

Development of a Large Language Model-based Multi-Agent Clinical Decision Support System for Korean Triage and Acuity Scale (KTAS)-Based Triage and Treatment Planning in Emergency Departments

Seungjun Han

*John Bapst Memorial High School
Bangor, USA*

shan25@johnbapst.org

Wongyung Choi

*Minerva University
San Francisco, USA*

woncho@uni.minerva.edu

Corresponding Author: Wongyung Choi

Copyright © 2025 Seungjun Han and Wongyung Choi. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract

Emergency department (ED) overcrowding and the complexity of rapid decision-making in critical care settings pose significant challenges to healthcare systems worldwide. While clinical decision support systems (CDSS) have shown promise, the integration of large language models (LLMs) offers new possibilities for enhancing triage accuracy and clinical decision-making. This study presents an LLM-driven CDSS designed to assist ED physicians and nurses in patient triage, treatment planning, and overall emergency care management.

We developed a multi-agent CDSS utilizing Llama-3-70b as the base LLM, orchestrated by CrewAI and Langchain. The system comprises four AI agents emulating key ED roles: Triage Nurse, Emergency Physician, Pharmacist, and ED Coordinator. It incorporates the Korean Triage and Acuity Scale (KTAS) for triage assessment and integrates with the RxNorm API for medication management.

The model was evaluated using the Asclepius dataset, with performance assessed by a clinical emergency medicine specialist. The CDSS demonstrated high accuracy in triage decision-making compared to the baseline of a single-agent system. Furthermore, the system exhibited strong performance in critical areas, including primary diagnosis, critical findings identification, disposition decision-making, treatment planning, and resource allocation.

Our multi-agent CDSS demonstrates significant potential for supporting comprehensive emergency care management. By leveraging state-of-the-art AI technologies, this system offers a scalable and adaptable tool that could enhance emergency medical care delivery, potentially alleviating ED overcrowding and improving patient outcomes. This work contributes to the growing field of AI applications in emergency medicine and offers a promising direction for future research and clinical implementation.

Keywords: Clinical decision support system, Emergency department overcrowding, Large Language Model, Multi-Agent system.

1. INTRODUCTION

Emergency Departments (EDs) are critical access points for urgent medical care but face unprecedented challenges that hinder timely and effective care. For instance, ED visits have risen significantly in countries like the United States and South Korea, where millions of patients rely on emergency care annually [1–3].

The consequences of ED overcrowding are severe, with Prolonged wait times leading to patient discomfort, delayed treatments, and potentially exacerbating medical conditions [4]. ED waiting times are associated with short-term mortality and hospital admissions in high-acuity patients. Furthermore, overcrowding has been linked to increased medical errors, adverse events, and patient mortality [5, 6].

For healthcare providers, overcrowding translates into immense pressure to make rapid decisions, often with incomplete information as well. This high-stress environment contributes to burnout among healthcare professionals, a problem that has been exacerbated in recent years [7]. Burnout rates among ED physicians are significantly higher than the average across all physician specialties [8].

In this challenging environment, the need for efficient triage and rapid, accurate clinical decision-making is paramount. Triage, the process of prioritizing patients based on the severity of their condition, is especially critical. The Korean Triage and Acuity Scale (KTAS) [9], adapted from the Canadian Triage and Acuity Scale [10], has been widely adopted in South Korea for standardizing emergency triage processes. While KTAS is effective, challenges such as variability in triage assessments among different nurses and the cognitive demands during peak hours highlight its limitations in emergency settings [11].”

Healthcare systems have increasingly turned to technology-driven solutions to address these challenges [12]. Clinical Decision Support Systems (CDSS) [13] has emerged as a promising approach to enhancing clinical decision-making and improving patient outcomes. Traditional CDSS, often rule-based or utilizing simple machine learning algorithms, have shown potential in improving clinical outcomes and reducing medical errors. For instance, Ehteshami et al. demonstrated that health information technologies, including CDSS, improved adherence to clinical guidelines and reduced medication errors in various healthcare settings [14]

However, the application of CDSS in emergency medicine presents unique challenges. The dynamic and unpredictable nature of emergency care, coupled with the need for rapid decision-making, often stretches the capabilities of traditional CDSS. These systems frequently struggle with the complexity and variability inherent in emergency medicine, leading to limited adoption and impact [15, 16].

Recent advancements in artificial intelligence, particularly in large language models (LLMs), offer new possibilities for enhancing CDSS capabilities. LLMs, such as GPT-4 and its successors [17], have demonstrated remarkable proficiency in understanding context [18], processing natural

language, and generating human-like responses across various domains. These models also show promise in medical applications, including clinical text summarization, medical question answering, and even simulating patient-doctor conversations [19–21].

