PREDICTING ALZHEIMER'S DISEASE BY SEGMENTING AND CLASSIFYING 3D-BRAIN MRI IMAGES USING CLUSTERING TECHNIQUE AND SVM CLASSIFIERS

BY

Sofia Matoug

A thesis submitted in partial fulfillment of the requirements for the degree

of Master of Science (M.Sc.) in Computational Sciences

The Faculty of Graduate Studies

Laurentian University

Sudbury, Ontario, Canada

THESIS DEFENCE COMMITTEE/COMITÉ DE SOUTENANCE DE THÈSE
Laurentian Université/Université Laurentienne
Faculty of Graduate Studies/Faculté des études supérieures

Title of Thesis Titre de la thèse	PREDICTING ALZHEIMER'S DISEASE BY SEGMENTING AND CLASSIFYING 3D-BRAIN MRI IMAGES USING CLUSTERING TECHNIQUE AND SVM CLASSIFIERS
Name of Candidate Nom du candidat	Matoug, Sofia
Degree Diplôme	Master of Science
Department/Program Département/Programme	Computational Sciences
Date of Defence August 31, 2015 Date de la soutenance	

APPROVED/APPROUVÉ

Thesis Examiners/Examinateurs de thèse:

Dr. Amr Abdel-Dayem
(Supervisor/Directeur(trice) de thèse)

Dr. Kalpdrum Passi
(Committee member/Membre du comité)

Dr. Julia Johnson
(Committee member/Membre du comité)

Dr. Sabah Mohammed
(External Examiner/Examinateur externe)

Approved for the Faculty of Graduate Studies
Approuvé pour la Faculté des études supérieures
Dr. David Lesbarrères
Monsieur David Lesbarrères
Acting Dean, Faculty of Graduate Studies
Doyen intérimaire, Faculté des études supérieures

ACCESSIBILITY CLAUSE AND PERMISSION TO USE

Abstract

Alzheimer's disease (AD) is the most common form of dementia affecting seniors age 65 and over. When AD is suspected, the diagnosis is usually confirmed with behavioural assessments and cognitive tests, often followed by a brain scan. Advanced medical imaging and pattern recognition techniques are good tools to create a learning database in the first step and to predict the class label of incoming data in order to assess the development of the disease, i.e., the conversion from prodromal stages (mild cognitive impairment) to Alzheimer's disease.

Advanced medical imaging such as the volumetric MRI can detect changes in the size of brain regions due to the loss of the brain tissues. Measuring regions that atrophy during the progress of Alzheimer's disease can help neurologists in detecting and staging the disease.

In this thesis, we want to diagnose the Alzheimer's disease from MRI images. We segment brain MRI images to extract the brain chambers. Then, features are extracted from the segmented area. Finally, a classifier is trained to differentiate between normal and AD brain tissues.

We discuss an automatic scheme that reads volumetric MRI, extracts the middle slices of the brain region, performs 2-dimensional (volume slices) and volumetric segmentation methods in order to segment gray matter, white matter and cerebrospinal fluid (CSF), generates a feature vector that characterizes this region, creates a database that contains the generated data, and finally classifies the images based on the extracted features. For our results, we have used the MRI data sets from the Alzheimer's disease Neuroimaging Initiative (ADNI) database[1].

We assessed the performance of the classifiers by using results from the clinical tests.

Keywords

"ADNI database, Image processing, Segmentation, Registration, Vector of attributes, classification, Machine learning, Training, Alzheimer's disease".

[1] http://adni.loni.ucla.edu/

Acknowledgments

I would like to express my appreciation and thanks to my advisor Dr. Amr Abdel-Dayem. I would also like to thank my committee members, Dr. Kalpdrum Passi, Dr. Julia Johnson for serving as my committee members. I also want to thank you for your brilliant comments and suggestions, thanks to you.

I would like to express appreciation to my beloved husband Brahim for his support and continues encouragement.

Table of Contents

List of Tables

List of Figures

List of Appendices

Chapter 1

Introduction

1.1 Introduction

Alzheimer's disease (AD) is the most common form of dementia affecting seniors age 65 and over. AD causes nerve cell death and tissue loss throughout the brain, resulting to brain tissue shrinking and larger ventricles (chambers within the brain that contain cerebrospinal fluid). When AD is suspected, the diagnosis is first confirmed with behavioural assessments and cognitive tests and often followed by a brain scan [1].

Advanced medical imaging with computed tomography (CT) or magnetic resonance imaging (MRI), and with single photon emission computed tomography (SPECT) or positron emission tomography (PET) can be used to help exclude other cerebral pathology or subtypes of dementia [1]. Moreover, it may predict conversion from prodromal stages (mild cognitive impairment) to Alzheimer's disease [1], which is the most critical brain disease for the senior population.

Medical image processing and machine learning tools can help neurologists in assessing whether a subject is developing the Alzheimer disease. A machine learning system has been developed in order to extract meaningful information from the ADNI database, where the ventricle chambers are extracted using a segmentation method based on the statistical and geometrical features of the region of interest. We performed an analysis to see if this region corresponds to a good marker.

Pattern recognition techniques are good tools to create a learning database in the first step and to predict the class label of incoming data in order to assess the development of the disease by

detecting changes in the size of brain regions due to the loss of the brain tissues. Measuring regions that atrophy during the progress of Alzheimer's disease can help neurologists in detecting and staging the disease.

We used the MRI data sets from the Alzheimer's disease Neuroimaging Initiative (ADNI) database[2]. ADNI data includes Alzheimer's disease patients, mild cognitive impairment subjects and elderly controls. ADNI database aims to assist the researchers in the progression of Alzheimer's disease by collecting, validating and using predictors for the disease such as MRI and PET images, cognitive tests and Cerebrospinal fluid (CSF).

The present thesis describes the whole process of pattern recognition where the following steps are performed: 1) accessing ADNI database, 2) describing the medical data, 3) reading the volumetric MRI, 4) extracting the middle slices of the brain region, 5) performing segmentation methods in order to detect the region of brain's ventricle, 6) generating a vector of attributes that characterizes this region, 7) creating a database that contains the generated data, 8) performing clustering to get the class labels and finally 9) performing some classification methods based on the clustering results.

1.2 Thesis Outline

In chapter 2, we describe the different image processing techniques including image pre-processing, image segmentation, feature extraction, and classification techniques.

In chapter 3, we include a literature review regarding the most used methods that describes and intends to diagnose Alzheimer's disease based on ADNI database and some other medical MRI scans.

In chapter 4, we describe briefly the organizational schema of the system implementation. Then, we describe the tools used to access the medical ADNI database, give an overview of the

[2] http://adni.loni.ucla.edu/

type of medical data files and discuss the problems encountered during the first step of accessing data.

In chapter 5, we include the segmentation methods used to extract the regions of interest and show the way they were used during the implementation process.

In chapter 6, we define the vectors of attributes and show their comprehensive statistical analysis. Then, we introduce the classification methods and show the results of the used database in addition to classical databases results. Finally, we assess the different classification techniques

Finally, in the conclusion, we summarize the different steps, assess the overall work and include recommendations and suggestions for future work.

Chapter 2
Pattern Recognition Techniques for Image Processing

2.1 Introduction

One of the ultimate goals of classification is to produce meaningful patterns from raw data, classify them into different groups based on their characteristics and predict new patterns based on previous knowledge. The purpose of this Chapter is to present some of the classical methods used in machine learning and pattern recognition and introduce some of the newest concepts in this domain. Since the different methods depend strongly on the application, most of the highlighted examples are taken from image processing domain.

Pattern recognition is the scientific discipline whose goal is the classification of objects into a number of categories or classes. Pattern recognition and machine learning were used in various applications such as speech recognition, face recognition, text analysis, image processing including medical images, space images, security domains, etc. All these domains share the same goal which is the extraction of patterns based on certain conditions and the separation of one class from the others [2] [3].

Different techniques based on classification rules and statistics where developed starting from linear and quadratic discriminates [4] (e.g. Fisher's linear discriminate analysis [5], principal component analysis (PCA) [6] and Karhunen-Loeve transform applied for the characterization of human faces [2]), to clustering techniques [7] (e.g. k-nearest neighbour classifiers [8], decision trees [9], etc.). To cope with the lack of meaningful information needed for the previous classifiers, new techniques were developed such as template matching [10] [11], Neural Networks [12], and more recently Support Vector Machine (SVM) [13] [14].

The current chapter describes the process of pattern recognition and some techniques related to each step of pattern recognition. Section **2.2** gives a schematic overview of the pattern recognition process. Section **2.3** brought out segmentation techniques named: thresholding, edge-based segmentations, and region-based segmentations, watershed and wavelet transform. Section **2.4** describes and discusses methods related to feature extraction and feature selection. Finally, Section **2.5** reveals some of the most "classical" pattern recognition methods such as classification methods and introduced some other new algorithms.

2.2 Pattern Recognition Methodology

Pattern recognition is a set of processes that aim to extract meaningful information or patterns from a set of data. The organizational chart in **Figure 2-1** shows supervised pattern recognition steps using classification techniques that predict categorical labels.

The first step of pattern recognition is the problem statement which is gathering the data and the background knowledge behind the application domain, making hypotheses and establishing which type of information is needed to be extracted from the data. Usually gathering the data is followed by a preprocessing step mostly to clean it and standardize it [15]. Once the data is well defined, the next step is the extraction and representation of the data features in the form of vectors followed by the creation of models of the classes through machine learning. Depending on the type of label output (categorical labels or real-valued labels), on whether learning is supervised or unsupervised, and on whether the pattern algorithm is statistical or non-statistical, the expert has to choose one of the pattern recognition techniques (algorithms) such as classification, clustering, regression, etc. Finally, we proceed to the performance evaluation of the pattern recognition algorithm results using evaluation metrics such as bootstrapping and cross-validation.

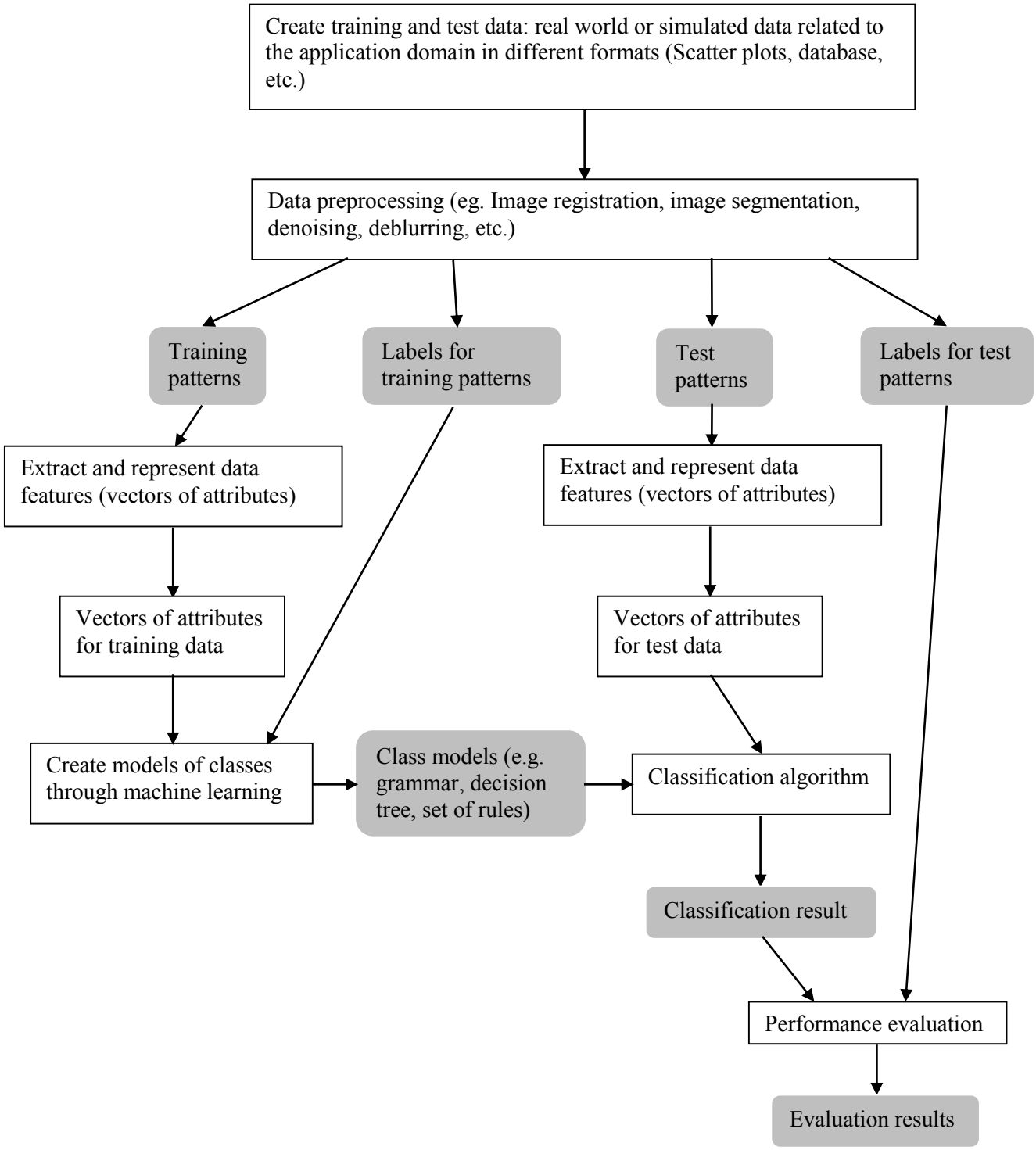


Figure 2-1 Pattern Recognition Process

2.3 Segmentation techniques

In computer vision and machine learning systems, image segmentation is intended to partition images into well-defined regions, where each region is a set of pixels that share the same range of intensities, the same texture or the same neighborhood. The purpose of segmenting images is to remove unwanted information in order to locate meaningful objects from the processed images. Many segmentation algorithms have been developed through the years and only some of them are highlighted in this chapter [16].

2.3.1 Thresholding

When images contain different contrasting objects, thresholding provides effective means for obtaining segmented images. Thresholding techniques are based on partitioning the intensities using global or local threshold calculations techniques such as Otsu [17] and Niblack methods [18], where each threshold classifies the voxels (or pixels) into different modes using a clustering criterion.

2.3.1.1 The Otsu Method

The Otsu method [17] is a clustering technique that tends to produce two tight clusters by minimizing their overlap (misclassified pixels). The threshold is adjusted dynamically by increasing the spread of one cluster and decreasing the spread of the other one. The goal then is to select the threshold that minimizes the combined spread. We define the within-class variance as the weighted sum of the variances of each cluster:

$$\sigma^2_{\text{within}} = \mathrm{n_B(T)}\sigma^2_{\mathrm{B}}(\mathrm{T}) + \mathrm{n_O(T)}\sigma^2_{\mathrm{O}}(\mathrm{T}) \quad (2.1)$$

$$\sigma^2_{\text{between}} = \mathrm{n_B(T)n_O(T)}\left(\mu^2_{\mathrm{B}}(\mathrm{T}) + \mu^2_{\mathrm{O}}(\mathrm{T})\right) \quad (2.2)$$

where:

$n_B(T) = \sum_{i=0}^{T-1} p(i)$: the number of pixels in the first cluster

$n_O(T) = \sum_{i=T}^{N-1} p(i)$: the number of pixels in the second cluster

$\sigma_B^2(T)$: the variance of the pixels in the background (below threshold)

$\sigma_O^2(T)$: the variance of the pixels in the foreground (above threshold)

$\mu_B(T)$: the mean of all pixels less than the threshold

$\mu_O(T)$: the mean of all pixels greater than the threshold

[0, N-1]: is the range of intensity levels.

Otsu algorithm

The optimal threshold is the one that maximizes the between-class variance (or, conversely, minimizes the within-class variance).

1. Calculate the histogram *h*.
2. Separate the pixels into two clusters (background and foreground) according to the threshold.
3. Find the mean of each cluster.

for T=1:255

4. Calculte the new background's number of pixels: nb = nb + h(T)
5. Calculte the new foreground's number of pixels: no = no - h(T)
6. Calculte the new background's mean: ub = (ub*nb + n*T) / nb
7. Calculte the new foreground's mean: uo = (uo*no - n*T) / nb
8. Calculate the between-class variance: sbetween(T) = nb*no*(ub - uo)^2

end

9. Select *T* that corresponds to the maximum between-class varianc

2.3.1.2 Niblack method

Niblack's algorithm [18] calculates a local threshold *T* for each pixel. The threshold T is computed by using the mean μ and standard deviation σ of all the pixels in the pixel neighborhood, and is denoted as: $T = \mu + k*\sigma$, where the parameter *k* is a constant, which determines how much of the total object is extracted, and is usually chosen between 0 and 1. The value of *k* and the size of the neighborhood influence the result of thresholding.

2.3.2 Edge detection

Other segmentation methods are based on edge detection techniques such as Canny [19], active contours or snakes using the technique of matching a deformable model to an image by means of energy minimization [20] [21].

2.3.2.1 Canny edge detection

Canny edge detection algorithm [19] aims to the following optimal properties:

- Good detection: the algorithm should detect as many real edges in the image as possible.
- Good localization: edges marked should be as close as possible to edges in the real image.
- Minimal response: a given edge in the image should only be marked once, and where possible, image noise should not create false edges.

Canny's algorithm is based on finding an optimal function as the first derivative of a Gaussian, originally described by the sum of four exponential terms. The effectiveness and cost of the algorithm depends on the size of the Gaussian filter and the hysteresis thresholds.

2.3.2.2 Active contours

The active contour [20] [21] is also sometimes called snake algorithm. Given an approximation of the boundary of an object in an image, an active contour model deforms the initial boundary to lock onto characteristic features within the region of interest. The contour is deformed iteratively until it matches the boundary of the region of interest by looking for the minimum of energy of a

given problem. The energy function is a weighted combination of internal and external forces depending on the shape of the snake and location within the image.

The integral energy function to be minimized is given by:

$$E^*_{snake} = \int_0^1 E_{snake}(v(s))ds \quad 2.3$$

$$= \int_0^1 [E_{int}(v(s)) + E_{image}(v(s)) + E_{con}(v(s))]ds \quad 2.4$$

where $E_{int} = \alpha(s)\left|\frac{dv}{ds}\right|^2 + \beta(s)\left|\frac{d^2v}{ds^2}\right|^2$ is the internal spline energy,

$\alpha(s)$ and $\beta(s)$ are the elasticity and stiffness of the snake respectively,

E_{image} is derived from the image data over which the snake lies and it is modeled as a weighted combination of different function, and

E_{con} comes from external constraints that force the snake toward or away from particular features.

The effectiveness of the active contour algorithm depends on the initial choice of the approximate shape and starting position. A priori information is then used to move the snake toward an optimal solution.

2.3.3 Region-based segmentation

Region-based segmentation uses different techniques such as seeded region-growing [22], split-and-merge [23], watershed [24] and Wavelet-based segmentation [25] which is based on mathematical concepts such as quadrature mirror filtering, sub-band coding, and pyramidal image processing.

2.3.3.1 Region-growing segmentation

Region-growing segmentation [22] starts with initial seed points chosen from the target region or without a priori knowledge, taken from the picks of the histogram. It checks the neighborhood

pixels and adds them to the region if they are similar to the chosen seeds using a similarity criteria (homogeneity predicates) based on a vector of characteristics (attributes) in the image such as the average, standard deviation, texture, etc.

2.3.3.2 Split-and-merge segmentation

Split-and-merge segmentation [23] consists of two different parts. The split process keeps dividing the image into smaller regions that do not respect a criterion of similarity. In the merge process, neighboring regions, resulting from the split process that respects a similarity criterion, using a vector of predicates, are merged into bigger regions.

2.3.4 Watershed segmentation

Meyer et al. [24], Beucher et al. [26] and most recently Gonzalez et al. [27] presented mathematical morphology methods based on two main tools: the watershed transform (WT) that segments an image into regions of interest (ROI), also called objects, and the homotopy modification that solves the over-segmentation problem by initializing markers of the images' ROI. S. Beucher compared gray level images to topographic reliefs, where the intensity of a pixel corresponds to the altitude. In watershed by flooding, a water source is placed into each regional minimum and barriers or dams are built where different water flood sources are meeting. The resulting set of dams is called watershed by flooding.

The watershed by flooding algorithm works on a gray scale image and is performed on the gradient image. The images must be pre-processed and the regions that satisfy a similarity criterion must be merged.

1. Choose a set of markers, with different labels (pixels where the flooding shall start).

2. The neighboring pixels of each marked area are inserted into a priority queue with a priority level corresponding to the gray level of the pixel.

3. The pixel with the highest priority level is extracted from the priority queue. If the neighbors of the extracted pixel that have already been labeled all have the same label, then

the pixel is labeled with their label. All non-marked neighbors that are not yet in the priority queue are put into the priority queue.

4. Redo step 3 until the priority queue is empty. The non-labeled pixels are the watershed lines.

2.3.5 Wavelet transform

In order to analyze physical situations, scientists, theoreticians and engineers represent data in a certain way that help them understand the meaning and the behaviour of the data. Many of them represent the data as a function of time because most of the signals in practice are time-domain, which is called time-domain representation. In another hand, in many cases, the most distinguished information is hidden in the frequency content of the signal.

Example: The following CT image in **Figure 2-2** is corrupted by a repeated noise (like a pattern) that is impossible to get removed by using the time-domain representation of the image (2-D signal) and time-domain filtering because the noise signal cannot be represented. In the opposite, from the frequency spectrum of the image, the noise signal is well represented by 3 pairs of impulses with horizontal, vertical and diagonal directions respectively and the filtering is more accurate since the information of the noise is well defined.

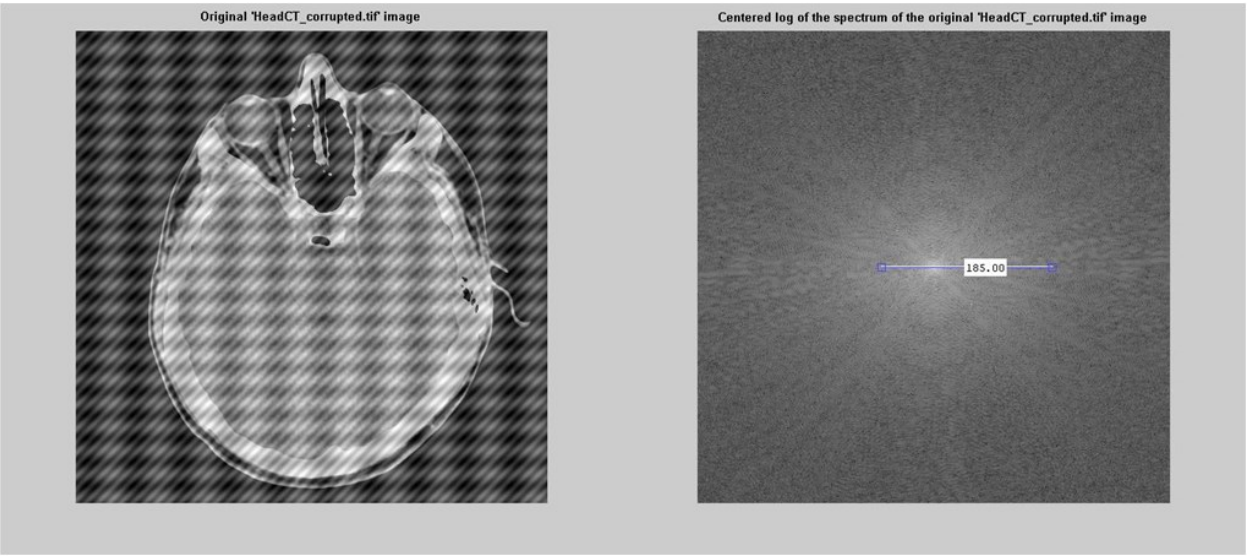


Figure 2-2 Original image "HeadCT_corrupted.tif" image (courtesy of **[119]**) and its centred spectrum log

The frequency spectrum of a signal shows what frequencies (or frequency components) exist in the signal and by general definition, the frequency shows the change in rate of a mathematical or physical variable, i.e. In the case of a variable that changes fast, we say that it has high frequency. A variable that changes smoothly, has a low frequency. If this variable does not change at all, then we say it has zero frequency, or no frequency [28].

However, Wavelet transform represents a signal in the time and frequency domain at the same time. Wavelets are mathematical functions that represent data (or signals) by dividing it into different frequency components, were each frequency component has a different scale, and then analyzing each frequency component with an adequate resolution. Unlike the Fourier transform, they can access the time-domain and frequency representations of the data in the same time and therefore, can analyze physical situations where the signal contains discontinuities and sharp spikes. [28]

In image processing, the Wavelets transforms are used to denoise the images, perform segmentation and compression of the signals. In the last decade, wavelet transform has been recognized as a powerful tool in a wide range of applications, including image/video processing, numerical analysis, and telecommunication. The advantage of wavelet over existing transforms such as Discrete Fourier Transform (DFT) and Discrete Cosine Transform (DCT) is that wavelet performs a multiresolution analysis of a signal with localization in both time and frequency [25]. In addition to this, functions with discontinuities and functions with sharp spikes require fewer wavelet basis vectors in the wavelet domain than sine-cosine basis vectors to achieve a comparable approximation. Wavelet operates by convolving the target function with wavelet kernels to obtain wavelet coefficients representing the contributions in the function at different scales and orientations. Wavelet or multiresolution theory can be used alongside segmentation approaches, creating new systems which can provide a segmentation of superior quality to those segmentation approaches computed exclusively within the spatial domain [29].

Discrete wavelet transform (DWT) can be implemented as a set of filter banks, comprising a high-pass and low-pass filters. In standard wavelet decomposition, the output from the low-pass filter can then be decomposed further, with the process continuing recursively in this manner.

2.4 Feature selection and feature extraction

Features should be easily computed, robust, compact, accurate, insensitive to various distortions and variations, and rotationally invariant.

2.4.1 Feature extraction

Feature extraction is a special form of dimensionality reduction that depends closely on the type of data and the application domain. For example, in image processing, the image contains meaningful objects characterized by their shape, textures, intensities, etc. Some of these attributes are summarized by Zhang et al. [30] where the authors classify the shape based on its contour attributes (e.g. chain-code, perimeter, compactness, etc.) and region attributes (e.g. area, Euler number, geometric and statistical moments, convex hull, etc.). Both types of attributes can be defined as structural or global. From the point of view of the authors, the structural approaches are too complex to implement compared to global approaches. However, they are useful in applications where partial matching is needed. Also, even though more popular, the contour shape descriptors are more sensitive to noise and variations than the region shape since they carry a smaller amount of information. Finally, for general shape applications, methods based on complex moments and spectral transforms are the best choices since they satisfy the six principles set by MPEG-7: good retrieval, accuracy, compact features, general application, low computation complexity, robust retrieval performance and hierarchical coarse to fine representation.

2.4.2 Feature selection

Feature extraction is usually followed by the selection of the optimal feature subset that reduces the cost of pattern recognition and provides better classification accuracy by reducing the number of features that need to be collected [31]. Some of the feature selection algorithms perform heuristic search through the whole space of attributes using methods such as hill climbing. Other algorithms divide the space of attributes into subspaces to have smaller combinations.

Jain, et al. [32] presented a review of feature selection by demonstrating its value in combining features from different data models. They presented potential difficulties of performing feature

selection for small size sample data, due to the curse of dimensionality. They also reproduced the results of Pudil, et al. [33] who demonstrated the quality of the floating search methods in case of nonmonotonicity of the feature selection criterion or for computational reasons. They used the Mahalanobis distance $D_M(x) = \sqrt{(x-\mu)^T S^{-1}(x-\mu)}$ (μ and S are respectively the mean and the covariance matrix of the x vector) between two classes as a criterion function to assess the "goodness" of a feature subset and evaluated fifteen feature selection algorithms such as max-min, SFS and SBS. They finally claimed that using feature selection for classification of known distributions and comparing the selected subsets with the true optimal subset resulted in a well quantified quality of the selected subset.

There are three types of feature selection methods: filter, wrapper and embedded approaches [34] [35]. Filters are the most widely used and are performed at the first stage of classification by selecting the best features according to some prior knowledge [36] [37]. Wrappers do not depend on the type of classifiers [38] [39]. An example of a wrapper method for nonlinear SVMs can be found in [39], where instead of minimising the classification error, the features are selected to minimise a generalisation error bound. Finally, embedded approaches simultaneously determine features and classifier during the training process.

2.5 Pattern recognition algorithms

Starting from the acquisition of data and its preprocessing to the extraction and selection of an optimal vector of attributes, we need to perform the most important step of pattern recognition which is the pattern recognition algorithm in form of classifiers, clustering, regression, etc.

2.5.1 Classification algorithms

Classification algorithms are supervised methods which means that the data is already labelled and they perform prediction of the classes by assigning a categorical label to the current class. In **Figure 2-1**, the classification process is performed in two steps, first we use sample data training to get the training attributes followed by the creation of class models through machine learning. The whole step is called learning. Simultaneously, a sample data test is being used to get the test vector of attributes. At this point, both data are being transported to a classifier algorithm that

should classify the test data based on the learning process. There are several classification techniques such as:

Maximum entropy classifier is a classic Generalized Iterative Scaling algorithm that allows diverse sources of data to be combined where for each source of data, we determine a set of constraints on the model and using an algorithm such as Generalized Iterative Scaling (GIS), a model can be found that satisfies all of the constraints, while being as smooth as possible [40].

Naive Bayes classifier is a very popular probabilistic approach classifier based on the Bayesian theorem which is suitable for high dimensional input data. Even though its probability estimation is poor, Zhang, et al. [41] compared naive Bayes with C4.4 algorithm for ranking, and some extensions of naive Bayes such as the selective Bayesian classifier (SBC) and tree-augmented naive Bayes (TAN) and found out that naive Bayes performs significantly better than C4.4 and comparably with TAN.

Decision trees, decision lists are classification methods where the input is the vector of attributes being classified and the output is the class label of the given tuple, where each node consists of a feature, and after each iteration, we go deeper through the tree till we get to the leaf that corresponds to the output label. One of the issues of this kind of classifier is to choose the right type since there are several types such as the ID3 and C4.5 [9].

Support vector machines is a classification method, originally invented by Vladimir Vapnik, that maps an n-dimensional input vector into a high dimensional (possibly infinite dimensional) feature space. This technique offers a possibility to train generalizable, nonlinear classifiers in high dimensional spaces using a small training set. However SVMs generalization error might get important due to the margin with which it separates the data [36] [39] [42].

Kernel estimation and K-nearest-neighbor (KNN) algorithms are statistical methods (a uniform kernel function produces the KNN technique) that have been applied to statistical classification by computing the PDFs of each class separately, using different bandwidth parameters, and then comparing them [43] [44].

Neural networks is a multi-level perceptron where the term ’Neural network’ has its origins in attempts to find mathematical representations of information processing in biological systems.

Bishop defines the classical framework of a Neural network system by considering the functional form of the network model, including the specific parameterization of the basis functions, and then discussing the problem of determining the network parameters within a maximum likelihood framework, which involves the solution of a nonlinear optimization problem. This requires the evaluation of derivatives of the log likelihood function with respect to the network parameters which can be obtained efficiently using the technique of error backpropagation. [12]

2.5.2 Clustering algorithms

Clustering algorithms are unsupervised algorithms that aim to create clusters from raw unlabelled data and to predict categorical labels. They are usually used in the first process of classification for data training in order to get the initial set of class models. These techniques are usually easily programmed but they present several issues such as:

-The nature of the data and the nature of the desired cluster.

-The kind of required input and tools.

-The size of the data set.

-The choice of the initial set of clusters. [45] [46]

Different clustering techniques have been established such as Categorical mixture models, K-means clustering, Hierarchical clustering which is agglomerative or divisive and Kernel principal component analysis (Kernel PCA) [43]

2.5.3 Other pattern recognition algorithms

In addition to the previous classical methods, other recent techniques have been developed such as the Regression algorithms which aim to predict real-valued labels. Some of the regression algorithms are supervised such as Linear regression and extensions, Neural networks and Gaussian process regression (kriging) and others are unsupervised such as Principal components analysis (PCA) and Independent component analysis (ICA). Categorical sequence labeling algorithms predict sequences of categorical labels and similar to the regression algorithms, they include

supervised and unsupervised techniques such as Hidden Markov models (HMMs), Maximum entropy Markov models (MEMMs) and Conditional random fields (CRFs). Real-valued sequence labeling algorithms predict sequences of real-valued labels such as Kalman filters and Particle filters. Parsing algorithms predict tree structured labels such as Probabilistic context free grammars (PCFGs). General algorithms predict arbitrarily-structured labels Bayesian networks such as Markov random fields. Ensemble learning algorithms are supervised meta-algorithms for combining multiple learning algorithms such as Bootstrap aggregating ("bagging"), Boosting, Ensemble averaging and Hierarchical mixture of experts [12] [36] [47].

Chapter 3

Literature Review

3.1 Introduction

Alzheimer's disease is manifested by progressive brain cell decay, the reason cells decay is still generally unknown. Research on new methods for earlier diagnosis is one of the most active areas in Alzheimer's scientific research domains that aim to generate future treatments that could target the disease in its earliest stages, before irreversible brain damage or mental decline has occurred. Different diagnosis techniques have been developed such as Biomarkers for earlier detection such as brain imaging/neuroimaging, cerebrospinal fluid (CSF) proteins, proteins in blood, Genetic risk profiling and mild cognitive impairment [48].

Magnetic resonance imaging (MRI) is a radiation free medical imaging technique that uses a magnetic field and radio waves to visualize detailed images of the internal structures (soft tissue) of the body producing cross-sectional gray level images of the body [49]. These images can be reconstructed into three-dimensional (3-D) images and processed using image processing techniques to denoise the images and extract meaningful information that might help the clinical diagnostic.

3.2 Alzheimer's Disease Neuroimaging Initiative data collection and MRI core Analysis

The collection of the Alzheimer's disease Neuroimaging Initiative (ADNI) database images was created under the LONI Image Data Archive (IDA) and has the objective of developing biomarkers to track both the progression of Alzheimer's disease and changes in the underlying pathology [50].

The IDA has developed many neuroimaging research projects across North America and Europe and accommodates MRI, PET, MRA, DTI and other medical imaging techniques.

3.3 Amyloid-imaging Positron Emission Tomography (PET) and Pittsburgh compound-B (PiB)

In the early nineties, The Consortium to Establish a Registry for Alzheimer's Disease (CERAD) has developed a standardized neuropathology protocol for the postmortem assessment of dementia and control subjects that provides common language of Alzheimer's disease and establishes a better diagnostic criteria, and resulted to a better interpretation of early subclinical changes of AD and normal aging. [51]

From that point, more researches were conducted establishing that the Alzheimer's disease was due to the presence of beta-amyloid plaques and neurofibrillary tangles.

In order to follow the progress of these proteins using medical imaging techniques, William E. Klunk and Chester A. Mathis, from the University of Pittsburgh, discovered a class of benzothiazoles (C_7H_5NS), heterocyclic compound derived from thioflavin T (Basic Yellow 1 or CI 49005). This biophysical dye included some compounds, used as an agent in positron emission tomography imaging. The first trials of the amyloid-imaging positron emission tomography (PET), using this new agent (tracer), were conducted in human research subjects in partnership with Uppsala University (Sweden) which named this compound Pittsburgh compound-B (PiB). In their study, mild AD patients expressed noticed retention of PIB in areas of frontal, parietal, temporal, occipital cortex and the striatum cortex where we assume to find large amounts of amyloid deposits in AD. Also, PIB retention was similar in AD patients and controls in unaffected areas (such as subcortical white matter). In the other hand, young people and older healthy control subjects showed a similar low PIB retention in cortical areas. [52]

Later on, they developed a quantitative imaging method for the measurement of amyloid deposition in humans (Kinetic modeling of amyloid binding) and included subjects with mild cognitive impairment (MCI). [53]

However they needed much more data to validate their results.

From there, Schroeter et al. [54] carried a systematic and quantitative meta-analysis (anatomical likelihood estimates) to identify patterns among study results, specifically neural correlates of Alzheimer's disease (AD) and early symptoms stage. Their results were based on 1351 patients and 1097 healthy control subjects with either atrophy or decreases in glucose utilization. The meta-analysis revealed that early AD affects the structure of (trans-)entorhinal and hippocampal regions, and the functionality of the inferior parietal lobules and precuneus. This could isolate predictive markers for future diagnostic systems.

3.4 Image Segmentation and processing techniques of ADNI data

One of the first brain tissue segmentations studies was conducted by Tina Kapur, in the mid-nineties, which presented a method for segmentation from magnetic resonance images using a parallel implementation of three existing computer vision techniques: expectation/maximization segmentation, binary mathematical morphology, and active contour models. [55]

In the same way, a more accurate technique was developed by W. M. Wells et al. [56] based on adaptive segmentation of MRI data in contrast to the intensity based techniques. This method used knowledge of tissue intensity properties and intensity inhomogeneities in addition to the expectation-maximization (EM) algorithm and carried the results of more than 1000 brain scans.

Held et al. [57] developed 3-D segmentation technique that classifies brain MR images into gray and white matters, cerebrospinal fluid (CSF), scalp-bone and background. They used Markov random fields (MRF's) by extracting three features related to the MR images, i.e., nonparametric distributions of tissue intensities, neighborhood correlations, and signal inhomogeneities.

Many other segmentation methods were applied in MR images afterwards. In 2000, an extensive survey of those methods was made by Pham et al. [58]. Those methods include:

- thresholding or multithresholding (based on the intensity values and the image histograms),
- region growing (based on intensity values and the image contours),
- region classification methods (supervised methods based on pattern recognition techniques such as the k-nearest neighbors, maximum-likelihood or Bayes classifier that use training data),
- clustering (similar to the classification techniques without the training data, including K-means, ISODATA algorithm, Fuzzy C-Mean algorithm, and the expectation-maximization EM algorithm),
- Markov Random Field Models (or MRF which is a statistical model that shows the spatial correlations between close pixels. MRF is combined with clustering algorithms to provide proper segmentation),
- artificial Neural Networks (or ANNs which are parallel networks of nodes that simulate biological learning)
- and other approaches including model-fitting, watershed algorithms, atlas guided approaches and deformable models.

A more recent review regarding the brain MRI image segmentation methods was published in 2010. Balafar et al. [59] added newest segmentation methods including fuzzy clustering algorithm (FCM), Gauss mixture vector, learning vector quantization (LVQ) that is a supervised competitive learning, self-organizing maps (SOM) which is an unsupervised clustering network, watersheds (gradient-based segmentation technique), region growing, active control model, double region based active control, multi region based active control, atlas-based segmentation and Markov random field (MRF).

Going back in time, Zhang et al. [60] suggested a HMRF-EM framework segmentation of brain MR images using a hidden Markov random field (HMRF) model and the expectation-maximization EM algorithm. The HMRF model is a random process produced by a MRF which can be modeled by estimating the observations. They chose the EM algorithm to match the HMRF model.

In 2002, Fischl et al. [61], developed an Automated Labeling technique, in addition to a registration procedure, that appoints a label value, from a 37 labels' training dataset, to each voxel of the neuroanatomical Structures in the Human Brain. The labels include left and right caudate, putamen, pallidum, thalamus, lateral ventricles, hippocampus, and amygdala. According to the authors, the results were accurate when they applied their procedure to detect volumetric changes in mild Alzheimer's disease.

Van Leemput et al. [62] demonstrated an enhanced statistical framework for partial volume segmentation (PV) using parametric statistical image model as a spatial prior knowledge and an expectation-maximization algorithm that estimates the model's parameters and performs a PV classification at the same time.

To overcome the disadvantages of using the watershed transform when segmenting MR images into gray matter/white matter, Grau et al. [63] used an enhanced version of the transform, by adding prior information and atlas registration.

Other researchers tried to automatically segment the brain MR images into more specific regions, e.g., cerebrospinal fluid (CSF), gray matter (GM), white matter (WM) and white matter lesions (WMAL). De Boer et al. [64] [65] used a trained k-nearest neighbor classifier with an extra step for the segmentation of white matter lesions.

In the same manner, Tu et al. [66] created a hybrid discriminative/generative classifier model. The learning process of their classifier used probabilistic boosting tree (PBT) framework and a high dimensional vector of attributes with different scales in order to extract different anatomical structures of 3D MRI volumes. The resulting information is introduced within a hybrid model and an energy function is minimized in order to perform the final segmentation process.

For the purpose of assisting the diagnosis of AD, Colliot et al. [67] used NINCDS-ADRDA criteria [68] for patients with AD and Petersen et al.'s criteria [69] for patients with mild cognitive impairment (MCI). Their purpose was to extract the hippocampus and the amygdale structures using competitive region-growing. Their algorithm started from known landmarks (positions) as a prior knowledge.

Zhang et al. [70] developed a new hybrid active contour model using level-set method whose energy function is not sensitive to image derivatives since it relied on both the object's contour and region information.

Concerning the work of Morra et al. [71], an auto context model (ACM) was created; to segment the hippocampus automatically in 3D T1-weighted MRI scans of subjects from the ADNI database. Their algorithm used 21 hand-labeled segmentations to learn a classification rule that classifies a hippocampus region from a non-hippocampus one using an AdaBoost method and a large vector of attributes (image intensity, position, image curvatures, image gradients, tissue classification maps of gray/white matter and CSF, and mean, standard deviation, and Haar filters of size $1 \times 1 \times 1$ to $7 \times 7 \times 7$). They employed the Bayesian posterior distribution of the labeling to recalculate the new system's attributes. Finally, they validated their algorithm by comparing their results with hand-labeled segmentations.

Following Adaboost algorithm, another popular classifier was applied to segment T1-weighted brain MRIs in order to extract the hippocampus region, i.e. the Support Vector Machine (SVM) as in Morra et al.'s work [72]. The authors compared the hierarchical AdaBoost, SVM with manual feature selection and hierarchical SVM with automated feature selection (Ada-SVM). They validated their results with the FreeSurfer brain segmentation package [73].

In the same manner, David W. Shattuck et al. [74] validated their brain segmentation methods by implementing a web-based test environment [75] using many datasets and a number of metrics to evaluate the segmentation's accuracy and the performance of skull-stripping (removal of extra-meningial tissues from the MRI volume) in T1-weighted MRI. According to the authors, their web-test framework had been satisfactory on 3 popular algorithms named: the Brain Extraction Tool [76], the Hybrid Watershed Algorithm [77], and the Brain Surface Extractor [78].

The segmentation based on edge detection was also used, e.g. Huang et al. [79] apply a geodesic active contour using the image edge geometry and the voxel statistical homogeneity in the purpose of extracting complex anatomical structures.

Since the subcortical grey matter structures (located in the deep brain region) are low in contrast, which delimitates the segmentation results, Helms et al. [80] proposed a semi-quantitative magnetization transfer (MT) imaging protocol that overcomes limitations in T1-weighted (T1w) magnetic resonance images.

Other authors were more inclined in using 3D segmentation in spite of the long computation problem. AlZu'bi et al. [81] suggested Multiresolution analysis segmentation using Hidden Markov Models (HMMs) and extracted the vector of attributes with the assistance of 3D wavelet and ridgelet.

To optimize the accuracy and speed of segmentation, Lötjönen et al. [82] created an optimised pipeline for multi-atlas brain MRI segmentation using different similarity measures. Additionally, they combined multi-atlas segmentation and intensity modelling through expectation maximisation (EM) and optimisation via graph cuts. For their results, they used two databases: IBSR data [83] and ADNI data [50].

Even though the segmentation of MR human brain images with multiple atlases was more successful, the method was less effective when it comes to the ventricular enlargement that is not caught by the atlas database. Heckemann et al. [84] added tissue classification information into the image registration and resumed their work into MAPER, multi-atlas propagation with enhanced registration [85]. They applied their algorithm on the subjects from the Oxford Project to Investigate Memory and Ageing (OPTIMA) [86] and the Alzheimer's Disease Neuroimaging Initiative (ADNI) [50].

As the MRIs of the brain present an intensity non-uniformity (INU) phenomenon which affects the segmentation results, Rivest-Hénault et al. [87] presented a new method that uses local linear region representative and embedded region models. They tested their method on the Internet Brain Segmentation Repository (IBSR) database [83].

3.5 Analysis and further classification techniques of ADNI data

The classification techniques were widely used to classify the MRIs of the human brain into regions of interest (ROIs) with the sole purpose of dividing each image into anatomical regions. They also have been used to create vectors of attributes of geometrical and statistical shapes that are embedded into a machine learning process, and associated with rules that are linked to the anatomical structures of the brain. Those rules should determine the corresponding brain's structure of the shape and indicate a possible health problem related to the shape, e.g. atrophy of the hippocampus due to an advanced AD stage.

Thus, Van Leemput et al. [88] described a model-based tissue classification of MRIs of the brain. According to the authors, starting from a digital brain atlas of prior expectations, their algorithm could segment multi-spectral MRIs, correct signal in-homogeneities, and add MRF's contextual information.

In order to estimate any modification of the size or the shape of the brain, a fully automated method of longitudinal (temporal change) analysis, SIENA [89] has been developed. Smith et al. [90] added improvements to the SIENA package concerning the cross-sectional (single time point) analysis. The package showed the extracted brain, executed registration and tissue segmentation, and estimated the atrophy of the brain.

Also, in order to get a robust brain MRI tissue classification, Cocosco et al. [91] created a pruning method that reduces incorrectly labeled samples in the training set (generated from prior tissue probability maps) using a minimum spanning tree graph-theoretic approach. The resulting set is associated with a supervised kNN classifier.

Since the hippocampus was one of the first structures affected by the AD, Chupin et al. [92] proposed classification-based segmentation of the brain into two main regions: hippocampus (Hc) and the amygdala (Am). They used region deformation based on stable local anatomical patterns and probabilistic prior information. They evaluated their segmentation method in patients with AD, MCI, and elderly controls from the ADNI database.

The ultimate purpose of classification is to make a diagnosis of the brains' MRIs and make a decision regarding the abnormality of the MRIs. Chaplot et al. [93] used the neural networks as a machine learning system with the wavelets as input and the support vector machine as the classification method. According to the authors, their classifier could classify the brain into normal or abnormal without specification of the abnormality. Another work was pursued by Klöppel et al. [94] who also used the support vector machines classifier in both learning process and classification process to separate patients with AD from healthy aging controls and to determine other forms of dementia.

From that point, researchers are more eager to detect the Alzheimer's disease in its first stage, which could prepare the patients and give more room to find possible cures. According to Polikar et al. [95], even though wavelets and neural networks gave promising results, the studies are still inconclusive. They defined a set of classifiers combined with multiple data source fusion and a modified weighted majority voting procedure. They used their LEARN algorithm as a voting procedure instead of Adaboost.

To diagnose subjects with possible AD, Vemuri et al. [96] aimed to develop and validate a diagnosis method using support vector machine (SVM) classification and a well characterized database. They applied three different classification models that use tissue densities and covariates and Include demographic and genetic information in the classification algorithm.

Similarly, Davatzikos et al. [97] segmented the MRIs into grey matter (GM), white matter (WM) and cerebrospinal fluid (CSF) regions. They studied patterns of the spatial distribution of GM, WM and CSF volumes using a pattern classification technique. Using Pearson correlation coefficient and a leave-one-out procedure, they built spatial patterns of good discriminators between normal and MCI groups and performed a watershed-based clustering method to determine brain regions with good discriminate attributes. Finally, a pruning method was applied to reduce the number of unnecessary attributes.

In the other hand, Magnin et al. [98] developed a classification method based on support vector machine (SVM). They first segmented the image into ROIs, using anatomically labelled template of the brain developed by Tzourio-Mazoyer et al. [99] to obtain probability masks for GM, WM, and CSF. Indeed, the histogram of each ROI showed 3 modes corresponding to the 3 probability

masks. The segmented ROI was modelled with a linear combination of three Gaussians. They use the SVM algorithm to classify the subjects and statistical procedures, based on bootstrap resampling, into AD subjects and elderly control subjects (CS).

Likewise, Robinson et al. [100] developed a machine learning approach that determines population differences in whole-brain structural networks from brain atlases. The authors aimed to classify subjects based on their patterns and identify the best features which distinguish between groups, i.e. ROIs are automatically generated by label propagation and followed by classifier fusion, connections are built between ROIs using probabilistic tracking, a vector of attributes is determined using mean anisotropy measurements along those connections and finally combined with the principal component analysis (PCA) and maximum uncertainty linear discriminant analysis.

Moreover, Zhang et al. [101] combined different modality of biomarkers to get complementary information for the diagnosis of AD and MCI. According to the authors, previous studies showed that structural MRI is suitable for brain atrophy measurement, functional imaging like FDG-PET is used for hypometabolism quantification, and CSF is best used for quantification of specific proteins. Henceforth, they propose to combine three modalities of biomarkers, i.e., ADNI baseline MRI, FDG-PET, and CSF biomarkers, to accurately distinguish between AD or MCI and healthy subject controls, using a kernel combination method. They extracted and labeled volumetric features from ROIs of each MR or FDG-PET image using atlas warping algorithm and used the original values of CSF biomarkers as direct additional features. They performed feature selection method to select the most discriminative MR and FDG-PET features and finally, they apply SVM method to evaluate the classification accuracy, using a 10-fold cross-validation.

Finally, Cuingnet et al. [102] performed an automatic classification between patients with Alzheimer's disease (AD) or mild cognitive impairment (MCI) and elderly controls (CN) from structural T1-weighted MRI and compared 10 methods based on ADNI database: five voxel-based methods, three methods based on cortical thickness and two methods based on the hippocampus. In another hand, the authors performed their classification methods on three groups: CN vs. patients with probable AD, CN vs. prodromal AD or MCI converters (MCIc) and MCI non-converters (MCInc) vs. MCIc.

The smallest part of data was used for the training process and the optimization of the parameters of the chosen mathematical model and the rest was used to obtain an unbiased estimate of the performance of the methods. They finally compared DARTEL [103] registration versus SPM5 unified segmentation results [104].

3.6 Present research

In the present thesis, a pattern recognition methodology has been applied to classify an ADNI database subject as *AD* or *Normal*.

