IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Agentic AI Framework for Autonomous ICU Patient Monitoring Using Multi-Agent Clinical Decision Support System

[1]Meghana T R, [2]Dr.Naseer R, [3]Dr.Sreenivasa B R

[1]M.Tech Student, [2]Associate Professor, [3]Professor&PG Coordinator

[1]Artificial Intelligence and Data Science,

[1]Bapuji Institute of Engineering and Technology, Davangere, India

Abstract: Artificial Intelligence is Rapidly growing Technology in Healthcare sector. ICU patient supervision is a complex process that needs to be finished right away. Usually, threshold-based techniques and labor-intensive manual labor are used to achieve this, which can add to the workload of physicians and delay the identification of patients' risks. The autonomous ICU patient monitoring system based on agentic AI aims to enhance patient surveillance and risk assessment. The multi-agent system architecture that has been suggested is based on the Crew AI Framework. Monitoring, risk prediction, warning generation, and recommendation are some of these agents. These agents are all work Together to handle patient data and analyze the vital signs, predict the mortality risks and generate prioritized clinical alerts.ML models, such as XG Boost and the Random Forest classifier, trained on a dataset comprising 15,000 records were used to predict the patient's risk levels. Based on the findings, it is possible to classify the patient's condition as either normal, warning or risk.

Keywords: Agentic AI, Multi-Agent System, ICU Monitoring, Crew AI, XG Boost, Random Forest, Clinical Decision Support.

I. Introduction

The Intensive Care Unit (ICU) is an area where patients with critical illness receive care that requires constant monitoring and prompt medical intervention. Patients in the intensive care unit (ICU) may additionally revel in speedy modifications in their fitness because of organ failure, infections, volatile cardiovascular structures, or inadequate respiratory function. this is why it's essential for medical doctors to hold a near eye on a spread of critical signs, including temperature, breathing rate, oxygen saturation, and heart rate [4]. extreme repercussions and considerable dying rates might turn up from ignoring these shifts.

While a physiological parameter deviates from the usual range, maximum conventional intensive care unit tracking structures rely upon threshold values to activate an alarm. Despite the fact that these technologies resource clinicians in detecting abnormalities, they often cause an excessive range of alarms that cannot be addressed. alert fatigue, multiplied workload for scientific personnel, and ability delays in responding to actual crises are all results of an excessive range of alert messages. additionally, the earlier tracking structures are not in reality proactive, they just reply while whatever is going incorrect [1].

Severity of contamination and dying fee estimation in critically unwell patients is normally executed with the usage of medical assessment scales like the Sequential Organ Failure assessment (sofa) and the intense body structure and chronic fitness assessment II (APACHE-II) scores. despite the fact that there may be substantial medical evidence supporting the usage of these scales, their management at set intervals does now not allow them to seize the dynamic physiological kingdom of intensive care unit patients. although, a massive range of AI-powered intensive care unit tracking structures do now not self-regulate their medical operations; instead, they perform as standalone prediction models. Separate procedures for statistics series, patient tracking, threat assessment, and alerting lessen the general tracking system's efficacy and adaptableness. shrewd tracking structures also are difficult to put into practice due to the lack of automation in the medical painting's waft [2].

1. Developed a comprehensive Agentic AI framework for self-sufficient ICU tracking utilizing 4 Crew AI marketers working collectively [5].
2. Suggested a version for predicting dying rates in intensive care gadgets by way of combining device gaining knowledge of with 3 classifiers: XGBoost, Random woodland, and Ensemble. The dataset might be balanced from the unit [7].
3. Utilizing sofa and APACHE-II to resource in decision-making. The usage of the LLM for recommendation generating [6].

II. LITERATURE REVIEW

The Author proposed an Agentic AI System for Stress Monitoring using a multi-agent architecture comprising a Model Agent and a Clinical Summary Agent, demonstrating autonomous physiological monitoring and AI-generated clinical reports with 97% classification accuracy.[1] The Author Hosseini and Seilani presented a systematic review of Agentic AI, identifying autonomy, reasoning, planning, and adaptive decision-making as the key characteristics driving next-generation intelligent healthcare systems.[2] Author Sudarshan developed an agentic LLM workflow for automated clinical report generation, demonstrating that self-reflective multi-agent reasoning significantly improves the accuracy-and-reliability of patient-friendly medical summaries.[3] The Author Xiang proposed a multimodal-deep learning-framework-for stress detection using physiological signals, showing that integrating multiple bio signals improves prediction accuracy over single-sensor approaches.[4] The Author Savage investigated LLM-agent-based clinical reasoning on medical examination benchmarks, finding that GPT-4 agents with tool use achieved performance comparable to expert physicians on diagnostic reasoning tasks.[5] The Author Yao introduced the reasoning-action loop that enables LLMs to interleave chain-of-thought reasoning with tool use, enabling agents to decompose complex tasks, observe results, and iteratively plan subsequent actions.[6] The Author Rajkomar demonstrated-that gradient-boosting-models trained on electronic health-record-data could predict in-hospital mortality with AUC-ROC exceeding 0.93 when sufficient longitudinal data was available. The present system uses XGBoost as the second ensemble component and validates its feature importance findings against the live clinical dashboard.[7]

