THE PENNSYLVANIA STATE UNIVERSITY
SCHREYER HONORS COLLEGE

DEPARTMENT OF ELECTRICAL ENGINEERING

FEATURE SELECTION AND CLASSIFICATION TO AUTOMATICALLY DETECT KNEE OSTEOARTHRITIS USING MRI

MATTHEW G KEFFALAS
FALL 2010

A thesis
submitted in partial fulfillment
of the requirements
for a baccalaureate degree
in Electrical Engineering
with honors in Electrical Engineering

Reviewed and approved* by the following:

David Miller
Professor
Thesis Supervisor

Jeffrey Schiano
Associate Professor
Honors Advisor

Kenneth L. Urish
Physician, Penn State Milton S. Hershey Medical Center
Thesis Co-Supervisor

*Signatures are on file in the Schreyer Honors College

ABSTRACT

Osteoarthritis (OA) is a common joint disease, affecting roughly half of people age 55 or older. Currently, there is no reliable, noninvasive method for OA diagnosis in advance of the onset of symptoms. Recently, MRI, as an alternative to radiography, has shown promise for identifying pre-radiographic disease signatures. In this work, textural features calculated from MRIs of knee cartilage are used to train an automatic classifier that is designed to predict changes due to OA years prior to both their symptomatic presentation and radiographic detection. Also, a feature selection algorithm is used to identify a smaller feature set that has comparable performance to the initial larger feature set. The algorithm is tested by repeatedly splitting the patient data into equally-sized training and test sets, then using the training set to train a classifier and perform feature selection while the test set is used to estimate the performance. This experiment is repeated for 100 trials, and the algorithm achieves an average accuracy of 74.6% with an average sensitivity of 79.2% and average specificity of 68.5% using the selected smaller feature set. In its present state, the algorithm described in this work presents a viable method for detecting OA in the early stages of the disease, and with further development the algorithm could become a significant tool for early clinical OA diagnosis and for identifying study populations for both epidemiological and drug studies.

TABLE OF CONTENTS

Chapter 1 Introduction 1

Chapter 2 Preprocessing 5

- T2 Map Creation 7
- Registration 7
- Segmentation 7
- Mask Division 8
- Voxel Extraction and Feature Calculation 10
  - Histogram Features 11
  - Gray Level Co-Occurrence Matrix (GLCM) 12
  - Gray Level Run-Length Matrix (GLRL) 13
  - Z-Score 14
- Feature Normalization 14

Chapter 3 Feature Selection and Classification 15

- Support Vector Machine (SVM) 15
- Margin-Based Feature Elimination (MFE) 19
- Classification Experiment 20

Chapter 4 Experimental Results 22

Chapter 5 Discussion and Future Work 28

Chapter 6 Conclusion 34

References 35

LIST OF FIGURES

Figure 2-1: Block Diagram of Preprocessing……6

Figure 2-2: MRIs and segmentation masks of a patient for each compartment……9

Figure 3-1: Diagram of support vectors and SVM margin ……..17

Figure 3-2: Separating hyperplane with maximized margin …….....18

Figure 3-3: Feature selection and classification experiment……..21

Figure 4-1: Results of 100 classification trials……...23

Figure 4-2: Classifier performance as features are removed ……24

Figure 4-3: Histogram of SVM score for a single trial ……25

Figure 4-4: Histogram of compartment feature partial sums……27

LIST OF TABLES

Table 2-1: Features from each compartment……..…...10

Table 4-1: Results of classification experiment……...22

Table 5-1: Average number of selected features (section)……...29

Table 5-2: Average number of selected features (feature category)……..30

Table 5-3: Average number of selected features (feature type)……31

ACKNOWLEDGEMENTS

I would like to thank Dr. David Miller and Dr. Jeffrey Schiano for their guidance and professional expertise throughout my Bachelor's program. Their help was invaluable and their approval of this thesis was greatly appreciated. Also, I would like to thank Dr. Kenneth Urish from Penn State Milton S. Hershey Medical Center for his medical expertise and guidance throughout this thesis project.

Chapter 1

Introduction

Knee osteoarthritis (OA) is a common degenerative joint disease. This disease causes cartilage loss and inflammation that results in pain and stiffness in the joint, which in some cases can be so severe that it inhibits daily activity. Roughly half of people age 55 and older are both symptomatic and show radiographic evidence of OA, but there is also a significant population of individuals with joint pain symptoms but no radiological signs of the disease [1]. This is partly because the disease is known to present heterogeneously, which has made it difficult to diagnose in its early stages. The most common method of diagnosis is a patient presenting symptoms and a physical examination. Currently, there is no reliable, non-invasive method to detect and diagnose cartilage damage at an early stage, before the presentation of symptoms.

The most commonly applied imaging technique for evaluating the progression of OA is radiography. However, radiographs are generally only useful for detecting large joint changes that do not occur until the later stages of OA, such as joint space narrowing and osteophytes. The more subtle microstructural and molecular changes in cartilage tissue that occur during earlier stages are not visible on a radiograph [2]. This greatly reduces the effectiveness of radiographs to detect early OA symptoms. An alternative to radiography is magnetic resonance imaging (MRI), which captures the cartilage structure and its molecular biochemistry in three spatial dimensions. A specific type of MR imaging called T2 mapping probes both cartilage water content and collagen fiber orientation, which can help to assess the structural integrity of the cellular matrix [3]. T2 intensity has also been used to measure molecular changes that are useful for OA detection, including anisotropy of water and proteoglycan content [4,5,6]. In this thesis, we will

demonstrate that T2 mapping can be used to perform fairly accurate longitudinal prediction of OA status, as early as three years prior to the onset of symptoms.

While T2 maps may possess a large amount of information relevant to OA detection, there is no simple "signature" of the disease that can be easily interpreted from the images. Instead, evidence of the disease may appear as subtle changes in image texture, which could occur anywhere within the large space of voxels in the T2 map. Also, human expert reading of these (and similar) images is both expensive and time-consuming. The use of automated statistical classification techniques is directly motivated by problems of this nature, where the data is high-dimensional and manual categorization is costly [7]. Other than providing automated diagnosis, there are two other benefits to applying statistical classification to medical image analysis: 1) When validating the classifier, one must measure sensitivity and specificity, which indicate the discriminative power of the image modality and associated features used by the classifier; 2) When designing the classifier, one usually performs a feature selection step to remove unnecessary features and possibly improve the generalization accuracy of the classifier. This feature selection step can help identify "biomarker" features which can give clues to the presentation of the disease, and the spatial location of these features can help localize the disease.

The features used in this study are primarily texture measurements derived from the T2 map of the femoral cartilage. Texture is chosen because, in general, these features represent the statistical distribution of T2 intensity as well as the spatial distribution of the intensity. A voxel's intensity and its relationship to its neighbors should provide insight into the interior structure of the cartilage. Since it is not known a priori what measurements and what locations are important, there are bound to be many redundant, noisy, and possibly unhelpful features. This characteristic leads to the "curse of dimensionality" (COD), which is the idea that the relatively large number of feature dimensions will lead to the classifier overfitting the training data and therefore having poor generalization accuracy. The COD is combated in two ways in this study: 1) The chosen

classification scheme is the support vector machine (SVM), which has built-in mechanisms to prevent overfitting [8]; 2) A feature selection method called margin-maximizing feature elimination (MFE) is integrated into the training process to reduce the dimensionality of the feature space.

This study differs from the existing literature in its specific use of imaging modality and machine learning algorithms. The study uses MR imaging of the knee combined with textural features, SVMs, and wrapper feature selection, and to the best of the author's knowledge no previous study has assessed knee OA longitudinally using these methods. However, these methods have been significantly studied separately in existing literature. Pattern classification, and specifically SVM, methods have been used before in medical research with success. For example, SVMs are often used to identify subsets of genes useful for cancer diagnosis, prognosis, and discovery [9,10] and it is being investigated for its ability to automatically identify Alzheimer's disease through the use of brain MR images [11,12]. In the OA literature, measurements derived from MRI signal intensities have been a recent topic of interest. For example, [13] discovered that T2 relaxation times have a significant correlation with OA severity, and [14] attempted to use T1 sequences to separate healthy and early OA patients. Cartilage textural measurements from MRIs have been a specific research interest [15,16]. Previous studies have been conducted that used feature selection and classification to automatically identify OA using biomedical imaging. Boniatis et al [17] assessed hip OA using radiographs, textural features, and an artificial neural network classifier. This study's patient data was relatively small (only 36 patients, 18 each for healthy/symptomatic) and it made no attempt to longitudinally study the disease. Shamir et al [18] investigated the automatic classification of knee radiographs using textural features (among many others), a filtering algorithm for feature selection, and a nearest neighbor classifier. This previous study's purpose is very similar to the methods proposed in this thesis, but our use of MRI, a more robust classifier, and a wrapper

feature selection algorithm could lead to better performance. In addition, the use of linear SVMs allows us to determine the effect of each spatial location and feature, which allows us to propose a method that can identify the area of the knee with the most OA damage. The methods proposed in this thesis could also be further developed to allow for multi-class classification or regression, allowing for automatic identification of OA in various stages.

