A data mining approach to predict conversion from mild cognitive impairment to Alzheimer's Disease

Luís Jorge Matias de Lemos
luis.lemos@ist.utl.pt

Instituto Superior Técnico, Lisboa, Portugal

October 2012

Abstract

Alzheimer's disease (AD) is a well known neurodegenerative disease causing cognitive impairment. Besides being one of the best studied diseases of the central nervous system, it remains incurable. Mild Cognitive Impairment (MCI) is currently considered to be an early stage of a neurodegenerative disease. Patients diagnosed with MCI are assumed to have higher risk to evolve to AD. In this context, the correct diagnosis of MCI and an effective assessment of its predictive value for the conversion to AD are crucial.

In this thesis neuropsychological data is used to distinguish patients with MCI from those already suffering from AD and to predict the evolution of MCI patients to AD. We analyse a dataset with patients labelled by the clinicians as MCI or AD. As most real clinical data, this dataset is strongly imbalanced and has a high percentage of missing values.

We use state of the art supervised learning techniques and perform a throughout study on the effect of class imbalance and missing values in their performance. Since the number of attributes is large, feature selection is studied and used to effectively decrease the dimensionality of the problem. A data mining methodology has been created to automatize the oversampling and parameters search. The created tool automatically creates the models and parametrize them, having in account the balance state of the data.

A decision support system was created making available to medical doctors the diagnosis and prognosis models developed in this model.

Keywords: Alzheimer's Disease, Data Mining, Temporal Windows, diagnosis, prognosis, prediction.

1. Introduction

Declines in cognitive and motors functions, together with other evidences of neurological degeneration, become increasingly likely as healthy people age. The fact is that everyone will experience altered brain functions, although some at an earlier age or at a faster rate than others. As such, distinguishing the motor and cognitive declines of normal ageing from those due to pathological processes and understanding the individualized disease diagnostic and prognostic patterns are ongoing research challenges [21]. In this context, Alzheimer's disease (AD), a well known neurodegenerative disease causing cognitive impairment, is amongst the best studied diseases of the central nervous system due to its devastating effect on patients and their family, and to the socio-economic impact in modern societies. Nevertheless, it remains uncurable.

Every year millions of new Alzheimer's disease(AD) cases are diagnosed. The result is dementia on elderly individuals; internment and expensive medical care are a common outcome. An early diagnostic and prognosis can improve the patient quality of life, minimizing the need for internment and expensive medical care, reducing the patient and the family's suffering and minimizing the social-economic effects on the society. In this context, finding out if and when a patient will progress from Mild Cognitive Impairment (MCI) to AD is of a major important to the timely administration of pharmaceutics and therapeutic interventions. Furthermore it can allow medical doctors to adjust periodicity of medical consults.

Mild Cognitive Impairment is currently considered to be an early stage of a neurodegenerative disease, particularly AD. Patients diagnosed with MCI are regarded with special attention since they are assumed to have higher risk to evolve to dementia, usually AD [25]. Under these assumptions, the correct diagnosis of MCI conditions and an effective assessment of its predictive value for the conversion to AD are thus of major importance. However, the definition of MCI and its diagnosis criteria are not yet consensual; the pathologic and molecular sub-

strate of people diagnosed with MCI is not well established [20]. Moreover, people considered to be suffering from preMCI, that is people having cognitive complains but not fullfiling the criteria for MCI, have recently been shown to have high risk of progression to MCI and AD [15]. This makes the diagnosis of MCI a dificult task in itself and consequently transforms the prediction of MCI to AD conversions into an even more complicated task.

Neuropsychological tests have been used by medical doctors mainly because they are cheaper and faster than PET Scans and biomarkers search. Furthermore, technology such as PET Scans and biomarkers are not globally available. The neuropsychological tests involve simple tasks such as those concerning orientation, memory, attention and language to evaluate the mental state of the patient.

This work aims to use this data to predict the conversion of MCI to AD. The use of data mining algorithms will allow to the extraction of knowledge or rules from the data in what regards the prediction of MCI to AD.

To study the relation between MCI and AD, researchers typically focus on three related but distinct problems [17] [3][12][6][13][24]: (1) distinguishing MCI from AD, (2) predicting the conversion from MCI to AD, and (3) predicting the time to conversion from MCI to AD. In this work, we tackle the problem of distinguishing patients with MCI, from those already suffering from AD, using neuropsychological data. This type of data has also been used by other authors [17, 3, 6, 12]. Giving the increasing difficulty of these three problems and the non-consensual classification of patients as MCI, distinguishing MCI from AD is an important problem in itself but it gains increasing relevance as a support for the feasibility of effectively tackling the conversion and time to progression problems.

In this context, this thesis tackles two related problems: (i) distinguishing the MCI patients form the AD patients; (ii) predicting if a MCI patient will evolve to AD. To achieve this goals we use state of the art machine learning techniques, such as SVMs, artificial neural networks, Nave Bayes, C4.5 Decision trees and k-nearest neighbour. we also talked the missing values and feature selection. The missing values were analysed to find a way to minimize the missing data effects on the learning process and the feature selection to reduce the data complexity.

