University of Reading

A Genetic Algorithm for the selection of structural MRI features for classification of Mild Cognitive Impairment and Alzheimer's Disease

Conference or Workshop Item

Accepted Version

Spedding, A. L., Di Fatta, G. and Cannataro, M. (2015) A Genetic Algorithm for the selection of structural MRI features for classification of Mild Cognitive Impairment and Alzheimer's Disease. In: The IEEE International Conference on Bioinformatics and Biomedicine (BIBM), 9-12 Nov 2015, Washington D.C., pp. 1566-1571. doi: https://doi.org/10.1109/BIBM.2015.7359909 Available at http://centaur.reading.ac.uk/51025/

Published version at: http://dx.doi.org/10.1109/BIBM.2015.7359909

To link to this article DOI: http://dx.doi.org/10.1109/BIBM.2015.7359909

Proceedings of the IEEE International Conference on Bioinformatics and Biomedicine (BIBM), Workshop on Machine Learning in Decision Making for Biomedical Applications, Nov. 9-12, 2015, Washington D.C. (in press)

A Genetic Algorithm for the Selection of Structural MRI Features for Classification of Mild Cognitive Impairment and Alzheimer's Disease

Alexander Luke Spedding*, Giuseppe Di Fatta*, Mario Cannataro† and the Alzheimer's Disease Neuroimaging Initiative
*School of Systems Engineering
University of Reading, UK
Email: a.l.spedding@pgr.reading.ac.uk, g.difatta@reading.ac.uk
†Bioinformatics Laboratory
Department of Medical and Surgical Sciences
Magna Graecia University of Catanzaro, Italy
Email: cannataro@unicz.it

***Abstract*—This work investigates the problem of feature selection in neuroimaging features from structural MRI brain images for the classification of subjects as healthy controls, suffering from Mild Cognitive Impairment or Alzheimer's Disease. A Genetic Algorithm wrapper method for feature selection is adopted in conjunction with a Support Vector Machine classifier. In very large feature sets, feature selection is found to be redundant as the accuracy is often worsened when compared to an Support Vector Machine with no feature selection. However, when just the hippocampal subfields are used, feature selection shows a significant improvement of the classification accuracy. Three-class Support Vector Machines and two-class Support Vector Machines combined with weighted voting are also compared with the former and found more useful. The highest accuracy achieved at classifying the test data was 65.5% using a genetic algorithm for feature selection with a three-class Support Vector Machine classifier.**

I. Introduction

The number of people diagnosed with Alzheimer's Disease (AD) is expected to rise over the next few decades, and by 2050 over 1% of the world's population are predicted to be suffering from AD [1]. AD is a neurodegenerative disease causing memory loss, disorientation and behavioural issues; all symptoms of AD get progressively worse as the disease advances. Mild Cognitive Impairment (MCI) is a brain function syndrome with similar symptoms to AD although to a lesser extent, and being diagnosed with MCI has been recognised as a risk factor for developing AD in the future [2].

A structural magnetic resonance imaging (MRI) scan - often the structural part is assumed when referring to MRIs - is a non-invasive imaging technique which generates a 3D image of the physical structure of a subject's brain. The 3D image is made up of voxels, each voxel has a position in the 3D space and an intensity value which determines whether the voxel is classed as white matter (WM), grey matter (GM) or cerebrospinal fluid (CSF).

The development of an algorithm to classify a subject based on their MRI scan as a healthy control (HC), or suffering from MCI or AD would be a great advantage as it would allow for diagnosis of AD and MCI in hospitals with limited resources as they would only need the MRI scanner and a computer to run the classification algorithm. However, the downside to this is that it is near impossible to have 100% accuracy and thus it will not be trusted as much as a clinical diagnosis based on neuropsychological tests. Thus it would be best to create a classifier to aid rather than replace human expertise and clinical tests.