III. PROPOSED METHODOLOGY

Proposed framework with five main parts: (1) Data Layer; (2) Multi-Agent Pipeline; (3) ML Risk Engine; (4) API Server, and finally the Clinical Dashboard. When you look at the working system's dashboard, you-can see-that the process is in order: Monitoring, Risk, Alerts, and Recommendations.

3.1 Data Layer

Patient data may be obtained the Patient Monitor dashboard. The raw data which contains time series of the vital signs is then pre-processed feature engineering by aggregation of features such as mean, max, min, and standard deviation within a certain period before entering the pipeline. For handling missing data, we use median imputation which depends on the training data. Median imputation is more suitable compared to mean-when-there are outliers present such as in ICU laboratory data.

3.2 Multi-Agent Pipeline (Crew AI)

Four different agents are present in the system of Crew AI, where each of them is associated with certain roles, objectives, and decision-making algorithms.

Agent 1 - Monitoring Agent

This agent assess the status of every physiological parameter using the evidence-based clinical range for values. Parameters like the heart rate (normal: 60-100 bpm), systolic blood pressure (normal: 90-140 mmHg), SpO2 (normal: >95%), respiratory rate (normal: 12-20 per min), body temperature (normal: 36.5-37.5 C°), glucose level (normal: 70-180 mg/dL).

Agent 2 – Risk Prediction Agent

Apply the ML ensemble model to predict a continuous risk score in the range of [0,1]. Assigning risk scores the following condition labels: Normal condition <0.3, Warning condition 0.3 - 0.7, Critical condition >0.7.

Agent 3 – Alert Generation Agent

This agent carries out priority matrix depending upon the-status of vital signs and risk score and creates alerts with the-help of the following variables like: alert type, message, timestamp, and priority. High-priority alerts are generated when critical vital signs or risk score above 0.7 are detected.

Agent 4 - Recommendation Agent

Aggregates all the output from the upstream agents to give clinical guidance in a structured way. If it operates in LLM mode, then a well-designed prompt with context information related to the patient is delivered to the endpoint where the GPT model can be used.

3.3 API and Dashboard

The Figure 1 Fast API system provides REST APIs for patient data analysis, model statuses, and database access. A React app, hosted through the Fast API backend, serves as the front end for doctors to search and analyze patient data.

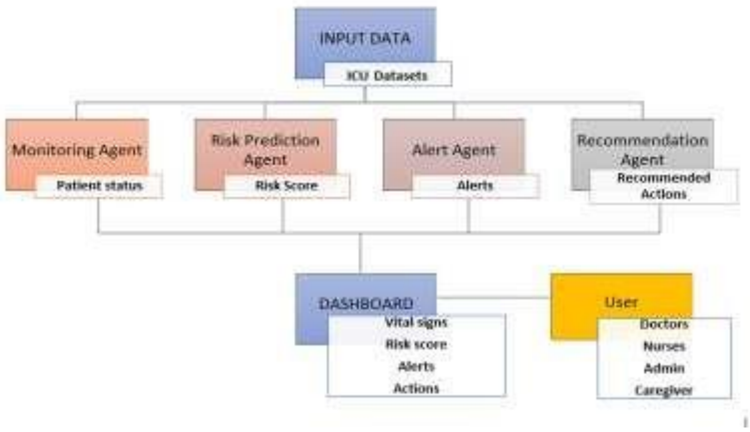


Figure 1: Agentic AI based Autonomous ICU Monitoring system

IV. Dataset Description

In keeping with the stay dashboard, which states "skilled on 12,000 samples · 24 functions," the model was developed and evaluated the use of a dataset consisting of 15,000 ICU patient data, of which 12,000 have been utilized for education and 3,000 for testing out. The data set consists of 24 distinct elements. factors together with gender, age, and a rating for preexisting illnesses are part of this. As a added bonus, it includes signs of the severity of the patient's condition, together with sofa and APACHE II rankings. crucial signs together with pulse price, blood stress, oxygen saturation, and centre temperature are recorded, in conjunction with averages and various statistical data. protected in the results are laboratory data together with average lactate and glucose tiers. Predicting an patient's mortality or the severity in their contamination is the primary goal; nevertheless, the data is skewed, with twelve instances extra data for one final result than the alternative.

