Multi-Label Transduction for Identifying Disease Comorbidity Patterns

Ehsan Adeli[1], Dongjin Kwon[1,2], and Kilian M. Pohl[2]

[1]Stanford University [2]SRI International

Abstract. Study of the untoward effects associated with the comorbidity of multiple diseases on brain morphology requires identifying differences across multiple diagnostic groupings. To identify such effects and differentiate between groups of patients and normal subjects, conventional methods often compare each patient group with healthy subjects using binary or multi-class classifiers. However, testing inferences across multiple diagnostic groupings of complex disorders commonly yield inconclusive or conflicting findings when the classifier is confined to modeling two cohorts at a time or considers class labels mutually-exclusive (as in multi-class classifiers). These shortcomings are potentially caused by the difficulties associated with modeling compounding factors of diseases with these approaches. Multi-label classifiers, on the other hand, can appropriately model disease comorbidity, as each subject can be assigned to two or more labels. In this paper, we propose a multi-label transductive (MLT) method based on low-rank matrix completion that is able not only to classify the data into multiple labels but also to identify patterns from MRI data unique to each cohort. To evaluate the method, we use a dataset containing individuals with Alcohol Use Disorder (AUD) and human immunodeficiency virus (HIV) infection (specifically 244 healthy controls, 227 AUD, 70 HIV, and 61 AUD+HIV). On this dataset, our proposed method is more accurate in correctly labeling subjects than common approaches. Furthermore, our method identifies patterns specific to each disease and AUD+HIV comorbidity that shows that the comorbidity is characterized by a compounding effect of AUD and HIV infection.

1 Introduction

Improvements in modern health-care together with the aging population caused populations with multiple conditions that require ongoing medical attention. The U.S. alone has approximately 75 million people living with 2 or more conditions [12] such as brain-related disorders. For brain-related conditions, disease comorbidity often leads to new cognitive impairments [8,9]. However, few studies have examined the potentially heightened burden of disease comorbidity. As an example, Alcohol Use Disorder (AUD) is common among individuals in the United States, and its co-occurrence in individuals with human immunodeficiency virus (HIV) infection is high [8], occurring at twice the rate as occurs in the

Supported in part by NIH grants AA017347, AA005965, AA010723, AA017168, and MH113406.

general population [9]. Both AUD and HIV infection reduce health-related quality of life. Adapting robust and multi-label technology can transform the mechanistic understanding of the compounding factors of such comorbidity.

Conventional models for identifying the compounding effects of disease comorbidity use binary (such as statistical tests [9]) or multi-class (*e.g.*, [15,12]) study designs. By nature, these methods assume samples are assigned to mutually-exclusive labels, and hence inaccurately model the compounding factors of disease comorbidity. Thus, they often lead to inconclusive or contradicting findings for the cases of comorbidity [14,16]. One can accurately model disease comorbidity by allowing multiple labels to be assigned to each subject, such as done by multi-label classification methods. However, there are two major challenges in applying multi-label methods to neuroimaging data: (1) neuroimaging data often contain several brain morphology measures that are highly correlated; (2) the measures are prone to noise due to inaccuracies in acquisition, preprocessing, and diagnosis of subjects. To overcome these challenges, we introduce a multi-label transductive (MLT) classification approach based on low-rank matrix completion (MC) [7,2] that models noise and overcomes the problem of feature redundancy/correlation by reducing the data to low-rank subspaces. In addition, our method identifies patterns related to each disease (label).

To gain a better understanding of MLT, consider a training data set of N_{tr} samples with l different labels (which lead to $c = 2^l$ number of different classes). Let d be the dimensionality of the feature space, $N = N_{tr} + N_{ts}$ the number of total samples (N_{ts} the number of testing samples). Our MLT then determines the labels of the testing samples by first combining all features (of training and testing) and the labels of the training data into a matrix. Here, we propose to add columns to the matrix, in which only the labels but not the features are defined. These allow us to compute disease specific patterns, which, together with the testing labels are computed by MLT through matrix completion.

