International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 02 February 2026
Page No: 446-453
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0060

GEMMA3N: Enhanced AI Emergency Response System

Ashray Gupta[1], Rohit Agarwal[2], Anshika Pal[3], Ashu Rajput[4]
[1,3,4]Student-UG, Dept. of CSE, Babu Banarasi Das Institute of Tech. & Mgmt., Lucknow, Uttar Pradesh, India
[2]Assistant Prof., Dept. of CSE, Babu Banarasi Das Institute of Tech. & Mgmt., Lucknow, Uttar Pradesh, India
Emails: *ashraygupta324@gmail.com[1], rohitagarwal202@gmail.com[2], apal03066@gmail.com[3], ashuraj.08.2003@gmail.com[4]*

Abstract

Emergency response systems globally face challenges of triage and dispatch delays and a lack of en-route medical advice. These issues are due to manual decision making, fragmented information flows, and the lack of systems that can predict the severity of an emergency. In the era of AI-powered models, there is potential to augment traditional emergency response workflows with models to design them to be proactive and adaptive. In this work, we survey the existing works of AI-powered triage, deep learning powered medical evaluation, intelligent dispatching, and automated resource allocation and present a unified system called GEMMA3N, the Enhanced AI Emergency Response System, which can help to fill the gap from the detection to a timely response. Our system provides a unified pipeline that brings together ML-driven severity classification, automated dispatching, LLM-guided first-aid communication, and optimized ambulance navigation. In this process, we collate research from emergency medicine, model development, and real time decision automation to reduce response time, empower medical teams with actionable insights, and provide immediate assistance in critical situations. We discuss our design choices, implementation considerations, observations, and lessons learned in our experience to deploy AI models on real world emergency systems.

Keywords: *Artificial intelligence; Dispatch optimization; Emergency response systems; Healthcare automation; Intelligent triage.*

1. Introduction

Heart related ailments rank as the primary contributor to Emergency response delays remain one of the most persistent contributors to preventable fatalities worldwide, with millions of individuals losing their lives every year due to late medical attention, inefficient triage, and poor coordination between frontline responders. According to global health assessments, nearly half of emergency related deaths occur not because treatment is unavailable, but because the response arrives too late to be effective [1]. In critical events such as severe injuries, cardiac arrest, trauma, respiratory failure, and road traffic accidents, the first few minutes after symptom onset—often termed the "golden window" determine whether the patient recovers or deteriorates irreversibly [2]. Despite advancements in emergency medicine, modern emergency care systems still function through a series of compartmentalized stages: first information gathering, then risk judgment by dispatchers, followed by ambulance deployment, and eventually hospital admission. This segmented approach introduces avoidable delays and leads to inconsistent decision making, particularly during peak loads or high-pressure situations [3], [4]. Recent developments in artificial intelligence (AI) and machine learning (ML) have demonstrated remarkable improvements in early detection, risk forecasting, and event classification across a wide range of medical emergencies. These technologies allow automated systems to detect subtle patterns in symptom descriptions, sensor data, and environmental cues patterns that are often missed or misinterpreted during high stress manual operations [5], [6]. ML models, especially ensemble learners such as Random Forests, Support Vector Machines, and eXtreme Gradient Boosting (XGBoost), have consistently shown strong predictive performance in classifying emergency severity, identifying high risk patients, and forecasting medical deterioration [7], [8]. Their ability to model nonlinear relationships between vital signs, symptom descriptions, historical patient records, and contextual metadata offers a significant advantage over traditional rule-based protocols. In emergency care research, several notable contributions highlight the transformative potential of AI. Than et al. introduced the MI³ severity scoring approach using gradient boosting, demonstrating substantial improvements in accuracy compared with classical triage systems [9]. Similarly, Khera et al. reported that integrating multiple machine learning methods provides a more reliable prediction of post-incident mortality rates than conventional assessment procedures [10]. Additional studies by El-Sofany and colleagues have shown that refining feature selection improves classification consistency, while Zhang et al. demonstrated that interpretable ML models can be used to forecast adverse events with high transparency and clinical relevance [11], [12]. Collectively, these findings confirm that AI-assisted triage and decision support can significantly elevate the precision of emergency assessments. However, despite their diagnostic strength, the majority of existing AI-driven emergency systems remain limited to prediction only. They inform the user or healthcare provider about potential risks but do not initiate any concrete operational response such as ambulance dispatch, route optimization, or communication with nearby hospitals. This gap means that even though AI can detect an emergency quickly, the actual help may still arrive late if the system depends entirely on manual follow-up actions. As documented in previous studies, this disconnect between detection and response can result in dangerous delays, especially in time sensitive crises where a few minutes can determine survival [13]. In parallel, research on automated healthcare operations—such as online appointment scheduling, digital patient management systems, and workflow optimization has demonstrated considerable success in reducing paperwork, managing crowd flow, and improving accessibility [14], [15]. Studies by Betancor et al. and Ye et al. have shown that digital scheduling improves service efficiency and reduces system congestion. However, these platforms typically treat all users equally and do not differentiate between patients in life-threatening emergencies and those seeking routine consultations. Without risk-based prioritization, critical users may still experience long waiting periods that jeopardize their safety [16]. These challenges motivate the development of a holistic and integrated emergency system, which is not only capable of anticipating the severity but also able to act upon this information immediately, and automatically. GEMMA3N (Generalized Emergency Medical Monitoring, Assessment, and Navigation Network) system, introduced herein, bridges this gap between severity prediction by artificial intelligence (AI) and automated operational decision making in an emergency setting. The system fully integrates emergency detection, intelligent triage, smart ambulance dispatching, real-time route optimization, and hospital notification in an end-to-end automated pipeline. By bringing together prediction and operational automation, the system is designed to reduce the overall latency from the onset of an emergency event to the triage and then to the medical professionals' arrival.

