ISyE 6740 - Summer 2021
Final Report

Team Member Names: Linh Dinh (GTID: 903651327) & Phuong Nguyen (GTID: 903650790)

Project Title: Predicting the 2-year Risk of Fall in Community-Dwelling Older Americans

Problem Statement

Falls are one of the main causes of morbidity and disability in the elderly. It is estimated that about one-third of adults older than 65 fall each year. However, falls are not necessarily an normal part of aging, and there are interventions of fall prevention targeting at both personal factors and environmental factors. The prediction of the risk of falls, therefore, is very important in tailoring more targeted interventions. In this project, we utilize data from the Health and Retirement Survey (HRS) to predict the 2-year fall risk of Americans aged 65 or older.

Data Source

A secondary data analysis is conducted using data from the HRS. HRS a nationally representative longitudinal survey of the 50+ U.S population, which has been conducted every two years since 1992. We limit our sample to only participants aged 65 or older at the wave in which the outcome was measured, and for whom the outcome value is non-missing. A total sample of 4108 participants is included in our analysis.

The main outcome of interest is whether a respondent had any falls since the last wave, and is obtained from the HRS wave 2018-2019 data. The predictors are derived from the HRS wave 2016-2017, and include established personal risk factors of the aging process and fall. The predictors are in several domains:

- Demographic variables: age, gender, whether the participant lives alone
- Physical health: self-rated health, self-rated vision, and self-rated hearing, feeling fatigue, feeling dizzy, frequency of doing vigorous activity and light activity, and whether the participant has ever had arthritis, obesity, high blood pressure, diabetes, cancer, lung problems, cardiovascular diseases, stroke, heart attack, angina, incontinence, and back pain. Under assumption that chronic diseases surveyed play an equal role in the risk of fall, we create a numerical variable "co-morbidity" as the total number of chronic diseases the participant has ever diagnosed.
- Physical measures: semi-tandem test result(sec), full tandem test result(sec), balance test result(sec), waist diameter, walking speed, and BMI.
- Cognitive functioning: immediate and delayed recall, self-rated memory, and whether the participant had ever diagnosed with Alzheimer, dementia.
- Mental health: using eight items of the CES-D scale ("was depressed," "everything was an effort," "sleep was restless," "was happy," "felt lonely," "enjoyed life," "felt sad" and "could not get going") to construct a single item indicating depression level of the participant.
- Disability/Physical functioning: the number of limitations in activities of daily living, and in mobility.
- Behavioral factors: smoking in present and in the past.

We also attempt to include other seemingly important predictors such as joint replacement, side-by-side test result(sec). However, the proportion of missingness in these variables are all greater than 20%, so they are not included in the current analysis. There are total 57 features to be used in our classifier model.

Methodology

Data Pre-processing

One of the main challenge in this project is to perform data cleaning. We include only predictors with the proportion of missingness not greater than 10%. Missing values are imputed by using MissForest algorithm [1], which flexibly impute mixed data. For composite variables ("comorbidity", "psych"), imputation is conducted for individual indicators before taking the sum to create the composite variables.

Although the majority of predictors in our analysis is categorical by nature, we treated only variables with equal or less than 3 categories as categorical variable, and the others are treated as numerical in modeling process. For computational stability, we scale numerical variables before modeling.

Data are partitioned into training and testing sets by ratio 80:20. Models are calibrated with training set, and then are used to predict outcome in the testing set and subsequently evaluate their performance. The proportion of participants having falls since last wave (the outcome) in our data accounts for 35.95% of the sample. Therefore, we create a balanced training set by oversampling the minority class.

Modeling

We tackle the research question as a classification problem, using supervised classification algorithm available in *sklearn* package in Python. Models considered are: logistic regression (LR), k-nearest neighbors (KNN), Support Vector Machine (SVM), random forest (RF), multilayer perceptron (MLP) with two hidden layers of size (20, 20), and an ensemble model of these learners. Hyper-parameters are tuned by using GridSearchCV.

Evaluation and Final Results

Evaluation of model performance As mentioned previously, we notice that our sample is slightly imbalanced. Therefore, model may tend to classify instances into non-positive class, and accuracy may not well represent model performance. Other performance metrics (precision, recall, F-1 score) would be more useful in evaluating model performance. Here, we report F-1 score together with accuracy.