While LLMs have been increasingly applied in healthcare, much of this research focuses on general clinical applications like diagnostics, record summarization, and patient communication. However, the unique, high-stakes nature of emergency medicine remains largely unaddressed by current LLM applications. EDs demand real-time decision-making and precise triage under time constraints, challenges that traditional LLM applications in healthcare have not fully explored. This gap in current research presents an opportunity to develop LLM-based solutions tailored to the dynamic and critical nature of EDs. While previous studies have explored LLM usage in patient triage systems [22] or AI-driven CDSS [23], they lack the specialized, multi-agent configuration necessary for emergency departments [24], where role-specific, real-time decision-making is critical. In contrast, our study leverages a multi-agent system tailored for ED settings, with each agent mirroring specific ED roles to enable rapid and coordinated care.

In this study, we address this gap by proposing a novel multi-agent CDSS powered by state-of-the-art large language models. Our system integrates KTAS-based triage with comprehensive emergency care management, leveraging specialized AI agents to support ED physicians and nurses in real-time decision-making, treatment planning, and resource management. By adapting LLMs for this high-pressure environment, our approach represents a new application of AI that aligns with the unique demands of emergency care.

2. MATERIALS AND METHODS

2.1 System Architecture

Our Clinical Decision Support System (CDSS) is built upon a multi-agent architecture powered by the Llama-3-70b large language model [25]. At its core, Llama-3-70b serves as the foundation, chosen for its open-source capability and advanced natural language understanding. The system's multi-agent framework is orchestrated by CrewAI, which enables the creation and management of multiple AI agents [26]. This framework allows us to simulate key emergency room roles, enhancing the system's ability to provide comprehensive and coordinated care recommendations. Each agent operates with a specific role, goal, and set of tools, mirroring the collaborative nature of real emergency department teams. FIGURE 1 provides a schematic representation of our model's workflow.

2.2 AI Agents

Our CDSS comprises four distinct AI agents, each emulating a crucial role in the emergency department. The Emergency Physician Agent initiates the process by diagnosing patients and developing treatment plans. This agent analyzes the patient's presenting information along with additional examination data to formulate a diagnosis and differential diagnoses, consider potential complications, and outline necessary follow-up care. Access to up-to-date medical information through an

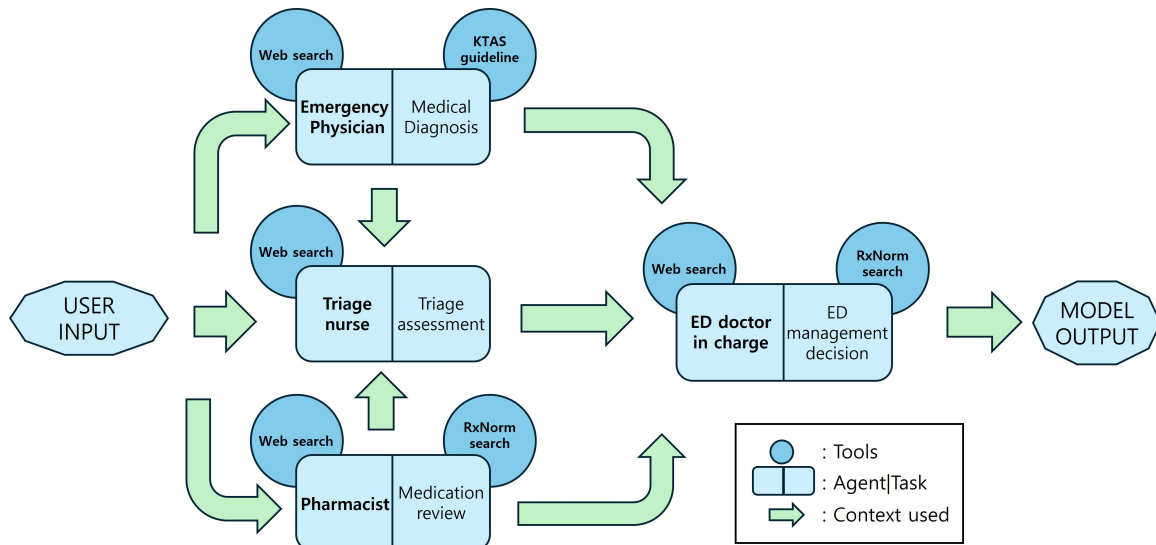


Figure 1: Model architecture.

integrated search function enhances the agent’s capabilities, allowing it to base its recommendations on the latest evidence-based practices.

