



HHS Public Access

Author manuscript

IJSE Trans Healthc Syst Eng. Author manuscript; available in PMC 2024 April 26.

Published in final edited form as:

IJSE Trans Healthc Syst Eng. 2023 ; 13(3): 215–225. doi:10.1080/24725579.2023.2202877.

A Novel Sparse Linear Mixed Model for Multi-Source Mixed-Frequency Data Fusion in Telemedicine

Wesam Alramadeen^{a,#}, Yu Ding^a, Carlos Costa^b, Bing Si, PhD^{a,*}

^aDepartment of Systems Science and Industrial Engineering, State University of New York at Binghamton, Binghamton, NY, USA 13902, USA

^bIBM T. J. Watson Research Center, Yorktown Heights, NY 10510, USA.

Abstract

Digital health and telemonitoring have resulted in a wealth of information to be collected to monitor, manage, and improve human health. The multi-source mixed-frequency health data overwhelm the modeling capacity of existing statistical and machine learning models, due to many challenging properties. Although predictive analytics for big health data plays an important role in telemonitoring, there is a lack of rigorous prediction model that can automatically predicts patients' health conditions, e.g., Disease Severity Indicators (DSIs), from multi-source mixed-frequency data. Sleep disorder is a prevalent cardiac syndrome that is characterized by abnormal respiratory patterns during sleep. Although wearable devices are available to administrate sleep studies at home, the manual scoring process to generate the DSI remains a bottleneck in automated monitoring and diagnosis of sleep disorder. To address the multi-fold challenges for precise prediction of the DSI from high-dimensional multi-source mixed-frequency data in sleep disorder, we propose a sparse linear mixed model that combines the modified Cholesky decomposition with group lasso penalties to enable joint group selection of fixed effects and random effects. A novel Expectation Maximization (EM) algorithm integrated with an efficient Majorization Maximization (MM) algorithm is developed for model estimation of the proposed sparse linear mixed model with group variable selection. The proposed method was applied to the SHHS data for telemonitoring and diagnosis of sleep disorder and found that a few significant feature groups that are consistent with prior medical studies on sleep disorder. The proposed method also outperformed a few benchmark methods with the highest prediction accuracy.

Keywords

linear mixed model; group lasso; multi-source mixed-frequency data; telemonitoring

*Corresponding Author: Bing Si, PhD; bsi@binghamton.edu.

#Wesam Alramadeen is currently an assistant professor in the Operations And Analytics Department at the University of Scranton, Scranton, PA, USA.

Disclosure

The authors declare no conflict of interest.

Consent and Approval

This study has been exempt from the requirement for approval by an institutional review board, because it is a secondary analysis of de-identified data with an approved data use agreement.

1. Introduction

Digital health technologies have been increasingly used to monitor, manage, and improve human health and well-being at individual and population levels, which provide timely and cost-effective solutions to many medical conditions (Nguyen, et al., 2022). Until a few years ago, digital health applications were restricted to the use of data obtained from Electronic Health Records (EHRs) and other clinical medical information, but in more recent years, the context of digital health has notably expanded with the advancements in technologies including and especially wearable sensors and devices. For example, Parkinson's Disease is known to impact patients' mobility and its progression can be closely and remotely tracked via built-in sensors in smartphones that can effectively measure the gait quality and deficits (Pierce et al., 2021; Hobert et al., 2019). Another example is the monitoring and management of migraine, a disabling, chronic, and complex neurological disorder, with the use of mobile apps that record and track migraine patients' disease-related risk factors and symptoms, such as dietary intake, exercise, work and life stress, and frequency and severity of migraine attacks for personalized disease management (Alves et al., 2021). In particular, sleep disorder is a prevalent cardiac syndrome that affects 10% of middle-aged women and 25% of middle-aged men. It is characterized by abnormal respiratory patterns during sleep and is known to be a significant contributor to short sleep duration among the United States population. According to Centers for Disease Control and Prevention (CDC), short sleepers are more likely to report chronic health conditions such as cardiovascular diseases, cancer, and depression, compared to those who got enough sleep (Consensus Conference Panel, 2015). Although effective treatment can be offered to treat sleep disorder and mitigate its risk sequelae, many patients with sleep disorder are under-diagnosed, mostly likely due to the costly and logistically inconvenient diagnostic approach for sleep disorder. The conventional diagnostic criteria are primarily based upon an overnight sleep study (Ahmadi et al., 2009) for long-term recording of patients' bio-signals such as ECG and EEG. The overnight sleep study requires patients' long-time physical presence in a specialized clinic that are costly and logistically inconvenient for both the patients and medical staff. Recently, the sleep study becomes available at home via wearable devices such as the use of the electrocardiogram (ECG) to record the cardiac activity (Fensli et al., 2005) and the electroencephalograph (EEG) to record brain activity (Askamp and van Putten, 2014), providing a promising and cost-efficient solution to remote monitoring and diagnosis of sleep disorder.

Digital health and telemonitoring result in a wealth of multi-source mixed-frequency health information. Take sleep disorder as an example. The disease severity of sleep disorder is potentially relevant to both clinical data collected at hospital and bio-signal data remotely collected at home. It is clear that bio-signal data such as an overnight recording of ECG and EEG are measured at a very high frequency. For example, with an epoch length of 30 seconds, the ECG recording for six hours may result in a total of 720 epochs for the same patient. In contrast, clinical data such as patients' age, gender, Body Mass Index (BMI), and other health conditions, would remain very similar if not exactly the same over all the epochs for obvious reasons. Thus, there is no need to repeatedly collect such clinical data very frequently, e.g., every 30 seconds. Multi-source mixed-frequency big data often

overwhelm the modeling capacity of existing statistical and machine learning models, due to many challenging properties such as the so-called “four Vs of big data” including Volume, Velocity, Variety, and Veracity (Yin and Kaynak, 2015). “Volume” refers to the amount of data being collected, which is the base of high-dimensional “big” data. “Velocity” refers to how quickly data is generated. High velocity is a common characteristic of high-frequency bio-signal data such as ECG and EEG, compared with low-frequency clinical data collected at hospital visits. “Variety” refers to the diversity of data types or data sources in which data standardization and integration are clear obstacles. “Veracity” refers to the quality and accuracy of data that may be inconsistent between data remotely collected via wearable devices at home and data collected in clinical settings.

Predictive analytics for big health data plays an important role in telemonitoring. As more and more patients’ health data become available, predictive analytics can be introduced to analyze the multi-source mixed-frequency data to facilitate cost-effective and precise prediction of patients’ health conditions, e.g., the so-called “Disease Severity Indicators” (DSIs). Despite of increasing interests in a variety of digital health applications, there are significant gaps in predictive analytics that can fully leverage big health data to automatically predict patients’ DSIs and is ready to be paired up with emerging digital health systems to enable automated disease telemonitoring and diagnosis. For example, the DSI of sleep disorder is the number of adverse respiratory events that occur during sleep. In current clinical settings, the DSIs need to be scored by the certified medical staff who review the recording of multi-channel bio-signals such as ECG and EEG and manually count the adverse respiratory events for each epoch. As a result, although wearable devices (Collop et al., 2007) are now available to administrate sleep studies at home, the manual scoring process to generate the DSI remains a bottleneck in automated monitoring and diagnosis of sleep disorder in telemonitoring.

The challenges for precise prediction of the DSI from high-dimensional multi-source mixed-frequency data are multi-folds. First, while mixed-frequency data contain rich health information, there is lack of rigorous method to integrate high-and low-frequency features of the same patient for precise prediction. Second, digital health applications typically enable high-dimensional features to be collected that are potentially relevant to patients’ health, but it is very likely that some features are not significantly predictive of patients’ DSIs and need to be eliminated from the prediction. Last but not least, conventional feature selection methods often overlook the relationships among features of multi-sources while it is more reasonable to assume grouping structures for multi-source features selection. To address these challenges, this paper proposes a sparse linear mixed model with group variable selection to simultaneously predict the DSI from multi-source mixed-frequency data with high prediction accuracy and select the significant feature groups to facilitate sparse variable selection and medical knowledge discovery. The rest of the paper is organized as follows: Section 2 reviews the relevant work; Section 3 presents the development of the proposed method; Section 4 discusses the application of proposed method for telemonitoring of sleep disorder; and Section 5 concludes the paper.

2. Relevant work

2.1. Statistical models

The proposed sparse linear mixed model is the combination of the conventional statistical linear mixed model and modern sparse selection technique. This section first reviews the two topics separately and then discusses their limited combinations in the field.

