FILTER FEATURE SELECTION ANALYSIS TO DETERMINE THE CHARACTERISTICS OF DEMENTIA

Savaş OKYAY[1*], Nihat ADAR[2]

[1]Eskişehir Osmangazi University, Faculty of Engineering and Architecture, Department of Computer Engineering, Eskişehir, Turkey, ORCID No: https://orcid.org/0000-0003-3955-6324
[2]Eskişehir Osmangazi University, Faculty of Engineering and Architecture, Department of Computer Engineering, Eskişehir, Turkey, ORCID No: https://orcid.org/0000-0002-0555-0701

Keywords
Alzheimer's disease
Dementia
Filter feature selection
ADNI
Freesurfer

Abstract

Dementias are known as neuropsychiatric disorders. Two-dimensional sliced brain scans can be generated via magnetic resonance imaging. Three-dimensional measurements of regions can be reached from those scans. Numerical brain features can be extracted through operating the Freesurfer tool. Parametrizing those features and demographic information in learning algorithms can label an unknown sample as healthy or dementia. On the other hand, some of the features in the initial set may be less practical than others. In this research, the aim is to decrease the input feature-count, a total of 2939 attributes, as a first step to determine the most distinctive dementia characteristics. To that end, a total of 2264 ADNI dataset samples (471 AD, 428 lMCI, 669 eMCI, 696 healthy controls) are divided into two sets: 65% training set (1464 samples) and 35% test set (800 samples). Various filter feature selection algorithms (Information Gain, Gain Ratio, Symmetrical Uncertainty, Pearson's Correlation, Correlation-based Feature Subset Selection) are tested over different parameters together with multiple Bayesian-based and tree-based classifiers. Test performance accuracy rates up to 76.50% are analyzed in detail. Instead of processing the whole feature set, the overall performance tends to increase with correctly fewer attributes taken.

DEMANS ÖZELLİKLERİNİN BELİRLENMESİ İÇİN FİLTRE ÖZNİTELİK SEÇİM ANALİZİ

Anahtar Kelimeler
Alzheimer hastalığı
Demans
Filtre öznitelik seçimi
ADNI
Freesurfer

Öz

Demans hastalıkları nöropsikiyatrik bozukluklar olarak tanımlanır. Manyetik rezonans görüntüleme teknikleri ile iki boyutlu dilimlenmiş beyin taramaları oluşturulabilir. Bu taramalar üzerinden bölgelerin üç boyutlu ölçümlerine ulaşılabilir. Sayısal beyin özellikleri Freesurfer aracı kullanılarak çıkarılabilmektedir. Bu özelliklerin ve demografik verinin öğrenme algoritmalarında parametreler olarak yer almasıyla, bilinmeyen bir örnek, sağlıklı veya demans olarak etiketlenebilir. Öte yandan, tüm özellik setindeki bazı öznitelikler diğerlerine göre daha az yararlı veya direkt etkisiz olabilir. Bu araştırmanın amacı, en belirgin demans özelliklerini belirlemek adına ilk adım olarak toplamda 2939 olan girdi özniteliklerinin sayısını azaltmaktır. Bu amaçla, ADNI veri setindeki toplam 2264 numune (471 AH, 428 gHBB, 669 eHBB, 696 sağlıklı kontrol), %65 eğitim seti (1464 numune) ve %35 test seti (800 numune) olmak üzere iki gruba ayrılmaktadır. Çeşitli filtre öznitelik seçim algoritmaları (Bilgi Kazanımı, Kazanç Oranı, Simetrik Belirsizlik, Pearson Korelasyonu, Korelasyona Dayalı Öznitelik Alt Kümesi Seçimi), Bayes tabanlı ve ağaç tabanlı sınıflandırıcılarla birlikte farklı parametreler üzerinden test edilmektedir. %76,50'ye varan test performans doğruluğu oranları ayrıntılı olarak analiz edilmektedir. Öznitelik setinin tamamını işlemek yerine, doğru şekilde daha az öznitelik alındığında genel performans artış eğilimindedir.

