Using Machine Learning Algorithms to Predict Alzheimer's disease

Team: Wentao Yuan, Zhi Zheng, Kainan Liu (all from 519, KongYao's Cohort Meeting)
Project Mentor TA: KongYao Chee

Abstract

Detecting and predicting Alzheimer's disease (AD) is crucial for patients. This project aims to use machine learning technology to build a model that can predict AD at an early stage. The project tries on four types of ML algorithms with different combinations of training feature sets. The result shows random forest with magnetic resonance imaging information as training features produce the most accurate predictions.

Github Link to Code: https://github.com/wentaoy/CIS519_Project.git

Introduction

Alzheimer's disease (AD) is one of the most common causes of dementia. Current available therapies for AD can only delay the advance of symptoms and have the greatest impact when provided at the earliest stage. Therefore, early diagnosis of AD is key to most patients who are at high risk. Mild cognitive impairment (MCI) is considered the most important signal at the early stage of AD. According to referral resources such as memory clinics and AD centers, approximately 10% to 15% of MCI patients will develop into AD annually (Moradia et al., 2015). We focus on early AD diagnosis by predicting the MCI-to-AD conversion using machine learning algorithms.

Related Prior Work

[3] developed a magnetic resonance imaging (MRI)-based method for predicting the MCI-to-AD conversion from one to three years before the clinical diagnosis. In particular, researchers first learned a separate MRI biomarker by unsupervised learning and then integrated it with age and cognitive measures with a random forest classifier. The resulting classifier achieved a 10-fold cross-validated score of 0.9020. [4] employed a sample of 550 MCI subjects whose diagnostic follow-up is available for at least 3 years after the baseline assessment. Information regarding sociodemographic, clinical characteristics, and neuropsychological tests were used to develop ML models. The final neural network model produced an accuracy of 0.88. [5] uses the Boruta algorithm as the feature selection method, which showed that random forest with Grid Search Cross-Validation outperformed with a 0.9439 accuracy. This article showed how to tune hyperparameters of the random forest by using Grid Search Cross-Validation, and Randomized Search Cross-Validation. Those articles above gave us a sense of selecting and examining features; selecting ML models and tune hyperparameters; as well as establishing a goal accuracy of our models to be around 90%.

Formal Problem Setup (TEP)

T(Task): Our objective is using machine learning techniques to predict Alzheimer's disease at an early stage. We mainly focus on the prediction of the conversion of mild cognitive impairment (MCI) to AD by using decision tree, random forest, and deep learning.

E(Experience): The models are trained with a set of clearly labeled clinical MCI data. The dataset has basic survey information, such as age, gender, socioeconomic status, educational level, and Mini-Mental State Examination score. And MRI information, such as MRI cross-sectional images, Estimated Total Intracranial Volume and Normalize Whole Brain Volume, and so on. The patients are classified into three groups: Demented, which means diagnosed as AD; Nondemented, which means diagnosed as non-AD; and Converted, which means diagnosed as non-AD originally, but then converted to AD. The models try to learn from samples by selecting which features are more important for making predictions and eventually be able to predict if the MCI patients can be diagnosed as AD, if not, they will be converted into AD.
P(Performance Metrics): This task is a multi-classification problem. Therefore, we use cross-entropy as loss function and accuracy classification score as model score. The accuracy classification score function computes subset accuracy: the set of labels predicted for a sample must exactly match the corresponding set of labels in y_true.

Methods

Data Acquisition and Preprocessing

The MRI-related data was generated by the Open Access Series of Imaging Studies (OASIS) project. We use OASIS-1 cross-sectional (only used by CNN) and OASIS-2 longitudinal MRI datasets. OASIS-1 consists of a cross-sectional collection of 416 subjects aged 18 to 96. For each subject, 3 or 4 individual T1-weighted MRI scans obtained in single scan sessions are included. OASIS-2 set consists of a longitudinal collection of 150 subjects aged 60 to 96. Each subject was scanned on two or more visits, separated by at least one year for a total of 373 imaging sessions. For OASIS-2 dataset, we remove all irrelevant features, such as patient ID, subject ID. Then we handle the missing value by fitting into the feature mean. For OASIS-1 cross-sectional dataset, we keep all the features and implement a custom dataset to combine MRI images with labels.