A general organization schema has been established that exhibits the overall steps of the pattern recognition methodology. Starting from a raw data that has been collected from ADNI data source, the system goes through image preprocessing steps in order to remove unwanted and/or noisy information. In the following step, the ventricles area have been extracted from the coronal view of the *3D* ADNI data using different segmentation techniques such as the active contour. From that point, every ventricles area that corresponds to one of the ADNI subjects, has been characterized using a unique set of attributes that characterizes the most the shape and morphology of the area. A learning step method has been added in order to generate class models by training an original set of data using unsupervised techniques such as the KNN and then generating a test data to be classified based on the class models created during the learning step and on the choice of the SVM classification algorithm.

Chapter 4
Processing Methodology for Predicting Alzheimer's Disease

4.1 Introduction

As mentioned in the introduction, AD causes brain tissue shrinking and larger ventricles [1] [105]. As a result, the ventricular enlargement is considered as a possible measure of Alzheimer's disease progression. In this work, the brain's ventricles image is extracted using image processing techniques such as image enhancement and segmentation methods. The extracted image for the object of interest is analyzed using characterization and classification techniques.

4.2 The Methodology

Figure 4-1 shows the steps required to analyze and predict the Alzheimer's disease. In Step 1, the ADNI data is accessed and stored in a database. In Step 2, it is reoriented for better interpretation and non-relevant information is removed. In Step 3, image segmentation is performed on the preprocessed 3D MRI neuroimaging brain data using different techniques in order to extract the ventricle's area. In Step 4, segmentation techniques are followed by attribute extraction such as surface area, centre of gravity, average intensity and standard deviation in order to analyze the shape of the ventricle. In Step 5, characterization is followed by classification/prediction methods in order to assess whether the patient is developing the Alzheimer's disease (AD).

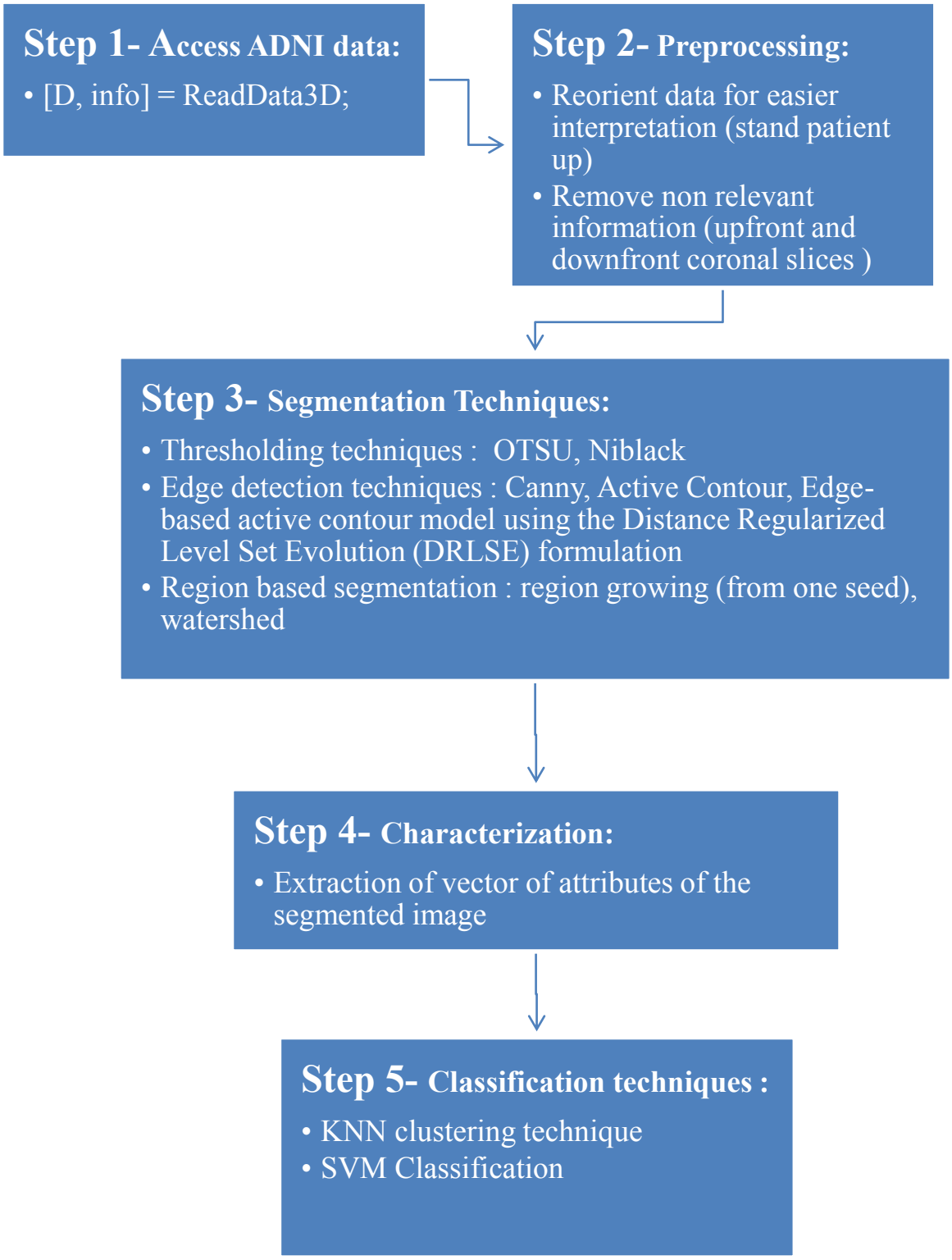


Figure 4-1 Organizational schema of the system implementation

4.3 Data access

The data was accessed from Alzheimer's Disease Neuroimaging Initiative Database (ADNI). ADNI is a multisite longitudinal clinical/imaging/genetic/biospecimen/biomarker study, whose goal is to determine the characteristics of AD as the pathology evolves from normal ageing to mild symptoms, to MCI, to dementia. It is a generally accessible data repository, which describes longitudinal changes in brain structure and metabolism.

ADNI uses several medical file formats such as the classical Analyse Format (hdr/img) that contains a header file and a separate 3*D* image and the next generation of medical images based on the Analyse Format, called NIFTI which is an *nii* structure containing both the header file and the 3*D* image.

The headers contain the information about the data such as the patient sex and age, the type of radiography, the view, size of voxels, etc, all of them stored into an info structure and the data itself as a 3D matrix usually of single type. All the nifti medical image files in ADNI database have the same standard which is: ADNI_pppp_S_ssss_Sequence_Sxxxx_Iyyyyy.nii where *pppp* is the patient ID, *ssss* is the site ID, *Sequence* is the Sequence and processing steps, *Sxxxx* is LONIUID and *yyyyy* is the Image ID.

In this thesis, Analyse and Nifti medical image formats were used. Both the formats contain the same information in the header files, even though the architecture of the structure is different in both the formats. The data is first read using a matlab gui called readData3D[3], which allows the user to open medical 3D files. It supports the following formats: Dicom Files (.dcm, .dicom), V3D Philips Scanner (.v3d), GIPL Guys Image Processing Lab (.gipl), HDR/IMG Analyze (.hdr), ISI Files (.isi) and NifTi Files (.nii) etc.

Using the Matlab statement: [D, info] = ReadData3D;

where D corresponds to the 3D MRI data and info is a structure containing the header information of the data such as the age, the sex, the view, etc. In case of Analyse 7.5 Format which uses

[3] http://www.mathworks.com/matlabcentral/fileexchange/29344

radiological orientation (LAS), data should be flipped for correct image display in MATLAB and reoriented for easier interpretation (stand patient up).

Some other packages were also used that allow 3*D* view and extraction of statistical information such as *Twfu_bpm* toolbox[4] for multimodal image analysis called biological parametric mapping (BPM), based on a voxel-wise use of the general linear model. It has a high degree of integration with the SPM[5] (statistical parametric mapping) software relying on it for visualization and statistical inference.

Medical data is accessed through different packages and stored into *3D* arrays where each element corresponds to a voxel with the three space coordinates and the intensity value. Depending on the chosen view, the three dimensions of each medical data were flipped or interchanged for correct interpretation of images. The next step is to extract meaningful information from the data using segmentation methods on the middle slice by taking the middle 2D array (or image) from the coronal section of the *3D* data.

[4] http://www.fmri.wfubmc.edu/cms/software
[5] http://www.fil.ion.ucl.ac.uk/spm/

Chapter 5

Segmentation of ADNI data

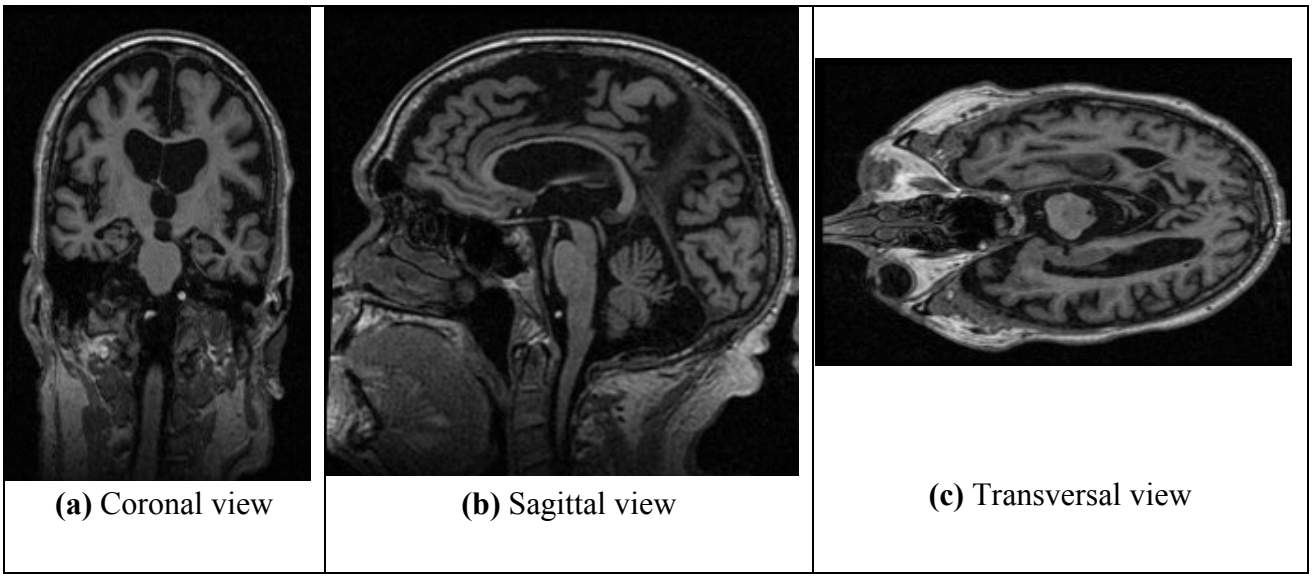
5.1 Introduction

In this chapter, 2D/3D image segmentation of original 3D MRI neuroimage brain data is performed. The theory about 2D segmentation is easily transported into 3D; however the cost of the algorithms is highly increased. The 3D segmentation transforms the original voxels in 3D images into 3D regions where each region, identified by a different label, represents meaningful physical behaviors defined by a vector of attributes (average, standard deviation, etc.). There are many existing segmentation techniques applied for medical image segmentation, including statistical methods, thresholding, edge detection, region-based techniques and more recently multi-resolution (using wavelets, ridgelets, etc.) techniques [106] [107]. The choice of the method depends on the type and quality of the image and the statistics of the extracted regions.

5.2 Preprocessing the ADNI Data

When reading ADNI data (**Figure 4-1**), the resulting *3D* data shows slices of images that can be visualized into three different views: coronal, transversal or sagittal (**Figure 5-1**). More relevant for the present work is visualization of the lateral ventricles of the brain. There are four ventricles in the brain, filled with cerebrospinal fluid (CSF), that are located within the brain parenchyma; two of them are called the lateral ventricles which are two curved shaped cavities located within the cerebrum in the middle region of the brain (**Figure 5-2**) [108]. In the preprocessing step, Analyse7.5 data is flipped for correct image display and to focus on the coronal view slices of the 3D data. To extract the lateral ventricles, some of the upfront and down-front coronal view slices,

spatially low positioned (below brain mass) are removed, in order to decrease the cost of the algorithms.



(a) Coronal view

(b) Sagittal view

(c) Transversal view

Figure 5-1 Three different views of a *3D* ADNI data

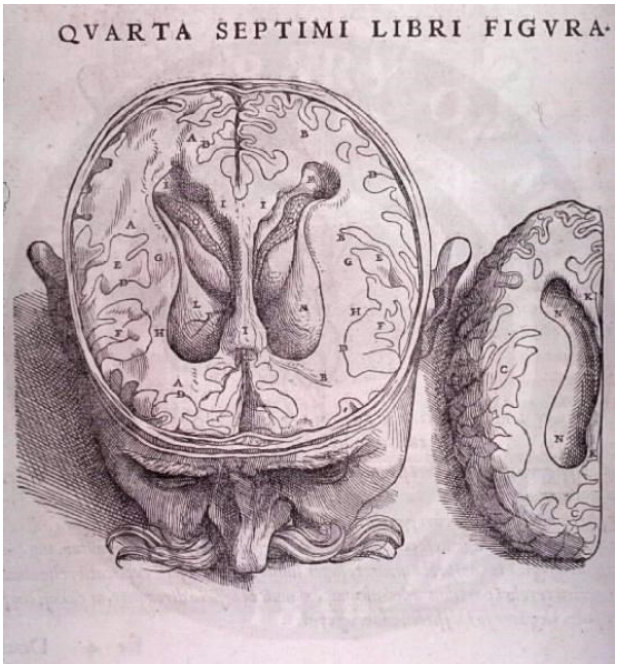


Figure 5-2 Lateral Ventricles of the Brain.
This classical woodcut is presented courtesy of the National Library of Medicine.
(http://www.gather.com/viewImage.action?fileId=3096224744546601)

Figure 5-3 shows the grey view of the coronal slices 1 to 255 every 10 slices of the ADNI AD subject" I60451.nii". It can be noticed that the brain tissue starts being displayed in slice number 31 and the lateral ventricles are visible from slice number 91 and they fade away completely starting at slice number 221. As a preprocessing step, the slices that provide little information about the shape of the lateral ventricles can be removed.

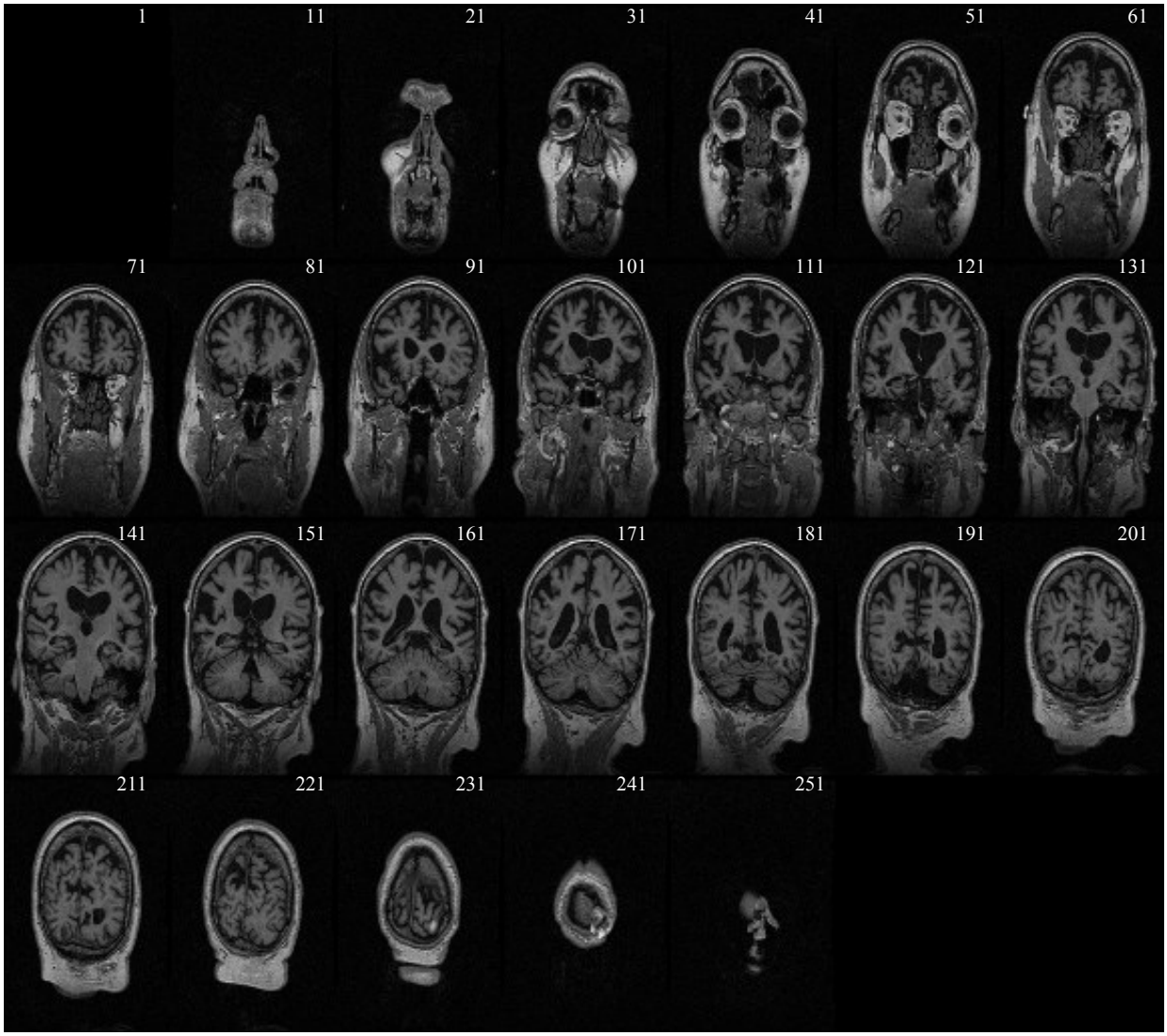


Figure 5-3 Grey view of the coronal slices (Slices 1 to 251 every 10 slices of the ADNI AD subject I60451.nii) from left to right, top to bottom

5.3 Segmentation of ADNI Data

5.3.1 Thresholding techniques

Following the preprocessing step discussed in the above section, some of the classical segmentation techniques were performed in order to extract the ventricles' region. The first performed segmentation method was the Otsu global thresholding algorithm (see **section 2.3.1.1**) by selecting initial threshold level values based on the histogram, i.e. the distribution of the image's pixels' intensities, of each coronal slice (**Figure 5-4)** and performing an adaptive thresholding based on the iterative threshold and the in-between variances. Since the selection of the initial threshold level value is based on the histogram of the image (**Figure 5-4)** and the image is dominated by a black background (high dark intensity frequency), it would be necessary to be careful when choosing the threshold. The resulting image in **Figure 5-5** gives the entire area of the brain based on the intensity value after segmenting the middle slice using Otsu method. The ventricle's region has been entirely extracted but some of the unwanted regions have been under or over segmented which will distort the final classification result.

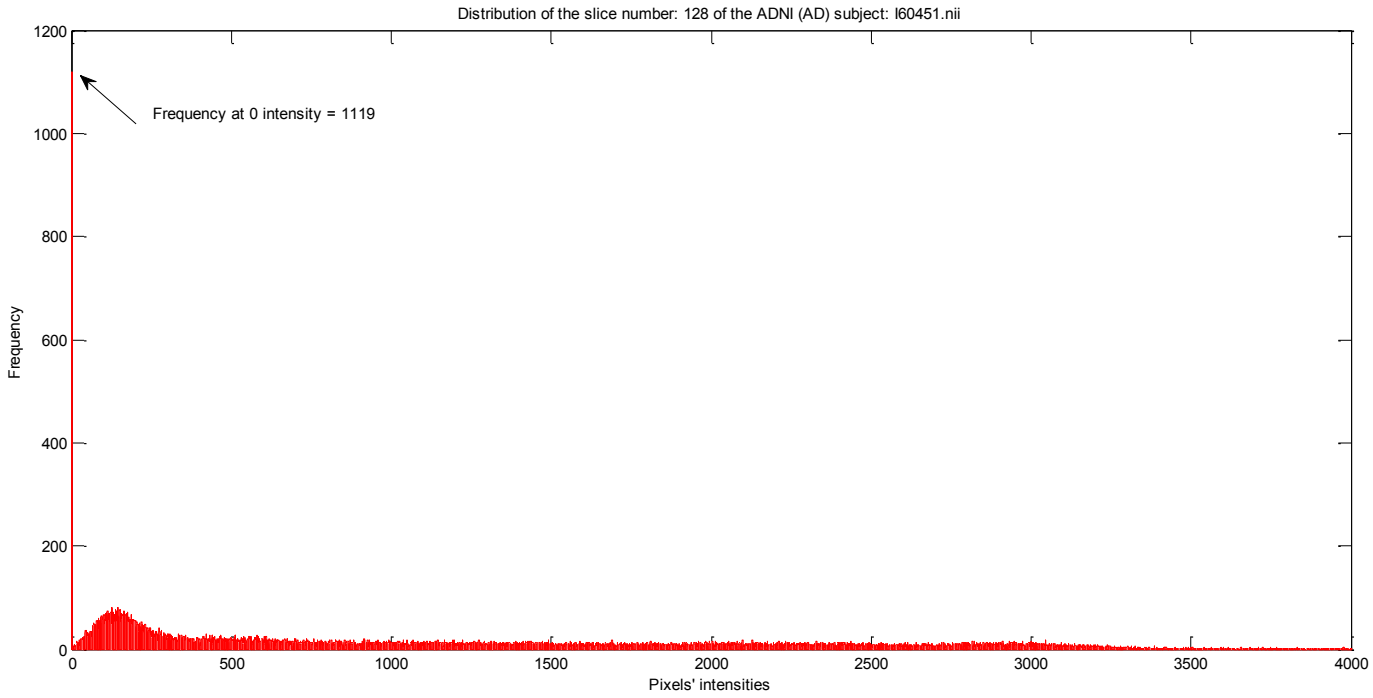


Figure 5-4 Histogram of the middle coronal slice of the ADNI AD subject "I60451.nii"

Figure 5-5 Middle slice of the ADNI AD subject "I60451.nii" after OTSU global thresholding

The Otsu method, as defined above, is a global thresholding method that calculates a single threshold based on the entire image intensity without much attention to the complexity of the image. However, these results can be used for selecting some markers as initial points to more sophisticated segmentation methods. Since the Otsu method fails on extracting precise regions, local thresholding methods were used afterward. Among those methods, the classic Niblack algorithm. Niblack's algorithm is presented in **Figure 5-6**. It is an adaptive or local thresholding method that depends on the choice of the parameter *k.* The latter is a coefficient that controls the standard deviation. However, it is also dependent on the size of the filter *N*. As can be seen in **Figure 5-7**, **Figure 5-8** and **Figure 5-9**, increasing the size of the filter N gives better results. In addition, smoothing filters where used to remove irrelevant details from the image.

1. Read the original slice image
2. Initialize the segmented image
3. Initialize the coefficient *k* that controls the standard deviation
4. Get the size of the filter *N*
5. for every pixel, calculate the average and the standard deviation of the neighbourhood of size $N \times N$
6. Set the segmented pixel image = Slice(i, j) > mu + k * sigma

Figure 5-6 Implementation of Niblack thresholding algorithm

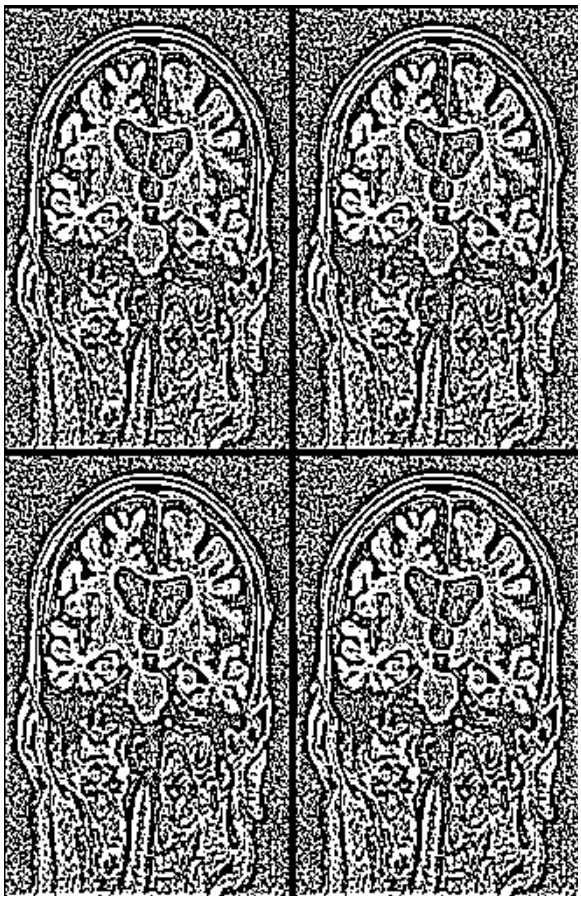


Figure 5-7 Results of using Niblack local thresholding segmentation using the middle slice of the ADNI AD subject "I60451.nii", N=5. Upper left image k=-0.01, Upper right image k=-0.02, Bottom left image k=-0.03, Bottom right image k=-0.04.

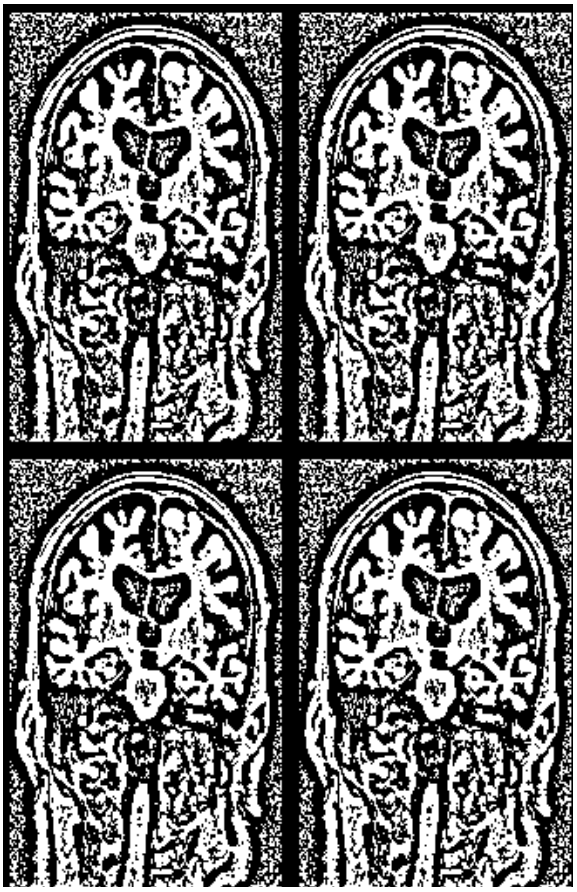


Figure 5-8 Results of using Niblack local thresholding segmentation using the middle slice of the ADNI AD subject "I60451.nii", N=11. Upper left image k=-0.01, Upper right image k=-0.02, Bottom left image k=-0.03, Bottom right image k=-0.04.

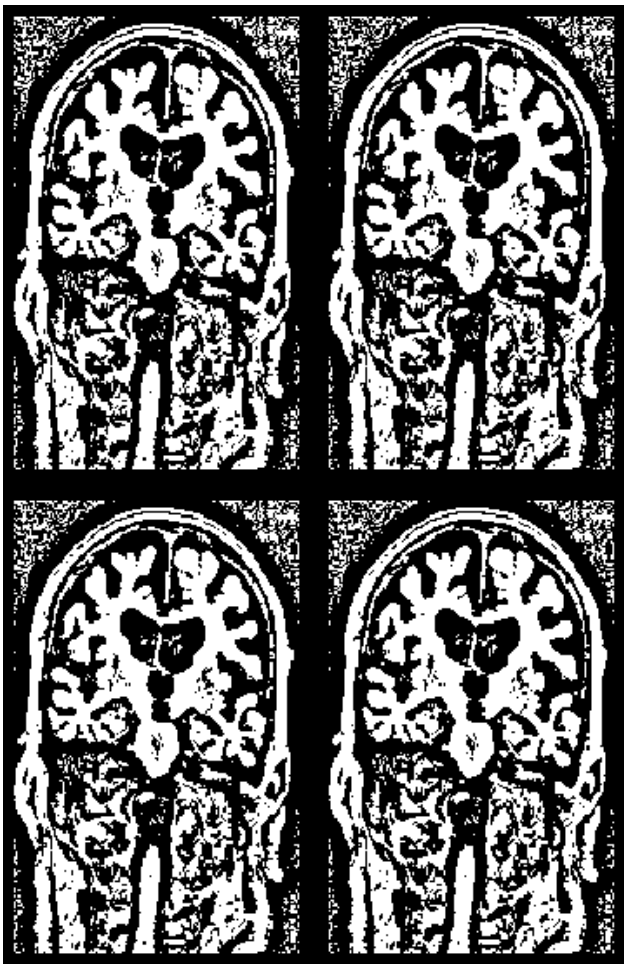


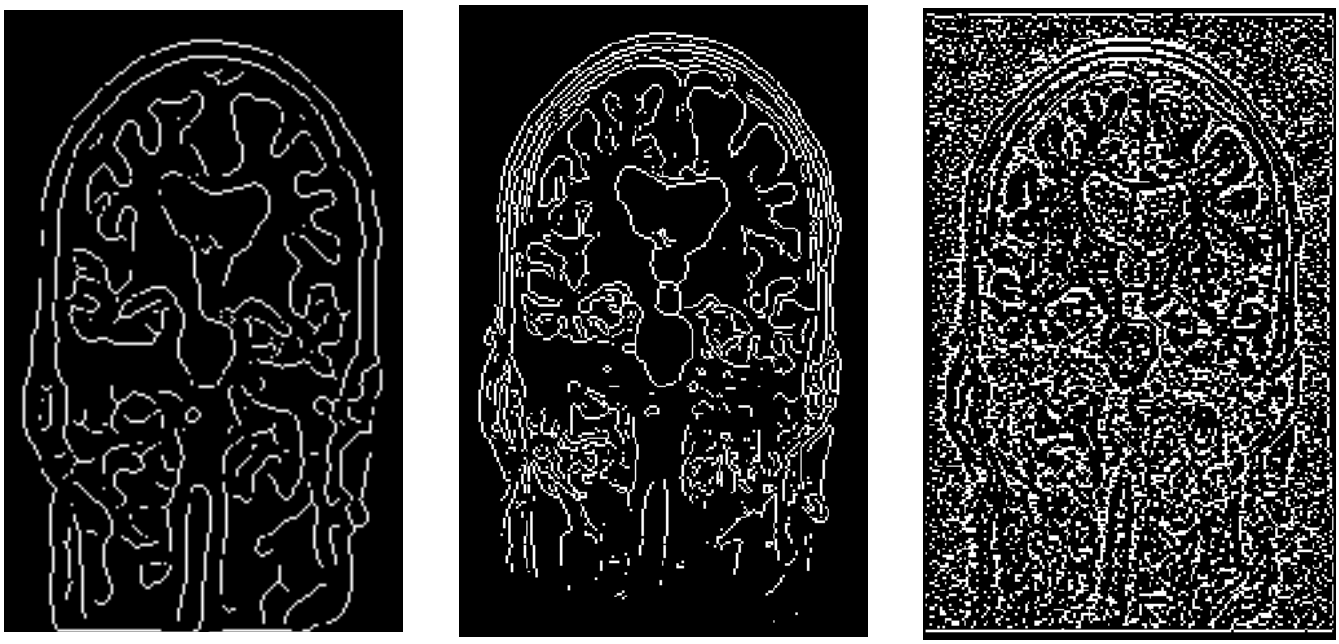
Figure 5-9 Results of using Niblack local thresholding segmentation using the middle slice of the ADNI AD subject "I60451.nii", N=17. Upper left image k=-0.01, Upper right image k=-0.02, Bottom left image k=-0.03, Bottom right image k=-0.04.

When using thresholding techniques, the results lacked in precision. That was due to the nature of the medical images which have very close intensity values. Consequently, low level features as in a simple thresholding technique, based only on intensity values, cannot capture the relationship between pixels.

5.3.2 Edge detection techniques

From this point, edge detection techniques were used in order to extract the contour of the ventricles' region. First of all, a classic and well established edge detector, known as Canny edge detector, was used. The edge detection using Canny edge detection algorithm [19] (see the

resulting images output in **Figure 5-10**) gave correct output contours with less noise compared to Marr-and-Hildreth or Sobel. Even though the ventricle's region is well detected and stable, some noise is introduced outside the region itself that must be thresholded a second time. The contours should also be closed in order to get well defined regions.



(a) Canny edge detection **(b)** Sobel edge detection **(c)** Marr-and-Hildreth edge detection

Figure 5-10 Canny, Sobel and Marr-and-Hildreth edge detection techniques using the middle slice of the ADNI AD subject "I60451.nii"

As the latter edge detection techniques lacked in precision regarding the extraction of the region of interest, the active contour technique was a better option to follow. The active contour, as explained in section **2.3.2.2**, gave much better results (**Figure 5-11**) than Canny's. However, the initialization is not automatic and is based on the current slice. Additionally, the initial contour should be around the region of interest in order to detect only the lateral ventricles. The cost has been reduced by initializing the contour to the same contour for every slice. This is acceptable for the images presenting a big area around the ventricle chambers and with similar brain dimensions.

However, a registration of the images should be proceeded to get similar dimensions and position of the brain in the case of a larger number of data (thousands), since the algorithm was applied on 121 medical ADNI data.

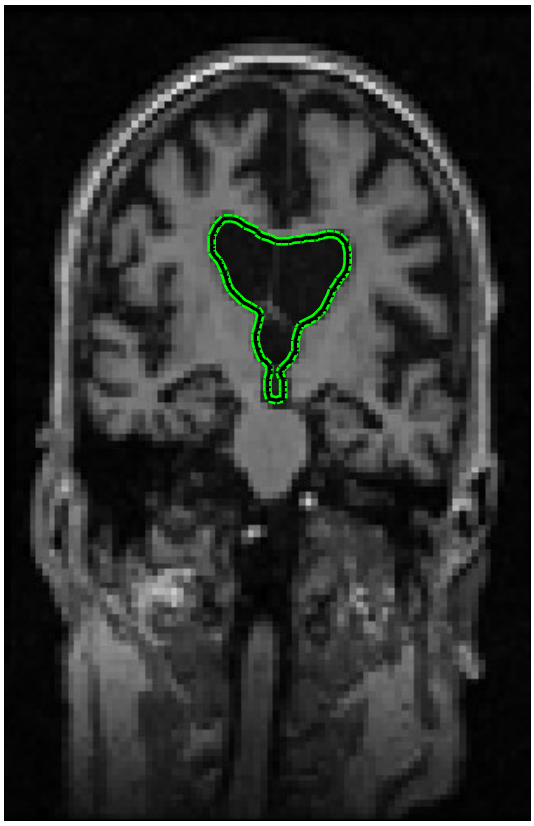


Figure 5-11 Active contour using a slightly reoriented and resized middle slice

Hence, for smaller areas, the latter algorithm was unsuccessful compared with the DRLSE segmentation which extracts the active contour using the Distance Regularized Level Set Evolution (DRLSE) formulation [109] (the Matlab code of the author can be found in *http://www.imagecomputing.org/~cmli/DRLSE/*). The parameters were changed as illustrated in **Figure 5-12**. The results of the same data image, using the DRLSE method, can be seen in **Figure 5-13** after 510 iterations.

1. Set the time step
2. Calculate the coefficient of the distance regularization term
3. Set the number of iterations
4. Set the coefficient of the weighted length term
5. Set the coefficient of the weighted area term
6. Set the parameter that specifies the width of the Dirac Delta function
7. Set the scale parameter in Gaussian kernel

Figure 5-12 Section from DRLSE matlab code **[109]** illustrating the parameter setting

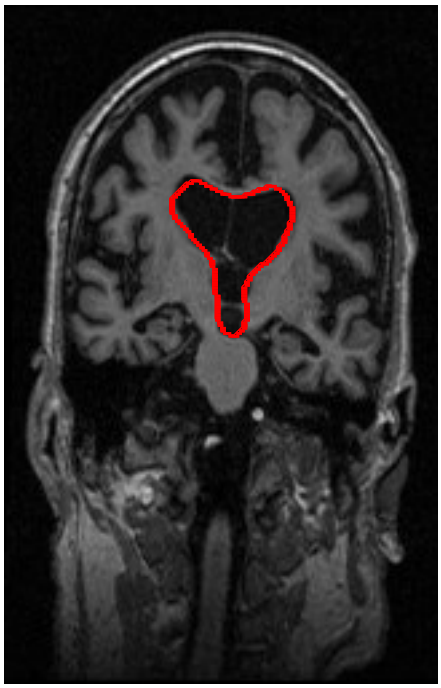


Figure 5-13 Edge-based active contour model using the Distance Regularized Level Set Evolution (DRLSE) formulation after 510 iterations

Even though DRLSE segmentation eliminates the need for re-initialization, the level set of the function was initialized by extracting correct position points and the same local points were used as a basis to the rest of slices using the Matlab statement: [BW, c, r] = roipoly(mat2gray(Slice)); In addition to the general and efficient initialization of the level set function, the algorithm reduced the number of iterations, while ensuring sufficient numerical accuracy. Nevertheless, the DRLSE algorithm was not applied on a large set of images to get a realistic idea on its results. More images should be tested.

5.3.3 Region growing

Following the edge detection techniques, region based segmentations were also used. One of the leading methods is the region growing technique that starts from one point called seed and start growing the region by adding neighbours according to a stopping criterion. Refer to section **2.3.3.1** and to **Figure 5-14** for further details.

```
% Initialize the Output to zero matrix of same size than the input
% Start the region with one pixel
% Create a large matrix  to store the current segmented region pixels' (neighbours)
and their coordinates
while(distance between region and possible new pixels is less than a certain
treshold)
    % Add new neighbors pixels
    for j=1:4, %four neighbours because it is 4 connectivity
        % Calculate the neighbour coordinate
        % Check if neighbour is inside or outside the image
        % Add neighbor if inside and not already part of the segmented area
    end

    % Add a new block of free memory
    % Add pixel with intensity nearest to the mean of the region, to the region
    % Calculate the new mean of the region
    % Save the x and y coordinates of the pixel
    % Remove the pixel from the neighbour (check) list
end
% Return the segmented area as logical matrix
```

Figure 5-14 Code Snippet of Region Growing method

The region growing algorithm gave good results for a small number of images. That was due to the choice of the initial seed which has to be more automatic and less prone to errors; the cost is much bigger since the seed is chosen for every slide. **Figure 5-15** shows the resulting segmented image for the coronal middle slice of the ADNI AD subject "I60451.nii". The size of the slice image is 256×166, the *distance*=700 pixels and the initial seed point has the coordinates x=85 and y=105. We apply a mathematical morphological preprocessing before the region growing using the following Matlab code:

```
n=3;

Slice = imclose(imopen(imclose(imopen(Slice, ones(n)), ones(n)), ones(n)),ones(n));
```

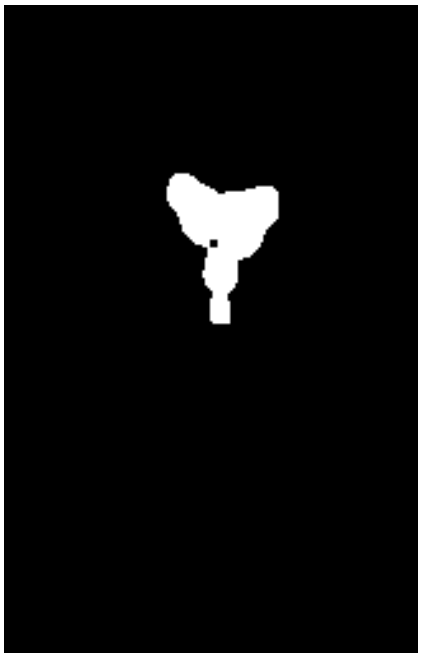


Figure 5-15 Segmentation of the middle slice using region growing method

5.3.4 Watershed method

Finally, the watershed method was tested (see section **2.3.4** for further details). The watershed method extracted the ventricle region but over-segmented the live tissue of the brain (**Figure 5-16**).

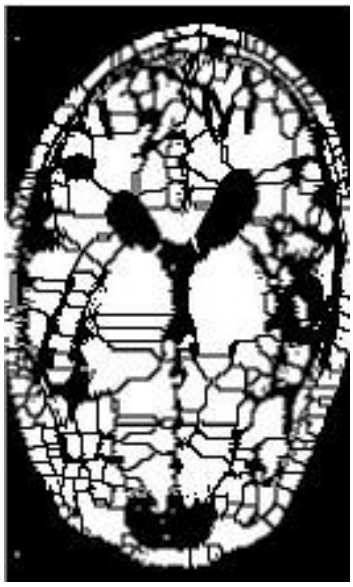


Figure 5-16 Segmentation of the middle slice using Watershed method

Chapter 6

Characterization and Classification Techniques for ADNI Data

6.1 Introduction

Up to this point, the process of pattern recognition as illustrated in **Figure 4-1** was performing several steps such as access ADNI database, describe the medical data, read the volumetric MRI, extract the middle slice of the brain region and perform segmentation methods in order to detect the region of brain's ventricle. In this chapter, the fourth and fifth steps were performed, where the system will generate a vector of attributes that characterizes this region, create a database that contains the generated data, perform clustering to get the class labels and finally perform some classification methods based on the clustering results.

The choice of attributes depends closely on the object's shape and statistics such as the statistical moments of order n, textures, geometric measures, etc. Once the attributes were extracted, a database, that is well defined and large enough to contain both training and testing data, was created. The first step of classification is to proceed to the learning process in order to produce the classes' categorical labels, and then to perform the classification. A step of accuracy measure of the classifier is added to assess the accuracy of the employed classifier.

6.2 Attribute selection

The AD disease can be assessed based on the shape of the ventricle chamber's area. Since AD causes the loss of brain mass due to molecules created in this area and spread all over the brain, it was assumed in this work that the whole brain can be assessed based on this specific area in order to decrease the cost of calculation. Having stated this, the ventricle's area was characterized based on its shape and morphology using statistical and geometrical attributes. The resulting attributes are respectively the surface area of the extracted region (**Surf**), the perimeter (**Per**), the first statistical moment (**Mean**), the second statistical moment (**Std**), 28 horizontal distances (**W01, W02, …, W28**), the height (**Height**) and the coordinates of the center of gravity of the region (**Gx, Gy**). The attributes are normalized into the range [0 1].

The surface (attribute **Surf**) of the extracted region corresponds to the number of pixels of the region divided by the size of the image (total number of pixels) to get normalized values.
The perimeter (attribute **Per**), or the contour of the extracted region, is the sum of the contour pixels divided by the size of the image (total number of pixels) to get normalized values.
The first statistical moment (attribute **Mean**), called the average or the mean, and the second statistical moment (attribute **Std)**, called the standard deviation, calculate respectively the average value of the region's intensity pixels and their standard deviation. **Mean** and **Std** are normalized into the range [0 1] using the following formulas:

$$\text{Mean} = \frac{(\text{Mean} - \text{Min})}{(\text{Max} - \text{Min})}$$

$$\text{Std} = \frac{(\text{Std} - \text{Min})}{(\text{Max} - \text{Min})}$$

where Max and Min are respectively the maximum and the minimum intensities of the extracted region.

W01 to **W28** are the normalized horizontal distances of the extracted region. The widths are normalized by dividing the original values by the number of columns of the image.

Height is the normalized height of the region. The height is normalized by dividing the original value by the number of rows of the image

Gx and **Gy** are the normalized coordinates of the center of gravity or center of mass statistic of the extracted region. The coordinates are normalized by dividing the original values by the size of the image.

The algorithm that calculates the vector of attributes is listed in **Algorithm 6-1**.

Algorithm 6-1 Extraction of vector attributes

1. Get the contour from the region.
2. Initialise the vector of attributes.
3. Add the first attribute (normalized area of the extracted region).
4. Add the second attribute (normalized perimeter of the extracted region).
5. Add the first and the second standardized statistical moments.
6. Add the standardized horizontal distances to the vector of attributes.
7. Add the standardized height to the vector of attributes.
8. Add the standardized center of gravity to the vector of attributes.

In order to make a first assessment of the behavioral trend of each attribute, a statistical analysis was performed, over a set of 121 patterns, of the 35 attributes using stacked columns where each attribute is summarized by a column bar containing five key data points (also called five-number summary) named *max, third quartile (Q3), median, first quartile* (Q1) and *min*. The majority of the attributes tend to vary clearly and shall induce some relevant information regarding the classification between the different medical image data base (**Figure 6-1**). A second assessment of the statistical trend of the attributes was performed by visualizing the average, the standard deviation and the mode. The attributes showed distinct behaviors as well (**Figure 6-2**). The latter behavior should lead to a good classification tool.

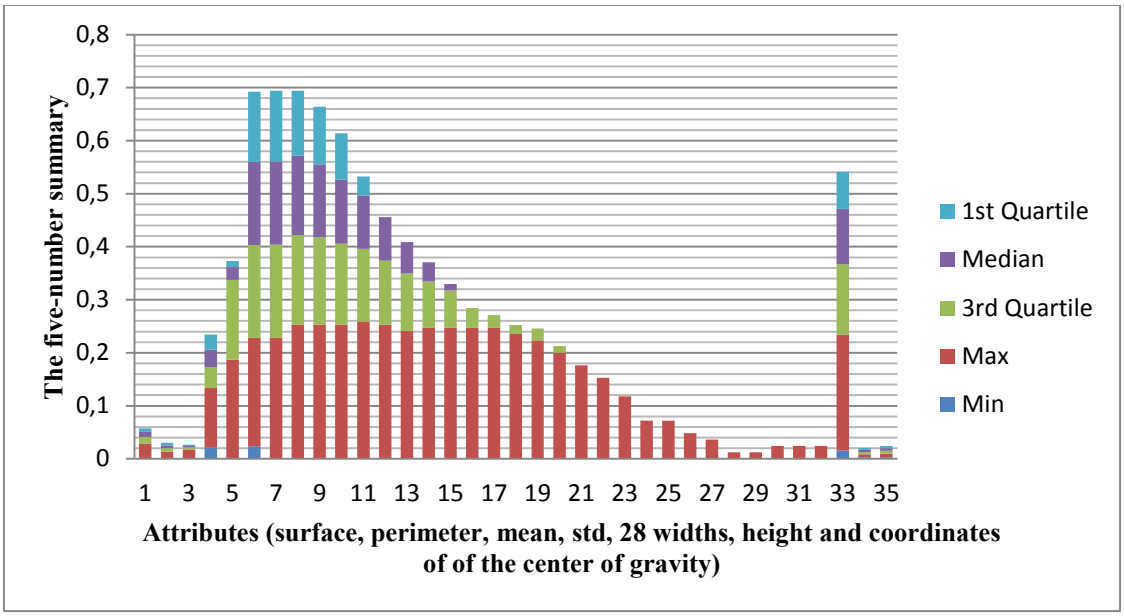


Figure 6-1 Statistical trend of 121 medical image data base based on the five-number summary (Min, 1st Quartile, Median, 3rd Quartile and Max)

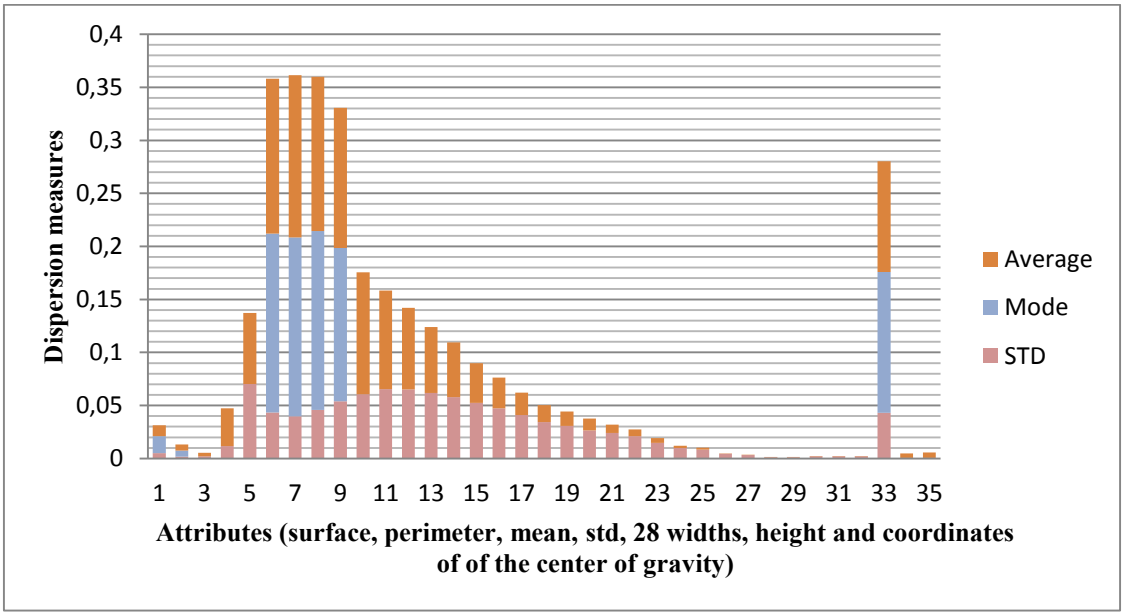


Figure 6-2 Statistical trend of 121 medical image data base based on the standard deviation, the average and the mode

6.3 Classification

After building a database of 121 patterns of 35 dimensions, the last step of the general framework, i.e. the classification, was performed. The clinical assessment of the patients imply 29 healthy (*Normal*) subjects and 92 *AD* subjects. The accuracy of the classification algorithms as well as the choice of the vector of attributes will be assessed, based on the latter statement. The K-Nearest Neighbors (KNN) clustering technique was first applied. Clustering is a procedure which aims to categorize objects. In particular, the objective is to categorize objects in groups that do not have class labels. Popular methods of clustering include; K-Nearest Neighbors, K-Means and K-Medoids [43], [44].

