Explainable Multimodal AI for Early ICU Deterioration Prediction

Mrs. K. Gomathi Lakshmi[1], Abdul Raheem F[2], Akshaya Priya T[3], Balaruban R[4], Kamalesh S[5]
[1]*Assistant Professor, K. Ramakrishnan College of Technology, Trichy, Tamil Nadu, India*
[2,3,4,5]*Student, Department of Artificial Intelligence and Data Science, K. Ramakrishnan College of Technology, Trichy, Tamil Nadu, India*

***Abstract*—Detecting early signs of deterioration in patients within Intensive Care Units can be a complex task. Threshold-based alerting methods may struggle to account for the relationship between different clinical parameters. In this work we propose an Explainable Multimodal AI framework for prediction of early deterioration based on physiological measurements, clinical notes and patient mobility/positioning. We use Machine Learning, Deep Learning, Biomedical Language Models and Computer Vision to extract meaningful features from each modality. These extracted features are then fused using a Multimodal approach to produce risk scores. The Shapley Additive exPlanations (SHAP) method was used to determine which key factors were driving the predictions. An interactive dashboard is presented showing the risk levels, factors driving risk and trends in patient data. This framework will provide healthcare professionals with a tool to help detect high risk patients in ICU earlier. This framework should be seen as a clinical decision support tool and not a replacement for medical expertise.**

***Index Terms*—ICU deterioration prediction, multimodal AI, explainable AI, clinical decision support, deep learning, SHAP, biomedical language models, patient monitoring.**

I. INTRODUCTION

ICU is a special setting of hospitals where critical care for severely ill patients who need constant monitoring and interventions becomes an integral part of healthcare practices. ICU patients can demonstrate rapid changes in their physiology, such as respiratory insufficiency, cardiopulmonary instability, sepsis, etc. Identifying a patient deterioration early on is critical in terms of better outcomes of treatment. It is not easy, however, as patients' conditions can become unstable fast enough, and signs of a possible deterioration can be dispersed among various sources of information.

The conventional systems for ICU surveillance typically incorporate predefined values for physiological variables or scoring system like the National Early Warning Score (NEWS2). While the techniques are effective, there is a possibility that their use may overlook the relationship that exists among different physiological factors, as well as clinical observations. [1]

The Artificial Intelligence and Machine Learning techniques create an avenue for the analysis of vast amounts of clinical data to detect the patterns related to the deterioration process. Deep Learning methods allow one to discover the temporal relations based on continuous physiological measurements. [2]

Nevertheless, the key drawback of most AI-driven clinical prediction tools is their inability to provide clear explanations. It is crucial for healthcare specialists not only to be aware of whether a certain patient poses any risks but also to know why this prediction was made.

In order to cope with such issues, this project suggests creating the Explainable Multimodal AI tool that will incorporate physiological data, clinical records, and information about the patient's physical condition. The tool will provide an early deterioration score and give the explanations through SHAP and other interpretable AI methods.

This project will help to enhance the process of early risk prediction and improve the decision-making process.

II. PROBLEM STATEMENT

The current ICU Deterioration Prediction Systems rely either on specific physiological data, pre-set thresholds, or traditional scoring systems in clinical settings. In most cases, the above approaches do not

sufficiently consider the interaction between various sources of clinical data or detect early changes which precede critical conditions in a patient.

Additionally, many of the AI prediction models do not give enough explanation on how they arrived at the risk scores.

As a result, there is a need for an intelligent and explainable multimodal system to be developed which will help to integrate different ICU patients' data, detect early deterioration patterns, assess the risk and explain the prediction of the system. This problem will be solved in the presented system by integrating physiological time series data, clinical text and patient posture data.

III. OBJECTIVES

The main goal of the present project is to build an interpretable multimodal AI model for the early prediction of the deteriorating condition of ICU patients.

