Journal of Intelligent Information Systems
https://doi.org/10.1007/s10844-020-00606-9

Healthcare predictive analytics for disease progression: a longitudinal data fusion approach

Yi Zheng[1] · **Xiangpei Hu**[1]

Received: 1 November 2019 / Revised: 20 April 2020 / Accepted: 22 April 2020 /
Published online: 23 May 2020

Abstract
Healthcare predictive analytics using electronic health records (EHR) offers a promising direction to address the challenging tasks of health assessment. It is highly important to precisely predict the potential disease progression based on the knowledge in the EHR data for chronic disease care. In this paper, we utilize a novel longitudinal data fusion approach to model the disease progression for chronic disease care. Different from the conventional method using only initial or static clinical data to model the disease progression for current time prediction, we design a temporal regularization term to maintain the temporal successivity of data from different time points and simultaneously analyze data from data source level and feature level based on a sparse regularization regression approach. We examine our approach through extensive experiments on the medical data provided by the Alzheimer's Disease Neuroimaging Initiative (ADNI). The results show that the proposed approach is more useful to simulate and predict the disease progression compared with the existing methods.

Keywords Healthcare predictive analytics · Longitudinal data fusion · Machine learning · Regression · Group lasso

1 Introduction

Improvements in healthcare in the past century have contributed to people living longer and healthier lives. However, this has also resulted in an increase in the number of people with non-communicable diseases, including Alzheimer's disease. Alzheimer's disease (AD), the most common type of dementia, is characterized by the progressive impairment of neurons and their connections resulting in loss of cognitive function and ultimately death (Khachaturian 1985). According to the World Health Organization (2012), most new cases and

✉ Yi Zheng
zhengyi8807@163.com

Xiangpei Hu
drhxp@dlut.edu.cn

[1] Institute of Systems Engineering, Dalian University of Technology, Dalian, 116024, People's Republic of China

mortalities of Alzheimer's disease occur in low- and middle- income countries. Current estimates indicate 31.2 million people worldwide are living with AD in 2015, and this number will almost double every 20 years. AD has caused substantial social and economic burden to every country, the total estimated worldwide cost of AD in 2015 is $600 billion (Prince 2015). The surging cases and expenses make patients, clinical experts, and health policymakers around the world believe that effective interventions are needed to prevent, detect, and manage Alzheimer's disease and their sequelae (OECD 2014).

With increased adoption of electronic health record (EHR) systems in clinical practices, EHR data analytics for advanced clinical decision support is attracting both scientific and practical interest (Agarwal et al. 2010). Clinical intelligence about a patient's mental status has been a critical element for effective decision making in care for Alzheimer's disease. Accurate evaluation of patients' mental status could enable clinicians to take preventive and personalized interventions, which in turn could reduce healthcare spending and improve their quality of life. However, the richness of neuroimaging records, such as correlations among test result, their longitudinal progression, and their highly professional and highly personalized characteristics (Fichman et al. 2011), makes it a difficult task for healthcare professionals to provide an accurate evaluation of patient' mental status after comprehensive medical check-up is performed.

The value of predictive analytics in healthcare has been repeatedly emphasized in previous information systems research. Chen et al. (2012) discussed the potential of EHR-based healthcare analytics for "smart health and wellbeing" from the perspective of business intelligence. Fichman et al. (2011) summarized the existing healthcare information systems research and suggested that another emerging avenue for knowledge discovery arose from using digital technology to enable new kinds of mathematical healthcare modeling and simulations. Developing and utilizing information technology artifacts, such as models, techniques, and systems, to address practical needs has been a focus of healthcare information systems research. The research motivations are often to obtain valuable insights through the development of advanced analytics techniques and the use of large and rich data source that was previously unavailable or underutilized.

Consistently with the machine learning paradigm and the recent information systems research on big data analytics principle, we developed a novel longitudinal medical data fusion formulation for predicting the disease progression of Alzheimer's disease measured by the clinical scores. Our goal is to improve clinical decision making and facilitate preventive and personalized care with data analytics. Specifically, we formulate the prediction of the current clinical score as a regression problem based on previous time points of medical data fusion. The disease progression modeling approach could augment healthcare provider's capability in accurately evaluating patients' health status for timely interventions.

The proposed longitudinal medical data fusion formulation is distinctly different from the existing diseases predictive model. Existing healthcare predictive analytics research often either focuses on modeling patients' health status by making use of one single time point data (Duchesne et al. 2009; Stonnington et al. 2010), or modeling disease progression of different time point with one single time point data based on multitask learning (Zhou et al. 2011; Zhou et al. 2012). However, it is believed that, especially in chronic care, data from different time points can depict the evolution of the disease progression. Modeling disease progression via longitudinal medical data fusion would provide healthcare professionals with more significant clinical insights toward a comprehensive and effective care plan. We will effectively make use of the intermediate information during the disease progression, which includes the serial medical data. The proposed Disease Progression via Longitudinal Data Fusion (DPLDF) model will jointly analyze the features from the data source level and

feature level based on sparse regularization regression approach. The regularization consists of two components including an $l_{2,1}$-norm penalty (Yuan and Lin 2006) on the regression weight vectors, which ensures that a small subset of features will be selected for the regression models at all time points, and a temporal regularization, which ensures different time points satisfy the temporal successivity from data source level aspect. In our article, temporal successivity indicates the weight factors of features from adjacent time points have the sequential characteristic, and features of recent time point have larger weight factors than the weight factors of early time points. The challenge for the sparse regularization regression approach is that the proposed models do not have an analytical solution. We also develop a numerical optimization method to fit the model.

The main contributions of our work can be summarized as follows:

- Taking into consideration the prediction model at different time points shares a common set of features, we propose a group feature selection approach, which can remove redundant and irrelevant feature from the feature space, thus improving the prediction accuracy and reducing the computational cost in data fusion.
- Taking into consideration the intermediate information during the disease progression, we propose a temporal regularization in longitudinal data fusion model to predict patient's health state. The effectiveness of our proposed method is verified on Alzheimer's Disease Neuroimaging Initiative (ADNI) database, the results demonstrate that our proposed method can achieve good performances.

The rest of the paper is organized as follows. In the next section, we review related research on data fusion in the context of healthcare analytics and disease progression modeling approach. We then describe the proposed DPLDF model and an efficient numerical optimization method to fit the model. Following that, we conduct a set of experiments and summary their results. Our evaluation results provide evidence that the proposed DPLDF approach demonstrate superior performance compared with related methods. In the final section, we discuss the contributions of this study to the information systems knowledge base and directions for future work.

2 Literature review

In this section, we summarize the previous studies of the healthcare predictive analytics. Table 1 presents the previous studies of chronic disease intelligent diagnosis, which we discuss in turn.

2.1 Data fusion in the context of healthcare predictive analytics

Healthcare predictive analytics aims to predict future health-related outcomes or events based on clinical and nonclinical patterns in the data (Lin et al. 2017). The results of interest in healthcare predictive analytics, such as medical diagnosis (Valmarska et al. 2018; Liu et al. 2019), hospital readmissions (Li et al. 2016), and patient mortality (Mayaud et al. 2013), treatment responses (Meyer et al. 2014), are often of great practical importance. Many studies are currently collecting multiple types and multiple time points of medical data and information from the same participants. Each medical data analytic method reports on a limited domain and is likely to provide some standard information and some unique information, which motivates the need for a joint analysis of these data.

