Cross-center Early Sepsis Recognition by Medical Knowledge Guided Collaborative Learning for Data-scarce Hospitals

Ruiqing Ding[1,2], Fangjie Rong[4], Xiao Han[*3], Leye Wang[*1,2]
[1]Key Lab of High Confidence Software Technologies (Peking University), Ministry of Education, China
[2]School of Computer Science, Peking University, Beijing, China
[3]School of Information Management and Engineering, Shanghai University of Finance and Economics, Shanghai, China
[4]School of Mathematical Science, Peking University, Beijing, China
ruiqingding@stu.pku.edu.cn,1900010621@pku.edu.cn,xiaohan@mail.shufe.edu.cn,leyewang@pku.edu.cn

ABSTRACT

There are significant regional inequities in health resources around the world. It has become one of the most focused topics to improve health services for data-scarce hospitals and promote health equity through knowledge sharing among medical institutions. Because electronic medical records (EMRs) contain sensitive personal information, privacy protection is unavoidable and essential for multi-hospital collaboration. In this paper, for a common disease in ICU patients, sepsis, we propose a novel cross-center collaborative learning framework guided by medical knowledge, *SofaNet*, to achieve early recognition of this disease. The Sepsis-3 guideline, published in 2016, defines that sepsis can be diagnosed by satisfying both suspicion of infection and Sequential Organ Failure Assessment (SOFA) greater than or equal to 2. Based on this knowledge, *SofaNet* adopts a multi-channel GRU structure to predict SOFA values of different systems, which can be seen as an auxiliary task to generate better health status representations for sepsis recognition. Moreover, we only achieve feature distribution alignment in the hidden space during cross-center collaborative learning, which ensures secure and compliant knowledge transfer without raw data exchange. Extensive experiments on two open clinical datasets, MIMIC-III and Challenge, demonstrate that *SofaNet* can benefit early sepsis recognition when hospitals only have limited EMRs.

CCS CONCEPTS

• **Applied computing** → **Health informatics**; • **Security and privacy** → Privacy protections.

KEYWORDS

healthcare representation learning, collaborative learning, early sepsis recognition

*Corresponding authors

WWW '23, April 30-May 4, 2023, Austin, TX, USA

ACM ISBN 978-1-4503-9416-1/23/04…$15.00
https://doi.org/10.1145/3543507.3583989

ACM Reference Format:
Ruiqing Ding, Fangjie Rong, Xiao Han and Leye Wang. 2023. Cross-center Early Sepsis Recognition by Medical Knowledge Guided Collaborative Learning for Data-scarce Hospitals. In *Proceedings of the ACM Web Conference 2023 (WWW '23), April 30-May 4, 2023, Austin, TX, USA*. ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/3543507.3583989

1 INTRODUCTION

Significant disparities in health resources have always existed, not only in developed and developing countries but even in different areas and among different ethnicities within the same country. Because of the restriction of health resources and service level, it is overwhelming for medical institutions in less developed areas to early diagnosis and timely clinical management of some noncommunicable diseases, e.g., sepsis, diabetes, and heart diseases. Enhancing health services in less developed regions is important to promote health equity.[1] By leveraging web and AI techniques, recent efforts have attempted to connect multiple medical institutions for knowledge sharing to improve the service for data-scarce hospitals [10, 13, 17, 38]. However, medical data involves individuals' private and sensitive information, and thus directly transmitting these datasets will inevitably lead to severe privacy violations [25, 31]. In essence, enhancing health equity for medical institutions lacking data resources remains a critical issue on a global scale.

In this paper, we use *early sepsis recognition* as the representative task to study how to improve health equity for medical institutions without sufficient data, considering that sepsis is one of the most serious medical conditions causing millions of deaths with significant regional disparity. Sepsis is a life-threatening organ dysfunction resulting from a dysregulated host response to infection [35]. If not detected early and treated promptly, it can result in septic shock, multiple organ failure, and even death. Owing to the complexity and importance of clinical sepsis diagnosis and treatment, there are multiple versions of sepsis consensus definitions and diagnosis guidelines, including Sepsis-1 (1991) [27], Sepsis-2 (2001) [19] and Sepsis-3 (2016) [35]. It was estimated that there were 48.9 million cases and 11 million sepsis-related deaths worldwide in 2017, accounting for almost 20% of all global deaths [33]. Moreover, significant regional disparities exist in sepsis incidence and mortality — approximately 85% of sepsis cases and sepsis-related deaths occurred in low- and middle-income countries, specifically with the highest burden in sub-Saharan Africa, Oceania, south Asia, east Asia, and southeast Asia [33]. Prior research has suggested that the sepsis mortality rate may increase by 7% for every one-hour delay

[1]https://www.who.int/health-topics/health-equity

in the administration of antibiotic treatment [16]. Therefore, early recognition is a crucial first step in the management of sepsis.