2. Background

2.1. Alzheimer's Disease

By the latest estimates, 25 million people currently suffer from dementia and as a consequence of population ageing, the number of persons affected by this condition is expected to climb, doubling every 20 years. Complains of cognitive matter are very common in aged individuals and can be the first sign of on-going neurologically disorders such as AD [18], which is the most common irreparable, progressive cause of dementia. AD can be described by a gradual loss of memory and cognitive skills. Every year over 5 million of new cases are reported and the incidence increases with the age of the individual. These numbers are likely to raise in consequence of the excepted lifetime increase [1]. The relation between age and AD incidence is evident, making age the most likely influential risk factor in the diagnosis of AD. It is possible to identify, from the people that have cognitive complains, those who are in risk to progress to dementia. These are those suffering from MCI. Since the MCI classification mandate the expression of a cognitive decline greater than excepted for the person's age and education level, neuropsychological testing is a fundamental element in the diagnostic [18]. Currently many efforts are being carried out to investigate AD pathology and develop appropriate treatment strategies. These strategies have center their attention on the long-term conservation of cognitive and functional abilities or slowing down the disease development along with reducing behavioral symptoms and maintaining the patient's quality of life. Nowadays, there is no treatment leading to the cure or the complete stop of AD progression. Nonetheless, a current medical objective is the diminution of symptoms that can delay the institutionalization of the patient, therefore reducing the caregiver costs [23].

To diagnose, determine the stage, assess and monitor AD, MCI and other dementia, the mental health of a patients is assessed though neuropsychological assessment using a common set of tests.

These tests aim to of identify and quantify cognitive, functional and behavioral symptoms. A number of test batteries have been developed by medical doctors to assess the mental health of patients. The more important batteries are: Mini-Mental State Examination (MMSE), Alzheimer's Disease Assessment Scale (ADAS) and, in our case the *Bateria de Lisboa para Avaliao de Demencia* (Lisbon Test Battery for Dementia Evaluation) (BLAD). Each battery is composed of multiples tests, where some batteries are composed of multiples other batteries.

MMSE is one the most widely used test batteries to perform a brief evaluation of cognitive status in adults [19] [8] [26]. ADAS was designed to measure the severity of the most important symptoms of AD. The ADAS-cog is the most popular cognitive testing instrument used in clinical trials of nootropics (drugs, functional foods, supplements, etc, that improve mental functions). It consists of 11 tasks measuring the disturbances of memory, language, praxis, attention and other cognitive abilities which

are often referred to as the core symptoms of AD [14].

BLAD [2] [26] is a comprehensive neuropsychological battery evaluating multiple cognitive domains and it has been validated for the Portuguese population. This battery includes tests for the following cognitive domains: attention (Cancellation Task); verbal, motor and graphomotor initiatives (Verbal Semantic Fluency, Motor Initiative and Graphomotor Initiative); verbal comprehension (a modified version of the Token Test); verbal and non-verbal abstraction (Interpretation of Proverbs and the Raven Progressive Matrices); visual-constructional abilities (Cube Copy) and executive functions (Clock Draw); calculation (Basic Written Calculation); immediate memory (Digit Span forward); working memory (Digit Span backward); learning and verbal memory (Verbal Paired-associate Learning, Logical Memory and Word Recall).

The data used in this work was obtained using BLAD. Each evaluation of a patient corresponds to an instance identified by the date of evaluation and the ID of the patient. The majority of patients have multiple evaluations. Each of them is associated with an evaluation date and a patient classification given by a medical doctor.

2.2. Classification

Classification can be described as a two-stage process [10]. In the first stage, a classifier describing a set of data classes is built. This stage is designated by learning step or training phase. In this stage the classification algorithm is "learning from" a training set composed by instances of data, which are made up of a n-dimensional attribute vector, X $=(x_1,...,x_n)$, and a class label. In this case X a set of attributes extracted from the neuropsychological data and the class label is the patient mental health given by a medical evaluation and categorized as MCI and AD. The attributes in vector X is can be numerical or categorical. The instances used to train the classification algorithm compose the training set. This type of process is known as supervised learning, since the class label attribute is provided to each X in distinction to the unsupervised learning algorithms that do not known the class label attribute or the number of classes to be learned in advance. In the context of classification, the set of n-dimensional attribute vector that represents an evaluation of the patient, and the respective class label attribute can be named as instances. In the second step, the model obtained is used to classify the test set. The test set is a subset of data, independent from the train set, that is used to measure the accuracy of the classification model. It should be noticed that in this we only use supervised methods.

The used classification algorithms are: Naïve Bayes, Support vector machines, k-Nearest-Neighbour, Decision Trees and Neural Networks.

2.3. Feature selection

A central problem in machine learning is to identify the set of features that best represent the data and can be used to construct a classification model for a particular model [9]. Feature selection is the process of selecting a subset of features for building robust learning models. Feature selection is particular important when the dataset has a large number of features. By removing redundant and irrelevant features from the dataset, the feature selection technique aims to improve the overall performance of the learning models. This improvement results from the reduction on the curse of dimensionality [9], increasing the generalization capability of the learning models (and decreasing overfitting), increasing the speed of the learning process and improving the model interpretability since less features are used.

2.4. Overcoming class imbalance

Imbalanced learning targets a significant amount of problems of interest by academics, industry and governmental agencies [11]. The main problem of learning from imbalanced data sets is the fact that this imbalance compromises the performance of most standard learning algorithms. The majority of learning algorithms assumes a balanced class distribution or equal misclassification costs. The typical result is a favouring of the predominant class which gives poor class predictions.