Table 1 Summary of typical previous research for chronic disease intelligent diagnosis

Studies	Data source type	Methods	Predicting outcome
Calhoun and Adali (2008)	fMRI, sMRI	Independent component analysis	Classification
Yuan et al. (2012)	MRI, PET, CSF	Multi-task learning and Group Lasso	Value
Chen et al. (2016)	Multiple time points of medical examination	Uncertain labels learning with feature-based representation	Multiobjective classification
Zhou et al. (2012)	One time point of MRI	Multi-task learning and Fused Lasso	Vector
Xie et al. (2016)	Two adjacent time points of MRI	Regression with Fused Lasso	Value

There is an increasing interest in the field of multiple types of data fusion. For instance, Calhoun and Adali (2008) presented a feature-based fusion approach that first preprocessed the data to compute features of interest. The features were then analyzed in a multivariate manner using independent component analysis. Finally, the linkage between the patterns of information from the individual's brain images and other biological measures was obtained. Yuan et al. (2012) proposed a multi-source feature learning framework for the joint analysis of incomplete multiple heterogeneous neuroimaging data. In their work, a feature-based fusion approach was used, samples were partitioned into multiple blocks based on combinations of data source available. A multi-task learning model was built, every learning task was trained by different multiple blocks, and tasks relatedness was considered to improve whole performance.

Very few previous studies of healthcare predictive analytics consider multiple time points of medical data fusion. Chen et al. (2016) converted data from multiple time points into a feature-based representation, i.e., by transforming a sequence into a vector of features. A time smoothing kernel was used to assign time weights to features at different time points to model the changes of importance over time. Finally, the health evaluation regression model was obtained from the transformed features, which leveraged data from multiple time points. Their work is different from ours in model design. We will see later in the model development section that our DPLDF approach does not have such a data transformation process, but instead uses a temporal regularization term to coordinate among data from different time points. As a result, our DPLDF method enables a more flexible and precise approach to advanced clinical decision support.

2.2 Disease progression modeling approach

Disease progression describes the change of disease status over time as a function of the disease process and treatment effects (Dubitzky et al. 2013). Different machine learning methods are used to characterize the linkage between disease status and medical data (Saggi and Jain 2018; Tai et al. 2019). For instance, Zhou et al. (2012) proposed a multi-task learning technique to predict disease progression. In the multi-task learning frame, the prediction of mental status at each time point was considered as a regression task, and each prediction task was based on baseline data. To improve the performance of the regression model, multiple prediction tasks of different time points were performed simultaneously to capture the temporal smoothness of the prediction models across different time points. Their method simultaneously selected a common set of biomarkers for all time points and picked a specific set of biomarkers at different time points, to identify the temporal patterns of biomarkers in disease progression. Only one time point of medical data is used in their disease progression model, but it is a common belief that by fusing different time points of medical data, one may use the evolution information of features in progress to provide more accurate information on health evaluation. Xie et al. (2016) followed a novel sequential learning framework to model the disease progression using data from two different time points. They designed a score-involved approach and made use of the sequential diagnosis information in different disease time points to jointly simulate the disease progression. Nie et al. (2016) proposed a novel and unified multitask learning scheme to coregularize the prior knowledge of source consistency and temporal smoothness.For the predicting task at each time point had features from multiple sources, and multiple tasks were related to each other in chronological order.

The existing disease progression modeling based on machine learning methods involve data from limited time points, which is not the case in chronic care. When the modeling problem involves data from many time points, how to depict the evolution characteristics of

data of different time points and maintain the temporal successivity of data from different time points has hardly ever discussed. Utilizing longitudinal medical data and developing disease progression modeling approach to predict patient's health status remains a research gap in the machine learning field.

3 Disease Progression via Longitudinal Data Fusion method

In this section, we will describe our proposed method. We will introduce the simple longitudinal data fusion model using ridge regression, and then explain the longitudinal clinical data fusion model with longitudinal regularization. As the proposed problem is numerically challenging, we will also present an efficient algorithm.

3.1 Preliminaries

In the longitudinal AD study, the target patients will receive regular MRI or PET scan in a fixed time interval, and their cognitive scores will be measured accordingly. The regression model simulates the relationship between the collected feature data and the target cognitive scores, so as to predict the patient's potential clinical score at the specific time point in future.

Consider a regression problem with n time points $\boldsymbol{x}_n, \boldsymbol{x}_{n-1}, \cdots, \boldsymbol{x}_1$, where every $\boldsymbol{x}_t \in R^d$ represents a set of clinical measures of time point t, and d represents the dimension of the data. There is a corresponding clinical score y_t $(t = n, n-1, \cdots, 1)$ measured at time point t. The target clinical score at the next time point $n+1$ is denoted as y.

In this paper, we employ linear models for the prediction. The cognitive score of future time point $n+1$ is predicted using the information of the previously recorded cognitive scores and the clinical measures. The motivation is that the existing clinical score can provide certain information about the current status of the patient, so it can facilitate to make more precise prediction for future progression. Specifically, we denote the feature data matrix by $\boldsymbol{X} = [\widetilde{\boldsymbol{x}}_n; \widetilde{\boldsymbol{x}}_{n-1}; \cdots, \widetilde{\boldsymbol{x}}_1]^T \in R^{n\times(d+1)}$. Here each $\widetilde{\boldsymbol{x}}_i = (\boldsymbol{x}_i, y_i) \in R^{d+1}$ is the extension of clinical feature data $\boldsymbol{x}_i$ with one more dimension by embedding the cognitive score y_i. We denote $\boldsymbol{W} = [\boldsymbol{w}_n; \boldsymbol{w}_{n-1}; \cdots, \boldsymbol{w}_1]^T \in R^{n\times(d+1)}$ as the weight matrix, and the prediction model for the future time point $n+1$ is given by $f(\boldsymbol{X}) = \sum_{t=1}^{n} \widetilde{\boldsymbol{x}}_t^T \boldsymbol{w}_t$. Suppose there are N samples, for sample i the feature matrix is $\boldsymbol{X}^i = [\widetilde{\boldsymbol{x}}_n^i; \widetilde{\boldsymbol{x}}_{n-1}^i; \cdots, \widetilde{\boldsymbol{x}}_1^i]^T$ and y^i is the corresponding clinical score for the future time point $n+1$. A straightforward approach is to estimate $\boldsymbol{W}$ by minimizing the following objective function:

$$\min_{\boldsymbol{W}} \sum_{i=1}^{N} \Big(\sum_{t=1}^{n} \widetilde{\boldsymbol{x}}_t^{i\,T} \boldsymbol{w}_t - y^i \Big)^2 + \theta \|\boldsymbol{W}\|_F^2 \tag{1}$$

where the first term measures the empirical error on the training data, the second penalty term controls the generalization error, $\theta > 0$ is a regularization parameter, and $\|\boldsymbol{W}\|_F = (\sum_{i=1}^{n} \sum_{j=1}^{d+1} w_{ij}^2)^{1/2}$.

One major limitation of the regression model above is that the model treats samples from different time points all the same and the temporal correlation is not involved, which is crucial for improving accuracy of longitudinal data analysis in chronic disease care.