Nowadays, machine learning techniques are broadly studied in early sepsis recognition and diagnosis, such as the linear model [34], neural network [7], GBDT [20], etc. These methods require a large amount of training data to guarantee performance. Unfortunately, a prior worldwide data challenge [32] has revealed that the early sepsis recognition model learned from a hospital's data may not work well for another hospital. However, it is unrealistic to have large-scale electronic medical records for every hospital, especially for sparsely populated areas where the admitted patients are limited. To cope with the limitations of small data, some academic studies have built models with data from multiple hospitals, called multicenter study [5, 39]. Nevertheless, most multicenter studies do not consider the potential privacy leakage when different centers' patient data are gathered.

In this research, to strengthen the ability of early sepsis recognition for medical institutions without sufficient data, we investigate two possible strategies: (i) incorporating domain knowledge in healthcare model design to relieve the data limitation, and (ii) enabling cross-center collaborations between medical institutions in a privacy-preserving manner. Accordingly, we propose a cross-center collaborative learning framework, *SofaNet*, to realize early sepsis recognition with two main components: (i) *the multi-channel recurrent neural network structure* to predict SOFA (Sequential Organ Failure Assessment) scores of multiple systems which are highly associated with sepsis diagnosis (according to the guidelines of Sepsis-3 [35]), and (ii) *the cross-center feature distribution alignment component* to achieve effective knowledge transfer without raw data sharing. Our contributions are as follows:

(i) To the best of our knowledge, this is one of the pioneering studies to design a privacy-preserving cross-center collaboration mechanism for early sepsis recognition by explicitly considering domain knowledge (i.e., multi-system SOFA scores).

(ii) By conducting the transfer experiments on two open clinical datasets, MIMIC-III and Challenge, we have validated taht *SofaNet* significantly and consistently outperforms the start-of-the-art methods without raw clinical data exchange. We release our code at https://doi.org/10.5281/zenodo.7625404.

2 RELATED WORK

Machine learning techniques are excellent at analyzing complex signals in data-rich environments which promise the effectiveness of early sepsis recognition. Most studies are carried out in the ICU [6, 14]. The systematic review and meta-analysis indicate that individual machine learning models can accurately predict the onset of sepsis in advance on retrospective data [20, 37]. The PhysioNet/Computing in Cardiology (CinC) Challenge 2019 focused on this issue and promoted the development of open-source AI algorithms for real-time and early recognition of sepsis [32]. However, there are few studies that concentrate on sepsis recognition without sufficient data and the common method is centralized learning (i.e., put data together to analysis [39]). Recently, transfer learning and multi-task learning are becoming popular to utilize knowledge shared by different datasets or tasks to achieve better model performance [29]. In healthcare, some work focuses on a specific disease to design transfer methods, such as blood pressure [18], heart disease [13], Covid-19 [22], etc.; there are also work focusing on privacy issues and designing corresponding algorithms [12].

3 PROBLEM FORMULATION

Early Sepsis Recognition. The objective is to use patients' electronic medical records (EMRs) to predict the risk of sepsis. Considering the early warning of sepsis is potentially life-saving, we recognize sepsis onset in the next 6 hours with patients' last 6-hour EMRs, including vital variables, laboratory variables and demographic information (details in Appendix). This setting is consistent with the PhysioNet Computing in Cardiology Challenge 2019 [9, 32] on *Early Prediction of Sepsis from Clinical Data*[2]. In brief, given n patients' variables, $\{\mathcal{X}_1, \mathcal{X}_2, \cdots, \mathcal{X}_n\}$, where the i-th patient's data is $\mathcal{X}_i = \{\boldsymbol{x}_{i,1}, \boldsymbol{x}_{i,2}, \cdots, \boldsymbol{x}_{i,m}\}$, $\boldsymbol{x}_{i,j}$ is the clinical features of j-th hour since patient i entered ICU. For each 6-hour records, $\{\boldsymbol{x}_{i,k}, \boldsymbol{x}_{i,k+1}, \cdots, \boldsymbol{x}_{i,k+5}\}$, we aim to predict whether sepsis will occur for patient i in the next 6 hours ,i.e., before $(k + 11)$-th hour.

Cross-center Early Sepsis Recognition. When there are only limited EMRs per hospital, it is difficult to guarantee the model performance. In this paper, we focus on the collaborative learning of two participants (i.e., hospitals), to generate better health status representations for sepsis recognition. From the machine learning perspective, it can be viewed as a multi-task learning task [41].

4 METHODOLOGY

4.1 Overview

SofaNet is proposed to achieve secure and compliant knowledge transfer when there is limited data in each hospital. Figure 1 shows the overall framework, which contains two key parts, (i) *Multi-channel GRU for health status representation learning* and (ii) *Privacy-preserving Cross-center Collaborative Learning*.

4.2 Health Status Representation Learning

Existing machine learning methods for early sepsis recognition usually input EMR variables and output a binary label on whether sepsis would occur [6, 34]. For hospitals with limited patients' EMRs, this learning strategy may not perform well as the supervision signals are restricted to the small number of sepsis labels.

