MRI Texture-Based Classification of Dystrophic Muscles. A Search for the Most Discriminative Tissue Descriptors

Dorota Duda[1](✉), Marek Kretowski[1], Noura Azzabou[2,3], and Jacques D. de Certaines[2,3]

[1] Faculty of Computer Science, Bialystok University of Technology, Wiejska 45a, 15-351 Bialystok, Poland
{d.duda,m.kretowski}@pb.edu.pl
[2] Institute of Myology, Nuclear Magnetic Resonance Laboratory, Paris, France
[3] CEA, I2BM, MIRCen, NMR Laboratory, Paris, France

Abstract. The study assesses the usefulness of various texture-based tissue descriptors in the classification of canine hindlimb muscles. Experiments are performed on T2-weighted Magnetic Resonance Images (MRI) acquired from healthy and Golden Retriever Muscular Dystrophy (GRMD) dogs over a period of 14 months. Three phases of canine growth and/or dystrophy progression are considered. In total, 39 features provided by 8 texture analysis methods are tested. Features are ranked according to their frequency of selection in a modified Monte Carlo procedure. The top-ranked features are used in differentiation (i) between GRMD and healthy dogs at each phase of canine growth, and (ii) between three phases of dystrophy progression in GRMD dogs. Three classifiers are applied: Adaptive Boosting, Neural Networks, and Support Vector Machines. Small sets of selected features (up to 10) are found to ensure highly satisfactory classification accuracies.

Keywords: Golden Retriever Muscular Dystrophy (GRMD) · Duchenne Muscular Dystrophy (DMD) · Texture analysis · Feature selection · Classification · MRI T2

1 Introduction

Duchenne Muscular Dystrophy (DMD) is a genetic disorder affecting approximately 1 in 3,600 boys worldwide [1]. It is caused by the absence of dystrophin, a protein that plays an essential role in supporting fiber strength, mainly in the skeletal and cardiac muscles. In affected individuals dystrophin is not synthesized normally, which results in progressive muscle degeneration. This leads to permanent progressive disability (decreased mobility, deformities, cardiomyopathy, and respiratory failure) and premature death [2]. No treatment can reverse the fatal muscle destruction and there is still no effective cure for DMD.

The Golden Retriever Muscular Dystrophy (GRMD) canine model is the most widely used in research on potential treatment of DMD in humans. It mimics the human DMD model in many aspects [3]. Important information about the progression of the disease and/or its response to therapy can be obtained, in an atraumatic manner, using Magnetic Resonance Imaging (MRI). However, the interpretation of image content is not a trivial task. Great hope is placed in computer-aided image recognition methods, especially those based on texture analysis (TA) [4,5]. The use of appropriately selected textural features in the tissue differentiation process could reduce the need for invasive diagnostic methods, such as those involving needle biopsies, which can considerably weaken already degenerated muscles.

Many different TA methods have been successfully employed for tissue characterization in the classification process [6–8]. Research has shown that not all of these methods provide equally useful features. The study [4] presents the theoretical basis for the suitability of various MRI-TA methods for muscular tissue characterization in healthy and GRMD dogs. The authors noted that muscle properties are not the only important factors in choosing the most appropriate TA method. The shape and size of the image regions analyzed (*Regions of Interest*, ROIs) are also critical. Many ROIs may be narrow and very small, which disqualifies some TA methods, particularly those that analyze pairs of pixels in which a long distance separates the two pixels in the pair. This problem was also observed in another study [9], which investigated the potential of different MRI-TA techniques for characterization of canine hindlimb muscles (GRMD and healthy). The study also showed that the use of all possible textural features does not always ensure the best classification results.

The aim of the present study is to find the most suitable tissue descriptors in the muscle differentiation process based on the T2-weighted MRI. Images are derived from GRMD and healthy dogs and correspond to three phases of canine growth and/or dystrophy progression [4]. Muscular tissue is characterized using features provided by eight TA methods. First, a modified Monte Carlo (MC) feature selection [10] is used to assess the relative importance of each feature in the tissue recognition process. Then the features are ranked. Finally, the top-ranked features are used to describe tissues in various classification tasks.

The next section gives a short overview of related work. Section 3 describes methods making it possible to find the most discriminative features. Next, the experimental setup is detailed. In Sect. 5 the results are presented and discussed. Conclusions and perspectives are outlined in the final section.

