Computers in Biology and Medicine 169 (2024) 107872

Contents lists available at ScienceDirect

Computers in Biology and Medicine

journal homepage: www.elsevier.com/locate/compbiomed

# An automated method for tendon image segmentation on ultrasound using grey-level co-occurrence matrix features and hidden Gaussian Markov random fields

Isabelle Scott [a,b,*], David Connell [c], Derek Moulton [a], Sarah Waters [a], Ana Namburete [d], Anurag Arnab [e], Peter Malliaras [c,f]

[a] Mathematical Institute, University of Oxford, Oxford, United Kingdom
[b] Orygen, The National Centre of Excellence in Youth Mental Health, University of Melbourne, Parkville, Melbourne, Australia
[c] Imaging at Olympic Park, Melbourne, Australia
[d] Oxford Machine Learning in Neuroimaging laboratory, OMNI, Department of Computer Science, University of Oxford, Oxford, United Kingdom
[e] Google Research, France
[f] Department of Physiotherapy, Monash University, Melbourne, Australia

ARTICLE INFO

*Keywords:*
Tendon
Co-occurrence matrix
Markov random field
Gaussian mixture model
Image segmentation
Expectation–maximisation algorithms

ABSTRACT

**Background:** Despite knowledge of qualitative changes that occur on ultrasound in tendinopathy, there is currently no objective and reliable means to quantify the severity or prognosis of tendinopathy on ultrasound.
**Objective:** The primary objective of this study is to produce a quantitative and automated means of inferring potential structural changes in tendinopathy by developing and implementing an algorithm which performs a texture based segmentation of tendon ultrasound (US) images.
**Method:** A model-based segmentation approach is used which combines Gaussian mixture models, Markov random field theory and grey-level co-occurrence (GLCM) features. The algorithm is trained and tested on 49 longitudinal B-mode ultrasound images of the Achilles tendons which are labelled as *tendinopathic* (24) or *healthy* (25). Hyperparameters are tuned, using a training set of 25 images, to optimise a decision tree based classification of the images from texture class proportions. We segment and classify the remaining test images using the decision tree.
**Results:** Our approach successfully detects a difference in the texture profiles of tendinopathic and healthy tendons, with 22/24 of the test images accurately classified based on a simple texture proportion cut-off threshold. Results for the tendinopathic images are also collated to gain insight into the topology of structural changes that occur with tendinopathy. It is evident that distinct textures, which are predominantly present in tendinopathic tendons, appear most commonly near the transverse boundary of the tendon, though there was a large variability among diseased tendons.
**Conclusion:** The GLCM based segmentation of tendons under ultrasound resulted in distinct segmentations between healthy and tendinopathic tendons and provides a potential tool to objectively quantify damage in tendinopathy.

## 1. Introduction

Tendinopathy is a painful and chronic condition of the tendon for which there is no gold standard imaging approach to quantify disease severity or prognosis [1]. Ultrasound (US) imaging is the most common and accessible imaging modality used to assess tendinopathy [2]. On US, tendinopathy is characterised by segments within the tendon displaying a deviation from the homogeneous grey scale texture seen in healthy tendons. A healthy tendon profile consists of parallel echogenic lines with thin hypoechoic spaces between [3]. Whilst the absence of these imaging changes is helpful in excluding a diagnosis of tendinopathy, conventional US methods are unable to reliably quantify microstructural changes which correlate with pain or prognosis [1]. Pain is known to be complex and there may be many contributors beyond microstructural changes which may explain the poor correlation. However, US imaging is affected by speckle and acoustic shadowing, is operator-dependent, and image analyses are classically performed subjectively [1]. Due to the current limitations of conventional US,

* Corresponding author at: Mathematical Institute, University of Oxford, Oxford, United Kingdom.
*E-mail address:* isabelle.scott@orygen.org.au (I. Scott).

https://doi.org/10.1016/j.compbiomed.2023.107872
Received 10 April 2023; Received in revised form 7 December 2023; Accepted 17 December 2023
Available online 20 December 2023

further investigation into quantitative and objective tools is warranted to refute or challenge the premise that imaging and pain are minimally correlated.

Ultrasound Tissue Characterisation (UTC) is the only method of quantitatively analysing and segmenting static US images of tendons which has been widely adopted by the clinical research community [4–6]. The algorithm for UTC compares grey level variability between adjacent transverse images and subsequently segments the tendon tissue on US into four distinct categories, based on the continuity of greyscale values across images [4]. A series of transverse images are taken along the entire length of the tendon at regular intervals, classically 0.2 mm apart. It is assumed that low variability in pixel intensities across corresponding points in contiguous transverse images represents healthy continuous collagen alignment [4,7]. Whilst UTC has become a popular method for tendon image analysis among clinical researchers, findings suggest limited diagnostic and prognostic capabilities [8,9]. In a study on 179 male football players, UTC provided no additional diagnostic or prognostic benefit over B-mode US [9]. A potential explanation for these findings may be that pixel intensity-based approaches for image segmentation, such as UTC, are often more sensitive to noise than filtering or statistical model-based approaches [10]. UTC also requires specialised and expensive US equipment, and a large number (100s) of images are required to span the entire tendon length [11].

Machine learning is an objective tool for classification and segmentation of images which has the potential to be applied in tendon; labels can be generated at the image level, providing a label of 'tendinopathic' or 'healthy' tendon, or at the pixel level, providing a segmentation of the image into 'tendinopathic' and 'healthy' regions. In the latter case, we can also quantify the proportion of tendinopathic tissue in an image. Neural networks, particularly convolution neural networks (CNNs), are a commonly used ML approach for medical image segmentation [12,13]. Neural networks have classically been used for supervised segmentation, requiring a gold standard segmentation of the training images. Jahanifar et al. [14],Alzyadat et al. [15] investigated the use of CNNs in the automatic segmentation of tendinopathic regions in the supraspinatus and Achilles tendon, respectively. Both authors utilised transfer learning and implemented a fully supervised approach in which a segmentation carried out by a radiologist was considered as ground truth, and hence the approach is unlikely to surpass the accuracy of the radiologist.

Recently CNNs have been used in a weakly supervised setting for medical image segmentation [16,17]. Weakly supervised segmentation involves assigning a label to each pixel whilst training the model on data with only coarse annotations, such as image level labels [18,19]. In the context of CNNs, a weakly supervised segmentation can be extracted through self-attention analysis [20]. Using an image-level labelled dataset of *tendinopathic* and *healthy* tendons, weakly supervised learning provides a means to extract distinguishing features and regions between healthy and tendinopathic tendons. This would enable the detection of pathology that may be missed by visual inspection and provide an automated means of identifying and quantifying pathology in the clinical setting, without requiring a set of segmented images for training. A key limitation of the application of CNNs to tendon, however, is the volume of data required for training [21]. Given healthy and pain-free tendons are not routinely scanned under US, acquiring a large sample of healthy tendon images poses a challenge. Instead, it may be more appropriate to consider a ML algorithm which uses targeted feature engineering based on domain knowledge, rather than deep learning approaches such as CNNs.

The grey level co-occurrence matrix (GLCM) is a commonly used approach for feature engineering in tendon image analysis. The GLCM describes the probability distribution of grey-level pairs at a specific distance and orientation within an image, or window within an image, and can be used to extract texture features [22]. Texture features generated from the GLCM, such as the *contrast*, have the advantage of being easily interpretable. The GLCM has been used as a feature extraction method to coarsely identify regions of interest (a square window over a large section of the image) in tendinopathic images [23,24] and for image classification [25,26]. The GLCM has shown promise in these applications but has not been investigated as a feature extraction method to segment tendon images at the pixel level within an unsupervised or weakly supervised framework. In contrast to identifying coarse regions of interest, providing a full segmentations of the image at the pixel level may provide direct insight into the proportion of tissue which has exhibited a loss of the normal texture seen in healthy tendons. Furthermore, for images with a high level of directionality, the GLCM orientation and distance can be chosen to capture specific microstructural changes in collagen fibre alignment [27], however, it has not yet been applied in this directed manner for tendinopathic images under US.

Filter bank approaches are an alternative method to the GLCM to generate texture-based features. Wavelet and curvelet filters have been applied to automatically segment tendon tears on US in animals [28] and humans [29] but have not been applied in tendinopathy. A key advantage of these approaches is that a rotationally invariant filter bank can be chosen such that the orientation of the image does not affect analysis [30]. However, filter banks also generate higher dimensional and less readily interpretable features than those derived from the GLCM [22,31,32].

Texture features can be paired with Bayesian model-based representations of the image for enhanced segmentation [33]. Gaussian mixture models (GMMs) are commonly used for image segmentation and clustering [34,35]. Unlike K-means clustering, a GMM allows for non-spherical clusters. A key limitation of GMMs is the lack of spatial information encoded. A Hidden Gaussian Markov random field (HGMRF) is a composite of a GMM and Markov random field which has been used for the segmentation of breast and brain lesions in medical images but has not been applied to tendons [36,37]. The Markov element introduces a set of spatial dependencies across the image. As a result, HGMRFs cope well with noisy images, providing a smooth segmentation. They also represent a flexible approach which can be combined with other techniques to derive improved segmentations of the image [37].

There is currently no gold standard technique for automated tendon image segmentation in the setting of tendinopathy. Collectively, quantitative approaches to tendon image analysis show promise but further research in this area is needed to generate a clinically reliable tool. In this work, we utilise a HGMRF model-based approach to segment a dataset of tendinopathic and healthy images based on GLCM texture features. A weakly supervised framework is used in which segmentation results are validated by assessing their use in determining image-level class labels. To our knowledge, we are the first to apply a HGMRF approach to tendinopathic images and to use the GLCM to extract features for a full segmentation of tendon images under US. Furthermore, in contrast to previous work, we specifically extract GLCM features utilising a specific orientation and distance that is biologically informed and consistent with visual changes observed by the radiologist in practice. We start, in Section 2, by introducing hidden Gaussian Markov random fields (HGMRFs) and grey-level co-occurrence matrix (GLCM) features. In Section 3 we outline how the GLCM-based texture features are calculated for a dataset of tendinopathic and healthy tendon images under US, and how these image features are captured in the HGMRF modelling framework to provide a segmentation of each image. The validation process is then explained, in which we use a decision tree approach to assess how well texture class proportions can be used to predict disease state from imaging. In Section 4 we demonstrate how the GLCM features differ between healthy and tendinopathic tendon images, and how this results in distinct texture segmentation profiles which correlate with disease state.

