이 개발되었으며, 내부 검증을 위한 혼동행렬(confusion matrix)은 Table 8 과 같다.

모델 1 에서 로지스틱 회귀의 전체 정확도는 60.7%로 가장 낮았으며, 귀가 예측에서 6,336 건 중 4,618 건(72.9%)을 정확히 분류하였다. 중환자실 입원(13.2%), 일반 병실 입원(39.1%), 전원(19.2%), 사망(33.3%)과 같은 나머지 클래스에 대한 정확도는 40% 미만이었다. 랜덤 포레스트의 경우 전체 정확도는 72.4%로, 로지스틱 회귀에 비해 높은 성능을 보였으며, 일반 병실 입원에서는 2,519 건 중 1,692 건(67.2%)을, 귀가에서는 6,336 건 중 5,053 건(79.8%)을 정확히 예측하였다. 그러나 중환자실 입원(6.9%), 전원(9.6%), 사망(0%)과 같은 소수 클래스에서는 낮은 정확도를 보였다. 그래디언트 부스트는 전체 정확도가 63.3%였으며, 귀가(64.7%)와 일반 병실 입원(66.3%)에서 상대적으로 높은 정확도를 보였다. 또한 중환자실 입원(31.4%)과 전원(31.9%)에서 랜덤 포레스트보다 높은 성능을 나타내었다.

간호 기록 데이터를 포함한 모델 2 는 모든 알고리즘에서 전반적인 정확도가 향상되었다. 로지스틱 회귀의 전체 정확도는 68.8%로 증가하였으며, 귀가에서는 6,336 건 중 5,028 건(79.4%)을, 일반 병실 입원에서는 2,519 건 중 1,293 건(51.3%)을 정확히 예측하였다. 중환자실 입원(22.0%)과 사망(50.0%)에서도 예측 정확도가 증가하였다. 랜덤 포레스트는 전체 정확도가 79.6%로 가장 높은 성능을 보였으며, 귀가에서는 6,336 건 중 5,254 건(82.9%)을, 일반 병실 입원에서는 2,519 건 중 2,149 건(85.3%)을 정확히 예측하였다. 하지만 중환자실 입원 (15.1%)과

전원(11.6%)에서는 성능이 낮았다. 그래디언트 부스트는 전체 정확도가 74.0%였으며, 귀가(76.1%)와 일반 병실 입원(76.3%)에서 높은 성능을 보였다. 특히 중환자실 입원(39.6%)과 전원(37.6%)에서는 모든 모델 중 가장 높은 정확도를 기록하였다.

혼동 행렬을 통해 결과 변수별로 예측 정확도를 확인한 결과, 귀가와 일반 병실 입원에 대한 예측 정확도는 각각 82.9%, 85.3%로 모델 2의 랜덤 포레스트 알고리즘이 가장 높은 성능을 보였다. 중환자실 입원(39.6%), 전원(37.6%)에서는 모델 2의 그래디언트 부스트 알고리즘이 가장 높은 성능을 보였으며, 사망에 대한 예측은 모델 2의 로지스틱 회귀 알고리즘이 50.0%로 가장 높은 정확도를 보였다.

Table 8. Confusion matrix of clinical outcome predictions

Model	Prediction modeling method	Predicted	Actual					Accuracy
			ICU admission (n=159)	GW admission (n=2,519)	Transfer (n=354)	Death (n=12)	Discharge (n=6,336)	
Model 1	Logistic regression	ICU admission	21	190	23	1	282	60.7%
		GW admission	45	985	63	5	802	
		Transfer	12	233	68	1	527	
		Death	29	139	39	4	107	
		Discharge	52	972	161	1	4,618	
	Random forest	ICU admission	11	27	3	0	21	72.4%
		GW admission	113	1,692	128	12	1,111	
		Transfer	5	71	34	0	151	
		Death	0	1	1	0	0	
		Discharge	30	728	188	0	5,053	
	XGBoost	ICU admission	50	234	20	3	205	63.3%
		GW admission	82	1,670	116	4	1,283	
		Transfer	10	198	113	3	742	
		Death	2	21	6	1	6	
		Discharge	15	396	99	1	4,100	
Model 2	Logistic regression	ICU admission	35	208	17	0	256	68.8%
		GW admission	43	1,293	85	4	612	
		Transfer	13	297	96	2	392	
		Death	24	105	31	6	48	
		Discharge	44	616	125	0	5,028	
	Random forest	ICU admission	24	53	6	3	37	79.6%
		GW admission	120	2,149	198	9	928	
		Transfer	1	48	41	0	117	
		Death	0	0	1	0	0	
		Discharge	14	269	108	0	5,254	
	XGBoost	ICU admission	63	200	20	2	186	74.0%
		GW admission	69	1,921	137	6	813	
		Transfer	9	205	133	2	517	
		Death	5	16	6	1	0	
		Discharge	13	177	58	1	4,820	

Note. Model 1 includes 21 predictors based on triage information only. Model 2 includes 33 predictors, integrating nursing documentation patterns with triage information.

Abbreviations. ICU = intensive care unit, GW = general ward, XGBoost = extreme gradient boosting

혼동행렬을 통해 확인된 소수 클래스와 다수 클래스 간의 분류 성능 차이를 보완하기 위해 AUROC, AUPRC, 민감도(재현율), 특이도, 양성예측도(정밀도), 음성예측도, F1 점수와 같은 종합적인 성능 지표를 분석하였다(Table 9).

모델 1의 경우, 로지스틱 회귀의 AUROC는 0.850, AUPRC는 0.641이었으며, 민감도(0.607)는 특이도(0.861)보다 낮아 다수 클래스에서 더 높은 성능을 나타냈다. 양성예측도(0.680)와 음성예측도(0.853)는 음성 클래스에서 더 높은 성능을 보였으며, F1 점수는 0.641로 나타났다. 랜덤 포레스트는 AUROC(0.941)와 AUPRC(0.813)가 가장 높은 성능을 보였다. 민감도(0.724)와 특이도(0.892)가 균형 잡힌 성능을 나타냈으며, F1 점수 0.725였다. 그래디언트 부스트는 AUROC는 0.893, AUPRC는 0.705로 로지스틱 회귀보다 높은 성능을 보였다. 민감도(0.633)와 특이도(0.891)는 다수 클래스에서 강점을 나타냈으며, F1 점수는 0.686이었다.

모델 2는 모든 알고리즘에서 모델 1 대비 전반적인 성능이 개선되었다. 로지스틱 회귀의 AUROC는 0.891, AUPRC는 0.715로 증가하였으며, 민감도(0.688)와 특이도(0.896)는 더 높은 균형을 나타냈다. F1 점수는 0.722으로 향상되었다. 랜덤 포레스트는 AUROC(0.964), AUPRC(0.888)로 가장 높은 성능을 나타냈다. 민감도(0.796)와 특이도(0.932)는 높은 수준의 성능을 나타냈으며, F1 점수는 0.802으로 모든 알고리즘 중 가장 높았다. 그래디언트 부스트는 AUROC(0.940)와 AUPRC(0.837)로 높은 성능을 보였다. 민감도(0.740)와 특이도(0.928)는 높은 수준의 성능을 나타냈으며, F1 점수는 0.781로 모델 1 대비 크게 향상되었다.

AUROC와 AUPRC를 기반으로 한 추가적인 분석을 통해 모델의 전반적인 성능을 시각적으로 비교하였다. 수신자 조작 특성 곡선(ROC Curve)에서 모델 1과 모델 2 모두 그래디언트 부스트와 랜덤 포레스트가 로지스틱 회귀에 비해 높은 성능을

나타냈다(Figure 5). 모델 1에서 랜덤 포레스트는 AUROC 0.941로 가장 우수한 성능을 기록하였다. 모델 2에서도 랜덤 포레스트는 AUROC 0.964로 가장 높은 성능을 보였다.

정밀도 재현율 곡선(Precision-Recall curve, PRC)을 통해 각 클래스별 예측 성능을 분석한 결과, 귀가 클래스의 평균 정밀도(Average Precision, AP)는 모델 2가 0.960으로 모델 1 (0.905)보다 높았다(Figure 6). 일반 병동 입원 클래스에서도 모델 2의 평균 정밀도는 0.713로, 모델 1 (0.601)보다 높았다. 중환자실 입원과 전원 클래스에서는 두 모델 모두 평균 정밀도가 낮았지만, 모델 2에서 다소 성능이 향상되었다.

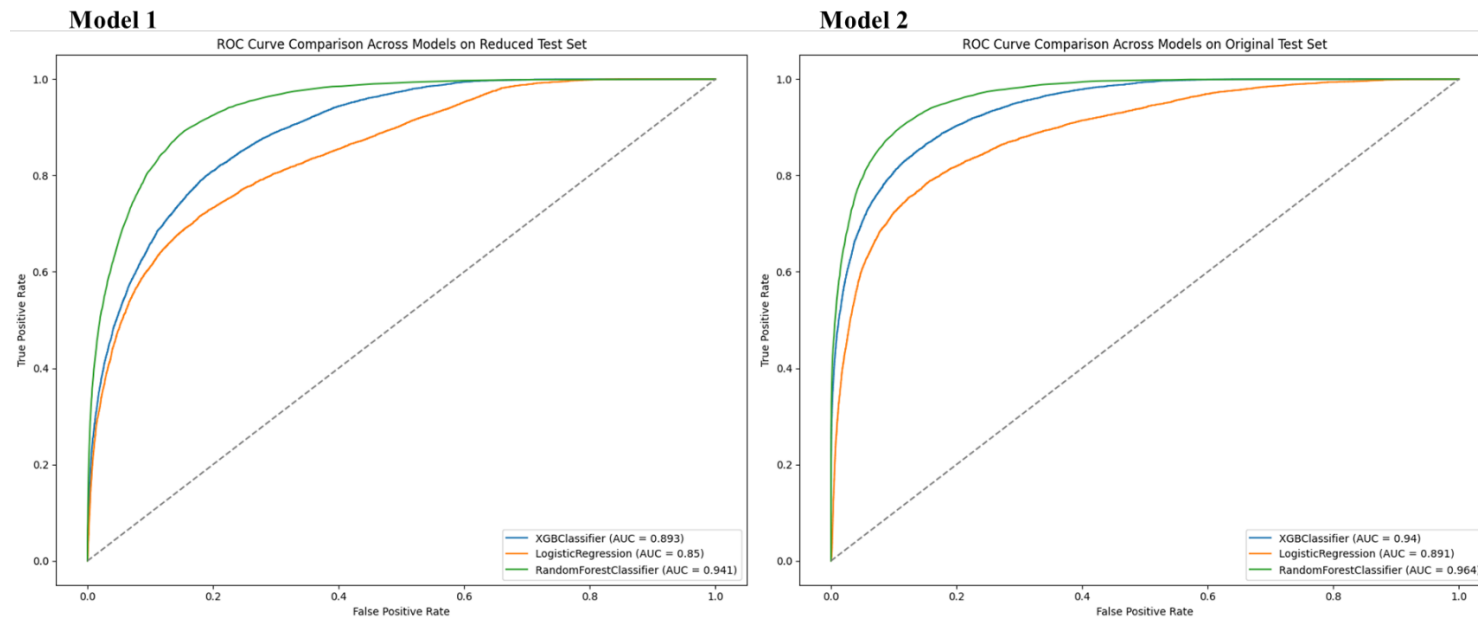
Table 9. Model performance of clinical outcome prediction models

Model	AUROC [95% CI]	AUPRC [95% CI]	Sensitivity	Specificity	PPV	NPV	F1 score
Model 1							
Logistic regression	0.850 [0.845, 0.855]	0.641 [0.630, 0.653]	0.607	0.861	0.680	0.853	0.641
Random forest	0.941 [0.938, 0.944]	0.813 [0.805, 0.821]	0.724	0.892	0.726	0.888	0.725
XGBoost	0.893 [0.890, 0.898]	0.705 [0.695, 0.715]	0.633	0.891	0.748	0.870	0.686
Model 2							
Logistic regression	0.891 [0.887, 0.896]	0.715 [0.705, 0.727]	0.688	0.896	0.760	0.884	0.722
Random forest	0.964 [0.961, 0.966]	0.888 [0.882, 0.894]	0.796	0.932	0.809	0.920	0.802
XGBoost	0.940 [0.937, 0.944]	0.837 [0.830, 0.845]	0.740	0.928	0.826	0.903	0.781

Note. Model 1 includes 21 predictors based on triage information only. Model 2 includes 33 predictors, integrating nursing documentation patterns with triage information.

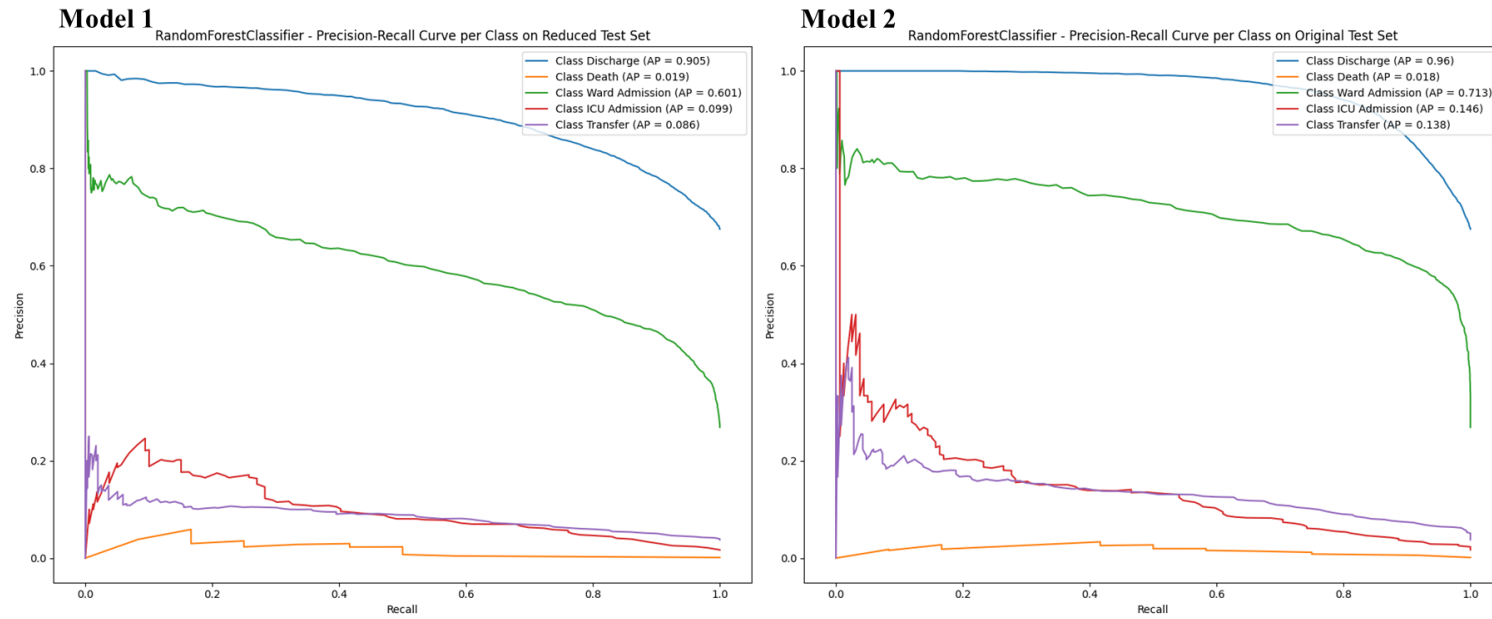
Abbreviations. AUROC = area under the receiver operating characteristic curve, AUPRC = area under the precision-recall curve, CI = confidence interval, PPV = positive predictive value, NPV = negative predictive value, XGBoost = extreme gradient boosting

Figure 5. Receiver operating characteristic curves of prediction models



Abbreviations. ROC = receiver operating characteristic curve, AUC = area under the curve

