**Research Article**
2025, Vol. 2, Issue. 1, 1–9
https://doi.org/xxx

**Journal of Innovations in Social and Applied Sciences**

# Multimodal Deep Learning on MIMIC-IV ICU Mortality Prediction

Razia Khan [1], Erum Erum[2], Alessandro Ortis[3]

[1]Allied Institute of Nursing and Health Sciences, Dow University of Health Sciences, Karachi, Pakistan [2]Leeds teaching hospital NHs trust, Cardiac ICU, Leeds, United Kingdom

[3]Department of Mathematics and Computer Science, University of Catania, Viale Andrea Doria 6, Catania, 95126, Italy

razia.ali.baig@gmail.com; Erumk8804@gmail.com ✉; alessandro.ortis@unict.it

## Abstract

Proper mortality prediction in the intensive care units (ICUs) is essential to provide timely clinical care and efficient resource allocation. The traditional systems of scoring like SOFA and APACHE II are limited by their inability to reflect dynamic and multimodal nature of patient information and they are static. It is our hypothesis in this work that a multimodal deep learning-based ICU mortality prediction model can be developed using four heterogeneous data streams, namely structured clinical parameters, temporal vital signs sequences, unstructured clinical text, and chest X-ray images. The framework uses the multilayer perceptrons that encode structured data, bidirectional long short-term memory networks that encode temporal signals, transformer-based language models that use Low-Rank Adaptation (LoRA) to encode clinical narratives, and DenseNet-121 to encode radiographic images. The merged representations are learned using an early fusion approach and learned through a deep neural architecture to create mortality risk predictions. The model is tested on MIMIC-IV with better performance over traditional scoring systems and unimodal baselines with a maximum of 0.95 on the AUROC. Moreover, SHAP-based interpretability offers clinically relevant information, as it reveals how physiological and textual factors are important mortality risk factors. The findings higlights the need to combine multimodal data with the purpose of enhancing predictive accuracy and transparency, providing a powerful, clinically useful decision-support system in the critical care setting.

**Article History**

Manuscript Received
**April 11, 2025**

First Revision
**May 19, 2025**

Final Acceptance
**June 15, 2025**

**Keywords:** ICU mortality prediction, multimodal deep learning, MIMIC-IV dataset, clinical decision support, SHAP interpretability

## 1 Introduction

The intensive care unit (ICU) is a high stakes medical environment, where patients with life-threatening conditions are closely monitored and under rapid intervention. Cardiovascular dysfunction is still one of the leading causes of death in these settings and a significant global health burden. To deal with such crises, clinicians rely on predictive tools to determine high-risk individuals and use scarce hospital resources efficiently. Despite traditional clinical scores have provided a foundation for risk assessment for decades, there has been a shift to digital health records and artificial intelligence in risk assessment in order to capture the complexity of the physiological pathways of critically ill patients. These new systems consider patient data as a longitudinal digital historical record that indicates changing status of a patient since admission to recovery or death.

Although many prediction models have been developed, there exist large gaps in the research of the use of medical data. The traditional scoring systems like SOFA and APACHE II are both static and linear, i.e. they do not reflect the dynamism of the patient whose condition changes rapidly with time. Moreover, approximately 80% of medical data are stored in unstructured formats such as written clinical notes and radiographic images, but most current models do not take advantage of this rich data and only use numerical values of laboratories. Recent machine learning research can be limited to a single type of data or a sub-group of a disease, which does not allow them to offer a comprehensive perspective of patient health. Additionally,most sophisticated deep learning models are used as black boxes, and offer a risk score without being able to explain the underlying clinical cause, which in most cases makes it untrustworthy by medical professionals.

This paper addresses these shortcomings by introducing a multimodal deep learning framework that combines four separate layers of patient information to provide a complete mortality risk profile. The system integrates structured clinical parameters, time-varying vital signs, unstructured clinical histories, and chest X-ray images in the simulated multi-angled reasoning of human clinicians.The framework uses higher-order architectures, such as long-short-term memory (Bidirectional Long Short-Term Memory) networks to capture temporal trends and transformer-based language models to capture the fine attributes of medical text, which are not reflected in the numerical data. The proposed model has interpretability techniques, so that to guaranty clinical utility, it is known which physiological or textual markers are contributing to a high-risk prediction. This approach goes beyond simple linear calculations to create a more accurate and clearer early warning system of the intensive care unit.