Sparse variable selection techniques are modern statistical machine learning developments that were motivated by the emerging high-dimensional data (Hastie et al., 2009). The basic idea of sparse selection is to add the lasso-type $l - 1$ penalty to the model coefficients in order to shrink the estimates of insignificant coefficients to be exactly zeros (Tibshirani, 1996). In recent years, various forms of penalties have been proposed to model different structures of the predictors to enable group or structured variable selection. For instance, Yuan and Lin (2006) developed a group lasso method with an $l - 21$ penalty that is capable of selecting a sparse set of groups by imposing an $l - 1$ penalty on the regression coefficients of predictors from each group. Jenatton et al. (2011) proposed a structured sparsity-inducing penalty by combining the lasso and group lasso. Yan and Bien (2017) also introduced a few variations of the lasso penalty to achieve desired structured sparsity relations among parameters. However, the integration of group variable selection techniques with linear mixed models is very limited.

The linear mixed model is a natural choice of statistical models to make prediction from mixed-frequency data in digital health and telemonitoring (Demidenko, 2013). It considers the relationship between the response variable and “clustered” variables in prediction and has been widely used in many domains (Magezi, 2015; Si et al., 2017). The longitudinal measurement of bio-signals over all epochs for the same patient is a classic example of clustered variables. While conventional statistical prediction models such as linear regression assume fixed effects only, i.e., regression coefficients, the linear mixed model assumes both fixed effects and random effects in prediction. Consequently, both low-and high-frequency features in the linear mixed model contribute fixed effects to predict the response variable, e.g., DSI, by assuming constant fixed coefficients between features and response across all the observations. In addition, high-frequency features also bring random effects into prediction by assuming a random distribution to account for within-patient variability. However, there are very few studies on sparse variable selection for linear mixed models, thus limiting its capacity to handle high-dimensional data.

Among a few existing efforts in sparse learning of linear mixed models, most studies can select fixed effects only, and sparse selection of random effects is more challenging. Indeed, there are significant differences in sparse variable selection between fixed effects and random effects. The fixed effects are characterized by regression coefficients that be easily selected by imposing an $l - 1$ penalty. In contrast, the random effects are characterized by a covariance matrix, Ψ , and it is not straightforward to penalize a covariance matrix using the lasso penalties. To address this challenge, the modified Cholesky decomposition was proposed to decompose the covariance matrix Ψ into $\mathbf{D}\mathbf{\Gamma}(\mathbf{D}\mathbf{\Gamma})^T$ in which \mathbf{D} is a diagonal matrix and $\mathbf{\Gamma}$ is lower triangular matrix with all the diagonal elements being ones (Ibrahim

et al., 2011; Bondell et al., 2010). If one diagonal element in \mathbf{D} is penalized to be zero, the corresponding row and column in the covariance matrix Ψ becomes zeros, eventually leading to the elimination of the corresponding random effect. However, none of these existing studies considered group or structured selection of fixed and random effects brought in by multi-source features.

2.2. Sleep disorder detection and prediction

Recently, there are emerging studies that utilize data analytical methods to detect adverse respiratory events from bio-signals collected during sleep and classify the participants into patients with sleep disorders and healthy individuals. Sharma et al. (2019) proposed an optimal two-band filter bank technique to split ECG signals into wavelet frequency bands for feature extraction and used various classifiers such as K nearest neighbor, decision tree, linear discriminant, logistic regression, and support vector machine to differentiate patients and healthy individuals with a 90.87% accuracy. Sheta et al. (2021) proposed a computer-aided diagnosis system to detect adverse respiratory events based on ECG with a number of classic machine learning and deep learning methods and achieved an accuracy of 86.25%. Wang et al. (2019) developed a time window artificial neural network that can account for the time dependence between ECG signal epochs that significantly outperformed traditional non-time window methods for sleep disorder prediction. Niroshana et al. (2021) applied a convolutional neural network to extract features from images created with ECG epochs and achieved an average accuracy of 92.4% for the fused images. Jarchi et al. (2020) proposed to extract features from ECG and EMG using entropy and statistical moments, and synchrosqueezed wavelet transform, respectively, and developed a deep learning framework to incorporate both ECG and EMG features for classification with a mean accuracy of 72%. Huysmans et al. (2021) developed a sleep-wake classifier for sleep time estimation and used the predicted sleep-wake patterns for healthy, mild, moderate, and severe patient classification.

Other types of bio-signals such as EEG have also been leveraged for sleep disorder prediction. Wang et al. used the infinite impulse response butterworth band pass filter to divide the EEG signals into different frequency sub-bands and applied Random forest, K-nearest neighbors, and bagging for classification resulting in an average accuracy of 90.43% (Wang et al., 2021). Zhao et al. used random forest, K-nearest neighbor, and support vector machine to classify EEG features extracted by the neighbor composition analysis resulting in an average accuracy of 88.99% (Zhao et al., 2021). Last, there is a recent study that combines EEG, ECG, and EMG signals to classify healthy individuals and patients by evaluating sleep healthy at different sleep stages, which achieves at least 96% for sensitivity, specificity, and accuracy (Moridani et al., 2019).

However, existing studies have a few limitations. First, most of these studies used either ECG or EEG only in the prediction without fully leveraging the complementary information from multi-source bio-signals including both ECG and EEG. More importantly, most existing studies focus on classification of patients with sleep disorder and healthy individuals, instead of providing more specific disease severity information among patients by predicting their DSIs, which limits the clinical utilization of the developed machine

learning methods. For example, patients with mild sleep disorder may be recommended to make lifestyle changes such as adapting to a healthy diet, while severe patients may need to be aggressively treated with a variety of therapeutic approaches (Chang et al., 2020). Therefore, compared to the classification of healthy individuals and patients with sleep disorder, it is more critical to predict the DSI of each patient so that different management and treatment approaches can be offered to patients at different disease severity for optimized outcomes.

3. Development of a sparse linear mixed model with group variable selection

3.1. Model formulation

The linear mixed model approaches the prediction of mixed-frequency features with a two-level model. In the epoch-level (or Level 1), DSI for each epoch is predicted from *high-frequency features*, i.e., bio-signal features, only. In the patient-level (or Level 2), the predictive relationship in Level 1 further depends on *low-frequency features*, i.e., each patient's health covariates such as age, gender, and other basic health information. Figure 1 exhibits a graphical illustration of the proposed linear mixed model.

Specifically, in the epoch-level (Level 1), we use y_{nt} and \mathbf{z}_m to denote the DSI and bio-signal features for patient n at epoch t , respectively, for $n = 1, \dots, N$ and $t = 1, \dots, T$. N is the total number of patients and T is the total number of epochs. Without loss of generality, we assume all patients have the same number of epochs to keep the notations simple. The epoch-level model predicts the DSI of patient n at epoch t (y_{nt}) from his or her bio-signal features (\mathbf{z}_m) and can be written as

$$y_{nt} = \alpha_n^0 + \mathbf{z}_m^T \boldsymbol{\alpha}_{n1} + \varepsilon_{nt}, \quad (1)$$

where $\varepsilon_{nt} \sim N(0, \sigma^2)$. The differences between model (1) and the classic linear regression model are: (a) the intercept and coefficients α_n^0 and $\boldsymbol{\alpha}_{n1}$ are patient-specific and depend on the patient index n , while the classic regression model assumes the same intercept and coefficients across all the observations; (b) observations in (1) are not independent of each other, because DSIs of the same patient across different epochs, i.e., y_{n1}, \dots, y_{nT} , are clearly dependent, thus violating the independent and identically distributed (i.i.d.) assumption of the linear regression model. In the patient-level (Level 2), the patient-specific intercept and coefficients α_n^0 and $\boldsymbol{\alpha}_{n1}$ can be further characterized by low-frequency features such as patient's age, gender, and other basic health information, denoted by \mathbf{w}_n , to describe how a patient's characteristics affect the relationship between patient's DSI and his or her bio-signal features at each epoch. Without loss of generality, we follow the widely used assumption of linear mixed models and write the patient-level model as

$$\alpha_n^0 = \mathbf{w}_n^T \boldsymbol{\beta}_x \text{ and } \boldsymbol{\alpha}_{n1} = \boldsymbol{\beta}_z + \tilde{\boldsymbol{\epsilon}}_n \text{ for } n = 1, \dots, N, \quad (2)$$

where $\tilde{\boldsymbol{\epsilon}}_n \sim N(0, \sigma^2 \boldsymbol{\Psi})$.

