



This MICCAI paper is the Open Access version, provided by the MICCAI Society. It is identical to the accepted version, except for the format and this watermark; the final published version is available on SpringerLink.

MELON: Multimodal Mixture-of-Experts with Spectral-Temporal Fusion for Long-Term Mobility Estimation in Critical Care

Jiaqing Zhang¹, Miguel Contreras², Jessica Sena², Andrea Davidson³, Yuanfang Ren³, Ziyuan Guan³, Tezcan Ozrazgat-Baslanti³, Tyler J. Loftus⁴, Subhash Nerella², Azra Bihorac³, and Parisa Rashidi²

¹ Department of Electrical and Computer Engineering, University of Florida, Gainesville, FL, USA

² J. Crayton Pruitt Family Department of Biomedical Engineering, University of Florida, Gainesville, FL, USA

³ Department of Medicine, University of Florida, Gainesville, FL, USA

⁴ Department of Surgery, University of Florida, Gainesville, FL, USA

Abstract. Patient mobility monitoring in intensive care is critical for ensuring timely interventions and improving clinical outcomes. While accelerometry-based sensor data are widely adopted in training artificial intelligence models to estimate patient mobility, existing approaches face two key limitations highlighted in clinical practice: (1) modeling the long-term accelerometer data is challenging due to the high dimensionality, variability, and noise, and (2) the absence of efficient and robust methods for long-term mobility assessment. To overcome these challenges, we introduce MELON, a novel multimodal framework designed to predict 12-hour mobility status in the critical care setting. MELON leverages the power of a dual-branch network architecture, combining the strengths of spectrogram-based visual representations and sequential accelerometer statistical features. MELON effectively captures global and fine-grained mobility patterns by integrating a pre-trained image encoder for rich frequency-domain feature extraction and a Mixture-of-Experts encoder for sequence modeling. We trained and evaluated the MELON model on the multimodal dataset of 126 patients recruited from nine Intensive Care Units. Experiments showed that MELON outperforms conventional approaches for 12-hour mobility status estimation with an overall area under the receiver operating characteristic curve (AU-ROC) of 0.82 (95% confidence interval 0.78-0.86). Notably, our experiments also revealed that accelerometer data collected from the wrist provides robust predictive performance compared with data from the ankle, suggesting a single-sensor solution that can reduce patient burden and lower deployment costs. Project repository: <https://github.com/iheallab/MELON>.

Keywords: Accelerometer · Multimodal · Spectrogram · Mobility · Intensive Care Unit · Mixture of Expert · Patient Monitoring.

1 Introduction

Patients in the Intensive Care Unit (ICU) often experience prolonged immobility, increasing the risk of developing ICU-acquired weakness (ICU-AW) and cognitive impairment, such as delirium [3]. Frequent patient mobility assessments are pivotal for facilitating timely interventions and mitigating adverse outcomes [13,2,19]. Several functional mobility tools have been developed and implemented in clinical practice, such as the ICU Mobility Scale [5] and the Johns Hopkins Highest Level of Mobility Scale [6]. However, these tools require manual administration by ICU staff, which can be time-consuming and prone to documentation errors. Additionally, manual assessments often lack the granularity needed to capture subtle changes in mobility over time. Autonomous sensor-based approaches offer a promising alternative, enabling continuous, objective, and high-resolution patient mobility monitoring while reducing the workload on healthcare providers.

Accelerometry-based sensors are a commonly used option for capturing physical activity levels, posture changes, and mobility trends over time in a non-invasive manner [12,21]. However, directly modeling raw time-series accelerometer data presents several challenges due to the inherent complexity of the signals. The data consists of long sequences, making it computationally demanding to process and analyze. Additionally, its non-stationary nature, where statistical properties change over time, complicates pattern recognition and model generalization. Traditional approaches that rely on statistical features extracted from raw accelerometry data often fail to capture the nuances of mobility fully. To address this, recent studies have explored converting sequential sensor data into image representations, leveraging the hierarchical feature extraction capabilities of pre-trained vision models [23,22,1]. For instance, recurrence plots—a method transforming temporal dynamics into 2D texture patterns—have been successfully adopted in works such as [9] and [8], achieving state-of-the-art performance in activity recognition tasks. Alternatively, frequency-domain representations like spectrograms generated via Short-Time Fourier Transform (STFT) have also proven effective. Studies by [7] and [11] demonstrate that CNN-based models trained on spectrograms can robustly identify complex activities through learned spectral-temporal patterns. However, most of these models are developed and validated in general domains, and they do not address the specific needs of the ICU setups, which exhibit different activity patterns [14] and long-term assessment.

