Statistical Machine Learning and Data Fusion with Application in Healthcare Systems

Yu (Adam) Ding

Systems Science and Industrial Engineering

Binghamton University State University of New York (SUNY)

Outline

- **Introduction**
  - Statistical machine learning and Data fusion for Precise Medicine and Public Health
- **Disease diagnosis and phenotyping**
  - Unsupervised learning of multi-faced medical data for phenotype discovery
- **Privacy-preserving telemedicine**
  - Federated learning of functional data for privacy-preserving telemedicine
- **Conclusions and Future Works**

Outline

- **Introduction**
  - Statistical machine learning and Data fusion for Precise Medicine and Public Health
- **Disease diagnosis and phenotyping**
  - Unsupervised learning of multi-faced medical data for phenotype discovery
- **Privacy-preserving telemedicine**
  - Federated learning of functional data for privacy-preserving telemedicine
- **Conclusions and Future Works**

Introduction

- Currently, 800 million people spend at least 10 percent of their household budgets on health expenses for themselves, a sick child or other family member. For almost 100 million people these expenses are high enough to push them into extreme poverty, forcing them to survive on just $1.90 or less a day.

  –World Health Organization

- The digital divide is now a matter of life and death for people who are unable to access essential health-care information. It is threatening to become the new face of inequality, reinforcing the social and economic disadvantages suffered by women and girls, people with disabilities and minorities of all kinds.

  - UN Secretary-General António Guterres in the pandemic period

Introduction

- How to implement statistical machine learning techniques to meet the general benefiting objective under the constraints of limited healthcare capacity and costs?
- Answer: Precise Medicine and Public Health by data mining on the patients' health records
- Challenge:

1. Privacy and Security
2. Mixed types, numerical, ordinal, categorical, functional, etc.
3. Lack of a framework to include and distill knowledges from different resources.

Outline

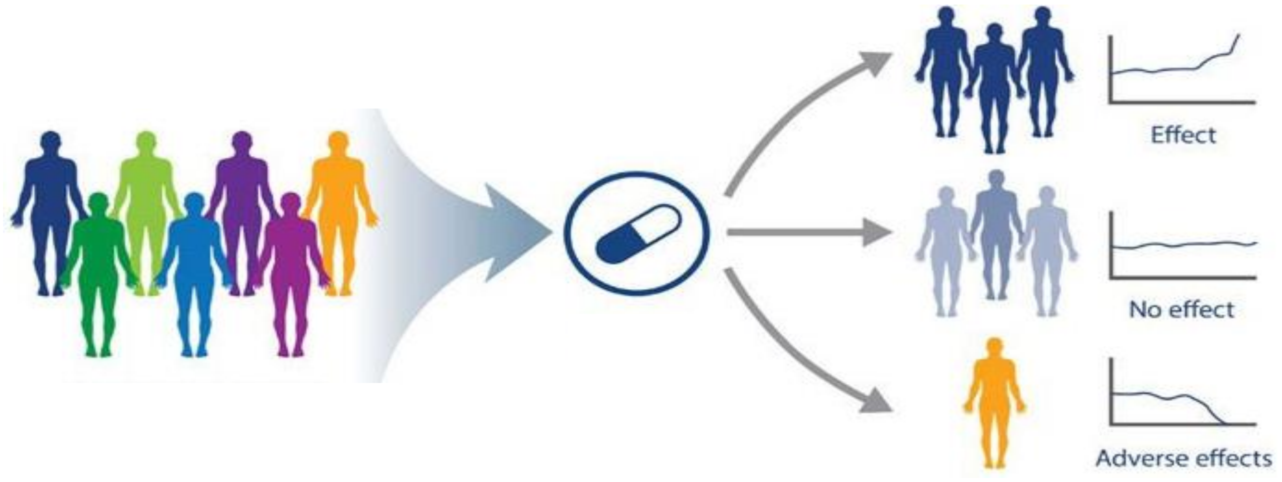
- Introduction
  - Statistical machine learning and Data fusion for Precise Medicine and Public Health
- **Disease diagnosis and phenotyping**
  - Unsupervised learning of multi-faced medical data for phenotype discovery
- **Privacy-preserving telemedicine**
  - Federated learning of functional data for privacy-preserving telemedicine
- **Conclusions and Future Works**

A novel sparse generalized structural equation modeling with structured sparsity for subgroup discovery from multi-modal mixed-type data

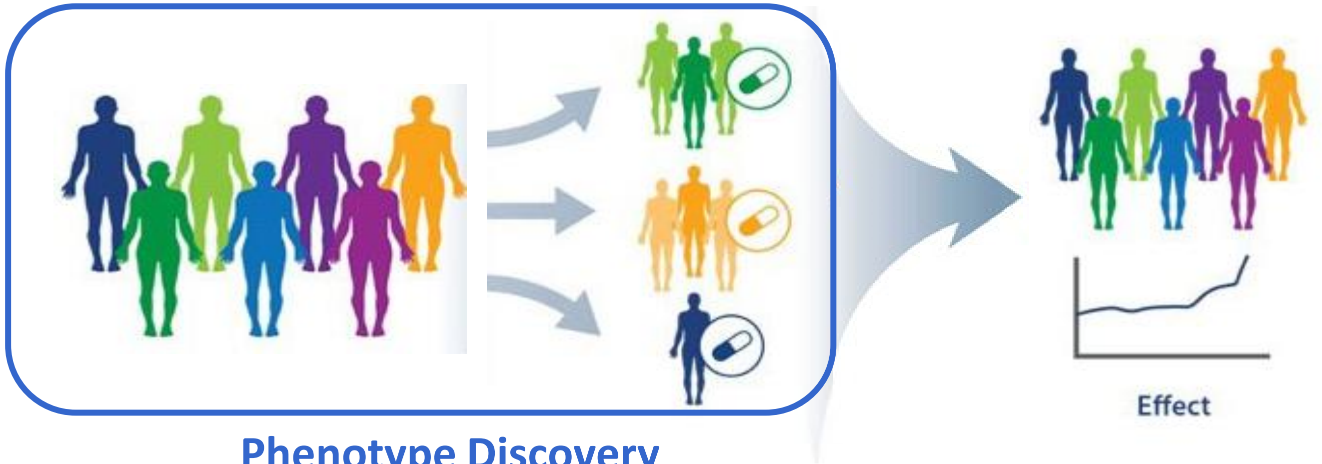
Yu Ding, Bing Si

Precision Medicine

Traditional Medicine
One Treatment Fits All



Precision Medicine

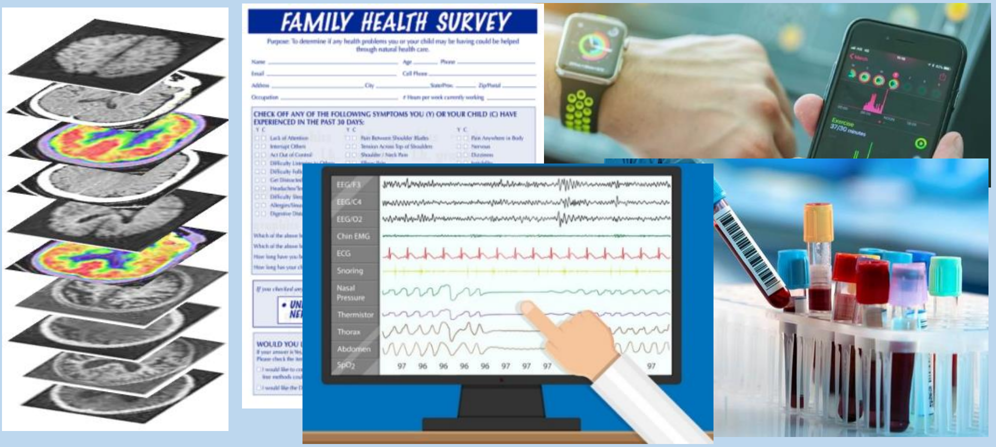


Phenotype Discovery

Precision Medicine

Multi-modal Health Data

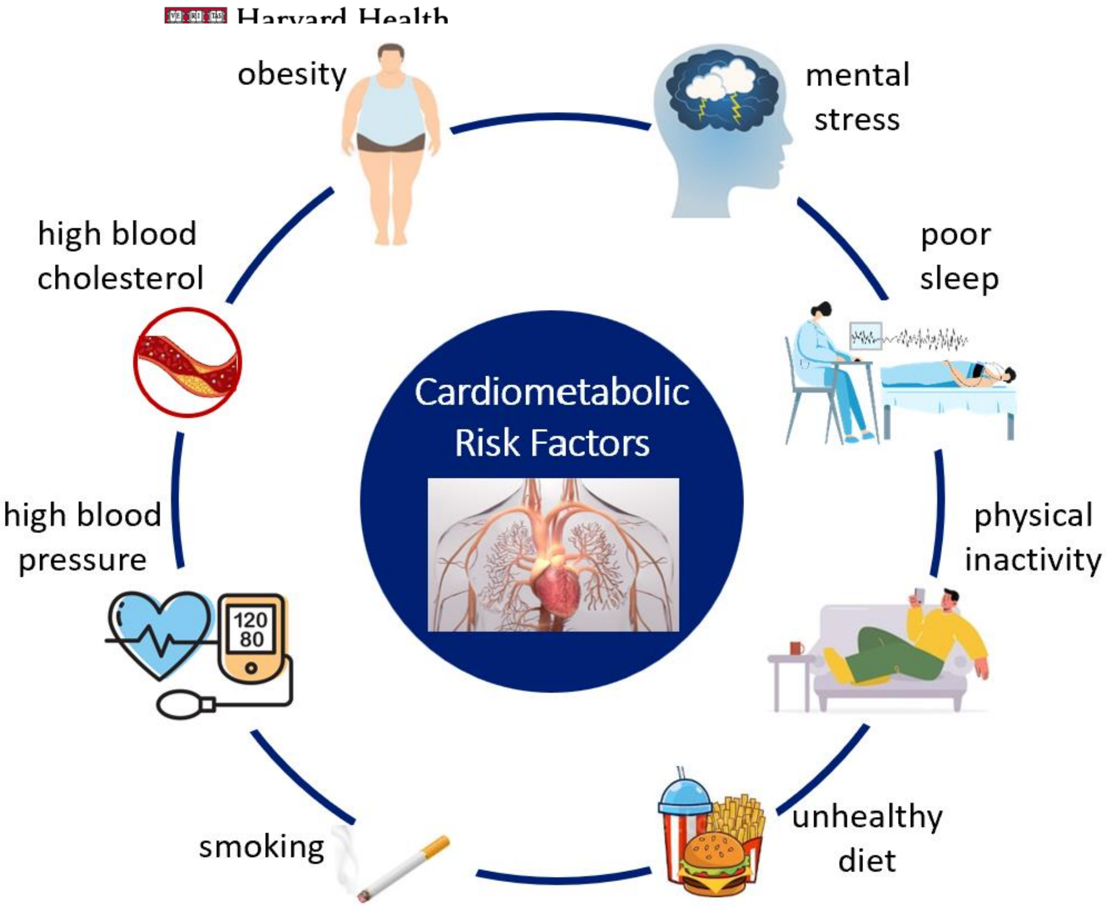
- Medical Imaging
- Electronic Health Records
- Health Surveys
- Smart Sensing data
- …

