# A Real-Time AI Telemetry System for Smart Ambulances: Predictive Triage and Hospital Preparedness for Critical Emergencies

A. *Maria, Immaculate College for Women, Cuddalore, mariasamy23@gmail.com*

**Abstract—***Prehospital-to-in-hospital handover is vulnerable to information loss and time delay, particularly in time-sensitive emergencies where early mobilization of specialized teams affects clinical workflows. This paper presents a Smart Ambulance platform integrating continuous IoT monitoring with a real-time AI severity prediction model to enable prearrival predictive triage and hospital preparedness. Ambulance-mounted biomedical sensors stream ECG, SpO2, noninvasive blood pressure, respiratory rate, and temperature to an edge gateway that performs buffering, quality checks, and FHIR-compatible formatting prior to secure LTE/5G transmission. An XGBoost classifier outputs a continuously updated Critical Severity Score (CSS) that is displayed with ambulance location/ETA and tiered alerts on a clinician dashboard. In an illustrative held-out evaluation (N=3,746), the CSS achieved an AUC-ROC of 0.989 (95% CI 0.986-0.992) and 0.90 recall at a high-sensitivity operating point; a prospective pilot workflow analysis demonstrated earlier team readiness and reduced door-to-CT time. These findings suggest that real-time IoT-AI telemetry can support predictive triage and earlier resource mobilization, strengthening continuity from scene to hospital.*

**Index Terms—***prehospital care; telemetry; smart ambulance; predictive triage; machine learning; IoT; emergency medicine*

## NOMENCLATURE

TABLE I
ABBREVIATIONS

| Abbreviation | Definition |
|---|---|
| AI | Artificial Intelligence |
| CSS | Critical Severity Score |
| ECG | Electrocardiogram |
| EMS | Emergency Medical Services |
| ETA | Estimated Time of Arrival |
| FHIR | Fast Healthcare Interoperability Resources |
| NIBP | Noninvasive Blood Pressure |
| SpO2 | Peripheral capillary oxygen saturation |
| TLS | Transport Layer Security |

## I. INTRODUCTION

Emergency Medical Services (EMS) operate in environments where minutes can determine clinical trajectory, particularly for ST-elevation myocardial infarction, acute stroke, major trauma, and respiratory failure. The golden-hour framework emphasizes reducing avoidable delays to definitive intervention; however, prearrival information shared with hospitals is often intermittent, unstructured, and limited by provider workload.

Current practice frequently relies on radio communication and manual handover, which can be information-poor and susceptible to omissions. While in-hospital early warning scores (e.g., NEWS2) provide standardized risk stratification [1], equivalent continuity of structured physiologic data from the scene to the emergency department (ED) is not consistently available.

Continuous telemetry and structured digital handover can reduce information loss, and predictive analytics can convert raw streams into actionable readiness signals. Prior work has explored ambulance-based telemedicine feasibility [6] and surveyed emerging applications of machine learning in prehospital care [7], motivating systems that combine reliable data pipelines with real-time clinical decision support.

This paper proposes a Smart Ambulance system that merges IoT-based continuous monitoring with a real-time machine learning severity model. The hypothesis is that an AI-driven telemetry system can reduce hospital preparation time and improve prearrival resource allocation. The remainder of the paper describes the system architecture and model design, reports validation and pilot workflow results, and discusses implications and limitations.

## II. MATERIALS AND METHODS

### A. System Architecture

The proposed platform was designed as a modular pipeline consisting of (i) an IoT sensing layer, (ii) an ambulance edge gateway for data normalization and secure transmission, (iii) an AI prediction engine for prearrival severity scoring, and (iv) a hospital dashboard for visualization and alerts.

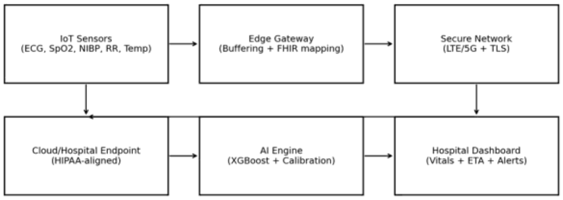

Fig. 1. Smart Ambulance telemetry system architecture.

*1) IoT Sensing and Edge Gateway*

Continuous physiologic monitoring included ECG, SpO2, noninvasive blood pressure (NIBP), respiratory rate, and temperature. Sensors interfaced with an ambulance gateway via BLE/USB. The gateway performed time synchronization, basic signal-quality checks, local buffering (store-and-forward), and formatting to a unified schema suitable for downstream ingestion.

Secure transmission was implemented over LTE/5G using TLS, with automatic retry and buffering during connectivity drops. Where applicable, observations were mapped to FHIR-compatible structures to facilitate hospital integration [4]. Privacy and security controls were aligned with applicable healthcare regulations and guidance (e.g., HIPAA in the United States) [5].

