JOURNAL OF CLINICAL MEDICINE OF KAZAKHSTAN (E-ISSN 2313-1519)
Original Article DOI: https://doi.org/10.23950/jcmk/17364

Diagnostic Performance of the STEPSS Score and Machine Learning Models for Predicting Outcomes in Pediatric Status Epilepticus: A Prospective Observational Study

Alapati Lakshmi Rushitha[1], Dinesh Kamaraj[1], Rangasamy Krishnamoorthi[1], Panneerselvam Periasamy[2], Arbind Kumar Choudhary[3]

[1]Department of Pediatrics, Vinayaka Mission's Kirupananda Variyar Medical College and Hospital, Vinayaka Mission's Research Foundation (Deemed to be University), Salem, Tamil Nadu, India
[2]Department of Physiology, Government Erode Medical College and Hospital, Erode, Tamil Nadu, India
[3]Department of Pharmacology, Government Erode Medical College and Hospital, Erode, Tamil Nadu, India

Received: 2025-08-07.
Accepted: 2025-10-06.

J Clin Med Kaz 2025; 22(6): 44–53

Corresponding author:
Dinesh Kamaraj
E-mail: doctordinesh1986@gmail.com.
ORCID: 0009-0003-1877-6720.

ABSTRACT

Background: Pediatric status epilepticus (SE) is a neurological emergency associated with significant morbidity. The Status Epilepticus Pediatric Severity Score (STEPSS) offers rapid bedside prognostication, but its performance relative to machine learning (ML) models has not been well studied in resource-limited settings.

Methods: We prospectively enrolled 100 children with SE (median age 3.4 years; 52 % male) admitted to a tertiary center in South India (2023–2024). Clinical features, investigations, and outcomes were recorded. Functional outcome was assessed at discharge using the Pediatric Overall Performance Category (POPC), with unfavorable outcome defined as POPC ≥3. Prognostic accuracy of STEPSS and three ML models—Logistic Regression (LR), Random Forest (RF), and XGBoost—was evaluated using sensitivity, specificity, predictive values, accuracy, and area under the ROC curve (AUC).

Results: Overall, 30 % of children had unfavorable POPC outcomes. At presentation, 40 % had altered consciousness and 45 % had high-risk seizure types. STEPSS ≥3 was associated with ICU admission, mechanical ventilation, and poor outcome. STEPSS demonstrated good discrimination (sensitivity 79 %, specificity 78 %, NPV 84 %, AUC 0.83). Additional predictors of poor outcome included low SpO_2, hyperglycemia, CT abnormalities (present in 41 % of those imaged), and delay to first AED EEG was performed in 53 % of patients, with abnormalities in 38 % .Among ML models, LR achieved performance similar to STEPSS (AUC 0.82), while RF (AUC 0.88) and XGBoost (AUC 0.91) outperformed it, with XGBoost achieving the highest accuracy (90 %) and the fewest misclassifications. Feature importance analysis highlighted CT abnormalities, treatment delay, blood glucose, and SpO_2 as dominant predictors, with STEPSS also contributing significantly.

Conclusion: STEPSS remains a practical and reliable bedside triage tool in pediatric SE, particularly in low-resource emergency settings. However, integrating additional clinical indicators into ensemble ML models, especially XGBoost, provides superior prognostic accuracy. A hybrid strategy—STEPSS combined with ML augmentation—offers both interpretability and precision, supporting early risk stratification and informed critical care decisions.

Keywords: Status epilepticus, pediatric neurology, STEPSS score, machine learning, outcome prediction, XGBoost, POPC score

Introduction

Status epilepticus (SE) is a critical pediatric neurological emergency characterized by seizures lasting longer than five minutes or recurrent episodes without full recovery of consciousness in between. It carries significant morbidity and mortality, especially when there are delays in recognition or treatment. Globally, pediatric SE incidence ranges from 10 to 27 per 100,000 children annually, with higher rates reported in low- and middle-income countries (LMICs) due to increased prevalence of CNS infections, delayed access to care, and limited emergency infrastructure [1]. Hospital-based data from India have reported pediatric SE rates as high as 35 per 100,000 per year, with mortality ranging from 3% to 11% and long-term neurodevelopmental impairment in 25%–40% of survivors [2].

Early identification of patients likely to experience poor outcomes is crucial in managing SE effectively. In this context, the Status Epilepticus in Pediatric Patients Severity Score (STEPSS) was proposed by Sidharth et al. to aid emergency physicians in predicting short-term functional outcomes. The score incorporates four easily obtainable bedside parameters: age under two years, altered level of consciousness, generalized seizure type, and a history of seizures [3]. In a prospective Indian cohort, STEPSS demonstrated high sensitivity (93%) and specificity (81%) for unfavorable outcomes using a cutoff score ≥3. An external validation study by Soydan et al. in a Turkish pediatric population confirmed these findings, reporting an AUC of 0.917 for poor outcome prediction .

Despite its simplicity, STEPSS does not include several key clinical variables known to influence outcomes, such as blood glucose levels, oxygen saturation, serum sodium, neuroimaging results, and treatment latency. These additional factors have been associated with increased morbidity and mortality in pediatric SE and are particularly relevant in LMICs, where such complications are prevalent [4].

Machine learning (ML) has emerged as a powerful approach for predictive modeling in clinical care. ML techniques can process and integrate multidimensional data, uncovering complex relationships that traditional statistical models may overlook. In pediatric epilepsy, ML models such as random forest and XGBoost have been successfully used to predict seizure outcomes after surgery and intensive care mortality These models can offer improved prediction performance and flexibility while preserving clinical interpretability [5].

The present study was designed to evaluate the real-world diagnostic utility of the STEPSS score and assess whether the integration of additional early clinical variables using supervised ML models—Logistic Regression, Random Forest, and XGBoost—could improve prediction of adverse short-term outcomes, including a Pediatric Overall Performance Category (POPC) score ≥3, ICU admission, and need for ventilator support. By comparing the baseline STEPSS model to enhanced data-driven approaches, this study aims to bridge bedside decision-making with computational triage.

Methods

Study Setting and Design

This was a prospective observational study carried out at the Pediatric Emergency and Intensive Care Units of Vinayaka Mission's Kirupananda Variyar Medical College, Salem , over a period of two years, from January 2023 to December 2024. The study was designed to evaluate the clinical performance of the STEPSS score in pediatric status epilepticus and to assess whether machine learning algorithms could enhance prognostic accuracy. Approval for the research protocol was obtained from the Institutional Ethics Committee prior to patient recruitment (IEC Approval No. VMKVMC&H /IEC/25/024) Written informed consent was obtained from the guardians of all eligible participants.

Patient Selection

All children between the ages of one month and twelve years presenting with clinical features of status epilepticus were screened for eligibility. Status epilepticus was defined as a continuous seizure lasting more than five minutes or two or more seizures occurring without full recovery of consciousness in between. Neonates and children with progressive neurodegenerative diseases were excluded to reduce confounding. In total, 122 children were screened, and 100 were included in the final analysis after applying exclusion criteria and ensuring complete clinical and outcome data.