* Corresponding author; e-mail : osavas@ogu.edu.tr

Araştırma Makalesi		Research Article	
Başvuru Tarihi	: 14.07.2020	Submission Date	: 14.07.2020
Kabul Tarihi	: 13.12.2020	Accepted Date	: 13.12.2020

1. Introduction

Dementia diseases are defined as neuropsychiatric disorders and are among the most significant problems of old age. It is stated that as the age gets older, the rate of individuals getting dementia diseases is quite high, and the number of occurrences of such disorders increases day by day (Alam, Kwon, & Initi, 2017). Despite being aware of the importance of Alzheimer's disease (AD), the well-known subtype of dementia, there is currently no cure for its healing effect. However, only prescription medications can help delay the progress of the condition (Lama, Gwak, Park, & Lee, 2017). Mild cognitive impairment (MCI) is another one of the subtypes of those disorders and characterized by memory impairment in the absence of dementia (Yao, Calhoun, Fu, Du, & Sui, 2018). MCI is also known as age-related cognitive decline and the transition phase to AD. Early (eMCI)/late (lMCI) may be defined as the clinically detectable early stages of progression to dementia or AD (Moradi et al., 2015).

Magnetic resonance imaging (MRI) is one of the visualization techniques that give the advantage to understand body anatomy. Two-dimensional sliced *dicom*-type (Digital Imaging and Communications in Medicine) scans can be generated via MRI devices. During medical examinations, clinical measurements such as cortical volume, thickness, and surface area are observed, and clinical decision about likely to have one of the dementia diseases is tried to be diagnosed. Technological refinements, either hardware or software, are always needed to facilitate clinical diagnosis procedures.

Machine learning is one of the most critical developments in the last decade in terms of computing workload within the scope of neuroimaging, as in other disciplines. It can offer neuroscientists, radiologists, and clinicians with tools for automatic and early diagnosis of brain diseases (Dimitriadis, Liparas, & Initi, 2018). Multivariate pattern analysis provides powerful options for creating useful image-based prediction models for classification. In theory, when the features are selected correctly, they may be benefited better than individual radiological evaluation to estimate the clinical score (Doan et al., 2017). By analyzing two-dimensional sliced brain scans, three-dimensional region measurements can be expressed, and desired features can be extracted.

In machine learning studies, some of the features in the initial set may be more valuable than others. On the contrary, some features may have no positive effect on the designed application. Accordingly, studies examining the sub-feature sets through blind search become essential, to facilitate the clinical diagnosis of doctors, and to make the health status prediction of an unlabeled sample more successful. Thus, such researches in this field may be described as much more helpful in order to give perspective to other studies to be conducted. Furthermore, feature analysis or selection algorithms gain great importance, especially when the features are desired to analyze within themselves. In some researches on this subject, it is observed that the brain characteristics are evaluated by weight coefficients via operating methods such as analysis and comparison (Dimitriadis, Liparas, Tsolaki, & Initia, 2018). However, when the beneficial attributes listed in the neuroimaging literature are examined, approaches using feature selection algorithms are not frequently encountered.

Feature selection approaches are needed in many aims, such as reducing the size of the input matrix, finding valuable subsets, or excluding ineffective ones. The wrapper approaches, in which classification algorithms are integrated with, appear in the literature. The filter feature selection approaches, such as regression, etc.,

Table 1

Summary of the Literature (The Most Similar Studies)

Publication	Dataset	Feat. Ext.	Initial Feature Set	Feat. Selection	Classifier	Performance
Yao et al., 2018 DOI	ADNI - 400 samples (60 HC, 60 eMCI, 60 lMCI, 60 AD, 40×4 unlabeled)	FS v5.3.0	426 morphometric, 3 demographic	Wrapper method (own proposal)	Various (SVM, NB, RF, ...)	54.38% (4-class Kaggle competition)
Sorensen et al., 2017 DOI	ADNI - 504 samples (169 HC, 234 MCI, 101 AD) and AIBL (...)	FS v5.1.0	cortical thickness, volumetric measures, hippocampal shape, hippocampal tissue	SFS	LDA	63% (Multi-class classification)
Dimitriadis et al., 2018 DOI	ADNI - 400 samples (60 HC, 60 eMCI, 60 lMCI, 60 AD, 40×4 unlabeled)	FS v5.3.0	thickness, surface area, cortical curvature, grey matter density, cortical and subcortical volume, hippocampus shape and area volumes, MMSE, age	RF	RF	61.9% (4-class Kaggle competition)
Ramirez et al., 2018 DOI	ADNI - 400 samples (60 HC, 60 eMCI, 60 lMCI, 60 AD, 40×4 unlabeled)	FS v5.3.0	429 morphometric and demographic	ANOVA feature selection, PLS feature reduction	Various (SVM, RF, ...)	56.25% (4-class Kaggle competition)

which are independent of classifiers and run quickly, are also popular.

