Emergency Medicine Australasia

ORIGINAL RESEARCH

The Sydney Triage to Admission Risk Tool With Artificial Intelligence (START-AI): Prediction of Inpatient Admission From Emergency Departments Using Ensemble Machine Learning

Michael Dinh[1,2] | Elizabeth Corbett[1,2] | Eliot Salmon[1,2] | Saleem Ahmed Khan[3] | Nicholas Moore[4,5] | Sreenadh Brahmapuram[3] | Radhika Seimon[1] | Naomi Derrick[1] | Saartje Berendsen Russell[1,2] | Thuy Truc Ngo[6] | Irena Koprinska[6]

[1]RPA Green Light Institute, Sydney Local Health District, Sydney, New South Wales, Australia | [2]Emergency Department, Royal Prince Alfred Hospital, Sydney Local Health District, Sydney, New South Wales, Australia | [3]Digital Health and Innovation, Sydney Local Health District, Sydney, New South Wales, Australia | [4]Emergency Department, Campbelltown Hospital, Sydney South West Local Health District, Sydney, New South Wales, Australia | [5]Macarthur Clinic School, Western Sydney University, Sydney, New South Wales, Australia | [6]School of Computer Science, Faculty of Engineering, The University of Sydney, Sydney, New South Wales, Australia

Correspondence: Radhika Seimon (radhika.seimon@health.nsw.gov.au)

Received: 20 November 2025 | **Revised:** 10 February 2026 | **Accepted:** 12 February 2026

Keywords: artificial intelligence | disposition | emergency department | ensemble machine learning

ABSTRACT

Objective: Use artificial intelligence (AI) to extend the Sydney triage to admission risk tool (START) and improve prediction of emergency department (ED) patient disposition.

Methods: The study was conducted at an inner-city tertiary referral hospital ED. Adult (age ≥ 16 years) presentations from 1 January 2023 to 30 June 2025 were included. Participants were excluded if dead on ED arrival or left ED prior to completing treatment. The primary outcome was admission to an inpatient ward. A sequential ensemble modelling approach was used. To predict patient disposition, the original START was combined (stacked) with vital signs, blood results and CT imaging orders using a gradient boosting decision tree algorithm (XGBoost) and a pre-trained transformer model for clinical free text.

Results: 162,915 cases were analysed with 27.31% overall inpatient admission rate. The final stacked meta-XGBoost model had an area under receiver operating curve (AUROC) of 0.88 (95% CI: 0.88, 0.89) with overall weighted accuracy of 0.84 (95% CI: 0.84, 0.85) and F1 score of 0.83 (95% CI: 0.83, 0.84) in the testing dataset. The model was adequately calibrated with R^2 of 0.92 (95% CI: 0.67, 0.99) with a drop-off in correlation at the highest predicted probability ranges (> 0.80). After classifying inpatient stays < 24 h as potential discharges, a sensitivity analysis demonstrated AUROC for the final model of 0.89 (95% CI: 0.88, 0.89).

Conclusions: An ensemble machine learning model was developed to accurately predict patient disposition from ED using structured and unstructured data. Prototype development and prospective evaluation of START-AI are required to assess model performance in clinical settings.

1 | Introduction

The decision to either admit or safely discharge a patient is one of the most important clinical decisions faced in the emergency department (ED). There has been increasing interest in the use of artificial intelligence (AI), including the use of unstructured free text data, to develop prediction models to aid with these decisions [1–9].

Sydney triage to admission risk tool (START) was a decision support tool developed in 2016 to estimate the probability of inpatient admission using only structured data available at triage [10, 11]. START has since been implemented into state-wide patient flow portals, validated prospectively and benchmarked against similar tools internationally, with 82% overall discriminatory accuracy [1, 2, 12]. In the present study, we aimed to improve the prediction capabilities of START by incorporating additional variables such as vital signs, blood tests, imaging and free text from clinical notes using AI. Such a model would enable real-time synthesis of complex data to support existing decision-making processes in ED and potentially flag those at risk of clinical deterioration.

2 | Methods

2.1 | Design

The study was a machine learning analysis using electronic health record (EHR) data from a single site.

2.2 | Study Setting

The study was conducted at an inner-city tertiary referral hospital in Sydney, Australia, with around 85,000 patients per annum and 25% inpatient admission rate.

2.3 | Participant Selection

Adult (aged ≥ 16 years) presentations to ED from 1 January 2023 to 30 June 2025 were included. Participants were excluded if the patient was dead on arrival to ED, or if the patient left prior to completing treatment in ED. Only the most recent ED presentation was used for each patient to avoid statistical bias and loss of independent observations due to frequent representations by a small number of patients.

2.4 | Data Sources and Variables

Data were extracted from the EHR by trained data analysts using reporting functions within Cerner FirstNet. Outputted variables included demographic (age, sex), administrative (date and time of ED arrival and departure, triage category based on Australasian Triage Scale [13], presenting problem, mode of arrival and separation, admission unit), free text (triage comment, first recorded ED nursing note, first recorded ED medical case history note), initial vital signs in EHR observation charts (heart rate [per minute], systolic blood pressure [mmHg], respiratory rate [/min], oxygen saturations [%], oxygen delivery, Glasgow Coma Score, temperature [°C]), initial blood tests (first recorded full blood count, biochemistry, troponin, venous blood gas) and any CT scans performed (regardless of CT result).

2.5 | Data Preprocessing and Imputation

START scores were calculated based on the original validation report [10] comprising age, triage category, presenting problem, mode of arrival, time of arrival and previous inpatient admission within 30 days. Flag variables were created to indicate any blood test, vital sign or CT performed as a proxy marker for the decision to request these. Vital signs were extracted from standard nursing observation charts (categorised according to established 'Between the Flags' thresholds) [14]. Laboratory results were extracted from existing fields in the EHR. For consistency, the following categories were used for vital signs and laboratory results: 1 = critically low, 2 = low, 3 = normal, 4 = high, 5 = critically high. Vital signs that were not recorded were imputed to be within normal range. Patient identifiers (names, medical record number, birthdate) were removed from the dataset pre-analysis. Missing ED medical case history notes were imputed with ED nursing notes. If ED nursing notes were missing, triage comment was used. ED medical case history notes were clipped to 150 words. Stop words, punctuation and special characters were retained to preserve contextual meanings within free text. Natural language processing was used to create flag variables based on the presence of major comorbidities or frailty in clinical free text based on Charlson Comorbidity Index [15] and Rockwood Frailty Scores, respectively [16].

