

Identifying the Presence and Severity of Dementia by Applying Interpretable Machine Learning Techniques on Structured Clinical Records

Akhilesh Vyas (✉ akhilesh.vyas@tib.eu)

Leibniz University Hannover

Fotis Aisopos

NCSR "Demokritos"

Maria-Esther Vidal

Leibniz University Hannover

Peter Garrard

Centre for Clinical Neuroscience, St George's University Hospital

George Paliouras

NCSR "Demokritos"

Research Article

Keywords: Dementia, Mini Mental Score; Machine learning, Data science; LIME; CAMCOG

Posted Date: March 22nd, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-336921/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

RESEARCH

Identifying the Presence and Severity of Dementia by Applying Interpretable Machine Learning Techniques on Structured Clinical Records

Akhilesh Vyas^{1,2*}, Fotis Aisopos^{3*}, Maria-Esther Vidal^{1,2*}, Peter Garrard⁴ and George Paliouras³

*Correspondence:

akhilesh.vyas@tib.eu;
fotis.aisopos@iit.demokritos.gr;
Maria.Vidal@tib.eu

¹L3S Research Center, Leibniz University Hannover, Hannover, Germany

³Software and Knowledge Engineering Laboratory, Institute of Informatics and Telecommunications, NCSR "Demokritos", Athens, Greece
 Full list of author information is available at the end of the article

Abstract

Background: Dementia occurs through declining of cognitive abilities, and its early detection stands essential for effective preventive measures. However, mainstream diagnostic tests and screening tools, such as Cambridge Cognition Examination (CAMCOG) and Mini-Mental State Examination (MMSE), often fail to detect dementia accurately. Various graph-based or feature-dependent prediction and progression models have been proposed. Whenever these models exploit information in the patients' electronic medical records, they represent promising options to identify the presence and severity of dementia more precisely.

Methods: The methods presented in this paper aim to address two problems related to dementia: (a) Basic diagnosis: identifying the presence of dementia in individuals, and (b) Severity diagnosis: predicting the presence of dementia, as well as the severity of the disease. We formulate these two tasks as classification problems and address them using machine learning models based on random forests, analysing structured clinical data from an elderly population cohort. We perform a hybrid data curation method where a dementia expert is in the loop to ensure meaningful curation decisions. We then employ the machine learning algorithms that classify individual episodes into a specific dementia class. Decision trees are also used for enhancing the explainability of decisions made by prediction models, allowing medical experts to identify the most crucial patient features and their threshold values for the detection of dementia.

Results: Our experiment results prove that clinical record features (such as age or comorbidities), along with baseline arithmetic or memory tests, can predict dementia and its severity with high accuracy. In specific, our prediction models have reached an average f1-score of 0.93 and 0.85 for problems (a) and (b), respectively. Moreover, the decision trees produced for the two issues empower the interpretability of the prediction models.

Conclusions: This study proves that there can be an accurate estimation of the existence and severity of dementia disease by analysing various clinical features and cognitive tests from the past episodes of the elderly population. Moreover, a set of rules represent building blocks in the efficiently patient classification. Relevant clinical and screening test features (e.g., simple arithmetic or animal fluency task) represent precise predictors without calculating the scores of mainstream cognitive tests such as MMSE and CAMCOG. They identify not only meaningful features but also justifications of classifications. As a result, the predictive power of machine learning models over curated clinical data is demonstrated, paving the way for accurate diagnosis of dementia.

Keywords: Dementia, Mini Mental Score; Machine learning, Data science; LIME; CAMCOG

Background

According to recent surveys, dementia is underdiagnosed. ICD codes alone cannot serve as a reliable gold standard for investigating the demographic characteristics or the clinical associations of the condition using electronic health records [1].

There is no single diagnostic test that can determine if a person has any form of dementia. However, clinicians employ various tools and tests to detect the presence of dementia, whether due to Alzheimer's disease or some other cause. The Mini-Mental State Examination (MMSE) is the most common test for measuring cognitive impairment [2]. Creavin et al. [3] attempt to determine the diagnostic accuracy of MMSE at various cut points for dementia in people aged 65 years and over. The authors conclude that MMSE contributes to a diagnosis of dementia in low prevalence settings. But this work also suggests that MMSE should not be used in isolation to confirm or exclude the disease.

An alternative approach is to use the Clock Drawing Test (CDT) [4], which provides a simple scoring system for the rapid screening for cognitive impairment in patients with mild cognitive impairment [5].

More recently, a variety of automatic speech-based tools have also been used to detect dementia. These approaches usually employ machine learning classifiers trained with various vocal features, derived from recorded data resulting from standard spoken tasks provided to individuals [6, 7].

Automatic speech-based tools can also use lexical and conversation analysis-inspired features derived from transcripts of recorded data [8] in conversations led by neurologists or intelligent virtual agents. As another example, the CogAware tool [9] provides textual analysis for transcripts of individuals describing the "cookie theft" picture [10] in order to automatically detect whether they are originated from a patient with dementia or a cognitively normal individual.

A powerful but underutilised resource that could be employed to rapidly and automatically detect dementia is a patient's electronic medical record (EMR). EMRs are increasingly available sources of information that contain large quantities of heterogeneous data. EMRs include MR images of the brain that demonstrate cortical atrophy. Moreover, EMRs comprise demographic and clinical information and cognitive test performance. This information has been used to train machine learning models to detect the presence and severity of dementia [11, 12].

Demographic and clinical information from EMRs have also been analysed by Shao et., al [1]; they utilise both structured and unstructured EMRs to define individual's risk scores for dementia. The authors also combine structured data features, consisted of standard codes and medications from EMRs, with topic features, extracted from a topic modelling approach on free-text clinical notes. Finally, they employ a logistic regression model using the selected features as predictors. A similar retrospective study is combined by Ford et., al [13], using structured data including medical diagnoses, primary care tests and investigations, lifestyle information and prescribing data. Their survey compared various machine-learning models with baseline epidemiological approaches to identify undetected dementia in UK primary care patients and concludes that logistic regression and random forest algorithms allow for important features to be exposed and may be the best approaches for this prediction task.

In this paper, we examine whether the values of various features in electronic medical records can consistently decide the patient's cognitive status, i.e. if the patient suffers from dementia and the level of its severity. This is realised in an automated way, by employing machine learning models that analyse a big dataset of EMRs, and validating the models' effectiveness.

For this purpose, we employ a set of EMRs that comprises structured information such as demographics, MMSE, and performance on the CAMCOG - a screening instrument for dementia. CAMCOG includes tests sensitive to different cognitive domains, and which forms part of the Cambridge Mental Disorders of the Elderly Examination (CAMDEX) [14]. The CAMCOG has high levels of sensitivity and specificity when used to distinguish individuals with mild dementia from those who are cognitively intact. We analyse these data using a random forests model, providing an automatic classifier that effectively discriminates dementia patients from control individuals and estimates the severity of the disease for the former.

However, machine learning algorithms usually work as a black-box tool, without the ability to interpret individual predictions. Thus, an emerging challenge is to achieve the explainability of decisions taken by such models, in order to provide clinicians with the ability to understand the rationale of the model. Towards this directions, we employ decision tree models that are able to visualise a set of configurable rules for predictions made. Since MMSE and CAMCOG scores can directly provide an estimation of dementia severity, we decided to exclude those from the training of our decision trees, in order to investigate other interesting features that are clinically useful.

Methods

Problem Statement

The problem addressed in this work is related to detecting the presence of a form of dementia in individuals based on a set of available demographic and psychometric features. A relevant extension of the problem above is accurately deciding the severity of the disease for patients with dementia. These two problems are formulated as follows:

- 1 **Problem (a): Predict Dementia, No Dementia:** This problem addresses the issue of identifying if a patient has 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)". We also identify relevant features for the classification decision.
- 2 **Problem (b): Predict No Dementia, Minimal or Mild Dementia, and Moderate or Severe Dementia:** We further refine the previous problem to have a better understanding of the severity of dementia. We classify dementia patients into two more classes "Minimal or Mild" and "Moderate or Severe". We discover important features behind the classification decision.

Approach

Our approach initially demands a data cleaning and curation process. Since all real-world clinical datasets contain a big amount of noise and missing values, we need to define a set of general rules, in order to be able to complete missing features and

ignore features that are sparse. This process includes the filling of stable parameters throughout the patient’s lifetime, such as height or gender, and removing parameters where the majority of values in patient episodes is meaningless (e.g. ‘not known’ or ‘not asked’).

The problems previously defined are formulated as simple classification tasks, addressed by machine learning models. These model’s algorithms are based on a supervised learning method Decision Tree (DT) to make the decisions easily interpretable by clinicians. They predict the class of patients by inferring decision rules from training data features.

A decision tree is composed of nodes and leaves. A node represents a dichotomous threshold for the value of some feature in the dataset. A leaf represents a patient subgroup in whom the likelihood of belonging to the positive class (in this case developing dementia) 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.

To understand the concept of decision trees, consider the example tree in Figure 1. In this tree, the goal is to decide whether an individual has some form of dementia or not (binary classification). As we can see, each intermediate node entails a discriminating feature and a threshold. Based on the value of this feature a clinician must follow the appropriate branch, until reaching a leaf node. Leaf nodes conclude to a decision (“Dementia” / “No Dementia”), based on the majority of cases in the training sample. To make our trees more informative and self-explainable, we have visualized the percentages in pie charts, the size of which depends on the number of cases falling in this leaf node. It is also used to find the class probability which is the fraction of the same class in a leaf.

