Algorithmic Triage in Emergency Departments: Designing AI-Enhanced Clinical Prioritization for Equity and Efficiency

James C. Ingram*
School of Data Science and Society
University of North Carolina at Chapel Hill
Chapel Hill, NC 27514
`jcingbme@unc.edu`

Abstract

This work presents a comprehensive framework for implementing AI-augmented triage support systems in emergency departments, designed to enhance throughput efficiency while mitigating disparities in clinical prioritization. The proposed solution integrates machine learning capabilities with human oversight mechanisms to address the dual challenges of operational efficiency and algorithmic equity in high-acuity clinical environments. Through systematic stakeholder analysis, ethical framework development, and robust governance structures, this approach provides a scalable model for responsible AI deployment in emergency medicine.

1 Introduction

In overcrowded emergency departments, patients may experience extended wait times that can adversely impact clinical outcomes and, in severe cases, result in preventable mortality [1],[2]. This work proposes an AI-augmented triage system designed to assist emergency department staff in optimizing care prioritization through integration of real-time patient data, historical outcomes, and symptom severity assessments. The implementation framework addresses critical considerations including ethical compliance, privacy protection, and equitable performance across diverse patient populations.

Data collection will utilize de-identified patient records subjected to rigorous preprocessing and fairness validation. The system architecture incorporates collaborative oversight between clinical staff and AI developers to ensure transparency and maintain clinical override capabilities. This research aims to simultaneously improve medical response efficiency and reduce healthcare access disparities through thoughtful governance, comprehensive stakeholder engagement, and transparent accountability mechanisms, providing a scalable and ethically sound solution to a pressing public health challenge.

2 Problem Statement

Urban emergency departments face persistent challenges related to excessive patient volume, prolonged wait times, and inconsistent triage practices [3],[4]. Contemporary human-centered triage systems, including the widely adopted Emergency Severity Index (ESI), demonstrate significant limitations including inter-rater variability, susceptibility to implicit bias, and inadequate integration

*(Acting) Public-sector healthcare consortium position focusing on ethical AI implementation in clinical environments.

32nd Conference on Neural Information Processing Systems (NeurIPS 2018), Montréal, Canada.

of real-time data sources. This research addresses these challenges through the development and implementation of an AI-augmented triage support system designed to enhance throughput efficiency while mitigating disparities in clinical prioritization. The proposed solution must effectively balance the computational advantages of machine learning with the imperative to prevent algorithmic inequity in clinical decision-making.

3 Literature Review

3.1 Limitations of Traditional Triage Systems

The Emergency Severity Index (ESI), the most commonly used triage system in the United States, has been shown to suffer from subjective variation, especially in high-stress or high-volume environments [5]. Studies have documented inter-nurse reliability issues, leading to inconsistencies in patient prioritization [6]. Furthermore, bias in triage decisions based on race, gender, or language has been reported [7], reinforcing the need for assistive technology.

3.2 Promise of AI in Emergency Medicine

Machine learning and deep learning models have demonstrated capacity to predict clinical deterioration, readmission risk, and mortality [8],[9]. Specifically, neural network-based models have been used to estimate acuity in prehospital settings with higher accuracy than manual methods [10]. AI models can integrate multimodal inputs (vitals, symptoms, structured EHR) to improve real-time decision support in triage.

3.3 Challenges in Fairness and Bias Mitigation

Studies reveal that predictive models trained on historical medical data often replicate racial, gender, and socioeconomic biases present in the healthcare system [11]. A landmark study found that a widely deployed algorithm used fewer resources on Black patients despite similar health status. This necessitates deliberate implementation of fairness constraints, subgroup calibration, and counterfactual reasoning in model design.

3.4 Ethical and Regulatory Frameworks

The FDA's evolving Software as a Medical Device (SaMD) framework and the proposed EU AI Act signal increasing scrutiny of healthcare AI [12]. Governance frameworks, such as the WHO Guidance on Ethics and Governance of AI for Health (2021), emphasize transparency, accountability, and public engagement [13]. Scholars have advocated for human-in-the-loop systems that augment rather than replace clinical judgment [14].

3.5 Data Quality and Clinical Informatics

Data heterogeneity, missingness, and lack of standardization across hospital systems present practical hurdles [15]. Techniques such as semantic harmonization, dimensionality reduction, and data augmentation (e.g., using synthetic records) are necessary to maintain data fidelity and model generalizability across sites.