Then, we can substitute (2) into (1) and obtain the two-level linear mixed model that predicts the DSI from mixed-frequency features as below:

$$y_{nt} = \mathbf{w}_n^T \boldsymbol{\beta}_x + \mathbf{z}_n^T \boldsymbol{\beta}_z + \mathbf{z}_n^T \tilde{\boldsymbol{\epsilon}}_n + \varepsilon_{nt}. \quad (3)$$

where $\boldsymbol{\beta}_x \in R^{I \times 1}$ and $\boldsymbol{\beta}_z \in R^{J \times 1}$ are considered as fixed effects and $\tilde{\boldsymbol{\epsilon}}_n$ are considered as random effects. The combined linear mixed model in (3) clearly shows that both high-and low-frequency features, i.e., \mathbf{w}_n and \mathbf{z}_n , contribute fixed effects to predict the response variable, i.e., y_{nt} , and only high-frequency features, i.e., \mathbf{z}_n , contribute random effects to the prediction. After reparameterization, we can rewrite the combined model in (3) as

$$y_{nt} = \mathbf{x}_n^T \boldsymbol{\beta} + \mathbf{z}_n^T \tilde{\boldsymbol{\epsilon}}_n + \varepsilon_{nt}, \quad (4)$$

where $\mathbf{x}_n = (\mathbf{w}_n^T, \mathbf{z}_n^T)^T$, $\boldsymbol{\beta} = (\boldsymbol{\beta}_x^T, \boldsymbol{\beta}_z^T)^T$, $\tilde{\boldsymbol{\epsilon}}_n \sim N(0, \sigma^2 \boldsymbol{\Psi})$, and $\varepsilon_{nt} \sim N(0, \sigma^2)$. The fixed effects are characterized by the coefficients $\boldsymbol{\beta}$ and random effects are characterized by the covariance matrix $\boldsymbol{\Psi}$. Then, we apply a modified Cholesky decomposition to the covariance matrix of random effects and have $\boldsymbol{\Psi} = \mathbf{D}\boldsymbol{\Gamma}(\mathbf{D}\boldsymbol{\Gamma})^T$ in which \mathbf{D} is a diagonal matrix and $\boldsymbol{\Gamma}$ is lower triangular matrix with all the diagonal elements being ones. To ease the following discussion, we let $\tilde{\boldsymbol{\epsilon}}_n = \mathbf{D}\boldsymbol{\Gamma}\mathbf{e}_n$ where $\mathbf{e}_n \sim N(0, \sigma^2)$ and Model (4) becomes

$$y_{nt} = \mathbf{x}_n^T \boldsymbol{\beta} + \mathbf{z}_n^T \mathbf{D}\boldsymbol{\Gamma}\mathbf{e}_n + \varepsilon_{nt}. \quad (5)$$

For patient n , we can stack up his or her data across all epochs and have

$$\mathbf{y}_n = \mathbf{X}_n \boldsymbol{\beta} + \mathbf{Z}_n \mathbf{D}\boldsymbol{\Gamma}\mathbf{e}_n + \boldsymbol{\varepsilon}_n, \quad (6)$$

where $\mathbf{y}_n = (y_{n1}, \dots, y_{nT})^T$, $\mathbf{X}_n = [\mathbf{x}_{n1}, \dots, \mathbf{x}_{nT}]^T$, $\mathbf{Z}_n = [\mathbf{z}_{n1}, \dots, \mathbf{z}_{nT}]^T$, and $\boldsymbol{\varepsilon}_n = (\varepsilon_{n1}, \dots, \varepsilon_{nT})^T$. By pooling the data across all patients, the complete linear mixed model becomes

$$\begin{pmatrix} \mathbf{y}_1 \\ \vdots \\ \mathbf{y}_N \end{pmatrix} = \begin{pmatrix} \mathbf{X}_1 \\ \vdots \\ \mathbf{X}_N \end{pmatrix} \boldsymbol{\beta} + \begin{pmatrix} \mathbf{Z}_1 & & \\ & \ddots & \\ & & \mathbf{Z}_N \end{pmatrix} \begin{pmatrix} \mathbf{D} & & \\ & \ddots & \\ & & \mathbf{D} \end{pmatrix} \begin{pmatrix} \boldsymbol{\Gamma} & & \\ & \ddots & \\ & & \boldsymbol{\Gamma} \end{pmatrix} \begin{pmatrix} \mathbf{e}_1 \\ \vdots \\ \mathbf{e}_N \end{pmatrix} + \begin{pmatrix} \boldsymbol{\varepsilon}_1 \\ \vdots \\ \boldsymbol{\varepsilon}_N \end{pmatrix}, \text{ and then} \quad (7)$$

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}(\mathbf{I} \otimes \mathbf{D})(\mathbf{I} \otimes \boldsymbol{\Gamma})\mathbf{e} + \boldsymbol{\varepsilon}.$$

In Equation (7), the parameters to be estimated include $\boldsymbol{\beta}$, \mathbf{D} , $\boldsymbol{\Gamma}$, and σ^2 that can be organized into a vector $\boldsymbol{\phi} = (\boldsymbol{\beta}^T, \mathbf{d}^T, \boldsymbol{\gamma}^T, \sigma^2)^T$. We assume the total numbers of fixed effects and random effects are $P = I + J$ and $Q = J$, respectively. Then, we have $\boldsymbol{\beta} \in R^{P \times 1}$, $\mathbf{d} \in R^{Q \times 1}$ consisting of all the diagonal elements in the diagonal matrix \mathbf{D} , and a vector $\boldsymbol{\gamma}$ consisting of all the non-zero elements in the lower triangular matrix $\boldsymbol{\Gamma}$. After dropping constant terms, the complete log-likelihood function of the linear mixed model in (7) can be written as

$$l(\boldsymbol{\phi} | \mathbf{y}, \mathbf{e}) = -\frac{NT + NQ}{2} \log(\sigma^2) - \frac{1}{2\sigma^2} (\|\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}(\mathbf{I} \otimes \mathbf{D})(\mathbf{I} \otimes \boldsymbol{\Gamma})\mathbf{e}\|^2 + \mathbf{e}^T \mathbf{e}). \quad (8)$$

The challenge, however, is that we don't know which fixed and random effects contribute to the prediction of DSIs so we propose to adopt sparse learning approaches such as lasso-based penalties to select fixed effects and random effects in model estimation. Moreover, we propose to add an $l - 21$ penalty based on features grouped by their sources to the complete log-likelihood function in (8), resulting in the following optimization problem:

$$\min_{\boldsymbol{\phi}} g(\boldsymbol{\phi}) = \min_{\boldsymbol{\phi}} \left\{ -l(\boldsymbol{\phi} | \mathbf{y}, \mathbf{e}) + \lambda \left(\sum_{g_1=1}^{G_1} \|\boldsymbol{\beta}^{g_1}\|^2 + \sum_{g_2=1}^{G_2} \|\mathbf{d}^{g_2}\|^2 \right) \right\}, \quad (9)$$

in which we refer to features' sources to divide $\boldsymbol{\beta}$ and \mathbf{d} into G_1 and G_2 groups, respectively. For example, all depression questions to measure patients' mental status are considered low-frequency features that bring in fixed effects only, and their fixed coefficients can be put in the same group, i.e., $\boldsymbol{\beta}^{g_1}$; while all the high-frequency ECG features bring in both fixed and random effects, and their fixed and random effects are penalized as two groups, i.e., $\boldsymbol{\beta}^{g_1}$ and \mathbf{d}^{g_2} . Note that Equation (9) applied the $l - 21$ penalty to all features to simplify the formula. However, we can choose to hold off the sparse selection for some features by adding different weights to each group. This flexibility of group lasso is very useful for healthcare applications in which some critical features such as age, gender, and race are known to be significant contributors to disease severity according to medical knowledge and should always be kept in the prediction models.

3.2. Model estimation by integrating EM algorithm with an efficient MM algorithm

The model estimation of the sparse linear mixed model results in the optimization problem as shown in Equation (9). There are two challenges in solving this optimization problem. First, the objective function in (9) contains unobserved latent variables \mathbf{e} in addition to observed data, and thus relies on the Expectation Maximization (EM) algorithm for the model estimation. EM algorithm is an iterative algorithm with E-steps and M-steps. In the $(\omega + 1)$ -th iteration, the E-step calculates the expectation of unobserved variables based on their conditional probabilities given observed data, i.e., $f(\mathbf{e} | \mathbf{y}; \boldsymbol{\phi}^{(\omega)})$, and then the subsequent M-step aims to maximize the expected log-likelihood function, i.e., $E_{\mathbf{e} | \mathbf{y}; \boldsymbol{\phi}^{(\omega)}}(g(\boldsymbol{\phi}; \boldsymbol{\phi}^{(\omega)}))$, to obtain an updated set of parameters $\boldsymbol{\phi}^{(\omega + 1)}$. The second challenge, however, is how to efficiently solve the non-smooth maximization problem with $l - 21$ penalty in M-steps, while the conventional maximization algorithms such as Block Coordinate Gradient Descent (Meier et al., 2008) and Nesterov's method (Liu et al., 2009) are computationally expensive. To address these challenges, this paper proposes to integrate the EM algorithm with a Majorization Maximization (MM) algorithm to efficiently solve the optimization problems in M-steps.