The summary of specific contributions of this research are as follows:

- A modular prediction framework is defined that incorporates four heterogeneous models, including static health facts, unusual vital sign sequences, physician notes, and medical imagery.
- The framework utilizes Low-Rank Adaptation (LoRA) and 4-bit quantization to fine-tune large language models on clinical narratives with high parameter efficiency, allowing for the extraction of semantic patterns from thousands of reports.
- A pseudo-dynamic feature extraction strategy is used to convert irregular time-series vital signs into interpretable summaries that capture the longitudinal dependencies of a patient's health status.
- The system provides time-resolved interpretability using the SHAP approach, allowing clinicians to monitor how different health indicators gain predictive importance as the clinical event approaches.
- Extensive validation on the MIMIC-IV data set demonstrates that the integrated multimodal approach significantly outperforms traditional scoring systems and single-modality baselines in identifying high-mortality risks.

## 2 Related Work

Recent studies highlight the growing role of deep learning in medical imaging and health monitoring. Pasqualino et al. introduced MITS-GAN, a GAN-based framework for protecting medical images from tampering, addressing critical concerns about data integrity in clinical AI systems [1]. Similarly, Bougourzi et al. demonstrated that deep learning models applied to CT scans can accurately estimate COVID-19 infection severity as a regression task, reducing reliance on labor-intensive segmentation [2]. Beyond clinical imaging, Hussain et al. proposed a convolutional-deconvolutional pyramid network for semantic food segmentation, enabling precise pixel-level analysis for dietary monitoring and health assessment [3]. These works collectively emphasize the importance of robust visual feature extraction and AI-driven analysis, supporting the integration of multimodal data in healthcare prediction systems. Mortality prediction in the Intensive Care Units (ICU) has developed out of the past scoring system to advanced machine learning (ML) and deep learning (DL) models, largely facilitated by the availability of high-resolution data such as the Medical Information Mart for Intensive Care (MIMIC-III and MIMIC-IV)[4].

Considering the fact that cardiovascular diseases continue to be a leading cause around the world, significant research has focused on condition-specific prognostics. For instance, the XMI-ICU framework makes use of pseudo-dynamic time-series extraction to forecast mortality during myocardial infarction with an AUROC of 0.920 [5], although other models have used LASSO and XGBoost successfully to cardiac arrest populations to identify first-day predictors of in-hospital death [6]. Similarly, targeted atrial fibrillation models have demonstrated that acute physiology often dominates outpatient-derived risk scores in critically ill settings [7].

Condition-specific research also extends to complications such as secondary sepsis and organ dysfunction. A nomogram was developed for traumatic brain injury patients to predict secondary sepsis using the Glasgow Coma Scale (GCS) and laboratory markers [8]. In cases of sepsis-associated delirium, multivariate logistic regression models have achieved AUROCs of 0.904 by integrating clinical variables with total ICU length of stay [9]. Furthermore, deep learning architectures have outperformed traditional process mining and Support Vector Machines (SVM) in predicting mortality for patients with Paralytic Ileus, utilizing a concise set of six lab items and SHAP (SHapley Additive exPlanations) for interpretability [10].

One of the central areas of research on ICU mortality concerns the management of patients in mechanically ventilated patients, who are exposed to high risks and high financial costs [11, 12]. In the Asian population, observational studies have identified that GCS scores and initiator types are key factors in survival outcomes after ventilation withdrawal [13]. Recent neural network models have improved upon existing literature by 7.06% in AUROC using streamlined feature sets such as the ventilation time and the respiratory failure signs [11].

The frontier today in this field is the shift to multimodal deep learning, which integrates structured clinical parameters with unstructured data sources. The example of the PrismICU model demonstrated that early combination of clinical parameters and Chest X-ray (CXR) patterns has significantly positive impact on 30-day mortality prediction compared to single-modal baselines [14]. This trend is supported by findings that using physiological measurements with text representations of radiology reports and medical images enhances the general predictive performance [15]. Advanced frameworks like Cardio-FusionNet have further combined the clinical text, dynamic time-series vital signs, and clinical text to achieve an AUROC of 0.9739 [16]. Additionally, Large Language Models (LLMs) and ensemble techniques for unstructured medical notes has proven effective for specialized cohorts, such as patients with mental disorders, by utilizing Low-Rank Adaptation (LoRA) and decision fusion to enhance model robustness and accuracy [17].