The clustering results of some pairs of attributes are visually summarized in Appendix A, where each class is marked by the same color. Since there is 35 attributes, there would be $\frac{35\times34}{2} = 1190$ pairs of attributes. As 1190 results cannot be shown, only the results of some attributes were shown, i.e. the surface, the perimeter, the average, the standard deviation, the first two widths, the height and the coordinates of the center of gravity. As there is 9 attributes, only $\frac{9\times8}{2} = 36$ pairs of attributes are shown. They are including the 2-dimensionnal clustering results, projected into the coordinates of the pair of attributes, using two-dimensional KNN clustering technique. In every KNN results image, the *AD* subjects were labeled as red stars and the *Normal* subjects as green stars. The classification performance of the latter KNN results for each pair of attributes were summarized in **Table 6.1**.

Clinical tests were used to assess the performance of the classifiers based on the following performance classification parameters [110]:

- The sensitivity (SN) which refers to the ability of identifying the AD patients.

$$\mathrm{SN} = \frac{\mathrm{TP}}{(\mathrm{TP} + \mathrm{FN})} \times 100\%$$

- The specificity (SP) which refers to the ability of identifying the normal or healthy people.

$$\mathrm{SP} = \frac{\mathrm{TN}}{\mathrm{TN} + \mathrm{FP}} \times 100\%$$

- The positive predictive value (PPV), also called precision or probability of correct positive prediction.

$$\mathrm{PPV} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}} \times 100\%$$

- The Negative predictive value (NPV), which is the probability of correct negative prediction.

$$\mathrm{NPV} = \frac{\mathrm{TN}}{\mathrm{TN} + \mathrm{FN}} \times 100\%$$

- The accuracy (ACC), which is the probability of both correct positive and negative predictions.

$$\mathrm{ACC} = \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{TP} + \mathrm{FP} + \mathrm{TN} + \mathrm{FN}} \times 100\%$$

Where the parameters TP, FP, TN and FN are defined as follows:

True positive (TP): the patient has the AD and the classification result is positive (AD).
False positive (FP): the patient is normal and the classification result is positive.
True negative (TN): the patient is normal and the classification result is negative (Normal).
False negative (FN): the patient has the AD but the test is negative.

Based on the latter parameters, as noticed in **Table 6.1**, good classification accuracy was achieved (almost 82%) for the pair of attributes *Surface* vs. *Gx*. For the same pair, 83.33% of subjects, with a positive test, have actually the *AD* disease and 66.67% of subjects, with a negative test, do not have the disease. A bad classification accuracy has been obtained for the pair of attributes *Gx*. vs. *Gy* (53.33%) with the least sensitivity value (59.57%). Additionally, whenever *Gy* was paired with another attribute, the classification tended to be less accurate (between 56.33% and 66.67%) with the exception of the *Surface* where the accuracy was actually 75%. For the rest of pairs of attributes, the accuracy was between 68.33% and 78.33%.

Table 6.1 KNN classification results for a pair of attributes

Pairs of attributes		SN	SP	PPV	NPV	ACC
Surf	Per	85.1064%	38.4615%	83.3333%	41.6667%	75%
Surf	Mean	89.3617%	38.4615%	84%	50%	78.3333%
Surf	Std	87.234%	38.4615%	83.6735%	45.4545%	76.6667%
Surf	W01	93.617%	23.0769%	81.4815%	50%	78.3333%
Surf	W02	82.9787%	23.0769%	79.5918%	27.2727%	70%
Surf	Height	78.7234%	30.7692%	80.4348%	28.5714%	68.3333%
Surf	Gx	95.7447%	30.7692%	83.3333%	66.6667%	81.6667%
Surf	Gy	87.234%	30.7692%	82%	40%	75%
Per	Mean	76.5957%	53.8462%	85.7143%	38.8889%	71.6667%
Per	Std	78.7234%	30.7692%	80.4348%	28.5714%	68.3333%
Per	W01	85.1064%	23.0769%	80%	30%	71.6667%
Per	W02	87.234%	15.3846%	78.8462%	25%	71.6667%
Per	Height	85.1064%	23.0769%	80%	30%	71.6667%
Per	Gx	85.1064%	30.7692%	81.6327%	36.3636%	73.3333%
Per	Gy	78.7234%	23.0769%	78.7234%	23.0769%	66.6667%
Mean	Std	87.234%	23.0769%	80.3922%	33.3333%	73.3333%
Mean	W01	89.3617%	23.0769%	80.7692%	37.5%	75%

Mean	W02	82.9787%	23.0769%	79.5918%	27.2727%	70%
Mean	Height	82.9787%	23.0769%	79.5918%	27.2727%	70%
Mean	Gx	89.3617%	15.3846%	79.2453%	28.5714%	73.3333%
Mean	Gy	72.3404%	30.7692%	79.0698%	23.5294%	63.3333%
Std	W01	82.9787%	30.7692%	81.25%	33.3333%	71.6667%
Std	W02	76.5957%	15.3846%	76.5957%	15.3846%	63.3333%
Std	Height	87.234%	30.7692%	82%	40%	75%
Std	Gx	87.234%	7.6923%	77.3585%	14.2857%	70%
Std	Gy	68.0851%	15.3846%	74.4186%	11.7647%	56.6667%
W01	W02	89.3617%	23.0769%	80.7692%	37.5%	75%
W01	Height	85.1064%	15.3846%	78.4314%	22.2222%	70%
W01	Gx	82.9787%	15.3846%	78%	20%	68.3333%
W01	Gy	74.4681%	38.4615%	81.3953%	29.4118%	66.6667%
W02	Height	78.7234%	30.7692%	80.4348%	28.5714%	68.3333%
W02	Gx	85.1064%	15.3846%	78.4314%	22.2222%	70%
W02	Gy	72.3404%	15.3846%	75.5556%	13.3333%	60%
Height	Gx	82.9787%	30.7692%	81.25%	33.3333%	71.6667%
Height	Gy	82.9787%	23.0769%	79.5918%	27.2727%	70%
Gx	Gy	59.5745%	30.7692%	75.6757%	17.3913%	53.3333%

As some pairs of attributes gave better results than others, all the attributes were used to see their impact on the classification results for 121 patterns. **Table 6.2** show the output label of each pattern and its "actual" clinical label. The results gave 76.6667% of classification accuracy with 85.1064% of sensitivity, 46.1538% of specificity, 85.1064% of precision and 46.1538% of probability of correct negative prediction. Even though the values obtained by pairing *Surface* with *Gx* were not reached, some good results were obtained, since 40 *AD* subjects over 47 have been correctly identified from 60 testing samples.

Table 6.2 Class labels of the KNN clustering technique using all the attributes

Pattern	Clinical label	Output label	Pattern	Clinical label	Output label	Pattern	Clinical label	Output label
1	'AD'	'AD'	42	'AD'	'Normal'	83	'AD'	'AD'
2	'AD'	'AD'	43	'AD'	'AD'	84	'AD'	'AD'
3	'AD'	'AD'	44	'AD'	'AD'	85	'AD'	'AD'
4	'AD'	'AD'	45	'AD'	'AD'	86	'AD'	'AD'
5	'AD'	'AD'	46	'AD'	'AD'	87	'AD'	'AD'
6	'AD'	'AD'	47	'AD'	'AD'	88	'AD'	'AD'
7	'AD'	'AD'	48	'AD'	'Normal'	89	'AD'	'AD'
8	'Normal'	'Normal'	49	'AD'	'AD'	90	'Normal'	'Normal'
9	'AD'	'Normal'	50	'Normal'	'AD'	91	'AD'	'AD'
10	'AD'	'AD'	51	'Normal'	'AD'	92	'AD'	'AD'
11	'AD'	'AD'	52	'Normal'	'Normal'	93	'AD'	'AD'
12	'AD'	'AD'	53	'Normal'	'AD'	94	'Normal'	'Normal'
13	'AD'	'AD'	54	'AD'	'AD'	95	'AD'	'AD'
14	'AD'	'AD'	55	'AD'	'AD'	96	'AD'	'AD'
15	'AD'	'AD'	56	'AD'	'AD'	97	'AD'	'AD'
16	'AD'	'Normal'	57	'AD'	'AD'	98	'AD'	'AD'
17	'Normal'	'AD'	58	'AD'	'AD'	99	'AD'	'AD'
18	'AD'	'AD'	59	'Normal'	'Normal'	100	'AD'	'AD'
19	'Normal'	'AD'	60	'Normal'	'Normal'	101	'AD'	'AD'
20	'AD'	'AD'	61	'Normal'	'Normal'	102	'AD'	'AD'
21	'AD'	'AD'	62	'AD'	'AD'	103	'AD'	'AD'
22	'Normal'	'Normal'	63	'AD'	'AD'	104	'AD'	'AD'
23	'AD'	'AD'	64	'AD'	'AD'	105	'AD'	'AD'
24	'AD'	'AD'	65	'Normal'	'Normal'	106	'Normal'	'Normal'
25	'AD'	'AD'	66	'Normal'	'Normal'	107	'Normal'	'Normal'
26	'AD'	'Normal'	67	'Normal'	'Normal'	108	'AD'	'AD'
27	'Normal'	'AD'	68	'AD'	'AD'	109	'Normal'	'Normal'
28	'AD'	'Normal'	69	'Normal'	'Normal'	110	'Normal'	'Normal'
29	'AD'	'AD'	70	'AD'	'AD'	111	'AD'	'AD'

30	'AD'	'AD'	71	'AD'	'AD'	112	'AD'	'AD'
31	'AD'	'AD'	72	'AD'	'AD'	113	'Normal'	'Normal'
32	'AD'	'Normal'	73	'AD'	'AD'	114	'AD'	'AD'
33	'AD'	'AD'	74	'Normal'	'Normal'	115	'AD'	'AD'
34	'AD'	'AD'	75	'AD'	'AD'	116	'AD'	'AD'
35	'AD'	'AD'	76	'Normal'	'Normal'	117	'AD'	'AD'
36	'AD'	'AD'	77	'AD'	'AD'	118	'Normal'	'Normal'
37	'AD'	'AD'	78	'AD'	'AD'	119	'AD'	'AD'
38	'AD'	'AD'	79	'AD'	'AD'	120	'AD'	'AD'
39	'AD'	'AD'	80	'AD'	'AD'	121	'Normal'	'Normal'
40	'Normal'	'AD'	81	'AD'	'AD'			
41	'Normal'	'Normal'	82	'AD'	'AD'			

The KNN classification technique performed fairly well but lacked accuracy when using some of the attributes. Also, even though the *AD* subjects have been correctly identified with a good percentage, some of the *Normal* subjects have been identified as diseased. Another classifier might resolve this problem and decrease the cost of treating healthy subjects. A similarity criterion, that classifies each data into the proper class name or label using the support vector machine (SVM) classification method, was created.

The SVM is a classifier that attempts to create a linear vector that segments the classes equally. In the event that the data is not linearly inseparable, the SVM algorithm lifts the space into a higher dimensional plane, until, on that plane, a vector can separate the classes. It is a particularly good classifier, albeit a slow one [36], [42], [39]. It supports various mathematical formulations. We used the C-Support Vector Classification or C-SVC [110], [111] that solves the following optimization problem

$$\min_{\omega,b,\xi} \quad \frac{1}{2}\omega^T \omega + C\sum_{i=1}^{l}\xi_i \qquad \textbf{6.1}$$

subject to the decision function $\quad y_i(\omega^T\phi(x_i) + b) \geq 1 - \xi_i$

$$\xi_i \geq 0, i = 1, \ldots, l,$$

given training vectors of attributes $x_i \in R^n$, *i=1, ...,l,* in two classes, and a label vector $y \in R^l$ such that $y_i \in \{1, -1\}$, to indicate the first or the second class

ω and b are adjustable parameters of the decision function which indicate, respectively, the weight and the bias

$\phi(x_i)$ are predefined functions of x , that map x_i into a higher-dimensional space

$C > 0$ is the regularization parameter.

We set the parameter C of class i to $weight*C$ in C-SVC in Equation **6.1**.

The SVM algorithm generates a support vector that defines the margin of largest separation between the two classes.

The totality of the SVM results could not be interpreted since they represent 35 dimensions. However, the results were examined visually by plotting the SVM results of only 2-D vector of attributes. The data is first divided into two groups. The first group consists of 61 training patterns and the second group refers to the testing patterns and consists of the remaining 60 patterns. The SVM classifier was applied on the training data followed by the testing data. Each plot includes training data, testing data and the support vectors. In the figures, support patterns are indicated with blue squares surrounding one of the patterns. To assess the classification results, the previous performance classification parameters were applied.

Using the first two attributes that correspond to the surface and perimeter of the objects, and by forcing a linearly separable solution, 75% of the patterns have been correctly classified. The patterns were more likely classified as *AD* than *Normal* with a sensitivity of 93.617% and a specificity of 7.6923% for 55 support vectors (SV). The classification results including the labeled patterns and support vectors were plotted in **Figure 6-3** showing a predominant *AD* class for the first two attributes. The *Normal* patterns implied smaller values of surface and perimeter. For the same pair of attributes, the SVM classification performance gave close results compared to the KNN.

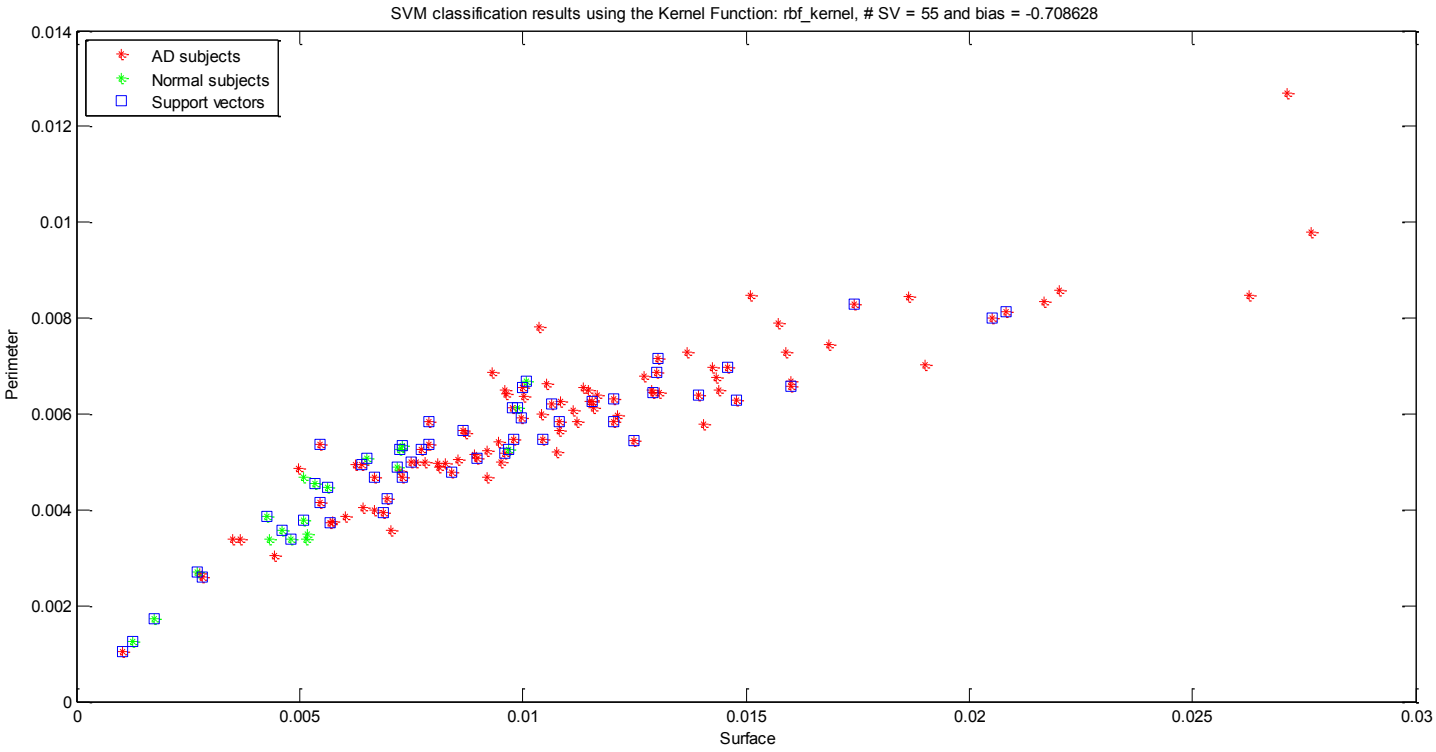


Figure 6-3 SVM classification results for the first two attributes (Perimeter vs. Surface) with linearly separable solution

As the SVM classifier helps selecting the best features that improve the classification results, the same work was performed for the remaining 35 pairs of attributes and summarized in Appendix B. **Table 6.3** summarizes the SVM classification results corresponding to 36 pairs of attributes by forcing a linearly separable solution. The results included the bias, the number of support vectors (SV), the sensitivity (SN), the specificity (SP), the positive predictive value (PPV), the negative predictive value (NPV) and the accuracy (ACC).

Table 6.3 SVM classification results for a pair of attributes with linearly separable solution including the bias, the number of support vectors (SV), the sensitivity (SN), the specificity (SP), the positive predictive value (PPV), the negative predictive value

Pairs of attributes		Bias	SV	SN	SP	PPV	NPV	ACC
Surf	Per	-0.7086	55	93.617%	7.6923%	78.5714%	25%	75%
Surf	Mean	-0.6841	58	95.7447%	23.0769%	81.8182%	60%	80%
Surf	Std	-0.6753	61	93.617%	15.3846%	80%	40%	76.6667%
Surf	W01	-0.6461	57	97.8723%	7.6923%	79.3103%	50%	78.3333%
Surf	W02	-0.6439	57	100%	7.6923%	79.661%	100%	80%
Surf	Height	-0.7132	59	97.8723%	0%	77.9661%	0%	76.6667%
Surf	Gx	-0.6694	57	97.8723%	7.6923%	79.3103%	50%	78.3333%
Surf	Gy	-0.6445	61	97.8723%	0%	77.9661%	0%	76.6667%
Per	Mean	-0.6593	57	93.617%	7.6923%	78.5714%	25%	75%
Per	Std	-0.6757	61	97.8723%	0%	77.9661%	0%	76.6667%
Per	W01	-0.6998	55	97.8723%	7.6923%	79.3103%	50%	78.3333%
Per	W02	-0.7075	58	97.8723%	0%	77.9661%	0%	76.6667%
Per	Height	-0.6704	57	95.7447%	0%	77.5862%	0%	75%
Per	Gx	-0.6251	57	97.8723%	15.3846%	80.7018%	66.6667%	80%
Per	Gy	-0.6271	59	97.8723%	0%	77.9661%	0%	76.6667%
Mean	Std	-0.7063	58	95.7447%	15.3846%	80.3571%	50%	78.3333%
Mean	W01	-0.6907	55	95.7447%	7.6923%	78.9474%	33.3333%	76.6667%

Mean	W02	-0.7108	60	100%	0%	78.3333%	NaN	78.3333%
Mean	Height	-0.6586	58	95.7447%	7.6923%	78.9474%	33.3333%	76.6667%
Mean	Gx	-0.7097	54	100%	0%	78.3333%	NaN	78.3333%
Mean	Gy	-0.6721	61	97.8723%	7.6923%	79.3103%	50%	78.3333%
Std	W01	-0.6295	56	100%	7.6923%	79.661%	100%	80%
Std	W02	-0.6406	57	95.7447%	0%	77.5862%	0%	75%
Std	Height	-0.6533	58	97.8723%	15.3846%	80.7018%	66.6667%	80%
Std	Gx	-0.6431	55	100%	7.6923%	79.661%	100%	80%
Std	Gy	-0.6959	60	97.8723%	0%	77.9661%	0%	76.6667%
W01	W02	-0.6699	54	93.617%	0%	77.193%	0%	73.3333%
W01	Height	-0.6829	56	91.4894%	0%	76.7857%	0%	71.6667%
W01	Gx	-0.6369	53	95.7447%	0%	77.5862%	0%	75%
W01	Gy	-0.7549	58	95.7447%	0%	77.5862%	0%	75%
W02	Height	-0.6495	56	93.617%	0%	77.193%	0%	73.3333%
W02	Gx	-0.6794	57	95.7447%	0%	77.5862%	0%	75%
W02	Gy	-0.6618	60	93.617%	0%	77.193%	0%	73.3333%
Height	Gx	-0.6602	58	100%	0%	78.3333%	NaN	78.3333%
Height	Gy	-0.6584	60	97.8723%	0%	77.9661%	0%	76.6667%
Gx	Gy	-0.6898	56	91.4894%	7.6923%	78.1818%	20%	73.3333%

The results shown in **Table 6.3**, imply an overall better classification performance than the KNN's. 31 pairs of attributes gave better results than the KNN, and for every pair of attributes, more than 93% of the *AD* subjects have been correctly identified. However, the *Normal* subjects have not been correctly classified due to the lower number of *Normal* subjects compared to the number of *AD* subjects. Also, the lower values of sensitivity occurred when the testing data happened to be mostly of *AD* subjects.

The results show that the SVM technique gave more accurate classification results. Additionally, the SVM was a good tool for feature selection since it showed good classification results for a large number of pairs of attributes.

To test the classifier for its performance, all the attributes were used subsequently. As the training patterns and the testing patterns were chosen randomly, the program was run several times (50 times) and only the results of 10 distinct trials were kept as seen in **Table 6.4**. Another trial was executed by taking the first 61 patterns as training data and the rest 60 patterns as testing data. The classification performance accuracy slightly decreased to the range 70%- 78.3333% and the *AD* subjects were correctly identified into the range 91.3043% -100%. However, the *Normal* subjects were not all the time identified, which is due to the smaller number of *Normal* subjects. In this trial, all the testing data was classified as *AD* (**Table 6.5**).

Table 6.4 SVM classification performance results using all the attributes

Trials	SN	SP	PPV	NPV	ACC
1	91.3043%	7.1429%	76.3636%	20.0000%	71.6667%
2	93.4783%	7.1429 %	76.7857%	25.0000 %	73.3333%
3	97.8261%	0 %	76.2712%	0%	75.0000%
4	93.4783%	0%	75.4386%	0%	71.6667%
5	100%	0%	76.6667%	NaN	76.6667%
6	93.4783%	0%	75.4386%	0%	71.6667%
7	95.6522%	0%	75.8621%	0%	73.3333%
8	95.6522%	7.1429%	77.193%	33.3333%	75%
9	91.3043%	0%	75%	0%	70%
10	97.8261%	7.1429%	77.5862%	50%	76.6667%
11	100%	0%	'78.3333%	NaN	78.3333%

Table 6.5 Class labels of the SVM classification technique using all the attributes

Pattern	Clinical label	Output label	Pattern	Clinical label	Output label	Pattern	Clinical label	Output label
1	'AD'	'AD'	42	'AD'	'Normal'	83	'AD'	'AD'
2	'AD'	'AD'	43	'AD'	'AD'	84	'AD'	'AD'
3	'AD'	'AD'	44	'AD'	'AD'	85	'AD'	'AD'
4	'AD'	'AD'	45	'AD'	'AD'	86	'AD'	'AD'
5	'AD'	'AD'	46	'AD'	'AD'	87	'AD'	'AD'
6	'AD'	'AD'	47	'AD'	'AD'	88	'AD'	'AD'
7	'AD'	'AD'	48	'AD'	'AD'	89	'AD'	'AD'
8	'Normal'	'AD'	49	'AD'	'AD'	90	'Normal'	'Normal'
9	'AD'	'AD'	50	'Normal'	'AD'	91	'AD'	'AD'
10	'AD'	'AD'	51	'Normal'	'AD'	92	'AD'	'AD'
11	'AD'	'AD'	52	'Normal'	'AD'	93	'AD'	'AD'
12	'AD'	'AD'	53	'Normal'	'AD'	94	'Normal'	'Normal'
13	'AD'	'AD'	54	'AD'	'AD'	95	'AD'	'AD'
14	'AD'	'AD'	55	'AD'	'AD'	96	'AD'	'AD'
15	'AD'	'AD'	56	'AD'	'AD'	97	'AD'	'AD'
16	'AD'	'AD'	57	'AD'	'AD'	98	'AD'	'AD'
17	'Normal'	'AD'	58	'AD'	'AD'	99	'AD'	'AD'
18	'AD'	'AD'	59	'Normal'	'AD'	100	'AD'	'AD'
19	'Normal'	'AD'	60	'Normal'	'AD'	101	'AD'	'AD'
20	'AD'	'AD'	61	'Normal'	'Normal'	102	'AD'	'AD'
21	'AD'	'AD'	62	'AD'	'AD'	103	'AD'	'AD'
22	'Normal'	'AD'	63	'AD'	'AD'	104	'AD'	'AD'
23	'AD'	'AD'	64	'AD'	'AD'	105	'AD'	'AD'
24	'AD'	'AD'	65	'Normal'	'Normal'	106	'Normal'	'Normal'
25	'AD'	'AD'	66	'Normal'	'Normal'	107	'Normal'	'Normal'
26	'AD'	'AD'	67	'Normal'	'Normal'	108	'AD'	'AD'
27	'Normal'	'AD'	68	'AD'	'AD'	109	'Normal'	'Normal'
28	'AD'	'AD'	69	'Normal'	'Normal'	110	'Normal'	'Normal'
29	'AD'	'AD'	70	'AD'	'AD'	111	'AD'	'AD'
30	'AD'	'AD'	71	'AD'	'AD'	112	'AD'	'AD'
31	'AD'	'AD'	72	'AD'	'AD'	113	'Normal'	'Normal'
32	'AD'	'AD'	73	'AD'	'AD'	114	'AD'	'AD'
33	'AD'	'AD'	74	'Normal'	'Normal'	115	'AD'	'AD'
34	'AD'	'AD'	75	'AD'	'AD'	116	'AD'	'AD'
35	'AD'	'AD'	76	'Normal'	'Normal'	117	'AD'	'AD'
36	'AD'	'AD'	77	'AD'	'AD'	118	'Normal'	'Normal'
37	'AD'	'AD'	78	'AD'	'AD'	119	'AD'	'AD'
38	'AD'	'AD'	79	'AD'	'AD'	120	'AD'	'AD'
39	'AD'	'AD'	80	'AD'	'AD'	121	'Normal'	'Normal'
40	'Normal'	'AD'	81	'AD'	'AD'			
41	'Normal'	'AD'	82	'AD'	'AD'			

To further test if better results could be achieved if a set of attributes were selected instead of selecting all the attributes, a dimensionality reduction algorithm is to be applied on the original data before applying the classification algorithms.

PCA was used for dimensionality reduction in order to achieve a more consistent attributes. PCA dimensionality reduction is performed by projecting the original data onto the eigenvectors corresponding to the largest eigenvalues of the covariance matrix. After reading the data which consists of n = 121 patterns of dimensionality m = 35 (121 row vectors of 35 attributes each), the PCA algorithm performs the following steps:

1. Center the data by subtracting the average value of the data from each row vector to ensure that every feature has zero mean.

2. Calculate the covariance matrix using the formula: $\Sigma = \frac{1}{m}\sum_1^m x(i)x(i)^T$ where the x vector has zero mean. Matlab statement : Sigma = cov(x);

3. Find the eigenvalues and their equivalent eigenvectors from the covariance matrix.

Matlab eigenvalues and eigenvectors function was applied on the covariance matrix as follows:

```
[V, D] = eig(Sigma);
```

where D is a diagonal matrix of eigenvalues and V is a full matrix whose columns are the corresponding eigenvectors so that Sigma*V = V*D.

The resulting principal components are the eigenvectors with largest eigenvalues. More precisely, as PCA defines a set of principal components (PC), the first PC indicates the direction of the greatest variability. The second PC, which is perpendicular to the first one, would demonstrate the second greatest variability and so on. Consequently, the old system of coordinates would be rotated to fit the new axis which correspond to the principal components, and the original data points would be projected into this new lower dimensional coordinates system while preserving as much information as possible. Therefore, the projection into the principal components will not only reduce the original data but it will also change the values and the position of the vectors of attributes into the data matrix.

In order to apply the above concept, the next two Matlab statements were added:

```
VReduced = V(:, k);

PCReduced = Vreduced*x;
```

where k is the reduced number of attributes, x is the standardized feature vector, and PCReduced is the reduced feature vector.

As the MATLAB *eig* function orders the principal components from last to first, the diagonal of matrix *D* contains the variances for the principal components. The proportion of each principal component relative to the total variance can be determined by extracting and normalizing the diagonal of matrix *D*:

```
cumsum(flipud(diag(D))) / sum(diag(D))
```

The following values were obtained:

0.4727, 0.6734, 0.7556, 0.8052, 0.8400, 0.8691, 0.8946, 0.9185, 0.9353, 0.9487, 0.9574, 0.9658, 0.9725, 0.9784, 0.9823, 0.9856, 0.9881, 0.9904, 0.9922, 0.9938, 0.9953, 0.9965, 0.9974, 0.9982, 0.9988, 0.9992, 0.9996, 0.9999, 1.0000, 1.0000, 1.0000, 1.0000, 1.0000, 1.0000 and 1.0000.

The first principal component contains 47.27% of the total variance of the original data, the first two principal components together contain 67.34% of the total variance of the original data, the first three principal components together contain 75.56% … etc. In addition to this, all of the variance can be reached using the first 29, 30, 31, 32, 33, 34 or 35 principal components together.

After applying the PCA dimension reduction technique, the KNN technique was reapplied on different ordered sets of features. The results are summarized in Appendix C. The performance parameters, including the accuracy, decreased slightly. It was also noticed that the accuracy was mostly non relevant when the first principal component was chosen. This is obvious, given that the variance is less important when only this principal component was chosen. However, in general, PCA did not work well. All the principal components might be needed to ensure a variability of the variance in the data.

Additionally, when the SVM classification technique was reapplied, after using the PCA dimension reduction technique on the same different ordered sets, the performance has generally remained the same with a few exceptions. From Appendix C, it can be noticed that more than 78% of the data was correctly identified and all the *AD* subjects were correctly identified.

Other techniques of classification have been used but did not produce better results due to the choice of the vector of attributes and the small size of the database.

The first one is the Bayesian Network which is a decision making model. This model makes decisions based on the dependencies of the previous conditions using probability theory. It takes a set of particular inputs and based on the probability of the occurrence of the inputs in the training set, it makes a decision of where the new input is supposed to be classed. The highest probability calculated for a particular output based on the inputs is the decision that is finally chosen as the output [12], [36], [41].

The second method is the neural network which is a learning system that creates interconnecting artificial neurons. It receives an input through its input layer and produces an output that depends on the architecture of the neural network, the values of the initial weights and the number of hidden layers. The hidden layer contains a set of neurons that have various weights to produce and output from each neuron. Once each neuron has a result, those results are aggregated with another set of weights to produce the final output [12].

Chapter 7

Conclusions and Future Work

The main contribution of this thesis was to develop a method based on different image processing and pattern recognition techniques that classifies a subject as AD or normal with a success rate of more than 75% measured with respect to clinical tests.

Image segmentation is very useful for MRI medical images, to detect regions of interest for radiotherapy planning and brain damage diagnostic. Several segmentation methods have been adapted and applied on *2D* medical images. Different algorithms have been tested on real medical data from ADNI database. The system was used to extract the ventricles from the middle slice of each patient's radiographic brain image, and these results were visually inspected and compared to the patient`s condition. Thresholding was applied to remove unnecessary regions and edge detection techniques as well as region segmentations were used to extract the ventricles. The results of the segmentation techniques including Otsu global thresholding, Canny Edge Detection, Niblack local thresholding and the Active contour, have been applied on 121 ADNI data including AD subjects and normal subjects. Active contours showed better results by extracting the exact region of the ventricles for the majority of the images.

After extracting the region of the ventricles of each slice, we characterized the extracted region using a vector of 35 attributes such as the surface area and the perimeter of the shape. Those attributes were stored into a large data matrix where each row corresponds to one pattern and each column corresponds to one attribute.

Once the database of 121 patterns was constructed, the data was classified using the KNN clustering technique for every pair of attributes at the beginning, and then, for the total 35 attributes. A maximum accuracy has been obtained using the pair of attributes *Surface* vs. *Gx*

(81.6667%) and a minimum accuracy for the pair *Gx* vs. *Gy* (53.3333%). When using all the attributes, 76.6667% of classification accuracy was achieved.

The SVM technique was applied in order to get a feature selection and better classification's accuracy. The SVM classifier showed more accurate results with a minimum accuracy of 73.3333% and a maximum accuracy of 80% depending of the chosen pair of attributes. The *AD* subjects were identified most of the time. However, the *Normal* subjects were less likely identified. By using all the attributes, all the *AD* subjects have been identified but the percentage of the *Normal* subjects dropped to 0% for many cases.

PCA dimensionality reduction technique was applied in order to get the attributes that could classify the data more accurately. The same classification techniques were reapplied to the new reduced dimensionality data. The SVM gave similar results. However, PCA did not work well for the KNN classifier. Indeed, even though the attributes were modified and rearranged from the most discriminate to the less discriminate, it was noticed that all the principal components were required to ensure a variability of the variance in the data.

The shape and the size of the ventricle, given by the vector of attributes, had a big impact on the final result. The SVM classifier was opted since it showed better accuracy compared to the KNN. A framework was successfully created to continue researching the use of SVMs towards classifying medical images. This framework has potential to make potent predictions based exclusively on the properties of a patient's hippocampus based on the strength of the SVM's ability to classify objects.

Future research could focus on investigating other regions that might be more affected, coupled with appropriate features set in order to characterize the new regions. It is suggested to focus on reducing the cost and improving the precision of the algorithms by using more intelligent algorithms such as adaptive seeds initialization and image registration in order to get initial contour for the active contour segmentation method. It is also suggested to focus on the enhancement of the classification algorithms and adding more data to the system.

Bibliography

[1] M. L. Schroeter, T. Stein, N. Maslowski and J. Neumann, "Neural correlates of Alzheimer's disease and mild cognitive impairment A meta-analysis including 1351 patients," *NeuroImage,* vol. 47, no. 4, pp. 1196-1206, 2009.

[2] M. Kirby and L. Sirovich, "Application of the Karhunen-Loeve Procedure for the Characterization of Human Faces," *IEEE Transactions on Pattern Analysis and Machine Intelligence* , vol. 12, no. 1, pp. 103-108 , January 1990.

[3] S. B. Kotsiantis, I. D. Zaharakis and P. E. Pintelas, "Machine learning: a review of classification and combining techniques," *Artificial Intelligence Review,* vol. 26, no. 3, pp. 159-190, November 2006.

[4] W. Wu, Y. Mallet, B. Walczak, W. Penninckx, D. L. Massart, S. Heuerding and F. Erni, "Comparison of regularized discriminant analysis linear discriminant analysis and quadratic discriminant analysis applied to nir data," *Analytica Chimica Acta,* vol. 329, no. 3, pp. 257-265, 20 August 1996.

[5] R. A. Fisher, "The use of multiple measurements in taxonomic problems," *Annals of Eugenics,* vol. 7, no. 2, pp. 179-188, September 1936.

[6] I. T. JOLLIFFE, Principal Component Analysis, New York: Springer-Verlag, 1986.

[7] J. C. Dunn, "Well separated clusters and optimal fuzzy-partitions," *Journal of Cybernetics,* vol. 4, no. 1, pp. 95-104 , January 1974.

[8] T. Hastie and R. Tibshirani, "Discriminant Adaptive Nearest Neighbor Classification," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 18, no. 6, pp. 607-616, June 1996.

[9] J. R. QUINLAN, "Induction of decision trees," *Machine Learning,* vol. 1, no. 1, pp. 81-106, March 1986.

[10] G. Vanderbrug and A. Rosenfeld, "Two-Stage Template Matching," *Computers, IEEE Transactions on,* Vols. C-26, no. 4, pp. 384-393, April 1977.

[11] J. P. Lewis, "Fast template matching," *Vision Interface,* vol. 95, pp. 120-123, 1995.

[12] C. M. Bishop and G. Hinton, Neural networks for pattern recognition, Oxford University Press, 1995.

[13] C. J. C. Burges, "A Tutorial on Support Vector Machines for Pattern Recognition," *Data Mining and Knowledge Discovery* , vol. 2, no. 2, pp. 121-167 , June 1998.

[14] C.-w. Hsu, C.-c. Chang and C.-j. Lin, "A practical guide to support vector classification," Taipei, 2003.

[15] E. Rahm and P. A. Bernstein, "A Survey of Approaches to Automatic Schema Matching," *The VLDB JOURNAL,* vol. 10, no. 4, p. 334–350, December 2001.

[16] R. M. Haralick and L. G. Shapiro, "Image segmentation techniques," *Computer Vision, Graphics, and Image Processing,* vol. 29, no. 1, pp. 100-132, 1 January 1985.

[17] N. Otsu, "A threshold selection method from gray-level histograms," *IEEE Transactionn on Systems, Man, and Cybernetics,* vol. 9, no. 1, pp. 62-66, January 1979.

[18] W. Niblack, An Introduction to Digital Image Processing, New Jersey: Prentice Hall, 1986.

[19] J. Canny, "A computational approach to edge detection," *Pattern Analysis and Machine Intelligence, IEEE Transactions on,* Vols. PAMI-8, no. 6, pp. 679 -698, 1986.

[20] V. Caselles, R. Kimmel and G. Sapiro, "Geodesic Active Contours," *Int. J. Comput. Vision,* vol. 22, no. 1, pp. 61-79, 1997.

[21] Kass, Michael, A. Witkin and D. Terzopoulos, "Snakes: Active contour models," *International Journal of Computer Vision,* vol. 1, no. 4, pp. 321-331, 1988.

[22] R. Adams and L. Bischof, "Seeded Region Growing," *Pattern Analysis and Machine Intelligence, IEEE Transactions on,* vol. 16, no. 6, pp. 641-647 , 1994.

[23] S. Horowitz and T. Pavlidi, "Picture Segmentation by a Directed Split and Merge Procedure," in *Proceedings of the 2nd International Joint Conference on Pattern Recognition*, Copenhagen, 1977.

[24] F. Meyer and S. Beucher, "Morphological Segmentation," *Journal of Visual Communication and Image Representation,* vol. 1, no. 1, pp. 21 - 46, 1990.

[25] S. G. Mallat, "A Theory for Multiresolution Signal Decomposition: The Wavelet Representation," *Pattern Analysis and Machine Intelligence, IEEE Transactions on,* vol. 11, no. 7, pp. 674-693, 1989.

[26] S. Beucher and F. Meyer, "The morphological approach to segmentation: the watershed transformation," *OPTICAL ENGINEERING,* vol. 34, pp. 433-481, 1992.

[27] M. A. Gonzalez, G. J. Meschino and V. L. Ballarin, "Solving the over segmentation problem in applications of Watershed Transform," *Journal of Biomedical Graphics and Computing,* vol. 3, no. 3, pp. 29-40, 2013.

[28] R. POLIKAR, "The Wavelet Tutorial," [Online]. Available: http://users.rowan.edu/~polikar/WAVELETS/WTpart1.html.

[29] K. Rajpoot, N. Rajpoot and M. J. Turner, "Hyperspectral colon tissue cell classification," in *Medical Imaging, Proceedings of SPIE*, 2004.

[30] D. Zhang and G. Lu, "Review of shape representation and description techniques," *Pattern Recognition,* vol. 37, no. 1, pp. 1-19, January 2004.

[31] A. K. Jain and B. Chandrasekaran, "39 Dimensionality and sample size considerations in pattern recognition practice," *Classification*

Pattern Recognition and Reduction of Dimensionality, Handbook of Statistics, vol. 2, p. 835–855, 1982.

[32] A. K. Jain and D. Zongker, "Feature selection: evaluation, application, and small sample performance," *Pattern Analysis and Machine Intelligence, IEEE Transactions on,* vol. 19, no. 2, pp. 153-158, February 1997.

[33] P. Pudil, J. Novovičová and J. Kittler, "Floating search methods in feature selection," *Pattern Recognition Letters,* vol. 15, no. 11, pp. 1119-1125, November 1994.

[34] P. Bradley and O. L. Mangasarian, "Feature selection via concave minimization and support vector machines," in *Machine Learning Proceedings of the Fifteenth International Conference*, San Francisco, 1998.

[35] I. Guyon and A. Elisseeff, "An introduction to variable and feature selection," *The Journal of Machine Learning Research* , vol. 3, pp. 1157-1182 , 3 January 2003.

[36] R. O. Duda, P. E. Hart and D. G. Stork, Pattern Classification, New York: Wiley, 2001 .

[37] L. Hermes and J. M. Buhmann, "Feature selection for support vector machines," in *Pattern Recognition, 2000. Proceedings. 15th International Conference on*, Barcelona, 2000.

[38] G. H. John, R. Kohavi and K. Pfleger, "Irrelevant features and the subset selection problem," in *Machine Learning: Proceedings of the Eleventh International Conference*, 1994.

[39] J. M. S. C. O. P. M. P. T. a. V. V. Weston, "Feature selection for svms," in *Advances in Neural Information Processing Systems 13* , Denver, 2000.

[40] J. N. DARROCH and D. RATCLIFF, "Generalized Iterative Scaling for Log-Linear Models," *The Annals of Mathematical Statistics,* vol. 43, no. 5, p. 1470–1480, 1972.

[41] H. Zhang and J. Su, "Naive bayes for optimal ranking," *Journal of Experimental & Theoretical Artificial Intelligence* , vol. 20, no. 2, pp. 79-93, 2008.

[42] V. N. Vapnik, The nature of statistical learning theory, New York: Springer-Verlag New York, Inc., 1995.

[43] M. Filippone, F. Camastra, F. Masulli and S. Rovetta, " 2008 A survey of kernel and spectral methods for clustering," *Pattern Recognition,* vol. 41, pp. 176-190., 2008.

[44] C. Taylor, "Classification and kernel density estimation," *Vistas in Astronomy,* vol. 41, no. 3, pp. 411-417, 1997.

[45] P. Bradley, U. Fayyad and C. Reina, "Scaling clustering algorithms to large databases," in *4th International Conference on Knowledge Discovery & Data Mining (KDD98)*, California, 1998.

[46] T. Zhang, R. Ramakrishnan and M. Livny, "BIRCH: A New Data Clustering Algorithm and Its Applications," *Data Mining and Knowledge Discovery,* vol. 1, no. 2, pp. 141-182, 1997.

[47] L. Breiman, "Bagging predictors," *Machine Learning,* vol. 24, no. 2, pp. 123-140, auguste 1996.

[48] Alzheimer's association, "What We Know Today About Alzheimer's Disease," [Online]. Available: http://www.alz.org/research/science/alzheimers_disease_causes.asp. [Accessed 21 July 2013].

[49] MedecineNet, "Magnetic Resonance Imaging (MRI Scan)," [Online]. Available: http://www.medicinenet.com/mri_scan/article.htm. [Accessed 21 July 2013].

[50] Alzheimer's Disease Neuroimaging Initiative (ADNI), "Sharing Alzheimer's Research Data with the World," [Online]. Available: http://adni.loni.ucla.edu/. [Accessed 21 07 2013].

[51] S. S. Mirra, A. Heyman, D. McKeel, S. Sumi, B. Crain, L. Brownlee, F. Vogel, J. Hughes, G. Van Belle, L. Berg and P. C. neuropathologists, "The Consortium to Establish a Registry for Alzheimer's Disease (CERAD) Part II. Standardization of the neuropathologic assessment of Alzheimer's disease," *Neurology,* vol. 41, no. 4, pp. 479-479, April 1991.

[52] W. E. Klunk, H. Engler, A. Nordberg, Y. Wang, G. Blomqvist, D. P. Holt, M. Bergström, I. Savitcheva, G.-F. Huang, S. Estrada, B. Ausén, M. L. Debnath, J. Barletta, J. C. Price and J. Sandell, "Imaging brain amyloid in Alzheimer's disease with Pittsburgh Compound-B," *Annals of Neurology,* vol. 55, no. 3, pp. 306-319, March 2004.

[53] J. C. Price, W. E. Klunk, B. J. Lopresti, X. Lu, J. A. Hoge, S. K. Ziolko, D. P. Holt, C. C. Meltzer, S. T. DeKosky and C. A. Mathis, "Kinetic modeling of amyloid binding in humans using PET imaging and Pittsburgh Compound-B," *Journal of Cerebral Blood Flow \& Metabolism,* vol. 25, no. 11, pp. 1528-1547, 8 June 2005.

[54] M. L. Schroeter, T. Stein, N. Maslowski and J. Neumann, "Neural correlates of Alzheimer's disease and mild cognitive impairment: A systematic and quantitative meta-analysis involving 1351 patients," *NeuroImage,* vol. 47, no. 4, pp. 1196-1206, 1 October 2009.

[55] T. Kapur, W. E. L. Grimson, W. M. Wells and R. Kikinis, "Segmentation of brain tissue from magnetic resonance images," *Medical Image Analysis,* vol. 1, no. 2, pp. 109-127, June 1996.

[56] W. M. I. Wells, W. E. L. Grimson, R. Kikinis and F. A. Jolesz, "Adaptive segmentation of MRI data," *Medical Imaging, IEEE Transactions on,* vol. 15, no. 4, pp. 429-442, August 1996.

[57] K. Held, E. Kops, B. Krause, W. I. Wells, R. Kikinis and H. Muller-Gartner, "Markov random field segmentation of brain MR images," *Medical Imaging, IEEE Transactions on,* vol. 16, no. 6, pp. 878-886, December 1997.

[58] D. L. Pham, C. Xu and J. L. Prince, "Current methods in medical image segmentation," *Annual review of biomedical engineering,* vol. 2, pp. 315-337, 2000.

[59] M. Balafar, A. Ramli, M. Saripan and S. Mashohor, "Review of brain MRI image segmentation methods," *Artificial Intelligence Review,* vol. 33, no. 3, pp. 261-274, March 2010.

[60] Y. Zhang, M. Brady and S. Smith, "Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm," *Medical Imaging, IEEE Transactions on,* vol. 20, no. 1, pp. 45-57, January 2001.

[61] B. Fischl, D. H. Salat, E. Busa, M. Albert, M. Dieterich, C. Haselgrove, A. van der Kouwe, R. Killiany, D. Kennedy, S. Klaveness, A. Montillo, N. Makris, B. Rosen and A. M. Dale, "Whole Brain Segmentation: Automated Labeling of Neuroanatomical Structures in the Human Brain," *Neuron,* vol. 33, no. 3, pp. 341-355, 31 January 2002.

[62] K. Van Leemput, F. Maes, D. Vandermeulen and P. Suetens, "A unifying framework for partial volume segmentation of brain MR images," *Medical Imaging, IEEE Transactions on,* vol. 22, no. 1, pp. 105-119, January 2003.

[63] V. Grau, A. U. J. Mewes, M. Alcaniz, R. Kikinis and S. Warfield, "Improved watershed transform for medical image segmentation using prior information," *Medical Imaging, IEEE Transactions on,* vol. 23, no. 4, pp. 447-458, April 2004.

[64] R. de Boer, F. van der Lijn, H. A. Vrooman, M. W. Vernooij, M. A. Ikram, M. M. Breteler and W. J. Niessen, "Automatic segmentation of brain tissue and white matter lesions in MRI," in *Biomedical Imaging: From Nano to Macro, 2007. ISBI 2007. 4th IEEE International Symposium on*, 2007.

[65] R. De Boer, H. A. Vrooman, F. Van der Lijn, M. W. Vernooij, M. A. Ikram, A. Van der Lugt, M. M. Breteler and W. J. Niessen, "White matter lesion extension to automatic brain tissue segmentation on MRI," *NeuroImage,* vol. 45, no. 4, pp. 1151-1161, 1 May 2009.

[66] Z. Tu, K. L. Narr, P. Dollar, I. Dinov, P. M. Thompson and A. W. Toga, "Brain Anatomical Structure Segmentation by Hybrid Discriminative/Generative Models," *IEEE Transactions on Medical Imaging,* vol. 27, no. 4, pp. 495-508, April 2008.

[67] O. Colliot, G. Chételat, M. Chupin, B. Desgranges, B. Magnin, H. Benali, B. Dubois, L. Garnero, F. Eustache and S. Lehéricy, "Discrimination between Alzheimer Disease, Mild Cognitive Impairment, and Normal Aging by Using Automated Segmentation of the Hippocampus," *Radiology,* vol. 248, no. 1, pp. 194-201, July 2008.

[68] G. McKhann, D. Drachman, M. Folstein, R. Katzman, D. Price and E. M. Stadlan, "Clinical diagnosis of Alzheimer's disease: Report of the NINCDS-ADRDA Work Group under the auspices of Department of Health and Human Services Task Force on Alzheimer's Disease," *Neurology,* vol. 34, no. 7, pp. 939-944, July 1984.

[69] R. C. Petersen, R. Doody, A. Kurz, R. C. Mohs, J. C. Morris, P. V. Rabins, K. Ritchie, M. Rossor, L. Thal and B. Winblad, "Current concepts in mild cognitive impairment," *Archives of Neurology,* vol. 58, no. 12, pp. 1985-1992, 2001.

[70] Y. Zhang, B. Matuszewski, L. Shark and C. Moore, "Medical Image Segmentation Using New Hybrid Level-Set Method," in *BioMedical Visualization, 2008. MEDIVIS '08. Fifth International Conference*, London, 2008.

[71] J. H. Morra, Z. Tu, L. G. Apostolova, A. E. Green, C. Avedissian, S. K. Madsen, N. Parikshak, X. Hua, A. W. Toga, C. R. Jack, M. W. Weiner and P. M. Thompson, "Validation of a fully automated 3D hippocampal segmentation method using subjects with Alzheimer's disease mild cognitive impairment, and elderly controls," *NeuroImage,* vol. 43, no. 1, pp. 59-68, October 2008.

[72] J. Morra, Z. Tu, L. Apostolova, A. Green, A. Toga and P. Thompson, "Comparison of AdaBoost and Support Vector Machines for Detecting Alzheimer's Disease Through Automated Hippocampal Segmentation," *Medical Imaging, IEEE Transactions on,* vol. 29, no. 1, pp. 30-43, January 2010.

[73] Laboratory for Computational Neuroimaging, "FreeSurfer," [Online]. Available: http://surfer.nmr.mgh.harvard.edu. [Accessed 12 March 2014].

[74] D. W. Shattuck, G. Prasad, M. Mirza, K. L. Narr and A. W. Toga, "Online resource for validation of brain segmentation methods.," *NeuroImage,* vol. 45, no. 2, pp. 431-439, 1 April 2009.

