

Quantitative texture analysis of T2- weighted  
MR images in polymyositis and  
dermatomyositis patients

By

Yuan Xie

Thesis

Submitted to the Faculty of the  
Graduate School of Vanderbilt University  
in partial fulfillment of the requirements

for the degree of

MASTER OF SCIENCE

in

Biomedical Engineering

August 11, 2017

Nashville, Tennessee

Approved:

Bruce M. Damon, Ph.D

E. Brian Welch, Ph.D

## LIST OF TABLES

Table	Page
Table 1. Univariate regression analysis .....	19
Table 2. Stepwise regression analysis.....	20

## LIST OF FIGURES

Figure	Page
Figure 1. Biasfield correction samples.....	18
Figure 2. Quantitative MR samples .....	18
Figure 3. Leave-one-out cross validation results .....	21

## TABLE OF CONTENTS

	Page
LIST OF TABLES.....	ii
LIST OF FIGURES .....	iii
Chapter 1. Introduction .....	1
1.1 Disease background .....	1
1.2 Disease Pathology .....	2
1.3 Diagnostic techniques .....	3
1.4 Quantitative characterization of PM and DM using MRI.....	4
1.5 Texture analysis .....	7
1.6 Goals of the project.....	13
1.7 Abstract .....	14
Chapter 2. Methods.....	15
2.1 MRI studies .....	15
2.2 Texture analysis studies .....	16
2.3 Statistical analysis .....	17
Chapter 3. Results .....	18
3.1 Sample results .....	18
3.2 Correlation analysis .....	19
3.3 Predicative models .....	20
Chapter 4. Discussion .....	22
4.1 Model selection .....	22
4.2 Robustness .....	23
4.3 Reproducibility .....	24
Chapter 5. Conclusion.....	26
BIBLIOGRAPHY .....	27

## Chapter 1. Introduction

### 1.1 Disease Background

Polymyositis (PM) and dermatomyositis (DM) are both idiopathic inflammatory myopathies. Although they have different immunologic etiologies, the two diseases are usually discussed together due to their similar pathological signs and symptoms in skeletal muscles. In the early stage of disease progression, patients typically experience proximal muscle weakness, muscle pain, fever and fatigue. Muscle damage causes serum creatine kinase levels to increase and inflammatory cell infiltration in the affected muscles. As the disease progresses, muscle cells are attacked and slowly progress to necrosis. Connective tissues start to form as the healing mechanism is activated. In the late stage of disease, necrosis occurs at a much larger spatial scale. Connective tissue and adipose tissue replace the necrotic muscle cells.

Despite the similar presentations in muscle, PM and DM still vary in their manifestations at other target regions. PM also affects other organs and systems. For example, PM patients are typically observed with abnormal electrocardiographs, suggesting atrial arrhythmias and congestive heart disease. Restriction of ventilation caused by PM on respiratory muscles also affects pulmonary functions. A high incidence of interstitial lung diseases is also observed for PM patients. DM typically affects the skin. For example, skin erythema has been long established as a recognizable symptom of DM. Rash and skin lesions are commonly observed for early stage DM patients, and skin erythemas are replaced by scaling with areas of pigmentation at the late stage of diseases.

Currently the cause of idiopathic PM and DM is unknown. There are extensive data linking certain human leukocyte antigens (HLA) with particular disease states, yet currently the HLA allele is not claimed to directly increase genetic risks. Although epidemiologic reports show that PM and DM can occur for patients at all ages, the majority of PM and DM cases happen to people between the ages of 50 to 60. There is also a juvenile form of DM that mainly affects children aging between 5

and 14. A female predominance is shown for both diseases in all age groups. The current incidence rate in the US ranges from 3~7 cases per million. (17)

## 1.2 Disease Pathology

The majority of patients with inflammatory muscle disease have PM or DM. The main pathologic feature of both diseases is muscle inflammation. The main tissue inflammation sites are perimysium, endomysium, and perimysial venules and arterioles. During muscle inflammation, mononuclear cells such as lymphocytes, plasma cells and histiocytes enter perimysial sites and infiltrate the perivascular walls; some progress into endomysial sites and infiltrate muscle fibers. Other pathologic symptoms include muscle cell necrosis, variation in muscle fiber diameter and connective tissue proliferation. (17) In either single or clusters of muscle fibers, inflammation initiated mononuclear cells target healthy muscle cells and release reactive oxygen species, causing necrosis of muscle cells. As the disease progresses, the chronic muscle necrosis causes muscle fibers to shrink, leaving empty spaces between fiber bundles. Following the muscle fiber destruction, collagen proliferates and attempts to heal muscle fibers. The three steps form a vicious cycle. As the disease progress to late stage, fibro-connective tissues progress to fatty infiltration and continuously replace muscle fibers.

The two myopathies also differ in many subtle ways regarding the systems and organs they affect. DM is known for its conspicuous skin lesions. The severity and appearance of lesion varies depending on the stages of diseases and the area of skins affected. In early stage DM, the epidermis starts thinning, making dermal blood vessels more observable. Edema is also observed in the superficial dermis, containing mononuclear cells such as lymphocytes, plasma cells and histiocytes. When the inflammatory cells infiltrates epidermis, collagen proliferates and connective tissues begin to form. As a result, the skin becomes thickened and pigmented. In the chronic stage of DM, skin cell necrosis starts and ulcers occur. Superficial capillaries dilate while subcutaneous blood

vessels membranes are also affected, causing significant vascular changes. This leads to another aspect of DM, which is its effect on membranes in various body systems. For example, in many cases, particularly of the juvenile DM, angiopathy is observed in intramuscular blood vessels. Focal capillary depletion starts in early stages of DM, and the severity increases as the disease progresses. Because of angiopathy, reduced nutrient delivery occurs. Resulting infarcts due to cell death and cell structure collapse eventually change tissue superstructures. The infarcts are often detected in juvenile form DM studies. Idiopathic inflammation is likely to be the cause of angiopathy as well, as immune complexes like IgG and IgM are found within the walls of intramuscular arteries and veins.

In contrast, PM doesn't associate with pathological symptoms in any particular body systems. PM patients are not affected by perifascicular atrophy, microvascular injury and endothelial microtubular inclusions. The immune phenotypic profile and autoaggressive character of the endomysial inflammatory cells also make PM different from DM patients.

