

# Multi-label Inductive Matrix Completion for Joint MGMT and IDH1 Status Prediction for Glioma Patients

Lei Chen<sup>1,2</sup>, Han Zhang<sup>2</sup>, Kim-Han Thung<sup>2</sup>, Luyan Liu<sup>3</sup>,  
Junfeng Lu<sup>4,5</sup>, Jinsong Wu<sup>4,5</sup>, Qian Wang<sup>3</sup>, and Dinggang Shen<sup>2(✉)</sup>

<sup>1</sup> Jiangsu Key Laboratory of Big Data Security and Intelligent Processing,  
Nanjing University of Posts and Telecommunications, Nanjing, China

<sup>2</sup> Department of Radiology and BRIC,  
University of North Carolina, Chapel Hill, USA  
dgshen@med.unc.edu

<sup>3</sup> School of Biomedical Engineering, Med-X Research Institute,  
Shanghai Jiao Tong University, Shanghai, China

<sup>4</sup> Department of Neurosurgery, Huashan Hospital,  
Fudan University, Shanghai, China

<sup>5</sup> Shanghai Key Lab of Medical Image Computing and Computer Assisted  
Intervention, Shanghai, China

**Abstract.** MGMT promoter methylation and IDH1 mutation in high-grade gliomas (HGG) have proven to be the two important molecular indicators associated with better prognosis. Traditionally, the statuses of MGMT and IDH1 are obtained via surgical biopsy, which is laborious, invasive and time-consuming. Accurate presurgical prediction of their statuses based on preoperative imaging data is of great clinical value towards better treatment plan. In this paper, we propose a novel Multi-label *Inductive* Matrix Completion (MIMC) model, highlighted by the online inductive learning strategy, to jointly predict both MGMT and IDH1 statuses. Our MIMC model *not only* uses the training subjects with possibly missing MGMT/IDH1 labels, *but also* leverages the unlabeled testing subjects as a supplement to the limited training dataset. More importantly, we learn inductive labels, instead of directly using transductive labels, as the prediction results for the testing subjects, to alleviate the overfitting issue in small-sample-size studies. Furthermore, we design an optimization algorithm with guaranteed convergence based on the block coordinate descent method to solve the multivariate non-smooth MIMC model. Finally, by using a precious single-center multi-modality presurgical brain imaging and genetic dataset of primary HGG, we demonstrate that our method can produce accurate prediction results, outperforming the previous widely-used single- or multi-task machine learning methods. This study shows the promise of utilizing imaging-derived brain connectome phenotypes for prognosis of HGG in a non-invasive manner.

**Keywords:** High-grade gliomas · Molecular biomarker · Matrix completion

## 1 Introduction

Gliomas account for approximately 45% of primary brain tumors. Most deadly gliomas are classified by World Health Organization (WHO) as grades III and IV, referred to as high-grade gliomas (HGG). Related studies have shown that O6-methylguanine-DNA methyltransferase promoter methylation (MGMT-m) and isocitrate dehydrogenases mutation (IDH1-m) are the two strong molecular indicators that may associate with better prognosis (i.e., better sensitivity to the treatment and longer survival time), compared to their counterparts, i.e., MGMT promoter unmethylation (MGMT-u) and IDH1 wild (IDH1-w) [1, 2]. To date, the identification of the MGMT and IDH1 statuses is becoming clinical routine, but conducted via invasive biopsy, which has limited their wider clinical implementation. For better treatment planning, non-invasive and preoperative prediction of MGMT and IDH1 statuses is highly desired.

A few studies have been carried out to predict MGMT/IDH1 status based on the preoperative neuroimages. For example, Korfiatis *et al.* extracted tumor texture features from a single T2-weighted MRI modality, and trained a support vector machine (SVM) to predict MGMT status [1]. Yamashita *et al.* extracted both the functional information (i.e., tumor blood flow) from perfusion MRI and the structural features from T1-weighted MRI, and employed a nonparametric approach to predict IDH1 status [2]. Zhang *et al.* extracted more voxel- and histogram-based features from T1-, T2-, and diffusion-weighted images (DWI), and employed a random forest (RF) classifier to predict IDH1 status [3].