The particular goals are:

1. To gather and pre-process ICU patient data that include physiological data, medical notes, and patient's position.
2. To conduct time-series analysis of the patients' vital signs by applying machine learning and deep learning algorithms.
3. To derive information of clinical importance from unstructured medical notes using NLP.
4. To detect the patterns of patient's movements and posture using computer vision or motion analysis methods.
5. To combine the data coming from different modalities via multimodal feature fusion.
6. To predict the risk of deterioration among patients and to classify the patients under appropriate risk categories.
7. To make the predictions made by the system understandable through SHAP and feature attribution approaches.
8. To create an interactive dashboard that would display patient risk scores, their factors, and physiological trends.
9. To evaluate the proposed system using suitable classification and early prediction measures.

IV. LITERATURE SURVEY

A. Conventional ICU Deterioration Prediction Methods

The conventional methods for ICU deterioration prediction use physiological parameters and scores, such as NEWS2. The conventional methods use various physiological parameters like heart rate, blood pressure, breathing rate, oxygen saturation level, and temperature to identify which patients can benefit from medical attention.

The simplicity and clinical understanding are the main advantages of using this method; however, it may miss the non-linear dependencies between physiological parameters.

B. Machine Learning for ICU Deterioration Prediction

The machine learning approaches such as Logistic Regression, Random Forest, Support Vector Machine, and Gradient Boosting have been used for ICU patient monitoring.

The approach uses structured data to discover patterns which are related to the adverse outcomes. The deep learning approach includes LSTM and Temporal Fusion Transformer (TFT), which is capable of analyzing sequential physiological data.

Nevertheless, the models which rely only on the structured data cannot make use of the information available in the unstructured form.

C. Clinical Natural Language Processing

Clinical records provide a lot of information regarding the symptoms, diagnosis, management plans, and clinical observations of the patients. Biomedical language models like BioClinicalBERT process the medical texts and create contextual embeddings.

The contextual embeddings might give extra information that is not directly available through structured measurements.

Nevertheless, the clinical texts might have some missing values, abbreviations, delays in documenting, and inconsistency in predicting deterioration.

D. Multimodal Artificial Intelligence

Multimodal artificial intelligence combines data from various sources in order to provide a more complete prediction. [5], [6]

With regard to intensive care unit applications, multimodal methods may involve physiological

measurements, clinical notes, medical images, and activity data of patients. [5], [6]
A combined use of multiple modalities may contribute to a better patient representation. Yet, the difference in data format, sampling rates, missing values, and time synchronization poses significant technical problems. [5], [6]

E. Explainable Artificial Intelligence in Healthcare
Explainable AI methods seek to ensure that the predictions of machine learning models are understandable to the user.
SHAP is an explanation technique that attempts to quantify the importance of each input in making a prediction. It may be applied in determining the physiological variables or other clinical factors which most influence the ICU deterioration risk score.[7]
Explainability is especially crucial in healthcare because doctors need to judge whether the predictions have been made based on sound clinical grounds.

F. Research Gap
Current systems have shown the significance of physiological data collection, machine learning, clinical text analysis, and explainable AI in the health care industry.
There are still some issues with the above methods, including:
1) Inadequate use of structured physiological data and unstructured clinical texts.
2) Problematic inclusion of movement or posture of patients in predicting their deterioration.
3) Issues associated with alignment of diverse sources of data over time.
4) Inadequacy in explaining complicated multimodal prediction models.

The suggested research aims to overcome the above issues by developing a multimodal AI system that uses physiological time-series data, clinical texts, and movement or posture of patients in a prediction process.

V. PROPOSED SYSTEM

The proposed system makes use of a multimodal AI-based approach for predicting ICU patient deterioration through processing of various clinical data streams.
The proposed system comprises five primary modules, which are listed below.

1. Physiological Data Processing
Physiological data is obtained in structured format using monitoring devices in ICU. The following physiological variables are recorded:
1) Heart Rate (HR)
2) Blood Pressure (BP)
3) Oxygen Saturation (SpO_2)
4) Respiratory Rate (RR)
5) Body Temperature

2. Clinical Text Processing
Clinical notes are extracted from patients' data and processed using Natural Language Processing methods.
BioClinicalBERT is used for creating contextual embeddings based on medical text. The embeddings include clinical information such as symptoms, observations, and treatment-related data.[4]

3. Analysis of Movement and Posture
Patient's movements and postures are analyzed by computer vision or movement-based features.
Such patterns as decreased movements, posture change, or unusual movement patterns may be found, which will add more context to patient's state.
This module is optional, depending on whether there is enough data for it.