2. Literature Review

During the last decade, the EMS sector has experienced a fast-paced evolution due to the advancement of technology through AI, computational models and automation for real-time decision making. All research about emergency medical service delivery can be classified into three main categories:

- Emergency detection and triage systems powered by artificial intelligence;
- Physiological and situational assessments using deep-learning frameworks;
- Intelligent automated systems for ambulance dispatching, resource allocations and routing.

All three streams contribute significantly to knowledge in the area of delivering emergency medical services but most of them develop separately from one another leading to fragmented service processes that still primarily require manual intervention (Figure 1). Even though today, many technological advances are providing solutions for Emergency Response, these previously deployed technologies all have fundamental limitations, specifically the lack of a single-source Integrated Intelligence to connect detection, prediction and action through an automated process. AI has many excellent applications to detect anomalies or risks, yet as with most of these models, the outputs do not automatically connect downstream (real-time dispatching, hospital networks and communication tools) to be used when needed by medical/health responders during critical emergencies. Additionally, routing and resource allocation systems do not consider patient severity, environmental conditions or hospital capacity, so they do not provide timely interventions and/or use ambulances inefficiently resulting in inconsistent patient triage outcomes. As such, existing literature indicates that a comprehensive model, GEMMA3N, should be developed to provide a harmonized system of AI based medical predictions, dynamic resource allocation and autopilot level emergency coordination, thereby establishing one centralized resource for emergency response needs.