We apply our model to identify the impact of alcohol use and HIV infection in the brain morphometry of individuals. We interpret the two conditions `AUD` and `HIV` as labels that encode four classes: `CTRL` (samples are assigned to none of the two labels), `AUD` (only assigned to the first label), `HIV` (only assigned to the second label), and `AUD+HIV` (having both labels). Fig. 1 shows the corresponding matrix whose missing values are computed by our method (see[1] for notations).

In summary, we make two contributions: we model disease comorbidity within a transductive multi-label setting that is robust to noise, and we identify the disease-specific patterns by modifying the original MC algorithm [7].

2 Multi-Label Transduction via Matrix Completion

MC is the process of recovering a matrix from a sampling of its entries. Specifically, MC recovers a data matrix $\mathbf{D}$ from a matrix $\mathbf{D}^0$ for which only the subset Ω

[1] Bold capital letters denote matrices (*e.g.*, $\mathbf{D}$), and bold small letters denote vectors (*e.g.*, $\mathbf{d}$). All non-bold letters are scalar variables. d_{ij} is the scalar in row i and column j of $\mathbf{D}$. $|A|$ denotes the number of elements in set A. $\|\mathbf{D}\|_*$ designates the nuclear norm (sum of singular values) of $\mathbf{D}$.

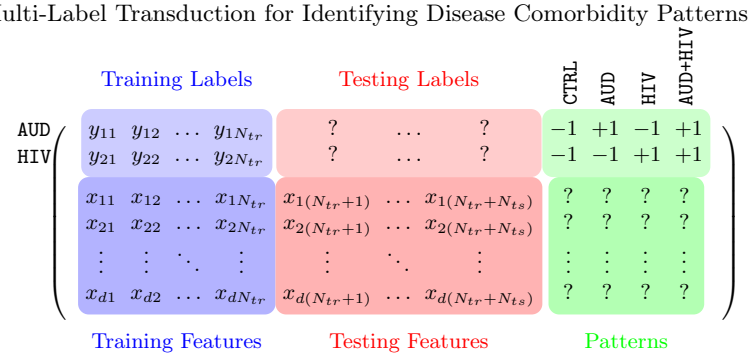

Fig. 1: Multi-label transduction (MLT) via matrix completion: each column represents one sample, the top rows hold the labels (two labels in this example, *i.e.*, AUD and HIV), while the rest are comprised of the features. The last columns include the representative features for each of the classes.

of its entries are defined (or observed). Knowing a sufficiently large number of measurements in the matrix, MC assumes that the relation between the elements of the matrix can be accurately described by a low-rank matrix [3,7,2]. Thus, MC estimates the missing entries in $\mathbf{D}$ by minimizing its rank (rank($\mathbf{D}$)) while $d_{ij} = d^0_{ij}, \forall (i,j) \in \Omega$. The rank function is a non-convex, non-smooth function that can be approximated by the nuclear norm of the matrix, *i.e.*, the sum of its non-vanishing singular values, σ_k: $\|\mathbf{D}\|_* = \sum_{k=1}^{d} \sigma_k(\mathbf{D})$.

To apply MC to classification, we note that classifiers generally learn the relation between the features, $\mathbf{X}_{tr} \in \mathbb{R}^{d \times N_{tr}}$, and the labels, $\mathbf{Y}_{tr} = \mathbb{R}^{l \times N_{tr}}$, of the training samples. Learning this relation while at the same time classifying or determining the labels $\mathbf{Y}_{ts} \in \mathbb{R}^{l \times N_{ts}}$ of the N_{ts} testing samples (*i.e.*, $\mathbf{X}_{ts} \in \mathbf{R}^{d \times N_{ts}}$) can again be interpreted as a Matrix Completion task [7] of a matrix that contains all training and testing data (see also Fig. 1). MC assumption about rank deficiency of the matrix is equivalent to assuming a linear relationship between the measurements and labels, which is a common assumption of classifiers. However, unlike inductive methods that learn a separate model and then apply it to testing data, MC classifies the testing samples by filling in the submatrix $\mathbf{Y}_{ts}$ of missing entries such that rank of the matrix is minimized. As discussed earlier (and illustrated in Fig. 1), we propose to extract class-specific patterns by adding columns (*i.e.*, submatrices $\mathbf{Y}_{pat} \in \mathbb{R}^{l \times c}$ and $\mathbf{X}_{pat} \in \mathbb{R}^{d \times c}$) with empty entries for the features to the matrix. As in [7,2], a row with the entries of all 1 (called '1' row) is added to the matrix modeling the bias term of linear classifiers (see Eq. (1)). In summary, $\mathbf{D}^0$ is formed by concatenating all features ($\mathbf{X}_{tr}$, $\mathbf{X}_{ts}$, $\mathbf{X}_{pat}$), labels ($\mathbf{Y}_{tr}$, $\mathbf{Y}_{ts}$, $\mathbf{Y}_{pat}$), and the '1' row.