The class imbalance in the dataset can damage the quality of the classification. If the minority class is hard to discriminate, the classifier can simply classify every instance as the majority class. In this case if the minority class only represented 1% of the data, the obtained accuracy would be 99%, although the classifier is useless in the discriminative task.

To overcome this problem, a set of methods have been developed. These techniques have the objective of minimizing the effect of using imbalanced data. Examples of this techniques are: sampling by removing, creating or duplicating instances; and cost sensitive methods, where different costs to the misclassification of the different classes are assigned.

2.5. Model Validation

In a specific classification problem one can ask the question on which method is the best. Using the train set to make the comparison would lead to misleading overoptimistic results due to an overspecialization of the learning algorithm to the data. To avoid this problem, the dataset is divided into train and test. The accuracy of a model is then evaluated

by comparing the results on the test dataset. The evaluation is done using metrics such as:

- **Accuracy**
  Accuracy is measured by the ratio between the numbers of correctly classified instances and the total number of instances.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

- **Sensitivity (True Positive Rate or Recall) and Specificity (True Negative Rate)**
  Sensitivity is the proportion of instances which were classified as class x, among all examples which truly have class x.

$$Recall = Sensitivity = \frac{TP}{TP + FN} \quad (2)$$

  Specificity is the True Negative Rate that is the proportion of instances which were classified as class x, but belong to a different class, between all instances which are not of class x.

$$Specificity = \frac{TN}{TN + FP} \quad (3)$$

- **Precision**
  Precision is the probability of obtaining a relevant result among the subset declared as positives (TP+ FP):

$$Precision = \frac{TP}{TP + FP} \quad (4)$$

- **F-Measure**
  The F-Measure is the harmonic mean of precision (P) and Recall (R).

$$F = \frac{2PR}{P + R} \quad (5)$$

- **Receiver Operating Characteristic(ROC) curves and $|TPR - FPR|$**
  ROC curves are a useful tool to compare two classifier models. The ROC curve show us the trade-off between the true positive rate (TPR) rate or sensitivity and the false positive rate (FPR) rate for a given classifier model.

  The Area Under the Curve (AUC) is also a good metric for imbalanced learning [11], but is harder to compute than the $|TPR-FPR|$ in probabilistic models. The AUC will not have in account the discriminative power of the classifier directly, since a classifier that has $AUC = 0$ has more discriminative power than one with $AUC = 0.5$.

  In a ROC space, the line TPR = FPR represents a random classier or a meaningless classifier. Then a classifier model, m, can be represented in this space as $m =(1 - Specificity, Sensitivity)$. If m is in the point $(0, 1)$ the classifier is perfect, classifying everything correctly. Or the other hand if $m = (1, 0)$, the classifier is always wrong. In this case by inverting the label we can get a perfect classifier. In general we are interested in a classification model that returns a point in the ROC space as far as possible from the random line ($TPR = FPR$). This metric will give as the discriminative score of the classifier. Calculating the normalized Euclidean distance from a point to the random line we get the expression: $|TPR - FPR|$, or $|Sensitivity - (1 - Specificity)|$. This value has a maximum of 1 that represents a perfect discrimination of the class and a minimum of 0 that represents a random classifier or a classifier without discriminative power. This metric gives an equal value of classification to both classes. One advantage of this metric is that it is insensitive to the data imbalacement, in opposition of the accuracy that overvalues the majority class. The advantage of this metric versus the analysis of the sensitivity and specificity, is that this discriminative score combines both in a single value, being more simple to analyse. Because of this, the metric is refereed in the literature as informedness [22].

2.6. Related Work

Table 1 synthesizes the related work together with the information about the test performed.

3. Differentiating MCI from AD (Diagnosis)

Correctly differencing MCI from AD is a key step in the process of predict the conversion from MCI to AD. In this context, this chapter addresses the diagnosis problem. Two feature selection techniques are used, techniques for overcoming missing values are studied, and the methodology described in this chapter is applied.

3.1. Formulation and Methodology

In this chapter the problem formulation and methodology is defined. With this we mean the formulation of the problems that we aim to solve, in this work, and by methodology the methods that we use to solve the formulated problems.

3.2. Data Description

The Cognitive Complaints Cohort [18, 16] is a prospective study conducted at the Institute of Molecular Medicine (IMM), Lisbon, to investigate the cognitive stability or evolution to dementia of subjects with cognitive complaints based on a comprehensive neuropsychological evaluation and other

Table 1: Synthesis of the related work.

	Problem	Data	Methods
Clifford Jr. et al	Time to progression	MRI , PIB PET and Neuropsychological tests	Statistical methods
Ewers et al	Time to progression and Prognostic	Neuropsychological tests and Biomarkers	Statistical methods, Logical regression analysis
Chapman et al	Prognostic	Neuropsychological tests	PCA, Statistical methods
Hinrich et al	Prognostic	Neuropsychological tests, FDG PET and MR images	MKL (Multi-Kernel Learning)
Maroco et al	Prognostic	Neuropsychological tests	Fisher's Linear Discriminant, Quadratic Discriminant Analysis, Linear Regression, MLP, SVM, Radial basis Functions, CART, CHAID, QUEST and Random Forests

biomarkers. The criteria for inclusion, exclusion, and diagnosis of the participants as MCI or AD during follow-up are described in detail in Dina et al. [25]. In this work, we used a revised and augmented version of this dataset and considered only neuropsychological data.