Decision Tree and Random Forest

We use Sklearn's train_test_split function to shuffle and split the OASIS-2 dataset into train and test sets, with 20% of the data to be used as the test set. We first train a decision tree and a random forest classifier by using Sklearn's DecisionTreeClassifier and RandomForestClassifier function respectively and with the Gini criterion. Both classifiers are trained with different sets of features and different max-depths to give us a sense about which features are more useful in this application and how depth affects prediction accuracy. After getting a fundamental sense, we train another random forest with different combinations of hyperparameters. We use the GridSearchCV and RandomizedSearchCV functions from Sklearn to tune hyperparameters of the random forest model. We provide functions with a list of each hyperparameter, and the functions will automatically tune and find the best combination. To get more diverse predictors, we also train the second random forest model with different sets of features.

Deep learning (DNN)

We use the torch.nn and torch.nn.Module functions to construct neural networks. We use torch.nn.Linear as layers, and torch.F.relu as activation function. torch.nn.CrossEntropyLoss is used as the loss function, and torch.optim function is used to try out different optimizers. Again, each neural network is trained with a different set of features; so we have multiple neural networks to compare with. Each neural network is tuned based on layer number, layer output-input number, learning rate, epoch number, and optimizer type to get the best result.

Convolutional Neural Network (CNN)

CNN model is created by Tensorflow, and it focuses on using OASIS-1 MRI images to predict AD. We implement two models with different input formats. The first one model uses the complete MRI image, which overall runtime is 8 minutes. In order to speed up the training without losing accuracy, we transform the original MRI image into only the hippocampus image., and construct the second CNN model. According to medical research, the hippocampus is highly relative to MCI and it's most indicative area for AD symptoms, therefore we extract the left hippocampus image from the brain MRI image, another employs the transformation of image (only takes the left hippocampus area). The proposed two CNN models have trained over 11 training epochs with batch size of 4. Binary cross-entropy (torch.nn.CrossEntropyLoss) is used during training, and different optimizer functions from torch have been tuned during the learning process. For more information, refer to the implementation details section.

Baseline approaches we compare against: The baseline methods are decision trees with different depths and random forest with optimized hyperparameters. They got highest accuracy 0.91 and 0.96 respectively. The models are neural network and convolutional neural network, and they got highest accuracy 0.92 and 0.86 respectively. Therefore, in our project, the baseline performs better than the models.

Implementation Details:
Dataset Separation: In order to examine whether the survey information or the MRI information is more important for the application, we make three sub-dataset: one with all features, one with only survey information, and one with only MRI features. With ethical consideration, gender is a sensitive feature. We thus further make two additional sub-dataset: all features with no gender, and only survey information with no gender. Thus, we end up with a total of five sub-dataset.
Decision Tree: We use Gini as criterion for each model. Each model is trained with five sub-dataset as mentioned above, to examine which features are more important and measure how much effect gender features have on prediction. For those five decision tree models, we change max-depth to have a more diverse result. The max-depth range from 1 to feature numbers. For example, the decision tree trained with all features sub-dataset has 9 features in total; and thus, the max-depth to try on is from 1 to 9.
Random Forest: The five sub-dataset are then used to train the Random Forest classifier. To tune the hyperparameters, we first change only maximum depth without optimizer. To search and find out the best set of hyperparameters, we use two types of cross-validation techniques: grid search cross-validation (GSCV) and randomized search cross-validation (RSCV). The set of hyperparameters for GSCV and RSCV to search is shown in the table below.