Clinical Assessment and Data Capture

A detailed clinical history was taken at admission, and all relevant clinical findings were documented systematically. Data collection focused on key demographic details, including age, sex, and body weight, as well as seizure-specific factors such as seizure type, estimated duration before arrival, and any known prior episodes. Vital signs including heart rate, respiratory rate, blood pressure, temperature, and oxygen saturation were recorded upon initial evaluation.

In addition to routine blood investigations, serum glucose, sodium, calcium, and white cell count were documented where available. Imaging, in the form of CT or MRI, was conducted selectively based on clinical judgment, particularly in children with focal seizures, prolonged altered consciousness, or a first seizure episode. EEG was performed in patients suspected to have non-convulsive seizures or in whom recovery of consciousness was delayed beyond the typical postictal period.

STEPSS Scoring and Expanded Variable Selection

Each patient's clinical profile was evaluated using the STEPSS scoring system. This scoring tool considers four parameters: age under two years, generalized seizure type, altered consciousness, and a prior history of seizures. Scores ranged from zero to four. In addition to STEPSS components, supplementary variables of clinical relevance—such as oxygen saturation levels, time to first antiepileptic drug, abnormal CT findings, and metabolic derangements—were recorded to allow for deeper exploratory analysis. These variables were considered due to their known associations with poor outcomes in pediatric neuro-emergencies and were intended to enhance predictive granularity.

Outcome Measurement

Short-term neurological outcome was measured using the Pediatric Overall Performance Category (POPC) scale at discharge. The POPC ranges from 1 (normal) to 5 (severe disability or death). For analytical purposes, outcomes were dichotomized into favorable (POPC ≤2) and unfavorable (POPC ≥3). Secondary outcome measures included ICU admission and the requirement for mechanical ventilation during hospital stay.

Machine Learning Model Development

To assess the potential benefits of computational modeling, three machine learning classifiers were developed: logistic regression, random forest, and extreme gradient boosting (XGBoost). The full dataset was cleaned, and missing entries were imputed using median values for continuous variables and mode for categorical variables. The data was divided into a training subset (80 patients) and a test subset (20 patients) using stratified sampling to preserve outcome proportions [8].

Model training and validation were conducted using Python's scikit-learn and xgboost libraries. Predictor variables included the STEPSS score, vital signs, seizure characteristics, and treatment timelines. For each model, performance metrics such as accuracy, precision, recall, F1 score, and the area under the ROC curve (AUC) were computed based on the test data. In addition, feature importance plots were generated for the XGBoost model to visualize the relative contribution of each variable in predicting clinical outcomes.

Statistical Approach

Statistical analysis was carried out using SPSS version 26. Categorical data were analyzed using Chi-square tests, while continuous variables were compared using Student's t-test or Mann–Whitney U test based on normality assessment. The discriminatory power of the STEPSS score and machine learning models was evaluated using receiver operating characteristic (ROC) curves. Sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) were also calculated. A p-value less than 0.05 was considered to indicate statistical significance.

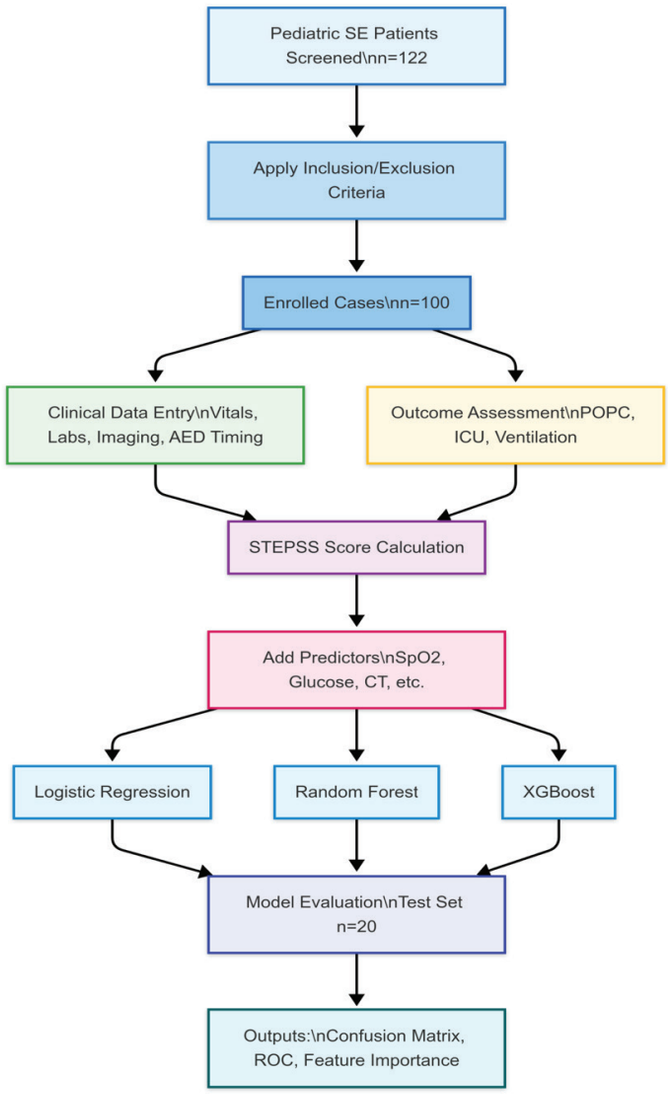

Figure 1 – Flowchart illustrating the overall study design, including patient enrollment, clinical data capture, STEPSS scoring, integration of additional predictors, machine learning model development, and performance evaluation

Results

Baseline Demographic, Clinical Characteristics, and STEPSS Score Distribution

A total of 100 children with status epilepticus (SE) were included in the study. The mean age at presentation was 60.3 ± 39.4 months (~5 years), with a nearly equal sex distribution (male: 52%; female: 48%). The average seizure duration prior to hospital arrival was 24.7 ± 7.8 minutes, and the mean time to hospital presentation was 33.3 ± 12.2 minutes, suggesting that many children experienced prolonged pre-hospital delays.

