A Self-Supervised Learning Framework for Domain Invariant Early Prediction of Sepsis

Neeresh Kumar Perla
University of Massachusetts Lowell
Lowell, USA
neereshkumar_perla@student.uml.edu

Yingzhe Qin
University of Massachusetts Dartmouth
Dartmouth, USA
yqin@umassd.edu

Yu Shrike Zhang
Harvard Medical School
Boston, USA
yszhang@bwh.harvard.edu

Ming Shao
University of Massachusetts Lowell
Lowell, USA
ming_shao@uml.edu

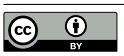
Abstract

Sepsis is a life-threatening condition that requires early and accurate detection to improve patient outcomes. While deep learning models have shown promise in predicting sepsis from high-resolution ICU time-series data, they often fail to generalize across hospitals due to domain shifts. To address this challenge, we propose GTN-SimMTM, a novel self-supervised domain adaptation framework for robust multivariate time-series modeling in early sepsis prediction. Our approach integrates a Gated Transformer Network (GTN), which effectively captures both temporal and channel-wise dependencies, with masked time-series modeling (SimMTM), a self-supervised contrastive learning strategy that enhances feature consistency across domains. Furthermore, we incorporate multiple domain adaptation techniques, including Deep Coral, DANN, SASA, and DSAN, to mitigate distributional discrepancies between hospital datasets. Extensive experiments on the PhysioNet dataset demonstrate that GTN-SimMTM consistently outperforms previous methods in utility score, AUROC, and AUPRC in multiple hospital settings. Notably, our framework achieves strong results on a large-scale test set of 16,000 highly imbalanced samples while eliminating the need for extensive feature engineering. These results highlight the power of self-supervised pre-training, contrastive learning, and domain adaptation to advance early sepsis detection for real-world clinical deployment.

Keywords

Early sepsis detection, Domain adaptation, Self-supervised learning, Healthcare Domain Shift

ACM Reference Format:
Neeresh Kumar Perla, Yingzhe Qin, Yu Shrike Zhang, and Ming Shao. 2025. A Self-Supervised Learning Framework for Domain Invariant Early Prediction of Sepsis. In *ACM/IEEE International Conference on Connected Health: Applications, Systems and Engineering Technologies (CHASE '25), June 24–26, 2025, New York, NY, USA.* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3721201.3725520



CHASE '25, New York, NY, USA

ACM ISBN 979-8-4007-1539-6/2025/06
https://doi.org/10.1145/3721201.3725520

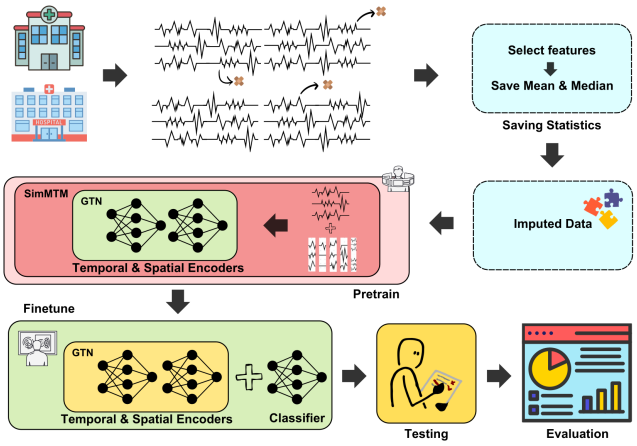


Figure 1: Overview of the proposed methodology, highlighting the integration of data statistics preservation, missing data handling, pre-training, fine-tuning, and evaluation processes within the SimMTM and GTN frameworks.

1 Introduction

Sepsis remains one of the leading causes of mortality in intensive care units (ICU), and early detection is critical for timely intervention and improved patient outcomes [10, 11]. Despite the widespread use of clinical scoring systems such as SOFA [16] and APACHE-II [5], these methods are inherently limited by their reliance on static thresholds and a narrow set of features [10, 11]. In contrast, high-resolution, multivariate time-series data exemplified by the PhysioNet/Computing in Cardiology Challenge 2019 dataset [12] — offers a promising alternative by enabling frame-by-frame classification of a patient's physiological state.