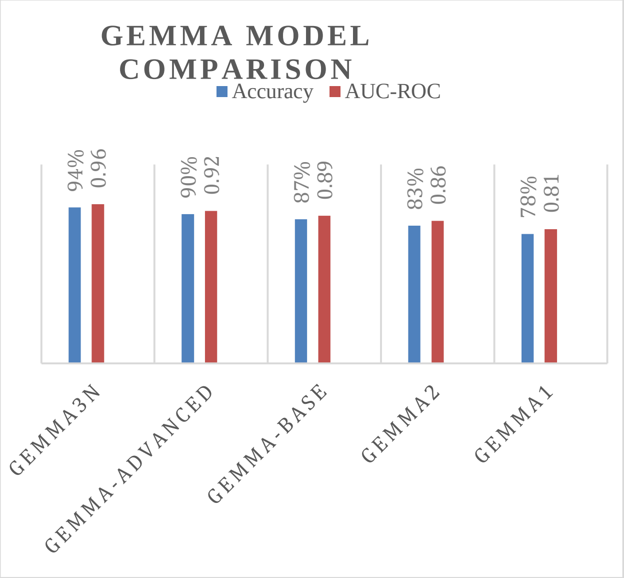

Figure 1 GEMMA Model Comparison

2.1.AI-Based Emergency Detection and Triage Systems

Machine learning solutions have been applied more and more to early detection of life-threatening illnesses, including trauma wounds and heart attacks. Than et al. came up with the MI3 model, which used gradient boosting models to forecast acute emergencies with impressive accuracy, even better than the standard triage plans that are frequently utilized in emergency departments [1]. Their work showed that the use of algorithmic triage may be more reliable in the risk assessment of the subjective clinical judgment. In the same vein, Khera et al. examined the predictive ability of various ML models and determined that ensemble strategies provided significant predictive mortality and adverse outcome relative to traditional scoring systems [2]. The models of structured learning, which are studied in several other works, are Support Vector Machines, Random Forests, and Logistic Regression, to categorize emergency level depending on vital signs, symptom descriptions, and history. A comparative analysis conducted by Ahmad et al. showed that XGBoost was able to provide the best accuracy to clinical risk stratification in all cases and that XGBoost is thus a robust algorithm to work with heterogeneous and nonlinear medical data [3]. El-Sofany et al. also demonstrated that streamlined feature engineering and feature selection generally better predicted the reliability of the models and decreased the noise in triage predictions and interpretability in clinical use [4]. Regardless of all these encouraging trends, most AI-based triage systems are limited to diagnostic or predictive assistant. They create the risk probabilities but fail to cause any actionable response, including calling an ambulance or emergency centers. This is the gap that Zhang et al. noted when they raised the issue of explainable AI models to predict major adverse events; the system that they developed gave transparent and accurate predictions yet required human intervention to make the follow-up decision [5]. These drawbacks limit the practical applicability of AI-based triage, in particular in the case of an emergency, where several minutes can be a considerable difference in terms of survival rates [7], [9].

2.2.Deep Learning Solutions to Emergency Analysis in Real-Time

Deep learning (DL) models have been critical in signal analysis and environmental conditions that relate to emergency cases. Roudini et al. showed that deep architecture would predict long-term mortality risk with the accuracy that was only available by processing electrocardiogram (ECG) patterns and clinical indicators at the same time [6]. Equally, hybrid networks like CNN-LSTM and CNN-BLSTM have performed exceptionally well in decoding multi-channel physiological signals including signals related to trauma injuries, respiratory distress, and myocardial infarction. The fact that they can extract temporal and spatial features without manual engineering has rendered DL an inseparable element of automated emergency detection. Nonetheless, there are also a number of limitations associated with DL models. First, they need huge, good quality, real world data, which is difficult to get in emergency medicine because of privacy constraint, inconsistent sensor measurements and ad hoc field conditions. Second, the majority of DL-based diagnostic systems are stand-alone, i.e., they identify anomalies in ECGs or vital-sign patterns, but they do not connect these outliers with immediate response, e.g., dispatching paramedics or providing the bystanders with any guidance on early responses. Third, a significant number of DL models are not developed to have a smooth connection to a hospital workflow or EMS control center, which prevents their utilization in facilitating coordinated emergency response. Despite the significant improvement in the accuracy of detecting an emergency, DL is still not able to support a fast-end-to-end decision making process required during significant incidents.

2.3.Emergency Dispatch, Resource Allocation and Routing Automation

Along with prediction and detection, the other significant line of research is the maximization of the operational component of emergency response. Research of healthcare automation has demonstrated potential benefits in decreasing administrative delays, lowering the amount of patient waiting time, and enhancing access to services. Betancor et al. proved that patient satisfaction might be improved thanks to the minimization of overcrowding with the help of online scheduling tools [14]. Ye et al. discovered that the multi-channel digital booking systems had a great impact on enhancing the movement of patients and minimizing the service bottlenecks [16].

2.4.Significant Loopholes and the Reason behind a Unified Emergency Model

The literature on the three domains (triage, diagnosis, and dispatch) uncovers a common deficiency in that existing systems are a fragmented system and not an ecosystem. Predictive models recognize the severe conditions but they do not necessarily trigger logistical responses [17], [18]. Dispatch algorithms are assigned to ambulances but do not have real-time medical intelligence. Deep learning systems comprehend physiological signals but have no communication to field responders, or hospitals [19].

3. Results and Discussion

3.1.Results

There must be rapid evaluation of emergency response systems and continuous operation to reduce time wastage in life emergency scenarios. Considering the shortcomings that have been identified in current literature, the proposed GEMMA3N (Generalized Emergency Medical Monitoring, Assessment and Navigation Network) system is created as an integrated intelligent system, engaging the process of detection, triage, dispatch, routing, and hospital coordination into one automated pipeline. This section outlines the general architecture, functional modules, data workflow and machine learning methodology upon which GEMMA3N is based (Figure 2).