[75] University of California, Los Angeles, "Segmentation Validation Engine," [Online]. Available: http://sve.bmap.ucla.edu/. [Accessed 12 March 2014].

[76] S. M. Smith, "Fast robust automated brain extraction," *Human Brain Mapping,* vol. 17, no. 3, p. 143–155, November 2002.

[77] F. Ségonne, A. Dale, E. Busa, M. Glessner, D. Salat, H. Hahn and B. Fischl, "A hybrid approach to the skull stripping problem in MRI," *NeuroImage,* vol. 22, no. 3, pp. 1060-1075, July 2004.

[78] D. W. Shattuck, S. R. Sandor-Leahy, K. A. Schaper, D. A. Rottenberg and R. M. Leahy, "Magnetic Resonance Image Tissue Classification Using a Partial Volume Model," *NeuroImage,* vol. 13, no. 5, pp. 856-876, May 2001.

[79] A. Huang, R. Abugharbieh and R. Tam, "A Hybrid Geometric Statistical Deformable Model for Automated 3-D Segmentation in Brain MRI," *Biomedical Engineering, IEEE Transactions on,* vol. 56, no. 7, pp. 1838-1848, 2009.

[80] G. Helms, B. Draganski, R. Frackowiak, J. Ashburner and N. Weiskopf, "Improved segmentation of deep brain grey matter structures using magnetization transfer (MT) parameter maps," *Neuroimage,* vol. 47, no. 1, pp. 194-198, August 2009.

[81] S. AlZu'bi and A. Amira, "3D medical volume segmentation using hybrid multiresolution statistical approaches," *Advances in Artificial Intelligence - Special issue on machine learning paradigms for modeling spatial and temporal information in multimedia data mining,* vol. 2010, no. 2, pp. 1-15, January 2010.

[82] J. M. Lötjönen, R. Wolz, J. R. Koikkalainen, L. Thurfjell, G. Waldemar, H. Soininen and D. Rueckert, "Fast and robust multi-atlas segmentation of brain magnetic resonance images," *NeuroImage,* vol. 49, no. 3, pp. 2352-2365, 1 February 2010.

[83] Neuroimaging Informatics Tools and Resources Clearinghouse (NITRC), "The Internet Brain Segmentation Repository (IBSR)," [Online].

Available: https://www.nitrc.org/projects/ibsr. [Accessed 13 March 2014].

[84] R. A. Heckemann, S. Keihaninejad, P. Aljabar, D. Rueckert, J. V. Hajnal and A. Hammers, "Improving intersubject image registration using tissue-class information benefits robustness and accuracy of multi-atlas based anatomical segmentation," *Neuroimage,* vol. 51, no. 1, pp. 221--227, 15 May 2010.

[85] R. A. Heckemann, S. Keihaninejad, P. Aljabar, D. Rueckert, J. V. Hajnal and A. Hammers, "Segmenting brain images with MAPER," [Online]. Available: http://www.soundray.org/maper/. [Accessed 13 March 2014].

[86] OPTIMA, "Oxford Project to Investigate Memory and Aging (OPTIMA)," [Online]. Available: http://www.medsci.ox.ac.uk/optima/information-for-patients-and-the-public. [Accessed 13 March 2014].

[87] D. Rivest-Hénault and M. Cheriet, "Unsupervised MRI segmentation of brain tissues using a local linear model and level set," *Magnetic Resonance Imaging,* vol. 29, no. 2, pp. 243-259, February 2011.

[88] K. Van Leemput, F. Maes, D. Vandermeulen and P. Suetens, "Automated model-based tissue classification of MR images of the brain," *Medical Imaging, IEEE Transactions on,* vol. 18, no. 10, pp. 897-908, October 1999.

[89] FMRIB Software Library (FSL) developpers, "SIENA - Analysis of Structural Brain MRI Data," [Online]. Available: http://fsl.fmrib.ox.ac.uk/fsl/fsl-4.1.9/siena/index.html. [Accessed 14 March 2014].

[90] S. M. Smith, Y. Zhang, M. Jenkinson, J. Chen, P. Matthews, A. Federico and N. De Stefano, "Accurate, Robust, and Automated Longitudinal and Cross-Sectional Brain Change Analysis," *NeuroImage,* vol. 17, no. 1, pp. 479-489, September 2002.

[91] C. A. Cocosco, A. P. Zijdenbos and A. C. Evans, "A fully automatic and robust brain MRI tissue classification method," *Medical Image Analysis,* vol. 7, no. 4, pp. 513-527, December 2003.

[92] M. Chupin, E. Gérardin, R. Cuingnet, C. Boutet, L. Lemieux, S. Lehéricy, H. Benali, L. Garnero and O. Colliot, "Fully automatic hippocampus segmentation and classification in Alzheimer's disease and mild cognitive impairment applied on data from ADNI," *Hippocampus,* vol. 19, no. 6, pp. 579-587, June 2009.

[93] S. Chaplot, L. Patnaik and N. Jagannathan, "Classification of magnetic resonance brain images using wavelets as input to support vector machine and neural network," *Biomedical Signal Processing and Control,* vol. 1, no. 1, pp. 86-92, January 2006.

[94] S. Klöppel, C. M. Stonnington, C. Chu, B. Draganski, R. I. Scahill, J. D. Rohrer, N. C. Fox, C. R. Jack, J. Ashburner and R. S. J. Frackowiak, "Automatic classification of MR scans in Alzheimer's disease," *Brain,* vol. 131, no. 3, pp. 681-689, 17 January 2008.

[95] R. Polikar, A. Topalis, D. Parikh, D. Green, J. Frymiare, J. Kounios and C. M. Clark, "An ensemble based data fusion approach for early diagnosis of Alzheimer's disease," *Information Fusion,* vol. 9, no. 1, pp. 83-95, January 2008.

[96] P. Vemuri, J. L. Gunter, M. L. Senjem, J. L. Whitwell, K. Kantarci, D. S. Knopman, B. F. Boeve, R. C. Petersen and C. R. Jack Jr, "Alzheimer's disease diagnosis in individual subjects using structural MR images: validation studies," *Neuroimage,* vol. 39, no. 3, pp. 1186-1197, 1 February 2008.

[97] C. Davatzikos, Y. Fan, X. Wu, D. Shen and S. M. Resnick, "Detection of prodromal Alzheimer's disease via pattern classification of magnetic resonance imaging," *Neurobiology of Aging,* vol. 29, no. 4, pp. 514-523, April 2008.

[98] B. Magnin, L. Mesrob, S. Kinkingnéhun, M. Pélégrini-Issac, O. Colliot, M. Sarazin, B. Dubois, S. Lehéricy and H. Benali, "Support vector machine-based classification of Alzheimer's disease from whole-brain anatomical MRI," *Neuroradiology,* vol. 51, no. 2, pp. 73-83, February 2009.

[99] N. Tzourio-Mazoyer, B. Landeau, D. Papathanassiou, F. Crivello, O. Etard, N. Delcroix, B. Mazoyer and M. Joliot, "Automated Anatomical Labeling of Activations in SPM Using a Macroscopic Anatomical Parcellation of the MNI MRI Single-Subject Brain," *NeuroImage,* vol. 15, no. 1, pp. 273-289, January 2002.

[100] E. C. Robinson, A. Hammers, A. Ericsson, A. D. Edwards and D. Rueckert, "Identifying population differences in whole-brain structural networks: A machine learning approach," *NeuroImage,* vol. 50, no. 3, pp. 910-919, 15 April 2010.

[101] D. Zhang, Y. Wang, L. Zhou, H. Yuan and D. Shen, "Multimodal classification of Alzheimer's disease and mild cognitive impairment," *Neuroimage,* vol. 55, no. 3, pp. 856-867, 1 April 2011.

[102] 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, 15 May 2011.

[103] J. Ashburner, "A fast diffeomorphic image registration algorithm," *NeuroImage,* vol. 38, no. 1, pp. 95-113, 15 October 2007.

[104] J. Ashburner and K. J. Friston, "Unified segmentation," *NeuroImage,* vol. 26, no. 3, p. 839–851, 1 July 2005.

[105] S. M. Nestor, R. Rupsingh, M. Borrie, M. Smith, V. Accomazzi, J. L. Wells, J. Fogarty, R. Bartha and t. A. D. N. Initiative, "entricular enlargement as a possible measure of Alzheimer's disease progression validated using the Alzheimer's disease neuroimaging initiative database," *Brain,* vol. 131, no. 9, pp. 2443-2454, 11 July 2008.

[106] H. Zaidi, T. Ruest, F. Schoenahl and M.-L. Montandon, "Comparative assessment of statistical brain MR image segmentation algorithms and their impact on partial volume correction in PET," *Neuroimage,* vol. 32, no. 4, pp. 1591-1607, 2006.

[107] N. Sharma and L. M. Aggarwal, "Automated medical image segmentation techniques," *J Med Phys.* , vol. 35, no. 1, pp. 3-14, 2010.

[108] H. Knipe, J. Jones and e. al., "Lateral ventricles," [Online]. Available: http://radiopaedia.org/articles/lateral-ventricles. [Accessed 23 May 2014].

[109] C. Li, C. Xu, C. Gui and M. D. Fox, "Distance Regularized Level Set Evolution and Its Application to Image Segmentation," *Image Processing, IEEE Transactions on,* vol. 19, no. 12, pp. 3243-3254, 2010.

[110] P. Baldi, S. Brunak, Y. Chauvin, . C. A. F. Andersen and H. Nielsen, "Assessing the accuracy of prediction algorithms for classification: an overview," *Bioinformatics,* vol. 16, no. 5, pp. 412-424, 2000.

[111] B. E. Boser, I. M. Guyon and V. N. Vapnik, "A training algorithm for optimal margin classifiers," in *Proceedings of the Fifth Annual Workshop on Computational Learning Theory*, 1992.

[112] C. Cortes and V. Vapnik, "Support-vector networks," *Machine learning,* vol. 20, no. 3, pp. 273-297, 1995.

[113] S. B. Kotsiantis, "Supervised Machine Learning: A Review of Classification Techniques.," *Informatica,* vol. 31, pp. 249-268, 2007.

[114] K. Fukunaga and W. L. G. Koontz, "Application of the Karhunen-Loève Expansion to Feature Selection and Ordering," *Computers, IEEE Transactions on,* Vols. C-19, no. 4, pp. 311-318, April 1970.

[115] T. Cover and P. Hart, "Nearest neighbor pattern classification," *IEEE Transactions on Information Theory* , vol. 13, no. 1, pp. 21-27, January 1967.

[116] National Institute for Health and Clinical Excellence, "Supporting people with dementia and their carers in health and social care," London, 2012.

[117] S. C. Neu and A. W. Toga, "Automatic Localization of Anatomical Point Landmarks for Brain Image Processing Algorithms," *Neuroinform,* vol. 6, no. 2, p. 135–148, 2008.

[118] R. Das, "A comparison of multiple classification methods for diagnosis of Parkinson disease," *Expert Systems with Applications,* vol. 37, no. 2, pp. 1568-1572, March 2010.

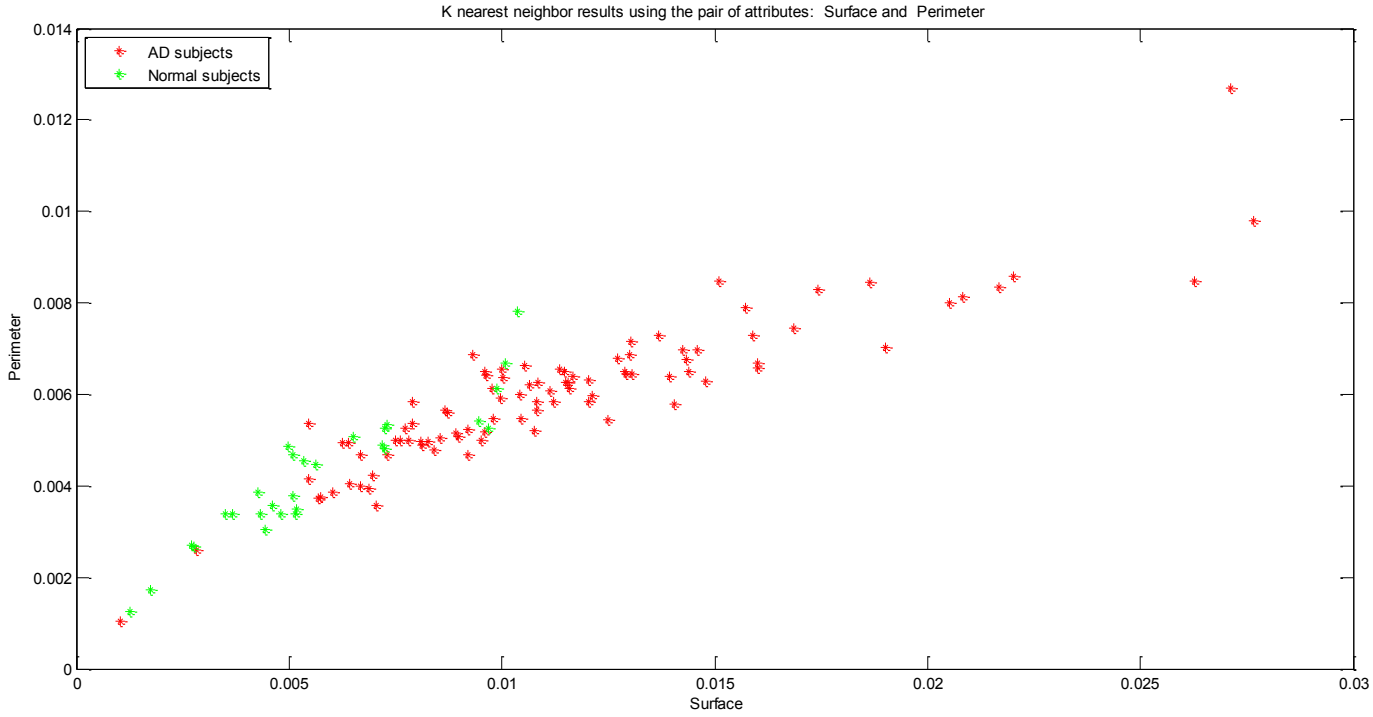
[119] B. Fischl, D. H. Salat, E. Busa, M. Albert, M. Dieterich, C. Haselgrove, A. van der Kouwe, R. Killiany, D. Kennedy, S. Klaveness, A. Montillo, N. Makris, B. Rosen and A. M. Dale, "Whole Brain Segmentation: Automated Labeling of Neuroanatomical Structures in the Human Brain," *Neuron,* vol. 33, no. 3, pp. 341-355, 31 January 2002.

[120] R. C. Gonzalez and R. E. Woods, "Image Processing Place," [Online]. Available: http://www.imageprocessingplace.com/root_files_V3/image_databases.htm.

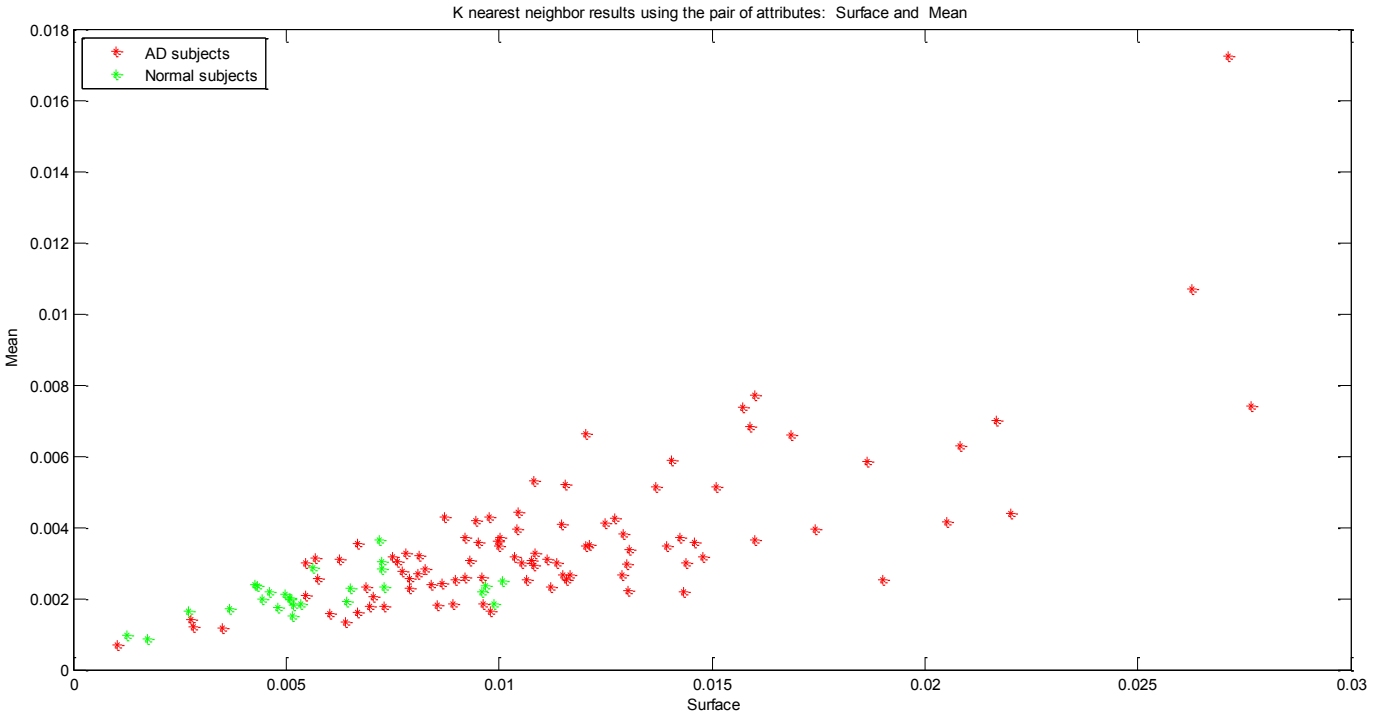
Appendices

Appendix A: Clustering results of pairs of attributes.

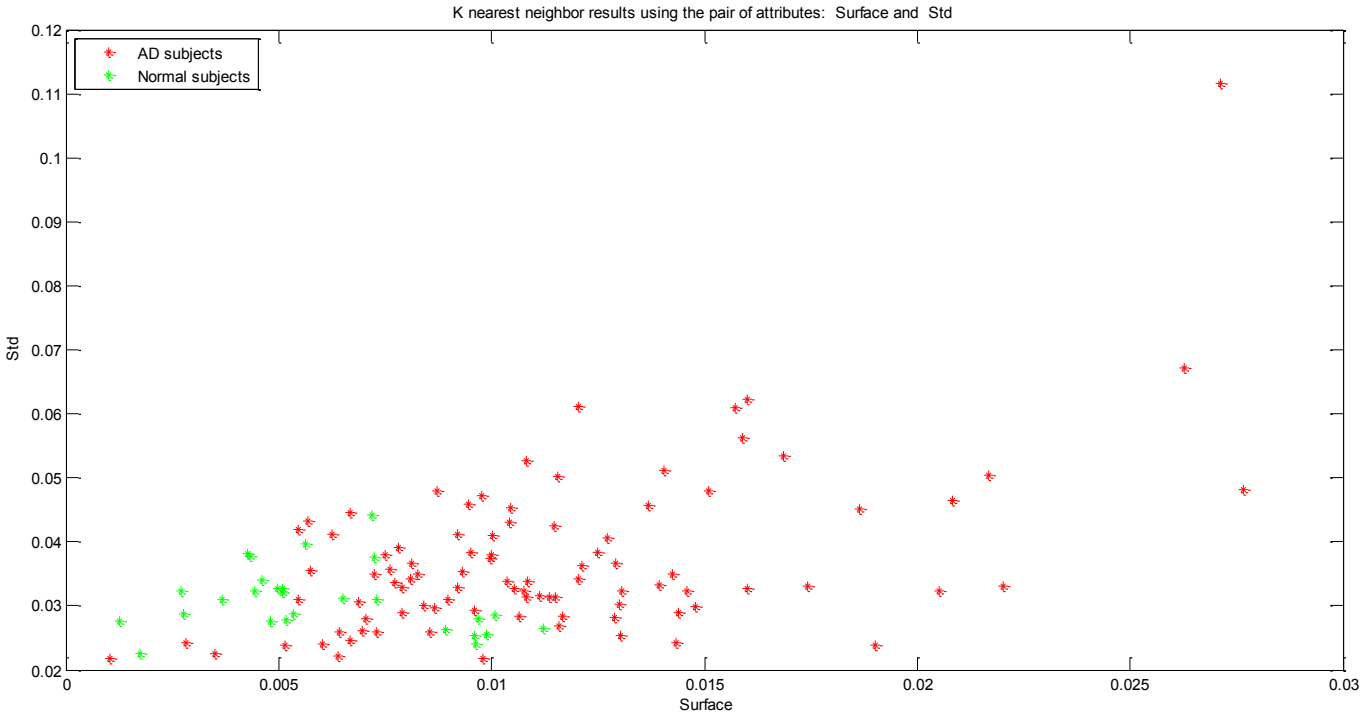
Clustering results using surface and perimeter attributes



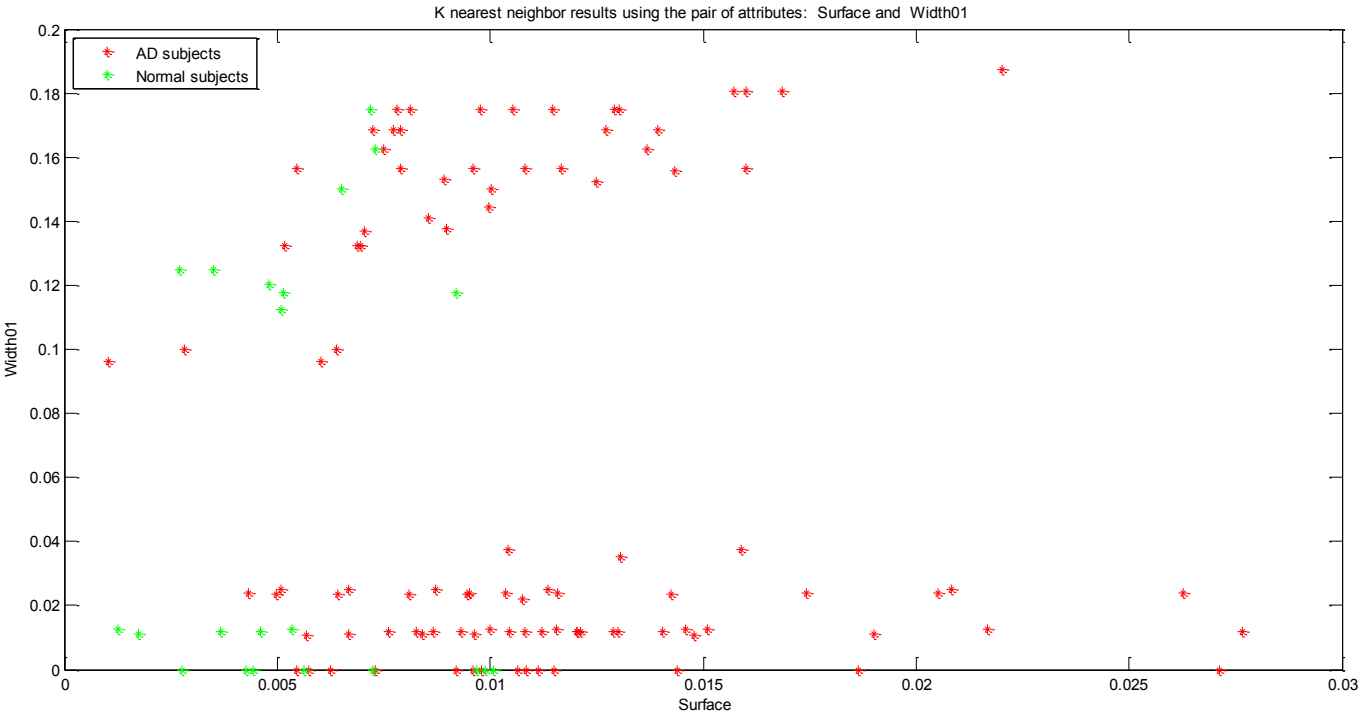
Clustering results using surface and average (mean) attributes



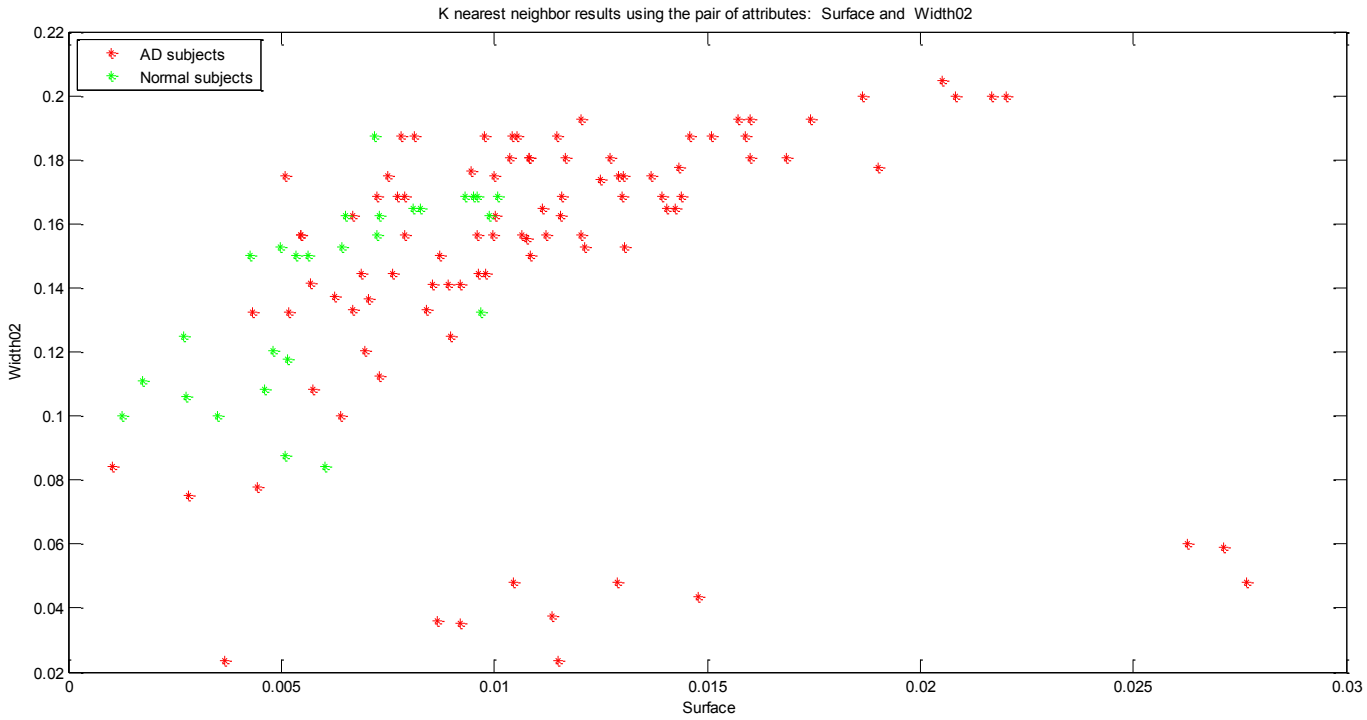
Clustering results using surface and standard deviation attributes



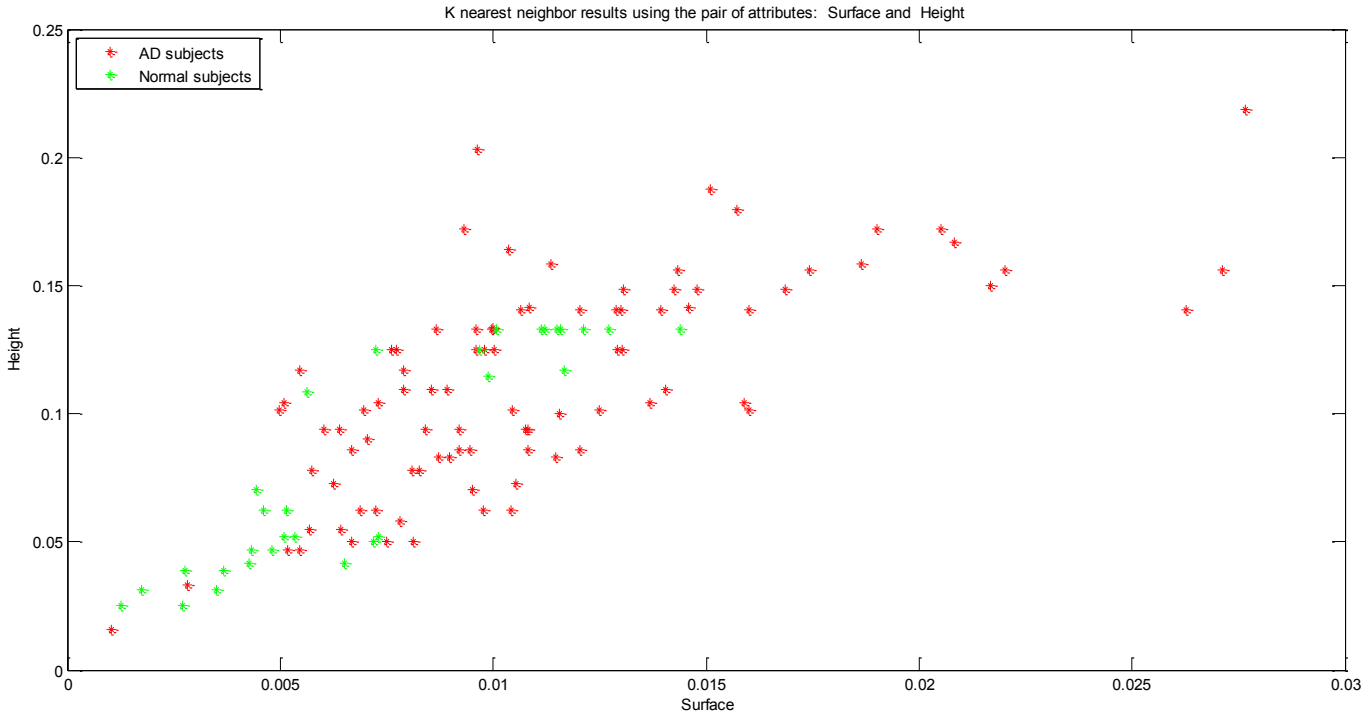
Clustering results using surface and width 1 attributes



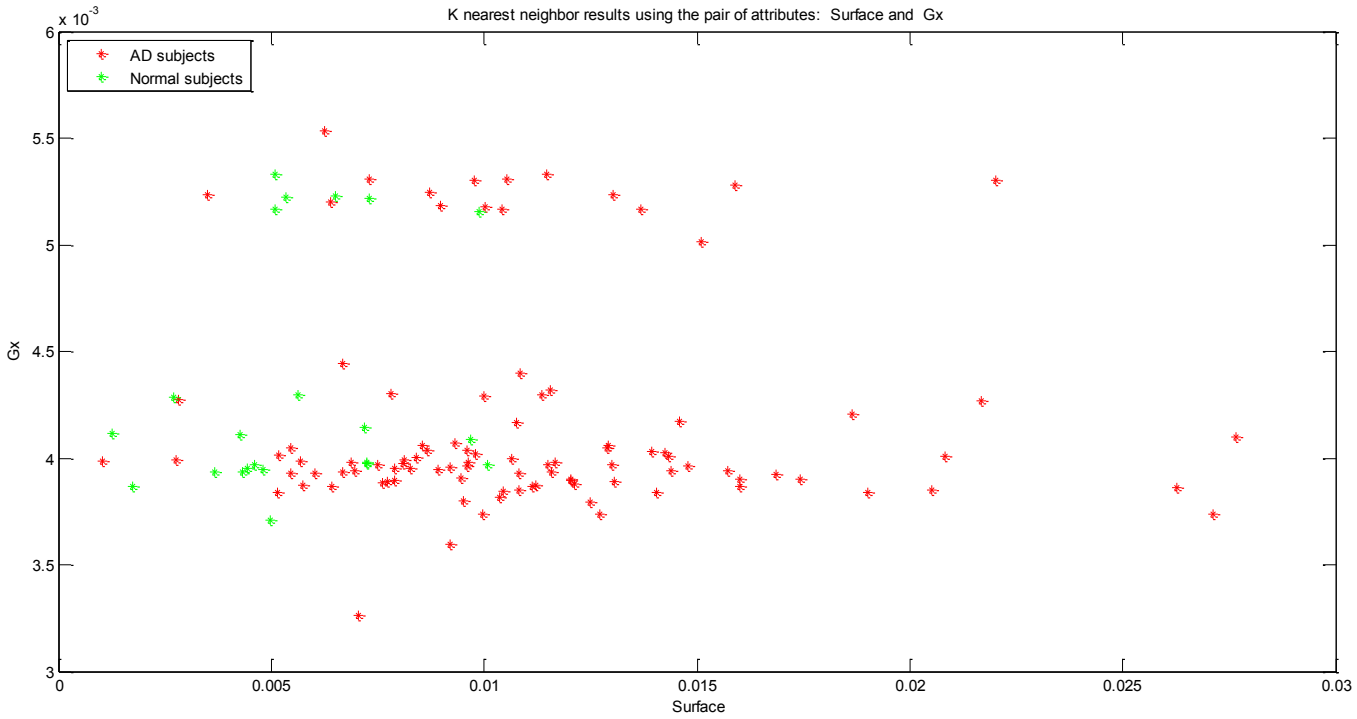
Clustering results using surface and width 2 attributes



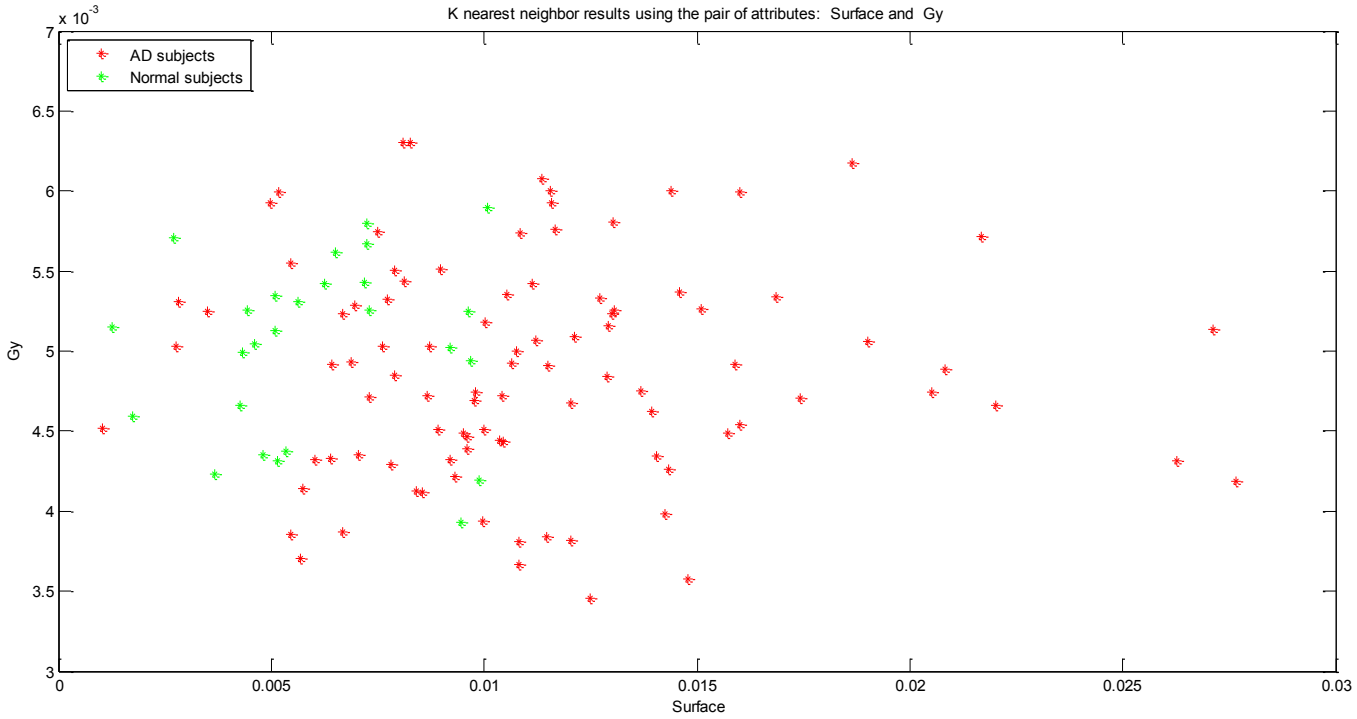
Clustering results using surface and height attributes



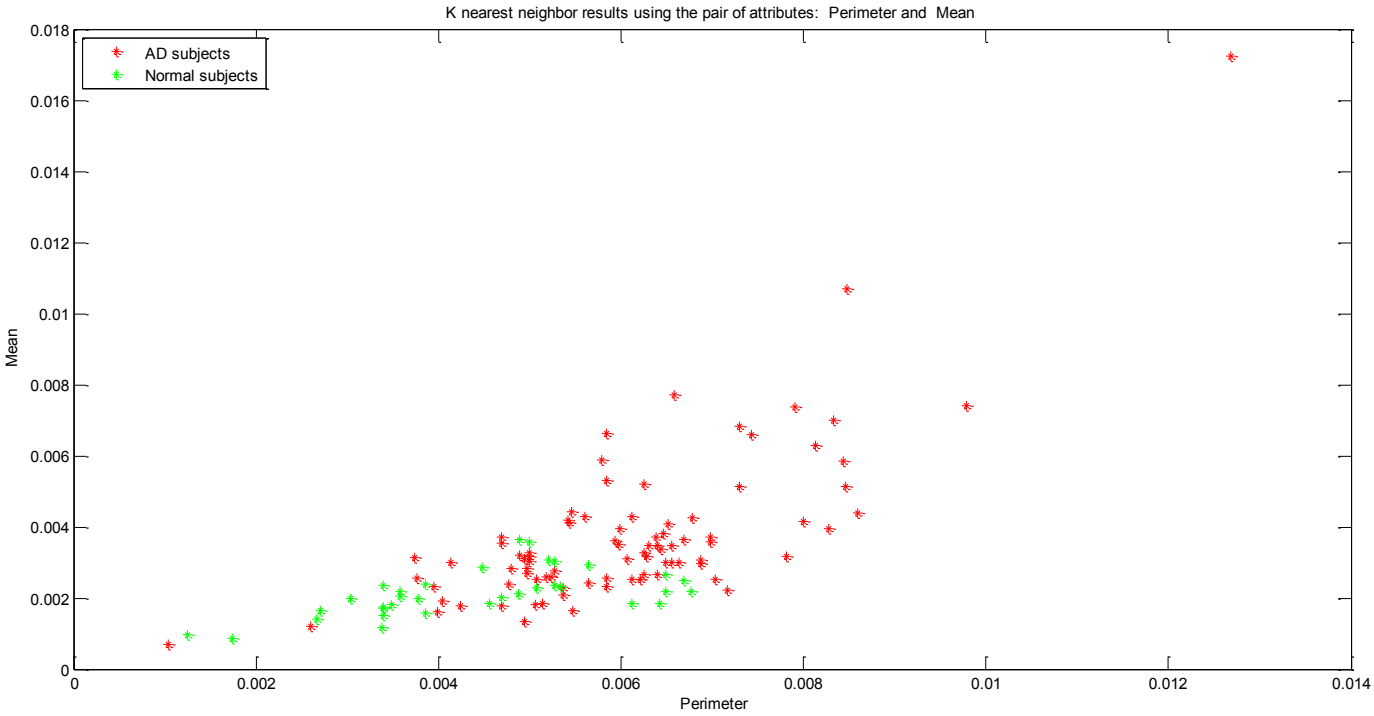
Clustering results using surface and center of gravity Gx attributes



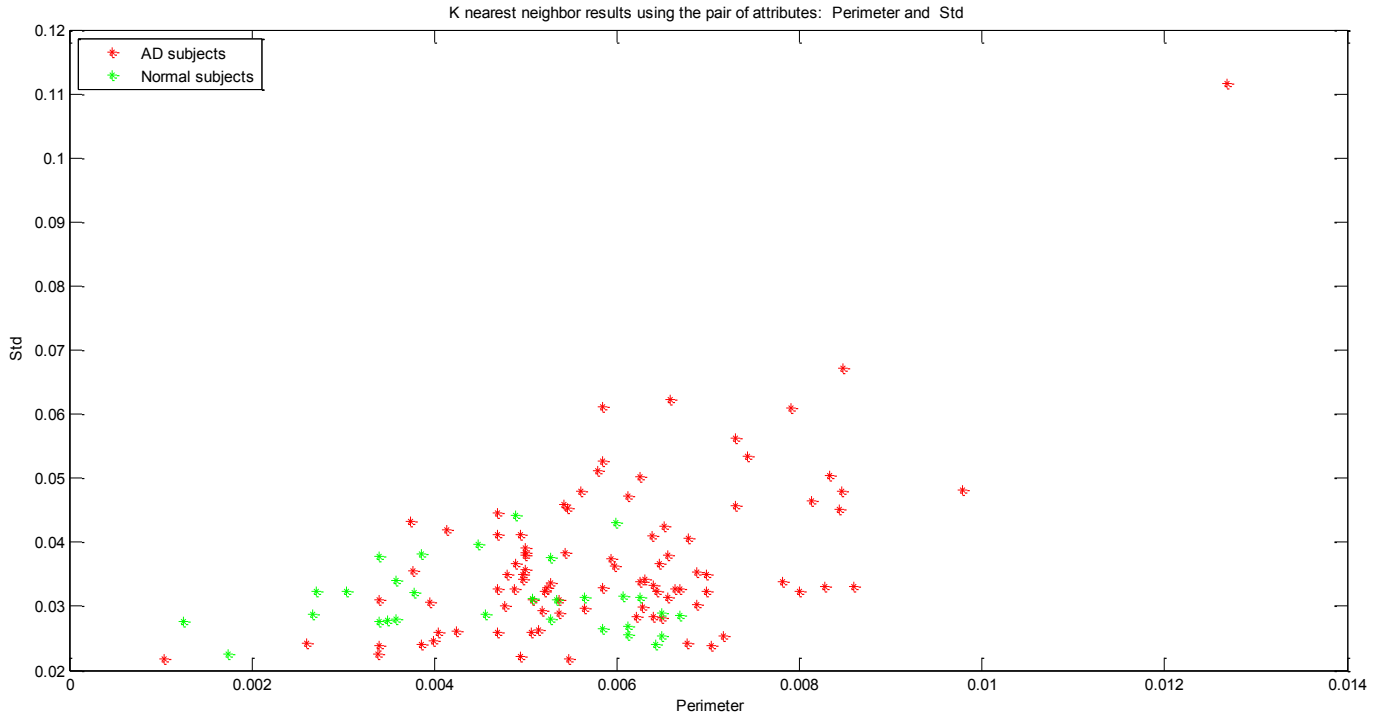
Clustering results using surface and center of gravity Gy attributes



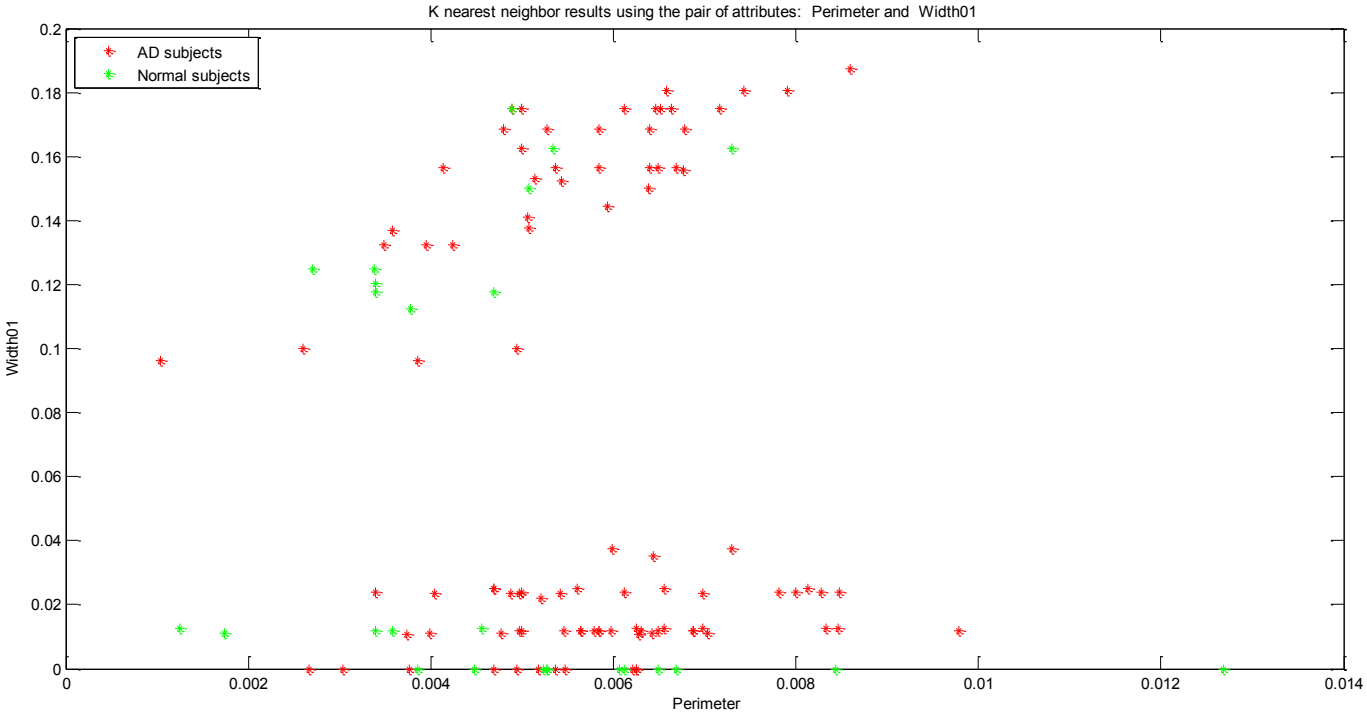
Clustering results using perimeter and average attributes



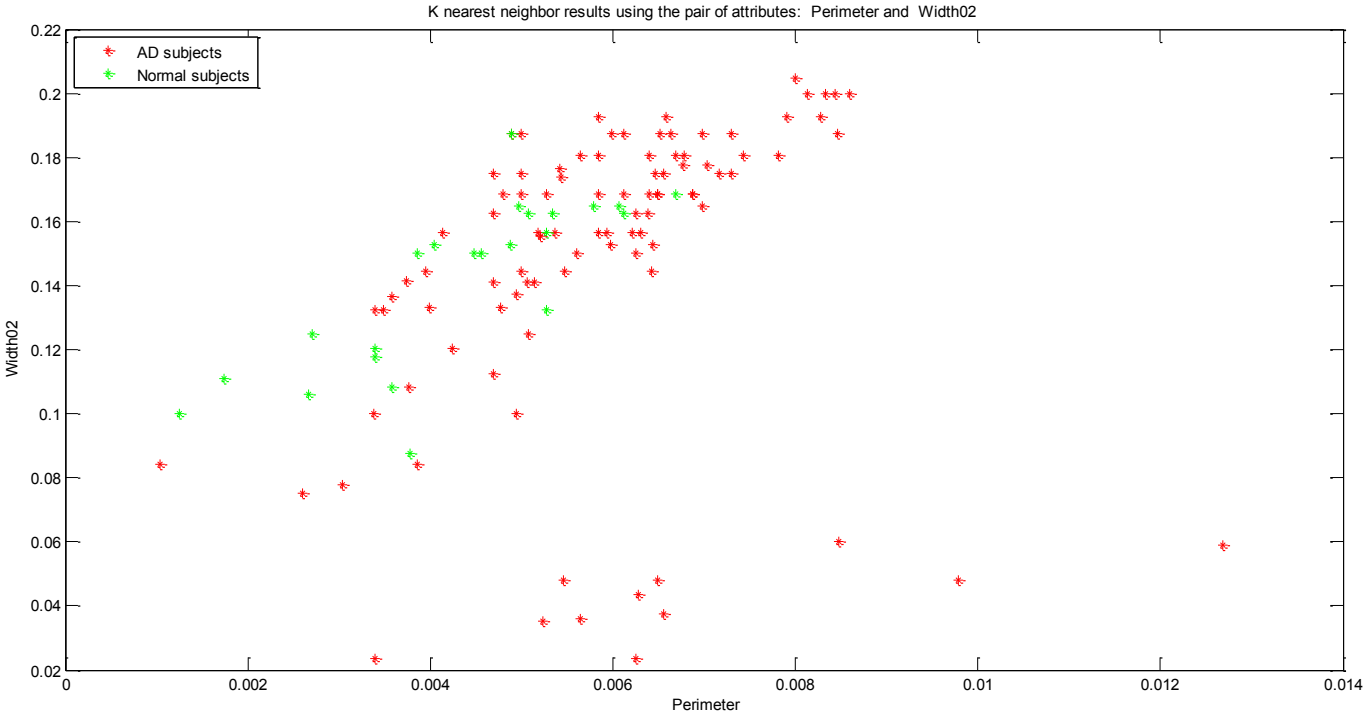
Clustering results using perimeter and standard deviation attributes



