Mobile Networks and Applications
https://doi.org/10.1007/s11036-021-01834-1

A Multi-modal Data Platform for Diagnosis and Prediction of Alzheimer's Disease Using Machine Learning Methods

Zhen Pang[1] · Xiang Wang[1] · Xulong Wang[1] · Jun Qi[2] · Zhong Zhao[3] · Yuan Gao[3] · Yun Yang[1] · Po Yang[4]

Accepted: 16 June 2021

Abstract
Alzheimer's an irreversible neurodegenerative disease with the most far-reaching impact, the most extensive, and the most difficult to cure in the world. It is also the most common disease of Alzheimer's disease. With the rapid rise of data mining, machine learning and other fields, they have penetrated various disciplines. In particular, research in the field of AD is developing rapidly and has demonstrated strong vitality. In terms of data, Alzheimer's Disease Neuroimaging Initiative (ADNI) researchers collect, verify and use a variety of data modalities as predictors of disease, including MRI and PET images, genetics, cognitive testing, cerebrospinal fluid and blood biomarkers, etc. Therefore, this paper uses a multi-task learning algorithm based on the ADNI data set to implement regression tasks and predict the cognitive scores of subjects in the next 3 years. This method can effectively assess the cognitive trends of patients in the future and aims to predict the progression of the disease. In addition, we used four different machine learning classification algorithms to conduct fusion research on AD multi-modal data, including MRI, PET, and cognitive scoring data. This method can determine the current patient's cognitive stage, to achieve the effect of assisting doctors in diagnosis. Finally, we designed a multi-modal data platform technical architecture to standardize management and sharing of ADNI data and data obtained by offline medical institutions to improve the utilization and value of data. The design of the technical architecture proposed in this article is more easily scalable and compatible with other neurological diseases. Nowadays, the large amount of data being generated by AD can provide valuable solutions for the research of disease progression prediction and auxiliary diagnosis.

Keywords Multi-modal data · Multi-task learning · Classification · Technical architecture · Disease progression prediction · Auxiliary diagnosis

1 Introduction

Alzheimer's disease (AD), is the most common type of dementia [1]. It is a progressive neurodegenerative disease with insidious onset, which is characterized by progressive damage to neurons and their connections leading to memory loss and cognitive decline. Clinically, it is characterized by comprehensive dementia such as memory impairment, executive dysfunction, and behavior changes [2, 3]. Mild cognitive impairment (MCI) [4] is also called cognitive impairment syndrome. It is a state between normal aging and dementia, and is widely regarded as the prodromal period of AD.

Recent statistics show that more than 50 million people worldwide suffer from Alzheimer's disease or other types of dementia. According to the World Health Organization, the number of people with dementia is expected to reach 82 million in 2030 and 152 million by 2050. Early detection

✉ Jun Qi
jun.qi@xjtlu.edu.cn

✉ Zhong Zhao
wasx-1128new@163.com

✉ Yun Yang
yang@ynu.edu.cn

[1] National Pilot School of Software for Yunnan University, Kunming, China

[2] Department of Computer Science and Software Engineering, Xi'an JiaoTong-Liverpool University, Suzhou, China

[3] Department of Neurology, the First People's Hospital of Yunnan Province, Kunming, China

[4] Department of Computer Science Faculty of Engineering, University of Sheffield, Sheffield, UK

Published online: 08 December 2021

of Alzheimer's disease is beneficial to affected individuals, their families, and society as a whole [5].

A cure for AD has no ideal solutions by far, but early intervention remains a primary requirement in prospective treatment and thus may significantly prolong the patients' lifespan [6]. Especially, the prodromal period of AD is of great importance for disease prediction. Therefore, in terms of data, regarding AD research, the data source commonly used in this article is from the Alzheimer's disease Neuroimaging Initiative (ADNI) [7]. It is an authoritative data source library shared by the whole world. It is also a multi-angle study with the focus on verticalization research and AD as the core. ADNI (http://adni.loni.usc.edu) aims to focus on imaging data, accompanied by genetic data, clinical data, biochemical biomarkers and other data modalities [8–10]. Its purpose is to detect and track the progress of AD as early as possible. In addition, some previous researchers have proposed a medical record management platform to help doctors make clinical decisions, analyze intractable diseases and effective medical care [11]. Some researchers put forward a model of integrated management of clinical big data, and designed a Hadoop-based platform to improve the sharing of medical data resources to serve the medical field more fully [12]. Other researchers focus on data collection and analysis, providing users with an effective and simple method way to classify data [13].

Previous work, nevertheless, either simply provides a visualization platform for physicians and patients or have more focused points on data analysis like early disease screening. The lack of a standardized and specialized disease diagnosis system provides technical support for the analysis process of disease prediction and auxiliary diagnosis. Specifically, in the data analysis part, few systems provide the prediction of the cognitive changes of patients in terms of dynamic time series.

Targeting these challenges, in this article, we propose an innovative technical architecture. Based on our previous work, we have carried out comprehensive upgrades and improvements [14]. It can realize the labeling, display, sharing, and visualization of AD data, and finally accurately provide the early prediction results of AD. The main contributions are as follows:

1) A multi-modal AD framework for integrating data from multiple sources is proposed. To start with, we obtained data modalities such as MRI, PET, and cognitive evaluation scores in clinical data from online ADNI, and then proceeded with data preprocessing and feature engineering.
2) Multi-Task learning (MTL) [15] considers the prediction of AD progression as manifold learning tasks which can be a general prediction task at a certain time point. Among these prediction tasks, all of them are assumed to be related to each other in the time domain with relevant temporal features (e.g., biomarkers in MRI). MTL algorithms for predicting cognitive ability of AD patients from their brain imaging scans, where the progress knowledge is shared and transferred among related subtasks to reinforce their own generalization ability [16]. The data sources employed are structured data (Extracted features from MRI like Volume of Hippocampus) and AD cognitive scores (MMSE or ADAS-cog) from selected AD patients repeatedly by multiple time points. Giving that the prediction of cognitive scores at a single time point (like 6, 12, or 18 months) as a regression task, the combined prediction of clinical scores at various forthcoming time points as a multi-task regression problem. MTL model weight matrix is trained and optimized through processing pre-extracted features from MRI and baseline cognitive scores. The overall idea is to process raw AD data and combine it with multi-task learning algorithms to finally provide patients with future cognitive changes. In other words, the MRI images in the baseline period are used to predict the changes in the patient's cognitive scores in the next few years, to achieve the effect of predicting disease progression.
3) In the application of classification algorithms, we first focus on the use of support vector machine algorithms for prediction. At the same time, we also conducted a comparative analysis on whether the MMSE score data was included, and found that the MMSE cognitive scale score data is one of the indispensable factors. Based on the existing MRI and MMSE modal data, a third modal data, namely PET, is introduced. This part of the data also plays an indispensable role in the diagnosis of AD disease. The data is extracted into three structured modal data, including MRI, PET, and MMSE. This part of the data is divided into two-by-two combinations, and then the three modes are combined into one group. The result is to divide it into four datasets. Next, train four classification algorithms (mainly XGBoost) models to perform the prediction accuracy of the three classifications, compare the differences between different modal datasets, and show the effect of multi-modal data fusion [17].