Figure 6. Precision-recall curves of prediction models



Abbreviation. AP = average precision

5. 최종 모델 선정 및 변수 중요도

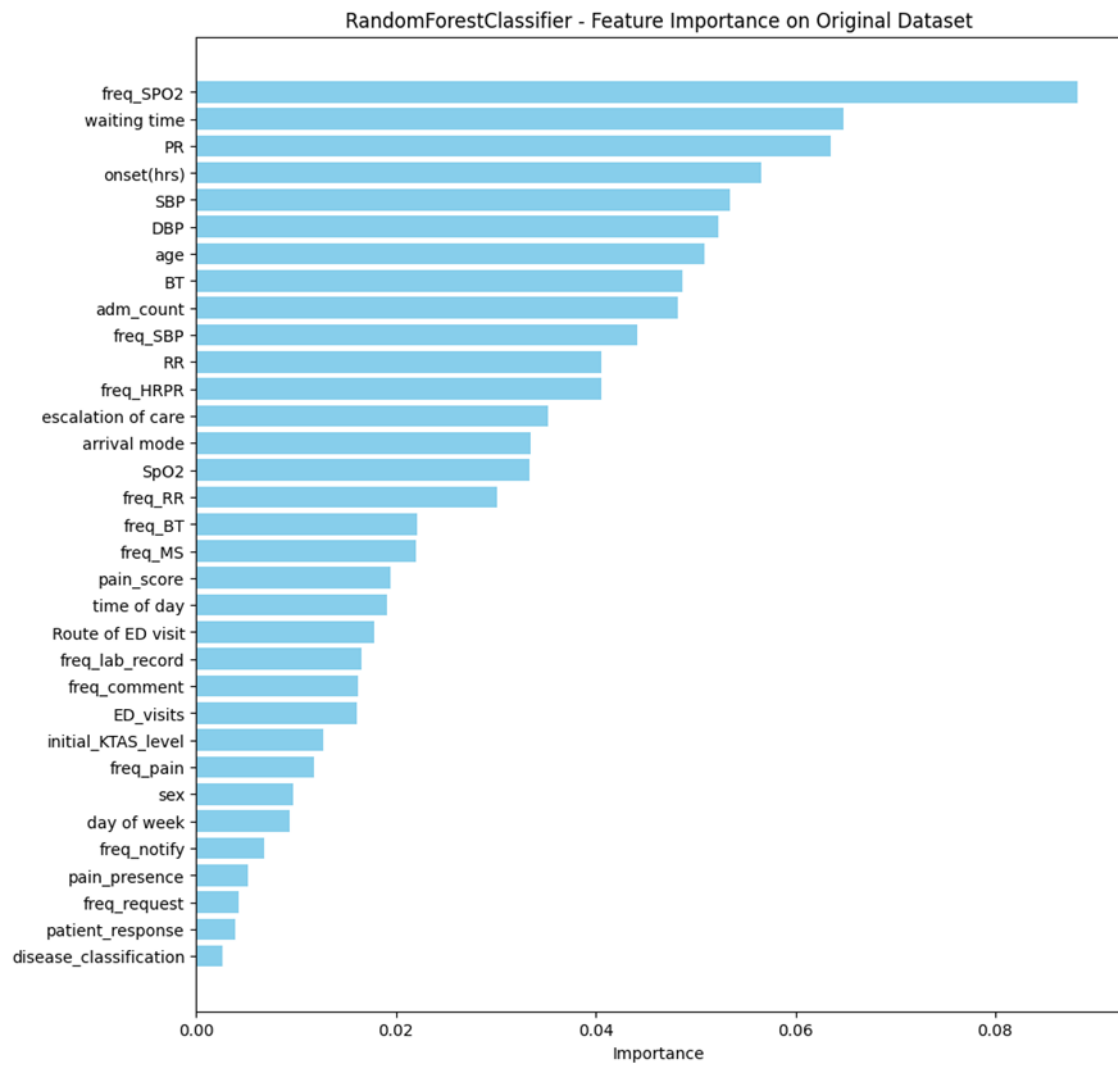
개발된 모델 성능을 비교한 결과, 랜덤 포레스트 모델 2가 모든 성능 지표에서 가장 우수한 결과를 나타냈다. 이 모델은 AUROC(0.964)와 AUPRC(0.888)에서 가장 높은 값을 나타내며, 전반적인 분류 성능이 높았다. 특히, 민감도와 특이도가 각각 0.796과 0.932로 가장 높은 값을 보여, 중요한 임상 결과(중환자실 입원, 전원, 사망)를 놓칠 가능성을 줄이는 동시에, 귀가에 대한 결과를 정확히 분류하는 성능을 보였다. 또한, F1 점수(0.796)와 정확도(0.796)에서도 가장 높은 성능을 나타내어 민감도와 정밀도 간의 균형이 있음을 확인하였다.

랜덤 포레스트 모델 1과 그래디언트 부스트 모델 2는 AUROC 및 AUPRC 지표에서 양호한 성능을 보였으나, 민감도와 F1 점수에서 랜덤 포레스트 모델 2에 비해 낮은 성능을 보였다. 특히 그래디언트 부스트 모델 2는 중요한 임상결과에 대한 예측에서 랜덤 포레스트 모델 2보다 낮은 민감도를 나타냈다.

따라서 본 연구에서는 랜덤 포레스트 모델 2가 다양한 성능 지표에서 가장 우수한 결과를 보였기 때문에 최적의 모델로 선정되었다. 최적 모델로 선정된 랜덤 포레스트 모델 2에서 Gini 중요도 분석을 통해 변수들의 상대적 기여도를 평가한 결과, 특정 변수들이 임상 결과 예측에 중요한 역할을 수행하는 것으로 나타났다(Figure 7). 변수 중요도는 모델 학습 과정에서 각 변수의 분기(split)에 기여한 정도를 기반으로 산출되었으며, 값이 높을수록 해당 변수가 모델의 예측 성능에 중요한 영향을 미쳤음을 의미한다.

가장 중요한 변수는 간호 기록 패턴 변수 중 산소포화도 측정 빈도로 나타났으며 그 외 대기 시간, 맥박수, 증상 발현부터 내원시까지 소요시간, 수축기 혈압, 이완기 혈압, 연령, 체온, 입원 횟수, 혈압 측정 빈도가 상위 10개 중요한 변수로 확인되었다.

Figure 7. Feature importance in the random forest model 2



변수 중요도에 대한 Gini 중요도 분석 결과를 통해 각 변수의 상대적 기여도를 확인한 데 이어, 모델의 예측 결과에 대한 보다 세부적인 해석을 위해 셰플리 가산 설명 기법(SHAP) 분석을 수행하였다. 랜덤 포레스트 모델 2의 변수별 예측 기여도를 SHAP summary plot 으로 분석한 결과는 다음과 같다(Figure 8).

중환자실 입원 예측에서 가장 중요한 변수는 1년 이내 입원 횟수와 초기 KTAS 분류 결과로 확인되었다. 과거 입원 경험이 많은 환자일수록, 초기 KTAS 레벨이 3인 경우 중환자실 입원 가능성이 증가하였다. 혈압을 비롯한 활력징후 측정 빈도는 모델 예측에 중요한 기여를 하였으며, 활력징후의 측정 빈도가 높을수록 중환자실 입원 가능성이 증가하였다. 내원 수단은 앰블런스를 통해 도착한 환자의 중환자실 입원 가능성을 증가시키는 주요 변수로 나타났다.

일반 병실 입원 예측에서는 산소포화도 측정 빈도와 1년 이내 입원 횟수가 가장 중요한 변수로 나타났다. 산소포화도 측정 빈도와 입원 횟수가 높을수록 일반 병실 입원 가능성이 증가하였다. 내원 수단 및 내원 경로도 모델 예측에 영향을 미쳤으며, 외래를 통해 내원하거나 타 의료기관으로부터 전원 온 경우 일반 병실 입원 가능성이 높아졌다.

다른 의료기관으로의 전원 예측에서는 1년 이내 입원 횟수와 산소포화도 측정 빈도가 중요한 변수로 나타났다. 입원 횟수가 많고, 산소포화도 측정이 빈번한 환자는 전원 가능성이 높았으며, 이는 복잡한 병력을 가진 중증 환자의 특징을 반영한다. 내원 수단과 통증 점수도 전원 가능성에 영향을 미쳤으며, 외부 병원으로 전원된 환자일 가능성이 높은 환자 특성을 나타냈다.

사망 예측에서 가장 중요한 변수는 내원 수단과 혈압, 산소포화도, 맥박 측정 빈도였다. 앰블런스로 도착한 환자와 혈압, 산소포화도, 맥박이 빈번히 모니터링된 환자에서 사망 가능성이 높았으며, 이는 초기 중증도가 사망 예측에 주요한 역할을

한다는 점을 보여주었다. 내원 시 호흡수, 맥박, 상위 진료구역으로의 이동도 사망 예측에 중요한 기여를 하였다.

귀가 예측에서는 산소포화도 측정 빈도와 내원 수단이 가장 중요한 변수로 확인되었다. 산소포화도 측정 빈도가 낮고, 도보로 도착한 환자에서 귀가 가능성이 높았으며, 초기 KTAS 레벨이 4 이고, 상위 진료구역으로의 이동이 필요하지 않은 환자는 귀가 가능성이 증가하였다. 1년 이내 입원 횟수와 통증 점수는 귀가 가능성을 평가하는 데 있어 보조적인 변수로 확인되었다.

Figure 8. SHAP summary plots for feature contributions in random forest model 2

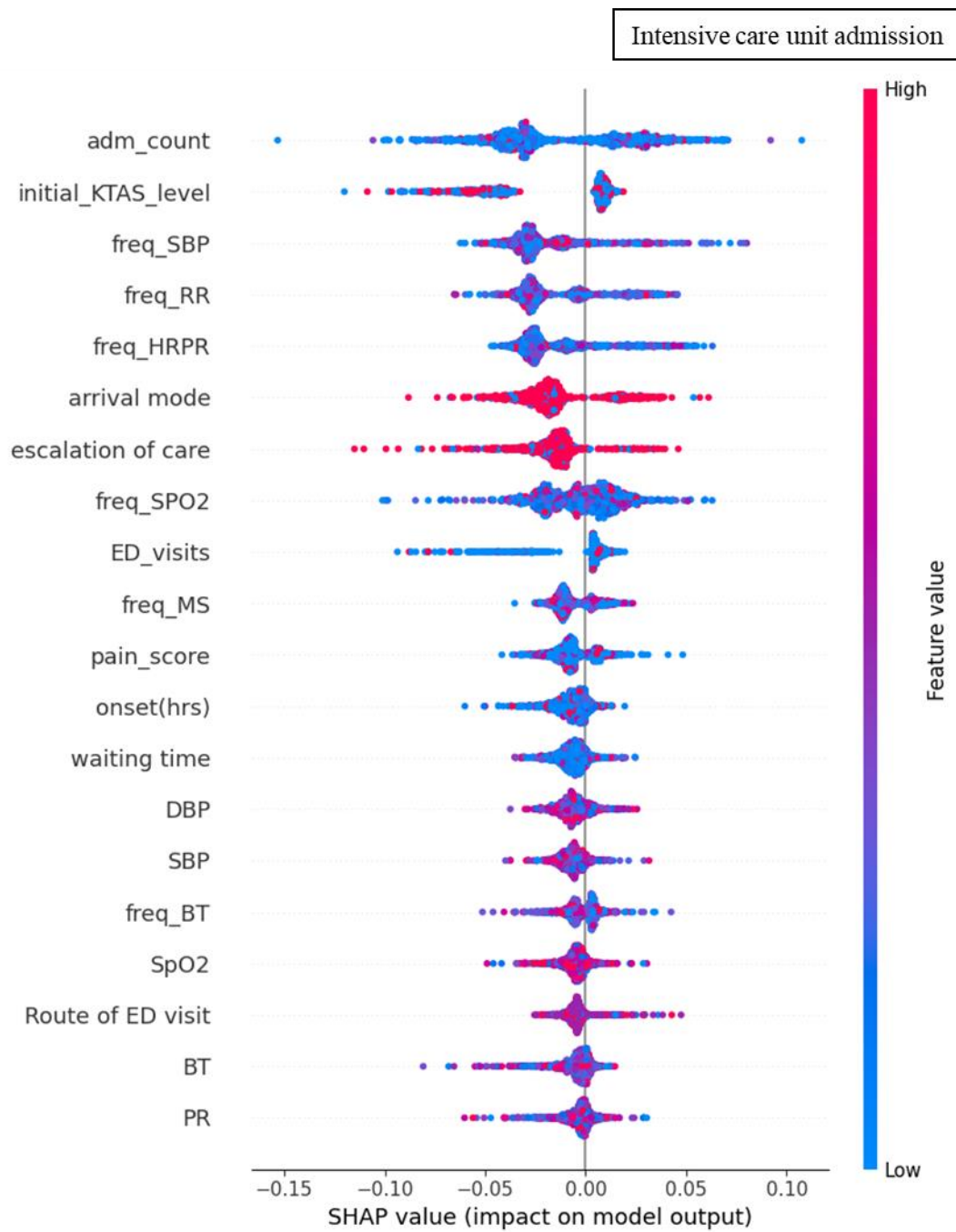


Figure 8. SHAP summary plots for feature contributions in random forest model 2 (continued)

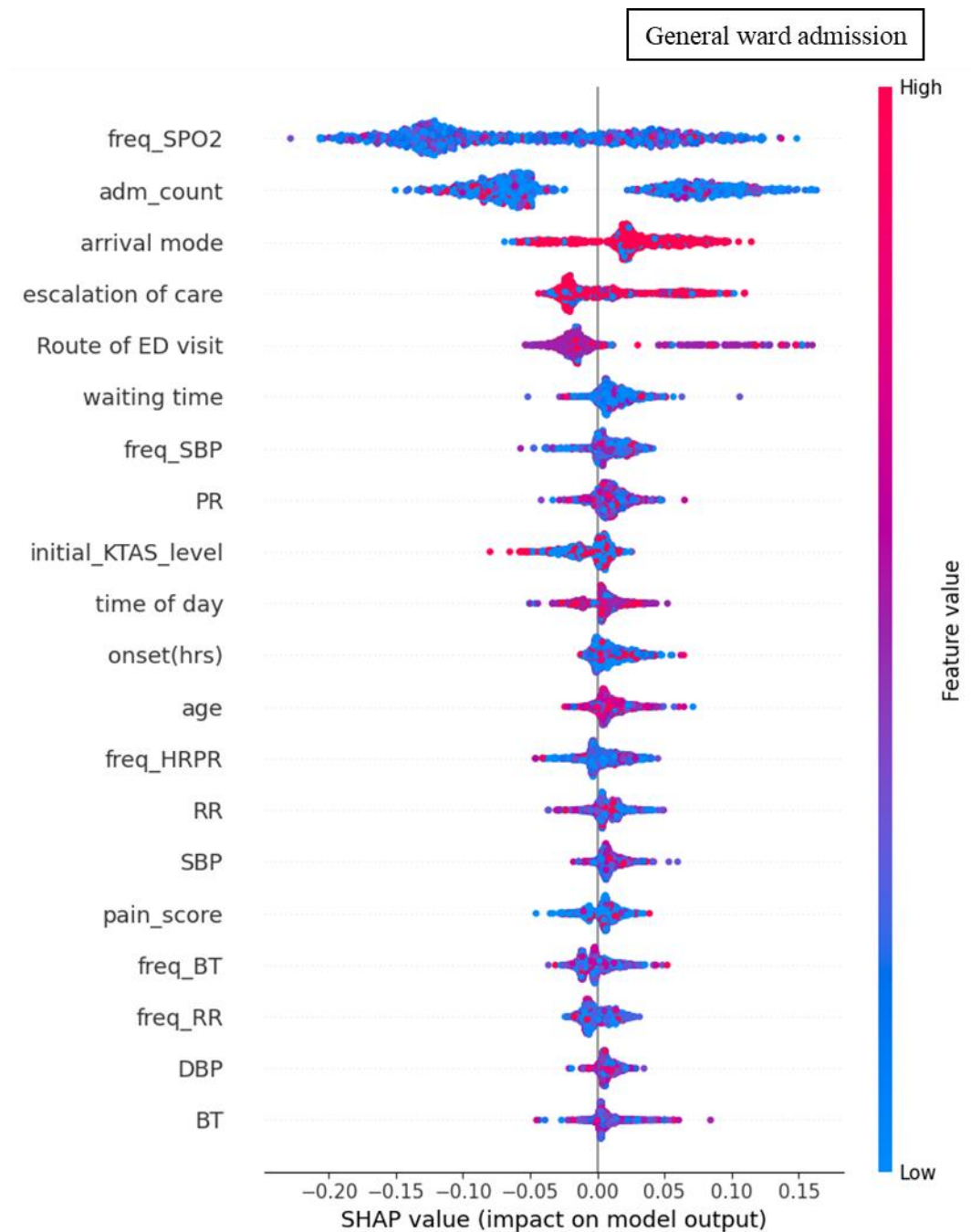


Figure 8. SHAP summary plots for feature contributions in random forest model 2 (continued)