The set of known entries in $\mathbf{D}^0$ is now defined by the 'feature' submatrix Ω_X (*i.e.*, all training $\mathbf{X}_{tr}$ and testing $\mathbf{X}_{ts}$ features) and the label submatrix Ω_Y (*i.e.*, all training labels $\mathbf{Y}_{tr}$ and the pattern-specific labels $\mathbf{Y}_{pat}$). Note, that the extracted neuroimaging measurements (used as features) may be contaminated by noise due to inaccuracies in acquisition or preprocessing. In addition, the brain

of a patient might not yet be impacted by a disease so that the patient looks more like a sample from the control cohort. This variability within cohorts can also be interpreted as noise associated with the training labels. We model those sources of noise in our MC approach by introducing the matrix $\mathbf{E}$ and defining $\mathbf{D}$ as sum between $\mathbf{D}^0$ and $\mathbf{E}$:

$$\mathbf{D} = \begin{bmatrix} \mathbf{D_Y} \\ \mathbf{D_X} \\ \mathbf{D_1} \end{bmatrix} = \begin{bmatrix} \mathbf{Y}_{tr} & \mathbf{Y}_{ts} & \mathbf{Y}_{pat} \\ \mathbf{X}_{tr} & \mathbf{X}_{ts} & \mathbf{X}_{pat} \\ & \mathbf{1}^\top & \end{bmatrix} + \begin{bmatrix} \mathbf{E}_{\mathbf{Y}_{tr}} & \mathbf{0} & \mathbf{0} \\ \mathbf{E}_{\mathbf{X}_{tr}} & \mathbf{E}_{\mathbf{X}_{ts}} & \mathbf{0} \\ & \mathbf{0}^\top & \end{bmatrix} = \mathbf{D}^0 + \mathbf{E}, \tag{1}$$

where $\mathbf{D_Y}$, $\mathbf{D_X}$, and $\mathbf{D_1}$ are the label, feature and '1' rows of $\mathbf{D}$, respectively. In other words, MC determines the testing labels $\mathbf{Y}_{ts}$, representative patterns for each class $\mathbf{X}_{pat}$, and the noise $\mathbf{E}$ such that the rank of $\mathbf{D} = \mathbf{D}_0 + \mathbf{E}$, *i.e.*, $\|\mathbf{D}\|_*$, is minimized. To further constrain the optimization problem and avoid trivial solutions, we introduce the squared loss function $\mathcal{L}_x(d_{ij}, d^0_{ij}) = \frac{1}{2}(d_{ij} - d^0_{ij})^2$ as a way to penalize large differences in features (*i.e.*, noise in features) between the two matrices and a log loss function $\mathcal{L}_y(d_{ij}, d^0_{ij}) = \frac{1}{\gamma}\log(1 + e^{-\gamma d_{ij} d^0_{ij}})$ to penalize difference in class assignment between the matrices (*i.e.*, noise in labeling). $\gamma = 1$ in our experiments. The complete minimization problem is then defined as

$$\begin{aligned} &\underset{\mathbf{D},\mathbf{E}}{\operatorname{argmin}} \ \|\mathbf{D}\|_* + \frac{\lambda_1}{|\Omega_X|} \sum_{ij \in \Omega_X} \mathcal{L}_x(d_{ij}, d^0_{ij}) + \frac{\lambda_2}{|\Omega_Y|} \sum_{ij \in \Omega_Y} \omega_{ij} \mathcal{L}_y(d_{ij}, d^0_{ij}) \\ &\text{subject to } \mathbf{D} = \mathbf{D}^0 + \mathbf{E}, \mathbf{D_1} = \mathbf{1}^\top, \end{aligned} \tag{2}$$