2   Related Work

There have been few studies on texture-based characterization of dystrophic muscles in GRMD (and healthy) dogs. However, some recent research has shown that texture analysis has great potential for characterization of dystrophy progression or differentiation between affected and healthy canine muscles. In the work [11], different MRI biomarkers derived from T2-weighted images were used to quantify

longitudinal disease progression and to differentiate between GRMD and healthy dogs at different phases of canine growth and/or dystrophy development. Three phases were considered: at 3, 6, and 9–12 months of age. Non-textural (MRI-based) and textural features were used to characterize seven types of muscles of the proximal pelvic limbs. The textural features were based on the gray-level histogram (GLH) and run length matrices (RLM) [12]. According to statistical tests, all the textural features were significantly different in the two classes of dogs (GRMD and healthy). This was observed at each phase of life. Moreover, classification experiments, run with Linear Discriminant Analysis (LDA), produced better results when textural features were used.

Yang *et al.* [13] attempted to follow texture changes in GRMD and healthy dogs, imaged over a period of 14 months. They focused on moment-based TA methods, applying Legendre and Zernike moments. Their feature vectors were analyzed by Principal Component Analysis (PCA) and classified by the SVM classifier [14]. Here too, three typical phases of canine growth were considered. The moment-based texture descriptors provided important discriminatory information for distinguishing between two dog classes at each phase.

The same database of images was analyzed in our previous work [9], as both studies are part of the European COST Action BM1304 project ("MYO-MRI") aimed at exploring strategies for muscle imaging texture analysis. We differentiated between GRMD and healthy dogs at each of three phases of canine growth. Textural features were extracted using statistical, model-based, and filter-based TA methods. Eight sets of features, each derived from a different method, were considered separately. The set of all features derived from all methods was also tested. Experiments involving five classifiers showed that highly satisfactory classification results can be obtained with certain (relatively small) sets of features, especially those based on RLM and co-occurrence matrices (COM) [15]. The work did not perform any feature selection nor attempt to differentiate between tissues in different phases of dystrophy development.

3   Evaluation of the Usefulness of Features

All the evaluations are performed on a *training set*. This set is created by characterizing each ROI with the same set of features (in our case textural). ROIs are labeled by assigning information to them about their "class". This information can be related to the presence or absence of disease or to the phase of disease progression. The relative importance of each textural feature in the tissue identification process is assessed in the study using Monte Carlo feature selection, initially proposed by Draminski *et al.* [10]. This procedure was chosen due to its proven effectiveness and reliability, and because it does not require any initial assumptions. It returns a ranking of features according to their importance. Further classification experiments allow us to identify the subsets of features that ensure the most satisfactory differentiation between ROI classes.

3.1 Modified Monte Carlo Feature Selection

The method initially described by Draminski *et al.* was slightly modified and adapted to our needs. Our proposal is schematized in Fig. 1. The number of objects in the initial data set and the number of tissue descriptors (features) for each observation are denoted by p and d, respectively. First, q observations are randomly chosen from the initial training set ($q < p$) and characterized by a small number (m) of randomly chosen features, $m \ll d$. This is followed by a single execution of the selection procedure, based on this truncated data set. These two steps are repeated many times (denoted by r), each time with a different subset of observations and a different subset of features. Next, the "incidence frequency rate" is calculated for each feature. This is the ratio between the number of cases in which the feature is selected and the number of times it occurs in the subsets of randomly chosen features describing the truncated data sets (subjected to selection). Finally, the features are ranked according to their incidence frequency rate, from the highest to the lowest rate. The top-ranked features are considered to be potentially the most important in the tissue recognition process.
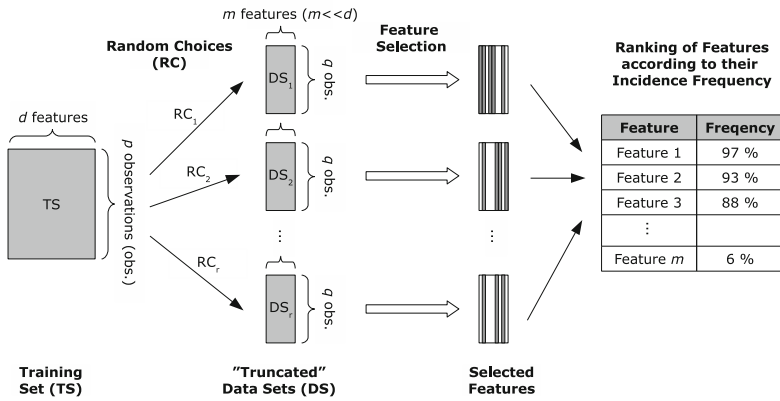
Fig. 1. Assessing the relative importance of each feature in the tissue identification process. The course of a modified Monte Carlo feature selection method

Reducing the number of features before proceeding to each repetition of the selection procedure is not only aimed at shortening the computation time. Guyon and Elisseeff [16] pointed out that some features may not prove informative separately, but only when combined with other features. By performing multiple repetitions of the selection procedure, each time with a different small subset of randomly chosen features at input, we can identify features which prove useful independently of the combination in which they are initially found.