Random Forest Hyperparameter Optimizer	Grid Search Cross Validation	Random Search Cross Validation
n_estimators	[1,2,4,8,16,32]	[1, 2, 4, 8, 16, 32, 64, 128, 500, 1000]
max_depth	[None,2,4,6,8]	[None, 1, 2,3, 4, 5, 6, 7, 8]
max_features	['auto', 'sqrt', 'log2']	['auto', 'sqrt', 'log2']
min_samples_split	[6, 8, 10, 12]	[6, 8, 10, 12]
Min_samples_leaf	[2, 4, 6, 8, 10]	[2, 4, 6, 8, 10]
bootstrap	[True, False]	[True, False]

Neural Network: From our experiment on the two baseline methods mentioned above, we find that gender features have a neglectable effect on prediction. With ethical consideration, we remove this feature from the neural network training. Thus, three sub-dataset are used to train three neural networks. Those sub-dataset are normalized before training. We use torch.nn.Linear as basic layer structure, F.relu as activation function since it is the newest activation function currently. With a multi-classification problem, we use cross-entropy as the loss function. The hyperparameters we modify are: first, the number of layers, we try from 3, 4, and 5 layers for each network; and also the input-output numbers for intermediate layers from 8, 16 to 32. Second, we also try different optimizers: Adadelta, Adagrad, Adam, and AdamW. Those hyperparameters generate very big differences in the final result. Thus, they are tuned first to get a reasonable high accuracy. Then, we tune the learning rate and the number of epochs to slightly increase accuracy and achieve the highest accuracy.

Convolutional Neural Network: We first train a convolutional neural network by using the complete MRI image. The original MRI image is 3-dimensional with [121, 145, 121]. We tune the learning rate (0.006 to 0.001) and optimizer (Adam and SGD), and we also train using different types and numbers of layers to find the optimal structure, which contains the convolutional sequence contains three groups of a convolutional 3d layer with kernel = 3 and padding = 1, a BatchNorm3d layer, LeakyReLU() activation function, and MaxPool3d layer. The dropout layer and fully connected linear layer are connected in the end. We maintain the same structure of neural networks, but with transformed image size [30,40,30], the running time can reduce from 8 minutes to 8 seconds without significantly dropping in accuracy.

Experimental Results

Questions: we aim to answer the following questions: (1) Which features should be used to predict AD? (2) Which kind of algorithm should be used to build up a prediction model? (3) What structure and hyperparameters should this model use? (4) How should users set up experiments and use our model?

All the accuracy is calculated based on a five-time experiment with a standard deviation marked as $\pm$ SD.

Decision Tree:

Max-depth		1	2	3	4	5	6	7	8	9
All features	gender	0.89 ± 0	0.89 ± 0	0.88 ± 0	0.84 ± 0.007	0.84 ± 0.015	0.82 ± 0.020	0.81 ± 0.013	0.82 ± 0.007	0.82 ± 0.020
	No gender	0.89 ± 0	0.89 ± 0	0.89 ± 0.007	0.85 ± 0.023	0.82 ± 0.007	0.83 ± 0.020	0.82 ± 0.020	0.83 ± 0.015	NA
Only survey information	gender	0.80 ± 0	0.78 ± 0.028	0.71 ± 0.005	0.77 ± 0.002	0.82 ± 0.011	NA	NA	NA	NA
	No gender	0.79 ± 0.005	0.79 ± 0.011	0.78 ± 0.004	0.76 ± 0.003	NA	NA	NA	NA	NA
Only MRI feature	No gender	0.91 ± 0.005	0.89 ± 0.002	0.89 ± 0.006	0.84 ± 0.006	NA	NA	NA	NA	NA

Random Forest:

		Grid Search	Random Search
All Features	gender	0.91 ± 0.01	0.93 ± 0.02
	no gender	0.96 ± 0.04	0.92 ± 0.02

Only Survey Information	Gender	0.83 ± 0.10	0.83 ± 0.03
	No gender	0.84 ± 0.08	0.37 ± 0.03
Only MRI feature	No gender	0.96 ± 0.01	0.96 ± 0.01

Neural Network: Since there are pretty many combinations of different hyperparameters, due to limiting space, here only presents the best result of a different number of layers. All the models below use Adagrad optimizer since it turns out to be the best one for all the models. The learning rate and the number of epochs are tuned to achieve the highest accuracy under different conditions.

