

# *A Machine Learning Framework for Predicting Dementia and Mild Cognitive Impairment*

Daniel Stamate, Wajdi Alghamdi, Jeremy Ogg  
Data Science & Soft Computing Lab, and  
Department of Computing,  
Goldsmiths, University of London, UK  
Email d.stamate@gold.ac.uk

Richard Hoile  
Department of Primary Care and Public Health,  
Brighton and Sussex Medical School, UK

Fionn Murtagh  
School of Computing and Engineering,  
University of Huddersfield, UK

**Abstract**— Dementia is one of the most feared illnesses that has a growing year-to-year negative global impact, having a health and social care cost higher than cancer, stroke and chronic heart disease, taken together. Without the availability of a cure, nor a standardised clinical test, the utilisation of machine learning methods to identify individuals that are at risk of developing dementia could bring a new step towards proactive intervention. This study’s goal is to carry out a precursor analysis leading to building classification models with enhanced capabilities for differentiating diagnoses of CN (Cognitively Normal), MCI (Mild Cognitive Impairment) and Dementia. The predictive modelling approach we propose is based on the ReliefF method combined with statistical permutation tests for feature selection, and on model training, tuning, and testing based on algorithms such as Random Forests, Support Vector Machines, Gaussian Processes, Stochastic Gradient Boosting, and eXtreme Gradient Boosting. Stability of model performances were studied in computationally intensive Monte Carlo simulations. The results consistently show that our models accurately detect dementia, and also mild cognitive impairment patients by only using the inclusion of baseline measurements as predictors, thus illustrating the importance of baseline measurements. The best results issued from Monte Carlo were achieved by eXtreme Gradient Boosting optimised models, with an accuracy of 0.88 (SD 0.02), a sensitivity of 0.93 (SD 0.02) and a specificity of 0.94 (SD 0.01) for dementia, and a sensitivity of 0.86 (SD 0.02) and a specificity of 0.9 (SD 0.02) for mild cognitive impairment. These results support in particular future developments for a risk-based method that can identify an individual’s risk of developing dementia.

**Keywords**— Dementia, Machine Learning, ReliefF, Statistical Permutation Tests, Support Vector Machines, Gaussian Processes, Stochastic Gradient Boosting, eXtreme Gradient Boosting, Monte Carlo Simulations

## I. INTRODUCTION

Being the most feared illness of people over the age of 55 [1], dementia has been described as “...a syndrome (essentially brain failure) affecting higher functions of the brain” [1]. By the year 2050

there will be a worldwide estimated 131.5 million people who will be living with dementia and by 2030 it will have an associated worldwide cost of 2 trillion US dollars [2]. For comparison, dementia has currently a health and social care cost higher than cancer, stroke and chronic heart disease, taken together.

Diagnosis is problematic because there is currently no standardized ‘dementia test’ and the diagnosis of dementia is a highly specified task based on the possible sub-types of dementia, as well as various means of cognitive assessment [1]. Additionally, those that are diagnosed with mild cognitive impairment (MCI), while having a substantially higher risk of developing dementia [3], can either become cognitively stable or return to a healthy cognitive norm [4]. These complexities have led to a large majority of people with dementia to go undiagnosed, and even with a successful diagnosis, there is no cure [1]. Therefore, the development of effective methods is crucial for early at-risk identification and proactive interventions [5,6].

Machine learning implementation in the healthcare sector can provide an efficient means of using complex information to accurately predict diagnoses. The size and convolution of DNA sequences have been increasing in recent years; however, supervised machine learning methods, like a Bayesian Hidden Markov model, have been used to interpret DNA sequences for cancer prediction [7]. Feature selection methods, in combination with classification models, have been

---

Daniel Stamate was partly funded by Alzheimer’s Research UK (ARUK) – ‘Predicting risk of dementia using routine primary care records’ grant, and Goldsmiths, University of London.

used to identify critical variables and produce high performing models that are able to distinguish between different ‘wheeze’ categories in young children, for asthma research [8]. Dementia shares these complexities and the combination of Medical and Computer Science specialities will benefit research greatly.