Concurrently, medication safety is ensured by the Pharmacist Agent, who reviews the patient’s medication history, checks for drug interactions, and provides crucial information about proper medication administration and potential side effects. This agent leverages the RxNorm dataset [27], a comprehensive source of standard names for clinical drugs, to verify medications, check for interactions, and ensure appropriate dosing.

Following these initial assessments, the Triage Nurse Agent conducts the patient assessment using the Korean Triage and Acuity Scale (KTAS) [9]. This agent synthesizes the information provided by the Emergency Physician and Pharmacist agents, along with the patient’s symptoms, vital signs, and medical history, to determine the urgency of care needed. The agent’s decision-making process is guided by KTAS guidelines, ensuring accurate and consistent triage classifications based on a comprehensive view of the patient’s condition.

Finally, the ED Doctor in Charge agent oversees the entire patient care process. This agent integrates information from all other agents to make critical decisions about patient management, including finalizing treatment plans, determining admission or discharge, and planning follow-up care. The agent’s role is to ensure coherent and comprehensive patient management, taking into account both clinical needs and resource utilization based on the collective insights of all preceding agents.

2.3 Data Integration and Tool Utilization

We integrated several key data sources and tools to enhance the system’s decision-making capabilities. The KTAS guidelines were implemented as a comprehensive prompt, enabling the Triage Nurse Agent to accurately classify patients based on their presenting symptoms and vital signs. The

RxNorm dataset (version 2023AB) serves as the backbone for the Pharmacist Agent's medication management capabilities, providing a standardized nomenclature for clinical drugs and facilitating accurate interaction checking and dosing recommendations. Every agent has the ability to search the web via the DuckDuckGo search tool [28].

2.4 System Workflow

The system employs a streamlined workflow that mirrors the collaborative nature of emergency department operations while optimizing patient care. At the core of this system, the Emergency Physician Agent initiates the process by analyzing the patient's presenting information to generate an initial diagnosis and treatment strategy. Simultaneously, the Pharmacist Agent conducts a thorough review of the patient's medication history, utilizing the RxNorm database to identify potential drug interactions and provide critical pharmacological insights. These preliminary assessments inform the subsequent actions of the Triage Nurse Agent, who performs a comprehensive evaluation of the patient's condition, assigns a KTAS level, and compiles a succinct yet detailed triage report. The workflow culminates with the ED Doctor in Charge Agent, which integrates all preceding analyses to formulate definitive care decisions and management strategies.

2.5 Evaluation Methodology

We evaluated our CDSS using the Asclepius dataset, which comprises simulated emergency department scenarios [29]. Utilizing randomly selected 43 cases, scenarios were carefully curated to represent a diverse range of emergency medical situations, providing a comprehensive test of the system's capabilities across various clinical presentations.

An experienced emergency physician reviewed the CDSS outputs for each scenario. He assessed the accuracy of KTAS classification, appropriateness of diagnosis and treatment recommendations, comprehensiveness of medication reviews, and overall clinical decision quality. This expert evaluation provides a clinically relevant assessment of the system's performance.

To quantify the benefits of our multi-agent approach, we conducted a comparative analysis against a single-agent system using Llama-3-70b alone, with ED Doctor in Charge agent prompt. Key performance metrics included the concordance rate with expert KTAS classification, a clinical appropriateness score for management recommendations, and a coherence improvement rate compared to the single-agent system.

3. RESULTS

3.1 Performance of Triage Assessment with KTAS

We evaluated the performance of our CDSS in classifying patients according to KTAS levels. The system's performance was compared to that of a single-agent system, which served as a control. Both systems were evaluated against the KTAS levels assigned by an experienced ED physician.

TABLE 1 and TABLE 2, show the results of KTAS assessment evaluation for multi-agent and single-agent systems, respectively.

Table 1: Confusion Matrix of KTAS Classification by our CDSS

Model's KTAS Prediction	KTAS Level by ED Physician				
	1	2	3	4	5
1	3	0	0	0	0
2	0	15	2	2	1
3	0	1	7	5	1
4	0	0	0	0	0
5	0	0	0	0	5

Table 2: Confusion Matrix of KTAS Classification by Single-agent System

Model's KTAS Prediction	KTAS Level by ED Physician				
	1	2	3	4	5
1	2	1	0	0	0
1 or 2	0	4	1	0	0
2	0	8	3	1	2
3	0	2	5	4	3
3 or 4	0	0	0	0	1
4	0	0	0	0	0
5	0	0	0	0	0
Not applicable	1	1	0	2	1

To assess the performance of our multi-agent CDSS and compare it with a single-agent system, we calculated sensitivity and specificity for each KTAS level.