with hyperparamters $0 \leq \lambda_1, \lambda_2 \leq 1$ controlling the influence of the loss functions in the minimization problem and the weight ω_{ij} of each training sample accounting for the imbalance in number of samples per cohort. This weight, computed as a processing step, weights the loss for each sample disproportional to the number of samples available in the training set with the same label y_{ij} of sample j.

Solving the optimization problem. As in [7,2,1], Eq. (2) is a constrained, convex optimization problem that is not smooth due to the nuclear norm (first term). We estimate its solution via the fixed-point continuation (FPC) [7], one of the interior point methods that can be applied to such problems. FPC iteratively alternatives between updating the gradient of the loss terms $\mathcal{L}_x(\cdot)$ and $\mathcal{L}_y(\cdot)$, and the singular value thresholding (SVT) [3], which is used for minimizing the nuclear norm. Cabral *et al.* [2] proved the constrained optimization (similar settings as in Eq. (2)) converges to the optimal solution.

Identifying Disease Patterns. As a result of completing the matrix in Eq. (1), the unknown entries in $\mathbf{X}_{pat}$ are determined. Since MC minimizes the rank of the matrix, it fills the entries in $\mathbf{X}_{pat}$ such that they have maximum correlations with their respective labels in $\mathbf{Y}_{pat}$ (*i.e.*, $(-1,-1), (-1,+1), (+1,-1), (+1,+1)$). The columns of $\mathbf{X}_{pat}$ can, hence, be considered as representative patterns for each class. Since we are interested in identifying how each of these classes are different from the control group, we simply compute the difference between the representative pattern of each class and the `CTRL` group, *i.e.*,

$$\mathbf{p}^{\texttt{AUD}} = |\mathbf{x}^{\texttt{CTRL}}_{pat} - \mathbf{x}^{\texttt{AUD}}_{pat}|, \ \mathbf{p}^{\texttt{HIV}} = |\mathbf{x}^{\texttt{CTRL}}_{pat} - \mathbf{x}^{\texttt{HIV}}_{pat}|, \text{ and } \mathbf{p}^{\texttt{AUD+HIV}} = |\mathbf{x}^{\texttt{CTRL}}_{pat} - \mathbf{x}^{\texttt{AUD+HIV}}_{pat}|,$$

where $\mathbf{p}^{\mathtt{C}}$ denotes the identified pattern for class $\mathtt{C}$, and $\mathbf{x}^{\mathtt{C}}_{pat}$ is the corresponding column in $\mathbf{X}_{pat}$ for class $\mathtt{C}$. We view small differences as noise and omit them from our findings by introducing the tolerance threshold ϵ and discarding the values $\mathbf{p}^{\mathtt{C}} < \epsilon$ ($\epsilon = 10^{-3}$ in our experiments).

3 Experiments

We now compare the accuracy of the proposed and alternative methods with respect to the multi-label AUD/HIV dataset. Alternative implementations include running MC separately for each label, denoted by Single Label MC (SL-MC) [7], and running our proposed method (similar to [3]) without considering the error matrix $\mathbf{E}$ (see Eq. (1); denoted as MLT-EMC). The comparison also includes the widely used multi-class SVM (MC-SVM) [4] and multi-label SVM (ML-SVM) [10,4]. For fair comparison, we ran the SVMs by weighing samples (similar to the weights ω_{ij} we used in our formulation) in the corresponding cost function according to the size of the associated class.

3.1 Dataset and Preprocessing

As summarized in Table 1, the dataset consists of the morphometric measurements extracted from the magnetic resonance images (MRIs) of 244 healthy controls (CTRL), 227 AUD subjects without HIV infection, 70 HIV-infected individuals that do not meet the criteria for AUD (HIV), and 61 subjects with both AUD and HIV infection (HIV+AUD). For additional details about data collection and preprocessing, please refer to [13].