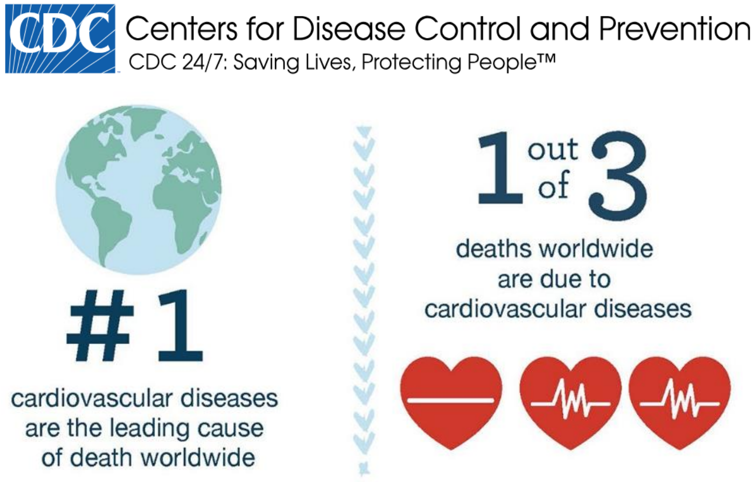


Precision Medicine

Phenotype Discovery

Cardiometabolic (CM) Health: The Leading Cause of Death





Heterogeneity in CM Risk Factors: A Major Challenge in CM Health Promotion

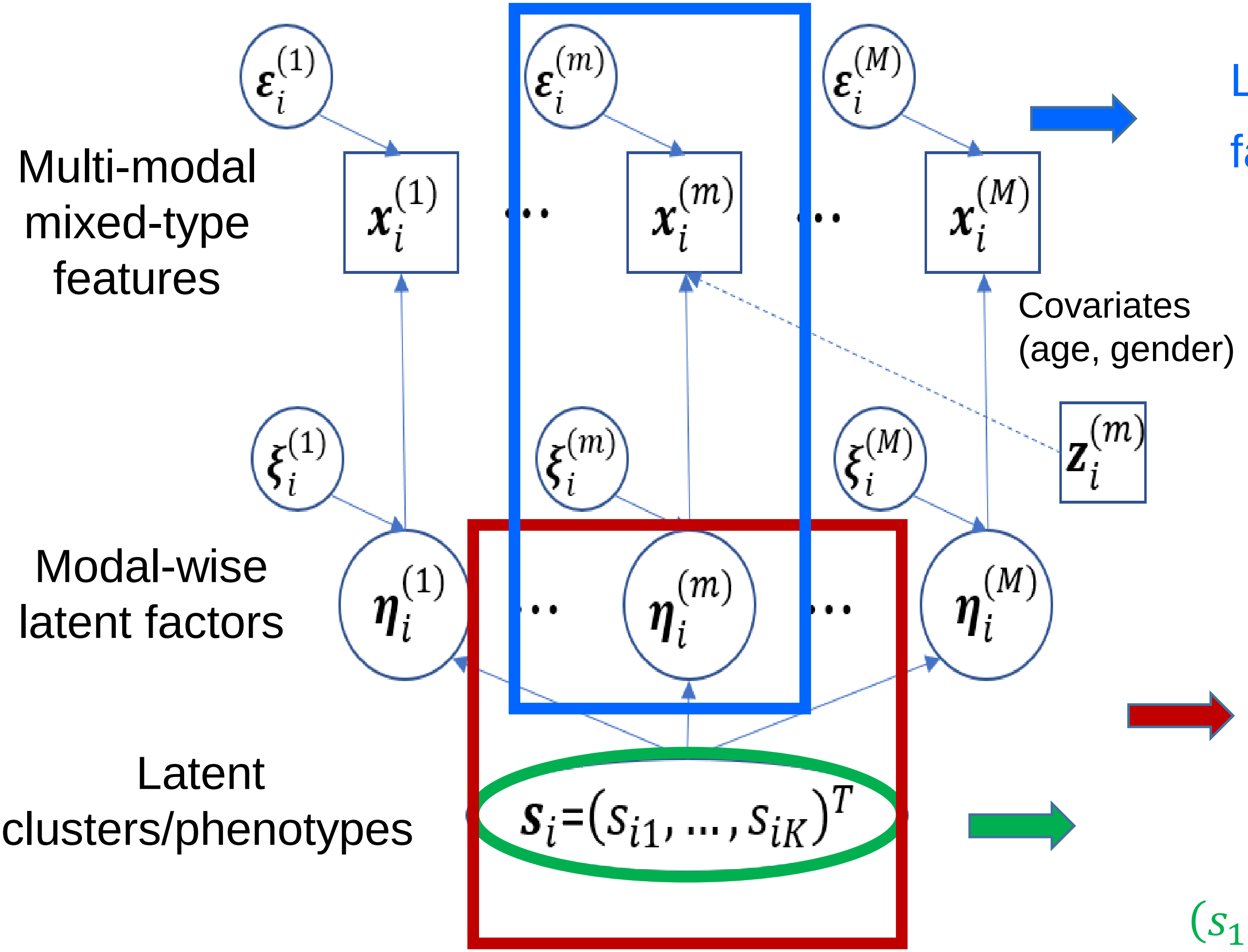
Limitations of Existing Research

- **Clinical models:** Specific to one risk modality
  - Aim to discover underlying mechanism between the risk factors and CM health
  - Lack a comprehensive consideration of multi-modal risk factors
- **Statistical clustering models**
  - Gaussian Mixture Model (GMM)
    - Not suitable for mixed-type data including continuous, nominal, and ordinal
    - Mixed-GMMs
      - Not consider latent factor structures
  - Factor Mixture Model (FMM)
    - Considers modality-specific latent factors for dimension reduction and knowledge discovery
    - Not consider either mixed-type data or sparse variable selection

Proposed: Multi-modal Mixed-type Factor Mixture Model with Hierarchical Selection (M2-FMM-hier)

- Develop a Multi-modal Mixed-type Factor Mixture Model capable of ***hierarchical selection*** **(**M2-FMM-hier) for modalities and features
  - If a modality is uninformative to clustering, all its features will be excluded.
  - Feature selection happens in the modalities that remain.
- Enable modality-specific latent factor extraction to increase model interpretability and facilitate medical knowledge discovery.
- Apply M2-FMM-hier for phenotype discovery from high-dimensional multi-modal mixed-type health data for targeted intervention and Precise Medicine.

Mathematical Formulation: M2-FMM-hier



Link modality-wise mixed-type features $\boldsymbol{x}_i^{(m)}$ with latent factors $\boldsymbol{\eta}_i^{(m)}$ and covariates $\boldsymbol{z}_i^{(m)}$:

$$\boldsymbol{x}_i^{(m)} = \mathbf{L}^{(m)}\boldsymbol{\eta}_i^{(m)} + \mathbf{B}^{(m)}\boldsymbol{z}_i^{(m)} + \boldsymbol{\varepsilon}_i^{(m)}$$

for $i = 1,\dots,N$ and $m = 1,\dots,M_1$.

$$log\left(\frac{\mathrm{P}\left(\boldsymbol{x}_i^{(m)} \le c\times\mathbf{1}_P\right)}{1-\mathrm{P}\left(\boldsymbol{x}_i^{(m)} \le c\times\mathbf{1}_P\right)}\right) = \boldsymbol{\alpha}_c^{(m)} + \mathbf{L}^{(m)}\boldsymbol{\eta}_i^{(m)} + \boldsymbol{B}^{(m)}\boldsymbol{z}_i^{(m)} + \boldsymbol{\varepsilon}_i^{(m)}$$

for $i = 1,\dots,N$ and $m = M_1+1,\dots,M_1+M_2$.

Link latent factors with latent phenotypes:

$$\boldsymbol{\eta}_i^{(m)} = \mathbf{U}^{(m)}\boldsymbol{s}_i + \boldsymbol{\xi}_i^{(m)}$$

Prior distribution of phenotypes:

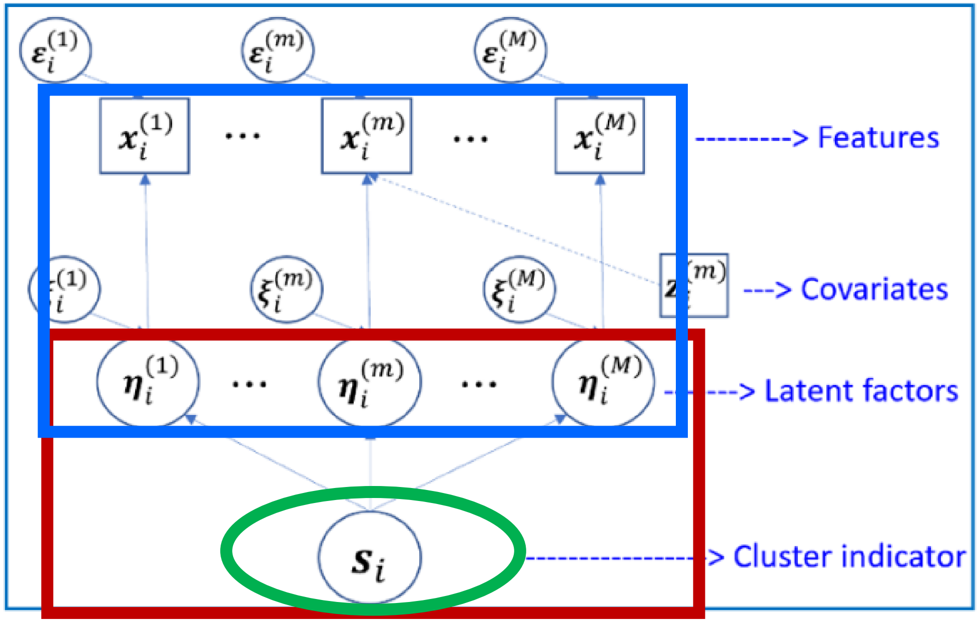
$(s_1,\dots,s_K)^T \sim$ multinomial $(w_1,\dots,w_K)$

Mathematical Formulation: M2-FMM-hier

- Complete-data (observed & latent) log-likelihood function:

$$l\left(f\left(\boldsymbol{\Theta}; \{\mathbf{X}_m, \mathbf{Z}_m, \mathbf{H}_m\}_{m=1}^{M}, \boldsymbol{s}\right)\right)$$

$$= \sum_{m=1}^{M_1} \log\left(f\left(\mathbf{X}_m, \mathbf{Z}_m | \mathbf{H}_m; \boldsymbol{\Theta}_1\right)\right) + \sum_{m=M_1+1}^{M_1+M_2} \log\left(f\left(\mathbf{X}_m, \mathbf{Z}_m | \mathbf{H}_m; \boldsymbol{\Theta}_2\right)\right) + \sum_{m=1}^{M} \log\left(f\left(\mathbf{H}_m | \boldsymbol{s}; \boldsymbol{\Theta}_3\right)\right) + \log\left(f\left(\boldsymbol{s}; \boldsymbol{\Theta}_4\right)\right).$$



