Golden Retriever Muscular Dystrophy
MRI-based tissue characterization, texture analysis
Monte Carlo feature selection, classification.

Dorota DUDA [1]

TEXTURE-BASED IDENTIFICATION OF DYSTROPHY PHASE. INDICATING THE MOST SUITABLE FEATURES FOR THERAPY TESTING

In this study, texture analysis (TA) is applied for characterization of dystrophic muscles visualized on T2-weighted Magnetic Resonance (MR) images. The study proposes a strategy for indicating the textural features that are the most appropriate for testing the therapies of Duchenne muscular dystrophy (DMD). The strategy considers that muscle texture evolves not only along with the disease progression but also with the individual's development. First, a Monte Carlo (MC) procedure is used to assess the relative importance of each feature in identifying the phases of growth in healthy controls. The features considered as age-dependent at a given acceptance threshold are excluded from further analyses. It is assumed that their application in therapies' evaluation may entail an incorrect assessment of dystrophy response to treatment. Next, the remaining features are used in differentiation among dystrophy phases. At this step, an MC-based feature selection is applied to find an optimal subset of features. Experiments are repeated at several acceptance thresholds for age-dependent features. Different solutions are finally compared with two classifiers: Neural Network (NN) and Support Vector Machines (SVM). The study is based on the *Golden Retriever Muscular Dystrophy* (GRMD) model. In total, 39 features provided by 8 TA methods (statistical, filter- and model-based) are tested.

1. INTRODUCTION

Duchenne muscular dystrophy (DMD) is a hereditary genetic disorder resulting from a deficiency of dystrophin, a protein that plays a key role in supporting fiber strength [12]. It is the most common and the most severe muscular dystrophy, affecting predominantly male children and young men. The disease is characterized by progressive muscle destruction, which implies decrease of mobility, deformities, cardiomyopathy, and respiratory failure. Despite extensive attempts to develop an effective therapy for DMD, there is still no cure, and affected individuals usually die in their second or third decade of life [24].

An important problem confronted while elaborating new therapeutic strategies is the choice of tools for the assessment of treatment effects. The need for multiple repetition of such an assessment, often over a short period of time, calls into question the application of generally effective invasive methods, such as histological examination of biopsy specimens. In fact, needle biopsies can aggravate the condition of muscles already damaged by the disease. Other measurement protocols, e.g. based on evaluation of motor function, respiratory function, muscle strength, or disability, are not entirely satisfactory and sometimes difficult to perform

[1]Faculty of Computer Science, Bialystok University of Technology, Wiejska 45a, 15-351 Bialystok, Poland.

in non-ambulant patients [7]. In this context, an increasing interest is being shown in the use of Magnetic Resonance Imaging (MRI), which is non-invasive and can provide relevant information about the dystropy progression [9]. However, the correct interpretation of image content is not a trivial task and still under investigation. The results obtained so far, indicate that this problem can be successfully managed with the use of texture analysis (TA) [2].

Several studies have already evaluated the usefulness of texture analysis in characterization of muscular disorders (they are discussed in the next section). All of them considered only the cases of untreated dystrophy, progressing along with the growth of an individual, from the earliest months of its life. If texture analysis was to be applied in testing the therapy's effects, it should be taken into account that muscle texture can evolve over time, as a result of two processes occurring simultaneously: the individual's development and the course of the disease. Identifying the stage of untreated dystrophy does not require evaluation how each of these processes separately influences the characteristics of muscle texture. When therapies are introduced, the second process can be slowed down, while the first one still advances. In that case, special attention should be paid to those textural features which values may evolve under the influence of the individual's development. Ignoring such an evolution during the assessment of treatment effects can alter the correct evaluation of tested therapy.

In this study, MRI texture analysis is used in classification-based system for identifying phases of dystrophy progression. The study focuses on indicating textural features that can be the most appropriate for the assessment of DMD therapies. First, it finds and eliminates features which values may evolve along with the individual's growth. This step is based on a Monte Carlo (MC) selection procedure (originally described in [3]) performed when stages of growth in healthy dogs are recognized. Features with a relatively high frequency of selections are considered to be age-dependent, as they contribute the most to the age identification. Further, the work assesses what is the best possible differentiation of dystrophy phases that can be achieved with only the use of remaining features, independent of age. At this step, two classifiers are used for evaluating different subsets of features: Neural Network (NN) [1] and Support Vector Machines (SVM) [27]. Moreover, the study compares results achieved at different acceptance thresholds for age-dependent features and attempts to find the optimal threshold. In total, 39 features calculated with 8 TA methods (statistical, filter- and model-based) are tested.

The study is based on the *Golden Retriever Muscular Dystrophy* (GRMD) canine model, commonly used in studies on DMD pathogenesis and treatment development [18]. Three phases of canine growth and/or dystrophy progression are distinguished in this model, in reference to histological changes in muscle structure [2]: the first phase (0-4 months of age), the second phase (more than 4 to 6 months), and the third phase (more than 6 months to death).

The next section includes an overview of related works. Section 3 describes the strategy proposed in this study. Next, details of the experimental setup are given. In Section 5 the results are presented and discussed. Conclusions and future work are outlined in the last section.