The original dataset has 1641 instances consisting of individual evaluations of 950 distinct patients during their follow-up at IMM. In each evaluation, each patient was classified by the medical doctors as Normal, preMCI, MCI and AD using clinical criteria. Only instances labeled as MCI and AD were considered, and from these, only those concerning patients with at least two evaluations were analyzed, since instances corresponding to patients without follow-up are more likely to be misclassified. All instances with a percentage of missing values of at least 90% were removed since these instances have little information. This yielded a dataset with 677 instances labeled as either MCI or AD, where each instance corresponds to a different evaluation of a set of 337 distinct patients. We note that, since we aim to distinguish between MCI and AD patients (or in the prognosis the evolution of MCI to AD), we can consider each evaluation of a patient as a different instance, meaning that we can learn from patients at different disease stages that are always diagnosed as MCI during follow-up and patients that convert to AD during follow-up.

After excluding non-informative features, such as "Patient ID", and features related with patient clinical history, such as "Follow-up Time", the analysed dataset is composed of 677 instances described by 157 features/attributes, which can be numerical, categorical/nominal or ordinal. This dataset is highly imbalanced, in the original classes, since approximately 86% of the instances are labelled as MCI. Moreover, missing values, which are around 50% in the overall data, are still an issue as we discuss below. Note that, 60% of all features have more than 40% of missing values.

3.3. Methodology

A single data mining methodology can be used used for all problems. In this methodology we include six classifiers: Nave Bayes, gaussian SVM, polynomial SVM, k-nearest neighbour, C4.5 Decision trees and Artificial neural networks using backpropagation. All classifiers used are implemented in WEKA .

The imbalance of the data is tackled with a synthetic oversampling technique (SMOTE) [5]. The classifier parameters and the percentage of oversampling is determined using 10-fold cross validation on a grid search approach. The percentage of oversampling and parameters are combined since SMOTE changes the dataset, and thus so the parameters founded with different SMOTE percentages may not be the same. The SMOTE algorithm is implemented in WEKA.

Each classifier has some type of parameter or a set of them. The best found SMOTE percentage can change for each problem, subset of features, for the classifier parameter or classifier itself. The best classifier parameters and SMOTE percentage must be determined having all this in account. Thus we cross all tested SMOTE percentages with all tested parameter sets in a grid search, using the model

represent in Figure2. This search is done for each feature selection method applied with a systematic process to evaluate the parameters and test SMOTE percentages on a defined space. Therefore, an automated tool was created to deal with this necessity, avoiding mistakes, and in order to optimize the grid search process. The metric used to compare the models for each parameter set and SMOTE is the $|TPR - FPR|$. This metric is obtained by calculating the normalized Euclidean distance from the point (FPR,TPR) from the random line (FPR=TPR).

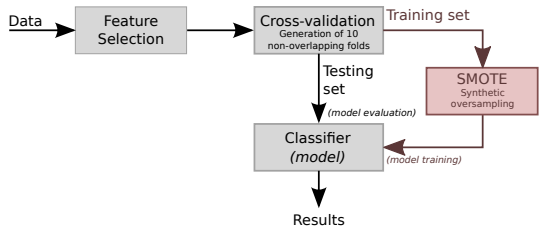

Figure 1: Data Flow used in the parameter grid search for finding the classifiers parameters. The SMOTE percentage is tested with 11 different values for each parameter combination.

For testing the obtained classification model a different data set is used, which was obtained by splitting the original dataset in 75% of patients for training and 25% of patients for testing. For this we apply stratification based on: (i) number of evaluations; (ii) age; (iii) sex; (iv) schooling years and (v) class. Spiting of patients was therefore made such as the distribution of the above variables is kept constant in the training and testing datasets.

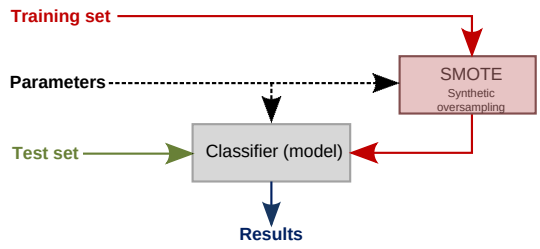

Figure 2: Data Flow used simulate the real world results.

This allows the test set to be used in all problems; diagnosis, prognosis and any future problem tacked in the NEUROCLINOMICS project. It should be noticed that the test set will not be used to find the best parameter set. It is only used to evaluate the final models created using the train set. These models only use the train set to find the best features and the best parameters for that specific data set. This will allow us to analyse the behaviour of trained models in a "real world" simulation, since the model has never been in contact with any instance of the test patients.

3.4. Results

For the diagnosis problem six triples, for each feature set, have been selected. These triples are the classifier, parameters set and the SMOTE percentage. These results are obtained using a grid search using only the train set.

Analysing the results (see Figure 3) we can see that for each classifier method, the best features can change, which means that a single feature set is not always the best feature set for all cases, but varies from model to model.