Mathematical Formulation: M2-FMM-hier

- Complete-data (observed & latent) log-likelihood function:

$$l\left(f\left(\boldsymbol{\Theta}; \{\mathbf{X}_m, \mathbf{Z}_m, \mathbf{H}_m\}_{m=1}^{M}, \boldsymbol{s}\right)\right)$$

$$= \sum_{m=1}^{M_1} \log\left(f(\mathbf{X}_m, \mathbf{Z}_m | \mathbf{H}_m; \boldsymbol{\Theta}_1)\right) + \sum_{m=M_1+1}^{M_1+M_2} \log\left(f(\mathbf{X}_m, \mathbf{Z}_m | \mathbf{H}_m; \boldsymbol{\Theta}_2)\right) + \sum_{m=1}^{M} \log\left(f(\mathbf{H}_m | \boldsymbol{s}; \boldsymbol{\Theta}_3)\right) + \log\left(f(\boldsymbol{s}; \boldsymbol{\Theta}_4)\right).$$

- Optimization with double l_{21} penalization:

$$\min_{\boldsymbol{\Theta}} \tilde{l}(\boldsymbol{\Theta}) = -l\left(f\left(\boldsymbol{\Theta}; \{\mathbf{X}_m, \mathbf{Z}_m, \mathbf{H}_m\}_{m=1}^{M}, \boldsymbol{s}\right)\right) + \lambda_1 \sum_{m=1}^{M} \left\|\mathbf{u}^{(m)}\right\|_2 + \lambda_2 \sum_{m=1}^{M} \sum_{p=1}^{P} \left\|\boldsymbol{l}_p^{(m)}\right\|_2$$

subject to $E\left(\boldsymbol{\eta}_i^{(m)}\right) = \mathbf{0}, Var\left(\boldsymbol{\eta}_i^{(m)}\right) = \mathbf{I}$ (identifiability constraints)

Novel Property of the optimization: hierarchical selection of modalities and features

$$\boldsymbol{x}_m | s_k = 1 \sim N\left(\mathbf{L}_m \boldsymbol{u}_{m,k} + \mathbf{B}_m \boldsymbol{z},\ \mathbf{L}_m \boldsymbol{\Sigma}_m \mathbf{L}_m^T + \boldsymbol{\Psi}_m\right)$$

m-th modality
k-th cluster

- If $\mathbf{u}^{(m)} = \mathbf{0}$, then the m^{th} modality is uninformative to phenotype clustering.
- If $\mathbf{u}^{(m)} \neq \mathbf{0}$ and $\boldsymbol{l}_p^{(m)} = \mathbf{0}$, then the p-th feature is uninformative to phenotype clustering.

Methodological Contribution: Gauss-Hermite Expectation-Majorization-Minimization (GH-EMM) Algorithm

- Traditional Expectation-Maximization (EM) framework does not suffice.
  - E-step: derive $Q(\boldsymbol{\Theta}; \boldsymbol{\Theta}^{(j-1)}) \triangleq E_{\{\mathbf{H}_m\}_{m=1}^M, \boldsymbol{s} | \{\mathbf{X}_m\}_{m=1}^M; \boldsymbol{\Theta}^{(j-1)}} \{\tilde{l}(\boldsymbol{\Theta}; \{\mathbf{X}_m, \mathbf{H}_m\}_{m=1}^M, \boldsymbol{s})\}$ **(*)**
  - M-step: minimize $-Q(\boldsymbol{\Theta}; \boldsymbol{\Theta}^{(j-1)})$

Methodological Contribution: Gauss-Hermite Expectation-Majorization-Minimization (GH-EMM) Algorithm

- Traditional Expectation-Maximization (EM) framework does not suffice.
  - E-step: derive $Q(\boldsymbol{\Theta}; \boldsymbol{\Theta}^{(j-1)}) \triangleq E_{\{\mathbf{H}_m\}_{m=1}^M, \boldsymbol{s} | \{\mathbf{X}_m\}_{m=1}^M;\, \boldsymbol{\Theta}^{(j-1)}} \{\tilde{l}(\boldsymbol{\Theta}; \{\mathbf{X}_m, \mathbf{H}_m\}_{m=1}^M, \boldsymbol{s})\}$ **(*)**

$$= \underbrace{\varphi_1\left(\left\{\mathbf{L}^{(m)}, \mathbf{B}^{(m)}, \boldsymbol{\Psi}^{(m)}\right\}_{m=1}^{M_1}\right)}_{\text{Numerical modalities}} + \underbrace{\varphi_2\left(\left\{\left\{\boldsymbol{\alpha}_c^{(m)}\right\}_{c=1}^{C-1}, \mathbf{L}^{(m)}, \mathbf{B}^{(m)}, \boldsymbol{\Psi}^{(m)}\right\}_{m=M_1+1}^{M_1+M_2}\right)}_{\text{Categorical modalities}} + \underbrace{\varphi_3\left(\left\{\left\{\boldsymbol{\mu}^{(m,k)}, \boldsymbol{\Sigma}^{(m,k)}\right\}_{k=1}^{K}\right\}_{m=1}^{M}\right)}_{\text{Latent factors}} + \varphi_4(\boldsymbol{w})$$

  - M-step: minimize $-Q(\boldsymbol{\Theta}; \boldsymbol{\Theta}^{(j-1)})$

Non-analytical terms

- Monte Carlo EM: computationally expensive
- *Gauss Hermite Quadrature*

$$\int \exp\{-x^2\} g(x) dx \approx \sum_{t=1}^{T} \omega_t g(x_t)$$

x_t: roots of the Hermite polynomial $H_T(x)$

$$\omega_t = \frac{2^{T+1} T! \sqrt{\pi}}{[H_{T+1}(x_t)]^2}$$

Methodological Contribution: Gauss-Hermite Expectation-Majorization-Minimization (GH-EMM) Algorithm

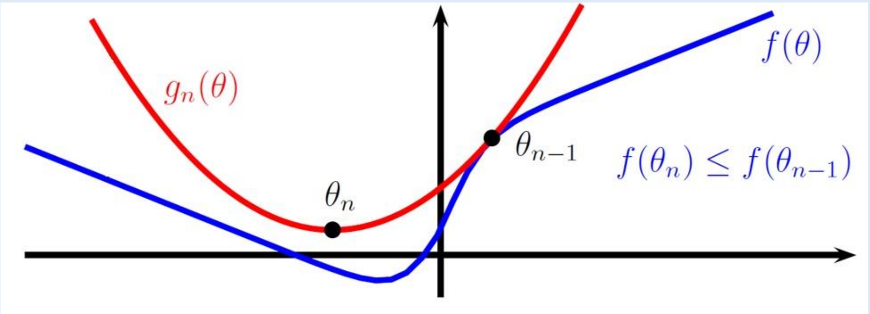
- Traditional Expectation-Maximization (EM) framework does not suffice.
  - E-step: derive $Q\left(\boldsymbol{\Theta};\boldsymbol{\Theta}^{(j-1)}\right) \triangleq E_{\{\mathbf{H}_m\}_{m=1}^{M},\boldsymbol{s}|\{\mathbf{X}_m\}_{m=1}^{M};\,\boldsymbol{\Theta}^{(j-1)}}\left\{\tilde{l}\left(\boldsymbol{\Theta};\ \{\mathbf{X}_m,\mathbf{H}_m\}_{m=1}^{M},\boldsymbol{s}\right)\right\}$ **(*)**

$$= \underbrace{\varphi_1\left(\left\{\mathbf{L}^{(m)},\mathbf{B}^{(m)},\boldsymbol{\Psi}^{(m)}\right\}_{m=1}^{M_1}\right)}_{\text{Numerical modalities}} + \underbrace{\varphi_2\left(\left\{\left\{\boldsymbol{\alpha}_c^{(m)}\right\}_{c=1}^{C-1},\mathbf{L}^{(m)},\mathbf{B}^{(m)},\boldsymbol{\Psi}^{(m)}\right\}_{m=M_1+1}^{M_1+M_2}\right)}_{\text{Categorical modalities}} + \underbrace{\varphi_3\left(\left\{\left\{\boldsymbol{\mu}^{(m,k)},\boldsymbol{\Sigma}^{(m,k)}\right\}_{k=1}^{K}\right\}_{m=1}^{M}\right)}_{\text{Latent factors}} + \varphi_4(\boldsymbol{w})$$

  - M-step: minimize $-Q\left(\boldsymbol{\Theta};\boldsymbol{\Theta}^{(j-1)}\right)$

l_{21}- penalized non-smooth optimization problems

- Conventional solvers (e.g., BCGD and Nesterov's) are slow.
- *Majorization-Minimization (MM)*



$f(\theta)$: Objective function

$g_n(\theta)$: Majorizing surrogate

E-step Enabled by Gauss-Hermite (GH)

Definition 1 (**Multivariate GH approximation**): Given vector $\boldsymbol{z}$ with $\text{rank}(\boldsymbol{z}) = Q$, and function $g \in \mathbb{C}^{2T}: \mathbb{R}^Q \to \mathbb{R}$ by applying Hermite interpolation , we have

$$\int_{\boldsymbol{S}} exp\{-\boldsymbol{z}^T \boldsymbol{z}\} g(\boldsymbol{z}) d\boldsymbol{z} \approx \sum_{t_1=1}^{T} \cdots \sum_{t_Q=1}^{T} \omega_{t_1} \cdots \omega_{t_Q} g(\boldsymbol{z}_t)$$

where $\boldsymbol{S} \in \mathbb{R}^Q$ is the integration set, $\boldsymbol{z}_t = (z_{t_1}, \cdots, z_{t_Q})^T$ and z_{t_q} are the roots of Hermite polynomial of order T, $H_T(x) = (-1)^T e^{x^2} \frac{d^T}{dx^T} e^{-x^2}$ for $t_q \in \{1, \cdots, T\}$ and $q \in \{1, \cdots, Q\}$. The weight, ω_{t_q}, is given by $\omega_{t_q} = \frac{2^{T+1} T! \sqrt{\pi}}{[H_{T+1}(z_{t_q})]^2}$.

Proposition 1 (**The GH approximate error is bounded**): If the integration set $\boldsymbol{S}$ in Definition 1 is closed, the GH approximation error, determined by $\frac{T! \sqrt{\pi}}{2^T (2T)!} g^{(2T)}(\xi)$, reduces to zero for a sufficiently large T.

E-step: The non-analytical terms can be explicitly approximated by GH Quadrature.