3.2 Texture-Based Classification

Once the ranking of features (according to their incidence frequency rates) is completed, we can determine how many top-ranked features would be sufficient

to ensure very good differentiation between tissue classes. At this stage, d experiments are conducted. In the ith experiment $(i = 1, \ldots, d)$, i top features from the ranking are used as tissue descriptors. In each of d experiments the usefulness of the considered subset of features is assessed by the quality of the classification it can provide. Classification accuracies are estimated by the standard 10-fold cross-validation procedure, repeated 10 times.

4   Experimental Setup

The images were gathered at the Nuclear Magnetic Resonance Laboratory of the Institute of Myology, Paris (France). Acquisitions were performed over a period of 14 months, using a 3T Siemens Magnetom Trio TIM imager/spectrometer. A precise description of all acquisition protocols was provided in [17]. Information concerning image pre-processing, completed prior to tissue characterization, can be found in our previous work [9]. Only T2-weighted spin echo sequences were chosen for our investigations. Images were collected from 10 dogs: 5 with GRMD and 5 healthy dogs. In total, 38 acquisitions were available (from 3 to 5 for each dog). Each acquisition was assigned to one of three phases of canine growth and/or dystrophy progression [4]. The first phase comprised the first four months of life, the second phase was the period from the fifth to the sixth month of life, and the third phase began at the age of seven months. The phases were represented in the study by 14, 9, and 15 acquisitions, respectively. Each acquisition provided a series of 12 to 14 images. The images had a size of 240×320 pixels, the in-plane resolution was $0.56 \, \mathrm{mm} \times 0.56 \, \mathrm{mm}$. ROIs defined on the images included four types of canine hindlimb muscles: the *Extensor Digitorum Longus* (EDL), the *Gastrocnemius Lateralis* (GasLat), the *Gastrocnemius Medialis* (GasMed), and the *Tibial Cranialis* (TC). Only ROIs of at least 100 pixels were considered. The numbers of ROIs (for each phase, tissue class, and muscle type) are given in Table 1. The average ROI sizes are presented in Table 2.

For each ROI, 39 features were extracted with the homemade application *Medical Image Processing* [18]. They were based on: the gray-level histogram (GLH), gradient matrices (GM) [19], co-occurrence matrices (COM), run-length matrices (RLM), gray level difference matrices (GLDM) [20], Laws' texture energy measures (LTE) [21], the fractional Brownian motion model (FB) [22],

Table 1. Numbers of suitable ROIs (for each phase, tissue class, and muscle type)

	Phase 1		Phase 2		Phase 3	
	GRMD	Healthy	GRMD	Healthy	GRMD	Healthy
EDL	45	52	56	48	73	136
GasLat	43	30	34	24	31	85
GasMed	64	60	43	37	60	113
TC	53	73	87	64	81	157

Table 2. Average ROI sizes in pixels (for each phase, tissue class, and muscle type)

	Phase 1		Phase 2		Phase 3	
	GRMD	Healthy	GRMD	Healthy	GRMD	Healthy
EDL	156	202	189	239	160	279
GasLat	189	161	220	184	199	220
GasMed	293	290	379	395	328	426
TC	165	205	250	255	236	316

Table 3. Textural features considered (and their abbreviations)

Method	Features
GLH	average (Avg), variance (Var), skewness ($Skew$), kurtosis ($Kurt$)
GM	average ($GraAvg$), variance ($GraVar$), skewness ($GraSkew$), kurtosis ($GraKurt$)
COM	angular second moment ($AngSecMo$), entropy ($Entr$), inverse difference moment ($InvDiffMo$), correlation ($Corr$), sum average ($SumAvg$), difference average ($DiffAvg$), sum variance ($SumVar$), difference variance ($DiffVar$), sum entropy ($SumEntr$), difference entropy ($DiffEntr$), contrast ($Contrast$)
RLM	short run emphasis ($ShortEm$), long run emphasis ($LongEm$), gray level non-uniformity ($GlNonUni$), run length non-uniformity ($RlNonUni$), fraction of image in runs ($Fraction$), low gray level runs emphasis ($LowGlrEm$), high gray level runs emphasis ($HighGlrEm$), run length entropy ($RlEntr$)
GLDM	average ($gAvg$), entropy ($gEntr$), contrast ($gContrast$), angular second moment ($gAngSecMo$), inverse difference moment ($gInvDiffMo$)
LTE	entropy of a ROI filtered with Laws' masks (E_3L_3, S_3L_3, S_3E_3, E_3E_3, S_3S_3)
FB	fractal dimension ($FractalDim$)
AC	autocorrelation ($Autocorr$)

and the autocorrelation model (AC) [23]. Full feature names and their abbreviations are given in Table 3.