## 2. Preliminaries

### 2.1. Hidden Gaussian Markov random fields

A multivariate Gaussian mixture model (GMM) describes a mixture population formed by a set of discrete multivariate Gaussian distributed subpopulations $k \in \{1, \dots, K\}$. We use $\{\boldsymbol{Y}_1, \dots, \boldsymbol{Y}_N\}$ to denote a set of independent variables belonging to the mixture population and $\{\boldsymbol{y}_1, \dots, \boldsymbol{y}_N\}$ to denote their associated realisations. The subpopulation of each variable is given by the realisation of an associated hidden random variable $X_i = x_i$. The distribution of $\boldsymbol{Y}_i$ conditioned on $X_i$, termed the emission distribution, is a multivariate Gaussian: $\boldsymbol{Y}_i | X_i = k \sim \mathcal{N}(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$ where ($\boldsymbol{\mu}_k$ and $\boldsymbol{\Sigma}_k$) are the mean vector and covariance matrix associated with the subpopulation $k$.

Hidden Gaussian Markov random fields (HGMRFs) can be considered as an extension of GMMs where the variables $\{X_1, \dots, X_N\}$, and subsequently $\{\boldsymbol{Y}_1, \dots, \boldsymbol{Y}_N\}$, are dependent. A Markov dependence structure is present in which the probability of $X_i$ having the label $k$, when conditioned on the labels assigned to the neighbouring hidden variables $X_j \in \mathcal{N}(i)$ where $i \neq j$, is independent of the remaining hidden variables. Mathematically this equates to the local Markov property

$$P(X_i = k | \boldsymbol{x}_{-\{i\}}) = P(X_i = k | \boldsymbol{x}_{\mathcal{N}(i)}) \tag{1}$$

where $\boldsymbol{x}_{-\{i\}}$ denotes the realisations of all variables excluding at the location $i$, whilst $\boldsymbol{x}_{\mathcal{N}(i)}$ denotes realisations for variables in the surrounding neighbourhood of $i$. The variables which constitute the neighbourhood $\mathcal{N}(i)$ are chosen dependent on the modelling task at hand, but, for images, usually constitute some square neighbourhood around each pixel.

The probability density function for the entire hidden field $\boldsymbol{X} = \boldsymbol{x}$ can be factorised across each clique $c \in C$ (all subgroups in which all variables within the subgroup share a direct dependency) such that the joint probability takes the form

$$p(\boldsymbol{x}) = \frac{\exp\left(-\sum_{c \in C} V_c(\boldsymbol{x}_c, \beta)\right)}{Z(\beta)}, \tag{2}$$

termed a Gibbs distribution. Here $V_c$ is termed the clique potential, $\boldsymbol{x}_c$ are the realisations of $X_i$ in the clique $c$, $\beta$ is the Gibbs parameter, and the function $Z$ is a partition (normalising) function given by

$$Z(\beta) = \sum_{\boldsymbol{x} \in \mathcal{X}} \exp\left(-\sum_{c \in C} V_c(\boldsymbol{x}_c, \beta)\right), \tag{3}$$

where $\boldsymbol{x} \in \mathcal{X}$ denotes all possible realisations of the hidden field. The joint distribution for the HGMRF is then given by

$$p(\boldsymbol{x}, \boldsymbol{y}) = p(\boldsymbol{x}) p(\boldsymbol{y} | \boldsymbol{x}) = \frac{e^{-\sum_{c \in C} V_c(\boldsymbol{x}_c, \beta)}}{Z(\beta)} \prod_{i=1}^{N} \mathcal{N}(\boldsymbol{y} | \boldsymbol{\mu}_{x_i}, \boldsymbol{\Sigma}_{x_i}) \tag{4}$$

where $\mathcal{N}(\boldsymbol{\mu}_{x_i}, \boldsymbol{\Sigma}_{x_i})$ is the multivariate Gaussian density function.

### 2.2. Grey-level co-occurrence matrix

For all images, the grey-level co-occurrence matrix (GLCM) is calculated for a window $W : w_{CM1} \times w_{CM2}$ centred around each pixel. Intensity values within the image are divided into $N_{CM}$ even width bins and pixels are assigned their corresponding grey-level value $I(x, y) \in \{1, \dots, N_{CM}\}$ where $(x, y)$ is the pixel row and column coordinate. The GLCM is an $N_{CM} \times N_{CM}$ matrix specifying the number of each unique grey-level pixel pairings ('co-occurrences') within the window that occur at a given offset. For an offset of $(\Delta x, \Delta y)$, each component $(i, j)$ of the normalised GLCM $\boldsymbol{C}$ is defined mathematically by

$$\boldsymbol{C}(i, j) = \frac{1}{|\boldsymbol{C}|} \sum_{x,y \in W} \mathbf{1}_0 \left(I(x, y) - i\right) \mathbf{1}_0 \left(I(x + \Delta x, y + \Delta y) - j\right) \tag{5}$$

where $\mathbf{1}_0(\cdot)$ is the $0 - 1$ indicator function and $|\boldsymbol{C}|$ is the total number of pixel pairings. The matrix entries represent estimates of the co-occurrence probabilities. In Fig. 1 an example of the GLCM calculations for a given image window are displayed.

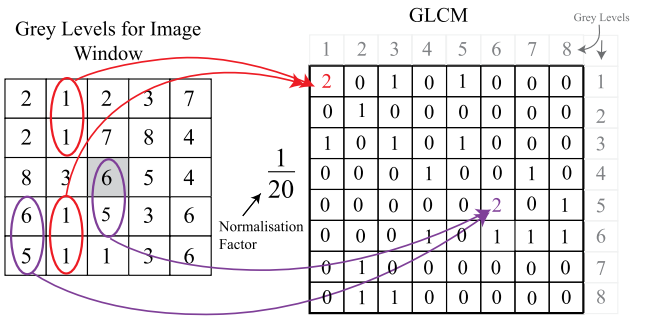


**Fig. 1.** Example of GLCM calculated for an image window with $N_{GM} = 8$. Left: Grey-level assigned to each pixel based on the pixel's intensity. Right: Corresponding normalised GLCM detailing the number of co-occurring pairs for the offset $(-1, 0)$. GLCM texture features are assigned to the central pixel (grey).

## 3. Methods

### 3.1. Image collection and pre-processing

Forty-nine images were provided by *Imaging at Olympic Park* (IOP), Melbourne. Images were obtained from patients who had been referred to IOP by their clinician and consented to their images being used for future research. Images were collected using a GE Logiq E9 US machine, using a 15Mhz linear transducer, and were downloaded as high resolution JPEG files (500 dpi). Twenty-four images were of tendinopathic tendons. Diagnosis was made by imaging features [38] and clinical presentation (by a sports physician or physiotherapist). Images were masked to include a 4 cm length along the midportion of the tendon. Grey scale values were normalised for each image and a Gaussian filter was applied, using the *MATLAB* function *imgaussfilt* (standard deviation of 1 pixel) to reduce noise.

### 3.2. GLCM features

In healthy tendons the texture pattern of bright parallel fibres interspersed with hypoechoic spaces [3] can be considered as a high-frequency signal parallel to the fibre axis. To capture disruptions to this signal, we calculate the local GLCM for each pixel with an offset of $(-1, 0)$ (one row down, same column). We use the following metrics calculated from the GLCM [22]:

$$\text{Contrast: } \sum_{i,j} |i - j|^2 C(i, j), \tag{6}$$

$$\text{Energy: } \sum_{i,j} C(x, y)^2, \tag{7}$$

as a set of texture features for each pixel. The contrast and energy have shown to correlate to disease state when the GLCM is applied at the image level [39]. We expect to see an increase in the energy and reduced contrast with loss of the high-frequency signal characteristic of healthy echotexture [39].

### 3.3. Hidden Gaussian Markov random field model

We model each $m \times n$ image as a hidden Gaussian Markov random field (HGMRF) whereby the GLCM features for each pixel $i \in \{1, 2, \dots, N = mn\}$ constitute the observable variables, with realisations given by a set of two-dimensional data points $\boldsymbol{Y}_i = \boldsymbol{y}_i$. We seek to approximate the associated unknown pixel labels $X_i = x_i$, representing the hidden field and providing a texture based segmentation of the image.

We assume that $\{Y_1, \dots, Y_N\}$ are conditionally independent given their associated pixel label $X_i \in \{1, \dots, K\}$: $\boldsymbol{Y_i} | \boldsymbol{X_i = k} \sim \boldsymbol{\mathcal{N}(\mu_k, \Sigma_k)}$. The pixel labels have a Markov dependence structure (Eq. (1)) where $\mathcal{N}(i)$

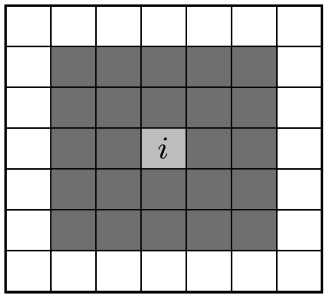


**Fig. 2.** Example of square Markov neighbourhood $\mathcal{N}(i)$. The label of pixel $i$ is directly dependent on the labels of surrounding (dark grey) pixels. A 5 × 5 neighbourhood size is displayed for visualisation.

represents a square neighbourhood surrounding each pixel (Fig. 2). The width of $\mathcal{N}(i)$ is a hyperparameter of the model, with a larger neighbourhood size leading to greater smoothing of the segmentation results. We let the clique potential in Eq. (2) have the common Ilsing form Winkler [40]

$$\sum_{c\in C} V_c(\boldsymbol{x}_c, \beta) = \sum_{i\in S}\sum_{j\in\mathcal{N}(i)} \beta \mathbf{1}_0(x_i - x_j) \tag{8}$$

which reflects prior knowledge that neighbouring pixels tend to share the same label. We refer to [40] for further details regarding how images can be viewed as hidden Gaussian Markov random fields.

### 3.4. HMRF-EM algorithm

For each image, given a set of GLCM features $\boldsymbol{y}$, we seek to determine the parameter set $\boldsymbol{\varphi} = \left((\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1), \ldots, (\boldsymbol{\mu}_K, \boldsymbol{\Sigma}_K)\right)$ which maximises the likelihood of observing the data

$$\hat{\boldsymbol{\varphi}} = \underset{\boldsymbol{\varphi}}{\operatorname{argmax}}\, p(\boldsymbol{y}|\boldsymbol{\varphi}), \tag{9}$$

which is termed the maximum likelihood estimate (MLE). We also seek to determine the most probable segmentation of the image $\boldsymbol{x}$ by approximating the maximum-a-posteriori (MAP) estimate

$$\overline{\boldsymbol{x}} = \underset{\boldsymbol{x}}{\operatorname{argmax}}\, p(\boldsymbol{x}|\boldsymbol{y}, \boldsymbol{\varphi}). \tag{10}$$

To approximate the MLE and MAP estimate, we utilise a modified version of the Expectation–Maximisation (EM) algorithm introduced by Zhang et al. [37]. The EM algorithm [41] is a popular choice to approximate the MLE in the setting of latent variables and operates by maximising over the expected log-likelihood (using the joint distribution $p(\boldsymbol{y}, \boldsymbol{x}|\boldsymbol{\varphi})$). The algorithm iterates between two steps:

*1. Expectation Step* Define the expected value of the complete log likelihood of $\boldsymbol{\varphi}$, with respect to the conditional distribution $\boldsymbol{x}|\boldsymbol{y}, \boldsymbol{\varphi}^t$:

$$Q(\boldsymbol{\varphi}; \boldsymbol{\varphi}^t) = E[\log p(\boldsymbol{y}, \boldsymbol{x}|\boldsymbol{\varphi})|\mathbf{Y} = \mathbf{y}, \boldsymbol{\varphi} = \boldsymbol{\varphi}^t], \tag{11}$$

where $\boldsymbol{\varphi}^t$ is the current parameter estimate at iteration $t$.