The multi-agent CDSS demonstrated strong sensitivity and specificity across KTAS levels, particularly excelling in KTAS levels 1 and 5. Sensitivity scores were 100%, 75%, 50%, 0%, and 100% for KTAS levels 1 through 5, respectively, while specificity scores were consistently high at 100%, 95.45%, 92.86%, 83.33%, and 94.59%. These values indicate the system's effectiveness in correctly identifying both high-acuity and low-acuity patients, with minor limitations in mid-acuity levels.

In contrast, the single-agent system exhibited lower sensitivity for KTAS levels 1 through 5, at 66.67%, 80%, 21.43%, 28.57%, and 100%, respectively, with specificity values of 100%, 65.63%, 73.91%, 95.65%, and 86.11%. Although the single-agent system achieved high specificity in certain levels, it displayed reduced sensitivity, particularly in mid-acuity levels, which could limit its effectiveness in the dynamic ED environment.

Beyond specificity and sensitivity, the multi-agent system accurately classified KTAS level 1 (most urgent), level 2, and level 5 (least urgent) cases. The system showed good performance, especially for KTAS level 2, which had a small number of misclassifications. However, the system encountered some challenges in differentiating between KTAS levels 3 and 5, suggesting an area for potential refinement.

In comparison, the single-agent system exhibited lower overall accuracy compared to the multi-agent system. A notable characteristic of the single-agent system was its tendency to provide range classifications (e.g., "1 or 2", "3 or 4") rather than specific KTAS levels. Moreover, in some cases, the single-agent system failed to predict any KTAS level, resulting in non-applicable (N/A) outputs.

The multi-agent approach demonstrated superior performance in several key aspects compared to the single-agent system. The multi-agent system consistently provided definitive KTAS classifications, while the single-agent system often resorted to range predictions or even did not predict the level (not applicable). Moreover, the multi-agent system exhibited higher overall accuracy, particularly for KTAS levels 1, 2, and 5. The multi-agent system's predictions also showed greater consistency with the ED physician's evaluations across all KTAS levels.

3.2 Evaluation on Clinical Decision Making

We evaluated the clinical decision-making capabilities of both the multi-agent and single-agent systems across various categories crucial to emergency care. The evaluation was conducted using two scales: a 5-point scale (1: very inaccurate, 5: very accurate) for primary diagnosis, critical findings, disposition decision, and justification; and a 1-point scale (0: inaccurate, 0.5: moderate, 1: accurate) for immediate action, medication, diagnostic test, consultation, and monitoring.

3.2.1 5-Point scale evaluation

TABLE 3 presents the results of the 5-point scale evaluation for both systems.

Table 3: 5-Point Scale Evaluation of Clinical Decision Making

Category	Multi-agent System					Single-agent System				
	1	2	3	4	5	1	2	3	4	5
Primary Diagnosis	0	0	0	2	41	0	0	0	2	41
Critical Findings	0	0	0	1	42	0	0	3	0	40
Justification	0	0	0	0	43	0	0	2	4	37

Both systems demonstrated high accuracy in primary diagnosis, with identical performance. However, the multi-agent system showed superior performance in identifying critical findings and providing justifications, with a higher proportion of cases rated as "very accurate" (5 points).

3.2.2 1-Point scale evaluation

TABLE 4 shows the results of the 1-point scale evaluation for both systems.

The multi-agent system consistently outperformed the single-agent system across all categories evaluated on the 1-point scale. Notably, the multi-agent system achieved perfect accuracy in imme-

Table 4: 1-Point Scale Evaluation of Clinical Decision Making

Category	Multi-agent System			Single-agent System		
	0	0.5	1	0	0.5	1
Disposition Decision	0	1	42	4	4	35
Immediate Action	0	0	43	1	4	38
Medication	0	0	43	0	3	40
Diagnostic Test	0	0	43	0	0	43
Consultation	0	0	43	1	0	42
Monitoring	0	0	43	0	0	43

mediate action, medication, diagnostic tests, consultations, and monitoring. While performing well, the single-agent system showed some inaccuracies and moderate performances across these categories.

These results indicate that the multi-agent system provides more consistent and accurate clinical decision-making support across various aspects of emergency care. The system's superior performance in critical areas suggests its potential to enhance the quality and safety of emergency care.

4. DISCUSSION

4.1 Model Performance

Our LLM-based multi-agent CDSS demonstrated superior performance compared to the single-agent system across multiple dimensions of emergency care decision-making. The system's effectiveness was particularly evident in KTAS-based triage and clinical decision-making.

