

Prioritization of Multi-Level Risk Factors for Obesity

Lu Wang^{*} § Ming Dong[§], Elizabeth Towner[†], and Dongxiao Zhu^{§*}

Abstract—Obesity has become a significant threat to health. Identifying and understanding the underlying obesity risk factors (ORFs) are crucial for optimizing prevention, intervention and treatment for obesity. Most existing methodological approaches to risk factor analysis are employed within the single task learning (STL) framework to learn a ranked list of ORFs for a whole population. However, obesity is a multi-faced health outcome. Some ORFs are highly specific to a certain subpopulation and others are universal to the entire population. Multi-task learning (MTL) framework offers a solution to connect multiple related tasks. Within the MTL framework, we implement two tailor-made models, i.e., multi-task feature learning (MTFL) and clustered multi-task learning (CMTL), to conduct ORFs analysis. The former is capable of finding the universal ORFs for all subpopulations without sacrificing the uniqueness of each subpopulation. The latter uncovers the grouping structure and conducts multi-level ORFs analysis simultaneously. Experiments on a public behavioral dataset demonstrate a superior performance of our methods in prioritizing multi-level ORFs.

Index Terms—Multi-level risk factor analysis, Multi-task feature learning, Clustered multi-task learning, Obesity.

I. INTRODUCTION

Nearly 38% of adults in the United States (U.S.) are obese [1], with rates rising stably and significantly over the past decade. Obesity places adults at risk for developing a plethora of serious medical comorbidities including cardiovascular disease, cancer [2] and premature death ([3], [4]). Identifying the salient risks for obesity and variance among subpopulations is imperative to optimize prevention efforts and treatment. Risk factor analysis is a common methodology to identify, rank and understand the underlying obesity risk factors (ORFs) [5], [6], [7] and to inform prevention and treatment of preventable physical and mental health conditions more broadly, e.g., ([8], [9], [10], [11], [12]).

In statistics, risk factor analysis examines the complicated relation between output and input variables [13], where they are the outcome and features, respectively. Traditional risk factor analysis methods are employed mainly through two approaches: 1) Explore the relationship between output and input variables using regression approaches, such as logistic regression [14] and linear regression ([15], [16]). 2) Distin-

guish differential factors using statistical hypothesis tests, e.g., chi-square test ([17], [18]) and t-test ([19], [20]).

These traditional risk factor analysis methods either build a global model at the population-level only or build a local model for each subpopulation. We consider this type of approaches as single-task learning (STL) risk factor analysis methods illustrated in Figure 1(a). These STL approaches have been shown effective initial efforts in the field of risk factor analysis. However, they have the following disadvantages: 1) A global model fails to capture the data heterogeneity in the population. 2) A local model fails to utilize the shared information among subpopulations. In addition, an over-parameterized model is susceptible to overfitting especially when the sample size of a subpopulation is small.

The multi-faced causes of obesity contain not only population-level ORFs but also subpopulation-level ORFs. In ([1], [21]), the authors consider that obesity may influence some subpopulations more than some others. Since people in various regions, ages and races are vastly different from each other, the subpopulations can be immensely distinguished in terms of ORFs. As a result, prioritization of multi-level ORFs, e.g., subpopulation and population-levels, is necessary in order to maximize the efforts of prevention and intervention for obesity.

Multi-task learning (MTL) framework is introduced to learn multiple related tasks simultaneously, which means MTL is capable of training multiple related models for all subpopulations at the same time by utilizing shared information among these subpopulations [12]. Thus, MTL can learn multiple ranked lists of ORFs simultaneously. To take into account both data heterogeneity and homogeneity, multi-task feature learning (MTFL) is implemented to build multiple related models along with an across-all-tasks penalty/regularization term, i.e., $l_{2,1}$ -norm, to ensure that the weight of each input feature is either small or large for all subpopulations [22], so that the ranked list of ORFs at population-level can be learned. We demonstrate the process of learning multiple ranked lists of ORFs using MTFL in Figure 1(b).

In the real-world scenario, grouping structure often exists in the multiple related tasks. Clustered multi-task learning (CMTL) is used to reveal the grouping structure of tasks and learn multiple related tasks simultaneously ([23], [24]). CMTL implements clustering technique within the MTL framework to combine diverse analyses including clustering, prediction and feature selection. We illustrate how CMTL works for learning multiple ranked lists of ORFs in Figure 1(c).

With both MTFL and CMTL, multi-level risk factor analysis

^{*}Corresponding Author

^{*}Dept. of Mechanical & Industrial Engineering, University of Toronto, Toronto, ON M5S 2X1, Canada. wanglu.wang@mail.utoronto.ca

[§]Dept. of Computer Science, Wayne State University, Detroit, MI 48202, USA. {lu.wang3, dzhu, mdong}@wayne.edu

[†]Dept. of Family Medicine and Public Health Sciences, Wayne State University, Detroit, MI 48202, USA. ekuhl@med.wayne.edu

(i.e., each subpopulation, each subgroup of the population and the whole population) is employed, where obesity is the research target. Note that, in this paper, each subpopulation is defined based on where people live in the U.S., i.e., the participants living in each state/district are one subpopulation. Each subgroup of the population is considered as a group of subpopulations, which is generated through clustering of all 54 U.S. states/districts using CMTL. The whole population in this paper represents all subpopulations. We summarize the main contributions of this paper as following:

- We take into account subpopulation variability and utilize the shared information among subpopulations using the MTL framework.
- We learn the population-level ORFs without sacrificing the unique characteristics of each subpopulation and learn a ranked list of ORFs for each subpopulation using MTFL.
- We perform clustering and ORFs ranking simultaneously using CMTL to uncover the group structure of subpopulations and learn ranked lists of ORFs for each subpopulation, each subgroup of the population and the whole population in the meantime.

We present the outline of this paper as: Section II summarizes traditional risk factor analysis methods and MTL framework. Section III describes the MTFL and CMTL models for risk factor analysis along with their optimization algorithms. In Section IV, we demonstrate the performance of MTFL and CMTL using a public behavioral analysis dataset for obesity. Finally, in Section V, this paper is concluded.

II. BACKGROUND

We briefly review the traditional risk factor analysis methods for obesity and then provide a brief background of multi-task learning (MTL) framework in this section.

A. Risk factor analysis for obesity

The conventional risk factor analysis approaches are implemented within single-task learning (STL) framework, which can be mainly categorized into two categories, i.e., regression methods and statistical hypothesis tests.

The most common regression methods used for risk factor analysis are univariate and multivariate modeling approaches based on generalized linear model, which is effective in working with a variety of targets [25]. For example, to minimize the difference between the observed and predicted values in obesity risk factor analysis, linear model using generalized least squares (LMGLS) is used to obtain a single ranked list of ORFs based on the feature weights.