The hypothesis of this study is that MRI and SVMs can be used to predict OA changes prognostically, several years prior to symptom onset. This thesis is set up as follows: Chapter 2 explains the preprocessing needed to prepare the patient data, Chapter 3 explains the feature selection and classification scheme in detail, Chapter 4 presents the results of the experiment, Chapter 5 discusses the results and potential future research areas, and Chapter 6 concludes the thesis.

Chapter 2

Preprocessing

Patient data was selected from the Osteoarthritis Initiative (OAI) database, and the dataset consisted of 89 healthy and 121 symptomatic patients for a total of 210. The method for assigning class labels is based on the Western Ontario and McMaster Universities Arthritis Index (WOMAC), which has been shown to be a reliable and efficient tool for OA clinical trials [19]. The healthy patients are taken exclusively from the OAI control cohort and they must have a baseline WOMAC score <= 5 and a change in score at the 3-year followup <= 5. The symptomatic patients are taken exclusively from the OAI incidence cohort and they must have a baseline WOMAC score <=10 and a change in score at the 3-year followup >= 10. Two different types of MR images of the right knee of each patient are used: dual echo steady-state (DESS) images and raw T2 echoes. More information about the OAI imaging procedure can be found in Peterfy et al [20]. Before the feature selection and classification experiments can be performed, the knee images must undergo a lengthy preprocessing process that includes registration, segmentation, segmented mask division, feature calculation, and feature normalization. The block diagram of this process is shown in Figure 2-1.

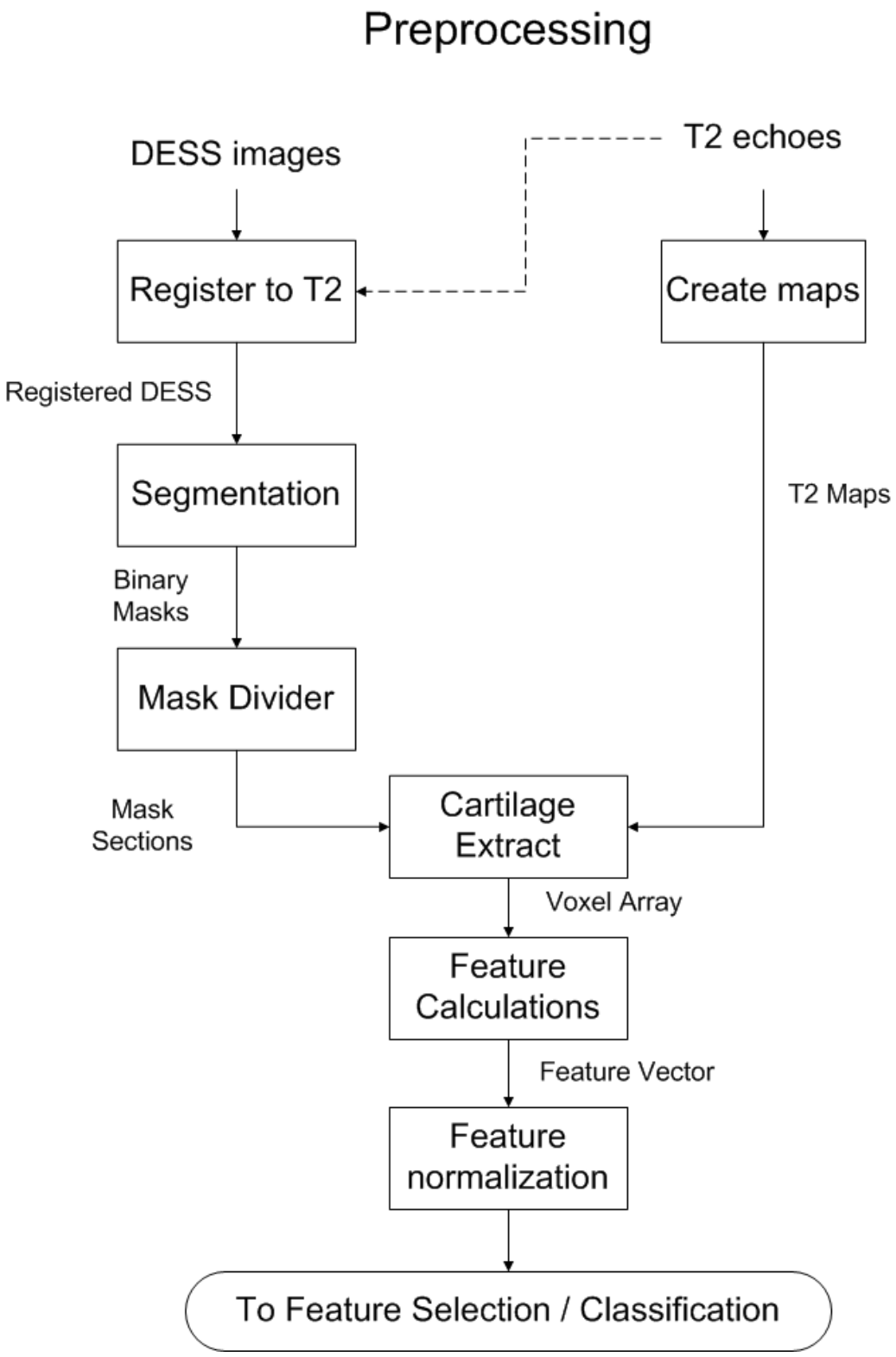


Figure 2-1: Block diagram of the preprocessing stage

T2 Map Creation

The OAI database only provides raw T2 echoes. These raw echoes were converted to T2 maps with the use of qMRI, developed at the Penn State Hershey Center for NMR Research [21]. The same qMRI settings were used for each patient. Each patient has a minimum of 27 T2 maps (taken at consecutive sagittal slices through the right knee) with some patients having as many as 33. Only the first 27 maps are used in the next preprocessing steps since it was found that almost all of the cartilage for every patient is contained in these first slices.

Registration

Even though the features used in this study are derived from the T2 map signal intensities, different tissues in the T2 images do not have a significant contrast from the cartilage structure. This would make segmentation using T2 maps extremely difficult. Instead, the DESS images are traditionally used for cartilage segmentation since there tends to be a better separation between the cartilage and surrounding tissue. However, the raw DESS and T2 images from the OAI database are not registered to each other, so registration must be performed so that the DESS image of the cartilage aligns with the T2 map. The registration is fully automated and uses the algorithm in [22].

Segmentation

After registration, the registered DESS images are used to create segmentation masks for the medial, lateral, and patella compartments of the femoral cartilage. The segmentation algorithm is semi-automatic (requiring user interaction) and is presented in [23]. The author of this thesis segmented all 210 patients. In general, there are a total of 27 registered DESS images

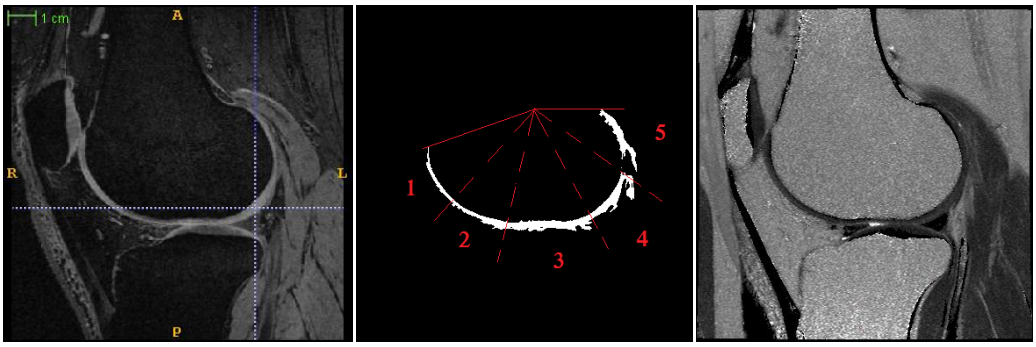
per patient (taken at consecutive sagittal slices through the right knee) and the majority of these images contain usable cartilage regions from at least one of the three compartments.

