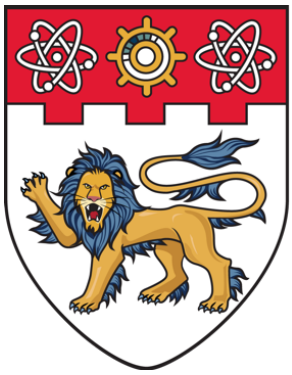

NANYANG TECHNOLOGICAL UNIVERSITY
SINGAPORE

BMDSIS Project title:

NOVEL ALGORITHMS FOR PREDICTING MENTAL HEALTH STATUS FROM SPEECH AND BEHAVIOR

Han Wenhao Han Wenhao

Goh Wen Bin Wilson Wilson goh

SCHOOL OF BIOLOGICAL SCIENCES

2023

Table of Contents

INTRODUCTION

Dementia is a general term for the impaired ability to remember, think, or make decisions that interfere with everyday activities. Alzheimer's disease(AD) is the most common type of dementia, accounting for 60% - 70% of all diagnoses[1]. AD usually starts slowly and progressively worsens, even though it mostly affects older adults, it is not a part of normal aging. Identification the AD in the early stage is important for dementia treatment and prevention.

The cause of Alzheimer's disease is poorly understood. There are many environmental and genetic risk factors associated with its development. The strongest genetic risk factor is from an allele of APOE[2][3]. Other risk factors include a history of head injury, clinical depression, and high blood pressure[4]. The conventional diagnostic methods are based on the history of the illness and cognitive testing with medical imaging and blood tests[5]. Initial diagnosis is often mistaken for normal brain aging[6] and requires collecting a large amount of data over a long diagnostic period with a possibility of misdiagnosis.

The rapid development of artificial intelligence(AI) technologies has enabled efficient and accurate early prediction of AD and dementia. The aim of this project is to develop a novel pipeline to mine the significant features in clinical testing datasets and blood cell gene expression datasets, and train AI models for predicting mental health status. There are two main research directions. The first is developing a method based on a hand-drawing task developed by CUHK counterparts, evaluating if performance can be enhanced by combining it with clinical Montreal Cognitive Assessment(MoCA) tasks to determine which part is meaningful for mental health status prediction.

The other is the modeling of dementia using blood-based gene expression data from ADNI[7] and ANMerge[8]. Developing a novel pipeline to remove the batch effect, significant gene selection, and prediction modeling. This project can lead to new digital health tools that can help in the early prediction of mental health and also cognitive disorders such as dementia and Alzheimer's.

MATERIALS AND METHODS

- **Materials**

MoCA is a widely used screening assessment for detecting cognitive impairment[9]. The test consists of 30 points and takes part in 10 minutes for the individual. The basics of this test include short-term memory, executable performance, attention, focus, and more. As Figure 1 shows, the test is performed in seven steps in total, it provides summary statistics rather than in-depth analysis.

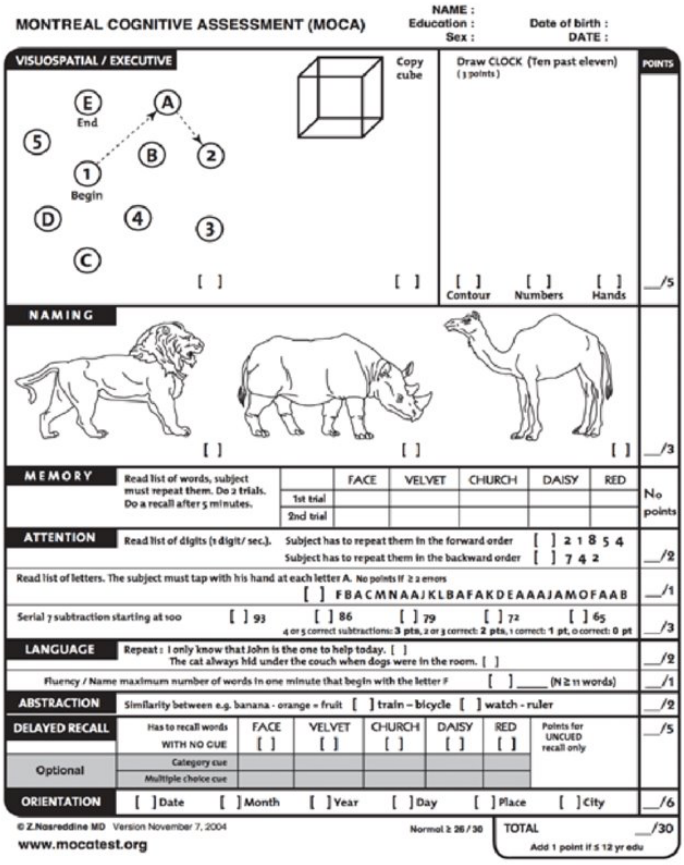

Figure 1. Montreal Cognitive Assesment (MoCA)

Drawing requires visuospatial skills, sustained attention, and executive function, dementia subjects tend to have longer thinking time and irregular drawing patterns which makes drawing a good material for dementia diagnosis. The hand-drawing features are generated based on a drawing test, the participants are asked to copy the given figure as similar as possible. Then, performing feature engineering to capture motion features in both pixel, stroke level, and pentagon levels like finger movement, time taken to draw each line, etc. As Figure 2 shows, adding the meta-features and genomic features to the features set makes 38 features in total.

