Asynchronous Multi-Modal Learning for Dynamic Risk Monitoring of Acute Respiratory Distress Syndrome in Intensive Care Units

Yidan Feng[1], Bohan Zhang[1], Sen Deng[1], Zhanli Hu[2], and Jing Qin[1](✉)

[1] Centre for Smart Health, School of Nursing, The Hong Kong Polytechnic University, Hong Kong, China
harry.qin@polyu.edu.hk

[2] Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences, Shenzhen, China

Abstract. Acute Respiratory Distress Syndrome (ARDS) is a critical adverse event with high modality rates, yet its recognition in ICU settings is often delayed. Clinicians face significant challenges in integrating asynchronous, multi-modal data streams with misaligned temporal resolutions during rapid deterioration. This work introduces a deep learning model for continuous ARDS risk monitoring, designed to dynamically integrate diverse ICU data sources and generate timely, actionable predictions of ARDS onset. We extend existing settings for ARDS detection from static, single-modality prediction to continuous, multi-modal monitoring that aligns with clinical workflows. To address the inherent complexities of this task, we propose tailored solutions for hierarchical fusion across irregular sampling points, heterogeneous data modalities, and sequential predictions, while ensuring robust training against dynamic, irregular inputs and severe class imbalance. Validated on 1,985 MIMIC-IV patients, our model demonstrates superior performance, achieving average AUROC scores of 0.94, 0.91, and 0.87 across 6, 24, and 48 hours pre-onset, respectively, outperforming previous models (AUROC 0.78–0.85). Furthermore, the model quantifies emergency level to aid in resource prioritization and identifies high-risk patients with peak relative risk reaching 25, demonstrating exceptional discrimination between cohorts. The code is publicly released at https://github.com/YidFeng/MICCAI25-ARDS-Risk-Prediction.

Keywords: ARDS · Multi-Modal Learning · Risk Prediction

1 Introduction

Risk prediction for critical adverse events, which are characterized by rapid onset and high mortality rates, is a crucial application of AI in clinical practice [16,13]. Among these events, Acute Respiratory Distress Syndrome (ARDS) represents a particularly urgent challenge in intensive care units (ICUs) especially after the COVID-19 pandemic [17], affecting 10–15% of ICU patients [2] with mortality rates alarmingly high (near 40%) [2,11]. A key contributor to this is the

frequent under-recognition of ARDS, which delays life-saving interventions [2]. This diagnostic gap is exacerbated by the overwhelming volume and complexity of ICU data, which often exceed the capacity of clinicians to synthesize actionable insights [3]. Therefore, effective exploitation of this wealth of continuous, multi-modal data streams to generate timely and accurate predictions of future ARDS risk could hold significant clinical value.

The early detection of ARDS relies on complementary clinical modalities, yet existing approaches remain fragmented in their integration. Existing studies have employed traditional statistical and machine learning methods on structured electronic health record (EHR) data, focusing on vital signs (VS) and laboratory results (LAB) within fixed observation windows. For instance, logistic regression models achieved AUROCs of 0.78–0.81 by analyzing EHR-derived features from the first 24–48 hours of ICU admission, though these methods largely ignored temporal dynamics and imaging evidence [21,19,8]. While chest X-rays (CXRs) provide critical diagnostic specificity by revealing lung-specific pathologies like bilateral opacities, their diagnostic value is constrained by sparse acquisition, interpretive subjectivity, and delayed radiographic manifestations of lung injury [15,1]. Recent advances in deep learning have demonstrated the power of CNNs to detect subtle CXR patterns predictive of ARDS (AUROC: 0.82–0.85) [1,14], yet these image-only models overlook the dynamic physiological context captured by high-frequency VS and irregular LAB measurements. In [12], an attempt was made to combine CXR with EHR data through late fusion strategies. However, this approach suffers from oversimplified temporal representations by reducing time-series data to summary statistics, and static prediction paradigms that rely on fixed time windows. This neglects both the asynchronous sampling patterns of ICU modalities and the clinical necessity for dynamic risk reassessment as patient conditions evolve. Consequently, current systems fail to address two critical gaps: 1) insufficient modeling of cross-modal dependencies across irregularly sampled data streams, and 2) rigid one-time predictions that misalign with the continuous monitoring workflow of ICUs, where clinicians require iteratively updated risk assessments to guide time-sensitive interventions.

