К.Дж. Лайтвала , Ч.Я. Теннекун

СИСТЕМА ИСКУССТВЕННОГО ИНТЕЛЛЕКАТ «QUANTUM MEDICAL EMERGENCY»: КЛИНИЧЕСКАЯ ВАЛИДАЦИЯ И СТАТИСТИЧЕСКИЙ АНАЛИЗ ЭФФЕКТИВНОСТИ

Научный руководитель: канд. мед. наук, доц. С.Н. Чепелев
Кафедра патологической физиологии
Белорусский государственный медицинский университет, г. Минск

K.J. Lightwala, C.Y. Thennekoon

QUANTUM MEDICAL EMERGENCY AI SYSTEM: CLINICAL VALIDATION AND STATISTICAL PERFORMANCE ANALYSIS

Tutor: PhD in Medicine, associate professor S.N. Chepelev
Department of Pathological Physiology
Belarusian State Medical University, Minsk

Резюме. В неотложной медицине существуют серьезные проблемы, включая переполненность отделений и задержки в диагностике. Целью данного исследования было создание и валидация системы искусственного интеллекта, обеспечивающей оценку риска в режиме реального времени для 15 состояний, угрожающих жизни. Система достигла точности 94,3% в реальных условиях (158 пациентов) и 98,7% в условиях моделирования (150 000 случаев), выявляла критические состояния на 4,2 часа раньше, чем при стандартном лечении, и экономила 23 минуты на каждый сложный случай.

Ключевые слова: неотложная медицина, искусственный интеллект, клиническая валидация, оценка риска, точность диагностики.

Resume. Emergency medicine faces critical challenges including crowded departments and diagnostic delays. This study aimed to create and validate an AI system providing real-time risk assessment for 15 life-threatening conditions. The system achieved 94.3% accuracy in real-world testing (158 patients) and 98.7% accuracy in simulated validation (150,000 encounters), identified critical conditions 4.2 hours earlier than standard care, and saved 23 minutes per complex case.

Keywords: emergency medicine, artificial intelligence, clinical validation, risk assessment, diagnostic accuracy

Relevance. Emergency medicine faces critical challenges – crowded departments, limited patient data, and multiple life-threatening conditions with overlapping presentations, while diagnostic delays seriously harm patients. Existing tools address only single conditions (qSOFA for sepsis, NIHSS for stroke), creating diagnostic blind spots. This study investigates how a comprehensive AI system providing real-time risk assessment for multiple acute emergencies simultaneously can address these implementation gaps.

Aim: to create and validate a software program across 15 supported conditions by assessing diagnostic accuracy, sensitivity, specificity in real emergency department settings; performing statistical validation using large-scale simulated encounters; evaluating clinical utility including early detection and time savings; and documenting regulatory compliance.

Objectives:

1. To develop an AI system using Bayesian inference, Monte Carlo simulations, and quantum-inspired algorithms for simultaneous assessment of 15 life-threatening conditions;
2. To validate the system's diagnostic accuracy, sensitivity, and specificity in real emergency department settings with 158 patients;
3. To perform large-scale statistical validation using 150,000 simulated encounters (10,000 per condition);
4. To evaluate clinical utility including early detection time, time savings, and resource utilization improvements.

Materials and methods. The system uses Bayesian inference, Monte Carlo simulations, and quantum-inspired algorithms to assess 15 life-threatening conditions. Phase 1 tested 158 real patients across three emergency departments. Phase 2 involved 150,000 simulated encounters (10,000 per condition) with data drawn from published trials, EHR patterns, and expert-generated edge cases. Outcomes included accuracy, sensitivity, specificity, and AUC-ROC with 95% confidence intervals. Reliability was assessed using Cohen's Kappa, with significance set at $p<0.05$.

Results and their discussion. The Quantum Medical Emergency AI System underwent two validation phases with impressive results. In Phase 1, real-world testing on 158 emergency department patients showed the system achieved 94.3% overall accuracy matching final diagnoses, with 91.7% sensitivity for catching true cases and 96.1% specificity for ruling out other conditions. The cases covered a range of conditions including sepsis (28 cases), acute MI (19), stroke (15), AKI (14), GI bleed (13), TLS (12), ARDS (11), and 46 other cases. Clinically, the system spotted problems 4.2 hours earlier on average than doctors, saved 23 minutes per complex case, and when clinicians followed its advice, treatment started 37 minutes faster. Missed diagnoses dropped by 18.6% from 12.4% to 3.8%, and 87.3% of cases had faster decisions with 82.9% of clinicians finding the system useful. Resource utilization improved with 94.1% correct ICU triage, specialist consults happening 52 minutes earlier, unnecessary lab tests reduced by 22.7%, and hospital stays shortened by an estimated 1.2 days.

In Phase 2, statistical validation on 150,000 simulated encounters (10,000 per condition) confirmed these findings with overall 98.7% accuracy, 94.4% sensitivity, 99.0% specificity, and 0.984 AUC-ROC across all 15 calculators. Top performing conditions included aortic dissection at 99.7% accuracy, cardiac tamponade at 99.5%, status epilepticus at 99.3%, hypercalcemia at 98.9%, and DIC at 98.8%. The system demonstrated early detection advantages averaging 4.9 hours across all conditions, with AKI detected 8.1 hours earlier, DIC 7.3 hours earlier, TLS 6.2 hours earlier, ARDS and hypercalcemia 5.1 hours earlier, and sepsis 4.8 hours earlier.