2.6 | Outcomes

The primary outcome was an inpatient admission ('admitted') defined as an ED mode of separation to an inpatient ward from ED (excluding ED short stay unit). Patients who were admitted under an inpatient team but discharged while they were still in ED were classified as discharged from ED. Deaths in ED were classified as an inpatient admission. The secondary outcome was a composite of admission to intensive care unit (ICU) or in-hospital death in the 'admitted patients' subset.

2.7 | Data Analysis and Modelling

A bivariate comparison of admitted and non-admitted groups was performed to provide initial exploratory data of distributions and unadjusted odds ratios (OR) for variables in the dataset. The whole dataset was randomly split 1:1 into training and testing. The training set was used to build and finetune ensemble machine learning and transformer-based models. The testing dataset was used to validate predicted probabilities outputted from training models.

A sequential ensemble modelling approach was used. Firstly, START scores were obtained using logistic regression based on the original validation study [10]. Secondly, START scores and categorical data from flag variables, vital signs and blood tests were modelled to predict inpatient admission using a gradient-boosted decision tree algorithm (XGBoost). Thirdly, free text clinical notes from triage comments and ED case history notes were tokenised and encoded using Clinical Bidirectional Encoder Representations from Transformers (ClinicalBERT), designed specifically for encoding and labelling clinical free text [17]. Outputted probabilities from XGBoost and ClinicalBERT models were then combined using a stacked XGBoost meta-model. Model hyperparameter tuning was performed with a random subset of the training data, using GridSearchCV with L2 (lambda) regularisation to minimise overfitting. The final tuned model was evaluated on the testing

dataset to obtain final model performance. A composite secondary outcome (ICU admission, death) was modelled using only START, vital sign and blood test data with a single-shot XGBoost model on the same validation dataset.

Gradient boosting is an ensemble algorithm which adds decision trees sequentially, where each tree minimises the error of the previous tree [18]. XGBoost is an optimised and highly scalable version of gradient boosting, which has provided state-of-the-art results in many applications [19].

2.8 | Performance Metrics

Model performance was assessed based on overall accuracy (proportion of cases correctly classified as admission or discharge), recall (sensitivity), precision (positive predictive value), F1 (harmonic mean of precision and recall) and AUROC. 95% confidence intervals were estimated using cross validation of 1000 bootstrap samples. Performance metrics were weighted to account for imbalanced datasets. Feature importance was plotted for the final fitted model based on mean change in AUROC associated with an individual variable. Calibration curves were used to summarise performance of the model over predicted probability ranges with R^2 derived from a simple linear regression used to summarise overall correlation between predicted and observed probabilities of inpatient admission.

2.9 | Sensitivity Analyses

Sensitivity analyses were conducted on the primary outcome model by classifying an inpatient ward (including ED short stay unit) length of stay <24 h as a 'discharge' to identify patients who may be suitable for short stay unit admission or expedited discharge planning.

2.10 | Software and Computing Requirements

Exploratory analyses were conducted using SAS Enterprise Guide (version 6.3 SAS Institute, Cary, NC, USA) and ensemble machine learning modelling was conducted using Python code (version 13.3) and programmed in Visual Studio Code (version 1.115) with the assistance of GitHub Pro Copilot (version 1.388.0). The final code was verified by the investigating team. A 12GB NVIDIA Graphics Processing Unit installed in a Windows 10 operated PC (Dell Precision Tower: 72GB RAM, Intel(R) Core i7-10700) was used for this study. Pre-trained ClinicalBERT tokenizers were manually downloaded for local use [20] and fine-tuned on a separate dataset of 10,000 ED case history notes.

2.11 | Ethics Approval

Approval was obtained from the Sydney Local Health District Human Research Ethics Committee. All data was stored and analysed locally. Ethics approval reference X25-0116 and 2025/ETH01006. A waiver of consent was granted. The Tripod AI guidelines were followed, where applicable [21].

3 | Results

3.1 | Study Population

A total of 162,915 cases between 1 January 2023 to 30 June 2025 was analysed with overall inpatient admission rate of 27.31% (see Table 1 for demographic data). The mean (standard deviation [SD]) age was 50.51 (21.13) years. 49.01% of the study population was male. Life-threatening cases (Australasian triage 1–2) comprised of 19.52%, with admission to ICU or deaths occurring in 706 (0.43%) of cases. Blood tests were recorded in 70.55% and CT in 24.49%. There were no duplicates in the dataset. Missing triage category or presenting problems were present in 196 (0.12%) cases. Five cases had missing ED medical case history notes after imputation. Cases with these missing variables were excluded from further ensemble machine learning analyses.

3.2 | Exploratory Data Analysis

Mean (SD) START scores were 12.62 (6.66) in discharged group and 19.31 (6.28) in admitted group ($p<0.001$). Abnormalities in vital signs and blood test results had predictable associations with inpatient admission. Of note, the presence of any blood test abnormality was associated with an eight-fold increase in unadjusted odds of inpatient admission (OR 8.81; 95% CI: 8.54, 9.09) and the presence of any CT result increased this by five times (OR 5.48; 95% CI: 5.34, 5.61).

3.3 | Ensemble Machine Learning Models in Training Datasets

Table 2 summarises performance metrics of START score when modelled using logistic regression, demonstrating it was associated with AUROC of 0.77 (95% CI: 0.76, 0.77). AUROC for other features modelled using XGBoost ranged from 0.86 (95% CI: 0.85, 0.86) for initial blood test results, including the decision to perform blood tests, to 0.73 (95% CI: 0.73, 0.73) associated with ED medical case history notes modelled using fine-tuned ClinicalBERT.

3.4 | Performance of Final Stacked Meta XGBoost Model

After hyperparameter tuning with a learning rate of 0.05, the final stacked meta XGBoost model had AUROC of 0.87 (95% CI: 0.87, 0.87) with overall weighted accuracy of 0.83 (95% CI: 0.82, 0.83) and F1 score of 0.82 (95% CI: 0.82, 0.82) in the testing (validation) dataset. Figure 1 shows the calibration curve for admission prediction for the final model. Calibration of the model demonstrated an R^2 of 0.92 (95% CI: 0.67, 0.99) with a drop-off in correlation observed at the highest predicted probability ranges (>0.80), indicating the model over-predicts admission for high-risk patients. Using the same modelling approach, but without clinical free text, the final weighted AUROC associated with the secondary outcome of ICU or death was 0.93 (95% CI: 0.93, 0.94) with a weighted recall of 0.89 (95% CI: 0.88, 0.89) and a precision of 0.96 (95% CI: 0.96, 0.96). Figure 2 compares AUROC curves for the original

TABLE 1 | Comparison of demographic and clinical characteristics in discharged and admitted patients.