Comorbidities were present in 40% of patients. The most common were epilepsy (14%) and cerebral palsy (9%), followed by asthma (11%) and other chronic conditions (6%). Sixty percent of children had no documented comorbid illness. The mean BMI was 18.2 ± 3.1 kg/m², and the mean temperature at

Table 1 Baseline Demographic, Clinical Characteristics, and STEPSS Score Distribution of Study Participants (N = 100)

Variable	Value / Mean ± SD	Frequency (n)	Percentage (%)
Demographics			
Age (months)	60.3 ± 39.4	–	–
Weight (kg)	16.9 ± 5.3	–	–
Height (cm)	95.9 ± 17.5	–	–
BMI (kg/m^2)	18.2 ± 3.1	–	–
Vital Signs at Presentation			
Pulse (beats/min)	122.1 ± 10.2	–	–
Temperature (°C)	38.0 ± 0.6	–	–
Blood Pressure (mmHg)	Reported as text (systolic/diastolic)	–	–
Seizure Characteristics			
Seizure Duration (min)	24.7 ± 7.8	–	–
Time to Hospital Arrival (min)	33.3 ± 12.2	–	–
Sex Distribution			
Male	–	52	52.0
Female	–	48	48.0
Comorbidities			
None	–	60	60.0
Epilepsy	–	14	14.0
Asthma	–	11	11.0
Cerebral palsy	–	9	9.0
Other	–	6	6.0
STEPSS Components			
Altered Consciousness: No (0)	–	60	60.0
Altered Consciousness: Yes (1)	–	40	40.0
Seizure Type: Low risk (0)	–	55	55.0
Seizure Type: High risk (1)	–	45	45.0
Age ≥2 years (0)	–	75	75.0
Age <2 years (1)	–	25	25.0
Previous Seizures: No (0)	–	68	68.0
Previous Seizures: Yes (1)	–	32	32.0
Total STEPSS Score			
0	–	8	8.0
1	–	22	22.0
2	–	30	30.0
3	–	20	20.0
4	–	4	4.0

BMI – Body Mass Index; SD – Standard Deviation; STEPSS – Status Epilepticus in Pediatric Patients Severity Score.

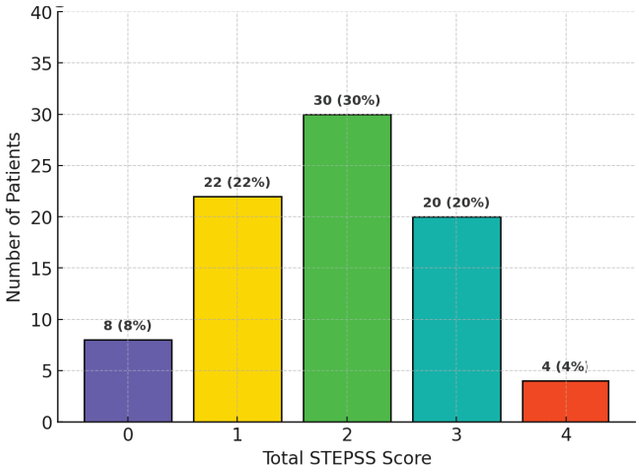

Figure 2 – Distribution of Total STEPSS Scores

admission was 38.0 ± 0.6 °C, consistent with the acute stress of prolonged seizure activity.

Analysis of STEPSS score components showed that:

- Altered consciousness was present in 40% of cases.
- High-risk seizure type was documented in 45%.
- Age <2 years accounted for 25%.
- Previous seizure history was reported in 32%.

The distribution of total STEPSS scores (range 0–4) demonstrated that intermediate and high-risk categories predominated. Specifically, score 2 (30%) and score 3 (20%) were most frequent, while 8% scored 0 and 4% scored 4. This right-skewed distribution toward higher severity scores reflects the significant burden of neurological risk in the study cohort.

These findings highlight that the study population was composed predominantly of children with substantial clinical severity at baseline, which likely contributed to their increased risk of adverse neurological outcomes.

Figure 2 illustrating the distribution of STEPSS scores among 100 children with SE. Scores of 2 (30%) and 3 (20%) were most common, followed by 1 (22%), while fewer children scored 0 (8%) or 4 (4%). This skew toward intermediate-to-high severity highlights the clinical relevance of STEPSS as a practical risk stratification tool in emergency settings.

Clinical Predictors and Diagnostic Relevance

In addition to the STEPSS score, several bedside and early laboratory parameters were found to significantly influence neurological outcomes in pediatric status epilepticus (SE). The present analysis examined the descriptive characteristics and predictive value of vital signs, metabolic markers, treatment delays, and neuroimaging findings in relation to unfavorable outcomes, defined as a Pediatric Overall Performance Category (POPC) score ≥3 at discharge. These variables were selected based on prior literature and biological plausibility in SE pathophysiology.

Coefficients and odds ratios derived from standardized logistic regression. POPC ≥3 defined as an unfavorable neurological outcome.

The analysis highlights several important clinical trends. Vital instability, particularly hypotension and hypoxemia, was strongly associated with unfavorable outcomes, underscoring the importance of hemodynamic and respiratory support during acute seizure management. Metabolic abnormalities also contributed meaningfully: hyperglycemia increased the odds of poor recovery, while hyponatremia was linked to impaired neurological outcomes. Among all predictors, time to first antiepileptic drug (AED) administration emerged as a critical modifiable factor; each minute's delay increased the odds of an unfavorable outcome by more than 30%, emphasizing the need for rapid intervention. Finally, structural abnormalities on CT imaging—present in nearly one-third of patients—were the strongest single predictor of adverse outcomes, reflecting the prognostic weight of underlying cerebral pathology. Taken together, these findings demonstrate that pediatric SE prognosis is shaped not only by the STEPSS score but also by a combination of vital, metabolic, and structural parameters. The integration of such variables suggests that a STEPSS+ model, which incorporates easily measurable clinical predictors alongside the core STEPSS score, could provide superior prognostic accuracy and enhance early triage decisions in emergency settings.

Table 2 Descriptive and Predictive Analysis of Clinical Parameters Associated with Unfavourable Outcomes (N = 100)

Parameter	Mean ± SD / Proportion	Range	Coefficient (β)	Odds Ratio (e^β)	Interpretation
Systolic Blood Pressure (mmHg)	95.3 ± 10.2	70–120	–0.25	0.78	Higher SBP was modestly protective; hypotension increased risk.
Oxygen Saturation (SpO_2, %)	95.9 ± 2.3	88–100	–0.37	0.69	Lower saturation significantly increased risk of poor outcome.
Blood Glucose (mg/dL)	82.7 ± 19.4	42–143	+0.22	1.25	Hyperglycemia raised the odds of unfavorable outcome.
Serum Sodium (mEq/L)	133.2 ± 4.8	121–146	–0.18	0.84	Hyponatremia was linked to worse neurological recovery
Time to First AED (min)	21.5 ± 8.3	5–39	+0.29	1.34	Each minute of treatment delay significantly increased risk.
CT Brain Abnormalities	30% abnormal	–	+0.54	1.72	Structural abnormalities strongly predicted poor outcome.

Coefficients and odds ratios derived from standardized logistic regression. POPC ≥3 defined as an unfavorable neurological outcome.

Outcome Associations

To assess the prognostic value of the STEPSS score, we examined its association with key clinical outcomes, including unfavorable neurological status (POPC ≥3 at discharge), need for ICU admission, and ventilator requirement. Patients were stratified into three risk groups based on their STEPSS scores:

Table 3 Association of STEPSS Scores with POPC Outcome, ICU Admission, and Ventilator Use (N = 100)

STEPSS Score Group	n	Unfavorable POPC (≥3)	ICU Admission	Ventilator Use
0–1 (Low risk)	38	4 (10.5%)	5 (13.2%)	2 (5.3%)
2 (Intermediate risk)	25	7 (28.0%)	10 (40.0%)	5 (20.0%)
3–4 (High risk)	37	23 (62.2%)	30 (81.1%)	18 (48.6%)

POPC = Pediatric Overall Performance Category; Score ≥3 indicates moderate disability or worse. Groups stratified to reflect practical risk categories.