To increase the supervision signals for such hospitals without sufficient data, our proposed *SofaNet* includes a novel set of auxiliary tasks by incorporating the medical expert knowledge from the sepsis diagnosis guideline [35]. Specifically, since SOFA >= 2 is one of the key conditions for sepsis diagnosis, SOFA score prediction has the potential to become an auxiliary task for early sepsis recognition. Furthermore, SOFA scores can be calculated for six criteria including respiratory, cardiovascular, hepatic, renal, coagulation, and neurological systems. Hence, multiple new supervision signals on SOFA scores can be provided.

With this prior medical knowledge, we adopt a multi-channel GRU structure to embed the health status of each system individually. Also, the dynamic changes of vital signs and laboratory

[2] https://physionet.org/content/challenge-2019/1.0.0/

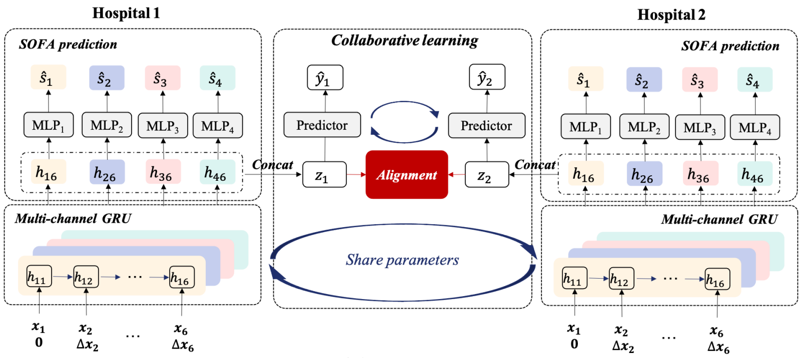


Figure 1: The *SofaNet* framework: (1) health status representation learning with SOFA prediction as the auxiliary task; (2) cross-center collaborative learning with model parameters sharing and feature distribution alignment in hidden space.

	MIMIC-III	Challenge
# patients	13379	4717
# septic patients	2688	1441
Sepsis prevalence (%)	20.09	30.55
# records	69946	18616
# records occur Sepsis in next 6 hours	8751	2476
records with sepsis (%)	12.51	13.30
missing rate (%)	20.83	63.07

Table 1: The Statistics of Datasets

variables can acutely reflect changes of a patient's status [23]. Therefore, given a 6-hour record of a patient, $\{x_1, x_2, \cdots, x_6\}$, we compute the differential features with $\Delta x_i = x_i - x_{i-1}$ when $i > 1$ and set $\Delta x_1 = \mathbf{0}$. As shown in Figure 1, we take the concatenation of original features and differential features as the input, i.e., $\{(x_1, \mathbf{0}), (x_2, \Delta x_2), \cdots, (x_6, \Delta x_6)\}$. Due to the missing variables in the datasets, SOFA scores can be precisely calculated for four systems (out of six), including coagulation, liver, cardiovascular, and renal [3]. Therefore, we build a 4-channel GRU feature extractor with the same input. For each channel, we take the last hidden state of GRU as the output, i.e., $\{h_{16}, h_{26}, h_{36}, h_{46}\}$. As the SOFA score prediction is the auxiliary task, the loss function for each hospital can be written as

$$\mathcal{L}_{local} = \mathcal{L}_{sepsis} + \alpha * \sum_{i=1}^{4} \mathcal{L}_{sofa_i} \tag{1}$$

where $\mathcal{L}_{sepsis}$ and $\mathcal{L}_{sofa_i}$ are cross entropy, and α is set to 0.5 [4]. In brief, in addition to the supervision signal of sepsis ($\mathcal{L}_{sepsis}$), we add four new supervision signals of SOFA scores ($\mathcal{L}_{sofa_i}, i = 1...4$) to improve the learning robustness for hospitals with scarce data.

4.3 Cross-center Collaborative Learning

In addition to auxiliary tasks on SOFA prediction, we introduce a cross-center collaborative learning procedure for multiple hospitals. Specifically, we expect that this collaborative procedure can enable data-scarce hospitals to benefit each other, so that they can be motivated to participate. Note that as direct data sharing may violate data protection regulations such as GDPR, we need to ensure that knowledge is shared while raw data is well protected.