To bridge these gaps, we redefine ARDS risk prediction as a continuous multi-modal monitoring task aligned with clinical workflows. Our framework advances conventional settings through two critical aspects: 1) dynamic reassessment of both risk score and time-to-onset urgency at regular intervals, and 2) comprehensive utilization of asynchronous multi-modal data streams (multi-view CXRs and EHR data) acquired throughout the patient's stay. By synchronizing risk updates with ICU workflow cycles, the system is expected to provide guidance during actionable intervention windows. However, this clinical alignment introduces algorithmic challenges in modeling cross-modal dependencies across irregular temporal resolutions and maintaining computational efficiency over extended monitoring periods. To address these, we devise a transformer-based architecture with two tailored modifications: The Staged Temporal-Modal (STM) fusion module decouples temporal and cross-modal interactions, explicitly contextualizes sparse imaging findings against continuous physiological trends, relaxing the

synchronicity assumption of existing multi-modal learning algorithms [6,4,20]; The Progressive Context Memory (PCM) enhances efficient sequential prediction of ICU data streams by incrementally integrating patient data into compact memory states, which allows adaptive integration of historical insights while minimizing redundant computations. Complementing these architectural advances, our training strategies, late batching and balanced sampling, support effective and robust learning under complex dynamic inputs and severe class imbalance. In conclusion, our contributions are as follows:

1. We advance ARDS detection by shifting from static, single-modality prediction to continuous multi-modal monitoring, contextualizing sparse imaging cues within physiological trends for earlier, actionable alerts.
2. Accordingly, we contribute tailored solutions to address the unique challenges in our setting, with empirical validation against alternatives confirming the effectiveness of our hierarchical fusion architecture and training strategies.
3. Our model demonstrates exceptional performance across $24h/48h$ (AUROC 0.91/0.87) pre-onset windows, surpassing prior methods (AUROC 0.78-0.85) in a more challenging setting. Notably, in the critical <6h pre-onset window, our system identifies 91% of ARDS cases with AUROC 0.94. Moreover, patient risk stratification by our model identifies high-risk cohorts that exhibit over 20-fold elevated ARDS incidence, while our emergency quantification (MAE< 0.6 level) directly supports prioritization of resource allocation.

2 Methods

2.1 Problem Setting

Asynchronous Multi-Modal Data. We focus on three modalities: chest X-rays (CXRs), vital signs (VS) and laboratory results (LAB), with VS and LAB defined based on a predefined parameter set $\mathcal{P}$. Unlike CXRs, which can be fully acquired at a single time point, the observation times for the various parameters in VS and LAB may differ. Each modality exhibits distinct sampling properties. Vital signs are typically measured at high frequency with regular intervals, and the parameters remain relatively consistent across samples. In contrast, laboratory tests are sampled at irregular intervals and cover a broad range of parameters. Chest X-rays, compared to both VS and LAB, are sampled much more sparsely and irregularly. These heterogeneous data are organized using a unified time coordinate, which formulates the asynchronous multi-modal input at prediction time t_i over period T as $I_{t_i} = \{D_t^{\cdot}, t_i - T < t < t_i\}$, where each data point is represented as D_t^m for $m = \text{CXR}$, or $D_t^{m_p}, p \in \mathcal{P}$ for $m \in \{\text{VS}, \text{LAB}\}$, and t represents the observation time during the patient's ICU stay.

Risk Monitoring. Instead of diagnosing ARDS from a fixed data window, our model dynamically monitors ARDS risk through frequent predictions at regular intervals. Specifically, for an ICU patient, the model begins predicting at time t_0 and continues to make predictions at each subsequent time point $t_i = t_0 + i \times 6$ hours, where $i = 1, 2, \ldots$. At each prediction time t_i, the