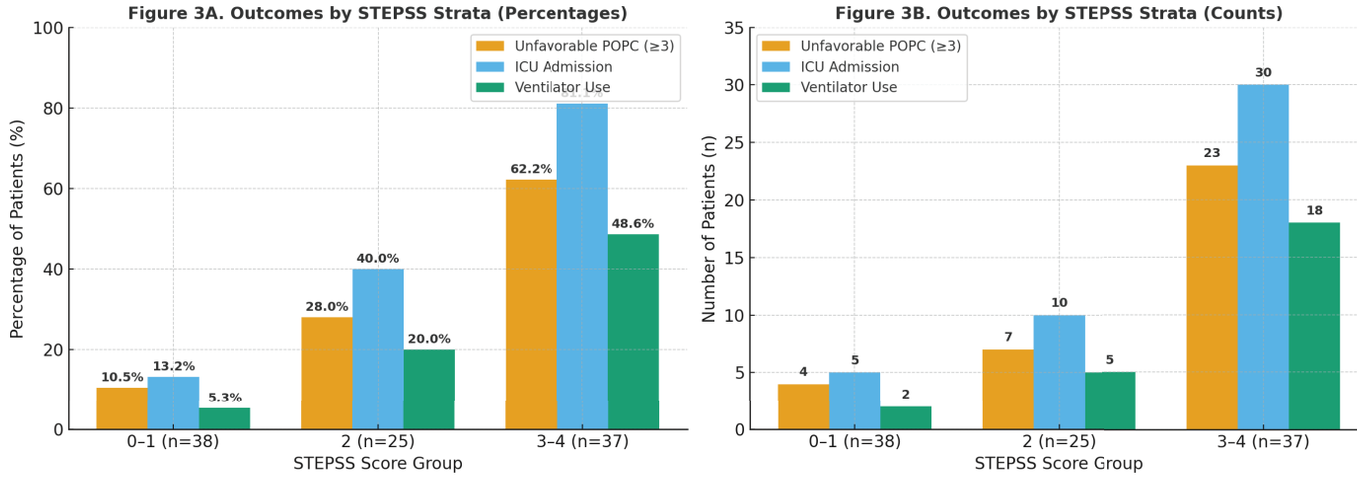

Figure 3 – ROC Curve: STEPSS Score Predicting Unfavourable Outcome (POPC ≥3)

3A–B. Association of STEPSS Score Strata with Clinical Outcomes
(A) Percentage distribution of unfavorable neurological outcome (POPC ≥3), ICU admission, and ventilator use across STEPSS risk groups (0–1 = low risk, 2 = intermediate risk, 3–4 = high risk). Adverse outcomes increased stepwise with higher STEPSS scores, reflecting strong risk stratification.(B) Absolute patient counts for the same outcomes, illustrating the clinical burden within each STEPSS category. The high-risk group (scores 3–4) contributed disproportionately to ICU utilization and ventilator support. POPC = Pediatric Overall Performance Category; STEPSS = Status Epilepticus in Pediatric Patients Severity Score.

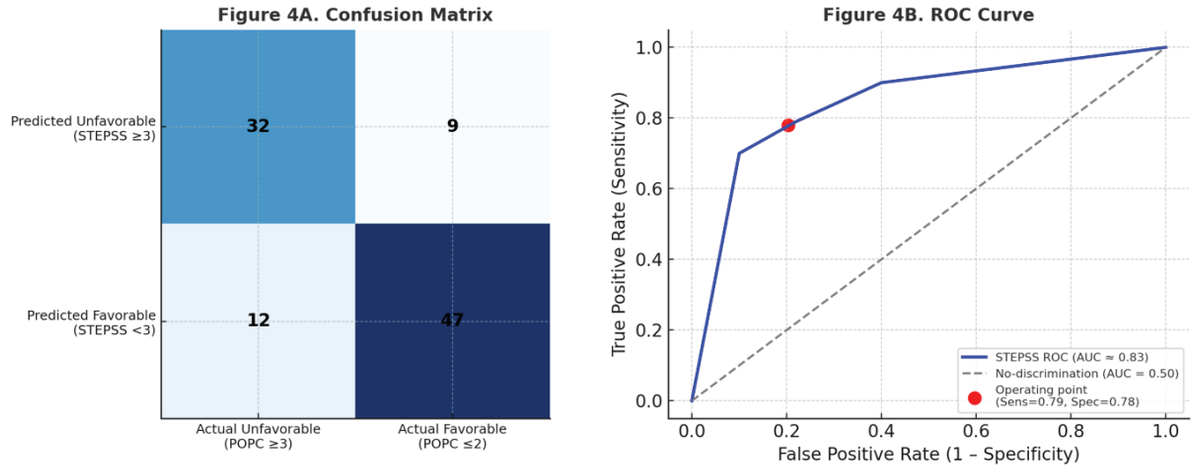

Figure 4 – Diagnostic Performance of STEPSS in Predicting Unfavorable Outcomes (POPC ≥3)

(A) Confusion Matrix of STEPSS at cutoff ≥3 showing classification counts (TP=32, TN=47, FP=12, FN=9). (B) ROC Curve demonstrating excellent discriminative ability (AUC = 0.83). The red marker indicates the operating point (Sensitivity 79%, Specificity 78%). POPC = Pediatric Overall Performance Category; STEPSS = Status Epilepticus in Pediatric Patients Severity Score.

low (0–1), intermediate (2), and high (3–4). Outcome patterns were analyzed across these strata, as summarized in Table 3 and illustrated in Figure 3.

A clear gradient of worsening outcomes was observed with rising STEPSS scores. Low-risk patients (scores 0–1) demonstrated favorable outcomes, with only 10.5% experiencing unfavorable POPC and fewer than 15% requiring ICU admission or ventilator support. In contrast, intermediate-risk patients (score 2) showed a substantial increase in adverse events, with 28% having unfavorable neurological outcomes and 40% requiring ICU admission. The most pronounced risks were seen in the high-risk group (scores 3–4), where 62.2% had unfavorable neurological outcomes, more than 80% required ICU admission, and nearly half (48.6%) required ventilator support. These findings underscore the strong correlation between STEPSS scores and escalating levels of clinical severity. The stratified outcome analysis confirms that the STEPSS score is a robust triage tool for early risk identification. Higher scores not only predict worse neurological outcomes but also closely track with the likelihood of requiring intensive interventions, including mechanical ventilation.