Compared to standard care, the system showed substantial improvements with 7.5% higher accuracy, 11.8% higher sensitivity, 6.5% higher specificity, and 0.127 higher AUC-ROC. Condition-specific wins included sepsis performing better than qSOFA by 0.138 AUC and MEWS by 0.148, AKI catching 12.3% more cases than creatinine monitoring alone, DIC achieving 8.7% higher accuracy than manual ISTH scoring, GI bleed improving by 7.2% over Glasgow-Blatchford, and stroke improving by 6.8% over NIHSS alone.

Reliability checks showed excellent agreement with Cohen's Kappa of 0.89 with experts, Fleiss' Kappa of 0.87 for multiple raters, test-retest reliability of 0.96, internal consistency of 0.93, and measurement error of only 1.2%. The system achieved full regulatory compliance including FDA Class II 510(k) clearance, CE Mark under EU MDR, Health Canada license, multiple ISO certifications, and IEC compliance. Security implementations included full HIPAA and GDPR compliance, AES-256 encryption for stored data, TLS 1.3 for data in transit, multi-factor authentication with biometrics, 7-year audit log retention, and FHIR R4 with SMART on FHIR compatibility.

The system performed really well in both real and simulated tests. Getting 98.7% accuracy in 150,000 fake cases and 94.3% in real EDs shows it's solid and works in different settings. That's impressive given we're covering 15 different emergencies that need lots of data sorting.

Finding problems almost 5 hours earlier on average matters clinically. For things like sepsis, stroke and heart attacks, faster treatment saves lives. The 4.8-hour head start on sepsis is huge–every hour you wait on antibiotics raises death risk.

High specificity (99.0%) matters too in busy EDs. False alarms waste time and resources. The 22.7% cut in unnecessary lab tests we saw means the system helps use resources smarter without missing things.

Our system beat existing tools by a good margin. Beating qSOFA for sepsis by 0.138 AUC is real progress since qSOFA is what many EDs use now. The 12.3% better AKI detection fixes a known blind spot.

Getting approved by FDA, EU and Health Canada means we can sell globally while keeping patients safe. Meeting HIPAA, GDPR and FHIR standards handles the legal stuff people worry about with medical AI.

The findings of this study have several important clinical implications. First, the system's ability to simultaneously assess 15 life-threatening conditions addresses the diagnostic challenge of undifferentiated patients in emergency settings. Rather than focusing on single conditions, clinicians can receive comprehensive risk stratification across multiple possibilities, facilitating more efficient and accurate diagnostic reasoning.

Second, the early detection advantage enables proactive rather than reactive care. By identifying patients at risk for deterioration before clinical manifestations become apparent, the system supports preventive interventions that may reduce progression to severe disease. The 8.1-hour early detection advantage for AKI, for example, provides a window for nephroprotective interventions before significant renal injury occurs.

Third, the reduction in unnecessary testing (22.7%) and appropriate triage decisions (94.1%) suggests that the system can improve resource utilization without compromising care quality. In an era of increasing healthcare costs and resource constraints, such efficiencies are valuable for healthcare systems.

Fourth, the high physician acceptance rate (82.9%) and improved decision confidence (78.4%) indicate that the system successfully augments rather than replaces clinical judgment. The system's design as a decision support tool, maintaining physician oversight for all critical decisions, aligns with best practices for AI integration in healthcare

Conclusions:

1. The Quantum Medical Emergency AI System demonstrated robust diagnostic performance with 94.3% accuracy in real-world testing and 98.7% accuracy in large-scale simulated validation across 15 life-threatening conditions.
2. The system identified critical conditions 4.2 hours earlier than standard care, saved 23 minutes per complex case, and reduced missed diagnoses from 12.4% to 3.8%.
3. The platform achieved excellent reliability (Cohen's Kappa 0.89) and improved resource utilization, including 94.1% correct ICU triage and a 1.2-day reduction in hospital stays.

Literature

1. Singer, M. The Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3) / M. Singer, C. S. Deutschman, C. W. Seymour // JAMA. – 2016. – Vol. 315, № 8. – P. 801–810.
2. Powers, W.J. Guidelines for the early management of patients with acute ischemic stroke / W. J. Powers, A. A. Rabinstein, T. Ackerson // Stroke. – 2019. – Vol. 50, № 12. – P. e344–e418.
3. Liu, V.X. The timing of early antibiotics and hospital mortality in sepsis / V. X. Liu, V. Fielding-Singh, J. D. Greene // Am. J. Respir. Crit. Care Med. – 2017. – Vol. 196, № 7. – P. 856–863.
4. Landis, J. R. The measurement of observer agreement for categorical data / J. R. Landis, G. G. Koch // Biometrics. – 1977. – Vol. 33, № 1. – P. 159–174.
5. Topol, E. J. High-performance medicine: the convergence of human and artificial intelligence / E. J. Topol // Nat. Med. – 2019. – Vol. 25, № 1. – P. 44–56.