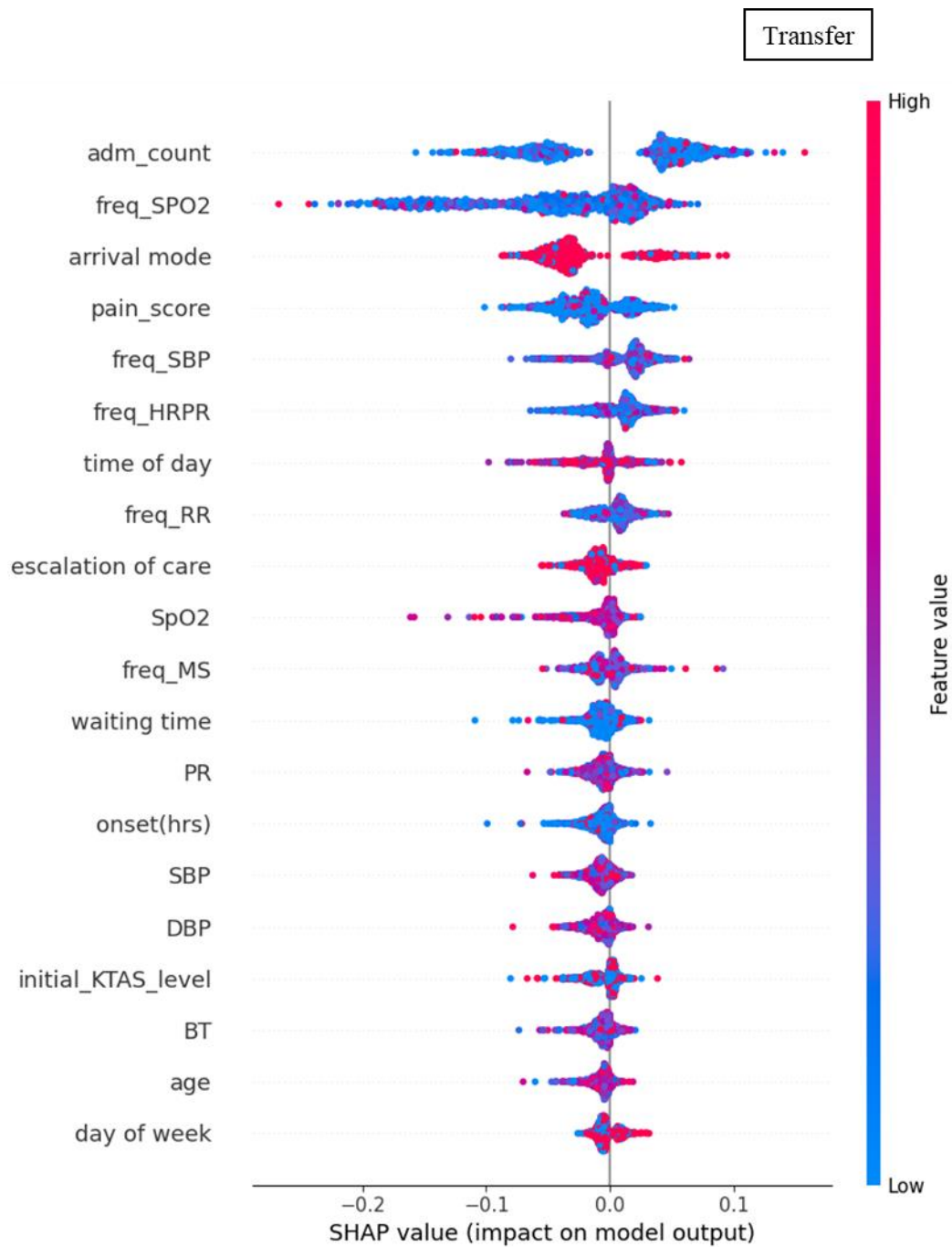
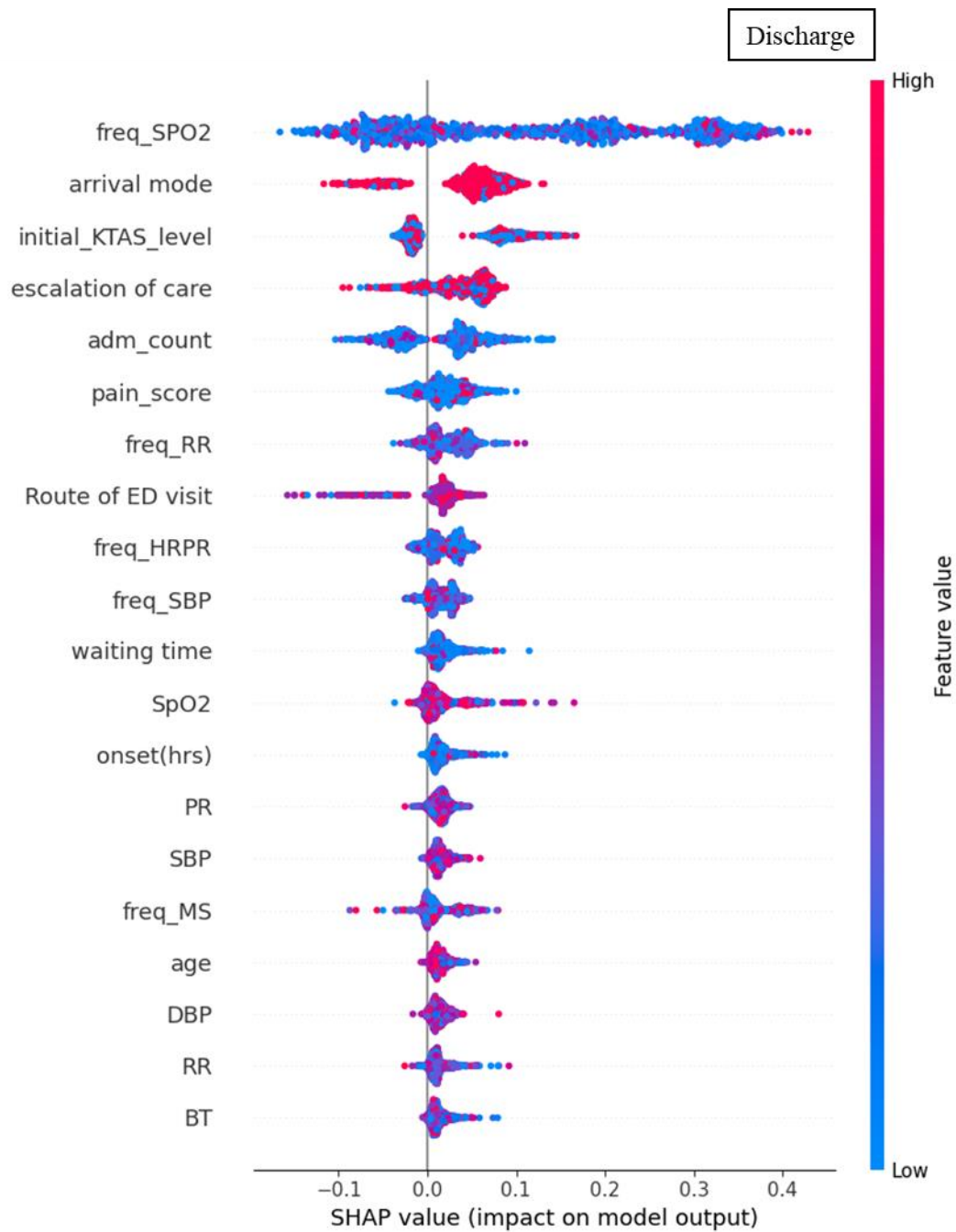


Figure 8. SHAP summary plots for feature contributions in random forest model 2 (continued)



Figure 8. SHAP summary plots for feature contributions in random forest model 2 (continued)



6. 보정 곡선 (Calibration plot)

최종 모델인 랜덤 포레스트 모델 2의 보정 곡선을 분석한 결과, 결과변수의 각 클래스별 예측 확률과 실제 관찰된 확률 간의 일치도를 시각적으로 확인할 수 있었다(Figure 9). 보정 곡선은 다섯 가지 주요 클래스에 대해 모델이 예측한 확률의 정확성을 나타내며, 이상적인 보정 상태는 대각선(점선)으로 표시된다.

귀가 클래스의 보정 곡선은 전반적으로 이상적인 보정선 위쪽에 위치하며, 이는 모델이 귀가 환자의 예측 확률을 과대평가(overconfidence)하고 있음을 나타낸다. 예측 확률이 0.2~0.8 구간에서 관찰 확률보다 높게 나타났으며, 이 구간에서 모델의 예측과 실제 확률 간의 불일치가 관찰되었다. 그러나 예측 확률이 높은 구간(0.8 이상)에서는 관찰 확률과의 차이가 줄어들어, 해당 구간에서 모델의 보정 상태가 상대적으로 개선되었다.

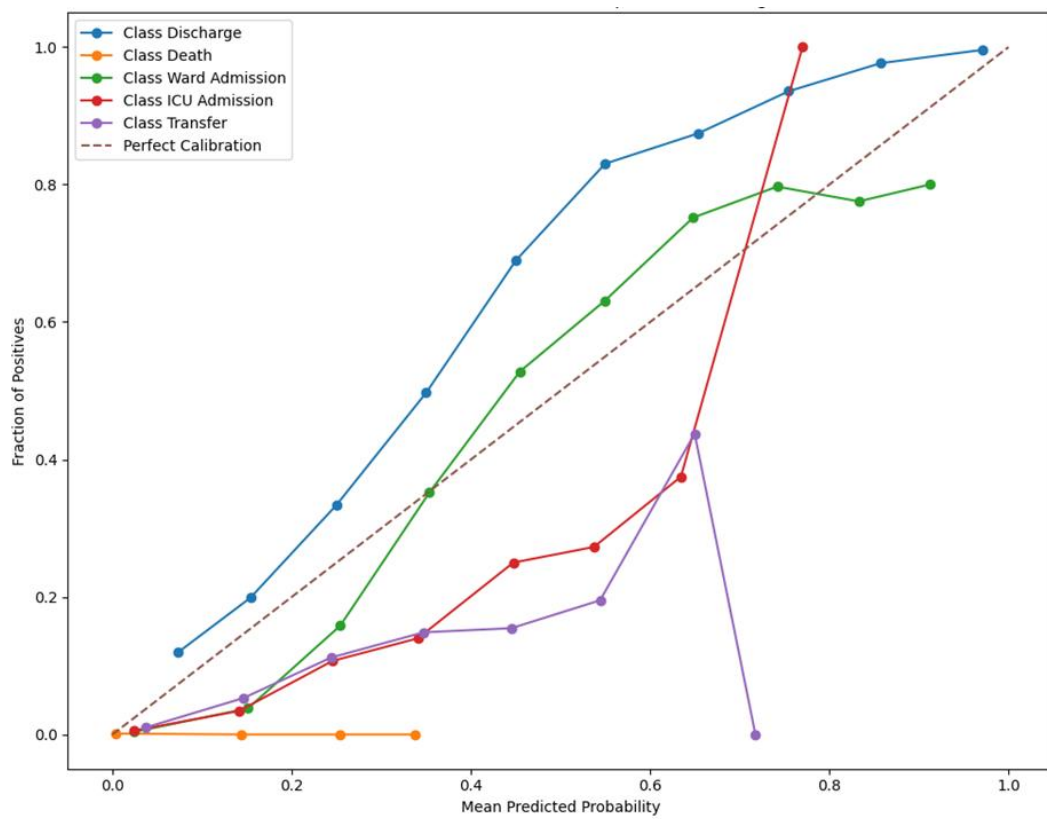
일반 병실 입원 클래스는 대체로 이상적인 보정선과 유사한 패턴을 보였으며, 예측 확률이 0.4 이상일 때 관찰 확률과의 일치도가 높게 나타났다. 그러나 예측 확률이 낮은 구간(0.2 미만)에서는 모델이 관찰 확률을 과소평가하는 경향이 확인되었다.

중환자실 입원 클래스의 보정 곡선은 전체적으로 이상적인 보정선과 크게 벗어나 있으며, 예측 확률이 낮은 구간(0.2 미만)에서 특히 관찰 확률과의 불일치가 두드러졌다. 이는 중환자실 입원 데이터의 불균형으로 인해 모델이 해당 클래스를 정확히 예측하지 못했음을 시사한다.

전원 클래스는 예측 확률이 중간 구간(0.4~0.6)에서 불안정한 곡선을 보이며, 관찰 확률과의 불일치가 확인되었다. 예측 확률이 높아질수록 관찰 확률과의 차이가 커져, 모델이 해당 클래스에 대해 적절히 보정되지 않았음을 보여준다. 사망 클래스의 보정 곡선은 이상적인 보정선과 크게 벗어나 있으며, 전반적으로 관찰 확률과 예측 확률

간의 불일치가 두드러지게 나타났다. 이는 사망 클래스가 극소수 클래스에 해당하며, 데이터 부족으로 인해 모델이 해당 클래스를 충분히 학습하지 못했을 가능성이 높다.

Figure 9. Calibration plot of the random forest model 2



VI. 논 의

본 연구는 응급실에 내원한 KTAS 3, 4단계 성인 환자를 대상으로, triage 정보와 간호 데이터를 활용하여 임상 결과를 예측하는 모델을 개발하고 검증한 후향적 서술 조사 연구이다. 연구에 사용된 데이터는 서울시 소재 상급종합병원의 임상 데이터 웨어하우스(CDW)에서 추출된 자료로, 모델 개발 및 내부 검증 과정에 적용되었다.

본 장에서는 연구 대상자인 KTAS 3, 4단계 응급 환자의 임상적 특성과 예측 모델의 성능을 중심으로 결과를 해석하고, 연구의 임상적 의의와 제한점을 고찰하고자 한다. 나아가 본 연구가 응급실 환자의 임상 결과 예측 모델 개발에 기여한 바를 평가하고, 향후 연구 및 실제 임상 적용 가능성에 대해 논의하고자 한다.

1. KTAS 3, 4단계 응급 환자의 임상결과 예측 모델 개발

본 연구는 2023년 한 해 동안 상급종합병원 응급실을 방문한 KTAS 3, 4단계 응급환자 46,900명을 대상으로 임상 결과를 분석하였다. 분석 결과, 귀가(67.5%)가 가장 큰 비율을 차지하였으며, 일반 병실 입원(26.9%), 전원(3.8%), 중환자실 입원(1.7%), 사망(0.1%) 순으로 나타났다. 특히 중환자실 또는 일반 병실로 입원한 환자의 비율은 KTAS 3단계에서 37.0%, KTAS 4단계에서 17.0%로 나타났다. 이러한 입원율은 같은 기간 발표된 2023년 응급의료 통계연보에서 보고된 전국 응급의료기관 KTAS 3단계 입원율 31.3%, KTAS 4단계 입원율 10.5%에 비해 높은 수준이다. 이는 본 연구의 대상 의료기관이 상급종합병원으로, 중증 환자의 진료 비중이 상대적으로 높아 평균적인 입원율보다 높은 결과를 보인 것으로 판단된다.

또한 주목할 점은, 중간 수준의 중증도와 긴급도를 나타내는 KTAS 3, 4단계에서 중환자실 입원(1.7%) 및 사망(0.1%)과 같은 중증 결과가 적지 않게 발생했다는 것이다. 특히, 동일한 KTAS 단계 내에서도 임상 결과가 다양하게 나타난 점은 초기 triage 단계에서의 오분류 가능성을 시사하며, 초기평가의 중요성이 강조된다. 선행 연구에서도 KTAS 단계가 응급실 재실 기간 중 조정된 환자군을 분석한 결과, KTAS가 조정된 환자가 변경되지 않은 환자에 비해 더 높은 입원률과 응급실 사망률을 보였다고 보고되었다(Chang et al., 2023). 이는 KTAS 정보만으로는 임상 결과를 충분히 예측하는 데 한계가 있음을 보여주는 동시에, 보다 정교한 예측 모델 개발의 필요성을 시사한다.

KTAS 3, 4단계 환자의 일반적 특성은 본 연구 대상 의료기관과 유사한 규모의 상급종합병원 응급실 환자를 대상으로 한 Choi et al.(2019)의 연구 결과와 유사한 경향을 보였다. 본 연구에서는 KTAS 3, 4단계 환자의 평균 연령이 58.7세로, Choi et al.(2019)에서 보고된 56~61세와 유사하였다. 성별 분포는 여성 50.4%, 남성 49.6%로, Choi et al.(2019)의 보고(KTAS 3단계 여성 52.3% 및 4단계 여성 54.7%)와 같이 여성 비율이 더 높았다. 내원 경로는 직접 내원이 71.8%로 가장 많았으며, 전원이 18.7%, 외래 경유가 9.5%로 나타나, Choi et al.(2019)에서 각각 77.2%, 15.0%, 7.8%로 보고된 결과와 유사하였다. 또한 본 연구의 분석 결과는 응급실 내원 환자의 다양한 활력징후 및 임상 특성이 임상 결과에 따라 유의미한 차이를 나타냄을 보여주었다. 특히, 사망 및 중증 환자군에서 활력징후의 이상 소견이 두드러지게 나타났다. 또한 사망 환자군에서 통증 점수가 낮게 나타난 것은 통증 여부를 측정할 수 없는 사례에서 통증 점수를 0점으로 대체한 결과로 판단된다. 이러한 결과는 통증 점수가 환자군에 따라 차이를 보일 수 있음을 시사하며, 통증 관련 정보의 수집과 해석에 있어 환자 상태와 측정 가능성에 대한 고려가 필요함을 보여준다.