Mask Division

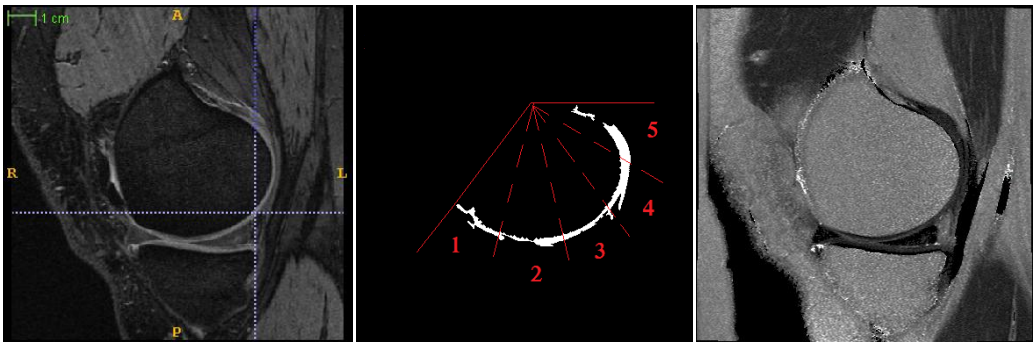
The segmented masks for a single slice in the lateral, patella, and medial compartments for a single patient are shown in Figure 2-2, as well as the corresponding registered DESS image and T2 map. From this figure one can see that the cartilage region for the lateral and medial compartments resembles an arc or semi-circle. In order to add an extra level of spatial localization ability to the feature set, the lateral and medial masks are divided into 5 sections each. The expected section boundaries are also shown in Figure 2-2, as well as the standard numbering scheme for the sections. Note that the medial and lateral masks are divided independently, so the section boundaries for the medial compartment are not necessarily the same as the ones for the lateral compartment. The automatic mask division algorithm is as follows:

1. Find all segmented masks within the current patient's current compartment (medial or lateral).
2. Superimpose all segmented masks onto a single image.
3. Find θ, the angle of the "arc" of cartilage.
4. Divide θ by 5 to find θ_S, and then draw section boundaries at intervals of θ_S.
5. Starting on the far left and rotating counterclockwise, number the sections 1-5.

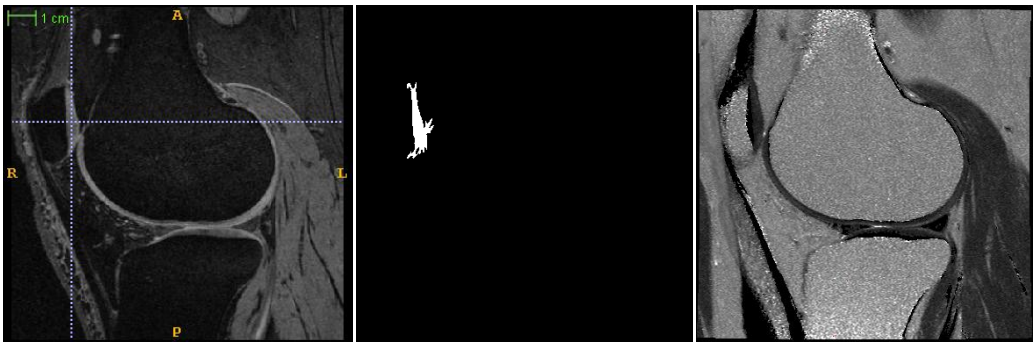
Note that the patella masks are usually much smaller than the lateral and medial masks, so they are not divided and are treated as a single section. Therefore, there are 5 medial sections, 5 lateral sections, and 1 patella section for a total of 11 sections. Each feature is measured independently in each section, so in general there are 11 instances of each feature.



(a)



(b)



(c)

Figure 2-2: Registered DESS, segmented mask, and T2 map for: a) lateral b) medial c) patella. The crosshairs in the DESS images point to the cartilage. The mask section boundaries are not drawn to scale.

Voxel Extraction and Feature Calculation

Each mask is used to extract the appropriate voxels from the corresponding T2 map. As mentioned before, there are 5 medial sections, 5 lateral sections, and 1 patella section for a total of 11 sections that voxels can originate from. Features are measured independently for the voxels from each section, creating 11 instances of the same feature, each from a different location in the right knee.

Primarily, textural features were chosen to represent the images. This is because texture can measure statistical properties and spatial distribution of the image intensities [24]. This fact makes textural features very attractive for predicting OA status from the T2 map, and many of the initial features have been used in previous OA studies. The initial feature set can be split into four categories: 1) histogram; 2) gray level co-occurrence matrix (GLCM); 3) gray level run-length matrix (GLRL); 4) z-score. All feature calculations are performed with custom-made Matlab functions. There are 725 total features in the initial feature set, and the distribution among the four categories and 11 sections is shown in Table 2-1. In the following sections, let $I_j(u,v,w)$ be the intensity of the pixel at index (u,v,w) in section j and S_j is the set of all voxel indices in section j.

	Histogram	**GLCM**	**GLRL**	**Z-Score**
Lateral	115	150	40	25
Patella	22	30	8	5
Medial	115	150	40	25
Total	252	330	88	55

Table 2-1: Total number of features in each category for each knee compartment.

Histogram Features

For each section, a 32-bin histogram is calculated from the intensity values of the T2 maps. The histogram in section j is calculated according to the equation:

$$p_j(x) = \frac{\text{number of voxels in section j with quantized gray level } q_x \text{ in } S_j}{\text{total number of voxels in } S_j}, \quad x = 0, 1, \ldots, 31 \qquad (2\text{-}1)$$

Note that within section j, the histogram p_j actually does not depend on the spatial location of the voxels. This means that all histogram features do not depend on the location of the voxels within their specified section, but instead the histogram features measure the statistical properties of the voxel intensities. Even though there is no spatial dependency, histogram features are still often used in texture measurement. Also, some level of spatial information is included in this feature category since each feature is measured independently in each section, so each feature instance is localized to one specific location in the knee. The following equations define the histogram features and are found in [25].

Mean: $$m_j = \sum_{x=0}^{31} x p_j(x) \qquad (2\text{-}2)$$

Variance: $$\sigma_j^2 = \sum_{x=0}^{31} (x - m_j)^2 p_j(x) \qquad (2\text{-}3)$$

Dispersion: $$\sum_{x=0}^{31} |x - m_j| \, p_j(x) \qquad (2\text{-}4)$$

Average Energy: $$\sum_{x=0}^{31} x^2 p_j(x) \qquad (2\text{-}5)$$

Energy: $$\sum_{x=0}^{31} [p_j(x)]^2 \qquad (2\text{-}6)$$

Entropy: $$-\sum_{x=0}^{31} p_j(x) \log_2(p_j(x)) \qquad (2\text{-}7)$$

Skewness: $$\sigma_j^{-3} \sum_{x=0}^{31} (x - m_j)^3 p_j(x) \qquad (2\text{-}8)$$

Kurtosis: $$\sigma_j^{-4} \left(\sum_{x=0}^{31} (x - m_j)^4 p_j(x) \right) - 3 \qquad (2\text{-}9)$$

In addition to these features, the median, mode, minimum value, maximum value, and range of values is calculated from the histogram. An 8-bin histogram is also calculated for each section j and the occupancy of each bin is used as a feature. Also, the following two additional

miscellaneous features are included in this category even though they are not calculated using the histogram.

Relative Size: $$\frac{\text{number of voxels in } S_j}{\text{number of voxels in corresponding compartment}} \quad (2\text{-}10)$$

L2 norm: $$\sqrt{\sum_{(u,v,w)\in S_j}\left[I_j(u,v,w)\right]^2} \quad (2\text{-}11)$$

Note that the "relative size" feature is not calculated for the patella section because there is only a single section in the patella compartment, which would make this feature equal to 1 for all patients.

Gray Level Co-Occurrence Matrix (GLCM)

Histogram features alone cannot completely characterize texture since they do not measure the spatial characteristics of the cartilage region. The second-order histogram, called the gray level co-occurrence matrix (GLCM), is a common tool for measuring cartilage [26]. In this study, the GLCM is calculated for a distance of 1 (the voxels must be immediate neighbors in the specified direction) and for direction $\theta = 0^o$, 45^o, 90^o, 135^o, and 90^o in the z (third dimension) direction. Before the GLCM is calculated, the intensities are quantized down to 8 gray levels, which results in each GLCM being an 8x8 matrix. The GLCM for section j is defined as:

$$h_{j,\theta}(x,y) = \#\text{ of pairs of voxels in } S_j \text{ in direction } \theta \text{ with intensities x and y} \quad (2\text{-}12)$$

If $h_{j,\theta}(x,y)$ is divided by the total number of neighboring pixels in section j, then the GLCM becomes an estimate of the joint probability $f_{j,\theta}(x,y)$. The following features are calculated from $f_{j,\theta}(x,y)$. These features were originally proposed in [26] and the following equations use notation from [24].

Angular Second Moment: $$\sum_{x=0}^{7}\sum_{y=0}^{7}\left[f_{j,\theta}(x,y)\right]^2 \quad (2\text{-}13)$$

Contrast: $\sum_{x=0}^{7}\sum_{y=0}^{7}(x-y)^2 f_{j,\theta}(x,y)$ (2-14)

Absolute Value: $\sum_{x=0}^{7}\sum_{y=0}^{7}|x-y| f_{j,\theta}(x,y)$ (2-15)

Inverse Difference: $\sum_{x=0}^{7}\sum_{y=0}^{7}\frac{f_{j,\theta}(x,y)}{1+(x-y)^2}$ (2-16)

GLCM Entropy: $-\sum_{x=0}^{7}\sum_{y=0}^{7} f_{j,\theta}(x,y)\log_2 f_{j,\theta}(x,y)$ (2-17)

Correlation: $\sum_{x=0}^{7}\sum_{y=0}^{7}\frac{xyf_{j,\theta}(x,y)-\mu_u\mu_v}{\sigma_u\sigma_v}$ (2-18)

Note that μ_u, μ_v and σ_u, σ_v are the means and standard deviations of the marginal distributions created by the row and column sums of the matrix, respectively. There are 5 different GLCMs (1 for each direction) calculated for each section, and since there are 11 sections this means that there are 55 total GLCMs calculated for a single patient. Each feature is measured independently for each GLCM.