### 1.3 Diagnostic Techniques

Typical clinical evaluations of myositis patients include strength testing, measurement of elevated serum levels of muscle enzymes, electromyography (EMG), and muscle biopsy. (2) A general investigation of a patient's disease history is typically performed before diagnosis. Depending on the severity of fatigue and past duration of muscle weakness, specific techniques are selected. The most commonly used is physical examination, which looks for characteristic lesions on skins or inflammatory responses in joints. Another examination technique is a strength test. Chronic myositis patients, even in early stages, typically perform poorly in proximal limb strength tests. Another common test, laboratory testing, looks for muscles enzymes levels in serum. The test includes creatine kinase, aldolase, antinuclear antibodies, and myositis-associated antibodies level. Traditionally, imaging techniques such as chest radiographs and computed tomography (CT) will

also be performed in all patients to help detect the presence of pulmonary involvement, such as interstitial lung disease. Recently, more advanced imaging techniques are also used in diagnosis.

Since the 1990s, contrast-based Magnetic Resonance Imaging (MRI) has been adopted as a useful clinical tool to evaluate myositis diseases. At first it was mainly used for guiding muscle biopsies and monitoring disease progression. (18) For two reasons, more advanced MRI modalities have been introduced for myositis diagnosis. First, MRI is capable of demonstrating areas of inflammation, edema with active myositis, fibrosis and calcification in muscle tissues. Second, compared to traditional invasive techniques such as biopsy, MRI can assess the muscle tissue over a large area, thereby avoiding problems with sampling error.

In images that create contrast based on between-tissue differences in the transverse relaxation time constant (T2), elevated T2 signal in muscles suggests inflammation. A MR image can therefore guide traditional biopsy for more accurate tissue acquisition. Despite multiple advantages in research, clinical applications of MRI have been mainly used as a qualitative assessment tool in disease diagnosis. However, realizing the potential value of quantitative information MRI can provide, more people are using quantitative MRI modalities in myositis research. (4)

#### 1.4 Quantitative characterization of PM and DM using MRI

A clinical biomarker is useful for quantitative characterization of any of the pathologic causes, signs, or symptoms that are related to a disease. Biomarkers take many forms, including the concentrations or activities of certain types of molecules, a histologic specimen, a physical quantity or an internal image. The selection of biomarkers is important for many reasons. A clear biomarker should have a well understood correlation with underlying physiological pathways, and should be sensitive to the progression of disease symptoms. By quantifying biomarkers, we can study drug

efficacy and observe the therapeutic responses at different stages of diseases, or we can in turn use biomarkers as a diagnostic tool for disease assessment. A biomarker also provides a universal standard that is translatable in different studies.

In PM and DM, the typical pathological features we are looking for in biomarker research are inflammation, fat infiltration, loss of membrane integrity and fibrosis. In the present MRI study, we focused on muscle inflammation and fat infiltration. Inflammation is well represented by the spin-lattice relaxation time (T1) and the spin-spin relaxation time (T2). T1 characterizes the time it takes for the longitudinal component of the magnetization vector to recover in an exponential decay towards its equilibrium value. T2 characterizes the time it takes for the transverse component of the magnetization vector to decay to its equilibrium value, which is zero. (23)

Fat composition in tissues can also be easily characterized by fat-water MRI. In this study the main things we are interested in are fat – water MRI (FWMRI). The main methods for FWMRI are signal thresholding based on T1 and T2 differences and chemical shift-based methods, such as fat signal suppression and Dixon imaging.

In healthy muscle, more than 70% of tissue mass is water. Water in biological tissues tends to have a higher T1 than fat tissue due to its smaller molecular size and higher molecular mobility. In T1-weighted images, these tissues with short T1 are brighter. Therefore when the majority of tissue content is fat and water, tissue will have a higher T1 signal when the fat content is low. Therefore by selecting a proper signal threshold, we can assign image pixels as fat if the signal is higher than threshold, and muscle if lower. The resulting binary images can also be assessed quantitatively using texture analysis methods (15). A similar method can be applied for T2-weighted signals, which tend to be lower in water and higher in fat.

The gold-standard for FWMRI is the Dixon method (5). When a magnetic field is applied, atomic nuclei align their magnetic moment parallel or anti-parallel to the field direction. The magnetic moments of the atomic nuclei precess around the applied magnetic field. The precession frequency of a nuclei is determined by the identity of the nucleus, the applied field strength, and the

local effects of molecule conformation. Protons in fat and water have different precession frequencies due to their different molecular properties. After a radio frequency (RF) pulse, the phase of all molecules and the magnetization across the sample is uniform; but if the precession frequencies differ, then there will be times when the signals from two populations of spins are in phase or out of phase. Dixon's original method requires two MR images to be taken, one at the time point where fat and water signals are in phase with each other, another when they are out of phase. (5) When in phase, the image signals are the sum of the water and fat complex signals. When out of phase, the image signals are the difference between the water and fat signals. Dixon showed that using in-phase and out-of-phase images we can acquire an estimate of the fat fraction ( $F_{fat}$ ) of tissue content in each pixel or voxel. To express in simplified math terms, if we set F as fat signal and W as the water signal. Then  $W+F$  is the signal from in phase image, while  $W-F$  is the signal from out of phase image. From these we can simply use  $\frac{1}{2}((W+F)+(W-F))$  to get image with only water signal W, and  $\frac{1}{2}((W+F)-(W-F))$  to get image with only fat signal F.

A disadvantage of the 2-point Dixon sequence is that it did not account for relaxation effects between the two echoes. Therefore, Glover et al. (16) introduced the 3-point Dixon sequence, in which two in phase echoes are acquired, at echo-times evenly spaced around the out-of-phase echo. This allowed a linear approximation of the in-phase signals, mitigating relaxation effects. In the work described in this thesis, a further evolved, 6-point Dixon sequence was used to acquire data, and  $B_0$  fitting and a multi-peak fat model were used to account for the errors caused by  $B_0$  inhomogeneity and the assumption of a single lipid resonance. Over recent years, due to its many advantages, the Dixon technique has become a gold standard to characterize fat tissue infiltration due to its robust fat signal suppression and the single acquisition -available images. On top of that, it is possible to combine it with all types of sequences such as gradient echo and spin echo, making it an ideal technique for musculoskeletal imaging. (7)

Beyond fat water content, other quantitative MR studies have also been done on PM and DM to evaluate biomarkers and disease progression. However, quantitative MRI methods like DTI,

structural imaging and multi-echo Dixon sequences are mainly carried out in research settings. Under clinical settings patients are usually scanned with contrast based MR images with low resolutions for qualitative diagnosis, which does not provide quantitative information for researchers.

In addition to Dixon imaging, T1 and T2 measurements have long been used in the non-invasive investigation of diseased muscle. Comparing to fat molecules, water tends to have longer T1 and T2 times. Therefore, in T1-weighted images the signal intensity tends to be higher for fat compared to healthy muscle tissue with less fat content. Though not quite as significant, the T2 value for fat is also longer than that of water, leading to higher signal for fat in T2-weighted images. These signal intensity differences are the basis for the texture analysis performed in this thesis.

