Automatic ROI Detection and Classification of the Achilles Tendon Ultrasound Images

Jamal Benrabha
School of Computing, Science and Engineering
University of Salford
M5 4WT, UK
J.S.Benrabha@edu.salford.ac.uk

Farid Meziane
School of Computing, Science and Engineering
University of Salford
M5 4WT, UK
f.meziane@salford.ac.uk

ABSTRACT

Ultrasound (US) imaging plays an important role in medical imaging technologies. It is widely used because of its ease of use and low cost compared to other imaging techniques. Specifically, ultrasound imaging is used in the detection of the Achilles Tendon (AT) pathologies as it detects important details. For example, US imaging is used for AT rupture that affects about 1 in 5,000 people worldwide. Decision support systems are important in medical imaging, as they assist radiologist in detecting probable diagnoses and lesions. The work presented in this paper concerns the development of a software application to detect changes in the AT ultrasound images and subsequently classify them into normal or abnormal. We propose an approach that fully automates the detection for the Region of Interest (ROI) in ultrasound AT images. The original image is divided into six blocks with 1 cm size in each direction. The blocks lie inside the vulnerable area considered as our ROI. The proposed system achieved an accuracy of 97.21%.

KEYWORDS

Medical image processing, Image classification, Region of Interest, Achilles tendon

1 INTRODUCTION

The AT, also known as the calcaneal tendon, is a tough band of fibrous tissue that connects the calf muscles to the heel bone. AT is the largest and strongest tendon in the human body and is subject to tensile forces of up to 12.5 times the body weight during sprinting and 6 to 8 times during athletic activities such as jumping or cycling. AT plays an important role in the movement and the flexibility of the foot. Just like any other organ, AT is associated with some medical conditions such as the Achilles rupture and Achilles tendonitis. According to [1], AT rupture affects about 1 in 5,000 people worldwide and most frequently (80% to 90% of cases) occurs 2 to 6 cm proximal to its calcaneal insertion.

AT diseases are effectively diagnosed through US images that also help in the detection of these conditions at early stages, thus improving the likelihood of recovery and avoiding AT complications and ruptures. In addition, they help in curtailing the cost of treatment and also the recovery time in some cases. In this research, machine learning techniques are used to develop a model in order to automatically distinguish between normal and abnormal tissues in AT images. To overcome the difficulty of handling the multiplicative noise of the images using gradient operator, Speckle Reduction Anisotropic Diffusion (SRAD) filter is selected. In order to enhance image contrast, we use Contrast Limited Adaptive Histogram Equalization (CLAHE) algorithm, which is an improved version of the adaptive histogram equalization.

Medical image segmentation continues to be a complex and challenging problem. In the current research, the segmentation approach used, is based on the literature that shows the most vulnerable areas in AT. Each image is therefore divided into sub images to determine the ROI. Statistical-based texture features are used as they are widely used in medical image analysis [2]. First and second-order description statistics are also applied to extract various texture features. Gray-Level Co-occurrence Matrix (GLCM) and Gray Level Run-Length Matrix (GLRLM) are employed in this stage to extract various features. In order to overcome the complexity of the dimensionality problem and reduce the features' dimensionality, the non nonlinear technique, Kernel Principal Component Analysis (KPCA), is used. To enhance the classification process results and selecting the most suitable classifier for data, different classifiers were tested. Decision Tree (DT), nonlinear Support Victor machine (SVM) and Ensembles classifiers are used to classify the dataset images using the extracted features. The dataset is divided into two groups: one for training and the other for testing. A comparison of the results of these classifiers is performed and is illustrated using the confusion matrix and Receiver Operating Characteristics (ROC) curves. The remaining part of the paper is organized as follows. The next section gives the necessary background on the ROI detection and the influence its size on classification. Section 3 introduces the proposed solution and section 4 its implementation. The results are discussed in section 5, and the paper concludes in section 6.

2 Related Work

As far as we are aware, this is the first study looking at the automatic identification of the medical condition and

classification of AT US images. Hence, our literature review on the ROI is based on works conducted on other organs and images to identify the best algorithm and processes that will produce the best classification system of the AT.

