Admission Interview

MPU PhD in Computer Applied Technology

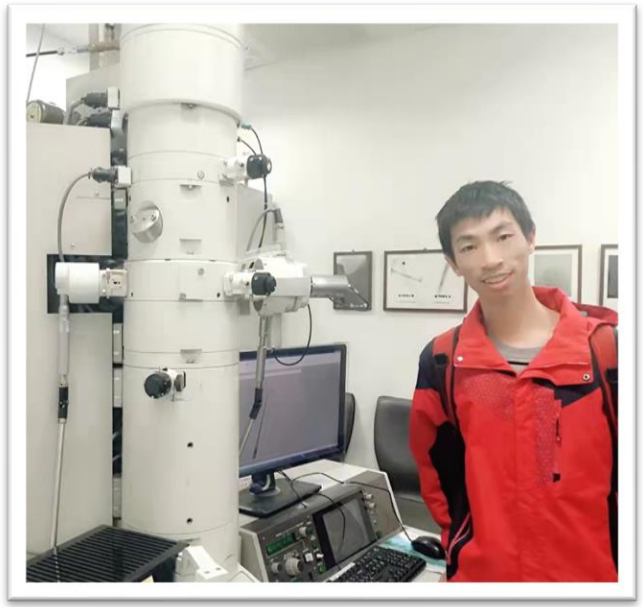
Applicant:

Chen Tianhao No. 430127

Table of Contents

- **Self-Introduction** (5min)
  1. Education Background
  2. Purpose of Study
  3. Research Area
  4. Skills Set
- **Research Proposal** (10min)
  - Novel AI for Biomedical Studies
    - **Part 1:** Stable Prediction Model in Medicine
    - **Part 2:** Biomarkers and Drug Discoveries

Education Background



At a Solid-State Physics Lab, **National Central University**, Taiwan, 2017

澳門大學
UNIVERSIDADE DE MACAU
UNIVERSITY OF MACAU

哈爾濱工程大學
HARBIN ENGINEERING UNIVERSITY

MSc, Data Science (Precision Medicine), **University of Macau**

- **Supervisor:** Prof. Xiaohua Douglas Zhang;
- **Thesis:** *Building a Shiny Web Application to Promote Data Management in a Multi-Center Project with Rodent Shrapnel Model* (full-text).

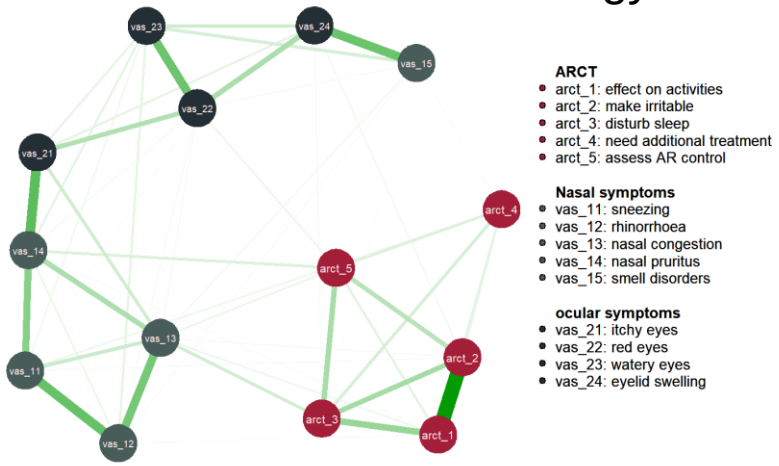
BSc, Materials Science and Engineering, **Harbin Engineering University (Project 211)**

- **Thesis:** *LabVIEW System for Acquisition and Analysis of Digital Welding Signals*
- Teaching Assistant, 201411041 College Physics I at Spring 2017

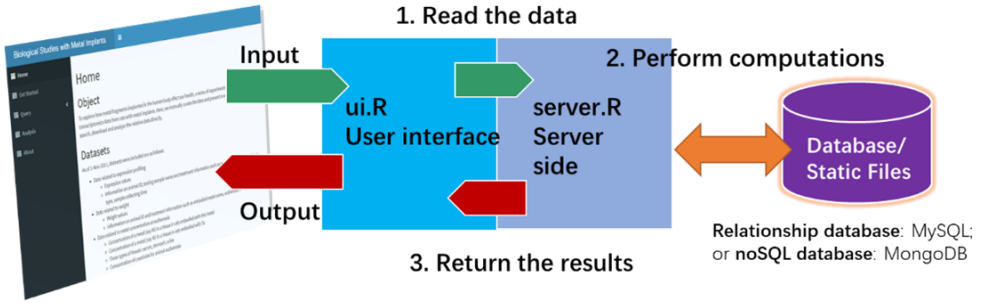
Purpose of Study: To Be An **Expert**

- Trained as a MSc student: have to be a **Dabbler**
  - Supporting various biomedical project
  - Investigating multiple scientific questions
- Trained as a PhD students: have to be an **Expert**
  - Push the boundaries Independently
  - Contributing Uniquely

Allergy Diseases

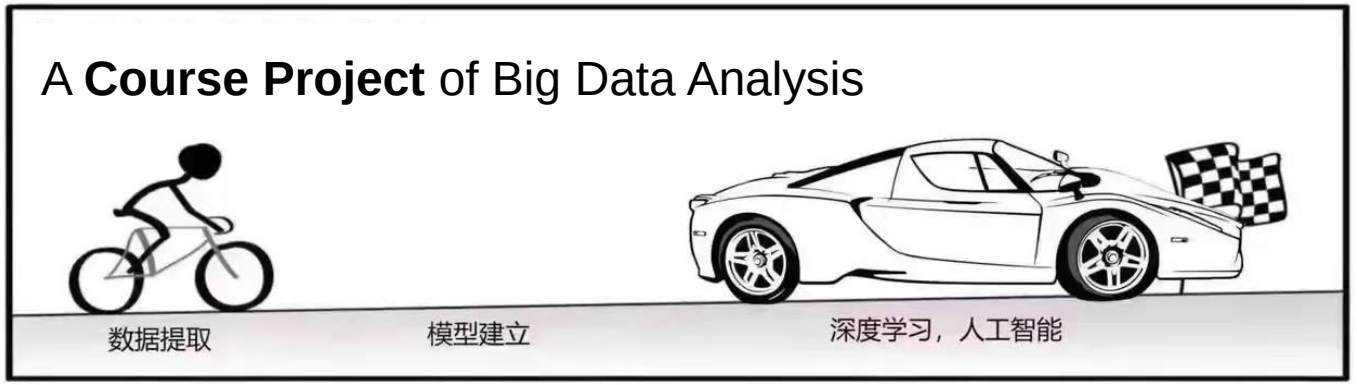


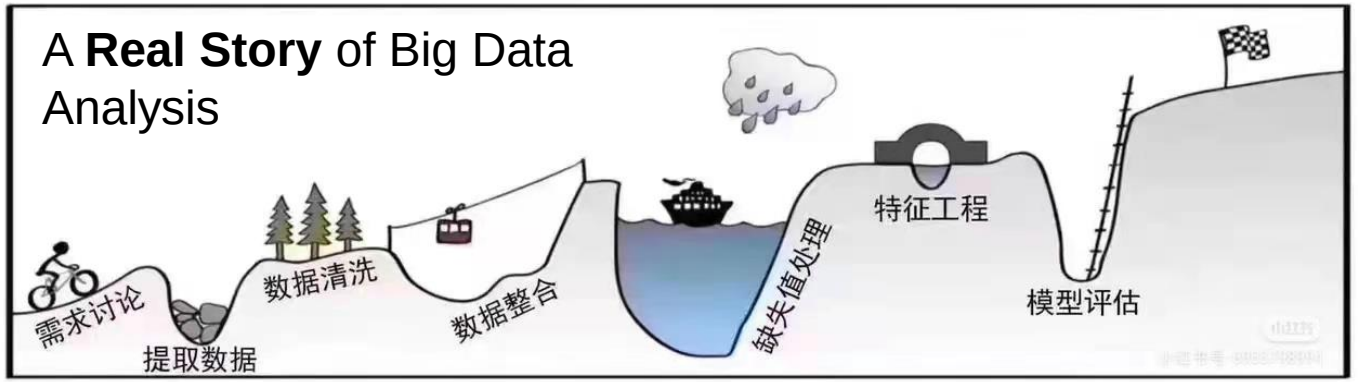
Shrapnel Injuries



Purpose of Study: To Be An **Expert**

- Properties of a good doctoral student in Computer domain:
  - Balancing real scientific questions and practical considerations.
  - Balancing independent work and teamwork.