Our multi-agent system exhibited high accuracy in KTAS classification, especially for the most urgent (KTAS level 1) and urgent (KTAS level 2) cases. This level of accuracy is crucial in emergency settings, where rapid and correct triage can significantly impact patient outcomes.

However, we observed a tendency for the system to overestimate urgency, particularly for KTAS levels 3 to 5. This bias towards higher urgency ratings contrasts with the typically more conservative approach of human ED physicians, who are trained to err on the side of caution. This discrepancy highlights an important difference between the decision-making processes of our LLM-based system and human clinicians. While the system's cautious approach might lead to over-triage in some cases, it could potentially reduce the risk of overlooking critically ill patients.

The multi-agent system's superiority over the single-agent system was particularly apparent in its decisiveness and consistency. While the single-agent system often provided range classifications or failed to predict a KTAS level, our multi-agent system consistently offered definitive classifications. This decisiveness is invaluable in real-world emergency departments where clear, rapid decisions are essential for efficient patient flow and resource allocation.

In terms of clinical decision-making, the multi-agent system demonstrated high accuracy across all evaluated categories. Notably, it achieved perfect scores in immediate action, medication, di-

agnostic tests, consultation, and monitoring on the 1-point scale evaluation. The system's strong performance in identifying critical findings and providing justifications for decisions, as evidenced by the 5-point scale evaluation, suggests its potential to enhance the quality and safety of emergency care delivery.

The consistent outperformance of the multi-agent system over the single-agent system across various aspects of emergency care decision-making underscores the effectiveness of our approach. This performance difference likely stems from the multi-agent system's ability to mimic the collaborative nature of real emergency department teams, where different specialists contribute their expertise to patient care.

4.2 Multi-agent System Effectiveness

Our research builds on previous work in AI-based CDSS applications but introduces a novel multi-agent system tailored specifically for emergency departments. By creating distinct agents that mirror key ED roles, our approach addresses the limitations identified in prior single-agent or non-specialized systems, facilitating a collaborative and adaptable CDSS that can better support emergency healthcare providers under pressure.

The superior performance of our multi-agent CDSS compared to the single-agent system underscores the potential of multi-agent strategies in leveraging LLMs for healthcare applications. This approach represents a significant advancement in the field of clinical decision support, particularly in the complex and time-sensitive environment of emergency departments.

Our multi-agent system's effectiveness stems from its ability to emulate the collaborative nature of real emergency department teams. By dividing tasks among specialized agents - the Triage Nurse, Emergency Physician, Pharmacist, and ED Doctor in Charge - the system mirrors the diverse expertise found in actual ED settings. This division of labor allows each agent to focus on its specific domain, leading to more accurate and comprehensive decision-making.

The Triage Nurse agent's proficiency in KTAS classification, coupled with the Emergency Physician agent's diagnostic capabilities, creates a robust initial assessment process. The Pharmacist agent's role in medication management adds a crucial layer of patient safety, while the ED Doctor in Charge agent synthesizes all inputs to make final decisions. This structure enables the system to handle the multifaceted nature of emergency care more effectively than a single-agent approach.

Moreover, the multi-agent strategy addresses one of the key challenges in applying LLMs to healthcare: the need for domain-specific expertise. While LLMs possess broad knowledge, they may lack the specialized understanding required for specific medical tasks. By creating agents with defined roles, we can fine-tune each agent or utilize specific biomedical-pertained models as agents to align more closely with the specialized knowledge and decision-making processes of different ED professionals.

4.3 Clinical Implications

The CDSS's strong performance in clinical decision-making across various categories (primary diagnosis, critical findings, disposition decision, immediate action, medication, diagnostic tests, consultation, and monitoring) suggests its potential to enhance the quality and consistency of care. The system could help standardize care practices by providing accurate, evidence-based recommendations, potentially reducing variability in treatment decisions and improving adherence to clinical guidelines.

The system's pharmacological recommendations, supported by integration with the RxNorm database, generally resulted in medically correct suggestions and have implications for medication safety. By providing an additional layer of checking for drug interactions and appropriate dosing, the CDSS could help reduce medication errors, a significant concern in emergency care settings. However, we observed instances where the system recommended medicines that were not widely used or prohibitively expensive. This highlights the importance of contextualizing the CDSS with hospital-specific formularies and considering medication costs. In real-world applications, integrating the hospital's own medication database and providing the agent with context about medication pricing could significantly enhance the practical utility of these recommendations.