Variable	Discharged ($n = 118,415$)	Admitted ($n = 44,500$)	Total ($n = 162,915$) (%)	Unadjusted odds ratio (CI)
Age (years)				
16–19	2233 (1.89)	333 (0.75)	2566 (1.58)	Ref.
20–39	50,894 (42.98)	9592 (21.56)	60,486 (37.13)	1.27 (1.13, 1.43)
40–59	32,289 (27.27)	10,179 (22.87)	42,468 (26.07)	2.12 (1.88, 2.38)
60–79	24,340 (20.55)	14,060 (31.60)	38,400 (23.57)	3.88 (3.45, 4.36)
80+	8659 (7.31)	10,336 (23.23)	18,995 (11.66)	8.02 (7.13, 0.04)
Gender				
Female	61,381 (51.84)	21,420 (48.13)	82,801 (50.82)	Ref.
Male	56,810 (47.98)	23,032 (51.76)	79,842 (49.01)	1.16 (1.14, 1.19)
Other	224 (0.19)	48 (0.11)	272 (0.17)	0.62 (0.45, 0.84)
Triage category				
1	712 (0.06)	1285 (2.89)	1997 (1.23)	30.40 (25.57, 36.14)
2	16,212 (13.69)	13,596 (30.56)	29,808 (18.30)	14.16 (12.20, 16.42)
3	58,089 (40.06)	24,691 (55.49)	82,780 (50.81)	7.18 (6.19, 8.31)
4	40,199 (33.95)	4734 (10.64)	44,933 (27.58)	1.98 (1.71, 2.31)
5	3201 (2.70)	189 (0.42)	3390 (2.08)	Ref.
Arrival				
Non-ambulance	86,790 (73.29)	20,845 (46.84)	107,635 (66.07)	Ref.
Ambulance	31,625 (26.71)	23,655 (53.16)	55,280 (33.93)	3.12 (3.05, 3.19)
Admission past 30 days				
No	113,007 (95.43)	39,834 (89.51)	152,841 (93.82)	Ref.
Yes	5408 (4.57)	4666 (10.49)	10,074 (6.18)	2.45 (2.35, 2.55)
Admission past 7 days				
No	11,643 (98.50)	43,192 (97.06)	159,835 (98.11)	Ref.
Yes	1772 (1.50)	1308 (2.94)	3080 (1.89)	2.0 (1.86–2.14)
Presenting problem				
Abdominal pain	13,883 (11.72)	7592 (17.06)	21,475 (13.18)	1.01 (0.97, 1.04)
Cardio-vascular	14,197 (11.99)	4362 (9.80)	18,559 (11.39)	0.56 (0.54, 0.59)
Non-specific[a]	14,762 (12.47)	8045 (18.08)	22,807 (14.00)	Ref.
Infection	2779 (2.35)	1648 (3.70)	4427 (2.72)	1.09 (1.02, 1.16)
Trauma	23,369 (19.73)	5515 (12.39)	28,884 (17.73)	0.43 (0.42, 0.45)
Respiratory	5127 (4.33)	4278 (9.61)	9405 (5.77)	1.53 (1.46, 1.61)
MSK/ortho	8958 (7.56)	1657 (3.72)	10,615 (6.52)	0.34 (0.32, 0.36)
Neuro	9409 (7.95)	4017 (9.03)	13,426 (8.24)	0.78 (0.75, 0.82)
Mental health	6067 (5.12)	2898 (6.51)	8965 (5.50)	0.88 (0.83, 0.92)
Toxicology	749 (0.63)	393 (0.88)	1142 (0.70)	0.96 (0.85, 1.09)
ENT	5298 (4.47)	498 (1.12)	5796 (3.56)	0.17 (0.16, 0.19)

(Continues)

TABLE 1 | (Continued)

Variable	Discharged ($n = 118,415$)	Admitted ($n = 44,500$)	Total ($n = 162,915$) (%)	Unadjusted odds ratio (CI)
Surgery	2486 (2.10)	371 (0.83)	2857 (1.75)	0.27 (0.25, 0.31)
Renal urology	3901 (3.29)	1231 (2.77)	5132 (3.15)	0.58 (0.54, 0.62)
Social	1 (0.00)	3 (0.01)	4 (0.00%)	5.5 (0.57, 52.94)
Endocrine	254 (0.21)	198 (0.44)	452 (0.28)	1.43 (1.19, 1.73)
Obstetrics gynaecology	3112 (2.63)	794 (1.78)	3906 (2.40)	0.47 (0.43, 0.51)
Dermatology	3562 (3.01)	570 (1.28)	4132 (2.54)	0.29 (0.27, 0.32)
Haematology	368 (0.31)	374 (0.84)	742 (0.46)	1.87 (1.61, 2.16)
Missing data	133 (0.11)	56 (0.13)	189 (0.12)	0.77 (0.57, 1.06)
Frailty				
No	106,141 (89.63)	34,170 (76.79)	140,311 (86.13)	Ref.
Yes	12,274 (10.37)	10,330 (23.21)	22,604 (13.87)	2.62 (2.54, 2.69)
Major comorbidity				
No	98,258 (82.98)	26,240 (58.97)	124,498 (76.42)	Ref.
Yes	20,157 (17.02)	18,260 (41.03)	38,417 (23.58)	3.39 (3.31, 3.48)
Abnormal vitals				
No	105,888 (89.42)	33,379 (75.01)	139,267 (85.48)	Ref.
Yes	12,527 (10.58)	11,121 (24.99)	23,648 (14.52)	2.81 (2.74, 2.90)
Abnormal bloods				
No	62,979 (53.18)	5086 (11.43)	68,065 (41.78)	Ref.
Yes	55,436 (46.82)	39,414 (88.57)	94,850 (58.22)	8.81 (8.54, 9.09)
Any bloods				
No	46,429 (39.21)	1557 (3.50)	47,986 (29.45)	Ref.
Yes	71,986 (60.79)	42,943 (96.50)	114,929 (70.55)	17.85 (16.94, 18.80)
WCC ($\times 10^9$/L)				
No data	47,489 (40.10)	1714 (3.85)	49,203 (30.20)	0.07 (0.07, 0.08)
0–1.9	170 (0.14)	314 (0.71)	484 (0.30)	3.68 (3.06, 4.44)
2.0–3.9	1879 (1.59)	1260 (2.83)	3139 (1.93)	1.34 (1.24, 1.44)
4.0–10.9	56,696 (47.88)	28,425 (63.88)	85,121 (52.25)	Ref.
11.0–24.9	12,008 (10.14)	12,270 (27.57)	24,278 (14.90)	2.04 (1.98, 2.10)
25+	173 (0.15)	517 (1.16)	690 (0.42)	5.96 (5.02, 7.09)
Hb (g/L)				
No value	47,617 (40.21)	1756 (3.95)	49,373 (30.31)	0.08 (0.08, 0.08)
0–79	373 (0.31)	1296 (2.91)	1668 (1.02)	7.49 (6.67, 8.41)
80–114 (female) 80–129 (male)	9768 (8.25)	12,943 (29.09)	22,711 (13.94)	2.85 (2.76, 2.93)
115–159 (female) 130–169 (male)	59,611 (50.34)	27,743 (62.34)	87,354 (53.62)	Ref.