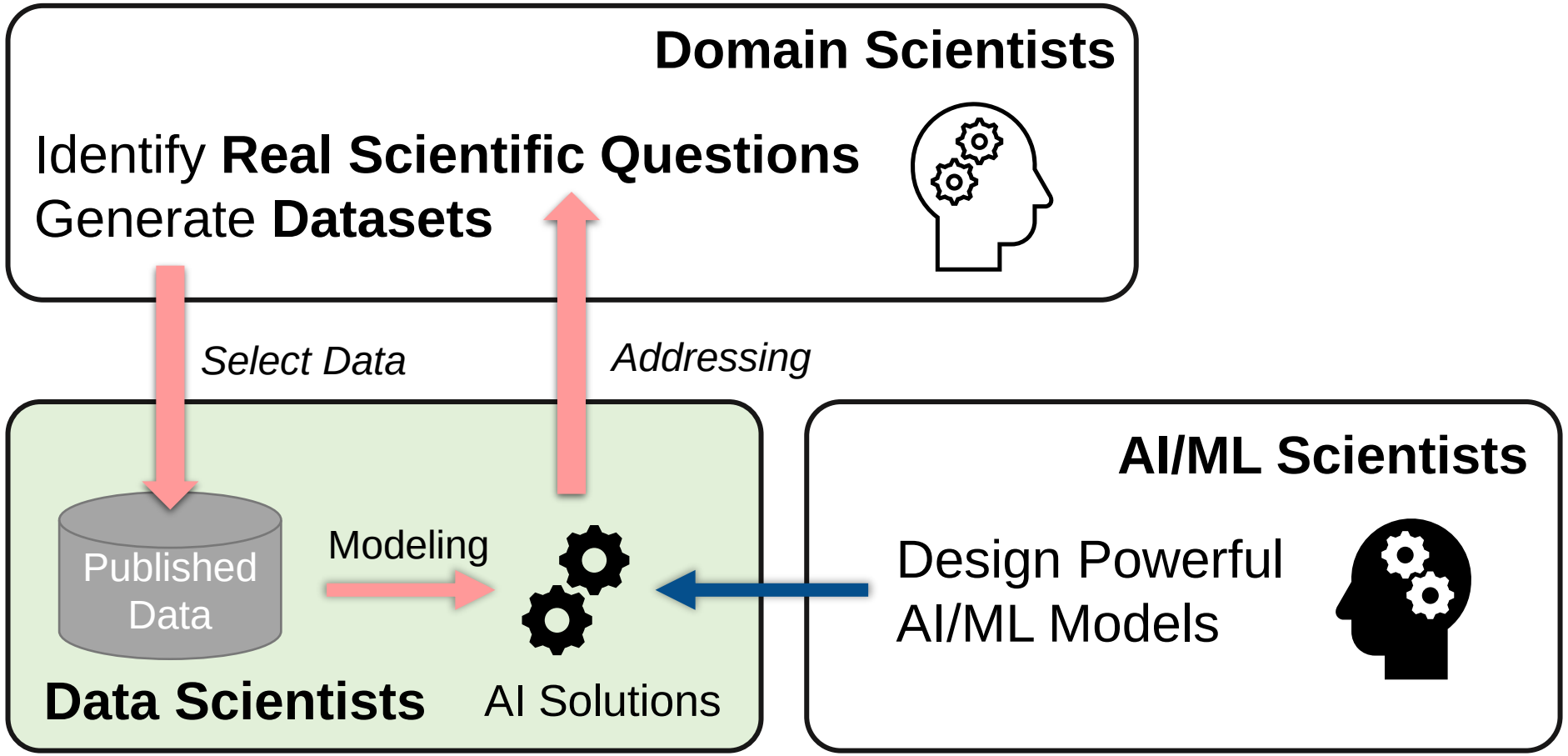




Purpose of Study: To Be An **Expert**

- Properties of a good doctoral student in Computer domain:
  - Balancing real scientific questions and practical considerations.
  - Balancing independent work and teamwork.