The system's comprehensive output, covering aspects from initial triage to disposition decisions, could facilitate better communication and handovers between different ED staff members. This improved information flow could lead to more coordinated care and potentially reduce errors related to miscommunication.

In terms of resource allocation, the CDSS's accurate predictions for necessary diagnostic tests, consultations, and monitoring could help optimize the use of hospital resources. This efficiency could be particularly beneficial in addressing the challenges of ED overcrowding, potentially reducing length of stay and improving patient throughput.

It's crucial to note that while the CDSS shows promise in enhancing various aspects of emergency care, it is designed to support, not replace, clinical decision-making. The system's recommendations should be used in conjunction with clinical expertise and judgment.

Importantly, the effective integration of the CDSS into clinical workflows remains a critical consideration. The system should complement rather than disrupt the established triage and treatment planning processes. Developing methodologies for seamless integration is essential for the system's success in real-world settings. This could involve creating an intuitive user interface (UI) that aligns with ED staff's needs and workflow patterns. Additionally, integration with the hospital's existing Electronic Medical Record (EMR) system could significantly enhance the CDSS's utility, allowing for real-time access to patient data and smoother incorporation of CDSS recommendations into patient care processes.

Furthermore, the clinical implications of this CDSS extend beyond immediate patient care. By providing a standardized approach to decision-making, the system could also serve as a valuable tool for training new ED clinicians, offering consistent, evidence-based guidance that can supplement experiential learning.

4.4 Ethical Considerations

The implementation of AI-driven clinical decision support systems in emergency medicine raises important ethical considerations and data privacy concerns that must be carefully addressed.

First and foremost, patient privacy and data protection are paramount. Our multi-agent CDSS was designed with these concerns in mind, utilizing the Llama-3-70b model, which can be run locally. This approach significantly reduces the risk of sensitive patient data being transmitted or stored externally, aligning with stringent healthcare data protection regulations such as HIPAA in the United States or GDPR in Europe [30, 31]. By keeping data processing on-premise, we mitigate the risks associated with data breaches or unauthorized access that could occur with cloud-based solutions.

However, the use of AI in clinical decision-making raises questions about accountability and transparency. While our system demonstrates high accuracy, it's crucial to maintain a "human-in-the-loop" approach where healthcare professionals retain ultimate responsibility for clinical decisions. The CDSS should be viewed as a tool to augment, not replace, clinical judgment. Clear guidelines must be established to delineate the roles of the AI system and human clinicians, ensuring that the ethical and legal responsibilities of care remain with healthcare professionals.

Transparency in AI decision-making processes is another critical ethical consideration. Although large language models like Llama-3-70b can provide highly accurate outputs, their decision-making processes can be opaque. This "black box" nature of AI can be problematic in healthcare settings where the rationale behind decisions is crucial. Future development should focus on improving the explainability of AI recommendations, possibly through the integration of explainable AI techniques.

Bias in AI systems is a significant ethical concern that requires ongoing attention [32]. While our multi-agent system aims to mimic the collaborative nature of ED teams, it's important to recognize that biases present in training data or model architecture could lead to systematic errors or unfair treatment of certain patient groups. Regular system performance audits across diverse patient populations will be necessary to identify and mitigate any biases.

The potential for AI systems to influence clinical decision-making also raises questions about patient autonomy and informed consent. Patients should be made aware when AI systems are being used to support their care, and consideration should be given to how this information is communicated without causing undue concern or eroding trust in healthcare providers.

4.5 Limitations and Future work

One significant limitation we encountered was related to the dataset used in the study. The Asclepius dataset, which is very descriptive and complex, may not fully represent real-world scenarios. Although comprehensive data is generally beneficial, real-world datasets in emergency medicine are often characterized by sparsity, imbalance, and a substantial proportion of missing values. These features pose significant challenges in developing robust models for clinical decision support. The discrepancy is particularly evident in emergency situations, such as those handled through emer-

gency telephone calls. In these scenarios, information is often incomplete, rapidly changing, and collected under high-stress conditions. Emergency responders and dispatchers have been dealing with such data limitations since the inception of emergency call systems, often making critical decisions based on partial information and relying heavily on experience and standardized protocols to fill in the gaps. However, the Asclepius database, which we used to evaluate our CDSS, contains comprehensive and well-structured data. While this allowed for a thorough initial assessment of our system's capabilities, it also means that our evaluation benchmarks may differ from what we would expect in real-world performance. The system's ability to handle incomplete or inconsistent data, a crucial skill in actual emergency settings, was not fully tested in this study. This limitation underscores the need for future research to bridge the gap between idealized datasets and the messy reality of emergency medicine, incorporating more realistic data scenarios in subsequent evaluations of our CDSS to better assess its robustness and applicability in real-world emergency departments.