However the traditional T1 and T2 techniques face several challenges. The first set of challenges includes the complications of B0 and B1 inhomogeneities. In contrast-based images, these inhomogeneities lead to inappropriate variations in the image grayscale known as the bias field. These inhomogeneities also lead to quantitative inaccuracy in T1 and T2 measurements. Another major problem is that T1 and T2 weighted images does not distinguish fat from other tissues with high T1, T2 signal intensities very well. An example in muscle is the T2-elevating effect of inflammation (Bryant 2014), a prominent feature of muscle disease.

## 1.5 Texture analysis

Originally developed as a satellite image analysis tool, texture analysis was created to classify or categorize pictorial data. Texture is an innate property of virtually all surfaces and is defined as any repeated visual pattern(s) in an image. (8) Depending on its size and intensity, the repeated pattern can be quantified and evaluated as being fine, coarse, smooth, irregular, etc.

Texture analysis is an intensity-based analysis in which first level and second level image

gray level information are used to describe an image's surface properties. First level information typically includes the average gray level intensity, the variance of gray level distribution, and the maximum difference between high gray level pixels and low gray level pixels, etc. Second level information describes not a direct property of the original image, but rather the property of an array or matrix that is used to quantify some aspect of the original image.

For the first level information describing the distribution of pixel intensities, it is easy to see that they have a direct, mathematical relationship to the underlying quantitative MR parameters. As a result, they are unlikely to contribute new information beyond those parameters. Second level information is extracted from matrices calculated based on the original image pixel intensities, and so this information should also reflect the underlying tissue and MRI parameters. Many matrix-based analysis tools have been developed in the past. The two main techniques used in this project are gray level co-occurrence matrix method (GLCM) and gray level run-length matrix method (GLRM). These matrices are described below.

However there are a few concerns. First, to mitigate the effects of noise, the matrices are reduced in the number of grayscale levels. By simplifying the original image we inevitably will suffer a loss of information, which may weaken the link between texture parameters and quantitative MR parameters. The second concern is that the GLCM and GLRM reflect not only the pixel intensities, but also the spatial proximity of similar and dissimilar pixel values. Therefore, although we know that the second level properties must relate in some way to the quantitative MRI and tissue properties, it is hard to tell whether the qualities they reflect on are the same. Demonstrating a direct relationship between the quantitative MRI parameters and second-level texture information requires us to identify the mathematical relationships between them.

The GLCM method is a way of extracting second order statistical texture features by computing the statistical distribution of observed combinations of intensities at specified positions relative to each other in the image. According to the number of pixels in each combination, statistics are classified into first-order, second-order and higher-order. In the GLCM, the number of

rows and columns is equal to the number of gray levels,  $G$ , in the image. The matrix element  $P(i, j|\Delta x, \Delta y)$  is the relative frequency with which two pixels, separated by a pixel distance  $(\Delta x, \Delta y)$ , occur within a given neighborhood, with intensity  $i$  and  $j$ . (1) Four statistics proposed by Haralick et al. can be extracted from the matrix. They are contrast, correlation, energy and homogeneity, each representing a certain aspect of the texture. Using the following notation:  $G$  is the number of gray levels used;  $\mu$  is the mean value of  $P$ ;  $\mu_x, \mu_y, \sigma_x$  and  $\sigma_y$  are the means and standard deviation of  $P_x$  and  $P_y$ ;  $P_x(i)$  is the  $i$ th entry in the marginal-probability matrix obtained by summing the rows of  $P(i, j)$ . The previously mentioned parameters are calculated as follows:

$$\text{Contrast} = \sum_{n=0}^{G-1} n^2 \left\{ \sum_{i=1}^G \sum_{j=1}^G P(i, j) \right\}, |i - j| = n \quad [1]$$

This measure of contrast or local intensity variation will increase if more points are distributed outside of the diagonal ( $i = j$ ), in which case the intensity contrast between a pixel and its neighbor is higher. Its value is 0 for a constant image.

$$\text{Energy} = \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} \{P(i, j)\}^2 \quad [2]$$

Energy is the summation of squared elements in the GLCM. It has a range of [0 1] and will equal 1 for a constant image. This measurement is also called angular second moment.

$$\text{Correlation} = \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} \frac{\{i*j\}*P(i, j) - \{\mu_x*\mu_y\}}{\sigma_x*\sigma_y} \quad [3]$$

Correlation is a measure of gray level linear dependence between the pixels at the specific positions relative to each other. It is 1 or -1 for a perfectly positively or negatively correlated image.

Correlation is infinite for a constant image because of dividing by zero.

$$\text{Homogeneity} = \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} \frac{1}{1+(i-j)^2} P(i, j) \quad [4]$$

Homogeneity calculates the closeness of the distribution of elements in the GLCM to the GLCM diagonal. It has a range of [0 1] and will equal 1 for a diagonal GLCM. It's influenced by the homogeneity of the image. Because of the weighting factor  $(1 + (i - j)^2)$ , homogeneity will get

small contributions from inhomogeneous areas. The result is a low homogeneity value for inhomogeneous images, and a relatively high value for homogeneous images.

The GLRM method calculates statistical parameters in a similar way in GLCM based on gray level run lengths. A gray level run is a set of consecutive, collinear pixels having the same gray level value (i.e. [1 1 1 1 1] or [3 3 3]). The length of the run is the number of pixels in the run. For a given image, we can calculate a gray level run length matrix for runs having any given direction, though commonly chosen to be 0, 45 and 90 degrees. The matrix element  $P(i, j)$  specifies the number of times that the image contains a run of length  $j$ , in a given direction, consisting of pixels having gray level  $i$ . (6)

Texture parameters similar to those derived from the GLCM can be calculated for the GLRM. Typically the original image is reduced to  $N_0$  gray levels (e.g. 64) to simplify calculation and mitigate noise effects. The total number of different run lengths that occur is  $N_r$ , so the run length matrix will have a size of  $N_0$  by  $N_r$ . The run-length features that we employed here included the Short Runs Emphasis (SRE), Long Runs Emphasis (LRE), Gray Level Non-uniformity (GLN), Run Length Non-uniformity (RLN), and Run Percentage (RP). (6) In addition, two other features including Low Gray-Level Run Emphasis (LGRE) and High Gray-Level Run Emphasis (HGRE), introduced by Chu et al (20), were also included.

$$SRE = \frac{1}{N_r} \sum_{i=1}^M \sum_{j=1}^N \frac{P(i, j)}{j^2} \quad [5]$$

SRE divides each run length value by the length of the run squared. This tends to emphasize short runs. The total number of runs  $N_r$  serves as a normalizing factor.