In this study, we propose MELON (Multimodal mixture-of-Experts with spectral-temporal fusion for Long-term mObility estimatioN), a novel dual-branch multimodal architecture for patient mobility assessment, Fig 1. The proposed architecture integrates two complementary input modalities: statistical accelerometer features sequence and spectrogram representation images of raw accelerometry data. To effectively process these modalities, we employ a pre-trained mixture-of-experts (Time-MoE) to model statistical feature sequences and an image encoder to extract patterns from frequency domain representations. By combining these modalities, we leverage their complementarity, statistical features offer

structured and fine-grained variations, whereas spectrograms enhance the ability to detect global temporal dependency, resulting in a more comprehensive representation of patient mobility. To the best of our knowledge, this is the first study that combines the spectrogram of accelerometer data and sequences of statistical features using a dual-branch multimodal architecture. Our main contributions are:

- (1) We propose MELON, a novel dual-branch multimodal framework designed to model accelerometer data by processing spectrogram images and their corresponding long-term statistical feature sequences using ResNet and a Mixture-of-Experts structure, respectively, for accurate prediction of patients’ mobility status in the ICU.
- (2) We performed extensive comparisons using data from 126 ICU patients to highlight the robustness of the proposed method. Our experiments demonstrated that the combination of spectral and temporal information of the accelerometer data could boost the prediction power.
- (3) We accessed multiple analyses to provide practical recommendations in evaluating patients’ mobility status.

2 Methods

In this study, we propose a multimodal dual-branch model, MELON, trained on spectrogram images I and their corresponding sequences of statistical features A extracted from raw accelerometer data to predict patients’ mobility measurements over a 12-hour long-term window. For mobility, we had four classes, i.e., completely immobile, very limited, slightly limited, and no limitation, extracted from the Braden scale [20]. Fig 1 illustrates the architecture of our proposed pipeline.

2.1 Accelerometer Data Preprocessing

Given the three-axis raw accelerometer data at 20 Hz for the spectral information, we first transformed the signal into a time-frequency representation using a Short-Time Fourier Transform (STFT). We set the sampling frequency to 20 Hz, a segment length of 64 samples, and an overlap of 32 samples to achieve a balanced resolution in both time and frequency domains. After computing the spectrogram S , we apply a logarithmic transformation to compress the dynamic range. The resulting logarithmic spectrogram is then normalized to the $[0, 255]$ range and converted to images I .

For the temporal information, first, we extracted five features following the prior work [24] which has been proven to be efficient to represent the temporal information: the mean and standard deviation of the vector magnitude, the mean and standard deviation of the angle between the x-axis and the vector magnitude, and the domain frequency. Then, we constructed 12-hour-long feature sequences