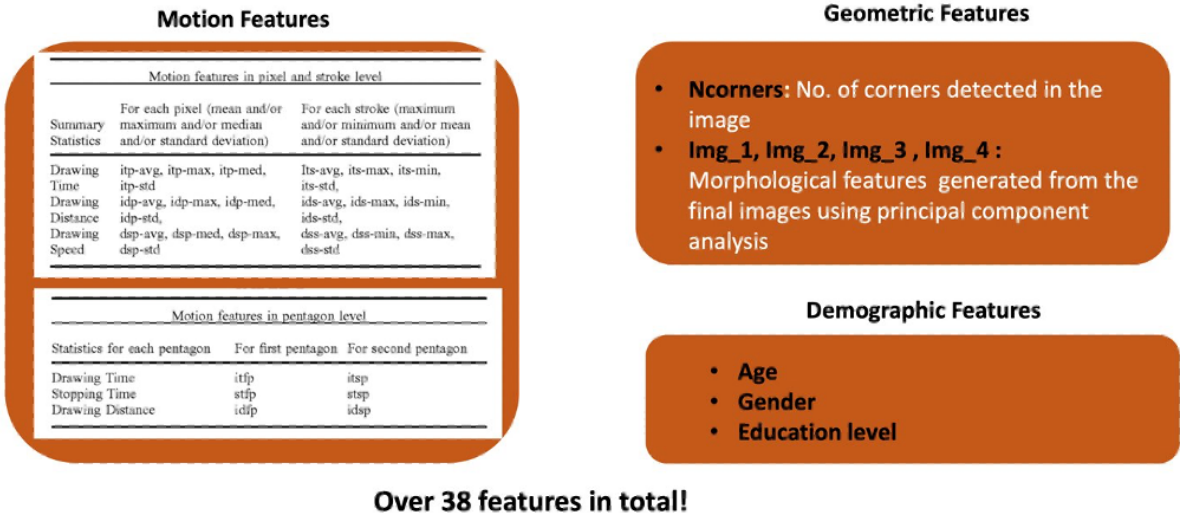

Figure 2. Hand-drawing feaatures

The ADNI is a multisite study that aims to improve clinical trials for the prevention and treatment of AD. Researchers at 63 sites in the US and Canada track the progression of AD in the human brain with neuroimaging, biochemical, and genetic biological markers[9][10]. To date, over 1000 scientific publications have used ADNI data and several other initiatives related to AD and other diseases have been designed and implemented using ADNI as a model[11][12], most of which apply ML/DL algorithms to predict the AD using MRI data. ADNI studies have shown that people who carry an APOE ε4 allele are at high risk for AD[13], we considered unlocking the prognostic potential of blood-based genetic data using local modeling based on the value of APOE ε4.

ANMerge is a comprehensive and accessible AD patient-level dataset, which is considered a new, improved, and updated version of AddNeuroMed. Although the shared AddNeuroMed involves more than 1,700 participants, it has only been cited about 65 times. In contrast, ADNI, which involves roughly 2,400 individuals, was cited more than 1,300 times. Compared to the impact ADNI has had on recent research activities, it seems AddNeuroMed has not reached its full potential.

The ANMerge dataset comprises four data modality-specific subtables, genotype data in PLINK format, and one combined table providing all preprocessed information as one. Respectively, one subtable was created for clinical data, proteo-mics, FreeSurfer calculated MRI features, and gene expression values[13]. The gene expression data in ANMerge is going to be applied to evaluate the generalization ability of the pipeline.

In total, the dataset comprises information on 1,702 patients, out of which 773, 665, and 264 originated from the AddNeuroMed, DCR, and ART cohorts, respectively (Table 1). Data on 4,585 individual participant visits are reported. At the study baseline, 512 participants had been diagnosed with AD, 397 with MCI, and 793 were non-cognitively impaired individuals. Table 1 describes the average characteristics of each diagnosis group at baseline.

Table 1 Summary statistics describing the ANMerge dataset at baseline[13]

Diagnosis	N	ANM	DCR	ART	Age (SD)	Female %	Education (SD)	*APOE ∈4* positive %
CTL	793	266	423	104	74.5 (6.4)	59	12.3 (4.3)	25
MCI	397	247	89	61	76.0 (6.5)	55	10.0 (4.3)	40
AD	512	260	153	99	78.6 (7.2)	63	9.4 (4.3)	54
Total	1702	773	665	264	76.4 (6.9)	59	10.9 (4.5)	39

- **Methods**

The MoCA dataset and hand-drawing datasets have been well-preprocessed. To explore the prognostic potential of the hand-drawing features, I helped to evaluate different ML models on the datasets with hyperparameter tuning, the model has been applied including decision tree, random forest, SVC with linear and rbf kernel, KNN, and Naïve Bayes, in the meantime, tried to dig out features playing a decisive role in prediction based on the feature importance. Finally, evaluate if the performance can be enhanced by combining the MoCA features with it.