We used some preliminary classification experiments to determine the best settings for each TA method. As a result, the number of image gray levels was reduced from 256 (initially) to 64 for the COM, RLM, and GLDM methods. Four standard directions of pixel arrangement ($0°$, $45°$, $90°$, and $135°$) were considered in applying the COM, GLDM, RLM, and AC methods. Only the smallest distances between pixels in pairs (1 and 2) were considered for the COM, GLDM, FB, and AC methods. Features calculated for different directions (and different distances between pixels) were averaged. The LTE method used only

3×3 zero-sum convolution kernels. Images obtained with kernel pairs consisting of a mask and its transposition were added.

Feature selection was performed with the Weka tool [24]. A single selection procedure was repeated $r = 200,000$ times. Each time, two-thirds of the available observations ($q = 2/3p$) were randomly chosen from the original data set. The proportions between observations representing each class in the truncated data set were the same as in the original data set. Observations were characterized each time by a randomly chosen set of 8 features, which was about 20 % of all the features initially calculated ($m = 0.2d$). For each selection procedure, the usefulness of each candidate subset of features was estimated by a wrapper (supervised) method *WrapperSubsetEval* combined with a C4.5 Decision Tree [25] classifier (called *J48* in Weka). This classifier was used because of its simplicity, good performance, and very short induction time. This last property is important when a single selection procedure must be repeated hundreds of thousands of times. Classification accuracies were assessed by a 10-fold cross-validation. The space of subsets of features was searched using the *BestFirst* strategy with the *Forward* searching direction.

Classification experiments were also conducted with the Weka tool. Three classifiers were utilized: Adaptive Boosting (AdaBoost) [26], back-propagation Neural Network (NN) [27], and nonlinear Support Vector Machines (SVM). The AdaBoost classifier was trained for 100 iterations and used the C4.5 tree as the underlying algorithm. The Neural Network used a sigmoidal activating function and had one hidden layer in which the number of neurons was equal to the average number of considered features and the number of tissue classes. Support Vector Machines used Platt's Sequential Minimal Optimization (SMO) algorithm [28] and a second-degree polynomial kernel.

Two problems were considered during the classification experiments. The first was to differentiate between GRMD and healthy dogs at each of the three phases of canine growth and/or dystrophy progression. The second was to differentiate between tissues at three phases of dystrophy progression in GRMD dogs.

5   Results and Discussion

The experiments were run separately for each type of muscle (EDL, GasLat, GasMed, and TC), and were repeated using three classifiers. Firstly, a modified MC feature selection was performed and the most frequently selected features were detected. Secondly, the number of features sufficient to ensure satisfactory differentiation between tissue classes was assessed by testing different subsets of top-selected features. Finally, the differences between classification qualities obtained with selected features and all possible features were examined.

5.1   The Most Frequently Selected Features

The most frequently selected features are listed in Table 4, separately for each of four muscle types and each classification problem (defined at the end of the the

previous section). Based on Table 4 a detailed comparison of the most appropriate textural features for each case could be performed.

Table 4. The 5 most frequently selected features for each classification task and each muscle type. The first three rows concern differentiation between GRMD and healthy dogs at different phases of canine growth. The last row concerns differentiation between tissues at three phases of dystrophy progression in GRMD dogs. Feature names are preceded by their incidence frequency rates [%] in the modified MC feature selection.