The most similar studies in neuroimaging literature by (Yao et al., 2018), (Sorensen et al., 2017), (Dimitriadis, Liparas, Tsolaki, et al., 2018), (Ramirez et al., 2018) are summarized in Table 1.

In this experimental research, it is aimed to decrease the feature-count of the dataset as a first step to determine the most valuable brain characteristics that distinguish dementia diseases from each other. Classification tests are carried out to determine more practical algorithms, also more correct parameters, where the number of features is reduced to reasonable levels. The preliminary results obtained in this article are intended to guide future studies.

The article is organized as follows. In the first section, the methodology of the aging problem and the neuroimaging field is introduced. Also, the literature is reviewed. The dataset is pointed out in Section 2. The virtual brain modeling, feature extraction, feature set preprocessing, and classification processes are theoretically explained in Section 3. In Section 4, experimental tests and the findings are reported. In the last section, the conclusion is drawn, and future work is planned.

2. The Dataset

Data used in the preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu, ADNI1Complete1Yr1.5T). The ADNI was launched in 2003 as a public-private partnership, led by Principal Investigator Michael W. Weiner, MD. The primary goal of ADNI has been to test whether serial MRI, positron emission tomography (PET), other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of mild cognitive impairment (MCI) and early Alzheimer's disease (AD). For up-to-date information, see http://www.adni-info.org. The dataset is described as free-access. Two-dimensional and sliced brain scans of anonymous patients are available for download to researchers without any charge. The use of the dataset does not pose any ethical problems (http://adni.loni.usc.edu/data-samples/access-data/).

Table 2
Distribution of the Class Labels and Genders

	Train	Test	Total
	♂ ♀	♂ ♀	♂ ♀
AD	308	163	**471**
	157 151	95 68	252 219
lMCI	280	148	**428**
	198 82	94 54	292 136
eMCI	434	235	**669**
	276 158	133 102	409 260
HC	442	254	**696**
	237 205	124 130	361 335
Total	**1464**	**800**	**2264**
	868 596	446 354	**1314 950**

Since the number of the individuals in the dataset is large enough, the samples are automatically separated into two for once, approximately 65% training and 35% test sets for each class. Next, all standalone tests are performed utilizing the same split data. In Table 2, information is given about the distribution of the class labels and genders in the dataset.

3. Methodology

In this study, research and publication ethics principles were followed. The methods applied in the substeps of this research are detailed in the following subsections.

3.1. Feature Extraction

The feature extraction step of this work is performed by executing the Freesurfer (FS) software tool for each sample of the dataset. FS (Fischl, 2012) is known as a software analysis tool widely preferred in medical researches, and it is detailly functional for virtual three-dimensional modeling of brain anatomy. It is briefly a set of algorithms that contain some image processing, numerical, etc. algorithms and facilitate the structural analysis of the human brain. The tool is developed at the Athinoula A. Martinos Biomedical Imaging Center for analysis of structural and functional neuroimaging data (Delgado et al., 2014; Fischl, 2012). It can be licensed for free by downloading the appropriate version from http://surfer.nmr.mgh.harvard.edu/fswiki (Delgado et al., 2014).

FS, for the first substep of the technical procedure, structurally combines the sliced medical image files of *dicom* or *nii* types, handling the header information. Then, virtual brains are generated in three-dimensions by managing image processing techniques numerically and iteratively (Reuter, Schmansky, Rosas, & Fischl, 2012). Nevertheless, the needed duration required for just one virtual modeling process to complete can take much time.

The measurements of brain regions on sliced scans were obtained as raw statistical files with parallel design study and reported in February 2018. A parallel approach to virtually 3D model all the samples in the dataset within a reasonable time was proposed. In other words, an algorithm was developed that could perform the feature extraction step for different individuals simultaneously. ADNI dataset containing 2292 samples were processed with carrying out the Message Passing Interface (MPI) library. With the launch of a 160-core

computer cluster in total, the entire workload could take about two consecutive years in theory, while the proposed algorithm was completed in just seven days. With only a few exceptions, all the morphometric features of the samples were successfully extracted (Okyay & Adar, 2018).

3.2. Initial Feature Set

Most of the studies, such as exemplified in the literature part, begin their works by choosing only certain features in the data preprocessing step (see Table 1). Besides, a large part of such studies evaluates the performance only with the accuracy metric. Within the scope of the research described in this article and its continuation, in the big picture, the most valuable attributes that distinguish diseases are tried to be determined. Therefore, 2939 different values obtained from a total of 2292 samples are analyzed through blind search logic. In detail, these features are all morphometric features extracted from the FS software tool and demographic information, including gender, age, and different test results such as CDR (Clinical Dementia Rating), MMSE (Mini-Mental State Examination), FAQ (Functional Assessment Questionnaire).