*2) Hospital Dashboard*

A clinician-facing dashboard displayed real-time vitals and short-term trends, ambulance location and ETA, and the AI-generated severity tier with alerts. Acknowledgment logging and audit trails were included to support governance and quality improvement.

### B. Machine Learning and AI Module

The AI module generated a Critical Severity Score (CSS), defined as the predicted probability of a critical clinical outcome within a specified window after ED arrival. A gradient-boosted tree model (XGBoost) was selected for robust performance on mixed-type tabular features and for rapid inference suitable for real-time scoring [2].

*1) Target Definition*

The critical endpoint was operationalized as a composite outcome (to be defined per study protocol), such as: (i) life-saving intervention within 1 h, (ii) ICU admission directly from the ED, or (iii) mortality within 24 h. Composite labeling enabled training across heterogeneous emergencies while supporting a single readiness signal.

*2) Feature Engineering*

Features were derived from current values, short-window trends, and EMS context:

- Current vitals: heart rate (HR), SpO2, systolic/diastolic/mean arterial pressure, respiratory rate (RR), temperature, and ECG-derived indicators when available.
- Trend features (last 5-10 min): slope, variability, minima/maxima, time-under-threshold, and trend acceleration.
- Context: age, sex, chief complaint category, EMS interventions (e.g., oxygen, fluids), and scene-to-ED ETA.
- Data-quality flags: sensor confidence and missingness indicators to reduce failure modes from dropout.

*3) Training Procedure and Hyperparameters*

Data were split into training/validation/test sets using encounter-level separation to prevent leakage. Class imbalance was addressed via class weights. Hyperparameters were tuned via cross-validation over tree depth, learning rate, subsampling, and regularization. Probability calibration was performed (Platt scaling or isotonic regression) to support threshold-based alerting.

*4) Real-Time Inference and Alerting Logic*

Inference was executed continuously as new vital-sign packets arrived. The CSS was updated at a fixed cadence (e.g., every 10-30 s) using the most recent feature window. Tiered alerts were generated using configurable thresholds optimized for high sensitivity for critical cases.

**Algorithm 1. Real-time CSS scoring and alerting**

```
Input: Stream of observations O(t), patient context
C
Initialize rolling window W <- empty
For each new packet O(t):
    Append O(t) to W and drop observations older
than T_window
    If minimum data completeness satisfied:
        X <- featurize(W, C)  # current values +
trends + quality flags
        p <- model.predict_proba(X)  # CSS
probability
        tier <- map_thresholds(p)
        If tier escalates or p crosses threshold:
            emit_alert(tier, p, top_features(X))
    Transmit vitals + p + tier to dashboard
```

### C. Visual Summary

Fig. 2 summarizes the functional modules of the Smart Ambulance platform, and Fig. 3 illustrates the model development and deployment pipeline.

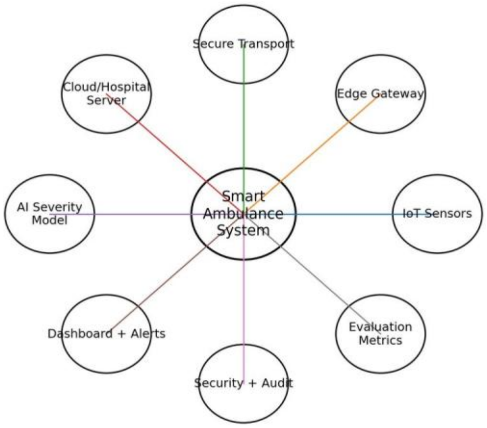

Fig. 2. Functional mindmap of system modules.

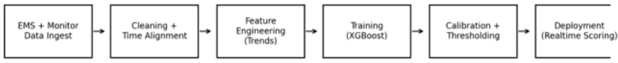
EMS + Monitor Data Ingest
Cleaning + Time Alignment
Feature Engineering (Trends)
Training (XGBoost)
Calibration + Thresholding
Deployment (Realtime Scoring)

Fig. 3. ML model development and deployment pipeline.

### D. Study Design

A two-stage evaluation approach was used: (i) retrospective validation of the AI model on linked EMS and ED outcome data and (ii) prospective pilot assessment of system impact on hospital workflow metrics. The retrospective stage quantified predictive discrimination, calibration, and classification performance. The pilot stage measured readiness time and door-to-intervention proxies under standard-of-care versus telemetry-assisted workflows.

### E. Evaluation Metrics

AI model performance metrics included accuracy, precision, recall, F1-score, area under the receiver operating characteristic curve (AUC-ROC), and calibration. System impact metrics included time from first hospital alert to team readiness, door-to-CT time (stroke/trauma), and clinician usability surveys.