KTAS 3, 4단계 환자의 간호기록 패턴 분석 결과, 임상결과에 따라 간호 기록의 빈도와 중재 내용에서 유의미한 차이가 관찰되었다. 특히, 활력징후 기록 중 산소포화도 측정이 가장 빈번하게 이루어졌으며, 사망 환자군에서 평균 14.73회로 가장 빈번하게 기록되었다. 이는 중증 환자의 상태 모니터링에서 산소포화도가 중요한 지표로 고려되었음을 보여준다. 또한, 의식 수준 및 통증 평가와 관련된 기록 역시 사망 환자와 중환자실 입원 환자군에서 높은 빈도로 나타났으며, flowsheet 에 작성된 코멘트 빈도 또한 중증 환자군에서 유의미하게 증가하였다. 이는 간호사가 중증 환자의 상태 변화를 인지하고 이를 직관적으로 판단하여 추가적인 기록을 남겼음을 유추해 볼 수 있다. 이러한 결과는 간호 기록의 중요성을 강조한 선행 연구와 맥락을 같이한다(Collins et al., 2013; Rossetti et al., 2019; Rossetti et al., 2024). Collins et al. (2013)은 간호 기록 패턴과 환자 사망률 간의 연관성을 제시하며, 사망 환자에서 선택적 코멘트와 활력징후 기록의 빈도가 생존 환자에 비해 유의하게 높았다고 보고하였다. 또한 Rossetti et al.(2024)의 CONCERN 모델은 간호사의 기록 패턴을 활용해 환자 악화를 조기에 감지하고 위험 수준을 분류하는 시스템을 제안하였는데 이 시스템은 중증 환자의 상태 변화를 조기에 파악하여 의료진이 적시에 중재할 수 있도록 지원하며, 병원 내 사망률을 35.6% 감소시키고, 평균 재원일수를 11.2% 단축시킨 것으로 보고되었다.

본 연구 결과는 생리적 지표뿐만 아니라 간호사의 지속적인 관찰과 상태 변화 기록이 중증 환자 관리에서 필수적인 역할을 한다는 점을 재확인하였다. 특히, 간호 기록 패턴에 대한 분석은 환자의 상태 변화에 대한 예측 가능성을 높일 수 있음을 확인하였다. 또한 flowsheet 에 확인할 수 있는 활력징후 기록 패턴뿐만 아니라 간호기록 내용 중 간호사의 우려가 반영되었을 것으로 추정되는 의사 보고 기록, 비정상적인 검사 결과에 대한 기록, 추가 검사를 요청한 기록 역시 중증 환자일수록

더 빈번하게 작성되었음을 확인하였다. 이는 간호 기록이 단순한 문서화 이상의 역할을 하며, 상태 악화를 조기에 인지하고 이에 대응하는 과정에서 중요한 역할을 한다는 점을 시사한다.

모델 성능 평가 결과, 간호 데이터를 포함한 모델 2는 모든 알고리즘에서 모델 1보다 성능이 향상되었다. 특히 랜덤 포레스트 알고리즘을 활용한 모델 2는 가장 높은 예측 성능을 나타냈다. 이는 간호 데이터가 triage 정보만으로는 설명되지 않는 임상적 맥락과 환자 상태를 반영함으로써 모델의 예측력을 향상시켰을 가능성을 시사한다. 이러한 결과는 중환자를 대상으로 간호 데이터(간호 노트, 간호 평가, 문서화 빈도 등)를 결합할 때 모델의 성능이 생리학적 데이터 단독 사용보다 향상된다고 보고한 Kim et al.(2023)의 선행 연구 결과와도 일치한다. 또한, 체계적 문헌고찰 연구에서 머신러닝과 같은 복잡한 알고리즘이 응급실 triage에서 기존 분류 도구보다 사망률, 중증 치료 필요성, 입원 여부 등 주요 결과를 더욱 정확히 예측할 수 있음을 강조하였는데(Sánchez-Salmerón et al., 2022), 본 연구에서도 간호 데이터를 포함한 머신러닝 모델이 기존 triage 정보만으로 구성된 모델보다 주요 임상 결과를 더욱 정확히 예측할 수 있음을 확인한 결과와 일치한다.

최적 모델로 선정된 랜덤 포레스트 모델(Model 2)은 AUROC 0.964, AUPRC 0.888, 민감도 0.796, 특이도 0.932, PPV 0.809, NPV 0.920, F1 score 0.802를 기록하며, 전반적으로 균형 잡힌 성능을 보였다. 특히 AUPRC(0.888)와 PPV(0.809)는 드문 사건인 중증 환자 분류에서 높은 정확도를 나타내며, 민감도와 특이도 간의 균형을 통해 과잉 및 과소 분류를 최소화하였다. 이러한 결과는 모델이 불균형 데이터 환경에서도 중증 환자와 비중증 환자를 효과적으로 구분하며 응급실 자원 배분 최적화와 중증 환자 조기 발견에 기여할 수 있음을 보여준다.

반면, 국내 대규모 데이터를 활용한 Kwon et al.(2018)의 사망 예측 모델은 AUROC 0.935로 높은 전체 성능을 기록했지만, AUPRC(0.264)와 민감도(0.24)가 낮아 소수 클래스인 예측에 한계를 보였다. 이는 정밀도와 재현율 간의 균형을 충분히 확보하지 못했음을 시사한다. Fernandes et al.(2020)의 사망 및 심정지 예측 모델은 AUROC 0.96, AUPRC 0.31, Specificity 0.94, Sensitivity 0.84, PPV 0.06, F1 score 0.12를 보고하며, 대다수 사례에서는 높은 성능을 보였으나, 낮은 AUPRC와 PPV로 인해 드문 사건 예측에서 거짓 양성률이 많아 임상적 활용이 제한될 수 있음을 보여준다. Raita et al.(2019)의 critical care 예측 모델은 AUROC 0.86, 민감도 0.80, 특이도 0.76으로 양성과 음성 분류에서 균형을 이루었으나, PPV(0.06)가 낮아 critical care 필요성을 예측한 사례 중 실제로 맞는 비율이 매우 낮았다. 또한, AUPRC가 보고되지 않아 불균형 데이터 환경에서 양성 사례 판별의 정확도가 평가되지 못했다.

본 연구의 모델은 선행 연구(Fernandes et al., 2020; Kwon et al., 2018; Raita et al., 2019)와 비교하여 AUPRC(0.888)와 PPV(0.809)에서 높은 성능을 보였으며, 민감도와 특이도 간의 균형을 통해 과대 및 과소 분류를 모두 줄일 수 있는 강점을 보였다. 특히 F1 score(0.802)는 민감도와 정밀도 간의 조화를 나타내며, 불균형 데이터 환경에서도 신뢰할 수 있는 예측 성능을 나타냈음을 의미한다. 이러한 결과는 본 모델이 응급실에서 중증 환자 조기 발견과 비응급 환자 선별을 동시에 수행하며, 응급실 운영의 효율성을 높일 수 있는 가능성을 보여준다.

본 연구에서는 간호 데이터인 산소포화도 측정 빈도, 도착 경로, 과거 입원 횟수 등이 주요 예측 변수로 확인되었으며, 이는 임상적 의사결정을 지원하는 데 활용 가능성을 보여주었다. SHAP 분석을 통해 모델의 변수별 기여도를 정량적으로 평가하여 모델의 해석 가능성을 높였고, 이를 바탕으로 의료진이 예측 결과를 실질적으로 활용할 수 있는 근거를 마련하였다. 중환자실 입원 예측에서 과거 입원

횡수와 초기 KTAS 레벨이 주요 변수로 도출되었으며, 이는 중증 환자의 과거 병력과 초기 평가가 환자 분류 및 자원 배치에서 중요한 역할을 한다는 점을 시사한다. 일반 병실 입원에서는 산소포화도 측정 빈도가 중요한 변수로 나타났으며, 이는 지속적인 모니터링이 환자의 상태를 파악하고 적절한 치료를 계획하는 데 필수적임을 시사한다. 전원 및 사망 예측에서는 내원 수단(앰블런스)과 활력징후 측정 빈도가 주요 변수로 나타나, 응급 상황에서 신속한 평가와 초기 대응이 중증 환자의 임상 경과에 중요한 영향을 미친다는 점을 강조한다. 귀가 가능성을 예측한 결과, 산소포화도 측정 빈도가 낮고 도보로 내원한 환자에서 귀가 가능성이 높은 것으로 나타났다. 이러한 결과는 경증 환자를 효율적으로 분류하여 응급실 자원을 최적화할 수 있는 가능성을 보여준다.

2. 연구의 제한점

본 연구는 KTAS 3, 4단계 응급 환자의 임상 결과를 예측하기 위한 모델을 개발하고 평가하였으며, 다음과 같은 제한점이 있다. 첫째, 본 연구는 단일 의료기관의 데이터를 기반으로 수행되었기 때문에 연구 결과의 외적 타당성이 제한될 수 있다. 단일 의료기관 데이터는 특정 의료 환경과 환자군의 특성을 반영하기 때문에, 이를 다양한 의료 환경에 일반화하기 어렵다.

둘째, 소수 클래스(사망, 중환자실 입원)에 대한 예측 성능이 낮게 나타난 것은 데이터 불균형 문제와 밀접한 관련이 있다. 본 연구에서는 대상자의 특성 상 클래스 불균형이 불가피했기 때문에 이를 해결하기 위해 데이터 증강 기법인 SMOTE, 비용 민감 학습 등을 적용하여 이를 완화하고자 했으나, 클래스 불균형 문제를 완전히 해결하지는 못하였다. 이로 인해 소수 클래스에서 모델의 신뢰도가 저하될 수 있다.

셋째, 간호 기록 데이터는 모델 성능 향상에 기여했으나, 간호 기록의 표준화 및 데이터 품질 관리 측면에서 한계가 존재한다. 본 연구의 의료기관 응급실에서는 전자간호기록 시스템에서 다빈도 작성 문구에 대한 기록 코드를 일부 활용하고 있지만, 상당수의 간호 기록이 비구조화된 서술 형태로 작성되었다. 이러한 비구조화된 기록은 표준화된 용어 체계의 부재로 인해 데이터의 일관성과 해석 가능성을 저하시켜 모델 개발 및 해석 과정에서 데이터 품질을 저해하는 요인으로 작용할 수 있다. 이와 관련하여, 선행 연구들은 HIRAID(Historical data, Identify red flags, Assessment, Interventions, Diagnostics, and Reassessment 및 Communication)와 같은 표준화된 응급 간호 평가 프레임워크가 간호 기록의 체계적 구조화 및 데이터 품질 향상에 기여할 가능성을 보여주고 있다(HIRAID Research Group., 2021; Munroe et al., 2015; Munroe et al., 2022). 따라서 향후 국내 응급실에서도 HIRAID 와 같은 표준화된 프레임워크를 개발하거나 도입하여 간호 기록의 구조화와 데이터 품질 관리 방안을 강화할 필요가 있다. 이를 통해 간호 기록이 데이터로서 가치와 활용도를 높일 수 있을 것으로 기대된다.

넷째, 본 연구는 개념적 기틀에서 개인적 조정 요인 중 간호사의 특성을 변수로 포함하지 못하였다. 간호사의 경험, 교육 수준, 근무 환경과 같은 특성은 환자 관리와 예후에 중요한 영향을 미칠 수 있으나, 이를 반영하지 못함으로써 모델의 설명력이 제한되었다.

다섯째, 주증상 데이터는 환자의 초기 상태를 반영하는 중요한 변수임에도 불구하고, 이를 세분화하여 분석하지 못하였다. 본 연구의 대상 의료기관에서 추출한 주증상 데이터는 정형 데이터(75.4%)와 비정형 데이터(24.6%)가 혼재되어 있으며, 40개 이상의 카테고리로 구성되어 있어 모델의 예측 변수로 활용할 경우 데이터의 복잡성과 모델의 계산 부하가 증가할 가능성이 있다. 특히, 비정형 데이터는 자연어

처리 기술을 통해 전처리가 요구되며, 이 과정에서 데이터의 정보 손실로 인해 예측 성능에 영향을 미칠 수 있다. 따라서, 주증상 데이터를 보다 체계적으로 정제하고, 정형 및 비정형 데이터를 효과적으로 통합할 수 있는 방법론을 개발하는 것이 필요하다. 이러한 접근은 주증상 데이터의 임상적 유용성을 극대화하고, 예측 모델의 성능을 향상시키는 데 기여할 수 있을 것으로 기대된다.

여섯째, Gini importance 와 SHAP 분석을 통해 각 결과변수 클래스별 중요한 변수를 도출하였으나, 변수 간 상호작용에 대한 분석은 수행하지 못한 한계가 있었다. 변수 간 상호작용은 특정 변수 조합이 임상 결과에 미치는 영향을 보다 정교하게 파악할 수 있는 중요한 요소이다. 예를 들어, 산소포화도 측정 빈도와 초기 KTAS 레벨의 조합이 입원과 귀가 클래스에 어떻게 다르게 작용하는지 분석한다면, 예측 모델의 해석력과 임상 적용 가능성을 더욱 높일 수 있을 것이다. 최근 Aziz et al.(2024)의 연구에서는 변수 간 상호작용을 고려한 규칙 기반 접근을 통해 해석 가능성을 강화한 사례를 제시하였다. 이 연구에서는 앰블런스 도착 여부와 중증도(Triage Acuity)를 조합하여 긴 체류 시간을 예측하는 간단한 규칙을 도출하였으며, 이를 통해 임상 현장에서 이해하기 쉽고 적용 가능한 결과를 제공하였다. 추후 연구에서 이러한 변수 간 상호작용을 분석하고 해석 가능한 규칙 기반 접근을 적용해 볼 수 있으며, 이러한 시도는 실제 임상 현장에서 더욱 효과적인 의사결정 지원 도구로 활용될 가능성을 높일 수 있을 것이다.

3. 연구의 간호학적 의의

1) 이론적 측면

본 연구는 전자 간호기록의 패턴과 특징을 예측 변수로 활용하여, KTAS 3, 4단계 환자의 임상결과를 보다 정확하게 예측하는 모델을 제시함으로써 간호기록의 임상적

가치를 높였다. 특히, 연구에서 적용한 HPM-ExpertSignals 프레임워크는 간호사들이 환자의 상태를 평가하고 반영하는 과정을 이론적으로 설명하며, 이를 통해 간호사의 판단과 중재가 임상 결과 예측에 중요한 역할을 한다는 점을 강조한다. 이 연구는 기존의 환자 중증도 분류 체계에서 간호기록의 활용 가능성을 제시하고, 간호 데이터가 예측 모델의 성능을 향상시킬 수 있다는 이론적 근거를 마련하였다.

2) 연구적 측면

본 연구는 임상결과 예측이 어려운 KTAS 3, 4단계 환자를 대상으로 예측 모델을 개발하고 평가했다는 점에서 중요한 연구적 의의를 가진다. KTAS 3, 4단계 환자는 증상의 중등도가 다양하고, 예측 변수가 불확실한 경우가 많아 기존 연구에서도 임상결과 예측 정확도가 낮은 경우가 많았다. 본 연구는 이러한 환자군을 대상으로 예측 모델을 구축하여 실제 임상에서의 적용 가능성을 높이는 데 기여하였다.

특히, 본 연구는 간호데이터를 활용하여 기존의 제한적인 예측 변수 기반 연구를 확장했다는 점에서 의미가 있다. 간호데이터는 환자의 초기 상태와 임상적 맥락을 보다 세밀히 반영할 수 있는 중요한 정보를 제공하며, 본 연구에서는 간호데이터를 포함한 모델이 예측 성능을 향상시키는 데 기여함을 확인하였다. 이러한 연구적 접근은 간호기록이 단순한 기록을 넘어서 환자 예후를 예측하는 중요한 데이터로서 활용될 수 있음을 입증하였으며, 향후 간호 데이터 기반의 예측 모델 연구에 중요한 기초 자료가 될 것이다.

또한, 데이터 불균형 문제를 해결하기 위해 SMOTE 및 비용 민감 학습 등의 기법을 시도한 점도 본 연구의 중요한 강점이다. 이러한 방법론은 소수 클래스의 예측 성능을 개선하는 데 기여했으며, 데이터의 구조적 문제를 보완하여 보다 균형 잡힌 모델을 개발할 수 있도록 하였다.

마지막으로, 개발된 모델의 해석 가능성을 높이기 위해 SHAP 분석을 적용한 점은 연구의 또 다른 의의로 볼 수 있다. SHAP 분석은 변수 중요도를 각 임상결과 클래스별로 세분화하여, 모델의 의사결정 과정을 명확히 하고 해석 가능성을 제고하였다. 이러한 접근은 단순히 높은 성능의 모델을 개발하는 것을 넘어, 임상 실무자들이 모델을 효과적으로 활용할 수 있는 기반을 마련했다는 점에서 중요하다.