Working on the hand drawing and MoCA features familiar me with the commonly used supervised learning classifier, the main procedure of hyperparameter tuning, and feature importance analysis. All experience gained was applied in the research of the blood-based gene expression dataset.

- **Main pipeline**

The blood-based gene expression dataset is applied to predict the conversion from Mild Cognitive impairment (MCI) to AD, and a novel pipeline including data cleaning, feature selection, modeling training, and evaluating is developed. The patients whose initial diagnosis is MCI and the final diagnosis is AD are labeled as the converter, and those whose final diagnosis remains MCI are labeled as non-converters. Figure 3 shows the main procedure of the pipeline and the methods applied in each step.

The pipeline starts with data checking, the aim of this step is to understand the data, deal with the null values, and remove the batch effect. Then, filtering the insignificant gene based on IQR, PCA variance, and quality control. Before modeling training, the Kruskal-Wallis test, Boruta, and Cox proportional hazard test are applied to feature selection. Finally, the significant genes are applied for the local modeling with nested cross-validation, and the oversampling strategies are applied in the hyperparameter.

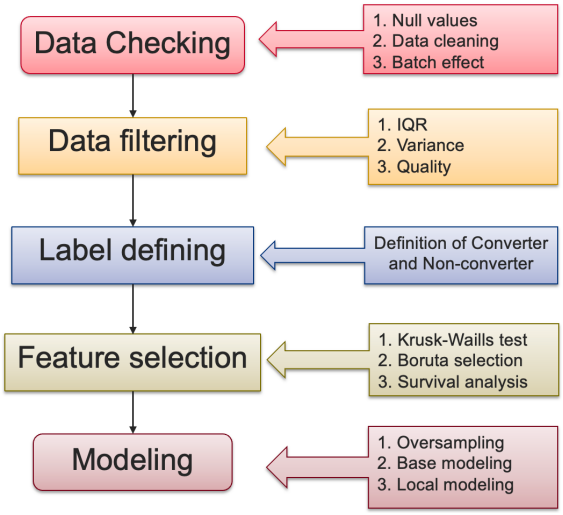

Figure 3. The main procedure and methods of the pipeline

- **Data preprocessing**

Data cleaning and batch effect removal are applied in the project.

Data cleaning is the process of identifying and correcting or removing errors, inconsistencies, and other issues in a dataset, which can help to improve the accuracy, completeness, and reliability of the data. The main steps of data cleaning:

1) Remove duplicates: Identify and remove any duplicate records in the dataset.
2) Handle missing value: Identify and handle the missing value.
3) Checking the outliers: Identify the outliers in the data that may be due to measurement errors, data entry errors, or other issues using the box plot or violin plot. Decided how to handle these outliers, which may involve removal or transformation.

   Interquartile Range (IQR) filtering is a pre-processing technique used to remove outliers from the dataset. The interquartile range is calculated as the difference between the third quartile (Q3) and the first quartile (Q1) of the data. To perform IQR filtering, one first calculates the interquartile range of the data. Any data point that falls outside of the range [Q1 - 1.5 * IQR, Q3 + 1.5 * IQR] is considered an outlier and can be removed from the dataset. This range is often referred to as the "IQR fence."

   It should be noted that IQR filtering can also remove valid data points that are not outliers, so it should be used with caution and in combination with data pre-processing techniques.
4) Handle inconsistencies: Identify any inconsistencies or errors in the data, such as conflicting values or unrealistic values. Decide how to handle these inconsistencies, which may involve manual correction or imputation.
5) Check data type: Check if the types of each variable in the dataset are appropriate and consistent with their intended use.
6) Standardize data: Standardize data formats and units, such as converting dates to a consistent format or converting units of measurement to a consistent scale.
7) Check for data quality: Check the overall quality of the data, including its completeness, accuracy, and reliability.

Missing values can have a significant impact on the accuracy and validity of statistical analyses and ML models. The common approaches to dealing with missing values in the dataset are listed below:

1) Delete the missing value. This approach will simply remove the observation with missing values. This approach will lead to information loss and introduce bias if the missing values are not missing completely at random.

2) Impute the missing value. Another approach is to estimate the missing values based on the available data. Commonly used imputation methods include mean imputation, median imputation, regression imputation, or multiple imputation.
3) Treat the missing values as a separate category. If the missing values represent a particular group, we may include a dummy variable to capture this information.
4) Using algorithms that can handle missing values. Such as decision trees, random forests, and deep learning models, can handle missing values without the need of imputation which use available data to make decisions.

It is important to carefully consider the result of the missing data, the choice of the method should be based on the nature of the data and the aim of the analysis.

PCA can be applied to identify batch effects in high-dimensional data by examining the clustering of samples in the first few principal components. If there is a batch effect present, samples from the same batch will tend to cluster together in the plot. To confirm that the observed clustering is due to batch effects and not some other confounding variable, we can examine the loadings of the principal components. Loadings indicate the contribution of each variable (e.g. gene expression) to the principal component. If a batch effect is present, the loadings of some variables are stronger in certain principal components, indicating that the variation in those variables is driving the batch effect.