(Continues)

TABLE 1 | (Continued)

Variable	Discharged ($n = 118,415$)	Admitted ($n = 44,500$)	Total ($n = 162,915$) (%)	Unadjusted odds ratio (CI)
160–199 (female) 170–199 (male)	1040 (0.88)	753 (1.69)	1793 (1.10)	1.56 (1.42, 1.71)
199+	7 (0.01)	9 (0.02)	16 (0.01)	2.76 (1.03, 7.42)
CRP (mg/L)				
No value	81,743 (69.03)	11,384 (25.58)	93,127 (57.16)	0.25 (0.25, 0.26)
0–4.9	18,340 (15.49)	10,076 (22.64)	28,416 (17.44)	Ref.
5.0–69.9	15,523 (13.11)	16,083 (36.14)	31,606 (19.40)	1.89 (1.93, 1.95)
70+	2809 (2.37)	6957 (15.63)	9766 (5.99)	4.51 (4.29, 4.74)
Platelets ($\times 10^9$/L)				
No value	47,688 (40.27)	1894 (4.26)	49,582 (30.43)	0.07 (0.07, 0.08)
0–49	174 (0.15)	517 (1.16)	691 (0.42)	5.44 (4.58, 6.47)
50–149	3896 (3.29)	4203 (9.44)	8099 (4.97)	1.98 (1.89, 2.07)
150–399	63,746 (53.83)	34,794 (78.19)	98,540 (60.49)	Ref.
400–799	2865 (2.42)	2993 (6.73)	5858 (3.60)	1.91 (1.82, 2.02)
800+	46 (0.04)	99 (0.22)	145 (0.09)	3.94 (2.78, 5.60)
Sodium (mmol/L)				
No value	47,642 (40.23)	1797 (4.04)	49,439 (30.35)	0.07 (0.07, 0.07)
0–119	11 (0.01)	153 (0.34)	164 (0.10)	25.72 (13.95)
120–134	4850 (4.10)	6526 (14.67)	11,376 (6.98)	2.49 (2.39, 2.59)
135–144	64,712 (54.65)	34,997 (78.64)	99,709 (61.20)	Ref.
145–154	1195 (1.01)	928 (2.09)	2123 (1.30)	1.44 (1.32, 1.57)
155+	5 (0.00)	89 (0.22)	104 (0.06)	36.61 (14.91, 89.91)
Potassium (mmol/L)				
No value	51,095 (43.15)	4262 (9.58)	55,357 (33.98)	0.15 (0.14, 0.15)
0–2.4	12 (0.01)	67 (0.15)	79 (0.05)	9.94 (5.38, 18.38)
2.5–3.4	2020 (1.71)	1969 (4.42)	3989 (2.45)	1.74 (1.63, 1.85)
3.5–5.1	64,018 (54.06)	35,961 (80.81)	99,979 (61.37)	Ref.
5.2–6.1	1191 (1.01)	1864 (4.19)	3055 (1.88)	2.79 (2.59, 3.00)
6.2+	79 (0.07)	377 (0.85)	456 (0.28)	8.50 (6.67, 10.83)
Urea (mmol/L)				
No value	48,429 (40.90)	1932 (4.34)	50,361 (30.91)	0.07 (0.07, 0.07)
0–4.4	26,373 (22.27)	11,529 (25.91)	37,902 (23.26)	0.76 (0.74, 0.78)
4.5–9.9	39,573 (33.42)	22,844 (51.33)	62,417 (38.31)	Ref.
10.0+	4040 (3.41)	8195 (18.42)	12,235 (7.51)	3.51 (3.37, 3.67)
Creatinine (micromol/L)				
No value	47,694 (40.28)	1855 (4.17)	49,549 (30.41)	0.07 (0.07, 0.08)
0–59	18,257 (15.42)	8966 (20.15)	27,223 (16.71)	0.93 (0.90, 1.0)

(Continues)

TABLE 1 | (Continued)

Variable	Discharged ($n = 118,415$)	Admitted ($n = 44,500$)	Total ($n = 162,915$) (%)	Unadjusted odds ratio (CI)
60–109	46,945 (39.64)	24,859 (55.86)	71,804 (44.07)	Ref.
110–349	5173 (4.37)	7713 (17.33)	12,886 (7.91)	2.82 (2.71, 2.93)
350+	346 (0.29)	1107 (2.49)	1453 (0.89)	6.04 (5.35, 6.82)
Troponin T (high sensitive) (ng/L)				
No value	100,254 (84.66)	32,299 (72.58)	132,553 (81.36)	1.20 (1.16, 1.25)
0–13	13,839 (11.69)	3716 (8.35)	17,555 (10.78)	Ref.
14–51	3812 (3.22)	5689 (12.78)	9501 (5.83)	5.56 (5.26, 5.87)
52+	510 (0.43)	2796 (6.28)	3306 (2.03)	20.42 (18.45, 22.59)
Venous pH				
No value	100,866 (85.18)	21,309 (47.89)	122,175 (74.99)	0.18 (0.18, 0.19)
0–7.19	132 (0.11)	548 (1.23)	680 (0.42)	4.56 (2.94, 4.32)
7.20–7.29	549 (0.46)	1563 (3.51)	2112 (1.30)	2.44 (2.21, 2.70)
7.30–7.39	10,084 (8.52)	11,746 (26.40)	21,830 (13.40)	Ref.
7.40–7.59	6720 (5.67)	9286 (20.87)	16,006 (9.82)	1.19 (1.14, 1.24)
7.60	64 (0.05)	48 (0.11)	112 (0.07)	0.64 (0.44, 0.94)
Lactate (mmol/L)				
No value	100,812 (85.13)	21,271 (47.80)	122,083 (74.94)	0.18 (0.18, 0.19)
0–1.9	12,700 (10.72)	14,789 (33.23)	27,489 (16.87)	Ref.
2.0–3.9	4116 (3.48)	6689 (15.03)	10,805 (6.63)	1.40 (1.33, 1.46)
4.0+	787 (0.66)	1751 (3.93)	2538 (1.56)	1.91 (1.75, 2.09)
Heart rate (/min)				
1–39	41 (0.03)	19 (0.04)	60 (0.04)	1.23 (0.72, 2.12)
40–49	574 (0.48)	166 (0.37)	740 (0.45)	0.77 (0.65, 0.91)
50–119	117,099 (98.89)	44,071 (99.04)	161,170 (98.93)	Ref.
120–139	656 (0.55)	200 (0.45)	858 (0.53)	0.81 (0.69, 0.95)
140+	45 (0.04)	44 (0.10)	89 (0.05)	2.60 (1.71, 3.94)
Respiratory rate (/min)				
0–4	8 (0.01)	5 (0.01)	13 (0.01)	1.68 (0.55, 5.11)
5–9	47 (0.04)	24 (0.05)	71 (0.04)	1.37 (0.84, 2.24)
10–24	117,818 (99.50)	44,021 (98.92)	161,839 (99.34)	Ref.
25–29	432 (0.36)	248 (0.56)	680 (0.42)	1.54 (1.31, 1.80)
30+	110 (0.09)	202 (0.45)	312 (0.19)	4.92 (3.90, 6.20)
Systolic BP (mmHg)				
0–89	143 (0.12)	291 (0.65)	434 (0.27)	5.49 (4.50, 6.71)
90–99	2724 (2.30)	1640 (3.69)	4364 (2.68)	1.63 (1.53, 1.73)
100–179	114,587 (96.77)	42,440 (95.37)	157,027 (96.39)	Ref.
180–199	885 (0.75)	118 (0.27)	1003 (0.62)	0.36 (0.30, 0.44)