A key innovation of the PhysioNet challenge is its unique utility score, which rewards correct predictions made within a specific time window (from 12 hours before to 3 hours after sepsis onset) and penalizes both excessively early or delayed predictions [11]. This evaluation criterion not only emphasizes prediction accuracy but also underscores the clinical value of early detection. However, as data are collected from different hospital systems, domain shift becomes a significant challenge. Recent studies have begun exploring uncertainty quantification and domain adaptation methods to address this issue (e.g., Deep Coral [13], SASA [7], and CDAN [9].

Domain adaptation methods are known for addressing the problems caused by domain shift. Domain shift refers to the variations in data distributions that occur when models trained on data from one setting are applied to data from a different setting. In this PhysioNet challenge, data is sourced from multiple hospital systems. These differences can lead to a domain shift, where the statistical properties of the data differ between hospitals. To mitigate the effects of domain shift, domain adaptation methods are employed. These methods aim to adjust the model to perform effectively across varying data distributions and enhance the model's generalizability and robustness, ensuring more accurate and reliable predictions in diverse clinical environments. Moreover, modern deep learning techniques — especially those incorporating self-supervised and contrastive learning strategies (e.g., Time-Frequency Consistency (TFC) [18]) — have further pushed the boundaries of representation learning for time-series data.

In our work, we propose a novel hybrid framework that adopts a "pre-training-then-fine-tuning" pipeline based on a Gated Transformer Network (GTN) integrated with a masked time-series modeling paradigm (SimMTM [2]). GTN, as introduced in [8], employs a two-tower encoder architecture that explicitly separates temporal and channel-wise feature extraction, effectively capturing both temporal and inter-channel correlations, thus leading to improved predictive accuracy on the large-scale ICU dataset. Our approach leverages self-supervised learning—including contrastive techniques such as TFC—to learn robust temporal and spatial representations while incorporating multiple domain adaptation strategies (e.g., Deep Coral [13], SASA [7], and CDAN [9]) to address the significant domain shift observed between hospital systems, as verified by improved performance across domains.

Our key contributions include:

- We propose a GTN-SimMTM framework for self-supervised pre-training on multivariate ICU time-series data, effectively capturing both temporal and channel correlations and leading to enhanced predictive accuracy.
- We integrate several contrastive learning strategies (including TFC) and systematically evaluate multiple domain adaptation methods (DSAN, Deep Coral, SASA, CDAN) to mitigate domain shift, as verified by improved performance across hospital systems.
- We perform extensive experiments on a large-scale test set (16,000 samples) using the official evaluation script, demonstrating the scalability and robustness of our approach.

The remainder of the paper is organized as follows. Section 2 reviews related work. Section 3 describes our methodology, including data preparation, pre-training, fine-tuning, and evaluation procedures. Section 4 presents our experimental results and analysis, and Section 5 concludes with discussions on clinical impact and future directions.

2 Related Work

A substantial body of research has focused on early sepsis prediction using ICU data. [11] developed an XGBoost-based model utilizing 107 clinical features from ICU records to predict sepsis onset six hours in advance. [14] later proposed a Transformer-based model that exploited temporal context from multiple pre-diagnosis windows, achieving a performance gain of approximately 20% over traditional RNNs. In parallel, [1] employed Temporal Convolutional Networks (TCNs) on ECG signals for sepsis detection, demonstrating that deep architectures can capture subtle physiological changes.

From a biomedical perspective, [15] investigated the prognostic utility of point-of-care serum proenkephalin (PENK) in septic shock patients, while [10] compared multiple biomarkers for mortality prediction in sepsis. [4] further refined the feature space for sepsis mortality prediction from the MIMIC-IV database, achieving an AUROC of 0.94 using a Random Forest model. [17] advanced the field by proposing SepsisLab — a system that integrates uncertainty quantification and active sensing to enhance early sepsis prediction. Additionally, fusion-based architectures, such as the Parallel LSTM-DNN model [3], have been explored to leverage complementary strengths of different deep learning paradigms.

While these works have significantly advanced the state of the art, they typically assume consistent data distributions. In real-world applications, however, ICU data often exhibit high missing rates, varying scales, and inconsistent sampling frequencies across different hospitals, rendering domain shift a more complex challenge than in typical computer vision tasks. Recent domain adaptation approaches, such as Deep CORAL [13], Sparse Associative Structure Alignment [7], and Conditional Domain Adversarial Networks (CDAN) [9], have been proposed to address these discrepancies; however, relatively few studies have applied these techniques to heterogeneous ICU time-series data.