	All features	Only survey information	Only MRI feature
3 Layers	0.90 ± 0.014	0.62 ± 0.045	0.92 ± 0.004
4 Layers	0.92 ± 0.004	0.56 ± 0.015	0.92 ± 0.003
5 Layers	0.88 ± 0.023	0.54 ± 0.004	0.90 ± 0.020

Convolutional Neural Network:

MRI image	Hippocampus image	Optimizer	Learning rate	Batch size	Accuracy	Time duration
Y		Adam	10**-4	4	0.8625	8m52s
	Y	Adam	10**-4	4	0.8253	9s
	Y	Adam	6**-4	4	0.8542	7s
	Y	SGD	6**-4	4	0.7412	10s
Y		Adam	6**-4	4	0.8486	9m33s

Conclusions and Future Work

From the experiment result, we can conclude that, first, MRI information is more useful on prediction than survey information, and gender features are fine to be removed from this process. Second, the baseline method, random forest, is better than DNN and CNN in this case. The potential reason could be that this prediction is not very complex, and is more suitable for basic ML algorithms. As we know, DNN and CNN typically aim for very complex situations. For future application in diagnosing, doctors may remove the survey process, and take MRI images for a patient. Then by applying the random forest model, it can predict if the patient will convert to AD with pretty high accuracy.

Ethical Considerations and Broader Impacts

All the features we use in this project are not sensitive, except the gender feature. And we prove that removing gender features has no impact on predicting accuracy. Thus, the final model can get rid of gender features. Other survey features are basic personal information, and MRI features are a type of measurement result, which have no worrying related ethical issues. With our project, doctors will have a basic method to use for diagnosing AD, or develop a more advancing method based on our project. If AD can be easily predicted by ML technology, a lot of people can benefit from that.

Prior Work / References / Projects that inspired us:

1. M. Tanveer, B. Richhariya, R. U. Khan, A. H. Rashid, P. Khanna, “Machine Learning Techniques for the Diagnosis of Alzheimer’s Disease: A Review”, Publication: ACM Transactions on Multimedia Computing, Communications, and Applications, April 2020
2. Grassi. Massimilianoa, Perna. Giampaoloa, Caldirola. Danielaa, Schruers. Koenb, Duara. Ranjane, Loewenstein. David, “A Clinically-Translatable Machine Learning Algorithm for the Prediction of Alzheimer’s Disease Conversion in Individuals with Mild and Pre Mild Cognitive Impairment”, Journal of Alzheimer's Disease, vol. 61, no. 4, pp. 1555-1573, 2018
3. Elaheh Moradia, Antonietta Pepeb, Christian Gaserc, Heikki Huttunena, Jussi Tohka, “Machine learning framework for early MRI-based Alzheimer's conversion prediction in MCI subjects”, NeuroImage, Volume 104, 1 January 2015
4. Massimiliano Grassi, Nadine Rouleaux, Daniela Caldirola, David Loewenstein, Koen Schruers, Giampaolo Perna, Michel Dumontier, “A Novel Ensemble-Based Machine Learning Algorithm to Predict the Conversion From Mild Cognitive Impairment to Alzheimer's Disease Using Socio-Demographic Characteristics, Clinical Information, and Neuropsychological Measures”, Frontiers in Neurology, 16 July 2019
5. Lee Kuok Leong, Azian Azamimi Abdullah, “Prediction of Alzheimer’s disease (AD) Using Machine Learning Techniques with Boruta Algorithm as Feature Selection Method”, 2019 Journal of Physics: Conference Series 1372 012065