Gray Level Run-Length Matrix (GLRL)

The gray level run-length matrix (GLRL) is another typical tool used in texture analysis. In fact, features derived from this matrix were used in a previous paper studying the application of an artificial neural network classifier to detect hip OA [17]. The GLRL $g_{j,\theta}(x,y)$ is defined as the number of runs of length y in the direction θ consisting of points with gray level x in section j [27]. Before the GLRL is calculated, the cartilage image intensities are quantized down to 8 gray levels. The GLRL is calculated for $\theta=0^{\circ},90^{\circ}$. The largest possible run length is the largest number of voxels that lie in direction θ, but the run lengths are quantized to 4 possible ranges. Let P_j be the total number of voxels in section j, and let $N_{j,\theta}$ be the sum of all elements in $g_{j,\theta}$. The following features are calculated from the GLRL $g_{j,\theta}(x,y)$. They were originally proposed in [27].

Short Runs Emphasis: $\sum_{x=0}^{7}\sum_{y=0}^{3}\frac{g_{j,\theta}(x,y)}{y^2}\Big/N_{j,\theta}$ (2-19)

Long Runs Emphasis: $\sum_{x=0}^{7}\sum_{y=0}^{3} y^2 g_{j,\theta}(x,y)/N_{j,\theta}$ (2-20)

Gray Level Nonuniformity: $\sum_{x=0}^{7}(\sum_{y=0}^{3} g_{j,\theta}(x,y))^2/N_{j,\theta}$ (2-21)

Run Percentage: $\sum_{x=0}^{7}\sum_{y=0}^{3} g_{j,\theta}(x,y)/P_j$ (2-22)

There are 2 different GLRLs per section, which means there are 22 total GLRLs calculated per patient. Each feature is calculated independently for each GLRL.

Z-Score

[13] proposed a special normalization procedure that produced features that had a significant correlation with OA damage. This normalization is defined as:

$$\widehat{I_j}(u,v,w) = \frac{I_j(u,v,w) - \mu_{j,control}}{\sigma_{j,control}} \quad (2\text{-}23)$$

where $\mu_{j,control}$ and $\sigma_{j,control}$ are the mean and standard deviation of section j for only the patients in the control group. This can be seen as normalization to "healthy" voxels, and therefore it may help the "symptomatic" voxels stand out more than normal. From the transformed intensities $\widehat{I_j}$, the features calculated are the mean, variance, minimum value, maximum value, and range of values. Similar to the histogram features, the z-score features do not have the ability to identify spatial characteristics of the voxels within section j. However, since the features are measured independently from each section, each feature instance is a localized measurement.

Feature Normalization

For each feature independently, each feature instance is normalized to the range [-1,1]. This is done because there might simultaneously be very large and very small feature values, which could cause numerical problems in the classifier training procedure.

Chapter 3

Feature Selection and Classification

Given the features, we need to calculate a classification decision for each patient. The classification decision is reduced to a binary decision (healthy or symptomatic) for simplicity. In this case, each patient is treated as a 725-dimensional feature vector and the classifier maps this vector into a binary value. Since the feature space is very high-dimensional (especially considering there are only 210 patients in this 725-dimensional space), decreasing the dimensionality of this space could potentially lead to better generalization accuracy. Feature selection and classifier training are performed using a training set of patients, and then the generalization accuracy is estimated on a patient test set that is disjoint from the training set. This chapter is set-up as follows: the following section describes the support vector machine (SVM) classifier and the reason it was chosen, then the next section describes the feature selection algorithm margin-based feature elimination (MFE), and then the final section explains the structure of the feature selection and classification experiment.

Support Vector Machine (SVM)

Each patient is represented as a data point in k-dimensional space, with k being the number of features. As mentioned previously, the number of initial features before the feature selection phase is 725 and the number of classes is 2. For now, assume these data points are linearly separable; in other words, the two classes can be separated using a single linear surface (hyperplane) of dimension k-1. This separating hyperplane essentially splits the k-dimensional space in two, with each subspace corresponding to one of the two classes. This hyperplane is

known as the linear discriminant function (LDF), and in the case of SVM this hyperplane has the form

$$0 = \mathbf{w}^{\mathbf{T}}\mathbf{x} + b \tag{3-1}$$

where **w** is the kx1 SVM weight vector, **x** is the kx1 feature vector, and b is a scalar bias term. In this case, **w** is a vector normal to the hyperplane and $|b|/\|\mathbf{w}\|$ is the perpendicular distance between the hyperplane and the origin where $\|\mathbf{w}\|$ is the Euclidean norm of the weight vector.

The choice of this separating hyperplane is not unique, and the choice of a particular hyperplane is the strength of SVM. A brief overview of the derivation is included below. For a more detailed explanation of SVM theory consult [28,29].

Since there are two classes, there are two class labels: -1 and +1, corresponding to healthy and symptomatic respectively. Let y_i be the class label for patient i and $\mathbf{x}_i$ be the feature vector for this patient. The SVM decision score for patient i is d_i and is written as

$$d_i = \sum_{j=0}^{k-1} w_j x_{i,j} + b \tag{3-2}$$

and the actual classification decision for patient i is

$$\widehat{d_i} = \mathrm{sgn}(d_i) \tag{3-3}$$

where sgn is the signum function. For all training patients in a linearly separable data set, the following inequality holds:

$$y_i d_i \geq 1 \tag{3-4}$$

It can be shown that the patients from each class closest to the hyperplane are a perpendicular distance $1/\|\mathbf{w}\|$ away from it. This distance is called the margin of the hyperplane, and these patients are called the support vectors. The structure of this separating hyperplane, support vectors, and the margin are illustrated in Figure 3-1 for the 2-dimensional case. The SVM procedure chooses a separating hyperplane such that the margin is maximized while satisfying the

inequalities in (3-4) for every training instance. This can be stated formally as an optimization problem:

$$\text{Minimize } \|\mathbf{w}\|^2$$

$$\text{subject to } y_i d_i \geq 1 \qquad (3\text{-}5)$$

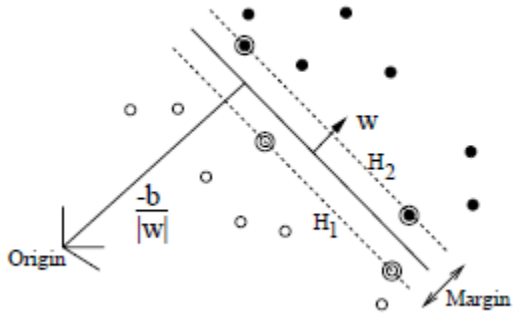


Figure 3-1: 2-Dimensional data space with a separating hyperplane, support vectors, and margin. The hyperplane is the solid line and the support vectors are circled. Originally appeared in [28].

The hyperplane is chosen to maximize margin because this hyperplane is proven to have better accuracy on unseen data points. Refer to Figure 3-2. The hyperplane on the left in this figure separates the training data perfectly, but the margin is very small. Since we do not know the true distribution of data points in the feature space, the unseen data points may easily cross the separator and therefore be classified incorrectly. The hyperplane on the right in the figure has a maximized margin. As you can see, this separator appears to account for the training set distribution more accurately and therefore this hyperplane should have a better generalization accuracy. This is why the separating hyperplane is chosen such that the margin between support vectors is maximized: the hyperplane with largest margin will typically account for the

distribution of the classes better than other separators, and therefore it should perform better on unseen test data.

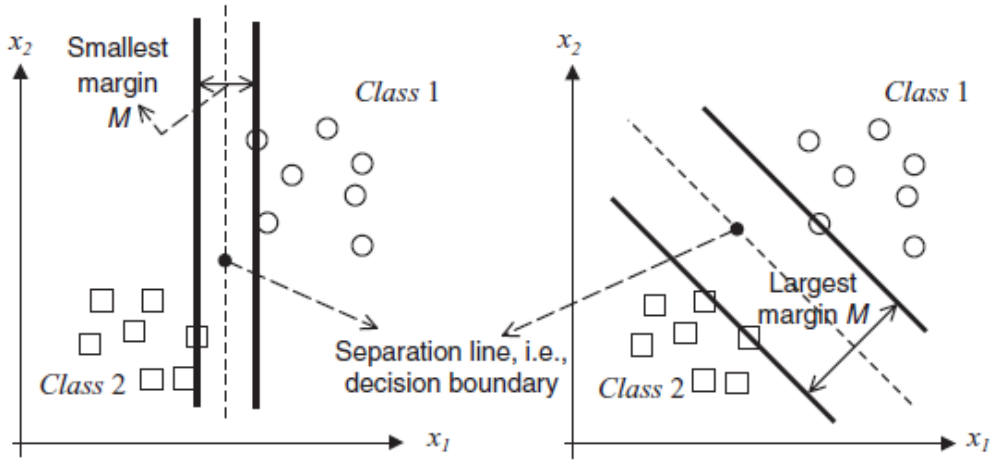


Figure 3-2: Two different separating hyperplanes. The hyperplane on the left has a small margin, while the one on the right has a maximum margin. Originally appeared in [29].