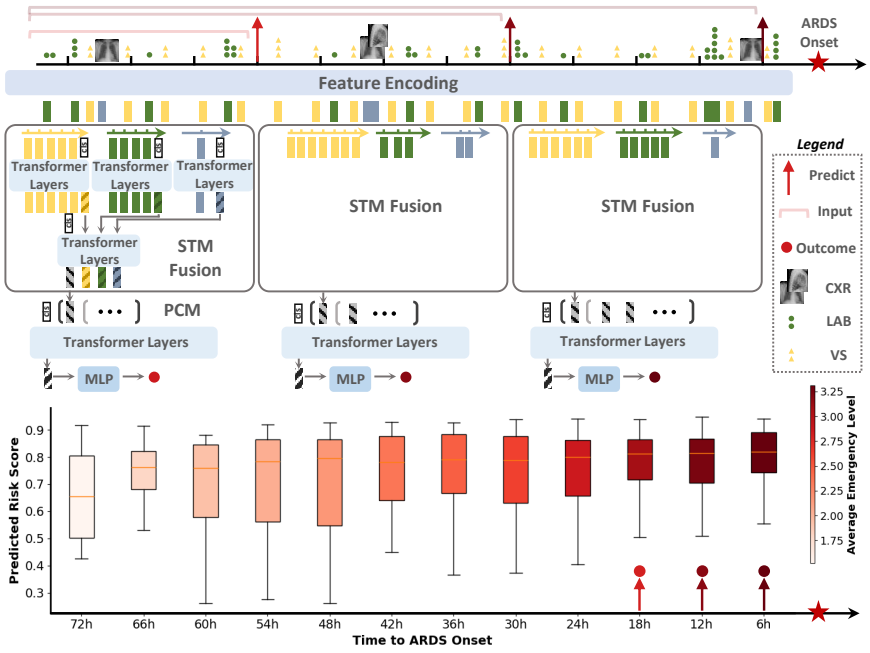

Fig. 1. Top: Network architecture for risk monitoring from asynchronous modalities (STM for Staged Temporal-Modal fusion and PCM for Progressive Context Memory, Transformer Layers consist of standard self-attention and feedforward network [18]). Bottom: Intuitive demonstration of monitoring results from ARDS patients (ARDS risk score and associated emergency level indicated in color).

model processes asynchronous multi-modal data I_{t_i} over a period T, where $T = \textbf{min}(72 \text{ hours}, t_i - \text{admission time})$. The model then outputs both the ARDS risk score ($r \in (0, 1)$), indicating the likelihood of ARDS, and the emergency level ($e \in \{1, 2, 3, 4\}$), which indicates the time to onset, providing guidance on appropriate interventions and resource allocation. The emergency level 4 represents the highest urgency, with ARDS onset expected within 12 hours, and level 3 for 12-24 hours, level 2 for 24-48 hours, level 1 for longer than 48 hours.

2.2 Network Architecture

Our network architecture is composed of three cohesive stages: 1) feature encoding, which converts raw data $\{D_t\}$ into unified latent features $\{f_t^m\}$; 2) feature fusion, which models the dependencies across asynchronous modalities and sequential predictions, and produces a summary feature $\hat{f}_i$; 3) the task head, which is a simple MLP that predicts the risk score and emergency level from $\hat{f}_i$.

Feature Encoding. Given irregularly sampled multi-modal inputs $\{D_t\}$, we encode each modality into unified latent features $\{f_t^m\}$ through specialized processing. For CXRs, we extract spatial features using ResNet-34 [7]. For Tabular

data (VS/LAB), we employ an adaptive sliding window algorithm to construct temporally cohesive parameter groups. Following [5], we embed windowed data through linear projection $\mathbf{W}_{\text{cont}} \in \mathbb{R}^{d\times 1}$ for continuous parameters, and embedding table $\mathbf{E} \in \mathbb{R}^{d\times|\mathcal{C}|}$ for categorical values, where $\mathcal{C}$ is the number of categories. Each parameter becomes a $d-$dimensional token, with attention masks suppressing missing values in Transformer layers. The CLS token aggregates cross-parameter interactions into feature $f_{\tilde{t}}^m$.