$$LRE = \frac{1}{N_r} \sum_{i=1}^M \sum_{j=1}^N P(i, j) * j^2 \quad [6]$$

LRE instead multiplies each run length value by the length of the run squared, which would in turn emphasis long runs.

$$GLN = \frac{1}{N_r} \sum_{i=1}^M (\sum_{j=1}^N P(i, j))^2 \quad [7]$$

GLN squares the number of run lengths for each gray level, and then normalizes the number by total number of runs in the image. When runs are equally distributed throughout the gray levels, the function has the lowest values.

$$RLN = \frac{1}{N_r} \sum_{j=1}^N (\sum_{i=1}^M P(i, j))^2 \quad [8]$$

RLN squares the number of runs for each length, and then normalizes the number by total number of runs in the image. It measures the non-uniformity of the run lengths. When the runs are equally distributed throughout the lengths, the function will have a low value. Large run counts contribute most to the function.

$$RP = \frac{N_r}{N_p} \quad [9]$$

RP is a ratio of the total number of runs to the total number of possible runs if all runs had a length of one. The value gets lower when the image has a more linear structure.

$$LGRE = \frac{1}{N_r} \sum_{i=1}^M \sum_{j=1}^N \frac{P(i, j)}{i^2} \quad [10]$$

$$HGRE = \frac{1}{N_r} \sum_{i=1}^M \sum_{j=1}^N P(i, j) * i^2 \quad [11]$$

Introduced by Chu (et al.), LGRE and HGRE were used to compensate for the inability of SRE and LRE to distinguish between images with similar run length features yet distinct gray values. LGRE have a higher value for images with mainly low gray level runs, while HGRE will have higher value for images with high gray level runs.

Texture analysis has been applied in conjunction with MRI for different diseases before. First introduced to CT as a classification tool, texture analysis has been tested in liver lesion classification (12) and breast cancer classification (9). In 1993, Lerski et al. first reviewed the potential usage of texture analysis for tissue characterization in MR image research. (21) At that point, with the rise of computing power, it was viewed as a potential statistical tool for automatic disease classification. In the prospective research conducted on patients with brain tumors, the best predictor for cerebral spinal fluid, white matter, gray matter, edema and tumor classification turned

out to be the average gray level. However in the early implementation the method still suffers heavily from the influences of artifacts, such as the bias field. Following the initial experiments, texture analysis was further studied as a diagnostic assistance tool by Herlidou (et al.) in 1999. Seventeen patients with skeletal muscle dystrophy were recruited for MRI studies. Transverse images of subjects' mid thighs were acquired. Eight selected muscles were segmented manually for each patient on T1 weighted images. Texture analysis methods including the histogram, GLCM, GLRM, and gradient matrix. They observed that the average pixel intensity had no value in discrimination between healthy and pathologic muscles. Using two texture parameters, contrast from GLCM and run length distribution from GLRM, the study was able to perform a correspondence factorial analysis. Generated factorial axis was able to classify 77% of the subjects correctly. (22)

To assess the efficacy in diagnosis, the study also did a single blind test on 9 senior radiologists by dividing patient image data into two groups. One group was provided with extra texture analysis information, while the other group only contained MR images. One type of texture analysis parameter was provided twice in the same group to also assess intra-observer variability. Results showed substantial inter-observer and intra-observer variability, making the evaluation not reproducible, and over all the sensitivity improvement was not significant. Despite the mixed outcome of the study, it did provide a solid framework for texture analysis application in MR studies. Some of the technical difficulties such as bias-field correction can be better addressed today, which should improve the outcome. Texture analysis was first shown as a powerful quantitative analysis tool, and potentially a disease classifier.

In 2015, Wang et al studied T2-weighted MR images of Golden Retriever muscular dystrophy (GRMD) as a model for Duschene Muscular Dystrophy (DMD) on golden retrievers using GLRM analysis. Ten GRMD dogs were scanned at 3, 6, and 9 months of age. The proximal limb muscles were segmented and GLRM parameters were collected. Histogram analysis results were also collected. A statistical model was established using the texture parameters. The model

was able to distinguish the normal and GRMD muscles at each age. (15) All the studies show that implementing high level texture analysis in muscle disease studies has been an evolving process. The outcomes in recent studies show that texture analysis can indeed be used as an efficient classification tool, and would be valuable in myositis research. However, the needs persists for a thorough analysis of the MRI and tissue parameters that contribute to disease-induced variations in texture parameters.

## 1.6 Goals of the project

The first goal of this project was to identify which quantitative MRI and physiological parameters explain inter-individual and inter-muscle variations in texture parameters. When applying texture analysis technique in medical images, we are interested in what physiological aspects of the disease can be explained by texture parameters. Quantitative MR, including fat saturated T1 (T1FS), fat saturated T2 (T2FS), and  $F_{\text{fat}}$  from fat water imaging using Dixon technique has been traditionally used to assess and quantify disease progressions and adipose tissue infiltrations in DM and PM patients. By establishing a relationship between quantitative MR parameters and texture analysis, we can explore the physiological meanings behind texture parameters.

The second goal of this project was to test the ability of texture parameters as tools to predict  $F_{\text{fat}}$  values derived from quantitative fat-water imaging.  $F_{\text{fat}}$  from Dixon imaging is the current established gold standard to determine fat infiltration within muscle tissue. However, under clinical settings, the main MRI method has been contrast based MR imaging. Though T1 and T2-weighted signals represent fat tissue content to a certain extent, they do not provide a truly quantitative description of the muscle's fat/water composition. This makes quantitative information of DM and PM patients hard to obtain from clinical MRI data. If we can use texture parameters as a tool to predict  $F_{\text{fat}}$ , we will be able to extract quantitative information from the large clinical

database of contrast based MR images to use in myositis research.

## 1.7 Abstract

Dermatomyositis (DM) and polymyositis (PM) patients experience intramuscular inflammation and necrosis, eventually progressing to fat infiltration. The gold standard for MRI assessment of fat tissue infiltration is quantitative fat-water MRI. Fat tissue is also detectable using standard contrast-based clinical MRI sequences; however, typical analyses of these data are qualitative. Texture analysis is a quantitative method for analyzing signal variations in contrast-based images. The goals of this study were to determine which MRI and tissue parameters explain variations in texture parameters and to use texture analysis of contrast-based MR images to predict the fat fraction ( $F_{\text{fat}}$ ), as determined by quantitative fat-water MRI. Fat signal-suppressed (FS) T1 and T2 maps,  $F_{\text{fat}}$  maps, and T2-weighted MR images were acquired from 5 DM patients, 8 PM patients, and 13 control subjects. Images were acquired at mid-thigh. The Grey Level Co-occurrence Matrix (GLCM) and Grey Level Run-length Matrix (GLRM) were calculated and used to derive 11 texture features. Regression analysis focused on the  $\log(\text{Energy})$  parameter, derived from the GLCM, and the High Gray-level Run-length Emphasis (HGRE), derived from the GLRM. 57.4% of the variance in  $\log(\text{Energy})$  was explained by  $F_{\text{fat}}$  variations. For HGRE, 68.6% of its variance was explained by  $F_{\text{fat}}$  variations. Finally, using HGRE, Low Gray level run emphasis, and Homogeneity as predictors, we were able to explain 70.3% of the variance in  $F_{\text{fat}}$ . These data show that HGRE primarily reflects fat tissue infiltration. Also, texture analysis can be used to predict  $F_{\text{fat}}$  from T2-weighted clinical MR images.