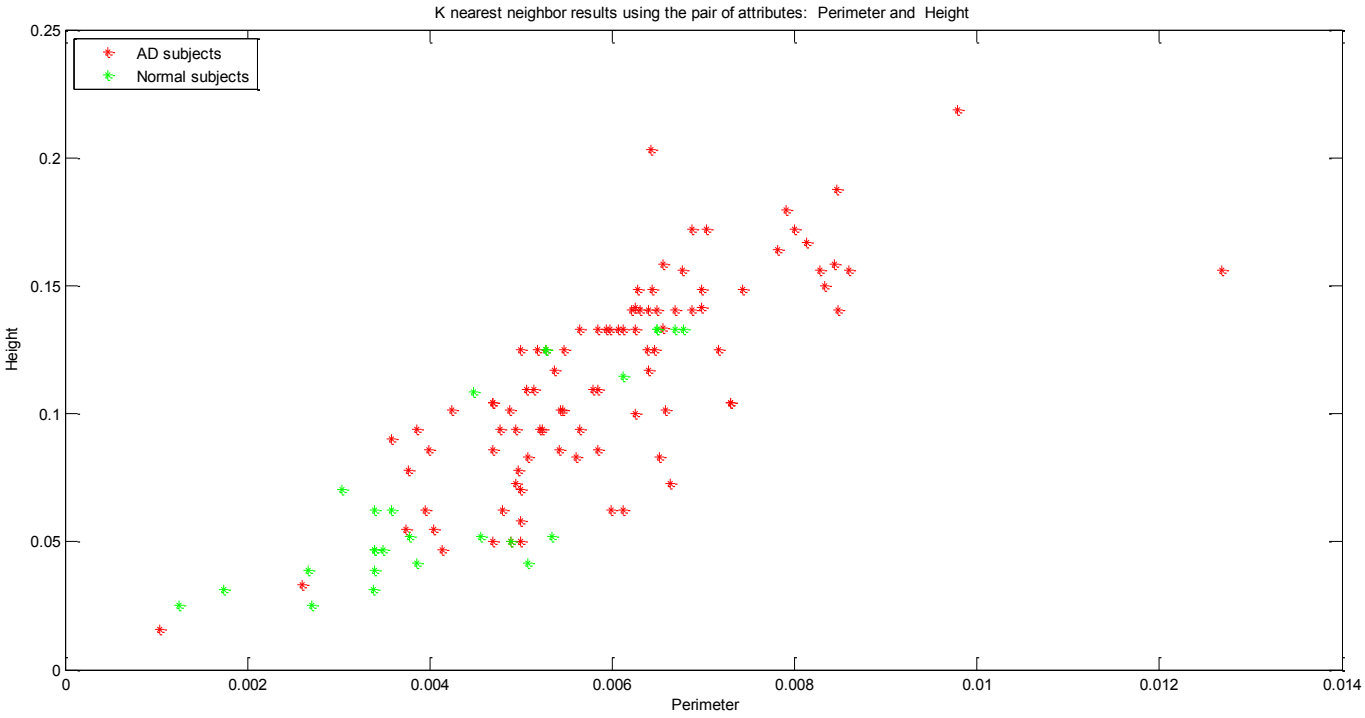
Clustering results using perimeter and Width 1



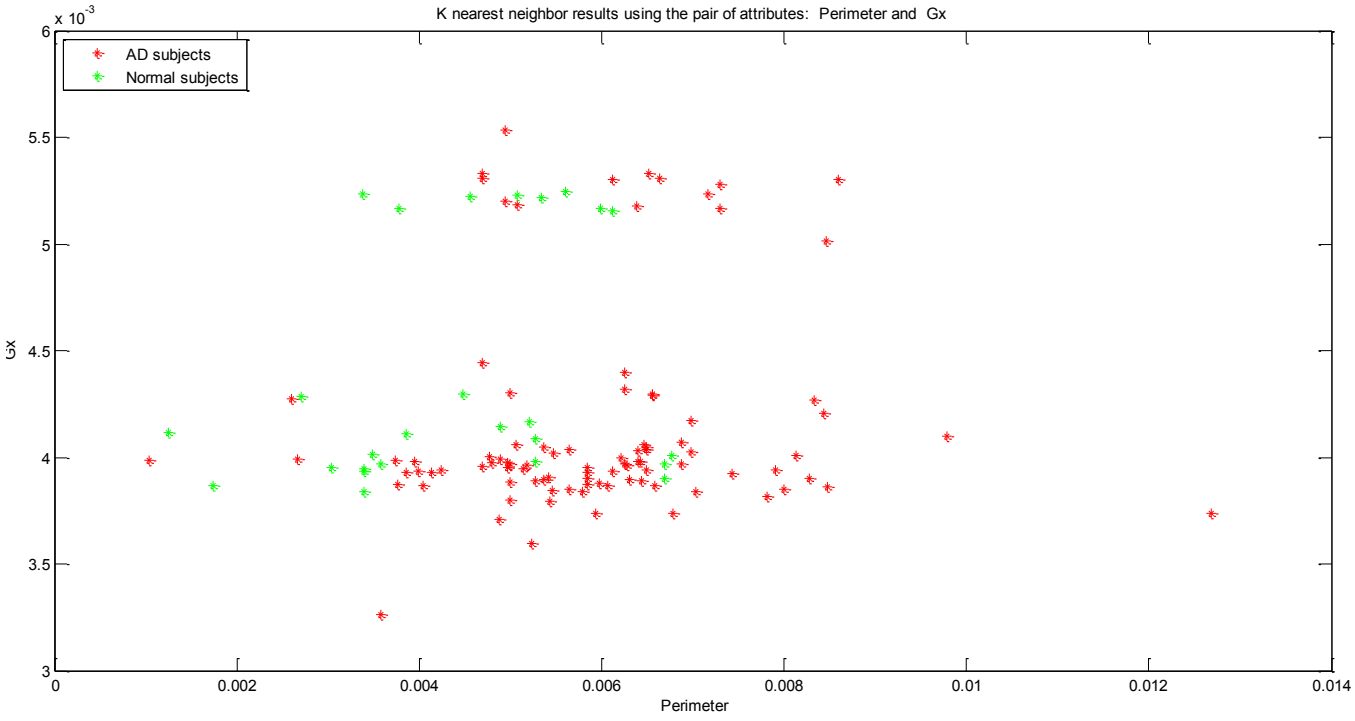
Clustering results using perimeter and width 2 attributes



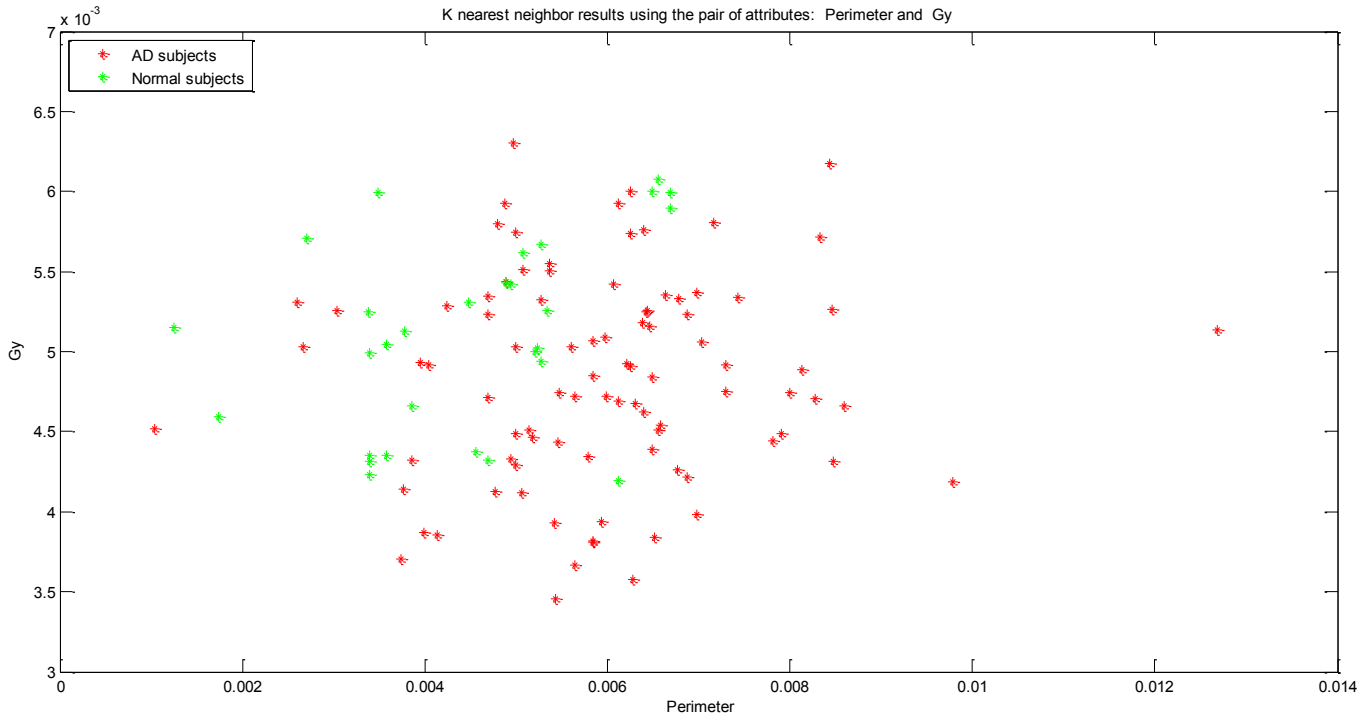
Clustering results using perimeter and height attributes



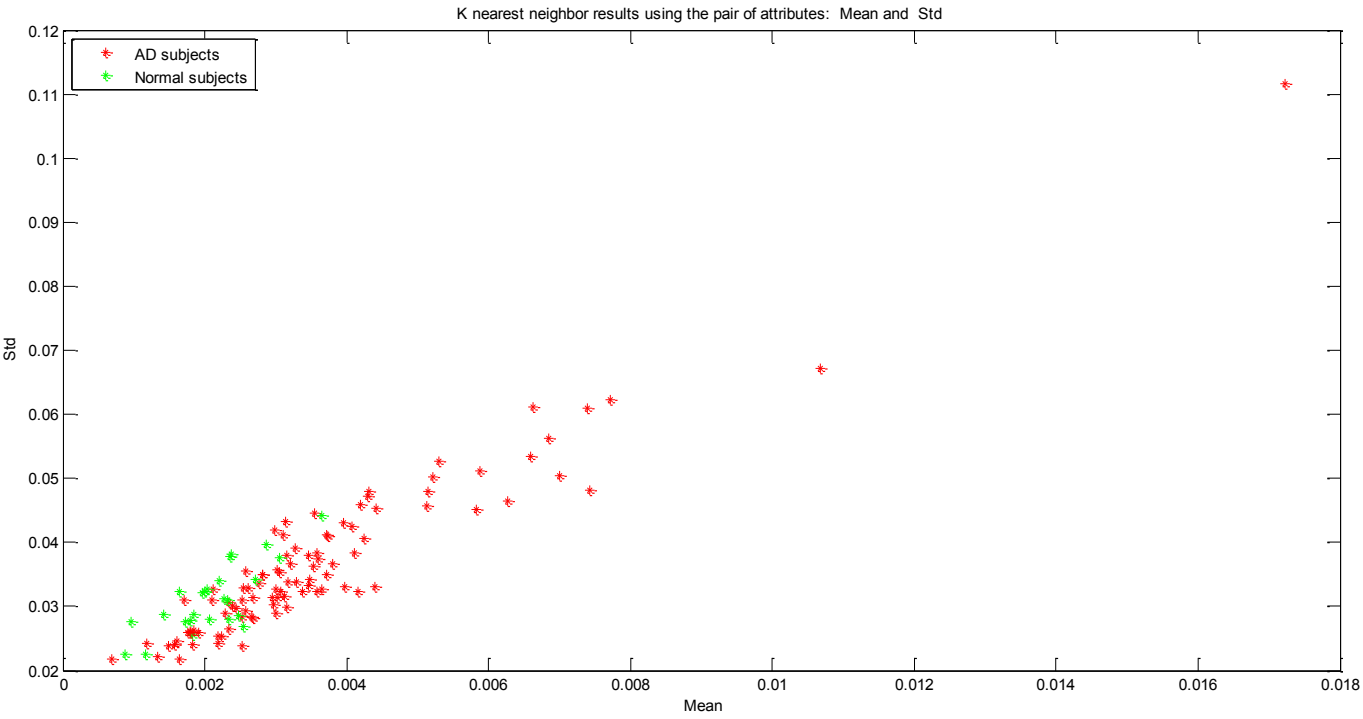
Clustering results using perimeter and center of gravity Gx attributes



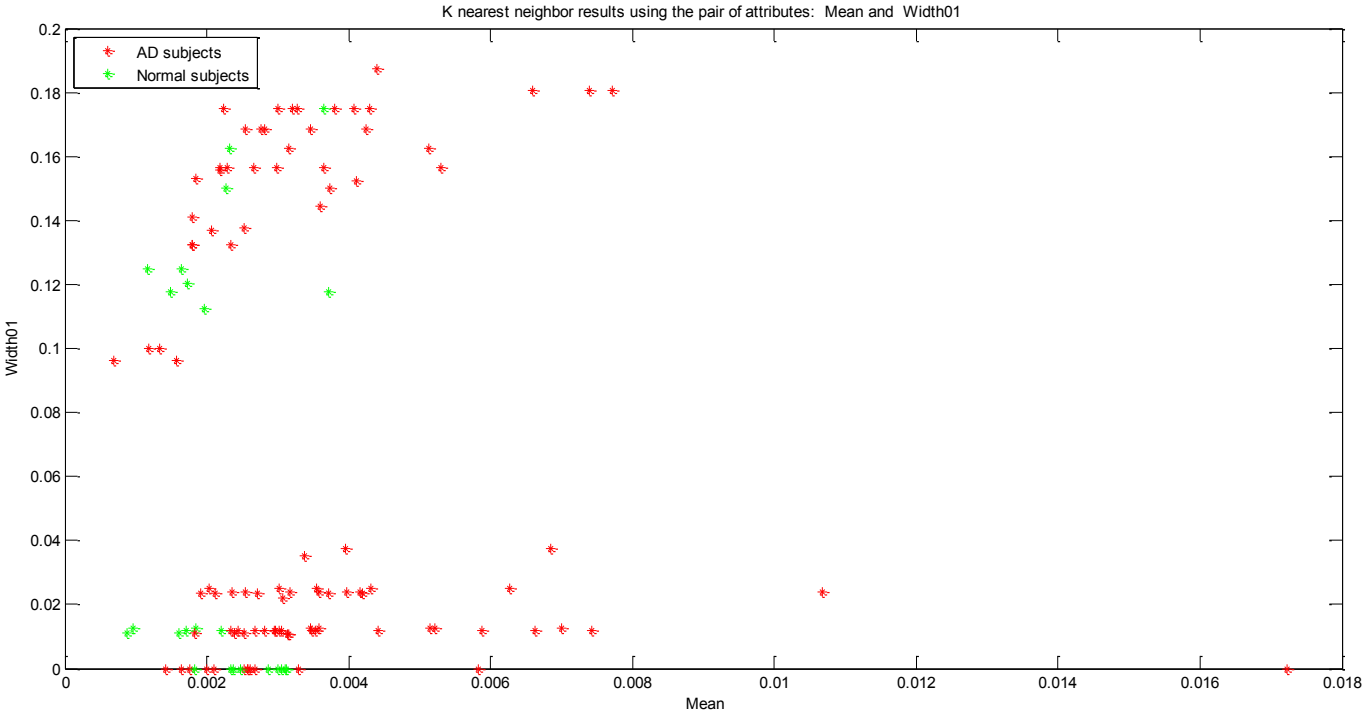
Clustering results using perimeter and center of gravity Gy attributes



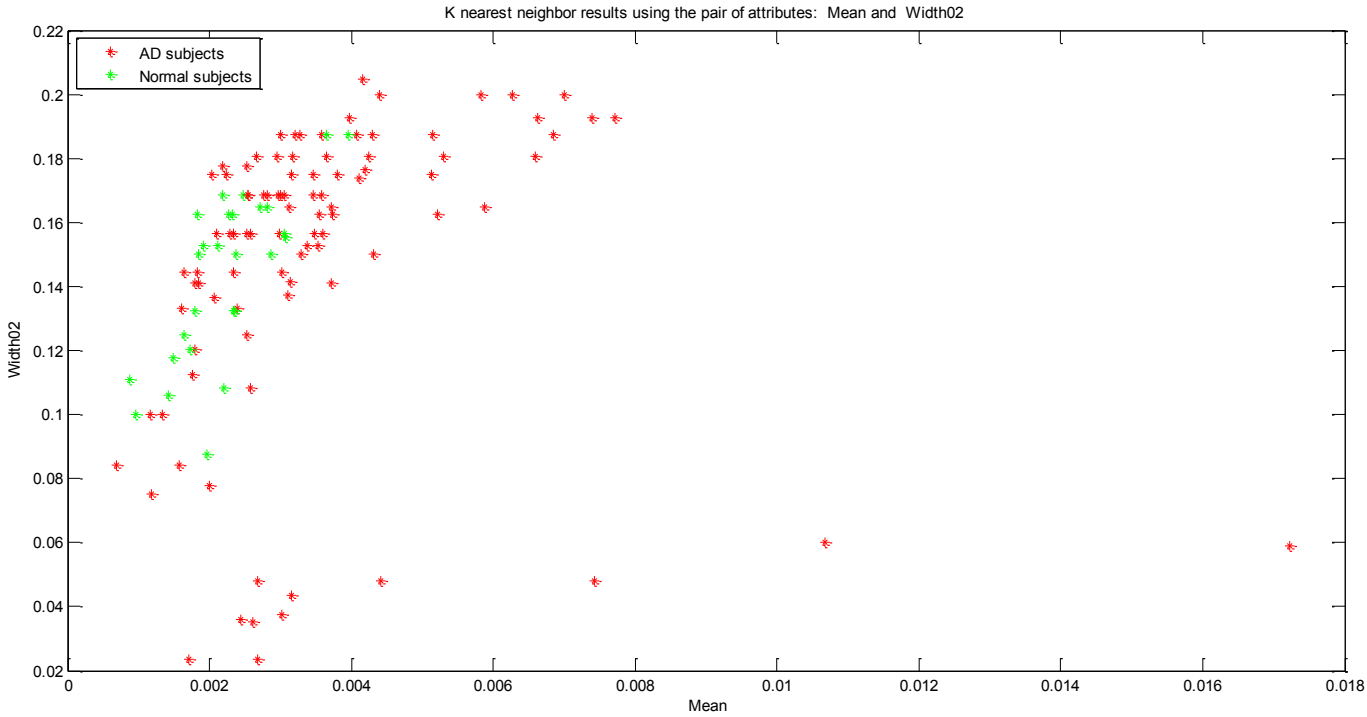
Clustering results using average and standard deviation attributes



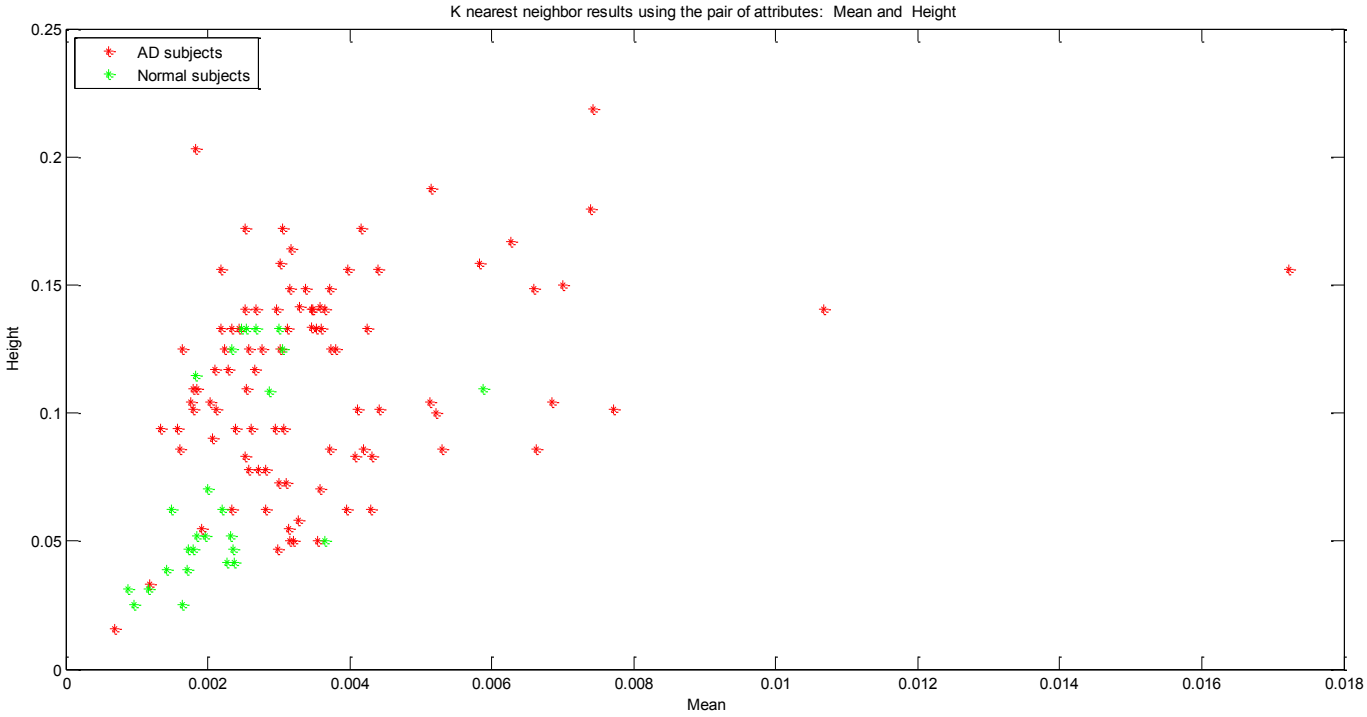
Clustering results using average and width 1 attributes



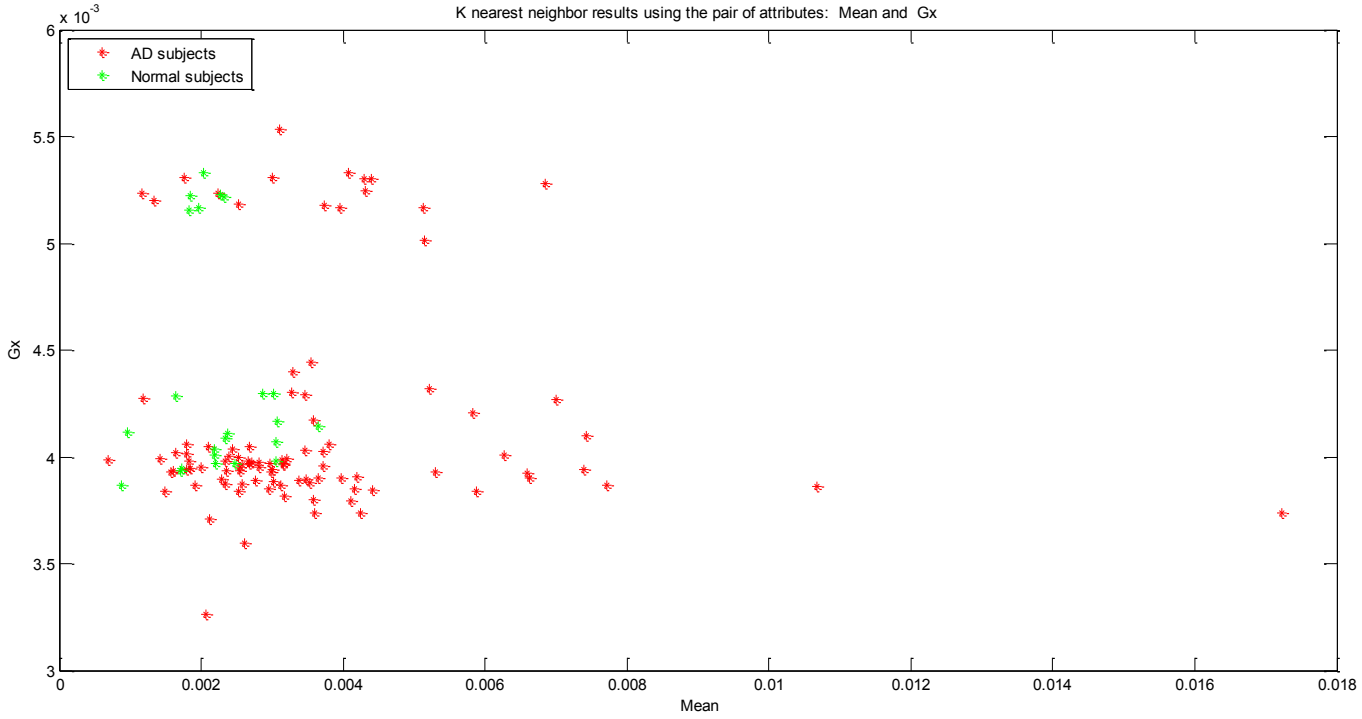
Clustering results using average and width 2 attributes



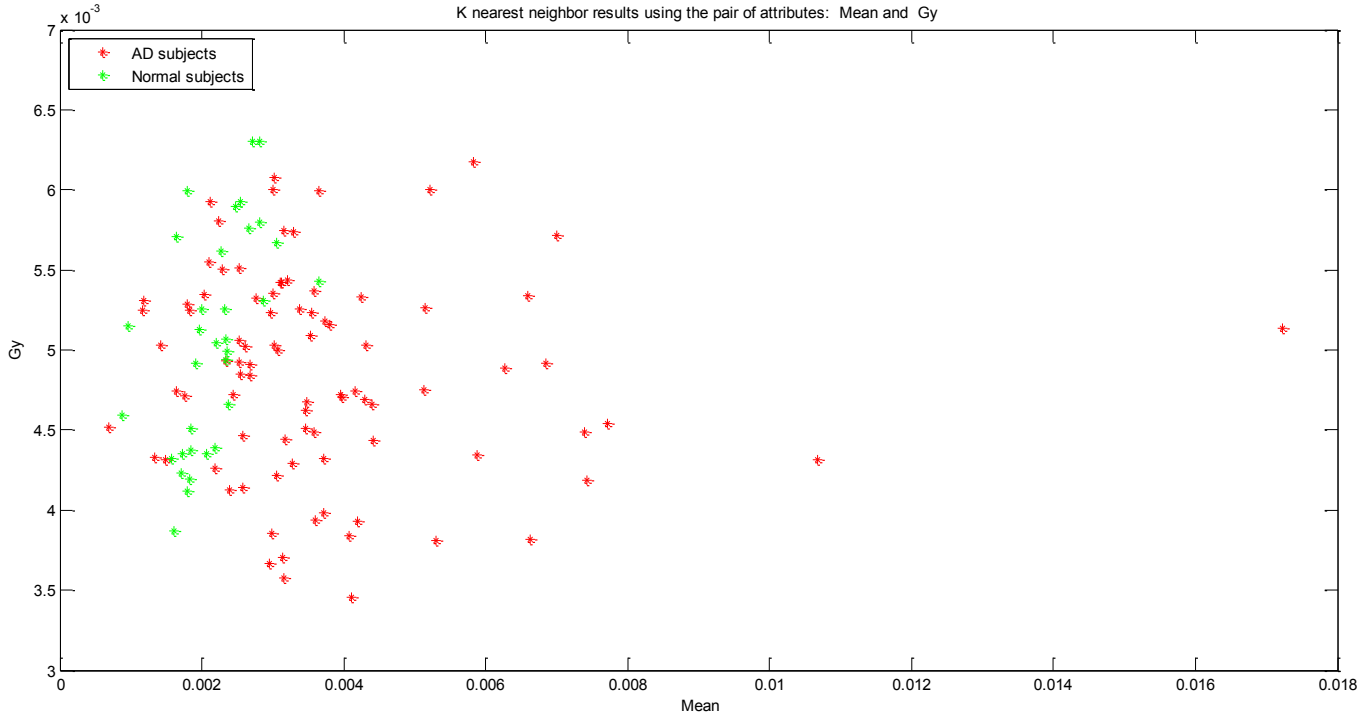
Clustering results using average and height attributes



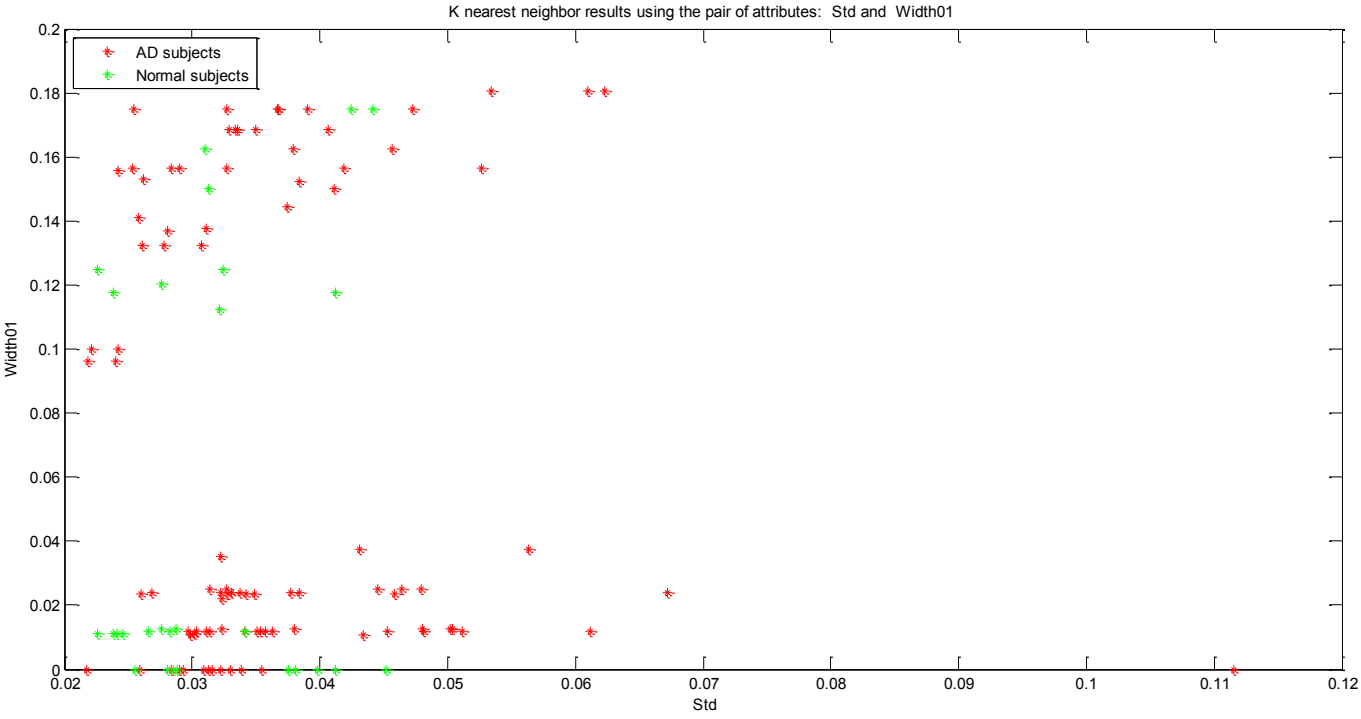
Clustering results using average and center of gravity Gx attributes



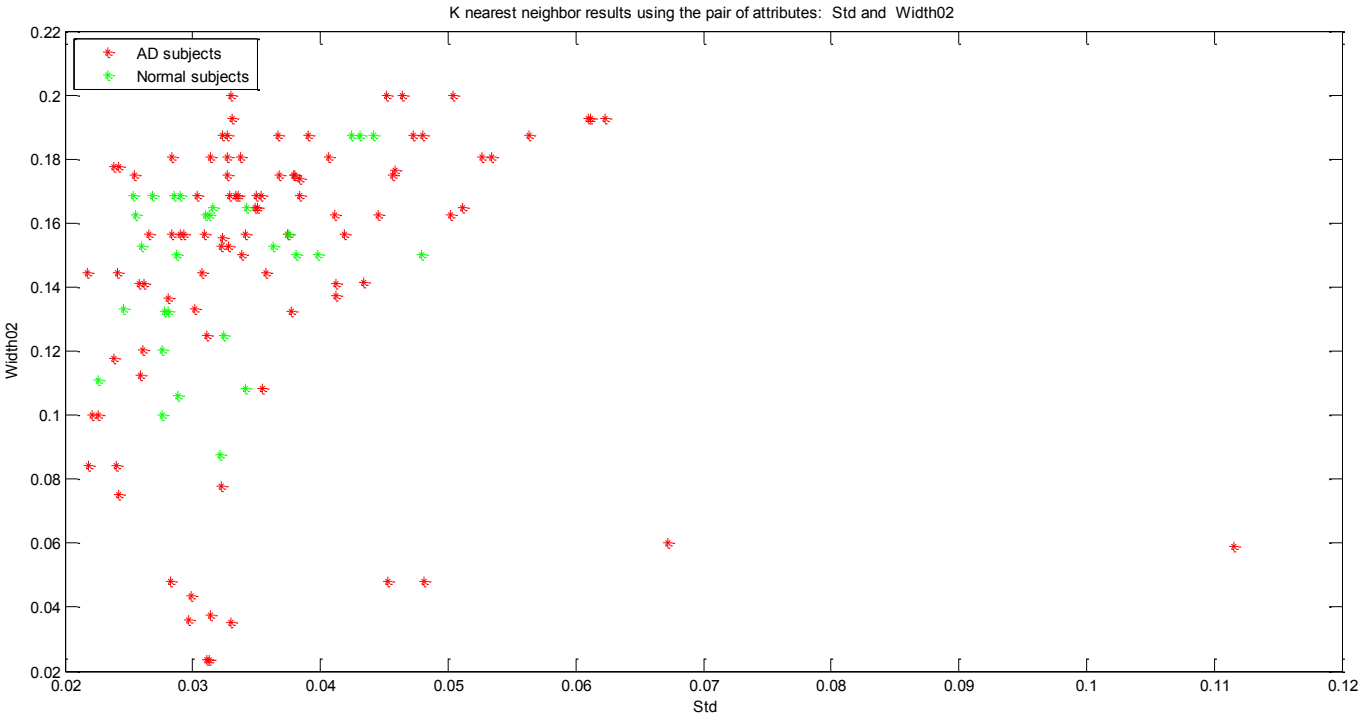
Clustering results using average and center of gravity Gy attributes



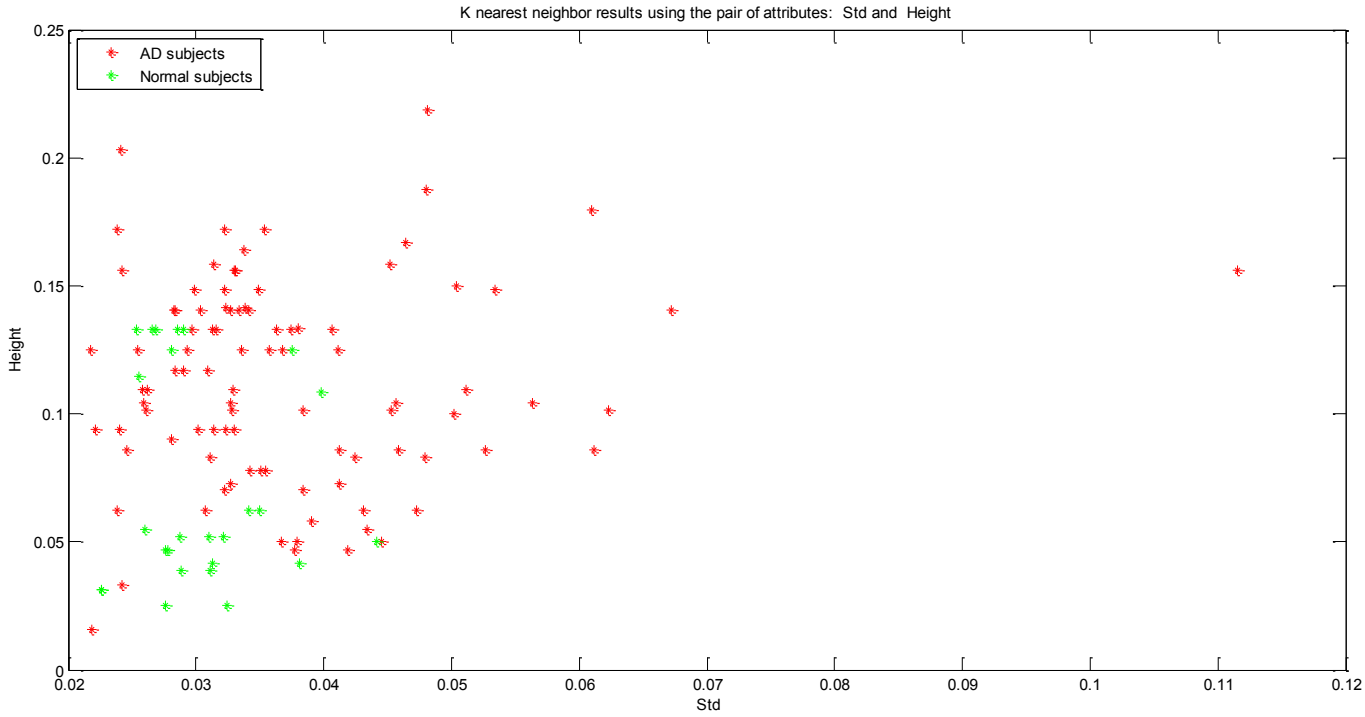
Clustering results using standard deviation and width 1 attributes



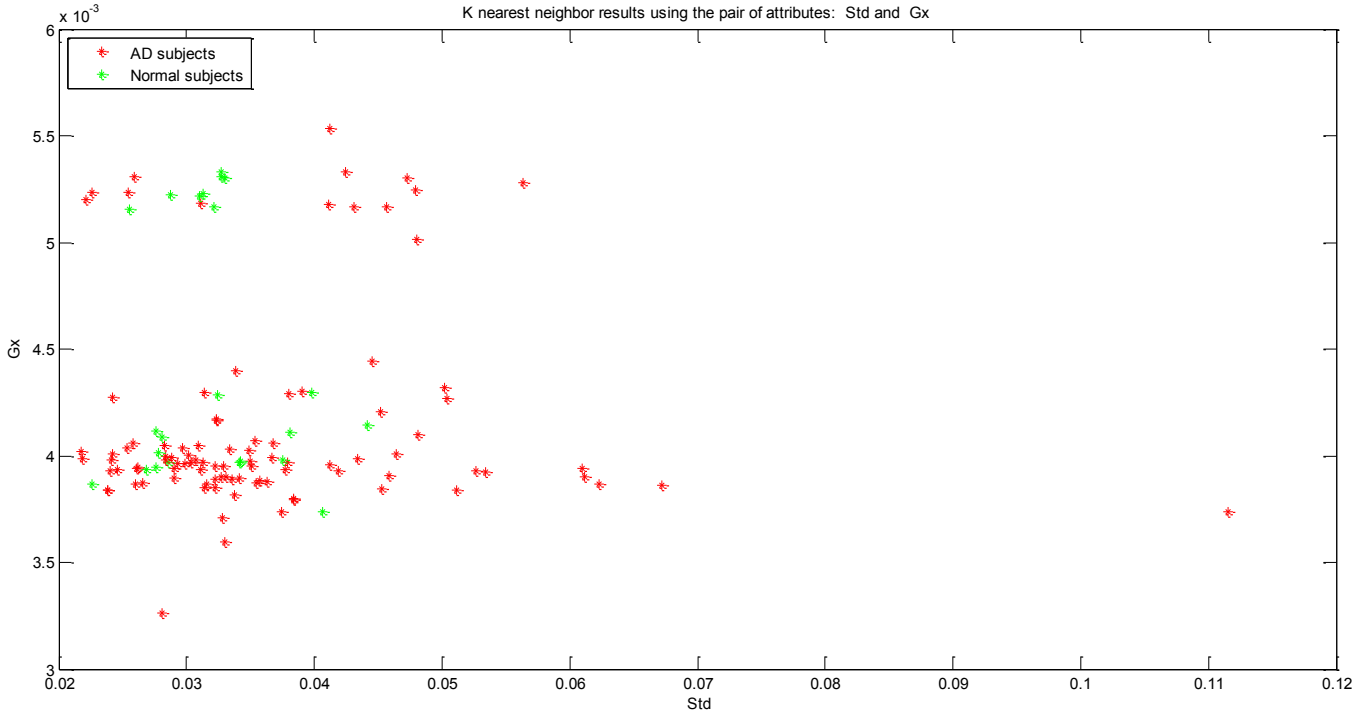
Clustering results using standard deviation and width 2 attributes



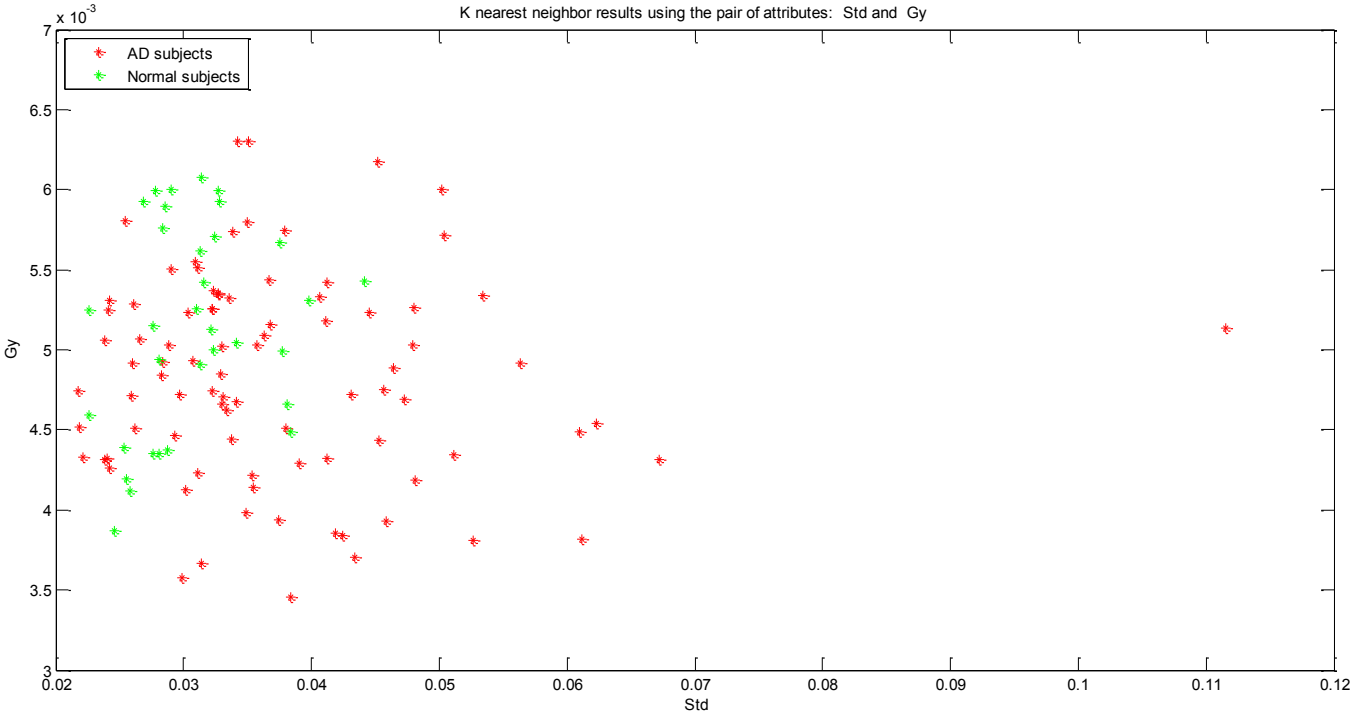
Clustering results using standard deviation and height attributes



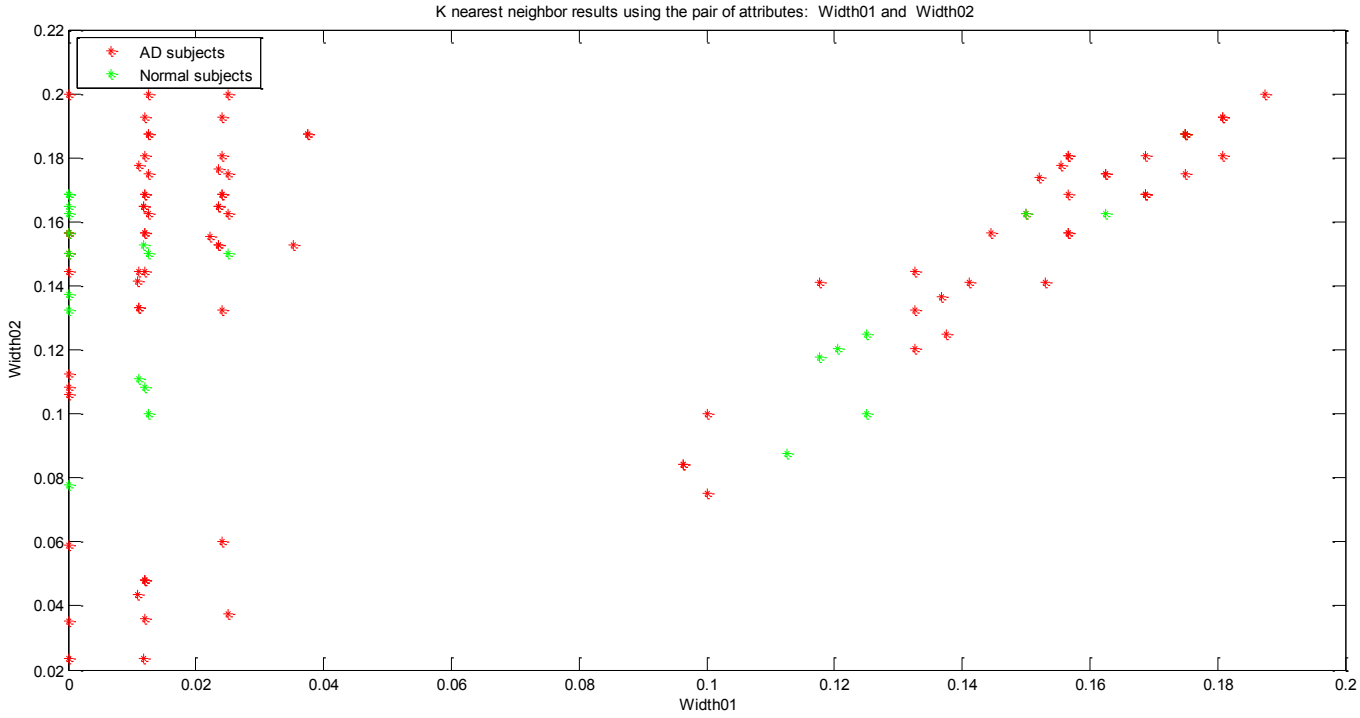
Clustering results using standard deviation and center of gravity Gx attributes



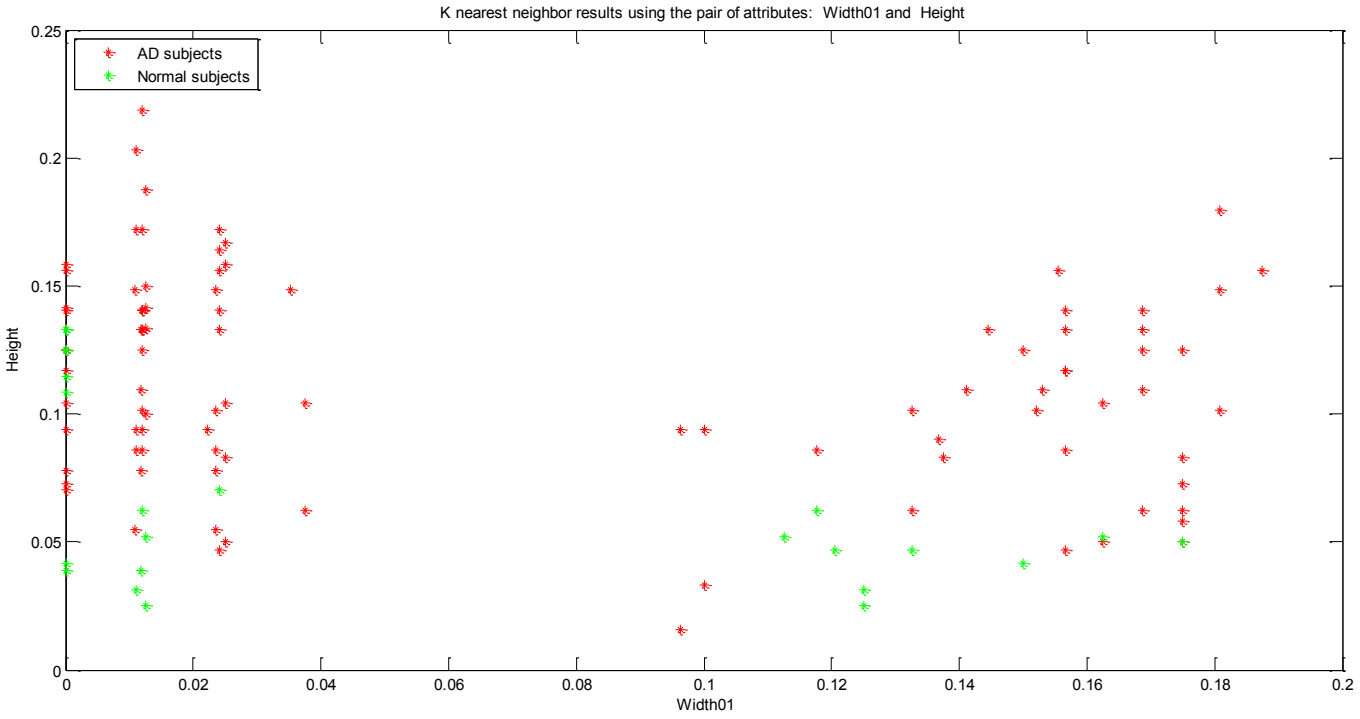
Clustering results using standard deviation and center of gravity Gy attributes