The structure of the rest of this paper is distributed as follows. Section II describes the role of the platform and the data processing method combined with the multi-task learning algorithm, and analyzes each sub-process in detail, and then explains the principle of the XGBoost classification algorithm. In the Section III, the implementation of the platform architecture is discussed. The results are shown in the Section IV. Conclusions and guidance for future work are presented in the Section V.

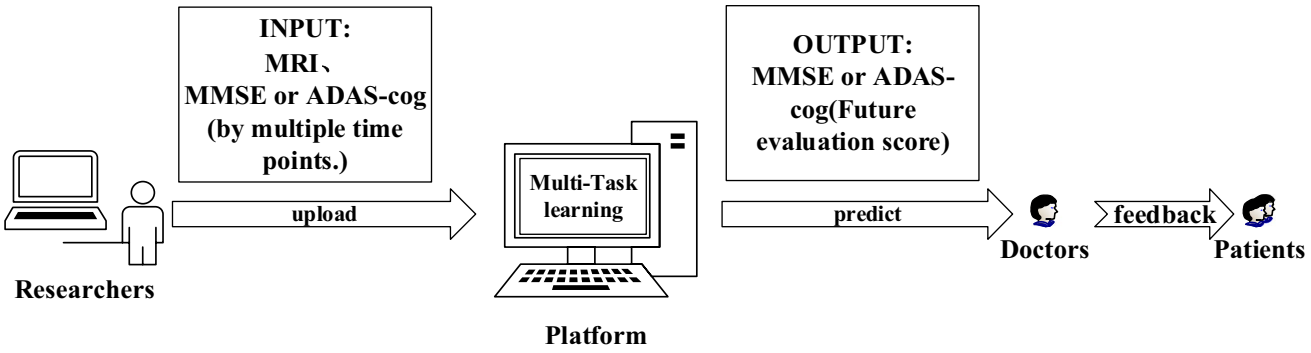

Fig. 1 The diagnosis process in a hospital environment

2 Methodology

The diagnosis of AD flowchart is shown in Fig. 1. This figure describes the detection process of patients with suspected AD after entering the hospital and integrates our platform to assist doctors in diagnosis. For example, after a suspected AD patient enters the hospital, he/she needs to have a doctor conduct a preliminary examination. Then check the equipment, such as biochemical indicators, MRI images and so on. According to medical experts, the most important of all patient test data are MRI image data and cognitive scores. After the doctor conducts a preliminary analysis of the patient's examination results, he draws a preliminary conclusion based on the doctor's experience. Then doctors use the platform to upload and input MRI image data, MMSE or ADAS-cog score data. The multi-task learning algorithm integrated into the platform is used to predict the change of the patient's cognitive score in the future for a while to assist in judging the change of the patient's future cognitive level. In addition, medical experts can upload three modal data (at least two types) such as MRI, PET and MMSE scores. The machine learning classification algorithm integrated into the platform trains classification models through different modal data to predict the final stage of the disease. The output results are submitted to the doctor for the summary. Finally, the doctor will feedback on the final prediction result to the patient. Next, we will introduce the theoretical part of the multi-task learning algorithm and the machine learning classification algorithm, and explain the experimental process in detail. The connection between multi-task learning and classification models is to analyze AD from different angles, in order to achieve the purpose of disease progression prediction and auxiliary diagnosis respectively. Both of the above are solutions to AD problems.

2.1 Multi-task learning

The purpose of MTL [18–20] is to learn a common set of features across all tasks and share them to improve the accuracy of all tasks. Among these learning tasks, a basic assumption of MTL is that one or more subsets are related to each other.

Consider an MTL of k tasks with n training samples of d features. Let $\{x_1, x_2, \ldots, x_n\}$be the input data for the samples, and $\{y_1, y_2,\ldots,y_n\}$be the predicted value for each sample, where each $x_i \in \mathbb{R}^d$ donates the feature data of an AD patient, and $y_i \in \mathbb{R}^k$ is the predicted value of cognitive score of different types of scales. The formulation for a linear regression question is given by $f_i(x) = x^T w_i$ where w_i is the weight vector of the model. In MTL, a matrix representation facilitates an intuitive understanding of algorithms and actual programming operations.

Then, let $X = [x_1, \cdots, x_n]^T \in \mathbb{R}^{n \times d}$ be the data matrix, $Y = [y_1,\cdot, y_n]^T \in \mathbb{R}^{n \times k}$ be the predicted matrix, and $W = [w_1, w_2, \ldots, w_k] \in \mathbb{R}^{d \times k}$ be the weight matrix. The process of establishing the MTL method is to calculate the value of W, which is the parameter to be estimated from the training samples.

Two common MTL models are presented to display their properties. Multi-task Lasso is a linear model that estimates sparse coefficients for multiple regression problems jointly. The constraint is that the selected features are the same for all the regression problems, also called tasks. The Fig. 2 compares the location of the non-zero entries in the coefficient matrix W obtained with a simple Lasso or a Multi-task Lasso. Mathematically, it consists of a linear model trained with a ℓ_{21}-norm for regularization. The objective function to minimize is:

$$\min_w \frac{1}{2n} \|XW - Y\|_F^2 + \alpha \|W\|_{21}. \quad (1)$$

where $\|\cdot\|_F$ donate the Frobenius norm $\|A\|_F = \sqrt{\sum_{i=1}^{m} \sum_{j=1}^{n} \left|a_{ij}\right|^2}$; and $\|W\|_{21}$ donate $\|W\|_{2,1} = \sum_{i=1}^{d} \sqrt{\sum_{j=1}^{t} W_{i,j}^2}$. The multi-task

