AI in EHR - AI capstone

Intro slide deck: AI in EHR - AI capstone.pdf
Video: https://youtube.com/playlist?list=PLYsINdwWInhwZreq7P_XLfV7Y3eh1Jo-3
Data: AI in EHR Dataset (Please login with your NYCU Google account)

Project Description:

1. Prediction tasks (choose one):
    a. **Task 1**: Patients who are admitted to the ICU may have sepsis, and **we would like to predict the sepsis onset in ICU at least 4 hours earlier** with the vital signs, lab results, diagnosis in prior inpatient visits, and demographic information collected at least 4 hr before the onset of sepsis.
    b. **Task 2**: Patients who are admitted to the ICU have a high mortality rate, and **we would like to predict the in-hospital mortality** with the vital signs, lab results, diagnosis in prior inpatient visits, and demographic information collected at the early stage (first 6 hours) of ICU admission.

2. Study cohort
    a. patients had at least one ICU stay
    b. first ICU stay, if the patients had multiple ICU stays in the dataset
    c. patients should have
        i. at least 4 hours of records before onset (task 1)
        ii. at least 6 hours of records before discharge (task 2).

    TODO 1:
    - Select the study cohort from the dataset we provide.
    - Draw a flow chart for the cohort selection. If you are not sure about the "flow chart", here are the examples.

3. Features
    a. Age & Gender
        i. in *Patient* table
    b. BMI
        i. in *chartevents* table
        ii. calculation needed
    c. Laboratory results from the ICU
        i. in *labevents* tables
        ii. "Suggested" important laboratory items are listed below
    d. Vital signs from the ICU
        i. in *chartevents* tables
        ii. "Suggested" important vital signs are listed below
    e. Diagnosis from the **previous** hospitalization (if available)

i. in *diagnosis* tables

f. Task targets

   i. Task 1: Sepsis onset or not: The definition is a bit complicated, so we have the labels for you!

   ii. Task 2: In-hospital mortality or not

TODO 2:

- Extract features listed above.
- Perform descriptive analysis across all features.

4. Data preprocess
   a. multiple measurements (temporal information)
   b. missing values
   c. outliers
   d. …others

TODO 3:

- Explain your strategies for data preprocessing.

5. Build a machine learning model to predict sepsis onset (task 1) or in-hospital mortality (task 2), with whatever algorithms.
   a. How to deal with multiple measurements?
   b. How to incorporate temporal information?

TODO 4:

- Describe the strategies of model development and evaluation
- Draw the proposed model architecture and describe it in detail
- Provide the evaluation results and state the conclusion (**The model's performance will not be perfect with only a limited number of patients)**

6. Share the codes, the results and descriptions from all the **TODOs**, and your conclusion based on the results.

References

About the data

1. MIMIC official site: https://mimic.mit.edu/
2. MIMIC tutorial: https://mimic.mit.edu/docs/iv/tutorials/video/
3. MIMIC IV paper: https://www.nature.com/articles/s41597-022-01899-x

About the task 1 (sepsis onset)

1. Moor, M., Bennett, N., Plečko, D., Horn, M., Rieck, B., Meinshausen, N., Bühlmann, P., & Borgwardt, K. (2023). Predicting sepsis using deep learning across international sites: a

retrospective development and validation study. *EClinicalMedicine*, *62*, 102124. https://doi.org/10.1016/j.eclinm.2023.102124
2. Shashikumar, S. P., Wardi, G., Malhotra, A., & Nemati, S. (2021). Artificial intelligence sepsis prediction algorithm learns to say “I don’t know.” *NPJ Digital Medicine*, *4*(1), 134. https://doi.org/10.1038/s41746-021-00504-6
3. Reyna, M. A., Josef, C. S., Jeter, R., Shashikumar, S. P., Westover, M. B., Nemati, S., Clifford, G. D., & Sharma, A. (2020). Early Prediction of Sepsis From Clinical Data: The PhysioNet/Computing in Cardiology Challenge 2019. *Critical Care Medicine*, *48*(2), 210–217. https://doi.org/10.1097/CCM.0000000000004145

About the task 2 (in-hospital mortality)