However, all these studies are limited to predict *either* MGMT *or* IDH1 status alone by using a single-task machine learning technique, which simply ignores the potential relationship of these two molecular expressers that may help each other to achieve more accurate prediction results [4]. It is desirable to use multi-task learning approach to jointly predict the MGMT and IDH1 statuses. Meanwhile, in the clinical practice, a complete molecular pathological testing may not always be conducted; therefore, in several cases there is only one biopsy-proven MGMT or IDH1 status, which leads to incomplete training labels or a missing data problem. Traditional methods usually simply discard the subjects with incomplete labels, which, however, further reduces the number of training samples. The recently proposed Multi-label *Transductive* Matrix Completion (MTMC) model is an important multi-task classification method, which can make full use of the samples with missing labels [5] and has produced good performance in many previous studies [5, 6]. However, it is difficult to be generalized to a study with a limited sample size due to its inherent overfitting; thus, many phenotype-genotype studies inevitably suffer from such a problem.

In order to address the above limitations, we propose a novel Multi-label *Inductive* Matrix Completion (MIMC) model by introducing an online inductive learning strategy into the MTMC model. However, the solution of MIMC is not trivial, since it contains both the non-smooth nuclear-norm and  $L_{21}$ -norm constraints. Therefore, based on the block coordinate descent method, we design an optimization algorithm to optimize the MIMC model. Note that, in this paper, we do not adopt the commonly used *radiomics* information derived from T1- or T2-weighted structural MRI, but instead use the *connectomics* information derived from both resting-state functional MRI (RS-fMRI)

and diffusion tensor imaging (DTI). The motivation behind this is that the structural MRI-based radiomics features are highly affected by tumor characteristics (e.g., locations and sizes) and thus significantly variable across subjects, which is undesirable for group study and also individual-based classification. On the other hand, brain connectome features extracted from RS-fMRI and DTI reflect the inherent brain connectivity architecture and its alterations due to the highly diffusive HGG, and thus could be more consistent and reliable as imaging biomarkers.

## 2 Materials, Preprocessing, and Feature Extraction

Our dataset includes 63 HGG patient subjects, which were recruited during 2010–2015. Each subject has at least one biopsy-proven MGMT or IDH1 status. We exclude the subjects without entire RS-fMRI or DTI, or with significant imaging artifacts as well as excessive head motion. Finally, 47 HGG subjects are used in this paper. We summarize subjects’ demographic and clinical information in Table 1. For simplicity, MGMT-m and IDH1-m are labeled as “positive”, respectively, and MGMT-u and IDH1-w as “negative”. This study has been approved by the local ethical committee at local hospital.

**Table 1.** Demographic and clinical information of the subjects involved in this study.

	MGMT	IDH1
Pos. labeled	26	13
Neg. labeled	20	33
Unlabeled	1	1
Age (mean/range)	48.13/23–68	
Gender (M/F)	26/21	
WHO III/IV	23/24	

In this study, all the RS-fMRI and DTI data are collected preoperatively with the following parameters. **RS-fMRI:** TR (repetition time) = 2 s, number of acquisitions = 240 (8 min), and voxel size =  $3.4 \times 3.4 \times 4 \text{ mm}^3$ . **DTI:** 20 directions, voxel size =  $2 \times 2 \times 2 \text{ mm}^3$ , and multiple acquisitions = 2. SPM8 and DPARSF [7] are used to preprocess RS-fMRI data and construct brain *functional* networks. FSL and PANDA [8] are used to process DTI and construct brain *structural* networks. Multi-modality images are first co-registered within the same subject, and then registered to the atlas space. All the processing procedures are following the commonly accepted pipeline [9]. Specifically, we parcellate each brain into 90 regions of interest (ROIs) using Automated Anatomical Labeling (AAL) atlas. The parcellated ROIs in each subject are regarded as nodes in a graph, while the Pearson’s correlation coefficient between the blood oxygenation level dependent (BOLD) time series from each pair of the ROIs are calculated as the functional connectivity strength for the corresponding edge in the graph. Similarly, the structural network is constructed based on

the whole-brain DTI tractography by calculating the normalized number of the tracked main streams as the structural connectivity strength for each pair of the AAL ROIs.