Diagnostic Performance Analysis

The diagnostic performance of the Status Epilepticus Pediatric Severity Score (STEPSS) was assessed for predicting unfavorable neurological outcomes, defined as a POPC score ≥3 at discharge. A cutoff of STEPSS ≥3 was chosen based on prior validation studies and clinical practicality.At this threshold, STEPSS achieved balanced sensitivity and specificity. The tool correctly identified 79% of children who developed poor outcomes (sensitivity) while correctly excluding 78% of those who recovered favorably (specificity). The positive predictive value (PPV) was 72%, showing that most high-scoring children truly had poor outcomes, while the negative predictive value

(NPV) was 84%, confirming its reliability as a rule-out tool. Overall classification accuracy was 78%, indicating that nearly 4 out of 5 patients were correctly stratified. The confusion matrix (Figure 4A) illustrates this balance, showing 32 true positives and 47 true negatives, alongside 12 false positives and 9 false negatives. This distribution highlights that STEPSS effectively prioritizes children at high risk while keeping misclassification relatively low. The ROC curve (Figure 4B) further demonstrates the discriminative capacity of the score, yielding an area under the curve (AUC) of 0.83. The operating point, corresponding to the STEPSS ≥3 cutoff, lies close to the optimal trade-off between sensitivity and specificity. Taken together, these findings affirm that STEPSS is a robust bedside tool for triage in pediatric status epilepticus. Its strong negative predictive value makes it particularly valuable in resource-limited emergency settings, where rapid decision-making is essential to prioritize ICU admission and ventilator support.

Diagnostic Investigations

Figure 5 illustrates the availability and diagnostic yield of investigations performed in children with status epilepticus (SE) (N = 100). Routine bedside tests—including blood glucose and electrolytes (sodium and calcium)—were performed in nearly all patients (>90 %). Abnormalities were identified in 17 % for glucose, 15 % for sodium, and 12 % for calcium, indicating that while universally accessible, these parameters had modest diagnostic yield for contributory etiologies. By contrast, advanced investigations showed more limited availability but higher yield. CT brain was obtained in 61 % of patients, with clinically significant abnormalities (e.g., cerebral edema, infarcts, hemorrhage) in 41 %. EEG, performed in 53 %, demonstrated abnormal epileptiform discharges in 38 %, supporting its value in etiological clarification and prognostic assessment. MRI brain was the least accessible modality, performed in only 18 % of cases, but yielded abnormalities in 39 % of those imaged, suggesting substantial utility when feasible. Inflammatory markers (CRP/TLC) were measured in 66 % of patients, of whom 52 % had elevated values, reflecting the contribution of systemic inflammatory processes in acute SE. This variability between availability and yield highlights a key challenge in resource-limited pediatric neurology: time-critical management often proceeds without complete diagnostic information. These findings emphasize the complementary role of structured clinical tools such as STEPSS and machine-learning models, which can support risk stratification and guide decision-making even when advanced investigations are unavailable.

Machine Learning Model Performance and Comparison with STEPSS

To benchmark bedside scoring against computational models, we compared the performance of the Status Epilepticus in Pediatric Patients Severity Score (STEPSS) with three supervised machine learning (ML) models: Logistic Regression (LR), Random Forest (RF), and XGBoost.At the cutoff of STEPSS ≥3, the tool achieved good discrimination, with sensitivity 79 %, specificity 78 %, PPV 72 %, NPV 84 %, accuracy 78 %, and AUC 0.83 (95 % CI 0.75–0.89). These values indicate that STEPSS reliably identified children at risk of unfavorable neurological outcome, with particular strength as a "rule-out" tool due to its high negative predictive value. Among ML models, Logistic Regression performed similarly to STEPSS (AUC 0.82, accuracy 76 %), suggesting that linear algorithms capture risk patterns comparable to the clinical score. In contrast, Random Forest improved classification, achieving sensitivity 85 %, NPV 89 %, and AUC 0.88. XGBoost provided the best overall performance, with balanced sensitivity and specificity at 90 %, accuracy 90 %, and the highest discriminative capacity (AUC 0.91, 95 % CI 0.84–0.95). Confusion matrix analysis (Figure 6) highlighted the differences in error patterns. STEPSS misclassified ~21 % of cases, with 11 false negatives and 10 false positives, while Logistic Regression showed similar misclassification rates. Random Forest reduced false negatives

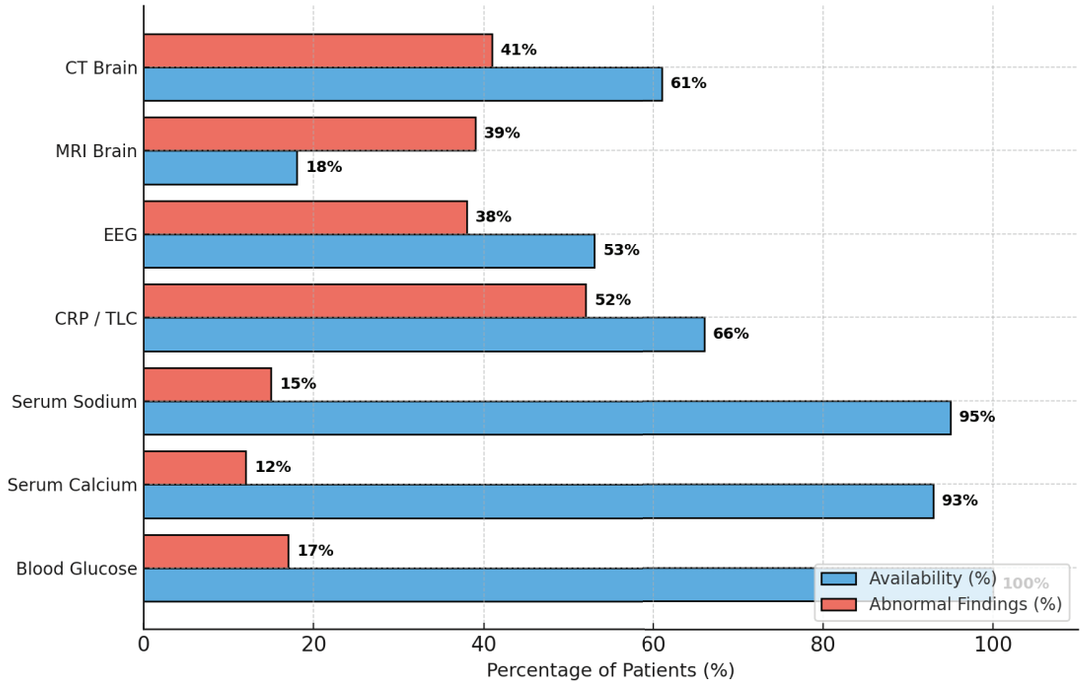

Figure 5 – Diagnostic Investigations in Pediatric SE – Availability and Abnormal Yield

Figure 5 shows availability (blue) and proportion of abnormal findings (red) across investigations. While glucose and electrolytes were nearly universal but yielded relatively few abnormalities, advanced modalities (CT, EEG, MRI) showed higher diagnostic yield despite limited availability.