3.3. Preprocessing

The preprocessing steps are carried out after a few steps. First of all, thirty-one inefficacious attributes of various types such as date, naming, and identifiers are manually eliminated from the 2292×2939 sized dataset. Afterward, the values of the attributes are examined, and more than two hundred characteristics found to be useless are automatically removed. Moreover, it is observed that twenty-eight samples do not contain enough information. As a result of all these elimination processes, a feature matrix of 2264×2691 elements is obtained. Since filtering approaches are performed to evaluate how attribute values distributed within the samples, no normalization technique is applied to the feature matrix.

3.4. Filter Feature Selection

In the filter feature selection approach, the data is analyzed independently of learning algorithms. In this study, various filtering methods are evaluated to reduce the data size, the feature vector length, at an appropriate threshold. The basis of some of the algorithms is based on the entropy calculation, which expresses the irregularity of the system and uses the possibilities in the relation of between the features and the class labels (1).

$$H(x) = -\sum_{i=1}^{N} (p(X_i) \times \log_2 p(X_i)) \quad (1)$$

The algorithms utilized in this study are detailed in the following subsections.

3.4.1. Information Gain

Calculates the weight of an attribute by measuring the gain of information according to its relationship with class values (2).

$$InfoGain(cls, feat) = H(cls) - H(cls|feat) \quad (2)$$

3.4.2. Gain Ratio

Calculates the weight of an attribute by measuring the rate of earnings based on class values (3).

$$GRatio(cls, feat) = \big(H(cls) - H(cls|feat)\big)/H(feat) \quad (3)$$

3.4.3. Symmetrical Uncertainty

Calculates the weight of an attribute by measuring the symmetrical uncertainty concerning the class values (4).

$$SymU(cls, feat) = 2 \times \frac{\big(H(cls) - H(cls|feat)\big)}{H(cls)} + H(feat) \quad (4)$$

3.4.4. Pearson's Correlation

Calculates the weight of an attribute by measuring the Pearson's correlation between it and the class values (5). The result up to `+1` means a positive correlation, while down to `-1` means a negative relationship.

$$PearsonCor(cls, feat) = \frac{\sum_i \big((cls_i - \overline{cls}) \times (feat_i - \overline{feat})\big)}{\sqrt{\sum_i (cls_i - \overline{cls})^2} \times \sqrt{\sum_i (feat_i - \overline{feat})^2}} \quad (5)$$

3.4.5. Correlation-based Feature Subset Selection

Calculates the worthiness of a feature subset through the predictive ability of each attribute along with the degree of redundancy between them. Unlike the other filter feature selection algorithms previously mentioned on subsections, subsets of highly correlated attributes with the class while having low intercorrelation are directly computed (Hall & Smith, 1998).

The highest scores obtained as the intermediate results of the filter feature selection algorithms are listed in Table 3. Since the CFS algorithm directly produces an optimal attribute subset result, all of its features are listed. `Top10` scores are selected for other methods.

The cut-off points at different thresholds of the filtered feature set and the elimination parameters are analyzed in the next section with the learning algorithms detailed in the following subsection.

Table 3
Top Scores of Filter Feature Selection Algorithms

Information Gain
Top 10

Feature	Unit	Ratio
CDR	test scr.	0.929
MMSE	test scr.	0.522
FAQ	test scr.	0.245
left *hippocam.*	volume	0.213
left *entorhinal*	thk avg	0.208
left *inferiorte.*	thk avg	0.205
left *hippocam.*	nm voxl	0.203
rght *middlete.*	thk avg	0.200
left *entorhinal*	thk avg	0.194
rght *hippoca.*	volume	0.192

Gain Ratio
Top 10

Feature	Unit	Ratio
CDR	test scr.	0.654
left *parsoperc.*	volume	0.242
left *parsoperc.*	nm voxl	0.242
rgt *occipitl ant*	nm vrtc	0.242
4. ventricle	nrm rng	0.235
left *cuneus*	thk std	0.233
left *cingl mar.*	mn crv	0.227
left *front inf*	srf area	0.219
left *temp. tran.*	srf area	0.208
right *chor. plx.*	nm vxl	0.200