	EDL	GasLat	GasMed	TC
Phase 1	(87.4) $InvDiffMo$	(59.4) $Entr$	(87.8) $FractalDim$	(71.6) $GlNonUni$
	(86.1) $gInvDiffMo$	(49.9) S_3L_3	(65.2) $GraAvg$	(68.7) $GraAvg$
	(65.2) $AngSecMo$	(39.0) $GraAvg$	(39.3) $gInvDiffMo$	(66.2) $gInvDiffMo$
	(39.8) $gAngSecMo$	(37.7) $gAngSecMo$	(39.1) $InvDiffMo$	(65.7) $InvDiffMo$
	(39.4) $gAvg$	(34.8) $SumEntr$	(29.4) $HighGlrEm$	(65.6) $LongEm$
Phase 2	(71.6) Avg	(77.6) $SumAvg$	(73.0) $SumAvg$	(81.3) $LowGlrEm$
	(69.0) $SumAvg$	(75.8) Avg	(72.7) Avg	(65.8) Avg
	(68.8) $LowGlrEm$	(71.5) $HighGlrEm$	(72.3) $HighGlrEm$	(65.4) $SumAvg$
	(66.9) $HighGlrEm$	(71.3) $LowGlrEm$	(66.0) $LowGlrEm$	(51.7) $HighGlrEm$
	(65.0) $GraVar$	(51.4) $RlEntr$	(25.5) $Entr$	(50.5) $gAvg$
Phase 3	(94.7) $GlNonUni$	(77.2) $LowGlrEm$	(80.5) $S3S3$	(96.4) $GlNonUni$
	(86.3) $RlNonUni$	(76.1) $HighGlrEm$	(79.9) $HighGlrEm$	(88.8) $GraAvg$
	(85.9) $LowGlrEm$	(66.4) Avg	(78.0) $LowGlrEm$	(88.7) $RlNonUni$
	(83.2) $HighGlrEm$	(62.7) $SumAvg$	(76.1) Avg	(85.8) $FractalDim$
	(77.2) S_3S_3	(49.8) Var	(73.1) $SumAvg$	(83.0) $LowGlrEm$
GRMD	(83.6) $LowGlrEm$	(60.8) $gEntr$	(79.3) $RlNonUni$	(83.5) $GraAvg$
	(78.1) $HighGlrEm$	(60.2) $DiffEntr$	(73.1) $Entr$	(78.9) $LowGlrEm$
	(75.9) $SumAvg$	(53.2) $gContrast$	(63.4) $GlNonUni$	(73.4) $Entr$
	(72.7) E_3E_3	(52.4) $Contrast$	(61.1) $GraAvg$	(71.3) $SumAvg$
	(71.1) $GraVar$	(48.5) $LongEm$	(51.6) $AngSecMo$	(69.2) Avg

First we will compare the results for the differentiation between healthy and GRMD dogs at each phase of canine growth. The sets of most frequently selected features can be seen to be different for each classification task. However, some TA methods proved more useful than others for each phase. At the beginning of canine growth (the first phase), when dogs are still very small and thus ROIs occupy small areas in the images, the COM and GLDM methods seem to be the most suitable. Note that in our experiments these methods take into account only small distances between pixels in pairs. Features selected fairly often, irrespective of muscle types, are $InvDiffMo$ (COM-based) and $gInvDiffMo$ (GLDM-based). These features are numerical descriptors of local image homogeneity and are inversely related to the contrast measure. Two other features encountered,

AngSecMo and *gAngSecMo* (COM- and GLDM-based, respectively), are also measures of local homogeneity.

In the second phase, the RLM method proves more accurate. Two RLM-based features, *LowGlrEm* and *HighGlrEm*, are top-ranked for each muscle type. The usefulness of such features could be explained by changes occurring in the muscles as dystrophy progresses. In healthy dogs, muscles display regular fibers, which may result in fairly regular pixel runs in the image. In GRMD dogs, necrotic fibers are abundant and fibers may even appear in clusters. These changes may result in frequent small darker and lighter regions in the image. In fact, the *LowGlrEm* and *HighGlrEm* features reflect the distribution of low and high gray level values, respectively. These distributions differ between the two dog classes. Two other frequently selected features, *Avg* (GLH), and *SumAvg* (COM), are also related to pixel gray-level distributions.

In the third phase, almost exclusively RLM-based features are at the top of all four rankings. In addition to the still common *LowGlrEm* and *HighGlrEm* features, *GlNonUni* and *RlNonUni* are encountered as well. The *Avg* and *SumAvg* features are also selected, but not so often as in the second phase.

Differentiation between the three phases of dystrophy progression in GRMD dogs requires the use of features derived from several TA methods. In this case identification of the best method is more difficult. At least one RLM- and one COM-based feature prove to be good tissue descriptor for each muscle type. The usefulness of RLM-based features could be explained again by already described changes in affected muscles. Continuous development of the dystrophy entails the disappearance of regular muscle structures (visible on MRI as homogeneous texture primitives and quite regular pixel runs). The same reasons determine the suitability of COM- and GM-based features, selected as well. Features obtained from above methods are various measures of the presence, frequency and size of different texture elements, changing over the dystrophy progression.

5.2   Estimation of a Sufficient Number of Selected Features

The plots in Fig. 2 show how the classification quality changes when consecutive features from the feature incidence frequency rankings are added to the set of tissue descriptors. We will present only the results obtained by the AdaBoost classifier for the TC muscle. Similar plots were observed for other classifiers and other muscle types.