(Continues)

TABLE 1 | (Continued)

Variable	Discharged ($n = 118,415$)	Admitted ($n = 44,500$)	Total ($n = 162,915$) (%)	Unadjusted odds ratio (CI)
200+	76 (0.06)	11 (0.02)	87 (0.05)	0.39 (0.21, 0.74)
Temperature (°C)				
0–34.4	18 (0.02)	49 (0.11)	67 (0.04)	7.25 (4.23, 12.45)
34.5–35.4	268 (0.23)	148 (0.33)	416 (0.26)	1.47 (1.20, 1.80)
35.5–38.4	117,959 (99.61)	44,270 (99.48)	162,229 (99.58)	Ref.
38.5–40.9	170 (0.14)	33 (0.07)	203 (0.12)	0.52 (0.36, 0.75)
Oxygen saturation (%)				
0–89	158 (0.13)	348 (0.78)	506 (0.31)	6.07 (5.03, 7.33)
90–94	1858 (1.57)	1923 (4.32)	3781 (2.32)	2.85 (2.67, 3.04)
95–100	116,399 (98.30)	42,229 (94.90)	158,628 (97.37)	Ref.
Oxygen delivery on arrival				
Room air	118,120 (99.75)	43,632 (98.05)	161,752 (99.29)	Ref.
Any O_2	295 (0.25)	868 (1.95)	1163 (0.71)	7.96 (6.98, 9.01)
GCS				
3–12	566 (0.48)	1175 (2.64)	1741 (1.07)	6.38 (5.76, 7.05)
13–14	5071 (4.28)	6601 (14.83)	11,672 (7.16)	4.00 (3.85, 4.15)
15	112,778 (95.24)	36,724 (82.53)	149,502 (91.77)	Ref.
Start score				
Mean (SD)	12.62 (6.66)	19.31 (6.28)	14.44 (7.21)	N/A
Interquartile range (Q1, Q3)	9 (8, 17)	9 (15, 24)	11 (9, 24)	N/A
CT				
CT brain	8029 (6.78)	7250 (16.29)	15,279 (9.38)	2.68 (2.59, 2.77)
CT stroke	1301 (1.10)	1089 (2.45)	2390 (1.47)	2.26 (2.08, 2.45)
CT head neck	1301 (1.10)	660 (1.48)	1961 (1.20)	1.36 (1.23, 1.49)
CT C spine	1165 (0.98)	591 (1.33)	1756 (1.08)	1.36 (1.22, 1.50)
CT chest	2138 (1.81)	3731 (8.38)	5869 (3.60)	4.98 (4.71, 5.26)
CT abdomen	3717 (3.14)	7533 (16.93)	11,250 (6.91)	6.29 (6.04, 6.56)
CT renal	1548 (1.31)	1394 (3.13)	2942 (1.81)	2.44 (2.27, 2.63)
CT ortho	4392 (3.71)	7947 (17.86)	12,339 (7.57)	5.64 (5.43, 5.87)
CT spine	419 (0.35)	448 (1.01)	867 (0.53)	2.86 (2.50, 3.27)
CT other	192 (0.16)	570 (1.28)	762 (0.47)	8.00 (6.78, 9.41)
Any CT	17,916 (15.13)	21,976 (49.38)	39,892 (24.49)	5.48 (5.34, 5.61)
Word count mean (SD)				
Triage comment	35.63 (10.61)	35.00 (11.68)	35.63 (10.92)	N/A
ED case history note	128.06 (47.13)	146.15 (29.22)	133.00 (43.73)	N/A

Note: All p values < 0.0001 except for the following: gender: other = 0.0025. Presenting problem: abdominal pain = 0.8531, infection = 0.0128; toxicology = 0.5549, social = 0.1396, endocrine = 0.0002, missing data = 0.1070. Heart rate: '1–39' 0.4532, '40–49' 0.0029, '120–139' 0.0093. Respiratory rate: '0–4' 0.3668, '5–9' 0.2131. Systolic BP: '200+' 0.0036. Temperature: '34.5–35.4' 0.0002, '38.5–40.9' 0.0005. Word count mean: Triage comment = 0.9266.

[a]Non-specific includes generalised pain without localisation, unspecified lethargy or malaise.

TABLE 2 | Modelling table.