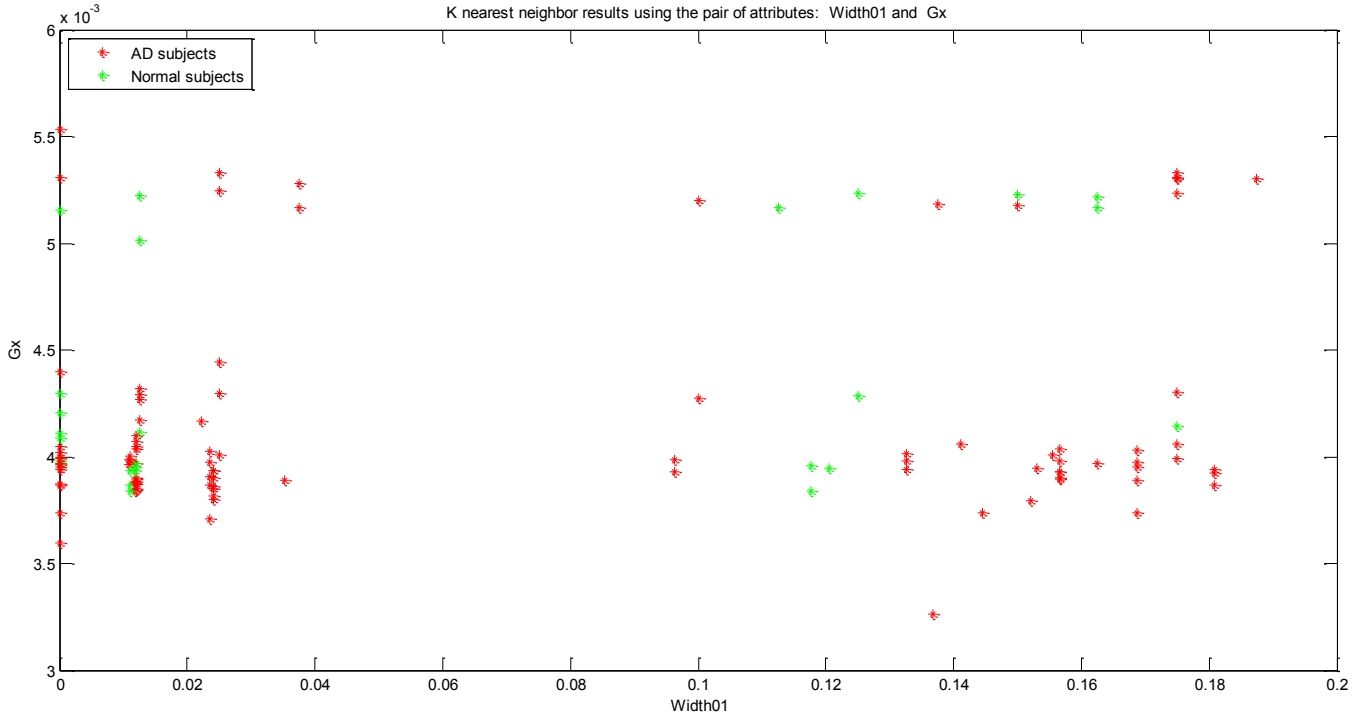
Clustering results using width 1 and width 2 attributes



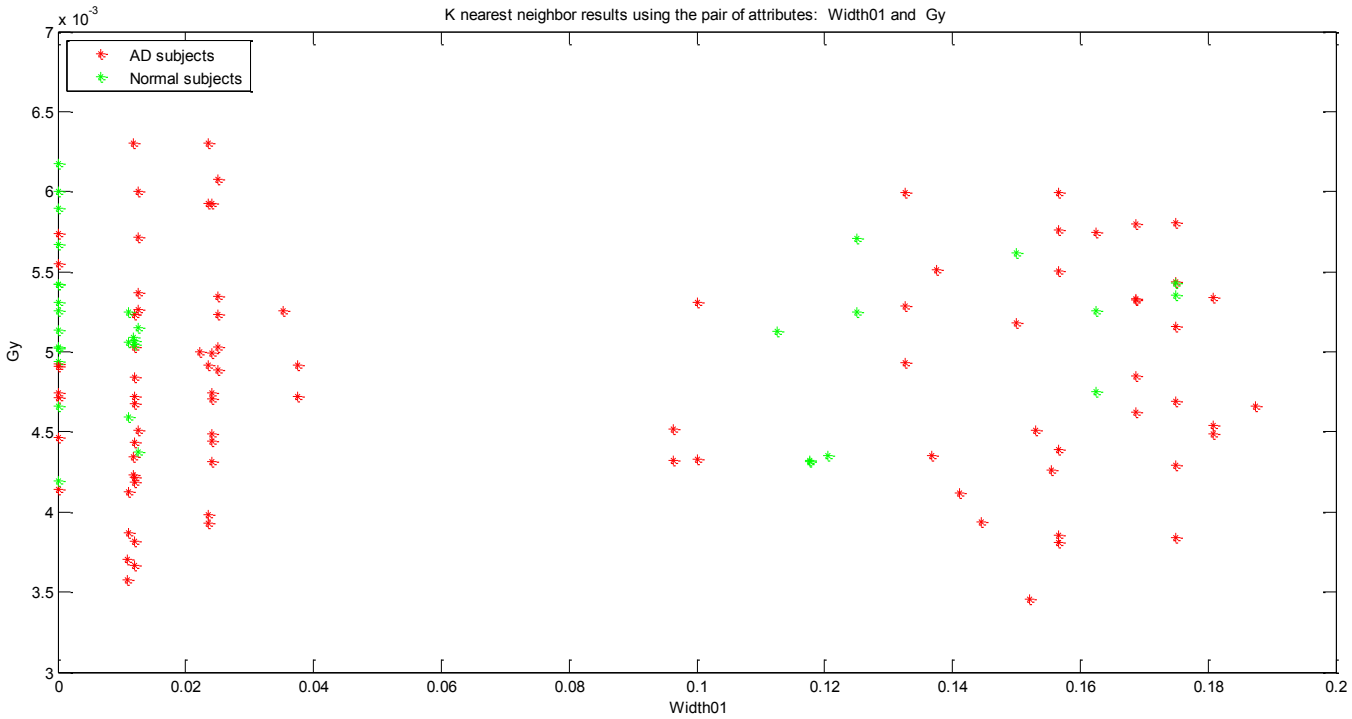
Clustering results using width 1 and height attributes



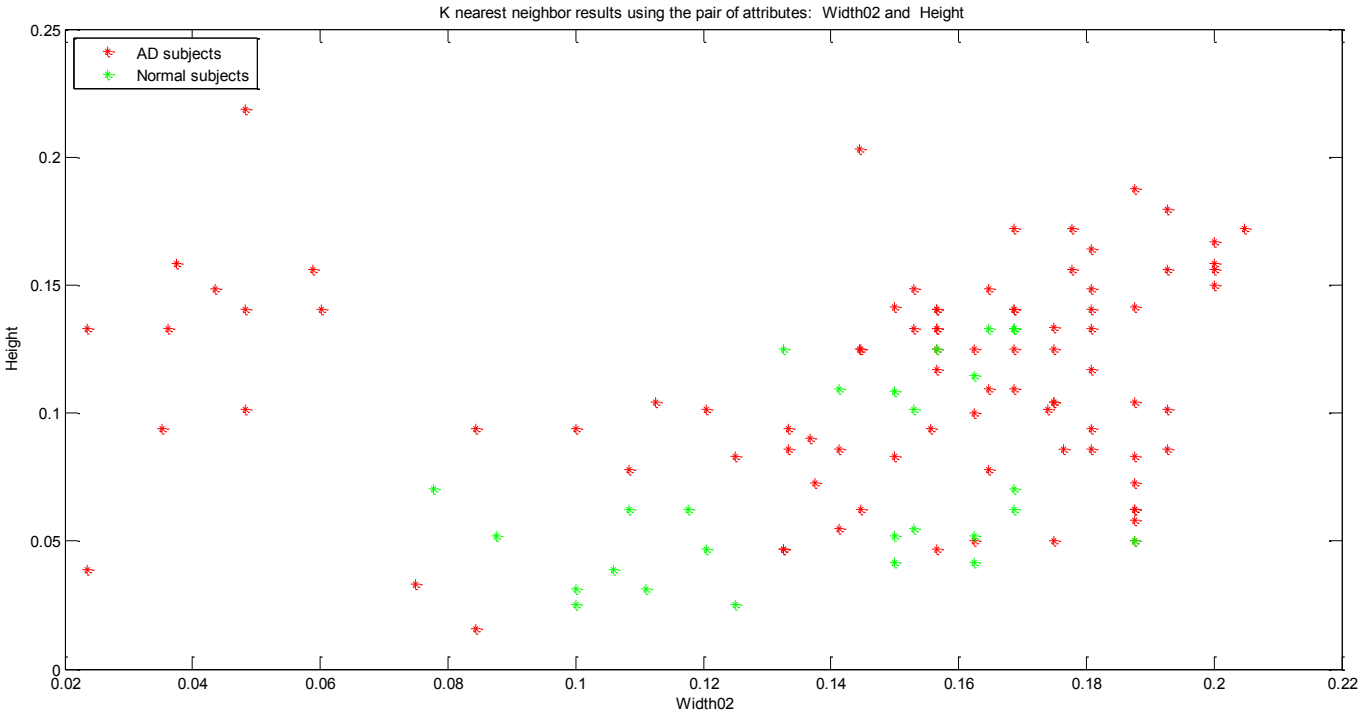
Clustering results using width 1 and center of gravity Gx attributes



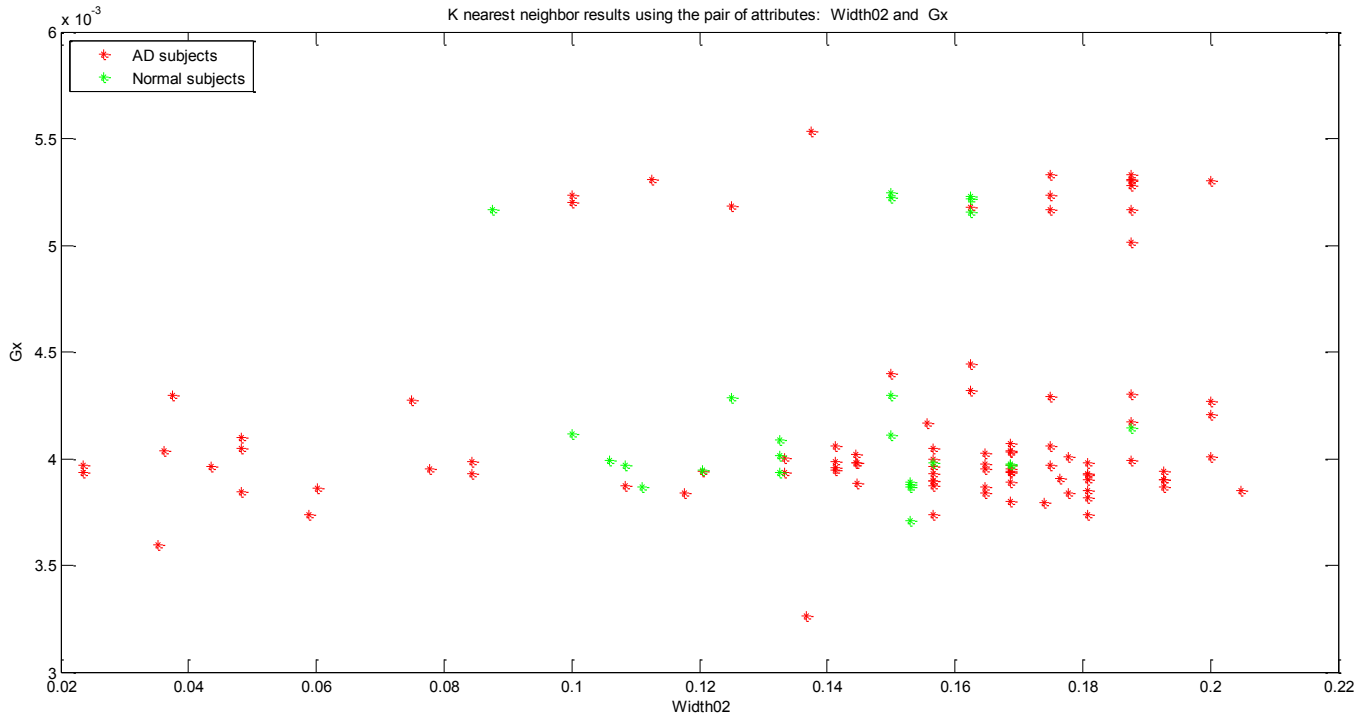
Clustering results using width 1 and center of gravity Gy attributes



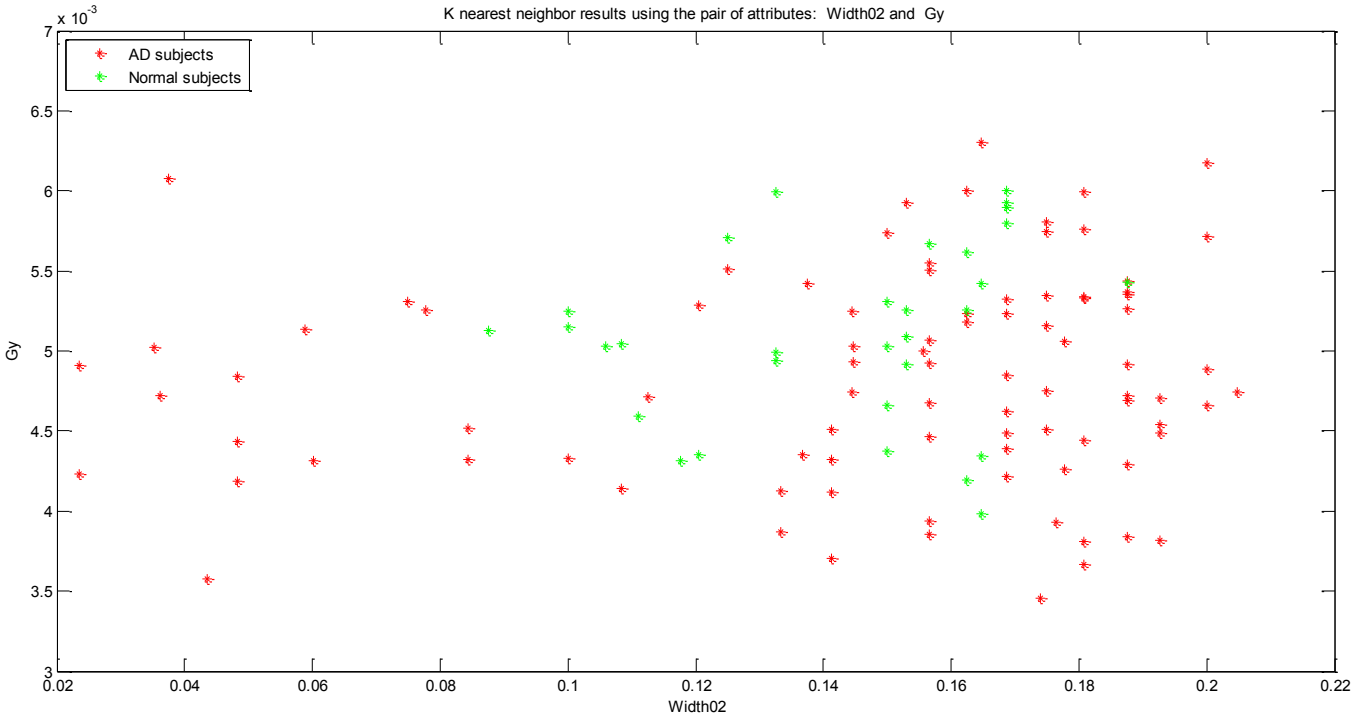
Clustering results using width 2 and height attributes



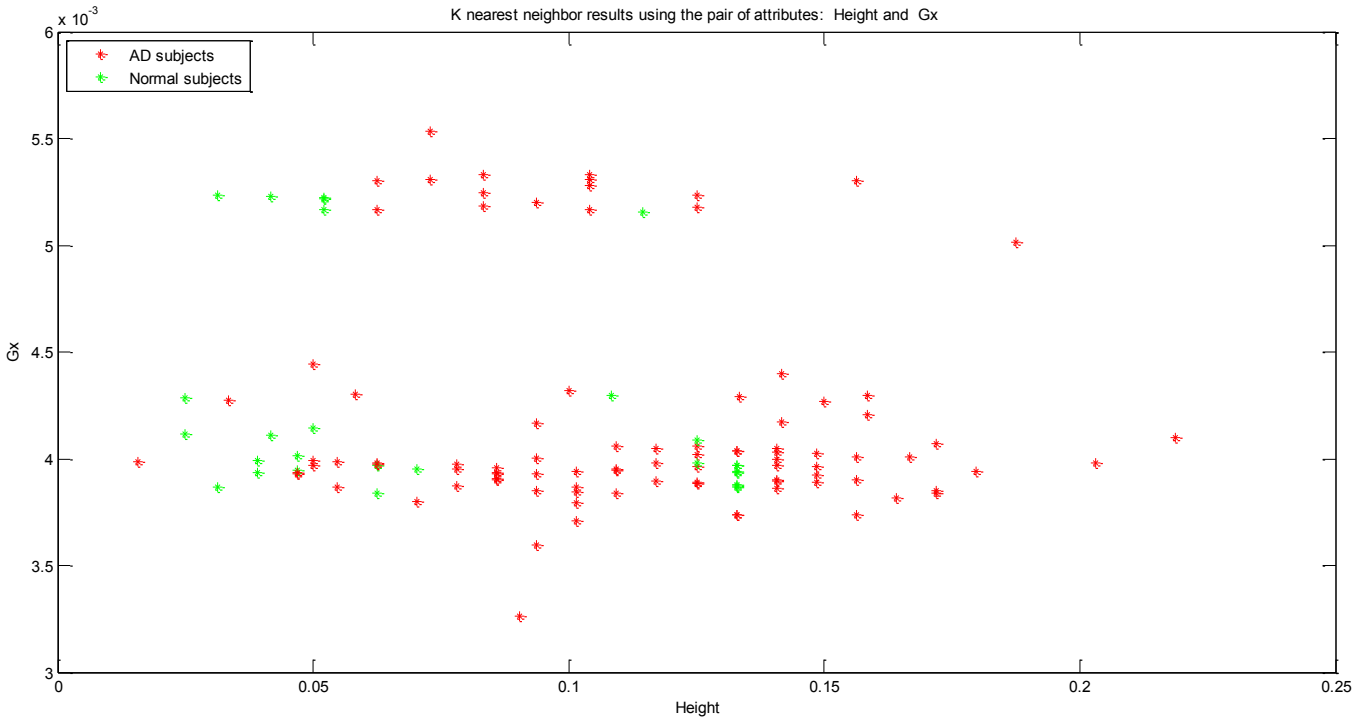
Clustering results using width 2 and center of gravity Gx attributes



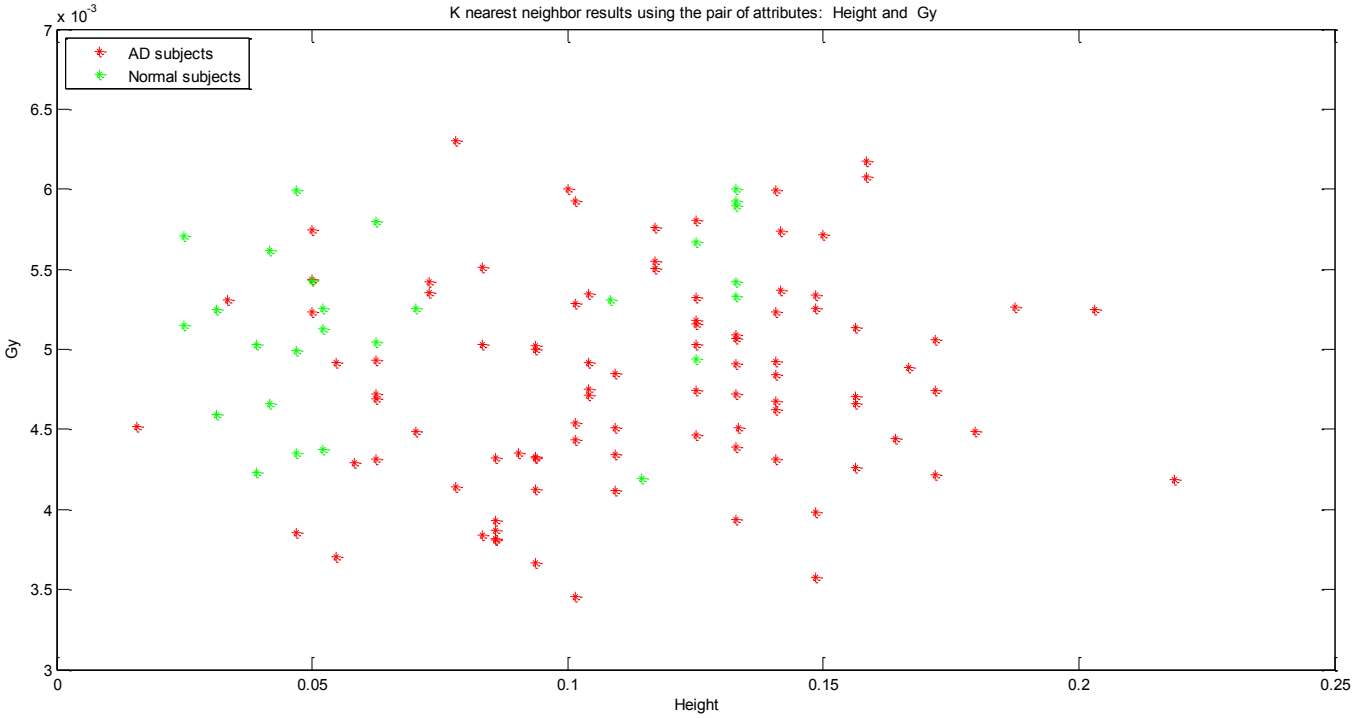
Clustering results using width 2 and center of gravity Gy attributes



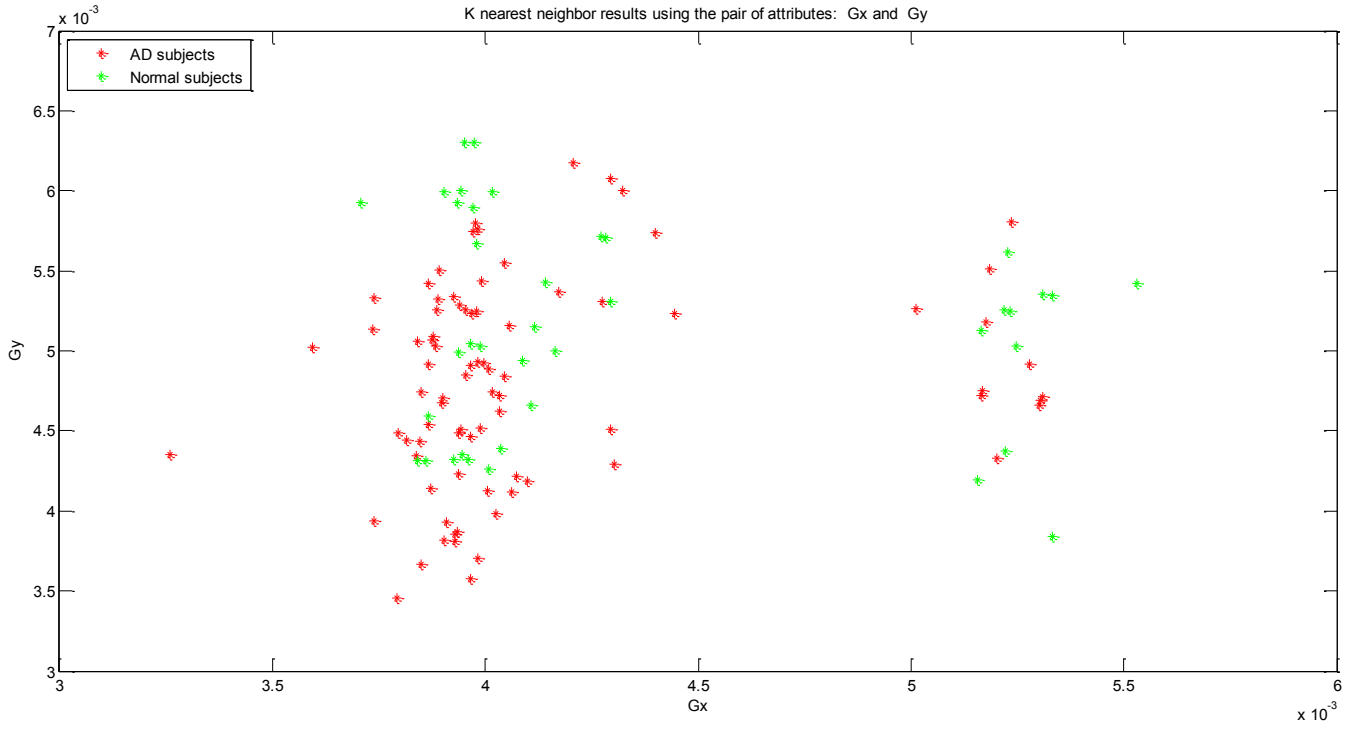
Clustering results using height and center of gravity Gx attributes



Clustering results using height and center of gravity Gy attributes

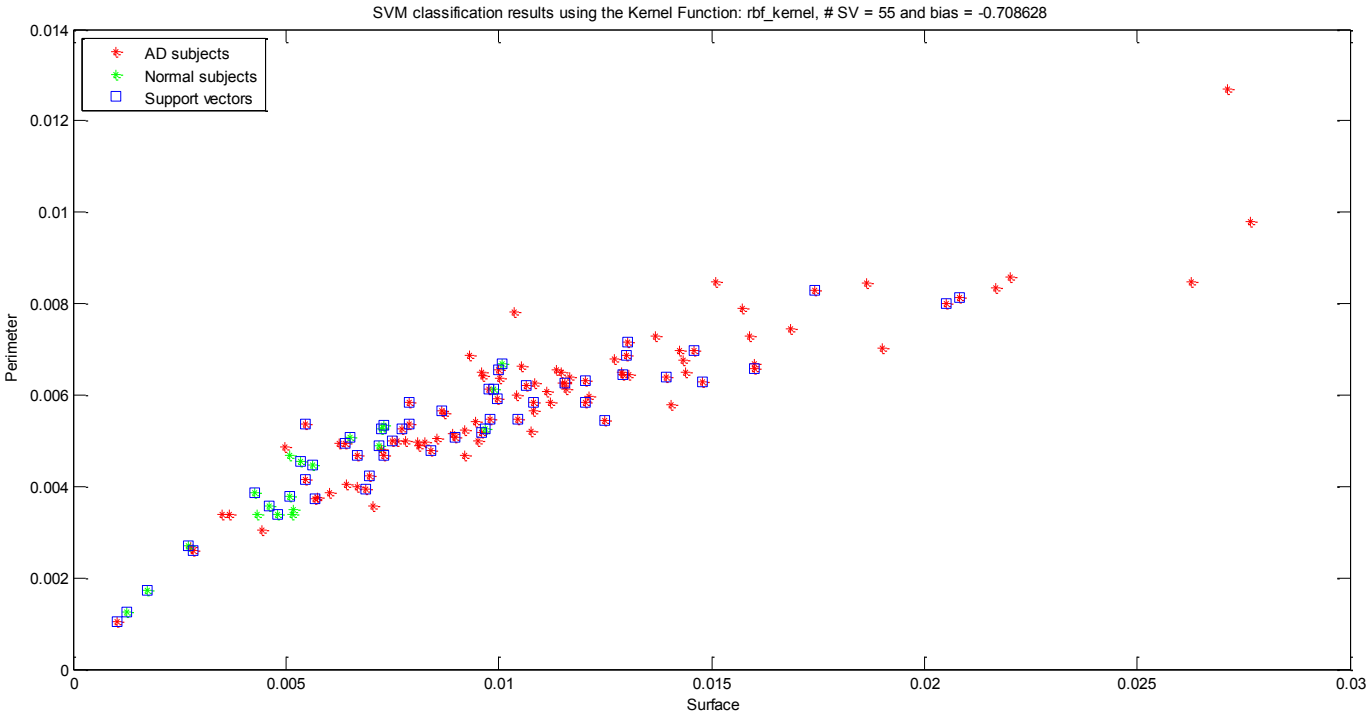


Clustering results using center of gravity Gx and center of gravity Gy attributes

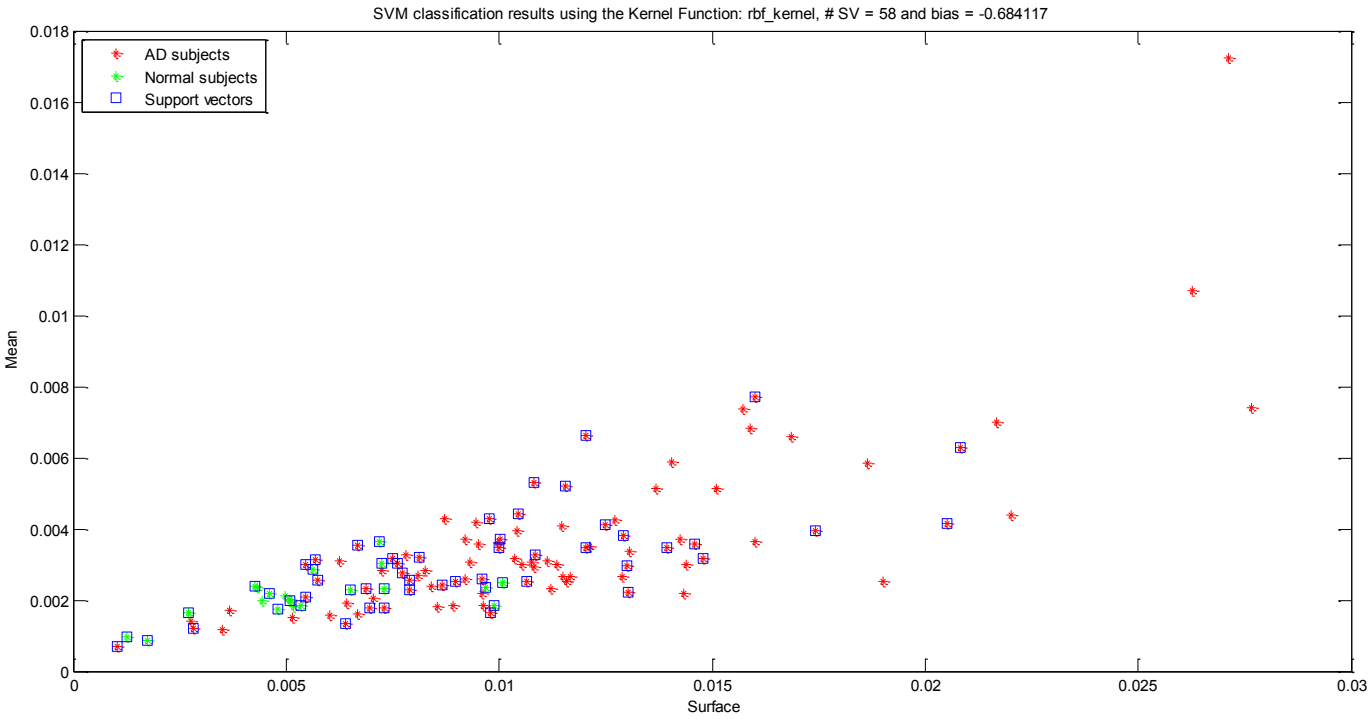


Appendix B: SVM results of pairs of attributes.

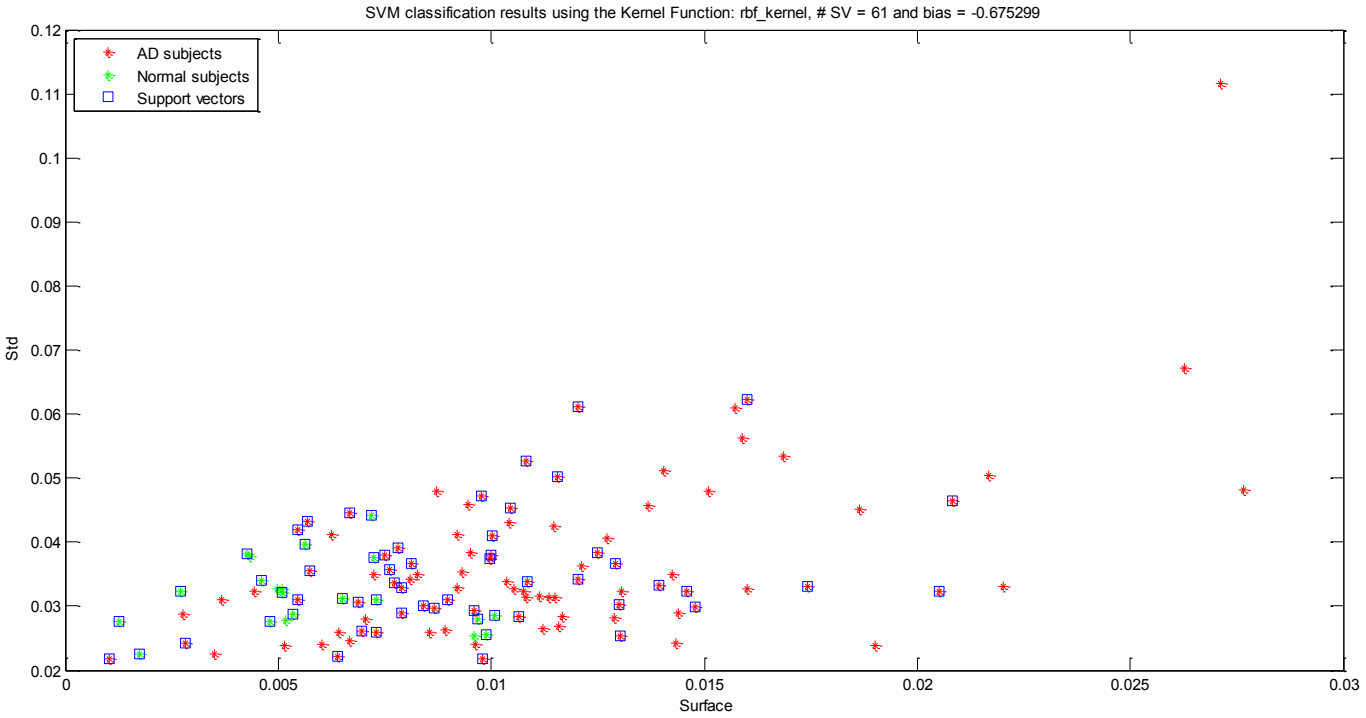
Support vector machine (SVM) results using surface and perimeter attributes



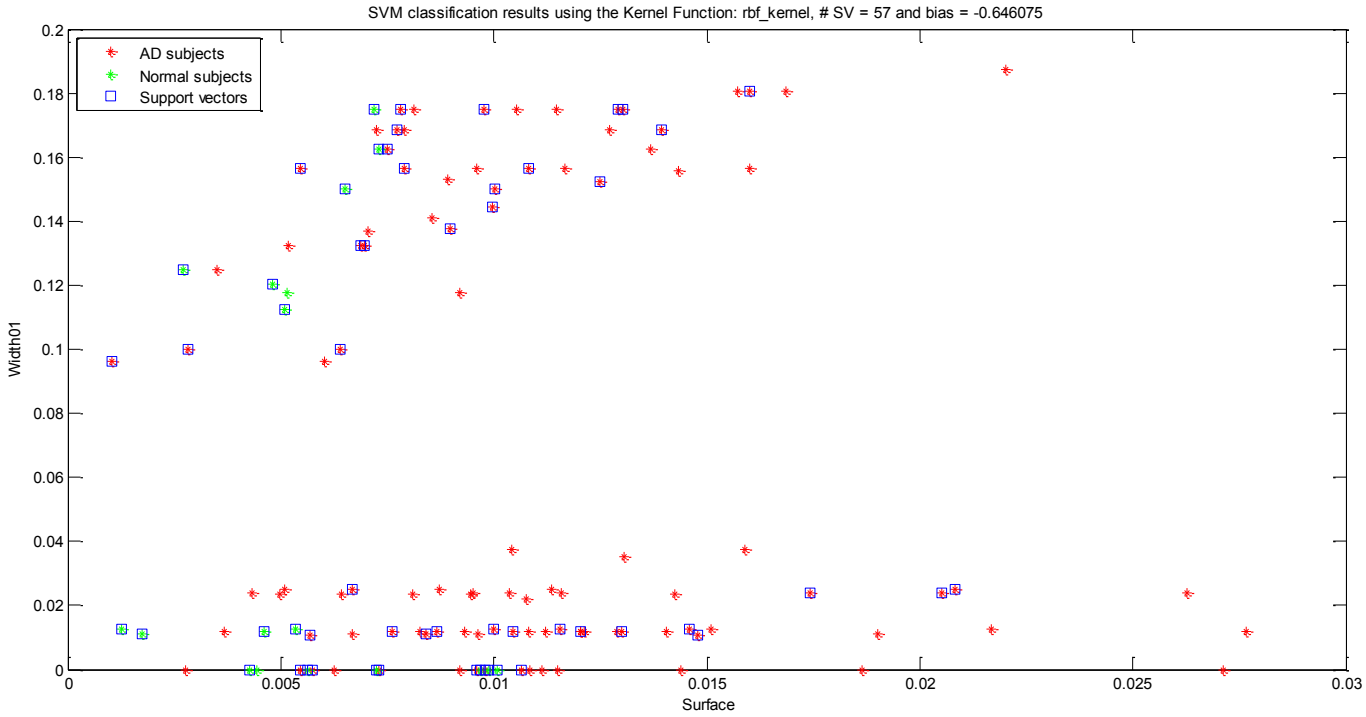
Support vector machine (SVM) results using surface and average attributes



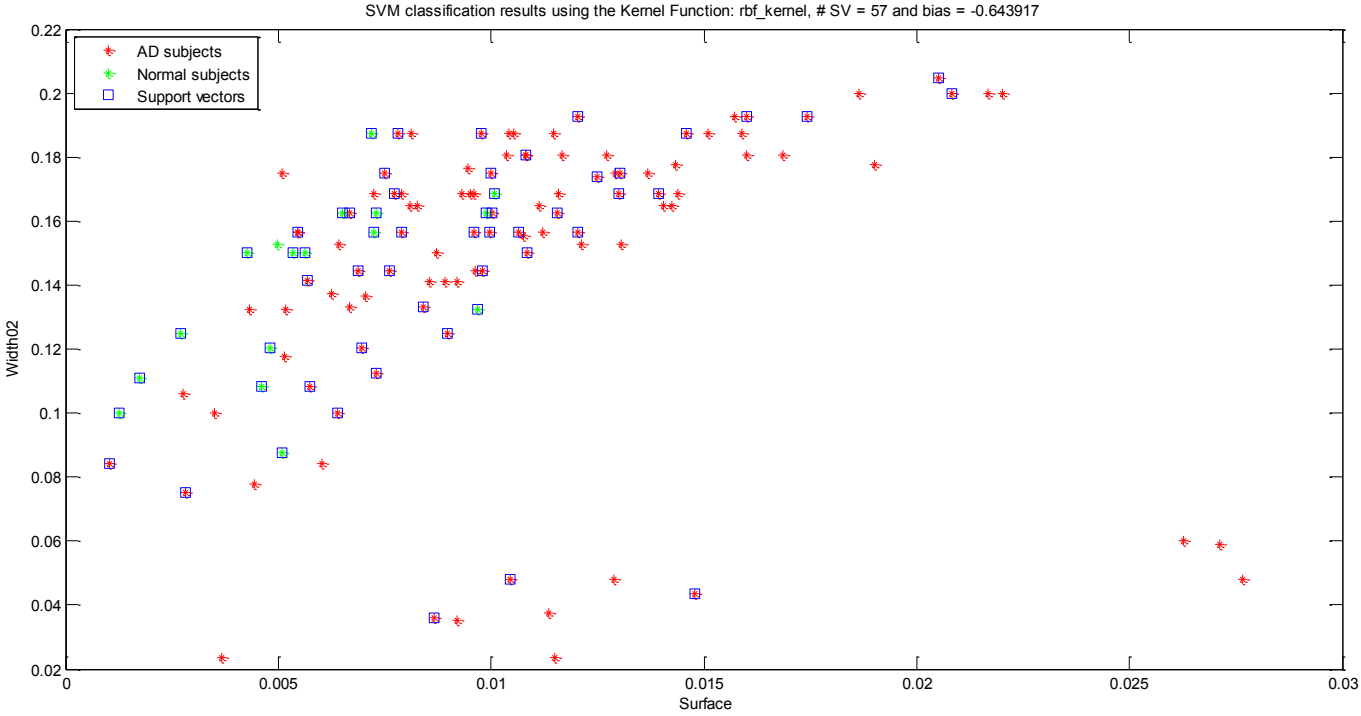
Support vector machine (SVM) results using surface and standard deviation attributes



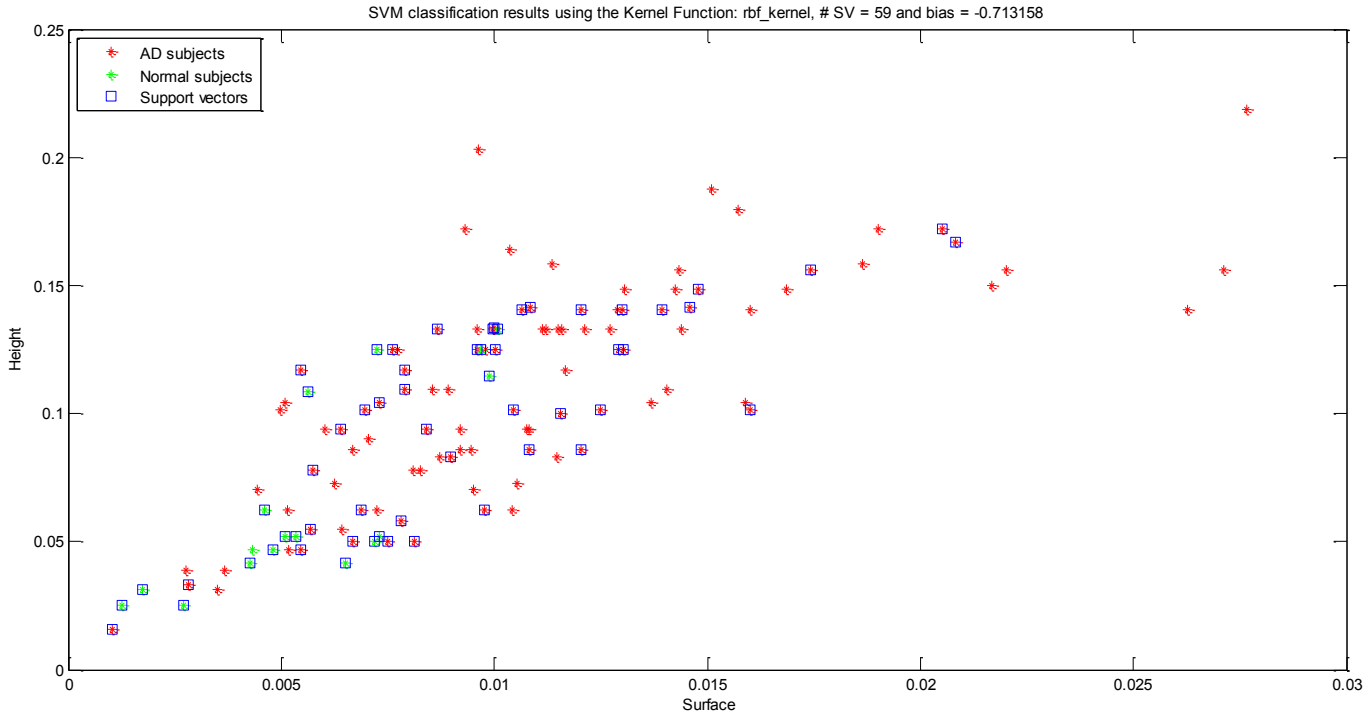
Support vector machine (SVM) results using surface and width 1 attributes



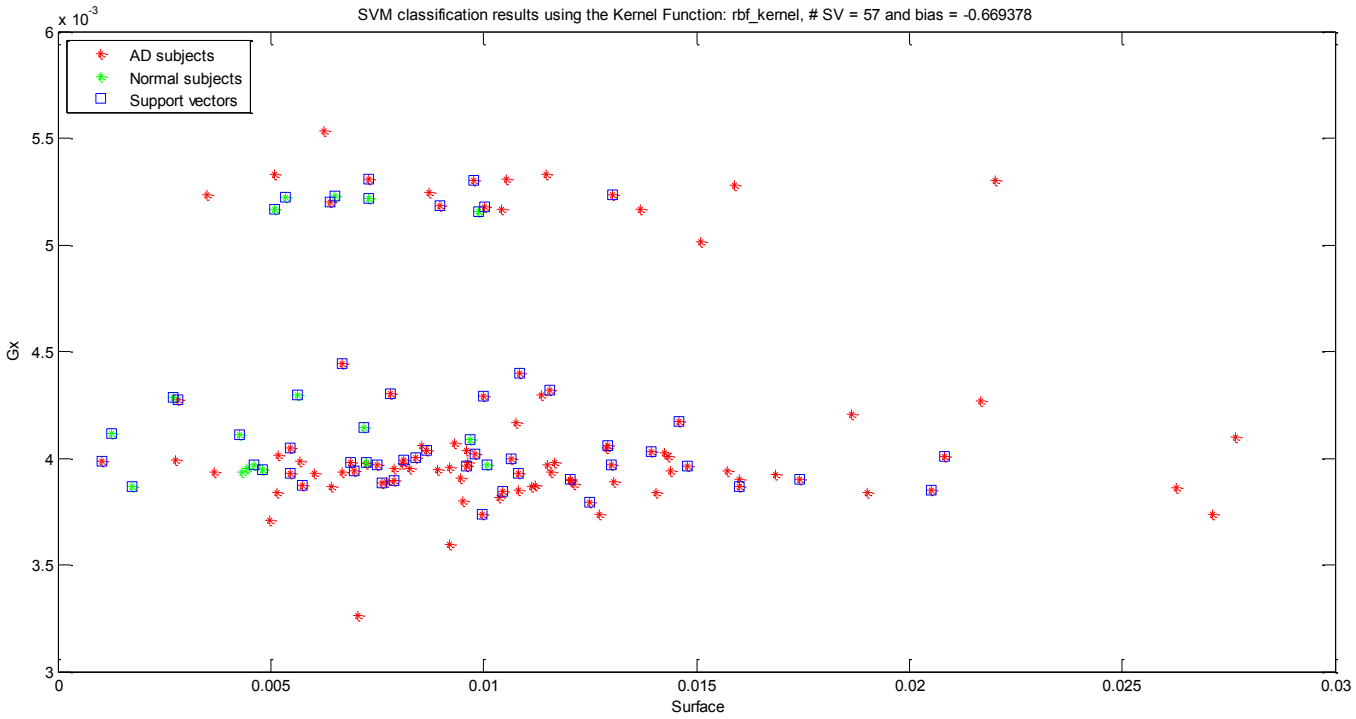
Support vector machine (SVM) results using surface and width 2 attributes



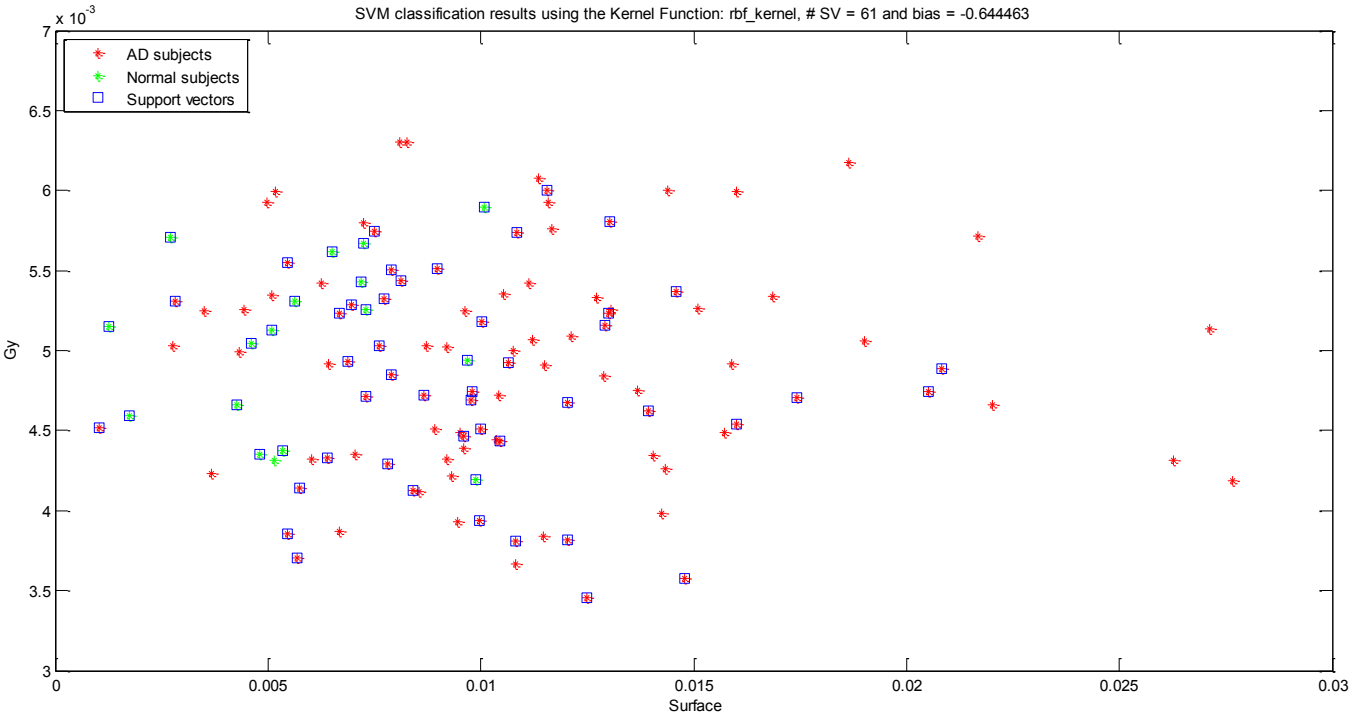
Support vector machine (SVM) results using surface and height attributes



