Big Data Supporting Public Health Policies

- SC1-PM-18-2016

Project **H2020-727658 / IASIS**

http://project-iasis.eu/

Dementia Prediction Models report

Dr. Peter Garrard, Eleftherios Samaras, Fotis Aisopos, George Katsimpras, William Scott, Akhilesh Vyas, Maria Esther Vidal

Status: Final (Version 1.0)

November 2020

Project

Project Ref. no	H2020-727658
Project acronym	IASIS
Project full title	Integration and Analysis of Heterogeneous Big Data for Precision Medicine and Suggested Treatments for Different Types of Patients
Project site	http://project-iasis.eu/
Project start	April 2017
Project duration	3 years
EC Project Officer	Dr. Jose Valverde Albacete

Deliverable

Deliverable type	Report
Distribution level	Public
Deliverable Number	D1.3 (Appendix of the Final Progress Report)
Deliverable Title	Dementia Prediction models report
Contractual date of delivery	
Actual date of delivery	
Relevant Task(s)	WP5/Task 5.6 WP7/Task 7.4
Partner Responsible	LUH, SGUL, NCSR
Other contributors	
Number of pages	18
Author(s)	Dr. Peter Garrard, Eleftherios Samaras, Fotis Aisopos, George Katsimpras, William Scott, Akhilesh Vyas, Maria Esther Vidal
Internal Reviewers	Fotis Aisopos
Status & version	Final
Keywords	Dementia prognosis, diagnosis, Mild Cognitive Impairment

Executive summary

In this report, we present the predictive models for identifying diagnostic results and estimating possible conversions to Dementia based on data coming from OPTIMA and Oxford UK-CRIS.

Prediction Model 1: Prognosis of Dementia in OPTIMA

Problem Statement

The goal of this predictive model is identifying possible conversion to Dementia. We aim to predict if a patient with Mild Cognitive Impairment will develop Dementia in 0-1 years and in 1-5 years. We define two possible target classes, class 1 denotes the possible development of dementia and class 0 denotes otherwise.

Data Pre-processing

To generate training samples for both classes we follow a specific methodology. For each of the identified dementia patients, we build a clinical background based on their dementia and non-dementia episodes, arranged in chronological order. As the first dementia episode of a patient we define the earliest dementia episode in their clinical background. As the first non-dementia episode of the patient we define the earliest episode that is classified as non-dementia in the time period defined by the specific task. For instance, for the 0-1 years task, the first non-dementia episode will be the earliest non-dementia episode within 1 year before the first dementia episode of the patient. Likewise, for the 1-5 years task, the first non-dementia episode will be the earliest non-dementia episode within 1 to 5 years before the first dementia episode of the patient. By collecting all first non-dementia episodes from each dementia patient we form our positive class instances. In a similar way, we build the negative class instances.

Feature Selection & Experiment Configuration

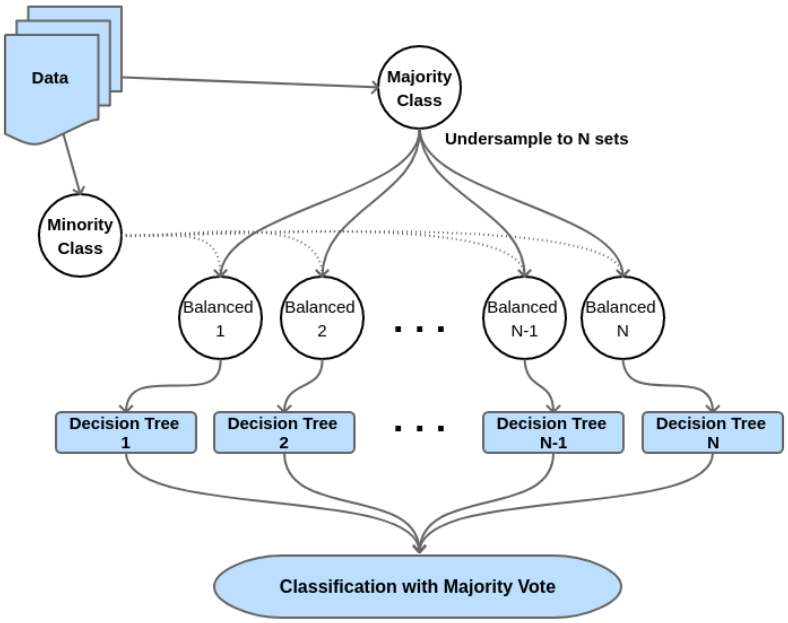
We manually identify the features that are highly-correlated (or provide the same information) which makes them redundant. In such cases, we can use one feature without losing information. For example, the value of feature “COGNITIVE EXAM 120-161: (137) IDENTIFIES OBJECTS” is the sum of features “COGNITIVE EXAM 120-161: (137) IDENTIFIES OBJECTS: PENCIL” and “COGNITIVE EXAM 120-161: (137) IDENTIFIES OBJECTS: WATCH”. In this case, we only keep the feature “COGNITIVE EXAM 120-161: (137) IDENTIFIES OBJECTS” and remove the other two. In this way, we eliminate redundancy but also the percentage of missing values. At a higher level, we observe that the CAMDEX CAMCOG scores (Orientation, Comprehension, Expression, Remote Memory, Recent Memory, Learning Memory, Attention, Praxis, Calculation, Abstract Thinking and Perception) represent the diagnostic results of two cognitive exams: COGNITIVE EXAM 120-161 and COGNITIVE EXAM 162-187. For instance, the Remote Memory score is produced by summing the features of COGNITIVE EXAM 148 to153 (namely, all REMEMBERS features). Similarly, the Praxis score is produced by summing the features of COGNITIVE EXAM 164 to 167, COGNITIVE EXAM 169 to 170 and COGNITIVE EXAM 172 to 174.