With the goal of moving away from a ‘one-size fits all’ approach to dementia prediction [13], the inclusion of machine learning methods enables the utilization of various data sources and predictive variables. There are hundreds of possible predictors, but they can generally be categorized based on the following applicable models: neuropsychological based models, health-based models, multifactorial models and genetic risk scores [13]. However, there is still large amount deviance in the application of these models [9-12]. The use of magnetic resonance imaging (MRI), in combination with multiplex neural networks, has been used to segregate healthy brains from progressive mild cognitive impairment (pMCI), based on the structural atrophy of the brain because of Alzheimer’s [9]. United Kingdom’s General Practitioner (GP) patient records have been used to develop a risk score for the purposes of estimating how at risk an individual may be of developing dementia [10]. Genetic markers have been used to create a ‘polygenic hazard score test’, where the results of the test indicate how likely a person is to developing Alzheimer’s over the course of the study [11]. Positron emission tomography (PET) scans and the regional analysis of the protein amyloid- $\beta$ , have been used by a Random Forest classifier to identify patients with age-related stable MCI and pMCI [12].

Risk scores are of use because these enable primary care facilitators to have an estimation of how at-risk an individual is of developing dementia. Once an individual is identified with a high probability of developing dementia, then proactive lifestyle interventions can be adopted to curb the development of this disease. Proven methods, such as the Finnish Geriatric Intervention Study to

Prevent Cognitive Impairment and Disability (FINGER), have illustrated how multidomain interventions (diet, exercise, vascular monitoring and cognitive training) could improve primary and secondary cognitive performance with age and reduce the risk of dementia [14]. This can be further extrapolated on with the advent of a mobile application that has users provide their information and yields the resulting risk score with intervention suggestions [15]. However, an idea such as this must be grounded in verifiable and validated methodologies.

The focus on risk score methodology development has resulted in a widespread effort with more than 50 different dementia risk scores in 2010 [16]. However, a systematic review of these models has concluded that no single method could be recommended for a generalized screening procedure due to methodological weaknesses (e.g. population biases and lack of external validation) [17]. This has spurred an emphasis on analysing existing research to inform future work’s feature selection, for instance, the inclusion of predictors such as: depression, anxiety, cognitive symptoms and others that are positively associated with dementia [18].

The work we propose here is the initial process for the development of a risk score methodology, which extrapolates on the feature selection process and is flexible enough to be generalized to different cohorts. This is accomplished by conducting the preliminary analysis on Alzheimer’s Disease Neuroimaging Initiative (ADNI) data and assessing the predictive power of the variables provided, as well as predictive capabilities for trained and optimised classification models that we propose. As such, this study’s goal is to carry out a precursor analysis that utilises feature selection to identify variables with large predictive capabilities and determining the capacity that classification models have on differentiating diagnoses of CN (Cognitively Normal), MCI (Mild Cognitive Impairment) and Dementia. Information gain and ReliefF methods are alternatively used for feature

selection. Model training, tuning, and testing are performed with cross-validation and are based on algorithms such as: Random Forests, Support Vector Machines, Gaussian Processes, Stochastic Gradient Boosting, and eXtreme Gradient Boosting. Stability of model performances are studied using Monte Carlo simulations. From the promising results that we present in this work, future research will be conducted for the adaptation of this framework into a generalized, individual risk score through methods like the Disease State Index (DSI) [19].

The rest of the paper is organised as follows. In Section II we describe the data and the machine learning based predictive modelling methodology that we devised for this study. Section III discusses the results, and section IV provides the conclusions and indicates future work directions.

## II. METHODS

### A. Alzheimer’s Disease Neuroimaging Initiative

Data used in this study was extracted from the Alzheimer’s Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). The ADNI study was launched in 2003 by the National Institute on Aging (NIA), the National Institute of Biomedical Imaging and Bioengineering (NIBIB), the Food and Drug Administration (FDA), private pharmaceutical companies and non-profit organizations, as a \$60 million dollar, 5-year, public-private partnership. The primary goal of ADNI study has been to test whether serial magnetic resonance imaging (MRI), positron emission tomography (PET), other biological markers, clinical and neuropsychological assessments could be combined to create impactful research into dementia.

Data used in this work was downloaded on 6th May 2018 from the ADNIMERGER R package [20]. The *adnimerge* dataset merges together several key variables from various case report forms and biomarker lab from all ADNI protocols (ADNI1, ADNIGO, and ADNI2). The integrated data consists of 113 variables and 13272

observations, which include multiple observations per participant, representing multiple visits to the study’s site for participant evaluation (up to 20 visits per participant). The data that is used in this study is a subset of the *adnimerge* data that contains only baseline variable measurements (i.e. the first visit for each participant) that have a diagnosis. The resulting data contains 1851 participants, 49 input attributes, and one output attribute. The outcome variable ‘DX’ has the three distinct classes, which are CN (Cognitive Normal), MCI (Mild Cognitive Impairment), and Dementia. In this paper and analysis, 75% of the data were used for building and optimising prediction models, and the remaining 25% comprised the testing set to evaluate the performance of the models.