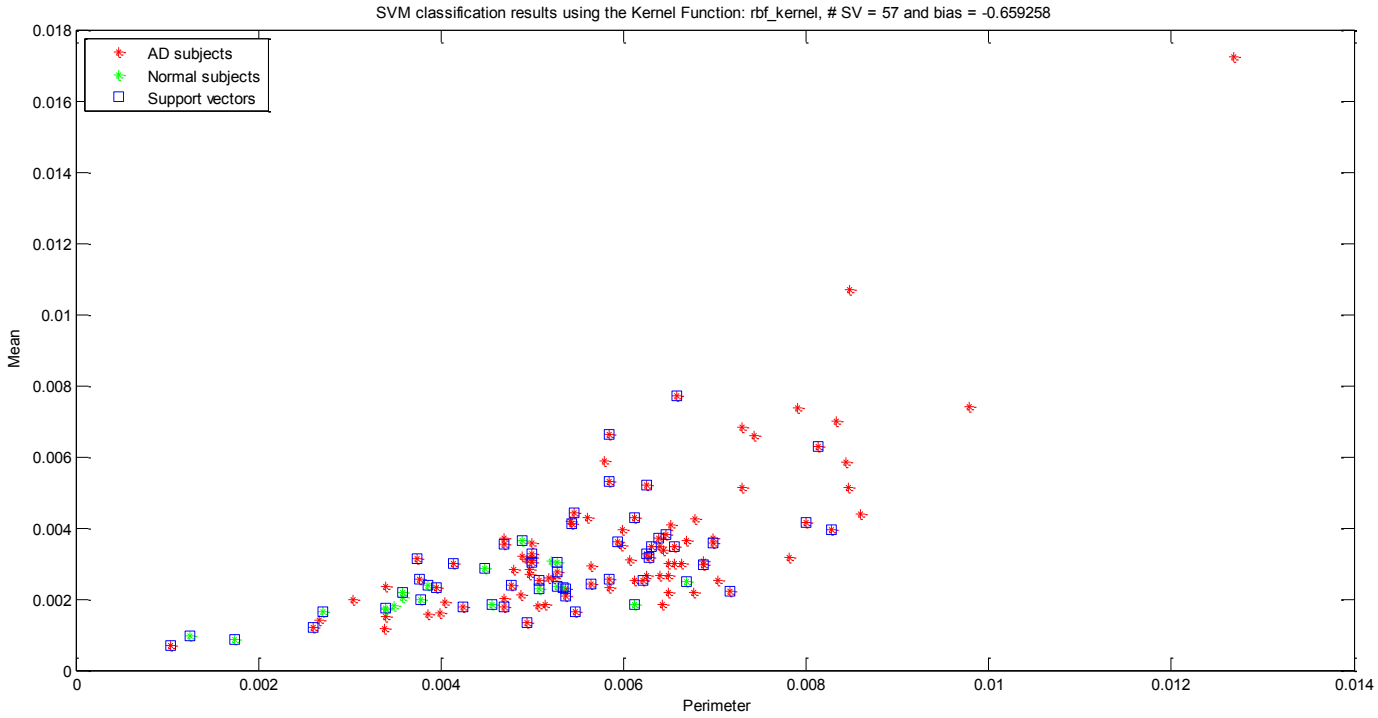
Support vector machine (SVM) results using surface and center of gravity Gx attributes



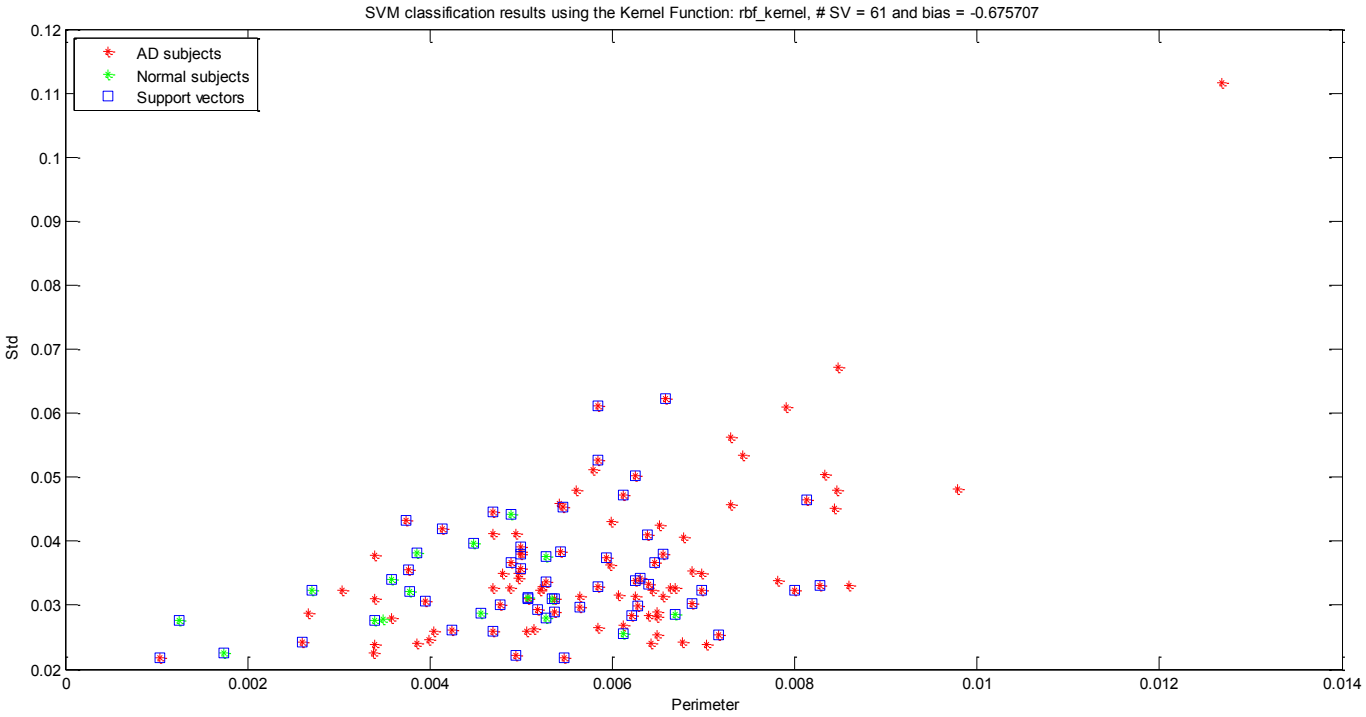
Support vector machine (SVM) results using surface and center of gravity Gy attributes



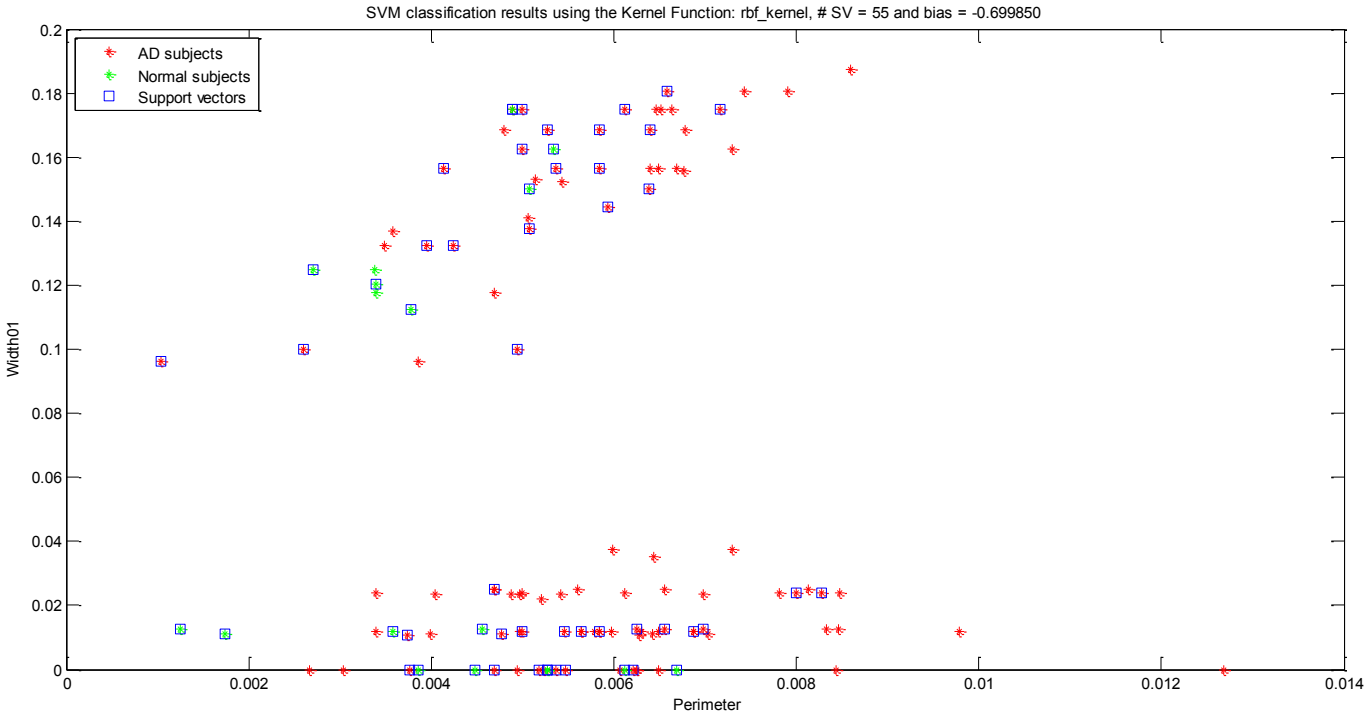
Support vector machine (SVM) results using perimeter and average attributes



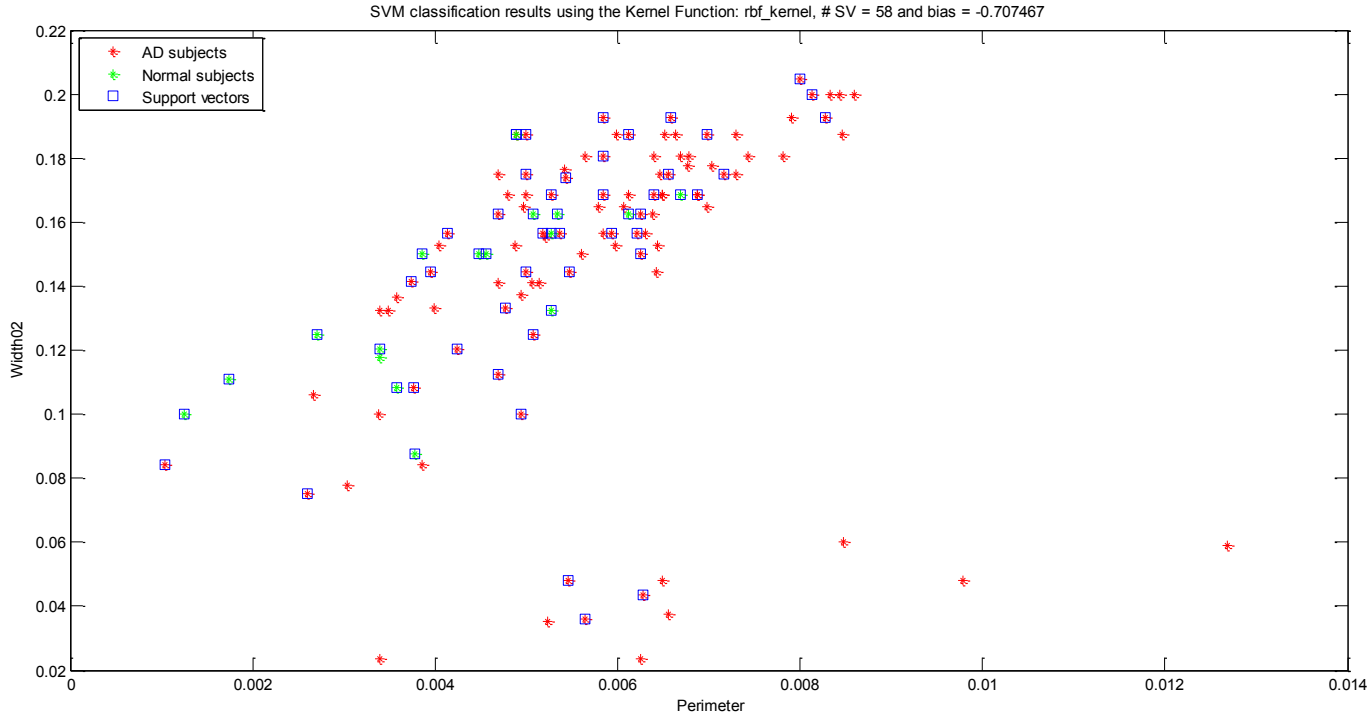
Support vector machine (SVM) results using perimeter and standard deviation attributes



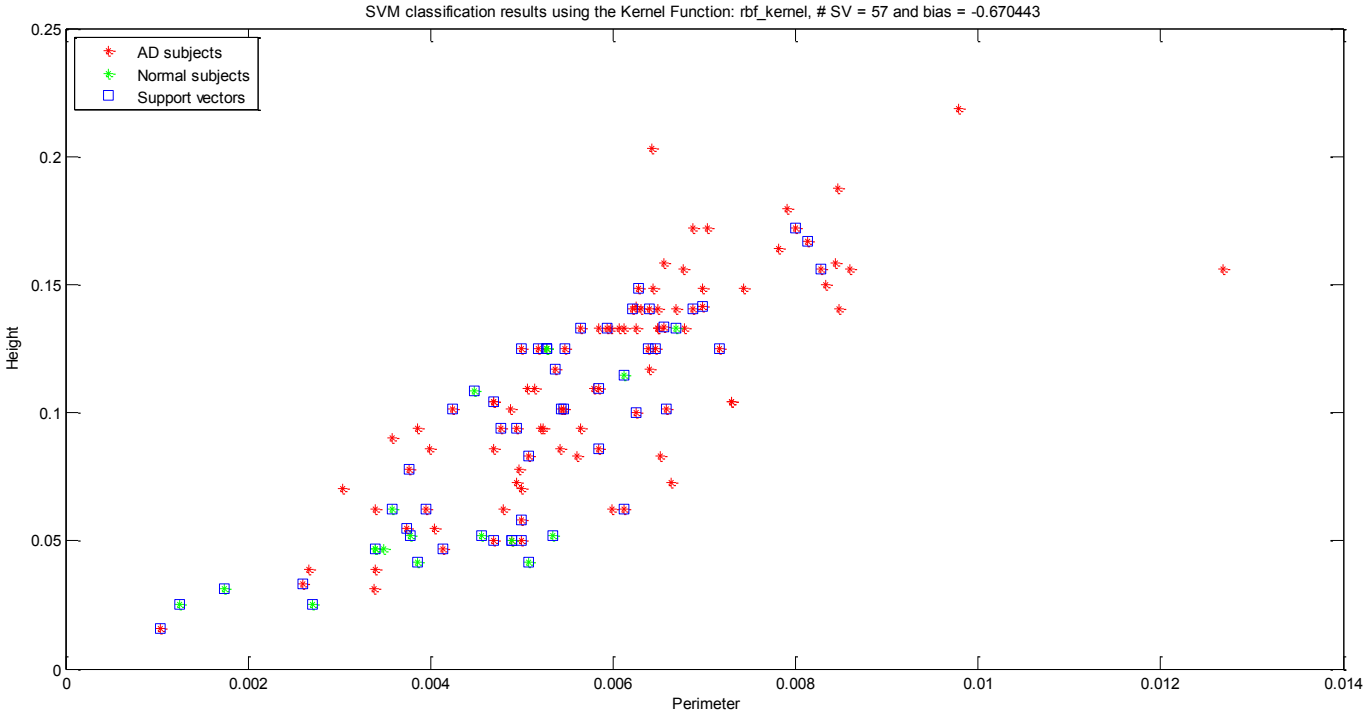
Support vector machine (SVM) results using perimeter and Width 1



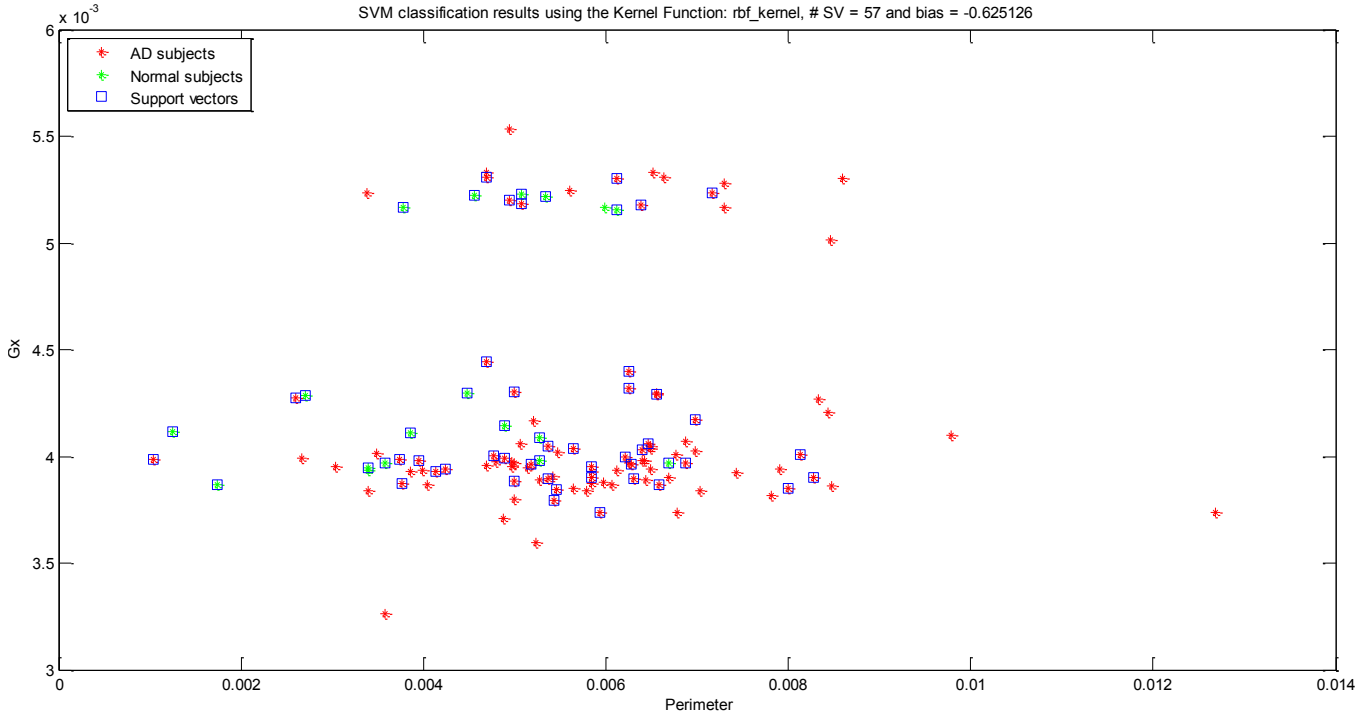
Support vector machine (SVM) results using perimeter and width 2 attributes



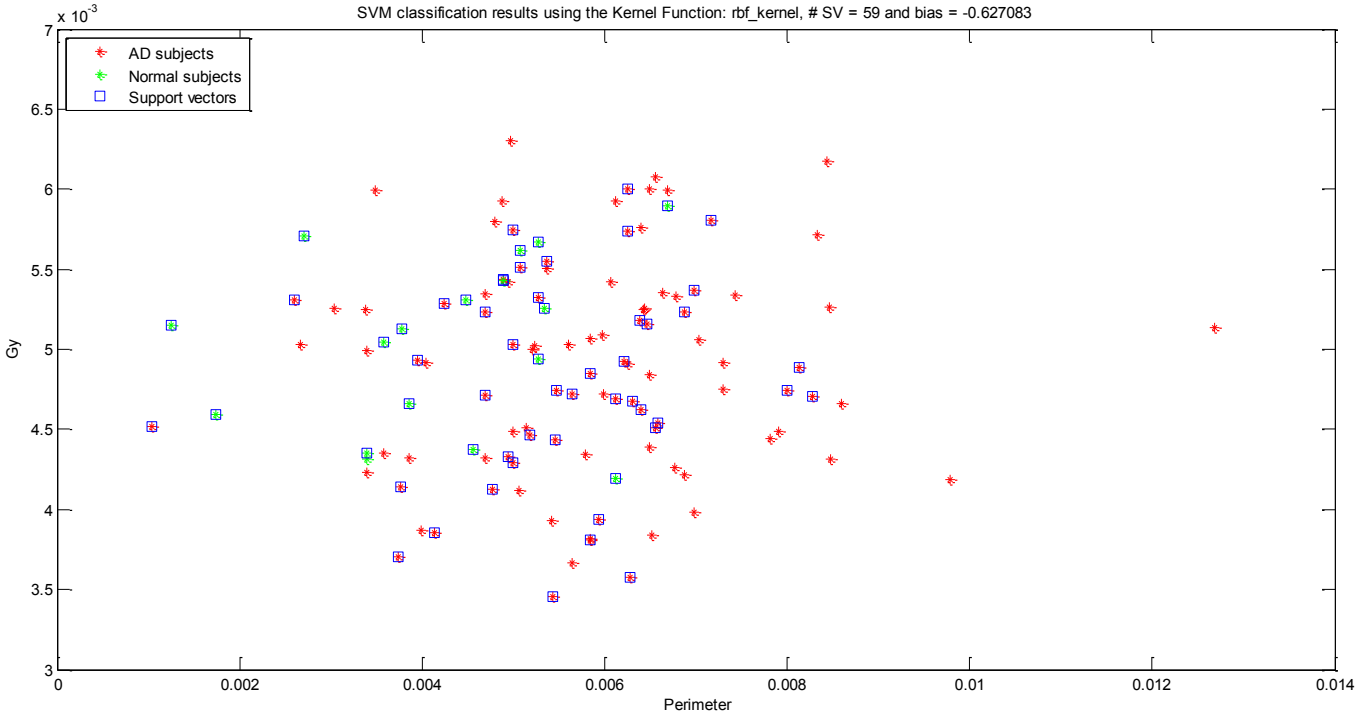
Support vector machine (SVM) results using perimeter and height attributes



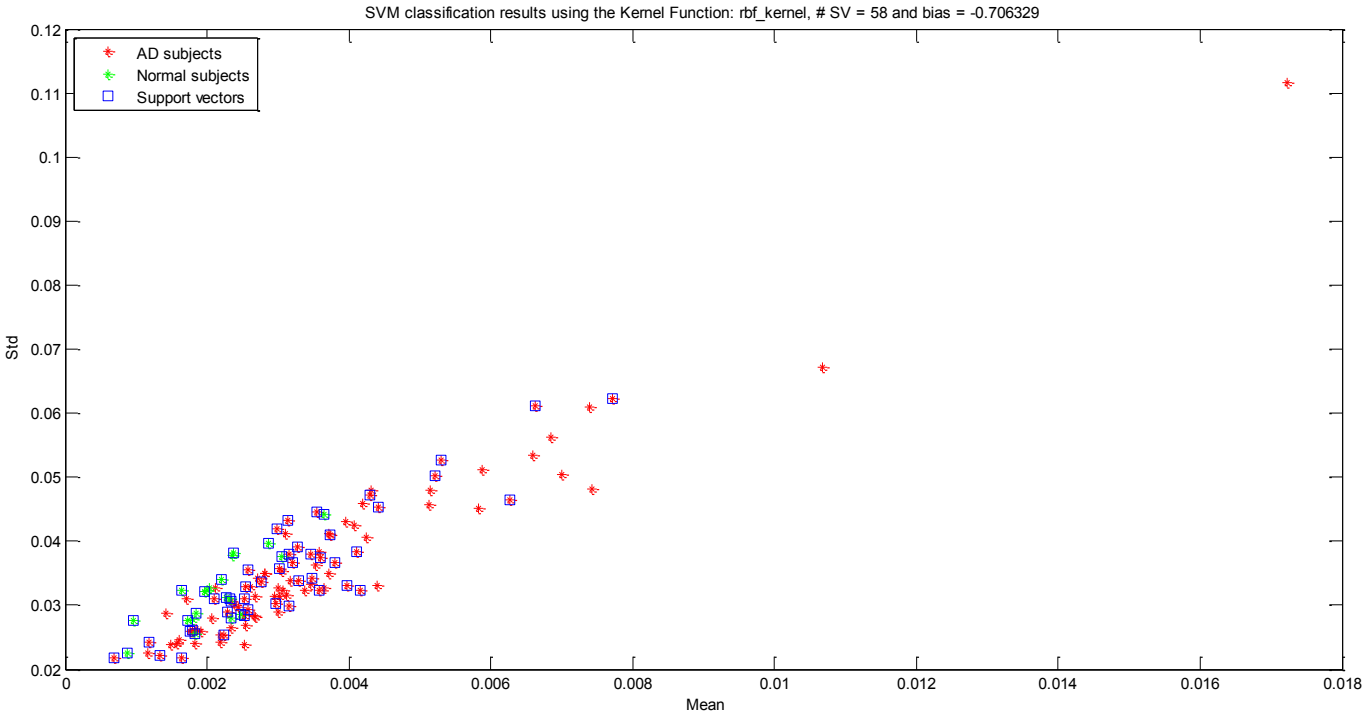
Support vector machine (SVM) results using perimeter and center of gravity Gx attributes



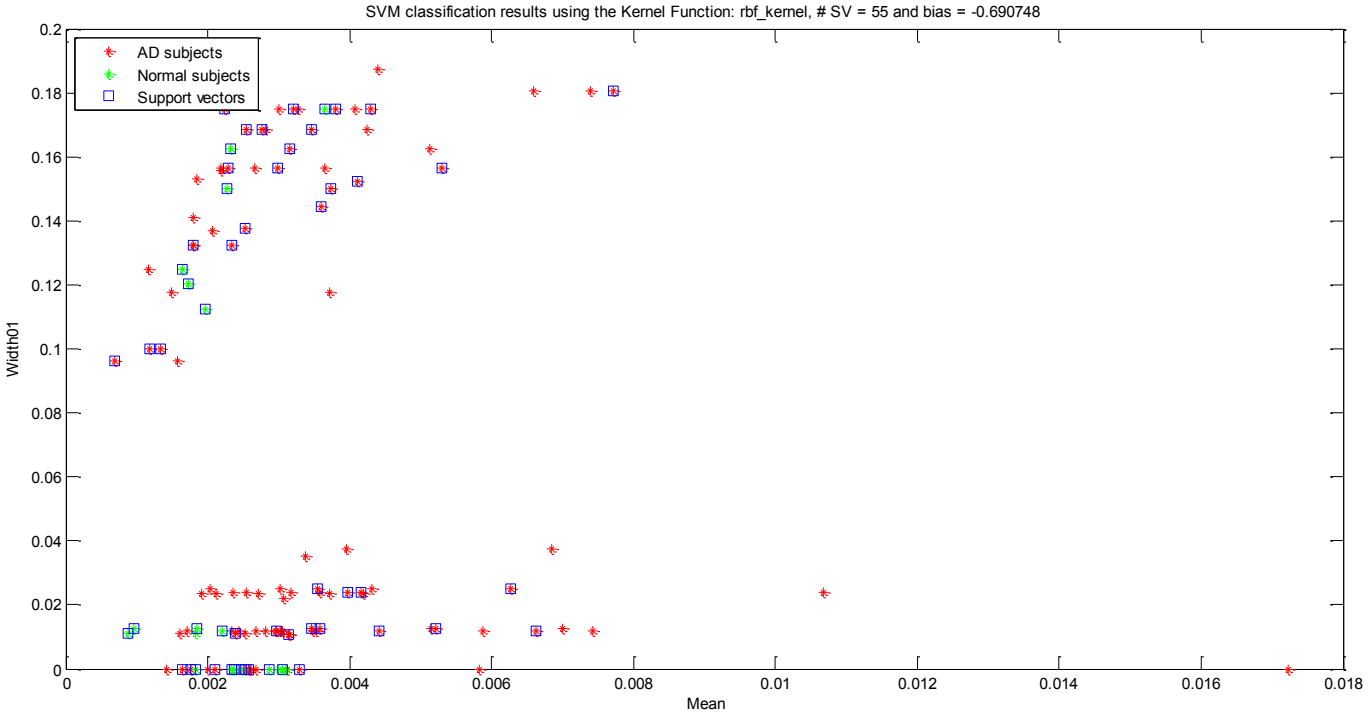
Support vector machine (SVM) results using perimeter and center of gravity Gy attributes



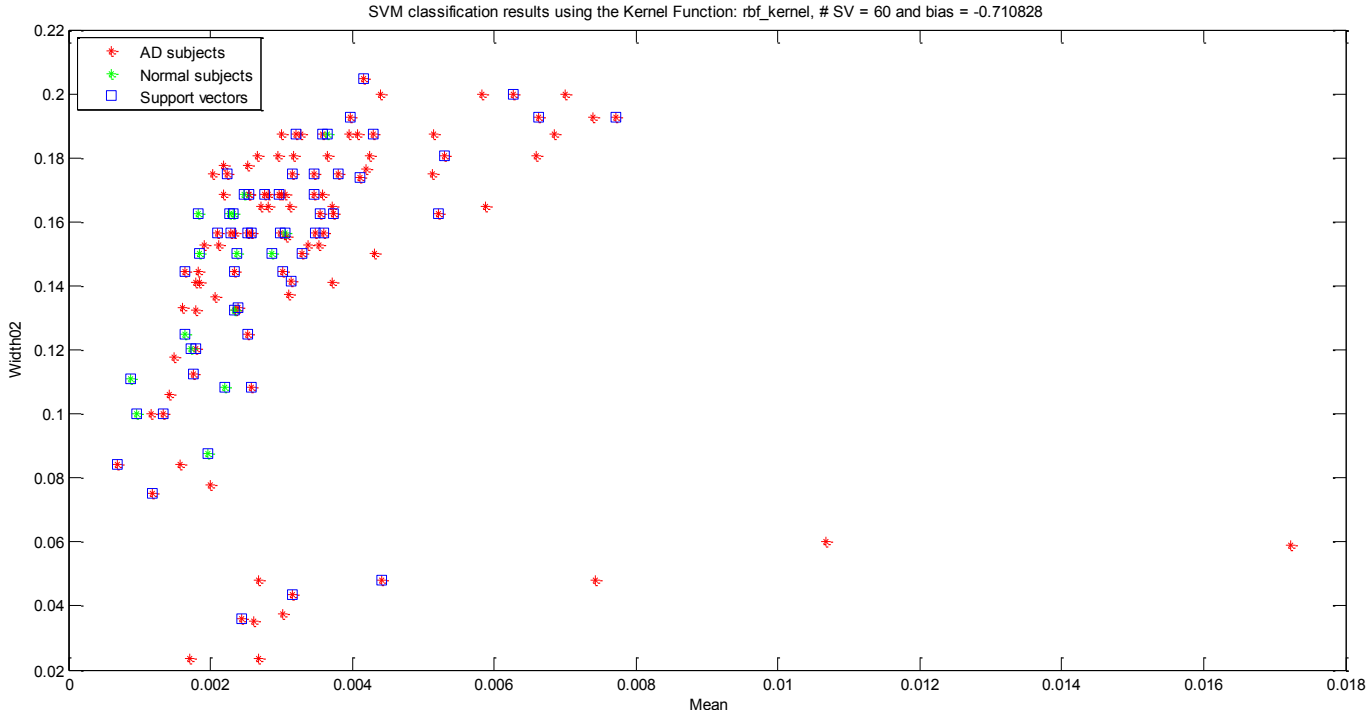
Support vector machine (SVM) results using average and standard deviation attributes



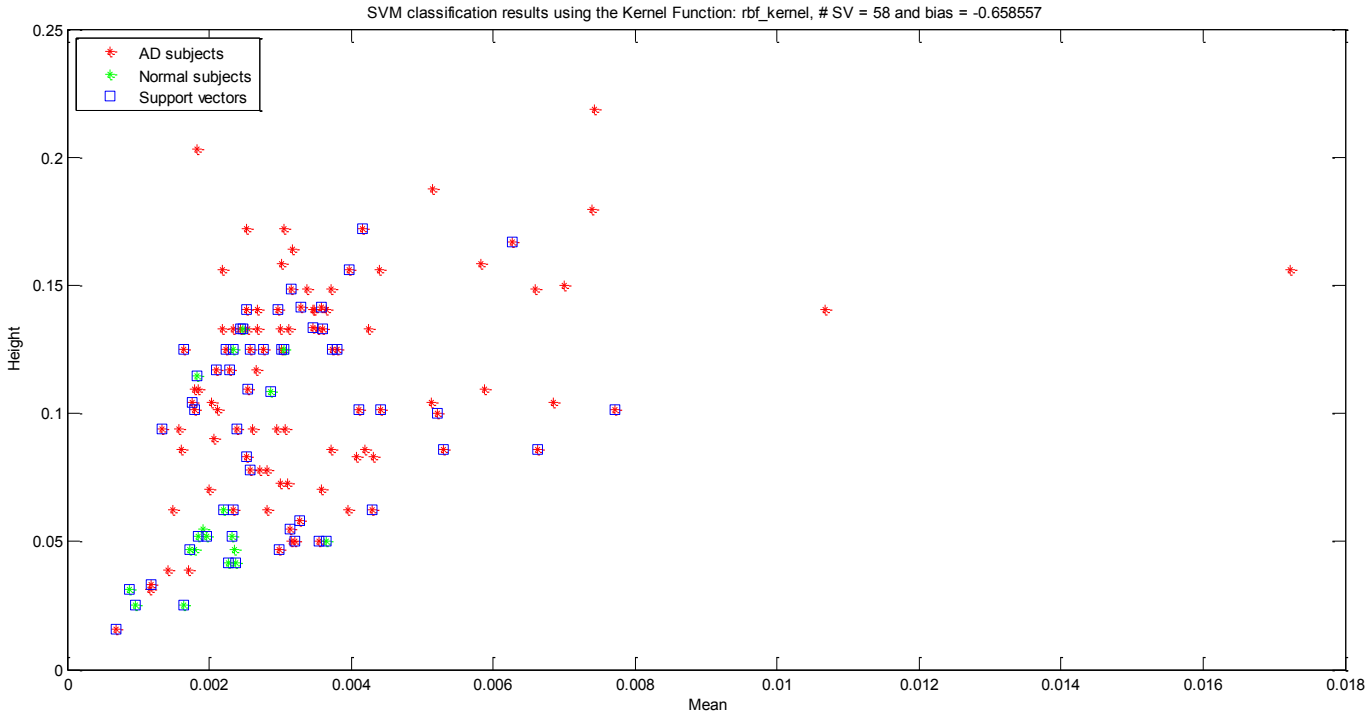
Support vector machine (SVM) results using average and width 1 attributes



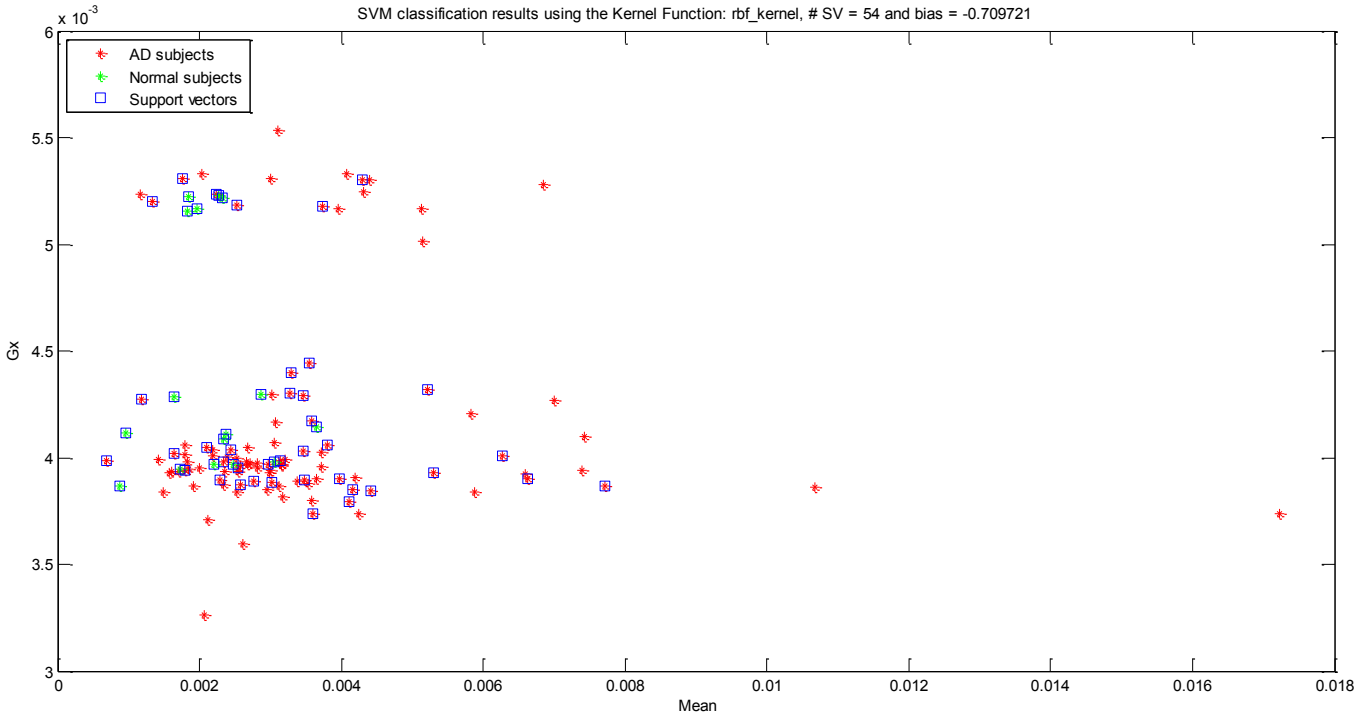
Support vector machine (SVM) results using average and width 2 attributes



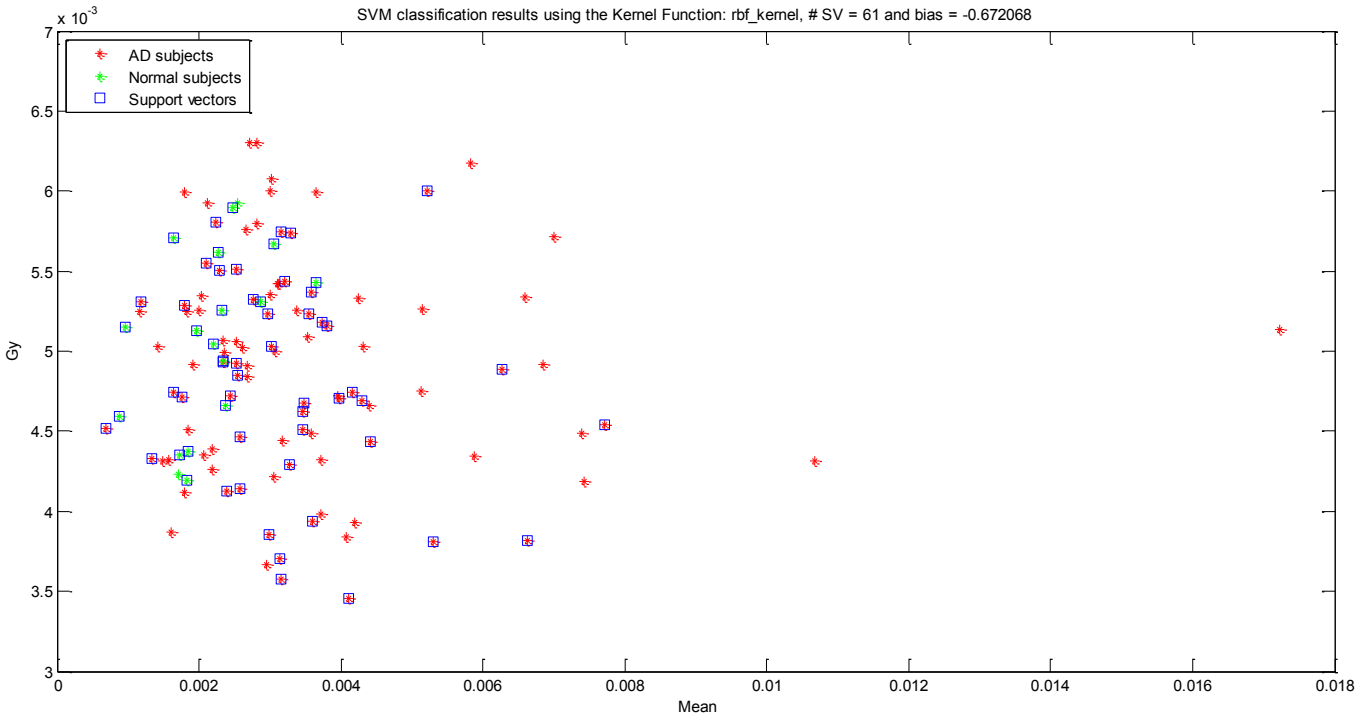
Support vector machine (SVM) results using average and height attributes



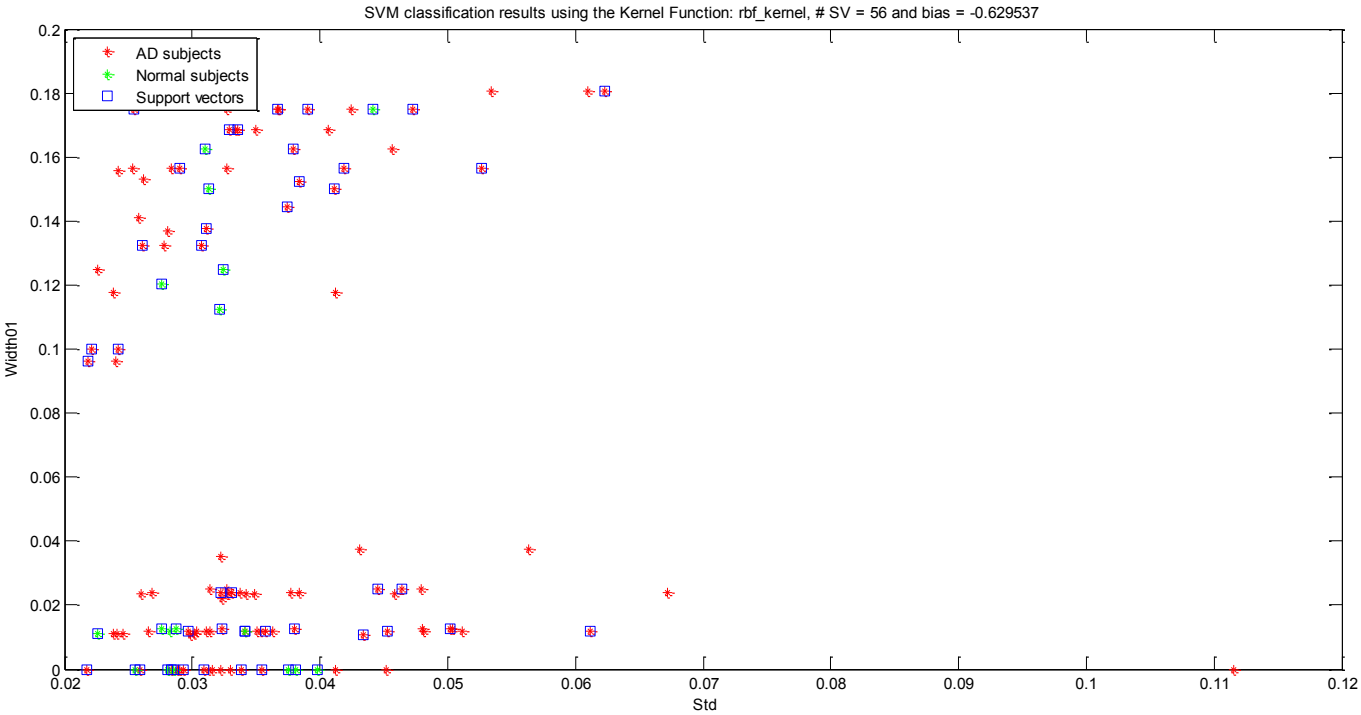
Support vector machine (SVM) results using average and standard deviation Gx attributes



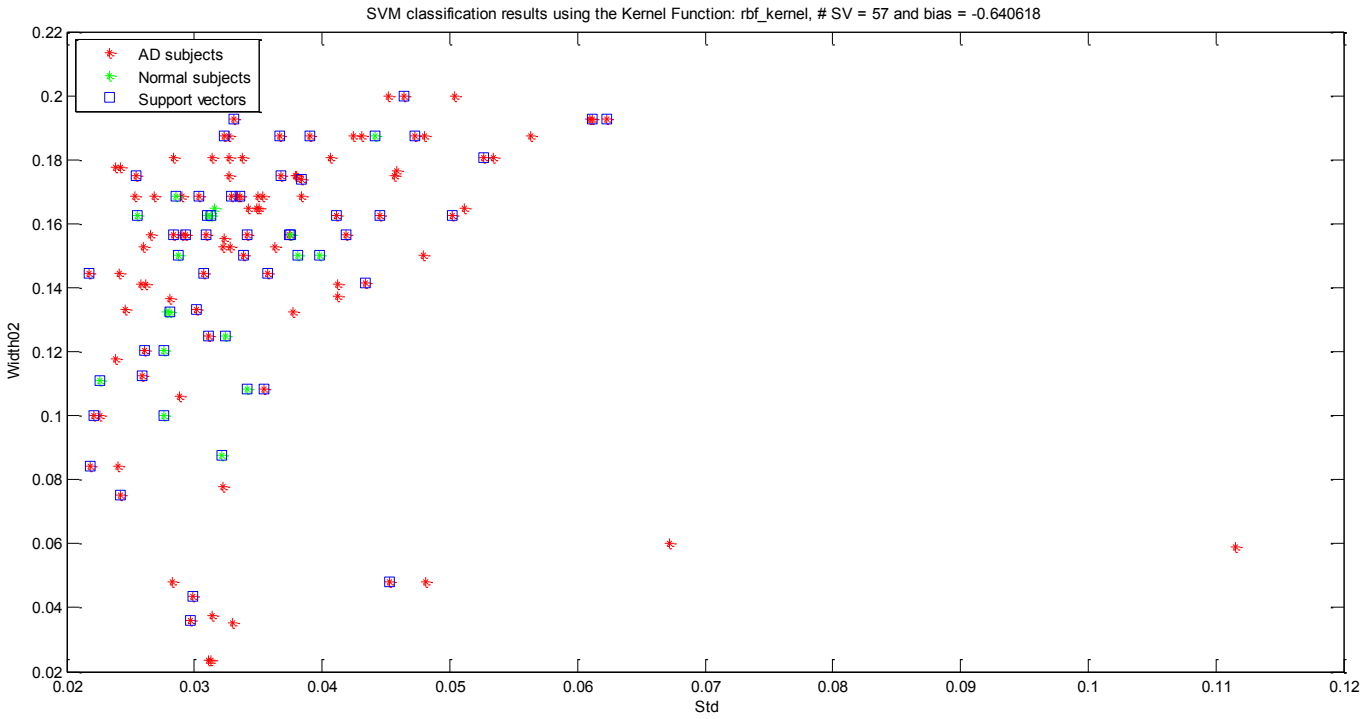
Support vector machine (SVM) results using average and standard deviation Gy attributes



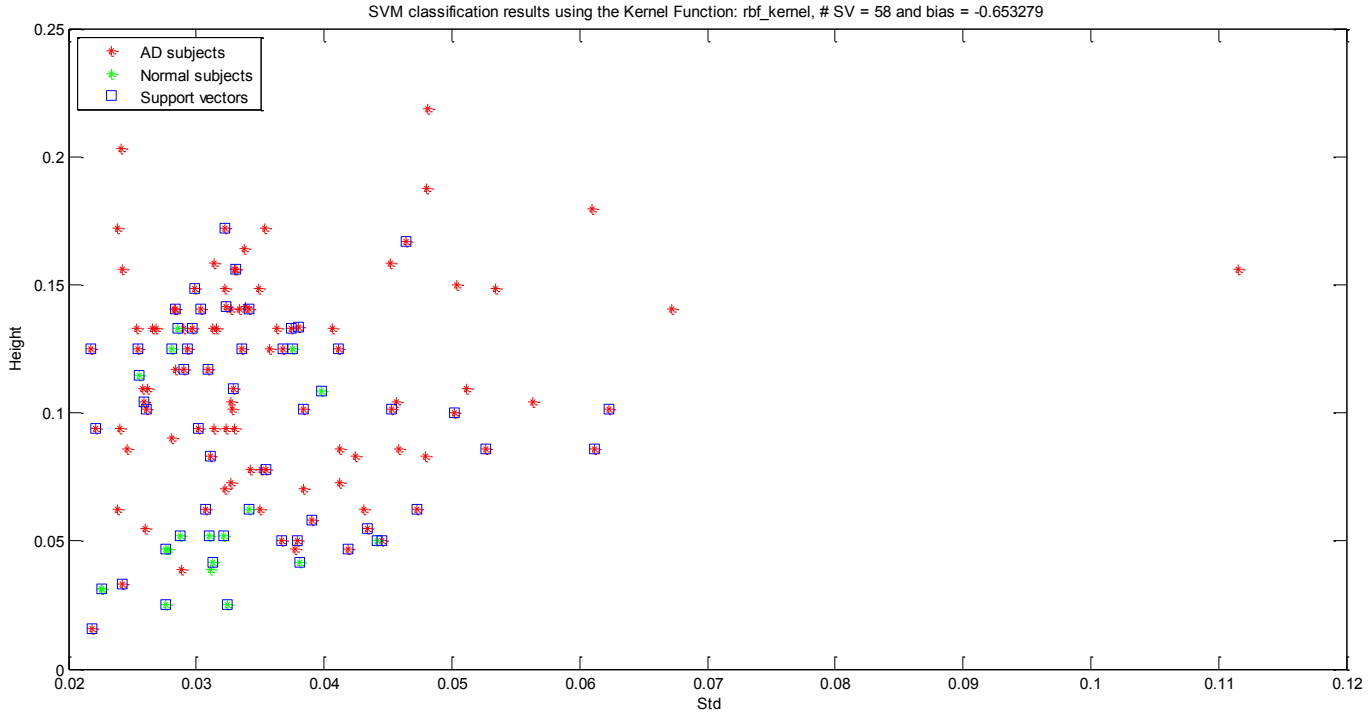
Support vector machine (SVM) results using standard deviation and width 1 attributes