If the batch effect is identified, a few ways can be applied to correct it: ComBat, a popular algorithm to correct for batch effects by adjusting the data based on the mean and variance of each batch; RUV (Remove unwanted variation), an algorithm that identifies and remove unwanted variation in high-dimensional data; Surrogate Variable Analysis (SVA), a statistical method that estimates the batch effect and other sources of variation in the data and then adjusts for them; Batch effect correction algorithms (BECAs). Persistent batch confounding effects are difficult to detect and remove.

RNA quantity (Figure 4) is one of the technical confounders in the data set (Figure 4).

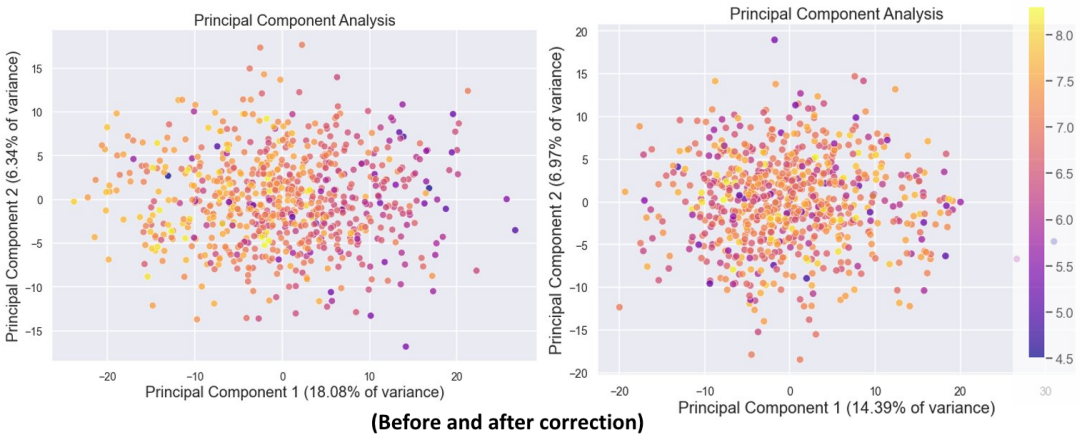

Figure 4. PCA plot of RNA integrity before and after correction

Figure 5 displays another batch effect factor – microarray plate number. According to the plot, the samples with the same plate number are grouped with each other which indicates the plate number is a factor of batch effect. After correction, the sample with different plate numbers overlaps with each other indicating the batch effect has been removed.

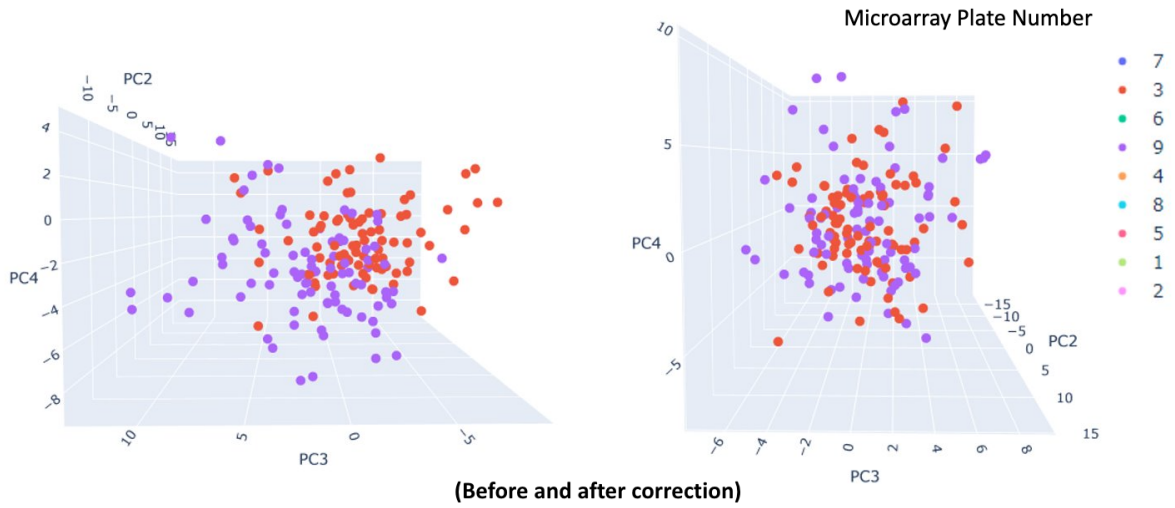

Figure 5. PCA plot for microarray plate number before and after correction

- **Survival analysis and Boruta feature selection**

To dig out the significant genes in the dataset for the prediction, Cox proportional hazard regression and Boruta were applied in the project for the feature selection.

Survival models related the time that passes, before some event occurs, to one or more covariates that may be associated with that quantity of time. In a proportional hazards model, the unique effect of a unit increase in a covariate is multiplicative with respect to the hazard rate[14].