*2. Maximisation Step* Find the parameter set $\boldsymbol{\varphi}$ which maximises $Q(\boldsymbol{\varphi}; \boldsymbol{\varphi}^t)$

$$\boldsymbol{\varphi}^{t+1} = \underset{\boldsymbol{\varphi}}{\operatorname{argmax}}\, Q\left(\boldsymbol{\varphi}; \boldsymbol{\varphi}^t\right). \tag{12}$$

For a HGMRF, the objective function $Q$ is given by

$$Q(\boldsymbol{\varphi}; \boldsymbol{\varphi}^t) = \sum_{\boldsymbol{x}\in\mathcal{X}} p\left(\boldsymbol{x}|\boldsymbol{y}, \boldsymbol{\varphi}^t\right) \log p\left(\boldsymbol{y}, \boldsymbol{x}|\boldsymbol{\varphi}^t\right).$$

The sum over $\mathcal{X}$ is intractable due to a combinatorial explosion of the possible segmentations ($k^N$ where $N$ is the number of pixels). We use an approximation of the marginal probability, introduced by Celeux et al. [42], in which the neighbouring hidden variables are held fixed:

$$p(\boldsymbol{x}) \approx \prod_{i\in S} p\left(x_i|\boldsymbol{x}^t_{\mathcal{N}(i)}\right). \tag{13}$$

The neighbours are fixed using an approximation of the maximum-a-posteriori (MAP) estimate

$$\overline{x}^t_i = \underset{x_i}{\operatorname{argmax}}\, p\left(x_i|y_i, \boldsymbol{x}^{t-1}_{\mathcal{N}(i)}, \boldsymbol{\varphi}^{t-1}\right). \tag{14}$$

A tractable approximation of our objective function is then

$$\begin{aligned} Q(\boldsymbol{\mu}, \boldsymbol{\Sigma}; \boldsymbol{\varphi}^t) \approx & \sum_{i=1}^{N}\sum_{x_i=1}^{K} p\left(x_i|\boldsymbol{x}^t_{\mathcal{N}(i)}, y_i, \boldsymbol{\varphi}^t\right) \log f_i\left(y_i|x_i, \boldsymbol{\varphi}\right) \\ & + \sum_{i=1}^{N}\sum_{x_i=1}^{K} p\left(x_i|\boldsymbol{x}^t_{\mathcal{N}(i)}, y_i, \boldsymbol{\varphi}^t\right) \log p\left(x_i|\boldsymbol{x}^t_{\mathcal{N}(i)}\right) \\ & = Q_1\left(\boldsymbol{\varphi}; \boldsymbol{\varphi}^t\right) + Q_2\left(\boldsymbol{\varphi}^t\right) \end{aligned} \tag{15}$$

where

$$p(x_i|\boldsymbol{x}^t_{\mathcal{N}(i)}, y_i, \boldsymbol{\varphi}^t) = \frac{f_i(\boldsymbol{y}_i|x_i, \boldsymbol{\mu}^t_{x_i}, \boldsymbol{\Sigma}^t_{x_i}) p(x_i|\boldsymbol{x}^t_{\mathcal{N}(i)})}{\sum_{x_i=1}^{K} f_i(\boldsymbol{y}_i|x_i, \boldsymbol{\mu}^t_{x_i}, \boldsymbol{\Sigma}^t_{x_i}) p(x_i|\boldsymbol{x}^t_{\mathcal{N}(i)})}. \tag{16}$$

Here $f_i(\cdot)$ represents the Gaussian density function. Combining Eqs. (2) with (8) we find

$$p(x_i|\boldsymbol{x}_{\mathcal{N}(i)}) = \frac{\exp\left(-\sum_{j\in\mathcal{N}(i)} \beta \mathbf{1}_0(x_i - x_j)\right)}{\sum_{x_i=1}^{K} \exp\left(-\sum_{j\in\mathcal{N}(i)} \beta \mathbf{1}_0(x_i - x_j)\right)}. \tag{17}$$

Maximising (15), the MLE estimates at iteration $t$ are given by

$$\hat{\boldsymbol{\mu}}_k = \frac{\sum_{i=1}^{N} P\left(x_i = k|\boldsymbol{x}^t_{\mathcal{N}(i)}, y_i, \boldsymbol{\varphi}^t\right) \boldsymbol{y}_i}{\sum_{i=1}^{N} P\left(x_i = k|\boldsymbol{x}^t_{\mathcal{N}(i)}, y_i, \boldsymbol{\varphi}^t\right)}, \tag{18}$$

$$\hat{\boldsymbol{\Sigma}}_k = \frac{\sum_{i=1}^{N} P\left(x_i = k|\boldsymbol{x}^t_{\mathcal{N}(i)}, y_i, \boldsymbol{\varphi}^t\right) (\boldsymbol{y}_i - \hat{\boldsymbol{\mu}}_k)(\boldsymbol{y}_i - \hat{\boldsymbol{\mu}}_k)^T}{\sum_{i=1}^{N} P\left(x_i = k|\boldsymbol{x}^t_{\mathcal{N}(i)}, y_i, \boldsymbol{\varphi}^t\right)} \tag{19}$$

for $k = 1, \ldots, K$.

Psuedocode for the initial clustering of GLCM features and subsequent HMRF-EM processing is provided in Appendix A.1.

### 3.5. Initial parameter estimation

The HMRF-EM algorithm is applied independently to each image. We use K-means clustering over the entire training set to generate the initial parameter values $\boldsymbol{\varphi}^{t=0}$ and segmentation for each image $\boldsymbol{x}^{t=0}$. Since the EM and K-means algorithms are sensitive to initialisation (both algorithms are only locally convergent), we perform three random initialisations of K-means and choose the optimal (in terms of the K-means objective function) result. For the test images, initial pixel labels are determined by assigning the pixels to the closest cluster centroid, calculated from the training set. After this initial K-means segmentation, texture class labels $\{1, 2, \ldots, K\}$ across all images refer to subpopulations from the same Gaussian mixture population. During HMRF-EM implementation, small updates to the mean and variance are made for each image individually, and hence labels between images no longer correspond to the exact same mixture population. However, since the HMRF-EM mainly serves to reduce noise and provide a more distinct segmentation, we show that corresponding labels between images to refer to approximately the same subpopulations. Hence labels across images are comparable for the purpose of classifying images and validating our segmentation results.

### 3.6. Classification and hyperparameter tuning

To validate our segmentation results we fit a decision tree to classify the images as *tendinopathic* or *healthy* based on texture class proportions within each image. The standard CART algorithm [43] in *MATLAB* was used and the number of splits was limited to two. Hyperparameters were tuned using grid-based optimisation of the decision tree results on 25 of the images, termed the training set. Because the HMRF-EM

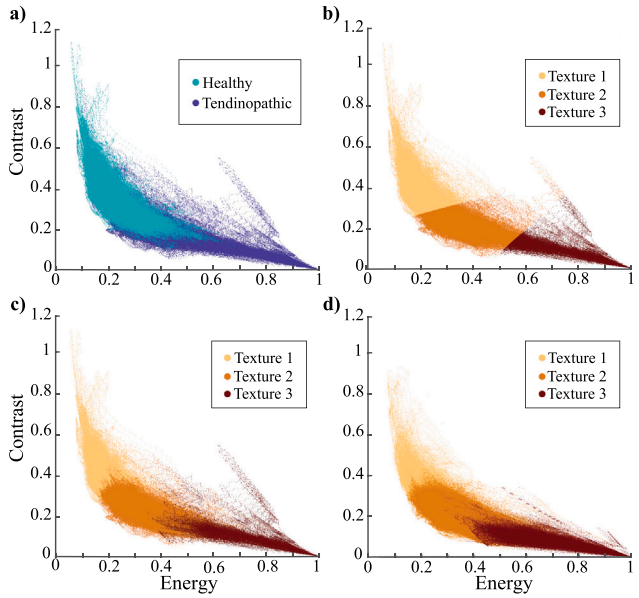


**Fig. 3.** Scatter plot of GLCM features for pixels within the training and test sets. (a) Different distributions between values for pixels within healthy and tendinopathic tendons for the training set. (b) Initial K-means class assignments for the training set. (c-d) HMRF-EM cluster assignments for the training and test sets.

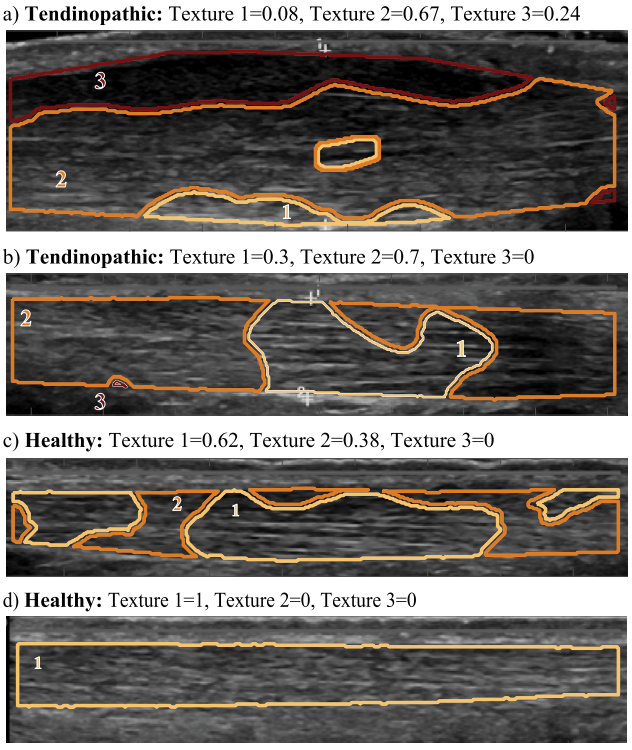


**Fig. 4.** Segmentation results for four images within the training set. Images (a) and (b) are of tendinopathic tendons, and (c) and (d) are of healthy tendons. Regions labelled as texture class one are outlined in yellow and constitute areas with high contrast and low energy. Areas assigned to texture class two are outlined in orange and refer to areas with moderate contrast and energy. Areas outlined in red are assigned to texture class three and constitute areas with low contrast and high energy.

component is sensitive to the accuracy of the initialisation, hyperparameters were divided into two sets: those related to the initial texton/GLCM feature generation and K-means clustering, and those related to the HMRF-EM algorithm, which were tuned sequentially (the decision tree was applied to only the K-means output for the former). The optimal hyperparameters are defined to be those that provide clear class proportion thresholds which can be used for classification. Thus we choose hyperparameters which produce a decision tree fit with the highest classification accuracy and minimum number of splits. The neighbourhood $\mathcal{N}(i)$ size was tested over the range $[1, 50]$ and had no impact on the classification accuracy. Hence, a neighbourhood size of $30 \times 30$ was chosen for qualitative purpose: this size provided sufficient smoothing of the segmentation to facilitate interpretation by the radiologist. Further details regarding the hyperparameter tuning and final values are provided in Appendix A.2.