SVMs are a state-of-the-art technique to classify high dimensional data as they are generally one of the best classifiers at obtaining a high accuracy. Research in [3] has found that feature selection does not improve the accuracy of the SVM due to the SVM's robustness of handling the features. However, while reducing the features of the input data may not increase the accuracy of the classifier the advantage it would have is producing a better model of the brain with AD. A smaller feature set being used to distinguish HC, MCI and AD subjects would mean that there is a model of the brain with a smaller number of features showing the main areas affected by MCI or AD. The advantage of this model would be to aid a medical centre in the diagnosis of patients who potentially suffer from MCI or AD. An MRI scan of the patient's brain can be taken as input to this automated classifier, this would then produce a diagnosis of the subject with a given confidence of what they suffer from (or if they are healthy), this can then be used to refer them onto further specialist treatment. The use for the model with a small number of features would be to produce a report about which areas of their brain are affected by the potential diagnosis of MCI or AD. This method would save time as it is an initial referral to screen patients saving the time of doctors with the specialist knowledge to analyse the MRI scans as they would only need to analyse the patients which have been referred by the automated diagnosis, which

would be a smaller amount than the initial amount of patients whose data are input to the classifier.

The program Freesurfer is used to analyse the MRI scans, it is a free and open source software suite which performs many tasks for processing and analysing MRI scans such as: image registration, subcortical segmentation, cortical thickness estimation and many others. In this paper, Freesurfer is used to extract measurements regarding different regions of interest (ROIs) throughout the brain and these measurements are then used to diagnose the subject as being healthy or suffering from MCI or AD. The curse of dimensionality is a term referring to a set of problems which occur when handling high-dimensional data: as the number of dimensions increase, the data becomes more sparse and therefore patterns are harder to find; it can also lead to overfitting of the training data making the classifier useless when applied to new data. A large number of measurements (356) are produced by Freesurfer's analysis and these must be reduced to achieve a model which can be easily interpreted by a human doctor; so this paper uses feature selection by a Genetic Algorithm (GA) to reduce the dimensionality of the problem this has the added benefit of reducing the sparsity of the data so that patterns will be easier to find.

A. Related Work

[4] used a GA for feature selection of a variety of neuropsychological tests, which were then input to a logistic regressor to predict conversion from HC to MCI or AD, and conversion from MCI to AD. They also showed the GA performed better than Stepwise Variable Selection, a commonly used feature selection method. In [5] a GA was used on multiple datasets for feature selection for a Support Vector Machine (SVM), they investigated a number of ways to evaluate the fitness of the algorithm and used the GA to search for optimal values for the SVM's hyperparameters. They found using a GA for feature selection for an SVM showed promising success.

Previous work has been done in [6] at solving the automated classification of AD, this paper intends to expand on this work in three ways. The first is that any data used will be raw MRI scans so the methods used to process the data (such as the Freesurfer version) can be controlled, as in [6] the training and test data was processed with a different version of Freesurfer which lead to some large differences in the values of the hippocampal subfields. This meant that the hippocampal subfields could not be used as features to diagnose the subjects. There will also be a larger amount of both training and test data used. Finally, a GA will be used for feature selection and while this has the downside of a longer computation time, it has the benefit of being able to search a larger search space of feature sets.

[7] also used extracted Freesurfer features from MRI where features were selected based on a priori knowledge, and features were also combined with each other to reduce the dimensionality. A high classification accuracy of 73% was achieved. [8] tested various methods to classify HC, MCI and AD based on MRI data; some of these methods included training classifiers on data extracted by Freesurfer. In particular they used a Parzen Window on the hippocampal volumes, achieving a sensitivity of 73% and a specificity of 74%. They concluded that feature selection increased the sensitivity and were more accurate at classification of problems where there are only a few ROIs in the brain.

By using a GA for feature selection the aim will be to develop an understanding of which features in the brain are the most useful for predicting MCI and AD without being biased by any previous knowledge of the brain's functionality nor by any restrictions in other feature selection methods as the majority are based on finding a locally optimal solution which could potentially mean that other important but less obvious patterns are not discovered.