2. RELATED WORK

So far, several studies on the possibility of applying TA for characterization of muscular disorders have been conducted using different animal models. They investigated the problem of automated classification of healthy and dystrophic muscles [5], [28], [30], differentiation among several phases of dystrophy progression in affected individuals [6], [20], or attempted to describe changes in muscle properties resulting from the disease development [8], [28]. They mainly used statistical methods (first- and second-order statistics), as well as different filter- and model-based approaches. These works demonstrated that texture analysis can outperform standard radiologists' clinical evaluation, and that a relatively high recognizability of considered

muscle groups can be achieved if textural features are properly selected. In general, effective features have turned out to be those derived from the co-occurrence matrices (COM) [15], the run length matrices (RLM) [11], or the gray level difference matrices (GLDM) [29].

Widely available literature does not provide a description of such a study that concerns a texture-based classification system for identifying dystrophy phase and, at the same time, focuses on finding (and eliminating) those features which values may evolve along with the individual's development. In fact, previous studies examined only the possibility of using TA to recognize a stage of untreated dystrophy, progressing with age, and did not consider its application for therapy testing. Some works, however, deserve special attention.

Martins-Bach et al. [20] investigated the potential of texture analysis in characterizing muscles in four mouse models of muscular dystrophy: the severely affected $Large^{myd}$ mouse, the worst double mutant $mdx/Large^{myd}$ mouse, the mildly affected mdx mouse, and normal mice (represented by 12, 9, 13, and 13 subjects, respectively). Various measurements derived from MRI T2-weighted images (e.g. transverse relaxation times of MRI contrast) were compared to selected textural features. In total, 371 features provided by the *MaZda* software [25] were analyzed. The best texture-based characteristics were those derived from the co-occurrence matrices: contrast- and entropy-based (30 features). They allowed to unambiguously identify all considered mouse strains and outperformed non-texture-based muscle T2 values.

The study [6] analyzed muscle textures on T2-weighted MRI images derived from 5 GRMD dogs. The relative importance of each textural feature in differentiating among three dystrophy phases was assessed by a Monte Carlo procedure. In total, 39 features obtained from 8 TA methods were tested. Four different types of muscles were considered. Three classifiers were used: adaptive boosting [10] with a *C4.5* tree [23], NN, and SVM. Experiments enabled to find the most discriminative features (mainly the COM-, RLM-, and GLDM-based ones) and demonstrated that the optimal set of features is different for each muscle. Moreover, they revealed that differentiation among three phases of dystrophy progression is quite a difficult task, as it was possible to recognize the dystrophy phase in a maximum of 71.3% cases.

Fan et al. [8] differentiated between GRMD and healthy dogs at three different dog's ages. Their work was performed on T2-weighted images with and without fat saturation, derived from 10 GRMD and 8 healthy dogs. Several MRI imaging biomarkers and three texture biomarkers were quantified in seven muscles. Three textural features were used: a gray level histogram-based entropy, and RLM-based short run emphasis and run length non-uniformity. The latter one performed best, giving statistically different values for GRMD and healthy dogs at each phase. Moreover, classification done with Linear Discriminant Analysis (LDA) showed better potential of textural features in comparison with other tested biomarkers. Finally, the study demonstrated that muscle texture may also evolve along with the dog's development. However, the research did not attempt to determine in which way the dog's development influences the summary changes in textural properties of affected muscles.

3. INDICATING THE MOST APPROPRIATE FEATURES FOR THERAPY TESTING

Three steps can be distinguished in the proposed strategy. The first one consists in indicating the features that demonstrate a relatively high usefulness in identifying phases of growth in healthy dogs. Such features (or their combinations) may potentially evolve under the influence of the individual's development. In this study, the feature usefulness for a given classification problem is determined by a frequency of feature selections in the modified Monte Carlo procedure [6]. This approach is based on the finding that a feature that is completely useless by itself can show its significant potential when taken with others [13]. Therefore, the MC procedure assesses the feature's ability to perform well regardless the combination in which

the feature is found. In fact, classifiers applied in the system for dystrophy identification, make their decisions basing on the combinations of many features and not on each feature separately.

In the second step, features previously demonstrating a relatively high frequency of selections (exceeding a given acceptance threshold, $accThr$) are considered as age-dependent and eliminated from further investigations. However, the choice of the threshold is not obvious. Accepting features too frequently selected (i.e. contributing too much to the age identification) may result in an inaccurate evaluation of dystrophy response to treatment. In turn, setting the threshold at a very low level may entail discarding of most of the features, also those that depend much more on the disease progression than on the individual's development. Due to this reason, several acceptance thresholds are tested within this study.

Finally, the remaining features (considered as age-independent at a given acceptance threshold) are used in differentiation among dystrophy phases in affected individuals. At this step, the modified Monte Carlo procedure is run again and features with the highest frequencies of selections are considered as the most useful in dystrophy identification. After completing the modified MC procedure, features are ranked according to their frequency of selections, from the most to the least selected, and different numbers of the top-ranked features are tested. The subset of features ensuring the best recognition of dystrophy phase is considered optimal.