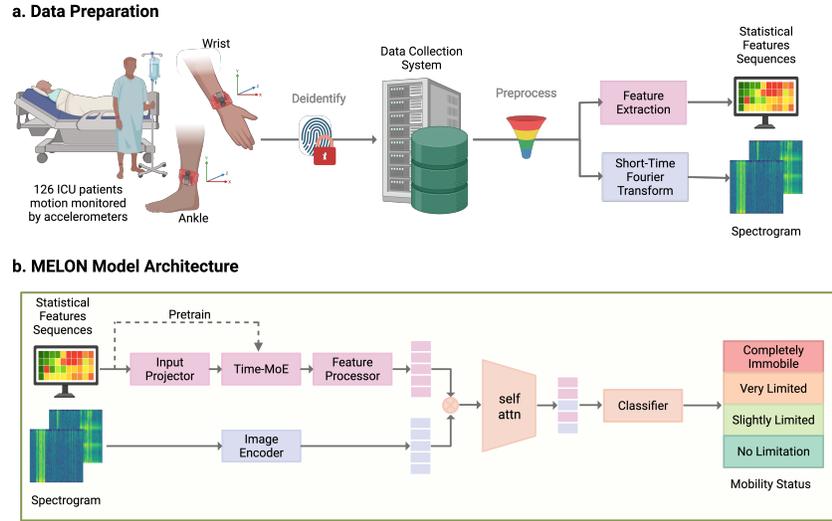


Fig. 1: In the data preparation stage (a), we collected data from two accelerometers positioned at the patient’s nondominant wrist and ankle. Data was then deidentified and stored in a secure server. We generated spectrograms using the Short-Time Fourier Transform and also extracted five features per minute with a 30-second overlap over 12 hours. The MELON model has a dual-branch structure, as shown in (b). The embeddings from the two branches are then fed into a self-attention layer to generate fusion embeddings. Then, the fused embeddings are processed in the classifier for mobility status prediction.

by (1) calculating these five features for each minute within the 12-hour window, (2) padding missing values when the accelerometer sensor was temporarily removed or the recording time was shorter than 12 hours, and (3) employing a 30-second overlap between adjacent segments. This process produced a feature sequence with 1440 time steps—each corresponding to one minute of data (with 30-second overlaps)—for each of the five features $A \in \mathbb{R}^{1440 \times 5}$.

2.2 Multimodal Dual-Branch Architecture

We provide the spectral feature images I and the temporal feature sequences A to the MELON model, which consists of three main components: (1) image encoder pre-trained on ImageNet dataset, (2) Time-MoE encoder pre-trained on our accelerometer statistical sequence data, and (3) attention fusion block and classifier.

Image Encoder To encode spectrogram images, we adapt a ResNet model pre-trained on ImageNet as the backbone image encoder. The original fully connected layer is replaced with a tailored projection module. This module consists of a linear layer that projects the high-dimensional feature vector into a 512-dimensional space, followed by a ReLU activation and dropout layer. The embedding of spectrogram is denoted as $I_{embed} \in \mathbb{R}^{1 \times \hat{D}}$.

Time-MoE Encoder Following Shi et al. [16]’s work, we utilize an input project f_{proj} and point-wise tokenization to generate embeddings of the accelerometer statistical sequences, ensuring the completeness of temporal information. Then, a SwiGLU layer is used to process each element of sequences [15]:

$$A_{hidden} = SwiGLU(A) = Swish(Wf_{proj}(A)) \otimes (Vf_{proj}(A)), \quad (1)$$

where $W \in \mathbb{R}^{D \times 1}$ and $V \in \mathbb{R}^{D \times 1}$ are two learnable parameters, $Swish$ is an activation function introduced by Ramachandran et al. [10], and $D = 128$ is the hidden size. Next, the projected A_{hidden} is fed into a MoE transformer block, which consists of a stack of transformer decoder layers augmented with a sparse MoE feed-forward network. In each decoder layer, the self-attention module computes attention scores using query, key, and value projections combined with rotary embeddings to effectively capture both local and long-range temporal dependencies [18]. Subsequently, the MoE feed-forward network employs a dynamic gating mechanism that routes token representations to one of four specialized temporal experts in addition to a shared expert. The MoE output for a token h is computed as:

$$E_{MoE}(h) = \sum_{i=1}^4 \alpha_i \cdot E_i(h) + \alpha_s \cdot E_s(h), \quad (2)$$

where the routing weights are given by $\alpha_i = softmax(W_g h)$ with $W_g \in \mathbb{R}^{4 \times D}$, and shared expert gating is computed as $\alpha_s = \sigma(W_s h)$ with $W_s \in \mathbb{R}^{1 \times D}$. Here, $E_i(h)$ represents the output of the i th specialized expert, and $E_s(h)$ is the output from the shared expert, with each expert applying an expansion via an up-projection, a non-linear SiLU activation, and a subsequent down-projection to restore the original dimensionality. Residual connections and RMS normalization are applied before the self-attention and after the MoE sub-layers to ensure stable gradient flow and robust feature integration. The final output embedding denoted $A_{embed} \in \mathbb{R}^{1 \times \hat{D}}$ is then obtained by aggregating the sequence of token representations, where $\hat{D} = 512$ is the embedding size.

