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

Yu Ding

Ph.D. Candidate

Systems Science and Industrial Engineering

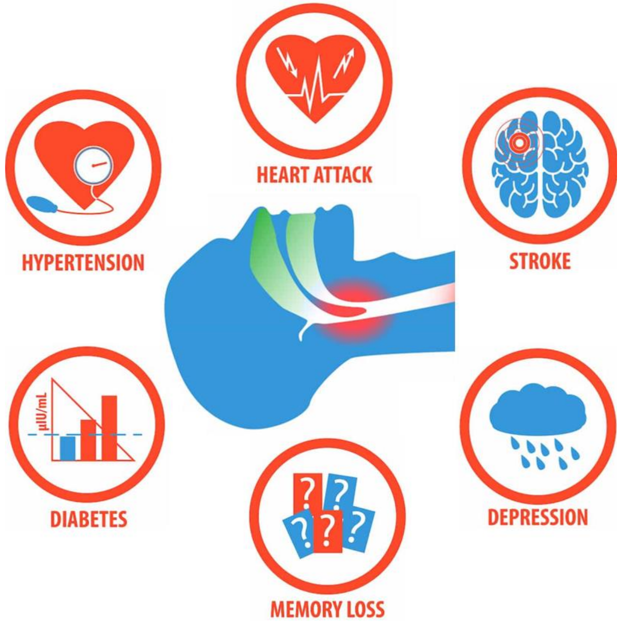
State University of New York (SUNY) at Binghamton