### B. Description of variables

The variables extracted from the original dataset were as follows:

- **Baselines Demographics:** age, gender, ethnicity, race, marital status, and education level were included as predictors. Demographic variables with nominal values (e.g. gender, ethnicity and race) were dummified to numeric format.

- **Functional Activities Questionnaire (FAQ)** is a test that can be used to assess the dependency on another person that a participant requires to carry out normal daily tasks. FAQ consists of questionnaires and multiple choice questions, which are given to yield an aggregate score from 0 to 30.

- **Mini-Mental State Exam (MMSE)** is used to estimate the severity and progression of cognitive impairment and to follow the course of cognitive changes in an individual over time.

- **PET measurements (FDG, PIB, AV45)** are participant’s brain function measurements.

- **MRI measurements (Hippocampus, intracranial volume (ICV), MidTemp, Fusiform, Ventricles, Entorhinal and WholeBrain)** are structural measurements of a participant’s brain.

- *APOE4* is an integer measurement representing the appearance of epsilon 4 allele of the APOE gene.

- *Variables 'ABETA', 'TAU', 'PTAU'* are cerebrospinal fluid (CSF) biomarker measurements.

- *Rey's Auditory Verbal Learning Test (RAVLT)* are neurophysiological tests evaluating an individual's episodic memory.

- *Everyday cognitive evaluations (Ecog)* are questionnaires that illustrates a participant's ability to carry out everyday tasks.

- *Logical Memory – Delayed Recall Total Number of Story Units Recalled (LDELTOTAL)* is a neuropsychological test that evaluates a person's ability to recall information after a prescribed amount of time.

- *Modified Preclinical Alzheimer Cognitive Composite (mPACC)* are tests that evaluate a person's cognition, episodic memory and timed executive function.

- *ADAS and MOCA* are generalized neuropsychological tests that evaluate a person's cognitive ability (e.g. memory, visuospatial, etc.).

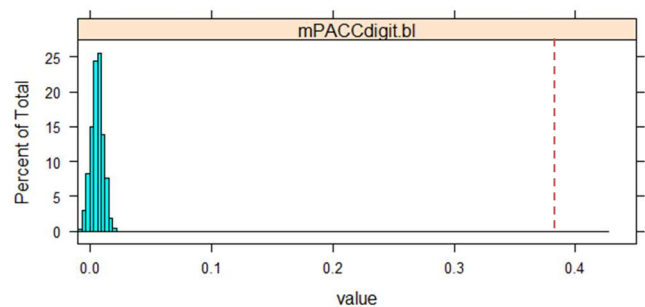
- *Variables 'Phase', 'RID', 'VISCODE'* are used as unique primary keys to merge the datasets.

### C. Feature selection

Feature selection is the process of selecting a subset of relevant features for the building of more accurate models. It helps to reduce overfitting [21], enables machine learning algorithms to train faster and reduces the complexity of models, which makes interpretation easier. Using all the 49 attributes in the experiments cost a negative impact on some machine learning algorithms such as Support Vector Machines. Therefore, to select a desired set of predictors, two feature selection methods were alternatively employed in this study:

- We simply select the attributes that score at least 0.01 in information gain.
- ReliefF feature selection combined by us with statistical permutation tests based on 500 random permutations of the labels, was employed similarly to the implementation in [8]. For instance, features with an observed Relief score corresponding to a distance of at least 1.96 standard deviations from the centre of the normal distribution built with the Relief scores, repeatedly calculated 500 times with permuted labels, were selected for further processing. This was based on the application of the permutation test with significance level  $\alpha=0.05$ . The example of such a variable (*mPACCdigit.bl*) is indicated in Figure 1. The analysis in our study was performed for each of the alpha values 0.01, 0.05, 0.1, and 0.2, to select the one leading to the best prediction performances.

FIGURE 1. THE DASHED LINE CORRESPONDING TO OBSERVED RELIEF SCORE=0.38 FOR MPACCDIGIT.BL VARIABLE IS SUFFICIENTLY FAR AWAY FROM THE CENTRE OF THE NORMAL DISTRIBUTION OF THE VARIABLE'S RELIEF SCORES BASED ON 500 PERMUTATIONS OF LABELS, IN ORDER TO INDICATE SIGNIFICANT PREDICTIVE POWER FOR MPACCDIGIT.BL.