The encoder is pre-trained using autoregression on our statistical features sequence (from the training set only to avoid data leakage). Given the sequential input, it predicts the next five features by adding a regression head at the end of the Time-MoE encoder. An early-stopping strategy has been set to avoid overfitting.

Table 1: Patient Characteristics (n=126) & Classes Distribution

	Train	Validation	Test
Basic information			
Number of patients, n	79	25	22
Age, mean (SD)	56.91 (14.92)	60.8 (17.55)	63.14 (11.66)
Female, n(%)	22 (28%)	7 (28%)	7(31%)
Length of stay (days), median (25 th , 75 th percentile)	13.43 (6.33, 30.04)	12.44 (8.16, 29.68)	15.00 (7.92, 22.24)
Number of samples, n	417	133	117
Classes (%)			
Completely immobile	3 (1%)	5 (4%)	2 (10%)
Very limited	138 (33%)	47 (35%)	59 (50%)
Slightly limited	236 (57%)	66 (50%)	53 (45%)
No limitation	48 (12%)	15 (11%)	3 (3%)

Abbreviations: n: number; SD: standard deviation

Fusion and Classification A self-attention mechanism is employed to effectively combine information from both modalities. First, we reshape the sum of the normalized accelerometer features and the normalized image features to act as the query, key, and value for the multi-head attention. Then, attention embedding was computed, and this yields an attention-enhanced feature vector $F_{attn} \in \mathbb{R}^{B \times \hat{D}}$, which was then passed through multi-head classifier blocks. Each classifier head comprises multiple fully connected layers with ReLU activations, culminating in a sigmoid function to produce a probability score.

3 Experiments and Results

3.1 Experimental Setup

Dataset We collected our ICU dataset of 126 adult patients who agreed to participate in this research study during admission to nine specialized ICUs (Cardiology, Cardiac, Medical, Neuromedicine, Neuro-vascular, Thoracic & Lung Transplant, Trauma, and Surgery) between 2019 and 2024. The study was approved by the University of Florida Institutional Review Board (IRB) under IRB201900354 and IRB202101013. Accelerometer sensor data for each participant was collected for at most 7 days, or until they were transferred or discharged from the ICU. Two types of devices were used in this study: 1) Shimmer ECG (Shimmer Sensing, Dublin, Ireland) and 2) Actigraph GTX3+ devices (Acti-Graph LLC, Pensacola, FL, USA). Participants were asked to wear the sensor on their wrists and/or ankles during the data collection period. Our care provider team recorded the times when the device was removed and reapplied (for example, during surgery or bathing). Additionally, mobility status data were extracted from the Electronic Health Records (EHR) and updated at each nursing shift, approximately every 12 hours.

We split the data into development and test sets by individual patient, with a ratio of 8:2, ensuring that the class distribution was preserved across all subsets. Then, we further partitioned the development set into a train and validation set for model selection and hyperparameter tuning with the same ratio. We computed and analyzed the distribution of patient characteristics and class distribution (Table 1).

Method Comparison We compared our model with multiple existing methods. We first benchmarked our proposed method against conventional approaches. Following the baseline setup from [4], we used the average activity counts along each axis (x , y , and z) during 12-hour windows as features to predict mobility using machine learning (ML) models. Additionally, we implemented GRU (Gated Recurrent Unit), the Transformer model, and Time-MoE 200M on the accelerometer sequences to provide further performance comparisons. Next, we performed an ablation study to evaluate the performance using only the spectrogram images and the performance when using only the sequential features. We also tested the performance using the accelerometer data gathered from the ankle. Our performances were evaluated by the area under the receiver operating characteristic curve (AUROC).