Following this process, we can keep the scores and remove the redundant features. Based on the above, and after evaluating many options, we choose a subset of 38 features, sufficient to the target concept/task and improves prediction performance (or at least does not significantly decrease it).

Our goal is to build predictive models that are efficient, yet are highly interpretable by the experts/clinicians. It is important to provide meaningful information about the logic involved in the decision making/classification outcome. Towards this direction, one major advantage of Decision Trees (DT) methods is their interpretability, as the classification outcome can be explained in terms of rules. For this reason, the prediction models are based on DT.

OPTIMA data present high imbalance, meaning that the data of one class (non-dementia instances) significantly outnumber the data of the other class (dementia instances). In such scenarios, where the

class of interest has a very small number of instances, there are different ways to address the class imbalance such as oversampling, undersampling or a combination. To avoid constructing synthetic data samples (oversampling), we choose an approach that combines bagging and undersampling. The idea is to train N different classifiers with **balanced** data by applying undersampling in the majority class. To make N balanced subsets we randomly sample the same number of majority class instances as there are minority class instances, ensuring equal numbers of instances for each class. Then, one classifier (DT) is learned for each of these N balanced datasets. Each classifier produces a decision (vote) and in order to classify a new instance we perform a majority vote scheme on all decisions/votes:



Evaluation Results

The complete dataset contains 1035 patients. Following the procedure described for data pre-processing, we identify 52 dementia patients (positive class) for the 0-1 years task, 72 dementia patients (positive class) for the 1-5 years task and 325 non-dementia patients (negative class) for both tasks. So, in total, there are 377 instances for the 0-1 years task and 397 instances for the 1-5 years task. A N-fold cross validation is applied on the data (N=10). This procedure divides the data into N different folds (subsets) of equal size and each time we use nine of them for training and the tenth for testing. Training and test data are different on each fold. For each fold, we perform the bagging model, described in the previous section, by undersampling the training data N times. The test data of the fold remains untouched. Therefore, we create N balanced training datasets from the original training data of the corresponding fold. A classifier is learned for each balanced dataset. The majority vote of the classifiers on the test data forms the decision/prediction for the corresponding fold.

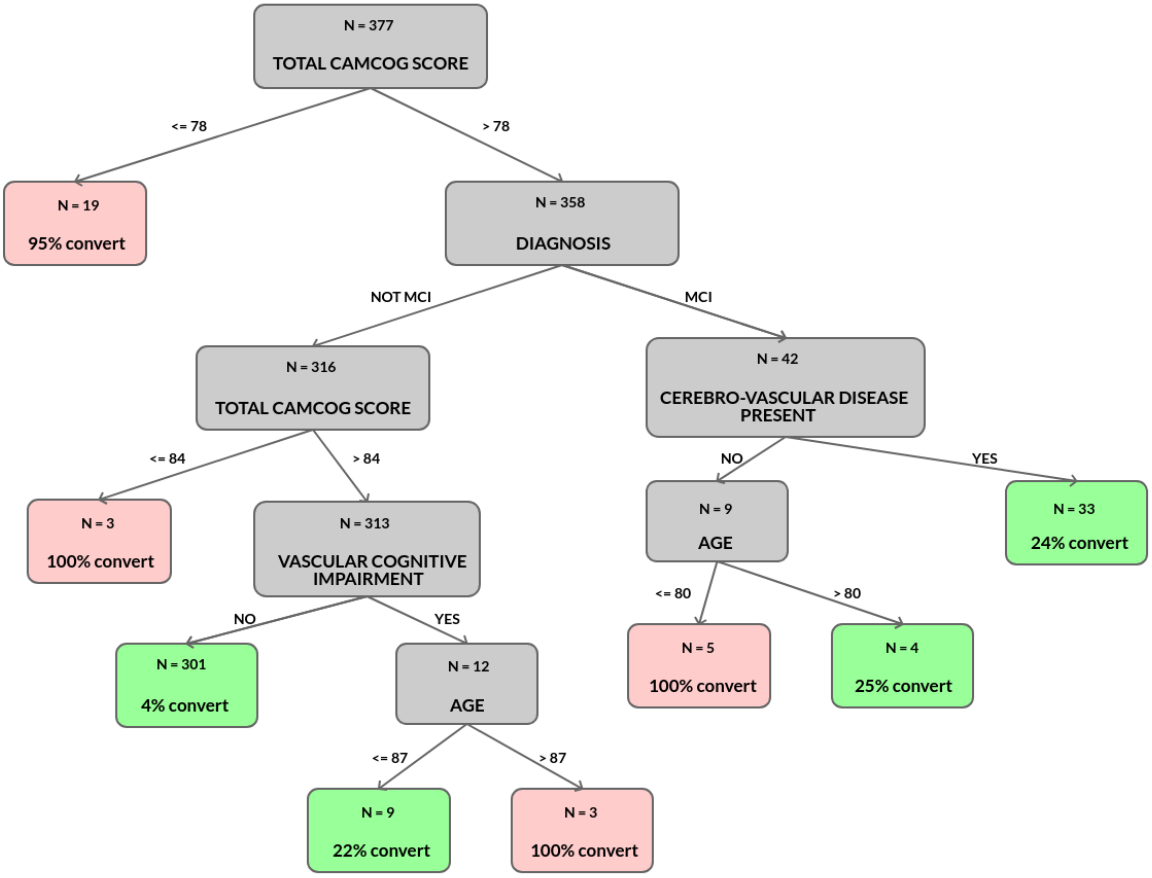
Moreover, we apply hyperparameter tuning at each fold. Specifically, we perform grid search on four hyperparameters to adjust the size and complexity of the decision tree (pruning). Applying grid search on each fold will produce ten different sets of values for the hyperparameters. While using different parameters for each fold results in better performance, it also limits the interpretability of the approach as we will end up with ten different decision trees. For this reason, we use the average of the hyperparameters out of all folds. By doing so, we face a slight drop in the performance but we can interpret the decision rules of the proposed model.