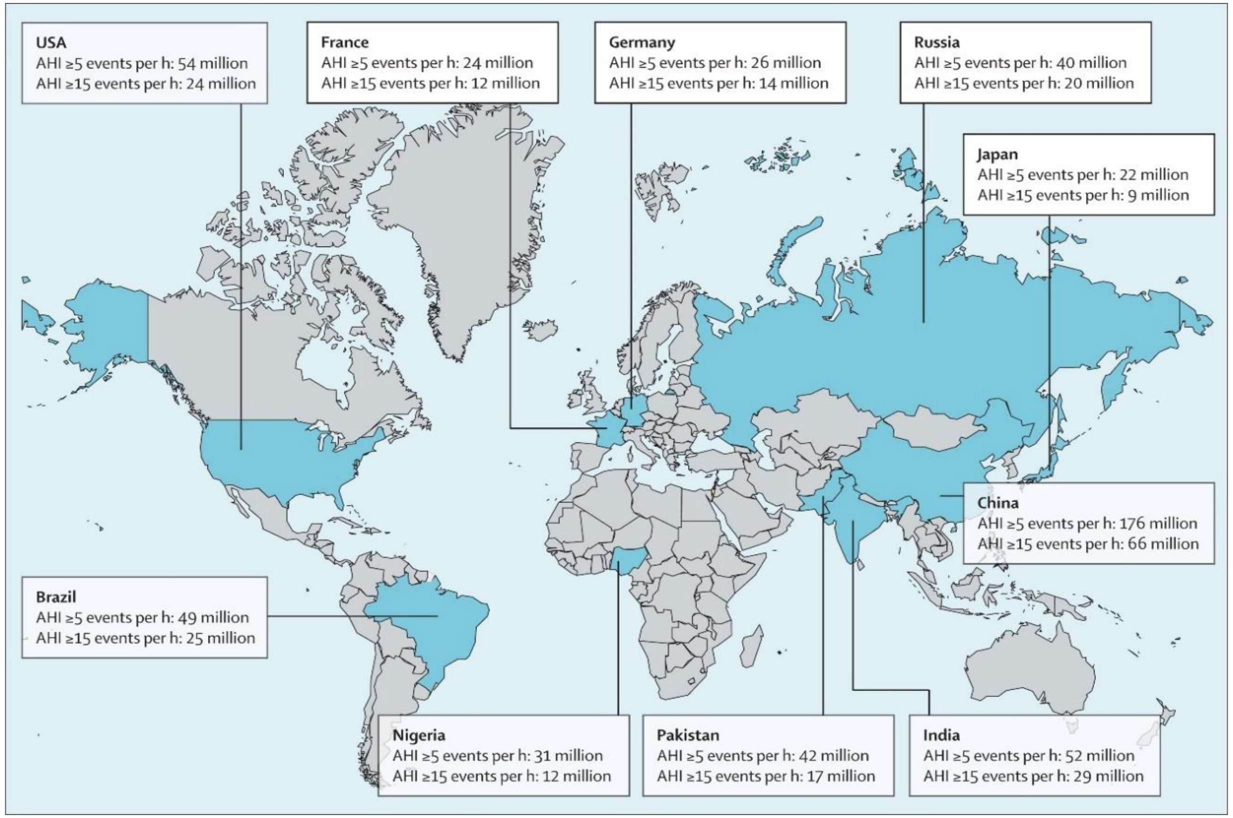
Advisor: Dr. Bing Si

Outline

- **Background**
  - Telemedicine for Obstructive Sleep Apnea (OSA)
- **Proposed Method**
  - Federated Gradient Boosting algorithm with the Least Squares Approximation
- **Results**
  - Simulation & Case Study
- **Conclusion and Future Work**

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

<u>Wearable devices</u>:

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

<u>Current diagnostic approach</u>:

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

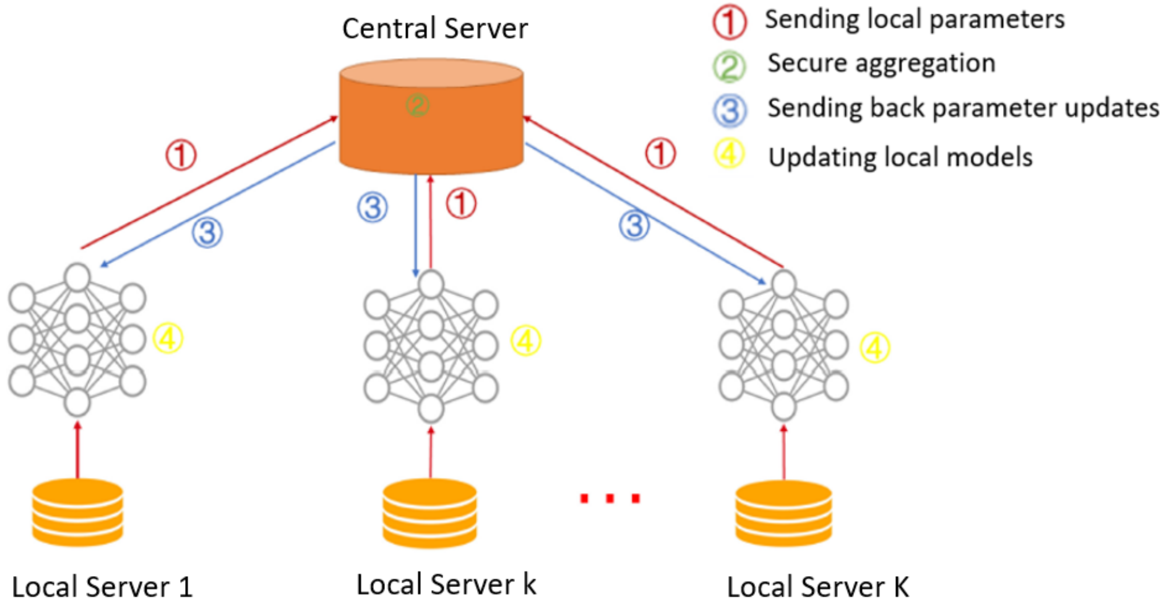
Research Problem:

- 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 $\boldsymbol{u}^{(\omega)}$

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

  - 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 $k = 1, \dots, K$; $S = \{1, \dots, N\} = \cup_{k=1}^{K} S_k$
- 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 covariance of the local estimator $\widetilde{\mathbf{B}}_{p,k}^{*}$ as $\mathbf{\Sigma}_{p,k}$, i.e., $\mathbf{\Sigma}_{p,k} = Cov(vec(\widetilde{\mathbf{B}}_{p,k}^{*}))$ and let $\mathbf{\Sigma}_{p} = \left(\sum_{k=1}^{K} \frac{N_k}{N} \Sigma_{p,k}{}^{-1}\right)^{-1}$. Assuming certain regularity conditions, we have $\sqrt{N}\left(vec(\widehat{\mathbf{B}}_p) - vec(\widetilde{\mathbf{B}}_p^{*})\right) \to_d N(0, \mathbf{\Sigma}_p)$.