Using the train set, the higher median value is $|TPR - FPR| \approx 0.6$ using SVM RBF with correlated-based feature selection (the results with others feature sets are almost similar). However when using the test set the model now has a $|TPR - FPR| \approx 0.5$. The maximum value, obtained in all algorithms, is $|TPR - FPR| \approx 0.6$. This maximum appears in 2 models that use correlation feature set, the Nave Bayes and Neural Network. In this case, we can compare the result, using the train set and using the test set (that simulate the real world with a fully independent sample) to analyse the consequence of only using the train set to pick the best model. The SVM RBF with correlation-based feature selection appears to have the best results. However, generalization is not so good since in contact with unknown instances its results drops.

4. Predicting conversion from MCI to AD (Prognosis)

The prognosis prediction of a patient is of great importance to the medical doctors. It allows for adequate medical care to the patient and support for the family. The prognosis prediction of Alzheimer's Disease (or other cognitive impairment) has also a role in the decisions of the patient about their future. For example if the conversion to AD occurs in a year, and this patient has a high responsibility job, e.g, as a company manager or a pilot, the patient can adjust his life to minimize the impact of his disease on the society.

4.1. Prognosis prediction approach

For prognosis prediction we use two different approaches. The first one, which is, normally used in similar problems [4] [7] [13] [12], consists in finding if a patient will ever convert to AD. This approach will be refereed in this work as First and Last Evaluation, since it looks for the first and last entry of the patient in the database to determine if a patient will ever evolve from MCI to AD. In this approach, each patient has only one single entry in the post-processed dataset.

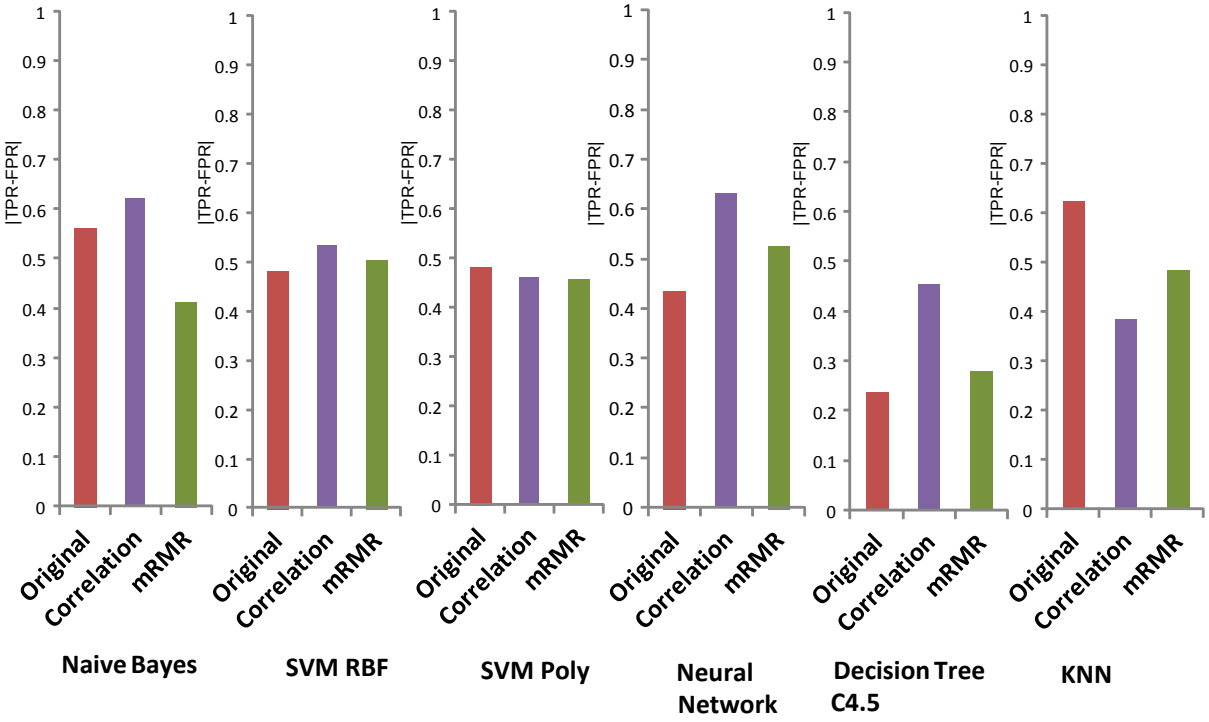

Figure 3: Results for the diagnosis using an independent test set, the scale is $|TPR - FPR|$ where higher is better.

The second approach looks at a given temporal window and tries to predict if a patient converts from MCI (at the beginning of the temporal window) to AD (at the end). For this, and according to Figure 4, a new set of labels has been created: evolution (Evol) and no evolution (noEvol) instances. The noEvol class is considered the positive class. To chose the temporal windows two factors were considered: (i) the instances distribution between classes (Evol/noEvol) and (ii) the medical relevance which was obtained by consulting the medical partners of the NEUROCLINOMICS project. For the latest case, a period of around 3 years was recommended.