## Chapter 2. Methods

### 2.1 MRI studies

This prospective study was approved by the Vanderbilt University IRB and was HIPAA-compliant. Thirteen physician-diagnosed patients (5 PM patients, 8 DM patients) and thirteen healthy persons participated in the study. The data collection was directed by Dr. Ke Li, PhD. All subjects provided written, informed consent. Subjects followed 24-hour restrictions against moderate and heavy exercise, alcohol use, drug use, and non-prescribed medications and a 6-hour caffeine restriction. Before the scan, patients were transferred to the imaging center by wheelchair. Following MRI, functional testing was conducted. The time requirement of the protocol was about 3 hours. All data were obtained on 3.0-T Philips Achieva MR scanners (Philips Healthcare; Best, the Netherlands). Radiofrequency excitation used a two-channel body coil; signal reception used a Philips six-channel sensitivity encoding (SENSE) cardiac coil array. All images had a field of view of  $256 \times 256$  mm<sup>2</sup> and slice thickness of 7 mm. The subjects laid in a supine, feet-first position; images were acquired at the midpoint of the right thigh.

Data analysis was performed in MATLAB 2015a (The Mathworks, Natick MA). Regions-of-interest (ROIs) were defined in the rectus femoris (RF), vastus lateralis (VL), vastus intermedius (VI), vastus medialis (VM), biceps femoris (BF), semitendinosus (ST), and adductor magnus/longus (AD) muscles. All ROIs were drawn on anatomical images by the candidate.

Anatomical images were acquired with a turbo-spin-echo sequence and repetition time/echo time (TR/TE) of 530/6.2 ms and 3000/100 ms for T1 and T2-weighting, respectively; acquired/reconstructed matrices,  $256 \times 256$  and  $512 \times 512$ ; and number of excitations ( $N_{EX}$ ), 1.

Fat/water imaging data were acquired with a 2D multi-slice, six-echo gradient-echo sequence with TR, 75 ms; TE<sub>1</sub>, 1.34 ms; echo spacing ( $\Delta TE$ ), 1.53 ms;  $\alpha$ , 22°; acquired/reconstructed matrices,  $128 \times 128$  and  $256 \times 256$ ; SENSE, 1.3; and  $N_{EX}$ , 1. The data were processed as previously described (10) and used to determine  $f_F$ .

Single slice, multiple spin-echo data with fat saturation were acquired using spectrally selective adiabatic inversion recovery of the  $-\text{CH}_2/\text{CH}_3$  resonances and saturation of the olefinic proton resonance; TR, 4000 ms; 20 echoes with  $\text{TE}_1 = 14$  ms and  $\Delta\text{TE} = 14$  ms; matrix,  $128 \times 128$ ; SENSE, 1.3; a “Version S” refocusing pulse; (Poon 1992) and  $N_{\text{EX}}$ , 2. The data were fitted to a mono-exponential decay with baseline model.

Inversion recovery T1 fat saturation data were acquired with 3D fast-low-angle-single-shot readout; a 1331 binomial water selective excitation pulse; TIs of 50, 100, 200, 500, 1000, 2000, and 6000 ms; pre-delay (TD), 1500 ms; and  $N_{\text{EX}}$ , 1. T1(FS) were calculated by fitting the data to an inversion-recovery with reduced TD model.

## 2.2 Texture analysis studies

Texture analyses were done on T2-weighted anatomical images for all 8 muscle groups from 12 patients and 13 healthy subjects. The N4ITK bias field correction method was applied to reduce the effect of B1 inhomogeneity (ANTs Advanced Normalization Tools from STNAVA). Before correction, subcutaneous fat were manually masked out to improve bias field correction outcome.

In each image type, regions of interest (ROIs) were defined for eight muscles: the vastus lateralis, intermedius, and medialis; rectus femoris; adductor muscle group; and the semitendinosus, semimembranosus, and biceps femoris. The GLCM and GLRM were calculated in each ROI for each subject. From the GLCM method, the Contrast, Energy, Correlation, and Homogeneity were calculated (MATLAB 2015a image analysis toolbox: graycomatrix and graycoprops). From the GLRM method, the SRE, LRE, GLN, RLN, RP, LGRE and HGRE were calculated using a corrected version of the MATLAB function written by Wout Oude Elferink and distributed at The Mathworks website (24).