3.2 Longitudinal regulation data fusion model building

To capture the temporal successivity of data from different time points, we propose a temporal regularization term in a regression model that penalize large time weights of early time points, which will ensure data from recent time points will get greater time weights than data from forwarding time points, resulting in the following formulation:

$$\min_{\boldsymbol{W}} \sum_{i=1}^{N} \Big(\sum_{t=1}^{n} \widetilde{\boldsymbol{x}}_t^{i\,T} \boldsymbol{w}_t - y^i \Big)^2 + \lambda \sum_{t=1}^{n} \|\boldsymbol{W}(1:t, 1:d)\|_F \tag{2}$$

where $\lambda > 0$ is a regularization parameter controlling the time weights, and $\boldsymbol{W}(1:t, 1:d)$ is the submatrix of $\boldsymbol{W}$ with $t \times d$ dimensions.

Because of the limited availability of subjects in the longitudinal AD study and a relatively large number of features from structural neuroimaging data, the prediction model (2) suffers from the so-called "curse of dimensionality". One practical approach is to reduce the dimensionality of the data. However, traditional dimension reduction techniques such as PCA, SVD, and M-CCA are not desirable since the resulting model is not interpretable, and regular feature selection algorithms are not suitable for longitudinal feature selection. In the proposed formulation, we employ group Lasso regularization based on $l_{2,1}$-norm penalty for feature selection (Yuan and Lin 2006), which assumes that a small set of features are predictive of the progression. The group Lasso regularization ensures that the regression model at different time points shares a common set of features. Together with the temporal penalty, we obtain the Disease Progression via Longitudinal Data Fusion (DPLDF) model:

$$\min_{\boldsymbol{W}} \sum_{i=1}^{N} \Big(\sum_{t=1}^{n} \widetilde{\boldsymbol{x}}_t^{i\,T} \boldsymbol{w}_t - y^i \Big)^2 + \lambda \sum_{t=1}^{n} \|\boldsymbol{W}(1:t, 1:d)\|_F + \mu \sum_{j=1}^{d} \|\boldsymbol{W}(:, j)\|_F \tag{3}$$

where $\boldsymbol{W}(:, j)$ is the j-column of $\boldsymbol{W}$, and μ is a regularization parameter. The weights of one feature over all time points are grouped using the l_2-norm, and all feature groups are penalized using the l_1-norm. Thus, the $l_{2,1}$-norm penalty tends to select features based on the strength of the feature over all time points, as shown in Fig. 1. The inputs of the method are longitudinal clinical data of a population linked to the geriatric medical examination database. The output is a real predicted value, reflecting personal health status. A linear prediction model is built, and the coefficients of some indicators corresponding to the clinical values at different time points are all 0, which makes the method select a common subset of features in the process of longitudinal clinical data fusion, thus improving the accuracy of health status prediction. Note that the method is designed for longitudinal clinical data sets that share the characteristics described in Section 2.2. Although in the following discussions we will use the ADNI data set as an example, the applicability of our proposed method is not limited to the data set.

3.3 Proximal gradient method for longitudinal regulation data fusion model

When the structure to be imposed in the penalty term has a relatively simple form, such as nonoverlapping groups over variables (e.g., lasso (Tibshirani 1996) or group lasso (Yuan and Lin 2006)), efficient optimization methods have been developed. For example, under group lasso, due to the separability among groups, a proximal operator associated with the penalty can be computed in closed-form; thus, some composite gradient methods (Beck and

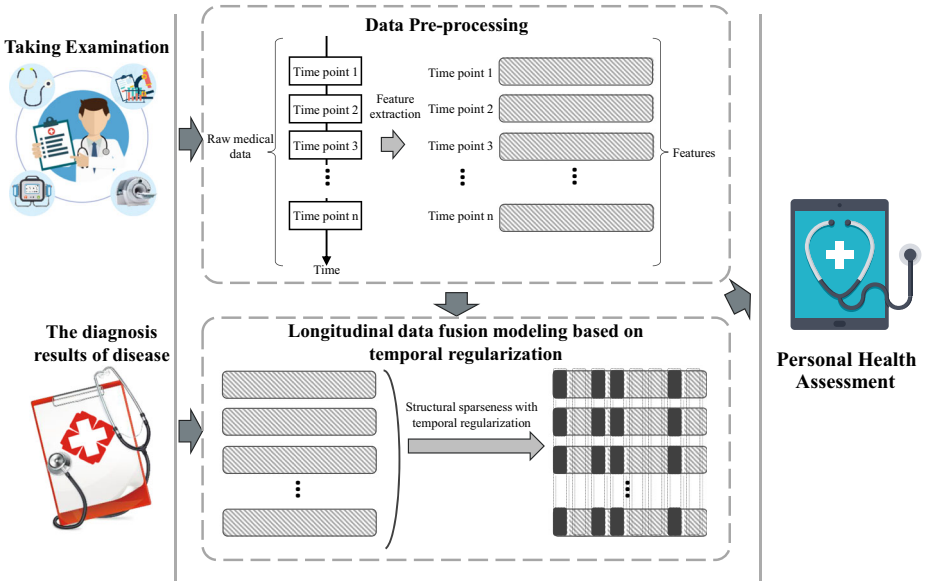

Fig. 1 Feature selection diagram of the proposed Disease Progression via Longitudinal Data Fusion method. We use longitudinal regulation to maintain the temporal successivity of data from different time points and simultaneously select a common subset of features (the selected features for all time points are highlighted)

Teboulle 2009) that leverage the proximal operator as a critical step can be directly applied. Unfortunately, these algorithmic advancements have been outpaced by the emergence of more complex structures one would like to impose in the penalty term.

The penalty terms of DPLDF model (3) are the composition of a temporal penalty and a group Lasso penalty. Since the group of features in the penalty terms with overlap structure, our model belongs to the overlapping group Lasso penalized problem. In this paper, we propose to solve it using the accelerated gradient descent method (Nesterov 2013a) because of its fast convergence rate.

We develop an efficient algorithm for the DPLDF model (3) via the accelerated gradient descent (AGD) method. The AGD method has the optimal order of convergence for the first-order black-box methods, which can achieve a convergence rate of $O(1/k^2)$ for k iterations. Note that, when directly applying the black-box first-order method for solving the non-smooth problem (3), one can only achieve a convergence rate of $O(1/\sqrt{k})$ (Nesterov 2013b), much slower than $O(1/k^2)$.

Equation (3) can be transformed into the following form:

$$\min_{\boldsymbol{W}} f(\boldsymbol{W}) = l(\boldsymbol{W}) + \sum_{i=1}^{n+d} \theta_i \|\boldsymbol{W}_{G_i}\|_F \tag{4}$$

where $l(\boldsymbol{W}) = \sum_{i=1}^{N} \left(\sum_{t=1}^{n} \widetilde{\boldsymbol{x}}_t^{i\,T} \boldsymbol{w}_t - y^i \right)^2$ is a smooth convex loss function, $\theta_i = \lambda, i = 1, \cdots, n$, $\theta_i = \mu, i = n+1, \cdots, n+d$. $\boldsymbol{W}_{G_i} = \boldsymbol{W}(1:t, 1:d), i = 1, \cdots, n$, $\boldsymbol{W}_{G_i} = \boldsymbol{W}(:, i-n), i = n+1, \cdots, n+d$, that is G_i contains the indices corresponding to the i-th group of features.