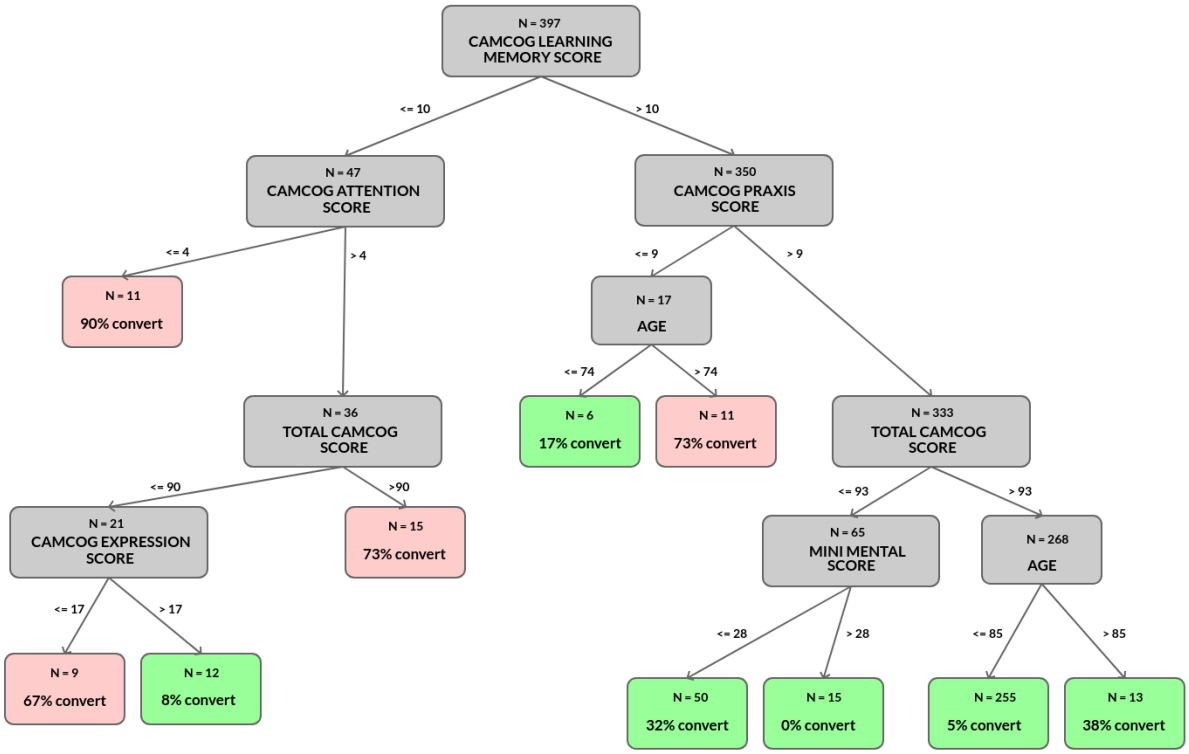
To evaluate the performance of the proposed model, we calculate the micro Precision, Recall and F1-score scores for the positive class only. The results of the proposed model are depicted in this Table:

Task	Method	Precision	Recall	F1-score
0-1 years	Logistic Regression	0.63	0.40	0.49
	Random Forest	0.74	0.38	0.51
	XGBoost	0.57	0.40	0.47
	Bagging	0.75	0.49	0.58
1-5 years	Logistic Regression	0.58	0.27	0.38
	Random Forest	0.51	0.28	0.36
	XGBoost	0.53	0.39	0.45
	Bagging	0.63	0.38	0.47

Decision Trees

In the following Figures, we illustrate the decision trees that were obtained for 0-1 year and 1-5 years prediction task:





A decision tree is composed of nodes and leaves. A node represents a dichotomous threshold for the value of some feature in the dataset, while a leaf represents a patient subgroup in whom the likelihood of belonging to the positive class (developing dementia within a time period) cannot be refined by any additional dichotomous test. Nodes and leaves are connected by branches, each of which represents an additional condition; any path through the decision tree represents the outcome of a series of conditional statements.

For example, in the first Figure, we have four leaves containing groups of cases of whom more than 50% belong to the positive class (coloured red) and four representing groups of cases of whom fewer than 50% belong to the positive class (coloured green). Paths leading to red-coloured leaves are the sets of conditions that place an individual at higher risk of converting to dementia within a year. The first of these leaves can be reached by following the left-hand branch of the root node, which represents a dichotomous division of the CAMCOG total score into those scoring 78 or below, and those scoring above 78. The former condition is met by 19 cases, 18 of whom are in the positive class: scoring below 79 on the CAMCOG can therefore be considered to be an outcome that indicates a high likelihood of conversion to dementia within a year. This does not mean that the 538 patients whose CAMCOG score was higher than 78 are all at low risk. A further series of conditions allows a further subgroup at high risk to be identified by selecting the 316 to whom the diagnostic label of 'Mild Cognitive Impairment' was deemed inapplicable (via the 'NOT MCI' branch), and then defining a further CAMCOG threshold (84), scoring at or below which indicates a high likelihood of conversion within a year.

It must be noted that the majority of patients without a diagnosis of MCI (313 out of 316) score above 84 on the CAMCOG, and that to identify those who are at high risk of converting, it is necessary to select out the twelve cases who met diagnostic criteria for Vascular Cognitive Impairment. Among these patients applying an age threshold of 87 selects out the majority of cases in which conversion to dementia will take place over the next twelve months.

A fourth subgroup of high risk patients exist among the 42 who scored above 78 on the CAMCOG and did meet criteria for a diagnosis of MCI. Following the path to the five patients at high risk from this node indicates that, in the absence of a diagnosis of cerebrovascular disease (the presence of which

does not, in itself, indicate a very high risk of conversion), those aged 80 or over form the final high-risk group. The characteristics of low risk groups (green leaves) can be identified in the same way, while the tree illustrated in the second Figure can also be completely traversed and interpreted in a similar fashion to discover the rules that lead to the identification of four groups of patients (n=46) whose risks of conversion to dementia within 5 years range between 67 and 90%.

Prediction Model 2: Dementia diagnosis

Problem Statement

This problem deals with classifying the patients' current state into different types of Dementia / No-Dementia. To consider a patients' current status we use all patient-episode data. Multiple sub problems are derived from the defined problem statement. They are as follows:

1. **Predict Dementia / No-Dementia**: This problem addresses the issue of identifying if a patient is suffering from dementia or not in a specific episode (correctly diagnose Dementia). In this task, we classify a patient-episode into Dementia (1) or No-Dementia (0).
2. **Predict Severity of Dementia / No-Dementia**: This problem expands the issue addressed by problem 1, by further identifying the severity of Dementia when the patient is suffering from Dementia. The classes involved here are: No-Dementia (0) and Dementia Severity (1, 2, 3, 4).
3. **Predict Severity of Dementia**: This problem deals with patients who already suffer from dementia and identifies their severity. The patient-episode is classified into 1, 2, 3 and 4 ranges of severity.
4. **Predict Low Dementia / High Dementia**: This problem addresses the issue of identifying severity on a weaker scale when the patient is already suffering from dementia. A stronger model is observed when the severity is in two ranges: Severity 1 and 2 are considered as low-dementia (1) and severity 3 and 4 from OPTIMA data are considered as high-dementia (2).
5. **Predict No-Dementia, Low-Dementia / High-Dementia**: This problem is further refined from problem 4 by further identifying if the patient does not suffer from dementia. No-dementia (0) is an added label to denote that the patient does not suffer from dementia.

Data Pre-processing

The predictive models learn on patient-episode data. To make the margin between a patient episode with and without dementia, we consider patient-episodes from patients who have no history of Dementia, for constructing a “No Dementia” class. Following this assumption, we found 183 patients (with 863 episodes) with at least one episode of no-dementia and at least one previous episode of dementia, so we ignore these patient-episodes. Also, due to inconsistencies in the data, a patient episode might fall under both Dementia and No-Dementia class, we resolve this issue by considering the particular patient-episode into the Dementia class.

Moreover, in case of inconsistencies among attributes showing the severity of dementia, we consider the maximum severity of those. The number of ambiguous dementia severity episodes is 231.