Cox proportional hazards regression can also be used for feature selection. The main steps to apply Cox proportional hazards regression for feature selection:

1) Collect and preprocess the data.
2) Split the data into training and validation sets and use the training set to fit the model.
3) Calculate the importance of features using a technique such as the Wald test or likelihood ratio test and rank the features based on their importance.
4) Select the top-rank features to use in the predictive model and validate the model using the validation set. Refine the model by adding or removing features as needed.

It should be noted that Cox proportional hazards regression assumes that the hazard ratio(HR) is constant over time, which may not be the case for all datasets. Besides, it is important to consider other factors such as the clinical relevance of the features when selecting features for predictive models.

Boruta is designed to select relevant features from a high-dimensional dataset. It is particularly useful when dealing with datasets with a large number of features, many of which may be irrelevant or redundant.

The Boruta algorithm works by comparing the importance of each feature to the importance of a set of 'shadow' features, which are created by permuting the values of the original features. If a feature is more important than its corresponding shadow feature, it is considered 'confirmed' as relevant. If a feature is not more important than its corresponding shadow feature, it is considered 'unconfirmed'. Features that are not confirmed after a certain number of iterations are considered 'rejected'.

It should be noted that the Boruta algorithm can be computationally expensive, especially for datasets with a large number of features. Besides, the Boruta may not work well for datasets with highly correlated features or for datasets where the relationship between the features and the outcome is non-linear.

- **Hyperparameter tuning with oversampling strategy**

Random forest and support vector machine are two supervised learning methods applied for the prediction. To find the optimal model, hyperparameter tuning was performed on both two models. The oversampling strategy was applied in the model training due to the small sample size.

The most important hyperparameters in the random forest:

1) The number of estimators: The number of decision trees in the random forest. Increasing the number of trees can improve the performance of the random forest, but it will increase the training time and the risk of overfitting.
2) Maximum depth: The maximum depth of each decision tree in the random forest. A deeper tree can model more complex relationships between features, but it also increases the risk of overfitting.
3) The minimum number of samples required to split an internal node. Increasing this parameter can help reduce overfitting, but it may also lead to underfitting.
4) Minimum samples leaf: The minimum number of samples required to be at a leaf node. Similar to the minimum samples split parameter, increasing this parameter can help reduce overfitting, but it may also lead to underfitting.
5) Maximum features: The maximum number of features to consider when splitting a node. This parameter can be used to control the complexity of the decision trees and prevent them from focusing too much on a small set of features.
6) Bootstrap samples: The number of samples to draw with replacement when building each decision tree. This parameter can be used to control the diversity of the decision trees in the random forest.

The number of estimators and maximum depth was applied for the tunning to get the optimal model in the project. Support vector classifier(SVC) is a popular supervised learning algorithm used for classification and regression analysis. It works by finding the best hyperplane that separates the data into different classes. The common hyperparameter of SVC are listed below:

1) C: This parameter controls the trade-off between maximizing the margin and minimizing the classification error. A smaller value of C will result in a wider margin, but more misclassifications, while a larger value of C will result in a narrower margin, but fewer misclassifications. This parameter is also known as the regularization parameter.
2) kernel: This parameter specifies the kernel function used to transform the input data into a higher-dimensional feature space. Common choices include linear, polynomial, radial basis function (RBF), and sigmoid kernels.
3) gamma: This parameter controls the shape of the kernel function and the degree of influence of each training example. A smaller value of gamma will result in a smoother decision boundary, while a larger value of gamma will result in a more complex decision boundary.
4) degree: This parameter is used for the polynomial kernel function and specifies the degree of the polynomial.

In the project, the SVC with linear and RBF kernel is applied for C and gamma tunning to get the optimal SVC. Due to the small sample size and imbalanced class label, the oversampling strategy is applied in model training SMOTE (Synthetic Minority Over-sampling Technique), Borderline SMOTE, and ADASYN (Adaptive Synthetic Sampling) are three popular techniques used for oversampling imbalanced datasets in machine learning.

1) SMOTE: SMOTE works by creating synthetic examples of the minority class by interpolating between neighboring examples. Specifically, for each minority class example, SMOTE selects k nearest neighbors and generates new examples by linearly interpolating between the original example and one of its k neighbors.
2) Borderline SMOTE: Borderline SMOTE is an extension of SMOTE that focuses on examples near the decision boundary. Specifically, it only generates synthetic examples for minority class examples that are misclassified or near the decision boundary. This can improve the robustness and generalization of the classifier.
3) ADASYN: ADASYN is another extension of SMOTE that adaptively generates synthetic examples based on the density distribution of each class. Specifically, it generates more synthetic examples for minority class examples that are in areas

of the feature space with low density, and fewer synthetic examples for minority class examples that are in areas with high density.

The optimal choice of technique depends on the specific problem and dataset. For example, SMOTE may work well when the decision boundary between classes is well-defined, while Borderline SMOTE may work better when the decision boundary is more ambiguous. ADASYN may work well when there is a large imbalance between classes and the density distribution of each class is highly variable.

The usage of oversampling should be careful, it may cause overfitting if not applied properly. The oversampling is performed on the training set in iteration to avoid biased or unrealistic results in the project.