Analysis of the plots in Fig. 2 reveals that high classification accuracies can be obtained with a small subset of the set of features originally considered. Moreover, adding the top-ranked features to the set of tissue descriptors results in fairly rapid improvement of classification accuracy. However, this improvement is not strictly proportional to the number of features. Beyond a certain number of features the quality of classification does not increase significantly, and may even decrease. This threshold usually does not exceed 10 features.
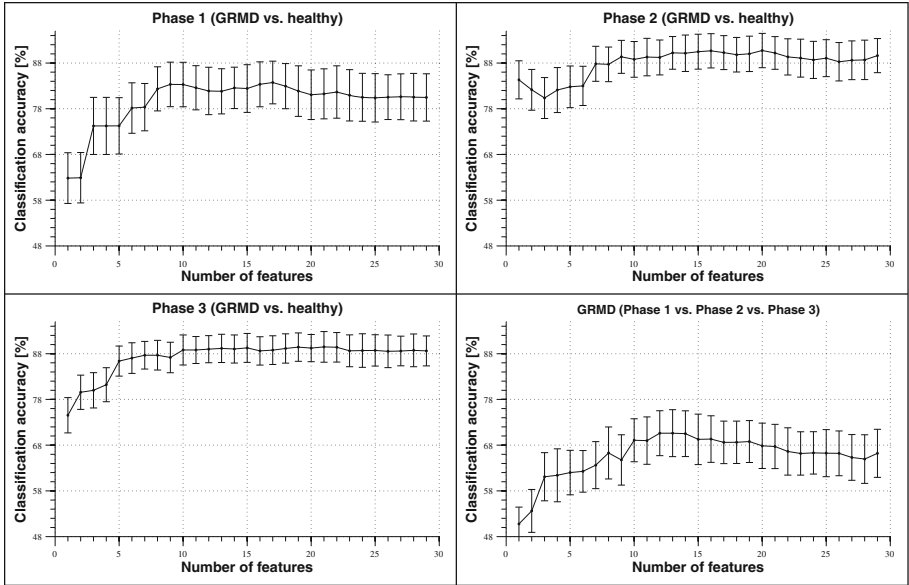
Fig. 2. Classification accuracy achieved with different numbers of the most frequently selected features. The results were obtained using the AdaBoost classifier, for the TC muscle. The first three plots concern differentiation between GRMD and healthy dogs at different phases of canine growth. The last plot concerns differentiation between tissues at three phases of dystrophy progression in GRMD dogs.

5.3 Classification Improvements Due to Feature Selection

Table 5 presents the best classification results obtained with sets of selected features. Results are given separately for each of four differentiation problems and for each of four muscle types. Each result is followed by the percentage by which the classification quality has been improved as compared to the quality obtained in the same differentiation task but considering all the available 39 features. It can be seen that using the best set of selected features always improves the classification accuracy in comparison with the case where all the possible textural features are applied as tissue descriptors. Sometimes the improvement is not significant, and feature selection mainly reduces the time and memory requirements for feature extraction and for training of classifiers. In many cases the improvement is considerable, even exceeding 10 %. The best classification results (more than 99 %) were observed in the first phase, when the presence or absence of disease was recognized in the EDL muscle (using the NN and SVM classifiers). Lower classification accuracies were achieved during differentiation between the three phases of dystrophy progression in GRMD dogs. This task seems to be the most difficult, as the best result was about 71.3 % correctly recognized cases.

Table 5. Classification accuracies [%] (and standard deviations) achieved with the best sets of selected features (for each muscle type). The first three columns concern differentiation between GRMD and healthy dogs at different phases of canine growth. The last column concerns differentiation between tissues at three phases of dystrophy progression in GRMD dogs. Results are followed by the percentage by which the classification quality was improved in comparison to the case where all possible features were used. The results were obtained with the AdaBoost, NN, and SVM classifiers.