Table 1: Details of the multi-label dataset ('svol' = supratentorial volume).

	Total	**sex**		**Age** (years)	**svol** ($\times 10^6$)
		F	M		
CTRL	244	122	123	45.59 ± 17.17	1.27 ± 0.13
AUD	227	67	159	48.49 ± 10.04	1.27 ± 0.11
HIV	70	20	45	51.81 ± 8.44	1.27 ± 0.15
AUD+HIV	61	23	43	50.97 ± 8.12	1.23 ± 0.14

We apply the cross-sectional approach of FreeSurfer 5.3.0 software to the skull-stripped T1w MRI of each subject in order to measure the *mean curvature (MeanCurv)*, *surface area (SurfArea)*, *gray matter volume (GrayVol)*, and *average thickness (ThickAvg)* of 34 bilateral cortical Regions Of Interest (ROIs), the volumes of 8 bilateral subcortical ROIs (*i.e.*, thalamus, caudate, putamen, pallidum, hippocampus, amygdala, accumbens, cerebellar cortex), the volumes of 5 subregions of the corpus callosum (posterior, mid-posterior, central, mid-central and anterior), the volume of all white matter hypointensities, the left and right lateral and third ventricles, and the supratentorial volume (svol). In addition to svol, each subject is thus represented by the z-scores of 298 morphometric features.

Confounding Factors. With respect to the CTRL group, *age*, *sex*, and *svol* significantly impact (p-value < 0.001) the morphemetric measurements according to the paired t-test between each demographic factor and feature. To omit their influence from the analysis, we capture the relationship between each feature

and the confounding factors by parameterizing a generalized linear model (GLM) [11] on the CTRL cohort of each training run within the cross-validation. After parameterizing GLM, the model is applied to the measurements of each subject to compute the residual score that are indifferent to the confounding factors. Note, the GLM model is only trained on the data from the training folds not to involve testing data in the preprocessing stage.

Evaluation. The classification accuracy of each method is measured via 10-fold nested cross-validation with the hyperparameters determined via 5-fold inner cross-validation. For MLT and SL-MC, the search space of the hyperparameters λ_1 and λ_2 is $\{0.001, 0.01, 0.1, 0.5, 0.9, 1\}$, and for MLT-EMC we do not have those hyperparameters as there is no error terms associated. The setting of FPC hyperparameters are set according to [2]. We also rely on the literature to set the search space for the hyperparameters of the alternative approaches, each based on their respective references, MC-SVM [4] and ML-SVM [10] (*e.g.*, the search space for the hyperparameter C of SVM is $\{0.01, 0.1, 1, 10, 100\}$).

We summarize the outcome of each approach through several accuracy scores. Specifically, for each class $\mathtt{C}$ with $P^{\mathtt{C}}$ positive and $N^{\mathtt{C}}$ negative samples, we compute for each approach the precision (Pre) and recall (Rec) based on the true positive (TP), true negative (TN), false positive (FP) and false negative (FN) of the classifier: $\text{Pre}^{\mathtt{C}} = \text{TP}^{\mathtt{c}}/(\text{TP}^{\mathtt{c}} + \text{FP}^{\mathtt{c}})$ and $\text{Rec}^{\mathtt{C}} = \text{TP}^{\mathtt{c}}/(\text{TP}^{\mathtt{c}} + \text{FN}^{\mathtt{c}})$. We also report the area under the ROC curve (AUC) and balanced accuracy (BAc) score for all methods: $\text{BAc}^{\mathtt{C}} = 1/2 \cdot (\text{TP}^{\mathtt{c}}/P^{\mathtt{c}} + \text{TN}^{\mathtt{c}}/N^{\mathtt{c}})$. To evaluate the performance with respect to each separate class, we also use true positive rate (TPR) metric, which shows the portion of subjects identified correctly: $\text{TPR}^{\mathtt{C}} = \text{TP}^{\mathtt{c}}/P^{\mathtt{c}}$.