Feature Selection & Experiment Configuration

Necessary data points are sampled from the required classes to build the training data for all models. The data is balanced before training and GridSearchCV is performed with cv=5 on 70-30 train-test split. The accuracies reported are the average of multiple such runs. To further simplify the information needed for models to reach a conclusion, feature importance of all the models are considered to select 20 attributes in total. The attributes are selected so as not to degrade the accuracy.

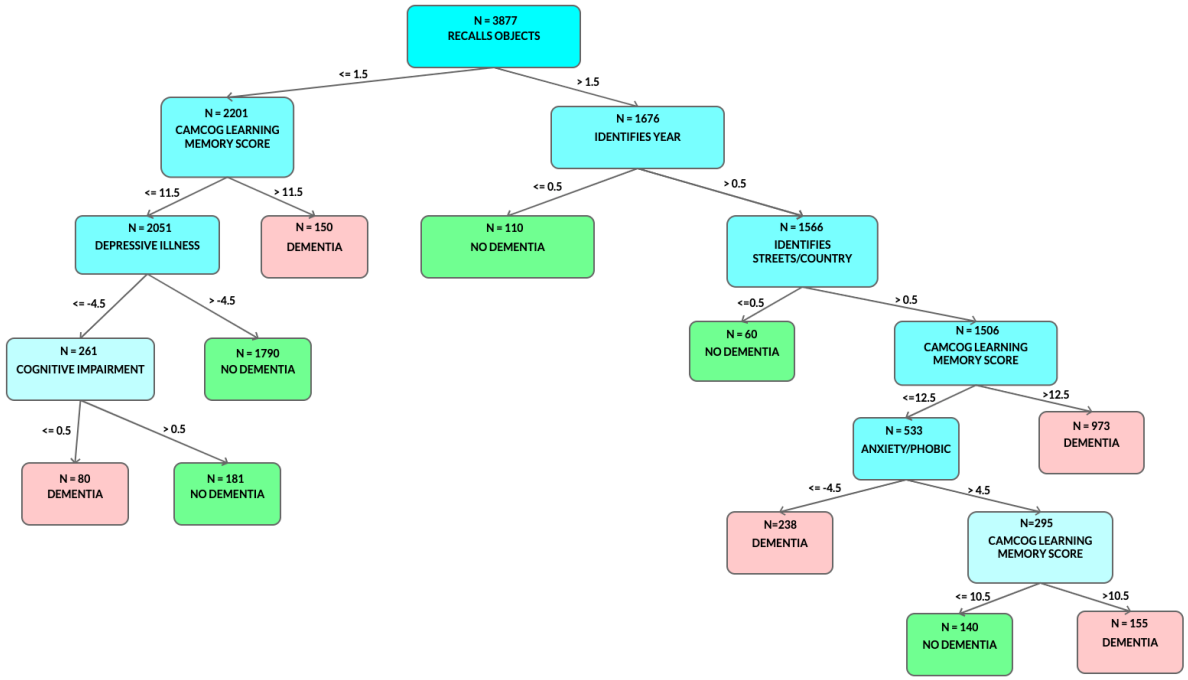
Evaluation

The following Table shows the evaluations performed on two supervised models using Random Forest and Logistic Regression on both F1 and Accuracy for all sub-problems mentioned above. We can observe that Random Forest performs better than Logistic Regression in both F1 and Accuracy.

Model	Random Forest		Logistic Regression	
	F1	Acc	F1	Acc
Sub-problem 1	95	95	92	92
Sub-problem 2	68	69	64	65
Sub-problem 3	62	62	59	59
Sub-problem 4	87	88	84	84
Sub-problem 5	87	87	86	86

Decision Tree

The following Figure illustrates the results of a Decision Tree classifier for sub-problem 1 (Predicting Dementia/ No-Dementia on a single-patient episode). DT classifiers may slightly decrease the performance of a model, but as said they provide a rationale behind decisions taken with only a small subset of variables.



In a similar way to the previous Prediction Model DTs, the decision tree above describes 5 ways in which a patient's performance could be analysed to support a diagnosis of dementia. To illustrate, a relatively good (1.5) visual memory score coupled with only partial orientation to time and place would justify a diagnosis of dementia (right branch). By contrast, a diagnosis of dementia in a patient with relatively poor visual and verbal memory performance would be justified unless the patient suffered from depression, in which case it would be reasonable to rule out dementia, except when there was other evidence of cognitive impairment (left branch).

Prediction Model 3: Mini-Mental State Examination miscalculation

Problem Statement

The Mini-Mental State Examination (MMSE) score of the patients is analysed in different episodes to determine their cognitive assessment and it is recorded in a structured way in OPTIMA. There are four categories based on the range of MMSE values namely Normal, Mild, Moderate and Severe. The range of values for each category is defined such as greater than 24 (Normal), [19, 24] (Mild), [14,18] (Moderate) and less than 14 (Severe). However, sometimes, these MMSE values are not accurate for a given episode of a patient. Thus, patient episodes can be classified into two classes as mis-classified (or misdiagnosed) 'YES'/'NO'. In this problem, we first calculate the baseline of the misdiagnosed episodes using a hand-crafted rule, based on an evaluation of iASiS medical experts.