4. EXPERIMENTAL SETUP

4.1. DATABASE

The experiments were conducted on a database of images provided by the Nuclear Magnetic Resonance Laboratory of the Institute of Myology in Paris, France. All the details concerning the acquisition protocols are presented in [26]. Images were acquired on a 3T Siemens Magnetom Trio TIM scanner with a standard, circularly polarized extremity coil. The in-plane resolution was 0.56 mm $\times$ 0.56 mm, the slice thickness was 3 mm, and the inter-slice gap was 7.5 mm. The slice orientation was axial with respect to the long axis of the muscle. As for the T2-weighted image series, the repetition time (TR) was 3,000 ms, the first echo time ($TE1$) was 6.3 ms and the second echo time ($TE2$) was 50 ms. Each series contained from 12 to 14 images. All images had a size of 240 $\times$ 320 pixels and were provided in *Analyze* format.

Due to a relatively high cost of the trial, only five GRMD and five healthy dogs were involved in the experiment. Each dog was imaged from 3 to 5 times over a maximum of 14 months. In total, 38 examinations were performed. Each examination was assigned to one of three mentioned phases of canine growth and/or dystrophy progression. In total, the first, second and third phase were represented by 14, 9, and 15 examinations, respectively.

Four types of muscles were considered: the *Extensor Digitorum Longus* (EDL), the *Gastrocnemius Lateralis* (GL), the *Gastrocnemius Medialis* (GM), and the *Tibial Cranialis* (TC). For each muscle, up to two *Regions of Interest* (ROIs) were delineated on each image – a maximum of one within each limb, left and right. Only ROIs of at least 100 pixels were analyzed. The numbers of used ROIs are given in Table 1, separately for each cohort type (GRMD, healthy), each muscle, and each phase. In turn, Table 2 contains the average ROIs' sizes.

4.2. TEXTURE-BASED TISSUE CHARACTERIZATION

First, images were pre-processed as described in [5] and converted from initial *Analyze* format to 8-bit BMP format with 256 gray levels. Then, muscular tissue within each ROI was characterized using 8 different texture analysis methods implemented by the author as a part of the *Medical Image Processing* application [4]. In total, 39 textural features were calculated:

Table 1. Numbers of accepted ROIs

phase	cohort type	Muscle			
		EDL	GL	GM	TC
first	healthy	52	30	60	73
	GRMD	45	43	64	53
second	healthy	48	24	37	64
	GRMD	56	34	43	87
third	healthy	136	85	113	157
	GRMD	73	31	60	81

Table 2. Average sizes of accepted ROIs

phase	cohort type	Muscle			
		EDL	GL	GM	TC
first	healthy	202	161	290	205
	GRMD	156	189	293	165
second	healthy	239	184	395	255
	GRMD	189	220	379	250
third	healthy	279	220	426	316
	GRMD	160	199	328	236

- Avg (average), Var (variance), $Skew$ (skewness), and $Kurt$ (kurtosis) – first order statistics, obtained from a gray level histogram (abbreviated GLH),
- $AngSecMom$ (angular second moment), $Contrast$ (contrast), $Corr$ (correlation), $SumSqr$ (variance), $InvDiffMom$ (inverse difference moment), $SumAvg$ (sum average), $SumVar$ (sum variance), $SumEntr$ (sum entropy), $Entr$ (entropy), $DiffVar$ (difference variance), and $DiffEntr$ (difference entropy), from the co-occurrence matrices (COM),
- $ShortEmph$ (short run emphasis), $LongEmph$ (long run emphasis), $GlNonUni$ (gray level non-uniformity), $RlNonUni$ (run length non-uniformity), $Fraction$ (fraction of image in runs), $LowGlrEmph$ (low gray level run emphasis), $HighGlrEmph$ (high gray level run emphasis), and $RlEntr$ (run length entropy), from the run length matrices (RLM),
- $dContrast$ (contrast), $dAngSecMom$ (angular second moment), $dEntr$ (entropy), $dMean$ (mean), and $dInvDiffMom$ (inverse difference moment), form the gray level difference matrices (GLDM),
- $EntrE3L3$, $EntrS3L3$, $EntrS3E3$, $EntrE3E3$, and $EntrS3S3$ – entropy of an image region filtered, respectively, with the following pairs of the Laws' masks: E3L3 and L3E3, S3L3 and L3S3, S3E3 and E3S3, E3E3 and E3E3, and S3S3 and S3S3 (LTE),
- $GradAvg$ (average), $GradVar$ (variance), $GradSkew$ (skewness), and $GradKurt$ (kurtosis), from the gradient matrix (GM),
- $FractalDim$ – fractal dimension based on the fractional Brownian motion model (FB),
- $AutoCorr$ – normalized autocorrelation coefficient (AC).

A detailed description of the above features and their reference to the properties of muscle tissue, as well as basis of corresponding TA methods can be found in [19].