In terms of self-supervised learning, masked time-series modeling has shown promise for learning robust representations from incomplete data. [18] introduced a self-supervised contrastive pre-training approach based on Time-Frequency Consistency (TFC), which enforces similarity between the time-based and frequency-based representations of an instance. This strategy is particularly beneficial for noisy physiological signals and motivates our integration of contrastive learning within our framework.

Furthermore, the choice of network architecture is crucial when dealing with multivariate time-series data. Unlike standard Transformers, the Gated Transformer Network (GTN) [8] utilizes a two-tower encoder structure that separately captures temporal and channel-wise correlations, leading to superior performance on multivariate time-series classification tasks. This architectural advantage makes GTN especially suitable for modeling the complex, high-dimensional ICU data encountered in sepsis prediction.

In summary, although previous studies [1, 3, 4, 6, 10, 11, 14, 15, 17] have demonstrated the feasibility of early sepsis prediction using advanced machine learning and deep learning techniques, the challenges posed by domain shift in heterogeneous ICU data and the effective use of self-supervised learning remain underexplored. Our work addresses these gaps by proposing a GTN-SimMTM framework that leverages self-supervised pre-training (including TFC) and multiple domain adaptation strategies to enhance early sepsis detection on a large-scale, multi-domain dataset.

3 Methodology

In this section, we introduce the dataset and explain how it is prepared for training, followed by details on the model's pre-training, fine-tuning strategies, and evaluation procedures. Specifically, we begin by describing how the data is processed to facilitate model training. Next, we explore methods for pre-training the base model and adapting it to the target domain using self-supervised learning. Additionally, we discuss techniques for aligning source and target datasets using domain adaptation methods to predict sepsis onset six hours in advance. Finally, the evaluation segment explains the metrics and procedures employed to assess the model's performance. Figure 2 illustrates the overall architecture of the proposed approach, and the subsequent sections elaborate on each step depicted in the figure.

3.1 Dataset Setup

We utilized data from the PhysioNet/Computing in Cardiology Challenge 2019 [12], which focuses on the early prediction of sepsis, to forecast sepsis onset six hours in advance. The challenge provides ICU records from two distinct hospital systems: one hospital consists of 20,336, while the other has 20,000 patient samples. To mimic the domain shift challenge, we use the first hospital data as the source and the second one as the target. Each patient sample spans an ICU stay, during which clinical variables are recorded hourly, resulting in rows of data that capture up to 40 features, including eight vital signs, 26 laboratory test results, and six demographic attributes.

3.2 Dataset Preparation

We experimented with various imputation techniques and found that relying solely on mean or median imputation does not ensure optimal results. Consequently, we identified which features would benefit most from mean or median imputation by checking whether the mean and median values for each feature fall within the observed minimum and maximum range of that feature. If the mean lies within these bounds, we use mean imputation for that feature; similarly, if the median is more appropriate, we apply median imputation. We then store the computed imputation values for each feature to apply consistently when preparing the fine-tuning and test sets. After calculating and storing these statistics, we rescale the FiO2 feature to its original range because some rows contained improperly scaled values.

Next, we addressed missing values in the Unit1 and Unit2 features, which serve as administrative identifiers for ICU units. Since a patient cannot be in two ICU units simultaneously, we rely on the current time step to determine whether Unit1 or Unit2 is active: if a patient is admitted to Unit1, then Unit2 is set to zero, and vice versa. We apply the same procedure to the fine-tuning and test sets, ensuring we check the status of Unit 1 and Unit 2 at each time step.

For any remaining missing entries, we use forward fill, backward fill, and finally, assign zeros if the values are still unavailable. Finally, once the dataset is finalized with all missing values imputed, we pad extra rows to ensure that all samples have the same sequence length. Surprisingly, GTN can handle raw multivariate time-series data without requiring normalization or scaling. Therefore, we adhere to the same experimental setup. To avoid data leakage into the fine-tuning and test sets, we only use the mean and median statistics calculated from the pre-training dataset. These values remain fixed to ensure a fair evaluation.