Predictive Models

Our objective is to predict the class of dementia patients as described in the problem statement based on data features by employing machine learning predictive models. We utilize the random-forests algorithm for class prediction and local interpretable model-agnostic(LIME) [15] for the decision explanation of any patient. LIME utilizes local surrogate models to explain the black-box behaviour of the machine learning model and its prediction. A random forest algorithm is an ensemble learning method that utilizes different decision trees over the sub-sample of dataset instances. The decision of the algorithms averages the results provided by different decision trees. It also controls the over-fitting by having the trained over sub-sample of the dataset. We train our decision tree classifier feeding important data features obtained by the random forest classifier model.

Calculating classes for the predictive model The predictive models learn on patient-episode data and for the rest of the paper, we consider a patient-episode to be a training point. A patient in an episode is considered with “Dementia”, if and only if, there is at least one episode where the person has dementia. We consider patient-episodes from patients who have no history of dementia. As a result, the margin between a patient episode with and without dementia is normalised in the “No Dementia” class.

Table 1 Features and their respective values to consider a patient in dementia class

| Features Considered | Values considered | Total Cases |
|--------------------------|---------------------------------|-------------|
| EST SEVERITY OF DEMENTIA | Minimal, Mild, Moderate, Severe | 2,552 |
| DEMENTIA CLOUDED | Present | 10 |
| CLOUDED DEMENTIA | Present | 2 |
| SEVERITY OF DEMENTIA | Minimal, Mild, Moderate, Severe | 814 |
| DEMENTIA PRESENT | Mild, Moderate, Severe | 951 |
| MIXED DEMENTIA | Yes | 172 |
| DSM-IIIIR | Dementia | 1 |
| CLINICAL DIAGNOSIS 1 | Dementia | 876 |

Table 2 Features and their respective values to consider a patient in no-dementia class

| Features Considered | Values considered | Total Cases |
|--------------------------|-------------------|-------------|
| EST SEVERITY OF DEMENTIA | No | 2,084 |
| SEVERITY OF DEMENTIA | No | 958 |
| DEMENTIA PRESENT | No | 2,185 |
| DSM-IIIIR | No Dementia | 1 |
| CLINICAL DIAGNOSIS 1 | No Dementia | 965 |

Table 3 Features and their respective values to consider the severity of dementia

| Features Considered | Values considered | Total Cases |
|--------------------------|---------------------------------|-------------|
| EST SEVERITY OF DEMENTIA | Minimal, Mild, Moderate, Severe | 2,552 |
| SEVERITY OF DEMENTIA | Minimal, Mild, Moderate, Severe | 814 |

As there are no direct features to suggest if a patient episode is considered to be in the “Dementia” or “No Dementia” classes, we consider the features with the respective values from Tables 1, 2, and 3. Features and values in Tables 1, 2 are used to identify if a patient episode falls under the “Dementia” or “No Dementia” classes, respectively. Due to data inconsistencies, a patient episode might fall under both “Dementia” and “No Dementia” classes. we resolve this issue by considering the particular patient-episode into the “Dementia” class. Following the above assumptions, we find patient’s episodes with at least one episode of no-dementia and at least one episode of dementia, so we ignore these patient-episodes. Features and values in Table 3 are used to find the severity of dementia class. In case of inconsistencies in the attributes of severity, the maximum severity of both is considered.

Initially, each patient episode is labelled into “No Dementia” or “Dementia”, based on the value of specific features described in tables 1, 2, and 3, in order to produce the ground truth for our problems. For each of the problem statements, the baseline classes are formed from the labelled data. Classes of severity (Minimal, Mild, Moderate, Severe) are considered into the “Dementia” class, whereas no subclass is considered into the “No Dementia” class.

Experiments

Dataset

In our experiments, we have employed clinical data from the OPTIMA database. The OPTIMA dataset is a collection of 9,584 episodes from 1,035 patients who underwent serial evaluation incorporating the CAMDEX (and CAMCOG), in order to acquire a set of 1,593 data features for each episode of the patient. The features are derived from various sources of information, including demographic characteristics, YES/NO questions to health and well-being related questions, rating scales, and cognitive test performances. We predict that, in this patient group, only a subset of these features will be important for the prediction tasks. An important issue in

OPTIMA, is that the dataset contains 62% missing values of all possible features values. This occurs due to a variety of reasons. For example, clinicians may note down only the most relevant feature values about the patient's current condition and neglect others. They may also ignore some repeated feature across episodes such as gender, demographics and diseases (diabetes or hypertension).

Data cleaning and pre-processing:

Various related works examined in [16] show that attributes or features related to demographics (e.g. age, gender, and education), health (e.g. BMI, diabetes, depression, high cholesterol, and traumatic brain injury), and lifestyle factors (e.g. smoking, alcohol, physical activity, cognitive activity, and fish intake) are essential in the diagnosis of dementia. We identify 242 such features from OPTIMA features. Applying the rule based techniques suggested by the clinicians, we complete few of the missing values on important features. These curated features include Petersen MCI, Depressive Illness, Cerebro Vascular Disease Present, and Anxiety/Phobic. Following these techniques, around 70% of the missing values are completed. We also used the OPTIMA dataset variable guide (It consists of data type of features including range of numerical features and categories of categorical feature) to find the corrupt values in the data features. We replace all such values with missing values (NULL). For example, IDENTIFIES YEAR feature should have 'YES' and 'NO' values only, so all other cases are considered as missing values. For dichotomous features, such as gender, values are transformed to 0 and 1. As data sparsity can cause improper learning to the models, we ignore features which have more than 20% sparsity and drop episodes or data instances containing missing values. This procedure leaves our models with the features mentioned in the lists 1 and 2 for the Problem (a) and Problem (b), respectively. Further, In the case of pre-processing, we embed categorical features using one-hot encoding for an explainable predictive model.

Model Training and Feature Selection

We employ a random forest classifier to train our predictive models with the default parameters of sklearn library [1] except the maximum depth of the tree (value:5). A random forest model utilises a stratified shuffle split cross-validator which splits data into 5 folds of train and test set with reshuffling; each fold preserves the percentage of samples for each class. After each fold, we record 50 most impurity based important features from the model training and store into a set to have distinct feature. These important features are utilised to train the decision tree classification model with maximum depth 5 of the tree to represent a generalised decision tree to avoid overfitting. Following a most common approach as per the empirical study, we choose the split ratio (70:30) between the training and testing set and utilise a stratified method to preserve class frequencies in both sets. Both problems represent the same experimental settings. However, the models are trained on different population sizes; 1,052 and 2,034 patient episodes for Problem (a) and Problem (b), respectively. Different population sizes are used for the two problems. Thus, we minimise the number of null values, and their impact on models' training

[1] <https://scikit-learn.org/stable/>

and performance. Tables (4, 6, 5 and 7) report the results of both the problems using random forest and decision tree algorithms.

Evaluation Metrics We measure the performance of the predictive models in terms of precision, recall, f1-score and accuracy. In Problem (a), precision is the ratio of correctly predicted patients in the "Dementia" class to the total patients predicated as dementia. Alternatively, recall represents the ratio between the number of patients correctly predicted in the "Dementia" class to all patients in the "Dementia" class. Lastly, f1-score is the harmonic mean of precision and recall. The same metrics are also utilised for the classification of patients into the "NO dementia" class. Accuracy which is the between total correctly predicted patients to the total patients, is the least significant compared to the f1-score. We use the same metrics for Problem (b) in the problem statement.

Results

The random forest predictive model's performance is measured after each fold in terms of macro-averaged precision, recall and f1-score of each class. We represent the results of the random-forest models in Tables 4 and 5 for Problem (a) and Problem (b), respectively. The performance of the predictive models is very similar for each fold concerning both the problems. Further, we also show the results of the decision tree models for the two problems in Tables 6 and 7, respectively. The decision tree predictive model for Problem (a) provides similar results for both classes. The decision tree predictive model for Problem (b) accomplishes better results for "No dementia" and "Minimal or Mild dementia" classes, compared to "Moderate or Severe dementia" classes. Comparing the results of both classification reports tables (6 and 7) shows that Problem (b) is more complex to solve than Problem (a) due to the division of the "Dementia" class into the classes "Minimal or Mild" and "Moderate or Severe" dementia classes for Problem (b).

Table 4 Evaluation results for Problem (a) (Predict No-Dementia and Dementia) with 5-fold cross validation after each iteration of random-forest in terms of macro-averaged precision, recall, and f1-score

| Iterations | Precision | Random Forest | | |
|------------|-----------|---------------|----------|----------|
| | | Recall | f1-score | Accuracy |
| 1 | 0.95 | 0.95 | 0.95 | 0.95 |
| 2 | 0.96 | 0.96 | 0.96 | 0.96 |
| 3 | 0.94 | 0.94 | 0.94 | 0.94 |
| 4 | 0.96 | 0.97 | 0.96 | 0.96 |
| 5 | 0.95 | 0.95 | 0.95 | 0.95 |

Evaluation results

The resulting Decision Trees for Problems (a) and (b) are provided in Figures 1 and 2, respectively.