The construction of the co-occurrence matrices, run length matrices, and gray level difference matrices was performed with a reduced number of gray levels, 64 instead of initial 256. Four standard directions of pixel runs ($0°$, $45°$, $90°$, and $135°$) were considered for the COM, RLM, GLDM, FB, and AC methods. Due to an irregular shape of many ROIs, only two smallest possible distances between pixels in pairs (1 and 2) were taken into account when applying the COM, GLDM, FB, and AC methods. If the same feature was calculated at several directions and (if applied) distances between pixels in pairs, its average value was used.

4.3. MONTE CARLO-BASED ASSESSMENT OF FEATURE IMPORTANCE

The modified Monte Carlo procedure, used in this study, consists in multiple repetitions of a single selection experiment, run on a "truncated" data set. Such data set is created by a random choice of a fixed number of observations from the initial data set, described with a relatively small part of initially used features. Each time, different subset of observations and features is chosen. After repeating this a relatively large number of times, a frequency of selections ("incidence frequency rate", IFR) is calculated for each feature. This is the ratio between the number of cases in which the feature was selected and the number of times it occurred in the subsets of randomly chosen features in truncated data sets, subjected to a selection.

A single selection experiment, executed within this procedure, was repeated 200,000 times. Truncated data sets were always composed of 2/3 of initial number of observations described by 20% of the initially used features (here, 8 features were always randomly chosen). The space of subsets of features was searched using the best-first strategy with the *Sequential Forward Selection* (SFS) algorithm [22]. During the search, a supervised wrapper method [17] combined with the *C4.5* decision tree classifier and a 10-fold cross-validation was applied for evaluation of candidate subsets of features. All the selection procedures were carried out with the *Weka* software [14]. Additional tools for generating truncated data sets, as well as creating feature incidence frequency rankings were implemented by the author in the C++ language.

4.4. CLASSIFICATION

Classification was also performed using *Weka*. Five classifiers were applied to assess the potential of different subsets of features that do not depend (or weakly depend) on age:

- *C4.5* Decision Tree (DT),
- Ensemble of Classifiers with an Adaptive Boosting voting scheme – *AdaBoostM1* (AB) [10] using the *C4.5* classifier as the underlying algorithm,
- Logistic Regression (LR) [16]
- back-propagation Neural Network with a sigmoidal activating function and one hidden layer wherein the number of neurons was equal to the average value of the number of features and the number of classes,
- nonlinear Support Vector Machines using the *Sequential Minimal Optimization* (SMO) algorithm [21] and a second-degree polynomial kernel.

The classification accuracies were assessed using 10-fold cross-validation, repeated 10 times.

5. RESULTS AND DISCUSSION

The course of experiments reflected the three steps enumerated and described in Section 3. All the experiments were carried out separately for each muscle: EDL, GL, GM, and TC.

5.1. DIFFERENTIATION AMONG PHASES OF GROWTH IN HEALTHY DOGS

For each feature its relative frequencies of selections obtained with the modified MC procedure are given in Table 3. It can be observed that these frequencies vary, often significantly, among different types of muscles. However, some features demonstrate their high importance regardless the muscle that they are calculated for. The most frequently selected features, and therefore considered as the most contributing to the age identification, are: Avg, from the gray level histogram, $LowGlrEmph$ and $HighGlrEmph$, from the run length matrices, and $SumAvg$, from the co-occurrence matrices. All of them were selected in more than 50% cases for each muscle. Several other features show their high usefulness only when derived from some muscles, for example $FractelDim$ (from the EDL, GM, and TC muscles) or RLM-based $GlNonUni$ and $RlNonUni$ (from the GM and TC muscles). It should be noted that the COM- and RLM-based features were considered as generally quite effective in dystrophy identification process in previously reported studies (e.g., in [5], [6], [8], [28]). Now, it was discovered that they may evolve along with the individual's development. Therefore, a "good ability to recognize a dystrophy phase", stated for them in above-cited works, could also be related to their changes caused by dog's growth (with which the dystrophy progresses as well).

The least frequently selected features, i.e. the least contributing to the age identification, with

Table 3. Percentage of feature selections obtained by the modified Monte Carlo procedure when phases of growth in healthy dogs were differentiated. Results are given separately for each muscle: EDL, GL, GM, and TC.

Feature \ Muscle	EDL	GL	GM	TC
Avg	57.90	77.49	78.68	67.48
Var	19.57	12.98	32.74	34.57
$Skew$	62.54	13.55	24.32	29.00
$Kurt$	14.11	12.12	25.22	51.54
$AngSecMom$	34.20	13.69	33.49	41.82
$Contrast$	25.00	12.44	23.44	30.71
$Corr$	12.20	7.81	38.90	33.47
$SumSqr$	29.76	13.45	43.06	29.50
$InvDiffMom$	26.66	11.74	25.88	36.11
$SumAvg$	57.03	78.77	82.52	60.55
$SumVar$	24.06	10.78	38.02	30.67
$SumEntr$	27.81	10.00	47.43	36.38
$Entr$	31.08	11.83	38.78	44.89
$DiffVar$	21.46	13.73	39.44	25.02
$DiffEntr$	29.04	9.89	26.60	25.13
$ShortEmph$	23.94	9.78	26.40	28.62
$LongEmph$	20.23	11.12	38.13	39.79
$GlNonUni$	11.95	33.74	73.96	98.20
$RlNonUni$	18.69	42.06	63.51	87.11
$Fraction$	19.98	8.13	38.41	28.00