Based on this idea, we design a cross-center collaborative learning mechanism, which achieves knowledge sharing by two ways: (i) *model parameter sharing* in each iteration, which can be seen as the simplified version of the privacy-preserving federated learning algorithm, *FedAvg* [24], since there are only two participants and the central server is unnecessary; (ii) *feature distribution alignment* in hidden space, like domain adaptation [28], to avoid diverged distributions between two hospitals, which may result in negative knowledge transfer. Various alignment methods can be implemented, such as maximum mean discrepancy (MMD) [36] and adversarial training [8, 30]. In our current implementation, we use the MMD method (denoted as $SofaNet_{mmd}$) as it performs generally well in our experiments. The loss function of $SofaNet_{mmd}$ is

$$\mathcal{L} = \mathcal{L}_{local_1} + \mathcal{L}_{local_2} + \mathcal{L}_{mmd} \tag{2}$$

Privacy Protection. In training $SofaNet_{mmd}$, two hospitals only need to transfer intermediate results (i.e.,, model parameters and hidden representations) instead of raw data, which follows the privacy protection criteria in state-of-the-art federated domain adaptation methods [30]. Meanwhile, recent studies show that such intermediate results may also indirectly reveal raw data under certain conditions [42]. In our future work, we plan to add advanced techniques, such as homomorphic encryption and differential privacy, to further protect the intermediate training results [21].

5 EXPERIMENTS

5.1 Experiment Setup

We conduct our experiments on two widely-used real-life Sepsis recognition datasets, **MIMIC-III** [11] and the PhysioNet Computing in Cardiology Challenge 2019 (**Challenge**) [32]. As some patients' records have a large number of missing values, we screen out the patients whose missing value ratio is less than 80%. For missing values, we fill in with data from the previous time point. If a feature at the initial time point is missing, we fill in with the mean value. Through such preprocessing, we obtain the final data for experiments and keep the records of 10% patients as test data. Some basic statistic information is enumerated in Table 1. We list the detailed features in Table 3 (Appendix), including 3 demographic variables, 6 vital sign variables and 18 laboratory variables. The experiment environment is described in Appendix.

5.2 Baselines

In our experiments, we assume that there is only x% ($x = 1$ by default) patients' records from MIMIC or Challenge for two hospitals, respectively, to simulate a data-scarce scenario. To compare with our method *SofaNet*, we implement two types of methods.

[3]The features and SOFA scoring standard can refer to Table 3 and Table 4 respectively in Appendix.

[4]This hyperparameter is determined according to the experimental results, and the difference is small when using different values (from 0.5 to 1.0).

- *Local Learning*: each hospital trains a classifier only with its local medical records.
- *Collaborative Learning*: two hospitals collaborate by techniques such as parameter sharing and finetuning.

For *local learning*, we introduce several classical models widely used for sepsis recognition, including, *Logistic Regression* (*LR*) [34], *Neural Network* (*NN*) [7], *XGBoost*[40], and *GRU*[2]. For *collaborative learning*, we mainly compare with the state-of-the-art finetuning methods without raw data exchange [13, 18, 22]. Following previous research using EMR data [1, 22, 23], model performance is assessed by the area under the receiver operating characteristic curve (AUROC), area under the precision-recall curve (AUPRC), and the minimum of precision and sensitivity (Min(Se,P+)).

5.3 Experimental Results

5.3.1 Main Result. As shown in Table 2, *SofaNet* outperforms in most metrics (except for the AUROC on MIMIC-III), demonstrating *SofaNet* can learn a better representation through knowledge transfer while protecting the raw data. Concretely, compared to the best local methods, *SofaNet* achieves a 0.9% higher AUROC, a 5.39% higher AUPRC, a 2.56% higher min(Se,P+) on MIMIC-III dataset, and achieves a 7.70% higher AUROC, a 50.22% higher AUPRC and a 25.39% higher min(Se, P+) on Challenge dataset.

Effectiveness of Collaborative Learning: By comparing the local methods and collaborative methods, we can conclude that learning knowledge from each other between different datasets can promote the prediction performance for their respective tasks. Also, we can observe that the improvement is more obvious on Challenge dataset, because of the relatively poorer data quality (i.e., smaller data size and higher missing rate shown in Table 1).

Effectiveness of Knowledge-guide Multi-channel GRU: Compared to *GRU* and $SofaNet_{lc}$ *w.o. MC* model, $SofaNet_{lc}$ achieves a better performance during local training on two datasets. This indicates that taking SOFA values prediction as the auxiliary task is helpful for data-scarce hospitals' early sepsis recognition.

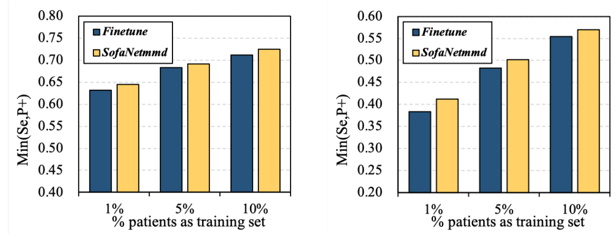
Future direction on combining *XGBoost* and *SofaNet*: Among local methods, *XGBoost* performs the best, outperforming all the deep learning methods including $SofaNet_{lc}$. This is actually consistent with the competition results of the PhysioNet Challenge, where the top three teams all use *XGBoost*-like ensemble methods [4, 26, 40]. A promising direction would be how to combine the local learning capacity of *XGBoost* and the collaboration power of $SofaNet_{mmd}$. We have attempted to use *XGBoost* on the collaboratively aligned representations of $SofaNet_{mmd}$ (i.e., z_1 and z_2 in Figure 1) and observed certain improvements (although not stable). This confirms the feasibility of combining these two methods, and we highly believe that more advanced combination techniques can be developed in the future for significant improvements.