After network constructions, we use GRETNA [10] to extract various network properties based on graph theoretic analysis, including degree, shortest path length, clustering coefficient, global efficiency, local efficiency, and nodal centrality. These complex network properties are extracted as the connectomics features for each node in each network. We also use 12 clinical features for each subject, such as patient’s age, gender, tumor size, tumor WHO grade, tumor location, etc. Therefore, each subject has 1092 (6 metrics  $\times$  2 networks  $\times$  90 regions + 12 clinical features) features.

### 3 MIMC-Based MGMT and IDH1 Status Prediction

We first introduce the notations used in this paper.  $\mathbf{X}^{(i)}$  denotes the  $i$ -th column of matrix  $\mathbf{X}$ .  $\mathbf{x}_{ij}$  denotes the element in the  $i$ -th row and  $j$ -th column of matrix  $\mathbf{X}$ .  $\mathbf{1}$  denotes all-one column vector.  $\mathbf{X}^T$  denotes the transpose of matrix  $\mathbf{X}$ .  $\mathbf{X}_{train} = [\mathbf{x}_1, \dots, \mathbf{x}_m]^T \in \mathbb{R}^{m \times d}$  and  $\mathbf{X}_{test} = [\mathbf{x}_{m+1}, \dots, \mathbf{x}_{m+n}]^T \in \mathbb{R}^{n \times d}$  denote the feature matrices associated with  $m$  training subjects and  $n$  testing subjects, respectively. Assume there are  $t$  binary classification tasks, and  $\mathbf{Y}_{train} = [\mathbf{y}_1, \dots, \mathbf{y}_m]^T \in \{-1, 1, ?\}^{m \times t}$  and  $\mathbf{Y}_{test} = [\mathbf{y}_{m+1}, \dots, \mathbf{y}_{m+n}]^T \in \{?\}^{n \times t}$  denote the label matrices associated with  $m$  training subjects and  $n$  testing subjects, where ‘?’ denotes the unknown label. Furthermore, for the convenience of description, let  $\mathbf{X}^{obs} = [\mathbf{X}_{train}; \mathbf{X}_{test}]$ ,  $\mathbf{Y}^{obs} = [\mathbf{Y}_{train}; \mathbf{Y}_{test}]$  and  $\mathbf{Z}^{obs} = [\mathbf{X}^{obs}, \mathbf{1}, \mathbf{Y}^{obs}]$  denote the observed feature matrix, label matrix, and stacked matrix, respectively. Let  $\mathbf{X}^0 \in \mathbb{R}^{(m+n) \times d}$  denote the underlying noise-free feature matrix corresponding to  $\mathbf{X}^{obs}$ . Let  $\mathbf{Y}^0 \in \mathbb{R}^{(m+n) \times t}$  denote the underlying soft label matrix, and  $\mathbf{sign}(\mathbf{Y}^0)$  for the underlying label matrix corresponding to  $\mathbf{Y}^{obs}$ , where  $\mathbf{sign}(\cdot)$  is the element-wise sign function.