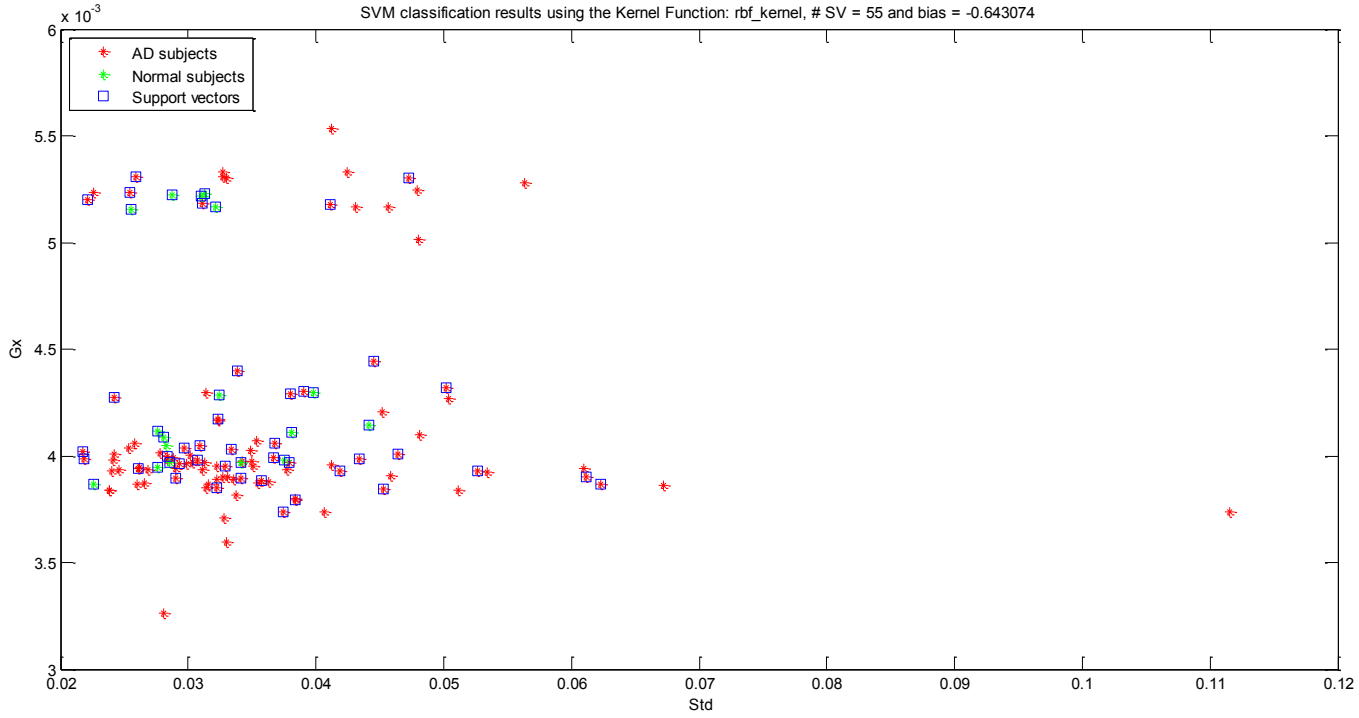
Support vector machine (SVM) results using standard deviation and width 2 attributes



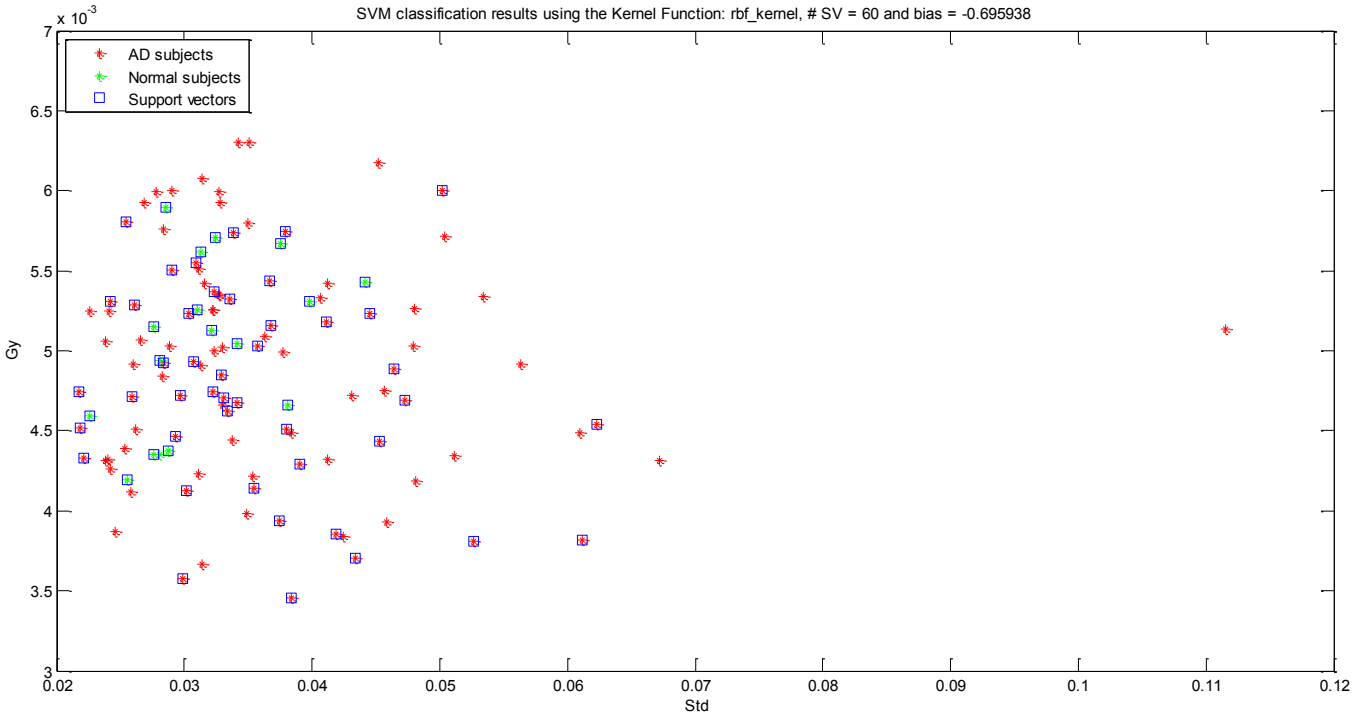
Support vector machine (SVM) results using standard deviation and height attributes



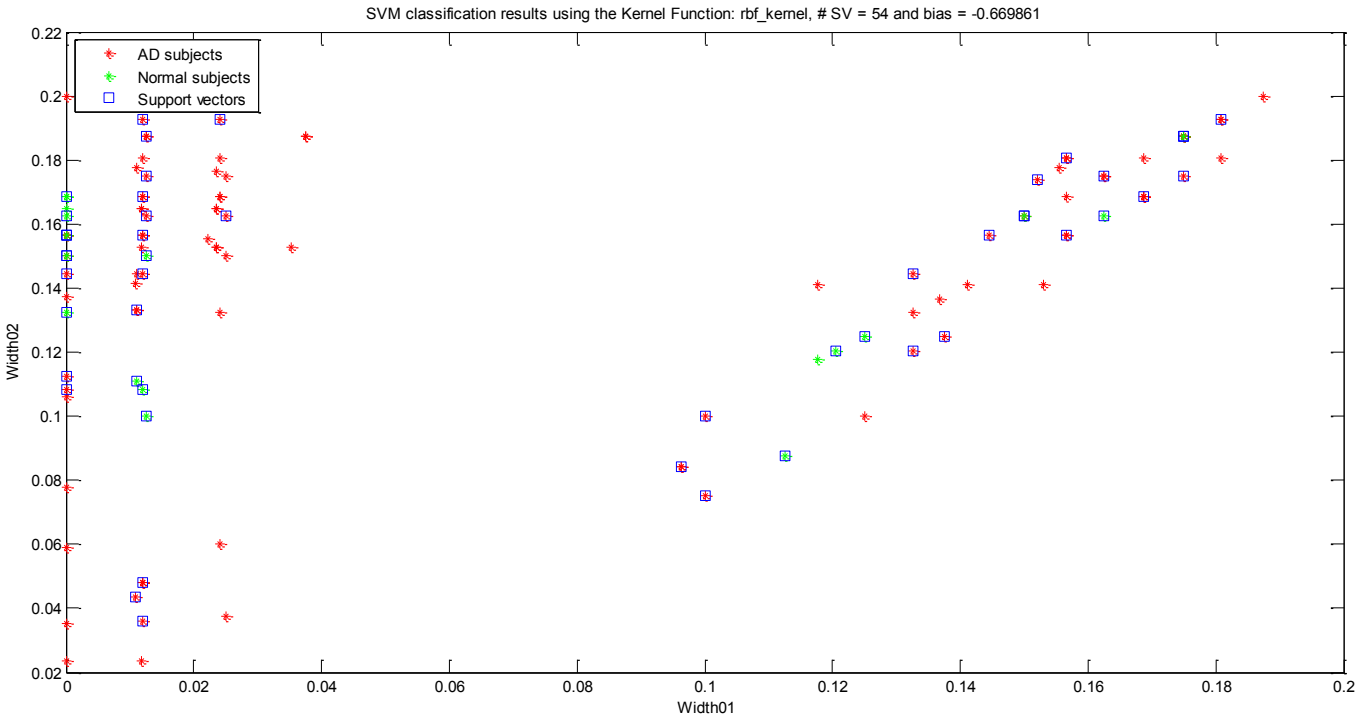
Support vector machine (SVM) results using standard deviation and Gx attributes



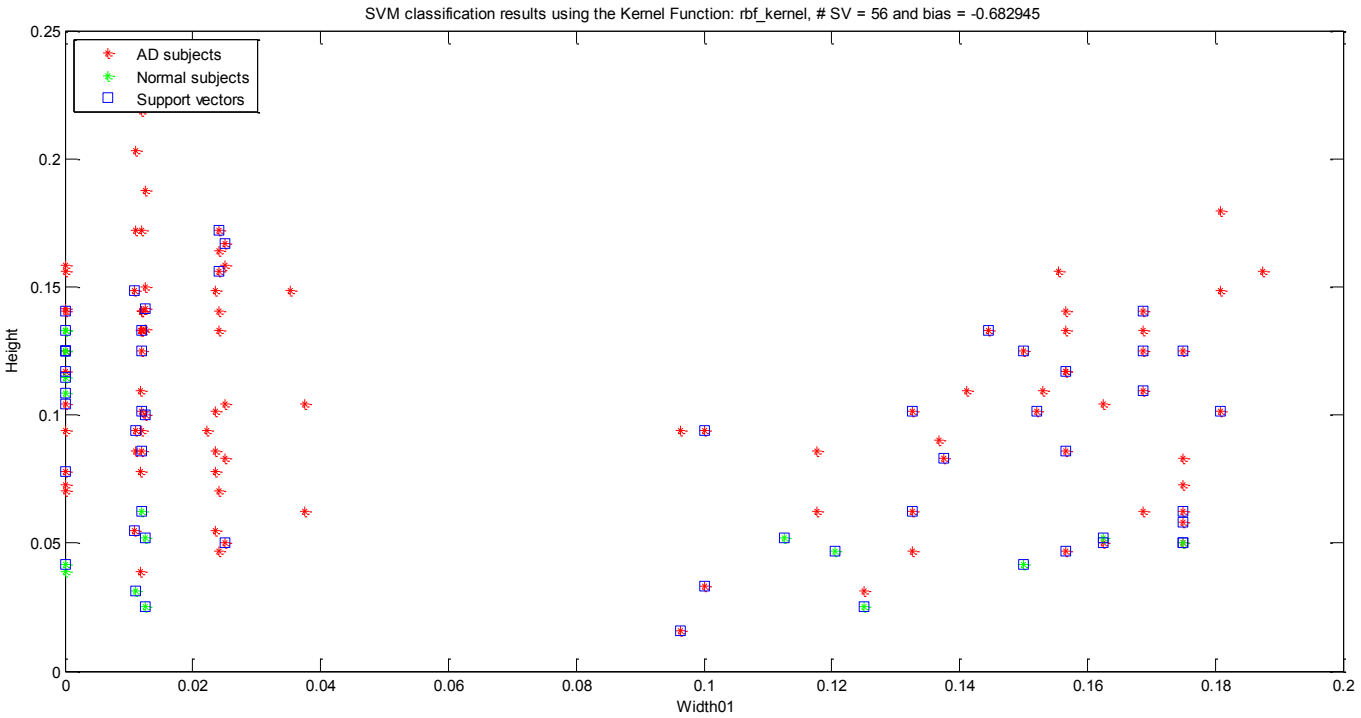
Support vector machine (SVM) results using standard deviation and Gy attributes



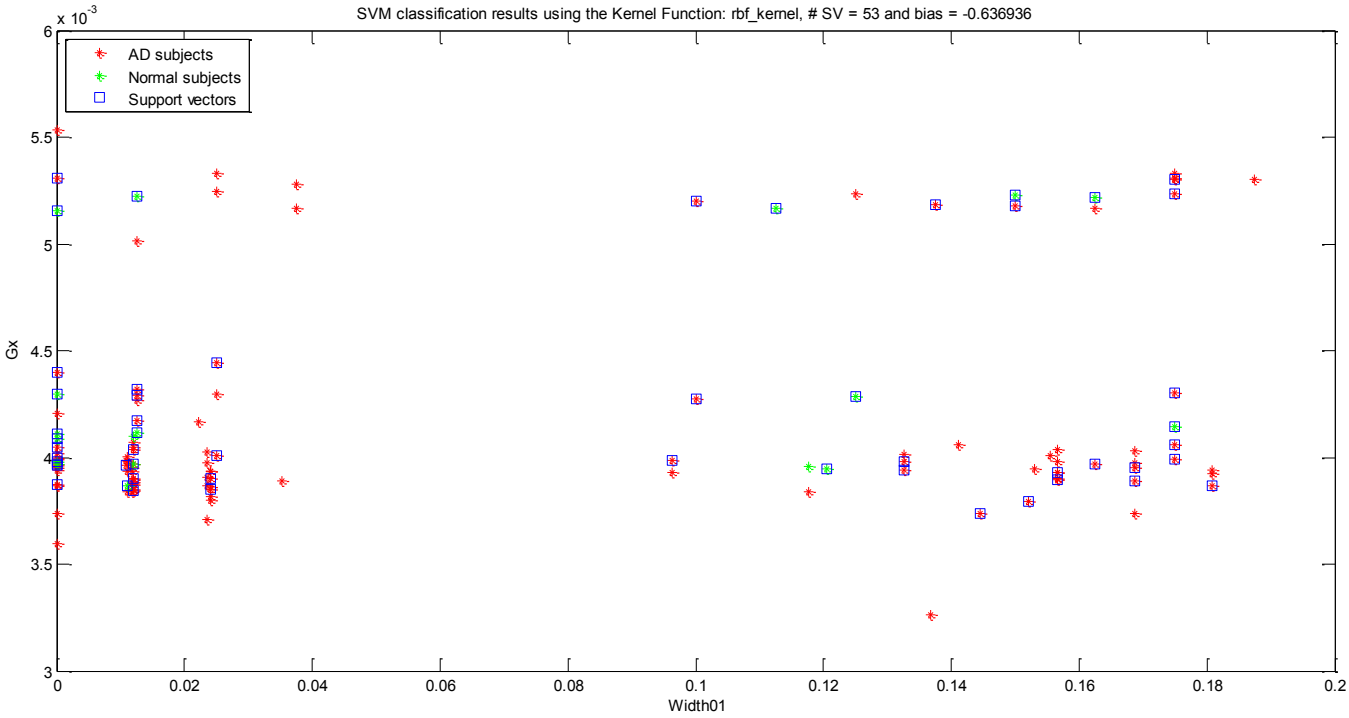
Support vector machine (SVM) results using width 1 and width 2 attributes



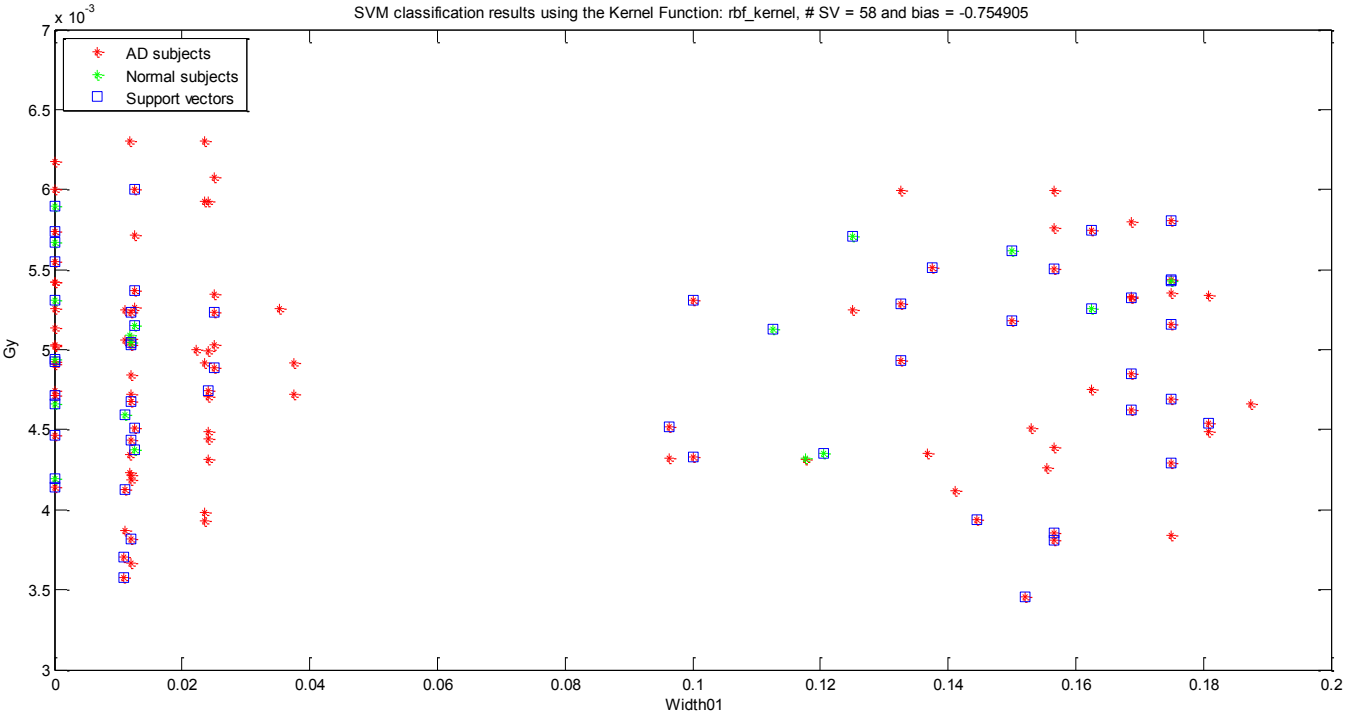
Support vector machine (SVM) results using width 1 and height attributes



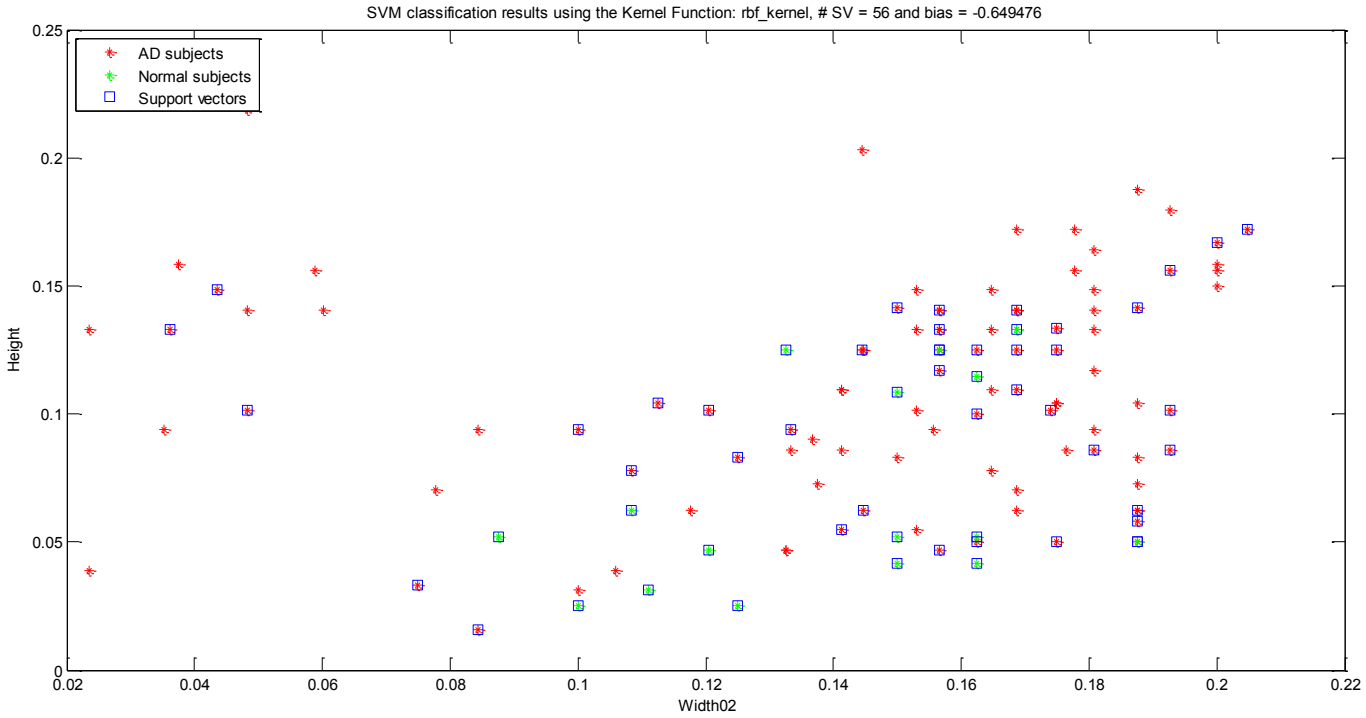
Support vector machine (SVM) results using width 1 and center of gravity Gx attributes



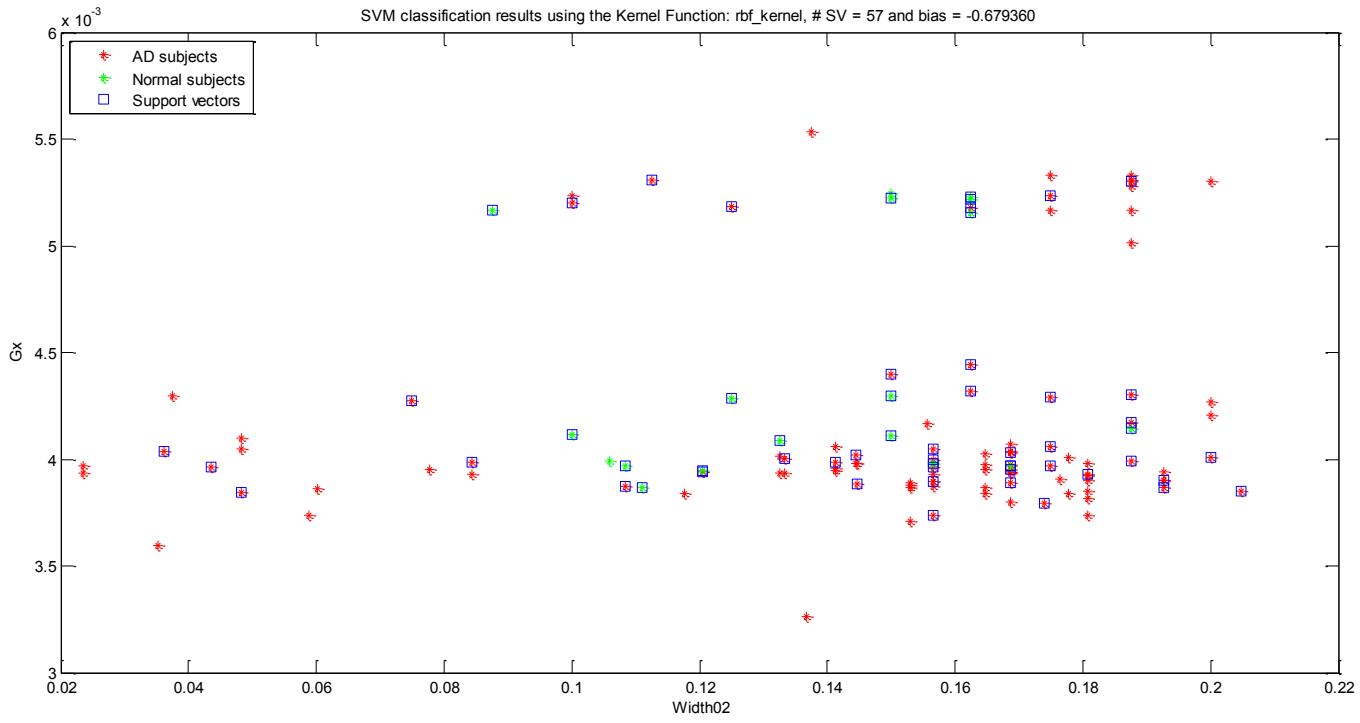
Support vector machine (SVM) results using width 1 and center of gravity Gy attributes



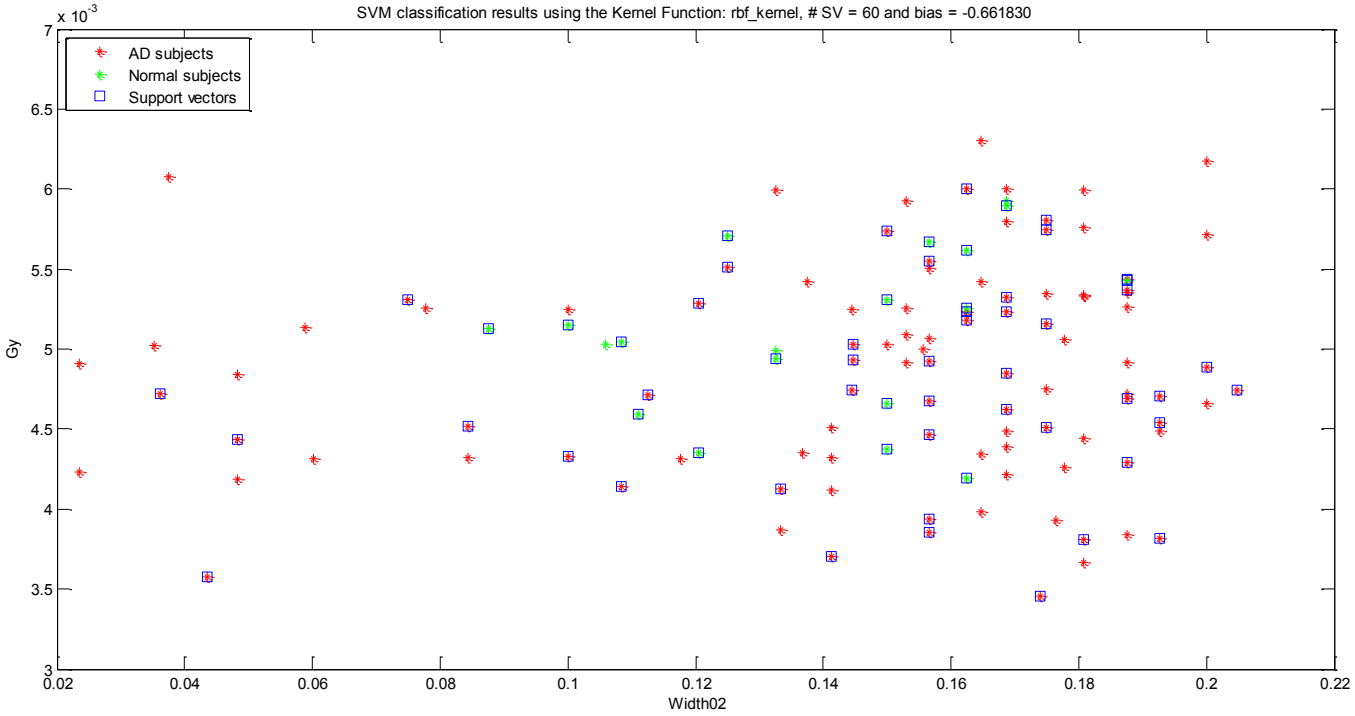
Support vector machine (SVM) results using width 2 and height attributes



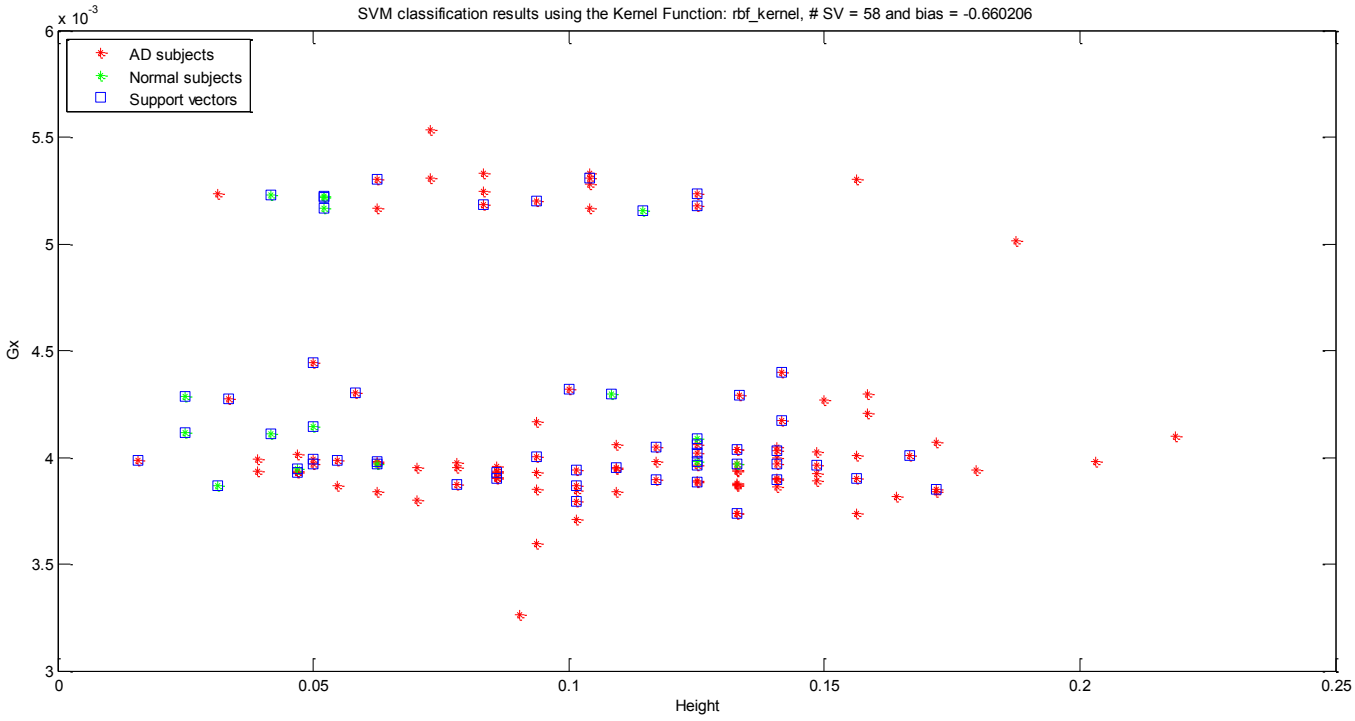
Support vector machine (SVM) results using width 2 and center of gravity Gx attributes



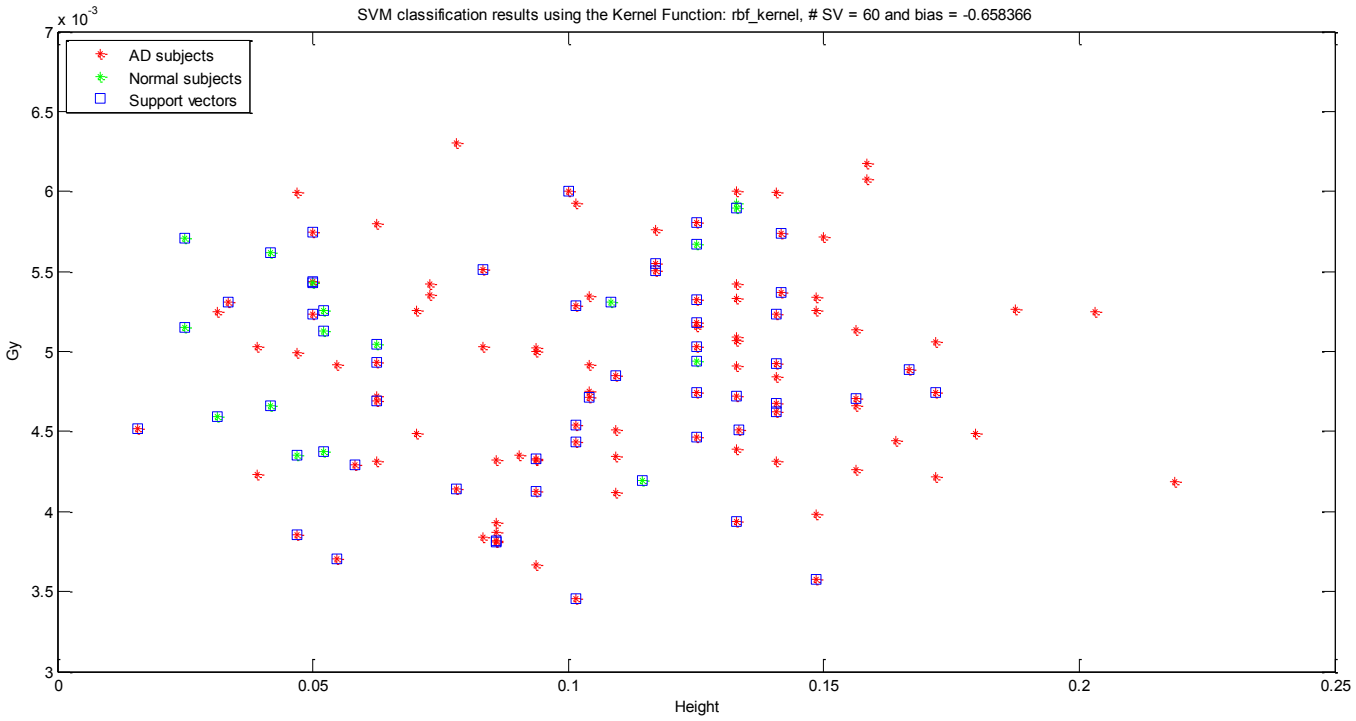
Support vector machine (SVM) results using width 2 and center of gravity Gy attributes



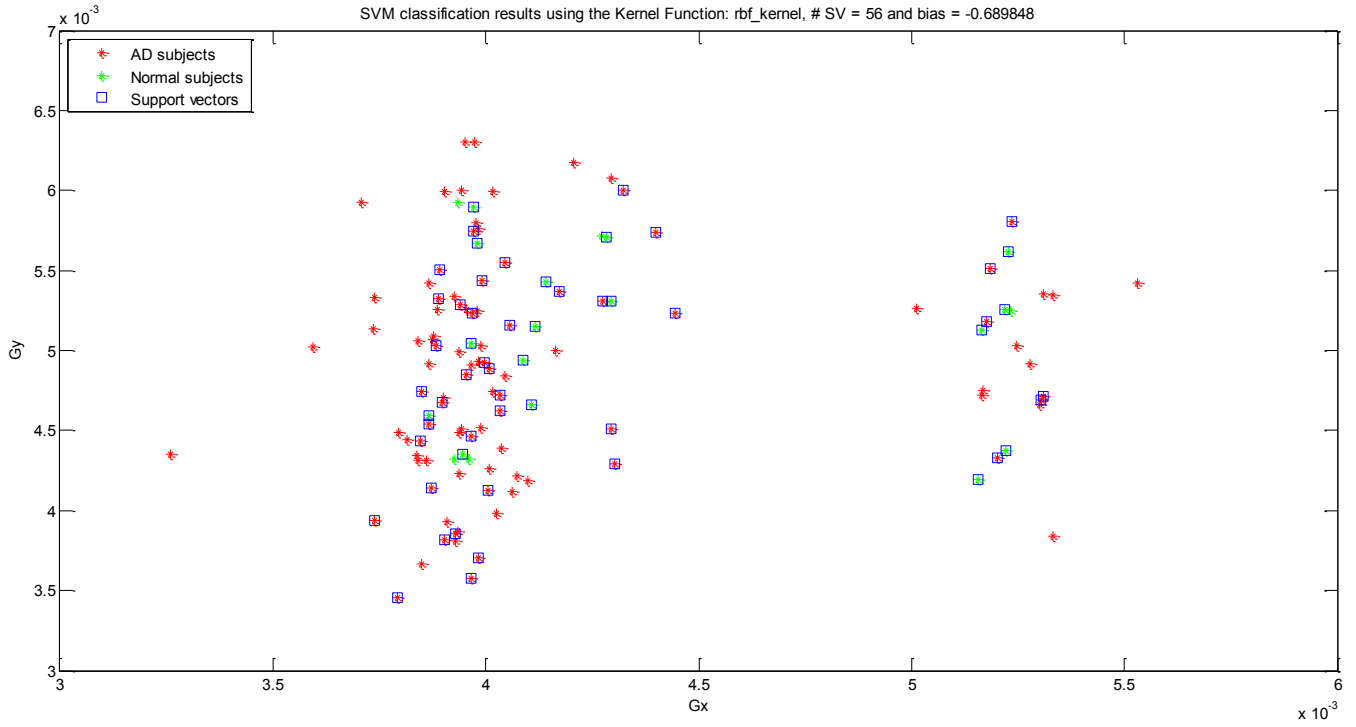
Support vector machine (SVM) results using height and center of gravity Gx attributes



Support vector machine (SVM) results using height and center of gravity Gy attributes



Support vector machine (SVM) results using center of gravity Gx and center of gravity Gy attributes



Appendix C: Performance assessment of the KNN and SVM classification results using sets of attributes, before and after applying PCA

C- 1 Performance assessment of the KNN classification technique including SN, SP, PPV, NPV and ACC classification's performance

	KNN performances				
	SN	**SP**	**PPV**	**NPV**	**ACC**
Before PCA	85.1064%	46.1538%	85.1064%	46.1538%	76.6667%
PCA : all attributes	85.1064%	30.7692%	81.6327%	36.3636%	73.3333%
Attributes 1 to 34	82.9787%	30.7692%	81.25%	33.3333%	71.6667%
Attributes 1 to 33	78.7234%	30.7692%	80.4348%	28.5714%	68.3333%
Attributes 1 to 32	74.4681%	30.7692%	79.5455%	25%	65%
Attributes 1 to 31	80.8511%	30.7692%	80.8511%	30.7692%	70%
Attributes 1 to 30	85.1064%	23.0769%	80%	30%	71.6667%
Attributes 1 to 29	87.234%	15.3846%	78.8462%	25%	71.6667%
Attributes 1 to 28	87.234%	38.4615%	83.6735%	45.4545%	76.6667%
Attributes 1 to 27	82.9787%	46.1538%	84.7826%	42.8571%	75%
Attributes 1 to 26	82.9787%	38.4615%	82.9787%	38.4615%	73.3333%
Attributes 1 to 25	78.7234%	23.0769%	78.7234%	23.0769%	66.6667%
Attributes 1 to 24	74.4681%	15.3846%	76.087%	14.2857%	61.6667%
Attributes 1 to 23	76.5957%	23.0769%	78.2609%	21.4286%	65%
Attributes 1 to 22	85.1064%	23.0769%	80%	30%	71.6667%
Attributes 1 to 21	85.1064%	23.0769%	80%	30%	71.6667%
Attributes 1 to 20	82.9787%	23.0769%	79.5918%	27.2727%	70%
Attributes 1 to 19	82.9787%	38.4615%	82.9787%	38.4615%	73.3333%
Attributes 1 to 18	76.5957%	23.0769%	78.2609%	21.4286%	65%
Attributes 1 to 17	78.7234%	38.4615%	82.2222%	33.3333%	70%
Attributes 1 to 16	74.4681%	30.7692%	79.5455%	25%	65%
Attributes 1 to 15	63.8298%	30.7692%	76.9231%	19.0476%	56.6667%

Attributes 1 to 14	78.7234%	38.4615%	82.2222%	33.3333%	70%
Attributes 1 to 13	78.7234%	53.8462%	86.0465%	41.1765%	73.3333%
Attributes 1 to 12	74.4681%	53.8462%	85.3659%	36.8421%	70%
Attributes 1 to 11	89.3617%	46.1538%	85.7143%	54.5455%	80%
Attributes 1 to 10	76.5957%	46.1538%	83.7209%	35.2941%	70%
Attributes 1 to 9	68.0851%	53.8462%	84.2105%	31.8182%	65%
Attributes 1 to 8	80.8511%	23.0769%	79.1667%	25%	68.3333%
Attributes 1 to 7	65.9574%	23.0769%	75.6098%	15.7895%	56.6667%
Attributes 1 to 6	68.0851%	38.4615%	80%	25%	61.6667%
Attributes 1 to 5	82.9787%	46.1538%	84.7826%	42.8571%	75%
Attributes 1 to 4	82.9787%	38.4615%	82.9787%	38.4615%	73.3333%
Attributes 1 to 3	80.8511%	30.7692%	80.8511%	30.7692%	70%
Attributes 1 to 2	72.3404%	23.0769%	77.2727%	18.75%	61.6667%
Attribute 1	70.2128%	53.8462%	84.6154%	33.3333%	66.6667%
Attributes 2 to 35	85.1064%	30.7692%	81.6327%	36.3636%	73.3333%
Attributes 3 to 35	85.1064%	30.7692%	81.6327%	36.3636%	73.3333%
Attributes 4 to 35	85.1064%	30.7692%	81.6327%	36.3636%	73.3333%
Attributes 5 to 35	85.1064%	30.7692%	81.6327%	36.3636%	73.3333%
Attributes 6 to 35	85.1064%	30.7692%	81.6327%	36.3636%	73.3333%
Attributes 7 to 35	85.1064%	30.7692%	81.6327%	36.3636%	73.3333%
Attributes 8 to 35	85.1064%	30.7692%	81.6327%	36.3636%	73.3333%
Attributes 9 to 35	85.1064%	30.7692%	81.6327%	36.3636%	73.3333%
Attributes 10 to 35	85.1064%	30.7692%	81.6327%	36.3636%	73.3333%
Attributes 11 to 35	85.1064%	30.7692%	81.6327%	36.3636%	73.3333%
Attributes 12 to 35	85.1064%	30.7692%	81.6327%	36.3636%	73.3333%
Attributes 13 to 35	85.1064%	30.7692%	81.6327%	36.3636%	73.3333%
Attributes 14 to 35	85.1064%	30.7692%	81.6327%	36.3636%	73.3333%
Attributes 15 to 35	82.9787%	30.7692%	81.25%	33.3333%	71.6667%
Attributes 16 to 35	82.9787%	30.7692%	81.25%	33.3333%	71.6667%
Attributes 17 to 35	82.9787%	30.7692%	81.25%	33.3333%	71.6667%

Attributes 18 to 35	82.9787%	30.7692%	81.25%	33.3333%	71.6667%
Attributes 19 to 35	82.9787%	30.7692%	81.25%	33.3333%	71.6667%
Attributes 20 to 35	82.9787%	30.7692%	81.25%	33.3333%	71.6667%
Attributes 21 to 35	82.9787%	30.7692%	81.25%	33.3333%	71.6667%
Attributes 22 to 35	82.9787%	30.7692%	81.25%	33.3333%	71.6667%
Attributes 23 to 35	82.9787%	30.7692%	81.25%	33.3333%	71.6667%
Attributes 24 to 35	82.9787%	30.7692%	81.25%	33.3333%	71.6667%
Attributes 25 to 35	80.8511%	30.7692%	80.8511%	30.7692%	70%
Attributes 26 to 35	80.8511%	38.4615%	82.6087%	35.7143%	71.6667%
Attributes 27 to 35	85.1064%	38.4615%	83.3333%	41.6667%	75%
Attributes 28 to 35	85.1064%	23.0769%	80%	30%	71.6667%
Attributes 29 to 35	82.9787%	30.7692%	81.25%	33.3333%	71.6667%
Attributes 30 to 35	78.7234%	30.7692%	80.4348%	28.5714%	68.3333%
Attributes 31 to 35	74.4681%	23.0769%	77.7778%	20%	63.3333%
Attributes 32 to 35	74.4681%	23.0769%	77.7778%	20%	63.3333%
Attributes 33 to 35	76.5957%	30.7692%	80%	26.6667%	66.6667%
Attributes 34 to 35	76.5957%	15.3846%	76.5957%	15.3846%	63.3333%
Attribute 35	70.2128%	30.7692%	78.5714%	22.2222%	61.6667%

C- 2 Performance assessment of the SVM classification technique including SN, SP, PPV, NPV and ACC classification's performance

	SVM performances				
	SN	**SP**	**PPV**	**NPV**	**ACC**
Before PCA	100%	0%	78.3333%	NaN	78.3333%
PCA : all attributes	100%	0%	78.3333%	NaN	78.3333%
Attributes 1 to 34	100%	0%	78.3333%	NaN	78.3333%
Attributes 1 to 33	100%	0%	78.3333%	NaN	78.3333%
Attributes 1 to 32	100%	0%	78.3333%	NaN	78.3333%
Attributes 1 to 31	100%	0%	78.3333%	NaN	78.3333%
Attributes 1 to 30	100%	0%	78.3333%	NaN	78.3333%
Attributes 1 to 29	100%	0%	78.3333%	NaN	78.3333%
Attributes 1 to 28	100%	0%	78.3333%	NaN	78.3333%
Attributes 1 to 27	100%	0%	78.3333%	NaN	78.3333%
Attributes 1 to 26	100%	0%	78.3333%	NaN	78.3333%
Attributes 1 to 25	100%	0%	78.3333%	NaN	78.3333%
Attributes 1 to 24	100%	0%	78.3333%	NaN	78.3333%
Attributes 1 to 23	100%	0%	78.3333%	NaN	78.3333%
Attributes 1 to 22	100%	0%	78.3333%	NaN	78.3333%
Attributes 1 to 21	100%	0%	78.3333%	NaN	78.3333%
Attributes 1 to 20	100%	0%	78.3333%	NaN	78.3333%
Attributes 1 to 19	100%	0%	78.3333%	NaN	78.3333%
Attributes 1 to 18	100%	0%	78.3333%	NaN	78.3333%
Attributes 1 to 17	100%	0%	78.3333%	NaN	78.3333%
Attributes 1 to 16	100%	0%	78.3333%	NaN	78.3333%
Attributes 1 to 15	100%	0%	78.3333%	NaN	78.3333%
Attributes 1 to 14	100%	0%	78.3333%	NaN	78.3333%
Attributes 1 to 13	100%	0%	78.3333%	NaN	78.3333%
Attributes 1 to 12	100%	0%	78.3333%	NaN	78.3333%
Attributes 1 to 11	100%	0%	78.3333%	NaN	78.3333%

Attributes 1 to 10	100%	0%	78.3333%	NaN	78.3333%
Attributes 1 to 9	100%	0%	78.3333%	NaN	78.3333%
Attributes 1 to 8	100%	0%	78.3333%	NaN	78.3333%
Attributes 1 to 7	100%	0%	78.3333%	NaN	78.3333%
Attributes 1 to 6	100%	0%	78.3333%	NaN	78.3333%
Attributes 1 to 5	100%	0%	78.3333%	NaN	78.3333%
Attributes 1 to 4	100%	0%	78.3333%	NaN	78.3333%
Attributes 1 to 3	100%	0%	78.3333%	NaN	78.3333%
Attributes 1 to 2	97.8723%	0%	77.9661%	0%	76.6667%
Attribute 1	93.617%	0%	77.193%	0%	73.3333%
Attributes 2 to 35	100%	0%	78.3333%	NaN	78.3333%
Attributes 3 to 35	100%	0%	78.3333%	NaN	78.3333%
Attributes 4 to 35	100%	0%	78.3333%	NaN	78.3333%
Attributes 5 to 35	100%	0%	78.3333%	NaN	78.3333%
Attributes 6 to 35	100%	0%	78.3333%	NaN	78.3333%
Attributes 7 to 35	100%	0%	78.3333%	NaN	78.3333%
Attributes 8 to 35	100%	0%	78.3333%	NaN	78.3333%
Attributes 9 to 35	100%	0%	78.3333%	NaN	78.3333%
Attributes 10 to 35	100%	0%	78.3333%	NaN	78.3333%
Attributes 11 to 35	100%	0%	78.3333%	NaN	78.3333%
Attributes 12 to 35	100%	0%	78.3333%	NaN	78.3333%
Attributes 13 to 35	100%	0%	78.3333%	NaN	78.3333%
Attributes 14 to 35	100%	0%	78.3333%	NaN	78.3333%
Attributes 15 to 35	100%	0%	78.3333%	NaN	78.3333%
Attributes 16 to 35	100%	0%	78.3333%	NaN	78.3333%
Attributes 17 to 35	100%	0%	78.3333%	NaN	78.3333%
Attributes 18 to 35	100%	0%	78.3333%	NaN	78.3333%
Attributes 19 to 35	100%	0%	78.3333%	NaN	78.3333%
Attributes 20 to 35	100%	0%	78.3333%	NaN	78.3333%
Attributes 21 to 35	82.9787%	7.6923%	76.4706%	11.1111%	66.6667%

Attributes 22 to 35	100%	0%	78.3333%	NaN	78.3333%
Attributes 23 to 35	100%	0%	78.3333%	NaN	78.3333%
Attributes 24 to 35	100%	0%	78.3333%	NaN	78.3333%
Attributes 25 to 35	100%	0%	78.3333%	NaN	78.3333%
Attributes 26 to 35	100%	0%	78.3333%	NaN	78.3333%
Attributes 27 to 35	100%	0%	78.3333%	NaN	78.3333%
Attributes 28 to 35	100%	0%	78.3333%	NaN	78.3333%
Attributes 29 to 35	100%	0%	78.3333%	NaN	78.3333%
Attributes 30 to 35	100%	0%	78.3333%	NaN	78.3333%
Attributes 31 to 35	100%	0%	78.3333%	NaN	78.3333%
Attributes 32 to 35	95.7447%	0%	77.5862%	0%	75%
Attributes 33 to 35	95.7447%	0%	77.5862%	0%	75%
Attributes 34 to 35	91.4894%	7.6923%	78.1818%	20%	73.3333%
Attribute 35	91.4894%	30.7692%	82.6923%	50%	78.3333%