M-step Integrated with Majorization Minimization (MM)

Theorem 1 (**Convergence of the MM algorithm**): Assume the following optimization problem where $\boldsymbol{l}$ denotes the parameters to be estimated and **D** denotes the data:

$$\min_{\boldsymbol{l}} -\varphi(\boldsymbol{l}|\mathbf{D}) + \lambda\|l\|_2$$

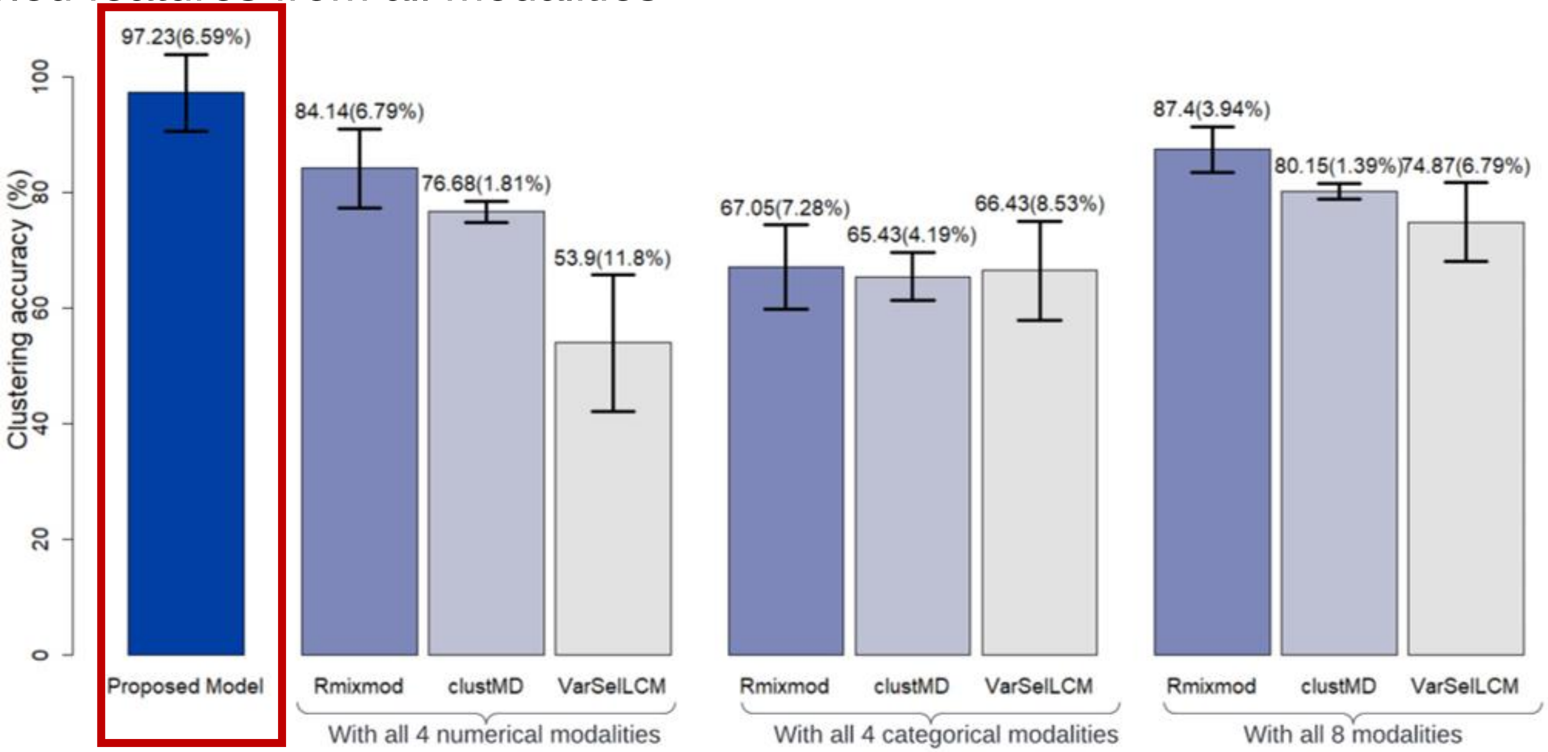
If the objective function $\varphi(\boldsymbol{l}|\mathbf{D})$ is differentiable and convex with respect to $\boldsymbol{l}$ and its first-order derivative $\varphi'(\boldsymbol{l}|\mathbf{D})$ is Lipschitz continuous, the MM algorithm with the first-order surrogate function is guaranteed to achieve the Karush–Kuhn–Tucker (KKT) conditions upon convergence.

Proposition 2 (**Lipschitz continuity**): The sub-optimization problems are jointly convex and their first-order derivatives are Lipschitz continuous with respect to $\{\mathbf{L}^{(m)}, \mathbf{B}^{(m)}\}$ for $m = 1, \ldots, M_1$, $\{\mathbf{L}^{(m)}, \mathbf{B}^{(m)}\}$ for $m = M_1 + 1, \ldots, M_1 + M_2$, and $\{\boldsymbol{\mu}^{(m,k)}\}_{k=1}^{K}$ for $m = 1, \ldots, M_1 + M_2$, respectively.

M-step: The objective functions satisfy the Lipschitz continuity condition and thus can be efficiently optimized by the MM algorithm.

Simulation Study – Clustering Accuracy

- Competing methods applied to cluster:
  - Pooled numerical modalities
  - Pooled categorical modalities
  - Pooled features from all modalities

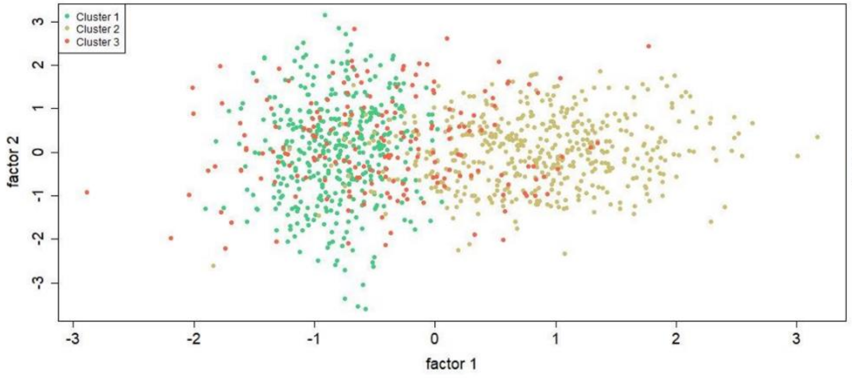


Application: CM Phenotype Discovery

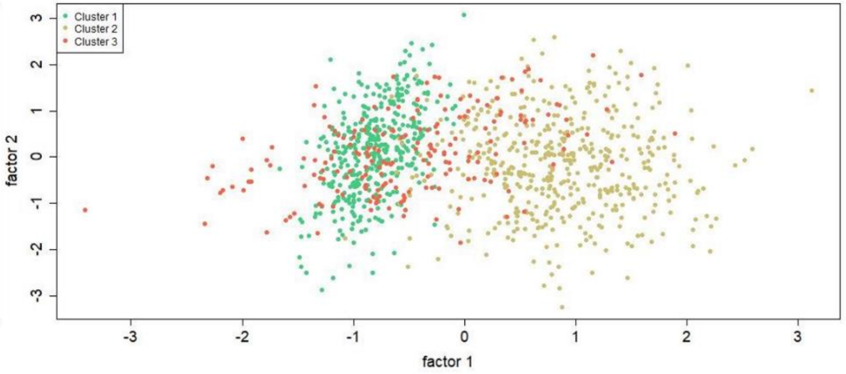
- 1,052 participants from the Hispanic Community Health Study (HCHS)
- 2 covariates: age and gender
- 10 continuous and categorical modalities of CM risk factors
    - Sleep Measures
    - Epworth Sleepiness Scales (ESS)
    - Women's Health Initiative Insomnia Rating Scales (WHIIRSs)
    - Alternative Healthy Eating Indices (AHEIs)
    - Global Physical Activity Questionnaire (GPAQ)
    - HCHS Acculturation
    - Center for Epidemiologic Studies Depression Scales (CES-D)
    - State-Trait Anxiety Inventories (STAIs)
    - Clinical Characteristics

Results

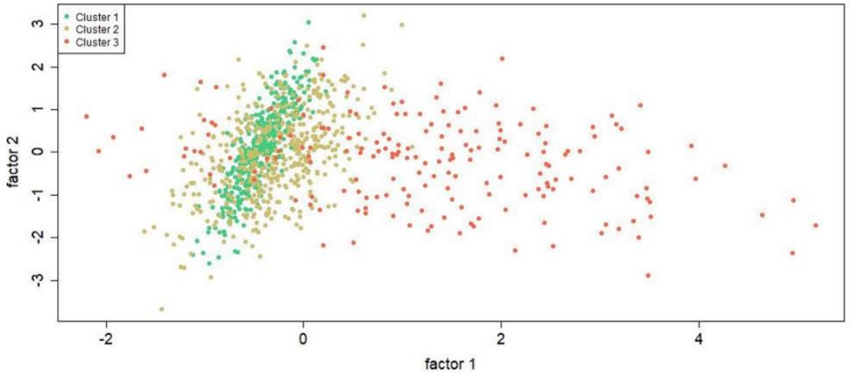
- Identified 3 CM phenotypes:



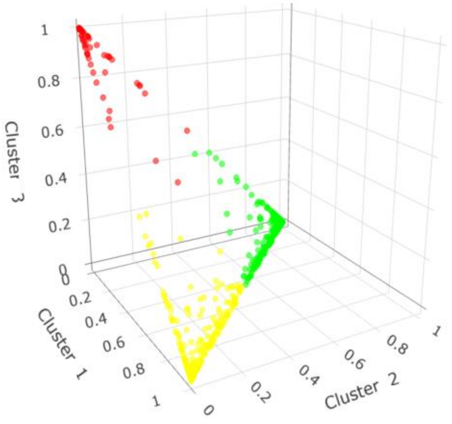
(a) Depression modality



(b) Anxiety modality



(c) Sleep monitoring modality

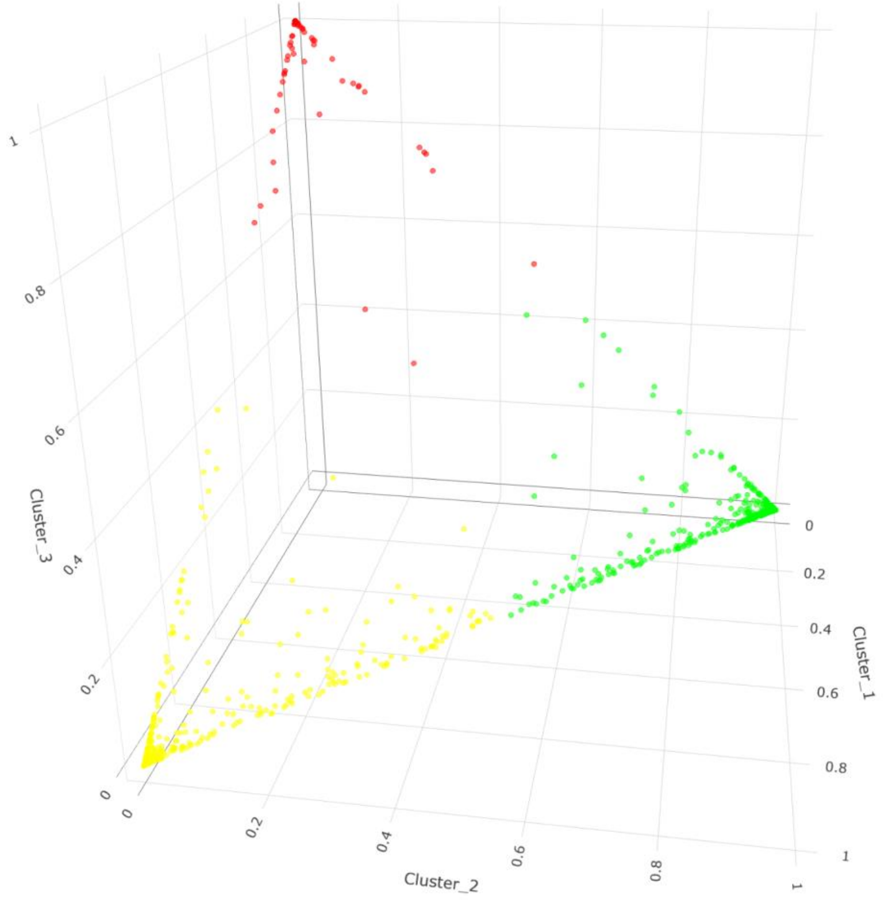


(d) Posterior probabilities of belonging to each cluster using all modalities

The multi-modal data achieves a better cluster separation.