3) 실무적 측면

본 연구의 결과는 다음과 같은 방식으로 실무에 활용될 수 있다. 첫째, 산소포화도 측정 빈도, 도착 경로, 초기 KTAS 수준과 같은 변수를 활용한 실시간 분석을 통해, 중환자실 입원 또는 사망 위험이 높은 고위험 환자를 신속히 식별할 수 있다. 이를 통해, 의료진은 조기에 중재를 실시하여 환자의 상태 악화를 예방하고, 치료 결과를 개선할 수 있다. 둘째, 예측 모델에서 중요한 변수로 도출된 결과를 기반으로 병실 배치 및 의료 자원 관리를 최적화할 수 있다. 입원이 필요한 환자를 보다 정확히 예측함으로써 병실 점유율을 효율적으로 조정하고, 의료 자원을 효과적으로 활용할 수 있다. 셋째, 귀가 가능성이 높은 환자를 조기에 식별하고 신속히 처치함으로써 대기 시간을 줄이고 응급실 혼잡을 완화할 수 있다. 이를 통해 환자의 만족도를 향상시키고, 응급실 운영 효율성을 높일 수 있다.

더 나아가, 본 연구는 임상 의사결정 지원 시스템 개발을 위한 기초자료와 근거로 활용될 수 있을 것이다. 본 연구에서 제안된 예측 모델과 중요한 변수들은 임상 의사결정 지원 시스템 설계 시 실무적 요구를 반영할 수 있는 데이터를 제공하며, 이를 통해 고위험 환자 식별, 자원 배분, 그리고 조기 퇴원 가능 환자 예측 등의 기능을 지원하는 시스템 개발에 유용하게 적용될 수 있다.

VII. 결 론

1. 결론

본 연구는 KTAS 3, 4단계에 해당하는 응급 환자의 임상 결과를 예측하기 위해 triage 정보와 간호 데이터를 활용한 예측 모델을 개발하고 성능을 평가하였다. 연구 결과, 간호 데이터를 포함한 모델 2는 triage 정보만을 기반으로 한 모델 1에 비해 전반적으로 높은 예측 성능을 보였으며, 특히 귀가와 일반 병실 입원과 같은 다수 클래스에서 현저한 성능 향상을 확인할 수 있었다.

간호 데이터를 포함함으로써 중요한 예측 변수로 산소포화도 측정 빈도, 내원 수단, 과거 입원 횟수 등이 도출되었으며, 이는 응급 환자의 상태를 세밀히 반영하고 임상적 의사결정을 지원하는 데 중요한 역할을 했다. 또한, SHAP 분석을 통해 모델의 해석 가능성을 높이고, 예측 결과에 대한 신뢰성을 강화하였다. 이를 통해 간호 데이터가 단순히 기록에 그치지 않고, 응급 환자의 임상 결과를 예측하고 의료 자원을 효율적으로 관리하는 데 핵심적인 데이터로 활용될 수 있음을 보여주었다.

결론적으로, 본 연구는 triage 정보와 간호 데이터를 통합한 다중 클래스 예측 모델이 KTAS 3, 4단계 환자의 임상 결과 예측에서 실질적이고 임상적으로 활용 가능함을 보여주었다.

2. 제언

본 연구는 소수 클래스(중환자실 입원, 사망)에서 낮은 예측 성능이라는 한계를 극복하기 위해 SMOTE 기반 데이터 증강과 비용 민감 학습을 적용하였다. 이러한 접근은 소수 클래스에서 성능 개선의 가능성을 확인하는 데 기여했으나, 일부 클래스에서는 여전히 예측 성능의 한계가 존재하였다. 향후 연구에서는 이러한 방법론을 더욱 고도화하거나 데이터 불균형 문제를 완화할 수 있는 추가적인 접근법을 도입하여 소수 클래스의 예측 성능을 더욱 강화할 필요가 있다.

또한, 주증상 데이터와 같은 비정형 데이터를 포함한 다양한 데이터의 체계적인 수집 및 관리 방안이 필요하다. 특히 간호 데이터를 임상적으로 유용하고 가치 있게 활용하기 위해 간호기록 체계의 표준화와 데이터 품질 관리에 대한 추가적인 노력이 요구되며, 이를 위한 기술적, 제도적 지원도 필요하다.

연구 결과의 일반화를 위해서는 다기관 연구를 통해 데이터의 다양성을 확보하고, 모델의 타당성과 신뢰성을 검증해야 한다. 특히, 다양한 의료 환경에서의 데이터 검증과 추후 연구를 통해 예측 모델의 일반화 가능성을 확인하고, 실질적 활용 가능성을 높일 필요가 있다.

마지막으로, 본 연구 결과를 바탕으로 임상 의사결정 지원 시스템을 개발하기 위한 추가 연구가 이루어져야 한다. 임상 의사결정 지원 시스템은 응급실에서 의료진이 환자의 상태를 신속하고 정확하게 평가하며, 자원을 효율적으로 배분할 수 있는 중요한 도구로 활용될 수 있다. 예를 들어, 귀가 가능성이 높은 환자를 조기에 식별하거나, 중환자실 입원이 필요한 환자를 신속히 경고하는 기능을 도입함으로써, 응급실의 혼잡을 완화하고 운영의 효율성을 향상시킬 수 있을 것이다.

참고 문헌

- Araf, M. (2024). Cost-sensitive learning for imbalanced medical data: A systematic review. *Artificial Intelligence Review*, 57(80). <https://doi.org/10.1007/s10462-023-10652-8>
- Araz, O. M., Olson, D., & Ramirez-Nafarrate, A. (2019). Predictive analytics for hospital admissions from the emergency department using triage information. *International Journal of Production Economics*, 208, 199–207. <https://doi.org/10.1016/j.ijpe.2018.11.024>
- Aziz, W., Nicalaou, A., Stylianides, C., Panayides, A., Kakas, A., Kyriacou, E., & Pattichis, C. (2024). Emergency department length of stay classification based on ensemble methods and rule extraction. *Studies in Health Technology and Informatics*, 316, 1812–1816. <https://doi.org/10.3233/SHTI240783>
- Barnett, V., & Lewis, T. (1984). *Outliers in statistical data*: Wiley.
- Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5-32. <https://doi.org/10.1023/a:1010933404324>
- Brink, A., Alsma, J., van Attekum, L., Bramer, W. M., Zietse, R., Lingsma, H., & Schuit, S. C. E. (2022). Predicting in-hospital admission at the emergency department: A systematic review. *Emergency Medicine Journal*, 39(3), 191-198. <https://doi.org/10.1136/emmermed-2020-210902>
- Cameron, A., Rodgers, K., Ireland, A., Jamdar, R., & McKay, G. A. (2015). A simple tool to predict admission at the time of triage. *Emergency Medicine Journal*, 32(3), 174-179. <https://doi.org/10.1136/emmermed-2013-203200>
- Chang, H., Yu, J. Y., Lee, G. H., Heo, S., Lee, S. U., Hwang, S. Y., ... Kim, T. (2023). Clinical support system for triage based on federated learning for the Korea triage and acuity scale. *Heliyon*, 9(8), e19210. <https://doi.org/10.1016/j.heliyon.2023.e19210>
- Chen, M. C., Huang, T. Y., Chen, T. Y., Boonyarat, P., & Chang, Y. C. (2023). Clinical narrative-aware deep neural network for emergency department critical outcome prediction. *Journal*

- of Biomedical Informatics*, 138, 104284. <https://doi.org/10.1016/j.jbi.2023.104284>
- Choi, S. W., Ko, T., Hong, K. J., & Kim, K. H. (2019). Machine learning-based prediction of Korean triage and acuity scale level in emergency department patients. *Healthcare Informatics Research*, 25(4), 305-312. <https://doi.org/10.4258/hir.2019.25.4.305>
- Collins, G. S., Moons, K. G. M., Dhiman, P., Riley, R. D., Beam, A. L., Van Calster, B., ... Logullo, P. (2024). TRIPOD+AI statement: Updated guidance for reporting clinical prediction models that use regression or machine learning methods. *BMJ*, 385, e078378. <https://doi.org/10.1136/bmj-2023-078378>
- Collins, S. A., Cato, K., Albers, D., Scott, K., Stetson, P. D., Bakken, S., & Vawdrey, D. K. (2013). Relationship between nursing documentation and patients' mortality. *American Journal of Critical Care*, 22(4), 306-313. <https://doi.org/10.4037/ajcc2013426>
- Collins, S. A., & Vawdrey, D. K. (2012). "Reading between the lines" of flow sheet data: Nurses' optional documentation associated with cardiac arrest outcomes. *Applied Nursing Research*, 25(4), 251-257. <https://doi.org/10.1016/j.apnr.2011.06.002>
- Dinh, M. M., Russell, S. B., Bein, K. J., Rogers, K., Muscatello, D., Paoloni, R., . . . Ivers, R. (2016). The Sydney triage to admission risk tool (START) to predict emergency department disposition: A derivation and internal validation study using retrospective state-wide data from New South Wales, Australia. *BMC Emergency Medicine*, 16(1), 46. <https://doi.org/10.1186/s12873-016-0111-4>
- Ekins, K., & Morphet, J. (2015). The accuracy and consistency of rural, remote and outpost triage nurse decision making in one Western Australia Country Health Service Region. *Australasian Emergency Nursing Journal*, 18(4), 227-233. <https://doi.org/10.1016/j.aenj.2015.05.002>
- Fernandes, M., Mendes, R., Vieira, S. M., Leite, F., Palos, C., Johnson, A., Finkelstein, S., Horng, S., & Celi, L. A. (2020). Risk of mortality and cardiopulmonary arrest in critical patients presenting to the emergency department using machine learning and natural language

- p processing.
- PLoS One*
- , 15(4), e0230876.
- <https://doi.org/10.1371/journal.pone.0230876>
- Fernández, A., García, S., Galar, M., Prati, R. C., Krawczyk, B., & Herrera, F. (2018). Imbalanced classification with multiple classes. *Learning from imbalanced data sets* (pp. 197–221). Springer. <https://doi.org/10.1007/978-3-319-98074-4>
- Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *The Annals of Statistics*, 29(5), 1189-1232. <https://doi.org/10.1214/aos/1013203451>
- Göransson, K. E., Ehrenberg, A., & Ehnfors, M. (2005). Triage in emergency departments: National survey. *Journal of Clinical Nursing*, 14(9), 1067-1074. <https://doi.org/10.1111/j.1365-2702.2005.01191.x>
- Graham, B., Bond, R., Quinn, M., & Mulvenna, M. (2018). Using data mining to predict hospital admissions from the emergency department. *IEEE Access*, 6, 10458–10469. <https://doi.org/10.1109/ACCESS.2018.2808843>
- Handly, N. T., Li, D. A., Chuirazzi, J., & Venkat, D. M. (2015). Evaluation of a hospital admission prediction model adding coded chief complaint data using neural network methodology. *European Journal of Emergency Medicine*, 22(2), 87-91. <https://doi.org/10.1097/mej.0000000000000126>
- Häyrinen, K., Saranto, K., & Nykänen, P. (2008). Definition, structure, content, use and impacts of electronic health records: A review of the research literature. *International Journal of Medical Informatics*, 77(5), 291-304. <https://doi.org/10.1016/j.ijmedinf.2007.09.001>
- HIRAID Research Group. (2021). The implementation of an emergency nursing framework (HIRAID) reduces patient deterioration: A multi-centre quasi-experimental study. *International Emergency Nursing*, 56, 100976. <https://doi.org/10.1016/j.ienj.2021.100976>
- Hong, W. S., Haimovich, A. D., & Taylor, R. A. (2018). Predicting hospital admission at emergency department triage using machine learning. *PLoS One*, 13(7), e0201016. <https://doi.org/10.1371/journal.pone.0201016>
- Hosmer, D., Lemeshow, S., & Sturdivant, R. (2013). *Applied Logistic Regression* (3rd ed.). Wiley.

- Imbalanced-learn Development Team. (n.d.). *Imbalanced-learn: A Python toolbox to tackle the curse of imbalanced datasets in machine learning*. Imbalanced-learn. <https://imbalanced-learn.org/stable/>
- Joseph, J. W., Leventhal, E. L., Grossestreuer, A. V., Wong, M. L., Joseph, L. J., Nathanson, L. A., Donnino, M. W., Elhadad, N., & Sanchez, L. D. (2020). Deep-learning approaches to identify critically ill patients at emergency department triage using limited information. *Journal of the American College of Emergency Physicians Open*, 1(5), 773–781. <https://doi.org/10.1002/emp2.12218>
- Kang, M. J., Dykes, P. C., Korach, T. Z., Zhou, L., Schnock, K. O., Thate, J., ... Rossetti, S. C. (2020). Identifying nurses' concern concepts about patient deterioration using a standard nursing terminology. *International Journal of Medical Informatics*, 133, 104016. <https://doi.org/10.1016/j.ijmedinf.2019.104016>
- Ketseridou, S., Logaras, E., Billis, A., Zouka, M., Fyntanidou, B., & Bamidis, P. (2021). AI-driven prediction for the disposition of medium-risk incidents visiting emergency departments. *25th Pan-Hellenic Conference on Informatics*. <https://doi.org/10.1145/3503823.3503912>
- Kim, D., Oh, J., Im, H., Yoon, M., Park, J., & Lee, J. (2021). Automatic classification of the Korean triage acuity scale in simulated emergency rooms using speech recognition and natural language processing: A proof of concept study. *Journal of Korean Medical Science*, 36(27), e175. <https://doi.org/10.3346/jkms.2021.36.e175>
- Kim, M., Park, S., Kim, C., & Choi, M. (2023). Diagnostic accuracy of clinical outcome prediction using nursing data in intensive care patients: A systematic review. *International Journal of Nursing Studies*, 138, 104411. <https://doi.org/10.1016/j.ijnurstu.2022.104411>
- Kim, S. W., Li, J. Y., Hakendorf, P., Teubner, D. J., Ben-Tovim, D. I., & Thompson, C. H. (2014). Predicting admission of patients by their presentation to the emergency department. *Emergency Medicine Australasia*, 26(4), 361-367. <https://doi.org/10.1111/1742-6723.12252>

- Klug, M., Barash, Y., Bechler, S., Resheff, Y. S., Tron, T., Ironi, A., Soffer, S., Zimlichman, E., & Klang, E. (2020). A gradient boosting machine learning model for predicting early mortality in the emergency department triage: Devising a nine-point triage score. *Journal of General Internal Medicine*, 35(1), 220–227. <https://doi.org/10.1007/s11606-019-05512-7>
- Kwon, J. M., Lee, Y., Lee, Y., Lee, S., Park, H., & Park, J. (2018). Validation of deep-learning-based triage and acuity score using a large national dataset. *PLoS One*, 13(10), e0205836. <https://doi.org/10.1371/journal.pone.0205836>
- Larburu, N., Azkue, L., & Kerexeta, J. (2023). Predicting hospital ward admission from the emergency department: A systematic review. *Journal of Personalized Medicine*, 13(5). <https://doi.org/10.3390/jpm13050849>
- Lee, J. H., Park, Y. S., Park, I. C., Lee, H. S., Kim, J. H., Park, J. M., . . . Kim, M. J. (2019). Over-triage occurs when considering the patient's pain in Korean Triage and Acuity Scale (KTAS). *PLoS One*, 14(5), e0216519. <https://doi.org/10.1371/journal.pone.0216519>
- Lee, J. T., Hsieh, C. C., Lin, C. H., Lin, Y. J., & Kao, C. Y. (2021). Prediction of hospitalization using artificial intelligence for urgent patients in the emergency department. *Scientific Reports*, 11(1), 19472. <https://doi.org/10.1038/s41598-021-98961-2>
- Lee, S. B., Kim, D. H., Kim, T., Kang, C., Lee, S. H., Jeong, J. H., Kim, S. C., Park, Y. J., & Lim, D. (2020a). Emergency department triage early warning score (TREWS) predicts in-hospital mortality in the emergency department. *American Journal of Emergency Medicine*, 38(2), 203–210. <https://doi.org/10.1016/j.ajem.2019.02.004>
- Lee, S. Y., Chinnam, R. B., Dalkiran, E., Krupp, S., & Nauss, M. (2020b). Prediction of emergency department patient disposition decision for proactive resource allocation for admission. *Health Care Management Science*, 23(3), 339–359. <https://doi-org/10.1007/s10729-019-09496-y>
- Levin, S., Toerper, M., Hamrock, E., Hinson, J. S., Barnes, S., Gardner, H., Dugas, A., Linton, B.,

- Kirsch, T., & Kelen, G. (2018). Machine-learning-based electronic triage more accurately differentiates patients with respect to clinical outcomes compared with the emergency severity index. *Annals of Emergency Medicine*, 71(5), 565–574.e2. <https://doi.org/10.1016/j.annemergmed.2017.08.005>
- Little, R. J., & Rubin, D. B. (2019). *Statistical analysis with missing data (Vol. 793)*. John Wiley & Sons.
- López, V., Fernández, A., García, S., Palade, V., & Herrera, F. (2013). An insight into classification with imbalanced data: Empirical results and current trends on using data intrinsic characteristics. *Information Sciences*, 250, 113-141. <https://doi.org/10.1016/j.ins.2013.07.007>
- Lucke, J. A., de Gelder, J., Clarijs, F., Heringhaus, C., de Craen, A. J. M., Fogteloo, A. J., Blauw, G. J., Groot, B., & Mooijaart, S. P. (2018). Early prediction of hospital admission for emergency department patients: A comparison between patients younger or older than 70 years. *Emergency Medicine Journal*, 35(1), 18–27. <https://doi.org/10.1136/emmermed-2016-205846>
- Munroe, B., Curtis, K., Murphy, M., Strachan, L., & Buckley, T. (2015). HIRAID: An evidence-informed emergency nursing assessment framework. *Australasian Emergency Nursing Journal*, 18(2), 83–97. <https://doi.org/10.1016/j.aenj.2015.02.001>
- Munroe, B., Curtis, K., Fry, M., Shaban, R. Z., Moules, P., Elphick, T. L., Ruperto, K., Couttie, T., & Considine, J. (2022). Increasing accuracy in documentation through the application of a structured emergency nursing framework: A multisite quasi-experimental study. *Journal of Clinical Nursing*, 31(19–20), 2874–2885. <https://doi.org/10.1111/jocn.16115>
- Park, J. B., Choi, H. J., Kang, B. S., Kim, C. S., Kang, H. G., & Lim, T. H. (2014). Administration ; A nationwide survey of Korean emergency department triage systems and scales; A first step towards reform of the emergency medical service system. *Journal of the Korean Society of Emergency Medicine*, 25(5), 499-508.