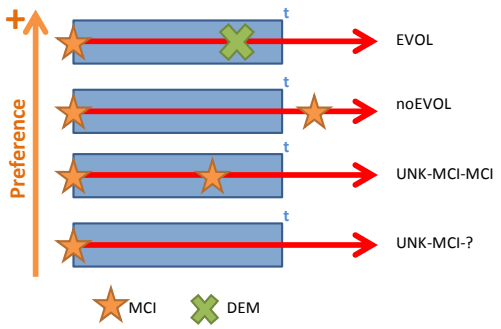

Figure 4: Graphical representation of the new class labels created for the temporal windows prognosis problem.

4.2. 3 Years Temporal Window

The best results, in all approaches is obtained using the 3 years temporal window. The classification methodology is equal of the diagnosis model. The grid search was done to each model to all temporal windows and three features sets are tested.

In this case the dataset is not unbalanced, but oversampling helped to delimit the decision boundaries. In the three years time window, test results (see Figure 5) show that all classifiers have a reasonable behaviour. The best result obtained in this time window is $|TPR - FPR| \approx 0.65$, a result that is closer to the perfect classifier that from the random one. We obtained this result using two models: the Nave Bayes and SVM RBF both using the mRMR dataset. The neural network, have also good results in all datasets, having a $|TPR - FPR| \approx 0.5$ in all cases although best results are achieved using the original set (all features). The kNN also have a good overall behaviour, having with the correlated dataset a result of $|TPR - FPR| \approx 0.55$. The worst results are achieved using the decision trees. This can be explained since high confidence was obtained using the train set which caused the model to overfit.

It should be noticed that the best results were obtained using no oversampling and using the mRMR feature set. As suspected, the oversampling has little influence when a balanced dataset is used. Nevertheless, in same cases, it is beneficial as it is the

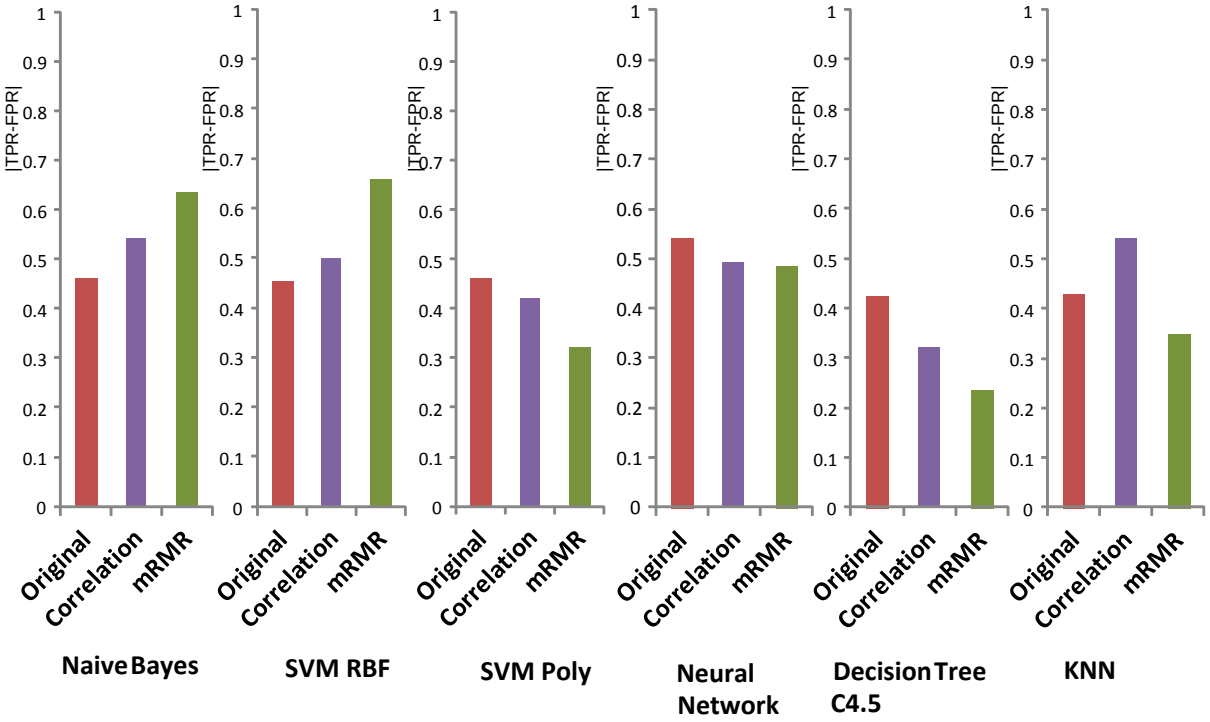

Figure 5: Test results of Prognosis using three years temporal window

case of the Nave Bayes. The best model is the radial SVM (SVM RBF) using the mRMR feature selection.

Using different metrics, we can observe that Nave Bayes achieves the highest accuracy, sensitivity ans ROC area. However, its specificity is significantly lower to compared that of radial SVM.

5. Decision Support System

The models created in chapter 3 and 4 showed an overall good performance in the tasks of discriminating MCI patients (diagnosis) and predicting the progression from MCI to AD (prognosis). But these models are unusable by the medical doctors. To bridge this usability impediment a solution was designed and implemented, to facilitate the use of the system by third parties. By integrating the models in an information system the medical doctors can now evaluate the system and models in real work situation. Since this work was done in the NEUROCLINOMICS project context, the integration of work done is of huge importance. Having this in mind, a DSS (Decision Support System) webservices architecture was applied. Enabling the integration with any other tool developed in the project, in particular the support information system for the AD, that is under development. The use of web services will allow models update without altering any other part of the system. This approach is highly modular. Beside the to inquiry the models with an instance to obtain a diagnosis or prognosis (a percentage referring to the model confidence in the predicted result), an auto-upgradable system was also created. That can be periodically updated with new information, more instances. This database will be automatically analysed, the models will be parametrized and the best oversampling percentage will be applied if needed, using the methodology created in this work.