Linear mixed effects model (LMEM) is an extension of linear regression that accommodates the data with both fixed and random effects [26], which is suitable for data with grouped features as the random effects. However, LMEM is not capable of uncovering data grouping structure, which needs to be predefined by its covariance structure.

Standard statistical hypothesis tests including chi-square test and t-test, which all assume the null hypothesis that output and

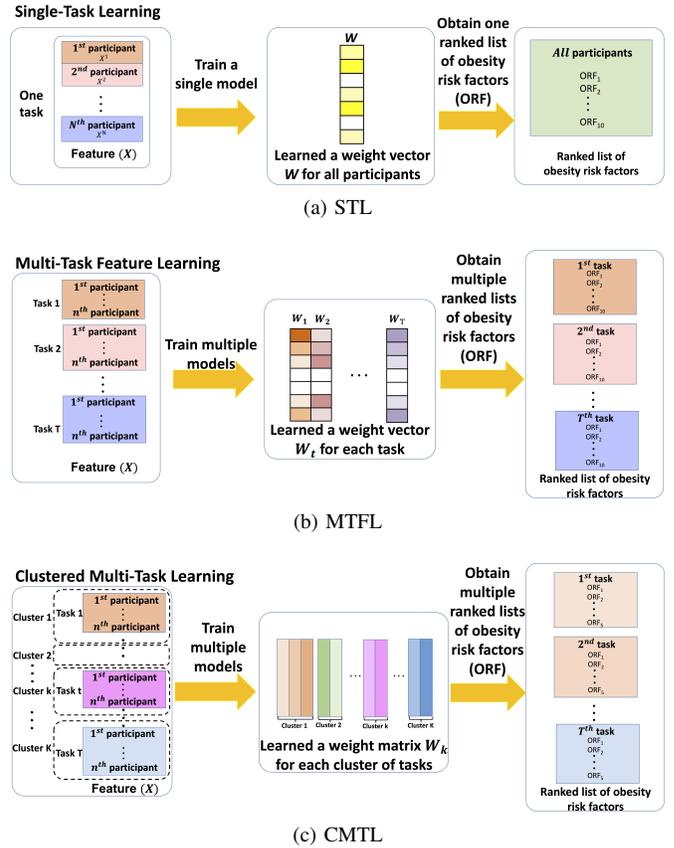


Figure 1: Risk factor analysis is implemented for obesity within the MTL framework: 1) MTFL trains multiple models simultaneously and learns a ranked list of ORFs for each subpopulation. 2) CMTL clusters subpopulations into several groups and obtains multiple ranked lists of ORFs for all subpopulations. But STL merely trains a global model that is one-size-fits-all at the population-level only. (Darker box in the weight vector/matrix means higher value of feature weight.)

input variables are independent. They all use low p -values to reject the null hypothesis and rank the input variables based on p -value of each input variable ([27], [28]).

B. Framework of multi-task learning

To accommodate the relatedness across the tasks and boost the performance, MTL is introduced as one of the inductive transfer learning frameworks to encourage knowledge transfer by simultaneously learning multiple relevant tasks. How task relatedness is formulated into the objective function is the central component of MTL. An earliest MTL approach uses an across-all-tasks regularization/penalty to connect the multiple related tasks [29]. This framework is capable of combining with massive algorithms to optimize the models, e.g., proximal gradient [30]. With these effective optimizing algorithms, the regularized MTL can efficiently handle complicated constraints and/or non-smooth terms in the objective function.

To reveal how the tasks are related, plentiful regularization/penalties have been devised over the last several years.

In [22], multi-task feature learning (MTFL) is introduced with the underlying assumption that there is a feature space shared by all tasks and modeled by a group sparse regularization, i.e., $l_{2,1}$ -norm.

Multiple tasks not only are related, but also can be clustered into groups. Clustered multi-task learning (CMTL) is introduced to uncover the grouping structure of tasks and learn multiple related tasks simultaneously. The objective function of CMTL is non-convex that is infeasible to be optimized using a traditional optimization approach. Thus, the conversion from non-convex to convex is needed. In [24], a convex formulation of CMTL using spectral norm is proposed under the assumption that the tasks with similar weight vectors are within one cluster. In [31], a convex relaxation CMTL regularization using block coordinate descent is introduced based on proving its equivalence with the alternating structure optimization. In [32], a CMTL model combining with hierarchical clustering method is proposed through fixing the latent cluster indicator variable.

In this paper, we employ a convex relaxed CMTL (crCMTL) to prioritize the multi-level risk factors for obesity when the grouping structure exists in the subpopulations. Otherwise, we develop MTFL to learn the ranked lists of risk factors across all tasks with joint sparsity that is implemented using $l_{2,1}$ -norm. Thus, we are capable of conducting the multi-level risk factor analysis to enable a precise prevention, invention and treatment plan.

III. METHOD

Here, we provide the details of two models for risk factor analysis along with their optimization algorithms, i.e., multi-task feature learning (MTFL) and clustered multi-task learning (CMTL) models.

A. MTL framework

Many independent tasks are rare compared with multiple related ones in most real-world applications, so that MTL is implemented to capture the information of multiple related tasks. We assume all the tasks are related with sharing the same feature space. To encode the task relatedness into the object function of MTL, a penalty/regularization term across all the tasks, denoted as $\Omega(\Phi)$, is used. Therefore, minimization of penalized empirical loss is expressed as the framework's objective formulation:

$$\min_{\Phi} \mathcal{L}(\Phi) + \Omega(\Phi), \quad (1)$$

where the empirical loss function is denoted as $\mathcal{L}(\Phi)$.

B. Algorithm of risk factor analysis using MTFL

1) *MTFL model*: The loss function in MTFL is formulated as:

$$\mathcal{L}(\Phi) = \frac{1}{2} \sum_{t=1}^T \|X_t \Phi_t^T - Y_t\|^2, \quad (2)$$

where T is the number of tasks and its corresponding index number is t . $\Phi \in R^{T \times J}$ is the weight matrix of J continuous input features. X is the input matrix and the t^{th} task has the

input matrix denoted as $X_t \in \mathbb{R}^{n_t \times J}$. Y_t denotes the output variable.

As we mentioned in Section I, $l_{2,1}$ -norm is the penalty/regularization term across all the tasks to encode the joint sparsity:

$$\begin{aligned} \Omega(\Phi) &= \|\Phi\|_{2,1}, \\ \|\Phi\|_{2,1} &= \sum_{j=1}^J \sqrt{\sum_{t=1}^T |\phi_{tj}|^2}, \end{aligned} \quad (3)$$

where j is the corresponding index number of continuous input features and ϕ_{tj} denotes weight scalar of the t^{th} task's j^{th} feature.

As a result, the object function of MTFL can be re-written as:

$$\min_{\Phi} \frac{1}{2} \sum_{t=1}^T \|X_t \Phi_t^T - Y_t\|^2 + \lambda \sum_{j=1}^J \sqrt{\sum_{t=1}^T |\phi_{tj}|^2}, \quad (4)$$

where $\lambda \geq 0$, called tuning parameter, can be used to adjust the penalty/regularization term and control the sparsity of feature weights matrix. It produces more sparse feature weights matrix when the value of λ is increasing.