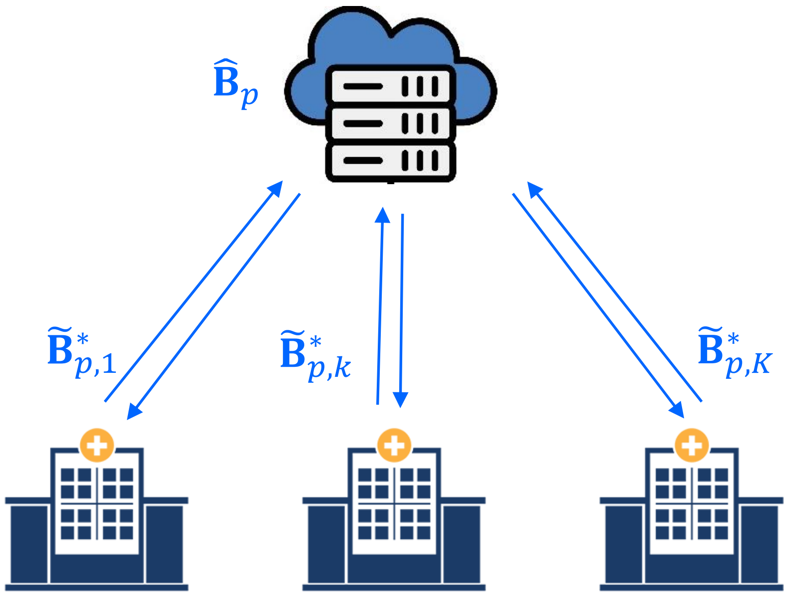
The proposed LSA estimator $\widehat{\mathrm{B}}_p$ achieves the same asymptotic normality as the global estimator $\widetilde{\mathrm{B}}_p^{*}$. $\Longrightarrow$ **Statistical Efficiency**

