Juhua Hu

Assistant Professor of Computer Science and Systems
Director of Center for Data Science
University of Washington, Tacoma
https://faculty.washington.edu/juhuah/

Research Focus:
Deep representation learning
Rare event prediction
Long-tailed learning
User-friendly multiple clustering

NSF

UW Medicine
HARBORVIEW
MEDICAL CENTER

advata

KenSci

Infoblox

W CENTER FOR DATA SCIENCE
UNIVERSITY *of* WASHINGTON | TACOMA
School of Engineering & Technology

W

Rare Event Prediction

Events of interest are often very very infrequent

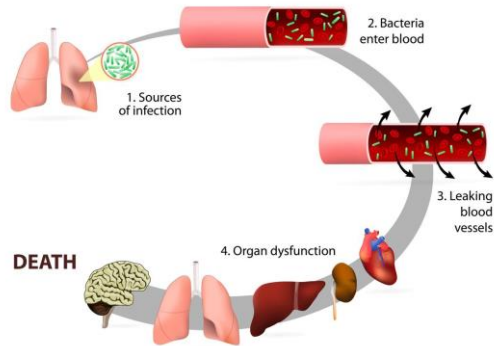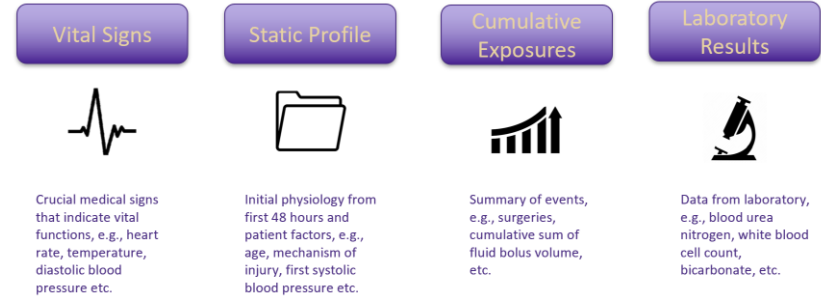**<0.9% patients developed Septic Shock on a daily basis** [data collected 2012-2019]

Figure 1. Sepsis process. Adapted from 'Disease profile: Sepsis' by Nicholas Parry, 2019, healthissuesindia.com.

Multi-modal Representation Learning

Vital Signs	Static Profile	Cumulative Exposures	Laboratory Results
Crucial medical signs that indicate vital functions, e.g., heart rate, temperature, diastolic blood pressure etc.	Initial physiology from first 48 hours and patient factors, e.g., age, mechanism of injury, first systolic blood pressure etc.	Summary of events, e.g., surgeries, cumulative sum of fluid bolus volume, etc.	Data from laboratory, e.g., blood urea nitrogen, white blood cell count, bicarbonate, etc.

T. Stewart, K. Stern, G. O'Keefe, A. Teredesai, and J. Hu. NPRL: Nightly Profile Representation Learning for Early Sepsis Onset Prediction in ICU Trauma Patients. In: Proceedings of the IEEE International Conference on Big Data (BigData'23), Sorrento, Italy, 2023, pp.1843-1852.

K. Ewig, X. Lin, T. Stewart, K. Stern, G. O'Keefe, A. Teredesai, and J. Hu. Multi-Subset Approach to Early Sepsis Prediction. In: Proceedings of the 2023 Congress in Computer Science, Computer Engineering, & Applied Computing (CSCE'23) - The 9th International Conference on Health Informatics and Medical Systems (HIMS'23), Las Vegas, NV, 2023, pp.1335-1341.
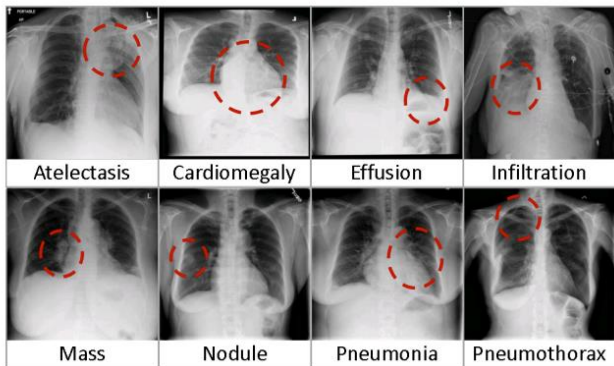
A. Teredesai, S. Huang, T. Stewart, J. Hu, A. Thakker, K. Stern, and G. O'Keefe. Sub-Sequence Graph Representation Learning on High Variability Data for Dynamic Risk Prediction in Critical Care. In: Proceedings of the IEEE International Conference on Big Data (BigData'22), Osaka, Japan, 2022, pp.2082-2092.