Other than maximizing margin between the classes and separating hyperplane, the strength of SVMs is the use of support vectors. In the derivation of the weight vector **w** and bias b, the support vectors are the only data points that affect **w** and b. This means that the classifier is uniquely determined by the choice of support vectors and only the support vectors. The number of model parameters in the classifier derivation depends on the number of support vectors. Therefore, unlike in other classifiers, the number of model parameters is not determined by the feature dimensionality and is instead bounded by the number of training instances, which causes the SVM to be more robust against overfitting. This characteristic makes SVMs a very attractive classification method and is the reason why SVMs have been used extensively in many different pattern recognition tasks.

Note that only the case where the training data is linearly separable is considered above. However, when the data is nonlinearly separable or when there are outliers in the training set that violate the decision boundary, this can be overcome by introducing "slackness" or nonlinear kernels. Slackness basically allows some training points to be misclassified, and kernels map the training points to a higher dimensional space to allow for a nonlinear decision boundary. In this study, the initial data is linearly separable and the feature selection process terminates just before the data loses separability, so slackness and kernels are not used for the SVM training process.

SVM training and classification is performed using the LIBSVM Matlab interface [30]. This software has been used extensively for SVM research and it is an accepted SVM software implementation.

Margin-Based Feature Elimination (MFE)

Feature selection is integrated into the classifier training process for this study. Selecting a subset of features from the initial feature set is necessary in order to eliminate redundant and non-informative features, and it is also possible to improve the generalization performance of the classification process. A review of feature selection methods can be found in [31].

Since margin-maximization is the goal of the SVM training procedure, a feature selection method was chosen that uses this margin as the criterion for removing features. This algorithm is called margin-based feature elimination (MFE), and it was shown to outperform other SVM-based algorithms, such as recursive feature elimination (RFE), on a number of UC Irvine datasets [32]. The goal of MFE is to maximize the SVM margin with each feature elimination step. This is a wrapper algorithm, meaning the trained classifier is used to determine the order of feature removal. In other words, classifier training is a part of the feature selection process, and the classifier is retrained many times to determine the usefulness of the features. Wrapper methods

were shown to outperform filter-based methods (features are removed before training) on some datasets [33]. Therefore, MFE represents a computationally-efficient feature elimination algorithm that works together with the goal of the SVM training procedure and has been shown to perform well on standard datasets.

The MFE algorithm for linear SVMs is found in [32]. The algorithm is implemented using a custom-made Matlab function. A simplified version of the algorithm is explained below.

1. Train a SVM using the current feature set.
2. Calculate the effect on the SVM decision function of removing each feature separately.
3. Remove the feature whose removal results in the largest SVM margin.
4. If linear separability is lost, stop.
5. Go to step 1, using the reduced feature set.

Note that in the version of MFE used in this study, the feature elimination is terminated when the training data becomes linearly nonseparable. MFE could be altered to continue removing features after this point, but this would require introducing slackness into the SVM training procedure. SVM training and MFE is simplified by not using slackness, and it was found experimentally that this method works well without considering slackness.

Classification Experiment

The experiment is designed to estimate the performance of the combined SVM/MFE procedure. We must do this by using the 210 fully-preprocessed patients. According to the suggestions in [34], the patient set used to test the classifier performance must be completely disjoint from the patient set that is used to train the classifier and perform feature selection. Therefore we test the performance of the classification and feature selection methods by splitting

the entire patient set into two equally-sized disjoint sets, training and test. The training set is formed by randomly sampling (without replacement) the full dataset. Each set has the same ratio of classes as the original dataset. The training set is used to select the sparse feature set and train the SVM, and the test set is used to estimate the generalization accuracy of the trained classifier and the selected feature set. This procedure is repeated for 100 trials using 100 different training sets (and therefore 100 different test sets), and the final performance of these methods are estimated by averaging the results from each trial. A block diagram of this experiment is shown in Figure 3-3.

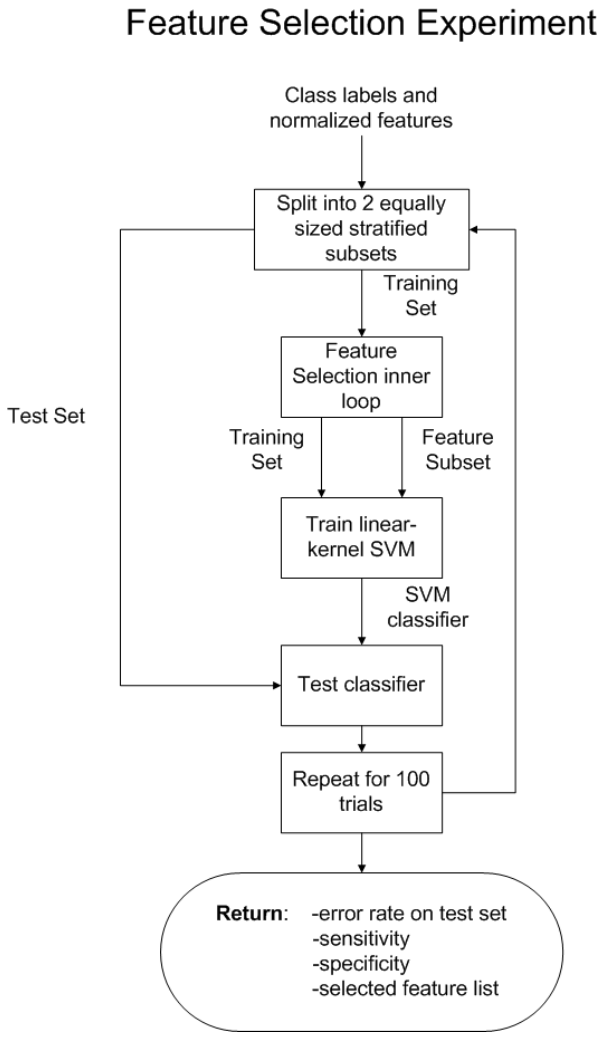


Figure 3-3: Classification and feature selection experiment.

Chapter 4

Experimental Results

Using all 725 features, the average accuracy was 79.1% with a standard deviation of 4.1%. There were 10.95 false positives and 10.95 false negatives on average, corresponding to an average sensitivity of 81.8% and an average specificity of 75.7%. After MFE was used to remove features, on average only 19.8 of the 725 features were retained. The average accuracy of the system with MFE feature selection dropped to 74.6% (standard deviation of 4.3%), with average sensitivity of 79.2% and average specificity of 68.5%. These results are summarized in Table 4-1. The accuracy, sensitivity, specificity, and number of features selected for each trial is shown in Figure 4-1.

	Using Full Feature Set	**Using MFE Feature Set**
Average # Features Used	725	19.8
Average Accuracy (+/- Stand. Dev.)	79.1% (+/- 4.1%)	74.6% (+/- 4.3%)
Average True Positives	49.05	47.53
Average False Positives	10.95	14.19
Average True Negatives	34.05	30.81
Average False Negatives	10.95	12.47
Average Sensitivity (+/- Stand. Dev.)	81.75% (+/- 5.7%)	79.2% (+/- 7.2%)
Average Specificity (+/- Stand. Dev.)	75.7% (+/- 7.2%)	68.5% (+/- 7.3%)

Table 4-1: Results for the 100 trial classification experiment

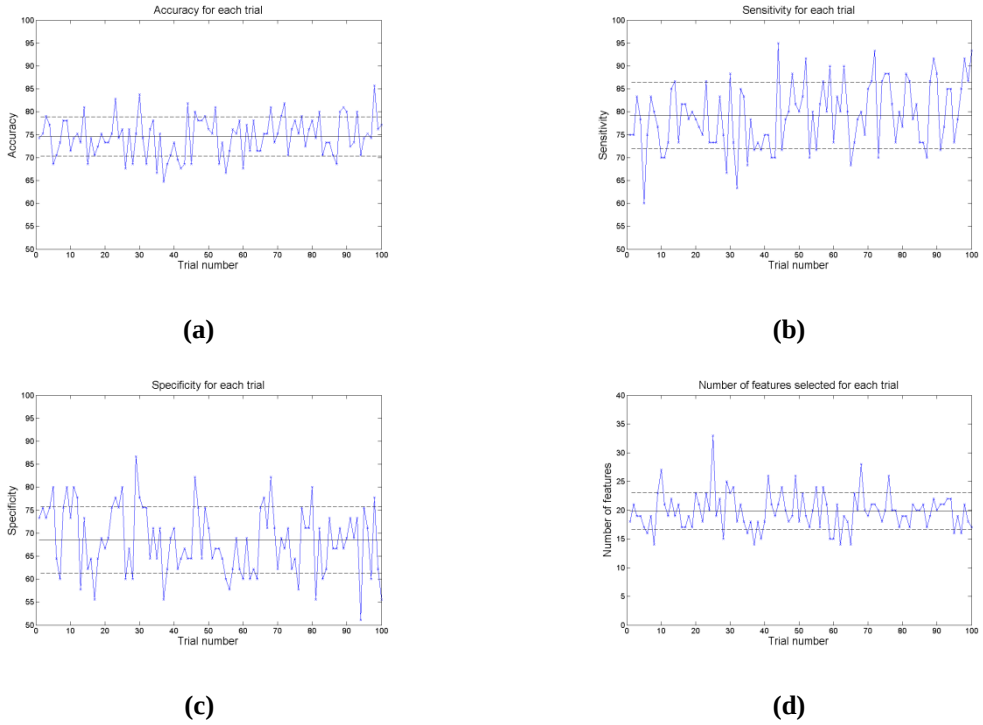


Figure 4-1: Results of each of the 100 trials. A) Accuracy B) Sensitivity C) Specificity D) # Features Selected. The solid black line in each plot is the mean of the result and the dashed lines represent 1 standard deviation away from the mean.