Variable group	Model	Accuracy (95% CI)	Precision	Recall	F1	AU ROC (discrimination)
Training dataset						
START score	Logistic regression	0.69 (0.68, 0.70)	0.45 (0.44, 0.47)	0.72 (0.69, 0.73)	0.56 (0.55, 0.56)	0.77 (0.76, 0.77)
Sex, readmission within 7 days, major Comorbidity, frailty	XGBoost	0.70 (0.69, 0.70)	0.45 (0.44, 0.45)	0.54 (0.53, 0.54)	0.49 (0.48, 0.49)	0.66 (0.66, 0.66)
Initial vital signs	XGBoost	0.74 (0.73, 0.74)	0.53 (0.52, 0.54)	0.25 (0.24, 0.25)	0.34 (0.33, 0.34)	0.61 (0.61, 0.61)
Initial blood tests	XGBoost	0.78 (0.76, 0.79)	0.57 (0.55, 0.59)	0.75 (0.72, 0.78)	0.64 (0.64, 0.65)	0.86 (0.85, 0.86)
CT	XGBoost	0.76 (0.75, 0.76)	0.56 (0.56, 0.57)	0.49 (0.48, 0.50)	0.52 (0.52, 0.53)	0.68 (0.68, 0.69)
Clinical free text						
ED MO case history notes	ClinicalBERT	0.67 (0.66, 0.67)	0.43 (0.42, 0.43)	0.66 (0.65, 0.67)	0.52 (0.51, 0.52)	0.73 (0.73, 0.73)
Triage comment	ClinicalBERT	0.64 (0.64, 0.65)	0.41 (0.40, 0.41)	0.66 (0.65, 0.57)	0.50 (0.50, 0.51)	0.70 (0.70, 0.71)
Stacked meta XGBoost model	Metamodel	0.85 (0.84, 0.86)	0.69 (0.67, 0.71)	0.86 (0.83, 0.87)	0.76 (0.75, 0.77)	0.94 (0.93, 0.94)
Testing dataset						
Stacked meta XGBoost model	Tuned metamodel	0.83 (0.82, 0.83)	0.82 (0.82, 0.82)	0.83 (0.82, 0.83)	0.82 (0.82, 0.82)	0.87 (0.87, 0.87)

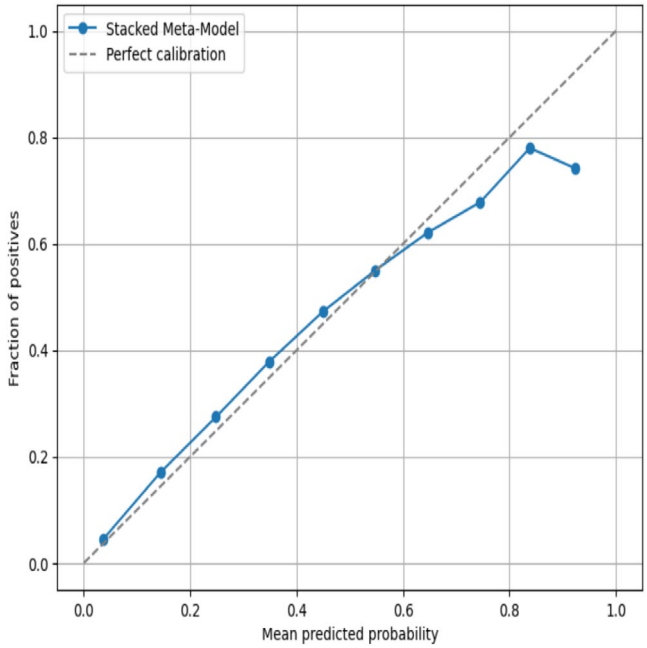

FIGURE 1 | Calibration curve for admission prediction for final stacked meta XGBoost model. *Y*-axis (fraction of positives) is equivalent to the actual/observed proportion of patients admitted.

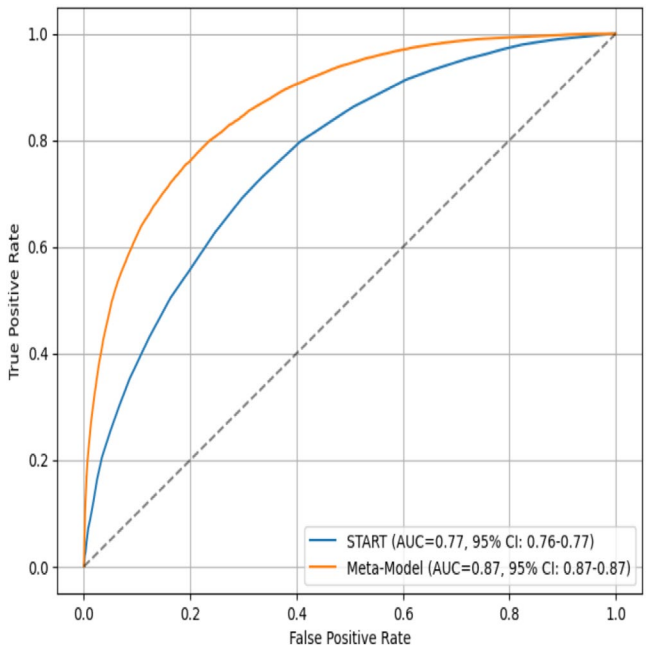

FIGURE 2 | AUROC curves: logistic regression for start score and final stacked meta XGBoost model.

START tool and final stacked meta XGBoost model in the testing (validation) dataset.

3.5 | Sensitivity Analysis

After reclassifying those with inpatient length of stay < 24 h (including ED short stay) as a potential discharge from ED for sensitivity analyses, inpatient admission rate was 22.42%. The final stacked meta XGBoost model was associated with an AUROC of 0.89 (95% CI: 0.88, 0.89).

3.6 | Feature Importance

The most important variables predicting inpatient admission with respect to mean change in AUROC using feature permutation were related to blood tests, with the presence of any abnormal white cell count or urea being most important (Figure 3a). With respect to ICU or death, the most important variables were serum venous lactate, presence of any vital sign abnormality, reduced GCS and START score (Figure 3b).

4 | Discussion

The present study built on the original START tool by employing ensemble machine learning AI algorithms to analyse a dataset comprising both structured and unstructured clinical data [10–12]. The final model demonstrated very good to excellent performance in predicting inpatient admission, based on commonly accepted thresholds for AUROC [22]. Model performance was robust in a sensitivity analysis and demonstrated adequate calibration across different ranges of predicted probabilities.

Since 2018, advances in data science and computing capabilities have made the analysis of large, unstructured clinical datasets more feasible [23]. Levin et al. used triage data, chief complaint and structured previous diagnoses from medical records in the United States to build an admission prediction model, with an AUROC for admission of 0.82–0.84 across different ED populations [8]. Tahayori et al. used ClinicalBERT to analyse free text in triage notes in Melbourne, Australia, achieving AUROC of 0.88 for disposition prediction [4]. Cusido et al. analysed structured triage data from over 3 million patients from 60 EDs in Spain, to develop an admission prediction model with an AUROC of 0.89 [9]. Recent studies using machine learning to develop ED disposition tools have demonstrated accuracies around 87% and highlighted the importance of additional data including pre-existing comorbidities and free text with natural language processing [2, 3]. A 2025 meta-analysis by Kuo and Chang found 81% sensitivity, 87% specificity and 0.87 AUROC for the pooled data of 39 studies assessing AI admission prediction models; however, there was variability regarding structured and unstructured text [5]. Similarly, in Shin et al.'s 2025 systematic review, 20 studies were identified that analysed triage data for admission prediction, with an AUROC from 0.80 to 0.89 [6].