II. Method

A. Data Acquisition

435 MRI scans were downloaded from the ADNI database and processed with Freesurfer version 5.3 with the standard cortical reconstruction process and the optional command to segment the hippocampal subfields. The hippocampal subfields were included since AD has been found to be prevalent in the hippocampal region [9] [10]. The scans used were the baseline scans for each subject - this is the initial scan taken and initial diagnosis given to the subject. Other criteria used to refine the MRI data was that the slice thickness of scan was 1.5mm and weighted in T1.

Freesurfer has been used in conjunction with GNU Parallel [11] to process the structural MRI data. All of the MRI data was processed using the same version of Freesurfer and the same version of the operating system (according to [12], differences in these two factors can affect the output).

The data mining software KNIME [13] using an improved version of the plug-in K-Surfer [14] was used to extract the required features from the processed MRI data. There are 356 features extracted in total - various thickness, volume and surface area measurements of regions of interest across the brain; these features were then merged with data ADNI provides about the subject which cannot be inferred from the brain data - the age and gender of the subject. ADNI also provides the diagnosis of the subject, whether they are a HC subject or suffering from MCI or AD. Thus there are a total of 358 features which can be used to predict the one output class.

Intra cranial volume (ICV) normalisation [15] is a process which alters the data for each subject to account for variations in head size (as this affects the size of ROIs within the brain) and is often used in classification of dementia from MRI data [16] [17]. Feature selection will both be tested with and without ICV normalisation as in [6] it was found in some situations the classification had a higher accuracy without ICV normalisation. Z-score normalisation is performed since SVM algorithms typically assume that the data is within a standard range; if the normalisation is not performed then the SVM can be adversely affected and misclassify the data.

TABLE I: Information about the subjects used

Dataset	Diagnosis	#Subjects	%Male	Age
T1	HC	117	49.1	74.4 ± 6.10
	MCI	117	50.0	73.9 ± 6.54
	AD	117	52.3	75.4 ± 7.93
T2	HC	28	50.0	72.6 ± 6.67
	MCI	28	46.7	73.7 ± 6.74
	AD	28	57.7	74.1 ± 7.53

The MRI data will be split into training data (T1) and test data (T2) - the classifier will be trained on the training data thus is will know the diagnosis of these subjects, it will be evaluated on the test data and will have to predict the classes of this data. Stratified sampling will be used so that the both sets have a similar proportion of the three classes. 351 subjects will be used as the training data and the remaining 84 will be used as test data, each data set is class balanced meaning that there are an equal number of subjects with one class as the other two classes, further information can be found in Table I.

B. Feature Selection Algorithm

GAs were pioneered by John Holland [18] and can be applied to numerous types of problems. They are based on principles of evolution such that solutions (chromosomes) are generated for a problem and a fitness value is calculated for how the given solution solves the problem. Then the solutions are bred with each other (an operation which takes elements from two solutions to generate a one or more solutions) to form a new solution which is then mutated (mutation involves randomly changing part of the solution). This repeats until the desired number of child solutions is met and a new generation is created. Then the solutions are evaluated and bred again, which continues until a termination criteria is met such as a certain number of generations has elapsed or a certain fitness value has been reached.

In this application, each potential solution is represented by a bit string and length is equal to the number of features available. The bit's value depends on whether the feature has been included or excluded for this solution. For example if the ith bit is 1, then the ith feature will be passed to the classifier and it will train using that feature; and if the jth bit is 0, then the jth feature will be ignored by it. Initially the chromosomes will be initialised randomly with a probability, p_I that the bit will be a 1. The crossover method used will be one-point crossover where two parents produce two children: a random index is chosen in the bit strings and the data beyond that index is swapped between each string producing two children. Mutation will be implemented via bit flipping, each bit in each string will be flipped with a probability, p_M.