Results

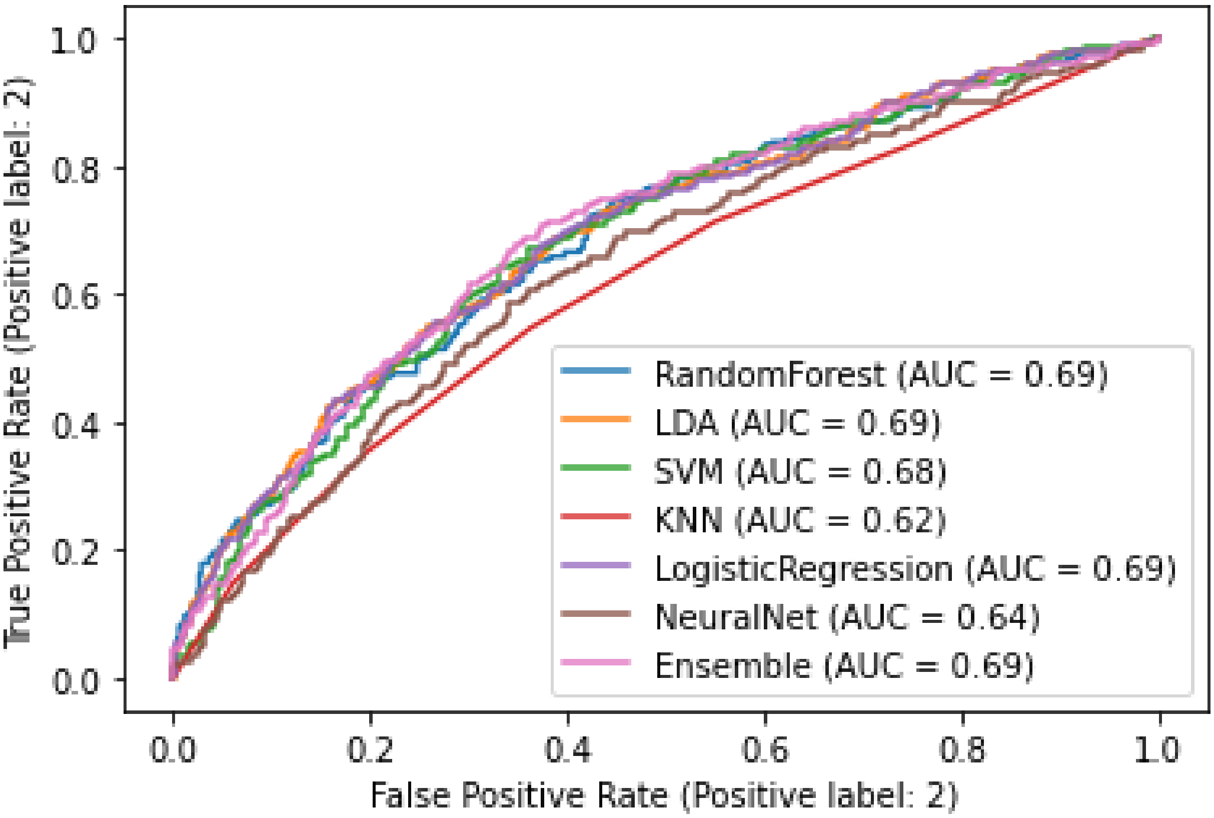
Table 1: Model performance

Model	Precision	Recall	F1-score	Accuracy
Logistic regression	0.50	0.67	0.580	0.64
Linear Discriminant Analysis	0.50	0.66	0.57	0.64
KNN	0.46	0.55	0.46	0.63
SVM	0.52	0.64	0.57	0.66
Random Forest	0.59	0.29	0.40	0.68
Multilayer Perceptron	0.49	0.47	0.48	0.64
Ensemble Model	0.53	0.57	0.57	0.67

Discussion

Using data from nationally representative survey, we are able to predict the 2-year risk of falls in the 65+ years old American population with accuracy of up to 0.68, and AUC up to 0.69. Across all classifiers, F-1 score, positive predictive value (precision), and sensitivity (recall) for the positive class is relatively low. Despite its highest accuracy, random forest does not provide