Symmetrical Uncertainty
Top 10

Feature	Unit	Ratio
CDR	test scr.	0.548
MMSE	test scr.	0.224
FAQ	test scr.	0.119
left *hippocam.*	volume	0.108
left *temp. inf.*	thk avg	0.108
left *inferiorte.*	thk avg	0.107
left *entorhinal*	thk avg	0.105
left *hippocam.*	nm voxl	0.104
rght *entorhinl*	thk avg	0.102
rght *oc-temp.*	thk avg	0.101

Pearson's Correlation
Top 10

Feature	Unit	Ratio
CDR	test scr.	0.407
MMSE	test scr.	0.320
left *inferiorte.*	thk avg	0.262
FAQ	test scr.	0.259
left *hippocam.*	volume	0.254
left *temp. inf.*	thk avg	0.249
left *hippocam.*	nm voxl	0.249
rght *entorhinl*	thk avg	0.247
left *entorhinal*	thk avg	0.247
left *middlete.*	thk avg	0.246

Correlation-based Feature Subset Selection
All Features

Feature	Unit	Feature	Unit	Feature	Unit
CDR	test scr.	left *temp. tra.*	srf area	left *entorhinl*	thk std
subject age	year	left *pole occi.*	thk avg	left *middlete.*	w-mean
posterior	norm avg	left *oc temp.*	thk avg	rght *inf-oper.*	gray vol
optic chiasm	norm min	left *caudalan.*	gaus curv	rght *occipital*	mean crv
rght *pallidum*	norm std	left *parsoper.*	gray vol	rght *subcallo.*	mean crv
rght *amygdala*	nm voxl	left *middlete.*	mean crv	rght *lat-fis-a.*	mean crv
left *precentrl*	gaus curv	left *cuneus*	srf area	rght *inf-angu.*	nm vert
left *interm*	gray vol	left *fusiform*	srf area	rght *temporal*	thk avg
left *subparietl*	mean crv	left *parahipp.*	thk avg		

3.5. Classification

Learning algorithms based on different principles are adopted to measure the success rate of variable reduced feature counts created with varying algorithms of filtering. In order to analyze the probability of being included in classes, performances are comparably tested with Bayesian-based Naïve Bayes and BayesNet classifiers. Furthermore, to extract rule-based relationships between attributes, decision tree-based Random Tree and J48 Tree learning algorithms are experimented.

3.5.1. NaïveBayes

This algorithm is also recognizable as a conditional classifier. Given in the Bayes rule, it is assumed that pure independence assumption and all features are independent of each other. In the classification stage, conditional probabilities are computed for all possible class labels in which the test sample can be put in. The class with the maximum likelihood determines the tag on which the individual sample will be involved (Rish, 2001).

3.5.2. BayesNet

This algorithm is known as statistical networks, and the edges that switch between different nodes are selected statistically. The method practices to procure a connectional graph that focuses on meeting all the criteria, including conditional dependencies in the data (Murphy, 2001).

3.5.3. Random Tree

The algorithm has no pruning and tries to operate class probabilities that can be contained in the test feature vector. During computations, it processes a pre-defined number of randomly selected features on nodes (Aldous, 1991).

3.5.4. J48 Tree

This algorithm is known as a simple C4.5 decision tree (Sharma & Sahni, 2011). The method ensures the functioning of the decision mechanism by constructing a binary tree (Patil & Sherekar). It is performed as a univariate decision tree in which a feature is assigned to each inner node (Bhargava, Sharma, Bhargava, & Mathuria, 2013).

4. Tests

Information acquisition and correlation methods are applied within the scope of the goal of reducing the number of features at a specific ratio or threshold. Numerical evaluation is completed first for each attribute, and then those computations are ranked according to the weight scores. When sequential weights for all algorithms are examined analytically, apart from all features (T0), three critical weight values are determined. These are `0` (T1), `0.01` (T2), and `0.05` (T3) weight score thresholds in order. Additionally, `Top500` features (T4), `Top250` features (T5), and `Top100` features (T6) are defined as experimental testing criteria.

The classification results achieved after employing Information Gain, Gain Ratio, Symmetrical Uncertainty, Pearson's Correlation, and Correlation-based Feature Subset Selection (CFS) filtering approaches for different parameters processing the same training and test samples are listed respectively in Table 4-8. The train and test performances for each learning algorithm, according to the corresponding cut-off or break-point (bp), are given in detail.