6. Conclusions

In this work we study the influence of the class imbalance, high dimensionality and missing values. To contract the influence of all these factors a unique approach was designed to create and evaluate the data. This unique model, uses a grid search that combines oversampling and multidimensional parameters search.

To evaluate the results without a bias in case of unbalance dataset. The $|TPR - FPR|$ metric is used in all models of this work. This metric is a trade-off between the sensibility and specificity.

In this work we tackle the diagnosis problem and create models that discriminate MCI and AD cases. We analysed the behaviour of a set of supervised data mining algorithms and we concluded that the Nave Bayes and Neural Networks have a better performance when in contact with a unknown test set. Those results are obtained using Original set of features (All Features) and a Correlated set. We can also conclude that the use of 10-fold-cross-validation provides an estimate of the goodness of

the result that is not analogue to the one obtained test set. This can be caused by an overfitting to all dataset, that leads to a low generalization of the models, similar results are obtained in all problems analysed. One of the best diagnostic model was obtained using Nave Bayes algorithm this model have a accuracy of 91%, a sensitivity of 93%, a specificity of 69%, a ROC Area of 0.85 and a $|TPR - FPR|$ of 0.62.

In the prognosis problem we present a new approach to predicting the conversion form MCI to AD. The standard method is to use the first and last evaluation of the patient. This approach in our opinion will not use important information, such as profiles that a patient can have in their evaluation history. In that 10 years a patient can pass for profiles that some other patient can also have. By using a temporal window approach we obtain better discriminative results. We concluded that the best models use the Nave Bayes and SVMs algorithms, and that the mRMR feature set showed us very good results, in generally better than those using the original set or the correlated set. The temporal window with higher discriminative power is the 3 years window. Using this window the best model was obtained using radial SVM algorithm this model have a accuracy of 82%, a sensitivity of 79%, a specificity of 86%, a ROC Area of 0.83 and a $|TPR - FPR|$ of 0.64.

Finally, we created a Decision support system, that uses the diagnosis and prognosis models. This system can help the medical doctors to evaluate in a short space of time the patients. This system was implemented using web services to integrate this work in the NEUROCLINOMICS project. The use of web services allows this work to be integrated in other works in the scope of the project since it uses a simple communication protocol.

As future work, we can use voting with the models to increase the discriminative power, to each problem.

Since exists missing data patterns, the use of bi-clusters algorithms to find bi-clusters, and then perform classification on those clusters. This approach can increase the discriminative power of the models. Also with those bi-cluster classifiers a voting system can be implemented or a system of reinforcement learning to adjust the importance of each bi-clusters.

Finally, approach the time to progress problem, this problem aims to find the time to the progression of a MCI patient to AD.

Acknowledgements

This work was partially supported by FCT - Fundaao para a Cincia e a Tecnologia under projects PTDC/EIA-EIA/111239/2009 (NEUROCLINOMICS - Understanding NEUROdegenerative diseases through CLINical and OMICS data integration).

References

[1] L. M. Bekris, C.-E. Yu, T. D. Bird, and D. W. Tsuang. Review article: Genetics of alzheimer disease. *Journal of Geriatric Psychiatry and Neurology*, 2010.

[2] G. C. *Doena de Alzheimer, problemas do diagnstico clnico.* PhD thesis, Faculdade de Medicina de Lisboa, 1984.

[3] R. Chapman, M. Mapstone, J. McCrary, M. Gardner, A. Porsteinsson, T. Sandoval, M. Guillily, E. DeGrush, and L. Reilly. Predicting conversion from mild cognitive impairment to alzheimer's disease using neuropsychological tests and multivariate methods. *Journal of Clinical and Experimental Neuropsychology*, 33(2):187–199, 2011.

[4] R. M. Chapman, M. Mapstone, J. W. McCrary, M. N. Gardner, A. Porsteinsson, T. C. Sandoval, M. D. Guillily, E. DeGrush, and L. A. Reilly. Predicting conversion from mild cognitive impairment to alzheimer's disease using neuropsychological tests and multivariate methods. *Journal of Clinical and Experimental Neuropsychology*, 2011.

[5] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer. SMOTE: Synthetic Minority Over-sampling TEchnique. *Journal of Artificial Intelligence Research 16*, 2002.

[6] M. Ewers, C. Walsh, J. Trojanowski, L. Shaw, R. Petersen, C. Jack Jr, H. Feldman, A. Bokde, G. Alexander, P. Scheltens, et al. Prediction of conversion from mild cognitive impairment to alzheimer's disease dementia based upon biomarkers and neuropsychological test performance. *Neurobiology of Aging*, 2010.