Figure 1: ROC curve comparison.



useful prediction, with very high number of fall negatives and very poor F1-score. Multilayer perceptron, though often highly acclaimed, does not perform well in our dataset. This is possibly due to the large number of hyperparameters in neural network, making it challenging in model specification. In overall, ensemble model shows the best model performance (accuracy = 0.67, AUC = 0.69, F1-score = 0.57). This highlights the well-acknowledged idea in machine learning that the combination of multiple weak learners often outperform a single strong learners. We also experiment model fitting with the original data (no oversampling applied), and notice that models that allow us to incorporate class weights in modeling process (i.e., logistic regression, SVM, random forest) show better performance. This indicates that the imbalanced nature of data plays a critical role in model performance.

Given that falls in the elderly is a function of the dynamic and complex interaction between genetic factors, behavior factors, and environmental factors, our prediction models using survey data of questionnaire items and physical measures only provide a reasonable performance. Similar behavior has been observed in previous studies of health outcome prediction. Fore example, in Lpez-Martnez et all (2019) [2], the authors used NHANES survey data to predict hypertension and achieved a model with sensitivity of only 40%, and precision of 57%, but their model still outperformed previously published models using the same survey data for hypertension prediction. As a saying goes, "Garbage in, garbage out", the prediction obtained from a model can be as good as the data at best. Without special data such as bio-markers or genome sequencing data, it is nearly impossible to predict health outcomes and achieve very low misclassification.

Our prediction model, however, can be useful in identifying older individuals at high risk of fall in the community and assisting fall prevention interventions. Predictors used for model building can be easily found in medical record and simple medical checkup, therefore, the model can find its application in various healthcare settings. In addition, as data fitted into the model is

Table 2: Confusion matrix

LR		Predicted class		LDA		Predicted class	
		Not fall	Fall			Not fall	Fall
True class	Not fall	330	196	True class	Not fall	328	198
	Fall	97	199		Fall	100	196

KNN		Predicted class		SVM		Predicted class	
		Not fall	Fall			Not fall	Fall
True class	Not fall	335	191	True class	Not fall	349	177
	Fall	134	162		Fall	106	190

RF		Predicted class		MLP		Predicted class	
		Not fall	Fall			Not fall	Fall
True class	Not fall	466	60	True class	Not fall	384	142
	Fall	209	87		Fall	158	138

Ensemble model		Predicted class	
		Not fall	Fall
True class	Not fall	376	150
	Fall	127	169

from nationally representative survey, the model is highly generalizable to the older American population.

Considering this project as an initial, exploratory step, our next steps could include: i) experiment different oversampling methods such as SMOTE, ii) tuning hyperparameters more precisely, iii) experiment with different approaches to combine models. Given time constraint of the project, we, unfortunately, are not able to do these in this report.

Evaluation and Final Results with PCA

Method: We use PCA for dimension reduction from 57 features to 2 features.

Evaluation of model performance: Using the transformed 2-dimensional data, we applied the above classifiers. Here, we report the F-1 score together with the accuracy. We also plotted the decision boundary for each classifier in the two-dimensional space in Fig. 2.

As being shown in Fig. 2, the labeled data cannot be separated using linear classifiers. It is interesting to note that the nonlinear classifiers, such as kernel SVM and neural networks, performed very well and close to the classifiers using the original data with all 57 features.

Table 3: Model performance using 2-dimensional data transformed by PCA.

Model	F1-score		Accuracy
	Not fall	Fall	
Naive Bayes	0.77	0.35	0.66
Logistic regression	0.77	0.34	0.66
KNN	0.72	0.389	0.62
Kernel SVM	0.78	0.30	0.67
Random Forest	0.72	0.43	0.63
Neural Network	0.76	0.39	0.66

Contributions

All team members contribute equally in conceiving ideas, collecting and cleaning data, modeling and writing report.