As part of our objectives, we would like to gain a deeper understanding of which variables have strong associations to the classes, to be used as practical clinical information. This was achieved by implementing, as previously mentioned, feature selection methods with several prediction models and the results were then compared.

#### D. Missing Value Imputation

Missing values are imputed by using the random forest imputation from the *randomForest* package [22]. Although this method is computationally expensive, it enhanced the predictive power of the final models.

#### E. Balancing Classes

Large class imbalance has often a negative impact on the performance of classification models, as the algorithms producing them tend to focus more on detecting the dominant class, and less on detecting the minority class. In other words, sensitivity for the smallest class is most often negatively affected. The class distribution for our data is as follows: CN: 617, MCI: 886, Dementia: 348, which suggests that without applying a remedy solution to treat the class imbalance, the minority class Dementia could be poorly detected by the classification models. The so called synthetic minority over-sampling technique SMOTE [23] was selected to treat this class imbalance problem. SMOTE chooses a data point randomly from the minority class, determines the K nearest neighbours to that point and then uses these neighbours to generate new synthetic data points using an interpolation of said neighbours. The synthetic data was added to the minority class to overcome the gap between the majority and minority classes in terms of number of instances. Our analysis used 5 neighbours. This technique is to be applied only on the training set as it shouldn't affect the distribution of classes in the test set.

#### F. Tuning Predictive Models

To develop optimised predictive models in our study, we controlled the parameter values for each of the considered algorithms using chosen grids. Predictive models have been fitted with a 10-fold cross-validation procedure on the training set, after pre-processing techniques were applied on the same training set. Models were then evaluated on the test set.

Models were based on the following algorithms: Random Forests, Support Vector Machines,

Gaussian Processes, Stochastic Gradient Boosting, and eXtreme Gradient Boosting.

Firstly, Random Forests models were tuned over the values 2, 4, 6, ..., 48 for 'Mtry' (the number of attributes competing in a node) and a fixed number of 500 trees. A value of 38 was found as optimal for the Mtry parameter.

Secondly, Support Vector Machines models were tuned with linear and radial kernels. The optimal Support Vector Machine models were obtained after tuning the parameters 'cost' and 'gamma' with over 32 distinct values. The optimal values for cost and gamma were 16 and 0.0161, respectively.

Thirdly, Gaussian Processes models were tuned with linear and radial kernels. The optimal Gaussian Processes models were obtained with a radial kernel and a value of 0.0312 for gamma, which was tuned over 10 values.

Fourthly, Stochastic Gradient Boosting models were tuned with a different number of trees to fit, maximum depth, minimum number of observations in the tree's terminal nodes and shrinkage (learning rate). The optimal values were, respectively, 100, 8, 10 and 0.1.

Finally, the tuning parameters for eXtreme Gradient Boosting were: the number of rounds for boosting iteration, maximum depth, shrinkage (learning rate), gamma (Minimum Loss Reduction), subsample ratio of columns, minimum sum of instance weight and the subsample percentage. The optimal values were 100, 4, 0.1, 0, 0.7, 1 and 0.5, respectively.

#### G. Monte Carlo simulation

The stability of well performing models was investigated using a Monte Carlo simulation. This method repeatedly involves randomly splitting the dataset in training and test datasets, and rebuilding and retesting the models to illustrate the variance in model performance. In our case we repeated the process 100 times. Performance metrics of: accuracy, Cohen's kappa statistic, sensitivity (with

respect to each class) and specificity (with respect to each class), were evaluated and recorded with each Monte Carlo iteration. In addition, in each Monte Carlo iteration we record also the area under the curve (AUC) for each class versus the rest.

Due to lack of space, the results were visualised only for class Dementia, using boxplots to capture model performance capability and stability, for the 3 best performing models. Performances for all classes issued from Monte Carlo were captured in tables, in an aggregated form, as average performance and standard deviation (SD)

#### H. Hardware and software

The Monte Carlo simulation in this study included model optimisation as part of a 10 cross-validation, which results in a computationally expensive procedure. This was implemented in a parallel processing which was performed on a data analytics cluster of 11 servers with Xeon processors and 832GB fast RAM. The R software was used with a number of packages, including *caret*, *pROC*, *xgboost*, *e1071*, *CORElearn*, *randomForest*, *plyr*, *data.table*, *AppliedPredictiveModeling*, *DMwR* and *doParallel*.