2) *Optimization*: Fast iterative shrinkage thresholding algorithm (FISTA) shown in Algorithm 1 is implemented to optimize the $l_{2,1}$ -norm regularization problem in Eq.(4) with the general updating steps:

$$\Phi^{(l+1)} = \pi_P(S^{(l)} - \frac{1}{\gamma^{(l)}} \mathcal{L}'(S^{(l)})), \quad (5)$$

where l is the iteration index, $\frac{1}{\gamma^{(l)}}$ is the possible largest step-size that is chosen by line search [33, Lemma 2.1, page 189] and $\mathcal{L}'(S^{(l)})$ is the gradient of $\mathcal{L}(\cdot)$ at search point $S^{(l)}$. $S^{(l)} = \Phi^{(l)} + \alpha^{(l)}(\Phi^{(l)} - \Phi^{(l-1)})$ are the search points for each task, where $\alpha^{(l)}$ is the combination scalar. $\pi_P(\cdot)$ is $l_{2,1}$ -regularized Euclidean projection shown as:

$$\pi_P(H(S^{(l)})) = \min_{\Phi} \frac{1}{2} \|\Phi - H(S^{(l)})\|_F^2 + \lambda \|\Phi\|_{2,1}, \quad (6)$$

where $H(S^{(l)}) = S^{(l)} - \frac{1}{\gamma^{(l)}} \mathcal{L}'(S^{(l)})$ is the gradient step of $S^{(l)}$. A sufficient scheme that solves Eq.(6) has been proposed as Theorem 1 [34].

Theorem 1: $\hat{\Phi}$'s primal optimal point in Eq.(6) can be calculated with λ as:

$$\hat{\Phi}_j = \begin{cases} \left(1 - \frac{\lambda}{\|H(S^{(l)})_j\|_2}\right) H(S^{(l)})_j & \text{if } \lambda > 0, \|H(S^{(l)})_j\|_2 > \lambda \\ 0 & \text{if } \lambda > 0, \|H(S^{(l)})_j\|_2 \leq \lambda \\ H(S^{(l)})_j & \text{if } \lambda = 0, \end{cases} \quad (7)$$

where $H(S^{(l)})_j$ is the j^{th} row of $H(S^{(l)})$ and $\hat{\Phi}_j$ is the j^{th} row of $\hat{\Phi}$.

From the 4th line to the 11th line in Algorithm 1, the optimal $\gamma^{(l)}$ is chosen by the backtracking rule. And $\gamma^{(l)} \geq b$, where b is the Lipschitz constant of $\mathcal{L}(\cdot)$ at search point $S^{(l)}$, which means $\gamma^{(l)}$ is satisfied for $S^{(l)}$ and $\frac{1}{\gamma^{(l)}}$ is the possible largest step size.

Algorithm 1: Fast iterative shrinkage thresholding algorithm (FISTA) for optimizing the $l_{2,1}$ -norm regularization problem.

Input: Input variables $\{X_1, X_2, \dots, X_T\}$, output variable Y across all T tasks, initialization of feature weights $\Phi^{(0)}$ and λ

Output: $\hat{\Phi}$

```

1 Initialize:  $\Phi^{(1)} = \Phi^{(0)}$ ,  $d_{-1} = 0$ ,  $d_0 = 1$ ,  $\gamma^{(0)} = 1$ ,  $l = 1$ ;
2 repeat
3   Set  $\alpha^{(l)} = \frac{d_{l-2}-1}{d_{l-1}}$ ,  $S^{(l)} = \Phi^{(l)} + \alpha^{(l)}(\Phi^{(l)} - \Phi^{(l-1)})$ ;
4   for  $j = 1, 2, \dots, J$  do
5     Set  $\gamma = 2^j \gamma_{l-1}$ ;
6     Compute  $\Phi^{(l+1)} = \pi_P(S^{(l)} - \frac{1}{\gamma^{(l)}} \mathcal{L}'(S^{(l)}))$ ;
7     Compute  $Q_\gamma(S^{(l)}, \Phi^{(l+1)})$ ;
8     if  $\mathcal{L}(\Phi^{(l+1)}) \leq Q_\gamma(S^{(l)}, \Phi^{(l+1)})$  then
9       |  $\gamma^{(l)} = \gamma$ , break ;
10    end
11  end
12   $d_l = \frac{1 + \sqrt{1 + 4d_{l-1}^2}}{2}$ ;
13   $l = l + 1$ ;
14 until Convergence of  $\Phi^{(l)}$ ;
15  $\hat{\Phi} = \Phi^{(l)}$ ;

```

At the 7th line in Algorithm 1, tangential line of $\mathcal{L}(\cdot)$ at search point $S^{(l)}$, denoted as $Q_\gamma(S^{(l)}, \Phi^{(l+1)})$, is computed by:

$$Q_\gamma(S^{(l)}, \Phi^{(l+1)}) = \mathcal{L}(S^{(l)}) + \frac{\gamma}{2} \|\Phi^{(l+1)} - S^{(l)}\|^2 + \langle \Phi^{(l+1)} - S^{(l)}, \mathcal{L}'(S^{(l)}) \rangle.$$

C. Algorithm of risk factor analysis using CMTL

1) *CMTL framework:* Multiple tasks in the real-world applications not only are related, but also show a more complicated grouping structure, which can be seen from that the estimated weights of tasks from the same group are closer than these from distinct groups. To reveal the grouping structure of multiple tasks, K -means clustering is employed by implementing CMTL. To encode the grouping structure of multiple tasks in the formulation, K -means's sum-of-square error (SSE) is used as the regularization term in the object function of CMTL.

We assume that T tasks can be clustered into K clusters, where $K < T$. The cluster's corresponding index number is k and the index set is defined as $\mathcal{I}_k = \{1, 2, \dots, K\}$. Let $\bar{\Phi}_k = \frac{1}{n_k} \sum_{k \in \mathcal{I}_k} \Phi_k$ be the mean function of the weight vectors in the k^{th} cluster, so that the SSE is calculated as:

$$\sum_{k=1}^K \sum_{k \in \mathcal{I}_k} \|\Phi_k - \bar{\Phi}_k\|_2^2 = \text{tr}(\Phi \Phi^T) - \text{tr}(\Phi O O^T \Phi^T), \quad (8)$$

where $\text{tr}(\cdot)$ is the trace norm of matrix and $O \in \mathbb{R}^{T \times K}$ is the cluster indicator matrix that is orthogonal:

$$O_{t,k} = \begin{cases} \frac{1}{\sqrt{n_k}} & \text{if } t \in \mathcal{I}_k, \\ 0 & \text{if } t \notin \mathcal{I}_k, \end{cases} \quad (9)$$