#### 3.1 Multi-label Transductive Matrix Completion (MTMC)

MTMC is a well-known multi-label matrix completion model, which is developed with two assumptions. First, linear relationship is assumed between  $\mathbf{X}^0$  and  $\mathbf{Y}^0$ , i.e.,  $\mathbf{Y}^0 = [\mathbf{X}^0, \mathbf{1}] \mathbf{W}$ , where  $\mathbf{W} \in \mathbb{R}^{(d+1) \times t}$  is the implicit weight matrix. Second,  $\mathbf{X}^0$  is also assumed to be low-rank. Let  $\mathbf{Z}^0 = [\mathbf{X}^0, \mathbf{1}, \mathbf{Y}^0]$  denote the underlying stacked matrix corresponding to  $\mathbf{Z}^{obs}$ , and then from  $\text{rank}(\mathbf{Z}^0) \leq \text{rank}(\mathbf{X}^0) + 1$ , we can infer that  $\mathbf{Z}^0$  is also low-rank. The goal of MTMC is to estimate  $\mathbf{Z}^0$  given  $\mathbf{Z}^{obs}$ . In the real application, where  $\mathbf{Z}^{obs}$  is contaminated by noise, MTMC is formulated as:

$$\min_{\mathbf{Z}^{(d+1)}=1} \mu \|\mathbf{Z}\|_* + \frac{1}{2} \|\mathbf{Z}_{\Delta \mathbf{X}} - \mathbf{X}^{obs}\|_F^2 + \gamma \sum_{(i,j) \in \Omega_Y} \mathbb{C}_Y(\mathbf{z}_{i(d+1+j)}, \mathbf{y}_{ij}^{obs}), \quad (1)$$

where  $\mathbf{Z} = [\mathbf{Z}_{\Delta\mathbf{X}}, \mathbf{Z}^{(d+1)}, \mathbf{Z}_{\Delta\mathbf{Y}}]$  denotes the matrix to be optimized,  $\mathbf{Z}_{\Delta\mathbf{X}}$  denotes the noise-free feature submatrix,  $\mathbf{Z}_{\Delta\mathbf{Y}}$  denotes the soft label submatrix,  $\Omega_Y$  denotes the subscripts set of the observed entries in  $\mathbf{Y}^{obs}$ ,  $\|\cdot\|_*$  denotes the nuclear norm,  $\|\cdot\|_F$  denotes the Frobenius norm, and  $\mathbb{C}_y(\cdot, \cdot)$  denotes the logistic loss function. Once the optimal  $\mathbf{Z}^{opt}$  is found, the labels  $\mathbf{Y}_{test}$  of the testing subjects can then be estimated by  $\mathbf{sign}\left(\mathbf{Z}_{\Delta\mathbf{Y}_{test}}^{opt}\right)$ , where  $\mathbf{Z}_{\Delta\mathbf{Y}_{test}}^{opt}$  denotes the optimal soft labels of the testing subjects. Based on the formulation of MTMC, we know that  $\mathbf{Z}_{\Delta\mathbf{Y}_{test}}^{opt}$  is implicitly obtained from  $\mathbf{Z}_{\Delta\mathbf{Y}_{test}}^{opt} = [\mathbf{X}_{test}^{opt}, \mathbf{1}] \mathbf{W}^{opt}$ , where  $\mathbf{X}_{test}^{opt}$  is the optimal noise-free counterpart of  $\mathbf{X}_{test}$ , and  $\mathbf{W}^{opt}$  is the optimal estimation of  $\mathbf{W}$ . Although  $\mathbf{W}^{opt}$  is not explicitly computed, it is implicitly determined by the training subjects and their known labels (i.e., in the third term of Eq. (1)). Therefore, for multi-label classification tasks with insufficient training subjects as in our case, MTMC will still have the inherent overfitting.