- Parker, C. A., Liu, N., Wu, S. X. Z., Shen, Y. Z., Lam, S. S. W., & Ong, M. E. H. (2019). Predicting hospital admission at the emergency department triage: A novel prediction model. *American Journal of Emergency Medicine*, 37(8), 1498-1504.
<https://doi.org/10.1016/j.ajem.2018.10.060>
- Patel, V. L., Gutnik, L. A., Karlin, D. R., & Pusic, M. (2008). Calibrating urgency: triage decision-making in a pediatric emergency department. *Advances in Health Sciences Education: Theory and Practice*, 13(4), 503-520. <https://doi.org/10.1007/s10459-007-9062-6>
- Raita, Y., Goto, T., Faridi, M. K., Brown, D. F. M., Camargo, C. A., Jr., & Hasegawa, K. (2019). Emergency department triage prediction of clinical outcomes using machine learning models. *Critical Care*, 23(1), 64. <https://doi.org/10.1186/s13054-019-2351-7>
- Riley, R. D., Ensor, J., Snell, K. I. E., Harrell, F. E., Jr., Martin, G. P., Reitsma, J. B., . . . van Smeden, M. (2020). Calculating the sample size required for developing a clinical prediction model. *BMJ*, 368, m441. <https://doi.org/10.1136/bmj.m441>
- Riordan, J. P., Dell, W. L., & Patrie, J. T. (2017). Can patient variables measured on arrival to the emergency department predict disposition in medium-acuity patients? *The Journal of Emergency Medicine*, 52(5), 769–779. <https://doi.org/10.1016/j.jemermed.2016.11.018>
- Rossetti, S. C., Dykes, P. C., Knaplund, C., Cho, S., Withall, J., Lowenthal, G., ... Cato, K. D. (2024). Multisite pragmatic cluster-randomized controlled trial of the CONCERN early warning system. *medRxiv*. <https://doi.org/10.1101/2024.06.04.24308436>
- Rossetti, S. C., Knaplund, C., Albers, D., Dykes, P. C., Kang, M. J., Korach, T. Z., . . . Cato, K. (2021). Healthcare process modeling to phenotype clinician behaviors for exploiting the signal gain of clinical expertise (HPM-ExpertSignals): Development and evaluation of a conceptual framework. *Journal of the American Medical Informatics Association*, 28(6), 1242-1251. <https://doi.org/10.1093/jamia/ocab006>
- Rossetti, S. C., Knaplund, C., Albers, D., Tariq, A., Tang, K., Vawdrey, D., Yip, N. H., Dykes, P. C., Klann, J. G., Kang, M. J., Garcia, J., Fu, L. H., Schnock, K., & Cato, K. (2020). Leveraging

- clinical expertise as a feature—not an outcome—of predictive models: Evaluation of an early warning system use case. *AMIA Annual Symposium Proceedings*, 2019, 323–332.
- Ruger, J. P., Lewis, L. M., & Richter, C. J. (2007). Identifying high-risk patients for triage and resource allocation in the ED. *American Journal of Emergency Medicine*, 25(7), 794–798. <https://doi.org/10.1016/j.ajem.2007.01.014>
- Salgado, C. M., Azevedo, C., Proenca, H., & Vieira, S. M. (2016a). Missing data. In M. I. T. C. Data(Ed.), *Secondary Analysis of Electronic Health Records* (pp. 143–162). Cham (CH): Springer International Publishing. https://doi.org/10.1007/978-3-319-43742-2_13
- Salgado, C. M., Azevedo, C., Proenca, H., & Vieira, S. M. (2016b). Noise versus outliers. In M. I. T. C. Data (Ed.), *Secondary Analysis of Electronic Health Records* (pp. 163–183). Cham (CH): Springer International Publishing. https://doi.org/10.1007/978-3-319-43742-2_14
- Sánchez-Salmerón, R., Gómez-Urquiza, J. L., Albendín-García, L., Correa-Rodríguez, M., Martos-Cabrera, M. B., Velando-Soriano, A., & Suleiman-Martos, N. (2022). Machine learning methods applied to triage in emergency services: A systematic review. *International Emergency Nursing*, 60, 101109. <https://doi.org/10.1016/j.ienj.2021.101109>
- Steyerberg, E. W. (2009). *Clinical prediction models*. Springer.
- Sun, Y., Heng, B. H., Tay, S. Y., & Seow, E. (2011). Predicting hospital admissions at emergency department triage using routine administrative data. *Academic Emergency Medicine*, 18(8), 844–850. <https://doi.org/10.1111/j.1553-2712.2011.01125.x>
- Teubner, D. J., Considine, J., Hakendorf, P., Kim, S., & Bersten, A. D. (2015). Model to predict inpatient mortality from information gathered at presentation to an emergency department: The Triage Information Mortality Model (TIMM). *Emergency Medicine Australasia*, 27(4), 300–306. <https://doi.org/10.1111/1742-6723.12425>
- Van Hoorde, K., Van Huffel, S., Timmerman, D., Bourne, T., & Van Calster, B. (2015). A spline-based tool to assess and visualize the calibration of multiclass risk predictions. *Journal of Biomedical Informatics*, 54, 283–293. <https://doi-org/10.1016/j.jbi.2014.12.016>

- Xie, F., Ong, M. E. H., Liew, J. N. M. H., Tan, K. B. K., Ho, A. F. W., Nadarajan, G. D., Low, L. L., Kwan, Y. H., Goldstein, B. A., Matchar, D. B., Chakraborty, B., & Liu, N. (2021). Development and assessment of an interpretable machine learning triage tool for estimating mortality after emergency admissions. *JAMA Network Open*, 4(8), e2118467. <https://doi.org/10.1001/jamanetworkopen.2021.18467>
- Yu, J. Y., Jeong, G. Y., Jeong, O. S., Chang, D. K., & Cha, W. C. (2020). Machine learning and initial nursing assessment-based triage system for emergency department. *Healthcare Informatics Research*, 26(1), 13-19. <https://doi.org/10.4258/hir.2020.26.1.13>
- Yun, H., Choi, J., & Park, J. H. (2021). Prediction of critical care outcome for adult patients presenting to emergency department using initial triage information: An XGBoost algorithm analysis. *JMIR Medical Informatics*, 9(9), e30770. <https://doi.org/10.2196/30770>
- Zachariasse, J. M., van der Hagen, V., Seiger, N., Mackway-Jones, K., van Veen, M., & Moll, H. A. (2019). Performance of triage systems in emergency care: A systematic review and meta-analysis. *BMJ Open*, 9(5), e026471. <https://doi.org/10.1136/bmjopen-2018-026471>
- Zhang, X. Y. K., Patzer, J., Pitts, R. E., Patzer, A., & Schrager, 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. <https://doi.org/10.3414/me17-01-0024>
- Zlotnik, A., Alfaro, M. C., Pérez, M. C. P., Gallardo-Antolín, A., & Martínez, J. M. M. (2016). Building a decision support system for inpatient admission prediction with the Manchester triage system and administrative check-in variables. *CIN: Computers, Informatics, Nursing*, 34(5), 224-230.
- 대한응급의학회 KTAS 위원회. (2021). *KTAS 한국형 응급환자 분류도구 매뉴얼 (개정판)*: 대한응급의학회.
- 백승연, 조은희, 김영아, & 최모나. (2019). 전자간호기록을 이용한 후향적 응급실 간호업무 분석. *성인간호학회지*, 31(5), 496-506.

<https://doi.org/10.7475/kjan.2019.31.5.496>

응급의료통계연보. (2023). *응급의료통계연보*. 서울: 보건복지부

이은실. (2019). *한국형 응급환자 분류도구(KTAS) 재평가 특성 분석*. 가천대학교 대학원, 인천. Retrieved from <https://www.riss.kr/link?id=T15374975> (국내석사학위논문)

이효진. (2022). *COVID-19 팬데믹 이후 응급실 임종 환자의 전자간호기록 분석*. 연세대학교 일반대학원, 서울. Retrieved from <https://www.riss.kr/link?id=T16069835> (국내석사학위논문)

APPENDICES

Appendix 1. TRIPOD+AI Checklist: Model development and validation

Section/Topic	Item	Development /evaluation	Checklist item	Reported on page
TITLE				
<i>Title</i>	1	D;E	Identify the study as developing or evaluating the performance of a multivariable prediction model, the target population, and the outcome to be predicted	vi
ABSTRACT				
<i>Abstract</i>	2	D;E	See TRIPOD+AI for Abstracts checklist	vi
INTRODUCTION				
<i>Background</i>	3a	D;E	Explain the healthcare context (including whether diagnostic or prognostic) and rationale for developing or evaluating the prediction model, including references to existing models	1-3
	3b	D;E	Describe the target population and the intended purpose of the prediction model in the context of the care pathway, including its intended users (e.g., healthcare professionals, patients, public)	1-3
	3c	D;E	Describe any known health inequalities between sociodemographic groups	-
<i>Objectives</i>	4	D;E	Specify the study objectives, including whether the study describes the development or validation of a prediction model (or both)	4
METHODS				
<i>Data</i>	5a	D;E	Describe the sources of data separately for the development and evaluation datasets (e.g., randomised trial, cohort, routine care or registry data), the rationale for using these data, and representativeness of the data	30
	5b	D;E	Specify the dates of the collected participant data, including start and end of participant accrual; and, if applicable, end of follow-up	31
<i>Participants</i>	6a	D;E	Specify key elements of the study setting (e.g., primary care, secondary care, general population) including the number and location of centres	30
	6b	D;E	Describe the eligibility criteria for study participants	31
	6c	D;E	Give details of any treatments received, and how they were handled during model development or evaluation, if relevant	-
<i>Data preparation</i>	7	D;E	Describe any data pre-processing and quality checking, including whether this was similar across relevant sociodemographic groups	37-39
<i>Outcome</i>	8a	D;E	Clearly define the outcome that is being predicted and the time horizon, including how and when assessed, the rationale for choosing this outcome, and whether the method of outcome assessment is consistent across sociodemographic groups	32
	8b	D;E	If outcome assessment requires subjective interpretation, describe the qualifications and demographic characteristics of the outcome assessors	-
	8c	D;E	Report any actions to blind assessment of the outcome to be predicted	-
<i>Predictors</i>	9a	D	Describe the choice of initial predictors (e.g., literature, previous models, all available predictors) and any pre-selection of predictors before model building	32
	9b	D;E	Clearly define all predictors, including how and when they were measured (and any actions to blind assessment of predictors for the outcome and other predictors)	35-36
	9c	D;E	If predictor measurement requires subjective interpretation, describe the qualifications and demographic characteristics of the predictor assessors	-
<i>Sample size</i>	10	D;E	Explain how the study size was arrived at (separately for development and evaluation), and justify that the study size was sufficient to answer the research question. Include details of any sample size calculation	34
<i>Missing data</i>	11	D;E	Describe how missing data were handled. Provide reasons for omitting any data	38

Appendix 1. TRIPOD+AI Checklist: Model development and validation (Continue)

METHODS				
<i>Analytical methods</i>	12a	D	Describe how the data were used (e.g., for development and evaluation of model performance) in the analysis, including whether the data were partitioned, considering any sample size requirements	40-42
	12b	D	Depending on the type of model, describe how predictors were handled in the analyses (functional form, rescaling, transformation, or any standardisation).	43-44
	12c	D	Specify the type of model, rationale ² , all model-building steps, including any hyperparameter tuning, and method for internal validation	43-44
	12d	D;E	Describe if and how any heterogeneity in estimates of model parameter values and model performance was handled and quantified across clusters (e.g., hospitals, countries). See TRIPOD-Cluster for additional considerations	-
	12e	D;E	12e D;E Specify all measures and plots used (and their rationale) to evaluate model performance (e.g., discrimination, calibration, clinical utility) and, if relevant, to compare multiple models	44
	12f	E	Describe any model updating (e.g., recalibration) arising from the model evaluation, either overall or for particular sociodemographic groups or settings	-
	12g	E	For model evaluation, describe how the model predictions were calculated (e.g., formula, code, object, application programming interface)	-
<i>Class imbalance</i>	13	D;E	If class imbalance methods were used, state why and how this was done, and any subsequent methods to recalibrate the model or the model predictions	39
<i>Fairness</i>	14	D;E	Describe any approaches that were used to address model fairness and their rationale	-
<i>Model output</i>	15	D	Specify the output of the prediction model (e.g., probabilities, classification). Provide details and rationale for any classification and how the thresholds were identified	40-41
<i>Training versus evaluation</i>	16	D;E	Identify any differences between the development and evaluation data in healthcare setting, eligibility criteria, outcome, and predictors	-
<i>Ethical approval</i>	17	D;E	Name the institutional research board or ethics committee that approved the study and describe the participant-informed consent or the ethics committee waiver of informed consent	45
OPEN SCIENCE				
<i>Funding</i>	18a	D;E	Give the source of funding and the role of the funders for the present study	-
<i>Conflicts of interest</i>	18b	D;E	Declare any conflicts of interest and financial disclosures for all authors	-
<i>Protocol</i>	18c	D;E	Indicate where the study protocol can be accessed or state that a protocol was not prepared	-
<i>Registration</i>	18d	D;E	Provide registration information for the study, including register name and registration number, or state that the study was not registered	-
<i>Data sharing</i>	18e	D;E	Provide details of the availability of the study data	30
<i>Code sharing</i>	18f	D;E	Provide details of the availability of the analytical code	-
PATIENT & PUBLIC INVOLVEMENT				
<i>Patient & Public Involvement</i>	19	D;E	Provide details of any patient and public involvement during the design, conduct, reporting, interpretation, or dissemination of the study or state no involvement.	-

Appendix 1. TRIPOD+AI Checklist: Model development and validation (Continue)

RESULTS				
<i>Participants</i>	20a	D;E	Describe the flow of participants through the study, including the number of participants with and without the outcome and, if applicable, a summary of the follow-up time. A diagram may be helpful.	46-47
	20b	D;E	Report the characteristics overall and, where applicable, for each data source or setting, including the key dates, key predictors (including demographics), treatments received, sample size, number of outcome events, follow-up time, and amount of missing data. A table may be helpful. Report any differences across key demographic groups.	50-54
	20c	E	For model evaluation, show a comparison with the development data of the distribution of important predictors (demographics, predictors, and outcome).	-
<i>Model development</i>	21	D;E	Specify the number of participants and outcome events in each analysis (e.g., for model development, hyperparameter tuning, model evaluation)	46-48
<i>Model specification</i>	22	D	Provide details of the full prediction model (e.g., formula, code, object, application programming interface) to allow predictions in new individuals and to enable third-party evaluation and implementation, including any restrictions to access or re-use (e.g., freely available, proprietary)	-
<i>Model performance</i>	23a	D;E	Report model performance estimates with confidence intervals, including for any key subgroups (e.g., sociodemographic). Consider plots to aid presentation.	60
	23b	D;E	If examined, report results of any heterogeneity in model performance across clusters. See TRIPOD Cluster for additional details.	58-63
<i>Model updating</i>	24	E	Report the results from any model updating, including the updated model and subsequent performance	-
DISCUSSION				
<i>Interpretation</i>	25	D;E	Give an overall interpretation of the main results, including issues of fairness in the context of the objectives and previous studies	77-82
<i>Limitations</i>	26	D;E	Discuss any limitations of the study (such as a non-representative sample, sample size, overfitting, missing data) and their effects on any biases, statistical uncertainty, and generalizability	82-84
<i>Usability of the model in the context of current care</i>	27a	D	Describe how poor quality or unavailable input data (e.g., predictor values) should be assessed and handled when implementing the prediction model	-
	27b	D	Specify whether users will be required to interact in the handling of the input data or use of the model, and what level of expertise is required of users	-
	27c	D;E	Discuss any next steps for future research, with a specific view to applicability and generalizability of the model	84-86