When the findings are examined, depending on the decrease in the length of the features, the training performance results of different methods are not stable enough and are unconditionally fluctuating. While the training performance is generally on the rise, it may abruptly be contrary. Or vice versa. However, it should be strictly noted that instead of processing all the features in the dataset, the prediction accuracy in the test set tends to increase with fewer attributes taken as input, especially from the threshold of the `Top250` features.

Comparing the learning algorithms with each other shows that rule-based learning algorithms perform considerably better than Bayesian-based ones. Furthermore, the J48 Tree algorithm slightly comes to the forefront.

For the filtering algorithms, the Correlation-based Feature Subset Selection algorithm works well with Bayesian-based classifiers compared to the others. The results of the Pearson's Correlation algorithm are not impressive. The Gain Ratio and Symmetrical Uncertainty filtering approaches ensure more correct results in general.

The experimental test performance results are already comparable to previous studies. When examined in depth after a feature-count reduction method such as the `Top250` focus threshold, follow-up researches promise more reliable correctness in the name of the neuroimaging literature.

Table 4

Classification Accuracy after Information Gain

		NaïveBayes		BayesNet		Random Tree		J48 Tree	
bp	Feat Cnt	Train %	Test %	Train %	Test %	Train %	Test %	Train %	Test %
T0	2691	53.21	46.00	54.17	45.38	100	52.63	97.54	72.88
T1	1575	52.80	45.00	54.03	45.50	100	57.25	97.61	72.63
T2	1473	51.91	44.38	53.96	45.63	99.86	50.13	96.99	73.25
T4	500	51.09	47.75	53.89	46.88	100	56.25	97.40	71.50
T5	250	51.30	49.00	53.21	48.75	100	56.50	96.79	74.50
T3	222	51.16	49.25	53.21	48.63	100	61.13	96.65	75.63
T6	100	53.07	50.88	54.10	50.25	99.73	59.38	95.90	75.50

Table 5

Classification Accuracy after Gain Ratio

		NaïveBayes		BayesNet		Random Tree		J48 Tree	
bp	Feat Cnt	Train %	Test %	Train %	Test %	Train %	Test %	Train %	Test %
T0	2691	53.21	46.00	54.17	45.38	100	52.63	97.54	72.88
T1	1575	52.80	45.00	54.03	45.50	100	58.75	97.61	73.00
T2	1568	52.73	45.00	54.03	45.50	100	55.38	97.75	72.00
T4	500	53.07	48.63	55.19	49.00	100	54.75	97.40	72.88
T3	356	53.89	50.37	54.20	50.13	99.86	57.38	96.93	72.50
T5	250	56.08	53.38	55.60	51.88	99.93	60.50	96.99	75.88
T6	100	62.16	59.25	63.80	62.50	100	69.38	96.31	74.13

Table 6

Classification Accuracy after Symmetrical Uncertainty

		NaïveBayes		BayesNet		Random Tree		J48 Tree	
bp	Feat Cnt	Train %	Test %	Train %	Test %	Train %	Test %	Train %	Test %
T0	2691	53.21	46.00	54.17	45.38	100	52.63	97.54	72.88
T1	1575	52.80	45.00	54.03	45.50	100	53.38	97.61	72.63
T2	1190	51.23	45.00	54.37	45.75	100	55.25	97.75	71.13
T4	500	51.71	48.00	53.76	47.25	100	59.38	97.06	71.63
T5	250	50.82	49.38	53.42	48.63	100	59.88	96.72	74.38
T3	109	52.66	51.50	53.96	50.38	100	61.25	96.11	76.50
T6	100	53.76	52.00	54.92	50.63	100	71.38	95.70	74.25

Table 7

Classification Accuracy after Pearson's Correlation

		NaïveBayes		BayesNet		Random Tree		J48 Tree	
bp	Feat Cnt	Train %	Test %	Train %	Test %	Train %	Test %	Train %	Test %
T0	2691	53.21	46.00	54.17	45.38	100	52.63	97.54	72.88
T1	2691	53.21	46.00	54.17	45.38	100	52.63	97.54	72.88
T2	2668	53.21	46.13	54.17	45.38	100	50.00	97.47	73.38
T3	1414	52.05	46.13	54.17	45.00	100	51.50	97.54	73.13
T4	500	50.68	47.63	51.98	46.25	100	53.63	96.99	71.25
T5	250	50.41	49.13	53.01	47.88	100	57.00	96.79	73.50
T6	100	52.32	50.75	53.42	49.50	100	57.50	94.88	74.50

Table 8

Classification Accuracy after CFS

		NaïveBayes		BayesNet		Random Tree		J48 Tree	
bp	Feat Cnt	Train %	Test %	Train %	Test %	Train %	Test %	Train %	Test %
T0	2691	53.21	46.00	54.17	45.38	100	52.63	97.54	72.88
-	26	65.16	60.88	74.25	69.50	100	66.88	96.17	75.50

5. Conclusion and Future Plans

Although there are many subtypes of dementia diseases, called neuropsychiatric disorders, the most familiar one is AD. Unfortunately, there is no cure method for its healing effect. In this direction, medical comments or computer-aided studies play a significant role in the analysis of the diseases. In neuroimaging science, it is substantial to establish relationships/rules between attributes and classes.