Implementation details We first pre-trained our Time-MoE encoder in an autoregressive manner on accelerometer statistical feature sequences. This approach leverages the temporal dependencies in the data, accelerating convergence and enhancing overall performance. For the image encoder, we used the pre-trained weights of ResNet.

All experiments, including pre-training, were conducted on an NVIDIA A100-SXM4-80GB GPU, with a batch size of 16 to fully utilize the available GPU memory. To prevent overfitting, we employed an early-stopping strategy with a patience of 7 epochs. The best model was selected based on the highest AUROC achieved on the validation set.

3.2 Results

We evaluated both conventional approaches, i.e., the baseline ML model (Logistic Regression), GRU, and the transformer (Table 2). Our proposed model MELON outperformed others in predicting "Completely Immobile", "Very Limited", "Slightly Limited", and "No Limitation" with AUROC (95% C.I.) of 0.83 (0.68-0.94), 0.81 (0.75-0.89), 0.80 (0.74-0.87), and 0.84 (0.67-0.96), respectively. "Very Limited" and "Slightly Limited" were the two dominant classes, resulting in stable performance across all experimental setups. In contrast, "Completely Immobile" and "No Limitation" were rare and imbalanced categories. Notably, our proposed model demonstrated robust performance in predicting these less-represented classes.

In our ablation study, we evaluated the model’s performance by separately removing either the accelerometer sequence features or the spectrogram features.

Table 2: **Mobility** Classification Performance Comparison

Method	AUROC (95% CI)				Overall
	Completely Immobile	Very Limited	Slightly Limited	No Limitation	
ML w/ activ. counts	0.65 (0.38-0.81)	0.68 (0.62-0.74)	0.61 (0.56-0.68)	0.50 (0.41-0.68)	0.61 (0.52,0.68)
GRU w/ accel seq.	0.70 (0.63-0.79)	0.75 (0.67-0.83)	0.76 (0.69-0.83)	0.35 (0.06-0.85)	0.65 (0.54,0.76)
Transformer w/ accel seq.	0.52 (0.42-0.62)	0.73 (0.64-0.82)	0.76 (0.69-0.84)	0.28 (0.18-0.44)	0.54 (0.45,0.64)
Time-MoE w/ accel seq.	0.64 (0.47-0.80)	0.71 (0.60-0.81)	0.73 (0.63-0.81)	0.81 (0.62-0.96)	0.72 (0.65,0.80)
MELON w/o Spec.	0.80 (0.67-0.91)	0.76 (0.66-0.85)	0.78 (0.69-0.85)	0.81 (0.64-0.94)	0.78 (0.72,0.83)
MELON w/o accel seq.	0.23 (0.04-0.45)	0.77 (0.69-0.84)	0.74 (0.67-0.83)	0.81 (0.69-0.91)	0.63 (0.56,0.71)
MELON	0.83 (0.68-0.94)*	0.81 (0.75-0.89)*	0.80 (0.74-0.87)*	0.84 (0.67-0.96)*	0.82 (0.75,0.88)*

Abbreviations: ML: machine learning; CI: Confidence interval; accel: Accelerometer; seq.: sequence; Spec.: spectrogram; *: p-value < 0.001 compared to baseline. P-values are based on pairwise Wilcoxon rank sum tests.

Our results showed a significant drop in performance when either modality was excluded. This further confirms that combining the global and fine-grained information enhances the model’s ability.

Additionally, we compared models trained on accelerometer data collected from the patient’s wrist and ankle. The results indicated that wrist-mounted data exhibited strong predictive power, effectively capturing mobility patterns. In contrast, ankle-mounted data failed to differentiate mobility differences.