3.3 GTN-SimMTM Hybrid Architecture

In this paper, we integrated the Gated Transformer Network (GTN) [8] as a backbone of a Simple pre-training framework for Masked Time-series Modeling (SimMTM) [2] for pre-training the model. GTN is a two-tower encoder architecture that is used to capture temporal and spatial information for a given multivariate time series. The temporal encoder uses self-attention with masks on each point across all the channels by calculating the pair-wise attention weights among all the time steps, whereas the spatial encoder calculates the attention weights among different channels across all the time steps. Then, a simple gating mechanism is used to learn the weight of each tower.

SimMTM, on the other hand, is a masked time-series modeling (MTM) paradigm in which random time points are masked, and the model is trained to reconstruct them using contextual information. However, rather than relying only on the unmasked segments of a single sequence, SimMTM leverages multiple masked instances of the same time series. This aggregation strategy enables the model to preserve essential temporal patterns more effectively during pre-training.

3.4 Pre-Training

During the pre-training phase, we use the proposed GTN-SimMTM for representation learning. Our GTN-SimMTM architecture is trained to reconstruct the original time series from multiple masked variations, thereby learning robust temporal and spatial representations. SimMTM's [2] contrastive approach and masked modeling help the model distinguish meaningful patterns from irrelevant noise, making it highly effective for downstream tasks such as classification.

During pre-training, we optimize three objectives — contrastive loss, reconstruction loss, and automatic weighted loss (AWL). The contrastive loss encourages time-series instances with similar structures to remain close in the learned feature space, while the reconstruction loss preserves step-wise and channel-wise dependencies. The AWL component balances the contrastive and reconstruction losses, preventing overfitting to either objective and ensuring high-quality feature extraction.

3.5 Fine-Tuning and DA Extension

After pre-training, the learned representations from GTN-SimMTM are fine-tuned for specific tasks. During fine-tuning, we did not freeze any layers. We attached a classification head (a 2-layer MLP) after passing the gating output through an MLP. We train the classification head utilizing the learned feature embeddings for supervised learning by optimizing a simple cross-entropy loss function.

Since GTN, along with SimMTM, provides a solid base framework for domain adaptation (DA) tasks, GTN can easily be extended to well-established DA models, including DeepCoral [13], CDAN [9], SASA [7], and DSAN [19] with an additional MLP layer. In particular, DA loss and constraint functions can be added to be jointly optimized with the GTN backbone in training the target

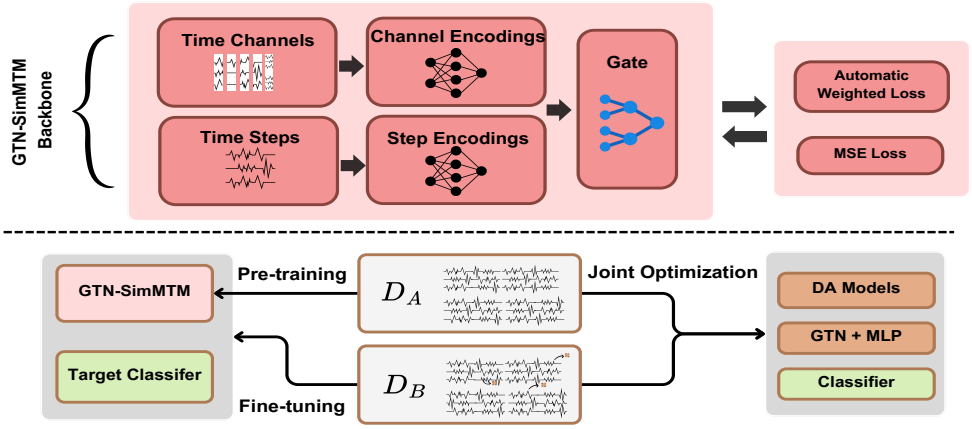


Figure 2: The proposed GTN-SimMTM framework follows a two-stage learning approach. In the pre-training stage, data from Hospital A (D_A) is encoded through Time Channels and Time Steps representations, followed by Channel and Step Encodings. A Gating Mechanism and MLP process these encodings, optimizing with Automatic Weighted Loss and MSE Loss to learn meaningful representations. In the fine-tuning stage, learned representations are transferred to a downstream task. A Target Classifier is attached to the pre-trained network, and the entire network is fine-tuned using data from Hospital B (D_B), leveraging the pre-trained model for improved classification performance. GTN can be readily extended to DA models with an additional MLP layer.

classifier. Note that we use an MLP after the gating operation in GTN [8] to reduce the number of features produced when the outputs of the towers are concatenated. This is primarily due to the high-dimensional output from the gating operation, and the proposed MLP layer would reduce the number of features and lower computational costs.