An important consideration in our research was the potential use of real-world emergency phone call data. Although we were unable to acquire such data for this study, our AI system has the potential to interpret emergency situations and provide crucial information to ED staff by integrating Speech-to-Text (STT) models. This capability could significantly enhance preparedness in the ED, allowing medical staff to anticipate incoming patients' needs more effectively. By providing early insights into the patient's condition, the system could facilitate the proactive preparation of resources such as treatment spaces, medications, diagnostic tests, and necessary medical equipment. This advance notice could streamline ED operations, potentially reducing wait times and improving the overall quality of emergency care by ensuring that appropriate resources are ready when the patient arrives.

While we equipped our AI with internet search capabilities via the DuckDuckGo tool, relying on internet searches for medical information is not optimal. The potential for information overload, exposure to false information, lack of source credibility, and the slight time delay associated with web searches could lead to misinformation, confusion, and inefficiencies in clinical decision-making. A more promising approach would be to utilize verified information from medical textbooks and databases, though this presents its own challenges in terms of copyright issues and credibility verification. To enhance the quality of our CDSS's responses, implementing standardized medical terminologies such as the Systematized Nomenclature of Medicine Clinical Terms (SNOMED CT) [33] and Logical Observation Identifiers Names and Codes (LOINC) [34] could be beneficial. Furthermore, integrating the CDSS into existing Electronic Medical Record (EMR) systems could significantly improve the productivity of ED physicians and nurses by assisting in the creation of medical records based on trained medical concepts. This integration would not only streamline workflow but also ensure that the AI's recommendations are grounded in verified, standardized medical knowledge, thereby enhancing the overall quality and reliability of the clinical decision support provided.

5. CONCLUSION

In conclusion, our multi-agent CDSS represents a significant step forward in the application of AI to emergency medicine. By demonstrating high performance across the spectrum of emergency care tasks, our system shows promise in enhancing emergency care quality, efficiency, and safety. As we move forward with further validation and refinement, this technology has the potential to significantly improve patient outcomes in these critical healthcare settings.

6. ACKNOWLEDGMENTS

We would like to thank Sangkuk Han, MD, Ph.D., for reviewing the model and evaluating the answers. Wongyung Choi is the corresponding author.

7. DATA AVAILABILITY

The data used in this study is publicly available through the Asclepius dataset, which can be accessed online.

8. CODE AVAILABILITY

The code used for this research is openly available in a GitHub repository at https://github.com/junhan51/LLM_ED_CDSS.

9. COMPETING INTERESTS

Not applicable.

References

- [1] Tang N, Stein J, Hsia RY, Maselli JH, Gonzales R. Trends and Characteristics of Us Emergency Department Visits 1997-2007. *JAMA*. Aug. 2010;304:664-670.
- [2] Jung HM, Kim MJ, Kim JH, Park YS, Chung HS, et. al. The Effect of Overcrowding in Emergency Departments on the Admission Rate According to the Emergency Triage Level. *PLOS ONE*. 2021;16:e0247042.
- [3] https://www.e-gen.or.kr/nemc/statistics_annual_report.do
- [4] Davenport PJ, O'Connor SJ, Szychowski JM, Landry AY, Hernandez SR. The Relationship Between Emergency Department Wait Times and Inpatient Satisfaction. *Health Mark Q*. Apr. 2017;34:97-112.
- [5] Plunkett PK, Byrne DG, Breslin T, Bennett K, Silke B. Increasing Wait Times Predict Increasing Mortality for Emergency Medical Admissions. *Eur J Emerg Med*. Aug. 2011;18:192-196.
- [6] Epstein SK, Huckins DS, Liu SW, Pallin DJ, Sullivan AF, et. al. Emergency Department Crowding and Risk of Preventable Medical Errors. *Intern Emerg Med*. 2012;7:173-180.
- [7] Watson AG, McCoy JV, Mathew J, Gundersen DA, Eisenstein RM. Impact of Physician Workload on Burnout in the Emergency Department. *Psychol Health Med*. 2019;24:414-428.