Strengths of the present study include the use of a validated risk tool as a foundation and an ensemble approach which improved the robustness of the model's performance. We used a range of clinically relevant features including blood test results and CT requests to replicate factors that influence usual clinical decision-making processes in ED. As the presence or absence of a particular variable was coded as part of the model, it was designed to incrementally calculate probabilities of outcomes and not rely on the presence of all variables to output a prediction. A prediction could therefore be

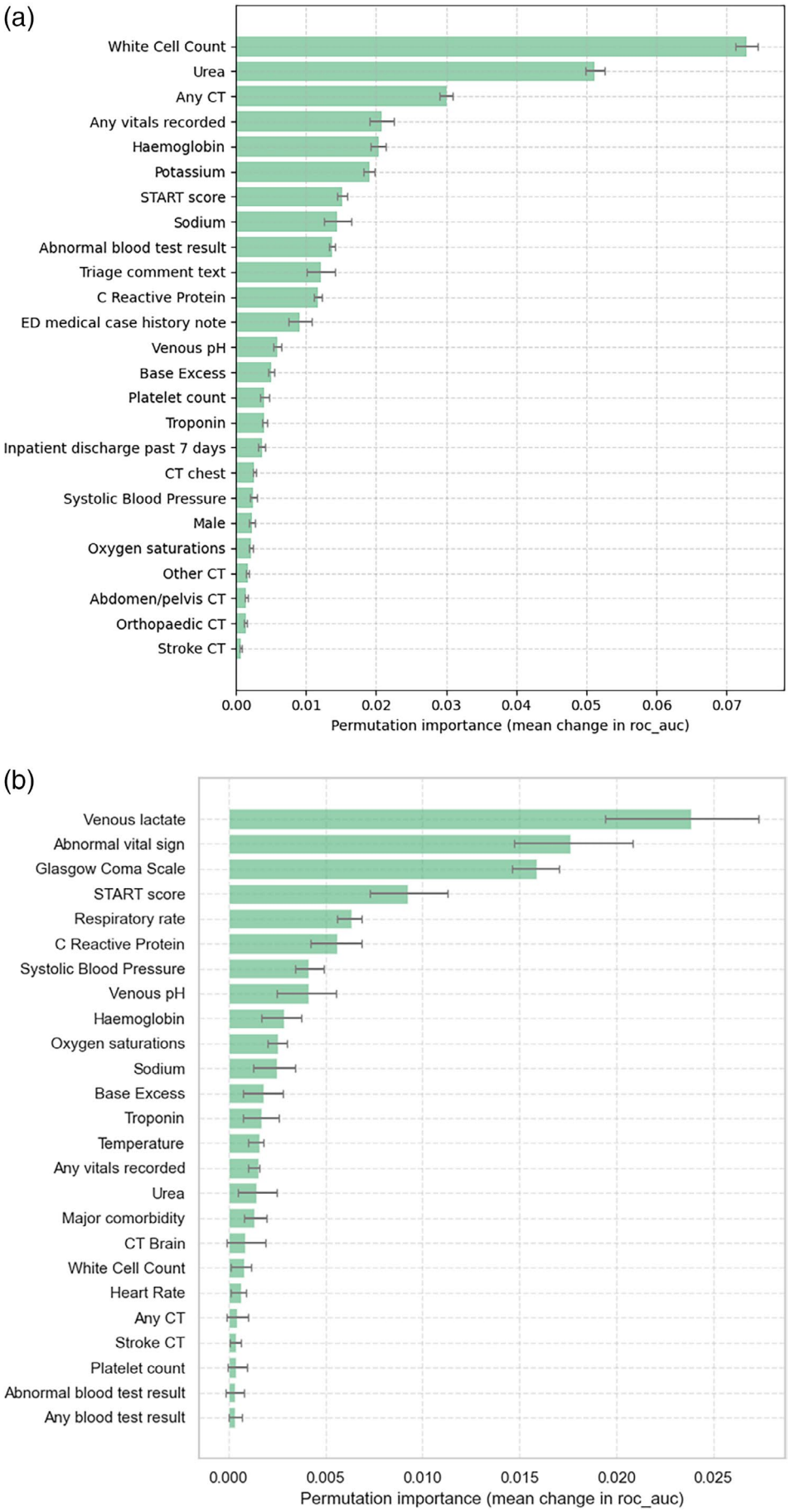

FIGURE 3 | Legend on next page.

FIGURE 3 | (a) Feature importances with 95% CI (permutation importance) of final stacked meta XGBoost model on testing data set for admission prediction. (b) Feature importances with 95% CI (permutation importance) of final stacked meta XGBoost model on testing data set for ICU or death prediction.

outputted with or without the presence of blood test results or free text clinical notes for instance. Our results provided evidence to support further development and prospective evaluation, similar to previous studies. A small implementation pilot of the original START demonstrated a reduction in ED length of stay (301 vs. 423 min; $p < 0.0001$) when the START score was used to assist decision making compared with matched controls [24].

4.1 | Limitations

There were several acknowledged limitations to our study. Firstly, this was a single centre analysis of EHR data. Implementation at scale would require analysis of data from a broad range of health services with different patient demographics, admission rates and proportion of arrivals by ambulance, to ensure validity and model performance across different sites and modes of care. Secondly, the final model was complex, utilising many clinical variables. For clinical use, the model would need to be integrated within clinical data warehouses, with incremental changes in outcome probabilities displayed dynamically on clinically facing systems as variables such as vital signs and blood tests become progressively available. We did not aim to compare this model against human performance as the use case for this model would be to support senior decision making in ED rather than replace it. Of note, Vaghasiya et al. assessed the accuracy of various groups, including triage nurses and ED doctors, in making admission predictions based on brief assessments at triage only, finding triage nurses were able to predict 445/490 (90.82%) discharges with a positive predictive value of 90.8% (95% CI: 87.8, 93.2) [25]. A prospective evaluation would involve measuring how START-AI could help to improve timeliness of human clinical decision making or reduce preventable inpatient admissions and representations to ED. Further work may include development of a prototype risk prediction tool for piloting and implementation. Other future directions include incorporation of imaging reports and medication summaries as part of the modelling process. These are likely to be computationally more intensive and will require access to significant cloud computing resources.

4.2 | Conclusions

START-AI was derived and internally validated to accurately predict patient disposition from ED in a single centre study. Prototype implementation and pilot prospective evaluation of START-AI will be required to assess model performance in the clinical setting.

Author Contributions

All authors contributed to the manuscript and study design. M.D., E.C., E.S., S.B., N.M., S.A.K., T.T.N., I.K. performed data analysis.

Acknowledgements

We acknowledge Kendall Bein, Emergency Physician, for his contribution to data processing.

Funding

The authors have nothing to report.