The fitness function (the function which is used to evaluate how well each chromosome performs at solving the problem) will be set to the classification accuracy achieved since the aim of the feature selection is to increase the successful classification rate of the problem; a second GA will be run with a modified fitness function whereby the chromosomes are penalised for having more than 20 features, for every feature over 20, the fitness is decreased by a value of 5 (this value of 5 is equivalent to a 5% drop in accuracy). This penalty is used to keep the number of features low as this is the aim of this research - to create a model with a small number of features. Parent selection is the mechanism which chooses which parents breed together to produce the offspring for the next generation, and the method used is stochastic universal sampling which removes the bias fitness proportionate methods have towards only selecting solutions with the highest fitness [19]. For the development and testing of the GA described in this paper, the programming language R [20] was used along with external packages for the GA [21] and the SVM [22].

C. Classifying the Data

The classification problem is a three-class problem, [23] found that classifiers had performed better when multi-class problems were split into binary-class problems, the classification result of each binary-class problem are then combined using an aggregation method to obtain a classification for the multi-class problem. Following on from this research, the ternary-class problem will be split into three binary-class problems and a GA will be trained for each of these binary-class problems. The Weighted Voting Strategy (WV) aggregation method in [23] for the SVM will be tested to see how well it performs at combining the three two-class SVMs against the single three-class SVM.

The fitness of each chromosome of the GA will be calculated from the accuracy of an SVM with a Radial Basis Function (RBF) kernel, $k(x, x') = \exp(-\sigma||x - x'||^2)$, using the given feature set, the SVM will be trained on the training data using 10-fold cross validation. Once the termination criteria is reached, the feature set which gave the highest accuracy will be trained on the entirety of the training data and then tested on the test data, and the accuracy of this will be a measurement of how well the final feature set chosen performs.

III. Results

A GA-based feature selection method was used in each of the 24 classification problems (e.g. one of these 24 problems is: HC vs. MCI with ICV normalisation using the cortical fields). The 24 cases include binary tests with and without ICV normalisation, multi-label classification with and without ICV normalisation, over three different initial sets of features. The All Fields subset contains both the cortical subfields and hippocampal subfields, gender and age of the subject aren't included in this subset as the aim is to select the best features of the MRI data not MRI data augmented with other features. The Cortical Fields subset contains all the fields generated by Freesurfer's recon-all command with the -all flag. The Hippocampal Subfields subset contains only the fields generated by the -hippo-subfields flag.

The parameters of the GA used are: a crossover rate of 0.6, a mutation rate of 0.02 (the probability that each bit of the

TABLE II: Results of the two-class SVMs to solve the binary classification problems using a GA for feature selection. Classification problems marked with a (p) is when the GA is run with a penalty for a feature set with more than 20 features; problems marked with (NoFS) are the result of the classification using no feature selection.

Classification Problem	**All Fields**						**Cortical Fields**						**Hippocampal Subfields**					
	ICV			*No ICV*			*ICV*			*No ICV*			*ICV*			*No ICV*		
	T1	*T2*	#Fea.	*T1*	*T2*	#Fea.	*T1*	*T2*	#Fea.	*T1*	*T2*	#Fea.	*T1*	*T2*	#Fea.	*T1*	*T2*	#Fea.
HC vs. MCI	65.7	62.5	186	64.6	67.9	184	64.7	58.9	182	66.6	67.9	169	64.5	71.4	7	67.3	48.2	11
HC vs. MCI (p)	67.6	67.9	16	69.7	66.1	20	67.6	64.3	10	73.9	55.4	11	65.9	57.1	7	68.0	37.5	7
HC vs. MCI (NoFS)	56.8	64.3	356	58.8	69.6	356	56.8	60.7	340	58.1	67.9	340	57.7	60.7	16	56.7	67.9	16
HC vs. AD	88.8	82.1	196	90.0	80.4	168	85.8	76.8	178	87.0	76.8	182	88.6	87.5	6	87.2	78.6	8
HC vs. AD (p)	89.3	78.6	15	88.8	75.0	20	88.3	80.4	13	83.8	73.2	7	88.5	89.3	6	86.2	76.8	8
HC vs. AD (NoFS)	85.7	82.1	356	86.1	80.4	356	82.7	82.1	340	84.0	76.8	340	84.7	87.5	16	83.0	75.0	16
MCI vs. AD	83.1	76.8	165	85.3	75.0	169	83.0	76.8	170	82.8	76.8	175	80.9	73.2	6	80.4	60.7	8
MCI vs. AD (p)	82.8	73.2	20	86.0	75.0	20	84.8	78.6	15	83.7	75.0	19	80.5	69.6	7	80.4	57.1	9
MCI vs. AD (NoFS)	80.6	78.6	356	81.0	75.0	356	79.2	82.1	340	79.7	73.2	340	75.3	71.4	16	76.2	62.5	16