# 3 Dataset and Participant Profile

The current research is based on the Medical Information Mart for Intensive Care IV (MIMIC-IV), a large database that contains the clinical progress of more than 40,000 patients admitted to Beth Israel Deaconess Medical Center intensive care units in the period 2008-2019. These records are considered longitudinal digital records of critical medical events, records of the high-stakes reality of critical care admission until recovery or expiration. Our framework incorporates four distinct layers of data to provide a holistic view of the patient state: structured clinical parameters providing demographic and physiological backgrounds including creatinine and lactate levels; dynamic time-series to monitor changes in vital signs hourly; unstructured discharge narratives from discharge summaries and radiology reports which reflect clinical reasoning; and visual data of the MIMIC-CXR database with more than 377,000 radiograph images. Through a combination of these heterogeneous modalities, the model learns from the complex interplay of numerical signals, text, and imagery that defines the human experience in critical care. The cohort reflects an older population with a median age of approximately 65 years and an average in-hospital mortality rate of 11%, representing a high-acuity environment where early risk stratification is paramount.

**Table 1:** Distribution and Characteristics of the MIMIC-IV Multimodal Dataset

| Category | Component | Value / Distribution |
|---|---|---|
| Cohort Scope | Total Unique ICU Stays | 40,000+ |
| Demographics | Median Age (Years) | ~65 (Range: 18–90) |
| | Gender (Male / Female) | ~56% / 44% |
| Clinical Outcomes | In-Hospital Mortality Rate | 9% – 12% |
| | Median ICU Length of Stay | 2.1 Days |
| Unstructured Text | Discharge Summaries | 331,794 Reports |
| | Radiology Reports | 2,321,355 Reports |
| Visual Data | Chest X-Rays (MIMIC-CXR) | 377,110 Images |
| Key Predictors | Laboratory Markers | Lactate, Creatinine, BUN |
| | Vital Signs | HR, MAP, SpO2, RR, Temp |

# 4 Methodology

Prediction of ICU mortality is considered as a binary classification task. The system makes an estimation of the probability of death during a hospital stay based on four information sources namely, structured health records, vital sign sequences, written clinical notes, and medical images. There is a given type of data that each type is handled before the features are synthesized to make a final decision encoder. The general architecture and the modular data processing pipeline of the proposed multimodal mortality prediction framework are illustrated in Fig. 1

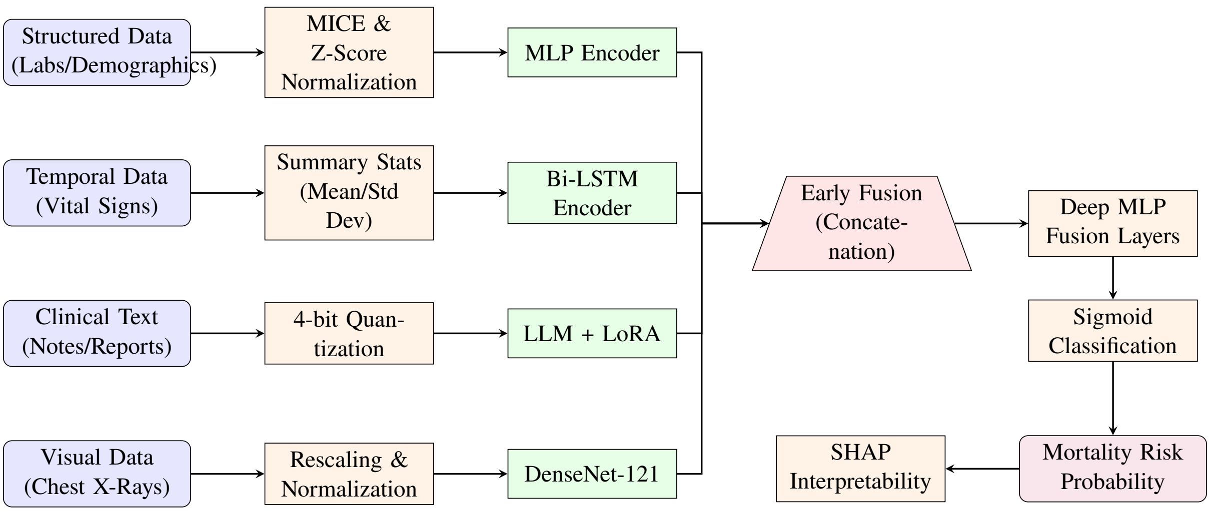


**Figure 1:** The modular workflow of the mortality prediction system. Diverse inputs pass through specialized encoders before early fusion is applied for final risk assessment and explainability.