## 4. Results

### 4.1. Feature generation and training

The GLCM feature generation for the training set took 460 s (> 10 s per image) whilst initial clustering of the training set, using the *kmeans* function in *MATLAB*, took 561 s. The computational time taken for subsequent analysis of the images by the HMRF-EM algorithm was fast, taking < 10 s per image.

### 4.2. Grey-level co-occurrence matrix features

A scatter plot of the GLCM features for pixels belonging to images within the training and test sets are given in Fig. 3. In Fig. 3a it is evident there is a distinct difference in the distribution of local GLCM features between tendinopathic (blue points) and healthy (green points) tendons, with pixels belonging to tendinopathic tendon images tending to have a lower contrast and higher energy. In Figs. 3b and 3c a plot of the same training data shows the texture label assigned for each pixel after K-means and subsequent HMRF-EM clustering. The data is clustered into three texture classes with yellow corresponding to high contrast and low energy values, orange corresponding to moderate contrast and energy values, and red corresponding to low contrast and high energy values. Comparing Figs. 3b and 3c, the reassignment of pixel class labels due to subsequent HMRF-EM processing occurred only for pixels that have GLCM values lying at the border of the clusters or that are outliers (low energy with moderate contrast). The overall distribution of GLCM values for the test set (Fig. 3d), as well as the distribution of pixel GLCM values within each class, is similar to that of the training set (Fig. 3c). Contrasting Figs. 3a and 3c, the pixels corresponding to tendinopathic images appear to be predominantly assigned to the second and third texture classes.

### 4.3. GLCM-based HMRF-EM segmentation

Fig. 4 displays the GLCM-based HMRF-EM segmentation for four examples from the training set. In Fig. 5 the corresponding GLCM features for each pixel within the four images are given. Figs. 4a and 4b give the segmentations for two tendinopathic tendons. In Fig. 4a a tendon segment has been assigned to the third texture class, with associated low contrast and high energy values occurring for the encompassed pixels (Fig. 5a). The region is visibly tendinopathic, with no discernible collagen fibres and very low echogenicity. In Fig. 4b the first and second texture classes are predominant. In contrast to Fig. 5a, more extreme high energy and low contrast values are not observed in Fig. 5b. However, the second texture – with moderate contrast and energy – accounts for the largest proportion of this image. Figs. 4c and 4d are the image segmentations for two healthy tendons. In Fig. 4c pixels were assigned to the second texture class but in lower proportions than seen in Fig. 4b. Fig. 4d displays a healthy tendon with a visibly consistent texture pattern throughout and, accordingly, only the first texture class is assigned. In Fig. 4b–d the first texture class

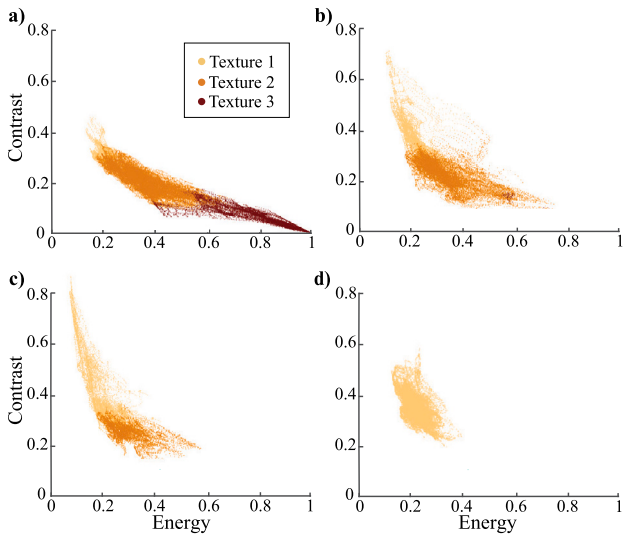


**Fig. 5.** GLCM features for each pixel for two tendinopathic (a,b) and two healthy (c,d) images within the training set.

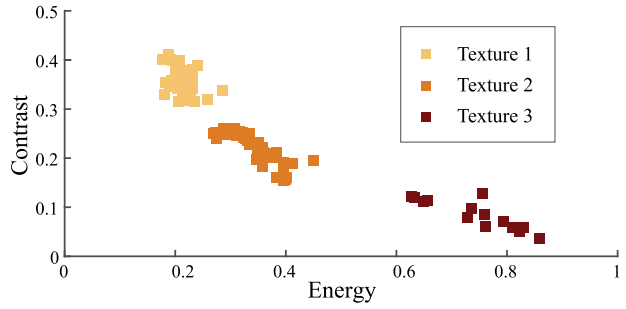


**Fig. 6.** Mean energy and contrast for images after GLCM-based HMRF-EM segmentation.

has been assigned to regions with clearly visible collagen fibres and darker hypoechoic spaces between. In Fig. 4a–c the second texture class has been assigned to regions with less clarity between the fibres and interfibre spacing, but without the level of hypoechogenicity of regions assigned to the third class. This qualitative pattern was consistent across the training set (Appendix B Fig. B.1).Across the training set, hypoechoic regions assigned to the third texture class occurred most commonly near the transverse posterior border.

In Table 1 the mean and standard deviation of texture proportions within the training set are given after GLCM-based clustering. Values after initialisation through K-means clustering, and following further HMRF-EM processing are displayed. Consistent with the visual segmentations shown in Appendix B Fig. B.1, there is a clear difference in texture class proportions between tendinopathic and healthy tendons. The first texture class appears predominantly in healthy tendons whilst the second class is more predominant in the tendinopathic tendons. The third class is only present in the tendinopathic tendons. As expected, following HMRF-EM implementation, there are only minimal changes to the mean texture class proportions.

In Table 1 it is also evident that there are only small differences in the texture class proportions between the K-means- and HMRF-EM-based segmentations. This is consistent with observable differences in the segmentations between the HMRF-EM-based (Appendix B Fig. B.1) and K-means-based (Appendix B Fig. B.2) segmentations of the training set, in which the HMRF-EM extension predominantly serves to reduce noise and provide a more distinct segmentation.

In Fig. 6 a plot of the means $\mu_k$ for the texture class labels $k \in \{1,2,3\}$, after implementation of the HMRF-EM algorithm, is displayed.

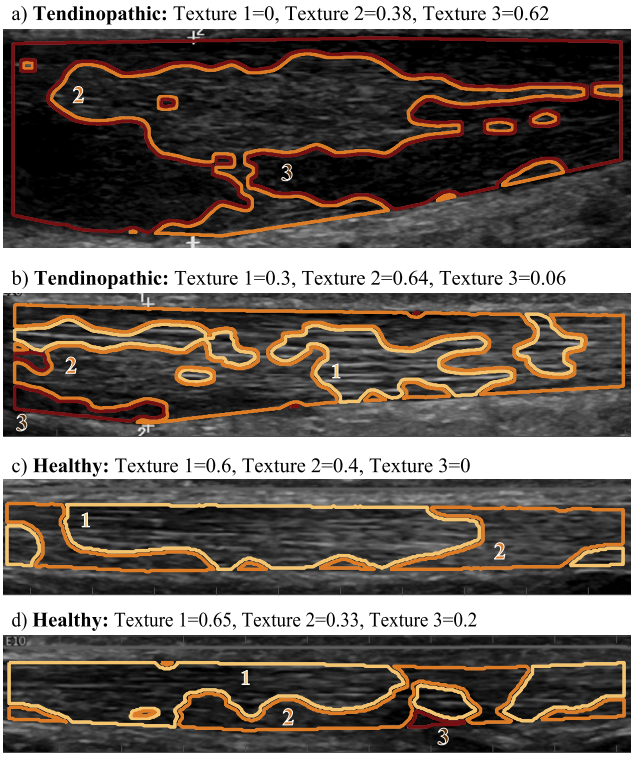


**Fig. 7.** Segmentation results for four images within the test set. Images (a) and (b) are of tendinopathic tendons, and (c) and (d) are of healthy tendons. Regions labelled as texture class one are outlined in yellow and constitute areas with high contrast and low energy. Areas assigned to texture class two are outlined in orange and refer to areas with moderate contrast and energy. Areas outlined in red are assigned to texture class three and constitute areas with low contrast and high energy.

**Table 1**
Mean and standard deviation of segmentation proportions for training set after K-means and HMRF-EM clustering of GLCM features.

| Image set | Texture class proportion: Mean (SD) | | |
|---|---|---|---|
| | 1 | 2 | 3 |
| K-means | | | |
| Tendinopathic | 0.15 (0.10) | 0.67 (0.16) | 0.18 (0.17) |
| Healthy | 0.69 (0.14) | 0.31 (0.14) | 0.00 (0.00) |
| HMRF-EM | | | |
| Tendinopathic | 0.12 (0.10) | 0.72 (0.19) | 0.16 (0.18) |
| Healthy | 0.73 (0.18) | 0.27 (0.18) | 0.00 (0.00) |

The mean is calculated for each individual image in the dataset. Across the images, the means for each label can be seen to form distinct clusters. Hence, the texture class labels $k \in \{1,2,3\}$ are comparable across images, referring to high contrast and low energy, moderate contrast and energy, and low contrast and high energy, respectively.