### III. RESULTS

The top performing 15 models were based on decision trees algorithms, such as: Random Forest, Stochastic Gradient Boosting, and eXtreme Gradient Boosting. Many models built with Gaussian Process, and Support Vector Machines algorithms also achieved satisfactory results. This confirms that there exists a significantly strong pattern in the data, as several different techniques were able to capture it.

The best performing feature selection technique was the ReliefF method with  $\alpha = 0.1$ ; therefore, it was our primary feature selection method in this study.

The best result was produced by an eXtreme Gradient Boosting model, with feature selection performed by the ReliefF method. This feature selection method included only variables with an

observed Relief score corresponding to a p-value lower than 0.1 in the statistical permutation test. The second-best result was achieved by a Stochastic Gradient Boosting model that used the same variables as the model just described. The third best performer was Random Forest, using the same variables mentioned. The performances of the best 3 models are displayed in Tables I, II, and III.

TABLE I. 3 TOP PERFORMING MODELS WRT ACCURACY CRITERION ACC.  $SENS_{DEM}$  AND  $SPEC_{DEM}$  ARE SENSITIVITY AND SPECIFICITY FOR DEMENTIA CLASS. KAPPA IS THE COHEN'S KAPPA STATISTIC.

Method	ACC	KAPPA	$SENS_{DEM}$	$SPEC_{DEM}$
eXtreme Gradient Boosting	0.91	0.86	0.91	0.96
Stochastic Gradient Boosting	0.91	0.85	0.91	0.96
Random Forest	0.91	0.85	0.93	0.95

TABLE II. MONTE CARLO 100 EXPERIMENTS: AVERAGE RESULTS AND STANDARD DEVIATION (SD) OF PERFORMANCES OF THE 3 TOP PERFORMING MODELS INCLUDING PERFORMANCES FOR DEMENTIA CLASS

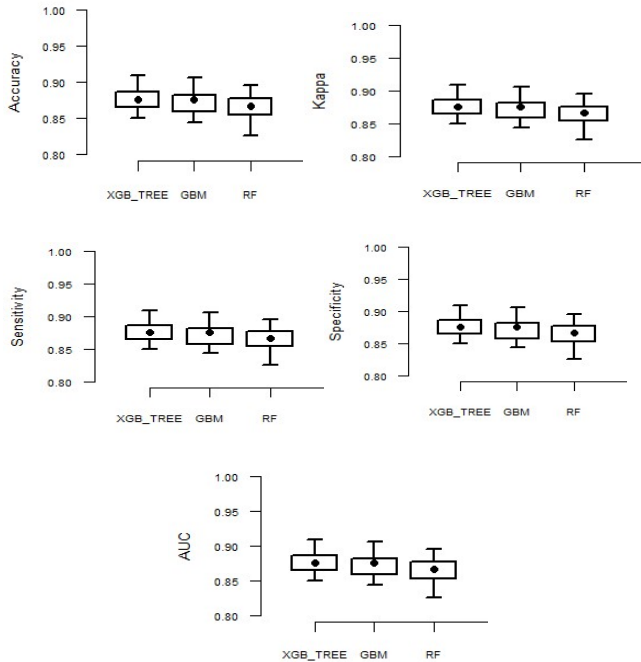
Method	ACC	KAPPA	$AUC_{DEM}$	$SENS_{DEM}$	$SPEC_{DEM}$
eXtreme Gradient Boosting	0.88 SD 0.02	0.8 SD 0.02	0.97 SD 0.01	0.93 SD 0.02	0.94 SD 0.01
Stochastic Gradient Boosting	0.87 SD 0.01	0.8 SD 0.02	0.96 SD 0.01	0.92 SD 0.03	0.95 SD 0.04
Random Forest	0.87 SD 0.01	0.79 SD 0.02	0.96 SD 0.01	0.92 SD 0.03	0.94 SD 0.04

TABLE III. MONTE CARLO 100 EXPERIMENTS: AVERAGE RESULTS AND STANDARD DEVIATION (SD) OF PERFORMANCES OF THE 3 TOP PERFORMING MODELS INCLUDING PERFORMANCES FOR MCI AND NC CLASSES

Method	$SENS_{MCI}$	$SPEC_{MCI}$	$SENS_{NC}$	$SPEC_{NC}$
eXtreme Gradient Boosting	0.86 SD 0.02	0.9 SD 0.02	0.89 SD 0.03	0.96 SD 0.01
Stochastic Gradient Boosting	0.86 SD 0.03	0.89 SD 0.02	0.87 SD 0.03	0.97 SD 0.01
Random Forest	0.86 SD 0.03	0.88 SD 0.02	0.86 SD 0.03	0.97 SD 0.01

The class-specific performances of the top 3 models for dementia, such as AUC, sensitivity and specificity, as well as the non-specific performances accuracy and kappa, are reflected also in Figure 2.