4 Methodology

4.1 Stakeholder Analysis

The implementation of an AI-augmented triage system involves a complex ecosystem of stakeholders with varying interests and concerns. Primary clinical stakeholders include emergency physicians and triage nurses who will directly interact with the system, as well as patients and patient advocacy groups who will be affected by algorithmic decisions. Administrative stakeholders encompass hospital administrators and medical boards responsible for institutional policy and risk management.

Technical implementation requires coordination with clinical informatics teams and data engineers who will develop and maintain the system infrastructure. Regulatory oversight involves Institutional

Review Boards (IRBs) and privacy officers ensuring compliance with ethical and legal standards. External governance includes public health oversight bodies such as HHS and FDA, as well as civil liberties organizations including the ACLU and EFF who monitor algorithmic fairness and transparency in public systems.

- Emergency physicians and triage nurses
- Patients and patient advocacy groups
- Hospital administrators and medical boards
- Clinical informatics teams and data engineers
- Institutional Review Boards (IRBs) and privacy officers
- Public health oversight bodies (e.g., HHS, FDA)
- Civil liberties organizations (e.g., ACLU, EFF)

4.2 Key Design Decisions

Critical implementation decisions center on several key technical and operational considerations. The selection of data modalities must balance clinical utility with privacy constraints, incorporating vital signs, patient speech patterns, and electronic health record history while maintaining data security. Model architecture choices involve weighing the trade-offs between black-box performance optimization and interpretable AI systems that provide clinical staff with transparent decision rationales.

Additional considerations include establishing appropriate thresholds for algorithmic confidence levels and defining clear protocols for clinical override authority. Pilot site selection requires careful attention to patient demographic balancing to ensure representative testing across diverse populations. Finally, the framework must incorporate robust post-deployment auditing mechanisms and dynamic retraining protocols to maintain system performance and adapt to evolving clinical practices.

- Selection of data modalities (vitals, patient speech, EHR history)
- Model architecture and explainability (black-box vs. interpretable AI)
- Thresholding for algorithmic confidence and override authority
- Pilot site selection and patient demographic balancing
- Post-deployment auditing and dynamic retraining protocols

5 Ethical Framework

The ethical complexity of implementing AI-augmented triage systems requires systematic analysis across multiple stakeholder perspectives. Table 1 presents a comprehensive ethical matrix that maps benefits, risks, ethical tensions, and proposed design responses for each key stakeholder group. This analysis reveals fundamental tensions between efficiency and equity, innovation and risk aversion, and transparency and proprietary technology interests. The matrix demonstrates that successful implementation requires nuanced design responses that address competing ethical imperatives while maintaining clinical efficacy and public trust.

6 Technical Implementation

6.1 Model Architecture

A hybrid deep learning architecture is recommended for triage prediction, combining structured EHR data with unstructured clinical notes or symptom descriptions. Figure 1 illustrates the proposed system workflow, showing how patient data flows through the AI-augmented triage process from initial input through clinical decision support.

The proposed architecture follows a dual-input design as depicted in Figure 1. The first input pathway processes numerical and categorical features including pulse, temperature, gender, and past medical

Table 1: Ethical Matrix for AI-Augmented Emergency Department Triage

Stakeholder	**Well-being**	**Autonomy**	**Fairness**	**Design Response**
Patients (High-Acuity)	Faster care; improved survival	Limited triage control during emergencies	Equal prioritization regardless of race/language	Human-in-the-loop; uncertainty thresholds; explainable AI
Patients (Low-Acuity)	Reduced wait frustration; better care flow	Freedom in choosing when and where to seek care	Avoid unjust deprioritization	Multi-metric scoring including distress; review pathways
Nurses & Clinicians	Reduced burnout; improved decision consistency	Discretion to override AI; clinical independence	Tool supports (not replaces) judgment	Override system; feedback loops; clear accountability
Hospital Administrators	Higher throughput; operational efficiency	Control over adoption, implementation pacing	Transparent ROI and model performance visibility	Phased rollout; embedded audit trail; governance reviews
AI Developers	Model validation; real-world deployment value	Freedom to implement ethical constraints	Fair credit for responsible innovation	Fairness-by-design; adversarial de-biasing; documentation
Public Health Authorities	System-wide care improvements; pandemic readiness	Regulatory authority and standard-setting	Equitable distribution across systems and populations	Open benchmarking; inclusive performance metrics
Civil Liberties Organizations	Strengthened patient rights; civic accountability	Voice in oversight, policy, and public participation	Equity in governance and transparency of tools	Publishable audits; advisory panels; consent standards
Marginalized Populations	Potential for reduction in triage bias	Right to data privacy and informed participation	Correction of structural disparities in care access	Subgroup calibration; co-design; opt-out options

history. The second input pathway handles free-text chief complaints and symptom descriptions through natural language processing techniques. These parallel streams converge through a dual-input model backbone with shared fusion layers, ultimately producing multi-class classification outputs corresponding to triage levels 1 through 5.