We construct a model for approximating $f(\cdot)$ at the point $\boldsymbol{W}$ as:

$$f_{L,\boldsymbol{W}}(\boldsymbol{U}) = l(\boldsymbol{W}) + \sum_{t=1}^{n}\left\langle \partial l(\boldsymbol{W})/\partial \boldsymbol{w}_t, \boldsymbol{u}_t - \boldsymbol{w}_t \right\rangle + \sum_{i=1}^{n+d}\theta_i \|\boldsymbol{U}_{G_i}\|_F + \frac{L}{2}\|\boldsymbol{U} - \boldsymbol{W}\|_F^2 \quad (5)$$

where $\boldsymbol{U} = [\boldsymbol{u}_n; \boldsymbol{u}_{n-1}; \cdots; \boldsymbol{u}_1]^T \in R^{n\times(d+1)}$. In the i-step, a search point $\boldsymbol{S}_i$ is computed based on the past solutions of the previous step by $\boldsymbol{S}_i = \boldsymbol{W}_i + \beta_i(\boldsymbol{W}_i - \boldsymbol{W}_{i-1})$. Then the new solution $\boldsymbol{W}_{i+1}$ is obtained via the minimization of the model at the current search point, that is, $\boldsymbol{W}_{i+1} = \arg\min_{\boldsymbol{U}} f_{L,\boldsymbol{S}_i}(\boldsymbol{U})$. This sub-problem is the critical component to the optimization, and we will give a detailed discussion of how to solve this sub-problem efficiently.

Denote the proximal operator associated with the overlapping group Lasso penalty as follows:

$$\pi_{\boldsymbol{\theta}}(\boldsymbol{U}) = \arg\min_{\boldsymbol{W}\in R^{n\times(d+1)}}\left\{ g(\boldsymbol{W}) = \frac{1}{2}\|\boldsymbol{U} - \boldsymbol{W}\|_F^2 + \sum_{i=1}^{n+d}\theta_i \|\boldsymbol{W}_{G_i}\|_F \right\} \quad (6)$$

where $\boldsymbol{\theta} = [\theta_1, \cdots, \theta_{n+d}]^T \in R^{n+d}$, which is a special case of (4).

We first reveal the relationship between the optimal solution of formula (5) and proximal operator (6).

Theorem 1 *Denote*

$$\boldsymbol{V} = \arg\min_{\boldsymbol{U}} f_{L,\boldsymbol{W}}(\boldsymbol{U}) \quad (7)$$

Then we have $\boldsymbol{V} = \pi_{\boldsymbol{\theta}/L}\big(\boldsymbol{U} - \frac{1}{L}\frac{\partial l(\boldsymbol{U})}{\partial \boldsymbol{U}}\big)$.

Proof

$$\begin{aligned}
&\pi_{\boldsymbol{\theta}/L}\left(\boldsymbol{U} - \frac{1}{L}\frac{\partial l(\boldsymbol{U})}{\partial \boldsymbol{U}}\right)\\
&= \arg\min_{\boldsymbol{W}\in R^{n\times d}}\left\{\frac{1}{2}\left\|\boldsymbol{U} - \boldsymbol{W} + \frac{1}{L}\frac{\partial l(\boldsymbol{U})}{\partial \boldsymbol{U}}\right\|_F^2 + \sum_{i=1}^{n\times(d+1)}\frac{\theta_i}{L}\|\boldsymbol{W}_{G_i}\|_F\right\}\\
&= \arg\min_{\boldsymbol{W}\in R^{n\times d}}\left\{\frac{1}{2}\|\boldsymbol{W} - \boldsymbol{U}\|_F^2 + \frac{1}{L}\sum_{t=1}^{n}\langle \partial l(\boldsymbol{U})/\partial \boldsymbol{u}_t, \boldsymbol{w}_t - \boldsymbol{u}_t\rangle\right.\\
&\quad \left. + \sum_{i=1}^{n\times(d+1)}\frac{\theta_i}{L}\|\boldsymbol{W}_{G_i}\|_F + \frac{1}{L^2}\left\|\frac{\partial l(\boldsymbol{U})}{\partial \boldsymbol{U}}\right\|_F^2\right\}\\
&= \arg\min_{\boldsymbol{W}\in R^{n\times d}}\left\{\frac{1}{2}\|\boldsymbol{W} - \boldsymbol{U}\|_F^2 + \frac{1}{L}\sum_{t=1}^{n}\langle \partial l(\boldsymbol{U})/\partial \boldsymbol{u}_t, \boldsymbol{w}_t - \boldsymbol{u}_t\rangle\right.\\
&\quad \left. + \sum_{i=1}^{n\times(d+1)}\frac{\theta_i}{L}\|\boldsymbol{W}_{G_i}\|_F\right\}\\
&= \arg\min_{\boldsymbol{W}\in R^{n\times d}}\left\{\sum_{t=1}^{n}\langle \partial l(\boldsymbol{U})/\partial \boldsymbol{u}_t, \boldsymbol{w}_t - \boldsymbol{u}_t\rangle + \sum_{i=1}^{n\times(d+1)}\theta_i\|\boldsymbol{W}_{G_i}\|_F\right.\\
&\quad \left. + \frac{L}{2}\|\boldsymbol{W} - \boldsymbol{U}\|_F^2\right\}
\end{aligned}$$

From here we come to our conclusions. □

As shown in Theorem 1 the optimal solution of formula (5) is the same as $\pi_{\boldsymbol{\theta}/L}\big(\boldsymbol{U} - \frac{1}{L}\frac{\partial l(\boldsymbol{U})}{\partial \boldsymbol{U}}\big)$. We use the results of Yuan et al. (2013) to solve (6). Algorithm 1 shows the full procedure for solving the DPLDF model.

Algorithm 1 The optimization algorithm.

Input: $L_0 > 0$, $\boldsymbol{W}_0$
Output: $\boldsymbol{W}$
Initialize $i = 1$, $\boldsymbol{W}_1 = \boldsymbol{W}_0$, $\alpha_{-1} = 0$, $\alpha_0 = 1$ and $L = L_0$
Repeat
Set $\beta_i = \frac{\alpha_{i-2}-1}{\alpha_{i-1}}$, $\boldsymbol{S}_i = \boldsymbol{W}_i + \beta_i(\boldsymbol{W}_i - \boldsymbol{W}_{i-1})$
Find the smallest $L = 2^j L_{i-1}$, $j = 0, 1, \cdots$ such that
$f(\boldsymbol{W}_{i+1}) \leq f_{L,\boldsymbol{S}_i}(\boldsymbol{W}_{i+1})$ holds, where $\boldsymbol{W}_{i+1} = \pi_{\boldsymbol{\theta}/L}\big(\boldsymbol{S}_i - \frac{1}{L}\frac{\partial l(\boldsymbol{S}_i)}{\partial \boldsymbol{S}_i}\big)$
Set $L_i = L$, $\alpha_{i+1} = \frac{1+\sqrt{1+4\alpha_i}}{2}$, $\boldsymbol{W} = \boldsymbol{W}_{i+1}$, $i = i + 1$
until $|f(\boldsymbol{W}_{i+1}) - f(\boldsymbol{W}_i)| \leq TOLERANCE * |f(\boldsymbol{W}_i)|$

The overall procedure for model development and evaluation is shown in Fig. 2. First, we extract features from medical data and then conduct feature standardization to assimilate clinical measurements of diverse scales. Accordingly, all features are rescaled so that they have the properties of a standard normal distribution with a mean of 0 and a standard deviation of 1. The full dataset is then split into a model development set (90%) and a testing set (10%) that is used for evaluating and comparing performances of competing models. The model development set is further split into training and validation sets. The training data was used to predict the responses for the observations in the validation set. This provided us with an unbiased evaluation of a model fit on the training dataset while tuning the hyperparameters of the model. For the validation procedure, we use the 5-fold cross-validation. The final model evaluation is conducted on a held-out testing set that has not been used prior, either for training the model or tuning the model's parameters.