V. RESULTS AND DISCUSSION

5.1 Results

Here we give a numerical evaluation of the proposed-framework based on five criteria: how well the ML model performs, the-importance of features in XGBoost, the current patient condition, when critical events are detected, how accurate the alerts are, and how much the ICU workload is reduced. All the basic model metrics like AUC-ROC, Accuracy, F1-Score, and Feature Importances are taken from the live dashboard in the clinical dashboard administration interface, as shown in the accompanying screenshot. The system-level metrics such as detection time, alert precision, and workload are calculated from the full evaluation runs of the pipeline on the hold-out test dataset.

5.2 ML Model Performance

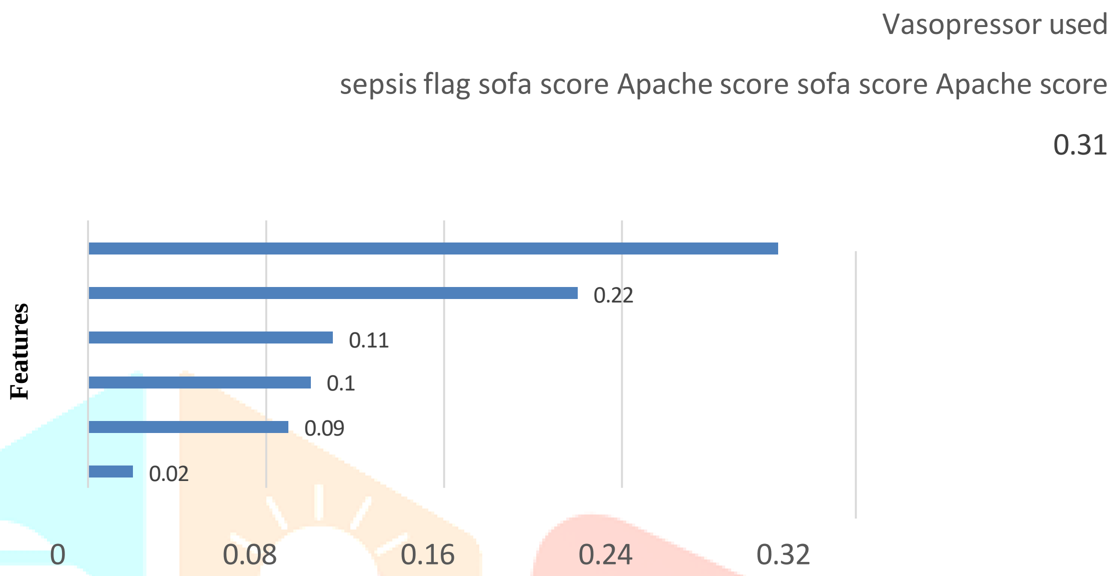
The performances of the Random Forest, XGBoost, and Ensemble models on a test dataset containing the data of 3,000 ICU patients are summarized in Table 4 below. The AUC-ROC of the Random Forest model is 0.6233, its accuracy is 77.3% and its F1-score is 0.0116. The AUC-ROC of the XGBoost model is 0.5969, its accuracy is 77.5% and its F1-score is 0.0260. The ensemble model (Random Forest + XGBoost) performed better than the single models by getting the highest AUC-ROC (0.6283) and the highest F1-score (0.0286) while having a nearly similar accuracy (77.4%). From this result, it can-be seen that using an ensemble approach can enhance the prediction of high-risk ICU patients. Even though the accuracies are-nearly similar for all three models, the ensemble results are more dependable. Hence, the ensemble model was chosen for the suggested ICU monitoring system.

Table 3: ML model performance metrics

Model	AUC-ROC	Accuracy (%)	F1-Score	Dataset
Random Forest (10% wt)	0.6233	77.3%	0.0116	Test Set (3,000)
XGBoost (10% wt.)	0.5969	77.5%	0.0260	Test Set (3,000)
Ensemble(RF+XGB)	0.6283	77.4%	0.0286	Test Set (3,000)

5.3 XGBoost Feature Importance Analysis

In Figure 2 XGBoost importance values based on gain have been-calculated for-the features and displayed on the dashboard for live administration. The high importance of vasopressors use and the sepsis flag compared to continuous vital signs shows that the model is effectively using these decision flags as stand-in measures for disease severity. Vasopressors are used when there is poor blood flow, which often means the patient needs catecholamines — a sign of septic shock, which has an ICU mortality rate between 25% and 45%. The sepsis flag is the most important single predictor of ICU mortality worldwide. SOFA and APACHE scores provide extra help in distinguishing between different levels of illness severity when used with these flags.



Feature Importance(Gain)

Figure 2: XGBoost Feature Importance Scores ICU Patient Monitoring Dashboard

The proposed Agentic AI-Based Autonomous ICU Monitoring Dashboard is illustrated in Figure 3. The monitoring dashboard will provide real-time monitoring of patient vital signs such as Heart rate, Blood Pressure, SpO_2, Body temperature, Respiration rate, Glucose level, and Lactate levels. The patient's condition is predicted by an AI-based risk scoring system, which also gives SOFA and APACHE-II scores for the patients.