TABLE III: Performance of the three-class SVM and the two-class SVMs combined using WV. Note that the combined two-class SVMs were not tested against T1, thus there are no results. The best results are shown in bold. The number of features for the combined binary classifiers is the length of the union of the features from the individual classifiers. Classification problems marked with a (p) are results from the GA run with a penalty. Results marked with a * are when the SVM predicted all subjects to be of one class (for example, all subjects were predicted to be AD).

Classification Problem	**All Fields**						**Cortical Fields**						**Hippocampal Subfields**					
	ICV			*No ICV*			*ICV*			*No ICV*			*ICV*			*No ICV*		
	T1	*T2*	#Fea.	*T1*	*T2*	#Fea.	*T1*	*T2*	#Fea.	*T1*	*T2*	#Fea.	*T1*	*T2*	#Fea.	*T1*	*T2*	#Fea.
HC vs. MCI vs. AD	62.1	59.5	170	62.8	58.3	187	60.9	65.5	171	65.2	33.3*	178	62.4	54.8	9	62.4	40.5	9
HC vs. MCI vs. AD (p)	61.5	58.3	19	65.3	52.4	9	61.0	63.1	20	61.6	33.3*	20	60.4	48.8	11	61.8	39.3	7
HC vs. MCI vs. AD (NoFS)	-	63.1	356	-	60.7	356	-	64.3	340	-	33.3*	340	-	50.0	16	-	39.3	16
Binary Combined	-	61.9	309	-	64.3	302	-	60.7	311	-	60.7	298	-	53.6	13	-	51.2	15
Binary Combined (p)	-	61.9	47	-	64.3	58	-	60.7	38	-	60.7	34	-	53.6	11	-	51.2	14
Binary Combined (NoFS)	-	61.9	356	-	64.3	356	-	60.7	340	-	60.7	340	-	53.6	16	-	39.3	16

bit string representation will flip), a population size of 50, single-point crossover and roulette wheel parent selection. At the start of the GA, each chromosome is initialised with a probability of 0.5 that a bit is a 1 instead of 0. The results of the GA feature selection for the two-class problems are in Table II, and the results of the GA feature selection for three-class problems and also the results of the two-class SVMs combined with WV are in Table III. Within these tables, T1 refers to the accuracy found of the best performing genotype when calculating its accuracy using 10-fold cross validation on the training dataset (351 subjects); and T2 is the accuracy found when the best features found from the GA are used to train an SVM on the entire training dataset and then used to evaluate the test dataset - a holdout method. The second GA - using the fitness function with the penalty for over 20 features, behaves similarly to the former GA; other than the fitness function the only other difference is the initialisation of the chromosomes. Every chromosome has a random ten features selected, all of which are set to be included (every other feature is excluded).

IV. Discussion

A. Binary Classification Problems

Feature selection increased the accuracy of the classifier performance for feature selection of the hippocampal subfields. For the binary classification problems, the highest accuracies achieved were 71.4% for HC vs. MCI using the hippocampal subfields with ICV normalisation and the standard GA for feature selection. The highest accuracy for HC vs. AD was 89.3% which used ICV normalisation and the GA with a penalty for feature selection. MCI vs. AD again used ICV normalised fields for its highest accuracy, 73.2% accuracy was achieved using the standard GA. ICV normalisation was an important factor in achieving the highest accuracies for classification based on the hippocampal subfields as usage of ICV normalisation created classifiers with a higher accuracy than whenever ICV normalisation was not used - thus the ICV is likely an important factor when dealing with the hippocampal regions of the brain.