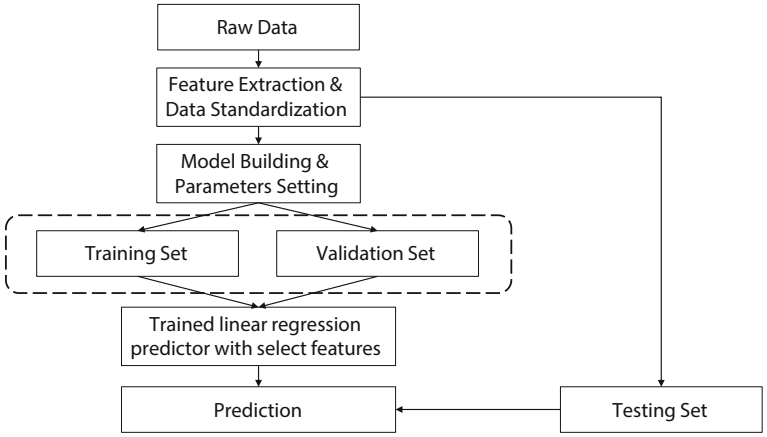

Fig. 2 Overview of the model development and validation procedure

4 Experimental study

In this section, we evaluate the proposed disease progression model on the data sets from the Alzheimer's Disease Neuroimaging Initiative (ADNI). The ADNI project is designed to collect the serial of MRI, PET and other clinical assessment scores to measure the progression of selected subjects, including Alzheimer's Disease patients (AD), Mild Cognitive Impairment patients (MCI) and Normal Controls (NC), and the subjects will be observed repeatedly and continuously over a 6-month or 1-year interval. For each observation, the MRI and PET scans will be collected, as well as other corresponding measurements, e.g., clinical scores such as MMSE. The mini-mental state examination (MMSE) provides a quick assessment of a patient's cognitive state. This test allows a healthcare provider to objectively assess a patient who may have cognitive impairments to determine their severity. Declining scores on the MMSE can be a sign that a patient is having neurological problems. These questions of MMSE determine the patient's level of orientation, both physically and mentally, and also assess memory and math skills. The healthcare provider can assign a score on the basis of one to 30, which will determine the patient's level of cognitive impairment. Scores of 25 or higher indicate the patient appears to be functioning well, without any problems. If the score falls between 20 and 24, it may indicate a mild level of cognitive impairment, while scores between 10 and 20 are considered moderate. Anything lower than 9 indicates severe impairment.

In our work, we use MRI scans to generate feature data, which are obtained from 126 subjects. Five types of MRI features are used in the work: white matter parcellation volume (Vol. WM.), cortical parcellation volume (Vol. C.), surface area (Surf. Area), cortical thickness average (CTA) and cortical thickness standard deviation (CTStd). The date when the patient performs the MRI screening in the hospital for the first time is called baseline (BL), and the time points of the following observations are denoted by the duration starting from the baseline. For example, "M06" means the screening taken at the time point 6 months after the first visit. All the subjects are under the repeated observations for up to 36 months. For each subject the sequence data is as follows:

$$\{M24, M18, M12, M06, BL\}.$$

In our experiments, the sequence data involves five time points of data. We predict MMSE scores of the time point 36 months after the first visit using various measurements from MRI scans. The sample size and dimensionality for features from the MRI scans are given in Table 2.

In our predictive regression test scenarios, 5-fold cross-validation is used to select model parameters for the sample data. We use 90% of the sample data for training and report the regression performance on the remaining test data. To measure the regression performance, we employ the mean squared error (MSE) calculated for the predicted values at the "M36" time point, and we also use correlation coefficient (R-value) given by the correlation between the predicted values and the true values (Duchesne et al. 2009; Zhou et al. 2012; Xie et al. 2016).

Table 2 The sample size and feature dimensionality of different time points used in the experiments

Target score	Sample size	Dimension of features from different time points
MMSE	126	346

We conduct five sets of evaluation to assess the proposed DPLDF approach. In the first set of evaluations, we aim to understand the utility of data fusion. We compare the DPLDF approach with basic data analysis counterparts-linear regression with lasso regularization. In the second set of evaluations, we aim to understand the performance of our DPLDF approach against other healthcare predictive analytics approaches in the literature. We hence compare the proposed method with Convex Fused Spares Group Lasso (cFSGL) based on multi-task learning (Zhou et al. 2012), regression model with transformed features based on point-based representation (Chen et al. 2016), incrementally sequential prediction (ISP) model based on sequential data learning (Xie et al. 2016). In the third set of evaluations, we compare the prediction errors of different approaches for patients with different conditions, so as to indicate that our proposed DPLDF has higher forecasting accuracy in predicting the condition of patients with cognitive scores below 20. In the last two sets of evaluations, we investigate the effects of parameters in the proposed DPLDF.

4.1 Comparison with linear regression with lasso regularization

In the first set of experiments, we compare DPLDF approach and linear regression with lasso regularization models to examine the utility of longitudinal data fusion over learning by linear regression.

We apply our proposed DPLDF method to the longitudinal data set including all five time points ranging from "M24" to "BL" for solving the problem of predicting the MMSE scores of time point "M36". For comparison purposes, we build linear regression with lasso regularization models to predict the MMSE scores of time point "M36" via the analysis of the same data sets. The third approach in comparison is the linear regression with lasso regularization models via the analysis of the data of time point "M24". For our proposed DPLDF method, six values ($2^1, 2^2, 2^3, 2^4, 2^5, 2^6$) are used for the regularization parameter λ, and six values ($2^{-5}, 2^{-4}, 2^{-3}, 2^{-2}, 2^{-1}, 2^0$) are used for the regularization parameter μ. For each linear regression with lasso regularization models, six values ($2^1, 2^2, 2^3, 2^4, 2^5, 2^6$) are used for the regularization parameter λ. 5-fold cross-validation is used to select model parameters, and the predictive performance is quantified by mean squared error (MSE). We first randomly select 90% portion of samples as the training set to learn the model, and then apply the model to predict the MMSE scores of time point "M36" on the remaining data, used as a non-overlapping test set. We repeat this process 30 times, the average performances are summarized in Table 3.