Research Area: **Real** AI Applications



Skills Set: As a Data Scientist

Reproducible Data Analysis

Development Tools

ML Tools

Math for Data Science

Calculus ***Linear Algebra***

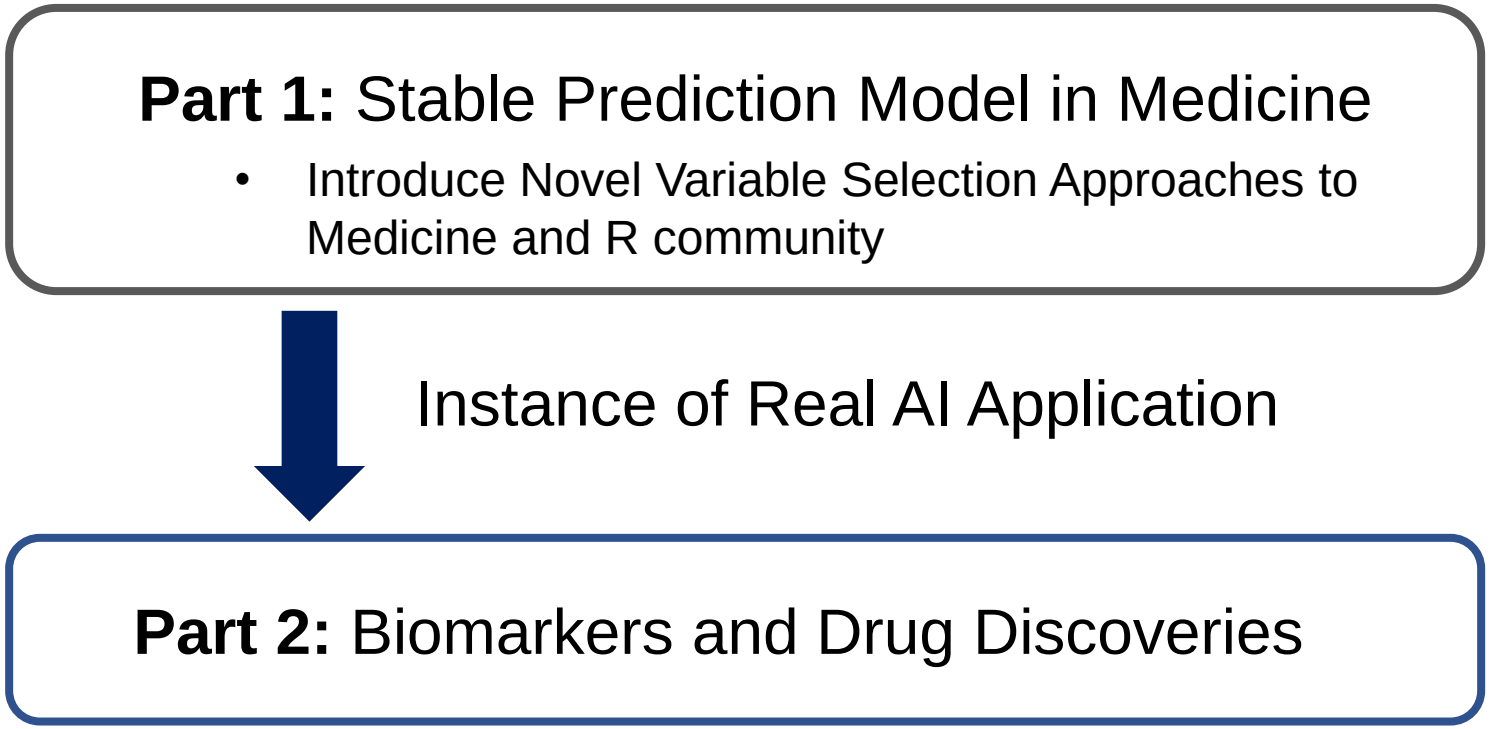
Probability ***Statistical Inference***

Domain Knowledge

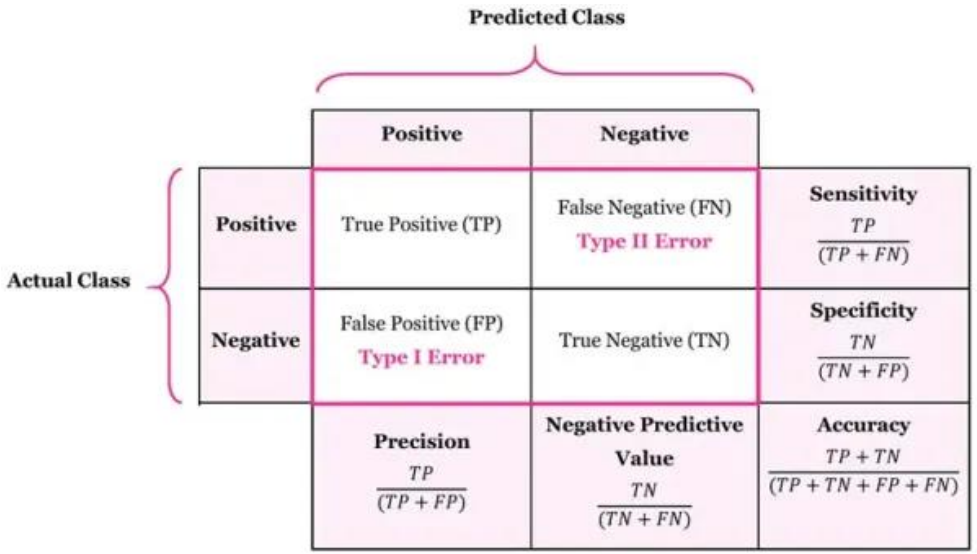
Allergy
Questionnaire
Bioinformatics
Epidemiology
Network Analysis
Clinical Study
Causal
Omics
Inference

Research Proposal (10min)

- Novel AI for Biomedical Studies



Part 1: Stable Prediction Model in Medicine



1. Background
2. Key Challenges
3. Research Gap and Objectives
4. Research Design

Background: Precision Medicine and AI

In big data era, *precision medicine* takes individual patient's specific information including **biomedical big data** (e.g., genomics data) into account to

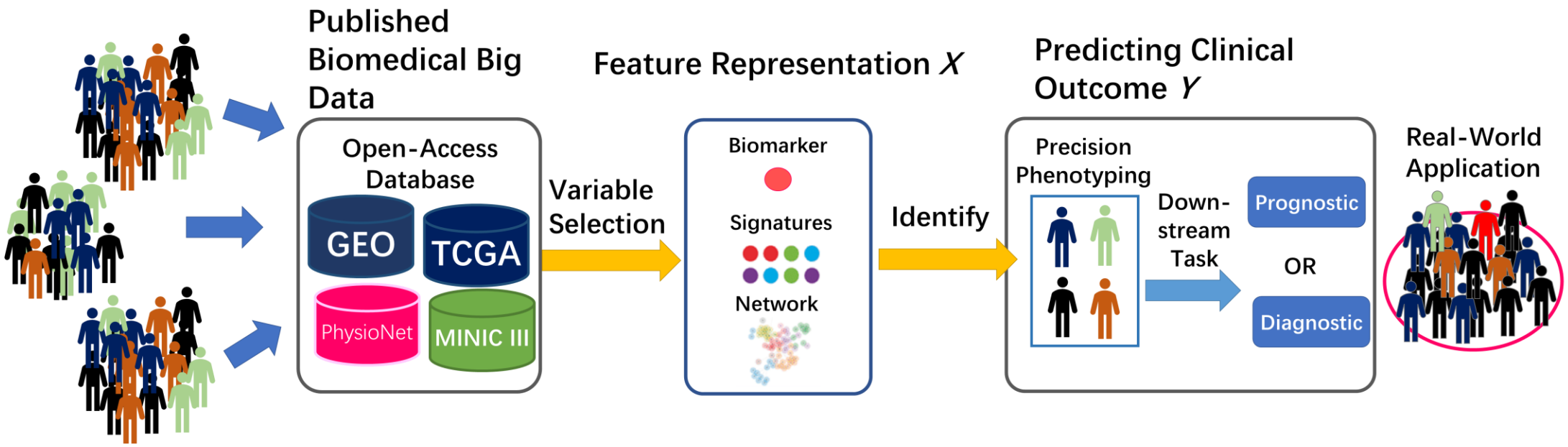
1. Predict the clinical outcome:
   - Prediction Model/Supervised Learning/Regression

$$\mathrm{E}(y|\boldsymbol{X})$$

2. Provide the optimal treatment strategy
   - Reinforcement Learning (with Potential Outcome Framework)

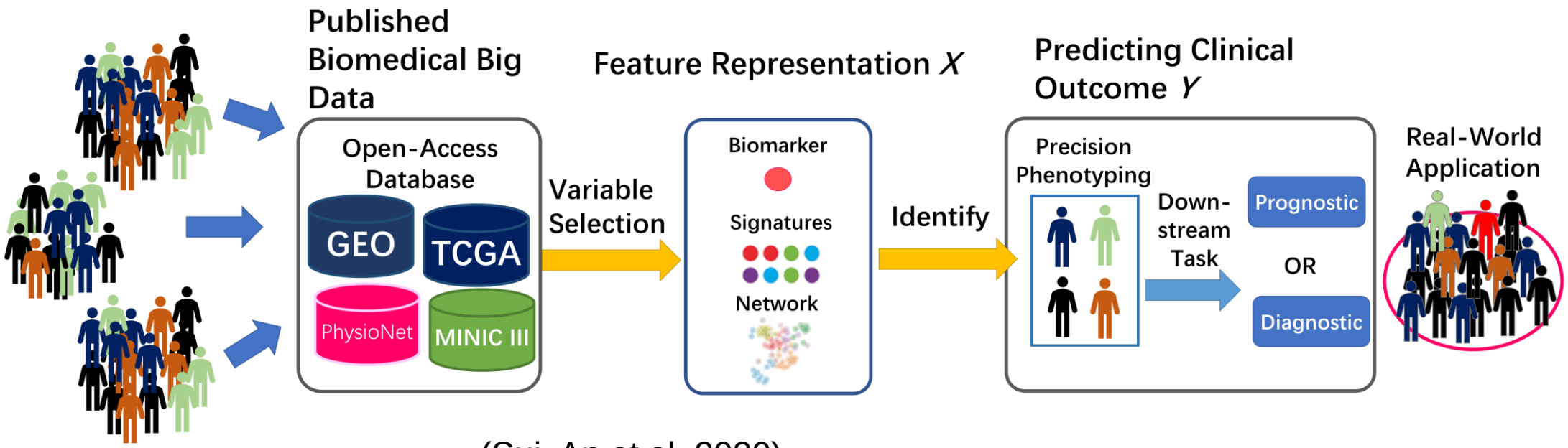
$$\mathrm{E}(y|do(\boldsymbol{T}),\boldsymbol{X})$$