2.1 ROI Detection

Many factors affect the results of feature extractions and subsequently classifications. Determining the appropriate ROI is an important parameter and is the step before starting the features extraction stage. Once the ROI is determined, the following tasks will use this region to classify the image into normal or abnormal.

Many methods and approaches for detecting the ROI are found in the literature [3] and these are divided into two types; manual/semi-automatic and fully automated. In manual and semi-automatic methodologies, the user will interact with the system to determine the ROI while in fully automated methodologies there is no user intervention. Stoitsis et al [4] proposed a manual system for ROIs selection by a specialized physician. Five texture features were extracted from each ROI and a genetic algorithm for feature selection was applied and achieved a 90.63% accuracy using a Neural Network. Manual segmentation is known to be time consuming and tedious. Hence, other alternatives are proposed such as that suggested by Hojjatoleslami and Kruggel [5] for lesion segmentation where a semi-automatic method using the region growing approach was used to segment a lesion from the surrounding tissues. An automatic segmentation method is performed in [6] and [7] using the active contours model, which is based on variable-weighted for local and global forces. However, according to Li, et al [8], semi-automatic computerized systems are used to allow the radiologist to adjust the segmentation parameters for an optimal performance. Therefore, a robust model is needed to assist radiologists to locate the suspicious areas automatically without user intervention.

2.2 ROI Size

There are several medical studies that demonstrate that the size of the ROI plays an important role in the classification process [9]. A study carried out by Basset and Mestas [10] on the influence of the ROI size on ultrasonic images of the prostate using mean of the co-occurrence matrices, reported that the size of the ROI affects the discrimination results. According to Singh, Singh and Gupta [11] the ROI size has a significant effect on the computed value of most tissue characterization features. However, as reported by Osicka, Freedman and Ahmed [12], the noise might have more effect on the measured texture from smaller ROI. On the other hand, if fewer features are extracted, this can decrease the vulnerability of these features to noise distortion [13]. Multiple ROIs was proposed by He et al [14] to enhance the classification accuracy. The segmentation of the lesion tissue into inside and border areas resulted in the building of new feature vectors that improved the classification accuracy.

3 Proposed Solution

3.1 Introduction

AT US images are first acquired from different normal and abnormal patients using specified US instruments. The task of the pre-processing stage is to reduce the amount of speckle noise without affecting the image features. However, in order to focus on the most vulnerable AT area, image dividing was applied to determine the ROI. Subsequently, different texture features were extracted from the ROI using GLCM and GLRM algorithms. These sets of features, along with histogram features, underwent a process of dimensionality reduction using the KPCA algorithm. In the classification stage, various classifiers have been tested in order to determine the most appropriate one for our application based on the final results. We applied SVM, k- Nearest Neighbors (k-NN), Decision Trees (DT) and Ensemble classifiers, and through their results we are able to evaluate their performance. Finally, the classifier-based efficiency is evaluated using different evaluation metrics. The proposed technique is for improving AT US images diagnosis. Generally, it involves six subsequent stages, namely, Image Acquisition, Pre-processing, Determining the ROI, Feature Extraction & Reduction, Classification, and finally, Evaluation, as shown in Figure (1).
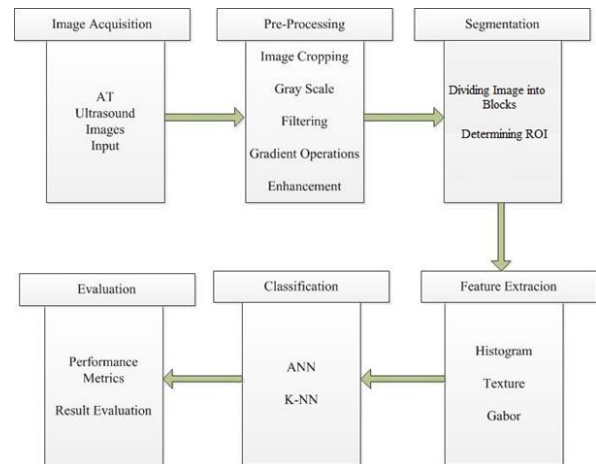

Figure 1: The proposed 6-stage AT classifier.