Feature \ Muscle	EDL	GL	GM	TC
$LowGlrEmph$	57.72	84.68	81.71	55.90
$HighGlrEmph$	55.59	73.40	78.95	56.91
$RlEntr$	27.09	9.44	57.94	33.03
$dContrast$	25.55	12.10	23.58	30.31
$dAngSecMom$	30.91	9.66	26.46	30.95
$dEntr$	27.54	9.87	26.00	25.43
$dMean$	29.01	12.36	26.27	28.19
$dInvDiffMom$	26.15	11.83	25.56	35.63
$EntrE3L3$	32.17	8.81	71.29	36.72
$EntrS3L3$	18.92	16.18	23.35	38.90
$EntrS3E3$	17.53	19.84	42.51	36.91
$EntrE3E3$	46.18	13.87	48.00	33.72
$EntrS3S3$	34.78	29.90	37.20	50.75
$GradAvg$	24.45	11.35	35.16	31.07
$GradVar$	20.20	18.10	49.97	27.61
$GradSkew$	25.21	23.02	23.59	38.65
$GradKurt$	24.63	15.56	19.06	28.13
$FractalDim$	77.73	7.13	52.63	56.19
$AutoCorr$	92.75	15.27	20.46	21.53

no more than 10% of selections, are exclusively derived from the GL muscle. These include: $Corr$, $SumEntr$, and $DiffEntr$ (COM-based), $ShortEmph$, $Fraction$, and $RlEntr$ (RLM-based), $dAngSecMom$ and $dEntr$ (GLDM-based), $EntrE3L3$ (LTE), and $FractalDim$. Many of them are based on different measures of texture entropy.

Finally, it could be deduced that setting the acceptance threshold ($accThr$) at 50% of feature selections implies discarding a relatively small number of features: 7, 4, 9, and 9 for the EDL, GL, GM, and TC muscles, respectively. The further consecutive reductions of this threshold by 5% cause rather small changes in obtained subsequent sets of discarded features. Nevertheless, setting this threshold at 25% results in a rejection of almost all features – 23, 32, and 38 (of 39) for the EDL, GM, and TC muscles, respectively. An exception is only observed for the GL muscle, the texture of which seems to show a relatively small dependence on age. The next experiments were repeated at three values of $accThr$: 50%, 40%, and 30%.

5.2. IDENTIFYING DYSTROPHY PHASE WITH AGE-INDEPENDENT FEATURES

After discarding features considered as age-dependent at a given acceptance threshold, the remaining ones were applied in identifying the dystrophy phase. As differentiation among the three considered phases turned out to be quite a difficult problem, experiments were also performed for the binary classification tasks, differentiating only between the adjacent dystrophy phases: (i) the first phase and the second phase, (ii) the second phase and the third phase.

For each classification task, the classification accuracies (with standard deviation) obtained for optimal subsets of accepted features are presented in Table 4. The results achieved with all tested features (including the age-dependent ones) are given here for reference. Although each of the five classifiers (DT, AB, LR, NN, and SVM) was applied in this experiment, only the results obtained by the last two classifiers are shown and discussed here. In fact, the *C4.5* Decision Tree provided the least satisfying classification qualities (they were up to 10% lower than those obtained with the use of the NN or SVM classifiers). In turn, the application of Logistic Regression and *AdaBoostM1* algorithm led to the results similar (but often slightly worse) to those obtained with the NN or SVM classifiers. After all, the conclusions drawn

from the results ensured the NN and SVM classifiers coincide closely with those provided by the LR and AB classifiers.

Table 4. Classification accuracy and standard deviation [%] achieved with all the features tested and those considered as age-independent at different acceptance thresholds, $accThr$. The number of features in optimal subset is given in square brackets. Results were obtained with the NN and SVM classifiers separately for each muscle: EDL, GL, GM, and TC.