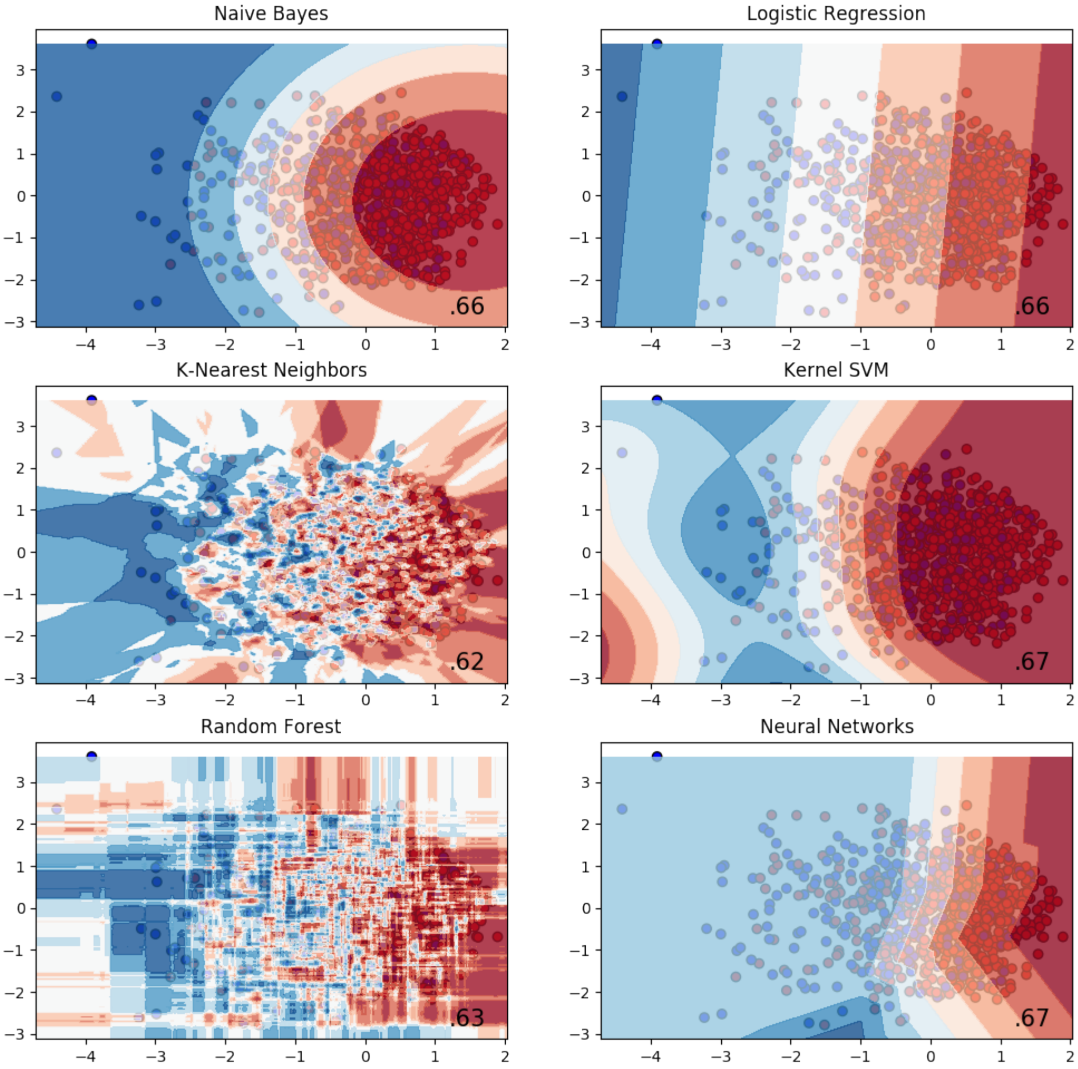


Figure 2: Data points and decision boundary for Naive Bayes, Logistic Regression, K-Nearest Neighbors, Kernel SVM, Random Forest, and Neural Networks classifiers in the two-dimensional PCA results.

References

[1] Daniel J. Stekhoven, Peter Bhlmann, MissForestnon-parametric missing value imputation for mixed-type data, Bioinformatics, Volume 28, Issue 1, 1 January 2012, Pages 112118, https://doi.org/10.1093/bioinformatics/btr597

[2] Lpez-Martnez, F., Nez-Valdez, E.R., Crespo, R.G. et al. An artificial neural network approach for predicting hypertension using NHANES data. Sci Rep 10, 10620 (2020). https://doi.org/10.1038/s41598-020-67640-z