The model employs weighted cross-entropy loss functions to address inherent class imbalance in emergency department triage data, where critical cases (levels 1-2) are naturally less frequent than lower-acuity presentations [16]. Fairness constraints are incorporated through adversarial de-biasing modules and post-hoc equalized odds constraints, ensuring that the system maintains equitable performance across demographic subgroups while preserving clinical accuracy.

6.2 Data Infrastructure

Aggregated, de-identified emergency department visit data from a metropolitan hospital system will be sourced under IRB-approved data-sharing agreements. The dataset will include vital signs, symptom keywords, demographic metadata, and triage outcomes as outlined in Table 2. This multi-institutional approach ensures sufficient sample size while maintaining geographic and demographic diversity necessary for robust model development.

The data curation process will involve dimensionality reduction, semantic normalization of symptom descriptors, and removal of mislabeled or conflicting triage levels. Balanced sampling across age, gender, and ethnicity groups will be enforced to prevent overfitting to dominant subpopulations. This preprocessing pipeline addresses the inherent challenges of electronic health record data quality while preserving the clinical validity of the underlying information.

Privacy and security protocols will adhere to HIPAA and GDPR standards where applicable [17], [18]. All data transfers will utilize end-to-end encryption, and differential privacy mechanisms will be evaluated to assess utility-privacy tradeoffs during model training. These measures ensure compliance with regulatory requirements while maintaining the data integrity necessary for effective algorithmic development.

Table 2: Patient ED record data schema for AI-augmented triage system

Field Name	Type	Description
patient_id	UUID	Unique identifier for patient
visit_timestamp	Timestamp	Arrival time
age	Integer	Patient age
gender	Categorical	Male / Female / Other
race_ethnicity	Categorical	Race/ethnicity classification
primary_language	Text	Patient-reported language
vitals_heart_rate	Float	Beats per minute
vitals_temperature	Float	Body temperature
vitals_blood_pressure	Text	Systolic/diastolic
triage_nurse_notes	Text	Chief complaint and presenting symptoms
historical_triage_score	Integer	ESI or prior triage score
final_triage_score	Integer	Target label (physician review)
diagnostic_outcome	Text	Final diagnosis (for validation)
resource_use_level	Integer	Number of clinical resources used

Several implementation challenges are anticipated, including institutional resistance to algorithmic decision support in clinical environments, ethical friction surrounding proprietary artificial intelligence in public institutions, algorithmic opacity concerns from frontline staff, and longitudinal drift in patient presentation patterns that may affect model retraining requirements. These hurdles will be addressed through comprehensive stakeholder engagement and transparent communication strategies.

6.3 Data Schema

Table 2 presents the proposed schema for the triage model's input features and metadata pipeline, designed to capture comprehensive patient information while maintaining privacy and enabling robust model training.

The data infrastructure incorporates comprehensive privacy protection measures including record pseudonymization with encryption protocols. Differential privacy techniques will be applied to any external data releases to ensure individual patient confidentiality while maintaining statistical utility. Additionally, stratified sampling methodologies will be employed to guarantee adequate representation across race, ethnicity, and gender categories for robust fairness testing and model validation.

7 Governance, Accountability, and Fairness

The proposed system will operate under comprehensive regulatory oversight, adhering to FDA Software as a Medical Device (SaMD) pre-certification requirements and submitting to recurring audits under the Clinical Decision Support Software (CDS) guidelines [19]. This regulatory framework ensures that the system meets established safety and efficacy standards while maintaining accountability to healthcare stakeholders and patients.

Model performance evaluation will employ disaggregated metrics including accuracy, sensitivity, and false-negative rates across key demographic strata. Advanced fairness evaluation tools such as IBM's AI Fairness 360, Google's What-If Tool, and SHAP will be deployed for comprehensive model interrogation and local explanation capabilities [20], [21]. This multi-tool approach enables thorough assessment of algorithmic bias and provides interpretable insights into model decision-making processes.

Transparency and accountability mechanisms will ensure that each triage recommendation is logged with confidence intervals and made reviewable through an intuitive visual dashboard. Clinical staff will have the capability to annotate outcomes and flag anomalies in real-time, which will trigger automated model drift alerts. This human-in-the-loop approach maintains clinical oversight while enabling continuous system improvement through feedback integration.