4. Multimodal Feature Fusion
The physiological, clinical, and movement features are fused together using a representation that is common to all.
The feature fusion approach uses the different information present in various modes to construct a representative patient feature set.

5. Risk Prediction and Interpretation
The fused representation is then input into a deterioration prediction model.
This model creates a risk score, which denotes the probability of deterioration within a certain prediction time period.
Features that contribute to the prediction are detected through SHAP. The dashboard shows the risk score, important contributing features, and the trends of the patient. [7]
The proposed system aims to help healthcare professionals in their decision making by providing early warning and interpretations.

VI. METHODOLOGY

A. Problem Formulation

Suppose that (X_v), (X_t), and (X_m) represent physiological time-series data, clinical text data, and movement or posture data, respectively.

The multimodal prediction model is expressed as follows:

$$\hat{y} = f(X_v, X_t, X_m) \quad (1)$$

Where:

X_v: Physiological features

X_t: Representation of clinical text data

X_m: Movement or posture features

f: Multimodal prediction model

$\hat{y}$: Estimated patient deterioration risk

The primary objective of this model is to estimate the likelihood of patient deterioration within a specified prediction horizon by combining information from all three data modalities.

$$P(Y = 1 \mid X_v, X_t, X_m) \quad (2)$$

Here, Y = 1 represents the occurrence of patient deterioration within the selected prediction horizon.

B. Datasets and Data Preprocessing

The task involves ICU patient data that includes physiological measurements, clinical documentation, and possibly movement/posture data where available.

Possible datasets could be:

1) MIMIC-IV: Clinical documentation and patient measurements for ICU.[8]
2) MIMIC-IV-Note: Clinical documentation for patient care.[9]
3) eICU Collaborative Research Database: Multi-center ICU clinical data.[10]

It needs to be determined separately whether there is any synchronized movement/posture data. If such a data is not available, then the first version of the project could use physiological and clinical text data only.

Data preprocessing includes:

1) Elimination of duplicate records.
2) Handling of missing physiological measurements.
3) Normalization of numeric features.
4) Temporal alignment of clinical measurements.
5) Cleaning and tokenizing clinical text.
6) Creating prediction windows and labeling of deterioration.

C. Extraction of Physiological Features

Measurements of physiology are ordered as time series.

An LSTM network could be used to train on dependency between subsequent measurements.[3]

Alternatively, a Temporal Fusion Transformer may be considered for multivariate time-series forecasting.

The physiological representation extracted is:

$$Z_v = f_v(X_v) \quad (3)$$

Where f_v is the feature extraction function for physiology.

D. Clinical Text Feature Extraction

Clinical notes are analyzed with BioClinicalBERT.[4]

Text encoder maps clinical descriptions to context-aware embedding.

$$Z_t = f_t(X_t) \quad (4)$$

The above-mentioned embeddings contain the extracted clinically valuable information from unstructured medical text.

E. Movement Feature Extraction

Data on movements and postures are obtained using computer vision/movement analysis methods.

Examples of possible features to be extracted may include the frequency of movements, postures, activities, etc.

$$Z_m = f_m(X_m) \quad (5)$$

Movement feature extraction is carried out when proper data is available.

F. Multimodal Feature Fusion

The features extracted from the input are fused via a feature fusion process.

One such fusion method based on concatenation is:

$$Z_f = \text{concat}(Z_v, Z_t, Z_m) \quad (6)$$

Where Z_f denotes the multimodal feature representation.

The fused features are then fed to the predictive model.

G. Probability of Deterioration Prediction

This is calculated using the following prediction model equation:

$$p = \sigma(WZ_f + b) \quad (7)$$

Where:

W: Weighting matrix for the model

b: Bias

σ: Sigmoid activation function

p: Predicted probability of deterioration

Risk probabilities may be classified as low, moderate, and high risk by applying threshold values chosen using training and validation data sets.