5.3.2 Varying the Size of Training Set. Fixing the test data, we adjust the data size of training set, i.e., sample 1%, 5%, 10% patients and use their medical records as the training data. Figure 2 shows the Min(Se, P+) values of MIMIC-III and Challenge test data under different training data sizes respectively. The Min(Se, P+) values rise as the training data size increases, and $SofaNet_{mmd}$ consistently outperforms *Finetune*. The performance gap between *SofaNet* and *Finetune* is more considerable in smaller datasets, which indicates the capability of our proposed collaborative learning mechanism to alleviate the data insufficiency problem.

Table 2: Early Sepsis Recognition Performance with Only 1% Patients as the Training Set

	MIMIC-III			Challenge		
	AUROC	AUPRC	Min(Se, P+)	AUROC	AUPRC	Min(Se, P+)
Local Learning						
LR	0.8987 (0.002)	0.6583 (0.003)	0.6022 (0.001)	0.5724 (0.002)	0.1715 (0.001)	0.2279 (0.002)
NN	0.8625 (0.008)	0.5986 (0.014)	0.5452 (0.006)	0.4745 (0.009)	0.1334 (0.004)	0.1457 (0.007)
XGBoost	0.9107 (0.003)	0.6829 (0.006)	0.6285 (0.009)	0.7588 (0.009)	0.2676 (0.007)	0.3292 (0.008)
GRU	0.8992 (0.006)	0.6639 (0.012)	0.5994 (0.008)	0.6283 (0.044)	0.2247 (0.019)	0.2667 (0.025)
$SofaNet_{lc}$ *w.o. MC*	0.9046 (0.006)	0.6726 (0.104)	0.6004 (0.009)	0.6647 (0.052)	0.2476 (0.043)	0.2941 (0.063)
$SofaNet_{lc}$	0.9067 (0.006)	0.6807 (0.017)	0.6171 (0.009)	0.6895 (0.036)	0.2609 (0.019)	0.2982 (0.015)
Collaborative Learning						
Finetune	**0.9216** (0.006)	0.7113 (0.028)	0.6314 (0.026)	0.7955 (0.010)	0.3161 (0.016)	0.3831 (0.020)
$SofaNet_{mmd}$	0.9187 (0.006)	**0.7197** (0.016)	**0.6446** (0.008)	**0.8172** (0.007)	**0.4020** (0.023)	**0.4128** (0.025)

1. Values in "( )" denote the standard deviation of five experiments' results;
2. **Bold** denotes the best-performed ones of the task;
3. Underline denotes the best-performed ones in local learning;
4. $SofaNet_{lc}$ means *SofaNet* without collaborative training;
5. $SofaNet_{lc}$ *w.o. MC* means $SofaNet_{lc}$ without multi-channel GRU.



(a) MIMIC-III (b) Challenge

Figure 2: Performance under different training data volume.

6 CONCLUSION

In this paper, we propose a privacy-preserving cross-center collaborative learning method, *SofaNet*, for early sepsis recognition. The experimental results on two datasets show the effectiveness of *SofaNet*, especially for hospitals with scarce data. We mainly achieve raw data protection through parameters sharing and health status representation alignment. Also, we find that *XGBoost* performs well in local training. It is worth thinking about how to further integrate *XGBoost* into *SofaNet* to get better results.

Limitation & Future Work. We focus on early sepsis recognition by two-hospital collaborative learning. However, this method can be easily extended to n-hospital scenario: first, every two hospitals can do collaborative learning for a better health status representation; second, one hospital can combine (e.g., concatenation) all the collaboratively learned representations with other hospitals for its own patients' early sepsis recognition. For the representative disease, sepsis, we used the Sepsis-3 guideline [35] to guide our method design. There are also diagnostic guidelines for different diseases, like KDIGO for acute kidney injury [3]. With the corresponding knowledge, we can design models like *SofaNet*. This idea to inject domain knowledge can be generalized to other domains.

ACKNOWLEDGMENTS

We appreciate Yanlin Chen for her discussion of the model design idea. This research was supported by National Key R&D Program of China (2020AAA0109401) and NSFC Grants no. 72071125, 61972008, and 72031001.

REFERENCES

[1] Edward Choi, Cao Xiao, Walter Stewart, and Jimeng Sun. 2018. Mime: Multilevel medical embedding of electronic health records for predictive healthcare. *Advances in neural information processing systems* 31 (2018).

[2] Junyoung Chung, Caglar Gulcehre, KyungHyun Cho, and Yoshua Bengio. 2014. Empirical evaluation of gated recurrent neural networks on sequence modeling. *arXiv preprint arXiv:1412.3555* (2014).