Figure 4-2 shows the accuracy, sensitivity, and specificity measured on the test set for a specific trial as features are removed. The performance estimates change very little during the first few hundred feature elimination steps. This suggests that many features are redundant or do not provide useful information and their removal does not significantly affect the classifier performance. At the point where there are roughly 100 or fewer features remaining, classification accuracy starts to change more and more with each additional feature eliminated. This suggests that these features are more important than the ones that were removed in the earlier stages. Note also that the MFE termination point (the loss of linear separability) coincides with the point where performance starts to degrade significantly. While the exact performance characteristic

varies (along with the order of feature elimination) for each trial, the general trend of classifier performance as a function of number of retrain features is well-represented by Figure 4-2.

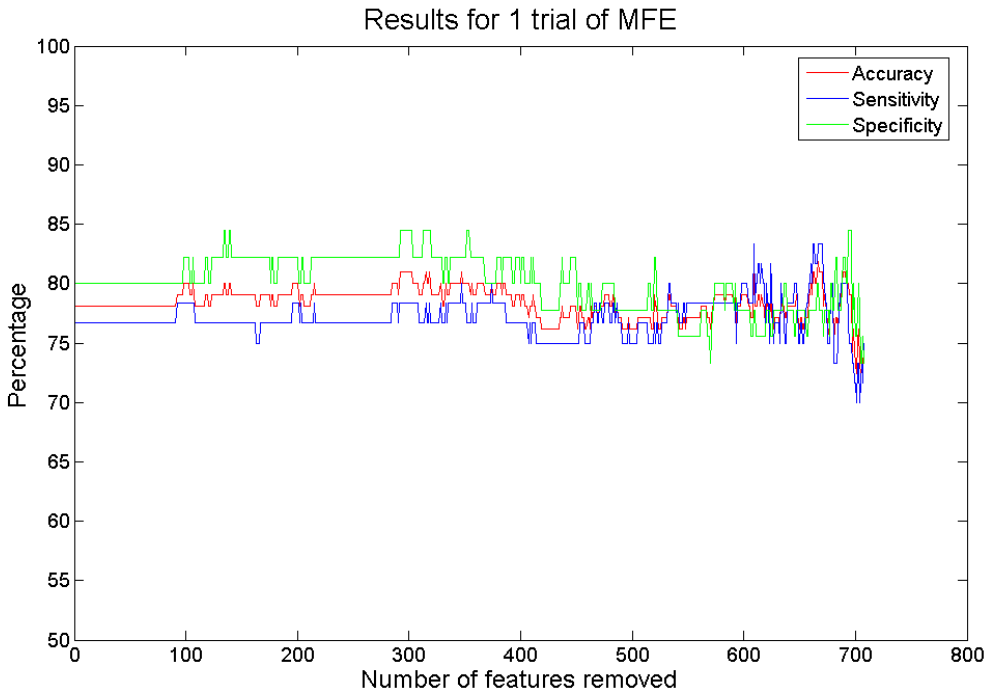


Figure 4-2: The accuracy, sensitivity, and specificity of the classifier on the test set as a function of the number of features removed.

Figure 4-3 shows histograms of the value of the SVM score function (equation 3-2) for healthy and symptomatic test set patients for a single trial. The score should be positive for symptomatic patients and negative for healthy patients. This figure visually illustrates the classification accuracy achieved by the system and the fact that most misclassified patients have scores that are close to the decision boundary (zero); a score close to zero may indicate further patient clinical evaluation is needed before making a definitive prognosis.

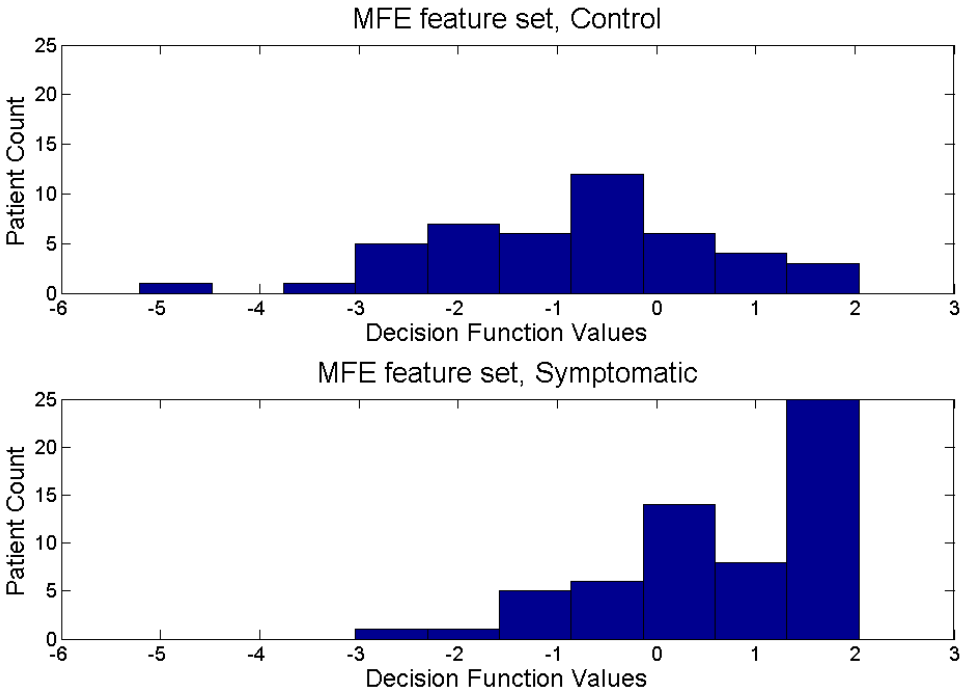


Figure 4-3: Histogram of the SVM score for test set patients in a single trial. The top graph is control patients and bottom graph is symptomatic patients.

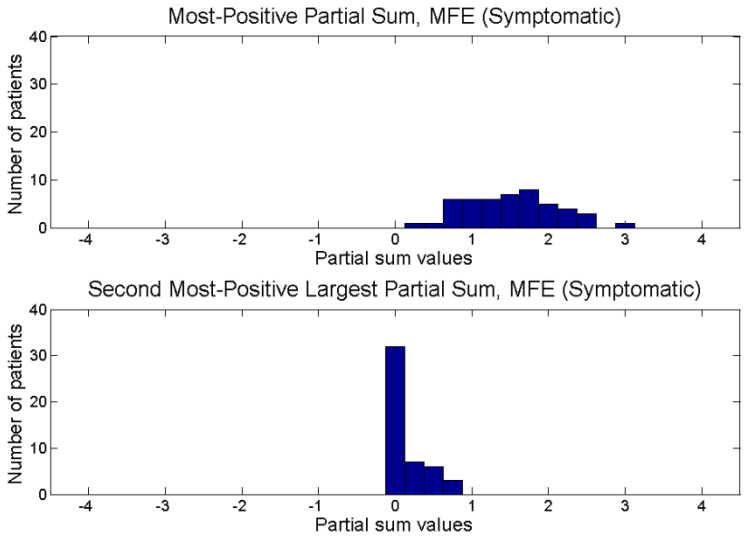
The SVM decision function is calculated by taking a linearly weighted sum of the features. By separately considering the features from each compartment (lateral, medial, patella) and finding the weighted partial sum for each section, we can determine the effective contribution of each compartment to the overall decision and thus to the diagnosis. Let F_{lat}, F_{pat}, and F_{med} be the selected features from the lateral, patella, and medial compartments. We can write equation (3-2) as

$$d_i = \sum_{j \in F_{lat}} w_j x_{i,j} + \sum_{j \in F_{pat}} w_j x_{i,j} + \sum_{j \in F_{med}} w_j x_{i,j} + b \qquad (4\text{-}1)$$