1. Gao, J., Lu, Y., Ashrafi, N., Domingo, I., Alaei, K., & Pishgar, M. (2024). Prediction of sepsis mortality in ICU patients using machine learning methods. *BMC Medical Informatics and Decision Making*, *24*(1), 228. https://doi.org/10.1186/s12911-024-02630-z
2. Iwase, S., Nakada, T.-A., Shimada, T., Oami, T., Shimazui, T., Takahashi, N., Yamabe, J., Yamao, Y., & Kawakami, E. (2022). Prediction algorithm for ICU mortality and length of stay using machine learning. *Scientific Reports*, *12*(1), 12912. https://doi.org/10.1038/s41598-022-17091-5
3. Hou, N., Li, M., He, L., Xie, B., Wang, L., Zhang, R., Yu, Y., Sun, X., Pan, Z., & Wang, K. (2020). Predicting 30-days mortality for MIMIC-III patients with sepsis-3: a machine learning approach using XGboost. *Journal of Translational Medicine*, *18*(1), 462. https://doi.org/10.1186/s12967-020-02620-5

About the data preprocess and modeling (EHR, in general)

1. Singh A, Nadkarni G, Gottesman O, Ellis SB, Bottinger EP, Guttag JV. Incorporating temporal EHR data in predictive models for risk stratification of renal function deterioration. J Biomed Inform. 2015 Feb;53:220–8. http://dx.doi.org/10.1016/j.jbi.2014.11.005
2. Meng Y, Speier W, Ong MK, Arnold CW. Bidirectional representation learning from transformers using multimodal electronic health record data to predict depression. IEEE J Biomed Health Inform. 2021 Aug;25(8):3121–9. Available from: http://dx.doi.org/10.1109/JBHI.2021.3063721
3. Solís-García, J., Vega-Márquez, B., Nepomuceno, J. A., Riquelme-Santos, J. C., & Nepomuceno-Chamorro, I. A. (2023). Comparing artificial intelligence strategies for early sepsis detection in the ICU: an experimental study. *Applied Intelligence*, *53*(24), 30691–30705. https://doi.org/10.1007/s10489-023-05124-z
4. Chen, Z., Tan, S., Chajewska, U., Rudin, C., & Caruna, R. (22 Jun--24 Jun 2023). Missing Values and Imputation in Healthcare Data: Can Interpretable Machine Learning Help? In B. J. Mortazavi, T. Sarker, A. Beam, & J. C. Ho (Eds.), *Proceedings of the Conference on Health, Inference, and Learning* (Vol. 209, pp. 86–99). PMLR. https://proceedings.mlr.press/v209/chen23a.html

5. Shashikumar, S. P., Josef, C. S., Sharma, A., & Nemati, S. (2021). DeepAISE - An interpretable and recurrent neural survival model for early prediction of sepsis. *Artificial Intelligence in Medicine*, *113*, 102036. https://doi.org/10.1016/j.artmed.2021.102036
6. Shukla, S. N., & Marlin, B. (2020, October 2). Multi-Time Attention Networks for Irregularly Sampled Time Series. *International Conference on Learning Representations*. https://openreview.net/pdf?id=4c0J6lwQ4_
7. Yang, Z., Mitra, A., Liu, W., Berlowitz, D., & Yu, H. (2023). TransformEHR: transformer-based encoder-decoder generative model to enhance prediction of disease outcomes using electronic health records. *Nature Communications*, *14*(1), 7857. https://doi.org/10.1038/s41467-023-43715-z

Hints of relatively important features

(just for your reference, you can use features not on this list, of course)**:**
Laboratory results from the ICU

Features	**table**	**itemid**
BUN	labevents	51006
Alkaline Phosphatase	labevents	50863
Bilirubin	labevents	50885
Creatinine	labevents	50912
Glucose	labevents	50931
Platelets	labevents	51265
Hemoglobin	labevents	51222

Vital signs from the ICU

Features	**table**	**itemid**
Heart Rate	chartevents	220045
Respiratory Rate	chartevents	220210
Mean Arterial Pressure	chartevents	220052
Temperature	chartevents	223762
Systolic Blood Pressure (SBP)	chartevents	220179

Pre-process done by TAs:

1. re-code the subject ID
2. re-code the admission ID, including admission ID (hadm_id) and ICU stay ID (stay_id)
3. edit the original "date"

Filter 30% of the original population data from all the tables you need in the Hospital and ICU modules to complete the following homework.

The same data can be applied to both tasks.

1. **Sepsis prediction task**

	Original Population		**Selected for AI Capstone**		**Percentage (B) / (A)**
	N (A)	**Percent**	**N (B)**	**Percent**	
Sepsis cases	25,323	38%	7,500	39%	30%
Non-Sepsis cases	40,916	62%	11,495	61%	28%
Total	66,239	-	18,995	-	29%

2. **Mortality prediction task**

	Original Population		**Selected for AI Capstone**		**Percentage (B) / (A)**
	N (A)	**Percent**	**N (B)**	**Percent**	
Non-survivors	6,966	11%	3,648	19%	52%
Survivors	59,273	89%	15,347	81%	26%
Total	66,239	-	18,995	-	29%