### 3.2 Multi-label Inductive Matrix Completion (MIMC)

In order to alleviate the overfitting, we employ an online inductive learning strategy to modify the MTMC model, and name the modified MTMC as Multi-label Inductive Matrix Completion (MIMC) model. Specifically, we introduce an explicit predictor matrix  $\tilde{\mathbf{W}} \in \mathbb{R}^{(d+1) \times t}$  into MTMC by adding the following constraint into Eq. (1):

$$\min_{\tilde{\mathbf{W}}} \lambda \|\tilde{\mathbf{W}}\|_{2,1} + \frac{\beta}{2} \|\mathbf{Z}_{\Delta\mathbf{Y}} - [\mathbf{X}^{obs}, \mathbf{1}] \tilde{\mathbf{W}}\|_F^2, \tag{2}$$

where  $\|\cdot\|_{2,1}$  denotes the  $L_{21}$ -norm, which imposes row sparsity on  $\tilde{\mathbf{W}}$  to learn the shared representations across all related classification tasks by selecting the common discriminative features. In addition, note also that, in the second term of Eq. (2), we use all the subjects (including the testing subjects) to learn the sparse predictor matrix  $\tilde{\mathbf{W}}$  based on the transductive soft labels  $\mathbf{Z}_{\Delta\mathbf{Y}}$ . In other words, we leverage the testing subjects as an efficient supplement to the limited training subjects, thus alleviating the small-sample-size issue of the training data that often causes the overfitting problem for training of the classifier. The final MIMC model is given as:

$$\min_{\substack{\mathbf{Z}, \tilde{\mathbf{W}} \\ \mathbf{Z}^{(d+1)} = \mathbf{1}}} \left\{ \mu \|\mathbf{Z}\|_* + \frac{1}{2} \|\mathbf{Z}_{\Delta\mathbf{X}} - \mathbf{X}^{obs}\|_F^2 + \gamma \sum_{(i,j) \in \Omega_Y} \mathbb{C}_y\left(\mathbf{z}_{i(d+1+j)}, \mathbf{y}_{ij}^{obs}\right) + \lambda \|\tilde{\mathbf{W}}\|_{2,1} + \frac{\beta}{2} \|\mathbf{Z}_{\Delta\mathbf{Y}} - [\mathbf{X}^{obs}, \mathbf{1}] \tilde{\mathbf{W}}\|_F^2 \right\}. \tag{3}$$

In this way, we can obtain the optimal sparse predictor matrix  $\tilde{\mathbf{W}}^{opt}$  by using our proposed optimization algorithm in Sect. 3.3 below, and estimate the labels  $\mathbf{Y}_{test}$  of the testing subjects  $\mathbf{X}_{test}$  by induction:

$$\mathbf{Y}_{test} = \mathbf{sign}\left([\mathbf{X}_{test}, \mathbf{1}]\tilde{\mathbf{W}}^{opt}\right). \quad (4)$$

Comparing with the overfitting-prone *transductive* labels  $\mathbf{sign}\left(\mathbf{Z}_{\Delta\mathbf{Y}_{test}}^{opt}\right)$ , the *inductive* labels in Eq. (4), which learned from more subjects (by including the testing subjects) and benefit from the advantage of joint feature selection (via  $L_{21}$ -norm), would give us more robust predictions, thus suffering less from the small-sample-size issue.

### 3.3 Optimization Algorithm for MIMC

The solution of MIMC is not trivial, as it contains the all-1-column constraint (i.e.,  $\mathbf{Z}^{(d+1)} = \mathbf{1}$ ) in Eq. (3), along with the fact that the  $L_{21}$ -norm and nuclear norm are the non-smooth penalties. Here, we employ the block coordinate descent method to design an optimization algorithm for solving MIMC. The key steps of this algorithm are to iteratively optimize the following two Subproblems:

$$\mathbf{Z}_k = \arg \min_{\mathbf{Z}^{(d+1)}=\mathbf{1}} \left\{ \begin{aligned} & \frac{1}{2} \|\mathbf{Z}_{\Delta\mathbf{X}} - \mathbf{X}^{obs}\|_F^2 + \gamma \sum_{(i,j) \in \Omega_{\mathbf{Y}}} \mathbb{C}_y\left(\mathbf{z}_{i(d+1+j)}, \mathbf{y}_{ij}^{obs}\right) \\ & + \mu \|\mathbf{Z}\|_* + \frac{\beta}{2} \|\mathbf{Z}_{\Delta\mathbf{Y}} - [\mathbf{X}^{obs}, \mathbf{1}]\tilde{\mathbf{W}}_{k-1}\|_F^2 \end{aligned} \right\}, \quad (5)$$