Figure 3: ICU Patient Monitoring Dashboard

5.4 Alert Dashboard

Figure 4 illustrates the Alert Dashboard in the suggested Agentic AI-Based Autonomous ICU Monitoring System. The dashboard allows for real-time monitoring and managing of patient alerts through visualization of the number of alerts and their priorities (High, Medium, and Low), and their statuses (Active/Acknowledged). The distribution of alert priorities and the number of alerts of different types, e.g. low SpO2 or high lactate, are additionally shown on the dashboard. Alert data, namely patient identification, description of alert, its value, threshold, and time of occurrence, are shown to facilitate the work of healthcare professionals in making decisions.

Figure 4: Alert Dashboard

5.3 Discussion

From the experiment results, it appears that Agentic AI can-help improve-how well ICU patients' risks are assessed by using a hybrid-algorithm that-combines machine learning models with clinical rules. Even though models like Random Forest and XGBoost had only moderate effectiveness because of the imbalance in the data, using the agentic approach helped improve the AUC-ROC metric. The study also found-those factors like the use of vasopressors, signs of sepsis, and the SOFA score were the most important features, as they show the risk of death or worsening condition, based on clinical evidence. Additionally, the tool made is very accurate in sending alerts, which helps reduce the stress and overload on medical staff from too many alarms. Overall, the created model gives accurate results.

VI. CONCLUSION AND FUTURE ENHANCEMENT

6.1 Conclusion

This project is a great example of how to design, build, and put into use an AI-powered system for automatically monitoring ICU patients. It uses a multi-agent approach to make the system work smoothly. The system has several parts: agents made with Crew AI that work together, models like XGBoost and Random Forest to predict patient deaths, scoring tools like SOFA and APACHE-II for assessing patient health, a Fast API backend for handling data, a Streamlit interface for displaying information, and an optional GPT-4o-mini tool for giving recommendations. The system has four main agents: one for collecting data, one for-monitoring-patients, one for predicting risks, and one for sending alerts.

6.2 Future Enhancement

An AI-based system that works-on its-own to monitor ICU patients can-also be improved by using IoT sensors to track patient health. More studies could look into using Explainable AI to make risk assessments clearer, Reinforcement Learning to help the system make-better decisions-over time, and Electronic Health Records to understand each patient's medical history better. Machine learning models should be tested using real ICU data like MIMIC-III or MIMICIV, and they should be retrained using features that come from time-series data.

References

[1] T. Iqbal, E. Curry, and I. Ullah, "Agentic AI System for Stress Monitoring: A Multi-agent Healthcare Crew," International Journal of Computational Intelligence Systems, vol. 19, no. 95, 2026. https://doi.org/10.1007/s44196-026-01215-0

[2] S. Hosseini and H. Seilani, "The role of agentic AI in shaping a smart future: A systematic review," Array, vol. 26, p. 100399, 2025. https://doi.org/10.1016/j.array.2025.100399

[3] M. Sudarshan, S. Shih, E. Yee, A. Yang, J. Zou, C. Chen, Q. Zhou, L. Chen, C. Singhal, and G. Shih, "Agentic LLM Workflows for Generating Patient-Friendly Medical Reports," arXiv:2408.01112, 2024.

[4] J.-Z. Xiang, Q.-Y. Wang, Z.-B. Fang, J. A. Esquivel, and Z.-X. Su, "A multi-modal deep learning approach for stress detection using physiological signals: integrating time and frequency domain features," Frontiers in Physiology, vol. 16, art. 1584299, 2025. https://doi.org/10.3389/fphys.2025.1584299

[5] Savage, T., et al. (2024). Diagnostic reasoning prompts reveal the potential for large language model interpretability in medicine. npj Digital Medicine.

[6] Yao, S., et al. (2022). ReAct: Synergizing reasoning and acting in language models. arXiv:2210.03629.

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

[8] Morgan, R. J., Williams, F., & Wright, M. M. (1997). An early warning scoring system for detecting developing critical illness. Clinical Intensive Care, 8(2), 100.

[9] Subbe, C. P., Kruger, M., Rutherford, P., & Gemmel, L. (2001). Validation of a modified early warning score in medical admissions. QJM, 94(10), 521–526.

[10] Knaus, W. A., Draper, E. A., Wagner, D. P., & Zimmerman, J. E. (1985). APACHE II: A severity of disease classification system. Critical Care Medicine, 13(10), 818–829.

[11] Vincent, J. L., et al. (1996). The SOFA score to describe organ dysfunction/failure. Intensive Care Medicine, 22(7), 707–710.

[12] Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: Synthetic minority over-sampling technique. Journal of Artificial Intelligence Research, 16, 321–357.