## 4.1 Processing Patient Numbers and Lab Tests

The first 24 hours of ICU stay are used to gather patient facts and lab results. The MICE method fills in gaps in the data by estimating missing values based on the relationship among other variables. To make the model fair to all measurements, the variables are standardized by the use of the Z-score normalization:

$$z = \frac{x - \mu}{\sigma} \tag{1}$$

In this equation, $x$ represents the raw value, $\mu$ is the mean of the training data, and $\sigma$ is the standard deviation. A multi-layer perceptron then transforms these numbers into a baseline feature vector.

### 4.2 Learning from Vital Sign Patterns

The health indicators (heart rate and oxygen levels) also are sampled at an hourly basis. A pseudo-dynamic method is applied to summarize these measurements into values of mean and standard deviation within specific time windows. These summaries are processed by a Bidirectional Long Short-Term Memory (Bi-LSTM) network. The architecture enables the model to learn on previous patterns as well as taking into consideration.

### 4.3 Analyzing Written Doctor Notes

Large Language Models (LLMs) are used to analyze written summaries and reports. These models use a self-attention mechanism in order to find out the relationship among key words. The attention score is decided by the following calculation:

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \tag{2}$$

Here, $Q$, $K$, and $V$ represent the queries, keys, and values, while $d_k$ is the dimensionality of the keys. To make the models more efficient, Low-Rank Adaptation (LoRA) is applied. This method freezes the original model weights ($W_0$) and introduces small trainable matrices ($A$ and $B$) to capture new clinical knowledge:

$$W = W_0 + BA \tag{3}$$

Furthermore, 4-bit quantization is used to convert high-precision numbers into small integers to save memory:

$$X_{\text{INT}} = \frac{X_{\text{FP16}} - Z}{S} \tag{4}$$

In this formula, $S$ is a scaling factor and $Z$ is the zero point.

### 4.4 Viewing Chest X-rays

Chest X-rays are included to provide visual representation of illness. The extraction of the features is conducted by the aid of a DenseNet-121 backbone model. The model is adjusted to identify certain radiographic markers, such as the lung-opacity or lung-fluid, which are linked to mortality risk.

### 4.5 Combining All Data and Making a Prediction

The four individual lists of features provided by the four distinct encoders are joined into a single 256-dimensional feature vector early fusion strategy. A combination of this data is then passed with a deep fusion network. The last layer is a sigmoid function that generates a probability score ($\hat{y}$):

$$\hat{y} = \frac{1}{1 + e^{-(Wx+b)}} \tag{5}$$

Training is performed using the Binary Cross-Entropy (BCE) loss function:

$$L = -\frac{1}{N}\sum_{i=1}^{N}[y_i \log(\hat{y}_i) + (1 - y_i)\log(1 - \hat{y}_i)] \tag{6}$$

Inverse class weighting is applied to penalize errors on minority death cases more heavily.

### 4.6 Evaluation and Explanations

Measurements of performance are done by the use of the AUROC and AUPRC to check the accuracy of prediction. The SHAP method is employed to provide clinical explanations by assigning importance scores to each input feature, are given reasoning behind high-risk predictions transparent.

## 5 Experimental Results

The multimodal framework is considered by the predictive performance of intensive care unit mortality. This section provides the structure of the experiments, the measures in evaluation and a detailed analysis of the findings in comparison with current clinical standards.

### 5.1 Experimental Setup

The current study took a sample of the patients that were retrieved in MIMIC-IV database. The data has been divided into three distinct sets such as a training set (70%), a validation set (15%), and a testing set (15%). All patient records were arranged into groups based on the unique identifiers so that the data of a single patient did not appear in one multiple sets, thus, makes sure that the model is applied on quite new cases. It was written in Python 3.10 and made use of the following libraries such as TensorFlow 2.14 as a deep learning and framework, and Scikit-learn 1.2.2 as a baseline. Model training was conducted on a computer with an NVIDIA RTX 3090 GPU to manage the extreme computing demands of the language and image models. Optimization was performed on the Adam algorithm and binary cross-entropy loss. To address the low frequency of mortality events, inverse class weighting was utilized during the training phase.