T. Stewart, B. Yu, A. Nascimento, and J. Hu. Enhancing Peak Network Traffic Prediction via Time-series Decomposition. arXiv preprint arXiv:2303.13529, 2023.
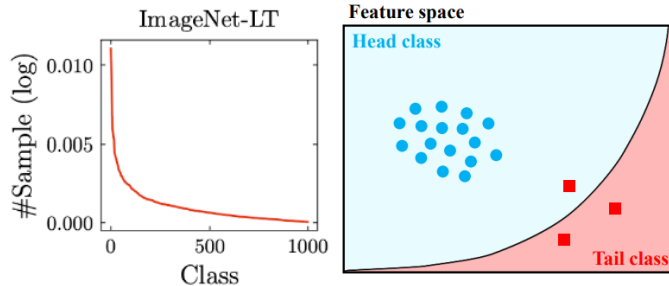
B. Yu, G. Graciani, A. Nascimento, and J. Hu. Cost-adaptive Neural Networks for Peak Volume Prediction with EMM Filtering. In: Proceedings of the IEEE International Conference on Big Data (BigData'19), Los Angeles, CA, 2019, pp.4208-4213.
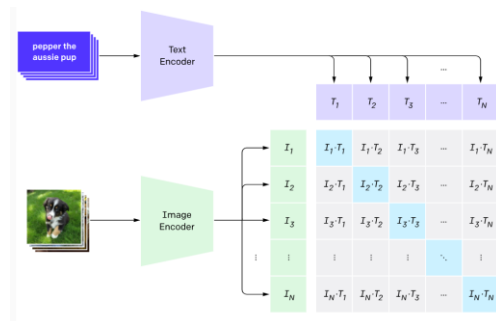
Long-tailed Learning

ImageNet-LT

Feature space

Head class

Tail class

## Language-Supervision

https://openai.com/index/clip/

Figure 1: Text-guided mixup allows semantically similar classes to be mixed more frequently (e.g., 'tiger' as a tail class is stretched to the directions of 'leopard' and 'cat', where 'cat' as a head class can help).

Atelectasis Cardiomegaly Effusion Infiltration

Mass Nodule Pneumonia Pneumothorax

> **Limited data for medical images**
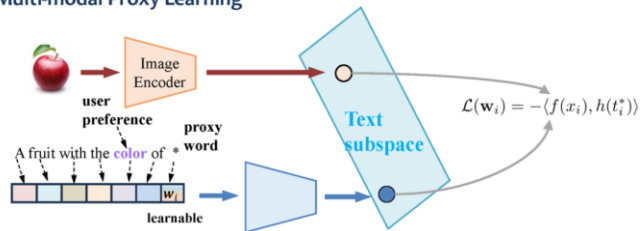> **Severely limited data for tail classes**

R. Franklin, J. Yao, D. Zhong, Q. Qian, J. Hu. Text-Guided Mixup Towards Long-Tailed Image Categorization. To appear in: Proceedings of the 35th British Machine Vision Conference (BMVC'24), Glasgow, UK, 2024.
J. Wang, Y. Xu, J. Hu, M. Yan, J. Sang, Q. Qian. Improved Visual Fine-tuning with Natural Language Supervision. In: Proceedings of the International Conference on Computer Vision (ICCV'23), Paris, France, 2023, pp.11865-11875.

User-friendly Multiple Clustering
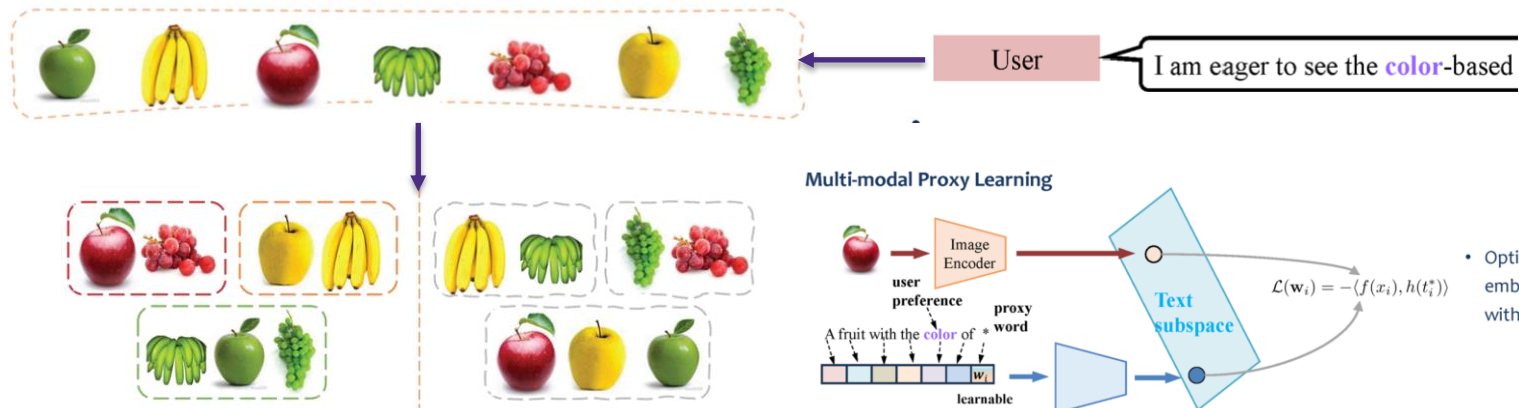
J. Yao, Q. Qian, and J. Hu. Multi-Modal Proxy Learning Towards Personalized Visual Multiple Clustering. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR'24), Seattle, WA, 2024, pp.14066-14075.
J. Yao and J. Hu. Dual-disentangled Deep Multiple Clustering. In: Proceedings of the SIAM International Conference on Data Mining (SDM'24), Houston, TX, 2024, pp.679-687.
J. Yao, E. Liu, M. Rashid, and J. Hu. AugDMC: Data Augmentation Guided Deep Multiple Clustering. Procedia Computer Science, 222 (2023): 571-580.