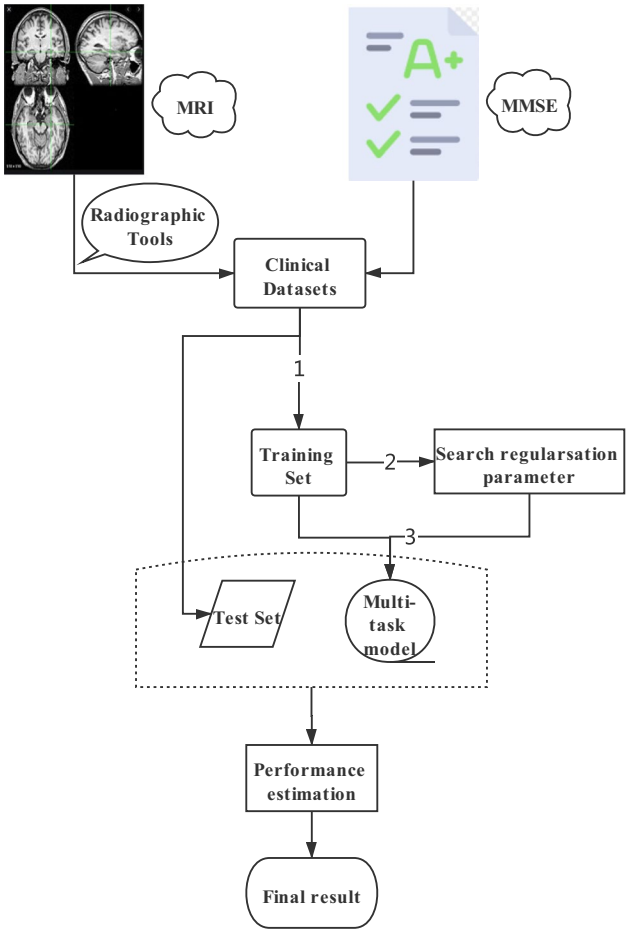

Fig. 2 Multi-task learning algorithm flowchart

lasso allows fitting multiple regression problems jointly enforcing the selected features to be the same across tasks.

For example, AD cognitive progress sequential measurements, each task is a time point, and the relevant features vary in amplitude over time while being the same. This makes feature selection by the Lasso more stable. However, when there are correlations between multiple features, the features will be randomly selected, especially when the brain region is regarded as a feature, there are some blocks with high correlation, such as atrophy of the cerebral cortex causes reduction in cortical volume and cortical thickness.

MTL confines the training process by using regularization terms and shares knowledge between tasks. Zhou et al. [21] proposed a convex fused sparse group Lasso methods including group Lasso and fused Lasso which considers the temporal patterns of the biomarkers and allows simultaneous joint feature selection for all tasks and selection of a specific set of features for each task. Cao et al. [22] improved sparse shared structure-based multi-task learning formulation including $L_{2,1}$ norm penalty, group $L_{2,1}$ norm penalty incorporating a hierarchical group sparsity and shared subspace uncovering regularization. Liu et al. [23] presented a multi-task sparse group lasso (MT-SGL) framework, which designs sparse features combined across tasks, and can satisfy loss functions associated with any Generalized Linear Models. Giving that a correlation sparse and low-rank constrained regularization, Wang et al. [24] propose a multi-task exclusive relationship learning model select the most discriminative features for different tasks and model the intrinsic relatedness among different time points.

In this paper, we concentrate on two AD progression prediction models proposed by Zhou: Temporal Group Lasso (TGL) and Convex Fused Sparse Group Lasso (cFSGL). Specifically, TGL contains a time smoothing term and a group Lasso term as constraints, which guarantees that each one regression fashions at exclusive time points share a not unusual place set of features. The TGL formulation solves the following convex optimization problem:

$$\min_W \|XW - Y\|_F^2 + \theta_1 \|W\|_F^2 + \theta_2 \|W\|_F^2 + \delta \|W\|_{2,1} \tag{2}$$

where the primary term measures the empirical mistakes at the training data, $\|W\|_F$ is the Frobenius norm, $\|WH\|_F^2$ is the temporal smoothness term, which ensures a small nonconformity between two regression models at consecutive instance points, and $\|W\|_{2,1}$ is the group lasso penalty, which ensures that a small subset of features will be chosen for the regression models at all instance points. cFSGL engages sparseness between tasks, which not only considers the common features at different points in time but also reflects the distinctive features to each task, which are effective to improve the overall performance of the model. cFSGL formulation solves the following convex optimization problem:

$$\min_W \|Y - XW\|_F^2 + \lambda_1 \|W\|_1 + \lambda_2 \left\|RW^T\right\|_1 + \lambda_3 \|W\|_{2,1} \tag{3}$$

where the first term estimates the observed error on the training data, $\|W\|_1$ is the lasso penalty, $\|RW^T\|_1$ is the fused lasso penalty, and $\|W\|_{2,1}$ is the group lasso penalty.

Lasso and group lasso combined employ are called sparse group lasso, which allows instantaneous selection of a public feature for all time points and internally makes sparse solutions in reaction to different time points. Fused lasso penalty having a given temporal smoothness, which makes chosen features at adjacent time points similar to each other. Besides, notice that cFSGL's formulation involves three non-smooth terms. The author suggested to employ the accelerated gradient descent method solving this optimization problem.

2.2 Pipeline of empirical protocol design

To use machine learning to pick out the most important features in the progression of Alzheimer's syndrome from MRI images, we utilize the structural data to train the model with the best prediction effect. Features selected by such models are convincing. Aiming at the machine learning model proposed by the current research of AD, the multi-task learning model embrace better performance. In order to verify this conclusion, we compared the two multi-task learning models cFSGL and TGL with the traditional regression models Ridge and Lasso under the same experimental conditions and selected the model with the best accuracy by comparing the experimental data.

The performance of the model is measured by evaluation metrics. Different evaluation metrics have different preferences. The regression performance metric often employed in multi-task learning is normalized mean square error (nMSE) and root mean square error (rMSE) is employed to measure the performance of each specific regression task. In particular, nMSE has been normalized to each task before evaluation, so it is widely used in multi-task learning methods based on regression tasks. In addition, weighted correlation coefficient (wR) which employed in a wide range of medical literatures to AD progress analysis problems [25, 26]. nMSE, rMSE and wR are defined as follows:

$$\mathrm{nMSE}(Y,\hat{Y}) = \frac{\sum_{i=1}^{t}\left\|Y_i-,\hat{Y}_i\right\|_2^2/\sigma\left(Y_i\right)}{\sum_{i=1}^{t} n_i} \tag{4}$$

$$\mathrm{rMSE}(y,\hat{y}) = \sqrt{\frac{\|y-\hat{y}\|_2^2}{n}} \tag{5}$$

$$\mathrm{wR}(\mathrm{Y},\hat{Y}) = \frac{\sum_{i=1}^{t}\mathrm{Corr}\left(\mathrm{Y}_i,\hat{Y}_i\right)n_i}{\sum_{i=1}^{t} n_i} \tag{6}$$

Our empirical protocol design is based on a pipeline shown in Fig. 2. The complete experimental process mainly includes 5 steps: (1) split the data set; (2) select the hyperparameters; (3) train the model; (4) evaluate the model using the test set; (5) iterate the above operations. Different colors donate the source or generation of different data, arrows indicate the flow of data, and serial numbers re the steps of the experiment.