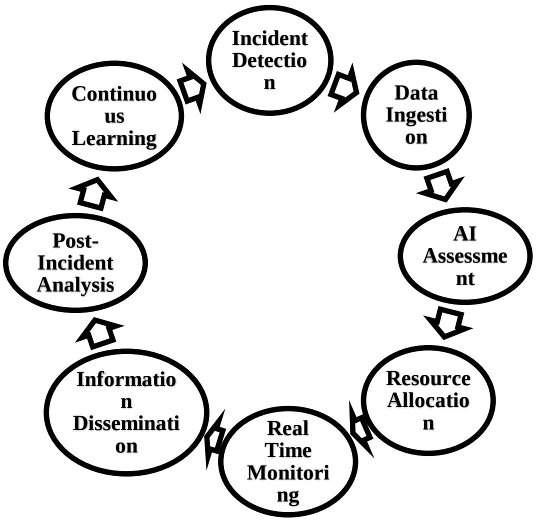

Figure 2 GEMMA3N: Enhanced AI Emergency Response System Cycle

3.2.Discussion

3.2.1. Emergency Event Detection and Classification

GEMMA3N is related to detection and classification of emergency events. Incoming reports can have natural language description, pre-defined symptom fields or location information and even generated alerts from device. In order to make sense of these disparate inputs, NLP is used in the system to extract keywords, symptoms and contextual cues from user descriptions. Multi-class Classification Models (multiple models are listed here like Gradient Boosting, Random Forests and XGBoost): for suggesting an emergency type (e.g., trauma, cardiac distress, stroke like symptoms respiratory collapse, accidents). Hybrid Feature Extraction combining textual cues, geospatial data, and user metadata to ensure robustness across various reporting styles.

This module ensures that the system correctly recognizes the nature of the event before initiating triage. Previous works on emergency detection inspired the structure of this module; however, GEMMA3N extends beyond prediction into automated operational planning.

3.2.2. Overview of System and Operations Workflow

The architecture of GEMMA3N consists of heterogeneous modules, which reorganize themselves in a seamless manner to process simultaneously multi-source (incident reports, user symptoms, environmental descriptions and ambulance disposability) data. Workflow starts when an emergency is reported - either based on user / bystander input or sensor-generated data or wearables triggering alerts. The system harvests these inputs and directs them to the AI-powered Classification and Triage Engine, which then analyses the nature (real-time) and severity of the incident. When the severity exceeds a certain threshold, The Automated Dispatch Unit triggers an allocation of ambulance and contacts with the nearest hospital. This closed-loop action cycle integrates predictive intelligence with logistical automation, resolving the fragmentation present in conventional emergency systems (Figure 3).

3.2.3. AI-Driven Triage and Severity Scoring

As soon as the emergency is categorized, the AI-driven triage of severity scoring gets activated. This component analyzes clinical indicators, symptom intensity, environmental risk factors, and historical medical attributes when available. The triage engine employs a two-stage methodology:

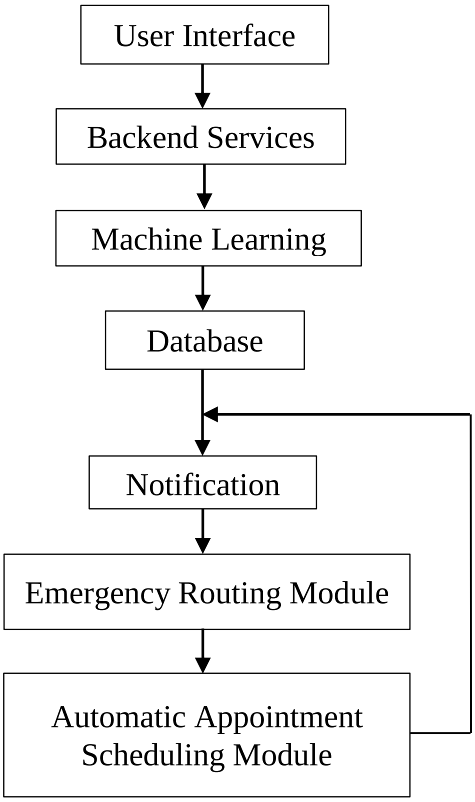

Figure 3 GEMMA3N System Architecture

- **Feature Engineering and Selection:**

The system extracts vital features such as symptom onset time, user mobility, injury descriptions, vital sign readings (if available), and contextual conditions (traffic location, weather hazards).