### 2.3 Statistical analysis

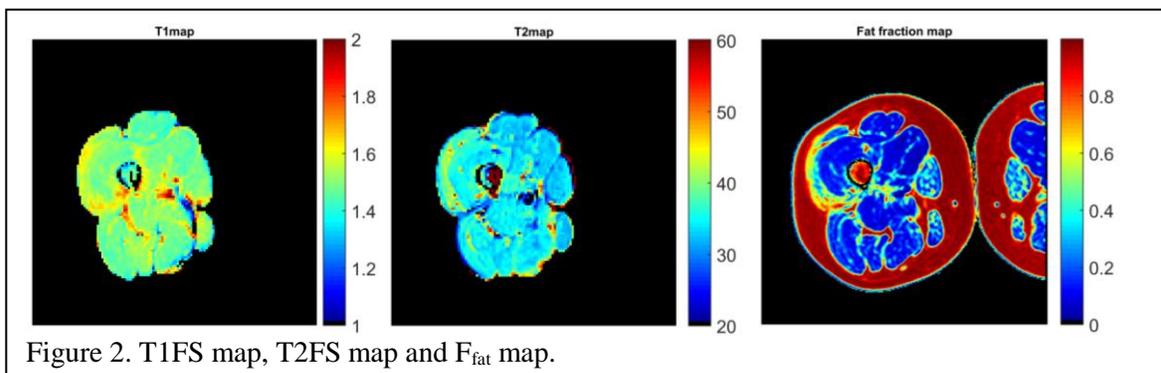
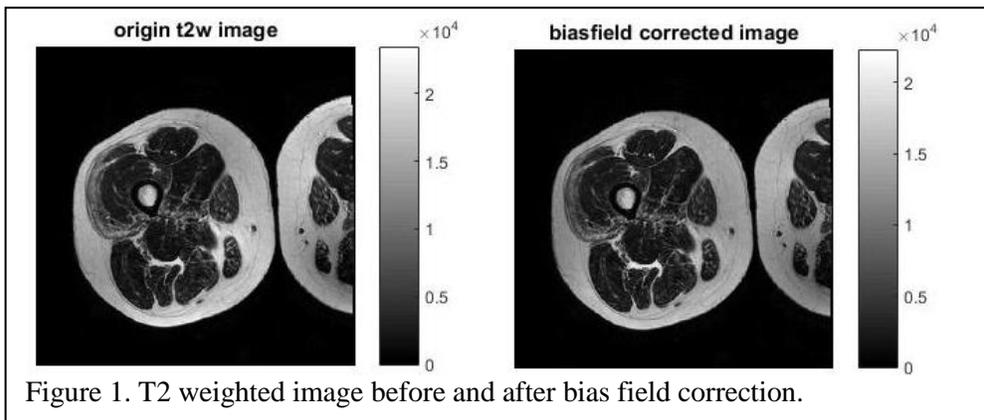
Multiple correlation analysis: The quantitative MRI parameters ( $F_{\text{Fat}}$ , T1FS and T2FS) were averaged over each muscle from each subject. The data from the eight muscles and all 13 subjects were combined into a single dataset ( $n=104$ ). For each individual texture parameter, a univariate correlation analysis was applied with respect to the above three quantitative MRI parameters. Statistical significance was accepted at  $p=0.05$ , and the  $R^2$  value was used to represent the proportion of shared variance. In addition, for each texture parameter, the partial correlation with each quantitative MRI parameter were calculated, while controlling for the effects of the other two parameters. The analyses were done in MATLAB 2015b.

Predictive model: First, a univariate regression model was constructed using the texture parameter with the highest shared variance with  $F_{\text{F}}$ . Then, a stepwise multiple linear regression model was used to construct a predictive model using multiple texture parameters to predict  $F_{\text{F}}$ . Leave-one-out cross-validation was used to estimate the predictive  $R^2$  of the model. This analysis was done in MATLAB 2015b.

## Chapter 3. Results

### 3.1 Sample results

Figure 1 shows example T2-weighted images from a myositis patient before and after bias field correction. Figure 2 shows the corresponding  $F_F$ , T1FS, and T2FS images. The values for each quantitative parameter in the vastus lateralis and vastus medialis are shown in Table 1.



The vastus lateralis (VL) region contains visible fat infiltration. For the patient's VL, T1FS was 1.57, T2FS was 37.2,  $F_{fat}$  was 0.374, HGRE was 415.5, LogE was 2.21.

Conversely, the Vastus Medialis (VM) region doesn't contain visible fat infiltration. For the patient's VM, T1FS was 1.48, T2FS was 33.3,  $F_{fat}$  was 0.065, HGRE was 124.5, LogE was 1.63.

### 3.2 Correlation analysis

Correlation analysis was performed for each texture parameter-quantitative MR parameter pair. Table 1 shows the Spearman rho coefficients. For the GLCM-derived parameters, the highest correlation existed between log(Energy) and  $F_{fat}$  ( $R = -0.574$ ) and T1FS ( $R = -0.681$ ). For the GLRM-derived parameters, the highest correlation existed between HGRE and  $F_{fat}$  ( $R = 0.686$ ) and T1FS ( $R = 0.638$ ). The partial correlation results showed that  $R = 0.248$  between  $F_{fat}$  and HGRE, when the correlations with T1FS and T2FS were accounted for, and that  $R = 0.335$  between T1FS and HGRE when the correlations with  $F_{fat}$  and T2FS were accounted for. Among all of the observed parameter pairs, the strongest correlation was found is between  $F_{fat}$  and HGRE.

Texture Parameters	T1FS	T2FS	$F_{fat}$
Contrast	0.105	0.251	0.145
Correlation	0.225	0.050	0.150
Homogeneity	-0.677	-0.343	-0.560
Log(Energy)	-0.681	-0.352	-0.574
SRE	0.677	0.337	0.578
LRE	-0.341	-0.191	-0.237
RP	0.667	0.333	0.562
LGRE	-0.348	-0.283	-0.526
HGRE	0.686	0.376	0.638
RLN	0.175	0.008	-0.066
GLN	-0.114	-0.009	-0.129

Table 1. Quantitative MR parameters were paired with texture analysis parameters for univariate correlation analysis. Significant correlations are indicated in bold type. For GLCM-derived parameters, significant correlations were observed for T1FS and log(Energy),  $F_{fat}$  and log(Energy). For GLRM-derived parameters, significant correlations were observed for T1FS and HGRE, as well as  $F_{fat}$  and HGRE.

### 3.3 Predictive Models

A univariate predictive model for  $F_{fat}$  is created using HGRE only. 93.9% of variation in  $F_{fat}$  were explained using this model:

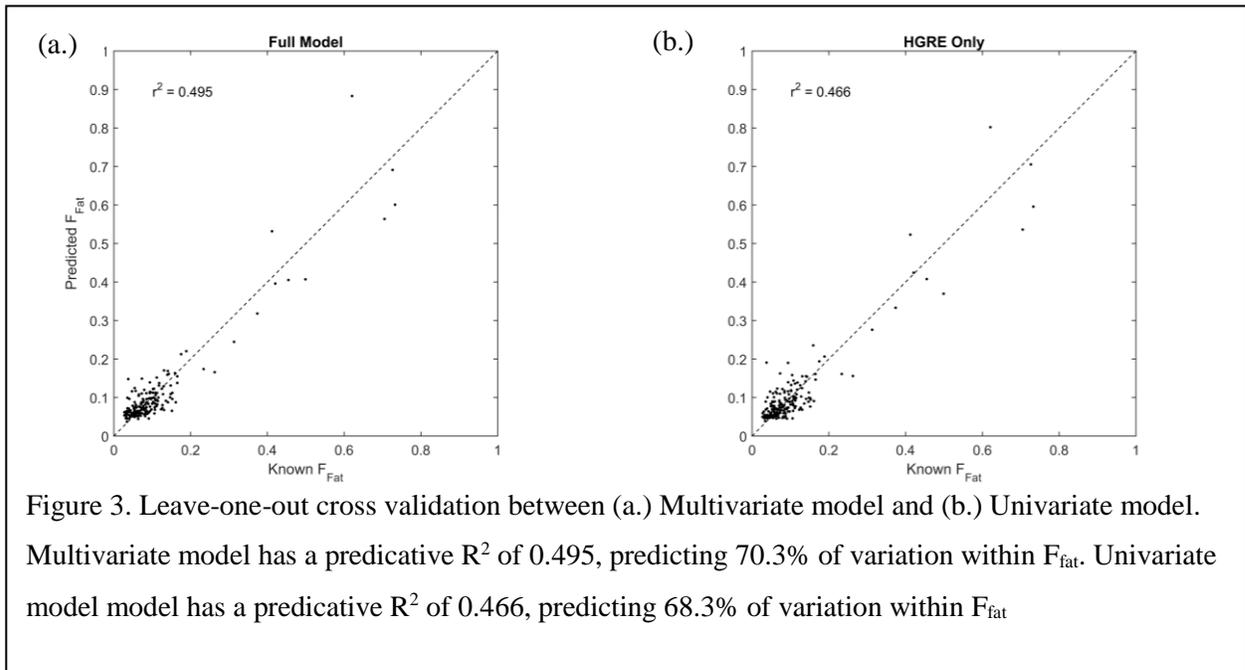
$$F_{fat} = 0.03 + 7.27 * 10^{-4} * HGRE$$