Results

- Identified 3 CM phenotypes:



	Mental Health (CES-D[1])	Sleep Condition[2]	Cardiometabolic Health (FRS[3])
Cluster 1 (456)	Healthy	Healthy	Healthy
Cluster 2 (449)	Worse**	Mild	Mild***
Cluster 3 (147)	Mild	Worse***	Worse***

1: The Center for Epidemiological Studies Depression Score
2: Measured by Apnea/Hypopnea Index, SpO2, heart rate, and time spent in loud snoring
3: Framingham Risk Score

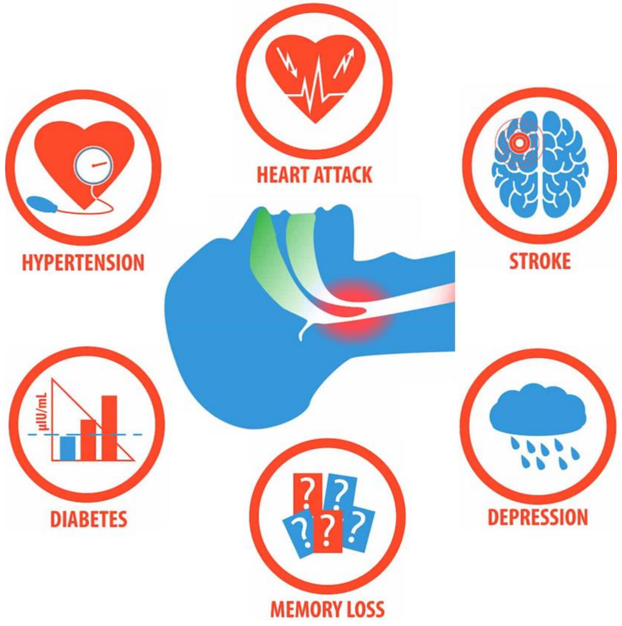
Outline

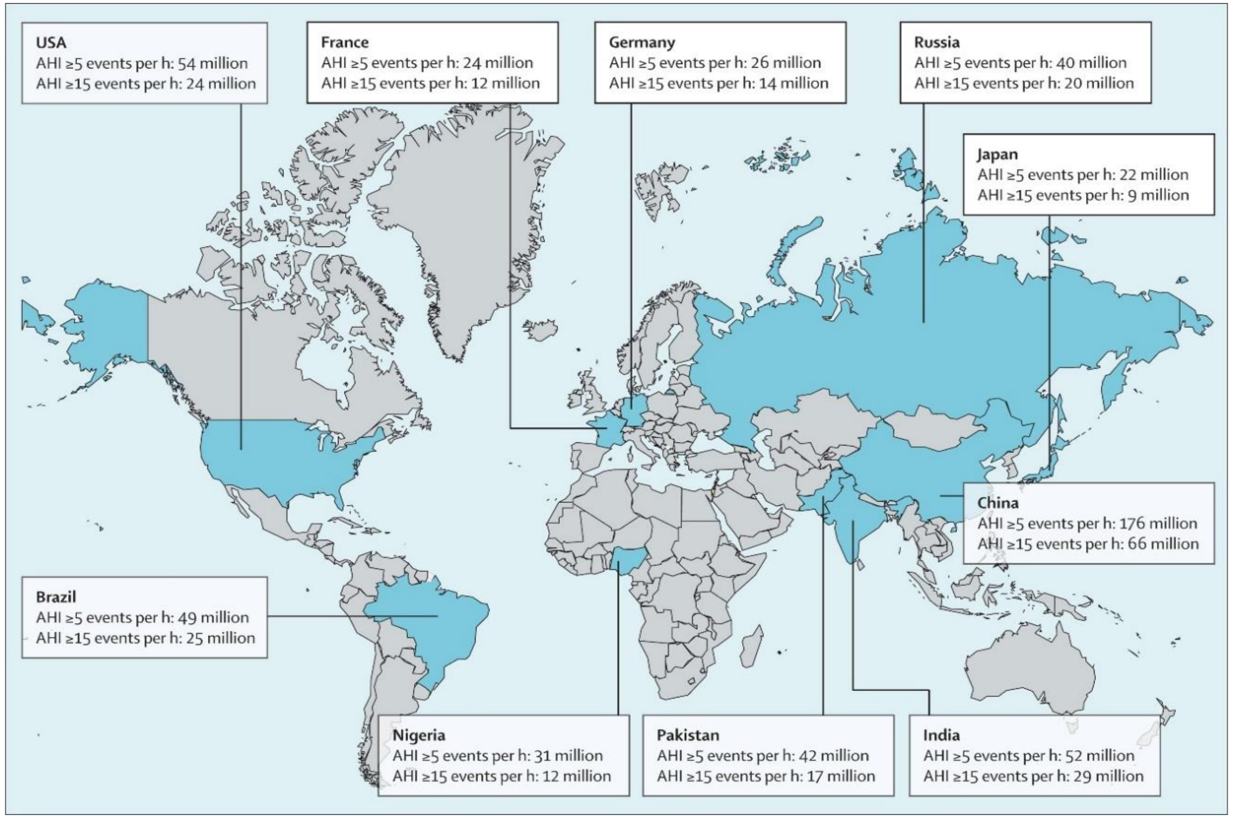
- Introduction
  - Statistical machine learning and Data fusion for Precise Medicine and Public Health
- Disease diagnosis and phenotyping
  - Unsupervised learning of multi-faced medical data for phenotype discovery
- **Privacy-preserving telemedicine**
  - Federated learning of functional data for privacy-preserving telemedicine
- Conclusions and Future Works

Federated function-on-function regression with an efficient gradient boosting algorithm for privacy-preserving telemedicine

Yu Ding, Carlos Costa, and Bing Si

Obstructive Sleep Apnea (OSA)





- Sleep-related breathing disorder
- Associated with neurocognitive and cardiovascular diseases

- OSA affects almost 1 billion people but is underdiagnosed in the population.

Dangers of Untreated Sleep Apnea: Memory, Cardiac, Diabetes, Quality of Life, Sleepiness, Depression, Automobile Accidents, and Death
Estimation of the global prevalence and burden of obstructive sleep apnoea: a literature-based analysis. The Lancet Respiratory Medicine. 2019 Aug 1;7(8):687-98.

OSA Telemedicine

Wearable devices:
- Make at-home sleep study feasible
- Offer opportunities for cost-effective telemedicine of OSA

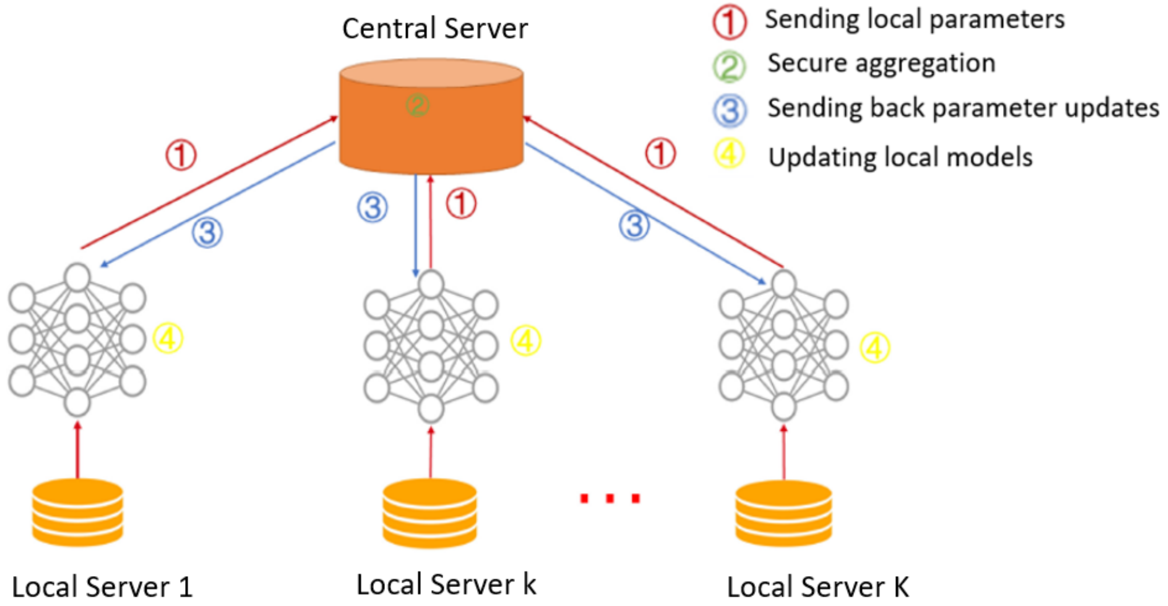
Current diagnostic approach:
- Manually scored by certified technicians
- Apnea-Hypopnea Index (AHI): frequency of adverse respiratory events
- Labor-intensive & subjective

Research Problem:
- Predict the functional AHI from functional bio-signal features and non-functional clinical characteristics
- Facilitate automated diagnosis and telemedicine of OSA

Privacy? **Efficiency?**

Limitations of Existing Work

- Functional Regression
  - Scalar-on-function (Müller & Yao, 2008; Wang et al., 2017)
  - Function-on-scalar (Zhang et al., 2022)
  - Function-on-function
    - No variable selection (Chiou et al., 2016; Iwaizumi and Kato, 2018)
    - Computationally expensive (Ivanescu et al., 2015; Sun et al., 2018; Luo and Qi, 2017)
    - Not privacy-preserving
- Federated Learning (FL)
  - Privacy-preserving
  - Not for function-on-function regression



Proposed: Federated Learning of Functional Regression

- Develop a novel Federated Learning (FL) algorithm for function-on-function regression with variable selection, coupled with an efficient optimization algorithm featuring two key innovations:
  - Gradient Boosting (GB) is leveraged for model estimation with variable selection, known to be **computationally-efficient**.
  - Least Squares Approximation (LSA) is deployed for FL, proven to be both **communicationally- & statistically-efficient**.
- Apply the proposed method to predict disease severity from functional and non-functional data, aiming to facilitate automated diagnosis and telemedicine for OSA.