Background: Prediction Model in Medicine

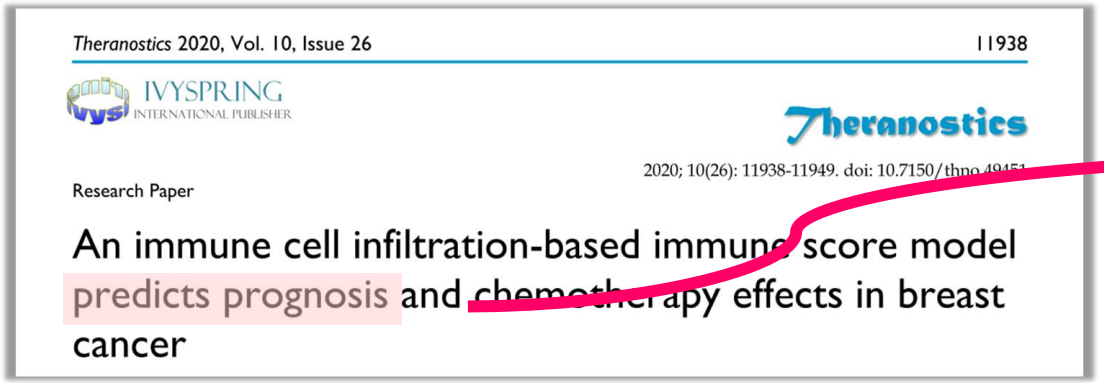


A typical supervised learning (regression) task: $\mathrm{E}(y|\boldsymbol{X})$

Background: Prediction Model in Medicine



(Sui, An et al. 2020)



Theranostics 2020, Vol. 10, Issue 26 11938

IVYSPRING INTERNATIONAL PUBLISHER

Theranostics

2020; 10(26): 11938-11949. doi: 10.7150/thno.49451

Research Paper

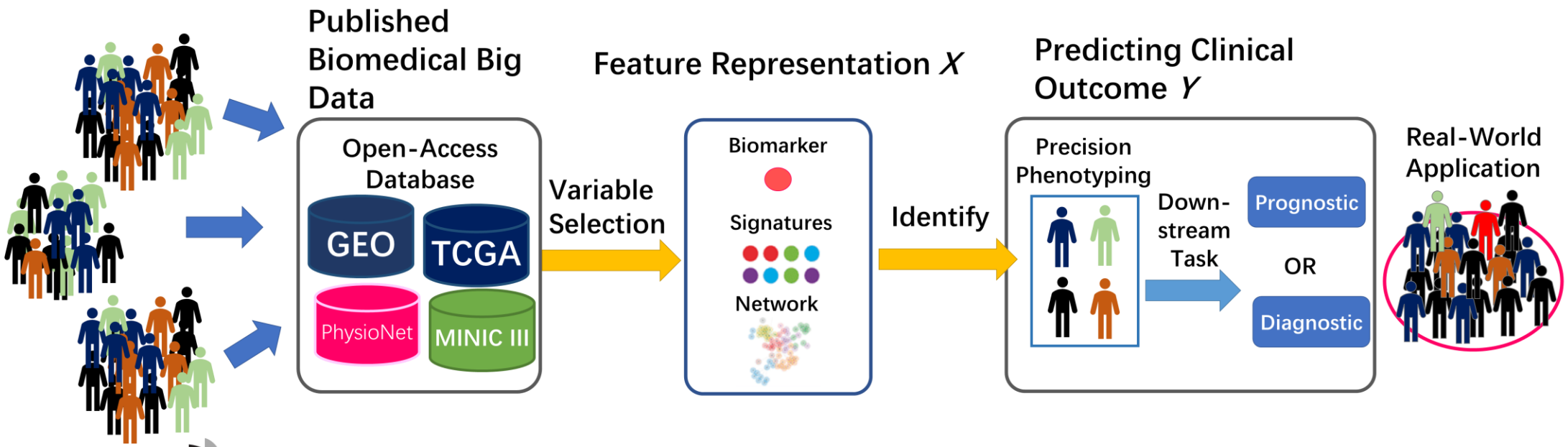
An immune cell infiltration-based immune score model predicts prognosis and chemotherapy effects in breast cancer

Impact Factor: 11.556

LASSO + COX Regression

Regression task $\mathrm{E}(y|\boldsymbol{X})$

Background: Prediction Model in Medicine



PubMed

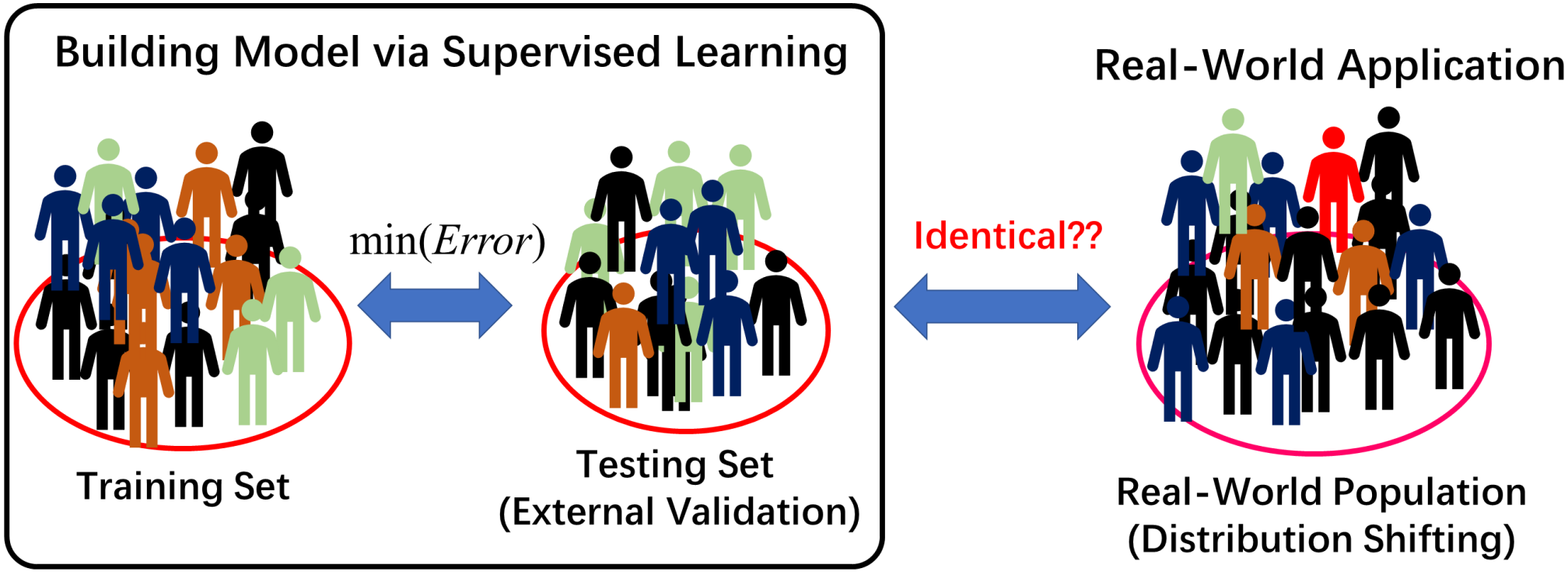
Query	Results	Time
Search: **(((prognostic) OR (diagnosis) OR (predict)) AND (LASSO)) AND (gene)** Filters: **from 2017 - 2022**	2,602	22:04:37

Got 2602 publications: Hot Topic!