In the E-step at the $(\omega + 1)$ -th iteration, we first derive the conditional distribute of latent variables e , i.e., $f(e | y; \phi^{(\omega)})$, which is a normal distribution with mean $\hat{e}^{(\omega)}$ and covariance $\Sigma^{(\omega)}$ as below:

$$\begin{aligned} & \hat{e}^{(\omega)} \\ &= \left((\mathbf{I} \otimes \Gamma^{(\omega)})^T (\mathbf{I} \otimes \mathbf{D}^{(\omega)})^T \mathbf{Z}^T \mathbf{Z} (\mathbf{I} \otimes \mathbf{D}^{(\omega)}) (\mathbf{I} \otimes \Gamma^{(\omega)}) + \mathbf{I} \right)^{-1} \left(\mathbf{Z} (\mathbf{I} \otimes \mathbf{D}^{(\omega)}) (\mathbf{I} \otimes \Gamma^{(\omega)}) \right)^T \\ & \quad (\mathbf{y} - \mathbf{X} \boldsymbol{\beta}^{(\omega)}); \end{aligned} \quad (10)$$

$$\Sigma^{(\omega)} = \widehat{\sigma}^2^{(\omega)} \left((\mathbf{I} \otimes \Gamma^{(\omega)})^T (\mathbf{I} \otimes \mathbf{D}^{(\omega)})^T \mathbf{Z}^T \mathbf{Z} (\mathbf{I} \otimes \mathbf{D}^{(\omega)}) (\mathbf{I} \otimes \Gamma^{(\omega)}) + \mathbf{I} \right)^{-1}. \quad (11)$$

Then, we can use (10) and (11) to calculate the expectation of the objective function with respect to the conditional distribution, i.e., $E_{e|x, \phi^{(\omega)}}(g(\phi; \phi^{(\omega)}))$. After dropping non-relevant terms and expect out the latent variables, the expectation can be explicitly written as

$$\begin{aligned} g'(\phi; \phi^{(\omega)}) &= \frac{1}{2\sigma^2} \left(\begin{bmatrix} \boldsymbol{\beta} \\ \mathbf{d} \end{bmatrix}^T \begin{bmatrix} \mathbf{X}^T \mathbf{X} & \mathbf{A}_{12} \\ \mathbf{A}_{12}^T & \mathbf{A}_{22} \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta} \\ \mathbf{d} \end{bmatrix} - 2\mathbf{y}^T [\mathbf{X} \ \mathbf{B}_2] \begin{bmatrix} \boldsymbol{\beta} \\ \mathbf{d} \end{bmatrix} \right) + \lambda \\ & \quad \left(\sum_{g_1=1}^{G_1} \|\boldsymbol{\beta}^{g_1}\|^2 + \sum_{g_2=1}^{G_2} \|\mathbf{d}^{g_2}\|^2 \right). \end{aligned} \quad (12)$$

Some of the terms in (12) are shown as below:

$$\mathbf{A}_{12} = \mathbf{X}^T \mathbf{Z} \text{Diag} \left((\mathbf{I} \otimes \Gamma^{(\omega)}) \hat{e}^{(\omega)} \right) (\mathbf{1}_N \otimes \mathbf{I}), \quad (13)$$

$$\mathbf{A}_{22} = (\mathbf{1}_N \otimes \mathbf{I})^T \left((\mathbf{Z}^T \mathbf{Z}) \circ \left((\mathbf{I} \otimes \Gamma^{(\omega)}) \left(\Sigma^{(\omega)} + \hat{e}^{(\omega)} \hat{e}^{(\omega)T} \right) (\mathbf{I} \otimes \Gamma^{(\omega)})^T \right) \right) (\mathbf{1}_N \otimes \mathbf{I}), \text{ and} \quad (14)$$

$$\mathbf{B}_2 = \mathbf{Z} \text{Diag} \left((\mathbf{I} \otimes \Gamma^{(\omega)}) \hat{e}^{(\omega)} \right) (\mathbf{1}_N \otimes \mathbf{I}), \quad (15)$$

where “ \circ ” is the Hadamard product operator.

In the subsequent M-step, all parameters in $\phi = (\boldsymbol{\beta}^T, \mathbf{d}^T, \boldsymbol{\gamma}^T, \sigma^2)^T$ can be iteratively estimated and $\boldsymbol{\gamma}$ and σ^2 have closed forms in each iteration. $\boldsymbol{\beta}$ and \mathbf{d} need to be jointly estimated through an efficient MM algorithm. Compared to conventional optimizers, the MM algorithm tackles the maximization of a complex objective function such as Equation (12) in our model by finding and maximizing a surrogate object function. Because the surrogate model is typically simpler than the original objective function, the maximization of the surrogate mode can often be analytically solved with a closed-form formula and therefore the MM algorithm is known to be more computationally efficient than most conventional optimizers such as Block Coordinate Gradient Descent method and Nesterov’s method. However, to apply the MM algorithm, the objective function in (12), specifically

its loss function without the $l - 21$ penalty, needs to satisfy a Quadratic Majorization (QM) condition.

To adopt the MM algorithm for the M-step, we first present the definition of QM conditions as below. Then, we derive the proposition to show that the unpenalized loss function in (12) satisfies the QM conditions so that the efficient MM algorithm is applicable to the proposed model.

Definition:

QM condition (Yang and Zou, 2015): We let t and D to denote the parameters and working data, respectively. The loss function $L(t | D)$ satisfies the QM condition if and only if the following two assumptions hold:

- i. $L(t | D)$ is differentiable as a function of t , i.e., $\nabla L(t | D)$ exists everywhere.
- ii. There exists a matrix H that only depends on the data D such that for any t and t^* ,

$$L(t | D) \leq L(t^* | D) + (t - t^*)^T \nabla L(t^* | D) + \frac{1}{2}(t - t^*)^T H(t - t^*). \quad (16)$$

Proposition:

The loss function without the $l - 21$ penalty in (12) of the proposed sparse linear mixed model satisfies the QM condition.

Proof: The proposed objective function in (12) has parameters β and d to be estimated. We denote $t = (\beta^T, d^T)^T$ and the loss function without the $l-21$ penalty in (12) becomes

$$L(t | D = \{\mathbf{A}, \mathbf{b}, \sigma^2\}) = \frac{1}{2\sigma^2}(\mathbf{t}^T \mathbf{A} t - 2\mathbf{b}^T t), \quad (17)$$

where $\mathbf{A} = \begin{bmatrix} \mathbf{X}^T \mathbf{X} & \mathbf{A}_{12} \\ \mathbf{A}_{12}^T & \mathbf{A}_{22} \end{bmatrix}$ and $\mathbf{b}^T = \mathbf{y}^T [\mathbf{X} \ \mathbf{B}_2]$.

For any t and $t^* \in R^{P+Q}$, we denote $t - t^* = v$ and define $k(s) = L(t^* + sv | D)$ so that $k(0) = L(t^* | D)$ and $k(1) = L(t | D)$. Based on the Mean Value Theorem (Rudin, 1976), we can find a value $a \in (0,1)$ such that

$$k(1) = k(0) + k'(0) + k''(a), \quad (18)$$

where $k'(0)$ and $k''(a)$ can be derived as

$$k'(0) = \frac{\partial L(t^* + sv | D)}{\partial s} \Big|_{s=0} = v^T \nabla L(t^* | D) = (t - t^*)^T \nabla L(t^* | D); \quad (19)$$

$$\begin{aligned}
k''(a) &= \left. \frac{\partial^2 L(t + s\mathbf{v} \mid \mathbf{D})}{\partial s^2} \right|_{s=a} = \frac{\partial(t + s\mathbf{v})^T}{\partial s} \frac{\partial^2 L(t + s\mathbf{v} \mid \mathbf{D})}{\partial(t + s\mathbf{v})^2} \frac{\partial(t + s\mathbf{v})}{\partial s} \bigg|_{s=a} \\
&+ \left(\frac{\partial L(t + s\mathbf{v} \mid \mathbf{D})}{\partial(t + s\mathbf{v})} \right)^T \frac{\partial^2(t + s\mathbf{v})}{\partial s^2} \bigg|_{s=a} = \mathbf{v}^T \frac{\mathbf{A}}{\sigma^2} \mathbf{v}.
\end{aligned} \quad (20)$$

To obtain Equation (16) in QM conditions' definition, we need to find a matrix \mathbf{H} such that

$$k''(a) = \mathbf{v}^T \frac{\mathbf{A}}{\sigma^2} \mathbf{v} \leq \frac{1}{2} \mathbf{v}^T \mathbf{H} \mathbf{v}. \text{ Note that this is trivial since the matrix } \mathbf{A} \text{ only depends on data } \mathbf{D}.$$