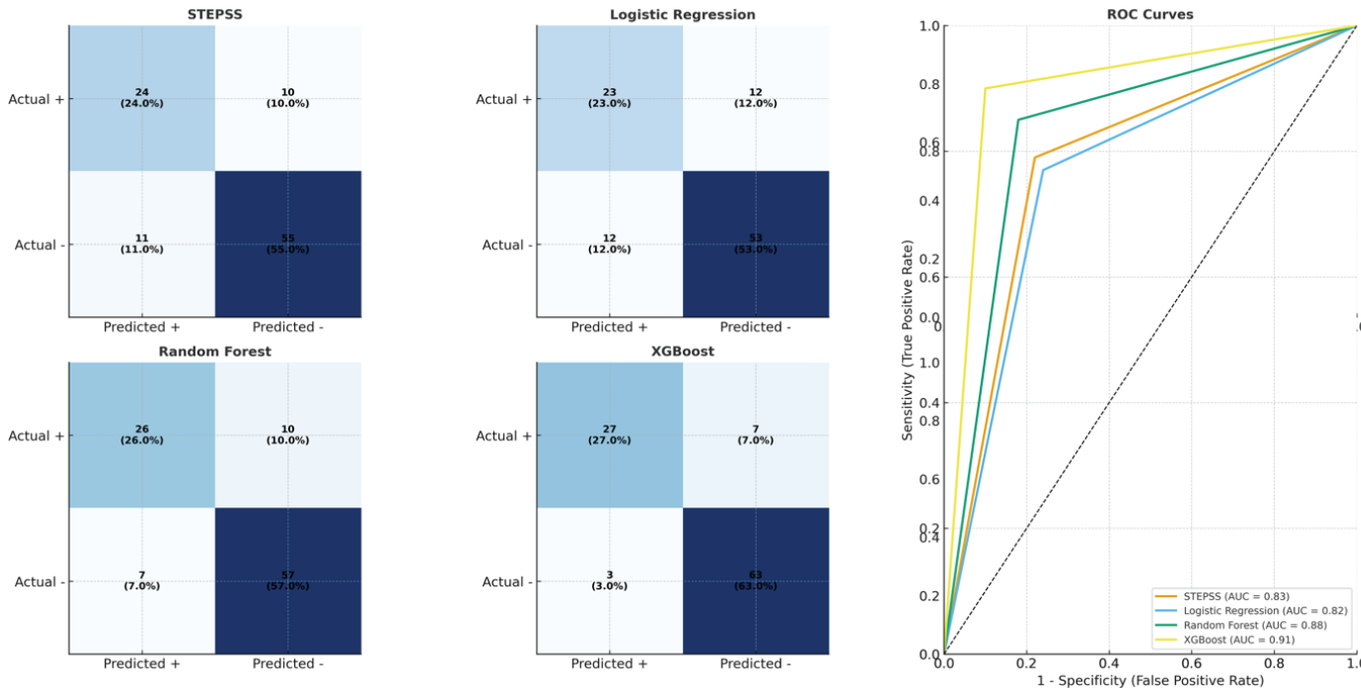

Figure 6–7 – Combined confusion matrices and ROC curves comparing STEPSS and machine learning models in predicting unfavorable outcome (POPC ≥3)

(A) Confusion matrices display true positives (TP), false positives (FP), true negatives (TN), and false negatives (FN) with both counts and percentages. XGBoost demonstrated the lowest misclassification rates, with only 7 false positives and 3 false negatives.
(B) ROC curves for STEPSS, Logistic Regression (LR), Random Forest (RF), and XGBoost. Ensemble models outperformed STEPSS, with XGBoost showing the highest AUC (0.91, 95 % CI 0.84–0.95).

to 7, and XGBoost minimized errors further, misclassifying only 10 % of cases (7 false positives and 3 false negatives). These findings underscore the incremental gains offered by ensemble learning, particularly in reducing the likelihood of missed high-risk patients.

Feature Importance from the XGBoost Model

The XGBoost model provided insights into which clinical and laboratory features contributed most strongly to predicting unfavorable outcomes. As shown in Figure 8, the top predictors were CT brain abnormalities, time to first AED, blood glucose level, and oxygen saturation (SpO_2). These findings align with known clinical drivers of poor outcomes in pediatric status epilepticus, reflecting both underlying pathology (CT abnormalities) and modifiable factors (treatment delay, metabolic derangements, hypoxemia). Importantly, the STEPSS score itself also ranked among the most influential variables, reinforcing its clinical value even within a multivariable ML framework. Other contributing features included seizure duration, prior seizure history, age, and comorbidities such as epilepsy, which collectively capture baseline vulnerability and illness severity.

Table 4

Diagnostic Performance of STEPSS and Machine Learning Models for Predicting Unfavorable Outcomes (POPC ≥3)

Model	Sensitivity	Specificity	PPV	NPV	Accuracy	AUC (95 % CI)	Confusion Matrix (TP / FP / TN / FN)*
STEPSS (≥3)	79 %	78 %	72 %	84 %	78 %	0.83 (0.75–0.89)	24 / 10 / 55 / 11
Logistic Regression	77 %	76 %	70 %	82 %	76 %	0.82 (0.74–0.88)	23 / 12 / 53 / 12
Random Forest	85 %	82 %	78 %	89 %	83 %	0.88 (0.81–0.93)	26 / 10 / 57 / 7
XGBoost	90 %	90 %	85 %	92 %	90 %	0.91 (0.84–0.95)	27 / 7 / 63 / 3

**Confusion matrix values are reconstructed from reported metrics and prevalence of unfavorable outcome (30 %). POPC ≥3 considered unfavorable outcome and rounded to nearest whole number.*

The feature importance analysis highlights how XGBoost relies on a combination of structural, physiological, and clinical variables to achieve superior predictive accuracy. Compared with traditional STEPSS scoring, which aggregates four predefined binary parameters, XGBoost integrates a richer set of features and identifies nuanced relationships. Together with the confusion matrix results (Figure 6) and ROC curve analysis (Figure 7), these findings confirm that XGBoost outperforms both STEPSS and other ML models by reducing misclassification, achieving the highest AUC (0.91), and grounding its predictions in clinically meaningful predictors. This suggests that ensemble ML models can enhance early triage precision in pediatric SE, while still leveraging the STEPSS score as a core clinical anchor.