Feature Fusion. Effective modeling of interactions across asynchronous modalities and sequential predictions is the key challenge in our task. To address this, we design a hierarchical fusion framework leveraging Transformer architectures [18,4], enhanced by two tailored designs: Staged Temporal-Modal Fusion (STM Fusion) for efficient temporal-modal integration and Progressive Context Memory (PCM) for sequential prediction optimization. STM Fusion processes asynchronous multi-modal features $\{f_{\tilde{t}}^m\}$ through decoupled temporal and modal phases. First, modality-specific temporal Transformer layers compress each modality's irregular sequence into a summary vector, where modality-aware positional encodings are employed to handle temporal asynchrony. For each feature $f_{\tilde{t}}^m$, the modality type m is encoded as categorical variable C_m, and the normalized relative timestamp is calculated as $R_{\tilde{t}} = (\tilde{t} - t_i)/\Delta T)$ within each observation window $[t_i, t_i + \Delta T]$. These two components were fused through a multilayer perceptron (MLP) with ReLU activation, generating a joint positional encoding $P = MLP(R_{\tilde{t}}, C_m) \in R^d$. Cross-modal interaction is then modeled by fusing all modality summaries using multi-modal transformer layers, with modality embeddings separating different modality inputs. For prediction tasks at $\{t_0, ..., t_i\}$, PCM explicitly models their dependencies through incremental encoding and memory-augmented attention. Specifically, at t_i, only new observations since t_{i-1} are processed to obtain $\Delta\hat{f}_i$. A compact memory bank $\mathcal{M} = \{\hat{f}_0, ..., \Delta\hat{f}_{i-1}\}$ dynamically stores historical context, and the current feature updates this bank by $\mathcal{M} \leftarrow \mathcal{M} \cup \Delta\hat{f}_i$. The model then applies a transformer layer over $\mathcal{M}$ to adaptively learn attention weights for context aggregation. This memory mechanism enables real-time inference while preserving long-range dependencies, which is crucial for clinical deployment.

2.3 Training

Late batching. To handle irregular sequences while preventing information leakage, we implement a late batching protocol: For each patient, we sequentially process all their prediction time points $\{t_0, ..., t_L\}$ through the network. Gradient accumulation is performed over B (batch size) patients before back propagation. This approach ensures natural handling of variable-length sequences within a patient and stable gradient statistics through accumulation across patients.

Balanced sampling The balanced sampler mitigates class imbalance by uniformly selecting half the batch size from positive and negative class indices, concatenating and shuffling them to ensure equal representation per batch. Implemented as a dynamic generator, it preserves dataset integrity while optimizing

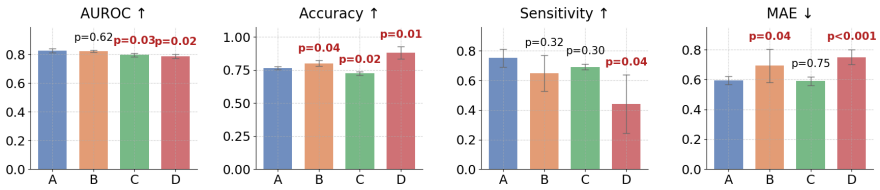

Fig. 2. Comparison results for different fusion architectures, with significant differences highlighted. **A** for ours, **B** w/o PCM, **C** w/o STM, and **D** w/o both.

gradient updates for this imbalanced learning task.

Training details. We use cross-entropy loss and MSE loss with equal contributions for risk score and emergency level, respectively. The optimization employs AdamW with an initial learning rate of 3e-4, scheduled by cosine annealing with linear warm-up. We train the model with a batch size of 16 for a total of 20,000 steps, and select the final model based on its balanced performance across both tasks on the validation set. All experiments were conducted using PyTorch on a single NVIDIA RTX 3090 Ti GPU with 24GB of memory.

2.4 Dataset and Pre-Processing

Our data was extracted from two publicly available databases 1) MIMIC-IV [10], containing EHR data from 53,130 ICU admissions between 2008 and 2019; 2) MIMIC-CXR [9], a linked dataset of 377,110 de-identified chest radiographs paired with free-text radiology reports acquired from 2011 to 2016. The initial exclusion criteria include transfer cases, multiple stays, age under 18, empty CXR record. Following [21,19], ARDS cases were identified based on diagnosis records and the Berlin criteria, with further exclusion of patients lacking clear onset timing or exhibiting onset within 6 hours of ICU admission. This yielded 469 ARDS cases among 9,921 eligible patients. All CXR images, regardless of view type, were preprocessed through intensity normalization, resizing, center-cropping to 224×224 pixels, and augmented using random horizontal flips and affine transformations, VS and LAB results were normalized to standard ranges [6]. The cohort was partitioned into training (70%), validation (10%), and test (20%) sets via stratified sampling to preserve outcome distribution, with final results reported on the test subset of 1,985 patients.

3 Results and Analysis

3.1 Design Choices

Fusion types. We validate our fusion architecture by: (1) STM fusion vs. interleaved fusion directly on long-concatenation of asynchronous multi-modal tokens, and (2) PCM-augmented attention vs. a naive baseline that indiscriminately processes all historical data. Performance is measured using ACC (Accuracy), SEN

Table 1. Effectiveness of training methods .