Then, we have

$$k(1) = k(0) + k'(0) + k''(a) = k(0) + k'(0) + k\mathbf{v}^T \frac{\mathbf{A}}{\sigma^2} \mathbf{v} \leq k(0) + k'(0) + \frac{1}{2} \mathbf{v}^T \mathbf{H} \mathbf{v}. \quad (21)$$

Since we have $k(1) = L(t \mid \mathbf{D})$, $k(0) = L(t^* \mid \mathbf{D})$, and $k'(0) = (t - t^*)^T \nabla L(t^* \mid \mathbf{D})$, Equation (21) can be rewritten as

$$L(t \mid \mathbf{D}) \leq L(t^* \mid \mathbf{D}) + (t - t^*)^T \nabla L(t^* \mid \mathbf{D}) + \frac{1}{2} (t - t^*)^T \mathbf{H} (t - t^*).$$

4. Application in telemonitoring of sleep disorder

4.1. Data description and bio-signal processing

This section illustrates the application of the proposed sparse linear mixed model with group variable selection using data collected in the Sleep Heart Health Study (SHHS) (Quan et al., 1997, Zhang et al., 2018). The SHHS is an epidemiological study on sleep disorder in the United States. After examining the quality and reliability of bio-signal recordings and data missingness, this study randomly selected 100 subjects with 20 epochs for each subject, resulting in a total number of 2,000 observations, i.e., $N = 100$, $T = 20$, and $N \times T = 2,000$. In this application of sleep disorder, the epoch length is chosen to be 5 minutes due to the characteristics of its DSIs, i.e., the frequency of adverse respiratory events that is typically between 5–30 per hour for patients with mild and moderate sleep disorder. A shorter epoch length may result in many observed epochs with DSIs being zeros and cause data imbalance, while a longer epoch length may overlook the longitudinal variation of DSIs for the same patient. In general, the epoch length can be customized based on the health conditions to be monitored, as long as the total number of patients and total number of epochs per patient are able to provide sufficient statistical power to estimate the fixed effects and random effects in the sparse linear mixed models. The SHHS dataset contains rich multi-source mixed-frequency features that are potentially predictive of the severity of sleep disorder. As described in Section 3, all the low-frequency features contribute fixed effects only, while the high-frequency features contribute both fixed and random effects. Next, we present the low-and high-frequency features included in this case study. An overview of features' descriptive statistics is depicted in Table 1.

The patient's low-frequency features contain basic health information including age, gender, ethnicity, and BMI, and other health conditions from different sources to measure the patient's mental health, sleep habits, and general health status. For example, the patient's mental status is known to be associated with his or her sleep health so that the SHHS study examined patient's depression status by administering a depression screening questionnaire with questions such as during the past 4 weeks, how much of the time: "Have you felt downhearted and blue"; "Have you been a very nervous person"; and "Did you feel tired?" The answer is on the five-Likert scale with 1 representing "never" and 5 representing "almost always". Then, all the depression questions are considered as one group in the sparse selection of the proposed linear mixed model. That is, if all the fixed effects brought in by the depression questions are penalized to be zeros, then it implies that the patient's depression status is not significantly predictive of his or her DSI in the presence of other predictors in the proposed model. It is worth noting that patient's basic information such as age, gender, ethnicity, and BMI are widely used as clinical indicators of disease severity, and should not be selected off in the proposed method. To incorporate this medical knowledge into prediction, the group lasso can be generalized to assign separate penalty weights to different groups that allow differential shrinkage for select feature groups. Therefore, without loss of generality, we can assign a separate weight with no shrinkage in group lasso to avoid the basic important information such as age, gender, ethnicity, and BMI being eliminated from the prediction model.

The patient's high-frequency data contain ECG-derived and EEG-derived features, along with the frequency of adverse respiratory events, i.e., DSIs, per epoch. For each patient, we first identified the epochs with both reliable ECG and EEG signals available and then collected features from both signals. The ECG signals were processed by the Heart Rate Variability (HRV) analysis to extract the morphology of the waves and intervals on the ECG curve (Qin et al., 2021). As shown in Figure 2 (a), the QRS complex is the most well-known waveform showing electrical activity inside the heart, and the interval between R peaks in two adjacent QRS complexes is defined as the "NN interval" (Normal-to-Normal interval to emphasize that the heartbeats are normal). HRV features included in this study are Average and Standard Deviation of NN intervals (AVNN and SDNN), NN < 10 ms counts divided by the total number of all NN intervals (pNN10), and other frequency domain features with Very low, low, and high frequencies (VLF, LF, and HF). All the six HRV features are considered as one group that captures the ECG information. The EEG signals were processed by the Power Spectral Analysis (PSD) (Stoica and Mosees, 2005; Hayes, 2009). First, we decompose the EEG signals into distinct frequency sub-bands including Slow (0.5–1 Hz), Delta (1–4 Hz), Theta (4–8 Hz), Alpha (8–12 Hz), Sigma (12–15 Hz), Beta (15–30 Hz), and Gamma (30+ Hz). Then, we used PSD to calculate the power distribution of EEG by estimating the area under the power density curve for each sub-band as shown in Figure 2 (b), resulting in seven EEG-derived features, i.e., mean of spectral band power for the seven sub-bands. All the seven PSD features are considered as one group that captures the EEG information.

4.2. Results and discussion

We applied the proposed sparse linear mixed model with group variable selection in Section 3 to the SHHS dataset. The regularized parameter λ for the group lasso in Equation (9) can be selected based on Bayesian Information Criterion (BIC) that balances the model fitness and complexity. BIC takes the form of $-2\tilde{l} + \log(N) \times df$ where \tilde{l} is the log-likelihood of the observed data and df is the degree of freedom that depends on the number of non-zero parameters in the model. Using the exhaustive grid search, we chose the optimal λ that minimizes BIC.

In the optimal model, both ECG- and EEG-derived features were selected as significant fixed and random effects in the prediction of DSIs at epoch-level, which were consistent with prior studies on the association between sleep disorder and multi-channel bio-signals (Stein and Pu, 2012; Zhou et al., 2020). Among the three types of low-frequency survey data, i.e., mental health, sleep habits, and daytime sleepiness, only the daytime sleepiness group was selected as significant fixed effects in predicting the severity of sleep disorder. Although existing studies showed that patients' depression and sleep habits may be related to their sleep health, the daytime sleepiness questionnaire, also referred to as the Epworth Sleepiness Scale (ESS), is widely used in the field of sleep medicine as a subjective measure of patients' sleepiness (Rosenthal and Dolan, 2008; Belgü et al., 2015). Finally, since no penalty were imposed on patients' age, gender, ethnicity, and BMI, all the four variables remained in the final prediction model. The estimates of their fixed effects are all positive (coefficients of 0.23, 0.02, 0.02, and 0.11 for age, gender, ethnicity, and BMI, respectively) indicating that subjects that are "older", "male", "non-Hispanic", or "with a higher BMI" are positively associated with a higher DSI, and thus may be more likely to have more severe sleep disorder, most of which are consistent with medical findings on sleep disorder (Deng et al., 2014; Jehan et al., 2017). The only exception is the finding that shows "non-Hispanic" people may be associated with more severity sleep disorder, while more and more medical studies show that Hispanic minorities may be more susceptible to sleep disorder (Redline et al., 2014). Although sleep disorder is more prevalent in Hispanics compared with the white majority, the Hispanic population is heterogeneous and Hispanics from different origins such as Cuban, Puerto Rican, and Mexican may have distinctly different lifestyles and health conditions (Merchant et al., 2015). Therefore, one possible explanation is that the origin of Hispanic participates in the SHHS study affects the association between sleep disorder severity and ethnicity, i.e., Hispanic versus non-Hispanic.

Moreover, we compare the prediction accuracy of the proposed model with a few benchmarks such as Regression Tree, Linear Regression, Support Vector Regression, and Random Forest. To avoid overfitting, we randomly split the SHHS dataset into a training set and a testing set with different splitting percentages from 30%, 50%, 70%, to 90%. We adopt different splitting percentages to test the model's performance with different sample sizes. The Mean Absolute Prediction Error (MAPE) is used to calculate the prediction accuracy. To account for randomness in data splitting, we repeatedly performed the similar comparison for 50 times by randomly splitting the dataset into a training set and a testing set using distinct random number generators. Both the mean value and standard deviation of MAPEs across the 50 replicates are reported in Figure 3.