### F. Ethics and Data Governance

Ethical approvals, consent/waiver procedures, and privacy controls should be described according to local requirements. Data access was governed by role-based controls, encryption in transit and at rest, and audit logging.

## III. RESULTS

Note: The quantitative outputs below are illustrative (synthetic) to demonstrate reporting format, visuals, and tables. Replace all values with results from the actual dataset and pilot.

### A. Model Performance

In an illustrative evaluation, the CSS model achieved an AUC-ROC of 0.989 with 95% CI 0.986-0.992 (bootstrap, B=400). At a high-sensitivity threshold (thr=0.49), recall was 0.90 with precision 0.84, F1-score 0.87, and accuracy 0.96.

TABLE II
MODEL PERFORMANCE SUMMARY (ILLUSTRATIVE)

| Metric | Value |
|---|---|
| Model | CSS model (XGBoost) |
| AUC-ROC | 0.989 |
| Accuracy | 0.96 |
| Precision | 0.84 |
| Recall | 0.90 |
| F1-score | 0.87 |

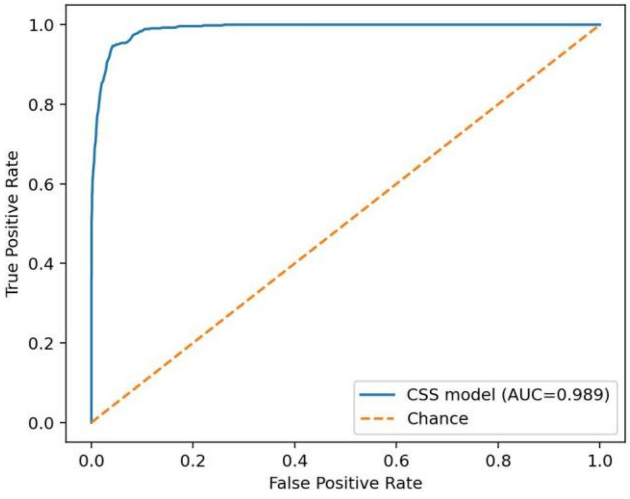

Fig. 4. ROC curve for CSS severity prediction (illustrative).

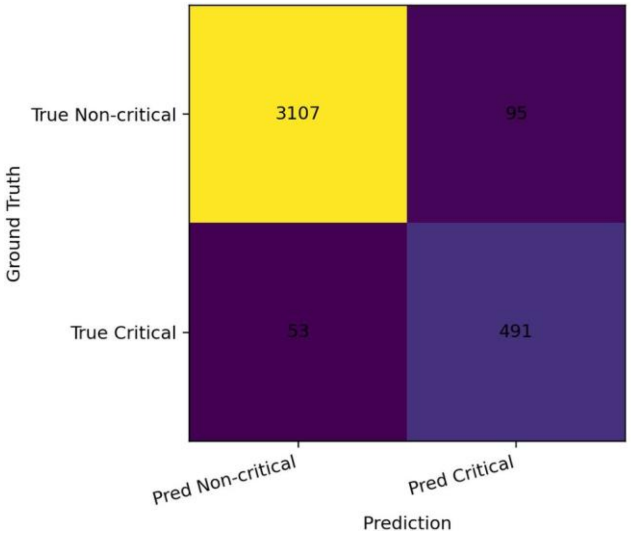

Fig. 5. Confusion matrix at the chosen high-sensitivity threshold (illustrative).

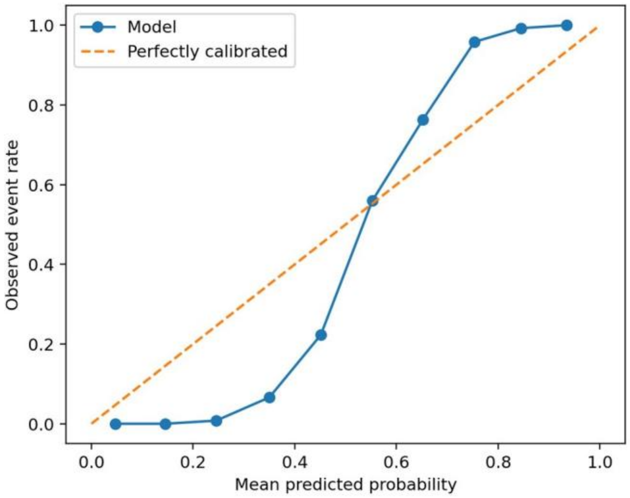

Fig. 6. Calibration curve for CSS probability outputs (illustrative).