$$\tilde{\mathbf{W}}_k = \arg \min_{\tilde{\mathbf{W}}} \lambda \|\tilde{\mathbf{W}}\|_{2,1} + \frac{\beta}{2} \|(\mathbf{Z}_{\Delta\mathbf{Y}})_k - [\mathbf{X}^{obs}, \mathbf{1}]\tilde{\mathbf{W}}\|_F^2. \quad (6)$$

We solve Subproblem 1 in Eq. (5) by employing the Fixed Point Continuation (FPC) method plus the projection technique, with its convergence being proven by Cabral et al. [6]. Specifically, it consists of two steps for each iteration  $t$ :

$$\begin{cases} (\mathbf{Z}_k)_t = \mathcal{D}_{\mu\tau_{\mathbf{Z}}}\left((\mathbf{Z}_k)_{t-1} - \tau_{\mathbf{Z}}\nabla\mathbf{G}\left((\mathbf{Z}_k)_{t-1}\right)\right) \\ ((\mathbf{Z}_k)_t)^{(d+1)} = \mathbf{1} \end{cases}, \quad (7)$$

where  $\tau_{\mathbf{Z}} = \min\left\{1, 4/\sqrt{32\beta^2 + 2\gamma^2}\right\}$  denotes the gradient step size,  $\mathcal{D}_{\mu\tau_{\mathbf{Z}}}(\cdot)$  denotes the proximal operator of the nuclear norm [6], and  $\nabla\mathbf{G}(\mathbf{Z})$  is the gradient of  $\mathbf{G}(\mathbf{Z})$ :

$$\mathbf{G}(\mathbf{Z}) = \left\{ \begin{aligned} & \frac{1}{2} \|\mathbf{Z}_{\Delta\mathbf{X}} - \mathbf{X}^{obs}\|_F^2 + \frac{\beta}{2} \|\mathbf{Z}_{\Delta\mathbf{Y}} - [\mathbf{X}^{obs}, \mathbf{1}]\tilde{\mathbf{W}}_{k-1}\|_F^2 \\ & + \gamma \sum_{(i,j) \in \Omega_{\mathbf{Y}}} \mathbb{C}_y\left(\mathbf{z}_{i(d+1+j)}, \mathbf{y}_{ij}^{obs}\right) \end{aligned} \right\}. \quad (8)$$

The Subproblem 2 in Eq. (6) is a standard  $L_{21}$ -norm regularization problem, which can be solved via the accelerated Nesterov's method with convergence proof in [11]. Specifically, it includes the following step for each iteration  $t$ :

$$(\tilde{\mathbf{W}}_k)_t = \mathcal{J}_{\lambda\tau_{\tilde{\mathbf{W}}}}\left((\tilde{\mathbf{W}}_k)_{t-1} - \tau_{\tilde{\mathbf{W}}}\nabla\mathbf{F}\left((\tilde{\mathbf{W}}_k)_{t-1}\right)\right), \quad (9)$$

where  $\tau_{\tilde{\mathbf{W}}} = 1/\sigma_{\max}(\beta[\mathbf{X}^{obs}, \mathbf{1}]^T[\mathbf{X}^{obs}, \mathbf{1}])$  denotes the gradient step size,  $\sigma_{\max}(\cdot)$  denotes the maximal singular value of matrix,  $\mathcal{J}_{\lambda\tau_{\tilde{\mathbf{W}}}(\cdot)}$  denotes the proximal operator of  $L_{21}$ -norm [11], and  $\nabla F(\tilde{\mathbf{W}})$  is the gradient of  $F(\tilde{\mathbf{W}})$ :

$$F(\tilde{\mathbf{W}}) = \frac{\beta}{2} \|(\mathbf{Z}_{\Delta\mathbf{Y}})_k - [\mathbf{X}^{obs}, \mathbf{1}]\tilde{\mathbf{W}}\|_F^2. \quad (10)$$