	Problem	Feature set	EDL	GL	GM	TC
NN classifier	differentiation among three phases	all tested	60.38 (6.01)	46.73 (7.18)	58.37 (5.49)	64.35 (4.84)
		$accThr$ = 50%	59.74 (6.18) [8]	57.89 (6.61) [6]	54.78 (5.58) [7]	67.36 (4.33) [6]
		$accThr$ = 40%	59.66 (5.65) [6]	57.89 (6.61) [6]	55.70 (6.01) [21]	63.41 (4.37) [6]
		$accThr$ = 30%	59.66 (5.65) [6]	57.89 (6.61) [6]	53.98 (6.18) [11]	57.21 (4.76) [4]
	first phase vs. second phase	all tested	72.45 (6.67)	67.80 (7.54)	78.75 (5.64)	75.00 (5.36)
		$accThr$ = 50%	75.91 (6.94) [18]	72.34 (6.79) [5]	77.93 (5.87) [1]	80.07 (5.28) [6]
		$accThr$ = 40%	76.99 (7.35) [6]	72.34 (6.79) [5]	77.93 (5.87) [1]	75.36 (5.35) [7]
		$accThr$ = 30%	77.22 (6.00) [22]	72.34 (6.79) [5]	72.52 (5.82) [6]	73.79 (5.51) [8]
	second phase vs. third phase	all tested	68.18 (6.61)	52.55 (9.22)	64.71 (7.54)	71.94 (5.01)
		$accThr$ = 50%	72.78 (6.10) [5]	62.98 (8.67) [19]	71.99 (6.62) [22]	72.14 (5.22) [6]
		$accThr$ = 40%	76.71 (5.81) [3]	62.48 (9.18) [10]	74.90 (6.93) [9]	72.34 (5.92) [28]
		$accThr$ = 30%	76.71 (5.81) [3]	62.48 (9.18) [10]	65.45 (8.12) [12]	67.11 (5.28) [3]
SVM classifier	differentiation among three phases	all tested	66.16 (5.61)	48.14 (6.55)	54.07 (5.65)	67.53 (4.84)
		$accThr$ = 50%	58.95 (5.99) [21]	59.21 (5.93) [13]	54.25 (4.90) [25]	69.79 (4.66) [21]
		$accThr$ = 40%	57.98 (5.53) [9]	59.21 (5.93) [13]	53.84 (5.46) [23]	66.22 (4.96) [28]
		$accThr$ = 30%	57.35 (5.96) [26]	59.21 (5.93) [13]	48.50 (4.52) [13]	58.35 (5.18) [10]
	first phase vs. second phase	all tested	75.25 (6.72)	70.34 (7.96)	81.55 (5.92)	76.00 (5.19)
		$accThr$ = 50%	80.34 (6.54) [27]	77.05 (7.54) [31]	78.23 (5.99) [1]	79.21 (5.36) [19]
		$accThr$ = 40%	80.32 (6.68) [26]	75.68 (7.72) [30]	78.23 (5.99) [1]	76.71 (5.94) [9]
		$accThr$ = 30%	81.02 (6.67) [22]	74.18 (6.68) [9]	74.12 (4.31) [3]	76.29 (5.09) [9]
	second phase vs. third phase	all tested	79.85 (5.30)	60.07 (9.32)	62.01 (7.64)	74.50 (4.83)
		$accThr$ = 50%	73.01 (5.23) [1]	65.07 (9.38) [13]	74.03 (7.34) [22]	74.58 (5.09) [30]
		$accThr$ = 40%	71.93 (5.05) [5]	65.07 (9.38) [13]	71.75 (6.59) [20]	73.62 (5.22) [28]
		$accThr$ = 30%	72.51 (5.71) [25]	65.07 (9.38) [13]	61.16 (6.80) [12]	68.21 (4.96) [4]

It can be seen that the results slightly differ between the NN and the SVM classifier, however, some regularities can be reported. In the majority of cases (23 of 36 for the NN, and 20 of 36 for the SVM), the optimal subset of age-independent features indicated by the proposed methodology, outperforms the set of all the features tested. This is observed regardless the adopted value for the acceptance threshold. In many other cases, this optimal subset ensures results only slightly inferior (by less than one percent) than those obtained with a full set of features. Such a worsening is not significant. In some cases, especially at a lower threshold value, the deterioration of classification accuracies is rather important, amounting to about 5%-9%. This is more frequently observed when using the SVM classifier, providing results generally not as good as the NN one.

The choice of the optimal acceptance threshold depends on the muscle type and the classification problem. In fact, the numbers of features considered as age-dependent and discarded at a given threshold vary among different muscles. At the same threshold, fewer features are eliminated for the GL and EDL muscles, more – for the GM and TC muscles. Indeed, it is for the GL and EDL muscles that generally smaller differences between the results at different acceptance thresholds are observed. Consequently, setting the threshold at the lowest considered level (30%) for these two muscles can be an acceptable solution, especially as in some cases such a threshold guarantees a better result than that obtained at $accThr$ = 50%. Different trend is observed for the GM and TC muscles. Here, setting the threshold at 30% leads to a greater worsening of the classification accuracies. Finally, the best results for the three-class classification problem were achieved for the TC muscle, at $accThr$ = 50%. They amounted to 67.36% and 69.79%, for the NN and the SVM classifier, respectively and were higher than those obtained with the full set of features, by 3.01% and 2.26%, respectively.

For the binary classification tasks, the results are more satisfactory. However, they also vary

among different muscles. Elimination of features that contribute the most to the age identification (at $accThr$ = 50% or $accThr$ = 40%) generally does not deteriorate the classification accuracies, in comparison to those obtained with all the features. More significant worsening of results (up to 7.43%) is sometimes observed at the lowest considered acceptance threshold. Nevertheless, it is just at $accThr$ = 30% that the best results for binary classification were observed: 81.02% (ensured by the SVM classifier) and 76.71% (by the NN classifier) for problem (i) – differentiating between the first and the second dystrophy phase, and problem (ii) – differentiating between the second and the third dystrophy phase, respectively. Both those results were obtained for the EDL muscle, which now turned out to be a bit more useful than the TC muscle.