2.3 Machine learning classification algorithm

In this section, we mainly explain the four classification algorithms used in the following experiments, of which the first three are the traditional classification algorithms that we are more familiar with, namely decision trees, random forests, and support vector machines. Next, I will mainly explain the classification algorithm with a higher degree of optimization. Its algorithm model is very powerful and mature. It is also the first choice for processing structured datasets.

eXtreme Gradient Boosting (XGBoost) is an extreme gradient boosting. It is often used in some competitions and its effect is remarkable. It is a tool for massively parallel boosted trees. The algorithm applied by XGBoost is an improvement of the Gradient Boosting Decision Tree for both classification and regression problems [27].

Like the random forest, gradient boosting is another technique used to perform supervised machine learning tasks. The implementation of this technology can use different names, the most common situation is that you encounter a Gradient Boosting computer (GBM) and XGBoost. Similar to "Random Forest", "Gradient Boosting" is a holistic learner. It creates the final model based on a collection of individual models. The predictive power of these individual models is very weak and easy to overfit, but merging many of these models together will bring greatly improved results overall.

The following describes the datasets and model:

Let training datasets $D = \left\{\left(\vec{x}_1, y_1\right), \cdots, \left(\vec{x}_n, y_n\right)\right\}$ where $\vec{x}_i \in R^m, y \in R$ is the final model of the entire XGBoost is:

$$\hat{y}_i = \phi\left(\vec{x}_i\right) = \sum_{k=1}^{K} f_k\left(\vec{x}_i\right), f_k \in F \tag{7}$$

where $F = \left\{f(\vec{x}) = w_{q(\vec{x})}\right\}$ is the decision tree space.

$q(\vec{x}) : R^m \to T$ $\vec{x}$ Map to a leaf;

$w \in R^T$ is the label for each leaf;

q is only the structure of a decision tree. The label on the specific leaf is determined by w, so w can be regarded as a vector $\vec{w} = \left(w_1, \cdots, w_T\right)$. The value of each dimension is a leaf label.

The XGBoost core algorithm is parallelizable so that it can be parallelized within a tree. XGBoost is usually used as a basic learner with a tree. The decision tree consists of a series of binary problems, and the final prediction occurs on the leaves. XGBoost is an integration method. The tree is to build iteratively until the stopping criterion is met. XGBoost uses the CART (Classification and Regression Tree) decision tree. CART is a tree that contains real values in each leaf, regardless of whether they are used for classification or regression. Then, if necessary, the real-valued scores can be converted into categories for classification.

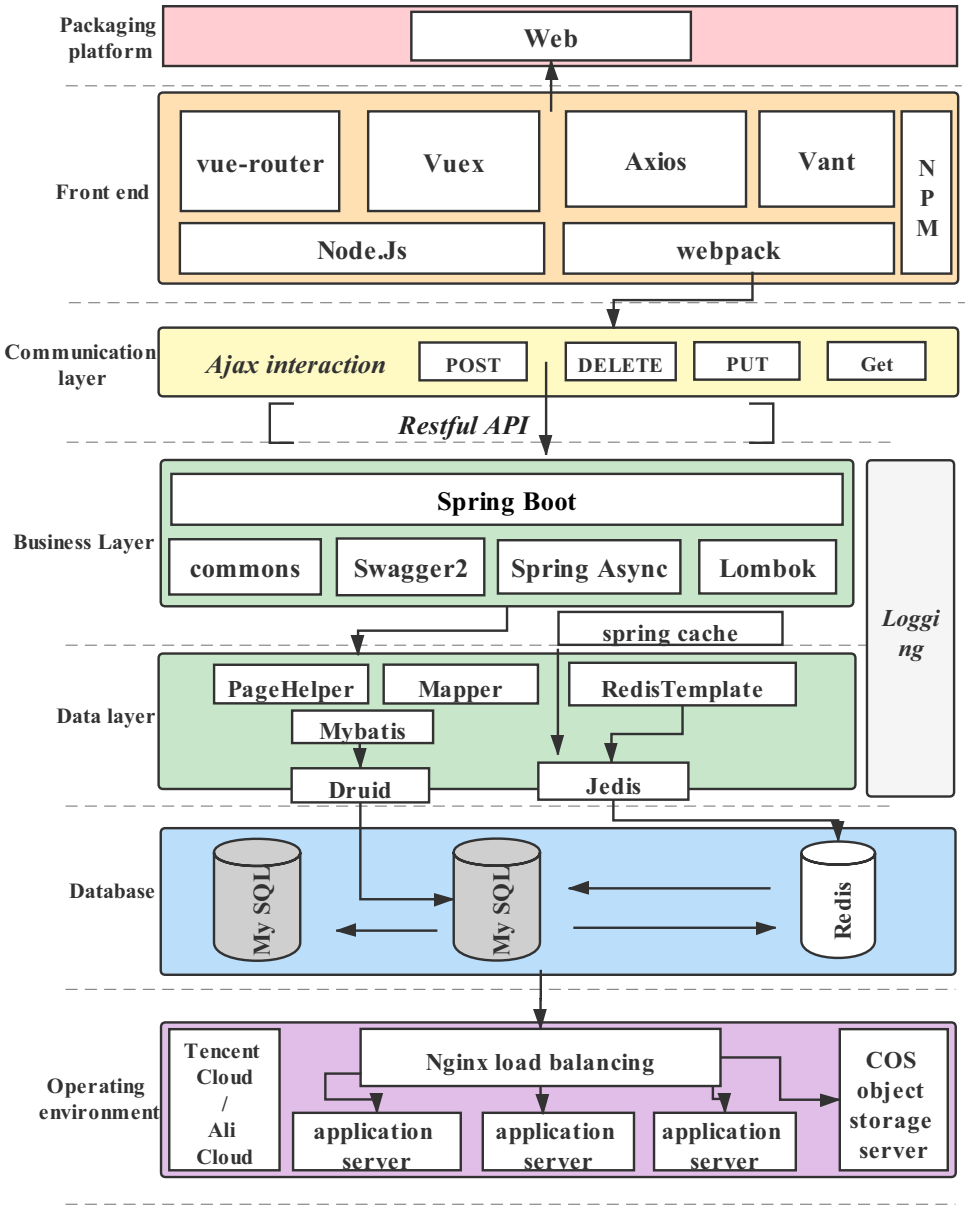

Fig. 3 The technical architecture diagram of the overall development (extended from [14])

XGBoost can be better prove the effect of multi-modality. Conventional gradient enhancement uses the loss function of the basic model (such as a decision tree) as a proxy to minimize the error of the overall model, while XGBoost uses the second derivative as an approximation. And advanced regularization (L1 & L2), which can improve the generalization ability of the model.