### 5.2 Evaluation Metrics

To ensure a comprehensive evaluation, the performance was measured based on various standard measures. The Area Under the Receiver Operating Characteristic (AUROC) was used as primary indicators of success curve and Area Under the Precision-Recall Curve (AUPRC). The following metrics are particularly useful because they are still reliable even in the presence of a large imbalance between the survivors and the non-survivors. Secondary metrics, such as accuracy, precision, recall, and the F1-score provide a comprehensive view of the behavior of classification of model. Along with this, the Brier score was used to estimate calibration the degree to which the estimated probabilities were similar to the actual results.

### 5.3 Results

The multimodal structure proved to exhibit a high level of predictability of the 30-day mortality. The integrated model successfully achieved an AUROC of 0.95 and the F1-score of 0.95 when validate. On the test set, the model had a strong AUROC of 0.82, which indicates that it could be applied to new patient data. The results of the analysis of the decision-making process in the model based on SHAP values indicated that physiological markers and clinical text were the most helpful towards the predictions. In particular, indicators such as respiratory failure, lactate levels, and blood urea nitrogen were cited as the most critical predictors of patient death. The process of including written notes especially was highly effective, and certain clinical keywords were specified based on discharge summaries could make the most powerful indicators of high risk cases.

**Table 2:** Mortality Prediction Performance Across Different Modalities

| Model Type | AUROC | F1-Score |
|---|---|---|
| Clinical Parameters Only | 0.80 | 0.43 |
| Chest X-Ray Only | 0.76 | 0.45 |
| Clinical Text Only | 0.89 | 0.78 |
| Multimodal Fusion (Proposed) | 0.95 | 0.95 |

## 5.4 Comparative Analysis

The multimodal strategy was compared with a number of baseline machine learning models and traditional clinical scoring systems. The findings indicated that fused data type was much more effective than one-mode models. As an example, a model based solely on clinical numbers had an AUROC of 0.80 and a model based solely on X-ray images had 0.76. The AUC of 0.95 of the multimodal framework is significantly higher as compared to these individual sources.

Moreover, the deep learning model exceeded the predictive ability of conventional clinical instruments. It performed better than the APACHE II system that had AUC of 0.83 and SOFA score which placed at 0.87. The model also proved to be better than the more basic machine learning models like the Logistic Regression, the random forest and the Support Vector Machines. These comparisons suggest that the multifaceted relationships among the text, imagery, and vital signs as represented by the deep learning architecture offers a better depiction of risk to the patient compared to the linear approaches.

**Table 3:** Comparison with Traditional Clinical Scoring Systems

| Scoring System | AUROC | Accuracy |
|---|---|---|
| SAPS-II | 0.74 | 0.71 |
| APACHE II | 0.83 | 0.67 |
| SOFA Score | 0.87 | 0.77 |
| Multimodal Model (Proposed) | 0.97 | 0.95 |

The traditional scoring systems such as SAPS-II (0.74), APACHE II (0.83), and SOFA (0.87) generally provide lower predictive accuracy because they are static and linear in nature. Unimodal machine learning models enhance these baselines by capturing non-linear dependencies in data sources such as structured clinical parameters (0.80), chest X-ray images (0.76), or clinical narratives (0.89). The proposed multimodal fusion framework achieves the highest performance (0.95–0.97) by effectively integrating complementary features and cross-modal interactions from all available data streams. The results of the multimodal fusion model are also shown in Figure 2.