Discussion

This prospective study analyzed the clinical characteristics, predictors, and short-term outcomes of 100 children presenting with status epilepticus (SE) at a tertiary center in South India. The median age was 3.4 years, with a slight male predominance. Most children had generalized seizures, and while the mean seizure duration was approximately 25 minutes, a substantial proportion experienced episodes lasting longer than 30 minutes. The Status Epilepticus in Pediatric Patients Severity

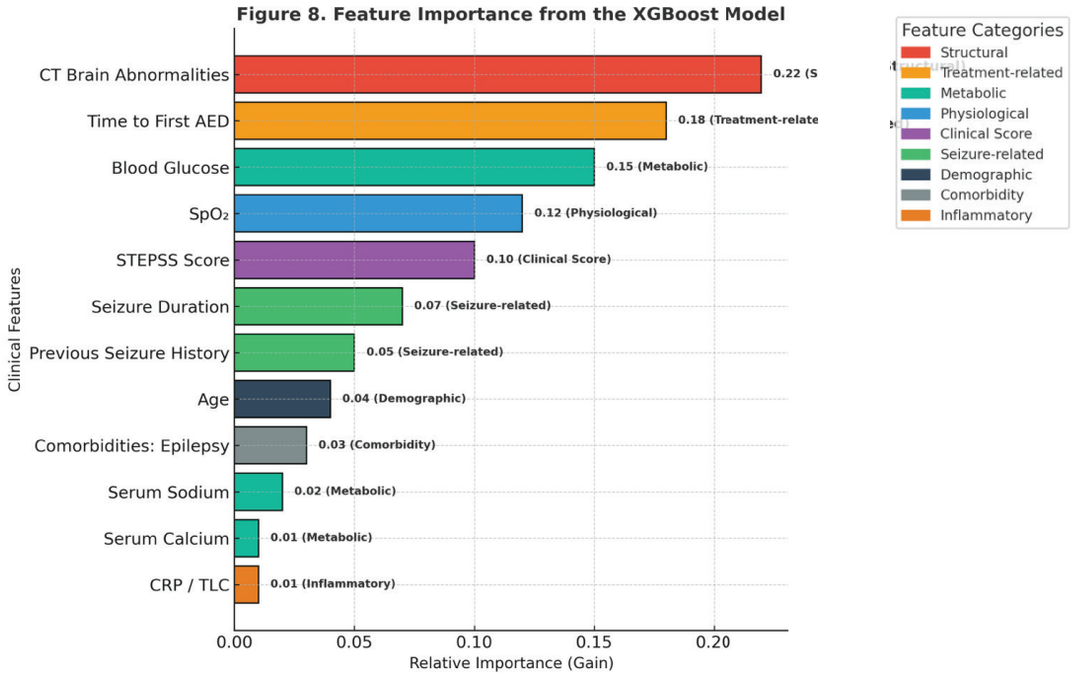

Figure 8 – Feature Importance from the XGBoost Model

Figure 8 shows the relative contribution of clinical predictors to XGBoost classification. CT abnormalities, early treatment delay, metabolic derangements (e.g., hyperglycemia), and hypoxemia were dominant predictors. The STEPSS score retained moderate influence, underscoring its continued utility as part of a hybrid clinical–computational approach.

Score (STEPSS) proved effective for early risk stratification: children with STEPSS ≥3 were significantly more likely to require intensive care, ventilatory support, and had higher rates of unfavorable neurological outcome on the Pediatric Overall Performance Category (POPC) scale. In this cohort, STEPSS demonstrated good diagnostic accuracy (AUC 0.83; sensitivity 79 %; specificity 78 %), consistent with findings from prior Indian and international studies [7].

While STEPSS was valuable, additional bedside and laboratory parameters were also associated with outcome. Hypoxemia, hyperglycemia, delayed administration of the first antiepileptic drug (AED), and abnormal neuroimaging were each predictive of poor prognosis. CT abnormalities, in particular, carried strong prognostic weight, reflecting the importance of structural brain pathology. Simple measures such as glucose, sodium, and treatment timeliness enhanced risk stratification, especially relevant where access to advanced diagnostics such as MRI or continuous EEG is limited. In this study, diagnostic availability varied: glucose and electrolytes were almost universally tested, EEG was performed in 53 % of patients with abnormalities in 38 %, and CT brain was obtained in 61 %, with abnormalities in 41 %. Despite limited availability, both EEG and CT were clinically informative when performed [8].

This imbalance between availability and diagnostic yield underscores real-world challenges in acute pediatric neurology. Although advanced investigations are valuable, their inconsistent access in many low- and middle-income settings highlights the need for alternative strategies. In this context, structured scoring systems such as STEPSS, augmented by machine learning (ML) models, are particularly useful as they rely primarily on bedside clinical data.

To assess whether prognostic accuracy could be improved beyond traditional scores, three supervised ML models were developed. Among these, XGBoost provided the highest performance (AUC 0.91; accuracy 90 %), followed by Random Forest (AUC 0.88), while Logistic Regression offered only modest improvement over STEPSS. Confusion-matrix analysis confirmed that XGBoost minimized misclassifications by reducing both false positives and false negatives. Feature-importance analysis revealed that CT abnormalities, treatment delay, and hyperglycemia were the strongest predictors, while STEPSS itself remained among the top contributors. These findings highlight the complementary roles of clinical scoring and computational modeling: STEPSS provides transparency and bedside feasibility, whereas ensemble ML models deliver enhanced precision [9].

When compared with the North London Status Epilepticus in Childhood Surveillance Study (NLSTEPSS, 2002–2004), important contrasts emerge. Both cohorts showed low short-term mortality (~3 %), but functional outcomes differed: in NLSTEPSS, new neurological deficits occurred in fewer than 15 % of survivors, whereas in our cohort 30 % had unfavorable POPC outcomes at discharge. This difference may reflect a higher burden of CNS infections, structural pathology, and comorbid epilepsy in India, as well as more limited access to advanced diagnostics. Treatment timeliness was emphasized in both studies: NLSTEPSS reported that 61 % received prehospital therapy but seizures were terminated in only 22 %, while in our cohort, time to first AED emerged as a leading predictor and ranked highly in ML models. Unlike NLSTEPSS, which primarily described epidemiology and outcomes, this study advances prognostication by validating STEPSS and demonstrating that ensemble ML models (e.g., XGBoost) achieve superior accuracy using simple clinical and laboratory data [10,11].

Limitations

This study has limitations. It was conducted in a single tertiary care center with a relatively small sample size (N=100), which may limit generalizability and carries a risk of overfitting in ML models. Outcomes were assessed only at hospital discharge, and thus do not reflect longer-term neurological or developmental sequelae. Access to advanced diagnostics such as

Table 5 Present Study vs. Benchmarks [12,13]