4 Experiments

In this section, we evaluate the proposed method on the PhysioNet/Computing in Cardiology Challenge 2019 dataset [12] to predict sepsis six hours before its clinical onset. The following subsections will detail the experimental setup, experimental results, and ablation study.

4.1 Setup

To mimic the domain adaptation setup, we use one hospital's data as the source dataset (D_A), while another one is the target dataset. The target dataset is stratified into D_B and D_C to preserve class imbalance. Our experiments involve three datasets:

- Source dataset of 20,336 samples (D_A) for pre-training
- Target dataset of 4,000 samples (D_B) for fine-tuning
- Target dataset of 16,000 samples (D_C) for test

The size of each dataset provides an ideal testbed for domain adaptation tasks. We also used the original evaluation script provided by the PhysioNet/ Challenge 2019 due to its practical value in early prediction compared to standard procedures. To our knowledge, this is the first study that allows the evaluation of sepsis early prediction under domain shift.

4.2 Model Configuration

To train GTN-SimMTM, we used a hidden dimension of 512 and an intermediate feedforward dimension of 1024, with 8 attention heads, 8 query/value dimensions, and 8 Transformer layers. We employed positional encoding and masking for improved sequential modeling and applied a dropout rate of 0.2 to mitigate overfitting. The input feature dimensionality was set to 336 with 40 channels, and the output space was constrained to two classes. We adopted the Adam optimizer β_1=0.9 and β_2=0.99 with a learning rate of 3×10^{-8} for both pre-training and fine-tuning.

The model was pre-trained for 20 epochs and fine-tuned for 9–10 epochs using a batch size of 32. Likewise, all domain adaptation methods were pre-trained for 40 epochs and fine-tuned for 20 epochs with the same batch size, while the self-supervised learning method was pre-trained for 20 epochs and fine-tuned for 15 epochs. All experiments shared the same data partitioning strategy: 100% source dataset D_A for pre-training, 20% of the target dataset D_B for fine-tuning, and the remaining 80% of the target dataset for testing.

4.3 Evaluation

We follow the challenge guidelines for evaluating the proposed framework, which uses a utility function that rewards early sepsis predictions and penalizes late or false ones. Specifically:

- **For True Sepsis Patients:** 1) The model earns a reward for predicting sepsis between 12 hours before and up to 3 hours after the onset time. 2) The maximum possible reward is 1.0. 3) Missing sepsis entirely or predicting it more than 12 hours too early incurs penalties. 4) Very early predictions carry a

Table 1: Comparison of various DA methods and our proposed GTN-SimMTM method. Bold fonts highlight the performance gain compared to the original DA methods without using GTN backbone.

Model	GTN as Feature Extractor (Ours)				Original Feature Extractor			
	Utility Score	AUROC	AUPRC	Accuracy	Utility Score	AUROC	AUPRC	Accuracy
DSAN	**0.22**	**0.67**	**0.05**	0.96	0.09	0.65	0.03	0.97
DeepCoral	**0.21**	0.58	0.03	0.93	0.16	0.59	0.03	0.95
CDAN	0.15	0.66	0.05	0.84	0.18	0.68	0.04	0.86
SASA	**0.14**	**0.66**	0.04	**0.84**	-0.04	0.64	0.04	0.73
GTN-SimMTM	**0.26**	0.72	0.01	0.92	-	-	-	-

penalty of up to 0.05, while late predictions can be penalized up to -2.0.

- **For Non-Sepsis Patients:** The model is penalized if it predicts sepsis at any point. The maximum false-alarm penalty is 0.05, and no penalty or reward is applied if the model correctly avoids predicting sepsis.

We use the official competition evaluation script, which reports the utility score, AUROC, AUPRC, and accuracy.

4.4 Experimental Results of GTN-SimMTM

Table 1 demonstrates the effectiveness of the proposed GTN-SimMTM approach compared to the state-of-the-art DA methods. For a fair comparison, GTN is also applied for representation learning purposes.