where n_k is the number of input instances/participants in cluster k . Since the orthogonal cluster indicator matrix O is non-convex that exhibits the above mentioned special structure, the SSE in Eq.(8) is hard to be minimized. To overcome this issue, we use a spectral relaxation approach [35], the latter is expressed as $O^T O = I_K$. Furthermore, a convex relaxation that relaxes the feasible domain of $O O^T$ into a convex set is proposed in [24], i.e., $\mathcal{C} = \{C | \text{tr}(C) = T, C \preceq I, C \in \mathbb{S}_+^T\}$, where \mathbb{S}_+^T is a subset of positive-semidefinite matrices. As a result, $O O^T$ can be approximated through the convex set \mathcal{C} . In conclusion, the previously mentioned two types of relaxation methods generate the convex relaxed CMTL (crCMTL) expressed as:

$$\begin{aligned} \min_{\Phi, C} \quad & \mathcal{L}(\Phi) + \rho_1 [\text{tr}(\Phi \Phi^T) - \text{tr}(\Phi C \Phi^T)] + \rho_2 \text{tr}(\Phi \Phi^T), \\ \text{s. t.} \quad & \text{tr}(C) = K, \Phi \preceq I, \Phi \in \mathbb{S}_+^T \end{aligned} \quad (10)$$

where $\text{tr}(\Phi \Phi^T) = \|\Phi\|_2^2$ is used to shrink the weights and relieve the multicollinearity, which is also called the square of Frobenius norm of Φ . A parameter η is introduced, which is defined as $\eta = \frac{\rho_2}{\rho_1} > 0$. Then with some simple allergic calculations, the crCMTL is reformulated as:

$$\begin{aligned} \min_{\Phi, C} \quad & \sum_{t=1}^T \frac{1}{N_t} \sum_{i=1}^{N_t} (X_i^t (C_i^t)^T - Y_i^t)^2 \\ & + \rho_1 \eta (1 + \eta) \text{tr}(\Phi (\eta I + C)^{-1} \Phi^T), \\ \text{s. t.} \quad & \text{tr}(C) = K, C \preceq I, C \in \mathbb{S}_+^T \end{aligned} \quad (11)$$

where N_t is the number of instances/participants in task t and i is the index of instance/participant in the t^{th} task.

2) *Optimization:* In Eq.(11), the equation is conjointly convex with respect to (w.r.t.) C and Φ , which is an convex unconstrained smooth optimization problem w.r.t. C . We iteratively update the gradient step of the aforementioned optimization problem in order to find the global optimum w.r.t. C :

$$G_\Phi = S - \frac{1}{\gamma} [\nabla \mathcal{L}(S_\Phi) + 2\rho_1 \eta (1 + \eta) (\eta I + C_S)^{-1} S^T], \quad (12)$$

where S_Φ is the search point of Φ that is defined as $S_\Phi^{(l)} = \Phi^{(l)} + \alpha^{(l)}(\Phi^{(l)} - \Phi^{(l-1)})$. The search point of C is denoted as C_S , which can be similarly updated as $C_S^{(l)} = C^{(l)} + \alpha^{(l)}(C^{(l)} - C^{(l-1)})$ at the l^{th} iteration. $\nabla \mathcal{L}(S)$ is the gradient of $\mathcal{L}(S)$ that is calculated as:

$$\nabla \mathcal{L}(S) = \left[\frac{l'(S_1)}{N_1}, \frac{l'(S_2)}{N_2}, \dots, \frac{l'(S_T)}{N_T} \right]. \quad (13)$$

Similarly, in the optimization of MTFM, FISTA is also implemented for optimizing the crCMTL, except the line 6 is replaced with the corresponding proximal operator that is solved by the following steps. To optimize the convex set \mathcal{C} , we need to solve a convex constrained minimization problem, which is formulated with its corresponding proximal operator and calculated using its gradient step, denoted as G_C , at the search point C_S :

$$\min_C \|C - G_C\|_F^2, \quad \text{s. t.} \quad \text{tr}(C) = K, C \preceq I, C \in \mathbb{S}_+^T. \quad (14)$$

We can compute the G_C by:

$$G_C = C_S + \frac{\rho_1 \eta (1 + \eta)}{\gamma} S^T S (\eta I + C_S)^{-2}. \quad (15)$$

In [36], a solution of Eq.(14) is proposed and summarized in the following theorem.

Theorem 2: Let $G_T = V \hat{\Sigma} V^T$ be the eigen-decomposition of gradient step $G_C \in \mathbb{S}^{T \times T}$, where $\hat{\Sigma} = \text{diag}(\hat{\sigma}_1, \dots, \hat{\sigma}_T) \in \mathbb{R}^{T \times T}$ and $V \in \mathbb{R}^{T \times T}$ is orthonormal. The optimization problem is formulated as:

$$\begin{aligned} \min_{\{\sigma_m\}} \quad & \sum_{t=1}^T (\sigma_t - \hat{\sigma}_t)^2. \\ \text{s. t.} \quad & \sum_{t=1}^T \sigma_t = K, \quad 0 \leq \sigma_t \leq 1, \quad \forall t = 1, \dots, T \end{aligned} \quad (16)$$

Let $\Sigma^* = \text{diag}(\sigma_1^*, \dots, \sigma_T^*) \in \mathbb{R}^{T \times T}$, so that the optimal solution of the above optimization problem is $\{\sigma_1^*, \dots, \sigma_T^*\}$. As a result, the proximal operator's optimal solution in Eq.(14) is calculated as $\hat{T} = V \Sigma^* V^T$.

IV. EXPERIMENTS AND RESULTS

In this section, we firstly provide the information of experimental setup and the public dataset we use for experiments. We then compare our methods with two STL based linear regression methods mentioned in Section II-A, in order to evaluate MTFL and CMTL's performance. At last, the results of obesity risk factor analysis are discussed.

A. Experiments setup

For our proposed methods, MTFL and crCMTL are implemented using Matlab [31]. For the two STL based linear regression methods mentioned in Section II-A, both are implemented in R using the package *nlme* [37]: 1) Linear model with generalized least squares (LMGLS) is trained using the *gls* function that permits correlated errors. 2) Linear mixed-effects model (LMEM) is trained using *lme* function that models fixed and random effects.

B. Dataset

Experiments are completed using 2016 data from the Behavioral Risk Factor Surveillance System (BRFSS)¹. BRFSS dataset is phone-based survey data collected from all the states/districts in the U.S. and filed by the Centers for Disease Control and Prevention (CDC). This dataset is state-specific and the participants are all adults. Table I provides the number of observations in each subpopulation. The original BRFSS dataset contains 486,303 instances and 275 variables. We remove the instances that the input variables with all cryptic information to generate a dataset containing 459,156 instances with 84 input variables and body mass index (BMI) as the outcome variable.