Training Method						
Late Batching	Balanced Sampling	Slope↑	Volatility↓	AUROC% ↑	ACC% ↑	SEN% ↑
✗	✗	5.48e-6	0.089	70.62	95.45	0
✓	✗	2.06e-5	0.078	78.32	95.45	0
✗	✓	3.76e-5	0.070	75.52	72.21	59.12
✓	✓	7.41e-4	0.046	82.69	76.55	75.20

Table 2. Effectiveness of using asynchronous modalities.

Modalities							
CXR	EHR	AUROC% ↑	ACC% ↑	SEN% ↑	SPE% ↑	MAE↓	MSE↓
✓	✗	73.71±1.73	79.92±1.17	45.38±2.34	82.15±1.18	1.12±0.03	1.85±0.11
✗	✓	79.03±1.19	78.04±0.96	67.03±1.80	79.13±0.98	1.26±0.04	2.42±0.12
✓	✓	83.26±0.96	76.43±0.29	73.70±0.20	76.53±0.08	0.58±0.01	0.70±0.02

(Sensitivity) and AUROC, as well as MAE for emergency level, averaged across all sequential prediction points with 3-fold cross-validation. Threshold is consistently fixed as 0.5. As shown in Fig. 2, STM contributes significantly to AUROC, while PCM contributes significantly to emergency level prediction and tends to generate robust results with small deviation, likely due to its context modeling. STM and PCM together ensures the sensitivity of ARDS patients, which is crucial for risk prediction models. Without effective fusion strategies, baseline **D** exhibits highest ACC but markedly lower SEN under severe class imbalance.

Effectiveness of training methods. The necessity of our tailored training methods is validated through ablation studies using absolute loss slope, volatility (relative mean absolute change), and diagnostic measures. Without balanced sampling, the negative class dominates, compressing predictions toward zero with minimal variance. Late batching alone improves AUROC by +8% by reducing excessive padding and masking for irregular data. Together, they ensure stable and effective training with largely improved diagnostic measures (Tab. 1).

Effectiveness of using asynchronous modalities. From Tab. 2, EHR surpasses CXR in risk prediction due to temporal sensitivity to early deterioration, while CXR marginally edges emergency prediction for positive cases via post-symptom pathological signatures. Their fusion has shown significant complementary synergy in this challenging task.

3.2 Performance Evaluation of Dynamic Risk Monitoring

Diagnostic performance of risk score predictions. Fig. 1 provides an intuitive demonstration of the monitoring results. Here, we further provide assessments on temporal prediction efficacy by ROC curves of risk scores across 6-hour windows within 24 hours preceding ARDS onset (Fig. 3(c-f)). For each window, predictions were derived from randomly sampled time points (negative cases: random windows; positive cases: pre-onset windows) over 10 experiments. AU-