- **Local Modeling with nested cross-validation evaluation**

Local modeling is a technique used in ML data analysis to build separate models for different subpopulations or local regions within a large dataset. In many cases, the underlying data distribution may be highly complex or heterogeneous, and a single global model may not be able to capture all the nuances and variability of the data. Local models can capture the specific patterns and relationships that are unique to each subplot or region and can lead to more accurate and interpretable predictions.

Nested cross-validation is a technique used in ML to evaluate the performance of a model and select the best hyperparameters of feature selection strategy.

The basic idea behind nested cross-validation is to use an outer loop and an inner loop to split the data into training and testing sets. The outer loop is used to evaluate the performance of the model on a held-out test set, while the inner loop is used to select the best hyperparameters or feature selection strategy using a separate validation set.

The outer loop is typically a k-fold cross-validation, where the data is divided into k non-overlapping folds, and each fold is used once as a test set while the remaining k-1 folds are used for training. This process is repeated k times, with each fold used once as the test set, and the performance of the model is averaged over the k iterations.

Within each iteration of the outer loop, the inner loop is used to select the best hyperparameters or feature selection strategy. This is typically done using another k-fold cross-validation, where the training set is further divided into k non-overlapping folds, and each fold is used once as a validation set while the remaining k-1 folds are used for training. This process is repeated k times, with each fold used once as the validation set, and the hyperparameters or feature selection strategy that results in the best performance on the validation set is selected.

The local modeling with nested cross-validation is applied in the project to select the optimal hyperparameter of feature selection and evaluate the performance of the pipeline. Figure 6 displays the process of local modeling and nest cross-validation in the project. The sample set is divided into 3 local groups based on the value of APOE ε4, the samples are split into five nonoverlapping folds in each group. In each iteration, four folds are combined together as the training set to carry out feature selection and hyperparameter tuning, the other fold is used to evaluate the optimal model. Finally, the mean of scores for five iterations will be used to evaluate the whole pipeline.

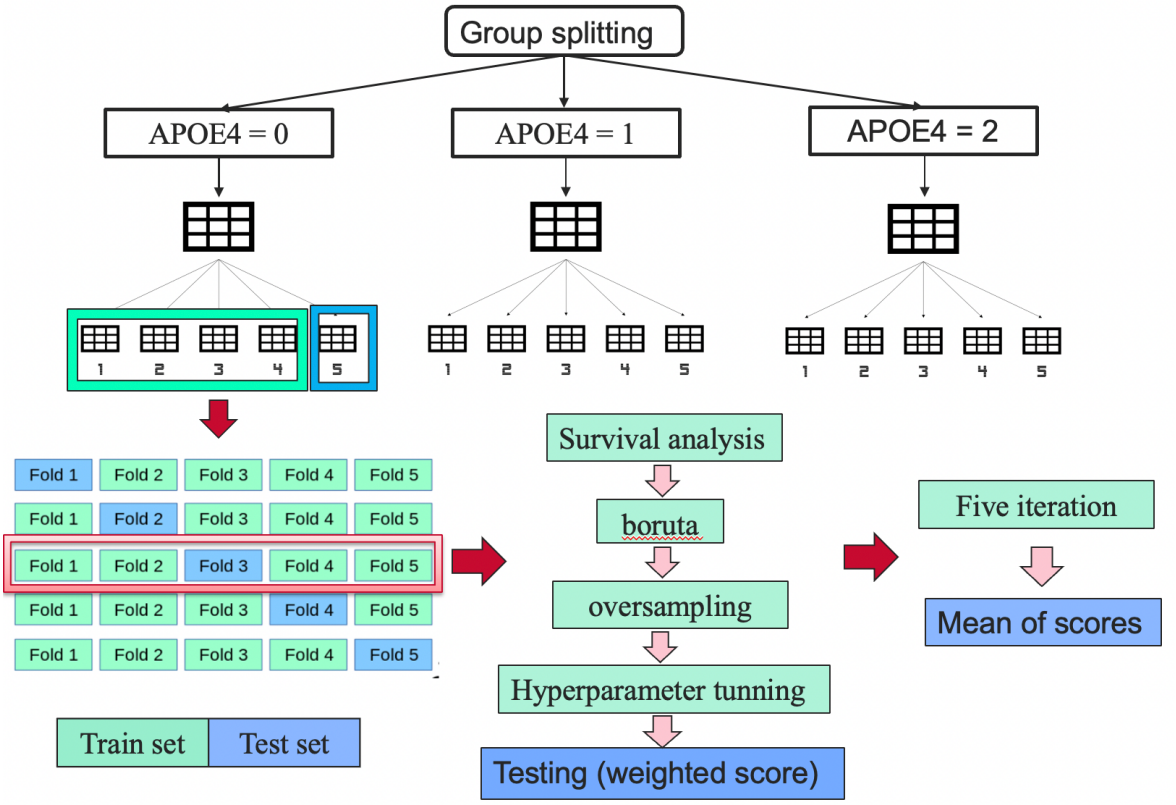

Figure 6. the process of local modeling and nest cross-validation

Six evaluation metrics are applied for evaluation including accuracy, balanced accuracy, precision, recall, f1-score, and AUC. the final weighted scores are the integration of three local modeling groups based on their group size.