As shown in Figure 3, the proposed method achieved the lowest MAPE compared with other competing methods under four different scenarios. As we increase the percentage of training data from 30%, 50%, 70%, to 90%, all the methods obtained smaller MAPEs and performed better as expected. The proposed method still consistently outperforms the competing methods with significant P-values of <0.001 , <0.001 , <0.001 , and 0.02 for the four scenarios, respectively, most likely because all the competing methods do not consider the heterogeneity in mixed-frequency features as what we did for the proposed sparse linear mixed model. Moreover, it is noteworthy that standard deviations of MAPEs increase for most of the models including the proposed model as more data is used for model training. In particular, when 90% of the data is used for training, the standard deviations of MAPEs in the testing data across the 50 replicates drastically increase for all the methods. This observation may indicate overfitting for many of the replicates in this extreme scenario that 90% of data is used for training and only 10% is left for testing.

5. Conclusion

Digital health and telemonitoring have resulted in a wealth of information to be collected to monitor, manage, and improve human health. However, the multi-source mixed-frequency health data overwhelmed the modeling capacity of existing statistical and machine learning models, due to many challenging properties. Therefore, although predictive analytics for big health data plays an important role in telemonitoring, there is a lack of rigorous prediction model that can automatically predicts patients' health conditions, e.g., DSIs, from the multi-source mixed-frequency data, which is ready to be paired up with emerging digital health systems to enable automated disease telemonitoring and diagnosis. Take sleep disorder as an example. Sleep disorder is a prevalent cardiac syndrome that is characterized by abnormal respiratory patterns during sleep. Although effective treatment can be offered to treat sleep disorder and mitigate its risk sequelae, many patients with sleep disorder are under-diagnosed, mostly likely due to the costly and logistically inconvenient diagnostic approach for sleep disorder. Although wearable devices are now available to administrate sleep studies at home, the manual scoring process to generate the DSI remains a bottleneck in automated monitoring and diagnosis of sleep disorder.

The challenges for precise prediction of the DSI from high-dimensional multi-source mixed-frequency data in sleep disorder are multi-folds. First, the severity of sleep disorder is potentially relevant to both clinical data collected at hospital and bio-signal data remotely collected at home. While bio-signal data such as an overnight recording of ECG and EEG are measured at a very high frequency, clinical data such as patients' age, gender, BMI, and other health conditions, would remain very similar if not exactly the same over all epochs and are considered low-frequency data. To leverage both high-and low-frequency data in prediction, we proposed to formulate the prediction model using a statistical linear mixed model. The second challenge, however, is how to simultaneously estimate the linear mixed model and select fixed and random effects from multi-source features groups. While the sparse selection of fixed effects, characterized by regression coefficients, is relatively simple, random effects are characterized by the covariance matrix and their sparse selection is not straightforward. To address this, we proposed to adopt the modified Cholesky decomposition with the group lasso penalties to enable joint group selection of fixed effects and random

effects. Last but not least, we developed an EM algorithm integrated with an efficient MM algorithm for model estimation of the proposed sparse linear mixed model with group variable selection.

Finally, the proposed method was applied to the SHHS data for telemonitoring and diagnosis of sleep disorder. The ECG and EEG signals are respectively processed by the HRV analysis and PSD method to prepare the ECG-and EEG-derived features groups, to be integrated with the low-frequency data including patients' age, gender, ethnicity, BMI, and other health conditions such as mental status, daytime sleepiness, and general health information. Then, we predicted patients' DSIs from their multi-source mixed-frequency data and found a few significant feature groups that are consistent with prior medical studies on sleep disorder. Moreover, the proposed method was compared with a few benchmarks and achieved the highest prediction accuracy.

Acknowledgements

The authors would like to thank the investigators and researchers of the SHHS for collecting, preprocessing, and sharing the data. The SHHS was supported by National Heart, Lung, and Blood Institute cooperative agreements U01HL53916 (University of California, Davis), U01HL53931 (New York University), U01HL53934 (University of Minnesota), U01HL53937 and U01HL64360 (Johns Hopkins University), U01HL53938 (University of Arizona), U01HL53940 (University of Washington), U01HL53941 (Boston University), and U01HL63463 (Case Western Reserve University).

Funding

This work was partially supported by the National Institutes of Health (NIH) under Grants R03HD108477 and R21HL161765; State University of New York (SUNY); and IBM.

Role of the funder

The opinions expressed in this document do not reflect the official position of NIH, U.S. Department of Health and Human Services, SUNY, or IBM.

References

- Ahmadi N, Shapiro GK, Chung SA, Shapiro CM. Clinical diagnosis of sleep apnea based on single night of polysomnography vs. two nights of polysomnography. *Sleep and Breathing*. 2009 Aug;13(3):221–6. [PubMed: 19067010]
- Alramadeen W, Rababa S, Costa C, Si B. Multi-level Multi-channel Bio-signal Analysis for Health Telemonitoring. In 2022 IEEE 18th International Conference on Automation Science and Engineering (CASE) 2022 Aug 20 (pp. 1539–1544). IEEE.
- Alves TV, da Hora Rodrigues KR, Ponti MA. Interactive protocol for acquisition of migraine diaries with a mobile app and machine learning data analysis. In Proceedings of the XX Brazilian Symposium on Human Factors in Computing Systems 2021 Oct 18 (pp. 1–9).
- Askamp J, van Putten MJ. Mobile EEG in epilepsy. *International journal of psychophysiology*. 2014 Jan 1;91(1):30–5 [PubMed: 24060755]
- Belgü AU, Erdo an B, San T, Gürkan E. The relationship between AHI, Epworth scores and sleep endoscopy in patients with OSAS. *European Archives of Oto-Rhino-Laryngology*. 2015 Jan;272(1):241–5. [PubMed: 25091323]
- Bondell HD, Krishna A, Ghosh SK. Joint variable selection for fixed and random effects in linear mixed-effects models. *Biometrics*. 2010 Dec;66(4):1069–77. [PubMed: 20163404]
- Chang HP, Chen YF, Du JK. Obstructive sleep apnea treatment in adults. *The Kaohsiung journal of medical sciences*. 2020 Jan;36(1):7–12. [PubMed: 31512369]

- Collop NA, Anderson WM, Boehlecke B, Claman D, Goldberg R, Gottlieb DJ, Hudgel D, Sateia M, Schwab R. Clinical guidelines for the use of unattended portable monitors in the diagnosis of obstructive sleep apnea in adult patients. *J Clin Sleep Med*. 2007 Dec 15;3(7):737–47. [PubMed: 18198809]
- Consensus Conference Panel:, Watson NF, Badr MS, Belenky G, Bliwise DL, Buxton OM, Buysse D, Dinges DF, Gangwisch J, Grandner MA, Kushida C. Joint consensus statement of the American Academy of Sleep Medicine and Sleep Research Society on the recommended amount of sleep for a healthy adult: methodology and discussion. *Journal of Clinical Sleep Medicine*. 2015 Aug 15;11(8):931–52. [PubMed: 26235159]
- Demidenko E *Mixed models: theory and applications* with R. John Wiley & Sons; 2013 Aug 26.
- Deng X, Gu W, Li Y, Liu M, Li Y, Gao X. Age-group-specific associations between the severity of obstructive sleep apnea and relevant risk factors in male and female patients. *PLoS One*. 2014 Sep 11;9(9):e107380. [PubMed: 25211035]
- Dong JG. The role of heart rate variability in sports physiology. *Experimental and therapeutic medicine*. 2016 May 1;11(5):1531–6. [PubMed: 27168768]
- Fensli R, Gunnarson E, Gundersen T. A wearable ECG-recording system for continuous arrhythmia monitoring in a wireless tele-home-care situation. In *18th IEEE Symposium on Computer-Based Medical Systems (CBMS'05)* 2005 Jun 23 (pp. 407–412). IEEE.
- Hastie T, Tibshirani R, Friedman JH, Friedman JH. *The elements of statistical learning: data mining, inference, and prediction*. New York: springer; 2009 Aug.
- Hayes MH. *Statistical digital signal processing and modeling*. John Wiley & Sons; 2009.
- Hobert MA, Nussbaum S, Heger T, Berg D, Maetzler W, Heinzel S. Progressive gait deficits in Parkinson's disease: a wearable-based biannual 5-year prospective study. *Frontiers in aging neuroscience*. 2019 Feb 13;11:22. [PubMed: 30814947]
- Huysmans D, Borzé P, Buyse B, Testelmans D, Van Huffel S, Varon C. Sleep Diagnostics for Home Monitoring of Sleep Apnea Patients. *Frontiers in digital health*. 2021:58.
- Ibrahim JG, Zhu H, Garcia RI, Guo R. Fixed and random effects selection in mixed effects models. *Biometrics*. 2011 Jun;67(2):495–503. [PubMed: 20662831]
- Jarchi D, Andreu-Perez J, Kiani M, Vysata O, Kuchynka J, Prochazka A, Sanei S. Recognition of patient groups with sleep related disorders using bio-signal processing and deep learning. *Sensors*. 2020 May 2;20(9):2594. [PubMed: 32370185]
- Jehan S, Zizi F, Pandi-Perumal SR, Wall S, Auguste E, Myers AK, Jean-Louis G, McFarlane SI. Obstructive sleep apnea and obesity: implications for public health. *Sleep medicine and disorders: international journal*. 2017;1(4).
- Jenatton R, Audibert JY, Bach F. Structured variable selection with sparsity-inducing norms. *The Journal of Machine Learning Research*. 2011 Nov 1;12:2777–824.
- Liu J, Ji S, Ye J. SLEP: Sparse learning with efficient projections. *Arizona State University*. 2009 Aug;6(491):7.
- Magezi DA. Linear mixed-effects models for within-participant psychology experiments: an introductory tutorial and free, graphical user interface (LMMgui). *Frontiers in psychology*. 2015 Jan 22;6:2 [PubMed: 25657634]
- Meier L, van de Geer S, Bühlmann P. The group Lasso for logistic regression. *Journal of the Royal Statistical Society. Series B*. 2008;70(0):53.
- Merchant G, Buelna C, Castañeda SF, Arredondo EM, Marshall SJ, Strizich G, Sotres-Alvarez D, Chambers EC, McMurray RG, Evenson KR, Stoutenberg M. Accelerometer-measured sedentary time among Hispanic adults: results from the Hispanic Community Health Study/Study of Latinos (HCHS/SOL). *Preventive medicine reports*. 2015 Jan 1;2:845–53. [PubMed: 26844159]
- Moridani MK, Heydar M, Behnam SS. A reliable algorithm based on combination of EMG, ECG and EEG signals for sleep apnea detection:(a reliable algorithm for sleep apnea detection). In *2019 5th Conference on Knowledge Based Engineering and Innovation (KBEI)* 2019 Feb 28 (pp. 256–262). IEEE.
- Nguyen NH, Martinez I, Atreja A, Sitapati AM, Sandborn WJ, Ohno-Machado L, Singh S. *Digital Health Technologies for Remote Monitoring and Management of Inflammatory Bowel Disease:*