[7] M. Ewers, C. Walsh, J. Q. Trojanowski, L. M. Shaw, R. C. Petersen, C. R. J. Jr., H. H. Feldman, A. L. Bokde, G. E. Alexander, P. Scheltens, B. Vellas, B. Duboisl, M. Weiner, and H. Hampel. Prediction of conversion from mild cognitive impairment to alzheimer's disease dementia based upon biomarkers and neuropsychological test performance. *Elsevier*, 2010.

[8] M. Guerreiro, B. Silva, A. P., L. M. A., A. Castro-Caldas, and C. Garcia. Adaptao a populao portuguesa da traduo do mini mental state examination (mmse). *Revista Portuguesa de Neurologia*, 1994.

[9] M. Hall. *Correlation-based feature selection for machine learning*. PhD thesis, The University of Waikato, 1999.

[10] J. Han and M. Kamber. *Data Mining:Concepts and Techniques*. Diane Cerra, 2 edition, 2006.

[11] H. He and E. Garcia. Learning from imbalanced data. *Knowledge and Data Engineering, IEEE Transactions on*, 21(9):1263 –1284, sept. 2009.

[12] C. Hinrichs, V. Singh, G. Xu, and S. Johnson. Predictive markers for ad in a multi-modality framework: An analysis of mci progression in the adni population. *NeuroImage*, 55(2):574–589, 2011.

[13] C. Jack Jr, H. Wiste, P. Vemuri, S. Weigand, M. Senjem, G. Zeng, M. Bernstein, J. Gunter, V. Pankratz, P. Aisen, et al. Brain beta-amyloid measures and magnetic resonance imaging atrophy both predict time-to-progression from mild cognitive impairment to alzheimer's disease. *Brain*, 133(11):3336–3348, 2010.

[14] E. Kolibas, V. Korinkova, V. Novotny, K. Vajdickova, and D. Hunakova. Adas-cog (alzheimer's disease assessment scale-cognitive subscale)–validation of the slovak version. *PubMed*, 2000.

[15] D. Loewenstein, M. Greig, J. Schinka, W. Barker, Q. Shen, E. Potter, A. Raj, L. Brooks, D. Varon, M. Schoenberg, et al. An investigation of premci: Subtypes and longitudinal outcomes. *Alzheimer's and Dementia*, 8(3):172–179, 2012.

[16] J. Maroco, D. Silva, M. Guerreiro, A. de Mendona, and I. Santana. Prediction of dementia patients: A comparative approach using parametric vs. non parametric classifiers. *XIX Congresso Anual da Sociedade Portuguesa de Estatstica*, 2011.

[17] J. Maroco, D. Silva, A. Rodrigues, M. Guerreiro, I. Santana, and A. de Mendonça. Data mining methods in the prediction of dementia: A real-data comparison of the accuracy, sensitivity and specificity of linear discriminant analysis, logistic regression, neural networks, support vector machines, classification trees and random forests. *BMC research notes*, 4(1):299, 2011.

[18] J. Maroco, D. Silva, A. Rodrigues, M. Guerreiro, I. Santana, and A. de Mendonça. Data mining methods in the prediction of dementia: A real-data comparison of the accuracy, sensitivity and specificity of linear discriminant analysis, logistic regression, neural networks, support vector machines, classification trees and random forests. *BMC research notes*, 4(1):299, 2011.

[19] F. MF, F. SE, and M. PR. "mini-mental state". a practical method for grading the cognitive state of patients for the clinician. *PubMed*, 1975.

[20] E. Mufson, L. Binder, S. Counts, S. DeKosky, L. deToledo Morrell, S. Ginsberg, M. Ikonomovic, S. Perez, and S. Scheff. Mild cognitive impairment: pathology and mechanisms. *Acta neuropathologica*, 123(1):13–30, 2012.

[21] F. Noorbakhsh, C. M. Overall, and C. Power. Deciphering complex mechanisms in neurodegenerative diseases: the advent of systems biology. *Trends in Neurosciences*, 32(2):88–100, January 2009.

[22] D. M. W. Powers. Evaluation : From precision , recall and f-measure to roc , informedness , markedness and correlation. *Journal of Machine Learning Technologies*, 2(1):37–63, 2011.

[23] P. Robert, S. Ferris, S. Gauthier, R. Ihl, B. Winblad, and F. Tennigkeit. Review of alzheimer's disease scales: is there a need for a new multi-domain scale for therapy evaluation in medical practice? *Alzheimer's Research & Therapy*, 2010.

[24] M. Samtani, M. Farnum, V. Lobanov, E. Yang, N. Raghavan, A. DiBernardo, V. Narayan, et al. An improved model for disease progression in patients from the alzheimer's disease neuroimaging initiative. *The Journal of Clinical Pharmacology*, 2011.

[25] D. Silva, M. Guerreiro, J. Maroco, I. Santana, A. Rodrigues, J. Bravo Marques, and A. de Mendonça. Comparison of four verbal memory tests for the diagnosis and predictive value of mild cognitive impairment. *Dementia and Geriatric Cognitive Disorders Extra*, 2(1):120–131, 2012.

[26] D. Silva, I. Santana, F. S. do Couto, M. G. J Maroco, and A. de Mendona. Cognitive deficits in middle-aged and older adults with bipolar disorder and cognitive complaints: Comparison with mild cognitive impairment. *INTERNATIONAL JOURNAL OF GERIATRIC PSYCHIATRY*, 2008.