Theoretically, for the jointly convex problem with the separable non-smooth terms, Tseng [12] has demonstrated that the block coordinate descent method is guaranteed to converge to a global optimum, as long as all Subproblems are solvable. In our MIMC model, obviously, the objective function in Eq. (3) is jointly convex for  $\mathbf{Z}$  and  $\tilde{\mathbf{W}}$ , and its non-smooth parts, i.e., both  $\mu\|\mathbf{Z}\|_*$  and  $\lambda\|\tilde{\mathbf{W}}\|_{2,1}$ , are separable. Based on this fact, our proposed optimization algorithm also has the provable convergence.

## 4 Results and Discussions

We evaluate the proposed MIMC by jointly predicting MGMT and IDH1 statuses using our HGG patients. Considering the limited number of 47 subjects, we use 10-fold cross validation to ensure a relatively unbiased prediction performance for the new testing subjects. We compare MIMC with the widely-used *single-task* machine learning methods (including SVM with RBF kernel [13] and RF [14]) and state-of-the-art *multi-task* machine learning methods (i.e., Lest\_L21 [11] and MTMC [5]). All the involved parameters in these methods are optimized by using the nested 10-fold cross validation procedure.

We measure the prediction performance in terms of accuracy (ACC), sensitivity (SEN), specificity (SPE), and area under the receiver operating characteristic curve (AUC). In order to avoid any bias introduced by randomly partitioning the dataset, each 10-fold cross-validation is independently repeated for 20 times. The average experimental results for MGMT and IDH1 status predictions are reported in Tables 2 and 3, respectively. The best results and those ones not significantly worse than the best results at 95% confidence level are highlighted in bold. Except that the Lest\_L21 achieves slightly higher specificity (but not statistically significant) than MIMC (i.e., 70.75% vs. 70.00%) in MGMT status prediction, MIMC consistently outperforms SVM, RF and MTMC in all performance metrics, which indicate that our proposed

**Table 2.** Performance comparison of different methods for MGMT status prediction.

Method		ACC (%)	AUC (%)	SEN (%)	SPE (%)
Single-task	SVM	68.04	72.43	71.35	63.75
	RF	65.54	70.62	62.50	69.50
Multi-task	Lest_L21	67.28	72.26	64.62	<b>70.75</b>
	MTMC	68.26	72.58	70.77	65.00
	<b>MIMC</b>	<b>71.74</b>	<b>77.21</b>	<b>73.08</b>	<b>70.00</b>

**Table 3.** Performance comparison of different methods for IDH1 status prediction.

Method		ACC (%)	AUC (%)	SEN (%)	SPE (%)
Single-task	SVM	75.54	82.47	61.92	80.91
	RF	72.50	77.94	68.46	74.09
Multi-task	Lest_L21	75.57	79.98	66.92	77.58
	MTMC	75.33	82.14	78.46	74.09
	<b>MIMC</b>	<b>83.26</b>	<b>88.47</b>	<b>84.62</b>	<b>82.73</b>

online inductive learning strategy can help improve the prediction performance of MIMC. In addition, we also find that all the *multi-task* machine learning methods consistently outperform the *single-task* RF method, but not outperform the *single-task* SVM method in terms of ACC. We speculate that this is mainly caused by the kernel trick of SVM, which implicitly carries out the nonlinear feature mapping. In future work, we will extend our proposed MIMC model to its nonlinear version by employing the kernel trick to further improve the performance of MGMT and IDH1 status predictions.