[3] KJKIS Disease. 2012. Improving global outcomes (KDIGO) acute kidney injury work group: KDIGO clinical practice guideline for acute kidney injury. *Kidney Int Suppl* 2, 1 (2012), 1–138.

[4] John Anda Du, Nadi Sadr, and Philip de Chazal. 2019. Automated prediction of sepsis onset using gradient boosted decision trees. In *2019 Computing in Cardiology (CinC)*. IEEE, Page–1.

[5] Ricard Ferrer, Antonio Artigas, Mitchell M Levy, Jesus Blanco, Gumersindo Gonzalez-Diaz, José Garnacho-Montero, Jordi Ibáñez, Eduardo Palencia, Manuel Quintana, María Victoria De La Torre-Prados, et al. 2008. Improvement in process of care and outcome after a multicenter severe sepsis educational program in Spain. *Jama* 299, 19 (2008), 2294–2303.

[6] Lucas M Fleuren, Thomas LT Klausch, Charlotte L Zwager, Linda J Schoonmade, Tingjie Guo, Luca F Roggeveen, Eleonora L Swart, Armand RJ Girbes, Patrick Thoral, Ari Ercole, et al. 2020. Machine learning for the prediction of sepsis: a systematic review and meta-analysis of diagnostic test accuracy. *Intensive care medicine* 46, 3 (2020), 383–400.

[7] Joseph Futoma, Sanjay Hariharan, and Katherine Heller. 2017. Learning to detect sepsis with a multitask Gaussian process RNN classifier. In *International Conference on Machine Learning*. PMLR, 1174–1182.

[8] Yaroslav Ganin and Victor Lempitsky. 2015. Unsupervised Domain Adaptation by Backpropagation. In *Proceedings of the 32nd International Conference on Machine Learning (Proceedings of Machine Learning Research, Vol. 37)*, Francis Bach and David Blei (Eds.). PMLR, Lille, France, 1180–1189.

[9] Ary L Goldberger, Luis AN Amaral, Leon Glass, Jeffrey M Hausdorff, Plamen Ch Ivanov, Roger G Mark, Joseph E Mietus, George B Moody, Chung-Kang Peng, and H Eugene Stanley. 2000. PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals. *circulation* 101, 23 (2000), e215–e220.

[10] Priyanka Gupta, Pankaj Malhotra, Jyoti Narwariya, Lovekesh Vig, and Gautam Shroff. 2020. Transfer learning for clinical time series analysis using deep neural networks. *Journal of Healthcare Informatics Research* 4, 2 (2020), 112–137.

[11] Alistair EW Johnson, Tom J Pollard, Lu Shen, H Lehman Li-Wei, Mengling Feng, Mohammad Ghassemi, Benjamin Moody, Peter Szolovits, Leo Anthony Celi, and Roger G Mark. 2016. MIMIC-III, a freely accessible critical care database. *Scientific data* 3, 1 (2016), 1–9.

[12] Ce Ju, Ruihui Zhao, Jichao Sun, Xiguang Wei, Bo Zhao, Yang Liu, Hongshan Li, Tianjian Chen, Xinwei Zhang, Dashan Gao, et al. 2020. Privacy-preserving technology to help millions of people: Federated prediction model for stroke prevention. *arXiv preprint arXiv:2006.10517* (2020).

[13] Mohammad Kachuee, Shayan Fazeli, and Majid Sarrafzadeh. 2018. Ecg heartbeat classification: A deep transferable representation. In *2018 IEEE international conference on healthcare informatics (ICHI)*. IEEE, 443–444.

[14] Anuj Karpatne, Imme Ebert-Uphoff, Sai Ravela, Hassan Ali Babaie, and Vipin Kumar. 2018. Machine learning for the geosciences: Challenges and opportunities. *IEEE Transactions on Knowledge and Data Engineering* 31, 8 (2018), 1544–1554.

[15] Diederik P Kingma and Jimmy Ba. 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980* (2014).

[16] Anand Kumar, Daniel Roberts, Kenneth E Wood, Bruce Light, Joseph E Parrillo, Satendra Sharma, Robert Suppes, Daniel Feinstein, Sergio Zanotti, Leo Taiberg, et al. 2006. Duration of hypotension before initiation of effective antimicrobial therapy is the critical determinant of survival in human septic shock. *Critical care medicine* 34, 6 (2006), 1589–1596.

[17] Gyemin Lee, Ilan Rubinfeld, and Zeeshan Syed. 2012. Adapting surgical models to individual hospitals using transfer learning. In *2012 IEEE 12th international conference on data mining workshops*. IEEE, 57–63.

[18] Jared Leitner, Po-Han Chiang, and Sujit Dey. 2021. Personalized blood pressure estimation using photoplethysmography: A transfer learning approach. *IEEE Journal of Biomedical and Health Informatics* 26, 1 (2021), 218–228.