The decision tree of Figure 1, first separates the group into two main groups based on the strength of their verbal recall. Group one (n=581) recalled two or more of the six items (e.g. a shoe, a typewriter, a set of scales, a suitcase, a barometer, and a lamp) pictures of which they were asked to name and remember at the beginning

Table 5 Evaluation results for Problem (b) (Predict No Dementia, Minimal or Mild Dementia and Moderate or Severe Dementia) with 5-fold cross validation after each iteration of random-forest in terms of macro-averaged precision, recall and f1-score

| Iterations | Precision | Random Forest | | |
|------------|-----------|---------------|----------|----------|
| | | Recall | f1-score | Accuracy |
| 1 | 0.88 | 0.85 | 0.86 | 0.89 |
| 2 | 0.86 | 0.81 | 0.83 | 0.87 |
| 3 | 0.88 | 0.84 | 0.85 | 0.89 |
| 4 | 0.88 | 0.84 | 0.85 | 0.88 |
| 5 | 0.89 | 0.85 | 0.86 | 0.89 |

Table 6 Classification Report: Evaluation results for Problem (a) (Predict No-Dementia and Dementia) using decision-tree model in terms of precision, recall and f1-score

| | Precision | Decision Tree Model | | |
|--------------|-----------|---------------------|----------|---------|
| | | Recall | f1-score | support |
| No-Dementia | 0.92 | 0.94 | 0.93 | 216 |
| Dementia | 0.94 | 0.93 | 0.94 | 236 |
| accuracy | | | 0.93 | 452 |
| macro avg | 0.93 | 0.93 | 0.93 | 452 |
| weighted avg | 0.93 | 0.93 | 0.93 | 452 |

Table 7 Classification Report: Evaluation results for Problem (b) (Predict Predict No Dementia, Minimal or Mild Dementia and Moderate or Severe Dementia) using decision-tree model in terms of precision, recall and f1-score

| | Decision Tree Model | | | |
|-----------------------------|---------------------|--------|----------|---------|
| | Precision | Recall | f1-score | support |
| No-Dementia | 0.94 | 0.95 | 0.94 | 343 |
| Minimal or Mild Dementia | 0.84 | 0.88 | 0.86 | 380 |
| Moderate or Severe Dementia | 0.80 | 0.70 | 0.75 | 145 |
| accuracy | | | 0.88 | 868 |
| macro avg | 0.86 | 0.84 | 0.85 | 868 |
| weighted avg | 0.88 | 0.88 | 0.88 | 868 |

of the CAMCOG. Group two (n=471) recalled one or zero of these items. Not surprisingly, a large majority (around 95%) of the latter group are judged to suffer from dementia, while in the former the dementia diagnosis rate is only 16%. These distributions can be refined slightly by comparing the outcomes in groups based on other aspects of the assessment. For instance, the probability of dementia in a member of group 2 is much lower (30%), if the patient was oriented to time and generates nineteen or more items in the ‘animal fluency’ task. Such patients would have had an isolated memory impairment that did not impair their independence, and would, therefore, probably have met the criteria for amnesic mild cognitive impairment (MCI). However, patients of group two may fall into a pure Dementia class by following the specific rules defined in paths A or B, illustrated in the decision tree 1. Patients in path A were not oriented to time and could not even identify the current year while also not recalling basic objects like a barometer. Patients in path B were oriented to time but could not list at least 19 Animals and remember public figures.

Similarly, those group 1, who failed a mental arithmetic task and generated less than 16 items on animal fluency, had an 81% chance of suffering from dementia. Contrary, the dementia risk was 45% in those who passed mental arithmetic but generated less than 13 animals. The patients in group 1 who had the lowest risk (5%)

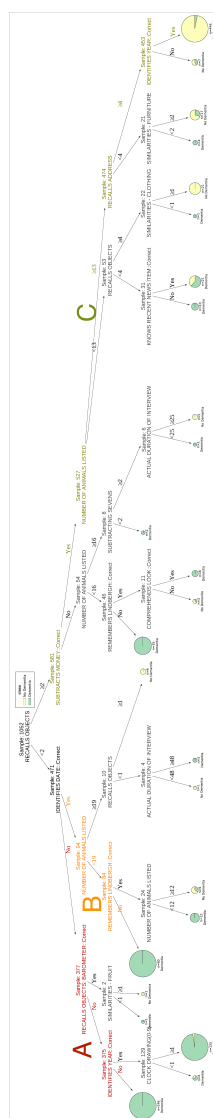


Figure 1 The decision tree predicts the dementia classes ("Dementia"/"No Dementia") of 1,052 patients for the Problem (a). The pie charts in leaves show class labels, the proportion of resulting classes and their support size. The branches demonstrate connections between features and their threshold values leading towards class labels. In the tree, some most significant paths A, B and C predict nearly pure decisive classes for large patients. The rules of path A and path B lead to a pure "Dementia" population of 246 and 60 patients, respectively, while the rules of path C lead to a large majority of "No Dementia" population comprised by 446 individuals.(see Additional file 1 in Supplementary Information)

of dementia were those who both passed the mental arithmetic test and generated 13 or more animals, as can be seen following path C.

In the decision tree of Figure 2, the outcome of the classification is changed such that diagnoses of dementia are divided into those with minimal or mild dementia and those with moderate to severe manifestations of the condition. As can be observed, the large majority of the moderate to severe cases are among the 1,076 participants on the left tree branch, scoring less than two on the verbal recall feature (A sub-test of the CAMCOG). However, the risk of severe dementia in these individuals ranges

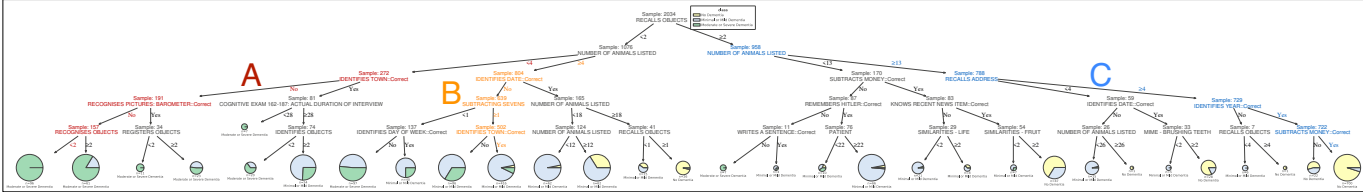


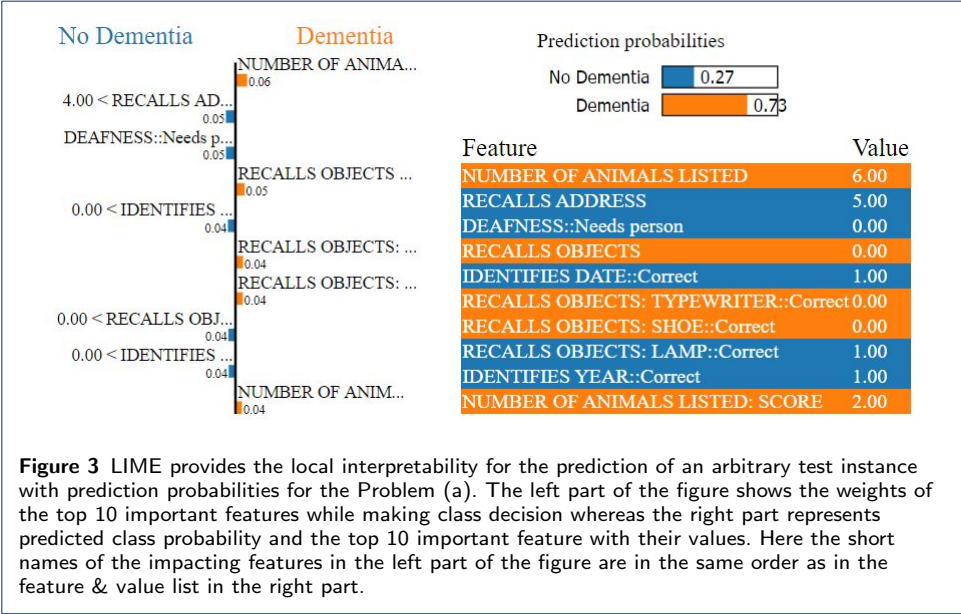
Figure 2 The decision tree predicts the dementia classes ("No Dementia" / "Minimal or Mild Dementia" / "Moderate or Severe Dementia") of 2,034 patients for the Problem (a). The pie charts in leaves show class labels, the proportion of resulting classes and their support size. The branches demonstrate connections between features and their threshold values leading towards class labels. In the tree, some most significant paths A, B and C predict nearly pure decisive classes for patients. The rules of path A and lead to a pure "Moderate or Severe Dementia" population of 96 patients. The rules of paths B and C lead to a large majority of "Mild or Minimal Dementia" and "No Dementia" populations, comprised by 416 and 700 individuals, respectively. (see Additional file 2 in Supplementary Information)

from 6% in those who list four or more animals and accurately perform simple mental arithmetic, to 87% when three or fewer animals are generated and the town that they are in cannot be recalled. For patients of the former group, path B rules lead to the highest possibility of minimal or mild dementia. For patients of the latter group, an absolute probability of moderate or severe dementia is provided if they are not able to identify their town and recognise simple objects and pictures, as shown in path A.