Our goal is two-fold:

a. Attempt to identify miscalculations, by classifying an episode in two classes (Misdiagnosed= YES/NO), based on the features of this episode only
b. Use a linear regression model for MMSE and apply retrospectively, in order to correct the percentage of the Misdiagnosed episodes

Data Cleaning & Preprocessing

Baseline plays an important role in planning of the selection of training dataset for the predictive models. As a first cleaning step, we remove the episodes from the dataset where MMSE values are not in the range [0,30]. Secondly, we apply a hand-crafted rule provided by our experts for the further selection of the dataset. The hand-crafted rule is defined as follows:

*A patient is considered to be misdiagnosed (**misdiagnosed='YES'**) if the change in MMSE value of two patient's episodes satisfies one of the following conditions:*

1. *MMSE value increases >5.0 in a period <=1 year*
2. *MMSE value increases >=3.0 in a period within 1-2 years*
3. *MMSE value increases >=0.0 in a period >2 years*

*If none of the above conditions apply, we consider the patient as **misdiagnosed='NO'**.*

Finally, we transform ordered categorical features by applying ordinal encoding (eg. size Small, Medium, Large) and categorical features by utilizing one-hot encoding[1].

Experiment Configuration

In total, models are trained on 1104 episodes of the patients. We define two classes for the misdiagnosed patient's episodes by utilising earlier defined hand-crafted rule misdiagnosed 'YES' (971 patients) and 'NO' (133 patients) for our classification model. In the context of these experiments, we try to improve rule-based baseline using machine learning models. The problem is addressed using two machine learning models classification model and regression & classification model.

[1] https://scikit-learn.org/stable/modules/preprocessing.html#preprocessing-categorical-features

Classification Model: The predictive model is implemented using random forest classifier with 5 fold-cross validation, and ensemble learning. Besides, employing an imbalance algorithm[2] for solving minority class imbalance issue in the dataset has a significant effect on the results of the model. We do our evaluation on different distribution of classes based on the imbalance algorithms as follows:

- Original-dataset: We make no changes in the dataset regarding imbalance of the classes.
- Oversampled-dataset: We oversample the minority class (Misdiagnosed='YES') data-points using SMOTE algorithm. The idea of the algorithm is to generate new data points for the minority class 'YES' equal to the majority class 'NO' for better evaluation of the model. However, this approach fails due to the very low minority class data-points.
- Undersampled-dataset: We undersample the majority class ('NO') data-points using cluster-centroids algorithm. The idea of the algorithm is to generate new data points for the majority class 'NO' equal to the minority class 'YES' for better evaluation of the model performance. It provides equal distribution of classes for unbiased prediction.
- Random-Undersampled-dataset: We undersample the majority class ('NO') data-points using random-sampler algorithm. The idea of the algorithm is to generate new data points for the majority class 'NO' equal to the minority class 'YES' for better evaluation of the model performance. It provides equal distribution of classes for unbiased prediction.

Classification Model Evaluation

The following Table shows that both original and oversampled datasets experience an imbalance and bias problem in the classes. Misdiagnosed='YES' class is not able to learn due to less number of data points compared to the 'NO' class in the original dataset. The model performs better on 'YES' class as a result of the dense cluster while generating data points for 'YES' class in oversampled dataset. The real performance of the model can be measured in undersampled datasets. In the case of the cluster-centroids, it makes use of centroid of the majority class datapoints for the classification. In the other case, the model is trained on a few randomly selected data points by random-under-sampler. The Table demonstrates a significant difference in the performance of undersampled datasets for F1-score.

[2] https://imbalanced-learn.readthedocs.io/en/stable/index.html

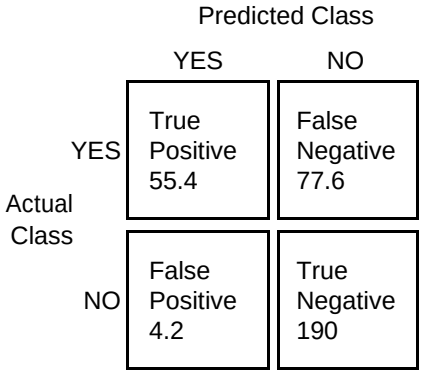
Original Dataset				
	Precision	Recall	f1-Score	Support
NO	0.88	0.99	0.93	971
YES	0.26	0.04	0.07	133
accuracy			0.87	1104
macro avg	0.57	0.51	0.50	1104
weighted avg	0.81	0.87	0.83	1104
Oversampled Dataset (SMOTE)				
	Precision	Recall	f1-Score	Support
NO	0.93	0.96	0.94	971
YES	0.96	0.93	0.94	971
accuracy			0.94	1942
macro avg	0.94	0.94	0.94	1942
weighted avg	0.94	0.94	0.94	1942
Undersampled Dataset (ClusterCentroids)				
	Precision	Recall	f1-Score	Support
NO	0.92	0.91	0.91	133
YES	0.91	0.92	0.91	133
accuracy			0.91	266
macro avg	0.91	0.91	0.91	266
weighted avg	0.91	0.91	0.91	266
Undersampled Dataset (RandomUnderSampler)				
	Precision	Recall	f1-Score	Support
NO	0.70	0.74	0.72	133
YES	0.72	0.68	0.70	133
accuracy			0.71	266
macro avg	0.71	0.71	0.71	266
weighted avg	0.71	0.71	0.71	266