¹https://www.cdc.gov/brfss/annual_data/annual_2016.html

Table I: Number of observations in each subpopulation of Behavioral Risk Factor Surveillance System (BRFSS) dataset. Note that, S/D and # are the abbreviation of U.S. state/district and the number of observations in each U.S. state/district.

S/D	#								
AL	6,276	FL	33,358	LA	4,760	NE	12,652	OK	6,224
AK	2,619	GA	4,873	ME	9,026	NV	3,904	OR	4,877
AZ	9,835	HI	7,294	MD	16,649	NH	5,770	PA	6,194
AR	4,767	ID	4,695	MA	7,582	NJ	6,897	RI	4,927
CA	10,313	IL	4,292	MI	10,810	NM	5,451	SC	10131
CO	13,493	IN	9,979	MN	15,275	NY	30,786	SD	5,202
CT	9,985	IA	6,527	MS	4,636	NC	5,880	TN	5,517
DE	3,653	KS	1,0951	MO	6,399	ND	5,132	TX	10,530
DC	3,462	KY	9,231	MT	5,337	OH	11,127	UT	9,855
								VT	5,920
								WA	8,109
								WV	6,392
								WI	4,765
								WY	4,049
								GU	1,436
								PR	5,232
								VI	1,153

C. Experimental results

The tasks are defined in BRFSS based on the geographic information, i.e, 54 states/districts in the U.S., so that there are 54 related tasks, which means there are 54 subpopulations.

In the MTFL, 54 models are trained simultaneously with $l_{2,1}$ -norm to encode the joint sparsity. Thus, one ranked list of ORFs is learned for each subpopulation. And then, we choose top 10 ORFs from each ranked list to summarize the results in Figure 2, where first column represents the names of ORFs and the other columns represent the abbreviations of 54 U.S. states/districts distinguished by different colors. To check the subpopulation-level ORFs in Figure 2, we can firstly find the abbreviation of a state/district from column two to the last column and then check the same row's first column. Population-level ORFs can be found by the counts of states/districts, e.g., the first three ORFs at the first column of Figure 2 are considered as population-level ORFs since the first two are shared by all the U.S. states/districts and the third one is shared 53 U.S. states/districts except IL.

In the crCMTL, clustering is implemented and 54 models are also trained simultaneously. Since there are four census regions² in the U.S., we set four clusters for clustering and the result is shown in Figure 3. Note that, only continental U.S. is shown in the Figure 3. More specifically, three states/districts are not included in the Figure 3, i.e., VI, PR and GU, and they are in the cluster with blue color.

We present the results of multi-level risk factor analysis for obesity using crCMTL in different formats: 1) Result of cluster-of-subpopulations-level, also named as subgroup-of-population-level, is shown in Table II. 2) Results of subpopulation and population-levels are shown in Figure 4. In Table II, the first, third and fifth columns represent the names of ORFs and the other columns are the cluster numbers that can be referred to Figure 3. In Figure 4, the organization is as the same as it is in Figure 2, where subpopulation and population-levels ORFs can be located through bridging the

²https://web.archive.org/web/20130921053705/http://www.census.gov/geo/maps-data/maps/pdfs/reference/us_regdiv.pdf

POORHLTH	AL	AK	AZ	AR	CA	CO	CT	DE	DC	FL	GA	HI	ID	IL	IN	IA	KS	KY
	LA	ME	MD	MA	MI	MN	MS	MO	MT	NE	NV	NH	NJ	NM	NY	NC	ND	OH
	OK	OR	PA	RI	SC	SD	TN	TX	UT	VT	VA	WA	WV	WI	WY	GU	PR	VI
SLEPTIM1	AL	AK	AZ	AR	CA	CO	CT	DE	DC	FL	GA	HI	ID	IL	IN	IA	KS	KY
	LA	ME	MD	MA	MI	MN	MS	MO	MT	NE	NV	NH	NJ	NM	NY	NC	ND	OH
	OK	OR	PA	RI	SC	SD	TN	TX	UT	VT	VA	WA	WV	WI	WY	GU	PR	VI
CDHOUSE	AL	AK	AZ	AR	CA	CO	CT	DE	DC	FL	GA	HI	ID	IL	IN	IA	KS	KY
	LA	ME	MD	MA	MI	MN	MS	MO	MT	NE	NV	NH	NJ	NM	NY	NC	ND	OH
	OK	OR	PA	RI	SC	SD	TN	TX	UT	VT	VA	WA	WV	WI	WY	GU	PR	VI
QLMENTL2	AL	AK	AZ	AR	CA	CO	CT	DE	ID	IL	NE	NV	NM	NY	ND	PA	SC	TN
	TX	UT	VA	WV	PR	VI												
X.DENVST2	DC	GA	ID	IN	MI	MN	MO	NM	NY	OH	PA	SC	UT	VA	WV	WY	PR	
DIABAGE	FL	IA	ME	NH	NC	RI	SC	VT	WY	WI	GU							
CHILDREN	AK	KS	MA	NJ	OK	SD	WA	PR										
USENOW3	HI	KY	MS	ND	TN	WI	GU											
MAXDRNKS	AL	DC	MT	OH	SD	VT												
SSBFRUT2	AZ	KS	NV	OK	WV													
PAINACT2	FL	LA	NJ	OR														
SSBSUGR2	DE	HI	NC	TX														
LSTBLDS3	AR	GA	OR															
MENTHLTH	IL	MD	RI															
PHYSHLTH	CO	MA	VI															
FALL12MN	IA	LA	NH															
ASTHMAGE	CT	MI																
ALCDAY5	KY	MN																
NOCOV121	ME	MS																
FEETCHK2	MD	MT																
HHADULT	CA	NE																

Figure 2: Obesity risk factor analysis result at subpopulation-level and population-level using MTL. Top 10 ORFs are selected from each subpopulation (i.e., the participants living in each state/district). Geographic information is represented by abbreviations of states/districts in various colors. Subpopulation-level ORFs can be found in the same row, where one interested state/district appears. For example, HHADULT, the number of adults per family, is the state-specific ORF for California and Nebraska shown at the last row in this figure.

names of ORFs and abbreviations of 54 U.S. states/districts. For example in Figure 4, the first three ORFs, at the first column are considered as population-level ORFs due to the first two are shared by all the U.S. states/districts and the third one is shared 53 U.S. states/districts except IL.

Table II: Obesity risk factor analysis result at subgroup-of-population-level using crCMTL, i.e., ORFs shared by each cluster of subpopulations. Note that, cluster 1, 2, 3, and 4 are equal to the clusters with colors blue, green, yellow and red in Figure 3, respectively.