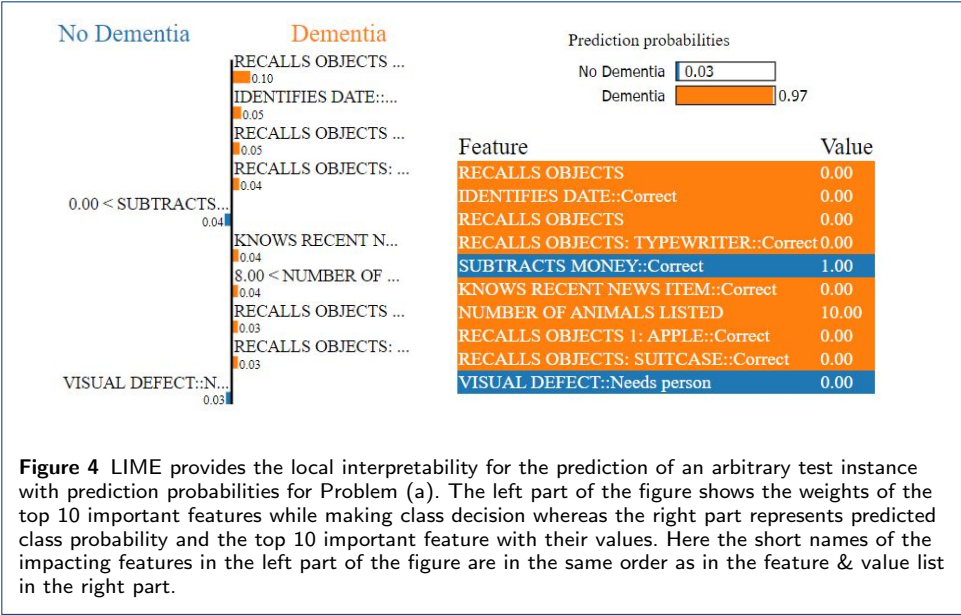
Likewise, the cases of severe dementia in the group with better verbal recall (two or more items) all occur in patients who list fewer than 13 items in the animal fluency task. Conversely, the group who are at least risk (less than 1%) of dementia of any degree are those with both good verbal fluency (13 or more animals generated) and accurate mental arithmetic, as can be seen following path C.

Model Explanation

To understand the machine learning model’s black-box behaviour, we perform local model interpretation over our random-forest model using LIME (Local Interpretable Model-Agnostic Explanations) [15]. We also make use of a decision tree classifier model for explaining the reasons behind our predictions. Here, we consider two examples from each problem’s testing set to understand the decision of our random forest model.



Figures 3, 4, 5, and 6 depict the contribution of the top 10 features for predicting the class for the respective problems. The right part of the figure represents the probabilities of the different classes and the names of the top 10 relevant features and their values. For example, in the right part of Figure 6, features in green colour support the "Moderate or Severe" class and features in other colour support "No Dementia" and "Minimal or Mild dementia" classes. The left part of Figure 6 measures the impact of these features in terms of weight while deciding for "NOT Moderate-Severe Dementia" (No-Dementia and Minimal or Mild Dementia) and "Moderate-Severe Dementia".

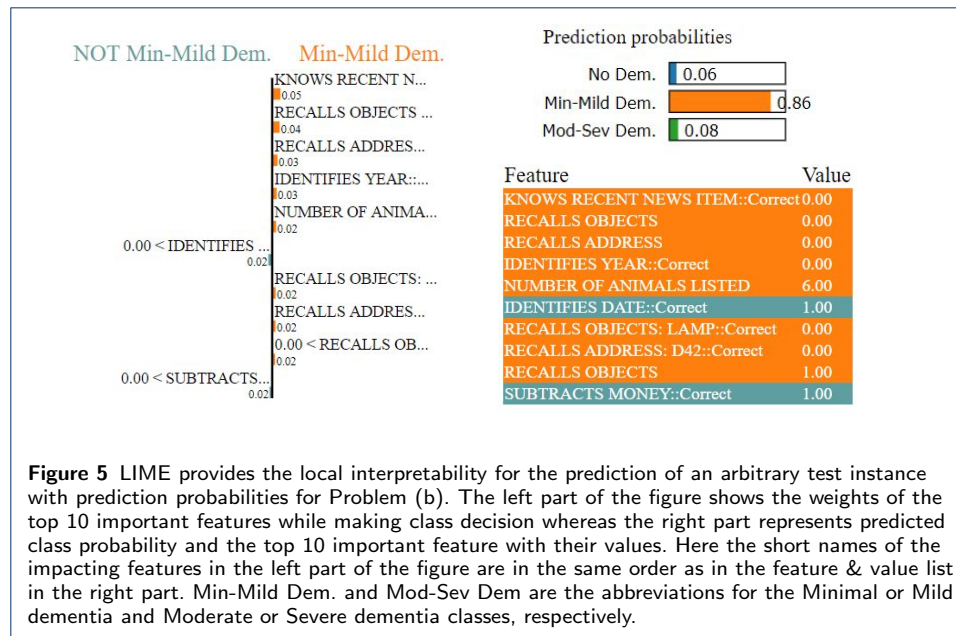


Discussion

It is axiomatic that the clinical management of all medical conditions needs to be tailored to the needs and characteristics of an individual patient. For example, drug doses always differ between age groups and are often treated by body weight. Long-acting, once-a-day drug formulations drug can help to maintain compliance in certain groups of patients. The choice of treatment in cancer often depends on the presence or absence of molecular markers in the tumour itself. As the basis of more and more of these individual differences are understood, drug treatments will become more personalised. Thus, the poor rates of response in common conditions such as depression, in which a ‘one size fits all’ approach is currently the norm (i.e. providing the same treatment independently of specific patient characteristics), should begin to improve.

The biology and pathophysiology of dementia and its many underlying causes (e.g, Alzheimer’s disease that is the most common, at least in later life) are diverse and subject to the influence of different factors (e.g., comorbidity, lifestyle, and genetics). They demand individualised and precise treatment to differentiate the conditions of each dementia patient. The approach described in this paper has shown the unpredictability that derives from a complex interplay of factors impacts on: i) the accuracy and efficiency of diagnosis, and ii) accuracy in prognostication. Improvements in all these areas will significantly enhance a clinician’s ability to offer the management plan most appropriate to an individual patient, both at present and when disease-modifying treatments become available for sporadic neurodegenerative dementias.

To be given a diagnosis of any untreatable condition is a life-changing event. When the condition is neurodegenerative dementia, patients need to make changes to their current lives and future plans, to consider legal rights appropriate to their future selves, and if possible maximise the utility of their residual cognitive resources. While the disease remains untreatable, there are negative as well as positive aspects to early diagnosis, and where accuracy is concerned, it is more important to

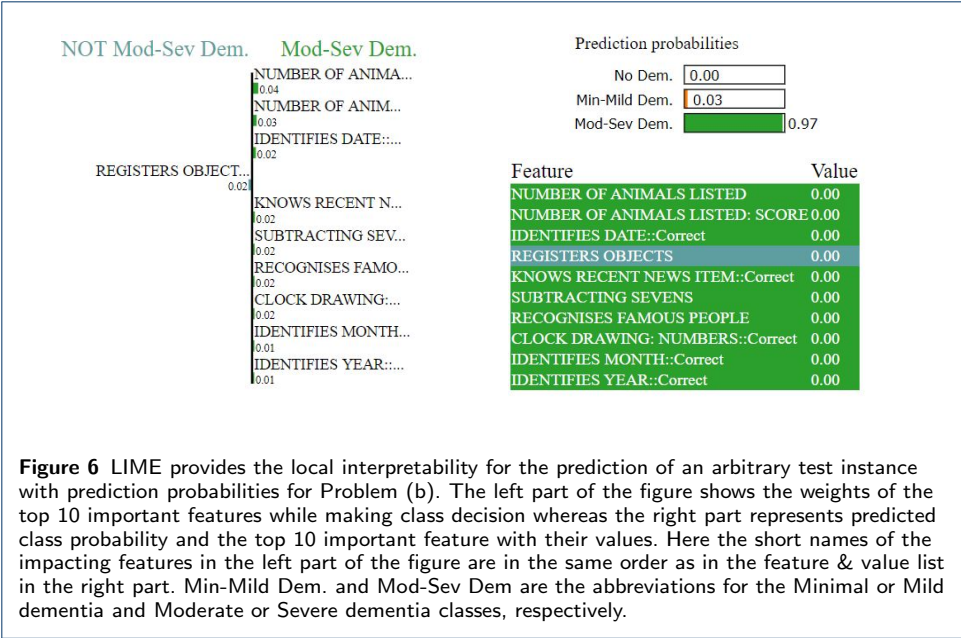


avoid false-positive than false-negative diagnoses. The advent of disease-modifying treatments, however, will change this: identifying a neurodegenerative condition as early as possible in the course of its evolution will limit the damage already done and therefore improve the outcome of treatment. At the same time, it will become as important not to miss a true positive as to misclassify a true negative.

Once diagnosed, a patient with Alzheimer's disease needs to be kept under regular clinical assessment. Whether conducted in the context of primary care or a specialist clinic, ongoing assessment aims to ensure that the patient's social and medical needs (including the choice and dose of symptomatic treatments) are optimally aligned with their cognitive abilities. Because current methods of assessment are both time-consuming and prone to inaccuracy and error, an a priori estimate of the likely trajectory of decline would alert the clinician to any anomalous results and therefore ensure that the best decisions are always made at the most appropriate time.

Conclusions and Future Directions

This paper presented automated prediction models for detecting the presence of dementia in the Electronic Medical Records of patients of a large ageing study, based on psychometric tests and demographic factors. Our study focused both on the accuracy, by employing different machine learning techniques, and interpretability, by visualising resulting models with the method of decision trees. The decision trees produced identified the most discriminating - and thus important - features for dementia detection as well as the disease severity classification. These rules include various memory (e.g. recollection of objects) and arithmetic (e.g. subtraction of money) tasks that patients have been asked to take, with their performance determining the sub-group in which they fall with certain probability. Sub-groups define either a specific severity level of the disease or a non-dementia condition (e.g. patients with MCI).