Classifier	Muscle	Phase 1	Phase 2	Phase 3	GRMD
AdaBoost	EDL	$97.9 \pm 2.5\,(1.7)$	$93.5 \pm 3.9\,(1.3)$	$88.1 \pm 3.4\,(0.7)$	$66.7 \pm 5.2\,(0.7)$
	GasLat	$81.7 \pm 6.3\,(12.6)$	$91.5 \pm 5.6\,(3.8)$	$89.7 \pm 4.0\,(3.2)$	$56.6 \pm 7.7\,(4.3)$
	GasMed	$84.7 \pm 5.2\,(2.0)$	$93.4 \pm 5.0\,(5.1)$	$88.6 \pm 3.7\,(2.2)$	$61.1 \pm 6.1\,(4.4)$
	TC	$83.7 \pm 4.7\,(4.3)$	$90.7 \pm 3.8\,(2.9)$	$89.5 \pm 3.4\,(1.3)$	$70.6 \pm 5.2\,(4.9)$
NN	EDL	$99.9 \pm 0.5\,(0.7)$	$93.6 \pm 3.9\,(1.8)$	$87.2 \pm 3.3\,(3.1)$	$63.7 \pm 4.9\,(3.3)$
	GasLat	$83.5 \pm 6.6\,(11.9)$	$97.9 \pm 3.1\,(2.5)$	$90.8 \pm 3.7\,(5.6)$	$56.5 \pm 6.7\,(9.8)$
	GasMed	$80.0 \pm 5.9\,(2.5)$	$96.3 \pm 3.1\,(4.4)$	$89.5 \pm 3.7\,(2.7)$	$60.2 \pm 5.7\,(1.9)$
	TC	$79.9 \pm 5.6\,(2.2)$	$91.7 \pm 3.6\,(3.8)$	$88.5 \pm 3.5\,(1.9)$	$70.0 \pm 5.1\,(5.7)$
SVM	EDL	$99.8 \pm 0.7\,(0.2)$	$95.0 \pm 3.3\,(5.0)$	$87.5 \pm 3.1\,(0.1)$	$70.3 \pm 5.1\,(3.9)$
	GasLat	$78.5 \pm 5.6\,(6.9)$	$98.1 \pm 2.7\,(3.1)$	$88.7 \pm 4.1\,(4.9)$	$58.4 \pm 5.8\,(10.2)$
	GasMed	$83.8 \pm 5.8\,(3.5)$	$96.3 \pm 3.1\,(4.8)$	$88.6 \pm 4.2\,(2.0)$	$62.0 \pm 5.1\,(7.9)$
	TC	$82.6 \pm 5.1\,(1.9)$	$92.7 \pm 3.3\,(1.5)$	$88.9 \pm 3.3\,(1.0)$	$71.3 \pm 5.2\,(3.8)$

6 Conclusion and Future Work

The aim of the study was to find the best texture-based tissue descriptors for four types of canine hindlimb muscles (EDL, GasLat, GasMed, and TC) at each of three identified phases of canine growth and/or dystrophy progression. A total of 39 textural features, derived from 8 TA methods, were analyzed. Classification experiments were conducted separately for each muscle type. They involved differentiation either between healthy and GRMD dogs at different phases, or between tissues at different phases in GRMD dogs only. The experiments enabled us to conclude the following. (i) The best discrimination (or nearly the best, and not significantly different from the best) can be obtained with a small set of selected features (up to 10). (ii) The best TA methods can be different for each phase of canine growth. The COM- and GLDM-based features can be the most useful in the first phase, the RLM-based features together with some of the GLH- and COM-based features in the second, and the RLM features in the third. (iii) Classification accuracy can be significantly improved when only a few selected features are used as tissue descriptors.

Differentiation between three phases of disease progression in GRMD dogs proved to be the most difficult task. However, it is important to find a satisfactory solution to this problem if texture analysis is to be applied to assessment of the canine response to treatment. In the future, experiments should be conducted on a larger database. Other TA methods could also be considered. It would be

interesting to analyze tissue descriptors derived from different muscle types at a time. Other image modalities could also be considered. Finally, it would be useful to describe how textural features change over the dystrophy progression.