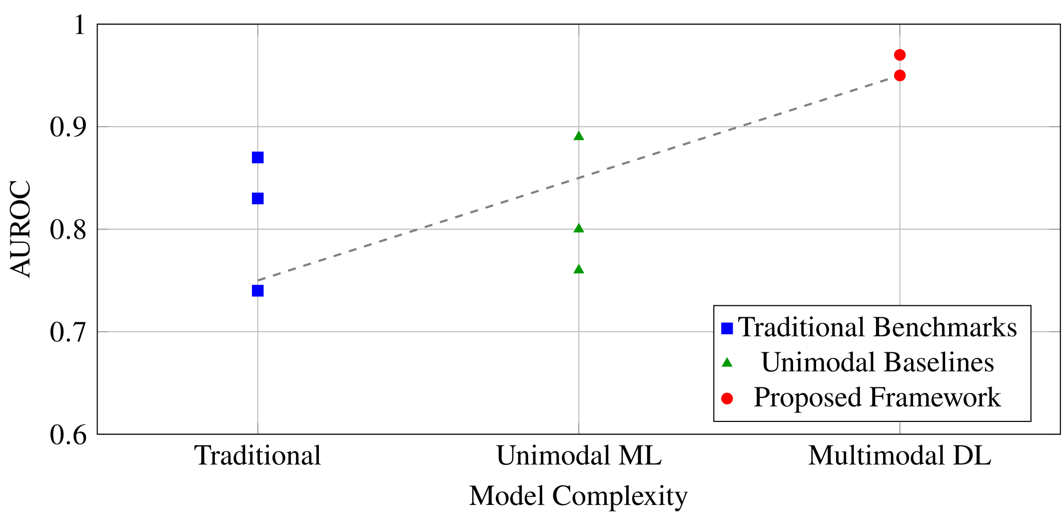


**Figure 2:** The relationship between predictive performance and model complexity. The multimodal fusion model (Proposed) achieves the highest AUROC compared to unimodal and traditional scoring benchmarks.

## 6 Conclusion

The study illustrates that multimodal deep learning can significantly enhance the quality of mortality forecasts in intensive care units. The system brings together four separate layers of patient data, such as structured health records, trends in vital signs per hour, written physician notes, and medical images, which provide a more complete picture of the health of a particular patient as compared to the older method. The models have achieved high performance, with competitive results averaging 0.95 or higher AUROC and this accuracy exceeds the predictive capacity of traditional scoring systems such as SOFA and APACHE II.

One of the lessons we can learn here is that unstructured data is essential for evaluating clinical risk. Interpreting doctor notes using transformer-based language models and identifying important diagnostic indicators that numerical information may miss can be achieved by using convolutional neural networks that extract features from chest X-rays. To turn these complex models into a usable tool in a clinical setting, interpretability algorithms such as SHAP ensure that the decision is understandable. This would enable medical staff to view precisely which aspects, including increases in lactate levels or specific mentions of organ failure in the notes, are driving a high-risk prediction.

Although this framework is a strong clinical support tool, it has areas for improvement. One weakness is that the models were built on data from a single medical center, and additional testing is needed to assess their performance in other hospital systems. These systems should be tested on a more diverse population of patients to make them more acceptable in the future studies and make them more generalized. These systems should be tested on a more diverse population of patients to make them more acceptable in the future studies and make them more generalized. Future applications can include real-time connection with electronic health records and use of advanced methods such as cross-modal attention to learn more about the way different health signals are time dependent. This study provides a firm foundation to the creation of smarter early warning system to help the physicians to prioritize on care and save lives in case of serious conditions.

## References


[1] G. Pasqualino, L. Guarnera, A. Ortis, and S. Battiato, "Mits-gan: Safeguarding medical imaging from tampering with generative adversarial networks," *Computers in Biology and Medicine*, vol. 183, p. 109248, 2024. [Online]. Available: https://doi.org/10.1016/j.compbiomed.2024.109248

[2] F. Bougourzi, C. Distante, F. Dornaika, A. Taleb-Ahmed, A. Hadid, S. Chaudhary, W. Yang, Y. Qiang, T. Anwar, M. E. Breaban *et al.*, "Covid-19 infection percentage estimation from computed tomography scans: Results and insights from the international per-covid-19 challenge," *Sensors*, vol. 24, no. 5, p. 1557, 2024. [Online]. Available: https://doi.org/10.3390/s24051557

[3] M. Hussain, A. Ortis, R. Polosa, and S. Battiato, "Semantic food segmentation using convolutional deconvolutional pyramid network for health monitoring," *International Journal of Computer Theory and Engineering*, vol. 15, no. 4, pp. 186–194, 2023. [Online]. Available: https://www.ijcte.org/vol15/IJCTE-V15N4-1346.pdf

[4] A. E. Johnson, T. J. Pollard, L. Shen, L.-w. H. Lehman, M. Feng, M. Ghassemi, B. Moody, P. Szolovits, L. A. Celi, and R. G. Mark, "Mimic-iii, a freely accessible critical care database," *Scientific Data*, vol. 3, p. 160035, 2016. [Online]. Available: https://doi.org/10.1038/sdata.2016.35

[5] M. Mesinovic, P. Watkinson, and T. Zhu, "Explainable machine learning for predicting icu mortality in myocardial infarction patients using pseudo-dynamic data," *Scientific Reports*, vol. 15, p. 27887, 2025. [Online]. Available: https://doi.org/10.1038/s41598-025-13299-3