3.2 Results and Comparison

Table 2 summarizes the accuracy scores with respect to the two defined labels (*i.e.*, `AUD` and `HIV`. BAc scores marked with a '†' are associated with a Fisher's exact test [6] that was significantly better than chance (*i.e.*, p-value < 0.001). Our proposed method obtains better results in terms of both balanced accuracy and AUC compared to all other methods. In comparison with the second best method (ML-SVM), our BAc scores are by at least 2% and our AUC scores by at least 3% better in both experiments. Also, our method obtaines a better balance between precision and recall, specially for the highly imbalanced case of `HIV`. The results confirm our intuition that the multi-label setting is better in modeling the problem compared to the multi-class methods, as both multi-label methods involved in the comparison (MLT and ML-SVM) lead to the best results. Both SL-MC and MC-SVM perform inferior to them in terms of both balanced

Table 2: Comparison for each single label. † sign indicates a p-value < 0.001 in a Fisher exact test.

	AUD				HIV			
Method	**BAc**	**Pre**	**Rec**	**AUC**	**BAc**	**Pre**	**Rec**	**AUC**
Ours (MLT)	**.68**†	**.67**	**.69**	**.69**	**.78**†	**.75**	.79	**.80**
SL-MC [7]	.61	.63	.58	.65	.72†	.68	.70	.70
MLT-EMC [3]	.52	.53	.51	.55	.59	.60	.56	.62
MC-SVM [4]	.63	.61	.63	.60	.72†	.68	.74	.70
ML-SVM [10]	.66†	.64	.65	.66	.75†	.69	**.80**	.74

accuracy and AUC. Furthermore, MLT-EMC, *i.e.*, the method without the noise term, obtains the worst results.

Table 3 summarizes the TPR with respect to all four classes (`CTRL`, `AUD`, `HIV`, and `AUD+HIV`) and the overall mean across the four classes. In a four-class classification problem, a TPR higher than 0.25 is considered better than chance. The scores agree with the findings of Table 2 in that the multi-label methods (*i.e.*, MLT and ML-SVM) lead to the best models. Our proposed method outperforms ML-SVM by 3% with respect to the overall mean TPR.

Table 3: Class specific TPR and overall mean.

Method	`CTRL`	`AUD`	`HIV`	`AUD+HIV`	**Mean**
Ours (MLT)	**.57**	**.54**	**.57**	**.59**	**.57**
SL-MC [7]	.38	.35	.32	.33	.35
MLT-EMC [3]	.37	.32	.33	.31	.33
MC-SVM [4]	.42	.44	.40	.44	.43
ML-SVM [10]	.55	.51	.52	.56	.54

Fig. 2 shows the identified patterns by our proposed method. These patterns are composed of approximately 20% of the 298 features for `AUD`, 16% for `HIV` and 40% for `AUD+HIV`, with approximately 5% of the features shared among all these cohorts. These identified patterns suggest that AUD and HIV infection are associated with deficits in cortical and subcortical regions, which agrees with the HIV and Alcohol literature [5,8,9,13]. Furthermore, several previous studies have reported cortical thickness and gray matter volume as important markers for AUD and HIV infection [9,13], which agrees with our findings. Subcortical regions, including hippocampus, thalamus and basal ganglia structures (*i.e.*, Caudate, Putamen, Palladium), are reported in the literature to be affected by HIV and AUD [5], which are also found important by our method. Specifically, the primary motor cortex region, and the basal ganglia subcortical structures are shown to be more severely affected as a result of comorbidity of AUD and HIV, compared to each single one of them.