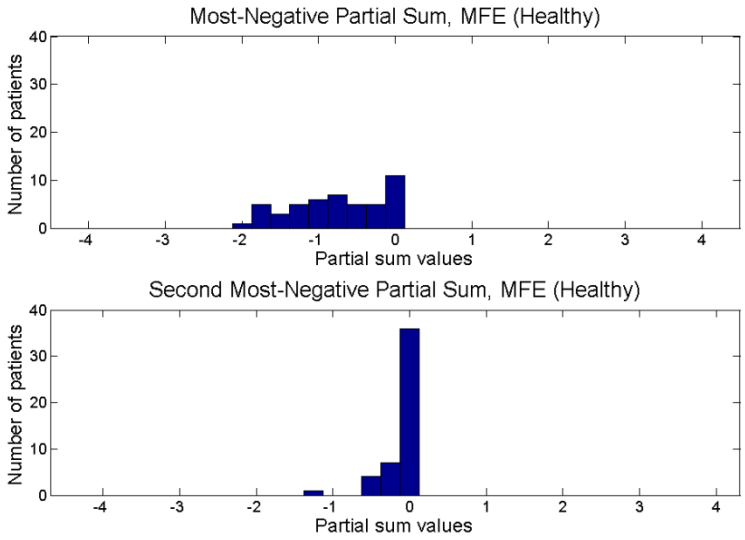
Equation (4-1) can be split into the partial sums of the features from each compartment in the following way:

$$d_{i,lat} = \sum_{j \in F_{lat}} w_j x_{i,j} \;, d_{i,med} = \sum_{j \in F_{med}} w_j x_{i,j} \;, d_{i,pat} = \sum_{j \in F_{pat}} w_j x_{i,j} \qquad (4\text{-}2)$$

For symptomatic patients that are correctly classified, the most positive partial sum amongst the three compartments contributes most to the correct decision. It can thus be inferred that the compartment with this partial sum is likely the one undergoing the most OA changes. Also, a significant disparity between the largest and second largest partial sums for an individual patient suggests OA changes may only be occurring in the knee compartment with the largest partial sum. For healthy patients, ideally all 3 partial sums would be negative (or have a very small positive magnitude). Figure 4-4 shows histograms of equation (4-2), for correctly classified symptomatic and healthy patients in the test set for a particular trial. The two most positive sums are shown for the symptomatic patients, and the two most negative sums are shown for the healthy patients. For now, the location of the most significant partial sums does not matter. Note the substantial separation between the two symptomatic histograms, which suggests that for many patients there is a dominant compartment which (primarily) influenced the correct (symptomatic) decision. This indicates that this knee compartment should perhaps be the focus of attention and early treatment. Note that there is a similar separation between the healthy patient histograms, and as mentioned above the two most negative partial sums are negative or very slightly positive.



(a)



(b)

Figure 4-4: Histograms of dominant compartment partial sums for: a) correctly classified symptomatic test patients; b) correctly classified healthy test patients.

Chapter 5

Discussion and Future Work

Based on the results in the previous section, SVM classification integrated with MFE feature selection using texture features provides a viable method for early onset detection and localization of OA using T2 maps. The initial feature set consisted of 725 features measured from the lateral, medial, and patella cartilage compartments, and this set was reduced to 19.8 features on average. This great reduction in the number of features caused only a modest reduction in average test set accuracy, from 79.1% to 74.6%. This suggests fairly accurate classification can be achieved using a small set of features. However, it was also observed that in every experimental trial, all three sections had features that were retained. Thus, taken over the whole population of patients, features from all three knee compartments appear to be necessary for making accurate OA status decisions. However, as previously discussed in the Results section, for individual correctly classified symptomatic patients there is generally one compartment that dominantly influences the classification decision. Therefore, it is possible that the classifier can also determine the area of the knee that has the most severe damage. This could help determine the best possible treatment for each patient.

One of the motivations for performing feature selection is the ability to identify a small subset of features that can identify OA. This subset can be thought of as a set of textural features that serve as a unique signature for OA. For each of the 100 feature trials, the selected features are recorded. Then, from the 19.81 features on average that are chosen, it is determined how many features from each group are selected on average. The features are grouped together by

section, by feature category, and by feature type. The grouping by knee section is shown in Table 5-1.

Section	Average # of features	Proportion of selected subset	Proportion of original set
Lateral Section 1	1.03	.052	.090
Lateral Section 2	2.34	.118	.090
Lateral Section 3	1.95	.098	.090
Lateral Section 4	1.59	.080	.090
Lateral Section 5	3	.151	.090
Patella	1.78	.090	.090
Medial Section 1	1.17	.059	.090
Medial Section 2	1.65	.083	.090
Medial Section 3	1.63	.082	.090
Medial Section 4	1.56	.079	.090
Medial Section 5	2.11	.107	.090

(a)

Compartment	Average # of features	Proportion of selected subset	Proportion of original set
Lateral	9.91	.499	.450
Patella	1.78	.090	.090
Medial	8.12	.410	.450

(b)

Table 5-1: Average number of features chosen from: a) each knee section; b) each knee compartment.

In Table 5-1b, note that the lateral, medial, and patella compartments occupy roughly the same proportion of the selected subset as they do in the original feature set, but the lateral compartment is slightly favored over the medial compartment. In Table 5-1a, note that there is not a significant difference between the proportions of each section, except for section 1 and 5 in both the lateral and medial compartments. The proportion of section 1 features in the selected subset is smaller than the proportion in the original set, suggesting that section 1 cartilage may have less importance in identifying OA. The opposite conclusion can be reached for section 5 for both compartments. However, the majority of sections have the same proportion in the selected and original feature sets, which suggests that OA classification cannot be done using a single section.

The grouping by feature category is shown in Table 5-2. The GLCM category is the only one whose proportion actually decreases. This is most likely because the GLCM category is the largest and consequently has many redundant or useless features. Also, the GLRL proportion increased the most, suggesting these features have a significant relationship with OA classification.

Category	**Average # of features**	**Proportion of selected subset**	**Proportion of original set**
Histogram	8.19	.413	.348
GLCM	5.05	.255	.455
GLRL	4.72	.238	.121
Z-score	1.85	.093	.076

Table 5-2: Average number of features chosen from each feature category.

Feature	Average # of Features	Proportion of selected subset	Proportion of original set
Max value	0	0	.0152
Min value	0	0	.0152
Relative size	.98	.0495	.0138
Energy	.41	.0207	.0152
Mean value	.35	.0177	.0152
Variance	.38	.0192	.0152
Dispersion	.23	.0116	.0152
Average energy	.15	.0076	.0152
Median value	.22	.0111	.0152
Mode value	.29	.0146	.0152
Range	0	0	.0152
Entropy	.41	.0207	.0152
Skewness	.21	.0106	.0152
Kurtosis	.54	.0273	.0152
L2 norm	.17	.0086	.0152
Hist bins	3.85	.194	.1214
GLCM energy	.86	.0434	.0759
Correlation	.8	.0404	.0759
Contrast	1.59	.0803	.0759
Absolute value	.66	.0333	.0759
Inverse difference	.96	.0485	.0759
GLCM entropy	.18	.0091	.0759
Short runs	1.87	.0944	.0303
Long runs	.71	.0358	.0303
GL nonuniform	.64	.0323	.0303
Run percent	1.5	.0757	.0303
Z mean	.1	.0050	.0152
Z variance	.29	.0146	.0152
Z min	.85	.0429	.0152
Z max	.39	.0197	.0152
Z range	.22	.0111	.0152

Table 5-3: Average number of features chosen for each feature type.

The grouping by feature type is shown in Table 5-3. The only features not selected in any trials are the max, min, and range of values calculated from the histogram. This is most likely because most of the T2 maps have at least one voxel with the largest and smallest possible intensity value in each section, and therefore these features contain no information that can be used to separate the patients. Kurtosis, relative size, and histogram bins are the histogram features whose proportion increases the most. Kurtosis measures the "sharpness" of peaks in the histogram, relative size measures the percentage of voxels that appear within each section

(relative to the number of voxels in the entire compartment), and histogram bins represent the statistical distribution of the T2 intensities. Of the GLCM features, the only feature whose proportion did not decrease is the contrast feature. Every GLRL features' proportion increased, with the short runs emphasis feature increasing the most. This feature measures the portion of shorter runs in the run-length matrices, where a run-length matrix with many short runs suggests the image does not have large uniform areas. From the Z-score features, the minimum and maximum value features perform the best. These features measure the maximum intensity difference from "healthy" cartilage.

Note that this particular classifier can only determine if the patient will develop OA symptoms within three years after the MRIs were taken. If it is determined that OA symptoms will develop within three years, we currently cannot predict when the OA symptoms began. If it is determined that OA symptoms will not develop within three years, we currently cannot make any prediction beyond this period. Also, the classification decision is binary, so even if OA is detected we are not predicting the severity of symptoms. However, the SVM decision score (see the histogram in Figure 4-3) may at least give some indication of the grade or stage of OA. A more direct approach to this problem may be investigated in future studies via multiclass (> 2 class) classification, regression, or a more sophisticated model that can predict both OA presence and severity at multiple time points.