3 Implementation of the platform

To implement the above models, this paper uses the current mainstream front-end separation technology to develop a data platform [28]. The technology selection includes Vue framework, Spring Boot framework, and MySQL database [29].

In the architecture where the front and back ends are separated, the front and back ends are separated in different projects. They interact with each other through API. The front end is mainly responsible for the view layer and the controller layer. The backend is only responsible for the Model layer and business/data processing. node.js is explored here for the scenarios with high concurrency, I/O intensive, and a small amount of business logic [30–32].

Whether it is the front-end separation model or other models, it is to more conveniently solve the needs, but they are only a "transit station." The front-end project and the back-end project are two projects that are placed on two different servers and need to be deployed

independently, two different projects, two different codebases, and different developers. The front end only needs to focus on the page style and dynamic data analysis and rendering, while the back end focuses on specific business logic.

The technical architecture diagram of the overall development is shown in Fig. 3. Our platform front-end uses the Vue framework [33]. The characteristics of Vue are lightweight framework (relatively speaking), two-way data binding, simple and easier to learn.

The backend of technical architecture uses the Spring Boot framework [34, 35]. The main reasons why we choose the Spring Boot framework for development are high efficiency, easy operation, more highly integrated configuration resource files, clearer coding, and better programming experience. Spring allows me to easily connect to database and queue services, such as MySQL.

The back-end database is currently based on MySQL, which is the most popular relational database management system. MySQL has excellent performance, and problems such as BUG downtime are rare. MySQL supports transactions, views, stored procedures, triggers, etc. High speed is a major feature of MySQL. It is very suitable for the development of small and medium-sized projects and is very suitable for our projects.

4 Results

4.1 Platform display

The homepage of the system is composed of three areas. The left area is the side navigation bar, which can retrieve data modal types and use function modules. The content presented at the top is the system name and basic information. The content displayed in the remaining area is detailed data information and functional parts, including the detailed content of the datasets, the number of views, downloads, version information, and so on (Table 1). The specific information is shown in Fig. 4. The description of Fig. 4 is shown in Table 2.

4.2 Model comparison

The data in Table 1are data from 429 subjects, the evaluation index is MMSE, and the number of features of each subject is 327 dimensions. It can be seen from Table 1 that the average and variance of the MTL model and single-task models like Ridge and Lasso in the nMSE and rMSE test indicators are smaller, indicating that their performance prediction is better, and a larger wR indicates that the correlation among tasks in the task model is stronger.

Comprehensive comparison, the performance of cFSGL is better in the multi-task model, so we chose it as the machine learning model for selecting biomarkers in this paper.

It shows that our experimental results under similar settings are quite close to Zhou's outcomes. It implies that combined selected structural regularization methods are all robust. Besides, MTL models (TGL and cFSGL) outperform the single task learning models (Ridge and Lasso) in terms of nMSE. This accords with our previous survey of features of MTL in dealing with data insufficiency cases. Notably, cFSGL performs the best in all 4 methods. It is probably because in the AD study, the model built by cFSGL has two levels of sparseness: (1) some common features shared within every part of tasks, (2) distinctive features for each point-in-time. The effect of cFSGL is significantly better than that of TGL. This may be due to the more restrictive sparse requirements imposed on the cFSGL. This proves that in the sparse optimization, the basic assumption of things that "only part of the key function of things" is correct, just like the learning process of new things and knowledge in human brain.

4.3 Determine features

Through the above comparative experiments, it is verified that the multi-task learning model cFSGL can indeed show the best performance when processing small sample large-dimensional data of AD. This is because the cFSGL model involves sparseness between tasks, which takes into account the common characteristics at distinctive instance in time and the specific features of every sub-task, and the key feature selected helps to enhance the general performance of the algorithm. In the end, the weight matrix W of the model gave important values $W_{i}j$ a larger value, and unimportant Wij was given a smaller value.

Previously, studies that have achieved better results in predicting AD use the trained model, that is, the weight matrix W is multiplied by the corresponding 327 features in the test set and then accumulated, and finally the accumulated value is corresponding to the label of the test subject. That is, the subject's actual MMSE and ADAS-cog scores during this period, the accuracy of the model was tested through three evaluation indicators of nMSE, rMSE, and wR, but the weight matrix W was not analyzed in detail.

The weight matrix W has a very strong representation of the features and numerically reflects the features. So,

Table 1 Comparison of AD disease progression model

	Ridge	Lasso	TGL	cFSGL
Target: MMSE				
nMSE	1.185 ± 0.286	0.641 ± 0.156	0.562 ± 0.106	**0.459 ± 0.095**
wR	0.545 ± 0.057	0.694 ± 0.034	0.734 ± 0.057	**0.777 ± 0.034**
M06 rMSE	2.770 ± 0.360	2.044 ± 0.472	1.853 ± 0.225	**1.845 ± 0.259**
M12 rMSE	3.029 ± 0.293	2.226 ± 0.466	1.972 ± 0.244	**1.873 ± 0.266**
M24 rMSE	3.375 ± 0.470	2.690 ± 0.664	2.544 ± 0.535	**2.374 ± 0.479**
M36 rMSE	4.533 ± 0.513	3.287 ± 0.584	3.060 ± 0.437	**2.932 ± 0.594**
Target: ADAS-cog				
nMSE	0.693 ± 0.116	0.417 ± 0.052	0.408 ± 0.073	**0.358 ± 0.057**
wR	0.660 ± 0.052	0.777 ± 0.034	0.789 ± 0.042	**0.809 ± 0.034**
M06 rMSE	4.517 ± 0.412	3.387 ± 0.496	3.500 ± 0.561	**3.319 ± 0.401**
M12 rMSE	3.387 ± 0.393	3.644 ± 0.462	**3.467 ± 0.437**	3.485 ± 0.473
M24 rMSE	5.519 ± 0.713	4.248 ± 0.828	4.260 ± 0.913	**3.553 ± 0.453**
M36 rMSE	7.655 ± 1.200	6.088 ± 1.077	**5.707 ± 0.824**	5.739 ± 1.037