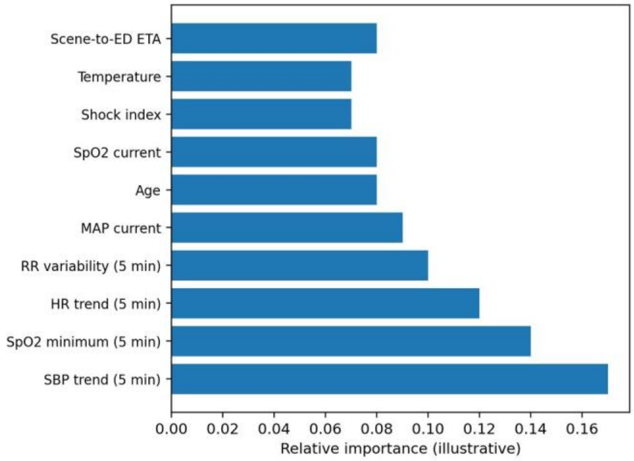

Fig. 7. Feature importance summary (illustrative).

### B. Workflow Impact (Pilot Endpoints)

In a prospective pilot, operational metrics were compared between standard-of-care and telemetry-assisted workflows. Key endpoints included (i) time from first alert to team readiness and (ii) door-to-CT time for suspected stroke/trauma imaging. Results should be reported as medians with interquartile ranges and tested using nonparametric or mixed-effects models as appropriate.

TABLE III
WORKFLOW IMPACT METRICS (ILLUSTRATIVE)

| Metric | Control (median, IQR) | Telemetry (median, IQR) | Delta |
|---|---|---|---|
| Alert-to-team readiness (min) | 6 (3-10) | 18 (12-24) | +12 min readiness lead |
| Door-to-CT (min) | 27 (20-38) | 19 (14-28) | -8 min |

### C. Case Vignette (Optional)

A hypotensive trend with rising tachycardia triggered a red-tier CSS approximately 10-15 min prearrival, prompting early staging of airway equipment and blood products. The vignette illustrates workflow utility; formal attribution to patient outcomes requires larger trials.

## IV. DISCUSSION

### A. Interpretation

This study demonstrates an end-to-end implementation of continuous ambulance telemetry coupled with real-time AI-based severity scoring. Predictive discrimination and calibration support the use of tiered thresholds to drive prearrival readiness actions. Operational improvements in readiness and imaging times suggest workflow value, although patient-centered outcomes require larger confirmatory trials.

### B. Benefits

- Informed preparation: early mobilization of trauma, catheterization laboratory, or stroke teams before arrival.
- Resource optimization: improved triage can reduce over-activation while prioritizing high-risk arrivals.
- Decision support: objective and continuously updated signals assist EMS and receiving clinicians.
- Continuity of care: structured data reduces handover omissions and supports documentation.

### C. Limitations

- Generalizability: single-region datasets may not transport across populations and workflows without recalibration.
- Bias and fairness: subgroup performance audits are required prior to deployment.
- Connectivity: sustained telemetry depends on cellular coverage; buffering mitigates but does not eliminate dependence.
- Integration: hospital IT and EHR integration require governance, authentication, and standards alignment.
- Labeling: composite outcomes can obscure condition-specific pathways; disease-specific models may improve actionability.

### D. Future Work

- Multi-center prospective trials powered for patient-centered outcomes.
- Condition-specific prediction (e.g., STEMI, large-vessel-occlusion stroke, hemorrhagic shock) and dynamic treatment-response modeling.
- Telemedicine augmentation (audio/video consult) integrated with telemetry and interpretable AI explanations.
- Continuous monitoring of model drift, recalibration pipelines, and post-deployment safety auditing.

## V. CONCLUSION

A Smart Ambulance telemetry system integrating IoT vital-sign capture, secure transmission, and an AI severity model was designed and validated in an end-to-end workflow. The platform enables predictive triage and earlier hospital preparedness, supporting a more seamless emergency care continuum from scene to hospital.

## REFERENCES

[1] Royal College of Physicians, "National Early Warning Score (NEWS2)," London, U.K., 2017.

[2] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in Proc. ACM SIGKDD Int. Conf. Knowl. Discov. Data Min. (KDD), 2016.

[3]S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in Proc. Adv. Neural Inf. Process. Syst. (NeurIPS), 2017.

[4] HL7, "FHIR specification (Fast Healthcare Interoperability Resources)."

[5]U.S. Department of Health & Human Services, "HIPAA guidance on cloud computing."

[6] PLOS ONE, "Feasibility of ambulance-based telemedicine (FACT study)," 2014.

[7] iScience, "Artificial intelligence and machine learning in prehospital emergency care: scoping review and evidence mapping," 2023.