H. Explainability using SHAP

SHAP is employed for explaining how the contribution of different features helps predict results. SHAP values of a feature indicate the contribution relative to the baseline prediction.[7]

The system detects:

1) The physiological features that contribute towards risk.
2) Clinical text features that contribute towards predicting the risk.
3) Any movement features that contribute towards the risk if present.
4) Features that reduce the prediction of risk.

I. Model Evaluation

The model is assessed based on the relevant classification and prediction measures.

The model assessment consists of comparison with other models, including Logistic Regression, Random Forest, and standard clinical scoring systems.[1],[2]

In addition, the assessment of the performance of different data sources is done with respect to the full multimodal model.

VII. EXPERIMENTAL SETUP

A. Hardware and Software Requirements

Table I
Proposed Hardware and Software Configuration

Component	Proposed Configuration
Processor	Intel Core i5 / equivalent
RAM	8 GB minimum, 16 GB recommended
GPU	NVIDIA GPU, optional
Programming Language	Python
Deep Learning	PyTorch / TensorFlow
Data Processing	Pandas, NumPy
Machine Learning	Scikit-learn
NLP	Transformers, BioClinicalBERT
Explainability	SHAP
Visualization	Matplotlib, Plotly
Dashboard	Streamlit
Database	SQLite / MySQL

VIII. RESULTS AND EVALUATION

A. Performance Metrics

The proposed system will be evaluated using the following metrics:

Table II
Performance Evaluation Metrics

Metric	Purpose
Accuracy	Overall correct predictions
Precision	Correct positive predictions
Recall / Sensitivity	Ability to identify deteriorating patients
Specificity	Ability to identify non-deteriorating patients
F1-score	Balance between precision and recall
AUROC	Discrimination between positive and negative cases
AUPRC	Performance under class imbalance
Lead Time	Time between prediction and deterioration

Recall and AUPRC are particularly important because missed deterioration events can have serious clinical consequences.

B. Model Architectures

The following architectures of the models can be analyzed:

1) Physiological only model.
2) Clinical text only model.
3) Combined physiological and clinical text multimodal model.
4) Multimodal model with movement/posture features included.

C. Benchmarking

The proposed model can be benchmarked against:

1) NEWS2/any other baseline.[1]
2) Logistic Regression.
3) Random Forest.
4) LSTM based prediction from physiological data.
5) Multimodal model without explanations.

The goal is to see whether feature combination helps with better prediction and whether the explanations contain valuable clinical information.

D. Modality Comparison

The contribution of each modality will be evaluated by comparing individual and combined model configurations.

Table III
Modality Comparison of Model Configurations

Model	Physiological Data	Clinical Notes	Movement Data
Model 1	✓	—	—
Model 2	—	✓	—
Model 3	✓	✓	—
Model 4	✓	✓	✓

E. Explainability Evaluation

The explanations generated by SHAP will be evaluated based on:

1. Identification of clinically relevant contributing features.
2. Consistency of explanations across similar patient cases.
3. Interpretability of patient-specific predictions.
4. Ability to identify important physiological trends.
5. Usefulness of explanations for clinical decision support.

F. Expected Experimental Outcome

The proposed system is expected to demonstrate whether combining physiological measurements and clinical text provides improved deterioration prediction compared with individual modalities.

The explainability component is expected to provide interpretable information about the major features contributing to patient risk.

The final results will be reported using quantitative metrics, comparison tables, and visualizations after completing model training and evaluation.

IX. DISCUSSION

This multimodal ICU deterioration prediction system will overcome some of the disadvantages associated with traditional ICU monitoring systems by combining various forms of clinical data.

Analysis of physiological time series gives the ability for the model to recognize temporal patterns as well as changes in the patient's condition. The analysis of clinical text will allow for information that cannot be captured by measurement data to be used. The movement and posture analysis can give further information about the patient's condition depending on availability of the right data.

Multimodal fusion will allow all these pieces of information to be combined together.

Explanation is an essential aspect of the proposed framework. The explanations based on SHAP can help the health care workers gain insight into the reasons for a particular prediction based on the features rather than only on a numeric risk score.

Nonetheless, the applicability of the system will depend on the quality of the clinical data, synchronization of modalities, accuracy of model predictions, and validation of the explanations.