Mathematical Formulation

- **Notations**
  - $y_n(t)$: Functional response for subject n
  - $\boldsymbol{x}_n = \{x_{n1}(t), \cdots, x_{np}(t), \cdots, x_{nP}(t)\}^T$: Functional or non-functional predictors for subject n
  - $\beta_p(s,t)$: Bivariate coefficient function for predictor p
  - N: Number of subjects; P: Number of predictors; T: Sampling period, i.e., $t \in T$
- **Assumption**
  - Double expansion of $\beta_p(s,t)$ on basis systems $\boldsymbol{\theta}$ & $\boldsymbol{\eta}$ with K_1 and K_2 functions:

$$\beta_p(s,t) = \boldsymbol{\theta}(s)^T \mathbf{B}_p \boldsymbol{\eta}(t) \qquad \mathbf{B}_p \in \boldsymbol{R}^{K_1 \times K_2}$$

  - Function-on-function regression for subject n:

$$y_n(t) = \sum_{p=1}^{P} \int_{s \in T} x_{np}(s)\, \beta_p(s,t) ds + \varepsilon_n(t)$$

$$= \sum_{p=1}^{P} \boxed{h_p(t)} + \varepsilon(t)$$

Base learner: $h_p(t) = \boldsymbol{z}_{np} \mathbf{B}_p \boldsymbol{\eta}(t)$ where $\boldsymbol{z}_{np} = \int_{s \in T} x_{np}(s)\, \boldsymbol{\theta}(s)^T ds$

Methodological Contribution: federated Gradient Boosting algorithm with the Least Squares Approximation (fed-GB-LSA)

Gradient Boosting (GB) for Function-on-function Regression

- GB aims to solve the following optimization:

$$f^* = argmin_f \sum_{n=1}^{N} \int_{t\in T} (y_n(t) - f(t, \boldsymbol{z}_n))^2 dt$$

- In the ω-th iteration:
  - Computes the **negative gradient** of the loss function with respect to f, i.e., $\boldsymbol{u}^{(\omega)} \in R^{N\times 1} = -\frac{\partial l}{\partial f}\Big|_{f = f^{[\omega-1]}}$
  - Fit **each base learner** $h_p(t) = \boldsymbol{z}_{np}\mathbf{B}_p\boldsymbol{\eta}(t)$ for $p = 1, \ldots, P$ to the negative gradient $\boldsymbol{u}^{(\omega)}$

$$\widehat{\mathbf{B}}_p^{(\omega)} = \underset{\mathbf{B}_p}{\operatorname{argmin}} \sum_{n=1}^{N} \int_{t\in T} \left(u_n^{(\omega)}(t) - \boldsymbol{z}_{np}\mathbf{B}_p\boldsymbol{\eta}(t)\right)^2 dt$$

  - Update the model using the **best learner** with the minimal residual $h_{p^*}^{(\omega)} = \boldsymbol{z}_{np}\widehat{\mathbf{B}}_{p^*}^{(\omega)}\boldsymbol{\eta}(t)$

$$f^{(\omega)}(t) = f^{(\omega-1)}(t) + \nu h_{p^*}^{(\omega)}$$

Methodological Contribution: federated Gradient Boosting algorithm with the Least Squares Approximation (fed-GB-LSA)

Proposition 1: Assume $\mathbf{Z} \in \boldsymbol{R}^{N \times K_1}$, $\mathbf{B} \in \boldsymbol{R}^{K_1 \times K_2}$, two functional vectors $\boldsymbol{u}(t)$ and $\boldsymbol{\eta}(t)$ where $\boldsymbol{u}(t) = \left(u_1(t), \cdots, u_N(t)\right)^T$ and $\boldsymbol{\eta}(t) = \left(\eta_1(t), \cdots, \eta_N(t)\right)^T$, and $J_{\eta\eta} = \int_{t \in T} \boldsymbol{\eta}(t)\boldsymbol{\eta}^T(t)\mathrm{d}t$. For the optimization problem

$$\mathbf{B}^* = \underset{\mathbf{B}}{\operatorname{argmin}} \int_{\mathrm{t} \in \mathrm{T}} \|\mathbf{u}(\mathrm{t}) - \mathbf{Z}\mathbf{B}\boldsymbol{\eta}(\mathrm{t})\|^2 \mathrm{dt},$$

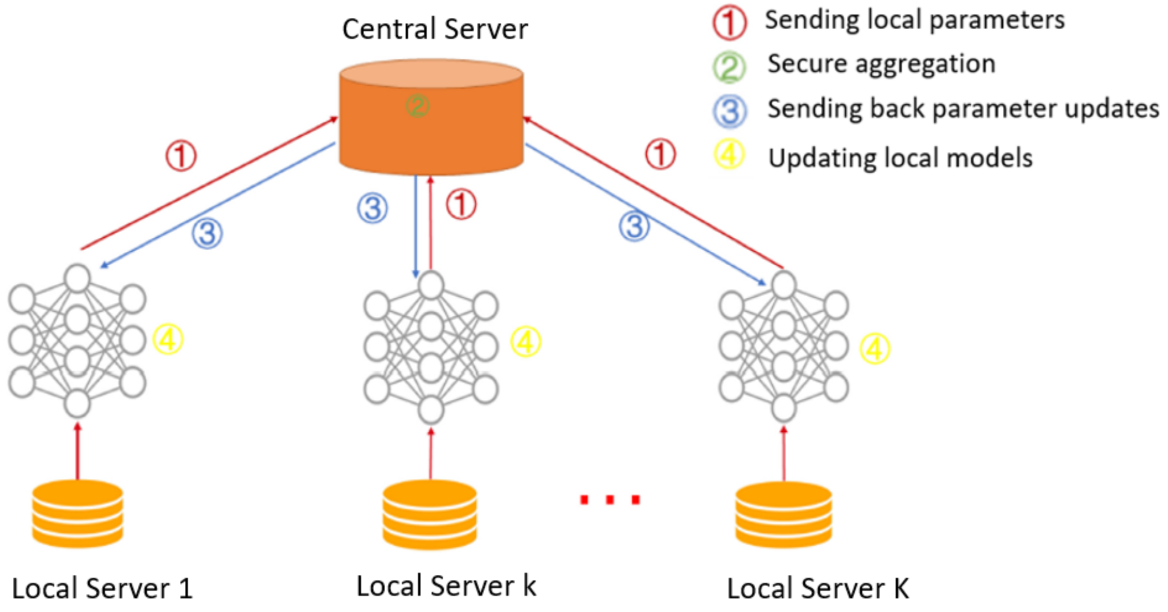
the optimal solution is

$$vec(\mathbf{B}^*) = \left(J_{\eta\eta} \otimes (\mathbf{Z}^T\mathbf{Z})\right)^{-1} vec\left(\mathbf{Z}^T \int_t \boldsymbol{u}(t)\boldsymbol{\eta}^T(t)dt\right).$$

In each GB iteration, the optimization problems can be solved analytically. ⟹ **Computational Efficiency**

Methodological Contribution: federated Gradient Boosting algorithm with the Least Squares Approximation (fed-GB-LSA)

- Notations:
  - K: Number of local servers; N: Number of subjects; N_k : Number of subjects in Server k
  - S_k contains subjects in Server k for $k = 1, \ldots, K$; $S = \{1, \ldots, N\} = \cup_{k=1}^{K} S_k$
- FL Model:



Methodological Contribution: federated Gradient Boosting algorithm with the Least Squares Approximation (fed-GB-LSA)

- Notations:
  - K: Number of local servers; N: Number of subjects; N_k : Number of subjects in Server k
  - S_k contains subjects in Server k for $k = 1, \dots, K$; $S = \{1, \dots, N\} = \cup_{k=1}^{K} S_k$
- FL Model:

Local models ($k = 1, \dots, K$)

$$\widetilde{\mathbf{B}}_{p,k}^{*} = \underset{\mathbf{B}_p}{\operatorname{argmin}}\, N_k^{-1} \sum_{n \in S_k} l_{n,p}(\mathbf{B}_p)$$

Global model

$$\widetilde{\mathbf{B}}_{p}^{*} = \underset{\mathbf{B}_p}{\operatorname{argmin}}\, N^{-1} \sum_{k=1}^{K} \sum_{n \in S_k} l_{n,p}(\mathbf{B}_p)$$

Taylor's Expansion at local optimal solution $\widetilde{\mathbf{B}}_{p,k}^{*}$

$$\approx N^{-1} \sum_{k=1}^{K} \sum_{n \in S_k} l_{n,p}(\widetilde{\mathbf{B}}_{p,k}^{*}) + N^{-1} \sum_{k=1}^{K} \sum_{n \in S_k} l'_{n,p}(\widetilde{\mathbf{B}}_{p,k}^{*})^{T} (\mathbf{B}_p - \widetilde{\mathbf{B}}_{p,k}^{*}) + N^{-1} \sum_{k=1}^{K} \sum_{n \in S_k} (\mathbf{B}_p - \widetilde{\mathbf{B}}_{p,k}^{*})^{T} l''_{n,p}(\widetilde{\mathbf{B}}_{p,k}^{*}) (\mathbf{B}_p - \widetilde{\mathbf{B}}_{p,k}^{*})$$

Not include $\mathbf{B}_p$ — Becomes zero — Least Squares Approximation (LSA)

$$\widehat{\mathbf{B}}_p = \left(\sum_{k=1}^{K} \frac{N_k}{N} \widehat{\Sigma}_{p,k}^{-1} \right)^{-1} \left(\sum_{k=1}^{K} \frac{N_k}{N} \widehat{\Sigma}_{p,k}^{-1} \widetilde{\mathbf{B}}_{p,k}^{*} \right)$$