ORFs	Clusters	ORFs	Clusters	ORFs	Clusters
AGE	1, 2, 3, 4	USENOW3	1, 2, 4	QLMENTL2	1
INCOME2	1, 2, 3, 4	X.SMOKER3	1, 2, 4	SSBFRUT2	2
SLEPTIM1	1, 2, 3, 4	ALHLTH2	1, 2	SSBSUGR2	2
DROCDY3	1, 2, 3, 4	QLACTLM2	1, 4	ASTHMAGE	3
MENTHLTH	1, 2, 3, 4	X.EDUCAG	1, 4	LSTCOVRG	3
PHYSHLTH	1, 2, 3, 4	X.INCOMG	3, 4	QLSTRES2	3
POORHLTH	1, 2, 3, 4	AVEDRNK2	1	RMVTETH3	3
CHILDREN	1, 2, 3, 4	DRNKANY5	1	ASNOSLEP	4
EDUCA	1, 2, 3, 4	PREDIAB1	1	NUMHHOL2	4
X.DRNKWEK	1, 2, 4	PAINACT2	1		

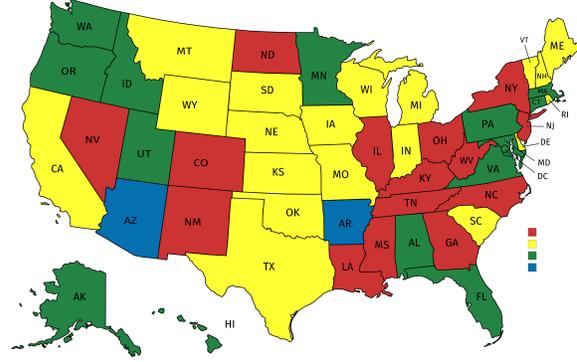


Figure 3: Clustering result of four clusters. Note that, different color represents different cluster.

Since STL trains model independently, it is not reasonable to train 54 independent models and then summarize these independent subpopulation-level ORFs results to obtain a population-level ranked list of ORFs. Thus, we only train a population-level model using each STL based regression method and compare with our methods in Table III. Top 10 population-level ORFs are selected from each method's population-level result shown in Table III. The population-level ORFs from MTL methods are ranked based on the number of U.S. states/districts that share the same ORF, while the ORFs from STL methods are ranked based on the weight of each variable. Note that, the first two ORFs in the results from MTL and crCMTL are shared by all 54 U.S. states/districts so that the ranking numbers of them are the same as 1 in Table III.

D. Discussion of results

1) *Results comparison:* MTL and crCMTL outperform STL based regression models, since they are capable of performing multi-level risk factor analysis and identifying more subpopulation-level ORFs, e.g., PAINACT is an ORF unique to Alaska (the only non-contiguous U.S. state on continental North America) identified by crCMTL but STL. The results of risk factor analysis for obesity using MTL and crCMTL also confirm that the multiple tasks are related since some ORFs are shared by all subpopulations or by some subpopulations.

The result of risk factor analysis for obesity using crCMTL is quite different from the one using MTL despite they all train multiple models simultaneously. For example, the number of selected ORFs in Figure 4 is more than the ones in Figure 2. More specifically, crCMTL generates more state-specific ORFs comparing with MTL. In addition, crCMTL can perform clustering as well, and hence is more suitable for multi-level risk factor analysis when the multiple tasks exhibit grouping structure and the number of tasks is large. Otherwise, MTL can be used for multi-level risk factor analysis with small number of tasks.

AGE	AL	AK	AZ	AR	CA	CO	CT	DE	DC	FL	GA	HI	ID	IL	IN	IA	KS	KY
	LA	ME	MD	MA	MI	MN	MS	MO	MT	NE	NV	NH	NJ	NM	NY	NC	ND	OH
INCOME2	OK	OR	PA	RI	SC	SD	TN	TX	UT	VT	VA	WA	WV	WI	WY	GU	PR	VI
	AL	AK	AZ	AR	CA	CO	CT	DE	DC	FL	GA	HI	ID	IL	IN	IA	KS	KY
SLEPTIM1	LA	ME	MD	MA	MI	MN	MS	MO	MT	NE	NV	NH	NJ	NM	NY	NC	ND	OH
	OK	OR	PA	RI	SC	SD	TN	TX	UT	VT	VA	WA	WV	WI	WY	GU	PR	VI
DROCDY3	AL	AK	AZ	AR	CA	CO	CT	DE	DC	FL	GA	HI	ID	IL	IN	IA	KS	KY
	LA	ME	MD	MA	MI	MN	MS	MO	MT	NE	NV	NH	NJ	NM	NY	NC	ND	OH
MENTHLTH	OK	OR	PA	RI	SC	SD	TN	TX	UT	VT	VA	WA	WV	WI	WY	GU	PR	VI
	AL	AK	AR	CA	CO	DE	DC	FL	GA	HI	ID	IL	IN	IA	KS	KY	LA	ME
PHYSHLTH	MD	MA	MI	MN	MS	MO	MT	NE	NV	NH	NJ	NM	NY	NC	ND	OH	OK	OR
	PA	RI	SC	SD	TN	TX	UT	VT	VA	WA	WV	WI	WY	GU	PR	VI		
POORHLTH	AL	AK	AR	CA	CO	DE	DC	FL	GA	HI	ID	IN	IA	KS	KY	LA	ME	MD
	MA	MI	MN	MS	MO	MT	NE	NV	NH	NJ	NM	NY	NC	ND	OH	OK	OR	PA
CHILDREN	RI	SC	SD	TN	TX	UT	VT	VA	WA	WV	WI	WY	GU	PR	VI			
	AL	AK	AR	CA	CO	DE	DC	FL	GA	HI	ID	IN	IA	KS	KY	LA	ME	MD
X.DRKNDRV	MA	MI	MN	MS	MO	MT	NE	NV	NH	NJ	NM	NY	NC	ND	OH	OK	OR	PA
	RI	SC	SD	TN	TX	UT	VT	VA	WA	WV	WI	WY	GU	PR	VI			
EDUCA	AL	AK	AR	CA	CO	DE	DC	FL	GA	HI	ID	IN	IA	KS	KY	LA	ME	MD
	MA	MI	MN	MS	MO	MT	NE	NV	NH	NJ	NM	NY	NC	ND	OH	OK	OR	PA
X.DRKNWEK	SC	SD	TX	UT	VT	VA	WA	WV	WI	WY	GU	PR	VI					
	AZ	AR	CA	CT	DE	DC	HI	IL	IN	IA	KS	KY	LA	MD	MI	MN	MS	MO
USENOW3	MT	NE	NV	NH	NM	NY	NC	ND	OR	PA	RI	SC	SD	UT	VA	WA	WV	WI
X.SMOKER3	WY	PR																
X.EDUCAG	AZ	CT	FL	GA	ID	MA	TN											
QLACTLM2	AZ	CT	IL	OH														
ALHLTH2	CT	IL																
X.INCOMG	CT	GU																
SSBFRUT2	OK	TN																
SSBSUGR2	AL																	
PAINACT2	AK																	
PREDIAB1	AZ																	
NUMHHOL2	IL																	
QLSTRES2	ME																	
RMVTETH3	MT																	
ASNOSLEP	NJ																	
LSTCOVERG	TX																	
ASTHMAGE	VT																	
AVEDRKN2	GU																	
DRNKANY5	GU																	
QLMENTL2	VI																	

Figure 4: Obesity risk factor analysis result using crCMTL at subpopulation-level and population-level.