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

The data that support the findings of this study are not publicy available due to third party restrictions.

References

1. A. Brink, J. Alsma, L. A. van Attekum, et al., "Predicting In-Hospital Admission at the Emergency Department: A Systematic Review," *Emergency Medicine Journal* 39, no. 3 (2022): 191–198, https://doi.org/10.1136/emermed-2020-210902.

2. W. S. Hong, A. D. Haimovich, and R. A. Taylor, "Predicting Hospital Admission at Emergency Department Triage Using Machine Learning," *PLoS One* 13, no. 7 (2018): e0201016, https://doi.org/10.1371/journal.pone.0201016.

3. Y. Change, Y.-C. Lin, F.-W. Huang, et al., "Using Machine Learning and Natural Language Processing in Triage for Prediction of Clinical Disposition in the Emergency Department," *BMC Emergency Medicine* 24 (2024): 237–255.

4. B. Tahayori, N. Chini-Foroush, and H. Akhlaghi, "Advanced Natural Language Processing Technique to Predict Patient Disposition Based on Emergency Triage Notes," *Emergency Medicine Australasia* 33 (2021): 480–484.

5. K.-M. Kuo and C. Chang, "A meta-Analysis of the Diagnostic Test Accuracy of Artificial Intelligence Predicting Emergency Department Dispositions," *BMC Medical Informatics and Decision Making* 25 (2025): 187.

6. H. Shin, H. Kang, and M. Choi, "Triage Data-Driven Prediction Models for Hospital Admission of Emergency Department Patients: A Systematic Review," *Healthcare Informatics Research* 31, no. 1 (2025): 23–36.

7. J. T. Lee, C. C. Hsieh, C. H. Lin, Y. J. Lin, and C. Y. Kao, "Prediction of Hospitalization Using Artificial Intelligence for Urgent Patients in the Emergency Department," *Scientific Reports* 11, no. 1 (2021): 19472, https://doi.org/10.1038/s41598-021-98961-2.

8. s. Levin, M. Toerper, E. Hamrock, et al., "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, no. 5 (2018): 565–574.e2.

9. J. Cudiso, J. Comalrena, H. Alavi, J. Cusidó, and L. Llunas, "Predicting Hospital Admissions to Reduce Crowding in the Emergency Department," *Applied Sciences* 12, no. 21 (2022): 10764, https://doi.org/10.3390/app122110764.

10. M. M. Dinh, S. B. Russell, K. J. Bein, et al., "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, no. 1 (2016): 46, https://doi.org/10.1186/s12873-016-0111-4.

11. A. A. Ebker-White, K. J. Bein, and M. M. Dinh, "The Sydney Triage to Admission Risk Tool (START): A Prospective Validation Study," *Emergency Medicine Australasia* 30, no. 4 (2018): 511–516, https://doi.org/10.1111/1742-6723.12940.

12. K. Rendell, I. Koprinska, A. Kyme, A. A. Ebker-White, and M. M. Dinh, "The Sydney Triage to Admission Risk Tool (START2) Using Machine Learning Techniques to Support Disposition Decision-Making," *Emergency Medicine Australasia* 31, no. 3 (2019): 429–435, https://doi.org/10.1111/1742-6723.13199.

13. Australasian College for Emergency Medicine, "Triage," accessed 6 November 2025, https://acem.org.au/Content-Sources/Advancing-Emergency-Medicine/Better-Outcomes-for-Patients/Triage.

14. Clinical Excellence Commission, "NSW Health Observation Charts," accessed 5 November 2025, https://www.cec.health.nsw.gov.au/keep-patients-safe/between-the-flags/observation-charts.

15. M. E. Charlson, P. Pompei, K. L. Ales, and C. R. MacKenzie, "A New Method of Classifying Prognostic Comorbidity in Longitudinal Studies: Development and Validation," *Journal of Chronic Diseases* 40, no. 5 (1987): 373–383, https://doi.org/10.1016/0021-9681(87)90171-8.

16. K. Rockwood, X. Song, C. MacKnight, et al., "A Global Clinical Measure of Fitness and Frailty in Elderly People," *CMAJ* 173, no. 5 (2005): 489–495.

17. W. Cui and J. Finkelstein, "Leveraging LLMs for Early Diagnosis: Comparing ClinicalBERT and GPT-4 to Address Diagnostic Uncertainty in the Emergency Department," *Studies in Health Technology and Informatics* 328, no. 26 (2025): 153–157, https://doi.org/10.3233/SHTI250692.

18. J. Friedman, T. Hastie, and R. Tibshirani, "Additive Logistic Regression: A Statistical View of Boosting (With Discussion)," *Annals of Statistics* 28 (2000): 337–374, https://doi.org/10.1214/aos/1016218223.

19. T. Chen and C. Guestrin, "XGBoost: A Scalable Tree Boosting System," in *Proceedings of the Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)* (2016), 785–794, https://doi.org/10.1145/2939672.2939785.

20. Hugging Face, "ClinicalBERT – Bio + Clinical BERT Model," accessed 6 November 2025, https://huggingface.co/emilyalsentzer/Bio_ClinicalBERT.

21. G. S. Collins, K. G. M. Moons, P. Dhiman, et al., "TRIPOD+AI Statement: Updated Guidance for Reporting Clinical Prediction Models That Use Regression or Machine Learning Methods," *BMJ* 385 (2024): e078378.

22. Ş. K. Çorbacıoğlu and G. Aksel, "Receiver Operating Characteristic Curve Analysis in Diagnostic Accuracy Studies: A Guide to Interpreting the Area Under the Curve Value," *Turkish Journal of Emergency Medicine* 23, no. 4 (2023): 195–198, https://doi.org/10.4103/tjem.tjem_182_23.

23. R. Petrella, "The AI Future of Emergency Medicine," *Annals of Emergency Medicine* 84, no. 2 (2024): 139–153.

24. A. Ebker-White, K. Bein, S. Berendsen Russell, K. J. Bein, and M. M. Dinh, "The Sydney Triage to Admission Risk Tool (START) to Improve Patient Flow in an Emergency Department: A Model of Care Implementation Pilot Study," *BMC Emergency Medicine* 19 (2019): 79, https://doi.org/10.1186/s12873-019-0290-x.

25. M. Vaghasiya Murphy, D. O'Flynn, M. R. Vaghasiya, M. Murphy, D. O'Flynn, and A. Shetty, "The Emergency Department Prediction of Disposition (EPOD) Study," *Australasian Emergency Nursing Journal* 17, no. 4 (2014): 161–166.

Supporting Information

Additional supporting information can be found online in the Supporting Information section. **Appendix 1** Start score from original START validation study. **Appendix 2:** Final stacked meta XGBoost model inputs.