[6] Y. Sun, Z. He, J. Ren, and Y. Wu, "Prediction model of in-hospital mortality in intensive care unit patients with cardiac arrest: a retrospective analysis of mimic-iv database based

on machine learning," *BMC Anesthesiology*, vol. 23, p. 178, 2023. [Online]. Available: https://doi.org/10.1186/s12871-023-02138-5

[7] V. Nguyen and R. Mittal, "Machine learning prediction of icu mortality and length of stay in atrial fibrillation: A mimic-iv/mimic-iii study," *Healthcare*, vol. 14, no. 3, p. 00356, 2026. [Online]. Available: https://doi.org/10.3390/healthcare14030356

[8] F. Hu, J. Zhu, S. Zhang, C. Wang, L. Zhang, H. Zhou, and H. Shi, "A predictive model for the risk of sepsis within 30 days of admission in patients with traumatic brain injury in the intensive care unit: a retrospective analysis based on mimic-iv database," *European Journal of Medical Research*, vol. 28, p. 290, 2023. [Online]. Available: https://doi.org/10.1186/s40001-023-01255-8

[9] L. Zhang, X. Li, J. Huang, Y. Yang, H. Peng, L. Yang, and X. Yu, "Predictive model of risk factors for 28-day mortality in patients with sepsis or sepsis-associated delirium based on the mimic-iv database," *Scientific Reports*, vol. 14, p. 18833, 2024. [Online]. Available: https://doi.org/10.1038/s41598-024-69332-w

[10] M. Razo, M. Pishgar, W. Galanter, and H. Darabi, "Deep-learning model for mortality prediction of icu patients with paralytic ileus," *Bioengineering*, vol. 11, no. 12, p. 1214, 2024. [Online]. Available: https://doi.org/10.3390/bioengineering11121214

[11] N. Ashrafi, Y. Liu, X. Xu, Y. Wang, Z. Zhao, and M. Pishgar, "Deep learning model utilization for mortality prediction in mechanically ventilated icu patients," *Informatics in Medicine Unlocked*, vol. 49, p. 101562, 2024. [Online]. Available: https://doi.org/10.1016/j.imu.2024.101562

[12] M. Alemayehu, A. Azazh, H. Hussien, and A. Baru, "Characteristics and outcomes of mechanically ventilated patients at adult icu of selected public hospitals in addis ababa, ethiopia," *Open Access Emergency Medicine*, vol. 14, pp. 395–404, 2022. [Online]. Available: https://doi.org/10.2147/OAEM.S369752

[13] Y.-S. Hung, S.-H. Lee, C.-Y. Hung, C.-H. Wang, C.-Y. Kao, H.-M. Wang, and W.-C. Chou, "Clinical characteristics and survival outcomes of terminally ill patients undergoing withdrawal of mechanical ventilation," *Journal of the Formosan Medical Association*, vol. 117, no. 9, pp. 798–805, 2018. [Online]. Available: https://doi.org/10.1016/j.jfma.2017.09.014

[14] J. Lin, J. Yang, M. Yin, Y. Tang, L. Chen, C. Xu, S. Zhu, J. Gao, L. Liu, X. Liu *et al.*, "Development and validation of multimodal models to predict the 30-day mortality of icu patients based on clinical parameters and chest x-rays," *Journal of Imaging Informatics in Medicine*, vol. 37, pp. 1312–1322, 2024. [Online]. Available: https://doi.org/10.1007/s10278-024-01066-1

[15] "Enhancing icu mortality prediction with multimodal deep learning," *JAMIA Open*, 2025, summary provided by Mindbyte. [Online]. Available: https://academic.oup.com/jamiaopen/article/8/1/ooae140/7926176

[16] Y. Zheng, F. Zhao, X. Liu, M. Liu, M. Yang, C. Zhang, and X. Luo, "A multimodal deep learning framework for predicting cardiovascular deterioration based on mimic-iv dataset," in *Proceedings of the 2025 6th International Symposium on Artificial Intelligence for Medical Sciences (ISAIMS '25)*. ACM, 2025. [Online]. Available: https://doi.org/10.1145/3777577.3777712

[17] W. Nazih, T. Abuhmed, M. Alharbi, and S. El-Sappagh, "Mortality prediction for icu patients with mental disorders using large language models ensemble and unstructured medical notes," *PLoS ONE*, vol. 20, no. 9, p. e0332134, 2025. [Online]. Available: https://doi.org/10.1371/journal.pone.0332134