Our work aims at thoroughly investigating and highlighting key characteristics that yield the presence and severity of dementia and creating an accurate prediction tool. Moreover, the decision tree approach ignores mainstream cognitive tests such as MMSE and CAMCOG, employed in most of the related work, in order to focus on simple rules, represented by simple arithmetic and memory tasks. Such a rule-based tool can easily assist clinicians in the early detection of dementia in primary care. To adopt this approach, the end-user of this tool would only have to assign the tasks depicted in the decision trees to her patients, in order to assess their condition with a certain confidence (based on the proportion of each sub-group in the tree). Thus, we hope that these results and tool can represent building blocks for individualised clinical decisions.

As for future challenges, we will validate the proposed prediction models, including random forests and interpretable decision trees, in other patient cohorts. An extensive evaluation across different populations would ensure the current approach does not suffer from a lack of scientific validity. More importantly, an extensive assessment will provide empirical proof of the generality of the properties of the proposed methods.

Abbreviations

BMI: Body mass index; CAMCOG: Cambridge Cognition Examination; CAMDEX: Cambridge Mental Disorders of the Elderly Examination; CDT: Clock Drawing Test; DT: Decision Tree; EMR: Electronic Medical Record; LIME: Local Interpretable Model-agnostic; MCI: Mild Cognitive Impairment; MMSE: Mini-Mental State Examination.

Supplementary Information

- Additional file 1
Decision Tree for the problem (a) in pdf format for better visibility.
- Additional file 2
Decision Tree for the problem (b) in pdf format for better visibility.

Declarations

Acknowledgements

The authors would like to acknowledge the project IASIS' consortium for their collaboration in this work.

Authors' contributions

All the authors defined the problems addressed in this paper. Fotis Aisopos, Akhilesh Vyas, and Maria-Esther Vidal devised the solution. Akhilesh Vyas implemented the machine learning models and executed all the experiments. Peter Garrard analysed and explained the results from the clinical perspective. Lastly, all the authors contributed to writing and editing the manuscript; they also read and approved the final manuscript.

Authors' information

Affiliations

L3S Research Center (Leibniz University Hannover), Scientific Data Management research group(TIB-Leibniz Information Centre for Science and Technology)

Prof. Maria-Esther Vidal & Akhilesh Vyas

Software and Knowledge Engineering Laboratory (Institute of Informatics and Telecommunications, NCSR "Demokritos")

Fotis Aisopos & Dr. George Paliouras

Centre for Clinical Neuroscience (St George's University Hospital)

Dr. Peter Garrard

Funding

This paper is supported by European Union's Horizon 2020 research and innovation programme under grant agreement No. 727658, project IASIS (Integration and analysis of heterogeneous big data for precision medicine and suggested treatments for different types of patients).

Availability of data and materials

Code is available on www.github.com/SDM-TIB/dementia_detection.git. The data that support the findings of this study were provided from the Oxford Project to Investigate Memory and Ageing (OPTIMA) via a bilateral agreement with the IASIS project and cannot be publicly shared.

Ethics approval and consent to participate

Permission to use the dataset supporting our work was granted by OPTIMA. Data collection from participants follows the ethics standards of the OPTIMA study.

Competing interests

The authors declare that they have no competing interests.

Consent for publication

Not applicable.

Author details

¹L3S Research Center, Leibniz University Hannover, Hannover, Germany. ²Scientific Data Management research group, TIB-Leibniz Information Centre for Science and Technology, Hannover, Germany. ³Software and Knowledge Engineering Laboratory, Institute of Informatics and Telecommunications, NCSR "Demokritos", Athens, Greece. ⁴Centre for Clinical Neuroscience, St George's University Hospital, London, UK.

References

- Shao, Y., Zeng, Q.T., Chen, K.K., Shutes-David, A., Thielke, S.M., Tsuang, D.W.: Detection of probable dementia cases in undiagnosed patients using structured and unstructured electronic health records. *BMC medical informatics and decision making* **19**(1), 128 (2019)
- Arevalo-Rodriguez, I., Smailagic, N., i Figuls, M.R., Ciapponi, A., Sanchez-Perez, E., Giannakou, A., Pedraza, O.L., Cosp, X.B., Cullum, S.: Mini-mental state examination (mmse) for the detection of alzheimer's disease and other dementias in people with mild cognitive impairment (mci). *Cochrane Database of Systematic Reviews* (3) (2015)
- Creavin, S.T., Wisniewski, S., Noel-Storr, A.H., Trevelyan, C.M., Hampton, T., Rayment, D., Thom, V.M., Nash, K.J., Elhamoui, H., Milligan, R., et al.: Mini-mental state examination (mmse) for the detection of dementia in clinically unevaluated people aged 65 and over in community and primary care populations. *Cochrane Database of Systematic Reviews* (1) (2016)
- Borson, S., Brush, M., Gil, E., Scanlan, J., Vitaliano, P., Chen, J., Cashman, J., Sta Maria, M.M., Barnhart, R., Roques, J.: The clock drawing test: utility for dementia detection in multiethnic elders. *Journals of Gerontology Series A: Biomedical Sciences and Medical Sciences* **54**(11), 534–540 (1999)
- Rakusa, M., Jensterle, J., Mlakar, J.: Clock drawing test: a simple scoring system for the accurate screening of cognitive impairment in patients with mild cognitive impairment and dementia. *Dementia and geriatric cognitive disorders* **45**(5–6), 326–334 (2018)
- Satt, A., Sorin, A., Toledo-Ronen, O., Barkan, O., Kompatsiaris, I., Kokonozi, A., Tsolaki, M.: Evaluation of speech-based protocol for detection of early-stage dementia. In: *Interspeech*, pp. 1692–1696 (2013)
- Satt, A., Hoory, R., König, A., Aalten, P., Robert, P.H.: Speech-based automatic and robust detection of very early dementia. In: *Fifteenth Annual Conference of the International Speech Communication Association* (2014)
- Mirheidari, B., Blackburn, D., Walker, T., Reuber, M., Christensen, H.: Dementia detection using automatic analysis of conversations. *Computer Speech & Language* **53**, 65–79 (2019)

9. Rentoumi, V., Paliouras, G., Danasi, E., Arfani, D., Fragkopoulou, K., Varlokosta, S., Papadatos, S.: Automatic detection of linguistic indicators as a means of early detection of alzheimer's disease and of related dementias: A computational linguistics analysis. In: 2017 8th IEEE International Conference on Cognitive Infocommunications (CogInfoCom), pp. 000033–000038 (2017). IEEE
10. Goodglass, H., Kaplan, E., Barresi, B.: Boston Diagnostic Aphasia Examination Record Booklet. Lippincott Williams & Wilkins, ??? (2001)
11. Möller, C., Pijnenburg, Y.A., van der Flier, W.M., Versteeg, A., Tijms, B., de Munck, J.C., Hafkemeijer, A., Rombouts, S.A., van der Grond, J., van Swieten, J., et al.: Alzheimer disease and behavioral variant frontotemporal dementia: automatic classification based on cortical atrophy for single-subject diagnosis. *Radiology* **279**(3), 838–848 (2016)
12. Shankle, W.R., Mani, S., Dick, M.B., Pazzani, M.J.: Simple models for estimating dementia severity using machine learning. *MedInfo* **98** (1998)
13. Ford, E., Rooney, P., Oliver, S., Hoile, R., Hurley, P., Banerjee, S., van Marwijk, H., Cassell, J.: Identifying undetected dementia in uk primary care patients: a retrospective case-control study comparing machine-learning and standard epidemiological approaches. *BMC medical informatics and decision making* **19**(1), 248 (2019)
14. Huppert, F.A., Jorm, A.F., Brayne, C., Gilling, D.M., Barkley, C., Beardsall, L., Paykel, E.S.: Psychometric properties of the camcog and its efficacy in the diagnosis of dementia. *Aging, Neuropsychology, and Cognition* **3**(3), 201–214 (1996)
15. Ribeiro, M.T., Singh, S., Guestrin, C.: "why should I trust you?": Explaining the predictions of any classifier. *CoRR abs/1602.04938* (2016). [1602.04938](https://arxiv.org/abs/1602.04938)
16. Warren-Gash, C.: Dementia risk prediction models what do policymakers need to know? Technical report, University of Cambridge (March 2019). <https://www.phgfoundation.org/documents/dementia-risk-prediction-models.pdf>