In both the feature selection of the cortical fields and feature selection of all fields, the accuracy was not improved by feature selection which was discovered in [3] where feature selection was tested with an SVM and very high dimensional data and found that when feature selection was performed, a lower accuracy was achieved. The joint best or best accuracy was achieved by using no feature selection whatsoever (except when ICV was applied to all fields and cortical fields of the HC vs. MCI classification problem - the GA using the penalty reached the highest accuracy here). Using feature selection on all of the fields the highest accuracies obtained were 69.6% for HC vs. MCI using no feature selection without ICV normalisation; 82.1% was the best for HC vs. AD achieved via no feature selection and also feature selection with the GA showing that 196 features can be used to obtain the same accuracy as all 356 features; and for MCI vs. AD, 78.6% was reached using no feature selection with ICV normalisation. Using feature selection of the cortical fields, the best accuracies achieved were: 67.9% for HC vs. MCI with both no feature selection without ICV normalisation and also 67.9% using the GA for feature selection also without ICV normalisation. The highest accuracy for both HC vs. AD and MCI vs. AD was 82.1% by using no feature selection with ICV normalisation.

B. Ternary Classification Problems

Of the entire 3-class classification problem, the best accuracy was achieved by a 3-class SVM with feature selection performed by a GA using the cortical fields with ICV normalisation, this achieved an accuracy of 65.5%. The 3-class SVM generally performed better than the combined 2-class SVMs, except in the cortical fields without ICV normalisation where the SVM always predicted one class for every subject. Feature selection again performed better for the hippocampal subfields with the best accuracy being achieved by the GA being used for feature selection on the ICV normalised data.

V. Conclusion

This work has applied GA-based feature selection in conjunction with an SVM to classify HC, MCI and AD patients from structural MRI brain data. The results have shown that a high accuracy can be achieved using just the hippocampal fields as a feature set. An SVM performs better with feature selection when feature selection is applied to cortical fields and also when it is applied to hippocampal fields; however, when applied to all of the fields, an SVM without feature selection performs better. Regarding the SVM performing better with feature selection of the cortical fields and hippocampal fields this could mean that there are some irrelevant features in both the cortical fields and hippocampal subfields that the SVM cannot handle and thus by using feature selection, these irrelevant features are removed and allow the SVM to perform better.

The hippocampal subfields are great predictors for distinguishing between HC, MCI and AD; only a small number of features are needed to achieve a fairly high accuracy, showing that the effects of AD and MCI are prevalent in the hippocampus (agreeing with other literature on this topic [9] [10]). From this finding, there is potential for the hippocampus subfields to be used to create a simple model to provide an understanding of why a patient was classified as HC, MCI or AD.

The accuracy obtained by the combined two-class SVMs was slightly lower than the single three-class SVMs, this could be down to the randomness of the GA in which feature subsets it evaluates - as in previous literature [23] the combined two-class classifiers performed better; another reason is that for this problem, a three-class SVM performs better than combined two-class classifiers. In general, the usage of ICV normalisation results in a classifier that performs better as from the results regarding the hippocampal subfields without ICV in Table II show that the accuracy on T2 is significantly lower than the accuracy on T1 suggesting that an SVM which overfits

the training data has been created. In summary, a predictive model to distinguish between the classes HC, MCI and AD, can be trained with a small number of features that provides near the same accuracy as the entire set of 356 features. This is useful for creating a report to show which areas of the patient's brain are most affected by MCI and AD. Future work would involve exploiting the power of the accuracy obtained using solely the hippocampal subfields to create a simple classifier which has rules that can be understood by both doctor and patient (provided the disease hasn't progressed too far).