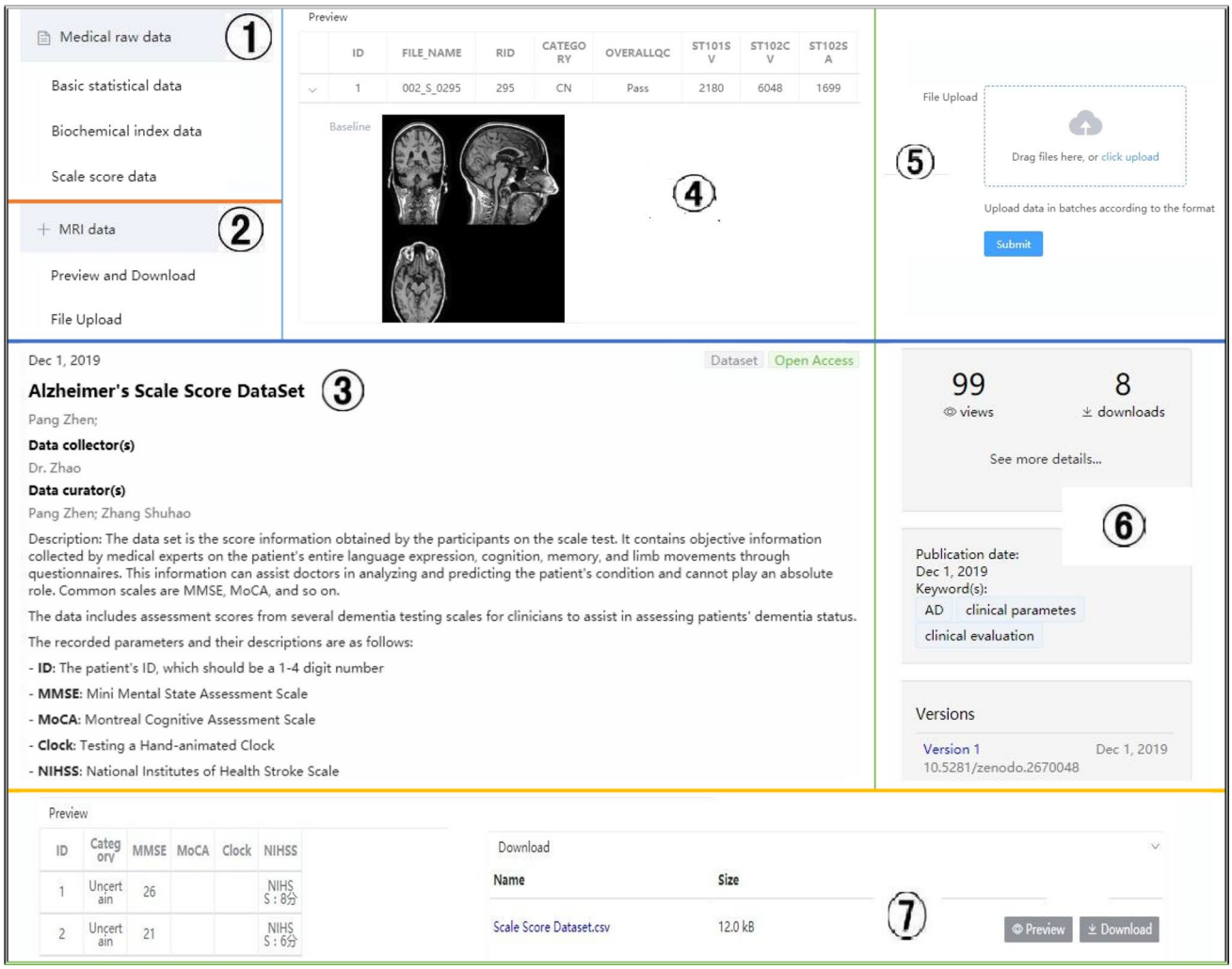

Fig. 4 The display of the platform and its functional modules (extended from [14]). The Table 2 is an interpretation

Table 2 Functionality of the platform

Label	Title	Description (associated with Fig. 4)
1	Offline data	Offline data is divided into three types of data
2	MRI	The integration of ADNI online datasets, including datasets download, upload and other functions
3	Datasets details	Classify and label offline data, and interpret and explain each type of data in detail
4	Three views of MRI	The three views of the MRI image correspond to the structured data extracted from it one-to-one
5	upload	Upload important data and perform algorithmic analysis on the data. Please refer to the next experimental part for the results
6	Background statistics	Basic information such as the number of times the datasets has been viewed and downloaded, the release date and version
7	Download and preview	Download and preview function of datasets

Table 3 Averaged results from executing SVM on ADNI data: 50 iterations, not including neurophysiological test mini mental state examination (MMSE) to eliminate bias

Task	NOT MMSE		
	Specificity (%)	Sensitivity (%)	Accuracy (%)
AD/MCI	47	72	65
MCI/CN	56	79	70
AD/CN	83	91	88

Table 4 Averaged results from executing SVM on ADNI data: 50 iterations, including neurophysiological (MMSE)

Task	MMSE		
	Specificity (%)	Sensitivity (%)	Accuracy (%)
AD/MCI	66	88	80
MCI/CN	67	90	80
AD/CN	98	98	98

we propose to analyze the weight matrix to determine the importance and particularity of the features.

4.4 Multi-modal data classification comparison

The datasets is divided into three categories, namely CN, MCI, and AD. What we have achieved is to classify these three categories to achieve the purpose of auxiliary diagnosis. We selected two types of datasets from the three types, and first performed a two-classification task. Among them, we did a comparative experiment, that is, whether to add MMSE data to the MRI single-modality data. The experimental results are shown in Tables 3 and 4. Here, we focus on using the SVM algorithm for prediction. At the same time, other classification algorithms can also be used for experiments [36–39]. This experiment concludes that adding the MMSE scale score data to the classification experiment has a significant effect.

Table 5 The accuracy obtained when various algorithms perform three classifications on different multi-modal data sets

Model Datasets	DT (%)	RF (%)	SVM (%)	XGBoost (%)
MRI + PET	71	63	53	63
MRI + MMSE	64	68	67	66
PET + MMSE	64	75	72	71
MRI + MMSE + PET	62	63	64	72

Based on the existing MRI and MMSE modal data, a third modal data, namely PET, is introduced. This part of the data also plays an indispensable role in the diagnosis of AD disease. We adopt the data extraction method introduced in Chapter 2, Extract the data into three structured modal data, divided into MRI, PET, MMSE. We combine this part of the data into two groups and combine the three modalities into one group, and divide them into four datasets. Next, the accuracy of three classification predictions is performed on the four algorithm models. Here mainly use decision trees, random forests, support vector machines, and the more famous XGBoost algorithm, a total of 16 sets of experiments. The detailed results are shown in the analysis of experimental results, as shown in Table 5.

Through the comparison of the results of the Tables 3 and 4, we can draw: (1) The MMSE cognitive score plays an important role in the binary classification experiment. (2) Relatively high sensitivity indicates a strong ability to diagnose the disease. (3) Since the disease always tends to be diagnosed on the more serious side, it needs to be improved. (4) The accuracy of the experimental results is ideal. Because when we add the modal data of MMSE, the accuracy rate can be increased to more than 80%. Compared with the single mode, the accuracy rate is greatly improved. This proves that our data processing is effective, and the performance of the selected algorithm is also very superior, which can achieve our expected results. Therefore, the platform can assist doctors in diagnosis.