Appendix

Lists

Listing 1 Complete list of selected attributes for Problem (a)

```
[ 'IDENTIFIES OBJECTS', 'RECALLS ADDRESS', 'SIMILARITIES — CLOTHING',
  'RECOGNISES FAMOUS PEOPLE', 'RECOGNISES OBJECTS', 'RECALLS OBJECTS',
  'NUMBER OF ANIMALS LISTED', 'SIMILARITIES — LIFE', 'PATIENT', 'SIMILARITIES
  — FURNITURE', 'REGISTERS OBJECTS', 'IDENTIFIES COIN', 'Age At Episode',
  'SUBTRACTING SEVENS', 'CLOCK DRAWING', 'COGNITIVE EXAM 162–187: ACTUAL
  DURATION OF INTERVIEW', 'MIME — BRUSHING TEETH', 'RECALLS OBJECTS', 'PRAXIS
  — PAPER', 'NUMBER OF ANIMALS LISTED: SCORE', 'SIMILARITIES — FRUIT',
  'dementia.range', 'RECALLS ADDRESS: BROWN::Correct', 'DEAFNESS::Needs aid',
  'DEAFNESS::Needs person', 'DEAFNESS::Normal', 'TREMOR::Needs aid',
  'TREMOR::Needs person', 'TREMOR::Normal', 'CLOCK DRAWING: CIRCLE::Correct',
  'RECOGNISES OBJECTS: PIPE::Correct', 'NAMES PICTURES: SUITCASE::Correct',
  'KNOWS RECENT NEWS ITEM::Correct', 'RECALLS OBJECTS 2: TABLE::Correct',
  'DIAGNOSIS 334–351: DEPRESSIVE ILLNESS::Absent', 'DIAGNOSIS 334–351:
  DEPRESSIVE ILLNESS::Present', 'PRAXIS — PAPER: RIGHT HAND::Correct',
  'PLANTAR REFLEXES::Both extensor', 'PLANTAR REFLEXES::Extensor l', 'PLANTAR
  REFLEXES::Extensor r', 'PLANTAR REFLEXES::Normal', 'RECOGNISES OBJECTS:
  CUP::Correct', 'NAMES PICTURES: BAROMETER::Correct', 'HEMIPARESIS::Both',
  'HEMIPARESIS::Left', 'HEMIPARESIS::None', 'HEMIPARESIS::Right', 'DIAGNOSIS
  334–351: ANXIETY/PHOBIC::Absent', 'DIAGNOSIS 334–351:
  ANXIETY/PHOBIC::Present', 'REGISTERS OBJECTS 3: PENNY::Correct',
  'COMPREHENDS TOUCH::Correct', 'REMEMBERS WW2 DATE::Correct', 'MANUAL
  DIFFICULTY::Correct', 'RECALLS OBJECTS 3: PENNY::Correct', 'RECALLS
  OBJECTS: SUITCASE::Correct', 'IDENTIFIES YEAR::Correct', 'RECOGNISES
  OBJECTS: PURSE::Correct', 'RECOGNISES OBJECTS: SPECTACLES::Correct',
  'RECOGNISES OBJECTS: SHOE::Correct', 'RECALLS OBJECTS: LAMP::Correct',
  'REMEMBERS MAE WEST::Correct', 'SHORTNESS OF BREATH::Correct', 'RECALLS
  OBJECTS: TYPEWRITER::Correct', 'READING COMPREHENSION 2::Correct', 'KNOWS
  PRIME MINISTER::Correct', 'COMPREHENDS NOD::Correct', 'DRAWS
  SPIRAL::Correct', 'CLOCK DRAWING: NUMBERS::Correct', 'IDENTIFIES
  STREETS/COUNTRY::Correct', 'GAIT::Normal', 'GAIT::Other', 'GAIT::Wide
  based', 'COMPREHENDS LOOK::Correct', 'READING COMPREHENSION 1::Correct',
  'COMPREHENDS VILLAGE::Correct', 'KNOWS MONARCH::Correct', 'RECALLS OBJECTS
  1: APPLE::Correct', 'KNOWS HEIR TO THRONE::Correct', 'IDENTIFIES OBJECTS:
  PENCIL::Correct', 'REMEMBERS WW1 DATE::Correct', 'NAMES PICTURES:
  SHOE::Correct', 'DEFINES HAMMER::Correct', 'SUBTRACTS MONEY::Correct',
  'MOBILITY::Needs aid', 'MOBILITY::Needs person', 'MOBILITY::Normal',
  'RECALLS OBJECTS: SCALES::Correct', 'RECALLS OBJECTS: BAROMETER::Correct',
  'RECOGNISES OBJECTS: TELEPHONE::Correct', 'NAMES PICTURES: LAMP::Correct',
  'RECALLS ADDRESS: BEDFORD::Correct', 'ABNORMAL EYE MOVEMENTS::Correct',
  'IDENTIFIES COUNTY::Correct', 'WRITES A SENTENCE::Correct', 'PRAXIS —
  PAPER: ON LAP::Correct', 'RECALLS ADDRESS: D42::Correct', 'DRAWS
  HOUSE::Correct', 'VISUAL DEFECT::Needs aid', 'VISUAL DEFECT::Needs person',
  'VISUAL DEFECT::Normal', 'IDENTIFIES SEASON::Correct', 'TENDON
```

```

REFLEXES::Correct', 'IDENTIFIES OBJECTS: WATCH::Correct', 'IDENTIFIES
DATE::Correct', 'COMPREHENDS TAP::Correct', 'IDENTIFIES TOWN::Correct',
'BLOOD PRESSURE::Abnormal', 'BLOOD PRESSURE::Normal', 'IDENTIFIES
FLOOR::Correct', 'RECALLS OBJECTS: SHOE::Correct', 'ADDS UP
MONEY::Correct', 'COUNTING BACKWARDS::> two errors', 'COUNTING
BACKWARDS::Correct', 'COUNTING BACKWARDS::One error', 'MIME –
WAVE::Correct', 'REPETITION::Correct', 'IDENTIFIES MONTH::Correct',
'REGISTERS OBJECTS 1: APPLE::Correct', 'CLOCK DRAWING: TIME::Correct',
'REMEMBERS HITLER::Correct', 'MIME – SCISSORS::Correct', 'MIME –
SCISSORS::Poor', 'RECALLS ADDRESS: JOHN::Correct', 'RECOGNISES PICTURES:
SCALES::Correct', 'REMEMBERS STALIN::Correct', 'COMPREHENDS
HOTEL::Correct', 'RECALLS ADDRESS: WEST::Correct', 'DICTATION::Correct',
'DICTATION::Poor', 'NAMES PICTURES: SCALES::Correct', 'IDENTIFIES DAY OF
WEEK::Correct', 'PRAXIS – ENVELOPE::Correct', 'REMEMBERS
LINDBERGH::Correct', 'IDENTIFIES PRESENT PLACE::Correct', 'DRAWS
PENTAGON::Correct', 'COMPREHENDS RADIO::Correct', 'RECOGNISES PICTURES:
BAROMETER::Correct', 'PRAXIS – PAPER: FOLDS::Correct', 'NAMES PICTURES:
TYPEWRITER::Correct', 'RECOGNISES PICTURES: SHOE::Correct']

```

Listing 2 Complete list of selected features for Problem (b)

```

['RECALLS OBJECTS', 'CLOCK DRAWING', 'SUBTRACTING SEVENS', 'SIMILARITIES –
FRUIT', 'NUMBER OF ANIMALS LISTED: SCORE', 'NUMBER OF ANIMALS LISTED',
'SIMILARITIES – LIFE', 'IDENTIFIES OBJECTS', 'MIME – BRUSHING TEETH',
'RECALLS ADDRESS', 'dementia_range', 'PRAXIS – PAPER', 'RECOGNISES
OBJECTS', 'IDENTIFIES COIN', 'SIMILARITIES – CLOTHING', 'PATIENT', 'RECALLS
OBJECTS', 'REGISTERS OBJECTS', 'RECOGNISES FAMOUS PEOPLE', 'SIMILARITIES –
FURNITURE', 'COGNITIVE EXAM 162–187: ACTUAL DURATION OF INTERVIEW', 'Age At
Episode', 'RECALLS ADDRESS: BROWN::Correct', 'PRAXIS – ENVELOPE::Correct',
'IDENTIFIES FLOOR::Correct', 'KNOWS PRIME MINISTER::Correct', 'COMPREHENDS
TAP::Correct', 'RECALLS OBJECTS: BAROMETER::Correct', 'RECOGNISES OBJECTS:
PIPE::Correct', 'REPETITION::Correct', 'REMEMBERS WW2 DATE::Correct',
'IDENTIFIES MONTH::Correct', 'DRAWS HOUSE::Correct', 'CLOCK DRAWING:
TIME::Correct', 'RECOGNISES PICTURES: SCALES::Correct', 'RECALLS OBJECTS:
SHOE::Correct', 'RECALLS ADDRESS: JOHN::Correct', 'DEFINES
HAMMER::Correct', 'RECOGNISES PICTURES: SHOE::Correct', 'IDENTIFIES
OBJECTS: PENCIL::Correct', 'PRAXIS – PAPER: FOLDS::Correct', 'WRITES A
SENTENCE::Correct', 'READING COMPREHENSION 2::Correct', 'REGISTERS OBJECTS
1: APPLE::Correct', 'DICTATION::Correct', 'DICTATION::Poor', 'COMPREHENDS
RADIO::Correct', 'RECOGNISES PICTURES: BAROMETER::Correct', 'IDENTIFIES
DATE::Correct', 'IDENTIFIES YEAR::Correct', 'IDENTIFIES
STREETS/COUNTRY::Correct', 'DRAWS PENTAGON::Correct', 'COMPREHENDS
VILLAGE::Correct', 'RECALLS OBJECTS: TYPEWRITER::Correct', 'RECOGNISE
PERSON::Correct', 'IDENTIFIES COUNTY::Correct', 'REMEMBERS MAE
WEST::Correct', 'KNOWS MONARCH::Correct', 'RECALLS ADDRESS: D42::Correct',
'RECALLS OBJECTS: SUITCASE::Correct', 'COUNTING BACKWARDS::> two errors',
'COUNTING BACKWARDS::Correct', 'COUNTING BACKWARDS::One error', 'KNOWS
RECENT NEWS ITEM::Correct', 'DIAGNOSIS 334–351: DEPRESSIVE
ILLNESS::Absent', 'DIAGNOSIS 334–351: DEPRESSIVE ILLNESS::Present',
'RECOGNISES OBJECTS: SHOE::Correct', 'COMPREHENDS NOD::Correct',
'RECOGNISES OBJECTS: TELEPHONE::Correct', 'REGISTERS OBJECTS 3:
PENNY::Correct', 'MIME – SCISSORS::Correct', 'MIME – SCISSORS::Poor',
'RECALLS ADDRESS: WEST::Correct', 'KNOWS HEIR TO THRONE::Correct', 'NAMES
PICTURES: LAMP::Correct', 'RECOGNISES OBJECTS: PURSE::Correct', 'CLOCK
DRAWING: NUMBERS::Correct', 'RECALLS OBJECTS 3: PENNY::Correct', 'CLOCK
DRAWING: CIRCLE::Correct', 'PRAXIS – PAPER: RIGHT HAND::Correct', 'READING
COMPREHENSION 1::Correct', 'REMEMBERS HITLER::Correct', 'PRAXIS – PAPER: ON
LAP::Correct', 'DIAGNOSIS 334–351: ANXIETY/PHOBIC::Absent', 'DIAGNOSIS
334–351: ANXIETY/PHOBIC::Present', 'RECALLS OBJECTS: LAMP::Correct',
'REMEMBERS LINDBERGH::Correct', 'NAMES PICTURES: TYPEWRITER::Correct',
'RECOGNISES OBJECTS: CUP::Correct', 'IDENTIFIES TOWN::Correct', 'IDENTIFIES
OBJECTS: WATCH::Correct', 'COMPREHENDS TOUCH::Correct', 'COMPREHENDS
HOTEL::Correct', 'NAMES PICTURES: SHOE::Correct', 'DRAWS SPIRAL::Correct',
'RECALLS ADDRESS: BEDFORD::Correct', 'SUBTRACTS MONEY::Correct', 'RECALLS
OBJECTS 2: TABLE::Correct', 'NAMES PICTURES: BAROMETER::Correct',
'IDENTIFIES SEASON::Correct', 'MIME – WAVE::Correct', 'RECOGNISES OBJECTS:
SPECTACLES::Correct', 'REMEMBERS STALIN::Correct', 'IDENTIFIES PRESENT
PLACE::Correct', 'ADDS UP MONEY::Correct', 'COMPREHENDS LOOK::Correct',
'NAMES PICTURES: SCALES::Correct', 'NAMES PICTURES: SUITCASE::Correct',
'REMEMBERS WW1 DATE::Correct', 'RECALLS OBJECTS 1: APPLE::Correct',
'IDENTIFIES DAY OF WEEK::Correct', 'RECALLS OBJECTS: SCALES::Correct']