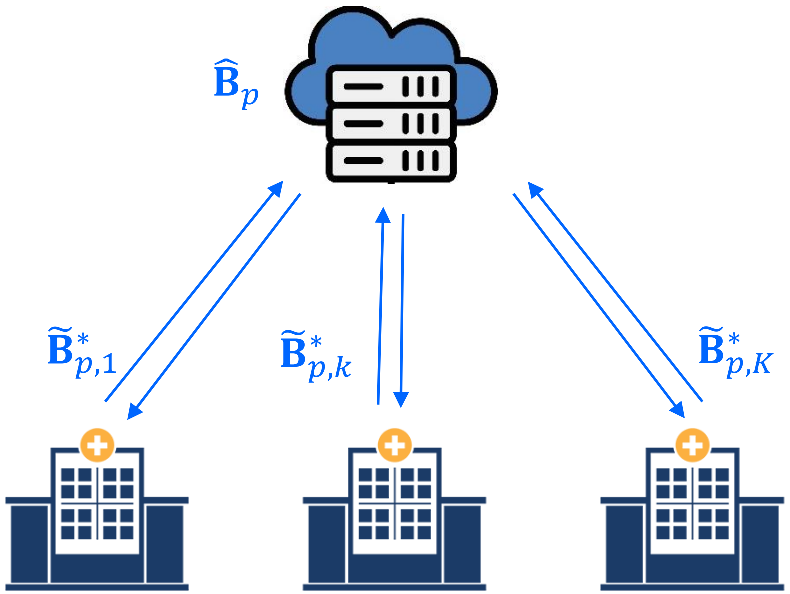
"One-shot" LSA-based aggregator for FL ⇒ **Communicational Efficiency**

Methodological Contribution: federated Gradient Boosting algorithm with the Least Squares Approximation (fed-GB-LSA)

Theorem 1 (Global asymptotic normality): We denote the asymptotic covariance matrix of the global estimator $\widetilde{\mathbf{B}}_p^*$ as $\boldsymbol{\Sigma}_p$. Given certain statistical regularity conditions and $K \ll \sqrt{N}$, we have $\sqrt{N}\left(vec(\widehat{\mathbf{B}}_p) - vec(\mathbf{B}_{p,0})\right) \rightarrow_d N(0, \boldsymbol{\Sigma}_p)$, which indicates that the proposed LSA estimator $\widehat{\mathbf{B}}_p$ achieves the same asymptotic normality as the global estimator $\widetilde{\mathbf{B}}_p^*$.

The proposed LSA estimator $\widehat{\mathrm{B}}_p$ achieves the same asymptotic normality as the global estimator $\widetilde{\mathrm{B}}_p^*$. $\Longrightarrow$ **Statistical Efficiency**

Methodological Contribution: federated Gradient Boosting algorithm with the Least Squares Approximation (fed-GB-LSA)



<u>**Iterate Until Convergence**</u>

Local Servers:

- Update local models (**Proposition 1**)
  - Closed-form GB estimators $\widetilde{\mathbf{B}}^*_{p,1}, \ldots, \widetilde{\mathbf{B}}^*_{p,K}$
  - **Computational efficiency**
- Send local parameters to the central server

Central Server:

- Global aggregation (**Theorem 1**)
  - LSA-based global aggregator $\widehat{\mathbf{B}}_p$
  - Global asymptotic normality: **Statistical efficiency**
  - One-shot: **Communicational efficiency**
- Send back parameter updates to local servers

Simulation Study - Setup

- Sample size: N = 1,000
- Number of predictors: P = 20 (5 effective & 15 dummy)

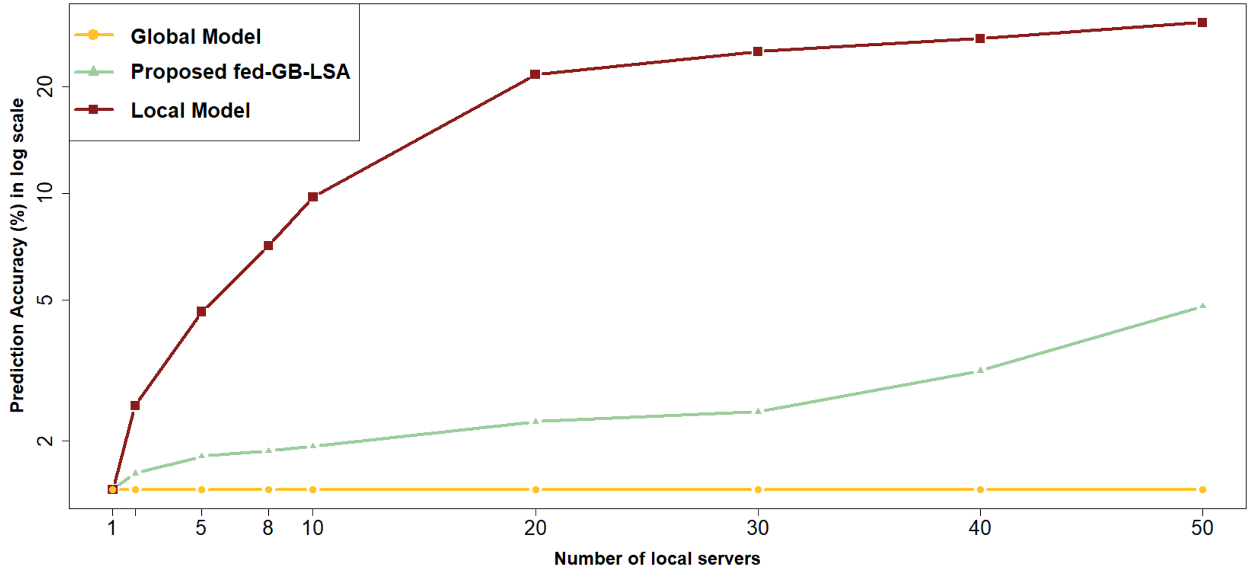
- Coefficient function: $\beta_p(s,t) = \boldsymbol{\varphi}(s)^T \mathbf{B}_p \boldsymbol{\varphi}(t)$
  - $\mathbf{B}_p$ is sampled from $N(1, 0.5)$ for effective predictor p.
  - $\mathbf{B}_p$ is set to be 0's for dummy predictor p.
- Functional predictors & response:
  - $x_{np}(t) = \sum_k c_{pk} \varphi_k(t)$ $\quad c_{pk} \sim U(-1,1) + e^{N(0.1\times p,\, 1)}$
  - $y_n(t_i) = \sum_{p=1}^{P} \sum_{i'} x_{np}(s_{i'}) \beta_p(s_{i'}, t_i) + \sum_k e_{pk} \varphi_k(t_i)$ $\quad e_{pk} \sim N(0,1)$

Simulation Study - Performance of the proposed fed-GB-LSA

- Mean Absolute Percentage Error (MAPE):

$$MAPE = \frac{100\%}{NT} \sum_{n=1}^{N} \sum_{t=1}^{T} \left| \frac{Y_{nt} - F_{nt}}{Y_{nt}} \right|$$

- We distribute the data (N = 1,000) across different numbers of servers.



Simulation Study - Compare fed-GB-LSA vs fed-GB-Average

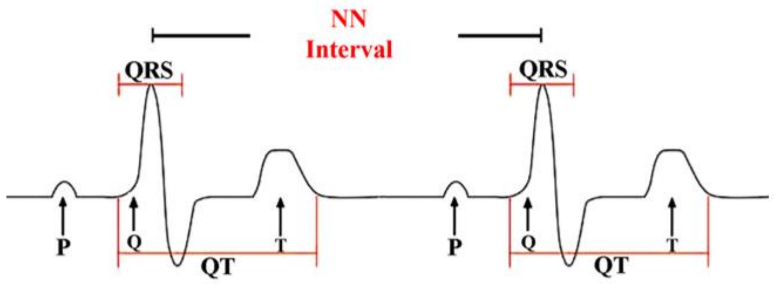
- fed-GB-LSA: $\hat{\mathbf{B}}_p = \left(\sum_{k=1}^{K} \frac{N_k}{N} \hat{\Sigma}_{p,k}^{-1}\right)^{-1} \left(\sum_{k=1}^{K} \frac{N_k}{N} \hat{\Sigma}_{p,k}^{-1} \tilde{\mathbf{B}}_{p,k}^{*}\right)$
- fed-GB-Average: $\hat{\mathbf{B}}_p = \sum_{k=1}^{K} \frac{N_k}{N} \tilde{\mathbf{B}}'_{p,k}$
- We increase # of servers (100 samples per server).

Table 2. Comparison of fed-GB-LSA and fed-GB-Average for MAPEs

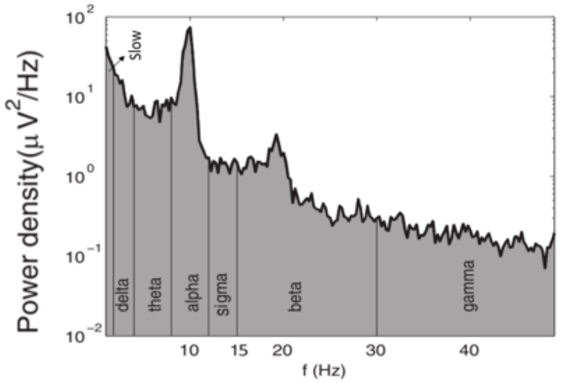
	K = 5	K = 10	K = 15	K = 20
Fed-GB-LSA	1.46%	1.19%	0.98%	0.83%
fed-GB-Average	1.74%	2.00%	2.05%	1.84%

Application for OSA telemedicine and diagnosis

- Data description
    - This dataset includes 408 subjects from the Sleep Heart Health Study (SHHS).
- Non-functional features
    - Age (year), gender (female or male), BMI (kg/m2), and ethnicity (Hispanic or not).
- Functional features
    - Bio-signal features extracted from the overnight sleep study
    - Each epoch includes 13 ECG-derived features 28 EEG-derived features.



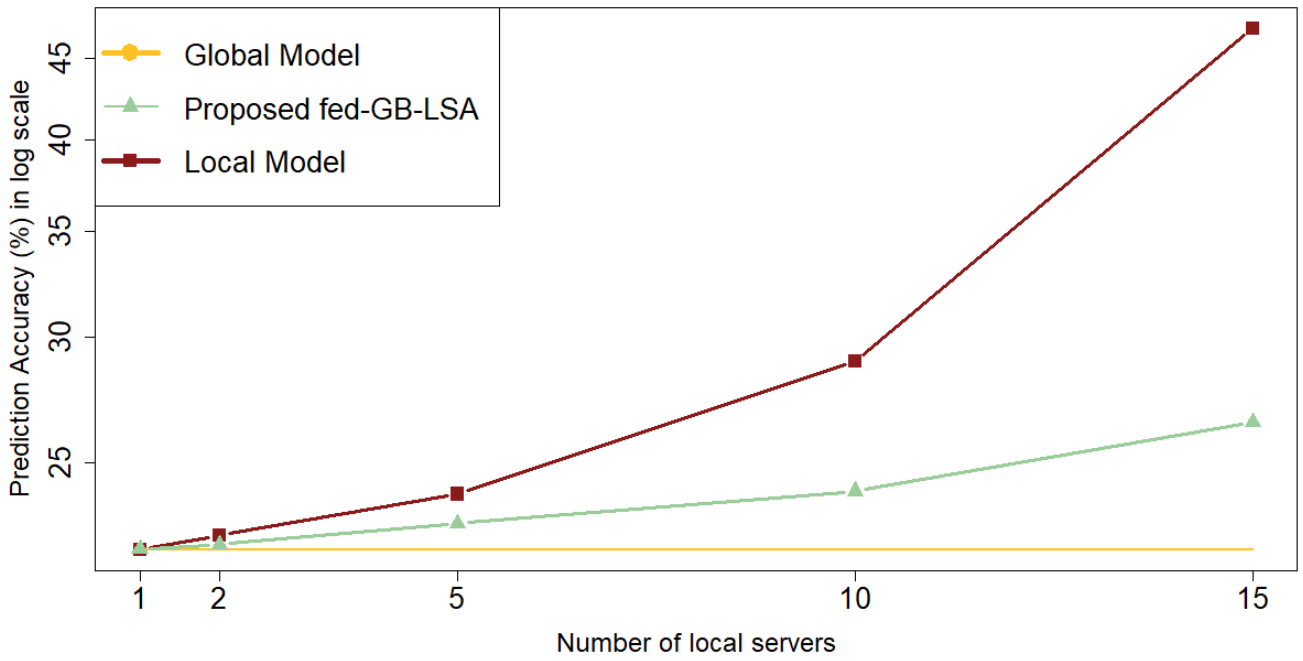
Heart Rate Variability (HRV) analysis for ECG signals