There are multiple ways to potentially improve our system's classification accuracy. Additional texture features, such as those derived from a wavelet or Fourier transforms, and other patient characteristics such as age, weight, and patient exercise/physical activities could improve accuracy. Alternative classification model and feature selection strategies (consult [7]) could also be evaluated. There is a theoretical limit to the accuracy of any classifier for a given domain; we know it is possible to achieve an accuracy of close to 80% for OA status prognosis, but it is unknown how close this is to the theoretical limit for the OA problem.

This study only used a linear SVM without slackness. One of SVMs greatest strength is the ability to use nonlinear kernels and slackness, which allows it to work well on data that is not linearly separable or data that is better represented with a nonlinear discriminant function. Including slackness or a nonlinear kernel may improve the accuracy of the final classifier and these more sophisticated methods should be explored in future work.

The methods presented in this study are completely automated, except for segmentation. Even though the segmentation process we used is much faster than manual hand segmentation, it is still necessary for the user to place several seed points on the cartilage for each slice. Since our system requires user interaction, different users may create slightly different segmentation masks for each patient, which introduces a potential source of variability in the classification results and their accuracy. It is unknown to what extent user-specific segmentation and mask quality affects classification decisions and their accuracy. A completely automatic segmentation, which would remove this source of variability and make our entire system fully automated, is a feasible option [35] that will be investigated in future. With these further developments beyond the present study, it may be possible to create a highly accurate, completely automated, and greatly informative objective method for detecting and localizing OA in its very early stages.

Chapter 6

Conclusion

Osteoarthritis (OA) is a prevalent joint disease affecting millions. Current imaging techniques are unable to detect OA in its early stages, especially before symptom presentation. In this study, we presented a MRI-based automatic classification technique for identifying OA up to three years before significant symptoms develop. This method has an average accuracy of 74.6% using an average of 19.8 textural features derived from T2 maps of knee cartilage. It also has an average sensitivity of 79.2% and an average specificity of 68.5%. This study presents a viable alternative to expert readings of radiographs and patient self-reporting of symptoms, and with further research and development of this system it could become an invaluable tool for early diagnosis and clinical trials. This paper lays the groundwork for an accurate and automatic computer-based technique for identifying a common degenerative joint disease.

References

[1] Felson D. Osteoarthritis of the Knee. New England Journal of Medicine. 2006; 354: 841-848.

[2] Blumenkrantz G, Majumdar S. Quantitative Magnetic Resonance Imaging of Articular Cartilage in Osteoarthritis. European Cells and Materials. 2007; 13: 75-86.

[3] Mosher T, Dardzinski B. Cartilage MRI T2 Relaxation Time Mapping: Overview and Applications. Seminars in Musculoskeletal Radiology. 2004; 8: 355-368.

[4] David-Vaudey E, et al. T2 Relaxation Time measurements in Osteoarthritis. Magnetic Resonance Imaging. 2004; 22: 673-682.

[5] Mosher T, Smith H, Dardzinski B, Schmithorst V, Smith M. MR imaging and T2 mapping of femoral cartilage: in vivo determination of the magic angle effect. American Journal of Roentgenology. 2001; 177: 665-669.

[6] Mosher T, et al. Age dependency of cartilage magnetic resonance imaging T2 relaxation times in asymptomatic women. Arthritis & Rheumatism. 2004; 50: 2820-2828.

[7] Duda R, Hart P, Stork D. Pattern Classification. John Wiley & Sons, 2nd edition, 2001.

[8] Vapnik V. Statistical Learning Theory. John Wiley & Sons, 1998.

[9] Guyon I, Weston J, Barnhill S, Vapnik V. Gene selection for cancer classification using support vector machines. Machine Learning. 2002; 46: 389-422.

[10] Brown M, et al. Knowledge-based analysis of microarray gene expression data by using support vector machines. Proceedings of the National Academy of Sciences of the United States of America. 2000; 97: 262-267.

[11] Kloppel S, et al. Automatic classification of MR scans in Alzheimer's disease. Brain. 2008; 131: 681-689.

[12] Davatzikos C, Resnick S, Wu X, Parmpi P, Clark C. Individual patient diagnosis of AD and FTD via high-dimensional pattern classification of MRI. Neuroimage. 2008; 41: 1220-1227.

[13] Dunn T, Lu Y, Jin H, Ries M, Majumdar S. T2 relaxation time of cartilage at MR imaging: comparison severity of knee osteoarthritis. Radiology. 2004; 232: 592-598.

[14] Qazi A, Folkesson J, Petterson P, Karsdal M, Christiansen C, Dam E. Separation of healthy and early osteoarthritis by automatic quantification of cartilage homogeneity. Osteoarthritis and Cartilage. 2007; 15: 1199-1206.

[15] Carballido-Gamio J, Link T, Majumdar S. New techniques for cartilage magnetic resonance imaging relaxation time analysis: texture analysis of flattened cartilage and localized intra- and inter-subject comparisons. Magnetic Resonance in Medicine. 2008; 59: 1472-1477.

[16] Blumenkrantz G, et al. The feasibility of characterizing the spatial distribution of cartilage T2 using texture analysis. Osteoarthritis and Cartilage. 2008; 16: 584-590.

[17] Boniatis I, Costaridou L, Cavouras D, Kalatzis I, Panagiotopoulos E, Panayiotakis G. Assessing hip osteoarthritis severity utilizing a probabilistic neural network based classification scheme. Medical Engineering & Physics. 2007; 29: 227-237.

[18] Shamir L, Ling S, Scott W, Hochberg M, Ferrucci L, Goldberg I. Early detection of radiographic knee osteoarthritis using computer-aided analysis. Osteoarthritis and Cartilage. 2009; 17: 1307-1312.

[19] Bellamy N, Buchana W, Goldsmith C, Campbell J, Stitt L. Validation study of WOMAC: a health status instrument for measuring clinically important patient relevant outcomes to antirheumatic drug therapy in patients with osteoarthritis of the hip or knee. Journal of Rheumatology. 1988; 15: 1833-1840.

[20] Peterfy C, Schneider E, Nevitt M. The osteoarthritis initiative: report on the design rationale for the magnetic resonance imaging protocol for the knee. Osteoarthritis and Cartilage. 2008; 16: 1433-1441.

[21] qMRI: software available at http://pennstatehershey.org/web/nmrlab/resources/software/qmri

[22] Urish K, Williams A, Durkin J, Chu C. Registration of DESS and T2 magnetic resonance image series for knee articular cartilage analysis. Pending publication. 2010.

[23] Durkin J. Segmentation of articular femoral knee cartilage using active shape models. Penn State University Senior Thesis. 2010.

[24] Materka A, Strzelecki M. Texture analysis methods – a review. Technical University of Lodz. 1998.

[25] Jain A. Fundamentals of digital image processing. Prentice-Hall Inc. 1989.

[26] Haralick R, Shanmugam K, Dinstein I. Textural features for image classification. IEEE Transactions on Systems, Man, and Cybernetics. 1973; 3: 610-621.

[27] Galloway M. Texture analysis using gray level run lengths. Computer Graphics and Image Processing. 1975; 4: 172-179.

[28] Burges C. A tutorial on support vector machines for pattern recognition. Data Mining and Knowledge Discovery. 1998; 2: 121-167.

[29] Kecman V. Support vector machines – an introduction. Studies in Fuzziness and Soft Computing. 2005; 177: 1-47.

[30] Chang C, Lin C. LIBSVM: a library for support vector machines. 2001. Software available at http://www.csie.ntu.edu.tw/~cjlin/libsvm.

[31] Guyon I, Elisseeff A. An introduction to variable and feature selection. Journal of Machine Learning Research. 2003; 3: 1157-1182.

[32] Aksu Y, Miller D, Kesidis G, Yang Q. Margin-maximizing feature elimination methods for linear and nonlinear kernel-based discriminant functions. IEEE Transactions on Neural Networks. 2010; 21: 701-717.

[33] Kohavi R, John G. Wrappers for feature subset selection. Artificial Intelligence. 1997; 97: 273-324.

[34] McLachlan G, Chevelu J, Zhu J. Correcting for selection bias via cross-validation in the classification of microarray data. Beyond Parametrics in Interdisciplinary Research: Festschrift in Honor of Professor Pranab K Sen. 2008; 1: 364-376.

[35] Dodin P, Pelletier J, Martel-Pelletier J, Abram F. Automatic human knee cartilage segmentation from 3D magnetic resonance images. IEEE Transactions on Biomedical Engineering. 2010; 57: 2699-2711.

ACADEMIC VITA of Matthew G Keffalas

Matthew G Keffalas

523 East Brady Street

Butler, PA 16001

mkeffalas@yahoo.com

Education: Bachelor of Science, Electrical Engineering, Penn State University, Fall 2010

Minor in Mathematics

Honors in Electrical Engineering

Thesis Title: Feature Selection and Classification to Automatically Detect Knee Osteoarthritis Using MRI

Thesis Supervisor: David J. Miller

Awards: Dean's List, Fall 2006 through Fall 2010

Lockheed Martin Corporation Scholarship, Fall 2008

National Honors Society, Fall-Spring 2006

Activities: Electrical Engineering Teaching Internship, Fall 2009 and Spring 2010

Student Space Programs Laboratory, Power Subsystem leader, Spring 2008