Regression & Classification Model: The Regression model is trained using an ensemble learning of Random Forest Regressor. Iteratively, we train our regression model using 838 Misdiagnosed='NO' data instances (selected 13 best features) out of a total 971 to predict the MMSE values (biased to learn 'NO' cases MMSE values only) and test on 133 data points of each class. We choose 838 and 133 data points of 'NO' class randomly for training and testing respectively in each iteration. This procedure is repeated 5 times to understand the performance of the model. After each process, predicted MMSE values of the data points are fed into the hand-crafted clinician rule to re-assess their class ('YES'/'NO').

Regression & Classification Model Evaluation

Average results of this evaluation process are summarized in the following Table and confusion matrix:

Average Classification Report				
	Precision	Recall	f1-Score	Support
NO	0.71	0.98	0.82	194.2
YES	0.93	0.42	0.58	133
accuracy			0.75	327.2
macro avg	0.82	0.70	0.70	327.2
weighted avg	0.80	0.75	0.72	327.2



The Table demonstrates that for previously misdiagnosed='YES' cases, the resulting classification has a low recall but significantly high precision. This practically means that although the regression model manages to 'correct' less than half of the misdiagnosed cases (55.4 True Positive in average), it does

so with very good precision. On the other hand, for previously misdiagnosed='NO' cases, the resulting classification has a very high recall (just 4.2 False Positives in average). This means that 98% of the correctly diagnosed cases did not change using the predicted MMSE scores. Therefore, we are able to correct a significant number of misdiagnosed cases comparing with the baseline rate, while not seriously affecting the misdiagnosed='NO' class.

Prediction Model 4: Prognosis of Dementia in UK-CRIS

Problem Statement

Our goal is to build a predictive model for identifying possible conversion from Mild Cognitive Impairment to Dementia in UK-CRIS, the same way as defined for Prediction Model 1 in OPTIMA.

Data Pre-Processing and cleaning

UK-CRIS patient data was provided in Excel tables. In the Patient table, patients' age was calculated in years from the date of birth column, gender was mapped as 1 for male and 0 for female and ethnicity was also mapped (e.g. Caucasian for every Caucasian-British, Caucasian-European subcategory and similarly for other ethnicities) resulting in the following ethnicity categories: Asian, Afro-Caribbean, Caucasian, Mixed, Other and Unknown. The MCI, Dementia and Comorbidities tables, were sorted in ascending order by patient id and diagnosis date and the earliest record was retained per table. Also, in the Dementia and Comorbidities tables the dementia type and psychiatric illness type codes were replaced with the corresponding text (i.e. F00.0 was replaced with 'Dementia in AD with early onset').

In the Smoking, Alcohol and Employment tables, apart from keeping only the earliest instance, we manually went through the text of each individual records and categorised it based on specific criteria. For alcohol status, the categories were nothing, occasional and excessive. For smoking status, the categories were non-smoker, ex-smoker and smoker. For employment status, the categories were retired, employed and other. In the Weight variable data, the earliest instance was extracted per patient. The Demographics table included the fields First Language and Marital Status, while various fields representing fall risks and unsteadiness were merged into one feature. Lastly, UK-CRIS included the Health of the Nation Outcome Scale (HONOS) dataset, providing various assessment scores (like physical illness score, other mental problems score, hallucinations /delusions score, etc.), which were split different variables, based on different score ranges.

To find patients with an MCI episode that also have an episode with Dementia, we merge the patient and Dementia tables on the all MCI patients' table resulting into 3779 cases, where the Dementia 'diagnosis date' and 'dementia type' null values are 2428 (64%), indicating patients that never converted to Dementia as well those that didn't convert yet. The Dementia 'diagnosis date' empty cells were filled with a recent date in order to calculate the conversion time (in months) per case, with the negative ones (where the MCI diagnosis date is more recent than the Dementia diagnosis date), being identified as misdiagnoses and removed from the dataset.

The MMSE scores of MCI patients (3664) were extracted by Oxford's UK-CRIS artificial intelligence and NLP team by using Natural Language Processing, with only 655 of those patients having an MMSE score. Based on how close to the MCI diagnosis date was the MMSE score assessment date, and by setting a threshold of less than 60 days, a smaller dataset was created. As the resulting dataset columns were quite sparse, rows with a minimum number of missing values were selected resulting in a dataset of 356 records. Missing values were replaced with the numerical value of -100 since data imputation tests were futile in improving the accuracy of the machine learning models explained below.