4 Discussion and Conclusion

Our study demonstrates that MELON, a novel multimodal dual-branch framework combining frequency domain spectrogram features with accelerometer statistical sequences, can effectively predict long-term mobility levels in ICU patients using wrist sensor data alone. Spectrograms represent the frequency-based features that emphasize periodicities, rhythms, and oscillations in the signal that are not immediately observable in the raw time domain. In contrast, statistical feature sequences capture structured and fine-grained variations over time, focusing on temporal variations such as signal amplitude changes, localized anomalies, bursts, and overall variability. This fusion of temporal and spectral pattern recognition holds promise for objective, continuous mobility monitoring in critical care settings.

In clinical practice, continuous monitoring of patient mobility is essential for guiding timely rehabilitation. Our work suggests that wearable wrist sen-

sors alone can provide reliable predictive power of long-term mobility. A single-sensor solution minimizes patient discomfort, simplifies sensor management, and reduces deployment costs, thereby facilitating widespread adoption in busy critical care environments.

However, the findings are constrained by the limited cohort (n=126) and class imbalance affecting rare activity detection. Additionally, while our model achieved high performance in controlled evaluations, its performance in broader, more heterogeneous ICU scenarios remains to be validated. Future work will focus on expanding our dataset through pre-training on larger public human activity recognition datasets and integrating additional modalities, especially depth imaging [17], to enhance robustness, prioritizing computational efficiency for real-world ICU deployment.

In conclusion, MELON offers a promising approach for the continuous and objective assessment of patient mobility in the ICU. By addressing current limitations and incorporating further clinical data and modalities, our framework could significantly contribute to personalized patient management and improved clinical outcomes in critical care.

Data availability The data used in this project is available upon request. For details and access instructions, please refer to our project repository.

Disclosure of Interest The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Azizi, S., Mustafa, B., Ryan, F., Beaver, Z., Freyberg, J., Deaton, J., Loh, A., Karthikesalingam, A., Kornblith, S., Chen, T., et al.: Big self-supervised models advance medical image classification. In: Proceedings of the IEEE/CVF international conference on computer vision. pp. 3478–3488 (2021)
2. Barr, J., Fraser, G.L., Puntillo, K., Ely, E.W., Gélinas, C., Dasta, J.F., Davidson, J.E., Devlin, J.W., Kress, J.P., Joffe, A.M., et al.: Clinical practice guidelines for the management of pain, agitation, and delirium in adult patients in the intensive care unit. *Critical care medicine* **41**(1), 263–306 (2013)
3. Chambers, M.A., Moylan, J.S., Reid, M.B.: Physical inactivity and muscle weakness in the critically ill. *Critical care medicine* **37**(10), S337–S346 (2009)
4. Davoudi, A., Malhotra, K.R., Shickel, B., Siegel, S., Williams, S., Ruppert, M., Bihorac, E., Ozrazgat-Baslanti, T., Tighe, P.J., Bihorac, A., et al.: Intelligent icu for autonomous patient monitoring using pervasive sensing and deep learning. *Scientific reports* **9**(1), 8020 (2019)
5. Hodgson, C., Needham, D., Haines, K., Bailey, M., Ward, A., Harrold, M., Young, P., Zanni, J., Buhr, H., Higgins, A., et al.: Feasibility and inter-rater reliability of the icu mobility scale. *Heart & Lung* **43**(1), 19–24 (2014)
6. Hoyer, E.H., Young, D.L., Klein, L.M., Kreif, J., Shumock, K., Hiser, S., Friedman, M., Lavezza, A., Jette, A., Chan, K.S., et al.: Toward a common language for measuring patient mobility in the hospital: reliability and construct validity of interprofessional mobility measures. *Physical therapy* **98**(2), 133–142 (2018)