The average mean squared errors of DPLDF, linear regression with lasso regularization models via analysis of all five time points and linear regression with lasso regularization models via analysis of "M24" time point are 3.212, 5.569, 4.316. The performance difference between two linear regression with lasso regularization models is likely due to more

Table 3 Comparison of our proposed DPLDF method with linear regression with lasso regularization models on MMSE scores prediction, in terms of average correlation coefficient (R) and mean squared error (MSE) for time point "M36". 90% portion of data is used as training data

Comparison item	DPLDF	Linear regression-Lasso via analysis of all five time points	Linear regression-Lasso via analysis of "M24" time point
Target: MMSE			
R	0.786±0.041	0.694±0.037	0.718±0.031
M36 MSE	3.212±0.941	5.569±1.365	4.316±1.091

features are involved in the model to make it overfitting (the empirical error measured by the mean squared errors of linear regression with lasso regularization models via analysis of all five time points and linear regression with lasso regularization models via analysis of "M24" time point are 3.694 and 4.130). We notice a greater performance difference between DPLDF and linear regression with lasso regularization models. This greater performance improvement from DPLDF is because in our proposed model temporal successivity of data from different time points is considered, which can better capture the evolution information of the disease. Overall, the result from evaluation 1 confirms our speculations that the utility of longitudinal data fusion can effectively exploit the progressive nature of the disease to improve predictive performance.

4.2 Comparison of DPLDF approach and other healthcare predictive analytics approaches

One of the strengths of the proposed formulation is that it facilitates the temporal regularization, which ensures data from different time points satisfy the temporal successivity from data source level aspect when using longitudinal data to model disease progression. To determine the standing of our DPLDF approach among other healthcare predictive analytics approaches, we conduct a head-to-head comparison of predictive performance with the convex fused sparse group Lasso (cFSGL) approach (Zhou et al. 2012), the incrementally sequential prediction (ISP) approach (Xie et al. 2016), and the feature-based approach (Chen et al. 2016). We apply these four methods to predict the MMSE scores of time point "M36" via analysis of data from all five time points ranging from "M24" to "BL". All other three of these alternative healthcare predictive analytics approaches require user-specified parameters. For cFSGL, we need to specify the weights for Lasso penalty, group Lasso penalty and fused Lasso penalty. For ISP, we need to specify the weights for Lasso penalty, group Lasso penalty and fused Lasso penalty. For feature-based approach, sequences were converted into a point-based representation based on time smoothing kernels; we establish linear regression-Lasso model by using point-based representation data to predict the MMSE scores of time point "M36". The weight for Lasso penalty of feature-based approach also needs to be determined. We identify the best parameter settings for these approaches through cross-validation before we conduct evaluation 2. We repeat this evaluation 30 times, the average performances are summarized in Table 4.

The average mean squared errors of DPLDF, cFSGL, ISP, Feature-based approach are 3.212, 4.383, 4.029 and 4.613 respectively. The prediction error of Feature-based approach is higher than the other three healthcare predictive analytics approaches because the weights of data sets form different time points are assigned by a given function which may not capture time correlation of different data sets. Due to the shared representation in parallel

Table 4 Comparison of our proposed DPLDF method and other healthcare predictive analytics approaches (cFSGL, ISP, Feature-based approach) on MMSE scores prediction, in terms of average correlation coefficient (R) and mean squared error (MSE) for time point "M36". 90% portion of data is used as training data

Comparison item	DPLDF	cFSGL	ISP	Feature-based approach
Target: MMSE				
R	0.786±0.041	0.713±0.036	0.691±0.039	0.702±0.032
M36 MSE	3.212±0.941	4.383±1.143	4.029±0.993	4.613±1.016

learning of the cFSGL approach, which utilizes the baseline clinical measures data and the cognitive scores of different time point, this can improve individual task prediction accuracy. The ISP approach can further improve the performance due to the feedback from the intermediate information. The result from evaluation 2 suggests that the proposed DPLDF method achieve the lowest prediction error because temporal successivity of data from deferent time points is modeled by the temporal penalty. More specifically, the DPLDF method can better depict the evolution characteristics of data from different time points in longitudinal data fusion, which in turn improve the prediction accuracy. We will provide further error analysis among different predictive analytics approaches in the following part.

4.3 Effect of patients' status on prediction performance

Based on the overall prediction comparison, we further provide a detailed error analysis to discover the effect of patients' status on prediction performance. For the prediction of MMSE scores of time point "M36", we divide the patients into different ranges (AD, MCI and NC) based on their actual clinical scores, and compare the prediction errors of different approaches for patients with different conditions, so as to examine the effect of the patients' actual scores on prediction performance.

Figure 3 shows the error analysis for MMSE prediction, and the mean squared error of cognitive scores prediction of patients with different conditions at time point "M36" are summarized. From the figure, we can conclude that the major prediction error is produced when predicting those patients with low cognitive measurement (For MMSE score, [0, 20] means AD patients). For patients with cognitive scores below 20, the cognitive scores of time point "M36" may vary greatly from time point "M24" and the prediction error of different healthcare predictive analytics approaches differs greatly. DPLDF, ISP and Linear regression-Lasso via analysis of the last time point are the three best methods with the lowest prediction error. Due to imprecise depiction of the evolution characteristics of data of different time points, the prediction accuracy of Feature-based approach and Linear

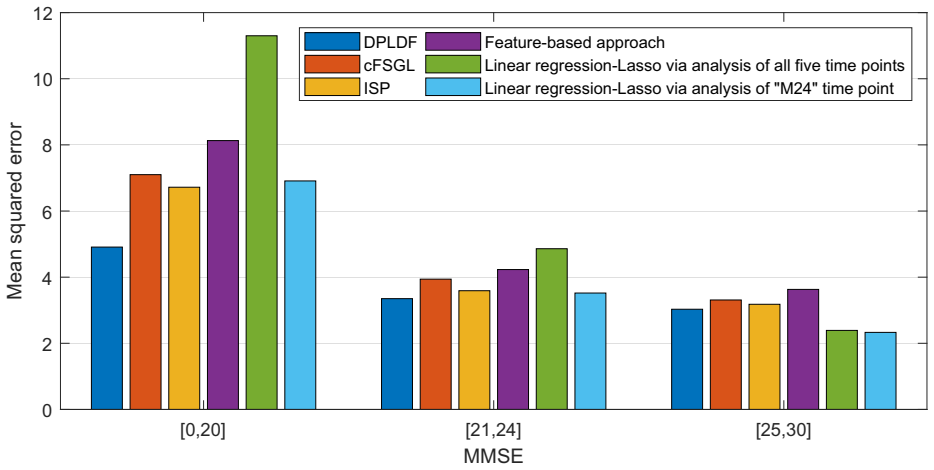

Fig. 3 Comparison of our proposed DPLDF method and other healthcare predictive analytics approaches (cFSGL, ISP, Feature-based approach, linear regression with lasso regularization) on MMSE scores prediction, in terms of mean squared error (MSE) for time point "M36" summarized according to patients' status. 90% portion of data is used as training data

regression-Lasso via analysis of all previous time points is lower than Linear regression-Lasso via analysis of the last time point. However, our proposed DPLDF method has the advantage of synthesizing information of data from different time points and characterizing the trend of features' dynamic change, which in turn makes our method has the highest accuracy for prediction of patient's potential clinical score at the specific time point in future. For those with higher cognitive scores (MCI or Normal Control), since their status is relatively stable, the prediction error is significantly decreased and the prediction accuracy of different healthcare predictive analytics approaches is close to each other.

Taken together, among the existing healthcare predictive analytics approaches, our proposed DPLDF can achieve better performance for predicting cognitive scores of patients with different conditions.