6. CONCLUSION AND FUTURE WORK

The study proposed a strategy for indicating the textural features that are the most appropriate in testing the therapies of Duchenne muscular dystrophy. The concept considers that muscle texture can evolve under the influence of both the course of the disease and the individual's development. The strategy was validated using the *Golden Retriever Muscular Dystrophy* canine model, in which three phases of canine growth and/or disease progression were identified. In total, 39 features derived from the T2-weighted MRI images were tested. First, a relative importance of each feature in identifying the phase of growth in healthy dogs was assessed by the modified Monte Carlo procedure. Features contributing too much to the age identification (considered as age-dependent at a given acceptance threshold) were excluded from further examinations. Next, the modified Monte Carlo procedure was run again, in order to find an optimal subset of features while dystrophy phases were differentiated.

Experiments have shown that many features potentially evolving with age do exist. Moreover, the feature importance in identifying phases of growth in healthy dogs may vary among different types of muscles. In general, elimination of features contributing the most to the age identification ($accThr$ = 50%) does not significantly deteriorate the maximum system's ability to recognize dystrophy phases. An important role may also be played here by further selection of features (from the features recognized as age-independent). Finally, setting the acceptance threshold at a very low level may result in elimination of a large number of features and therefore – may lead to an unsatisfactory identification of dystrophy phase.

A certain disadvantage of the database used is undoubtedly its small size. However, it should be taken into account that hardly any research on image-based follow-up of the disease have been performed so far, and there is still no large open database available for non-commercial studies. Any other work on this subject is also based on rather small data repositories. The latter fact can be related, among others, to a relatively high cost of conducting regular experiments on dystrophic individuals. In the future, if the proposed strategy is to be applied in practice, experiments should be repeated employing a larger number of subjects. More textural features can be introduced, in particular model- and transform-based ones. Furthermore, as feature importance in identifying phases of canine growth or dystrophy development can vary among different muscles, a reasonable concept seems to be analyzing simultaneously features derived from several muscles. Other image sequences can also be used, especially the T1-weighted ones. Textural features corresponding to different image sequences can be combined as well. Other methods can be applied for assessing the usefulness of features in testing the therapy's effects, especially at the first step of the adopted methodology. Here, one of the ideas may be to perform statistical tests to indicate features with significant differences among all the considered phases of the individual's growth, e.g. using the analysis of variance. Finally, a model describing the texture evolution under dystrophy progression is worth elaborating.

ACKNOWLEDGEMENT

I would like to thank Noura Azzabou from the Institute of Myology, Nuclear Magnetic Resonance Laboratory, Paris, France for preparing and providing the database of images and ROIs on which the experiments were performed within this study. I would also like to thank Jacques D. de Certaines, Marek Kretowski, and other participants of the European COST Action BM1304, MYO-MRI for valuable comments and discussions.

This work was supported by grant S/WI/2/18 (from the Bialystok University of Technology, Bialystok, Poland), founded by the Polish Ministry of Science and Higher Education.