A Systematic Review. *The American journal of gastroenterology*. 2022 Jan 14;117(1):78–97 [PubMed: 34751673]

- Niroshana SI, Zhu X, Nakamura K, Chen W. A fused-image-based approach to detect obstructive sleep apnea using a single-lead ECG and a 2D convolutional neural network. *Plos one*. 2021 Apr 26;16(4):e0250618. [PubMed: 33901251]
- Pierce A, Ignasiak NK, Eiteman-Pang WK, Rakovski C, Berardi V. Mobile phone sensors can discern medication-related gait quality changes in Parkinson's patients in the home environment. *Computer Methods and Programs in Biomedicine Update*. 2021 Jan 1;1:100028.
- Qin H, Steenbergen N, Glos M, Wessel N, Kraemer J, Penzel T. The Different Facets of Heart Rate Variability in Obstructive Sleep Apnea. *Frontiers in Psychiatry*. 2021;12:1128.
- Quan SF, Howard BV, Iber C, Kiley JP, Nieto FJ, O'Connor GT, Rapoport DM, Redline S, Redline S, Sotres-Alvarez D, Loredo J, Hall M, Patel SR, Ramos A, Shah N, Ries A, Arens R, Barnhart J, Youngblood M. Sleep-disordered breathing in Hispanic/Latino individuals of diverse backgrounds. The Hispanic community health study/study of Latinos. *American journal of respiratory and critical care medicine*. 2014 Feb 1;189(3):335–44. [PubMed: 24392863]
- Robbins J, Samet JM, Wahl PW. The sleep heart health study: design, rationale, and methods. *Sleep*. 1997 Dec 1;20(12):1077–85. [PubMed: 9493915]
- Rosenthal LD, Dolan DC. The Epworth sleepiness scale in the identification of obstructive sleep apnea. *The Journal of nervous and mental disease*. 2008 May 1;196(5):429–31. [PubMed: 18477888]
- Rudin W *Principles of mathematical analysis*. New York: McGraw-hill; 1976 Jan.
- Sharma M, Raval M, Acharya UR. A new approach to identify obstructive sleep apnea using an optimal orthogonal wavelet filter bank with ECG signals. *Informatics in Medicine Unlocked*. 2019 Jan 1;16:100170.
- Sheta A, Turabieh H, Thaher T, Too J, Mafarja M, Hossain MS, Surani SR. Diagnosis of obstructive sleep apnea from ecg signals using machine learning and deep learning classifiers. *Applied Sciences*. 2021 Jul 19;11(14):6622.
- Si B, Lamb G, Schmitt MH, Li J. A multi-response multilevel model with application in nurse care coordination. *IISE Transactions*. 2017 Jul 3;49(7):669–81.
- Stein PK, Pu Y. Heart rate variability, sleep and sleep disorders. *Sleep medicine reviews*. 2012 Feb 1;16(1):47–66. [PubMed: 21658979]
- Stoica P, Moses RL. *Spectral analysis of signals*. Upper Saddle River, NJ: Pearson Prentice Hall; 2005 May.
- Tibshirani R Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*. 1996 Jan;58(1):267–88.
- Van Albada SJ, Robinson PA. Relationships between electroencephalographic spectral peaks across frequency bands. *Frontiers in human neuroscience*. 2013 Mar 4;7:56. [PubMed: 23483663]
- Wang T, Lu C, Shen G. Detection of sleep apnea from single-lead ECG signal using a time window artificial neural network. *BioMed research international*. 2019 Dec 23;2019.
- Wang Y, Wang X, Li W, Ji S, Yang T, Zhuangwen X, Zhao X, Wang J. An Automatic Classification Method of Sleep Apnea Events Based on EEG Frequency Sub-band Division. 2021.
- Yan X, Bien J. Hierarchical sparse modeling: A choice of two group lasso formulations. *Statistical Science*. 2017 Nov;32(4):531–60.
- Yang Y, Zou H. A fast unified algorithm for solving group-lasso penalize learning problems. *Statistics and Computing*. 2015 Nov;25(6):1129–41.
- Yin S, Kaynak O. Big data for modern industry: challenges and trends [point of view]. *Proceedings of the IEEE*. 2015 Mar 24;103(2):143–6.
- Yuan M, Lin Y. Model selection and estimation in regression with grouped variables. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*. 2006 Feb;68(1):49–67.
- Zhang GQ, Cui L, Mueller R, Tao S, Kim M, Rueschman M, Mariani S, Mobley D, Redline S. The National Sleep Research Resource: towards a sleep data commons. *Journal of the American Medical Informatics Association*. 2018 Oct;25(10):1351–8. [PubMed: 29860441]

- Zhao X, Wang X, Yang T, Ji S, Wang H, Wang J, Wang Y, Wu Q. Classification of sleep apnea based on EEG sub-band signal characteristics. *Scientific Reports*. 2021 Mar 12;11(1):1–1. [PubMed: 33414495]
- Zhou G, Pan Y, Yang J, Zhang X, Guo X, Luo Y. Sleep electroencephalographic response to respiratory events in patients with moderate sleep apnea-hypopnea syndrome. *Frontiers in neuroscience*. 2020 Apr 21;14:310. [PubMed: 32372906]