FIGURE 2. BOXPLOTS SHOWING PERFORMANCE RESULTS FOR DEMENTIA OF THE TOP 3 MODELS IN MONTE CARLO 100 EXPERIMENTS



#### IV. DISCUSSION AND CONCLUSION

Several machine learning methods and subsequent models were explored in this study. All of them were able to recognize patterns differentiating the three classes to a certain level, which indicate that machine learning does have the capacity to accurately predict dementia and mild cognitive impairment, and as such further exploration of these methods is deemed necessary. In this paper these models' performance variation was assessed with Monte Carlo experiments, and they consistently yielded adequate predictive power and significant model performance stability. The best performing model in term o accuracy was eXtreme Gradient Boosting, achieving as high as 91% accuracy in some cases and an average performance of 88% accuracy in Monte Carlo simulations with 100 repetitions.

By evaluating the discriminative power of variables across models, we found that several features show high predictive capabilities. Feature analysis was carried out on the training data only and Table IV shows the top results of observed Relief score that correspond to a p-value lower than 0.1 in the permutation test performed with 500 permutations of labels. Table V also shows the top results as scored by information gain. Although the rank of the predictors is not identical, the two methods seem to concord significantly in this case. However, we tend to value more the ReliefF based procedure as the ReliefF method accounts for some interaction between features (which Information gain doesn't), and secondly it is combined with a statistical permutation test by permuting labels 500 times whose impact is a selection of features whose observed Relief index is higher than what is obtained by randomly re-combining records with labels. The downside of this procedure is that it becomes more expensive computationally. However, compared with the Information gain simple method, it led to better results in terms of performances of the predictive models.

TABLE VI. THE TOP 10 FEATURURES ACCORDING TO RELIEFF WITH PRMUTATION TEST, AND INFORMATION GAIN METHODS

Score	Relieff	Information gain
1	mPACCdigit.bl	LDELTOTAL.bl
2	mPACCtrailsB.bl	mPACCdigit.bl
3	LDELTOTAL.bl	mPACCtrailsB.bl
4	ADAS13.bl	MMSE.bl
5	FAQ.bl	FAQ.bl
6	MMSE.bl	ADAS13.bl
7	ADAS11.bl	EcogSPTotal.bl
8	EcogSPTotal.bl	ADASQ4.bl
9	EcogSPMem.bl	ADAS11.bl
10	EcogSPPlan.bl	EcogSPMem.bl

Ongoing work envisages multiple directions for extending this work, in particular: (i) translating the problem in 2-class classifications and extending the methodology with techniques such as ROC cross-evaluation [8], which also involves model optimisation in post-processing, and (ii) utilising

the best models resulting from this study and from its forthcoming extension, in the creation of individualized dementia-risk scores.