4.4 Effects of different λ in the proposed DPLDF

The purpose of this experiment is to discuss how the proportion of selected features and predictive performance vary when different λ values are chosen in our proposed DPLDF method. In the MMSE score predictions for time point "M36" via analysis of data from all five time points ranging from "M24" to "BL", we report the proportions of selected features and predictive performances when we use different choices of λ values (2^1, 2^2, 2^3, 2^4, 2^5, 2^6) fixing the value of temporal regularization parameter μ. We first randomly select 90% portion of samples as the training set to learn the model and then apply the model to predict the MMSE scores of time point "M36" on the remaining data, used as a non-overlapping test set. We repeat this process 30 times, for different choices of λ values the average proportions of selected features and predictive performances of DPLDF are shown in Fig. 4.

We can observe from the figure, as we increase λ fixing the value of temporal regularization parameter μ, the number of features selected gradually decrease, from about 94% of the features to 18%, meanwhile the predictive performance increase first and then decrease. Parameter λ is considered as a tradeoff factor between sparsity and predictive performance, and we can also observe that the best choice of λ lies in the middle of the region (in this example 2^4), which achieve a good balance between sparsity and predictive performance.

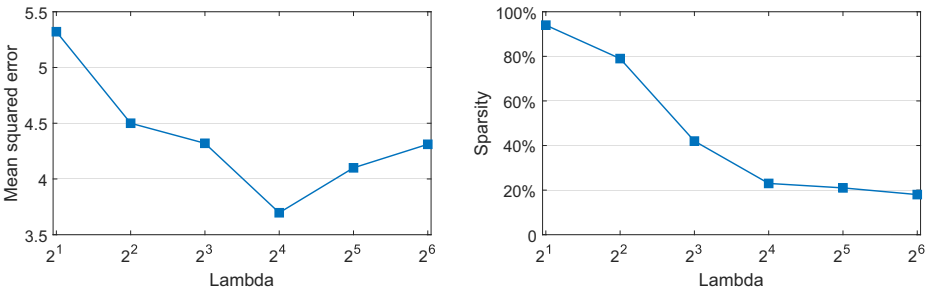

Fig. 4 Illustration of the results obtained using different λ in our proposed DPLDF method. The MMSE score prediction for time point "M36" via analysis of data from all five time points ranging from "M24" to "BL" problems are used, and the average performances (mean squared error) are reported. We vary the λ values from 2^1 to 2^6 (x-axis) and report the performances obtained (y-axis) in the left figure. In the right figure, we report the proportions of selected features (sparsity, y-axis) when we increase λ from 2^1 to 2^6 (x-axis)

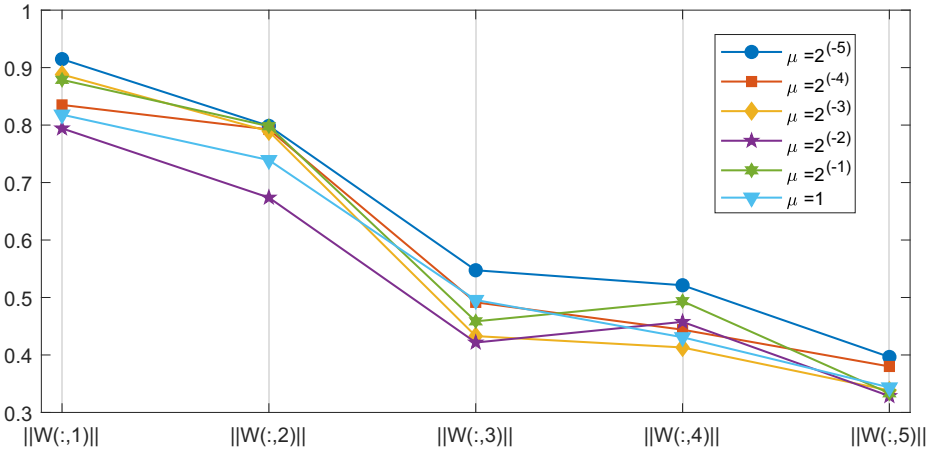

Fig. 5 Illustration of the l_2-norms change of the weight vectors corresponding to data vectors from different time points using different μ in our proposed DPLDF method. We vary the μ values from 2^{-5} to 1 (different polyline) and report the l_2-norms change (y-axis) of the weight vectors corresponding to data vectors from different time points (x-axis) using different μ

4.5 Effects of different temporal regularization parameter μ in the proposed DPLDF

The purpose of this experiment is to discuss how the weight values change when different temporal regularization parameters μ are chosen in our proposed DPLDF method. In the MMSE score predictions for time point "M36" via analysis of data from all five time points ranging from "M24" to "BL", we report the l_2-norm changes of weight vectors of the proposed DPLDF corresponding to data vectors from different time points when we use different choices of μ values (2^{-5}, 2^{-4}, 2^{-3}, 2^{-2}, 2^{-1}, 2^{0}) fixing the value of parameter λ. We first randomly select 90% portion of samples as the training set to learn the model and then apply the model to predict the MMSE scores of time point "M36" on the remaining data, used as a non-overlapping test set. For different choices of μ values, the l_2-norms of the weight vectors corresponding to data vectors from different time points are illustrated in Fig. 5. The predictive performances of different choices of μ values are also listed in Table 5.

It can be observed from the figure, for the MMSE score prediction for time point "M36" problems, that the polylines possess a downward trend and the l_2-norm of the weight vector ($W(:, 1)$) corresponding to data vector of current time point ("M24") has the maximum value. We can observe from Table 5 that the best choice of μ is 2^{-3}, which achieve a good predictive performance. When μ equals 2^{-3}, the corresponding polyline has the temporal successively characteristics, which in our article indicates the feature vectors of recent time

Table 5 The predictive performances corresponding to different choices of μ values

μ	2^{-5}	2^{-4}	2^{-3}	2^{-2}	2^{-1}	2^{0}
MSE	3.992	4.104	3.212	3.687	3.923	4.615

points have larger weight vector norms than that of early time points. This makes our proposed DPLDF method could use longitudinal medical data and guarantees that data from recent time points have greater weights, thus ensureing the high predictive performance.

5 Conclusions

Our study makes several research contributions. First, we propose a Disease Progression via Longitudinal Data Fusion (DPLDF) formulation for modeling disease progression. DPLDF allows the simultaneous analysis of features form data source level and feature level based on a sparse regularization regression approach. Based on the linear regression model, we introduce the temporal regularization term to maintain the temporal successivity of data from different time points. Due to the proposed model belongs to a class of overlap group Lasso model, we employ accelerated gradient descent method to solve the formulation. Second, we evaluate the proposed approach with real-world EHR data. We obtain empirical evidence that introducing the evolution characteristics of medical data of different time points in modeling disease progression improves predictive performance. That is, a longitudinal data fusion framework can indeed offer better clinical insights than other disease progression modeling methods. Finally, we recognize that sequence data length has an impact on effeteness of disease progression modeling. Our evaluation result further suggests that DPLDF outperforms the alternative healthcare predictive analytics approaches in disease progression modeling.