Among DA methods, DSAN achieves the highest utility score (0.22) and accuracy (0.96), outperforming other domain adaptation techniques. Additionally, AUROC increases to 0.67, indicating improved model robustness when handling distributional shifts across datasets. These findings suggest that DSAN successfully mitigates domain discrepancies and enhances model performance for early sepsis detection. In contrast, CDAN and SASA exhibit comparable AUROC values (0.66) but lower utility scores (0.15 and 0.14, respectively). While these methods demonstrate strong class separation, their lower utility scores suggest potential limitations in early sepsis prediction. DeepCoral achieves a moderate balance between a utility score (0.21) and AUROC (0.58), with an accuracy of 0.93. SimMTM with MLP achieves a high accuracy (0.94) and AUROC (0.64) and a lower utility score (0.20).

Table 1 also highlights the role of GTN in feature extraction. The GTN-based feature extractor outperforms the original extractor in utility score, which is a critical metric for early sepsis prediction. For instance, DSAN achieved a utility score of 0.22 with GTN, compared to only 0.09 with the original extractor. Similarly, DeepCoral's performance improved from 0.16 to 0.21. Notably, SASA performed poorly with the original extractor (-0.04 utility score) but achieved 0.14 with GTN, which shows its reliance on effective feature representations. Accuracy trends indicate that the original feature extractor slightly outperforms GTN in some cases, but this does not translate into improved utility scores. DSAN, for example, achieved 97% accuracy with the original extractor versus 96% with GTN, but its utility score was much lower (0.09 vs. 0.22). This suggests that while the original extractor may yield higher classification accuracy, it is less effective in making clinically meaningful predictions, which the utility score captures.

4.5 Ablation Study

Table 2: Performance comparison under different training/testing setups.

Model	Utility Score	AUROC	AUPRC	Accuracy
GTN ($D_A \rightarrow D_C$)	-0.02	0.58	0.04	0.73
GTN ($D_B \rightarrow D_C$)	-0.02	0.59	0.04	0.72
GTN ($D_A + D_B \rightarrow D_C$)	**0.06**	0.64	0.01	**0.77**
Challenge 2019 (Best)	-0.5	0.81	0.07	0.76

4.5.1 GTN with domain shift. First, we demonstrate the efficacy of GTN fine-tuning under domain shift. We applied no pre-training strategy at all, representing the raw performance of the GTN model. The following scenarios are included in the Table 2:

- $D_A \rightarrow D_C$: GTN pre-trained on D_A but not fine-tuned on D_B and tested directly on D_C.
- $D_B \rightarrow D_C$: GTN trained directly on D_B and tested on D_C.
- $D_A + D_B \rightarrow D_C$: GTN pre-trained on D_A and fine-tuned on D_B , then tested on D_C without SimMTM.
- Challenge 2019 (Best): A cross-dataset result (best) from PhysioNet Challenge 2019.

We compare our experimental results with those of the top-performing team in the competition, as shown in the last row of Table 2. This demonstrates that fine-tuning is necessary under domain shift, while this was never considered by the existing methods. The GTN model performs competitively, particularly when pre-trained on D_A and fine-tuned on D_B and outperforms when trained solely on D_A or D_B. Specifically, the utility score improves from -0.5 to 0.06, and accuracy is increased from 0.76 to 0.77. We observe a distinction between the utility score and the remaining metrics. Unlike other metrics, the utility score considers the sequential nature of time-series data. Specifically, during evaluation, the model is queried multiple times for each patient, corresponding to the total number of recorded time steps. For instance, if a test sample contains 40 hours of recordings, the model is queried 40 times. The process unfolds as follows:

- In the first iteration, the model is queried with only the first time step.
- In the second iteration, it is queried with the first two time steps.
- In the third iteration, it is queried with the first three time steps, and so on.

Therefore, we believe that the utility score is crucial for evaluating a sepsis prediction model designed to detect onset six hours in advance, alongside other performance metrics.

Table 3: Performance comparison of GTN under different pre-training strategies.