2) *Results interpretation*: The ORFs can be mainly classified into three categories: 1) Health conditions (e.g., sleep, asthma, diabetes). 2) Social behaviors (e.g., phone usage, drinking, smoking). 3) Demographic characteristics (e.g., age, family size, educational levels, income, employment). Please refer to the codebook of BRFSS³ for the detailed description of ORFs. In particular, we interpret the ORFs learned from four methods as follows:

- MTFL: Six out of 10 ORFs are within the 1st category, i.e., health conditions, such as sleep and diabetes. Three ORFs fall into the 2nd category, i.e., social behaviors, such as drinking and smoking behaviors. Only one ORF falls into the 3rd category, i.e., demographic characteristics.
- crCMTL: Four out of 10 ORFs fall into the category

Table III: Top 10 selected ORFs and their corresponding category numbers from our proposed MTL methods and two STL methods (please refer to Section IV-D2 for the details of corresponding categories of ORFs). Note that, category numbers are shown within parenthesis under the ORFs and their descriptions can be referred to Section IV-D2. \mathbb{R} means the ranking number of each ORF at population-level.

\mathbb{R}	MTL		\mathbb{R}	STL	
	MTFL	crCMTL		LMGLS	LMEM
1	POORHLTH (1)	AGE (3)	1	POORHLTH (1)	MENTHLTH (1)
1	SLEPTIM1 (1)	INCOME2 (3)	2	X.DENVST2 (1)	X.ASTHMS1 (1)
3	CDHOUSE (1)	SLEPTIM1 (1)	3	MENTHLTH (1)	ASATTACK (1)
4	QLMENTL2 (1)	DROCDY3 (2)	4	USENOW3 (2)	X.AGE80 (3)
5	X.DENVST2 (1)	MENTHLTH (1)	5	SLEPTIM1 (1)	CDHELP (1)
6	DIABAGE (1)	PHYSHLTH (1)	6	LSTBLDS3 (1)	TETANUS (1)
7	CHILDREN (3)	POORHLTH (1)	7	FALL12MN (1)	ALHLTH2 (1)
8	USENOW3 (2)	CHILDREN (3)	8	PREDIAB1 (1)	X.DUALUSE (2)
9	MAXDRNKS (2)	X.DRKNDRV (2)	9	FEETCHK2 (1)	CDSOCIAL (1)
10	SSBFRUT2 (2)	EDUCA (3)	10	QLACTLM2 (1)	DIABAGE (1)

of demographic characteristics and two of them are the population-level ones. Four ORFs fall into the category of health conditions. The other two are within the category of social behavior.

- LMGLS: Nine out of 10 ORFs fall into the category of health conditions. The other one is within the category of social behaviors.
- LMEM: Eight out of 10 ORFs fall into the category of health conditions. The other two fall into the categories of social behaviors and demographic characteristics.

From the summary and interpretation above, we can see that the category of health conditions plays the major role in obesity. Demographic characteristics and social behaviors also have profound impacts as identified by our methods. As obesity is a multi-faced outcome, the ORFs are also diverse. Our proposed MTFL and crCMTL models' results provide more categories of ORFs comparing with STL methods', which proves the assumption that obesity is a multi-faced outcome.

V. CONCLUSION

Obesity is well studied using STL based approaches. However, this type of approaches fails to model the heterogeneity of subpopulations and grouping structure of the population. To overcome the aforementioned limitations, we formulate two concrete models, i.e., MTFL and CMTL, to conduct the multi-level risk factor analysis within the MTL framework.

³https://www.cdc.gov/brfss/annual_data/2016/pdf/codebook16_llcp.pdf

ACKNOWLEDGMENT

The authors would like to thank the U.S. National Science Foundation for supporting our work with grants CNS-1637312 and CCF-1451316.