3.2 Image Pre-processing

To overcome the problem of having speckle noise in the AT image, the SRAD de-speckle method was used. It is considered better than the classical anisotropic reduction method when the image contains speckle noise. SRAD filter has the ability to preserve the edges, and also smoothed the rest of the image while reducing the noise. A sliding window with a number of iterations is defined as a parameter for the SRAD method in order to get good results. To evaluate the performance of the used filter a comparison between SRAD filter and some other popular filters was performed. The performance of the SRAD filter was assessed with the common quality metrics. According to Table (1) SRAD

is performing better than the other de-noising filters. Figures (2) and (3) show an AT image before and after applying SRAD respectively.
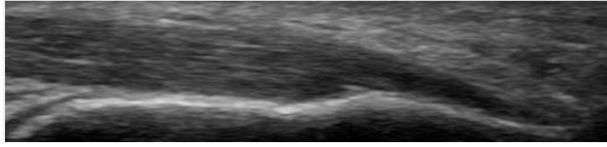

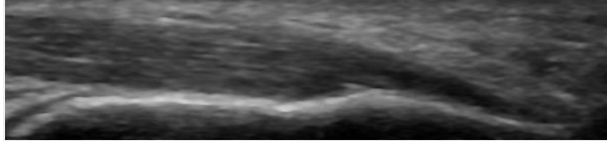**Figure 2: AT image before applying any filter.**

Figure 3: AT image after applying SRAD filter.

Table 1: Results of applying different filters to the noisy AT image

	PSNR	SNR	RMSE	MSE	IQ	CC
Mean	23.82	10.57	0.64	0.004	0.8	0.8
Median	22.42	9.181	0.07	0.005	0.9	0.7
Lee	22.67	9.425	0.07	0.005	0.8	0.8
Frost	32.54	19.30	0.02	0.001	0.9	0.9
Kuan	13.34	0.09	0.21	0.046	0.0	0.3
Wavelet	23.82	10.57	0.64	0.004	0.8	0.8
SRAD	35.19	21.95	0.01	0.000	0.8	0.9

To improve the image contrast without over-enhancing noise content or introducing artefacts to the processed image we used CLAHE. It works on small regions in the image called tiles rather than the entire image. This way, the contrast in the homogeneous region can be limited to its minimum and it also limits the appearance of artefacts and noises. Figure (4) shows the AT image after applying CLAHE histogram enhancement. It can be clearly seen from the Figure that image details and edges are preserved.
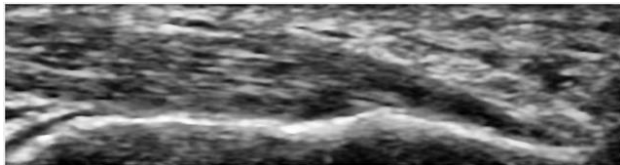

Figure 4: AT image after contrast was enhanced using CLAHE algorithm.

3.3 Dividing the Image into Blocks

Determining the ROI plays a vital role in many biomedical image applications and represents the foundation for the next stages. If any shortage occurs at this stage, the next stages will be affected. We intended to locate the suspicious area in AT US image by subdividing the image into component regions. There are many algorithms used for image segmentation, and no single one can be considered as superior due to their dependence on different conditions and constraints. There are various kinds of detection methods and approaches available, such as manual, semi-automatic, or fully automatic detection.

In this approach we tried to achieve one of the research goals which is to fully automate the identification of the ROI area. In the first instance, the system divides the AT image into six equal non-overlapping blocks. These blocks are equal in size and are in adjacent positions. Dividing the AT images is done automatically and each block is saved as a separate image. Hence, our dataset is now composed of 215 blocks obtained from the 57 AT images after excluding 114 untargeted blocks and 13 distorted blocks. In Figure (5) the first block (a) starts from the Achilles insertion in the calcaneal followed by the second block (b), and together will occupy approximately 2.0 cm, starting from the calcaneal insertion point. These blocks are untargeted blocks in all AT images. The last four blocks starting from (c) to (f) in Figure (5) contain the most vulnerable AT area according to [1]. More focus is placed on these blocks in extracting various features and obtaining the classification results.