Model	Utility Score	AUROC	AUPRC	Accuracy
GTN-SimMTM (w/o MLP)	**0.26**	**0.72**	**0.01**	**0.92**
GTN-SimMTM (w/ MLP)	0.20	0.63	0.01	0.94
GTN-TFC	-0.03	0.53	0.03	0.79

4.5.2 Pre-training strategies. In this set of experiments, we evaluated the efficacy of the pre-training strategy used in GTN-SimMTM. We explored the model performance without incorporating an MLP layer after the gating operating in GTN. The results are summarized in Table 3 by considering the following two scenarios:

- GTN-SimMTM (w/ and w/o MLP): The model was pre-trained on D_A for 9 epochs and fine-tuned on D_B for 15 epochs.
- GTN-TFC: The model was pre-trained on D_A for 20 epochs and fine-tuned on D_B for 19 epochs.

The results highlight the significant impact of SimMTM pre-training on model performance. Compared to the baseline GTN model, pre-training with SimMTM substantially improves the utility score (from 0.06 to 0.26), and AUROC (from 0.64 to 0.72). This suggests that the masked time-series modeling approach effectively enhances the model's ability to learn robust temporal representations, leading to better early sepsis detection. Additionally, accuracy increases from 0.77 to 0.92, demonstrating the model's improved predictive reliability.

In contrast, the TFC pre-training strategy does not yield competitive results, as reflected in its lower AUROC (0.53) and utility score (-0.03). This indicates that while contrastive learning techniques like TFC can capture useful temporal dependencies, they are not as robust as SimMTM in handling masked and incomplete time-series data for sepsis prediction.

5 Conclusions

In this study, we proposed a novel hybrid framework, GTN-SimMTM, for early sepsis prediction using the PhysioNet/Computing in Cardiology Challenge 2019 dataset to predict sepsis six hours before its clinical onset. By leveraging the strengths of Gated Transformer Networks (GTN) and masked time-series modeling (SimMTM), our approach effectively captured both temporal and channel-wise correlations, eliminating the need for extensive feature engineering techniques and improving utility score along with the predictive accuracy compared to top champion results in the challenge. We further integrated multiple contrastive learning techniques and domain adaptation methods, such as DSAN, DeepCoral, CDAN, and SASA to address domain shift challenges inherent in heterogeneous hospital datasets. Our experimental results demonstrated that GTN-SimMTM significantly enhances early sepsis detection compared to baseline models, achieving notable improvements in utility score, AUROC, and accuracy. The introduction of an MLP layer post-gating further optimized performance by reducing computational complexity while maintaining model robustness.

6 Acknowledgment

This material is based upon work supported by the National Science Foundation under Grant No. 2225818 and 2225698.

References

[1] Merve Apalak and Kamran Kiasaleh. 2024. Advancing Early Detection of Sepsis With Temporal Convolutional Networks Using ECG Signals. *IEEE Access* (2024).

[2] Jiaxiang Dong, Haixu Wu, Haoran Zhang, Li Zhang, Jianmin Wang, and Mingsheng Long. 2023. SimMTM: A Simple Pre-Training Framework for Masked Time-Series Modeling. arXiv:2302.00861 [cs.LG] https://arxiv.org/abs/2302.00861

[3] Hadj Ali Elmerahi, Baghdad Atmani, Fatiha Barigou, Belarbi Khemliche, Badreddine Errouane, and Mohammed Bousmaha. 2024. Parrallel LSTM-DNN Fusion Model for Early Prediction of Sepsis in Intensive Care Units. In *2024 4th International Conference on Embedded & Distributed Systems (EDiS)*. IEEE, 43–48.

[4] Jiayi Gao, Yuying Lu, Negin Ashrafi, Ian Domingo, Kamiar Alaei, and Maryam Pishgar. 2024. Prediction of Sepsis Mortality in ICU patients using machine learning methods. *BMC Medical Informatics and Decision Making* 24, 1 (2024), 228.

[5] William A Knaus, Elizabeth A Draper, Douglas P Wagner, and Jack E Zimmerman. 1985. APACHE II: a severity of disease classification system. *Critical care medicine* 13, 10 (1985), 818–829.

[6] Christopher Kok, V Jahmunah, Shu Lih Oh, Xujuan Zhou, Raj Gururajan, Xiaohui Tao, Kang Hao Cheong, Rashmi Gururajan, Filippo Molinari, and U Rajendra Acharya. 2020. Automated prediction of sepsis using temporal convolutional network. *Computers in Biology and Medicine* 127 (2020), 103957.