In this study, which is the first stage of detecting the most valuable features that distinguish dementia diseases, filtering feature selection approaches are utilized with various types of learning algorithms. All the demographic and morphometric characteristics that are extracted via Freesurfer for all individuals in the ADNI1Complete1Yr1.5T dataset are evaluated through the blind search logic. The dataset is preprocessed first, and filtered top-weighted features are listed comparably. After eliminating the features at various thresholds with low scores with different filtering feature selection algorithms, it is observed that the reduced feature sets tend to give more successful results. As a result, rather than handling the entire feature set, the overall evaluation tends to be more accurate when lesser attributes are correctly taken.

Many continuation plans exist in future work. Much more performance can be obtained in line with further detailed analysis to be made through the features specified with the threshold values reported in this article. Moreover, the statistical data among the morphometric features will be eliminated, and only physical measurements of the brain parts in the dataset will be analyzed. Besides, the determination of the relationship between measurement data and statistical data by repeating the tests on the relevant set can be specified as an essential research topic. In order to examine the dementia characteristics in more detail, studies with wrapper feature selection methods through the results obtained by filtering approaches are also in mind. With the realization of the planned studies, it is aimed to identify and list the most beneficial features that distinguish dementia diseases best. It is inevitable to produce more accurate performance prediction models with symbolic literature contributions to be made after determining the most valuable features.

Author Contributions

Both Savaş OKYAY and Nihat ADAR researched scientific publications, performed algorithm-based model design and tests, then completed the writing of the manuscript along with the analysis and interpretation of the findings, and finally reviewed and revised the article.

Conflict of Interest

No conflict of interest was declared by the authors.

Acknowledgment

Data collection and sharing for this project was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904) and DOD ADNI (Department of Defense award number W81XWH-12-2-0012). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: AbbVie, Alzheimer's Association; Alzheimer's Drug Discovery Foundation; Araclon Biotech; BioClinica, Inc.; Biogen; Bristol-Myers Squibb Company; CereSpir, Inc.; Cogstate; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; EuroImmun; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.; Fujirebio; GE Healthcare; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Lumosity; Lundbeck; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; NeuroRx Research; Neurotrack Technologies; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Takeda Pharmaceutical Company; and Transition Therapeutics. The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health (www.fnih.org). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's Therapeutic Research Institute at the University of Southern California. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California.

References

Alam, S., Kwon, G. R., & Initi, A. s. D. N. (2017). Alzheimer disease classification using KPCA, LDA, and multi-kernel learning SVM. *International Journal of Imaging Systems and Technology, 27*(2), 133-143. doi: http://dx.doi.org/10.1002/ima.22217

Aldous, D. (1991). The continuum random tree. I. *The Annals of Probability*, 1-28.

Bhargava, N., Sharma, G., Bhargava, R., & Mathuria, M. (2013). Decision tree analysis on j48 algorithm for data mining. *Proceedings of International Journal of Advanced Research in Computer Science and Software Engineering, 3*(6).

Delgado, J., Moure, J. C., Vives-Gilabert, Y., Delfino, M., Espinosa, A., & Gomez-Anson, B. (2014). Improving the Execution Performance of FreeSurfer. *Neuroinformatics, 12*(3), 413-421. doi: http://dx.doi.org/10.1007/s12021-013-9214-1

Dimitriadis, S. I., Liparas, D., & Initi, A. s. D. N. (2018). How random is the random forest ? Random forest algorithm on the service of structural imaging biomarkers for Alzheimer's disease: from Alzheimer's disease neuroimaging initiative (ADNI) database. *Neural Regeneration Research, 13*(6), 962-970. doi: http://dx.doi.org/10.4103/1673-5374.233433