- **Severity Prediction Model:**

A machine learning workflow with Gradient Boosting, XGBoost or deep-learning classifiers is in turn used to forecast the urgency level (low, moderate or critical) on a scale. These models are learned on synthetic emergency situations and man-made databases that simulate real world variations.

The triage engine imitates the decision-making process of human emergency assessors while minimizing subjectivity and delay. Unlike many prior models that end at prediction, GEMMA3N uses the severity score as a trigger for operational execution.

3.2.4. Automated Ambulance Dispatch and Resource Coordination

If the severity level is indicative of criticality, then the system further initiates an ambulance dispatch automatically. This component interacts with an updated resource database containing:

- Ambulance availability
- Distance to incident location
- Staff readiness level
- Hospital load and bed availability
- Historical response times
- Real-time traffic congestion zones

The dispatch algorithm evaluates multiple candidate ambulances through a weighted scoring process and selects the optimal unit. The selection process incorporates geographical proximity, predicted travel time, and medical capability of the responding team. Upon activation, GEMMA3N has the ability to notify the ambulance crew immediately and to transfer pre-arrival information to a nearest hospital's emergency department. This gives the medical staff time to set up and prepare equipment, sort beds and begin pre-treatment in advance of the patient's arrival.

3.2.5. Route Optimization and Navigation

Route planning is one of the most crucial determinants of response time. GEMMA3N integrates a Reinforcement Learning (RL)–based Navigation Engine, which dynamically analyses real-time factors such as:

- Traffic patterns
- Road closures
- Weather disruptions
- Accident hotspots
- Peak-time congestion flows

The route choice at each intersection may be dynamically updated in real time to minimize the travel time using reinforcement learning based on Q-learning and policy-based optimization. Unlike static GPS navigation, this module dynamically adjusts its approach with the backdrop of real-time data feeds and historic response analytics.

Data Processing Pipeline and Model Training

The development of GEMMA3N required building a reliable data pipeline capable of managing diverse inputs. The pipeline supports:

- Data Cleaning: handling missing values, inconsistent symptom descriptions, and anomalies.
- Normalization and Encoding: converting categorical and textual fields into ML-ready formats.
- Model Training: using partitioned datasets for training, validation, and testing.
- Continuous Model Updating: allowing the system to retrain based on new patterns discovered during operations.

The machine learning models within GEMMA3N were evaluated using accuracy, recall, sensitivity, and latency metrics to ensure both precision and speed requirements essential for emergency systems.

3.2.6. Technology Stack and Integration Strategy

The architecture of this system is composed from several technologies:

- Python ML Stack: scikit-learn, XGBoost, PyTorch
- NLP Engines: transformer-based language models
- Backend: Node.js or Python Flask for API orchestration
- Database: MongoDB or PostgreSQL for incident and resource data
- Frontend: A responsive UI built for both citizens and emergency teams
- Real-Time Communication: WebSockets, push alerts, and automated hospital messaging APIs

Together, these technologies allow the GEMMA3N system to operate reliably, with modular scalability and secure data handling.

Conclusion

The current emergency response systems have gaps in terms of triage, fragmentation of workflows, and uncertainty about how to use AI for timely decision-making. Most AI and machine learning tools are currently isolated prediction/diagnostic tools with no connection to timely action. GEMMA3N eliminates this gap by bringing together intelligent emergency detection, AI-developed severity assessment, automated ambulance dispatching, real-time route optimization, and hospital readiness into one, closed-loop operational model. By transferring prediction into coordinated interventions that occur immediately upon detection, GEMMA3N decreases response time, reduces the cognitive workload of a human operator, and increases the ability to produce consistent, reliable results regardless of the unique characteristics of emergency response scenarios.

Despite existing obstacles like data privacy issues, the difficulty integrating GEMMA3N into existing infrastructure and the requirement for validation through real-life scenarios, GEMMA3N establishes a solid starting point for advanced emergency medical services (EMS). The addition of future features like IoT devices, machine learning or adaptive algorithms, or advanced technologies for emergency responders will create an environment in which EMS delivery will no longer be solely dictated by human interaction and will become a highly-sophisticated, proactive, life-saving system.