## 5 Conclusion

In this paper, we focus on addressing the tasks of predicting MGMT and IDH1 statuses for HGG patients. Considering strong correlation between MGMT promoter methylation and IDH1 mutation, we formulate their prediction tasks as a Multi-label *Inductive* Matrix Completion (MIMC) model, and then design an optimization algorithm with provable convergence to solve this model. The promising results by various experiments verify the advantages of the proposed MIMC model over the widely-used *single-* and *multi-task* classifiers. Also, for the first time, we show the feasibility of molecular biomarker prediction based on the preoperative multi-modality neuroimaging and connectomics analysis.

**Acknowledgments.** This work was supported in part by NIH grants (EB006733, EB008374, MH100217, MH108914, AG041721, AG049371, AG042599, AG053867, EB022880), Natural Science Foundation of Jiangsu Province (BK20161516, BK20151511), China Postdoctoral Science Foundation (2015M581794), Natural Science Research Project of Jiangsu University (15KJB520027), and Postdoctoral Science Foundation of Jiangsu Province (1501023C).

## References

1. Korfiatis, P., Kline, T., et al.: MRI texture features as biomarkers to predict MGMT methylation status in glioblastomas. *Med. Phys.* **43**(6), 2835–2844 (2016)
2. Yamashita, K., Hiwatashi, A., et al.: MR imaging-based analysis of glioblastoma multiform: estimation of IDH1 mutation status. *AJNI Am. J. Neuroradiol.* **37**(1), 58–65 (2016)
3. Zhang, B., Chang, K., et al.: Multimodal MRI features predict isocitrate dehydrogenase genotype in high-grade gliomas. *Neuro-oncology* **19**(1), 109–117 (2017)



4. Noushmehr, H., Weisenberger, D., et al.: Identification of a CpG island methylator phenotype that defines a distinct subgroup of glioma. *Cancer Cell* **17**(5), 510–522 (2010)
5. Goldberg, A., Zhu, X., et al.: Transduction with matrix completion: three birds with one stone. In: *Proceedings of NIPS*, pp. 757–765 (2010)
6. Cabral, R., et al.: Matrix completion for weakly-supervised multi-label image classification. *IEEE Trans. Pattern Anal. Mach. Intell.* **37**(1), 121–135 (2015)
7. Yan, C., Zang, Y.: DPARSF: a MATLAB toolbox for “pipeline” data analysis of resting-state fMRI. *Front. Syst. Neurosci.* **4**, 13 (2010)
8. Cui, Z., Zhong, S., et al.: PANDA: a pipeline toolbox for analyzing brain diffusion images. *Front. Hum. Neurosci.* **7**, 42 (2013)
9. Liu, L., Zhang, H., Rekić, I., Chen, X., Wang, Q., Shen, D.: Outcome prediction for patient with high-grade gliomas from brain functional and structural networks. In: Ourselin, S., Joskowicz, L., Sabuncu, M.R., Unal, G., Wells, W. (eds.) *MICCAI 2016*. LNCS, vol. 9901, pp. 26–34. Springer, Cham (2016). doi:[10.1007/978-3-319-46723-8\\_4](https://doi.org/10.1007/978-3-319-46723-8_4)
10. Wang, J., Wang, X., et al.: GRETNA: a graph theoretical network analysis toolbox for imaging connectomics. *Front. Hum. Neurosci.* **9**, 386 (2015)
11. Liu, J., Ji S., Ye, J.: Multi-task feature learning via efficient L2,1-norm minimization. In: *Proceedings of UAI*, pp. 339–348 (2009)
12. Tseng, P.: Convergence of a block coordinate descent method for non-differentiable minimization. *J. Optim. Theory Appl.* **109**(3), 475–494 (2001)
13. Cortes, C., Vapnik, V.: Support-vector networks. *Mach. Learn.* **20**(3), 273–297 (1995)
14. Breiman, L.: Random forests. *Mach. Learn.* **45**(1), 5–32 (2001)