Dimitriadis, S. I., Liparas, D., Tsolaki, M. N., & Initia, A. D. N. (2018). Random forest feature selection, fusion and ensemble strategy: Combining multiple morphological MRI measures to discriminate among healhy elderly, MCI, cMCI and alzheimer's disease patients: From the alzheimer's disease neuroimaging initiative (ADNI) database. *Journal of Neuroscience Methods, 302*, 14-23. doi: http://dx.doi.org/10.1016/j.jneumeth.2017.12.010

Doan, N. T., Engvig, A., Zaske, K., Persson, K., Lund, M. J., Kaufmann, T., . . . Initi, A. s. D. N. (2017). Distinguishing early and late brain aging from the Alzheimer's disease spectrum: consistent morphological patterns across independent samples. *Neuroimage, 158*, 282-295. doi: http://dx.doi.org/10.1016/j.neuroimage.2017.06.070

Fischl, B. (2012). FreeSurfer. *Neuroimage, 62*(2), 774-781. doi: http://dx.doi.org/10.1016/j.neuroimage.2012.01.021

Hall, M. A., & Smith, L. A. (1998). Feature subset selection: A correlation based filter approach. *Progress in Connectionist-Based Information Systems, Vols 1 and 2*, 855-858.

Lama, R. K., Gwak, J., Park, J. S., & Lee, S. W. (2017). Diagnosis of Alzheimer's Disease Based on Structural MRI Images Using a Regularized Extreme Learning Machine and PCA Features. *Journal of Healthcare Engineering*. doi: http://dx.doi.org/10.1155/2017/5485080

Moradi, E., Pepe, A., Gaser, C., Huttunen, H., Tohka, J., & Initi, A. s. D. N. (2015). Machine learning framework for early MRI-based Alzheimer's conversion prediction in MCI subjects. *Neuroimage, 104*, 398-412. doi: http://dx.doi.org/10.1016/j.neuroimage.2014.10.002

Murphy, K. P. (2001). Active learning of causal Bayes net structure.

Okyay, S., & Adar, N. (2018). *Parallel 3D brain modeling & feature extraction: ADNI dataset case study*. Paper presented at the 2018 14th International Conference on Advanced Trends in Radioelecrtronics, Telecommunications and Computer Engineering (TCSET), Lviv-Slavske, Ukraine. doi: http://dx.doi.org/10.1109/TCSET.2018.8336172

Patil, T., & Sherekar, S. (2013). Performance Analysis of Naive Bayes and J48 Classification Algorithm for Data Classification. *International Journal Of Computer Science And Applications, ISSN: 0974, 1011*.

Ramirez, J., Gorriz, J. M., Ortiz, A., Martinez-Murcia, F. J., Segovia, F., Salas-Gonzalez, D., . . . Initia, A. D. N. (2018). Ensemble of random forests One vs. Rest classifiers for MCI and AD prediction using ANOVA cortical and subcortical feature selection and partial least squares. *Journal of Neuroscience Methods, 302*, 47-57. doi: http://dx.doi.org/10.1016/j.jneumeth.2017.12.005

Reuter, M., Schmansky, N. J., Rosas, H. D., & Fischl, B. (2012). Within-subject template estimation for unbiased longitudinal image analysis. *Neuroimage, 61*(4), 1402-1418. doi: http://dx.doi.org/10.1016/j.neuroimage.2012.02.084

Rish, I. (2001). *An empirical study of the naive Bayes classifier.* Paper presented at the IJCAI 2001 workshop on empirical methods in artificial intelligence.

Sharma, A. K., & Sahni, S. (2011). A comparative study of classification algorithms for spam email data analysis. *International Journal on Computer Science and Engineering, 3*(5), 1890-1895.

Sorensen, L., Igel, C., Pai, A., Balas, I., Anker, C., Lillholm, M., . . . Init, A. s. D. N. (2017). Differential diagnosis of mild cognitive impairment and Alzheimer's disease using structural MRI cortical thickness, hippocampal shape, hippocampal texture, and volumetry. *Neuroimage-Clinical, 13*, 470-482. doi: http://dx.doi.org/10.1016/j.nicl.2016.11.025

Web of ADNI (2017). Access Data and Samples. Retrieved from http://adni.loni.usc.edu/data-samples/access-data/ [Access Date: Jan 9, 2020]

Yao, D. R., Calhoun, V. D., Fu, Z. N., Du, Y. H., & Sui, J. (2018). An ensemble learning system for a 4-way classification of Alzheimer's disease and mild cognitive impairment. *Journal of Neuroscience Methods, 302*, 75-81. doi: http://dx.doi.org/10.1016/j.jneumeth.2018.03.008