```

Figures

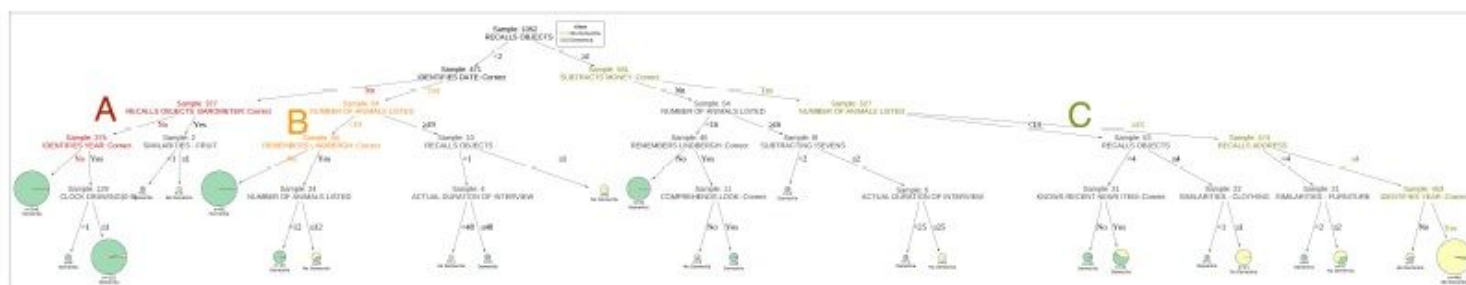


Figure 1

The decision tree predicts the dementia classes ("Dementia"/"No Dementia") of 1,052 patients for the Problem (a). The pie charts in leaves show class labels, the proportion of resulting classes and their support size. The branches demonstrate connections between features and their threshold values leading towards class labels. In the tree, some most significant paths A, B and C predict nearly pure decisive classes for large patients. The rules of path A and path B lead to a pure "Dementia" population of 246 and 60 patients, respectively, while the rules of path C lead to a large majority of "No Dementia" population comprised by 446 individuals.(see Additional le 1 in Supplementary Information)

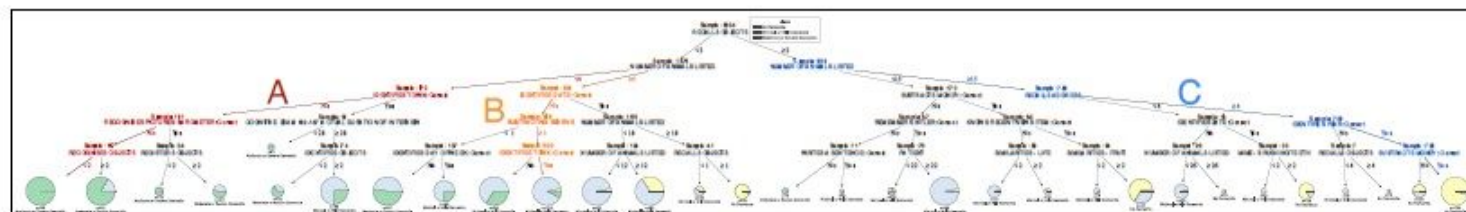


Figure 2

The decision tree predicts the dementia classes ("No Dementia"/ "Minimal or Mild Dementia" / "Moderate or Severe Dementia") of 2,034 patients for the Problem (a). The pie charts in leaves show class labels, the proportion of resulting classes and their support size. The branches demonstrate connections between features and their threshold values leading towards class labels. In the tree, some most significant paths A, B and C predict nearly pure decisive classes for patients. The rules of path A and lead to a pure "Moderate or Severe Dementia" population of 96 patients. The rules of paths B and C lead to a large majority of "Mild or Minimal Dementia" and "No Dementia" populations, comprised by 416 and 700 individuals, respectively. (see Additional file 2 in Supplementary Information)

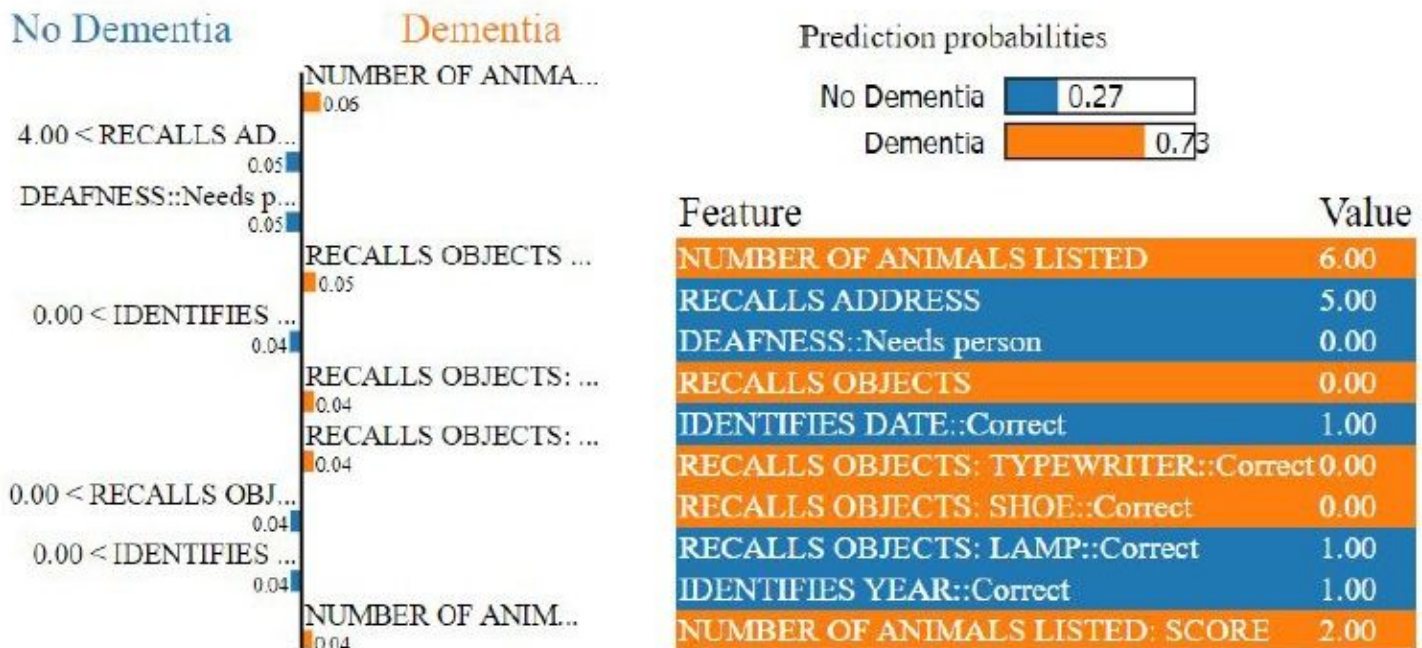


Figure 3

LIME provides the local interpretability for the prediction of an arbitrary test instance with prediction probabilities for the Problem (a). The left part of the figure shows the weights of the top 10 important features while making class decision whereas the right part represents predicted class probability and the top 10 important feature with their values. Here the short names of the impacting features in the left part of the figure are in the same order as in the feature & value list in the right part.