LASSO + Prediction Task

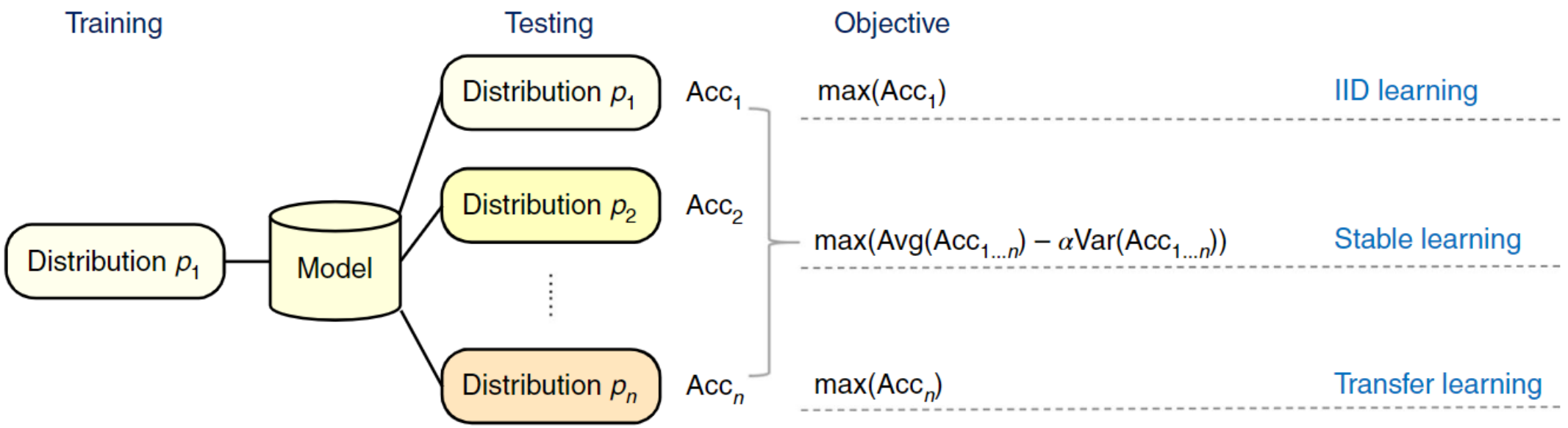
Regression task $\mathrm{E}(y|\boldsymbol{X})$

Key Challenge: Dose I.I.D. Hold?



Objective: to select "stable" features with **biological significance**.
Simulation study: regularizers may estimate larger coefficients on the "unstable" features. (Kuang, Xiong et al. 2020)

Key Challenge: Stable Prediction Task

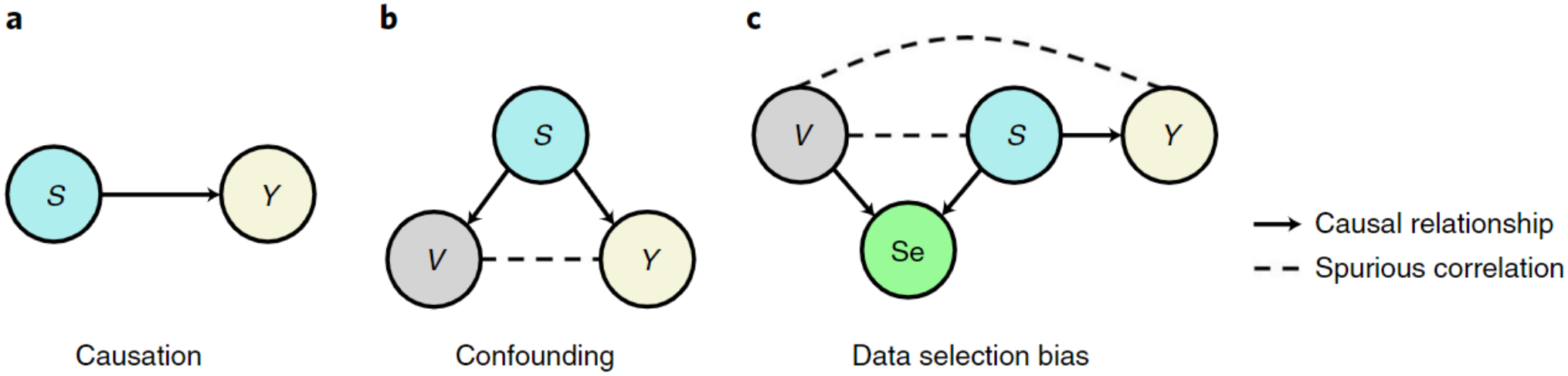


In stable learning (Kuang, Cui et al. 2018, Cui and Athey 2022)**:**

- Do not assume the availability of multiple environments in training data, while there are multiple distributions in the testing dataset, requiring a higher standard for a model's generalization ability.
- **Optimize objective:** Maximize the average accuracy and minimize the variance of it across different distributions.

Key Challenge: Stable Prediction Task

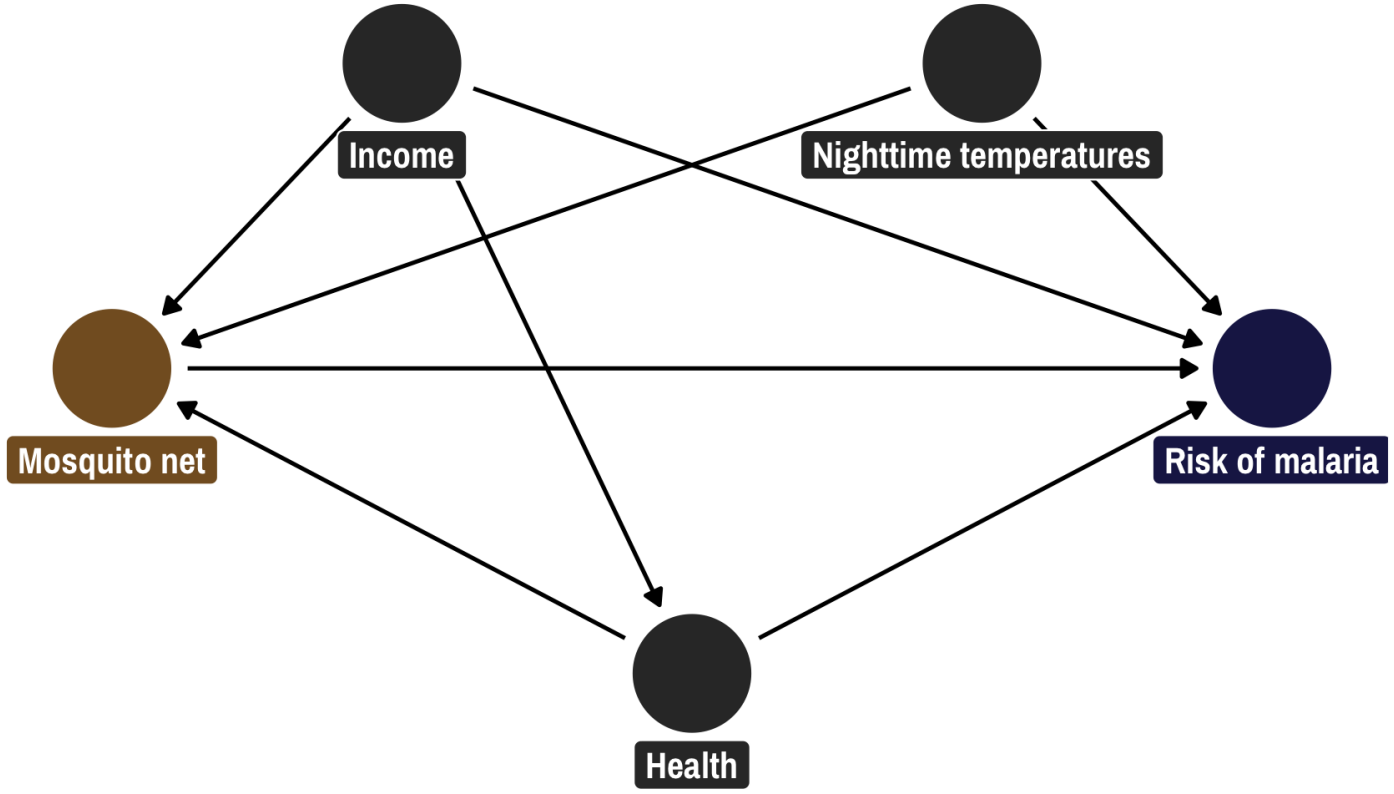
- **To be "Stable": Identify variables with biological significance**