[7] Zijian Li, Ruichu Cai, Jiawei Chen, Yuguan Yan, Wei Chen, Keli Zhang, and Junjian Ye. 2022. Time-Series Domain Adaptation via Sparse Associative Structure Alignment: Learning Invariance and Variance. arXiv:2205.03554 [cs.LG] https://arxiv.org/abs/2205.03554

[8] Minghao Liu, Shengqi Ren, Siyuan Ma, Jiahui Jiao, Yizhou Chen, Zhiguang Wang, and Wei Song. 2021. Gated Transformer Networks for Multivariate Time Series Classification. arXiv:2103.14438 [cs.LG] https://arxiv.org/abs/2103.14438

[9] Mingsheng Long, Zhangjie Cao, Jianmin Wang, and Michael I Jordan. 2018. Conditional adversarial domain adaptation. *Advances in neural information processing systems* 31 (2018).

[10] Sonia Luka, Adela Golea, Raluca Mihaela Tat, Eugenia Maria Lupan Mureșan, George Teo Voicescu, Ștefan Cristian Vesa, and Daniela Ionescu. 2024. Biomarkers as Predictors of Mortality in Sepsis and Septic Shock for Patients Admitted to Emergency Department: Who Is the Winner? A Prospective Study. *Journal of Clinical Medicine* 13, 19 (2024), 5678.

[11] Fahim Mohammad, Lakshmi Arunachalam, Samanway Sadhu, Boudewijn Aasman, Shweta Garg, Adil Ahmed, Silvie Colman, Meena Arunachalam, Sudhir Kulkarni, and Parsa Mirhaji. 2024. Early prediction of onset of sepsis in Clinical Setting. arXiv:2402.03486 [cs.LG] https://arxiv.org/abs/2402.03486

[12] Matthew A Reyna, Christopher S Josef, Russell Jeter, Supreeth P Shashikumar, M Brandon Westover, Shamim Nemati, Gari D Clifford, and Ashish Sharma. 2020. Early prediction of sepsis from clinical data: the PhysioNet/Computing in Cardiology Challenge 2019. *Critical care medicine* 48, 2 (2020), 210–217.

[13] Baochen Sun and Kate Saenko. 2016. Deep coral: Correlation alignment for deep domain adaptation. In *Computer Vision–ECCV 2016 Workshops: Amsterdam, The Netherlands, October 8-10 and 15-16, 2016, Proceedings, Part III 14*. Springer, 443–450.

[14] Yan Tang, Yu Zhang, and Jiaxi Li. 2024. A time series driven model for early sepsis prediction based on transformer module. *BMC Medical Research Methodology* 24, 1 (2024), 23.

[15] Christos Verras, Sofia Bezati, Vasiliki Bistola, Ioannis Ventoulis, Dionysis Matsiras, Sotirios Tsiodras, John Parissis, and Effie Polyzogopoulou. 2024. Point-of-Care Serum Proenkephalin as an Early Predictor of Mortality in Patients Presenting to the Emergency Department with Septic Shock. *Biomedicines* 12, 5 (2024), 1004.

[16] J L Vincent, Rui Moreno, Jukka Takala, Sheila Willatts, Arnaldo De Mendonça, Hajo Bruining, CK Reinhart, PeterM Suter, and Lambertius G Thijs. 1996. The SOFA (Sepsis-related Organ Failure Assessment) score to describe organ dysfunction/failure: On behalf of the Working Group on Sepsis-Related Problems of the European Society of Intensive Care Medicine (see contributors to the project in the appendix).

[17] Changchang Yin, Pin-Yu Chen, Bingsheng Yao, Dakuo Wang, Jeffrey Caterino, and Ping Zhang. 2024. Sepsislab: Early sepsis prediction with uncertainty quantification and active sensing. In *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*. 6158–6168.

[18] Xiang Zhang, Ziyuan Zhao, Theodoros Tsiligkaridis, and Marinka Zitnik. 2022. Self-supervised contrastive pre-training for time series via time-frequency consistency. *Advances in Neural Information Processing Systems* 35 (2022), 3988–4003.

[19] Yongchun Zhu, Fuzhen Zhuang, Jindong Wang, Guolin Ke, Jingwu Chen, Jiang Bian, Hui Xiong, and Qing He. 2020. Deep subdomain adaptation network for image classification. *IEEE transactions on neural networks and learning systems* 32, 4 (2020), 1713–1722.