The above experiments ensure that the data sets has changed under the premise that the algorithm model remains unchanged, and shows the accuracy of the three classifications of AD, MCI, and CN. From the results of the Table 5, we can see: (1) Decision tree, the effect of prediction is different in different modalities, and the data sets for MRI and PET are more preferred. (2) Random forest prefers PET and MMSE data sets, which have higher accuracy. The effects of the three fusions are better than those of the decision tree. (3) SVM is more prone to be overfitting for these four types of datasets. We compared the learning effects of the test set and the training set in the experiment and found that the training set has good effects, and the test set has a mediocre effect. (4) When there are a large number of training samples, ideally, more than 1000 training samples and less than 100 features, or we can say that the number of features is less than the number of training samples. And when there is a mixture of classification features and digital features or only digital features, the XGBoost algorithm is preferred. So we use this algorithm to check the classification effect, and we can conclude that the combination of the three modal data sets is better than the two modal data sets. This also reflects the importance of diversification of modal data types in high-performance algorithms.

5 Conclusion and future work

The large amount of data generated by Alzheimer's disease can be used to provide valuable assistance for disease diagnosis and progression prediction. In terms of data, as mentioned above, ADNI is a multi-angle study with a focus on verticalization and AD as the core. The purpose of the research is to focus on imaging data, accompanied by genetic data, clinical data, biochemical biomarkers, and other data modalities. Its purpose is to detect and track the progress of Alzheimer's disease (AD) as early as possible. We conduct multi-modal heterogeneous data fusion research based on the ADNI datasets.

This article focuses on the research of multi-modal data fusion about AD. The progress of the work is rough as follows: investigating and consulting related literature on AD diseases, understanding the different data modalities that affect AD diseases, collecting and sorting out important data sets related to assisting decision-making and diagnosis of diseases, and labeling and integrating these original data. Next, feature extraction is performed on the original data of different modalities. The most important thing is for the two major modal data of MRI and PET because the data are all original image data, we need to extract the structured data we use to connect with the algorithm model. The extracted structured datasets is subjected to preprocessing and feature engineering operations, and finally, we perform statistical analysis and visualization operations on the collated data to make the presentation of the data more intuitive. Under the premise that the control algorithm remains unchanged, we conduct comparative experiments on single-mode and multi-modal data and summarize the experimental results. Due to the complicated process of labeling and collecting AD multi-modal data, we have designed a technical architecture for displaying, labeling and storing data. It not only improves the value utilization of data, but also provides a solution for data modalities from multiple sources. On the other hand, we use data from the ADNI database to support multi-task learning algorithm models to obtain better evaluation scores to predict disease progression.

In the course of the experiment, we also discovered some problems of future machine learning in diagnosing AD diseases, including the generalization ability of the algorithm model to avoid serious overfitting. When we use SVM to train MRI and MMSE data sets, it is prone to overfitting. This is also one of the issues that need to be considered in the field of AD diagnosis in the future.

For the future, the correlation between data modalities can be deeply explored, so that different modal data can have a better fusion effect. Algorithms only constantly approach the upper limit, and data, especially multi-modal data, determines the upper limit of machine learning. The different modal types and complexity of AD data are the primary problems to be solved in the study of AD. In the future, multi-modal data fusion will be the top priority in the research of AD and its diagnosis. At the same time, we also consider adding multi-modal fusion algorithms to enhance our experimental results, and add more comparative experiments to study the importance of multi-modality.

Currently, there are still missing data sets and data imbalances in the data sets that can be researched. AD data sets also involve ethical and moral issues, which will greatly increase the difficulty of collection and collection. The labeling, integration, and collection of data sets will become more and more important in future machine learning research. Only good data will make our research more convincing and produce ideal results.

The data we study mainly come from the ADNI database. So far, the data in this database is collected in a controlled environment. After that, we can consider adding data collected in an uncontrolled external environment, such as a sports bracelet [40–43].

The technical architecture should be expanded with more functions in the future. We dig deep into the scientific value and commercial value, and make it suitable for the research of other diseases. From the perspective of physicians and artificial intelligence experts, it should be continuously improved and optimized so that it can serve more scholars.

Acknowledgements This research process is very grateful for the strong support of Kunming Hospital and this work was supported by the Yunnan University's Research Innovation Fund for Graduate Students (No. 2020228).