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

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

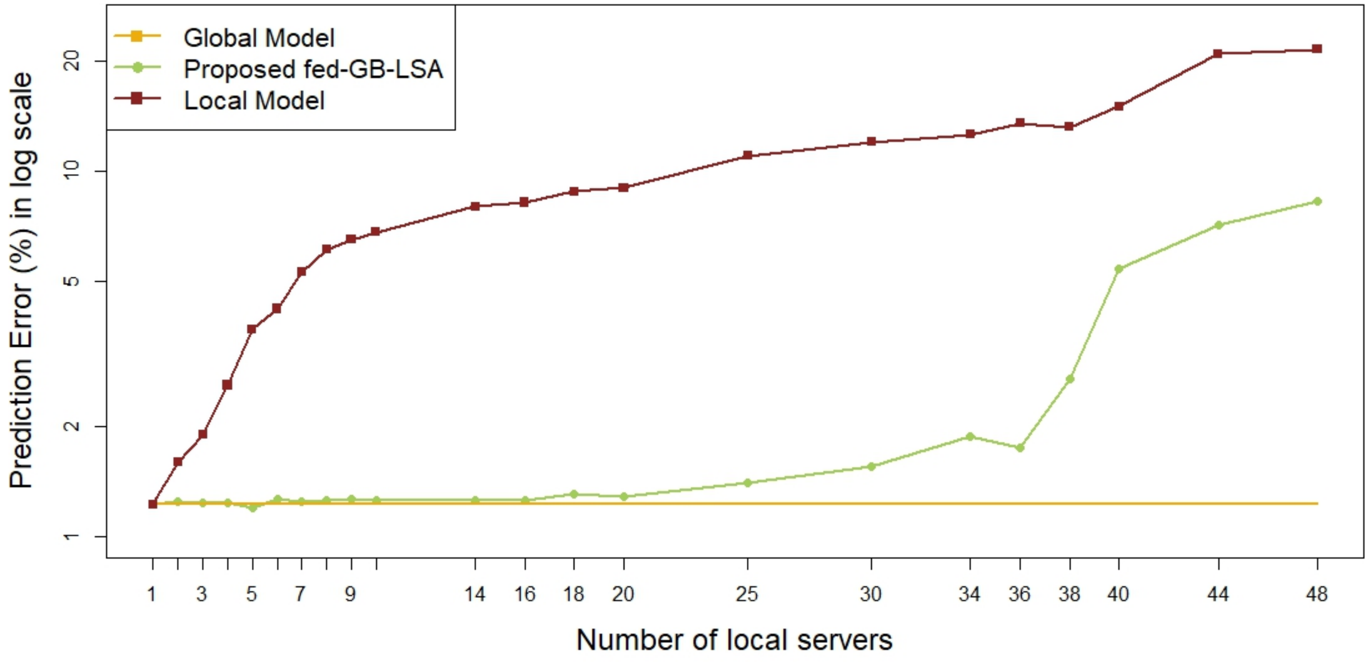
- Coefficient function: $\beta_p(s,t) = \boldsymbol{\varphi}(s)^T \mathbf{B}_p \boldsymbol{\varphi}(t)$
  - $\mathbf{B}_p$ is sampled from $N(1, 0.5)$ for effective predictor p.
  - $\mathbf{B}_p$ is set to be 0's for dummy predictor p.
- Functional predictors & response:
  - $x_{np}(t) = \sum_k c_{pk}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

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

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

	MAPE						Selection Accuracy				Computational Runtime (min)	
	Mean		Standard Deviation		Worst Case		Sensitivity		Specificity			
	LSA	Avg	LSA	Avg	LSA	Avg	LSA	Avg	LSA	Avg	LSA	Avg
K=2	2.59	2.87	0.40	0.32	3.78	3.59	0.85	0.69	0.77	0.81	1.7	2.4
K=4	2.25	2.82	0.49	0.22	3.47	3.38	0.85	0.63	0.86	0.83	3.6	5.1
K=6	2.13	2.76	0.44	0.19	3.17	3.22	0.89	0.65	0.85	0.83	5.9	8.6
K=8	1.90	2.81	0.44	0.17	3.08	3.33	0.90	0.77	0.85	0.82	8.5	12.3
K=10	1.90	2.77	0.41	0.15	3.22	3.06	0.93	0.77	0.85	0.82	10.9	16.8

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

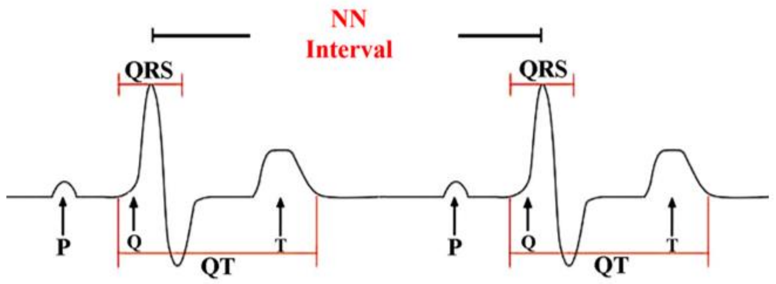
- fed-GB-LSA: $\widehat{\mathbf{B}}_p = \left(\sum_{k=1}^{K} \frac{N_k}{N} \widehat{\Sigma}_{p,k}^{-1}\right)^{-1} \left(\sum_{k=1}^{K} \frac{N_k}{N} \widehat{\Sigma}_{p,k}^{-1} \widetilde{\mathbf{B}}_{p,k}^{*}\right)$
- fed-GB-Average: $\widehat{\mathbf{B}}_p = \sum_{k=1}^{K} \frac{N_k}{N} \widetilde{\mathbf{B}}_{p,k}^{*}$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