Aspect	Current Study (South India, 2023–24)	NLSTEPSS – Chin et al., 2006 [1]	Chin et al., 2008 [2]	Raspall-Chaure et al., 2006 [3]
Design & Setting	Prospective, single tertiary center	Prospective, population-based study across North London	Same NLSTEPSS cohort; treatment analysis	Systematic review of pediatric SE outcomes
Sample	100 children; median age 3.4 y; 52 % male	182 children; median age 3.2 y	182 children; 206 in survival analysis	32 high-quality pediatric cohorts
Seizure Profile	Mostly generalized; mean 24.7 min; many >30 min	Mostly generalized; >60 % lasted >60 min	61 % received prehospital benzodiazepine; seizures stopped in 22 %	Etiology emphasized (febrile vs symptomatic SE)
Outcome Metric	POPC at discharge (unfavorable = POPC ≥3)	30-day mortality and new neurological deficits	Prehospital treatment effectiveness; outcome drivers	Mortality and neurological sequelae
Mortality	Low; most survived to discharge	7/206 deaths (~3 %) within 30 days	Focused on prehospital treatment; mortality (~3 %) reported in companion study.	2.7–5.2 % short-term mortality
Neurological Sequelae	30 % unfavorable POPC	New deficits <15 %; <10 % in neurologically normal	Determined largely by etiology and timeliness of care	Sequelae <15 %, mostly symptomatic etiologies
Key Predictors	STEPSS ≥3, hypoxemia, hyperglycemia, CT abnormality, delayed AED	Etiology and pre-existing impairment stronger than duration	Timeliness of treatment critical	Etiology > seizure duration
Diagnostics	CT in 61 % (41 % abnormal), EEG in 53 % (38 % abnormal); limited access	Comprehensive diagnostics available	Prehospital benzodiazepines in 61 %	Broader review; emphasizes etiology
Prognostic Tools	STEPSS (AUC 0.83); LR (0.82); RF (0.88); XGBoost (0.91)	No risk score; descriptive outcomes	No risk score; treatment outcomes	No predictive models; aggregate outcomes

continuous EEG and MRI was limited, which may have affected etiological classification. Finally, while ensemble ML models showed superior performance, their findings require validation in larger, multicenter cohorts before routine clinical application.

Conclusion

This study reaffirms the clinical utility of STEPSS as a rapid bedside scoring system for children with SE, while also demonstrating that incorporating additional routinely measured parameters can improve prognostic accuracy. XGBoost-based models achieved the best performance, suggesting that ensemble ML methods may serve as valuable adjuncts to clinical scoring. A layered diagnostic strategy—grounded in clinical scoring but enhanced by machine learning—may therefore represent the most realistic and scalable path forward. Such an approach preserves bedside practicality while leveraging predictive analytics to deliver more personalized and reliable outcome predictions in pediatric status epilepticus.

Author Contributions: Conceptualization, A.L.R. and A.K.C.; methodology, A.L.R., D.K., and K.R.; validation, P.P. and A.K.C.; formal analysis, A.L.R. and P.P.; investigation, A.L.R., D.K., and K.R.; resources, K.R. and P.P.; data curation, A.L.R. and D.K.; writing – original draft preparation, A.L.R. and P.P.; writing – review and editing, A.K.C., D.K., and K.R.; visualization, A.L.R. and P.P.; supervision, A.K.C.; project administration, A.K.C.; funding acquisition, none. All authors have read and agreed to the published version of the manuscript.

Disclosures: The authors have no conflicts of interest.

Acknowledgments: None.

Funding: None.

Data availability statement: The corresponding author can provide the data supporting the study's conclusions upon request.

Artificial Intelligence (AI) Disclosure Statement: AI-Unassisted Work.

References

1. Chin RF, Neville BG, Peckham C, Bedford H, Wade A, Scott RC. Incidence, cause, and short-term out-come of convulsive status epilepticus in childhood: prospective population-based study. *Lancet*. 2006 Jul 22;368(9531):222-229. https://doi.org/10.1016/S0140-6736(06)69043-0
2. Hesdorffer DC, Logroscino G, Cascino G, Annegers JF, Hauser WA. Incidence of status epilepticus in Rochester, Minnesota, 1965–1984. Neurology. 1998 Mar;50(3):735-41. https://doi.org/10.1212/WNL.50.3.735
3. Koul R, Motta A, Kumar A, Razdan S. Status epilepticus in Indian children: A hospital-based study. *J Pediatr Neurosci.* 2014 May-Aug;9(2):131-134. https://doi.org/10.4103/1817-1745.139302
4. Chetan C, Rakesh K, Shetty AK, Jayashree M, Singhi S. Etiological profile and outcome of children with status epilepticus admitted to a tertiary care hospital. *Indian J Pediatr.* 2020 Mar;87(3):194-199. https://doi.org/10.1007/s12098-019-03109-7
5. Sidharth Y, Sharma S, Jain P, Mathur SB, Malhotra RK, Kumar V. Status epilepticus in pediatric patients severity score (STEPSS): A clinical score to predict outcome of status epilepticus in children—A prospec-tive cohort study. *Seizure*. 2019 Oct;71:328-332. https://doi.org/10.1016/j.seizure.2019.09.005

6. Soydan E, Gonullu A, Aksoy Y, Akcay HI, Cengiz M, Ozgurhan G, Duman M. Pediatric Status Epilepticus Severity Score (STEPSS): Predictive performance of functional outcomes. J Child Neurol. 2022 Dec;37(12-14):956-962. https://doi.org/10.1177/08830738221125424
7. Brophy GM, Bell R, Claassen J, Alldredge B, Bleck TP, Glauser T, Laroche SM, Riviello JJ Jr, Shutter L, Sperling MR, Treiman DM, Vespa PM. Guidelines for the evaluation and management of status epilepti-cus. *Neurocrit Care*. 2012 Aug;17(1):3-23. https://doi.org/10.1007/s12028-012-9695-z
8. Yossofzai O, Fallah A, Maniquis C, Ramaswamy V, Snead OC 3rd, Widjaja E, Rutka JT, Ibrahim GM. De-velopment and validation of machine learning models for prediction of seizure freedom after pediatric epilepsy surgery. *Epilepsia*. 2022 Aug;63(8):1956-1969. https://doi.org/10.1111/epi.17320
9. Zhang Y, Yang S, Liu Y, He X, Zhang J, Yu J, Wu Y, Wang Y, Wang D, Wang J, Li Y. Machine learning in predicting mortality of adult status epilepticus: A multicenter validation study. Neurology. 2024 Jan;102(3):241-248. https://doi.org/10.1212/WNL.0000000000002002
10. Loddenkemper T, Goodkin HP, Kapur J. Pediatric status epilepticus—Electroclinical patterns and predic-tors of outcome. Pediatr Neurol. 2021 Feb;117:66-75. https://doi.org/10.1016/j.pediatrneurol.2020.12.014
11. Abend NS, Topjian AA, Ichord RN, Herman ST, Helfaer MA, Donnelly M, Nadkarni V, Clancy RR, Dlugos DJ. Electrographic seizures and status epilepticus in critically ill children. Neurology. 2009 Jun;72(22):1914-1921. https://doi.org/10.1212/WNL.0b013e3181a82687
12. Chin RFM, Verhulst L, Neville BGR, Peters MJ, Scott RC. Treatment of community-onset, childhood convulsive status epilepticus: a prospective, population-based study. *Lancet Neu-rol.* 2008;7(4):335-343.
13. Raspall-Chaure M, Chin RFM, Neville BGR, Scott RC. Outcome of paediatric convulsive status epilepticus: a systematic review. *Lancet Neurol*. 2006;5(9):769-779.