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

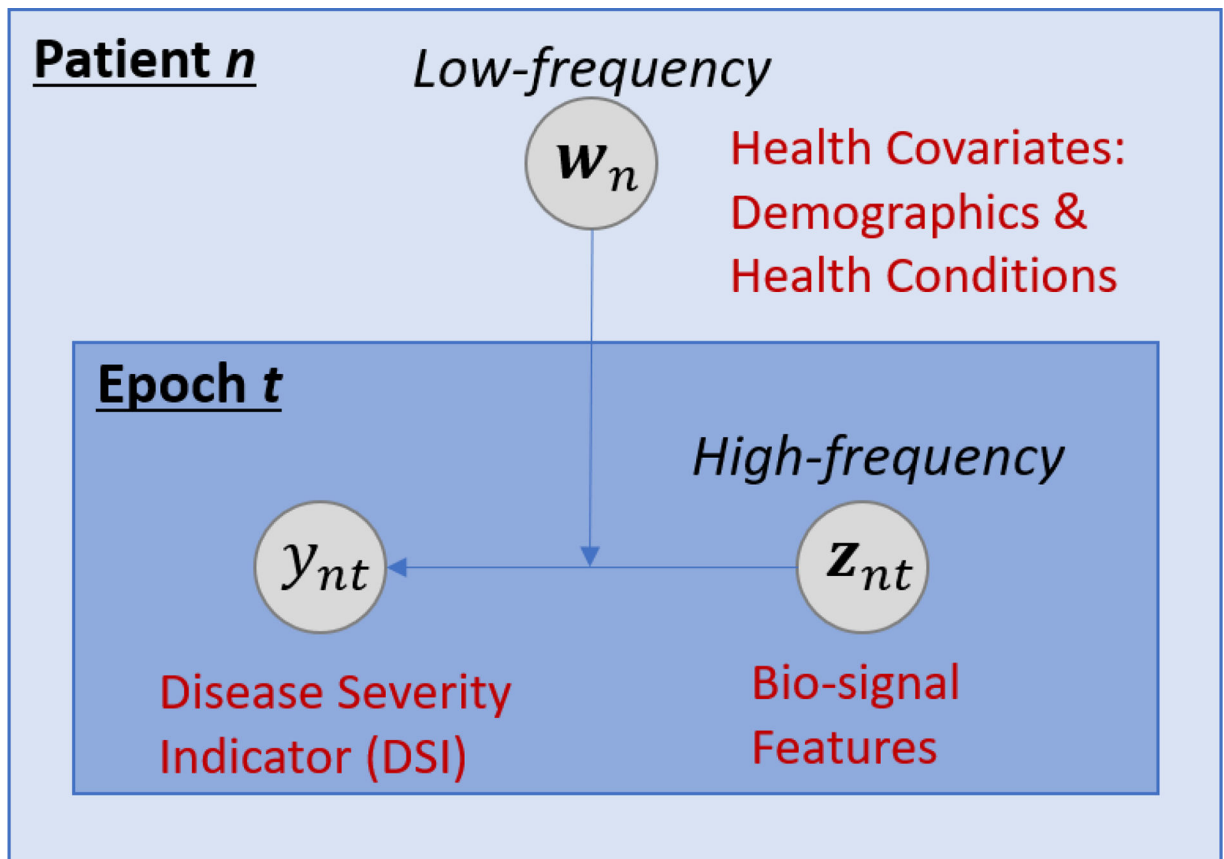


Figure 1:
Graphical illustration of the mixed-frequency data

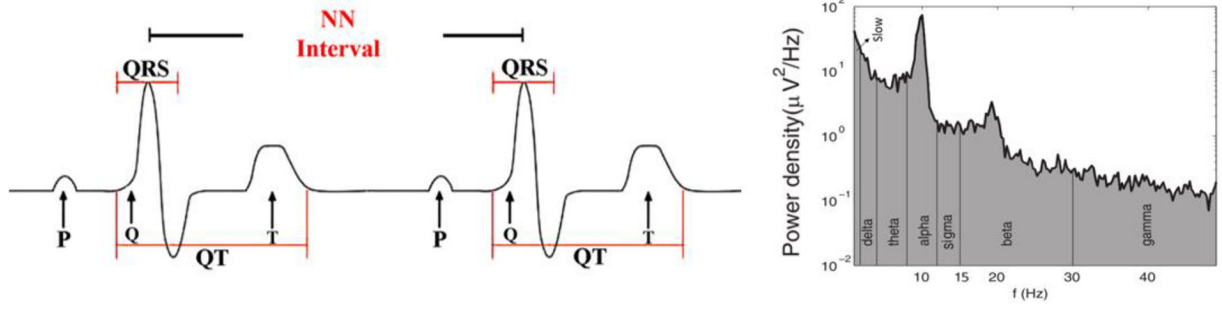


Figure 2:
 Bio-signal processing methods.
 (a) ECG as the basis of measuring HRV. (b) Example of an EEG spectrum and its sub-bands
 (Adopted and revised from (Dong, 2016; van Albada and Robinson, 2013))

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

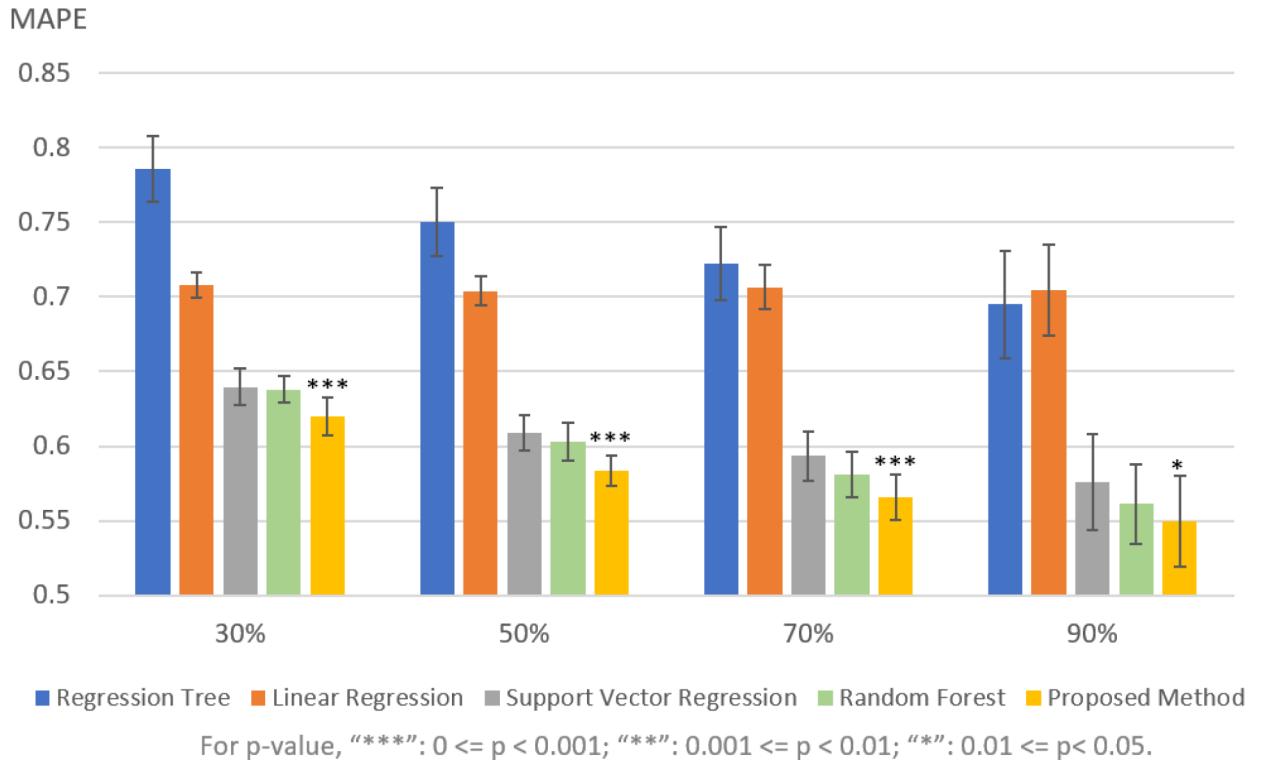


Figure 3: MAPEs on testing data with 30%,50%,70%, and 90% of the data used for model training

Table 1:

Description of variables included in the study

Variables		Summary statistics
<i>Low-frequency independent variables</i>		
Age (Unit: year)		53.5 ± 10.0
Gender (Female: 0; Male: 1)		42.0% / 58.0%
Ethnicity (Hispanic: 0; Non-Hispanic: 1)		9.0% / 91.0%
BMI (Unit: kg/m ²)		26.6 ± 4.5
Depression survey (9 variables; 5-Likert scale from 1 to 6)		4.0 ± 0.3*
Sleep habit survey (7 variables; 5-Likert scale from 1 to 5)		2.3 ± 0.6*
Daytime sleepiness survey (10 variables; 5-Likert scale from 1 to 4)		1.7 ± 0.5*
<i>High-frequency independent variables</i>		
ECG features (6 variables)	Average of all NN intervals (Unit: ms)	931.5 ± 135.6
	Standard deviation of NN intervals (Unit: ms)	56.4 ± 34.46
	NN<10ms counts divided by the total number of NN intervals	(65.0 ± 19.1) %
	Relative spectral power for very low frequency (0.003–0.04 Hz), low frequency (0.04–0.15 Hz), and High frequency (0.15–0.4 Hz)	**
EEG features (7 variables)	Relative spectral power for Slow (0.5–1 Hz), Delta (1–4 Hz), Theta (4–8 Hz), Alpha (8–12 Hz), Sigma (12–15 Hz), Beta (15–30 Hz), and Gamma (30+Hz)	**
<i>Dependent variable</i>		
DSI (Number of adverse events per epoch)		2.47 ± 2.78

* Due to the space limitations, we first calculate a composite score for each survey by taking the average of all separate questions and then report the average composite score across all participants.

** Relative spectral power of different frequency sub-bands are percentages and sum to 100% across all the sub-bands.