[19] Mitchell M. Levy, Mitchell P. Fink, John C. Marshall, Edward Abraham, Derek Angus, Deborah Cook, Jonathan Cohen, Steven M. Opal, Jean-Louis Vincent, Graham Ramsay, and for the International Sepsis Definitions Conference. 2003. 2001 SCCM/ESICM/ACCP/ATS/SIS International Sepsis Definitions Conference. 29, 4 (2003), 530–538.

[20] Xiang Li, Xiao Xu, Fei Xie, Xian Xu, Yuyao Sun, Xiaoshuang Liu, Xiaoyu Jia, Yanni Kang, Lixin Xie, Fei Wang, et al. 2020. A Time-Phased Machine Learning Model for Real-Time Prediction of Sepsis in Critical Care. *Critical Care Medicine* 48, 10 (2020), e884–e888.

[21] Lingjuan Lyu, Han Yu, Xingjun Ma, Lichao Sun, Jun Zhao, Qiang Yang, and Philip S Yu. 2020. Privacy and robustness in federated learning: Attacks and defenses. *arXiv preprint arXiv:2012.06337* (2020).

[22] Liantao Ma, Xinyu Ma, Junyi Gao, Xianfeng Jiao, Zhihao Yu, Chaohe Zhang, Wenjie Ruan, Yasha Wang, Wen Tang, and Jiangtao Wang. 2021. Distilling knowledge from publicly available online EMR data to emerging epidemic for prognosis. In *Proceedings of the Web Conference 2021*. 3558–3568.

[23] Liantao Ma, Chaohe Zhang, Yasha Wang, Wenjie Ruan, Jiangtao Wang, Wen Tang, Xinyu Ma, Xin Gao, and Junyi Gao. 2020. Concare: Personalized clinical feature embedding via capturing the healthcare context. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 34. 833–840.

[24] Brendan McMahan, Eider Moore, Daniel Ramage, Seth Hampson, and Blaise Aguera y Arcas. 2017. Communication-efficient learning of deep networks from decentralized data. In *Artificial intelligence and statistics*. PMLR, 1273–1282.

[25] Riccardo Miotto, Fei Wang, Shuang Wang, Xiaoqian Jiang, and Joel T Dudley. 2018. Deep learning for healthcare: review, opportunities and challenges. *Briefings in bioinformatics* 19, 6 (2018), 1236–1246.

[26] James Morrill, Andrey Kormilitzin, Alejo Nevado-Holgado, Sumanth Swaminathan, Sam Howison, and Terry Lyons. 2019. The signature-based model for early detection of sepsis from electronic health records in the intensive care unit. In *2019 Computing in Cardiology (CinC)*. IEEE, Page–1.

[27] American College of Chest Physicians, Society of Critical Care Medicine Consensus Conference Committee, et al. 1992. American College of Chest Physicians/Society of Critical Care Medicine Consensus Conference: definitions for sepsis and organ failure and guidelines for the use of innovative therapies in sepsis. *Crit. Care Med* 20 (1992), 864–874.

[28] Sinno Jialin Pan, Ivor W Tsang, James T Kwok, and Qiang Yang. 2010. Domain adaptation via transfer component analysis. *IEEE Transactions on Neural Networks* 22, 2 (2010), 199–210.

[29] Sinno Jialin Pan and Qiang Yang. 2009. A survey on transfer learning. *IEEE Transactions on knowledge and data engineering* 22, 10 (2009), 1345–1359.

[30] Xingchao Peng, Zijun Huang, Yizhe Zhu, and Kate Saenko. 2019. Federated Adversarial Domain Adaptation. In *International Conference on Learning Representations*.

[31] W Nicholson Price and I Glenn Cohen. 2019. Privacy in the age of medical big data. *Nature medicine* 25, 1 (2019), 37–43.

[32] Matthew A Reyna, Chris Josef, Salman Seyedi, Russell Jeter, Supreeth P Shashikumar, M Brandon Westover, Ashish Sharma, Shamim Nemati, and Gari D Clifford. 2019. Early prediction of sepsis from clinical data: the PhysioNet/Computing in Cardiology Challenge 2019. In *2019 Computing in Cardiology (CinC)*. IEEE, Page–1.

[33] Kristina E Rudd, Sarah Charlotte Johnson, Kareha M Agesa, Katya Anne Shackelford, Derrick Tsoi, Daniel Rhodes Kievlan, Danny V Colombara, Kevin S Ikuta, Niranjan Kissoon, Simon Finfer, et al. 2020. Global, regional, and national sepsis incidence and mortality, 1990–2017: analysis for the Global Burden of Disease Study. *The Lancet* 395, 10219 (2020), 200–211.

[34] Supreeth P Shashikumar, Matthew D Stanley, Ismail Sadiq, Qiao Li, Andre Holder, Gari D Clifford, and Shamim Nemati. 2017. Early sepsis detection in critical care patients using multiscale blood pressure and heart rate dynamics. *Journal of electrocardiology* 50, 6 (2017), 739–743.