Three ways of generating correlations. (Cui and Athey 2022)
Y: outcome, *S*: "Stable" variable, *V*: "Unstable" variable (covariate), Se: selection bias

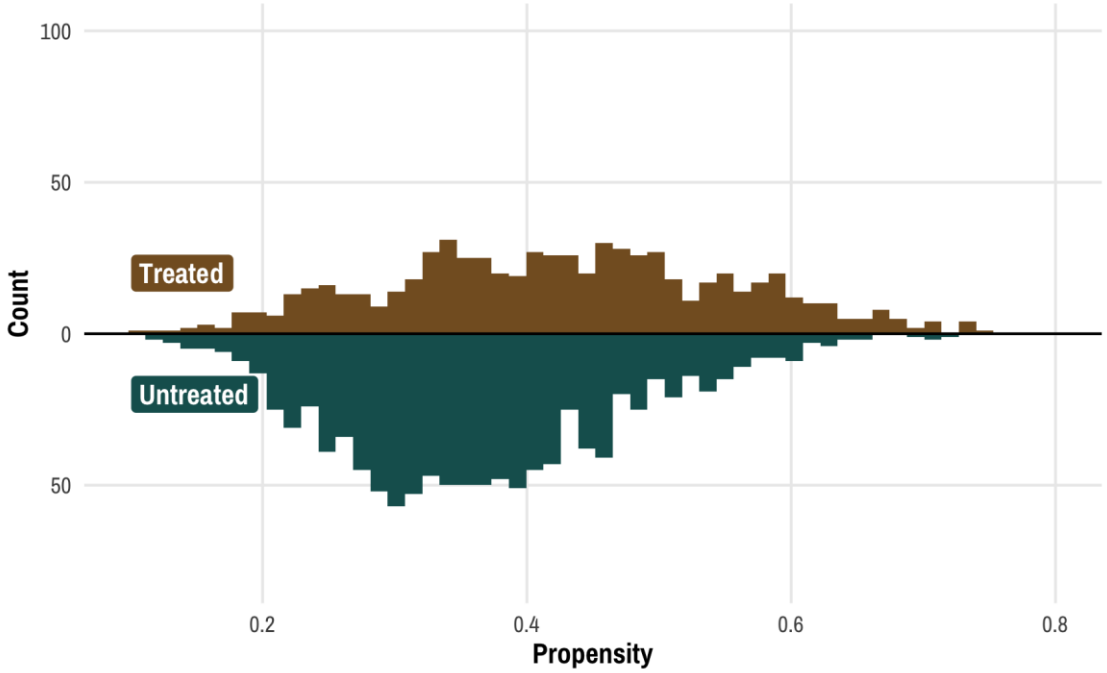
Key Challenge: Stable Prediction Task

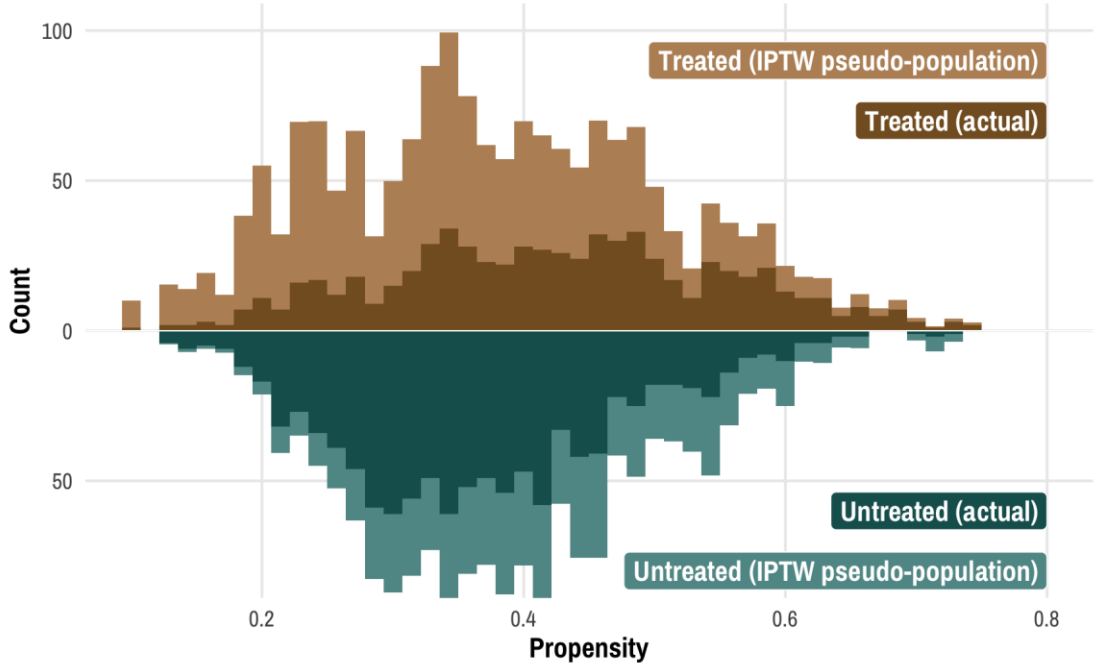
- An Instance: Covariate Balancing Strategies



Key Challenge: Stable Prediction Task

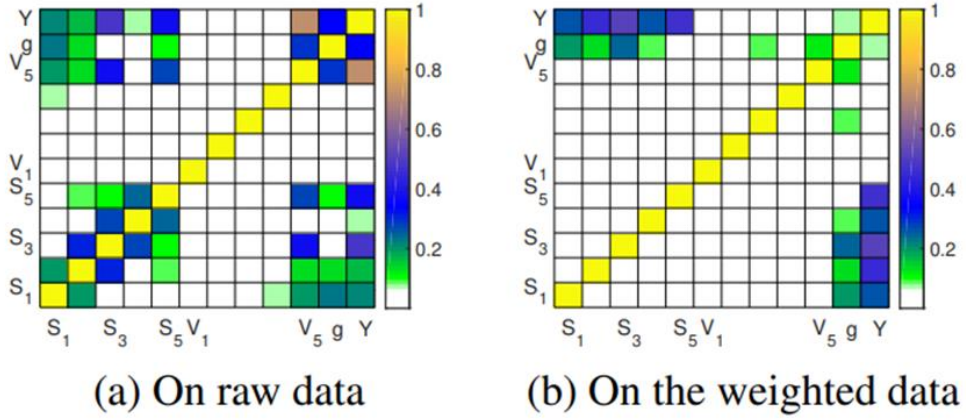
- Covariate Balancing Strategies





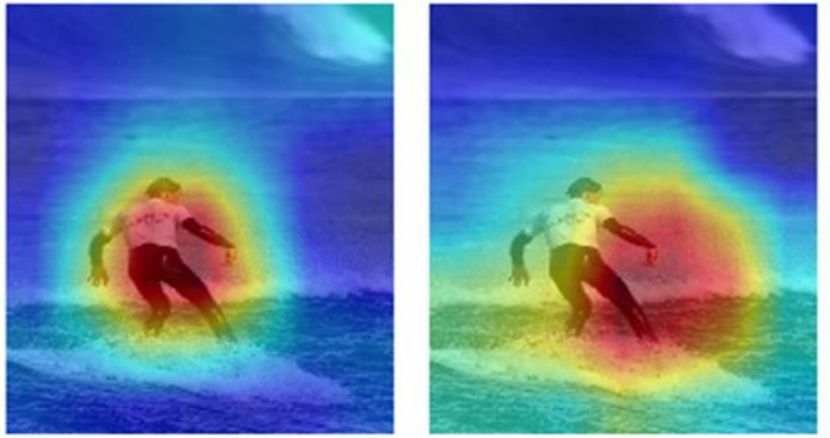
Key Challenge: Stable Prediction Task

- Latest advances: **Globally Balancing** for high-dimensional big data (e.g., images data)