## REFERENCES

- [1] E. Barrett and A. Burns, "Dementia Revealed. What Primary Care Needs to Know," Department of Health, 2014.
- [2] M. Prince, A. Wimo, M. Guerchet, G.C. Ali, Y.T Wu, and M. Prina, "The Global Impact of Dementia An analysis of prevalence, incidence, cost and trends," World Alzheimer Report, 2015.
- [3] R. Petersen, R. Roberts, D. Knopman, B. Boeve, Y. Geda, R. Ivnik, et al. "Mild cognitive impairment: ten years later," Arch Neurol. 2009
- [4] A.J. Mitchell and M. Shiri-Feshki, "Rate of progression of mild cognitive impairment to dementia—meta-analysis of 41 robust inception cohort studies," Acta Psychiatr Scand, 2009.
- [5] M. Prince, A. Comas-Herrera, M. Knapp, M. Guerchet, and M. Karagiannidou, "Improving healthcare for people living with dementia coverage, Quality and costs now and in the future," World Alzheimer Report, 2016.
- [6] M. Prince, E. Albanese, M. Guerchet, and Prina, M., "Dementia and Risk Reduction AN ANALYSIS OF PROTECTIVE AND MODIFIABLE FACTORS," World Alzheimer Report, 2014.
- [7] G. Manogaran, V. Vijayakumar, R. Varatharajan, P.M. Kumar, R. Sundarasekar, and C.H. Hsu, "Machine Learning Based Big Data Processing Framework for Cancer Diagnosis Using Hidden Markov Model and GM Clustering," Wireless Pers Commun, 2017.
- [8] D. Belgrave, R. Cassidy, D. Stamate, A. Custovic, L. Fleming, A. Bush, and S. Saglani, "Predictive Modelling Strategies to Understand Heterogeneous Manifestations of Asthma in Early Life," in 16th IEEE International Conference on Machine Learning and Applications (IEEE ICMLA'17), 2017.
- [9] Nicola Amoroso, Marianna La Rocca, Stefania Bruno, Tommaso Maggipinto, Alfonso Monaco, Roberto Bellotti, et al., "Brain structural connectivity atrophy in Alzheimer's disease," arXiv:1709.02369 [physics.med-ph], 2017.
- [10] K. Walters, S. Hardoon, I. Petersen, S. Iliffe, R.Z. Omar, I. Nazareth, and G. Rait, "Predicting dementia risk in primary care: development and validation of the Dementia Risk Score using routinely collected data," BMC Medicine, 2016.
- [11] R.S. Desikan, C.C. Fan, Y. Wang, A.J. Schork, H. Cabral, L.A. Cupples, et al., "Genetic assessment of age-associated Alzheimer disease risk: Development and validation of a polygenic hazard score," PLOS Medicine, 2017.
- [12] D. Brauser, "Machine Learning' Tool May Predict Dementia Development Up to 10 Years Later.," Medscape, 2016.
- [13] L. Robinson and M. Trenell, "Dementia: risk reduction," Newcastle University Institute of Ageing, 2016.
- [14] M. Kivipelto, A. Solomon, S. Ahtiluoto, T. Ngandu, J. Lehtisalo, R. Antikainen, et al., "The Finnish Geriatric Intervention Study to Prevent Cognitive Impairment and Disability (FINGER): Study design and progress," Alzheimer's Association, vol. 9, issue 6, pp. 657-665, January 2013.
- [15] S. Sindi, E. Calov, J. Fokkens, T. Ngandu, H. Soininen, J. Tuomilehto, and M. Kivipelto, "The CAIDE Dementia Risk Score App: The development of an evidence-based mobile application to predict the risk of dementia," Alzheimers & Dementia: Diagnosis, Assessment & Disease Monitoring, vol. 1, issue 3, pp. 328-333, July 2015.
- [16] S. C.M. Blossom, C. Brayne, C. Dufouil, T. Kurth and F. Matthews, "Dementia risk prediction in the population: are screening models accurate?" Nat. Rev. Neurol, 2010.
- [17] S. C.M. Blossom, E. Tang, G. Muniz-Terrera, "Composit risk scores for predicting dementia", Curr Opin Psychiatry, vol. 29, no. 2, March 2016.
- [18] Elizabeth Ford, Nicholas Greenslade, Priya Paudyal, Stephen Bremner, Helen E. Smith, Sube Banerjee, et al., "Predicting dementia from primary care records: A systematic review and meta-analysis," PLoS ONE, March 2018.
- [19] T. Pekkala, A. Hall, J. Lötjönen, J. Mattila, H. Soininen, T. Ngandu, et al., "Development of a Late-Life Dementia Prediction Index with Supervised Machine Learning in the Population-Based CAIDE Study," Journal of Alzheimers Disease, vol. 55, no. 3, pp. 1055–1067, 2017.
- [20] ADNIMERGE, Alzheimer's Disease Neuroimaging Initiative (2018). ADNIMERGE: Alzheimer's Disease Neuroimaging Initiative. R package version 0.0.1
- [21] M. Kuhn and K. Johnson, Applied Predictive Modeling. Springer New, York, 2013.
- [22] A. Liaw and M. Wiener (2002). Classification and Regression by randomForest. R News 2(3), 18–22
- [23] N. Qazi, K. Raza, Effect of Feature Selection, SMOTE and Under Sampling on Class Imbalance Classification, 2012 UKSim 14th, 2012. pp. 145-150.