[35] Mervyn Singer, Clifford S. Deutschman, Christopher Warren Seymour, Manu Shankar-Hari, Djillali Annane, Michael Bauer, Rinaldo Bellomo, Gordon R. Bernard, Jean-Daniel Chiche, Craig M. Coopersmith, Richard S. Hotchkiss, Mitchell M. Levy, John C. Marshall, Greg S. Martin, Steven M. Opal, Gordon D. Rubenfeld, Tom van der Poll, Jean-Louis Vincent, and Derek C. Angus. 2016. The Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3). *JAMA* 315, 8 (02 2016), 801–810. https://doi.org/10.1001/jama.2016.0287

[36] Eric Tzeng, Judy Hoffman, Ning Zhang, Kate Saenko, and Trevor Darrell. 2014. Deep domain confusion: Maximizing for domain invariance. *arXiv preprint arXiv:1412.3474* (2014).

[37] Franco van Wyk, Anahita Khojandi, and Rishikesan Kamaleswaran. 2019. Improving Prediction Performance Using Hierarchical Analysis of Real-Time Data: A Sepsis Case Study. *IEEE Journal of Biomedical and Health Informatics* 23, 3 (2019), 978–986. https://doi.org/10.1109/JBHI.2019.2894570

[38] Kuba Weimann and Tim OF Conrad. 2021. Transfer learning for ECG classification. *Scientific reports* 11, 1 (2021), 1–12.

[39] Jianfeng Xie, Hongliang Wang, Yan Kang, Lixin Zhou, Zhongmin Liu, Bingyu Qin, Xiaochun Ma, Xiangyuan Cao, Dechang Chen, Weihua Lu, et al. 2020. The epidemiology of sepsis in Chinese ICUs: a national cross-sectional survey. *Critical Care Medicine* 48, 3 (2020), e209–e218.

[40] Morteza Zabihi, Serkan Kiranyaz, and Moncef Gabbouj. 2019. Sepsis prediction in intensive care unit using ensemble of XGboost models. In *2019 Computing in*

Cardiology (CinC). IEEE, Page–1.
[41] Yu Zhang and Qiang Yang. 2021. A survey on multi-task learning. *IEEE Transactions on Knowledge and Data Engineering* (2021).
[42] Ligeng Zhu, Zhijian Liu, and Song Han. 2019. Deep leakage from gradients. *Advances in neural information processing systems* 32 (2019).

A SUPPLEMENT INFORMATION

In this section, we list the feature description in detail in Table 3 and the standard of Sequential Organ Failure Assessment (SOFA) in Table 4 which is given in Sepsis-3 [35].

Table 3: Feature Description

Type	Features
Demographic variables	Age, Gender, ICU_hours
Vital sign variables	Heart rates, Temperature, Systolic BP, Mean arterial pressure, Diastolic BP, Respiration rate
Laboratory variables	FiO_2, SaO_2, pH, AST, BUN, Calcium, Chloride, Creatinine, Glucose, Potassium, Total Bilirubin, Hct, Hgb, PTT, WBC, Platelets, BUN/CR, SaO_2/FiO_2

Table 4: Sequential [Sepsis-Related] Organ Failure Assessment Score [35]

System	Score				
	0	**1**	**2**	**3**	**4**
Respiration PaO_2/FiO_2, mmHg (kPa)	>= 400(53.3)	< 400(52.3)	< 300(40)	<200(26.7) with respiratory support	<100 (13.3) with respiratory support
Coagulation Platelets, $\times 10^3/\mu L$	>=150	< 150	< 100	<50	<20
Liver Total Bilirubin, mg/dL ($\mu mol/L$)	<1.2(20)	1.2-1.9 (20-32)	2.0-5.9 (33-101)	6.0-11.9 (102-204)	> 12.0 (204)
Cardiovascular	MAP >= 70 mmHg	MAP <70 mmHg	Dopamine < 5 or dobutamine (any dose)	Dopamine 5.1-15 or epinephrine <= 0.1 or norepinephrine <= 0.1	Dopamine > 15 or epinephrine > 0.1 or norepinephrine > 0.1
Central nervous system Glasgow Coma Scale score	15	13-14	10-12	6-9	<6
Renal Creatinine, mg/dL ($\mu mol/L$) Urine output, mL/d	< 1.2 (110)	1.2-1.9 (110-170)	2.0-3.4 (171-299)	3.5-4.9 (300-440) < 500	>5.0 (440) < 200

B EXPERIMENT ENVIRONMENT

The experiment is conducted on a server with AMD Ryzen 9 3900X 12-Core Processor, 64 GB RAM and GeForce RTX 3090. The code is implemented based on Pytorch 1.8.0. To train the model, we use Adam [15] with the batch size of 32, and the learning rate is set to 1e-3. We repeat each experiment for 5 times (i.e., 5 seeds) and record the average results.