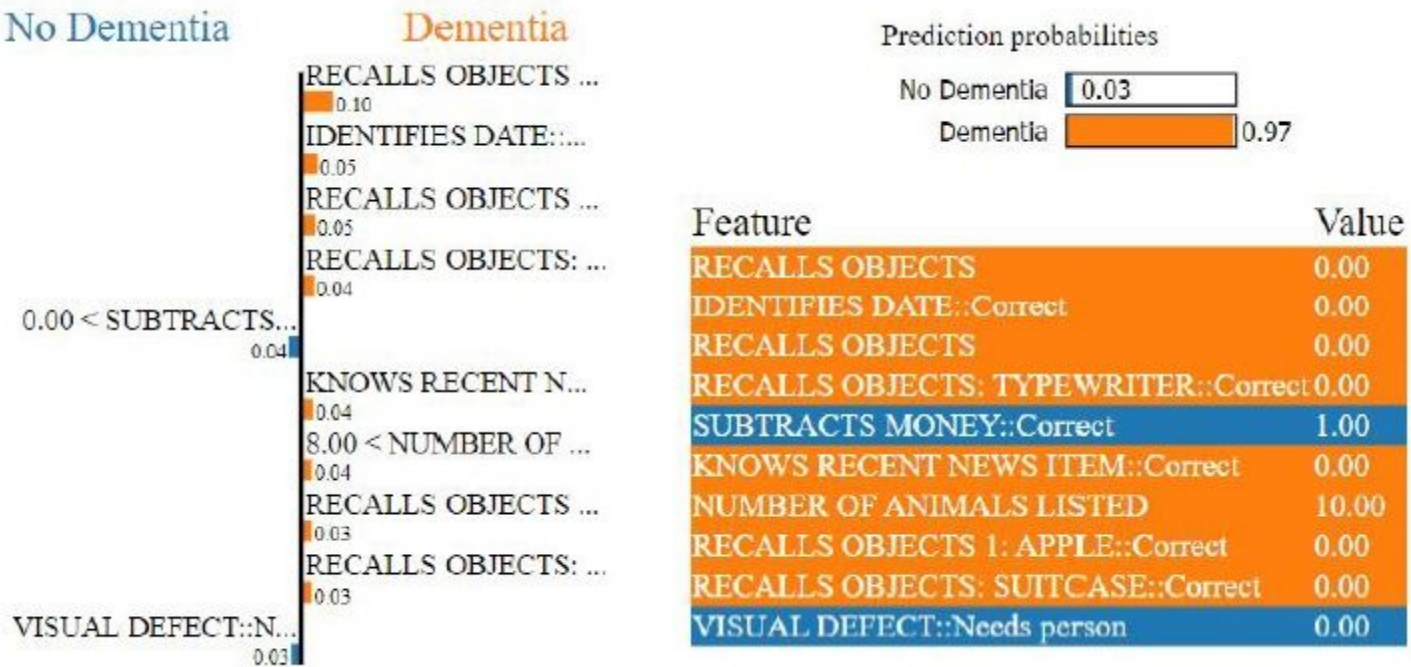


Figure 4

LIME provides the local interpretability for the prediction of an arbitrary test instance with prediction probabilities for Problem (a). The left part of the gure shows the weights of the top 10 important features while making class decision whereas the right part represents predicted class probability and the top 10 important feature with their values. Here the short names of the impacting features in the left part of the gure are in the same order as in the feature & value list in the right part.

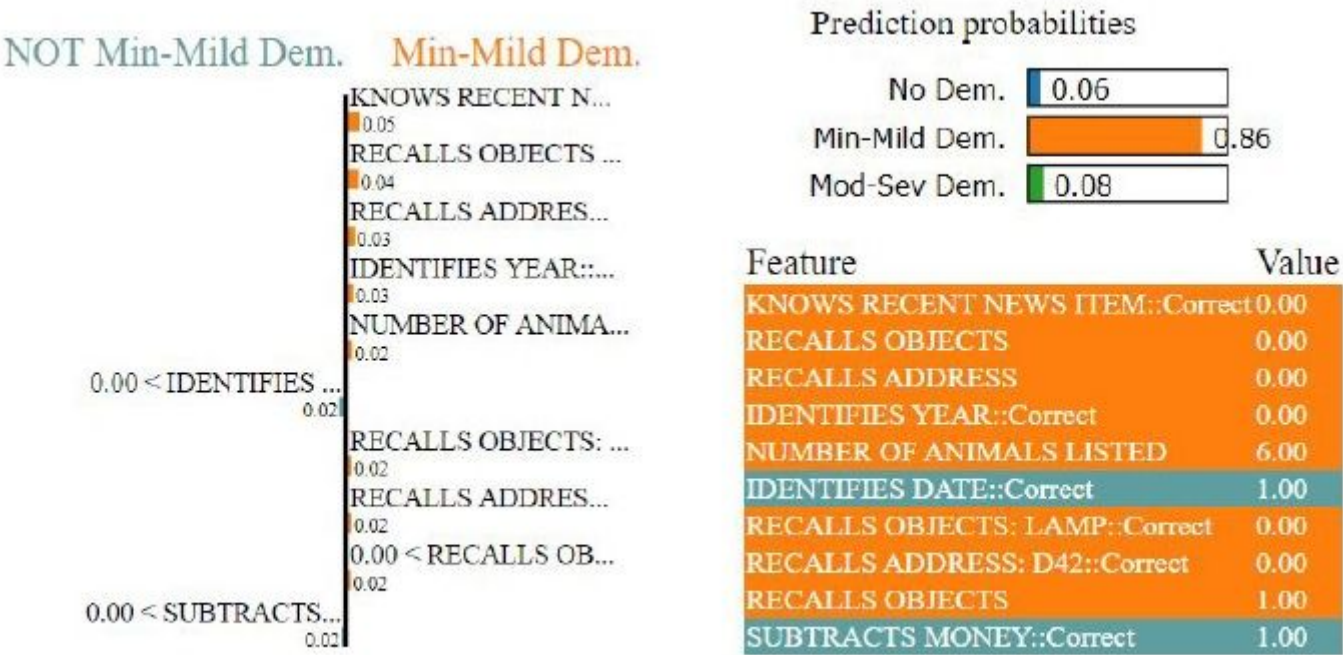
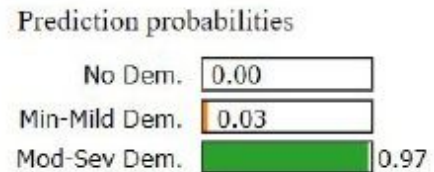
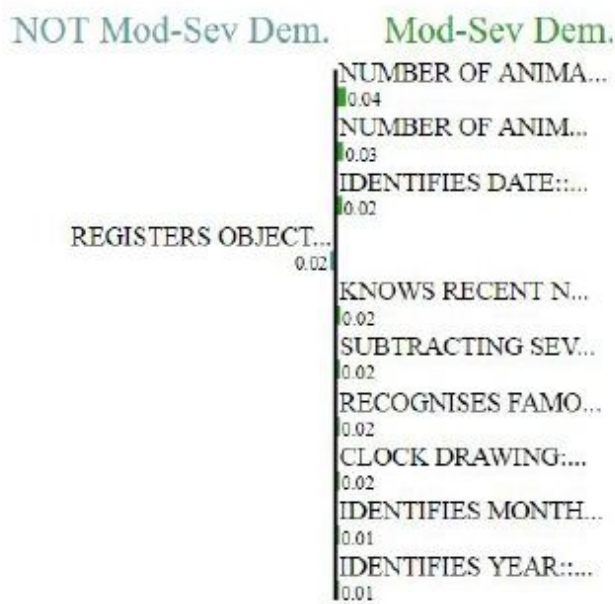


Figure 5

LIME provides the local interpretability for the prediction of an arbitrary test instance with prediction probabilities for Problem (b). The left part of the gure shows the weights of the top 10 important features while making class decision whereas the right part represents predicted class probability and the top 10 important feature with their values. Here the short names of the impacting features in the left part of the gure are in the same order as in the feature & value list in the right part. Min-Mild Dem. and Mod-Sev Dem are the abbreviations for the Minimal or Mild dementia and Moderate or Severe dementia classes, respectively.



| Feature | Value |
|---------------------------------|-------|
| NUMBER OF ANIMALS LISTED | 0.00 |
| NUMBER OF ANIMALS LISTED: SCORE | 0.00 |
| IDENTIFIES DATE::Correct | 0.00 |
| REGISTERS OBJECTS | 0.00 |
| KNOWS RECENT NEWS ITEM::Correct | 0.00 |
| SUBTRACTING SEVENS | 0.00 |
| RECOGNISES FAMOUS PEOPLE | 0.00 |
| CLOCK DRAWING: NUMBERS::Correct | 0.00 |
| IDENTIFIES MONTH::Correct | 0.00 |
| IDENTIFIES YEAR::Correct | 0.00 |

Figure 6

LIME provides the local interpretability for the prediction of an arbitrary test instance with prediction probabilities for Problem (b). The left part of the figure shows the weights of the top 10 important features while making class decision whereas the right part represents predicted class probability and the top 10 important feature with their values. Here the short names of the impacting features in the left part of the figure are in the same order as in the feature & value list in the right part. Min-Mild Dem. and Mod-Sev Dem are the abbreviations for the Minimal or Mild dementia and Moderate or Severe dementia classes, respectively.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [model1DTupd.pdf](#)
- [model4DTupd.pdf](#)