RESULTS AND DISCUSSION

- **Results**
  - **Dementia patient prediction on hand-drawing dataset**

Table 2. Dataset Stratified splitting

label	0	1	percentage
train	821	297	70%
test	353	127	30%

Table 3. Comparison of model performance on all features and important features

	Accuracy		precision		recall		specificity		f1-score	
	all	important	all	important	all	important	all	important	all	important
Decision Tree	0.7	0.65	0.91	0.91	0.64	0.57	0.83	0.85	0.75	0.7
Random Forest	0.64	0.66	0.87	0.89	0.6	0.6	0.77	0.81	0.71	0.72
KNN	0.73	0.73	0.76	0.74	0.9	**0.95**	0.28	0.16	**0.83**	**0.83**
Naive Bayes	**0.74**	**0.74**	0.83	0.82	0.81	0.82	0.56	0.52	0.82	0.82
SVM	0.62	0.59	0.91	**0.92**	0.52	0.47	0.87	**0.9**	0.66	0.62

○ Blood-based gene expression dataset from ADNI dataset

Table 4. (a) Distribution of class label (training set)

label	Fold 0	Fold 1	Fold 2	Fold3	Fold 4	percentage
0	187	187	187	187	188	71%
1	76	76	76	76	76	29%

Table 4. (b) Distribution of class label (testing set)

label	Fold 0	Fold 1	Fold 2	Fold3	Fold 4	percentage
0	47	47	47	47	46	71%
1	19	19	19	19	19	29%

Table 5. Base modeling VS. local modleing (with data leakage)

Modeling	Classifier	Accuracy	Balanced accuracy	precision	recall	F1-score	AUC
Base modeling	**RF**	0.87	0.90	0.85	0.87	0.95	0.87
	SVC (rbf)	0.90	0.86	**0.95**	0.90	**0.97**	0.90
	SVC (linear)	**0.92**	**0.91**	0.92	**0.92**	**0.97**	**0.92**
Local modeling	**RF**	0.89	0.90	0.87	0.88	0.95	0.89
	SVC (rbf)	0.90	0.88	0.92	0.90	**0.97**	0.90
	SVC (linear)	0.88	0.85	0.89	0.87	**0.97**	0.88

Table 6. The performance of prediction for local modeling (without data leakage)

Overampling	Classifier	Accuracy	Balanced accuracy	precision	recall	F1-score	AUC
SMOTE	**RF**	0.66	0.5	0.31	0.2	0.21	0.49
	SVC (rbf)	0.63	0.5	0.3	0.27	**0.27**	0.48
	SVC (linear)	0.63	0.49	0.29	0.25	0.25	0.48
ADASYN	**RF**	**0.69**	**0.52**	**0.34**	0.22	0.24	0.48
	SVC (rbf)	0.56	0.49	0.22	**0.39**	0.26	0.43
	SVC (linear)	0.65	0.5	0.3	0.24	0.25	**0.51**

- **Discussion**

Table 2 compares the performance of dementia prediction between using all features and important features selected based on permutation importance in the hand-drawing dataset. Although adding the class weight to reduce the effect of class imbalance, the models are unable to correctly predict control normal and dementia simultaneously. For example, the KNN model can identify most of the control normals but can hardly identify dementia while SVC can well predict dementia, but can only identify half of the healthy individuals. Even though the performance of the prediction didn't reach satisfactory scores, the high recall of some models does reveal the prognostic potential of the hand-drawing dataset. Applying oversampling strategy on the training set might be a way to improve the performance prediction. The result of the combination of MoCA and hand drawing features is not displayed here, but there has been enough evidence to show that the combination will enhance the performance of prediction not only for dementia but also patients at risk.

Table 5 compares the prediction scores of both base modeling and local modeling with data leakage in the process. The survival analysis and Boruta feature selection are performed before the dataset is spitted into three local groups, in the meantime, the oversampling strategy is performed on the testing set which no doubt leads to wrong positive scores.

By contrast, Table 6 displays the result without data leakage in the process. Sadly, the preprocessing pipeline and local modeling seems not working on the ADNI blood-based gene expression dataset. But the result in the table are not reliable enough for two reasons: The first is that definition of the converter and non-converter doesn't make sense because some non-converters haven't got their final diagnosis yet when the study ends, if we simply assign them as non-converters and use this label to do model training will introduce the wrong information to the model; The other reason is that the sample size is too small, for example, there are only 6 testing samples in the local group with APOE ε4 = 2.

Data leakage in machine learning can occur when information that should not be available to the model is inadvertently included in the training data or during the evaluation of the model's performance. This can result in overfitting, where the model performs well on the training data but poorly on new data, or inaccuracy, where the model is biased towards certain features or variables.

Here are some measurements that could lead to data leakage in machine learning:

1) Including target variables in the training data: If the target variable (the variable the model is trying to predict) is included in the training data, the model can simply memorize the target variable instead of learning the underlying patterns in the data.
2) Including derived or calculated variables: If derived or calculated variables are included in the training data, the model can accidentally learn the relationship between the derived variable and the target variable.
3) Using data from the future: If the model is trained on data that includes information from the future (e.g., stock prices for tomorrow), it will not be able to accurately predict outcomes in the real world.