(a) On raw data (b) On the weighted data

Globally Balancing for Variables Decorrelation as Data Pretreatment Approaches (Kuang, Xiong et al. 2020, Shen, Cui et al. 2020)



(a) DCDAN (b) Resnet-50

Globally Balancing plus Deep Learning for Domain Adaption (Zhang, Cui et al. 2021)

Research Gap and Objectives

- **Research Gap:**
  - Several approaches are proposed for stable prediction tasks (Kuang, Cui et al. 2018, Kuang, Cui et al. 2019, Kuang, Xiong et al. 2020, Ren, Huang et al. 2020, Shen, Cui et al. 2020, Xu, Cui et al. 2020, Cui and Athey 2022), but none of them applied in the biomedicine domain.
- **Objectives:**
  1. Adaptive stable learning to the clinical prediction models
  2. Consider **variable selection issues** at first
  3. “Translate” the latest algorithms into R packages

Research Design

- **Simulation Study**
  - SOTA: Sample Reweighted Decorrelation Operator (SRDO) (Shen et al., 2020) as data pretreatment method, plus linear regression models (cox and logistic) as downstream tasks.
  - Baseline: OLS, Elastic Net/LASSO + linear models
  - Evaluation: distributions shifting should be considered.
- **Case Study:**
  - Biomarkers Identification for Comorbid Depression and Obesity
- **Other contributions:**
  1. Design criteria for variable selection and sample size.
  2. Building R package supporting R-powered biomedical studies.

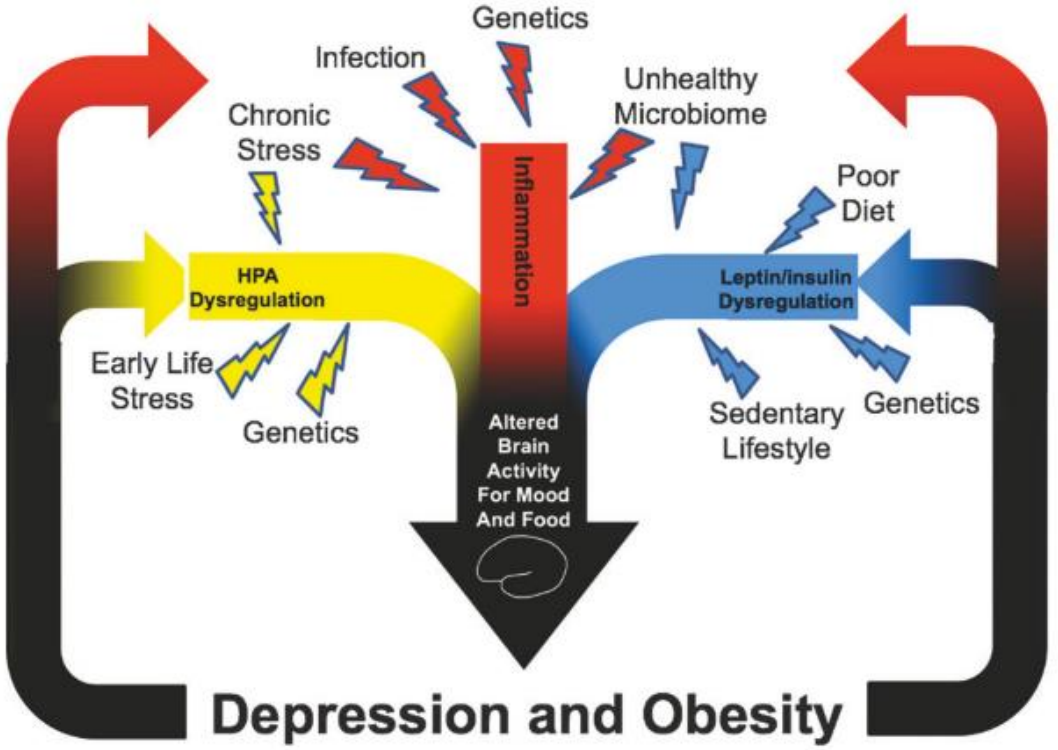
Part 2: Biomarkers and Drug Discoveries

- Biomarkers and Drug Discoveries for Comorbid Depression and Obesity: Multi-Omics Approaches

1. Background
2. Research Design

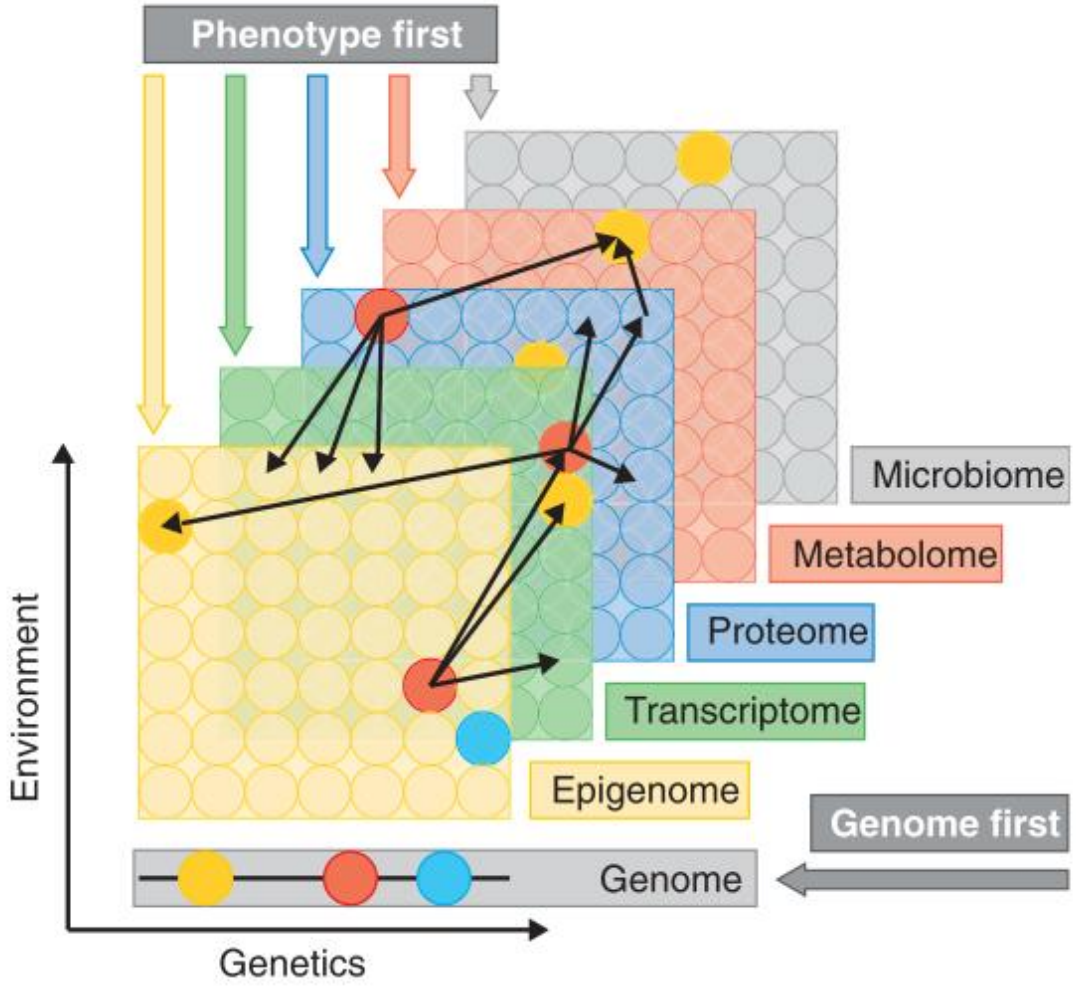
Background: Comorbid Depression and Obesity

- Evidence reveals a bidirectional relationship between depression and obesity.
- **Research Gap**
  1. integrating multi-level pathways for uncovering key molecular mechanisms
  2. handling the heterogeneity of patients.