The scalability framework incorporates containerized architecture for modular deployment across diverse healthcare settings. Initial pilot implementations will generate lessons learned that will be abstracted into open-source templates, facilitating adaptation for rural and global health contexts. This approach ensures that the benefits of the system can be democratized while respecting local clinical practices and resource constraints.

- Regulation: FDA's SaMD (Software as a Medical Device) guidelines; internal review board oversight.
- Fairness Testing Tools: IBM AI Fairness 360 or Google's What-If Tool to monitor for racial/ethnic disparities.
- Transparency: Patients informed about AI involvement; model decisions explained through interpretable AI methods (like SHAP).
- Scalability: Feasible across large hospital networks with minimal infrastructure change.

8 Discussion

This project represents a comprehensive approach to implementing ethical AI in healthcare settings. By addressing stakeholder concerns through structured ethical analysis (Table 1), establishing robust governance frameworks, and maintaining transparency throughout the development process (Figure 1), we aim to create a triage support system that enhances both efficiency and equity in emergency medical care. The data schema presented in Table 2 provides a foundation for responsible data collection and model training.

The implementation challenges span technical, ethical, and organizational domains. Technical challenges include ensuring model robustness across diverse patient populations and maintaining performance during distribution shifts in patient presentations. Ethical challenges center on balancing efficiency gains with fairness requirements, particularly for vulnerable populations who may be disproportionately affected by algorithmic bias. Organizational challenges involve securing institutional buy-in from clinical staff while addressing concerns about automation and professional autonomy.

Future research directions should focus on longitudinal evaluation of system performance, exploration of federated learning approaches to enable multi-institutional collaboration without compromising privacy, and development of more sophisticated bias detection and mitigation techniques. The success of this initiative will depend on continuous stakeholder engagement, rigorous bias testing, and adaptive governance structures that can evolve with technological and social changes.

9 Conclusion

This research presents a comprehensive framework for implementing ethical AI in emergency medical settings through systematic integration of stakeholder perspectives, robust governance mechanisms, and transparent development processes. The structured ethical analysis (Table 1), combined with the proposed technical architecture (Figure 1) and data infrastructure (Table 2), provides a foundation for developing triage support systems that simultaneously enhance operational efficiency and promote equitable care delivery.

The implementation success of such systems depends critically on sustained stakeholder engagement, rigorous bias detection and mitigation protocols, and adaptive governance structures capable of evolving with technological advancements and changing social expectations. Future research directions should prioritize longitudinal performance evaluation, federated learning approaches that enable multi-institutional collaboration while preserving privacy, and development of advanced bias detection methodologies. This work contributes to the growing body of literature on responsible AI deployment in healthcare while providing practical guidance for institutions seeking to implement similar systems.

References

[1] Bernstein, S.L. et al. (2009) The effect of emergency department crowding on clinically oriented outcomes. *Academic Emergency Medicine*

[2] Guttmann, A. et al. (2011) Association between waiting times and short term mortality and hospital admission after departure from emergency department: population based cohort study from Ontario, Canada.

[3] Moskop, J.C. et al. (2009) Emergency department crowding, part 1—concept, causes, and moral consequences. *Annals of Emergency Medicine*

[4] Derlet, R.W. & Richards, J.R. (2000) Overcrowding in the nation's emergency departments: complex causes and disturbing effects. *Annals of Emergency Medicine*

[5] Travers, D.A. et al. (2009) Reliability and Validity of the Emergency Severity Index. *Academic Emergency Medicine*

[6] Van der Wulp, I. et al. (2008) Reliability and validity of the Manchester Triage System. *International Emergency Nursing*

[7] Green, A.R. et al. (2007) Implicit bias among physicians and its prediction of patient-centered communication. *Patient Education and Counseling*

[8] Rajkomar, A. et al. (2018) Scalable and accurate deep learning with electronic health records. *npj Digital Medicine*

[9] Esteva, A. et al. (2019) A guide to deep learning in healthcare. *Nature Medicine*

[10] Q. Chen et al., "Enhancing performance of the National Field Triage Guidelines using machine learning: Development of a prehospital triage model to predict severe trauma," *Journal of Medical Internet Research*, vol. 26, Sep. 2024. doi:10.2196/58740

[11] Obermeyer, Z., Powers, B., Vogeli, C., & Mullainathan, S. (2019) Dissecting racial bias in an algorithm used to manage the health of populations. *Science*

[12] Topol, E. (2019) *Deep Medicine: How Artificial Intelligence Can Make Healthcare Human Again*. Basic Books.