To prevent data leakage in machine learning, it is important to carefully clean and preprocess the data, remove any variables that may leak information about the target variable, and properly split the data into training and testing sets. Additionally, it's essential to carefully evaluate the performance of the model on new data to ensure it is generalizing well to unseen examples. Besides, when applying statistical-based or ML-based feature selection methods, it should be noted that all methods can be only applied to the training set, the training and testing splitting should be carried out ahead of feature selection. Last but not least, it should be very careful when doing local modeling on the dataset, all methods should be performed on each local group respectively or the information of different subsets will be shared with each other.

Another thing that should be taken care of is that oversampling should not be applied to the test set. The test set should be a representative sample of real-world data and should not be manipulated in any way to avoid introducing bias or overfitting. Using an oversampled test set can lead to overly optimistic results and may not reflect the performance of the classifier in the real world.

When evaluating the performance of the classifier on the test set, it is important to use the original, unbalanced test set to obtain an accurate estimate of the performance on new, unseen data. Oversampling is typically applied only to the training set to balance the class distribution and improve the performance of the classifier.

CONCLUSION

The hand-drawing features possess the prognostic potential for dementia and performance can be enhanced by combining it with clinical MoCA tasks.

The ADNI blood-based gene expression data may not be informative enough as a predictive biomarker for AD. A novel pipeline consisting of preprocessing, features selection, and local modeling for AD prediction is developed, and its prognostic potential still waiting to be explored in ANMerge and other datasets.

REFERENCE

[1] "Dementia Fact sheet". World Health Organization. September 2020.

[2] Long JM, Holtzman DM (October 2019). "Alzheimer Disease: An Update on Pathobiology and Treatment Strategies". Cell. **179** (2): 312–339.

[3] "Study reveals how APOE4 gene may increase risk for dementia". National Institute on Aging. Retrieved 17 March 2021.

[4] Knopman DS, Amieva H, Petersen RC, et al. (May 2021). "Alzheimer disease". doi:10.1038/s41572-021-00269-y. PMC 8574196. PMID 33986301

[5] "Dementia diagnosis and assessment" (PDF). National Institute for Health and Care Excellence (NICE). Archived from the original (PDF) on 5 December 2014. Retrieved 30 November 2014.

[6] Burns A, Iliffe S (February 2009). "Alzheimer's disease". *BMJ*. **338**: b158. doi:10.1136/bmj.b158. PMID 19196745. S2CID 8570146.

[7] ADNI. https://adni.loni.usc.edu/

[8] Birkenbihl C, Westwood S, Shi L, et al. ANMerge: a comprehensive and accessible Alzheimer's disease patient-level dataset[J]. Journal of Alzheimer's Disease, 2021, 79(1): 423-431.

[9] Weiner, Michael W.; Aisen, Paul S.; Jack, Clifford R.; Jagust, William J.; Trojanowski, John Q.; Shaw, Leslie; Saykin, Andrew J.; Morris, John C.; Cairns, Nigel (2010-05-01). "The Alzheimer's disease neuroimaging initiative: progress report and future plans". *Alzheimer's & Dementia: The Journal of the Alzheimer's Association*. **6** (3): 202–211.e7.

[10] Jones-Davis, Dorothy M.; Buckholtz, Neil (2015-07-01). "The impact of the Alzheimer's Disease Neuroimaging Initiative 2: What role do public-private partnerships have in pushing the boundaries of clinical and basic science research on Alzheimer's disease?". *Alzheimer's & Dementia: The Journal of the Alzheimer's Association*. **11** (7): 860–864.

[11] Weiner, Michael W.; Veitch, Dallas P.; Aisen, Paul S.; Beckett, Laurel A.; Cairns, Nigel J.; Cedarbaum, Jesse; Green, Robert C.; Harvey, Danielle; Jack, Clifford R. (2015-06-01). "2014 Update of the Alzheimer's Disease Neuroimaging Initiative: A review of papers published since its inception". Alzheimer's & Dementia: The Journal of the Alzheimer's Association. **11** (6): e1–120. doi:10.1016/j.jalz.2014.11.001. ISSN 1552-5279.

[12] Weiner, Michael W.; Veitch, Dallas P.; Aisen, Paul S.; Beckett, Laurel A.; Cairns, Nigel J.; Cedarbaum, Jesse; Donohue, Michael C.; Green, Robert C.; Harvey, Danielle (2015-07-01). "Impact of the Alzheimer's Disease Neuroimaging Initiative, 2004 to 2014". *Alzheimer's & Dementia: The Journal of the Alzheimer's Association*. **11** (7): 865–884.

[13] Weiner, Michael (2017). "Recent publications from the Alzheimer's disease neuroimaging initiative: reviewing progress toward improved AD clinical trials". *Alzheimer'*

s & Dementia. **13** (5): 561–571. doi:10.1016/j.jalz.2016.10.006. PMC 5536850. PMID 27931796.

[14] Proportional hazrds model, Wikipedia.