Acknowledgment

Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at: http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf

References

[1] R. Brookmeyer, E. Johnson, K. Ziegler-Graham, and H. M. Arrighi, "Forecasting the global burden of alzheimer's disease," *Alzheimer's & Dementia*, vol. 3, no. 3, pp. 186 – 191, 2007. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S155252600700475X

[2] M. Grundman, R. C. Petersen, S. H. Ferris, and et al, "Mild cognitive impairment can be distinguished from alzheimer disease and normal aging for clinical trials," *Archives of Neurology*, vol. 61, no. 1, pp. 59–66, 2004. [Online]. Available: +http://dx.doi.org/10.1001/archneur.61.1.59

[3] R. Nilsson, J. Peña, J. Björkegren, and J. Tegnér, "Evaluating feature selection for svms in high dimensions," in *Machine Learning: ECML 2006*, ser. Lecture Notes in Computer Science, J. Fürnkranz, T. Scheffer, and M. Spiliopoulou, Eds. Springer Berlin Heidelberg, 2006, vol. 4212, pp. 719–726. [Online]. Available: http://dx.doi.org/10.1007/11871842_72

[4] P. Johnson, L. Vandewater, W. Wilson, P. Maruff, G. Savage, P. Graham, L. S. Macaulay, K. A. Ellis, C. Szoeke, R. N. Martins, C. C. Rowe, C. L. Masters, D. Ames, and P. Zhang, "Genetic algorithm with logistic regression for prediction of progression to alzheimer's disease," *BMC Bioinformatics*, vol. 15(Suppl 16):S11, no. 6, p. e38234, 12 2014. [Online]. Available: http://www.biomedcentral.com/1471-2105/15/S16/S11

[5] H. Frohlich, O. Chapelle, and B. Scholkopf, "Feature selection for support vector machines by means of genetic algorithm," in *Tools with Artificial Intelligence, 2003. Proceedings. 15th IEEE International Conference on*, Nov 2003, pp. 142–148.

[6] A. Sarica, G. D. Fatta, G. Smith, M. Cannataro, and J. D. Saddy, "Advanced feature selection in multinominal dementia classification from structural mri data," *Proc MICCAI Workshop Challenge on Computer-Aided Diagnosis of Dementia Based on Structural MRI Data*, pp. 82–91, 2014.

[7] L. Sørensen, A. Pai, C. Anker, I. Balas, M. Lillholm, C. Igel, and M. Nielsen, "Dementia diagnosis using mri cortical thickness, shape, texture, and volumetry," *Proc MICCAI Workshop Challenge on Computer-Aided Diagnosis of Dementia Based on Structural MRI Data*, pp. 111–118, 2014.

[8] R. Cuingnet, E. Gerardin, J. Tessieras, G. Auzias, S. Lehéricy, M.-O. Habert, M. Chupin, H. Benali, and O. Colliot, "Automatic classification of patients with alzheimer's disease from structural mri: A comparison of ten methods using the ADNI database," *NeuroImage*, vol. 56, no. 2, pp. 766 – 781, 2011, multivariate Decoding and Brain Reading. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1053811910008578

[9] K. Jin, A. L. Peel, X. O. Mao, L. Xie, B. A. Cottrell, D. C. Henshall, and D. A. Greenberg, "Increased hippocampal neurogenesis in alzheimer's disease," *Proceedings of the National Academy of Sciences*, vol. 101, no. 1, pp. 343–347, 2004. [Online]. Available: http://www.pnas.org/content/101/1/343.abstract

[10] M. Ball, V. Hachinski, A. Fox, A. Kirshen, M. Fisman, W. Blume, V. Kral, H. Fox, and H. Merskey, "A new definition of alzheimer's disease: A hippocampal dementia," *The Lancet*, vol. 325, no. 8419, pp. 14 – 16, 1985, originally published as Volume 1, Issue 8419. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0140673685909651