[13] WHO (2021) *Ethics and Governance of Artificial Intelligence for Health*. Geneva: World Health Organization.

[14] Morley, J. et al. (2020) Ethics of AI in Health Care: A Mapping Review. *Journal of Medical Internet Research*

[15] Johnson, A. et al. (2016) MIMIC-III, a freely accessible critical care database. *Sci Data 3*, 160035. https://doi.org/10.1038/sdata.2016.35

[16] Kirubarajan, A. et al. (2020) Artificial intelligence in emergency medicine: A scoping review. *Journal of Emergency Medicine*

[17] U.S. Department of Health and Human Services (2013) *Summary of the HIPAA Privacy Rule*. Washington, DC: HHS Office for Civil Rights.

[18] European Union (2016) *General Data Protection Regulation (GDPR)*. Official Journal of the European Union L119/1.

[19] U.S. Food and Drug Administration (2017) *Software as a Medical Device (SaMD): Clinical Evaluation*. FDA Guidance Document.

[20] R. K. E. Bellamy et al. (2019) AI Fairness 360: An extensible toolkit for detecting and mitigating algorithmic bias, *IBM Journal of Research and Development*, doi: 10.1147/JRD.2019.2942287

[21] Lundberg, S.M. & Lee, S.I. (2017) A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems*

[22] Chen, J.H. & Asch, S.M. (2017) Machine learning and prediction in medicine—beyond the peak of inflated expectations. *New England Journal of Medicine*

[23] Rajkomar, A., Hardt, M., Howell, M.D., Corrado, G., & Chin, M.H. (2018) Ensuring fairness in machine learning to advance health equity. *Annals of Internal Medicine*

10 Appendix

10.1 AI-Augmented Triage System Workflow

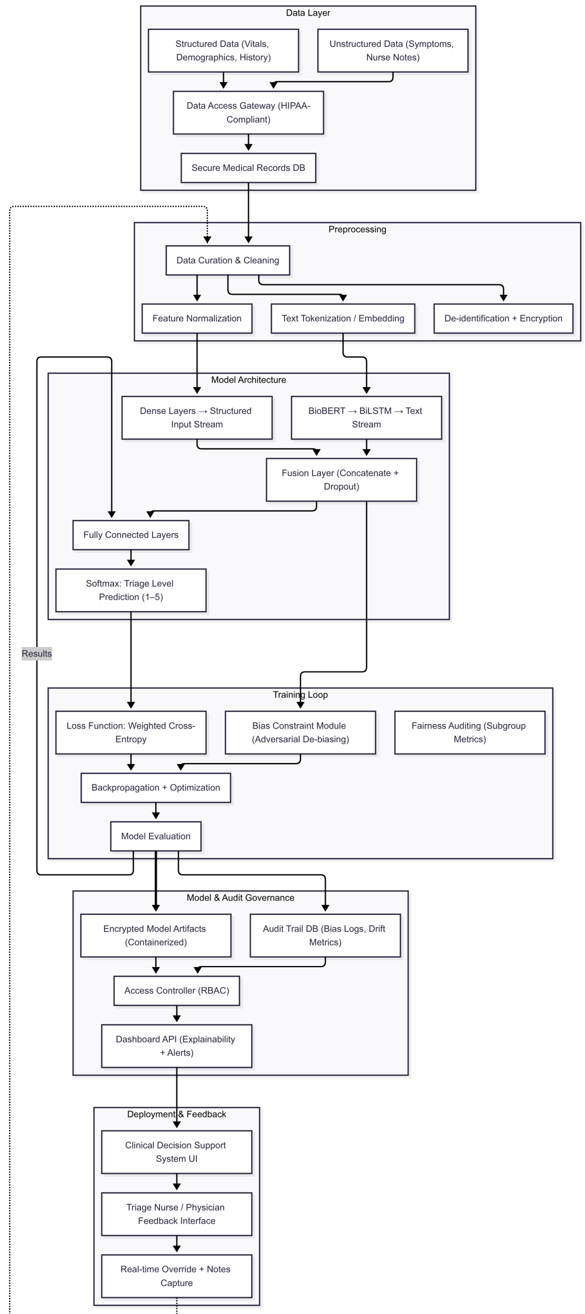

Figure 1: Proposed AI-augmented triage system workflow showing the integration of patient data inputs, machine learning processing, and clinical decision support outputs with human oversight mechanisms.