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

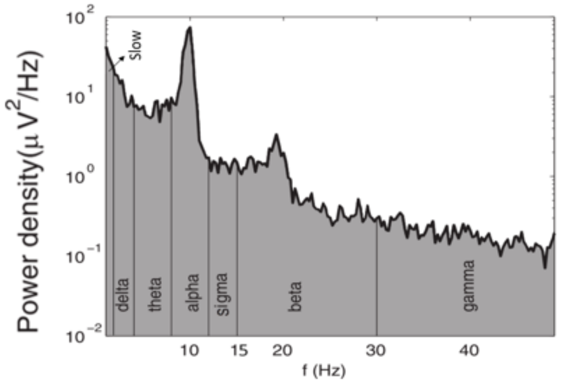
	Prediction Accuracy		Selection Accuracy				Computational	
	MAPE		Sensitivity		Specificity		Runtime (min)	
	LSA	Avg	LSA	Avg	LSA	Avg	LSA	Avg
K=2	2.59	2.87	0.85	0.69	0.77	0.81	1.7	2.4
K=4	2.25	2.82	0.85	0.63	0.86	0.83	3.6	5.1
K=6	2.13	2.76	0.89	0.65	0.85	0.83	5.9	8.6
K=8	1.90	2.81	0.90	0.77	0.85	0.82	8.5	12.3
K=10	1.90	2.77	0.93	0.77	0.85	0.82	10.9	16.8

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.



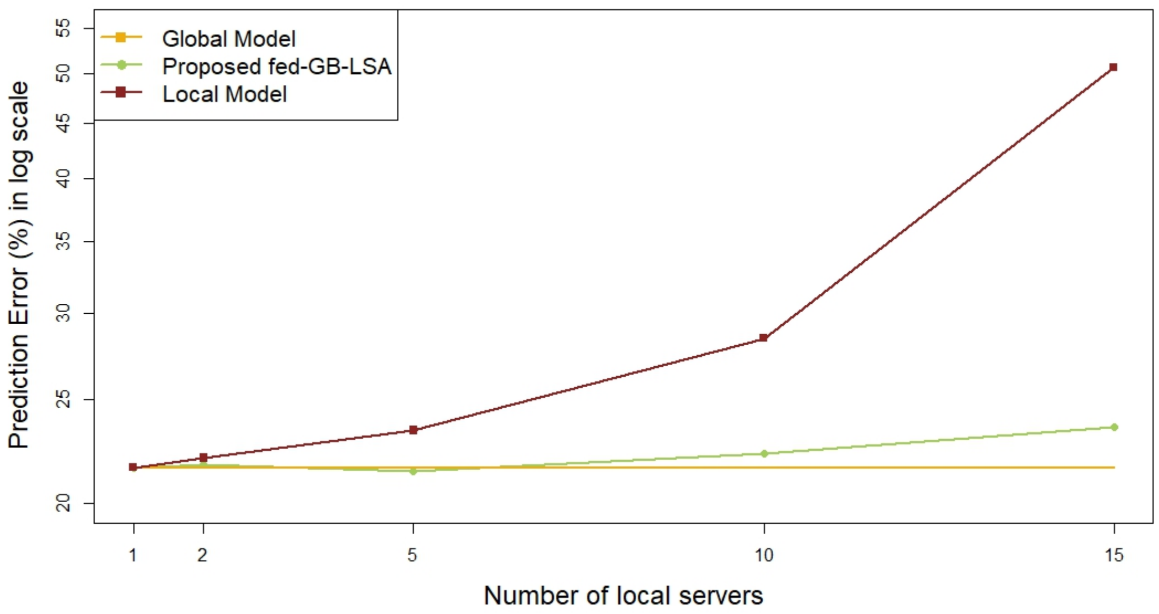
Heart Rate Variability (HRV) analysis for ECG signals