References

1. Haldeman-Englert, C.: Duchenne Muscular Dystrophy: MedlinePlus Medical Encyclopedia. Medline Plus. U.S. National Library of Medicine (2014). https://www.nlm.nih.gov/medlineplus/ency/article/000705.htm
2. Sarnat, H.B.: Muscular dystrophies. In: Kliegman, R.M., Stanton, B.F., Geme, J.W., Schor, N.F., Behrman, R.E. (eds.) Nelson Textbook of Pediatrics, 19th edn. Saunders Elsevier, Philadelphia (2011)
3. Kornegay, J.N., Bogan, J.R., Bogan, D.J., Childers, M.K., Li, J., et al.: Canine models of Duchenne muscular dystrophy and their use in therapeutic strategies. Mamm. Genome **23**(1–2), 85–108 (2012)
4. De Certaines, J.D., Larcher, T., Duda, D., Azzabou, N., Eliat, P.A., et al.: Application of texture analysis to muscle MRI: 1-What kind of information should be expected from texture analysis? EPJ Nonlinear Biomed. Phys. **3**(3), 1–14 (2015)
5. Lerski, R.A., de Certaines, J.D., Duda, D., Klonowski, W., Yang, G., et al.: Application of texture analysis to muscle MRI: 2-Technical recommendations. EPJ Nonlinear Biomed. Phys. **3**(2), 1–20 (2015)
6. Castellano, G., Bonilha, L., Li, L.M., Cendes, F.: Texture analysis of medical images. Clin. Radiol. **59**(12), 1061–1069 (2004)
7. Hajek, M., Dezortova, M., Materka, A., Lerski, R.A. (eds.): Texture Analysis for Magnetic Resonance Imaging. Med4Publishing, Prague (2006)
8. Nailon, W.H.: Texture analysis methods for medical image characterisation. In: Mao, Y. (ed.) Biomedical Imaging, pp. 75–100. InTech Open (2010)
9. Duda, D., Kretowski, M., Azzabou, N., de Certaines, J.D.: MRI texture analysis for differentiation between healthy and golden retriever muscular dystrophy dogs at different phases of disease evolution. In: Saeed, K., Homenda, W. (eds.) CISIM 2015. LNCS, vol. 9339, pp. 255–266. Springer, Heidelberg (2015)
10. Draminski, M., Rada-Iglesias, A., Enroth, S., Wadelius, C., Koronacki, J., Komorowski, J.: Monte Carlo feature selection for supervised classification. Bioinformatics **24**(1), 110–117 (2008)
11. Fan, Z., Wang, J., Ahn, M., Shiloh-Malawsky, Y., Chahin, N., et al.: Characteristics of magnetic resonance imaging biomarkers in a natural history study of golden retriever muscular dystrophy. Neuromuscul. Disord. **24**(2), 178–191 (2014)
12. Galloway, M.M.: Texture analysis using gray level run lengths. Comput. Graph. Image Process. **4**(2), 172–179 (1975)
13. Yang, G., Lalande, V., Chen, L., Azzabou, N., Larcher, T., et al.: MRI texture analysis of GRMD dogs using orthogonal moments: a preliminary study. IRBM **36**(4), 213–219 (2015)
14. Vapnik, V.N.: The Nature of Statistical Learning Theory, 2nd edn. Springer, New York (2000)

15. Haralick, R.M., Shanmugam, K., Dinstein, I.: Textural features for image classification. IEEE Trans. Syst. Man Cybern. Syst. **SMC−3**(6), 610–621 (1973)
16. Guyon, I., Elisseeff, A.: An introduction to variable and feature selection. J. Mach. Learn. Res. **3**, 1157–1182 (2003)
17. Thibaud, J.L., Azzabou, N., Barthelemy, I., Fleury, S., Cabrol, L., et al.: Comprehensive longitudinal characterization of canine muscular dystrophy by serial NMR imaging of GRMD dogs. Neuromuscul. Disord. **22**(Suppl. 2), S85–S99 (2012)
18. Duda, D.: Medical image classification based on texture analysis. Ph.D. thesis, University of Rennes 1, Rennes, France (2009)
19. Lerski, R., Straughan, K., Shad, L., Boyce, D., Bluml, S., Zuna, I.: MR image texture analysis - an approach to tissue characterization. Magn. Reson. Imaging **11**(6), 873–887 (1993)
20. Weszka, J.S., Dyer, C.R., Rosenfeld, A.: A comparative study of texture measures for terrain classification. IEEE Trans. Syst. Man Cybern. **6**(4), 269–285 (1976)
21. Laws, K.I.: Textured image segmentation. Ph.D. thesis, University of Southern California, Los Angeles, CA, USA (1980)
22. Chen, E.L., Chung, P.C., Chen, C.L., Tsai, H.M., Chang, C.I.: An automatic diagnostic system for CT liver image classification. IEEE Trans. Biomed. Eng. **45**(6), 783–794 (1998)
23. Gonzalez, R.C., Woods, R.E.: Digital Image Processing, 2nd edn. Addison-Wesley, Reading (2002)
24. Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., Witten, I.H.: The WEKA data mining software: an update. SIGKDD Explor. **11**(1), 10–18 (2009)
25. Quinlan, J.: C4.5: Programs for Machine Learning. Morgan Kaufmann, San Francisco (1993)
26. Freund, Y., Shapire, R.: A decision-theoretic generalization of online learning and an application to boosting. J. Comput. Syst. Sci. **55**(1), 119–139 (1997)
27. Rojas, R.: Neural Networks. A Systematic Introduction. Springer, Berlin (1996)
28. Platt, J.C.: Fast training of support vector machines using sequential minimal optimization. In: Scholkopf, B., Burges, C.J.C., Smola, A.J. (eds.) Advances in Kernel Methods - Support Vector Learning, pp. 185–208. MIT Press, Cambridge (1998)