In Fig. 7 the GLCM-based HMRF-EM segmentation for two tendinopathic and two healthy images from the test set are given. Fig. 8 depicts the corresponding scatter plots of GLCM features of pixels within each image. In Figs. 7a and 7b two tendinopathic images are displayed. In Figs. 7c and 7d two healthy images are displayed. There is good qualitative agreement with the segmentations given for the training examples in Figs. 4 and 5, in which the third texture is assigned to hypoechoic regions with lower contrast and minimal visible fibres, and the first texture is assigned to regions with higher contrast and more clearly visible fibres. Similar to the segmentation results for the training set (Fig. 4a and Appendix B Fig. B.1), across the test set, hypoechoic regions assigned to the third texture class occurred most commonly near the transverse border of the tendon and there is an overall loss

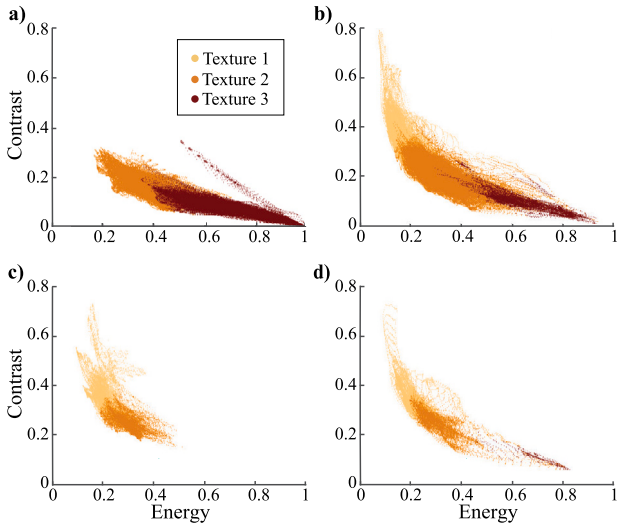


**Fig. 8.** GLCM features for each pixel for two tendinopathic (a,b) and two healthy (c,d) images within the test set.

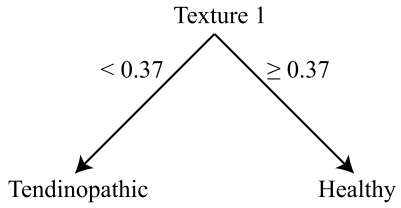


**Fig. 9.** Decision tree for classifying tendon images as tendinopathic or healthy based on GLCM-based HMRF-EM segmentation proportions.

of collagen fibre visibility demonstrated by a greater proportion of the second texture class also being present (Appendix B Fig. B.3).

Overall the HMRF-EM algorithm segmented and quantified the proportion of clearly hypoechoic regions that represent obvious tendinopathic damage, assigning these regions to the third texture class. The algorithm also segmented and quantified the proportion of regions which exhibited some loss of clearly defined collagen fibres but were not clearly hypoechoic, as can be seen by comparing texture class two to one in Fig. 5b. This provided a delineation of regions that were difficult to identify a-priori by visual inspection, but nonetheless could be visualised to show a loss of the clearly defined collagen fibres after the segmentation was applied to highlight these regions. The segmentations for all tendinopathic tendons in the training and test sets (Appendix B Figs. B.1 and B.3) show a high proportion of the second texture class and only a small proportion of the first, demonstrating an overall lost off collagen fibre visibility.

### 4.4. Classification of healthy and tendinopathic tendons

The decision tree produced from the training set is displayed in Fig. 9. Images are classified as tendinopathic if the proportion of the first texture, calculated for the masked 4 cm region, is less then 0.37. Otherwise, the tendon is considered healthy. Using this simple binary cut-off, the decision tree was able to classify all training images accurately. Using the decision tree on the test set, 22 of the 24 images were accurately classified. All tendinopathic images were successfully classified, whilst two healthy images were misclassified as tendinopathic due to low proportions of the first texture. Fig. 10a displays the tendinopathic image from the test set which exhibits the highest proportion of the first texture (0.33) and was correctly classified. This is contrasted with the two misclassified healthy images from the test

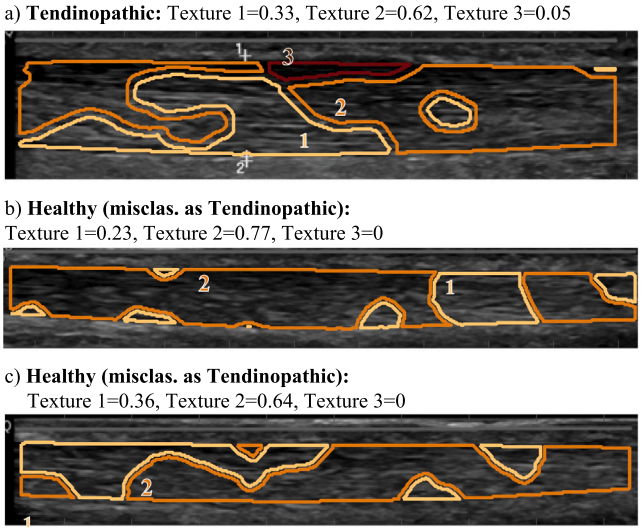


**Fig. 10.** Segmentation examples, and associated GLCM features for each pixel, for three images within the test set. (a) Successfully classified tendinopathic image with the largest proportion of texture one of all tendinopathic test images. (b, c) Misclassified healthy images from the test set.

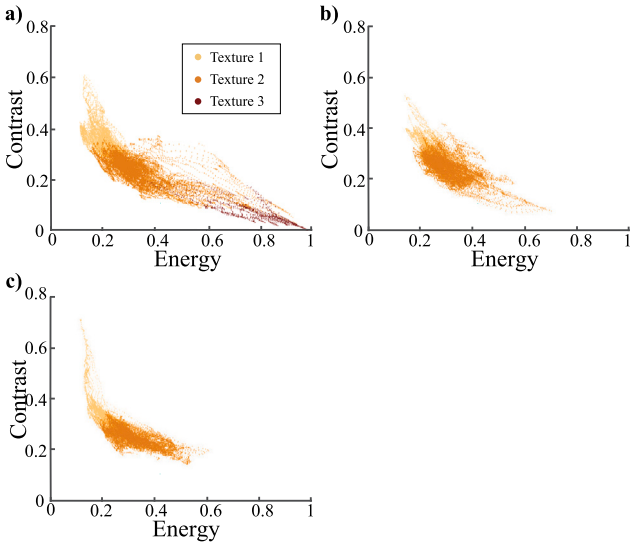


**Fig. 11.** GLCM features for each pixel for three images within the test set. (a) Successfully classified tendinopathic image with the largest proportion of texture one of all tendinopathic test images. (b, c) Misclassified healthy images from the test set.

set in Figs. 10b and 10c. Comparing, the associated scatter plots of GLCM features (Fig. 11), despite the misclassification, there is still a difference in the GLCM distributions between the tendinopathic and healthy tendons. The tendinopathic tendon has a greater spread of contrast and energy values amongst its' pixels than the misclassified healthy images. The tendinopathic tendon also exhibits a small region classified as the third texture whilst the other two images do not have a region with such low contrast and high energy values.

The classifications results were the same for the K-means-based segmentation of the images, except the texture proportion cut off was 0.39.

## 5. Discussion

Local texture features based on the grey-level co-occurrence matrix (GLCM) were analysed as a means for providing a segmentation

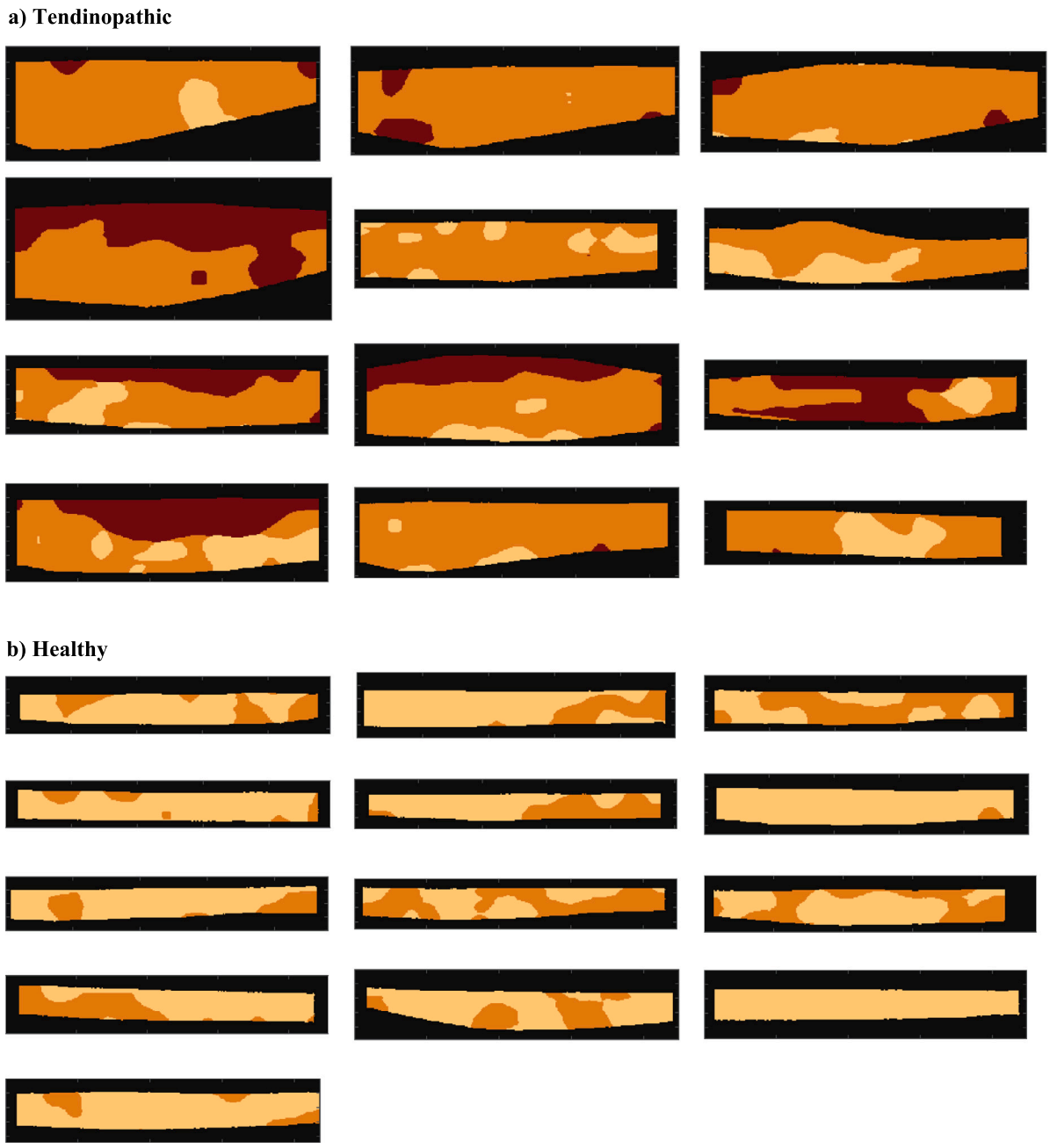


**Fig. B.1.** GLCM-based HMRF-EM segmentation results for the training set. Regions labelled as texture class one are coloured yellow and constitutes areas with high contrast and low energy. Areas assigned to texture class two are coloured in orange and refer to areas with moderate contrast and energy. Areas coloured in red have been assigned to texture class three and constitute areas with low contrast and high energy.