Feature Selection & Experiment Configuration

Similarly to Prediction Model 1, for each classification task, two classes are defined. For the 0-1 years task, Class 1 (YES) is for those that converted from MCI to Dementia in less than 1 year and Class 0 (NO) for those that didn't, resulting in 142 YES and 214 NO instances. For the 1-5 years task, 1 (YES) is for

those that converted from MCI to Dementia between 1 and 5 years and 0 (NO) for those that didn't, resulting in 118 YES and 238 NO.

Dummy variables were created for the categorical variables using the 'get dummies' function of Python's pandas library. Normalization (centering and scaling) of the continuous variables was achieved by using the scale() function of Python's scikit-learn library. The dataset was split to training and test (70:30 percent) set using the 'train test split' function of Python's scikit-learn library.

We tried to approach the two classification tasks using 3 different algorithms: k-NN, Logistic Regression and Random Forest. For the K-Nearest Neighbors classifier algorithm, the hyperparameter grid was used to optimise the number of neighbours along with GridSearchCV (cv=5), the confusion matrix and classification report were also applied. For the Logistic Regression classifier algorithm, the hyperparameter grid was used to optimise the logspace and 'C' along with GridSearchCV (cv=5), the confusion matrix and classification report were also applied. For the Random Forest classifier algorithm, the hyperparameter grid was used to optimise the 'max depth' and 'random state' along with GridSearchCV (cv=5), the confusion matrix and classification report were also applied.

Evaluation

The evaluation results of the two classification tasks for the classifiers used can be found below:

Task	Classifier	Precision	Recall	F1-score
0-1 years	KNN	0.73	0.43	0.54
0-1 years	LogisticRegression	0.77	0.52	0.62
0-1 years	RandomForest	0.94	0.36	0.52
0-1 years	DecisionTrees	0.62	0.52	0.57
1-5 years	KNN	0.41	0.36	0.39
1-5 years	LogisticRegression	0.53	0.48	0.51
1-5 years	Random Forest	0.52	0.39	0.45
1-5 years	DecisionTrees	0.41	0.58	0.48

As can be observed, the Logistic Regression algorithm out performs others in both tasks, in terms of F1. Note that we report the scores for the positive class only.

Decision Trees

In the following Figures, we illustrate the decision trees that were obtained for 0-1 and 1-5 years tasks.

The first decision tree suggests that a large proportion of non-native English-speaking patients develop dementia within a year of presenting with MCI, a proportion that is most likely to be an artefact, perhaps caused by delayed referral of these individuals to memory clinics because of difficulty establishing their clinical status accurately in the community. The next branch point selects out a small sub-population who experience hallucinations. Whilst this could be a sign of advanced degenerative disease that has been misdiagnosed as MCI, it is more likely to indicate the presence of additional mental health problems, as the conversion rate in this group is similar to the average annual rate of conversion in longitudinal studies of MCI. The tree divides the remainder into a large group who do not report disabilities related to cognition, and a smaller group who do. It is unexpected that the latter group should exist within a population of MCI patients, as the diagnosis presupposes the absence of cognition related disability. However, the fact that more than half of this group are non-native English speakers raises the possibility that their original classification may have been inaccurate. The employment branch-point shows that it is only those who were not in employment at the time of MCI diagnosis who convert to dementia within a year, a finding that is probably related to age, as this appears at each of the following four branch points, first to select out a small number in the lowest age bracket (under 76.5) who appear to be at higher risk of conversion. Of the remainder, the 50% who already experience daily living difficulties (not due to cognitive problems, but probably indicative

of worse overall health status) convert to dementia within a year. The highest conversion risks, however, are associated with older age, which confers much higher than average conversion rates (57% if over 91.5, and 45% if aged between 83.5 and 89.5). These risks do not appear to be modified by social isolation though those with worse overall health status are at relatively lower risk, possibly because they receive supportive care and treatment for other conditions.

As mentioned, around half of all people with MCI can be expected to convert to dementia within five years of diagnosis.

The total number of patients identified as converting over this time period in the decision tree shown in the next Figure (the sum of the absolute numbers of converters in each of the end leaves) is 118, or 33% of the total, a rate which, although lower than average, is within the range of reported results. Age is identified as an important risk factor for conversion, as 30% of those aged under 77.5 convert, while 70% of those over this age do so. Among the latter the highest conversion rate occurs among individuals who already show signs of cognition-related disability (arguably a surrogate marker of already being on the cusp of conversion at the time of receiving a diagnosis of MCI). Among these individuals with probable mild dementia, neither social isolation (living alone) nor age appear to be standalone modifiers of the overall conversion risks. However, within the socially isolated group the presence of a behavioural disorder is associated with a higher conversion rate, and among those under 81.5, remaining in employment appears to increase the risk of conversion. Likewise, the presence of psychiatric symptomatology does not affect the overall conversion rate in this group, though of those with such symptomatology, depression (rather than any other symptom) confers the highest risk.