- [8] Estryng-Behar M, Doppia MA, Guetarni K, Fry C, Machet G, et. al. Emergency Physicians Accumulate More Stress Factors Than Other Physicians-Results From the French SESMAT Study. *Emerg Med J.* 2011;28:397-410.
- [9] Park J, Lim T. Korean Triage and Acuity Scale (Ktas). *J Korean Soc Emerg Med.* 2017;28:547-551.
- [10] J Murray MJ. The Canadian Triage and Acuity Scale: A Canadian Perspective on Emergency Department Triage. *Emerg Med (Fremantle).* 2003;15:6-10.
- [11] Park JB, Lee J, Kim YJ, Lee JH, Lim TH. Reliability of Korean Triage and Acuity Scale: Interrater Agreement Between Two Experienced Nurses by Real-Time Triage and Analysis of Influencing Factors to Disagreement of Triage Levels. *J Korean Med Sci.* 2019;34:e189.
- [12] Rajpurkar P, Chen E, Banerjee O, Topol EJ. AI in Health and Medicine. *Nat Med.* Jan. 2022;28:31-38.
- [13] <https://library.oapen.org/bitstream/handle/20.500.12657/22918/1/1007243.pdf#page=153>
- [14] Kasaei M, Rezaei P, Tavakoli N, Ehteshami A. The Role of Health Information Technology in Reducing Preventable Medical Errors and Improving Patient Safety. *Int J Health Syst Disaster Manag.* 2013;1:195-199.
- [15] Xiao Y, Hunter WA, Mackenzie CF, Jefferies NJ, Horst RL, et. al. Special Section: Task Complexity in Emergency Medical Care and Its Implications for Team Coordination. human factors. *J Hum Factors Ergon Soc.* Dec. 1996;38:636-645.
- [16] Perry WM, Hossain R, Taylor RA. Assessment of the Feasibility of Automated Real-Time Clinical Decision Support in the Emergency Department Using Electronic Health Record Data. *BMC Emerg Med.* 2018;18:1-6.
- [17] OpenAI, Achiam J, Adler S, Agarwal S, Ahmad L, et al. Gpt-4 Technical Report. arXiv preprint arXiv:2303.08774. 2023 Mar 15.
- [18] Li M, Sun J, Tan X. Evaluating the Effectiveness of Large Language Models in Abstract Screening: A Comparative Analysis. *Syst Rev.* 2024;13:219.
- [19] Tu T, Palepu A, Schaekermann M, Saab K, Freyberg J, et. al. Towards Conversational Diagnostic AI. 2024. ArXiv preprint: <https://doi.org/10.48550/arxiv.2401.05654>
- [20] Thirunavukarasu AJ, Ting DS, Elangovan K, Gutierrez L, Tan TF, et. al. Large Language Models in Medicine. *Nat Med.* 2023;29:1930-1940.
- [21] Van Veen D, Van Uden C, Blankemeier L, Delbrouck JB, Aali A, et al. Adapted Large Language Models Can Outperform Medical Experts in Clinical Text Summarization. *Nat Med.* 2024;30:1134-1142.
- [22] Williams CY, Zack T, Miao BY, Sushil M, Wang M, et. al. Use of a Large Language Model to Assess Clinical Acuity of Adults in the Emergency Department. *JAMA Netw Open.* 2024;7:e248895.
- [23] Karabacak M, Margetis K. Embracing Large Language Models for Medical Applications: Opportunities and Challenges. *Cureus.* 2023;15:e39305.

- [24] Guo T, Chen X, Wang Y, Chang R, Pei S, et. al. Large Language Model Based Multi-Agents: A Survey of Progress and Challenges. 2024. Arxiv Preprint: <https://arxiv.org/abs/2402.01680>
- [25] Touvron H, Lavril T, Izacard G, Martinet X, Lachaux MA, et. al. Llama: Open and Efficient Foundation Language Models. 2023. ArXiv preprint: <https://arxiv.org/pdf/2302.13971>
- [26] <https://github.com/crewAIInc/crewAI>
- [27] Nelson SJ, Zeng K, Kilbourne J, Powell T, Moore R. Normalized Names for Clinical Drugs: RxNorm at 6 Years. *J Am Med Inform Assoc.* 2011;18:441-448.
- [28] <https://pypi.org/project/duckduckgo-search/>
- [29] <https://physionet.org/content/asclepius-r/1.1.0/>
- [30] <https://pharmaceutical-journal.com/article/news/essential-guide-to-the-general-data-protection-regulation-gdpr>
- [31] Moore W, Frye S. Review of HIPAA Part 1: history protected health information and privacy and security rules. *J Nucl Med Technol.* 2019;47:269-272.
- [32] Ferrara E. Should ChatGPT Be Biased? Challenges and Risks of Bias in Large Language Models. 2023. ArXiv preprint: <https://arxiv.org/pdf/2304.03738>
- [33] <http://www.snomed.org/>
- [34] <https://loinc.org/>