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”:

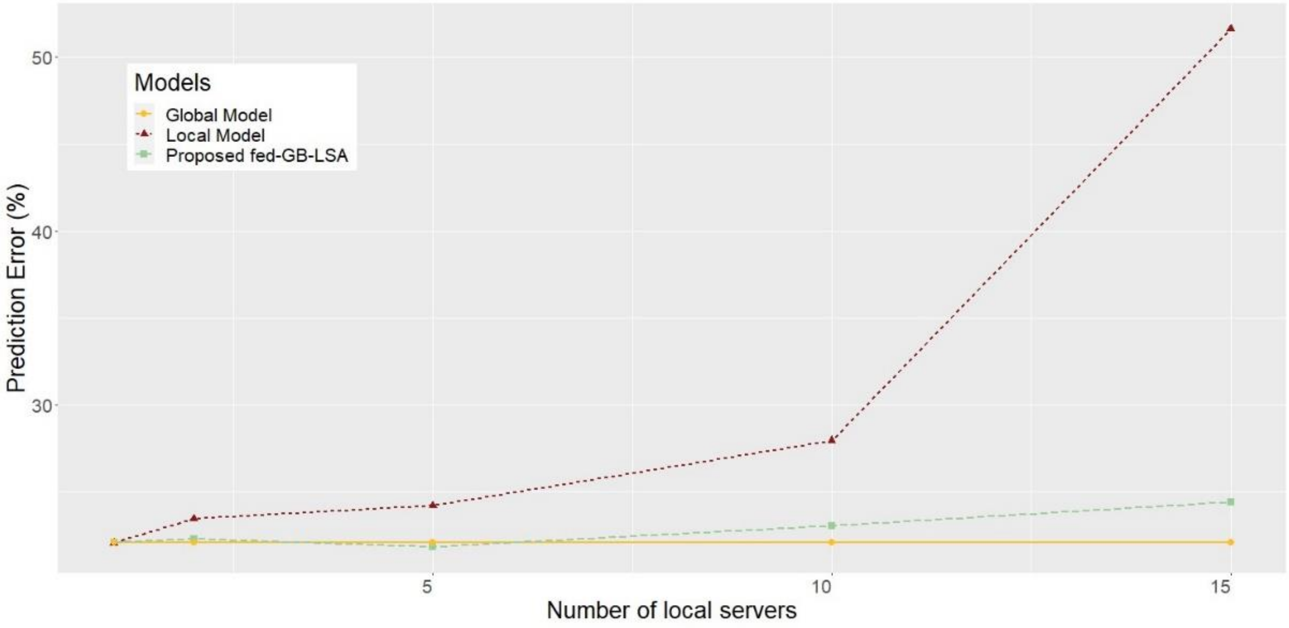


The proposed fed-GB-LSA sheds light on OSA diagnosis and telemedicine with privacy-preservation.

Increased healthcare accessibility
Improved public health

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

Conclusion

- Developed a novel Federated Learning (FL) algorithm for function-on-function regression with variable selection.
- Proposed the first-of-its-kind federated Gradient Boosting algorithm with the Least Squares Approximation (fed-GB-LSA) for efficient FL.
    - Computationally, communicationally, and statistically efficient
- Applied the proposed method to predict disease severity from functional and non-functional data, aiming to facilitate automated diagnosis and telemedicine for OSA.
- Federated functional regression with heterogeneity among local servers awaits explorations.

Conclusion and Future Work

- Developed a novel Federated Learning (FL) algorithm for function-on-function regression with variable selection.
- Proposed the first-of-its-kind federated Gradient Boosting algorithm with the Least Squares Approximation (fed-GB-LSA) for efficient FL.
    - Computationally, communicationally, and statistically efficient
- Applied the proposed method to predict disease severity from functional and non-functional data, aiming to facilitate automated diagnosis and telemedicine for OSA.
- In the future, we plan to generalize the current framework to explore federated function regression under heterogeneous settings to tackle this common challenge in FL.

Q&A

Thank you!