**Table A.2**
Final model hyperparameters and tuning ranges explored.

| Hyperparameter | Range | Value |
|---|---|---|
| $N_{CM}$: GLCM grey level bins | [4, 10] | 8 |
| $W_{CM}$: GLCM window size (px) | [10, 30]x[10, 30] | 10 × 20 |
| $K$: Texture classes | [2, 4] | 3 |
| $\beta$: Gibbs parameter | [0.5, 2] | 1 |
| $\mathcal{N}(i)$: Neighbourhood size (px) | [1, 30]x[1, 30] | 30 × 30 |

which could differentiate between tendinopathic and healthy tendons on imaging. Only two features were required to capture difference between tendinopathic and healthy tendon images: the contrast and energy, a measure of the uniformity. A continuum of changes in the contrast and energy appear to occur with tendinopathy. Collectively, the GLCM-based HMRF-EM segmentation results suggest that a reduction in the contrast, and an increase in the energy, of a region is indicative of potential tendinopathic changes. The results also suggest that regions with the lowest contrast ($< 0.2$) and highest energy ($> 0.6$) may represent a higher level of damage. Visually these findings correlated with a loss of fibre clarity, extending to a complete loss of fibre visibility. This could suggest a loss of collagen fibres in that region or an accumulation of denser substances such as fluid (for example due to proteoglycan deposition preventing fluid expulsion) which obscure the fibres on US. There was a wide variety of topological presentations, however, there was a general trend that these more extreme low contrast and high energy regions occurred closer to the transverse tendon boundaries (Appendix B Figs. B.1 and B.3).

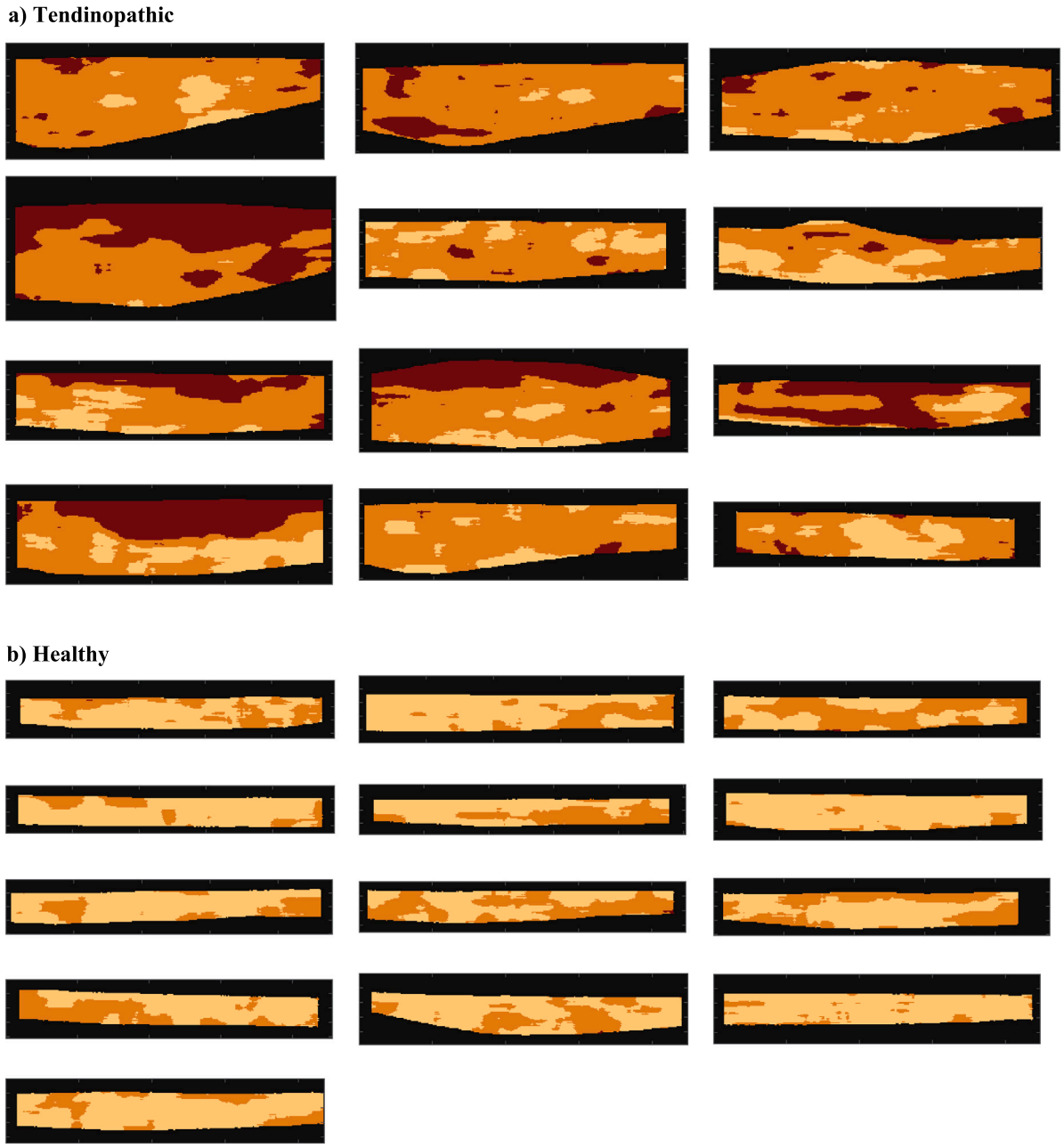


**Fig. B.2.** GLCM-based K-means segmentation results for the training set. Regions labelled as texture class one are coloured yellow and constitutes areas with high contrast and low energy. Areas assigned to texture class two are coloured in orange and refer to areas with moderate contrast and energy. Areas coloured in red have been assigned to texture class three and constitute areas with low contrast and high energy. The two healthy tendons which were misclassified during testing are labelled.

Whilst the second texture in the GLCM-based HMRF-EM segmentation, corresponding to moderate contrast and energy values, was predominantly present in tendinopathic tendons, it was also present in lower proportions amongst healthy tendon images. These regions may still correspond to early tendinopathic changes, as several longitudinal studies have demonstrated that tendinopathic changes on imaging often precede pain and suggest a higher risk for later symptom development [44–46]. This could also be a result of using K-means for initial clustering. In the setting of overlapping clusters, K-means allocates points in a way that roughly equalises the spatial variance of the clusters. This can be problematic if the true optimal clustering solution is one in which the clusters have a large difference in the number of points. Facilitating larger differences in cluster sizes may be necessary to provide an even better distinction between tendinopathic and healthy tendons.

The K-means and HMRF-EM approaches yielded similar segmentations and the classification accuracy between healthy and tendinopathic tendons was not impacted by including the additional HMRF-EM extension on top of the initial K-means segmentation. However, the HMRF-EM extension required minimal extra time computational run time and provided a smoother segmentation of the images that may be useful for the application of this tool as an adjunct to the radiologist. Because the distribution of GLCM features showed clear differences between tendinopathic and healthy tendons, threshold based segmentation techniques [47], using the GLCM features, provide an alternative approach to the algorithm presented here. However, these techniques can be sensitive to noise and even more computationally intensive than K-means [48], and hence consideration needs to be taken in terms of the training set size and quality.

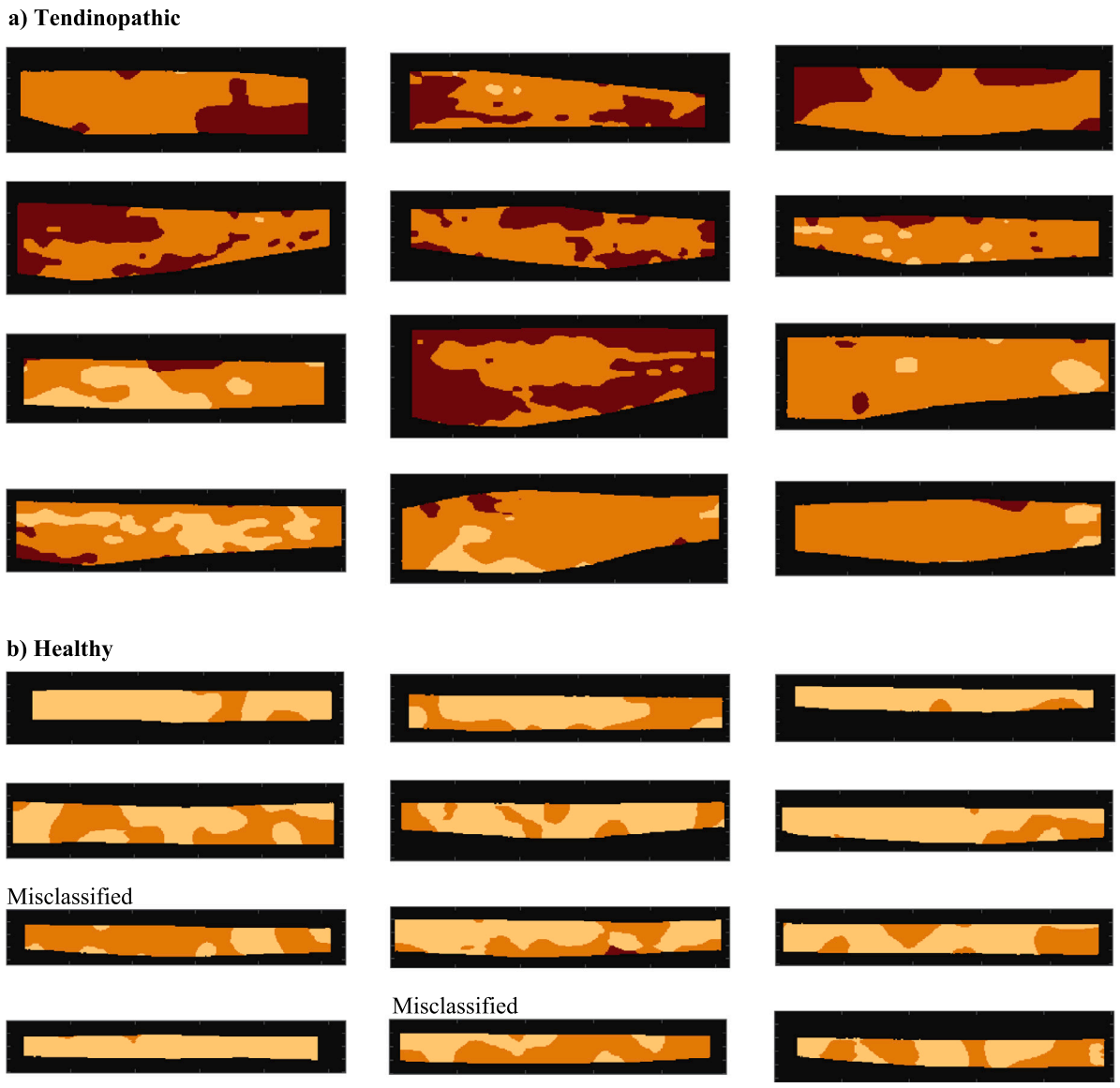


**Fig. B.3.** GLCM-based HMRF-EM segmentation results for the test set. Regions labelled as texture class one are coloured yellow and constitutes areas with high contrast and low energy. Areas assigned to texture class two are coloured in orange and refer to areas with moderate contrast and energy. Areas coloured in red have been assigned to texture class three and constitute areas with low contrast and high energy. The two healthy tendons which were misclassified during testing are labelled.