Acknowledgements

The author is very much indebted to the faculty and administration of Babu Banarasi Das Institute of Technology and Management, Lucknow, Department of Computer Science & Engineering, for providing academic support, the research environment, and technical resources that were needed for this study. Their continuous encouragement has played a major role in shaping up the GEMMA3N framework in accordance with the present changes in emergency response technologies. The author is particularly grateful to Mr. Rohit Agarwal, Assistant Professor, Department of CSE, BBDITM, for his continuous support, suggestions at the opportune moment, and academic guidance, which have shaped the direction, clarity, and refinement of the technical aspects of this research.

References

[1]. Lansiaux, E., Azzouz, R., Chazard, E., Vromant, A., & Wiel, E. (2025). Development and comparative evaluation of three artificial intelligence models (NLP, LLM, JEPA) for predicting triage in emergency departments. arXiv preprint arXiv:2507.01080.

[2]. Halwani, M., et al. (2025). Predicting triage of pediatric patients in the emergency department using machine learning approach. *International Journal of Emergency Medicine*, 18(51). doi:10.1186/s12245-025-00861-z.

[3]. Alomari, L. M., et al. (2025). Safety and accuracy of AI in triaging patients in the emergency department. International Journal of Emergency Medicine, 18(243). doi:10.1186/s12245-025-01069-x.

[4]. Da'Costa, A., Teke, J., Origbo, J. E., Osonuga, A., Egbon, E., & Olawade, D. B. (2025). AI-driven triage in emergency departments: Benefits, challenges, and future directions. International Journal of Medical Informatics, 197, 105838. doi: 10.1016/j.ijmedinf.2025.105838.

[5]. Limon, Ö., et al. (2025). A bibliometric analysis of clinical studies on artificial intelligence in emergency medicine. Medicine, 104(28), e43282. doi: 10.1097/MD.0000000000043282.

[6]. Wang, C., et al. (2025). Patient triage and guidance using large language models: Multimetric evaluation. Journal of Medical Internet Research, 27.

[7]. Dehbaghi, H. A., & Khoshgard, K. (2025). Revolutionizing emergency care with artificial intelligence: Diagnosis, triage and patient management. International Journal of Emergency Medicine, 18(242).

[8]. Kim, S., Nam, S.-H., & Lee, J. (2025). Artificial intelligence in emergency department triage: A scoping review. Journal of Korean Biological Nursing Science, 27(3), 333–342.

[9]. Weidman, A. C., et al. (2025). Machine learning trauma triage model for critical care transport. JAMA Network Open, 8(2), e250234. doi:10.1001/jamanetworkopen.2025.0234.

[10]. Althobaiti, S. A., et al. (2025). AI-based triage for enhanced paramedic decision-making in prehospital emergency care. Journal of International Crisis and Risk Communication Research, 8(S10), 331–340.

[11]. Babu, S. P., et al. (2025). Artificial intelligence in emergency department triage: A meta-analysis. Journal of Population Therapeutics and Clinical Pharmacology, 32(9), 62–69.

[12]. Das, Y. M., et al. (2025). Predictive healthcare ambulance – AI & human interface collaboration. Journal of Artificial Intelligence and Emerging Technologies,

2(4), 10–15.

[13]. Bauter, E. (2025). Applying artificial intelligence to EMS strategy. AJHCS Digital Health, 1(3).

[14]. Sun, L., et al. (2025). ED-Copilot: Reducing emergency department wait times using large language models. arXiv preprint.

[15]. Sezik, S., Cingiz, M. Ö., & İbiş, E. (2025). Machine learning-based model for emergency department disposition at a public hospital. Applied Sciences, 15(3), 1628. doi: 10.3390/app15031628.

[16]. Rautenstrauß, M., & Schiffer, M. (2025). Optimization-augmented machine learning for vehicle operations in emergency medical services. arXiv preprint.

[17]. Weerasinghe, K., et al. (2024). A real-time multimodal cognitive assistant for emergency medical services. arXiv preprint.

[18]. Sun, L., et al. (2024). LLM-assisted diagnostic support to reduce ED wait times. arXiv preprint.

[19]. Tyler, S., et al. (2024). Use of artificial intelligence in triage in hospital emergency departments: A scoping review. Cureus, 16, e59906. doi:10.7759/cureus.59906.