[11] O. Tange, "Gnu parallel - the command-line power tool," *;login: The USENIX Magazine*, vol. 36, no. 1, pp. 42–47, Feb 2011. [Online]. Available: http://www.gnu.org/s/parallel

[12] E. H. B. M. Gronenschild, P. Habets, H. I. L. Jacobs, R. Mengelers, N. Rozendaal, J. van Os, and M. Marcelis, "The effects of freesurfer version, workstation type, and macintosh operating system version on anatomical volume and cortical thickness measurements," *PLoS ONE*, vol. 7, no. 6, p. e38234, 06 2012. [Online]. Available: http://dx.doi.org/10.1371%2Fjournal.pone.0038234

[13] M. Berthold, N. Cebron, F. Dill, T. Gabriel, T. Kötter, T. Meinl, P. Ohl, C. Sieb, K. Thiel, and B. Wiswedel, "Knime: The konstanz information miner," in *Data Analysis, Machine Learning and Applications*, ser. Studies in Classification, Data Analysis, and Knowledge Organization, C. Preisach, H. Burkhardt, L. Schmidt-Thieme, and R. Decker, Eds. Springer Berlin Heidelberg, 2008, pp. 319–326. [Online]. Available: http://dx.doi.org/10.1007/978-3-540-78246-9_38

[14] A. Sarica, G. Di Fatta, and M. Cannataro, "K-surfer: A knime-based tool for the management and analysis of human brain mri freesurfer/fsl data," *Frontiers in Neuroinformatics*, no. 3. [Online]. Available: http://www.frontiersin.org/neuroinformatics/10.3389/conf.fninf.2014.18.00003/full

[15] J. L. Whitwell, W. R. Crum, H. C. Watt, and N. C. Fox, "Normalization of cerebral volumes by use of intracranial volume: implications for longitudinal quantitative MR imaging," *AJNR Am J Neuroradiol*, vol. 22, no. 8, pp. 1483–1489, Sep 2001.

[16] L. L. Chao, S. T. Buckley, J. Kornak, N. Schuff, C. Madison, K. Yaffe, B. L. Miller, J. H. Kramer, and M. W. Weiner, "ASL perfusion MRI predicts cognitive decline and conversion from MCI to dementia," *Alzheimer Dis Assoc Disord*, vol. 24, no. 1, pp. 19–27, 2010.

[17] Y. Fan, S. M. Resnick, X. Wu, and C. Davatzikos, "Structural and functional biomarkers of prodromal alzheimer's disease: a high-dimensional pattern classification study," *Neuroimage*, vol. 41, no. 2, pp. 277–285, 2008.

[18] J. H. Holland, *Adaptation in Natural and Artificial Systems*. Cambridge, MA, USA: MIT Press, 1992.

[19] J. E. Baker, "Reducing bias and inefficiency in the selection algorithm," in *Proceedings of the Second International Conference on Genetic Algorithms on Genetic Algorithms and Their Application*. Hillsdale, NJ, USA: L. Erlbaum Associates Inc., 1987, pp. 14–21. [Online]. Available: http://dl.acm.org/citation.cfm?id=42512.42515

[20] R Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2015. [Online]. Available: http://www.R-project.org/

[21] L. Scrucca, "Ga: A package for genetic algorithms in r," *Journal of Statistical Software*, vol. 53, no. 4, pp. 1–37, 4 2013. [Online]. Available: http://www.jstatsoft.org/v53/i04

[22] A. Karatzoglou, A. Smola, K. Hornik, and A. Zeileis, "kernlab – an S4 package for kernel methods in R," *Journal of Statistical Software*, vol. 11, no. 9, pp. 1–20, 2004. [Online]. Available: http://www.jstatsoft.org/v11/i09/

[23] M. Galar, A. Fernández, E. Barrenechea, H. Bustince, and F. Herrera, "An overview of ensemble methods for binary classifiers in multi-class problems: Experimental study on one-vs-one and one-vs-all schemes," *Pattern Recognition*, vol. 44, no. 8, pp. 1761 – 1776, 2011. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0031320311000458