This work has some limitations. First, the "no free lunch" theorem (Wolpert et al. 1997) implies that there will never be a learning method that can guarantee to outperform another method on every possible data set. Our evaluations are based on an EHR data set from ADNI. While we have employed cross-validation to train and test models, the better performance of DPLDF methods may still be limited to the data set under consideration. Future research may experiment the DPLDF approach on different data sets and explore the condition in which it is effective. Second, we assume a linear relationship between input features and output values in the proposed DPLDF approach. We note that this is a limitation to all existing disease progression modeling approaches in the literature. This is due to the relationship between input features and output values are not clear, and the non-linear model may make the learning optimization model non-convex, so a global optimal solution cannot be obtained through existing optimization methods. Despite these limitations, this study provides a new way to conduct longitudinal data fusion and healthcare predictive analytics for modeling disease progression.

Our future work includes the implementation of the method on an actual intelligent clinical decision support system, and evaluation of the approach under realistic conditions. Furthermore, we are planning to improve the performance of the present method by analyzing more patient information data, such as gender, age, family history, to reach the smaller examination cost and deal with more complex dynamic disease diagnosis and early warning situations.

Acknowledgements This work was partially supported by the Science Fund for Creative Research Groups of the National Natural Science Foundation of China[grant number 71421001]. We are grateful for this support. We also would like to thank the anonymous reviewers for their insightful and constructive comments, which greatly improved this paper.

References

Agarwal, R., Gao, G., DesRoches, C., Jha, A.K. (2010). Research commentary-The digital transformation of healthcare: Current status and the road ahead. *Information Systems Research*, *21*(4), 796–809.

Beck, A., & Teboulle, M. (2009). A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM Journal on Imaging Sciences*, *2*(1), 183–202.

Calhoun, V.D., & Adali, T. (2008). Feature-based fusion of medical imaging data. *IEEE Transactions on Information Technology in Biomedicine*, *13*(5), 711–720.

Chen, H., Chiang, R.H., Storey, V.C. (2012). Business intelligence and analytics: from big data to big impact. *MIS Quarterly*, *36*(4), 1165–1188.

Chen, L., Li, X., Yang, Y., Kurniawati, H., Sheng, Q.Z., Hu, H.Y., Huang, N. (2016). Personal health indexing based on medical examinations: a data mining approach. *Decision Support Systems*, *81*(1), 54–65.

Dubitzky, W., Wolkenhauer, O., Yokota, H., Cho, K.H. (2013). *Encyclopedia of Systems Biology*. New York: Springer-Verlag.

Duchesne, S., Caroli, A., Geroldi, C., Collins, D.L., Frisoni, G.B. (2009). Relating one-year cognitive change in mild cognitive impairment to baseline MRI features. *NeuroImage*, *47*(4), 1363–1370.

Fichman, R.G., Kohli, R., Krishnan, R. (2011). The role of information systems in healthcare: Current research and future trends. *Information Systems Research*, *22*(3), 419–428.

Khachaturian, Z.S. (1985). Diagnosis of Alzheimer's disease. *Archives of Neurology*, *42*(11), 1097–1105.

Li, C., Rana, S., Phung, D., Venkatesh, S. (2016). Hierarchical Bayesian nonparametric models for knowledge discovery from electronic medical records. *Knowledge-Based Systems*, *99*(9), 168–182.

Lin, Y.K., Chen, H., Brown, R.A., Li, S.H., Yang, H.J. (2017). Healthcare predictive analytics for risk profiling in chronic care: A Bayesian multitask learning approach. *Mis Quarterly*, *41*(2), 473–A3.

Liu, N., Qi, E.S., Xu, M., Gao, B., Liu, G.Q. (2019). A novel intelligent classification model for breast cancer diagnosis. *Information Processing & Management*, *56*(3), 609–623.

Mayaud, L., Lai, P.S., Clifford, G.D., Tarassenko, L., Celi, L.A.G., Annane, D. (2013). Dynamic data during hypotensive episode improves mortality predictions among patients with sepsis and hypotension. *Critical Care Medicine*, *41*(4), 954.

Meyer, G., Adomavicius, G., Johnson, P.E., Elidrisi, M., Rush, W.A., Sperl-Hillen, J.M., O'Connor, P.J. (2014). A machine learning approach to improving dynamic decision making. *Information Systems Research*, *25*(2), 239–263.

Nesterov, Y. (2013a). Gradient methods for minimizing composite functions. *Mathematical Programming*, *140*(1), 125–161.

Nesterov, Y. (2013b). Introductory lectures on convex optimization, vol 87. Springer.

Nie, L., Zhang, L., Meng, L., Song, X., Chang, X., Li, X. (2016). Modeling disease progression via multisource multitask learners: A case study with Alzheimer's disease. *IEEE Transactions on Neural Networks and Learning Systems*, *28*(7), 1508–1519.

OECD (2014). Unleashing the power of big data for Alzheimer's disease and dementia research.

Prince, M.J. (2015). World Alzheimer Report 2015: the global impact of dementia: an analysis of prevalence, incidence, cost and trends. Alzheimer's Disease International.

Saggi, M.K., & Jain, S. (2018). A survey towards an integration of big data analytics to big insights for value-creation. *Information Processing & Management*, *54*(5), 758–790.

Stonnington, C.M., Chu, C., Klöppel, S., Jack, J.r. C.R., Ashburner, J., Frackowiak, R.S. (2010). Predicting clinical scores from magnetic resonance scans in Alzheimer's disease. *NeuroImage*, *51*(4), 1405–1413.

Tai, A.M., Albuquerque, A., Carmona, N.E., Subramanieapillai, M., Cha, D.S., Sheko, M., Lee, Y., Mansur, R., McIntyre, R.S. (2019). Machine learning and big data: Implications for disease modeling and therapeutic discovery in psychiatry. *Artificial intelligence in medicine*, *99*(7), 101704.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society:, Series B (Methodological)*, *58*(1), 267–288.

Valmarska, A., Miljkovic, D., Lavrač, N., Robnik-Šikonja, M. (2018). Analysis of medications change in parkinson's disease progression data. *Journal of Intelligent Information Systems*, *51*(2), 301–337.

Wolpert, D.H., Macready, W.G., et al. (1997). No free lunch theorems for optimization. *IEEE Transactions on Evolutionary Computation*, *1*(1), 67–82.

World Health Organization (2012). Dementia: a public health priority. World Health Organization.

Xie, Q., Wang, S., Zhu, J., Zhang, X. (2016). S Disease Neuroimaging Initiative A Modeling and predicting ad progression by regression analysis of sequential clinical data. *Neurocomputing*, *195*(25), 50–55.

Yuan, L., Wang, Y., Thompson, P.M., Narayan, V.A., Ye, J. (2012). Alzheimer's Disease Neuroimaging Initiative Multi-source feature learning for joint analysis of incomplete multiple heterogeneous neuroimaging data. *NeuroImage*, *61*(3), 622–632.

Yuan, L., Liu, J., Ye, J. (2013). Efficient methods for overlapping group lasso. *IEEE transactions on pattern analysis and machine intelligence*, *35*(9), 2104–2116.

Yuan, M., & Lin, Y. (2006). Model selection and estimation in regression with grouped variables. *Journal of the Royal Statistical Society:, Series B (Statistical Methodology)*, *68*(1), 49–67.

Zhou, J., Yuan, L., Liu, J., Ye, J. (2011). A multi-task learning formulation for predicting disease progression. In *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 814–822): ACM.

Zhou, J., Liu, J., Narayan, V.A., Ye, J. (2012). Modeling disease progression via fused sparse group lasso. In *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 1095–1103): ACM.