Then a stepwise multiple regression method is used to produce a predicative model for  $F_{fat}$  using all texture parameters. Contrast, Homogeneity, LGRE, HGRE and RLN were selected for the model. 94.8% of variation in  $F_{fat}$  were explained using this model:

$$\begin{aligned}
 Ffat = & \\
 & 0.073(0.02) - 0.056(0.03) * Homogeneity \\
 & + 1.02 * 10^{-5} (3.81 * 10^{-6}) * LGRE \\
 & + 7.9 * 10^{-4} (2.49 * 10^{-5}) * HGRE \\
 & - 2.88 * 10^{-7} (5.19 * 10^{-8}) * LGRE * HGRE
 \end{aligned}$$

	Estimate	SE	tStat	pValue
(Intercept)	0.072638	0.019961	3.6389	0.000351
Homogeneity	-0.05596	0.030417	-1.8396	0.067345
LGRE	1.02E-05	3.81E-06	2.686	0.007855
HGRE	0.00079	2.49E-05	31.781	4.74E-79
LGRE *HGRE	-2.88E-07	5.19E-08	-5.554	9.06E-08

Table 2. Using stepwise regression of  $F_{fat}$  on all texture parameters improved the explanation rate from 93.9% to 94.8%. Homogeneity, LGRE and HGRE were selected for the model.



Leave-one-out cross validation shows that multivariate model predicts 70.3% of the variation within  $F_{fat}$  and the univariate model predicts 68.3% of the variation in  $F_{fat}$ , as shown in figure 3.

## Chapter 4. Discussion

### 4.1 Model selection

The first goal of this project was to identify which quantitative MRI and physiological parameters explain inter-individual and inter-muscle variations in texture parameters. Our univariate correlation showed a strong relationship between the HGRE, derived from the GLRM, and  $F_{\text{fat}}$ . Another high R value was also observed between HGRE and T1FS. Because the T1FS and  $F_{\text{fat}}$  data were highly correlated, we completed a partial correlation analysis as well. This analysis showed significant partial correlations between  $F_{\text{fat}}$  and HGRE and between T1FS and HGRE. From the GLCM analysis, most significant  $R^2$  values can be found in the  $\log(E)$  with T2FS and with  $F_{\text{fat}}$ . Because these correlations were lower than that for HGRE, HGRE was used in the remainder of the studies.

The second goal of this project was to test the ability of texture parameters as tools to predict  $F_{\text{fat}}$  values derived from quantitative fat-water imaging. Looking at the results from the partial correlation analysis, the most promising approach would be to invert the model formed above and predict FF from the HGRE. Indeed, this approach resulted in a good predictive model. But two questions further needed to be addressed. First, can we build a model better than this? Second, how do we avoid overfitting and determine the actual predictive  $R^2$ ?

To address the first question, we started by incorporating all eleven available texture parameters into a multivariate analysis and conducted a stepwise regression. The regression selected five texture parameters as predictor variables. Compared to the single regression model with only HGRE, the new model showed a marginal improvement. This indicates that even in the more complex model, the predictive ability is still mainly driven by HGRE.

The risk exists that the mild improvement in the more complex model is due to overfitting. To test this possibility, we completed a leave-one-out cross-validation analysis on both models. In

the case of the original univariate model, the correlation was better preserved than in the case of the multivariate model. Also, one data point was poorly predicted in the cross-validation of the multivariate model. Together, these data suggest that the slightly improved correlation in the multiple regression model is mainly due to overfitting and that the single correlation model using HGRE out-performs any other models using the observed texture parameters to predict  $F_f$ .

## 4.2 Robustness

The HGRE model has built a good foundation for  $F_{fat}$  prediction. Yet there are still unaddressed concerns regarding the robustness of the model. A future goal of this work is to extract quantitative data from any clinical database using the contrast based MR images provided. However the provided images are likely to differ in terms of their slice location, in-plane resolution, and TE/TR times due to the various needs at every clinical MR session. Just like changes in quantitative MR parameters, these factors also directly affect the image signal intensity and contrast. Since texture analysis outcomes are heavily based on the image intensity of the original image, the predicted  $F_{fat}$  using our current model will almost certainly vary from the actual results once switched to a different image database. Therefore the robustness of this prediction model to these anticipatable variations in clinical imaging data requires further validation.

Slice location directly changes the image content. In our study, all images were acquired at the mid-thigh for each patient. For studies conducted at different thighs location or even multiple thigh locations, the results may change accordingly. Because polymyositis and dermatomyositis progress differently at differently thigh locations, it is equally possible for a patient to be affected at higher thigh locations and at a lower thigh locations. While the disease manifestation may be similar, the muscle content could potentially change drastically, causing varied results between muscles. Segmentation will also become challenging due to shifting muscle locations. It should be noted, however, that this challenge is not unique to texture analysis. Before applying the model, it

is important to use a group of images with mixed slice locations and  $F_{\text{fat}}$  results to test for potential deviations and outliers. It will however prove to be difficult since segmentations of certain muscles will be challenging at different slice locations.

Repetition time and echo time (TR, TE) affect the image contrast more directly, as the signal intensity for a spin echo sequence is typically determined by TR and TE. T1 effects are mainly connected to TR and T2 effects to TE. As TE increases, the effect of T2 diminishes, while if TR increases, the influence of T1 diminishes as well. Therefore if we use the same model on a different database with longer TE time, the contrast is anticipated to be generally lower for T2w images. Therefore the predicted  $F_{\text{fat}}$  would be lower comparing to our current database. The same concerns applies for T1 weighted images as our raw image data.