References

1. Desai AK, Grossberg GT (2005) Diagnosis and treatment of Alzheimer's disease. Neurology 64(12 suppl 3):S34–S39
2. Khachaturian ZS (1985) Diagnosis of Alzheimer's disease. Arch Neurol 42(11):1097–1105
3. Rosen WG, Mohs RC, Davis KL (1984) A new rating scale for Alzheimer's disease. Am J Psychiatry 141(11):1356–1364
4. Petersen RC, Smith GE, Waring SC, Ivnik RJ, Tangalos EG, Kokmen E (1999) Mild cognitive impairment: clinical characterization and outcome. Arch Neurol 56(3):303–308
5. "2019 Alzheimer's disease facts and figures," Alzheimer's & Dementia, 15: 321–387. https://doi.org/10.1016/j.jalz.2019.01.010
6. Cummings JL, Doody R, Clark C (2007) Disease-modifying therapies for Alzheimer disease challenges to early intervention. Neurology 69(16):1622–1634
7. Petersen RC, Aisen PS, Beckett LA, Donohue MC, Gamst AC, Harvey DJ, Jack CR, Jagust WJ, Shaw LM, Toga AW, Trojanowski JQ, Weiner MW (2010) Alzheimer's Disease Neuroimaging Initiative (ADNI) clinical characterization. Neurology 74(3):201–209
8. Jack CR, Petersen RC, O'Brien PC, Tangalos EG (1992) MRI-based hippocampal volumetry in the diagnosis of Alzheimer's disease. Neurology 42(1):183–188
9. Coleman RE (2007) Positron emission tomography diagnosis of Alzheimer's disease. PET Clin 2(1):25–34
10. Holtzman DM (2011) CSF biomarkers for Alzheimer's disease: current utility and potential future use. Neurobiol Aging 32(Supplement):1
11. Ogescu C, Plaisanu C, Bistriceanu D (2008) "Web based platform for management of heterogeneous medical data," 2008 IEEE International Conference on Automation, Quality and Testing, Robotics, Cluj-Napoca, pp. 257–260, 2008
12. Lyu D, Tian Y, Wang Y, Tong D, Yin W, Li J (2015) "Design and Implementation of Clinical Data Integration and Management System Based on Hadoop Platform," 2015 7th International Conference on Information Technology in Medicine and Education (ITME), Huangshan, pp. 76–79, 2015
13. Lizarraga G, Cabrerizo M, Duara R, Rojas N, Adjouadi M, Loewenstein D (2016) "A Web Platform for data acquisition and analysis for Alzheimer's disease," Southeast Con 2016, Norfolk, pp. 1–5
14. Pang Z, Zhang S, Yang Y, Qi J, Yang P (2020) "Interoperable Multi-Modal Data Analysis Platform for Alzheimer's Disease Management," 2020 IEEE Intl Conf on Parallel & Distributed Processing with Applications, Big Data & Cloud Computing, Sustainable Computing & Communications, Social Computing & Networking (ISPA/BDCloud/SocialCom/SustainCom), pp. 1321–1327. https://doi.org/10.1109/ISPA-BDCloud-SocialCom-SustainCom51426.2020.00196
15. Thung KH, Wee CY (2018) A brief review on multi-task learning. Multimed Tools Appl 77(22):29705–29725
16. Zhou J, Yuan L, Liu J, Ye J (2011) "A multi-task learning formulation for predicting disease progression," in Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 814–822
17. Qi J, Yang P, Newcombe L, Peng X, Yang Y, Zhao Z (2020) An overview of data fusion techniques for internet of things enabled physical activity recognition and measure. Inf Fusion 55:269–280
18. Gong P, Ye J, Zhang C (2013) "Multi-stage multi-task feature learning," J Mach Learn Res
19. Argyriou A, Evgeniou T, Pontil M (2007) "Multi-task feature learning,"
20. Argyriou A, Evgeniou T, Pontil M (2008) "Convex multi-task feature learning," Mach Learn
21. Zhou J, Liu J, Narayan VA, Ye J, (2012) "Modeling disease progression via fused sparse group lasso," in Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 1095–1103
22. Cao P, Shan X, Zhao D, Huang M, Zaiane O (2017) Sparse shared structure based multi-task learning for MRI based cognitive performance prediction of Alzheimer's disease. Pattern Recognit 72:219–235
23. Liu X, Goncalves AR, Cao P, Zhao D, Banerjee A (2018) Modeling Alzheimer's disease cognitive scores using multi-task sparse group lasso. Comput Med Imaging Graph 66:100–114
24. Wang M, Zhang D, Shen D, Liu M (2019) Multi-task exclusive relationship learning for alzheimer's disease progression prediction with longitudinal data. Med Image Anal 53:111–122
25. Ito K, Corrigan B, Zhao Q et al (2011) Disease progression model for cognitive deterioration from Alzheimer's disease neuroimaging initiative database. Alzheimer's Dement 7(2):151–160
26. Stonnington CM, Chu C, Klöppel S, Jack CR, Ashburner J, Frackowiak RSJ (2010) Predicting clinical scores from magnetic resonance scans in Alzheimer's disease. Neuroimage 51(4):1405–1413
27. Chen T, Guestrin C (2016) "XGBoost: a scalable tree boosting system," In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '16). Association for Computing Machinery, New York, pp. 785–794. https://doi.org/10.1145/2939672.2939785
28. Nan F, Yang P, Meng Q, Xie Y, Zhang D, Muhammad K (2019) GAN-based semi-supervised learning approach for clinical decision support in health-IoT platform. IEEE Access 7:8048–8057
29. Győrödi C, Győrödi R, Pecherle G, Olah A, (2015) "A comparative study: MongoDB vs. MySQL,"2015 13th International Conference on Engineering of Modern Electric Systems (EMES), Oradea.
30. Taylor RH, Rose F, Toher C, Levy O, Yang K, Nardelli MB, Curtarolo S (2014) "A RESTful API for exchanging materials data in the AFLOWLIB.org consortium," Computational Mater Sci, 93.
31. Chen X, Fang X, Lin X (2012) "Ajax-based Positioning System for Coal Miners," 2012 Third World Congress on Software Engineering, Wuhan.
32. Agocs A, Goff JL (2018) "A web service based on RESTful API and JSON Schema/JSON Meta Schema to construct knowledge graphs," 2018 International Conference on Computer, Information and Telecommunication Systems (CITS), Colmar.
33. Wang S, Li X, Duan S, Bu Z, Jian X, He C (2019) "Modeling and Simulation of Radar Klystron Based on the System Vue,"2019 International Conference on Meteorology Observations (ICMO), Chengdu.
34. Ying-kui D, Yang W, Ping G, Yue P, LiJuan Z, Shu L (2019) "Cloud Data Monitoring Management and Visual Application System Based on Spring Boot," 2019 IEEE 4th Advanced Information Technology, Electronic and Automation Control Conference (IAEAC), Chengdu.
35. Guntupally K, Devarakonda R, Kehoe K (2018) "Spring boot based REST API to improve data quality report generation for big scientific data: ARM data center example," 2018 IEEE International Conference on Big Data (Big Data), Seattle.
36. Zhang Y, Guo Y, Yang P, Chen W, Lo B (2020) Epilepsy seizure prediction on EEG using common spatial pattern and convolutional neural network. IEEE J Biomed Health Inform 24(2):465–474

37. Moore B, Berger T, Song D (2020) "Validation of a Convolutional Neural Network Model for Spike Transformation Using a Generalized Linear Model," 2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), pp. 3236–3239, https://doi.org/10.1109/EMBC44109.2020.9176458
38. Fan D, Yang J, Zhang J, Lv Z, Huang H, Qi J, Yang P (2018) "Effectively measuring respiratory flow with portable pressure data using back propagation neural network," in IEEE Journal of Translational Engineering in Health and Medicine, vol. 6, pp. 1–12, Art no. 1600112
39. Xin R, Zhang J, Shao Y (2020) Complex network classification with convolutional neural network. Tsinghua Sci Technol 25(4):447–457. https://doi.org/10.26599/TST.2019.9010055
40. Deng Z, Yang P, Zhao Y, Zhao X, Dong F (2015) "Life-Logging Data Aggregation Solution for Interdisciplinary Healthcare Research and Collaboration," 2015 IEEE International Conference on Computer and Information Technology, pp. 2315–2320.
41. Cinel G, Tarim EA, Tekin HC (2020) Wearable respiratory rate sensor technology for diagnosis of sleep apnea. Med Technol Congress (TIPTEKNO) 2020:1–4. https://doi.org/10.1109/TIPTEKNO50054.2020.9299255
42. Mohsen S, Zekry A, Abouelatta M, Youssef K, (2020) "A self-powered wearable sensor node for IoT healthcare applications," 2020 8th International Japan-Africa Conference on Electronics, Communications, and Computations (JAC-ECC), pp. 70–73, https://doi.org/10.1109/JAC-ECC51597.2020.9355925
43. Qi J, Yang P, Waraich A, Deng Z, Zhao Y, Yang Y (2018) Examining sensor-based physical activity recognition and monitoring for healthcare using internet of things: a systematic review. J Biomed Informatics. https://doi.org/10.1016/j.jbi.2018.09.002