7. Ito, C., Cao, X., Shuzo, M., Maeda, E.: Application of CNN for human activity recognition with fft spectrogram of acceleration and gyro sensors. In: Proceedings of the 2018 ACM international joint conference and 2018 international symposium on pervasive and ubiquitous computing and wearable computers. pp. 1503–1510 (2018)
8. Lew, C.H., Lim, K.M., Lee, C.P., Lim, J.Y.: Human activity classification using recurrence plot and residual network. In: 2023 IEEE 11th Conference on Systems, Process & Control (ICSPC). pp. 78–83. IEEE (2023)
9. Lu, J., Tong, K.Y.: Robust single accelerometer-based activity recognition using modified recurrence plot. *IEEE Sensors Journal* **19**(15), 6317–6324 (2019)
10. Ramachandran, P., Zoph, B., Le, Q.V.: Searching for activation functions. arXiv preprint arXiv:1710.05941 (2017)
11. Sassi, M., Haleem, M.S., Pecchia, L.: Spectrogram-based approach with convolutional neural network for human activity classification. In: Mediterranean Conference on Medical and Biological Engineering and Computing. pp. 387–401. Springer (2023)
12. Schwab, K.E., To, A.Q., Chang, J., Ronish, B., Needham, D.M., Martin, J.L., Kamdar, B.B.: Actigraphy to measure physical activity in the intensive care unit: a systematic review. *Journal of intensive care medicine* **35**(11), 1323–1331 (2020)
13. Schweickert, W.D., Hall, J.: Icu-acquired weakness. *Chest* **131**(5), 1541–1549 (2007)
14. Sena, J., Mostafiz, M.T., Zhang, J., Davidson, A.E., Bandyopadhyay, S., Nerella, S., Ren, Y., Ozrazgat-Baslanti, T., Shickel, B., Loftus, T., et al.: Wearable sensors in patient acuity assessment in critical care. *Frontiers in Neurology* **15**, 1386728 (2024)
15. Shazeer, N.: Glu variants improve transformer. arXiv preprint arXiv:2002.05202 (2020)
16. Shi, X., Wang, S., Nie, Y., Li, D., Ye, Z., Wen, Q., Jin, M.: Time-moe: Billion-scale time series foundation models with mixture of experts. arXiv preprint arXiv:2409.16040 (2024)
17. Siegel, S., Zhang, J., Bandyopadhyay, S., Nerella, S., Silva, B., Baslanti, T., Bihorac, A., Rashidi, P.: Leveraging computer vision in the intensive care unit (icu) for examining visitation and mobility. arXiv preprint arXiv:2403.06322 (2024)
18. Su, J., Ahmed, M., Lu, Y., Pan, S., Bo, W., Liu, Y.: Roformer: Enhanced transformer with rotary position embedding. *Neurocomputing* **568**, 127063 (2024)
19. Tipping, C.J., Harrold, M., Holland, A., Romero, L., Nisbet, T., Hodgson, C.L.: The effects of active mobilisation and rehabilitation in icu on mortality and function: a systematic review. *Intensive care medicine* **43**, 171–183 (2017)
20. Valiani, V., Chen, Z., Lipori, G., Pahor, M., Sabbá, C., Manini, T.M.: Prognostic value of braden activity subscale for mobility status in hospitalized older adults. *Journal of hospital medicine* **12**(6), 396–401 (2017)
21. Yang, C.C., Hsu, Y.L.: A review of accelerometry-based wearable motion detectors for physical activity monitoring. *Sensors* **10**(8), 7772–7788 (2010)
22. Yu, B., Gong, K.: Adaptive whole-body pet image denoising using 3D diffusion models with controlnet. arXiv preprint arXiv:2411.05302 (2024)
23. Yu, B., Ozdemir, S., Dong, Y., Shao, W., Pan, T., Shi, K., Gong, K.: Robust whole-body pet image denoising using 3D diffusion models: evaluation across various scanners, tracers, and dose levels. *European Journal of Nuclear Medicine and Molecular Imaging* pp. 1–14 (2025)

24. Zhang, J., Contreras, M., Bandyopadhyay, S., Davidson, A., Sena, J., Ren, Y., Guan, Z., Ozrazgat-Baslanti, T., Loftus, T.J., Nerella, S., et al.: Mango: Multimodal acuity transformer for intelligent icu outcomes. arXiv preprint arXiv:2412.17832 (2024)