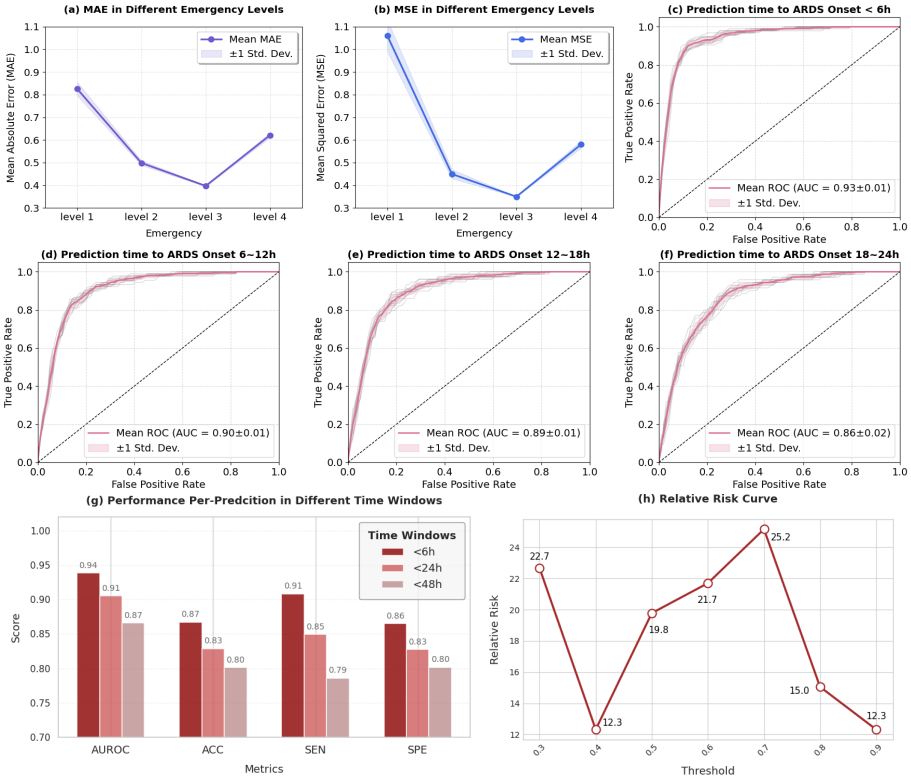

Fig. 3. Comprehensive evaluation of dynamic risk monitoring.

ROC decreased gradually from 0.93 ($< 6h$) to 0.86 (18-24h) with lower clinical urgency. Aggregated performance across broader intervals (Fig. 3(g)) achieved AUROC of 0.91 ($< 24h$) and 0.87 ($< 48h$) via 10 repeated samplings of all eligible prediction points, significantly outperforming prior studies (AUROC: 0.78-0.85).

Fidelity of emergency level predictions. The model's predicted emergency levels escalate progressively as ARDS onset approaches (Fig. 1). Stratification by true urgency levels (Fig. 3a-b) reveals that Level 1 ($> 48h$) predictions maintain clinically acceptable error margins (< 1 level) avoiding resource misallocation, while Levels 2-3 exhibit significantly improved accuracy yet cluster conservatively in mid-range values (predicted: 1.5–3.3 vs. true: 1–4; Fig. 1). This systematic under-prediction of extremes necessitates recalibration or loss function refinement to enhance discriminative capacity at critical urgency thresholds.

Patient risk stratification. Patient-level analyses compared disease incidence between between high-risk (any threshold-exceeding) and low-risk groups using relative risk (RR). In Fig. 3 (h), RR values consistently exceeded 12 in thresholds (0.3–0.9), demonstrating robustness in risk stratification. The RR score dropped at threhold of 0.4, which indicates a transition of domination from

low-risk group to high-risk group. Peak discriminatory performance occurred at a threshold of 0.7 (RR= 25.16), with stable RRs (19.78–25.16) observed in the moderate threshold range (0.5–0.7), effectively identifying high-risk patients, enabling efficient resource allocation.

4 Conclusion

This paper advances ARDS detection settings to continuous future risk monitoring leveraging asynchronous multi-modal ICU data. We propose tailored solutions for hierarchical fusion and effective training, validated through ablation studies. Notably, our results demonstrate markedly higher AUROCs compared to existing work, with potential clinical value for early ARDS detection and resource prioritization supported by emergency indication and patient risk stratification.

Acknowledgments. The study was funded by an Innovation and Technology Fund under Hong Kong Innovation and Technology Commission (No. ITS/202/23); a Shenzhen-Hong Kong-Macao Science and Technology Plan Project (Category C Project) under Shenzhen Municipal Science and Technology Innovation Commission (No. SGDX2023082 1092359002)

Disclosure of Interests. Author Yidan FENG has received research grants from Innovation and Technology Fund under Hong Kong Innovation and Technology Commission (No. ITS/202/23), and Shenzhen-Hong Kong-Macao Science and Technology Plan Project (Category C Project) under Shenzhen Municipal Science and Technology Innovation Commission (No. SGDX20230821092359002).

References

1. Arora, M., Davis, C.M., Gowda, N.R., Foster, D.G., Mondal, A., Coopersmith, C.M., Kamaleswaran, R.: Uncertainty-aware convolutional neural network for identifying bilateral opacities on chest x-rays: A tool to aid diagnosis of acute respiratory distress syndrome. Bioengineering **10**(8), 946 (2023)
2. Bellani, G., Laffey, J.G., Pham, T., Fan, E., Brochard, L., Esteban, A., Gattinoni, L., Van Haren, F., Larsson, A., McAuley, D.F., et al.: Epidemiology, patterns of care, and mortality for patients with acute respiratory distress syndrome in intensive care units in 50 countries. Jama **315**(8), 788–800 (2016)
3. Clark, B.J., Moss, M.: The acute respiratory distress syndrome: Dialing in the evidence? JAMA **315**(8), 759–761 (2016)
4. Feng, Y., Gao, B., Deng, S., Qiu, A., Qin, J.: Unified multi-modal learning for any modality combinations in alzheimer's disease diagnosis. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 487–497. Springer (2024)
5. Gorishniy, Y., Rubachev, I., Khrulkov, V., Babenko, A.: Revisiting deep learning models for tabular data. Advances in Neural Information Processing Systems **34**, 18932–18943 (2021)