The purpose of this work was to identify a segmentation technique that could be used as a quantitative tool to measure disease severity. Since a prognostic labelled dataset was not available, we correlated the segmentation results with disease state to first provide a proof of concept. A simple decision tree approach was chosen to validate whether the segmentation could be used to accurately classify images, as we desired a segmentation that would provide readily interpretable quantifiable predictors for the clinician. A simple cut-off value for one texture allowed us to classify 47 of 49 images correctly, with 100% accuracy for the tendinopathic images. Evaluation of GLCM features for the misclassified healthy images suggested there was still a difference between the distribution of GLCM features of misclassified healthy tendons and the tendinopathic set. A more refined decision tree may improve the classification accuracy, such as if we had used K-fold cross-validation[1] in training the model and allowed for a higher number of splits within the tree. However, the focus of this work was in providing a novel segmentation approach, and using the classification step to simply validate the segmentation approach. For this reason, and due to the computational runtime associated with generating the initial clustering of GLCM features, we did not further optimise our approach to achieve the highest classification accuracy. Another useful extension to our approach, and one that may improve accuracy, would be to consider extracting additional metrics from the segmentation results – other than texture class proportions – that could be utilised in the decision tree. In this case, it would be desirable to have a larger sample size to prevent over fitting.

The GLCM-based HMRF-EM segmentation algorithm presented here has several advantages over Ultrasound Tissue Characterisation (UTC) [7]. In contrast to UTC, which captures the continuity of fibres in the longitudinal direction, the GLCM-based HMRF-EM segmentation algorithm also captures information about the relationship of collagen fibres in the transverse direction. Our approach can also be used on the output of standard B-mode US devices, whilst UTC requires expensive specialised equipment. Furthermore, the HMRF-EM segmentation algorithm requires only one masked longitudinal image, in contrast to UTC

[1] The dataset if split into 'K' groups, with one used as the test set and the others used as the training set. The model is trained and tested, with the process repeated using each of the unique groups as the test set.

which requires a hundredfold greater number of masked transverse images to span the length of the tendon. .

## 6. Conclusions and limitations

We have developed a segmentation approach which strongly correlates with disease state in tendons. We utilised a Hidden Gaussian Markov Random Field model to segment each image based on a set of local grey-level co-occurrence features. The model was fit using a modified version of the HMRF-EM algorithm developed by Zhang et al. [37]. The segmentation output from this approach provides a tool which can be used as an adjunct to the radiologist. It provides both quantifiable and interpretable information regarding the texture profiles within the image, as well as a segmentation of the image which the radiologist can use to ensure they have detected all segments of interest. In agreement with the literature [1,49], our segmentation results and analysis of texture features suggest that tendinopathic changes occur as continuum and early changes structural changes may precede pain. The results also gave insight into the topology of changes which occur in tendinopathy, demonstrating that more severe tendinopathic changes are most likely to occur nearest to the transverse tendon boundary.

A key weakness of this study in the limited size of the dataset and the lack of additional clinical features to correlate with the segmentation results, such as pain levels. This would allow for more accurate hyperparameter tuning and allow us to confirm whether the texture profiles exhibited correlate with disease severity, as well as disease state. With an increase in dataset size, modifications to the initial K-means clustering may be required. We also note that the hyperparameters were tuned according to the resolution of the dataset available. The resolution between images in our dataset was similar and hence the same hyperparameters values were used throughout. To ensure generalisability, future work should consider identifying a resolution dependent scaling for the hyperparameters, such as GLCM window size. Finally, there is a trade off in developing an approach that only requires one longitudinal image: unlike approaches such as UTC which span the whole tendon, the output from our approach may not provide as comprehensive coverage of tendon changes. Future work should seek to compare this approach with UTC. Despite the above limitations, the results from this study are promising and support the collection of a larger prognostic dataset. The segmentation approach we have developed here has the potential to provide a readily accessible tool for the quantitative analysis of tendon images in both the research and clinical settings.

## CRediT authorship contribution statement

**Isabelle Scott:** Conceptualization, Formal analysis, Funding acquisition, Methodology, Validation, Visualization, Writing – original draft, Writing – review & editing. **David Connell:** Data curation, Funding acquisition. **Derek Moulton:** Supervision, Writing – review & editing. **Sarah Waters:** Supervision, Writing – review & editing. **Ana Namburete:** Methodology, Supervision, Writing – review & editing. **Anurag Arnab:** Methodology, Supervision, Writing – original draft, Writing – review & editing. **Peter Malliaras:** Data curation, Funding acquisition, Supervision, Writing – review & editing.

## Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgements

We are grateful to the Australasian Musculoskeletal Imaging Group for funding this work.

## Appendix A

### A.1. HMRF-EM algorithm pseudocode

**Algorithm 1** GLCM Feature Generation and K-Means Clustering

**Inputs:**
$N_{img}$ greyscale images.
$N_{img}$ corresponding image masks.
Number of grey levels for GLCM generation $N_{CM}$.
Window size for GLCM generation
$W_{CM1} = w_{CM2} \times w_{CM}$.
Number of texture class labels $K$.

**for each** image **do**
  **for each** pixel **do**
    Using *graycomatrix*: generate GLCM for surrounding window $W_{CM}$.
    Using *graycoprops* on GLCM: Calculate (energy, contrast)$^T \leftarrow \boldsymbol{y}_i$.

**for all** images and pixels **do**
  Cluster $\boldsymbol{y}_i$ into K clusters.
  Use cluster label $k \in \{1, ..., K\}$ to provide initial segmentation $k \leftarrow x_i^{t=0}$.

**Algorithm 2** HMRF-EM Segmentation

**Inputs:**
Texton histogram for each pixel $\boldsymbol{y}_i$.
Initial pixel class labels $x_i^{t=0}$.
Number of clusters for segmentation $K$.
Max iterations $T$.
Gibbs parameter $\beta$.

**for** texture class labels k=1:K **do**
  Compute initial mean and variance:
  $\boldsymbol{\mu}_k^0 = \sum_{i:\, x_i=k} \boldsymbol{y}_i \Big/ \sum_{i:\, x_i=k} 1.$
  $(\boldsymbol{\Sigma}_k^0)^2 = \sum_{i:\, x_i=k} (\boldsymbol{y}_i - \boldsymbol{\mu}_k)^2 \Big/ \sum_{i:\, x_i=k} 1.$

**repeat**
  *Neighbour Assignment*
  **for** pixels i=1:n **do**
    **for** class labels k=1:K **do**
      Calculate P(class label assignment):
      $P_G(x_i = k | y_i, \boldsymbol{x}_{\mathcal{N}(i)}^{t-1}, \boldsymbol{\varphi}^{t-1})$ .
      Assign pixel label:
      $\overline{x}_i^t \leftarrow \underset{x_i}{\text{argmax}}\ P_G(x_i | y_i, \boldsymbol{x}_{\mathcal{N}(i)}^{t-1}, \boldsymbol{\varphi}^{t-1})$.
  *Expectation*
  **for** pixels i=1:n **do**
    **for** class labels k=1:K **do**
      Calculate P(pixel label | neighbour labels):
      $P(x_i = k | y_i, \boldsymbol{x}_{\mathcal{N}(i)}, \boldsymbol{\varphi}^{t-1}) =$
      $f(y_i | x_i, \boldsymbol{\varphi}^{t-1}) P(x_i | \boldsymbol{x}_{\mathcal{N}(i)}^t, \beta) \Big/ f(y_i | \boldsymbol{\varphi}^{t-1})$.
  *Maximisation*
  **for** class labels k=1:K **do**
    Compute mean and variance: $(\boldsymbol{\mu}_k^t, \boldsymbol{\Sigma}_k^t) =$
    $\underset{(\mu_k, \Sigma_k)}{\text{argmax}} \sum_{i \in S} \sum_{x_i \in \mathcal{X}} P_{\overline{\boldsymbol{x}}^t}(x_i | y_i, \boldsymbol{\varphi}^t) \log f(y_i | x_i, \boldsymbol{\mu}, \boldsymbol{\Sigma})$.

**until** $\boldsymbol{\varphi}$ converges or iterations $t > T$.

### A.2. Hyperparameter tuning

A GLCM offset of $(-1, 0)$ is used to capture increases in fibre spacing consistent with tendinopathy [3]; this offset is perpendicular to the fibre axis in a healthy longitudinal tendon image. To tune the remaining hyperparameters, a grid-based search is used with parameter ranges based on previous values utilised in the literature, as well as the scale of physical units within the image. The optimal hyperparameters are defined to be those that provide clear class proportion thresholds that can be used for classification. Thus we chose hyperparameters which produce a decision tree fit with the highest classification accuracy and minimum number of splits. The model hyperparameter ranges explored in the grid-search, and final values chosen, are given in Table A.2.

The GLCM parameter ranges are based off previous work [50] and partially informed by knowledge of the scale of physical units within the image, prescribed to capture several fibres within a window. The parameter range for the Gibbs parameter is informed by previous applications to image segmentation [51].

## Appendix B

See Figs. B.1, B.2 and B.3

## References


[1] S. Docking, C.C. Ooi, D. Connell, Tendinopathy: is imaging telling us the entire story? J. Orthop. Sports Phys. Therapy 45 (11) (2015) 842–852.

[2] W. Grassi, E. Filippucci, A. Farina, C. Cervini, Sonographic imaging of tendons, Arthritis Rheum 43 (5) (2000) 969–976.

[3] R. Hodgson, P. O'Connor, A. Grainger, Tendon and ligament imaging, Br. J. Radiol. 85 (1016) (2012) 1157–1172.

[4] H. Van Schie, R.J. de Vos, S. de Jonge, E. Bakker, M. Heijboer, J. Verhaar, J. Tol, H. Weinans, Ultrasonographic tissue characterisation of human Achilles tendons: quantification of tendon structure through a novel non-invasive approach, Br. J. Sports Med. 44 (16) (2010) 1153–1159.

[5] S. Docking, S. Rosengarten, J. Cook, Achilles tendon structure improves on UTC imaging over a 5-month pre-season in elite a ustralian football players, Scand. J. Med. Sci. Sports 26 (5) (2016) 557–563.

[6] E. Wezenbeek, N. Mahieu, T. Willems, D. Van Tiggelen, M. De Muynck, D. De Clercq, E. Witvrouw, What does normal tendon structure look like? New insights into tissue characterization in the achilles tendon, Scand. J. Med. Sci. Sports 27 (7) (2017) 746–753.

[7] H. van Schie, E. Bakker, A. Jonker, P. van Weeren, Computerized ultrasonographic tissue characterization of equine superficial digital flexor tendons by means of stability quantification of echo patterns in contiguous transverse ultrasonographic images, Am. J. Vet. Res. 64 (3) (2003) 366–375.

[8] M. van Ark, E. Rio, J. Cook, I. van den Akker-Scheek, J. Gaida, J. Zwerver, S. Docking, Clinical improvements are not explained by changes in tendon structure on UTC following an exercise program for patellar tendinopathy, Am. J. Phys. Med. 97 (10) (2018) 708–714.