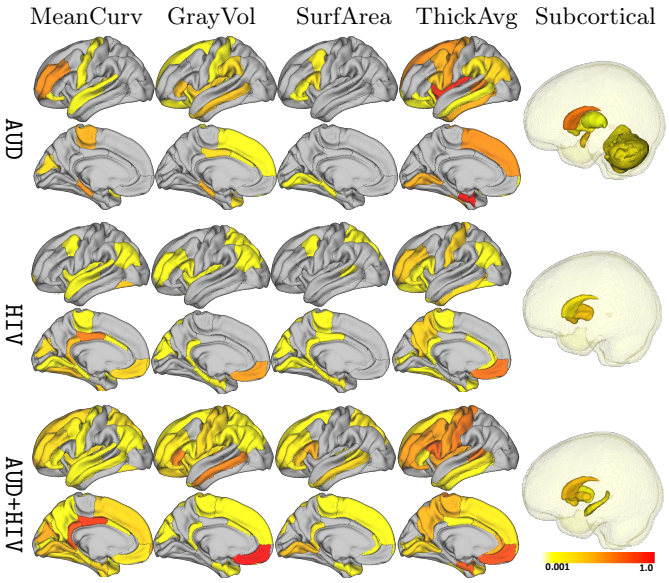

Fig. 2: Identified patterns for `AUD`, `HIV`, and `AUD+HIV`.

Discussions. The identified patterns of `AUD+HIV` record the largest number of relevant regions, which can document a compounding effect of `AUD` and `HIV`. Another interesting finding of our results is that our method avoids underestimating the impact of the disease to a small number of brain regions as commonly done by sparse classifiers. In addition, since our method spans the data matrix to a

low-rank subspace, it can implicitly alleviate the redundancy and correlation among the features that are (linearly) correlated.

4 Conclusion

In this paper, we introduced a multi-label transductive classifier that not only classifies the data into multiple labels but also can identify the patterns specific to each disease (labels) and their comorbidity. We experimented on a large set of data with control samples and subjects with alcohol use disorder, HIV infection or both. Our model led to the superior accuracy scores in comparison to state-of-the-art methods. Our method was also able to identify disease-related (and their comorbidity) patterns, which revealed that the comorbodity was characterized by a compounding effect of the two disorders.

References

1. Adeli, E., Fathy, M.: Non-negative matrix completion for action detection. Image Vis Comput. 39, 38–51 (2015)
2. Cabral, R., Torre, F., Costeira, J., Bernardino, A.: Matrix completion for weakly-supervised multi-label image classification. IEEE TPAMI 37(1), 121–135 (2015)
3. Candès, E.J., Recht, B.: Exact matrix completion via convex optimization. Found. Comput. Math. 9(6), 717–772 (2009)
4. Chang, C.C., Lin, C.J.: LIBSVM: A library for support vector machines. ACM Trans. IST 2(27), 1–27 (2011)
5. Fama, R., et al.: Impairments in component processes of executive function and episodic memory in alcoholism, hiv infection, and hiv infection with alcoholism comorbidity. Alcohol Clin Exp Res 40(12), 2656–2666 (2016)
6. Fisher, R.A.: The logic of inductive inference. Journal of the Royal Statistical Society 98(1), 39–82 (1935)
7. Goldberg, A., Recht, B., Xu, J., Nowak, R., Zhu, X.: Transduction with matrix completion: Three birds with one stone. In: NIPS. pp. 757–765 (2010)
8. Gongvatana, A., et al.: A history of alcohol dependence augments hiv-associated neurocognitive deficits in persons aged 60 and older. Neurovirology 20(5) (2014)
9. Justice, A., Sullivan, L., Fiellin, D., et al.: Hiv/aids, comorbidity, and alcohol: Can we make a difference? Alcohol Research & Health 33(3), 258 (2010)
10. Li, X., Guo, Y.: Active learning with multi-label SVM classification. In: IJCAI (2013)
11. Madsen, H., Thyregod, P.: Introduction to general and generalized linear models. CRC Press (2010)
12. Parekh, A.K., Barton, M.B.: The challenge of multiple comorbidity for the us health care system. Jama 303(13), 1303–1304 (2010)
13. Pfefferbaum, A., et al.: Accelerated aging of selective brain structures in human immunodeficiency virus infection. Neurobiology of aging 35(7), 1755–1768 (2014)
14. Tsoumakas, G., Katakis, I.: Multi-label classification: An overview. International Journal of Data Warehousing and Mining 3(3) (2006)
15. Wang, X., Wang, F., Hu, J.: A multi-task learning framework for joint disease risk prediction and comorbidity discovery. In: ICPR. pp. 220–225. IEEE (2014)
16. Wosiak, A., Glinka, K., Zakrzewska, D.: Multi-label classification methods for improving comorbidities identification. Comput. Biol. Med (2017)