6. Hayat, N., Geras, K.J., Shamout, F.E.: Medfuse: Multi-modal fusion with clinical time-series data and chest x-ray images. In: Machine Learning for Healthcare Conference. pp. 479–503. PMLR (2022)
7. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 770–778 (2016)
8. Jiang, Z., Liu, L., Du, L., Lv, S., Liang, F., Luo, Y., Wang, C., Shen, Q.: Machine learning for the early prediction of acute respiratory distress syndrome (ards) in patients with sepsis in the icu based on clinical data. Heliyon **10**(6) (2024)
9. Johnson, A., Pollard, T., Mark, R., Berkowitz, S., Horng, S.: Mimic-cxr database. PhysioNet10 **13026**, C2JT1Q (2024)
10. Johnson, A.E., Bulgarelli, L., Shen, L., Gayles, A., Shammout, A., Horng, S., Pollard, T.J., Hao, S., Moody, B., Gow, B., et al.: Mimic-iv, a freely accessible electronic health record dataset. Scientific data **10**(1), 1 (2023)
11. Meyer, N.J., Gattinoni, L., Calfee, C.S.: Acute respiratory distress syndrome. The Lancet **398**(10300), 622–637 (2021)
12. Pai, K.C., Chao, W.C., Huang, Y.L., Sheu, R.K., Chen, L.C., Wang, M.S., Lin, S.H., Yu, Y.Y., Wu, C.L., Chan, M.C.: Artificial intelligence–aided diagnosis model for acute respiratory distress syndrome combining clinical data and chest radiographs. Digital Health **8**, 20552076221120317 (2022)
13. Placido, D., Yuan, B., Hjaltelin, J.X., Zheng, C., Haue, A.D., Chmura, P.J., Yuan, C., Kim, J., Umeton, R., Antell, G., et al.: A deep learning algorithm to predict risk of pancreatic cancer from disease trajectories. Nature medicine **29**(5), 1113–1122 (2023)
14. Reamaroon, N., Sjoding, M.W., Gryak, J., Athey, B.D., Najarian, K., Derksen, H.: Automated detection of acute respiratory distress syndrome from chest x-rays using directionality measure and deep learning features. Computers in biology and medicine **134**, 104463 (2021)
15. Sjoding, M.W., Taylor, D., Motyka, J., Lee, E., Claar, D., McSparron, J.I., Ansari, S., Kerlin, M.P., Reilly, J.P., Shashaty, M.G., et al.: Deep learning to detect acute respiratory distress syndrome on chest radiographs: a retrospective study with external validation. The Lancet Digital Health **3**(6), e340–e348 (2021)
16. Tomašev, N., Glorot, X., Rae, J.W., Zielinski, M., Askham, H., Saraiva, A., Mottram, A., Meyer, C., Ravuri, S., Protsyuk, I., et al.: A clinically applicable approach to continuous prediction of future acute kidney injury. Nature **572**(7767), 116–119 (2019)
17. Tran, T.K., Tran, M.C., Joseph, A., Phan, P.A., Grau, V., Farmery, A.D.: A systematic review of machine learning models for management, prediction and classification of ards. Respiratory Research **25**(1), 232 (2024)
18. Vaswani, A.: Attention is all you need. Advances in Neural Information Processing Systems (2017)
19. Xu, C., Zheng, L., Jiang, Y., Jin, L.: A prediction model for predicting the risk of acute respiratory distress syndrome in sepsis patients: a retrospective cohort study. BMC Pulmonary Medicine **23**(1), 78 (2023)
20. Yao, W., Liu, C., Yin, K., Cheung, W.K., Qin, J.: Addressing asynchronicity in clinical multimodal fusion via individualized chest x-ray generation. In: The Thirty-eighth Annual Conference on Neural Information Processing Systems (2024), https://openreview.net/forum?id=uCvdw0IOuU
21. Zeiberg, D., Prahlad, T., Nallamothu, B.K., Iwashyna, T.J., Wiens, J., Sjoding, M.W.: Machine learning for patient risk stratification for acute respiratory distress syndrome. PloS one **14**(3), e0214465 (2019)