Features are extracted from the last four blocks, which represents the ROI. These features are reduced using the nonlinear reduction method KPCA, and then the returned predictors are used as input for the classifiers.
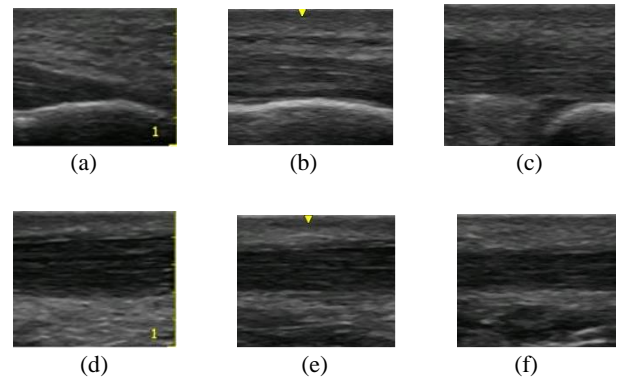

Figure 5: Dividing the AT ultrasound image into 6 equal blocks. (a), (b), and (c) are AT ultrasound image lower part. (d), (e) and (f) AT ultrasound image middle part.

3.4 Extracting Features

Extracting image features plays a crucial role in this research, as it provides the crude data for the subsequent stages. The decision as to whether the AT is normal or abnormal mainly relies on the features extracted. In this research, various features are extracted from the AT US image using different feature extraction methods. Most image features have been constructed based on

histogram and texture features [15]. Statistical descriptor, co-occurrence matrices for four different angles, and run length encoding algorithm were used for extracting different texture features. Furthermore, different histogram features such as skewness and kurtosis are extracted from the samples after different segmentation methods were applied. However, all these features are subject to filtering using a reduction method in the next stage.

For the histogram features extraction the spatial distribution in the ROI is not taken into account, but rather the distribution of the individual pixels value is calculated in each region. Mean, variance, standard deviation, skewness and kurtosis are calculated for all image datasets, yielding 5 features for each ROI in each sub image. GLCM is a two dimensional function determining the relationship between a pixel and its neighbors over an entire image in different directions. Fourteen texture features are computed with d=1 where d is the distance in pixels and four directions θ of 0^0, 〖45〗^{0}, 〖90〗^{0} and 〖135〗^{0}. Thus, the total texture feature vectors are 14 measures * 4 directions resulting in 56 features for each ROI in each image. Figure (6) shows the GLCM matrix with distance d=1 and the direction θ=0.
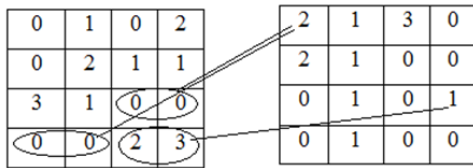

Figure 6: GLCM of 4 x4 image with d = 1 and θ=0.

GLRLM is a statistical approach describing the texture information of a gray level image region. It is represented by run length matrix P(i, j|θ), where i is the number of gray level that runs in length j in different directions, often in θ=0^0, 〖45〗^{0}, 〖90〗^{0}, 〖135〗^{0}. Figure (7) shows an example of GLRLM method.
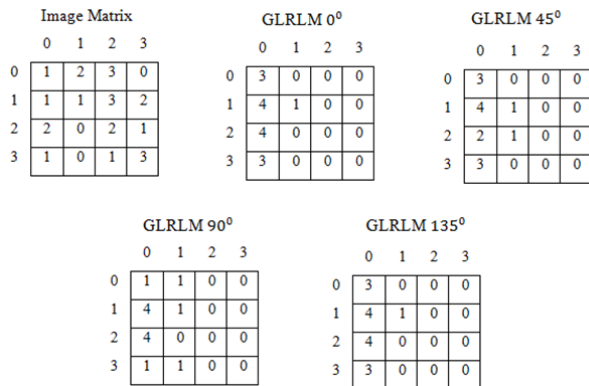

Figure 7: Calculating GLRLM elements values for 0, 〖45〗, 〖90〗 and 〖135〗 orientation.

By using gray level matrix seven texture features were computed for each image (SRE, LER, GLN, RLN, RP and LGRE. These features capture both structural and statistical information from the texture in the ROI [16].