A good way to test the model's sensitivity to TR and TE is to apply the model on raw image data with different TR and TE value and cross validate the result with true  $F_{\text{fat}}$ . Despite the two factors, proton density also plays an important role in the contrast of an acquired image. A similar approach may be applied to validate its influence on the model outcome.

### 4.3 Reproducibility

In this study, reproducibility should also be a priority so that the model may be useful in all situations. One major influence on reproducibility would be B1 inhomogeneity, or the bias field, which directly alters the intensity of raw contrast based image. The bias field also differs from dataset to dataset and even from image to image. Therefore choosing a valid bias field correction method was our priority concern.

Typical bias field correction methods such as a smooth surface estimation in combination with a fuzzy c-mean correction will suffice. The N4 bias field correction method yielded better results for this study. The main concern however, is that bias field correction lacks a standardized criteria that quantifies the image improvement, not to mention a quantitative comparison between

two bias field correction methods. The criteria tends to vary from study to study, especially in the cases where a true original image is unavailable.

However this criteria is still limiting. First the assumption does not account for all the other potential noise sources. In a different dataset, background noise or image artifact could be the main source of noise, while the bias field barely have any influence on the image quality. Another problem is that by using the outer ring of subcutaneous fat as criteria it limits the types of images we can analyze. If the image is preprocessed and the subcutaneous fat is removed, then this process will not apply. Therefore it is crucial to figure out criteria that applies universally for bias field correction, so that the correction process can be reproduced in every imaging database under different scenarios.

Another problem worth considering is the segmentation method. In this study all segmentations were done by a trained specialist. The process is time consuming and nearly impossible for a larger dataset. An automatic segmentation method is necessary for a more robust and efficient model. Though a main concern would be reproducibility of the segmentation results, as the infiltrated fat content can be mistakenly recognized as membrane, intramuscular fat or blood vessels in contrast MR images. There is currently no ideal segmentation method for muscle imaging, and some preprocessing may be required to achieve any meaningful results.

## Chapter 5. Conclusions

Overall, this study shows that HGRE has a strong correlation with  $F_{\text{fat}}$ . The constructed prediction model is capable of predicting variations within  $F_{\text{fat}}$  in this dataset. Future work is needed to test the models' sensitivity to other acquisition parameters so that the method is more robust. Still, these results strongly support the potential for texture analysis to be used for the quantitation of clinical imaging data.

## BIBLIOGRAPHY

1. Albreghsen, F. (2008). "Statistical Texture Measures Computed from Gray Level Cooccurrence Matrices."
2. Bohan , A. and J. B. Peter (1975). "Polymyositis and Dermatomyositis." *New England Journal of Medicine* 292(8): 403-407.
3. Castellano, G., et al. (2004). "Texture analysis of medical images." *Clinical Radiology* 59(12): 1061-1069.
4. Damon, B. M., et al. (2016). Chapter 41 - Magnetic resonance imaging of skeletal muscle disease. *Handbook of Clinical Neurology*. C. M. Joseph and R. G. González, Elsevier. Volume 136: 827-842.
5. Dixon, W. T. (1984). "Simple proton spectroscopic imaging." *Radiology* 153(1): 189-194.
6. Galloway, M. M. (1975). "Texture analysis using gray level run lengths." *Computer Graphics and Image Processing* 4(2): 172-179.
7. Guerini, H., et al. (2015). "Fat Suppression with Dixon Techniques in Musculoskeletal Magnetic Resonance Imaging: A Pictorial Review." *Semin Musculoskelet Radiol* 19(04): 335-347.
8. Haralick, R. M., et al. (1973). "Textural Features for Image Classification." *Systems, Man and Cybernetics, IEEE Transactions on SMC-3*(6): 610-621.
9. Holli, K. K., et al. (2010). "Texture analysis of MR images of patients with Mild Traumatic Brain Injury." *BMC Medical Imaging* 10: 8-8.
10. Kassner, A. and R. E. Thornhill (2010). "Texture Analysis: A Review of Neurologic MR Imaging Applications." *American Journal of Neuroradiology* 31(5): 809-816.

11. Kijowski, R., et al. (2009). "Improved fat suppression using multipeak reconstruction for IDEAL chemical shift fat-water separation: application with fast spin echo imaging." *J Magn Reson Imaging* 29(2): 436-442.
12. Mayerhoefer, M. E., et al. (2010). "Texture-based classification of focal liver lesions on MRI at 3.0 Tesla: A feasibility study in cysts and hemangiomas." *Journal of Magnetic Resonance Imaging* 32(2): 352-359.
13. O'Connell, M. J., et al. (2002). "Whole-Body MR Imaging in the Diagnosis of Polymyositis." *American Journal of Roentgenology* 179(4): 967-971.
14. Poon, C. S. and R. M. Henkelman (1992). "Practical T2 quantitation for clinical applications." *Journal of Magnetic Resonance Imaging* 2(5): 541-553.
15. Wang, J., et al. (2013). Statistical texture analysis based MRI quantification of Duchenne muscular dystrophy in a canine model.
16. Glover, G. H. and E. Schneider (1991). "Three-point dixon technique for true water/fat decomposition with B0 inhomogeneity correction." *Magnetic Resonance in Medicine* 18(2): 371-383.
17. Fishman, R. A. (1994). "Myology: Basic and clinical, second edition, vols 1 and 2. Edited by Andrew G. Engel and Clara Franzini-Armstrong, New York, McGraw-Hill, 1994, 1937 pp. illustrated. " *Annals of Neurology* 36(5): 808-808.
18. Park, J. H., et al. (1990). "Dermatomyositis: correlative MR imaging and P-31 MR spectroscopy for quantitative characterization of inflammatory disease." *Radiology* 177(2): 473-479.
19. Bryant, N. D., et al. (2014). "Multi-parametric MRI characterization of inflammation in murine skeletal muscle." *NMR in biomedicine* 27(6): 716-725.
20. Chu, A., et al. (1990). "Use of gray value distribution of run lengths for texture analysis."

Pattern Recognition Letters 11(6): 415-419.

21. Lerski, R. A., et al. (1993). "MR image texture analysis--an approach to tissue characterization." *Magn Reson Imaging* 11(6): 873-887.
22. Herlidou, S., et al. (1999). "Comparison of automated and visual texture analysis in MRI: Characterization of normal and diseased skeletal muscle." *Magnetic Resonance Imaging* 17(9): 1393-1397.
23. Hashemi, Ray H, Bradley, William G, Lisanti, Christopher J and Ovid Technologies, Inc *MRI : the basics* (2nd ed). Lippincott Williams & Wilkins, Philadelphia, 2004.
24. <https://www.mathworks.com/matlabcentral/fileexchange/52640-gray-level-run-length-image-statistics>, Elferink, 2015.