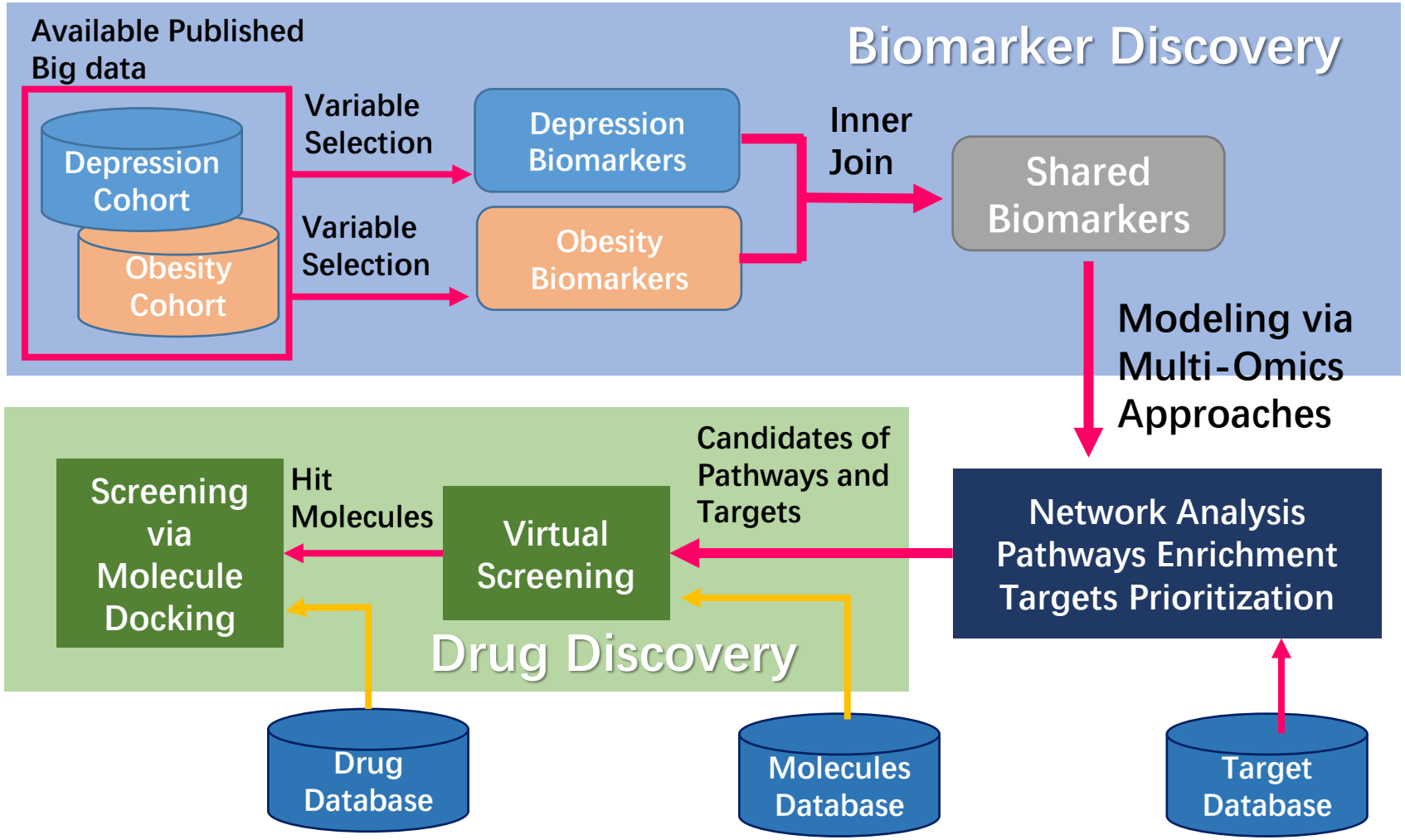
The shared biological mechanisms between the two diseases (Milaneschi, Simmons et al. 2019)

Background: Multi-Omics Approaches



- **Omics:** Biological Informatics Objects
- Multiple omics approaches attempt to integrate information across multiple levels of the biological systems. (Hasin, Seldin et al. 2017)

Research Design



Summary

- Introduce stable prediction approaches for the first time, which allows selecting “stable” variables, to promote the real application of clinical prediction models.
- Illustrate how stable prediction methods and other AI approaches contribute to drug discoveries for comorbid depression and obesity, addressing the real biological questions.
- My proposal could contribute uniquely to the biomedical community without the need for clinical resources.

Reference

Cui, P., & Athey, S. (2022). Stable learning establishes some common ground between causal inference and machine learning. *Nature Machine Intelligence, 4*(2), 110-115. doi:10.1038/s42256-022-00445-z

Faith, M. S., Butryn, M., Wadden, T. A., Fabricatore, A., Nguyen, A. M., & Heymsfield, S. B. (2011). Evidence for prospective associations among depression and obesity in population-based studies. *Obesity Reviews, 12*(5), e438-e453. doi:10.1111/j.1467-789x.2010.00843.x

Hasin, Y., Seldin, M., & Lusis, A. (2017). Multi-omics approaches to disease. *Genome Biology, 18*(1). doi:10.1186/s13059-017-1215-1

Kuang, K., Cui, P., Athey, S., Xiong, R., & Li, B. (2018). *Stable Prediction across Unknown Environments*. Paper presented at the Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, London, United Kingdom. https://doi.org/10.1145/3219819.3220082

Kuang, K., Xiong, R., Cui, P., Athey, S., & Li, B. (2020). Stable Prediction with Model Misspecification and Agnostic Distribution Shift. *Proceedings of the AAAI Conference on Artificial Intelligence, 34*(04), 4485-4492. doi:10.1609/aaai.v34i04.5876

Lee, Y.-H., Bang, H., & Kim, D. J. (2016). How to Establish Clinical Prediction Models. *Endocrinology and Metabolism, 31*(1), 38. doi:10.3803/enm.2016.31.1.38

Milaneschi, Y., Simmons, W. K., van Rossum, E. F. C., & Penninx, B. W. J. H. (2019). Depression and obesity: evidence of shared biological mechanisms. *Molecular Psychiatry, 24*(1), 18-33. doi:10.1038/s41380-018-0017-5

Shen, Z., Cui, P., Zhang, T., & Kunag, K. (2020). Stable Learning via Sample Reweighting. *Proceedings of the AAAI Conference on Artificial Intelligence, 34*(04), 5692-5699. doi:10.1609/aaai.v34i04.6024

Subramanian, I., Verma, S., Kumar, S., Jere, A., & Anamika, K. (2020). Multi-omics Data Integration, Interpretation, and Its Application. *Bioinformatics and Biology Insights, 14*, 117793221989905. doi:10.1177/1177932219899051

Sui, S., An, X., Xu, C., Li, Z., Hua, Y., Huang, G., . . . Li, M. (2020). An immune cell infiltration-based immune score model predicts prognosis and chemotherapy effects in breast cancer. *Theranostics, 10*(26), 11938-11949. doi:10.7150/thno.49451

Xu, R., Cui, P., Kuang, K., Li, B., Zhou, L., Shen, Z., & Cui, W. (2020, 2020-08-23). *Algorithmic Decision Making with Conditional Fairness.* Paper presented at the Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining.

Zhang, X., Cui, P., Xu, R., Zhou, L., He, Y., & Shen, Z. (2021, 20-25 June 2021). *Deep Stable Learning for Out-Of-Distribution Generalization.* Paper presented at the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR).

Zhao, X., & Cui, L. (2020). A robust six-miRNA prognostic signature for head and neck squamous cell carcinoma. *Journal of Cellular Physiology, 235*(11), 8799-8811. doi:10.1002/jcp.29723

Discuss and Q&A (15min)

- Many thanks for this opportunity!