Note. D=items relevant only to the development of a prediction model, E=items relating solely to the evaluation of a prediction model, D;E=items applicable to both the development and evaluation of a prediction model

Appendix 2. Data definition document

Name of variable		Definition	CDW table	CDW data
성별		환자의 성별	간호초기평가-응급	성별
나이		내원일시 기준 연령	간호초기평가-응급	내원일시 (연령)
1 년 이내 입원		내원일시 기준 365 일 이내 입원 횟수	입원정보	입원일자 (-365 일~0 일) & 횟수
1 달 이내 응급실 내원		내원일시 기준 31 일 이내 응급실 내원 횟수	응급정보	내원일시 (-31 일~0 일) & 횟수
내원 시간대		내원일시 기준 3 가지 시간대로 구분 Day(08:00~16:00), Evening(16:00~24:00), Night(00:00~08:00)	간호초기평가-응급	응급실내원일시 변환 함수
내원 요일		내원일시 기준 주중(월~금), 주말(토, 일)로 구분	간호초기평가-응급	응급실내원일시 변환 함수
내원경로		Ambulance(119 구급차, 기타구급차, 의료기관 구급차), Others(기타자동차, 도보, 경찰차등 공공차량, 항공이송, 기타, 미상) 으로 구분	간호초기평가-응급	내원경로
내원수단		Direct(직접내원), Transfer(전원), From OPD(외래, 검사실) 로 구분	간호초기평가-응급	내원수단
진료대기시간		위치값이 예진 이외의 구역으로 변경된 최초일시와 내원일시 사이의 소요시간을 분단위로 계산한 값	응급환자위치변경이력	위치변경(예진;not 예진)
증상발생~내원시 소요시간		발병일시-내원일시 사이의 시간을 분단위로 계산한 값 또는 발병시기(예, 2 주전)를 분단위로 환산한 값	간호초기평가-응급	발병일시 발병시기
수축기 혈압		내원 시 수축기 혈압 수치	간호초기평가-응급	V/S_SBP
이완기 혈압		내원 시 이완기 혈압 수치	간호초기평가-응급	V/S_DBP
맥박		내원 시 맥박 수치	간호초기평가-응급	V/S_PR
호흡수		내원 시 호흡 수치	간호초기평가-응급	V/S_RR

Name of variable		Definition	CDW table	CDW data
체온	내원 시 체온 수치		간호초기평가-응급	V/S_temp 내용
산소포화도	내원 시 산소포화도 수치		간호초기평가-응급	V/S_SpO2 값
의식상태	내원 시 의식상태를 Alert, Verbal response, Pain response, Unresponsive outcome 으로 구분		간호초기평가-응급	환자반응
주증상 종류	질병분류 상 질병 또는 질병 외로 구분		간호초기평가-응급	질병구분
초기 KTAS 결과	내원 시 초기 평가된 KTAS 결과값		간호초기평가-응급	최초 KTAS 레벨
최종 KTAS 결과	퇴실 전 마지막으로 입력된 KTAS 결과값		간호초기평가-응급	최종 KTAS 레벨
통증여부	내원 시 평가한 통증 유,무 값		Flowsheet 검색항목	항목그룹 (143), 항목 (1329)
통증점수	내원 시 평가한 통증 점수(NRS, FLACC 등) 값		Flowsheet 검색항목	항목그룹 (142), 항목 (375 376 377 378 380 1248 379)
활력징후 측정 빈도(SBP)	내원일시 이후부터 6 시간 이내에 기록된 Flowsheet 상 수축기혈압(NIBP, ABP) 측정 빈도(단, 퇴실일시 이전 기록만 포함)		Flowsheet 검색항목	항목 (166 138)/(136 164)
활력징후 측정 빈도(HR PR)	내원일시 이후부터 6 시간 이내에 기록된 Flowsheet 상 맥박 또는 심박수 측정 빈도(단, 퇴실일시 이전 기록만 포함)		Flowsheet 검색항목	항목 (160 171)
활력징후 측정 빈도(RR)	내원일시 이후부터 6 시간 이내에 기록된 Flowsheet 상 호흡수 측정 빈도(단, 퇴실일시 이전 기록만 포함)		Flowsheet 검색항목	항목 (174)
활력징후 측정 빈도(BT)	내원일시 이후부터 6 시간 이내에 기록된 Flowsheet 상 체온 측정 빈도(단, 퇴실일시 이전 기록만 포함)		Flowsheet 검색항목	항목 (182)
활력징후 측정 빈도(SpO2)	내원일시 이후부터 6 시간 이내에 기록된 Flowsheet 상 산소포화도 측정 빈도(단, 퇴실일시 이전 기록만 포함)		Flowsheet 검색항목	항목 (177)
의식상태 측정 빈도	내원일시 이후부터 6 시간 이내에 기록된 Flowsheet 상 의식상태(AVPU, GCS) 측정 빈도(단, 퇴실일시 이전 기록만 포함)		Flowsheet 검색항목	항목 (2633 1394 88)

Name of variable		Definition	CDW table	CDW data
통증사정 빈도		내원일시 이후부터 6 시간 이내에 기록된 Flowsheet 상 통증사정 측정 빈도(단, 퇴실일시 이전 기록만 포함)	Flowsheet 검색항목	항목그룹(143), 항목(1329)
간호기록 -notifying a doctor		내원일시 이후부터 6 시간 이내에 기록된 간호기록내용 중 다음의 조건에 해당하는 기록의 빈도(단, 퇴실일시 이전 기록만 포함) 간호기록내용 ("noti" or "알림" or "노티" or "notify") AND ("의사" or "주치의" or "전문의" or "Dr" or "pf")	환자간호기록	간호기록내용 (noti 알림 노티 notify) & (의사 주치의 전문의 Dr pf)
간호기록 -abnormal lab results		내원일시 이후부터 6 시간 이내에 기록된 간호기록내용 중 다음의 조건에 해당하는 기록의 빈도(단, 퇴실일시 이전 기록만 포함) 간호기록내용 ("검사일시" or "검사명" or "검사결과") AND not("COVID")	환자간호기록	간호기록내용 (검사일시 검사명 검사결과) & not(COVID)
간호기록 -request for additional tests		내원일시 이후부터 6 시간 이내에 기록된 간호기록내용 중 다음의 조건에 해당하는 기록의 빈도(단, 퇴실일시 이전 기록만 포함) 간호기록내용 ("의뢰" or "요청") AND ("검사" or "ABGA" or "EKG" or "추가" or "인턴" or "DI")	환자간호기록	간호기록내용 (의뢰 요청) & (검사 ABGA EKG 추가 인턴 DI)
치료 단계 상승		응급환자위치변경이력에서 처음 진료구역보다 최종 진료구역의 수준이 더 높은 구역으로 변경된 경우(예: ambulatory zone -> bed zone -> critical care zone -> resuscitation zone)를 구분함.	응급환자위치변경이력	위치변경일시, 구역
flowsheet Event 특이사항		내원일시 이후부터 6 시간 이내에 기록된 Flowsheet 상 Event, Comment 기록의 빈도(단, 퇴실일시 이전 기록만 포함)	Flowsheet 검색항목	항목 (2640 144 145 146 147 2653 1315 3146)
임상결과		응급진료결과(입원, 전원, 사망, 귀가)와 입원구분(일반병실로 입원, 중환자실로 입원)을 조합하여 "일반병실로 입원", "중환자실로 입원", "전원", "사망", "귀가"의 5 개 값으로 구분함.	간호초기평가-응급	응급진료결과 입원구분

Appendix 3. Approval from the institutional review board

DocNo : SMC202407063-HE001

Full Accredited

통지서



QR코드를 읽어 문서 진위여부를 확인할 수 있습니다.

※ 본 과제의 문서보존기간은 3 년입니다.

수신	지원(의뢰)기관	내부과제				
	연구책임자	차원철				
File No.	SMC IRB 2024-07-063	신청유형	신규신청서	통지일자	2024.07.19	
과제명	국문	한국형 응급환자 분류도구(KTAS) 3, 4단계 응급 환자의 임상결과 예측 모델 개발 및 평가				
	영문	Development and Validation of a Predictive Model for Clinical Outcomes in Emergency Patients Classified as KTAS Levels 3 and 4				
Study Nick Name						

연구구분	임상시험 외 연구	
대상자 수	전체	40000명
	국내	40000명
	본원	0명
연구승인기간		
지원(의뢰) 기관	내부과제	
제출서류 목록	1. 연구계획서 2. 응례기록서 3. 연구자서약서 4. CV-차원철, 신현아	
심사일	2024.07.18	심사종류 심사면제
중간보고 시기		
심사결과	● 심사면제에 해당 ○ 심사면제에 해당하지 않음	
내용	<p>본 연구는 KTAS 3, 4단계 응급환자의 특성 및 임상 결과를 확인하고, 임상 결과 예측 모델을 개발하여 타당도를 평가하고자하는 후향적 연구로, 본원 CDW(Clinical Data Warehouse)로부터 연구대상자의 개인식별정보를 제거하여 익명화된 자료만을 제공받아 이용하므로 IRB 심사면제 요건 Category 6에 해당합니다.</p> <p>연구대상자의 동의를 받는 것이 연구 진행과정에서 현실적으로 불가능하고, 연구대상자의 동의 거부를 추정할 만한 사유가 없고, 동의를 면제하여도 연구대상자에게 미치는 위험이 극히 낮은 경우에 해당하므로 동의면제가 가능할 것으로 판단합니다.</p> <ul style="list-style-type: none">본 연구는 지속성의 및 그 외의 사항에 대해 추가로 보고할 의무가 없으며, 심사면제의 허용범위를 벗어나는 연구계획의 변경을 하여서는 안됩니다.심사면제 허용범위를 벗어나는 연구계획의 변경이 필요한 경우 신규심사로 다시 심의를 신청하여야 합니다.공중보건상 긴급한 조치가 필요한 상황에서 국가 또는 지방자치단체가 직접 수행하거나 위탁한 연구인 경우, 연구종료 전 공공기관생명윤리위원회에 연구의 진행상황이 통보되어야 합니다.	
별첨		

삼성서울병원 Institutional Review Board



개정본 2024.07.01

ABSTRACT

Development of a clinical outcome prediction model for emergency patients
classified in level 3 and 4 of the Korean Triage and Acuity Scale (KTAS)
using electronic nursing records

Shin, Hyun A

Department of Nursing

The Graduate School of Yonsei University

Background: Predicting clinical outcomes in emergency department (ED) patients is essential for efficient resource allocation and timely clinical decision-making. Patients categorized as Korean Triage and Acuity Scale (KTAS) levels 3 and 4 often present heterogeneous and unpredictable clinical states, necessitating advanced predictive tools. This study aimed to develop and evaluate a predictive model for clinical outcomes in KTAS level 3 and 4 adult patients by integrating triage information and nursing data using the HPM-ExpertSignals conceptual framework.

Methods: This retrospective study utilized anonymized data from 46,900 KTAS level 3 and 4 adult patients who visited a 2,000-bed tertiary hospital in Seoul, Korea, between January 1 and December 31, 2023. Two predictive models were developed: Model 1 utilized triage information only, while Model 2 integrated both triage information and nursing data. Clinical outcomes included intensive care unit (ICU) admission, general ward (GW) admission, transfer, discharge, and mortality. Three machine learning algorithms—multinomial logistic regression, random forest, and gradient boosting—were employed. Model performance was assessed using accuracy, AUROC, AUPRC,

sensitivity, specificity, F1 score, and calibration curves. To address class imbalance, one-vs-rest SMOTE (Synthetic Minority Over-sampling Technique) and cost-sensitive learning approaches were employed.

Results: Model 2 consistently outperformed Model 1 across all performance metrics. The random forest algorithm applied to Model 2 achieved the highest predictive performance, with an AUROC of 0.964. Key predictive variables included oxygen saturation measurement frequency, mode of ED arrival, and previous admission history. SHAP (Shapley Additive Explanations) analysis further highlighted the relative importance of these variables, enhancing the model's transparency and applicability. Although SMOTE and cost-sensitive learning improved predictive performance for minority classes (ICU admission and mortality), class imbalance remained a limitation.

Conclusion: The integration of nursing data with triage information significantly enhanced the predictive performance and interpretability of the model, highlighting its potential as a clinical decision-support system. The random forest-based model demonstrated robust performance in predicting clinical outcomes for KTAS level 3 and 4 patients. Future research should focus on validating the model using multi-center datasets to improve generalizability and explore innovative methods to address class imbalance in predictive modeling.

Keywords: KTAS, Emergency Department, Predictive Model, Nursing Data, SHAP, Machine Learning, Class Imbalance, Clinical Decision Support

영문요약

Introduction

Emergency department (ED) overcrowding, worsened by the increasing number of critically ill patients, remains a significant challenge. Providing prompt and optimal medical services to all patients in overcrowded EDs is highly challenging. Therefore, early recognition and timely intervention for patients with life-threatening conditions are crucial (Lee et al., 2019).

The process of triage refers to the decision-making procedure used to categorize the severity of a patient's condition and allocate limited medical resources efficiently to provide appropriate emergency care within the "golden hour" for patients presenting with various acute conditions of differing severity (Patel et al., 2008). Triage is predominantly conducted by nurses in both domestic and international settings (Göransson et al., 2005; Park et al., 2014).

In South Korea, the Korean Triage and Acuity Scale (KTAS), developed based on CTAS (Canadian Triage and Acuity Scale), has been the standard triage tool since 2016. KTAS classifies patients into five levels based on severity and urgency, with recommended intervention times ranging from immediate for Level 1 ("resuscitation") to within two hours for Level 5 ("non-urgent"). (Korean Society of Emergency Medicine KTAS Committee, 2021). Although patients in Levels 1 and 2 often receive prompt care, those in Levels 3 and below may experience delayed treatment if under-triaged, potentially leading to worsened

outcomes. Conversely, over-triage can result in unnecessary consumption of human and material resources (Ekins & Morphet, 2015).

According to 2022 statistics, KTAS Level 1 and 2 patients comprised 1.3% and 5.8% of ED visits, respectively, while Level 3 (“urgent”) and Level 4 (“semi-urgent”) patients accounted for 43.4% and 39.4%, representing over 80% of all visits (Emergency Medical Statistics Annual Report, 2023). This group, while crucial for the efficient utilization of medical resources, poses challenges for accurate prediction of clinical outcomes compared to higher acuity patients in Levels 1 and 2.

Multiclass outcome modeling provides detailed predictions and improves resource allocation. While Riordan (2017) reported limited performance with a binary model (AUROC 0.730), Lee et al. (2020) demonstrated the effectiveness of a multiclass approach for predicting diverse clinical outcomes.

Nursing records, particularly those within electronic medical record (EMR) systems, have demonstrated their value in predicting clinical deterioration and mortality (Collins et al., 2013). Nurses frequently document vital signs and unstructured clinical observations in response to patient condition changes, providing critical insights (Collins & Vawdrey, 2012). While numerous studies have developed predictive models for ED patient outcomes, most have focused on triage information or laboratory results, with limited attention to nursing documentation (Brink et al., 2022; Larburu et al., 2023).

This study aims to address these gaps by developing and evaluating predictive models for clinical outcomes in KTAS Levels 3 and 4 patients, incorporating electronic nursing

records alongside traditional triage data. This approach seeks to improve the accuracy of clinical outcome predictions, facilitate efficient resource allocation, and enhance the overall quality of emergency care. Furthermore, by leveraging nursing data as key predictors, this study contributes to the body of evidence supporting the clinical value and scientific foundation of nursing documentation in ED settings.

Conceptual Framework

This study utilizes a modified HPM-ExpertSignals framework to develop and evaluate predictive models for KTAS Level 3 and 4 ED patients, incorporating triage data and nursing records. The framework emphasizes the role of nursing assessments (e.g., vital signs, consciousness, KTAS level) and nursing record patterns (e.g., frequency of vital sign monitoring, interventions like notifying medical staff or documenting abnormalities) in reflecting patient conditions. Additionally, individual factors (e.g., age, prior hospitalizations) and environmental factors (e.g., arrival mode, waiting time) are integrated. By leveraging electronic nursing records, the study aims to enhance predictive accuracy, optimize ED resource allocation, and improve patient outcomes, highlighting the critical value of nursing documentation in clinical decision-making.

Methods

Study Design

This study is a retrospective descriptive analysis aimed at investigating the characteristics and patterns of nursing records for KTAS Level 3 and 4 ED patients and their associations with clinical outcomes.

Data Source

The data for this study were collected from patients who visited the ED of a 2,000-bed tertiary hospital in Seoul, South Korea, between January 1 and December 31, 2023. The ED handles over 60,000 patient visits annually, and all data were extracted from the hospital's Clinical Data Warehouse (CDW) system, DARWIN-C.

Study Population

The study included ED visits by patients aged 18 years or older with an initial KTAS level of 3 or 4, excluding visits with canceled registrations, non-treatment visits, patients who left against medical advice or without being seen, dead on arrival (DOA) cases, and those with missing data. Each ED visit was analyzed as a separate event.