REFERENCES

- [1] K. M. Flegal, D. Kruszon-Moran, M. D. Carroll, C. D. Fryar, and C. L. Ogden, "Trends in obesity among adults in the united states, 2005 to 2014," *Jama*, vol. 315, no. 21, pp. 2284–2291, 2016.
- [2] M. Abdelaal, C. W. le Roux, and N. G. Docherty, "Morbidity and mortality associated with obesity," *Annals of translational medicine*, vol. 5, no. 7, 2017.
- [3] P. S. Collaboration *et al.*, "Body-mass index and cause-specific mortality in 900 000 adults: collaborative analyses of 57 prospective studies," *The Lancet*, vol. 373, no. 9669, pp. 1083–1096, 2009.
- [4] A. Berrington de Gonzalez, P. Hartge, J. R. Cerhan, A. J. Flint, L. Hannan, R. J. MacInnis, S. C. Moore, G. S. Tobias, H. Anton-Culver, L. B. Freeman *et al.*, "Body-mass index and mortality among 1.46 million white adults," *New England Journal of Medicine*, vol. 363, no. 23, pp. 2211–2219, 2010.
- [5] R. W. Jeffery, J. Baxter, M. McGuire, and J. Linde, "Are fast food restaurants an environmental risk factor for obesity?" *International Journal of Behavioral Nutrition and Physical Activity*, vol. 3, no. 1, p. 2, 2006.
- [6] B. J. Lee, K. H. Kim, B. Ku, J.-S. Jang, and J. Y. Kim, "Prediction of body mass index status from voice signals based on machine learning for automated medical applications," *Artificial intelligence in medicine*, vol. 58, no. 1, pp. 51–61, 2013.
- [7] C.-J. Tseng, C.-J. Lu, C.-C. Chang, G.-D. Chen, and C. Cheewakriangkrai, "Integration of data mining classification techniques and ensemble learning to identify risk factors and diagnose ovarian cancer recurrence," *Artificial intelligence in medicine*, vol. 78, pp. 47–54, 2017.
- [8] J. C. Franklin, J. D. Ribeiro, K. R. Fox, K. H. Bentley, E. M. Kleiman, X. Huang, K. M. Musacchio, A. C. Jaroszewski, B. P. Chang, and M. K. Nock, "Risk factors for suicidal thoughts and behaviors: A meta-analysis of 50 years of research," *Psychological Bulletin*, vol. 143, no. 2, p. 187, 2017.
- [9] J. Xie, H. Wang, J. Zhang, C. Meng, Y. Kong, S. Mao, L. Xu, and W. Zhang, "A novel hybrid subset-learning method for predicting risk factors of atherosclerosis," in *2017 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*. IEEE, 2017, pp. 2124–2131.
- [10] X. Li, J. Yu, M. Li, and D. Zhao, "Discover high-risk factor combinations using bayesian network from national screening data in china," in *2017 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*. IEEE, 2017, pp. 1047–1051.
- [11] Y. Yang, X. Wang, Y. Huang, N. Chen, J. Shi, and T. Chen, "Ontology-based venous thromboembolism risk factors mining and model developing from medical records," in *2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*. IEEE, 2018, pp. 1669–1672.
- [12] L. Wang, D. Zhu, E. Towner, and M. Dong, "Obesity risk factors ranking using multi-task learning," in *Biomedical & Health Informatics (BHI), 2018 IEEE EMBS International Conference on*. IEEE, 2018, pp. 385–388.
- [13] A. Aussem, S. R. De Moraes, and M. Corbex, "Analysis of nasopharyngeal carcinoma risk factors with bayesian networks," *Artificial intelligence in Medicine*, vol. 54, no. 1, pp. 53–62, 2012.
- [14] M. Miravittles, T. Guerrero, C. Mayordomo, L. Sánchez-Agudo, F. Nicolau, and J. L. Segú, "Factors associated with increased risk of exacerbation and hospital admission in a cohort of ambulatory copd patients: a multiple logistic regression analysis," *Respiration*, vol. 67, no. 5, pp. 495–501, 2000.
- [15] J. A. Vita, C. B. Treasure, E. G. Nabel, J. M. McLenachan, R. D. Fish, A. C. Yeung, V. I. Vekshtein, A. P. Selwyn, and P. Ganz, "Coronary vasomotor response to acetylcholine relates to risk factors for coronary artery disease," *Circulation*, vol. 81, no. 2, pp. 491–497, 1990.
- [16] M. M. Finucane, G. A. Stevens, M. J. Cowan, G. Danaei, J. K. Lin, C. J. Paciorek, G. M. Singh, H. R. Gutierrez, Y. Lu, A. N. Bahalim *et al.*, "National, regional, and global trends in body-mass index since 1980: systematic analysis of health examination surveys and epidemiological studies with 960 country-years and 9·1 million participants," *The Lancet*, vol. 377, no. 9765, pp. 557–567, 2011.
- [17] Y. Li, Y. Cui, C. Wang, X. Liu, and J. Han, "A risk factor analysis on disease severity in 47 premature infants with bronchopulmonary dysplasia," *Intractable & rare diseases research*, vol. 4, no. 2, pp. 82–86, 2015.
- [18] L. Anderson, N. Oldridge, D. R. Thompson, A.-D. Zwisler, K. Rees, N. Martin, and R. S. Taylor, "Exercise-based cardiac rehabilitation for coronary heart disease: Cochrane systematic review and meta-analysis," *Journal of the American College of Cardiology*, vol. 67, no. 1, pp. 1–12, 2016.
- [19] D. K. Matthaiou, G. Dimopoulos, F. S. Taccone, P. Bulpa, A. Van den Abeele, B. Misset, W. Meersseman, H. Spapen, T. Cardoso, P. Charles *et al.*, "Elderly versus nonelderly patients with invasive aspergillosis in the icu: a comparison and risk factor analysis for mortality from the aspico cohort," *Medical mycology*, 2017.
- [20] A. C. Lopilato, C. T. Doligalski, and C. Caldeira, "Incidence and risk factor analysis for gastrointestinal bleeding and pump thrombosis in left ventricular assist device recipients?" *Artificial organs*, vol. 39, no. 11, pp. 939–944, 2015.
- [21] E. F. Devèvre, M. Renovato-Martins, K. Clément, C. Sautès-Fridman, I. Cremer, and C. Poitou, "Profiling of the three circulating monocyte subpopulations in human obesity," *The Journal of Immunology*, vol. 194, no. 8, pp. 3917–3923, 2015.
- [22] G. Obozinski, B. Taskar, and M. Jordan, "Multi-task feature selection," *Statistics Department, UC Berkeley, Tech. Rep*, vol. 2, 2006.
- [23] B. Bakker and T. Heskes, "Task clustering and gating for bayesian multitask learning," *Journal of Machine Learning Research*, vol. 4, no. May, pp. 83–99, 2003.
- [24] L. Jacob, J.-p. Vert, and F. R. Bach, "Clustered multi-task learning: A convex formulation," in *Advances in neural information processing systems*, 2009, pp. 745–752.
- [25] M. Gönen, *Analyzing receiver operating characteristic curves with SAS*. SAS Institute, 2007.
- [26] A. Galecki and T. Burzykowski, "Linear mixed-effects model," in *Linear Mixed-Effects Models Using R*. Springer, 2013, pp. 245–273.
- [27] J.-H. Liu, S. J. Jones, H. Sun, J. C. Probst, A. T. Merchant, and P. Cavicchia, "Diet, physical activity, and sedentary behaviors as risk factors for childhood obesity: an urban and rural comparison," *Childhood Obesity (Formerly Obesity and Weight Management)*, vol. 8, no. 5, pp. 440–448, 2012.
- [28] D. A. Dev, B. A. McBride, B. H. Fiese, B. L. Jones, and H. Cho, on behalf of the STRONG Kids Research Team, "Risk factors for overweight/obesity in preschool children: an ecological approach," *Childhood obesity*, vol. 9, no. 5, pp. 399–408, 2013.
- [29] T. Evgeniou and M. Pontil, "Regularized multi-task learning," in *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2004, pp. 109–117.
- [30] X.-T. Yuan, X. Liu, and S. Yan, "Visual classification with multitask joint sparse representation," *IEEE Transactions on Image Processing*, vol. 21, no. 10, pp. 4349–4360, 2012.
- [31] J. Zhou, J. Chen, and J. Ye, "Malsar: Multi-task learning via structural regularization," *Arizona State University*, vol. 21, 2011.
- [32] A.-A. Liu, Y.-T. Su, W.-Z. Nie, and M. Kankanhalli, "Hierarchical clustering multi-task learning for joint human action grouping and recognition," *IEEE transactions on pattern analysis and machine intelligence*, vol. 39, no. 1, pp. 102–114, 2017.
- [33] A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM journal on imaging sciences*, vol. 2, no. 1, pp. 183–202, 2009.
- [34] J. Liu, S. Ji, and J. Ye, "Multi-task feature learning via efficient l2, 1-norm minimization," in *Proceedings of the twenty-fifth conference on uncertainty in artificial intelligence*. AUAI Press, 2009, pp. 339–348.
- [35] H. Zha, X. He, C. Ding, M. Gu, and H. D. Simon, "Spectral relaxation for k-means clustering," in *Advances in Neural Information Processing Systems*, 2002, pp. 1057–1064.
- [36] J. Zhou, J. Chen, and J. Ye, "Clustered multi-task learning via alternating structure optimization," in *Advances in neural information processing systems*, 2011, pp. 702–710.
- [37] J. Pinheiro, D. Bates, S. DebRoy, and D. Sarkar, "R core team (2014) nlme: linear and nonlinear mixed effects models. r package version 3.1-117," Available at <http://CRAN.R-project.org/package=nlme>, 2014.