[9] S. Docking, E. Rio, J. Cook, D. Carey, L. Fortington, Quantification of achilles and patellar tendon structure on imaging does not enhance ability to predict self-reported symptoms beyond grey-scale ultrasound and previous history, J. Sci. Med. Sport 22 (2) (2019) 145–150.

[10] L. Brattain, B. Telfer, M. Dhyani, J. Grajo, A. Samir, Machine learning for medical ultrasound: status, methods, and future opportunities, Abdom. Radiol. 43 (4) (2018) 786–799.

[11] H. Van Schie, S. Docking, J. Daffy, S. Praet, S. Rosengarten, J. Cook, Ultrasound tissue characterisation, an innovative technique for injury-prevention and monitoring of tendinopathy, Br. J. Sports Med. 47 (9) (2013) e2.

[12] Q. Han, H. Wang, M. Hou, T. Weng, Y. Pei, Z. Li, G. Chen, Y. Tian, Z. Qiu, HWA-SegNet: Multi-channel skin lesion image segmentation network with hierarchical analysis and weight adjustment, Comput. Biol. Med. 152 (2023) 106343.

[13] F. Uslu, A. Bharath, TMS-net: A segmentation network coupled with a run-time quality control method for robust cardiac image segmentation, Comput. Biol. Med. 152 (2023) 106422.

[14] M. Jahanifar, N. Tajeddin, M. Hasani, B. Shekarchi, K. Azema, Automatic recognition of the supraspinatus tendinopathy from ultrasound images using convolutional neural networks, 2020, arXiv preprint arXiv:2011.11777.

[15] T. Alzyadat, S. Praet, G. Chetty, R. Goecke, D. Hughes, D. Kumar, M. Welvaert, N. Vlahovich, G. Waddington, Automatic segmentation of achilles tendon tissues using deep convolutional neural network, in: Machine Learning in Medical Imaging: 11th International Workshop, MLMI 2020, Held in Conjunction with MICCAI 2020, Lima, Peru, October 4, 2020, Proceedings 11, Springer, 2020, pp. 444–454.

[16] M. Wyburd, L. Hesse, M. Aliasi, M. Jenkinson, A. Papageorghiou, M. Haak, A. Namburete, Assessment of regional cortical development through fissure based gestational age estimation in 3D fetal ultrasound, in: Uncertainty for Safe Utilization of Machine Learning in Medical Imaging, and Perinatal Imaging, Placental and Preterm Image Analysis: 3rd International Workshop, UNSURE 2021, and 6th International Workshop, PIPPI 2021, Held in Conjunction with MICCAI 2021, Strasbourg, France, October 1, 2021, Proceedings 3, Springer, 2021, pp. 242–252.

[17] N. Arun, N. Gaw, P. Singh, K. Chang, M. Aggarwal, B. Chen, K. Hoebel, S. Gupta, J. Patel, M. Gidwani, J. Adebayo, M. Li, J. Kalpathy-Cramer, Assessing the trustworthiness of saliency maps for localizing abnormalities in medical imaging, Radiol. Artif. Intell. 3 (6) (2021) e200267.

[18] Y. Xu, J.Y. Zhu, I. Eric, C. Chang, M. Lai, Z. Tu, Weakly supervised histopathology cancer image segmentation and classification, Med. Image Anal. 18 (3) (2014) 591–604.

[19] A. Kolesnikov, C. Lampert, Seed, expand and constrain: Three principles for weakly-supervised image segmentation, in: European Conference on Computer Vision, Springer, 2016, pp. 695–711.

[20] N. Dinsdale, E. Bluemke, V. Sundaresan, M. Jenkinson, S. Smith, A. Namburete, Challenges for machine learning in clinical translation of big data imaging studies, Neuron (2022).

[21] J. Cho, K. Lee, E. Shin, G. Choy, S. Do, How much data is needed to train a medical image deep learning system to achieve necessary high accuracy? 2015, arXiv preprint arXiv:1511.06348.

[22] R. Haralick, K. Shanmugam, I. Dinstein, Textural features for image classification, IEEE Trans. Syst. Man Cybern. (6) (1973) 610–621.

[23] F. Schneebeli, G. Vincenzo, C. Cescon, M. Barbero, Test-retest reliability of echo intensity parameters in healthy achilles tendons using a semi-automatic tracing procedure, Skeletal Radiol. 46 (2017) 1553–1558.

[24] J. Benrabha, F. Meziane, Automatic ROI detection and classification of the achilles tendon ultrasound images, in: Proceedings of the 1st International Conference on Internet of Things and Machine Learning, 2017, pp. 1–7.

[25] M.-J. Nadeau, A. Desrochers, M. Lamontagne, C. Larivière, D.H. Gagnon, Quantitative ultrasound imaging of achilles tendon integrity in symptomatic and asymptomatic individuals: reliability and minimal detectable change, J. Foot Ankle Res. 9 (1) (2016) 1–17.

[26] M. Lalumiere, C. Larivière, P. Nadeau, M. Lamontagne, F. Desmeules, D. Gagnon, Proposing a minimal data set of musculoskeletal ultrasound imaging biomarkers to inform clinical practice: an analysis founded on the achilles tendon, Ultrasound Med. Biol. 46 (9) (2020) 2222–2235.

[27] W. Hu, H. Li, L. Fu, C. Wang, S. Gou, Characterization of collagen fibers by means of texture analysis of second harmonic generation images using orientation-dependent gray level co-occurrence matrix method, J. Biomed. Opt. 17 (2) (2012) 026007.

[28] N.D. Kim, L. Booth, V. Amin, J. Lim, S. Udpa, Ultrasonic image processing for tendon injury evaluation, in: Proceedings of the IEEE-SP International Symposium on Time-Frequency and Time-Scale Analysis (Cat. No. 98TH8380), IEEE, 1998, pp. 241–244.

[29] R. Gupta, I. Elamvazuthi, S. Dass, I. Faye, P. Vasant, J. George, F. Izza, Curvelet based automatic segmentation of supraspinatus tendon from ultrasound image: a focused assistive diagnostic method, Biomed. Eng. Online 13 (1) (2014) 1–18.

[30] T. Ahonen, M. Pietikäinen, Image description using joint distribution of filter bank responses, Pattern Recognit. Lett. 30 (4) (2009) 368–376.

[31] S. Prasad, J. Domke, Gabor filter visualization, J. Atmos. Sci. 13 (2005) 2005.

[32] M. Unser, T. Blu, Wavelet theory demystified, IEEE Trans. Signal Process. 51 (2) (2003) 470–483.

[33] H. Permuter, J. Francos, I. Jermyn, A study of Gaussian mixture models of color and texture features for image classification and segmentation, Pattern Recognit. 39 (4) (2006) 695–706.

[34] D. Pham, C. Xu, J. Prince, Current methods in medical image segmentation, Annu. Rev. Biomed. Eng. 2 (1) (2000) 315–337.

[35] D. Comaniciu, P. Meer, Mean shift: A robust approach toward feature space analysis, IEEE Trans. Pattern Anal. Mach. Intell. 24 (5) (2002) 603–619, http://dx.doi.org/10.1109/34.1000236.

[36] R. Azmi, N. Norozi, A new markov random field segmentation method for breast lesion segmentation in MR images, J. Med. Signals Sensors 1 (3) (2011) 156.

[37] Y. Zhang, J. Brady, S. Smith, Hidden Markov random field model for segmentation of brain MR image, in: Medical Imaging 2000: Image Processing, Vol. 3979, International Society for Optics and Photonics, 2000, pp. 1126–1137.

[38] K. Khan, B. Forster, J. Robinson, Y. Cheong, L. Louis, L. Maclean, J. Taunton, Are ultrasound and magnetic resonance imaging of value in assessment of achilles tendon disorders? A two year prospective study, Br. J. Sports Med. 37 (2) (2003) 149–153.

[39] J. Collinger, B. Fullerton, B. Impink, A. Koontz, M. Boninger, Validation of greyscale-based quantitative ultrasound in manual wheelchair users: relationship to established clinical measures of shoulder pathology, Am. J. Phys. Med. Rehabil. Assoc. Acad. Physiatr. 89 (5) (2010) 390.

[40] G. Winkler, Image Analysis, Random Fields and Markov Chain Monte Carlo Methods: A Mathematical Introduction, Vol. 27, Springer Science & Business Media, 2003.

[41] A. Dempster, N. Laird, D. Rubin, Maximum likelihood from incomplete data via the EM algorithm, J. R. Stat. Soc. Ser. B Stat. Methodol. 39 (1) (1977) 1–22.
[42] G. Celeux, F. Forbes, N. Peyrard, EM procedures using mean field-like approximations for Markov model-based image segmentation, Pattern Recognit. 36 (1) (2003) 131–144.
[43] L. Breiman, J. Friedman, R. Olshen, C. Stone, Classification and Regression Trees, CRC Press, New York, 1984.
[44] U. Fredberg, L. Bolvig, Significance of ultrasonographically detected asymptomatic tendinosis in the patellar and achilles tendons of elite soccer players: a longitudinal study, Am. J. Sports Med. 30 (4) (2002) 488–491.
[45] K. Khan, J. Cook, Z. Kiss, P. Visentini, M. Fehrmann, P. Harcourt, B. Tress, J. Wark, Patellar tendon ultrasonography and jumper's knee in female basketball players: a longitudinal study., Clin. J. Sport Med. Off. J. Can. Acad. Sport Med. 7 (3) (1997) 199–206.
[46] J. Cook, K. Khan, Z. Kiss, C. Purdam, L. Griffiths, Prospective imaging study of asymptomatic patellar tendinopathy in elite junior basketball players, J. Ultrasound Med. 19 (7) (2000) 473–479.
[47] A. Qi, D. Zhao, F. Yu, A. Heidari, Z. Wu, Z. Cai, F. Alenezi, R. Mansour, H. Chen, M. Chen, Directional mutation and crossover boosted ant colony optimization with application to COVID-19 X-ray image segmentation, Comput. Biol. Med. 148 (2022) 105810.
[48] A. Kumar, A. Tiwari, A comparative study of otsu thresholding and k-means algorithm of image segmentation, Int. J. Eng. Technol. Res. 9 (2019) 2454–4698.
[49] J. Cook, E. Rio, C. Purdam, S. Docking, Revisiting the continuum model of tendon pathology: what is its merit in clinical practice and research? Br. J. Sports Med. 50 (19) (2016) 1187–1191.
[50] X.Z. Li, S. Williams, M. Bottema, Texture and region dependent breast cancer risk assessment from screening mammograms, Pattern Recognit. Lett. 36 (2014) 117–124.
[51] Q. Wang, HMRF-EM-image: implementation of the hidden markov random field model and its expectation-maximization algorithm, 2012, arXiv preprint arXiv: 1207.3510.