3.5 Features Reduction

At this stage the extracted features from the previous stage are reduced into meaningful representations through a dimensionality reduction method. By this reduction the complexity of dimensionality is mitigated and we might eliminate some features as just noise. In case of complex nonlinear data, nonlinear techniques have shown superiority over their linear counterpart. Figure (8) highlights the data scatter using nonlinear PCA, or kernel PCA. In this research we adopt Kernel PCA.
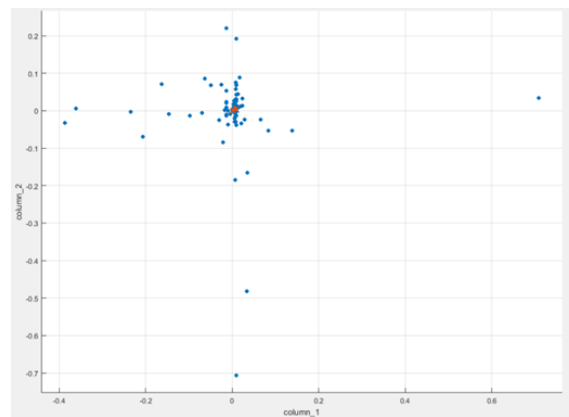

Figure 8: Data scatter using KPCA reduction method.

3.6 Classification

The efficiency of this stage depends on two important aspects. Firstly, the efficiency of the algorithms had chosen in the previous stages, before we apply the classification. All these stages have significant impacts on the classification algorithm results and it is important to tune all of them. Secondly, choosing the correct classifier is a very important part in the pattern classification system, because different classification results may be obtained depending on the choice of the classifiers.

In medical image processing supervised classification methods are used, where the data are already labelled according to their predefined categories. In this research, for all the proposed classifiers we use the same data matrix, which is extracted from AT US images. The classification process is carried out in two stages namely learning and testing. During the learning stage the classifier is trained on the extracted features to create a model, which is then used in the testing stage. In the training stage the input vectors are named a 'training set'. After the learning stage is completed the classifier is ready to be tested on new data that have not been used for its training before.

In order to separate all samples into training and testing sets we used the K-fold cross validation method. In this method the dataset is divided into k subsets called folds. For each k experiment we use k-1 folds for training and the remaining one for testing. Then, the average error across all k trails is computed. Although this approach can be computationally expensive it is recommended for small datasets.

4 System Implementation

4.1 Database Creation

Due to the lack of research on US AT images, it was not possible to find medical AT image datasets in the commonly used machine learning repositories on the internet. For this reason, an image database needs to collect for both normal and abnormal patients from radiology departments. A total of 57 AT US images were collected from normal and abnormal cases in the JPEG format from the radiology department at the University of Salford. Each of these images is already diagnosed by medical experts and labelled as normal or abnormal for the purpose of training and testing the classifiers in later stages. From the whole set of 57 AT US images, 49 of them are labeled as normal and the remaining 8 as abnormal. A portable Venue 40 musculoskeletal US system (GE Healthcare, UK) with a 5–13 MHz wideband linear array probe with 12.7 mm _ 47.1 mm footprint area was used for scanning. The scans were performed independently by different operators according to a reliable scan protocol and within the same session [17]. Good contact was maintained between probe and skin without applying excessive pressure. Three assessments were taken at each site of the AT with the probe removed between each recording. Each subject lay in the prone position for scanning.

4.2 Software Tools

Through a computer application built using MATLAB environment an expert user can interact with the system to perform all the basic tasks of image processing. All the algorithms explained in the proposed solution implemented in such application. Image cropping, resizing, de-noising and enhancing can be done easily. Dividing the AT images is done automatically and each block is saved as a separate image. The application will determine the ROI and extract the required texture features automatically without any user intervention. A feature matrix is built from all the extracted features, which is 5 histogram features, 56 GLCM features and 7 GLRLM. So, the total number of features that were extracted from the mentioned algorithms was 68 features for each ROI in each AT sub image block. Also, the classification results from the specified classifier will be shown at the end of the whole process.