Variables

The potential predictors for clinical outcomes in KTAS Level 3 and 4 patients were identified through a comprehensive literature review and categorized according to the

conceptual framework into triage information, nursing data patterns, individual adjustment factors, and environmental adjustment factors.

Triage information included 11 variables collected during the nursing initial assessment, such as vital signs, level of consciousness, chief complaints, initial KTAS classification, pain status, and pain score. Nursing record patterns comprised 12 variables designed to capture clinical observation and intervention activities. Frequency patterns included the number of recorded instances for vital sign measurements, consciousness assessments, and pain evaluations. Intervention-related factors included documentation of physician notifications, abnormal test results, additional test requests, transfers to higher-acuity areas, and specific notes regarding events in the flowsheet.

Individual adjustment factors included four variables: sex, age, history of hospital admissions within the past year, and ED visits within the past month. Environmental adjustment factors included six variables: time of ED arrival, day of the week, route of ED arrival, mode of arrival, waiting time for medical evaluation, and the time elapsed from symptom onset to ED arrival.

The outcome variable was categorized into five clinical outcomes: general ward admission, ICU admission, transfer to another facility, death, discharge.

Data Construction

Identification of Data Sources for Analysis Variables

Data sources for analysis variables were identified from the CDW. Nursing data were extracted from tables such as Nursing Initial Assessment – Emergency, Flowsheet Search Items, Patient Nursing Records, and Emergency Patient Location Change History.

Development of a Variable Definition Document

The findings from the investigation of data sources for each variable were compiled into a variable definition document. This document systematically outlined the data sources, extractable data fields, and associated codes for each variable.

Extraction of Raw Data

After finalizing the variable definition document and obtaining reviews from at least two experts, the document was further refined with input from CDW specialists. Based on the finalized definitions, the researcher directly extracted the necessary raw data from the CDW according to the defined variables.

Data Preprocessing

To ensure data reliability, preprocessing involved integrating and cleaning the dataset. Steps included outlier detection, handling missing data through imputation or exclusion, and addressing data imbalance using a combination of One-vs-Rest SMOTE (Synthetic Minority Over-sampling Technique) and cost-sensitive learning.

Development and Validation of Predictive Models

As proposed in the conceptual framework, this study developed two models to evaluate the impact of nursing record pattern variables on model performance: one excluding nursing record pattern variables (Model 1) and one including them (Model 2). The performance of these two models was subsequently compared.

To prevent overfitting and ensure model reliability, the dataset was randomly split into a training set (80%) and a test set (20%). The training set was further processed to address class imbalance using a combination of One-vs-Rest SMOTE and cost-sensitive learning. This study employed three machine learning algorithms to develop the predictive models: multinomial logistic regression, random forest, and Extreme Gradient Boosting (XGBoost). The models were trained using stratified 10-fold cross-validation. The performance of the developed models was evaluated through internal validation using the test set. After identifying the optimal model, variable importance was assessed to provide further insights.

Data Analysis

Data preprocessing, model development, and validation were conducted using Python version 3.10.12. Descriptive statistics were presented as means and standard deviations for continuous variables and frequencies with percentages for categorical variables. Relationships between predictor and outcome variables were analyzed using chi-squared tests for categorical variables and the Kruskal-Wallis test for non-normal continuous variables, with Dunn's test and Bonferroni correction applied for post-hoc analysis.

Hyperparameter optimization was performed using GridSearchCV, and model performance was evaluated using metrics such as AUROC, AUPRC, accuracy, sensitivity, specificity, PPV, NPV, and F1 score. Initial variable importance was assessed using the Gini importance method in the Random Forest model, and SHAP (Shapley Additive Explanations) analysis provided a deeper evaluation, with a summary plot visually displaying the relative importance of variables across outcome classes. Calibration plots were used to assess the alignment between predicted probabilities and actual outcomes, ensuring model reliability and interpretability.

Ethical Considerations

This study was conducted with the approval of the Institutional Review Board (IRB) of Samsung Medical Center, with an exemption granted (IRB No. SMC 2024-07-063).

Results

Cohort Selection and Characteristics

Between January 1 and December 31, 2023, a total of 50,251 patients visited the ED, resulting in 71,000 ED visit records. Of these, 24,100 records (33.9%) were excluded based on the inclusion and exclusion criteria, leaving 46,900 records (66.1%) from 33,885 ED patients for the final analysis.

Exclusions based on inclusion criteria included patients under 18 years of age ($n = 10,395$) and patients classified as KTAS Levels 1, 2, or 5 ($n = 5,373$). Exclusions based on

exclusion criteria included canceled registrations (n = 6,132), non-treatment-related visits (n = 1,706), patients who left against medical advice (n = 408), patients who left without being seen (n = 24), and records with missing data (n = 62). The proportion of excluded data due to missing values was less than 0.1%.

The eligible data were randomly allocated, with 80% assigned to the training set (n = 37,520) and 20% to the test set (n = 9,380). The training set was used for model development, while the test set was employed for internal validation of the model.

Characteristics and Clinical Outcomes of KTAS Level 3 and 4 Patients

Analysis of 46,900 ED visits showed that 67.5% of patients were discharged, 26.9% were admitted to a general ward, 1.7% to the ICU, 3.8% were transferred, and 0.1% resulted in death, with an overall hospital admission rate of 28.6%. Significant differences in clinical outcomes were observed: ICU and deceased patients were older, had abnormal vital signs, and higher ED visit rates. General ward admissions had the longest waiting times, while transfers reported the highest pain levels. Most ICU and deceased patients were KTAS Level 3, while discharged patients were evenly distributed between Levels 3 and 4. Severe outcomes were more common in males and ambulance arrivals.

Analysis of Nursing Record Patterns by Clinical Outcomes

The analysis of nursing record patterns revealed significant differences across clinical outcomes. Vital signs, including BP, PR, RR, BT, and SpO₂, were recorded most frequently for deceased patients, while discharged patients had the lowest recording frequencies.

Nursing interventions, such as physician notifications and abnormal test documentation, were also more frequent for deceased patients and ICU admissions. Treatment escalation occurred in 48.3% of deceased patients and 39.9% of general ward admissions, compared to 16.3% of discharged patients. Additionally, flowsheet comments were most frequent for deceased patients, emphasizing the importance of nursing documentation in managing severe clinical outcomes.

Development and Performance Evaluation of Predictive Models

This study evaluated two predictive models: Model 1, which included 21 predictors, and Model 2, which incorporated nursing record data for a total of 33 predictors. Using three machine learning algorithms (multinomial logistic regression, random forest, and gradient boosting), six models were developed and validated. Model 2 consistently outperformed Model 1 across all algorithms, highlighting the positive impact of including nursing data.

For Model 1, random forest achieved the highest overall accuracy (72.4%), followed by gradient boosting (63.3%) and logistic regression (60.7%). However, minority classes such as ICU admissions and transfers showed poor predictive performance across all algorithms, with accuracies below 40%. Model 2 showed substantial improvement, with random forest achieving the highest overall accuracy (79.6%), followed by gradient boosting (74.0%) and logistic regression (68.8%). Notably, Model 2 demonstrated significant gains in minority class predictions, particularly with gradient boosting, which achieved the best performance for ICU admissions (39.6%) and transfers (37.6%).

Comprehensive metrics further confirmed the improved performance of Model 2. Random forest in Model 2 showed the highest AUROC (0.964) and AUPRC (0.888), with balanced sensitivity (0.796) and specificity (0.932). Gradient boosting also demonstrated enhanced performance in Model 2, achieving an AUROC of 0.940, AUPRC of 0.837, and an F1 score of 0.771. Logistic regression in Model 2 showed modest improvements, with an AUROC of 0.891 and an F1 score of 0.720.

Visual analyses of AUROC and precision-recall curves highlighted consistent gains with Model 2. Discharge predictions achieved the highest average precision (0.960 in Model 2 vs. 0.905 in Model 1), while general ward admissions also improved significantly (0.713 in Model 2 vs. 0.601 in Model 1). Although ICU admissions and transfers continued to show lower precision, Model 2 demonstrated incremental improvements, underscoring the value of incorporating nursing data for clinical outcome prediction.

Final Model Selection and Variable Importance

Random forest Model 2, with the highest AUROC (0.964), AUPRC (0.888), sensitivity (0.796), and specificity (0.932), demonstrated superior performance across all metrics, making it the optimal predictive model.

Variable importance analysis using Gini importance identified the most influential predictors in random forest Model 2. The most significant variable was the frequency of oxygen saturation measurements, followed by waiting time, pulse rate, time from symptom onset to arrival, systolic and diastolic blood pressure, age, body temperature, prior hospitalizations, and frequency of blood pressure measurements.

SHAP analysis provided additional insights into the contribution of each variable to the model's predictions. For ICU admissions, key predictors included prior hospitalizations, initial KTAS levels, and frequency of vital sign measurements, reflecting higher monitoring for severe cases. For general ward admissions, oxygen saturation measurement frequency and prior hospitalizations were most critical. Transfers and deaths were influenced by factors such as oxygen saturation, mode of arrival (e.g., ambulance use), and frequency of vital sign monitoring, indicating the importance of initial severity. Conversely, discharges were associated with lower oxygen saturation measurement frequency, arrival on foot, and less need for higher-acuity care. These findings highlight the nuanced role of nursing data and triage variables in predicting diverse clinical outcomes.

Calibration Plot

The calibration plot for random forest Model 2 showed varying alignment with observed probabilities across classes. Discharge and general ward admissions were better calibrated, though discharges showed overestimation in the 0.2–0.8 range. ICU admissions, transfers, and mortality classes demonstrated poor calibration, particularly due to class imbalance and data scarcity, highlighting challenges in predicting minority outcomes.

Discussion

This study developed and validated predictive models for clinical outcomes of KTAS Level 3 and 4 adult patients using triage data and nursing records. Analysis of 46,900 ED

visits revealed a high proportion of discharge cases (67.5%), followed by general ward admissions (26.9%), ICU admissions (1.7%), transfers (3.8%), and deaths (0.1%). The findings emphasized the heterogeneity of outcomes within the same KTAS levels, highlighting the limitations of KTAS in accurately predicting clinical outcomes and underscoring the need for advanced predictive models. Additionally, the study confirmed that abnormal vital signs, such as oxygen saturation levels and frequent monitoring, were prominent in severe cases, aligning with prior research on the significance of nursing observations in early detection of clinical deterioration.

The inclusion of nursing data improved model performance across all algorithms, particularly in Model 2, which demonstrated superior accuracy, AUROC, and AUPRC compared to Model 1. The random forest algorithm in Model 2 achieved the best overall performance, with high sensitivity (0.796) and specificity (0.932), effectively identifying severe cases while minimizing unnecessary classifications. This result supports the value of incorporating nursing records, such as oxygen saturation measurement frequency, arrival mode, and prior hospitalizations, which provide critical clinical context beyond triage information alone. These findings align with previous studies showing that combining nursing data with physiological measures enhances model accuracy.

SHAP analysis further validated the importance of key predictors, such as prior hospitalizations, initial KTAS levels, oxygen saturation monitoring, and arrival mode, in explaining model predictions. The analysis highlighted how these variables support clinical decision-making, particularly in ICU admission and mortality prediction. For discharge

cases, predictors like low oxygen saturation monitoring frequency and arrival on foot indicated a higher likelihood of discharge, suggesting the potential of these models to optimize ED resource allocation. Overall, the study demonstrates the clinical value of integrating nursing data into predictive models, offering actionable insights to improve ED efficiency and patient outcomes.

This study developed and evaluated predictive models for KTAS Level 3 and 4 emergency patients, but several limitations warrant attention in future research. First, reliance on single-institution data limits the external validity and generalizability of the results to diverse clinical settings, highlighting the need for multicenter validation. Second, class imbalance, particularly for minority outcomes such as deaths and ICU admissions, constrained prediction performance despite applying SMOTE and cost-sensitive learning. Expanding datasets and employing advanced augmentation techniques could address this issue. Third, the lack of standardized nursing documentation posed challenges for data quality and model reliability, underscoring the need for standardized systems. Fourth, the exclusion of nurse-specific factors, such as experience and education, may have reduced the model's explanatory power. Additionally, initial symptoms, a critical indicator of patient status, were not categorized in sufficient detail, limiting predictive precision. Lastly, overlapping variables across outcome classes, such as oxygen saturation measurement frequency, reduced immediate clinical applicability. Adopting rule-based approaches to simplify variable interactions could enhance model interpretability and utility in real-world applications.

Conclusion

This study developed and evaluated predictive models for KTAS level 3 and 4 emergency patients using triage information and nursing data. The model incorporating nursing data outperformed the triage-only model, with significant improvements in predicting outcomes such as discharge and general ward admission. Key variables, including oxygen saturation frequency and mode of arrival, enhanced clinical decision support, while SHAP analysis improved model interpretability and reliability. These findings highlight the potential of nursing data to optimize resource allocation and predict clinical outcomes, demonstrating the practical utility of integrating triage and nursing data in predictive modeling.

References

- Aziz, W., Nicalaou, A., Stylianides, C., Panayides, A., Kakas, A., Kyriacou, E., & Pattichis, C. (2024). Emergency Department Length of Stay Classification Based on Ensemble Methods and Rule Extraction. *Studies in health technology and informatics*, 316, 1812–1816. <https://doi.org/10.3233/SHTI240783>
- Brink, A., Alsma, J., van Attekum, L., Bramer, W. M., Zietse, R., Lingsma, H., & Schuit, S. C. E. (2022). Predicting in-hospital admission at the emergency department: A systematic review. *Emergency Medicine Journal*, 39(3), 191-198. <https://doi.org/10.1136/emmermed-2020-210902>
- Collins, S. A., Cato, K., Albers, D., Scott, K., Stetson, P. D., Bakken, S., & Vawdrey, D. K. (2013). Relationship between nursing documentation and patients' mortality. *American Journal of Critical Care*, 22(4), 306-313. <https://doi.org/10.4037/ajcc2013426>
- Collins, S. A., & Vawdrey, D. K. (2012). "Reading between the lines" of flow sheet data: Nurses' optional documentation associated with cardiac arrest outcomes. *Applied Nursing Research*, 25(4), 251-257. <https://doi.org/10.1016/j.apnr.2011.06.002>
- Ekins, K., & Morphet, J. (2015). The accuracy and consistency of rural, remote and outpost triage nurse decision making in one Western Australia Country Health Service Region. *Australasian Emergency Nursing Journal*, 18(4), 227-233. <https://doi.org/10.1016/j.aenj.2015.05.002>
- Göransson, K. E., Ehrenberg, A., & Ehnfors, M. (2005). Triage in emergency departments: National survey. *Journal of Clinical Nursing*, 14(9), 1067-1074. <https://doi.org/10.1111/j.1365-2702.2005.01191.x>
- Korean Society of Emergency Medicine KTAS Committee. (2021). *KTAS: Korean Triage and Acuity Scale manual (Revised edition)*. Korean Society of Emergency Medicine.
- Larburu, N., Azkue, L., & Kerexeta, J. (2023). Predicting hospital ward admission from the

- emergency department: A systematic review. *Journal of Personalized Medicine*, 13(5).
<https://doi.org/10.3390/jpm13050849>
- Lee, J. H., Park, Y. S., Park, I. C., Lee, H. S., Kim, J. H., Park, J. M., . . . Kim, M. J. (2019). Over-triage occurs when considering the patient's pain in Korean Triage and Acuity Scale (KTAS). *PLoS One*, 14(5), e0216519. <https://doi.org/10.1371/journal.pone.0216519>
- Lee, S. Y., Chinnam, R. B., Dalkiran, E., Krupp, S., & Nauss, M. (2020). Prediction of emergency department patient disposition decision for proactive resource allocation for admission. *Health Care Management Science*, 23(3), 339–359. <https://doi-org/10.1007/s10729-019-09496-y>
- Ministry of Health and Welfare. (2023). *Emergency medical statistics annual report*. Seoul, South Korea: Ministry of Health and Welfare.
- Patel, V. L., Gutnik, L. A., Karlin, D. R., & Pusic, M. (2008). Calibrating urgency: triage decision-making in a pediatric emergency department. *Advances in Health Sciences Education: Theory and Practice*, 13(4), 503-520. <https://doi.org/10.1007/s10459-007-9062-6>
- Park, J. B., Choi, H. J., Kang, B. S., Kim, C. S., Kang, H. G., & Lim, T. H. (2014). Administration ; A nationwide survey of Korean emergency department triage systems and scales; A first step towards reform of the emergency medical service system. *Journal of the Korean Society of Emergency Medicine*, 25(5), 499-508.
- Riordan, J. P., Dell, W. L., & Patrie, J. T. (2017). Can patient variables measured on arrival to the emergency department predict disposition in medium-acuity patients? *The Journal of Emergency Medicine*, 52(5), 769–779. <https://doi.org/10.1016/j.jemermed.2016.11.018>