The proposed framework has to be thoroughly examined prior to its clinical application.

X. LIMITATIONS

The following are some of the limitations that should be considered with respect to the suggested project.

Data availability: There might be no synchronized data of physiological parameters, clinical texts, and movement in a single public ICU dataset.

Incomplete data: ICU data might have missing values and incomplete measures along with unequal intervals of data sampling and inconsistent documentation of clinical parameters.

Class imbalance: The events of deterioration in patients' health status might be rare in comparison with stable patients.

Temporal alignment: It is difficult to align different modalities since they might have been collected at different time intervals.

Computational limitations: Biomedical language models and multimodal deep learning models might demand high computational power.

Explainability limitations: SHAP helps with explaining feature attributions, but it cannot prove causality between features and deterioration.

Clinical generalizability: One dataset might not generalize to another hospital or population.

Clinical validation: Clinical validation of the system is required before its application in practice.

XI. CONCLUSION

The present work suggests an Explainable Multimodal AI for predicting ICU patient deterioration in its early stage, where several sources such as physiological time series data, clinical notes and patient posture or activity are used together. The suggested architecture

employs deep learning methods as well as Biomedical NLP (Natural Language Processing) to obtain valuable representations from various clinical data sources. Then, the fusion module fuses these representations to obtain the estimation of the patient deterioration probability.

The SHAP approach is used for making this solution explainable and able to show what were the most important features that were used for prediction.

The goal of this project is to be able to detect those ICU patients who require additional attention in order to avoid negative outcomes.

Further research will cover multimodal integration improvement, real-world dataset validation, prediction lead time estimation and the usage of uncertainty-aware predictions.

The current project shows how multimodal AI and explainable machine learning can be combined to create more transparent and intelligent solutions for ICU patient monitoring.

REFERENCES

[1] B. Sadanandan, "Multimodal deep learning for early prediction of patient deterioration in the ICU: Integrating time-series EHR data with clinical notes," *arXiv preprint* arXiv:2603.14719, 2026.

[2] M. Momenzadeh and A. Oshaghi, "Lightweight LLM summarization with LSTM for early ICU deterioration prediction," *Scientific Reports*, 2026.

[3] S. Jagadeesan, P. K. Panda, R. V. Sandarage, and E. C. Tsai, "Development and internal validation of an EHR machine learning model for clinical deterioration in ICU-linked inpatients," *Scientific Reports*, 2026.

[4] B. Jandoubi and M. A. Akhloufi, "MultiCardioNet: A multimodal deep learning model for early cardiovascular deterioration prediction in ICU patients," *Bioengineering*, vol. 13, no. 9, Art. no. 993, 2026.

[5] Bakumenko, D. H. Smith, and J. Hoelscher, "Early-horizon multimodal ICU mortality prediction without retraining," in *Proc. 11th Mach. Learn. Healthcare Conf.*, PMLR, vol. 340, pp. 214–240, 2026.

[6] "Comparison of multimodal deep learning approaches for predicting clinical deterioration in ward patients: Observational cohort study," *Journal of Medical Internet Research*, vol. 27, Art. no. e75340, 2025.

[7] "An interpretable machine learning algorithm enables dynamic 48-hour mortality prediction during an ICU stay," 2025.

[8] E. W. Johnson *et al*., "MIMIC-IV, a freely accessible electronic health record dataset," *Scientific Data*, vol. 10, Art. no. 1, 2023.

[9] E. W. Johnson *et al*., "MIMIC-IV-Note: Deidentified free-text clinical notes," PhysioNet, 2022.

[10] Soenksen *et al*., "Integrated multimodal artificial intelligence framework for healthcare applications," *npj Digital Medicine*, 2022.

[11] "ICU mortality prediction via multimodal learning on MIMIC-IV clinical data," *Bioinformatics Advances*, 2026.

[12] J. M. López Alcaraz *et al*., "A multimodal deep learning framework for predicting ICU deterioration: Integrating ECG waveforms with clinical data and clinician benchmarking," *arXiv preprint* arXiv:2601.06645, 2026.