BIBLIOGRAPHY

[1] BISHOP C. M. Neural networks for pattern recognition. 1995. Clarendon Press, Oxford, United Kingdom.
[2] DE CERTAINES J. D., LARCHER T., DUDA D., AZZABOU N., ELIAT P.-A., ESCUDERO L. M., ET AL. Application of texture analysis to muscle MRI: 1-what kind of information should be expected from texture analysis? EPJ Nonlinear Biomed. Phys., 2015, Vol. 3. pp. 1–14.
[3] DRAMINSKI M., RADA-IGLESIAS A., ENROTH S., WADELIUS C., KORONACKI J., KOMOROWSKI J. Monte Carlo feature selection for supervised classification. Bioinformatics, 2008, Vol. 24. pp. 110–117.
[4] DUDA D. Classification d'images médicales basée sur l'analyse de texture. 2009. Rennes, France.
[5] 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. Computer Information Systems and Industrial Management. CISIM 2015, 2015, Vol. 9339 of *Lecture Notes in Comput. Sci.* Springer, Cham, pp. 255–266.
[6] DUDA D., KRETOWSKI M., AZZABOU N., DE CERTAINES J. D. MRI texture-based classification of dystrophic muscles. A search for the most discriminative tissue descriptors. Computer Information Systems and Industrial Management. CISIM 2016, 2016, Vol. 9842 of *Lecture Notes in Comput. Sci.* Springer, Cham, pp. 116–128.
[7] EUROPEAN MEDICINES AGENCY (EMA). COMMITTEE FOR MEDICINAL PRODUCTS FOR HUMAN USE (CHMP). Guideline on the clinical investigation of medicinal products for the treatment of Duchenne and Becker muscular dystrophy. 2015.
[8] FAN Z., WANG J., AHN M., SHILOH-MALAWSKY Y., CHAHIN N., ELMORE S., ET AL. Characteristics of magnetic resonance imaging biomarkers in a natural history study of golden retriever muscular dystrophy. Neuromuscul. Disord., 2014, Vol. 24. pp. 178–191.
[9] FINANGER E. L., RUSSMAN B., FORBES S. C., ROONEY W. D., WALTER G. A., VANDENBORNE K. Use of skeletal muscle MRI in diagnosis and monitoring disease progression in Duchenne Muscular Dystrophy. Phys. Med. Rehabil. Clin. N. Am., 2012, Vol. 23. pp. 1–10.
[10] FREUND Y., SCHAPIRE R. E. A decision-theoretic generalization of on-line learning and an application to boosting. J. Comput. Syst. Sci., 1997, Vol. 55. pp. 119–139.
[11] GALLOWAY M. M. Texture analysis using gray level run lengths. Comput. Graph. Image Process., 1975, Vol. 4. pp. 172–179.
[12] GUIRAUD S., AARTSMA-RUS A., VIEIRA N. M., DAVIES K. E., VAN OMMEN G.-J. B., KUNKEL L. M. The pathogenesis and therapy of muscular dystrophies. Annu. Rev. Genomics Hum. Genet., 2015, Vol. 16. pp. 281–308.
[13] GUYON I., ELISSEEFF A. An introduction to variable and feature selection. J. Mach. Learn. Res., 2003, Vol. 3. pp. 1157–1182.
[14] HALL M., FRANK E., HOLMES G., PFAHRINGER B., REUTEMANN P., WITTEN I. H. The WEKA data mining software: an update. SIGKDD Explor. Newsl., 2009, Vol. 11. pp. 10–18.
[15] HARALICK R., SHANMUGAM K., DINSTEIN I. Textural features for image classification. IEEE Trans. Syst. Man Cybern., 1973, Vol. SMC-3. pp. 610–621.
[16] HOSMER D. W. J., LEMESHOW S., STURDIVANT R. X. Applied logistic regression. 3rd ed., 2013. John Wiley & Sons, Inc., Hoboken, NJ, USA.
[17] KOHAVI R., JOHN G. H. Wrappers for feature subset selection. Artif. Intell., 1997, Vol. 97. pp. 273–324.
[18] KORNEGAY J. N. The golden retriever model of Duchenne muscular dystrophy. Skelet. Muscle, 2017, Vol. 7. pp. 1–21.
[19] LERSKI R. A., DE CERTAINES J. D., DUDA D., KLONOWSKI W., YANG G., COATRIEUX J. L., ET AL. Application of texture analysis to muscle MRI: 2 – technical recommendations. EPJ Nonlinear Biomed. Phys., 2015, Vol. 3. pp. 1–20.
[20] MARTINS-BACH A. B., MALHEIROS J., MATOT B., MARTINS P. C. M., ALMEIDA C. F., CALDEIRA W., ET AL. Quantitative T2 combined with texture analysis of nuclear Magnetic Resonance Images identify different degrees of muscle involvement in three mouse models of muscle dystrophy: mdx, Large(myd) and mdx/Large(myd). PLOS ONE, 2015, Vol. 10. pp. 1–16.
[21] PLATT J. C. Fast training of Support Vector Machines using Sequential Minimal Optimization. Advances in Kernel Methods – Support Vector Learning, 1998. MIT Press, Cambridge, MA, USA, pp. 185–208.
[22] PUDIL P., NOVOVICOVA J., KITTLER J. Floating search methods in feature selection. Pattern Recog. Lett., 1994, Vol. 15. pp. 1119–1125.

[23] QUINLAN J. R. C4.5: Programs for machine learning. 1993. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA.
[24] SALMANINEJAD A., VALILOU S. F., BAYAT H., EBADI N., DARAEI A., YOUSEFI M., ET AL. Duchenne muscular dystrophy: an updated review of common available therapies. Int. J. Neurosci., 2018, Vol. 5. pp. 1–11.
[25] SZCZYPINSKI P. M., STRZELECKI M., MATERKA A., KLEPACZKO A. Mazda–a software package for image texture analysis. Comput. Methods Programs Biomed., 2009, Vol. 94. pp. 66–76.
[26] THIBAUD J., AZZABOU N., BARTHELEMY I., FLEURY S., CABROL L., BLOT S., CARLIER P. G. Comprehensive longitudinal characterization of canine muscular dystrophy by serial NMR imaging of GRMD dogs. Neuromuscul. Disord., 2012, Vol. 22, Suppl. 2. pp. S85–S99.
[27] VAPNIK V. N. The nature of statistical learning theory. 2nd ed., 2000. Springer, New York, USA.
[28] WANG J., FAN Z., VANDENBORNE K., WALTER G., SHILOH-MALAWSKY Y., AN H., ET AL. A computerized MRI biomarker quantification scheme for a canine model of Duchenne muscular dystrophy. Int. J. Comput. Assist. Radiol. Surg., 2013, Vol. 8. pp. 763–774.
[29] WESZKA J. S., DYER C. R., ROSENFELD A. A comparative study of texture measures for terrain classification. IEEE Trans. Syst., Man Cybern., 1976, Vol. SMC-6. pp. 269–285.
[30] YANG G., LALANDE V., CHEN L., AZZABOU N., LARCHER T., DE CERTAINES J. D., ET AL. MRI texture analysis of GRMD dogs using orthogonal moments: A preliminary study. IRBM, 2015, Vol. 36. pp. 213–219.