Power Spectral Density (PSD) analysis for EEG signals

Results

- Global function-on-function regression model:
  - 21.6% MAPE with 10-fold Cross Validation
- To mimic the FL setting, the dataset is randomly partitioned into several “local servers”:



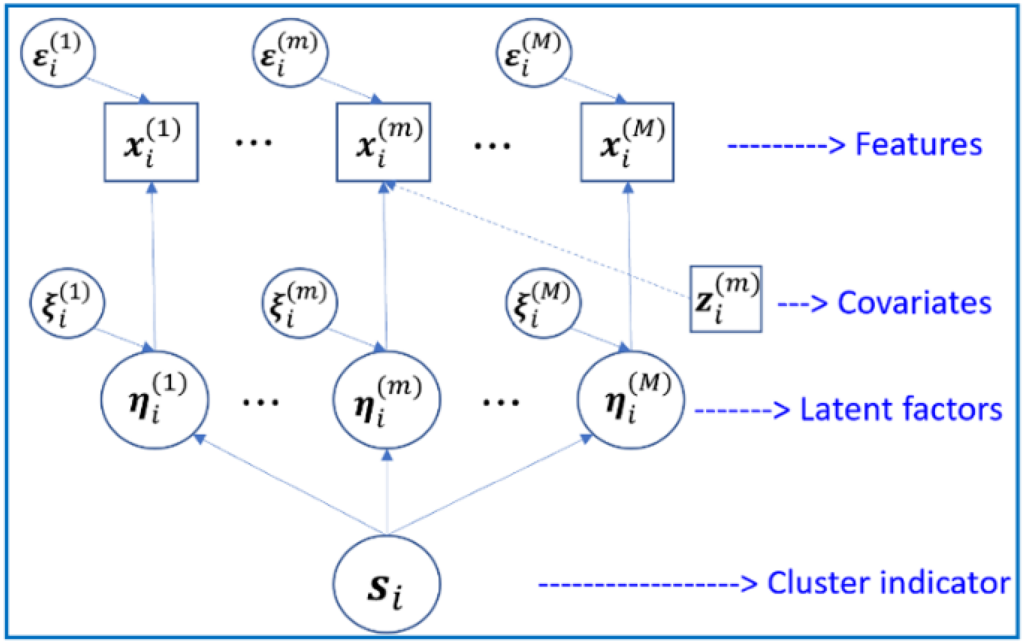
The proposed fed-GB-LSA sheds light on OSA diagnosis and telemedicine with privacy-preservation.

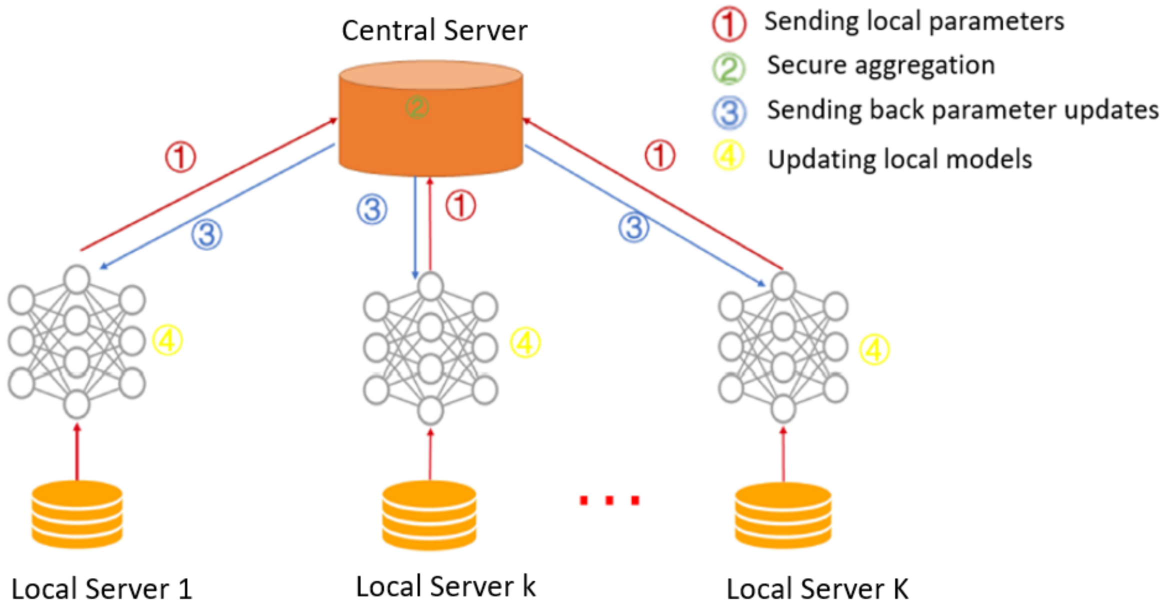
Increased healthcare accessibility
Improved public health

Outline

- Introduction
  - Statistical machine learning and Data fusion for Precise Medicine and Public Health
- Disease diagnosis and phenotyping
  - Unsupervised learning of multi-faced medical data for phenotype discovery
- Privacy-preserving telemedicine
  - Federated learning of functional data for privacy-preserving telemedicine
- **Conclusions and Future Works**

Conclusions

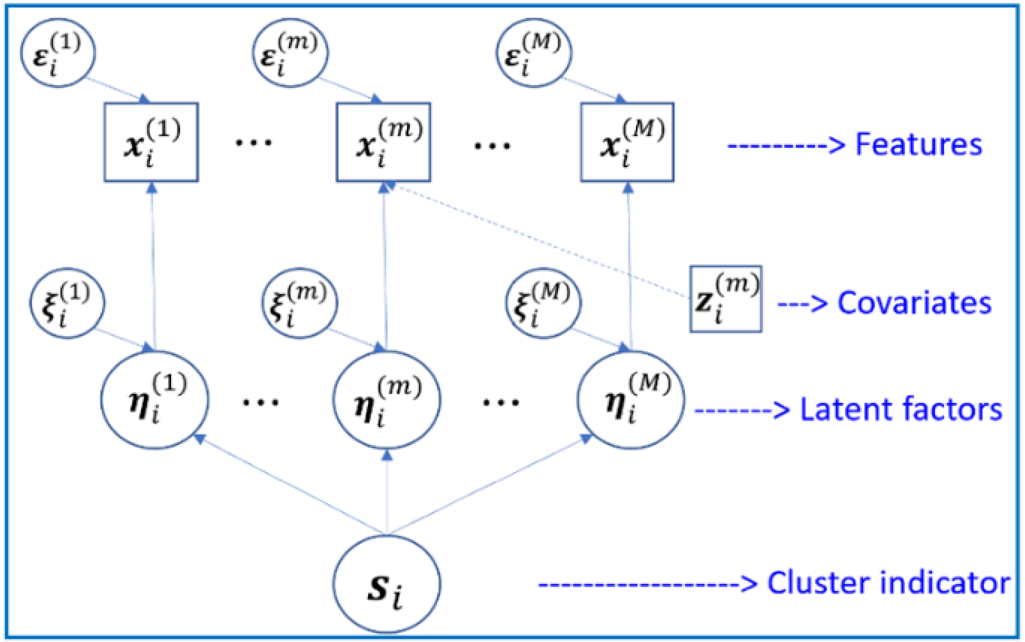


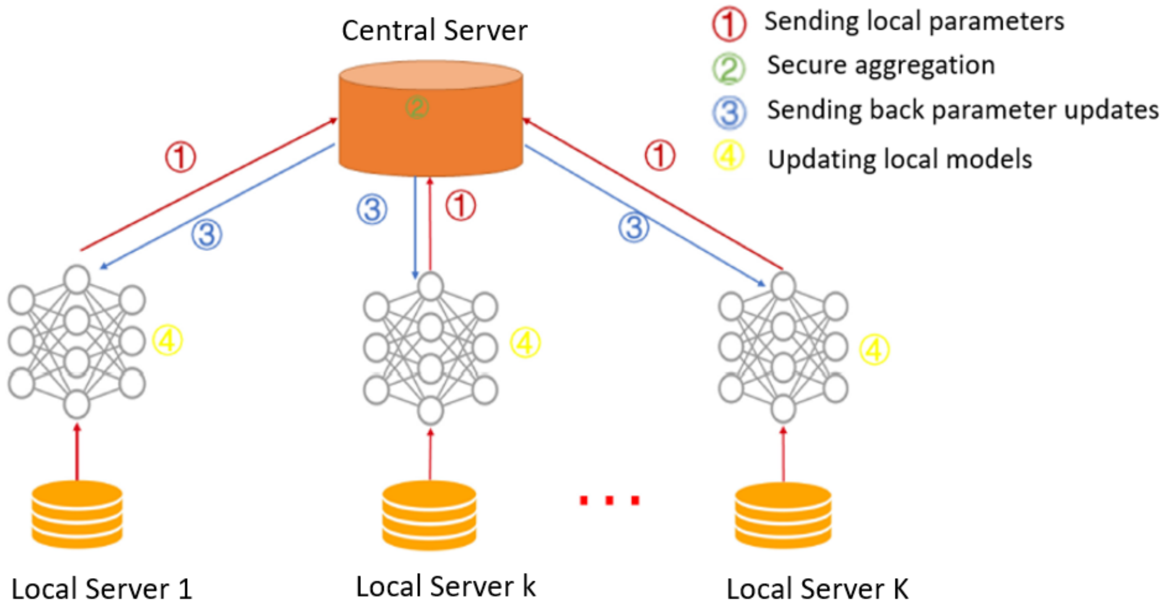


Challenges:

1. Privacy and Security
2. Mixed types, numerical, ordinal, categorical, functional, etc.
3. Lack of a framework to include and distill knowledges from different resources.

Future Works





1. Multi-Modal Functional Structural Equation Modeling
2. Functional Gaussian Graphical Model with latent factors
3. Phenotype discovery in Federated Learning

Q&A

Thank you!