5 Results and Discussion

Sensitivity and specificity are two important characteristics used to evaluate clinical tests. Sensitivity is the probability of AT images that are normal and identified by the classifier as normal. On the other hand, specificity is the proportion of AT images that are abnormal and diagnosed by the classifier as abnormal, while the accuracy is the proportion of the total number of predictions that were correct. Due to a few numbers of AT images in this research and to overcome the fluctuation in the classification results, the experiment has been repeated 50 times and the average of all these results has been taken. This iteration process has been done in order to be more accurate when calculating sensitivity, specificity and accuracy ratio.

Confusion matrix is a D X D matrix where D is the number of classes. In this matrix the number of True Positive (TP), means the samples that are abnormal and the test claims that they are abnormal. If some are normal because they have certain medical problems and classified by the test as abnormal these are called False Negative (FN). There are some cases where they are abnormal and classified as normal − this is called False Positive (FP). Finally, when the images are normal and the test says that they are, this is called True Negative (TN). On the other hand, ROC curves are two dimensional graphs highlighting the relationship between true positive rates on the Y axis and false positive rates on the X axis. The Area Under the ROC Curve (AUC) is often used to measure the quality of the classification model. Any realistic classifier has an AUC between 0.5 and 1 [18]. Random classifier has an AUC of less than 0.5, while AUC for the perfect classifier is equal to 1.

In this approach, we tried to achieve one of the research goals, by preventing the user getting involved in determining the ROI. Hence, automating the model by divide the AT image into six equal non-overlapping blocks. These blocks are equal in size and are in adjacent positions. To estimate the efficiency of this approach, features are extracted from the last four blocks, which represents the ROI. These features are reduced using nonlinear reduction method KPCA, and then the returned predictors are considered as input for the classifiers. A Comparison between different classifiers' performance is shown graphically in Figure (9).
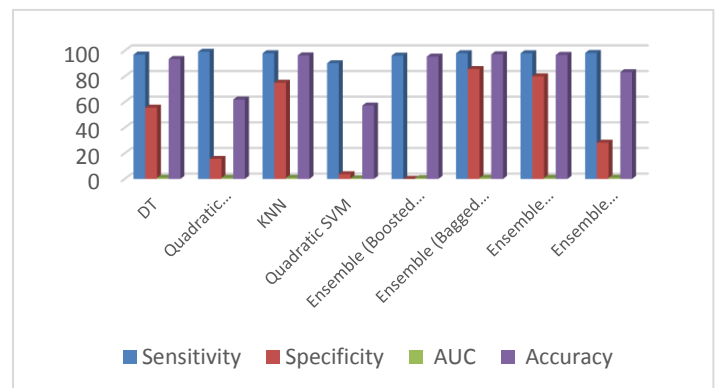

Figure 9: Graph showing Sensitivity, Specificity, AUC and Accuracy for different classifiers.

Table (2) uses four indicators to show the ability of each classifier in terms of prediction and decision making of AT US images. It gives the summary of various classifiers sensitivity, specificity,

AUC and accuracy. In terms of accuracy it can be seen that the classifiers accuracy is in the range between 57.21% for the quadratic SVM to 97.21% for the Bagged tree ensemble classifier.

Table 2: Various classifiers' results for AT ultrasound image classification

Classifiers	Sensitivity	Specificity	AUC	True Positive	True Negative	Acuuracy
DT	96.95	55.56	0.88	191	10	93.49
Quadratic Discriminant	99.16	15.62	0.87	118	15	61.86
KNN	97.99	75.00	0.86	195	12	96.28
Quadratic SVM	90.23	3.66	0.42	120	3	57.21
Ensemble (Boosted Trees)	96.10	80,00	0.61	197	8	95.35
Ensemble (Bagged tree)	98.01	85.71	0.91	197	12	97.21
Ensemble (Subspace KNN)	98.00	80.00	0.86	196	12	96.74
Ensemble (RUSBoosted Trees)	98.22	28.26	0.90	166	13	83.26

It is noticeable from the figures that the sensitivity and specificity has higher values with most of ensemble classifiers. Where the higher value of sensitivity means a higher chance to identify a patient with certain diseases as positive. On the other hand, a higher value of specificity means a higher chance to identify a patient without disease as negative. It is clear that the proposed method gives better results when focusing on the most vulnerable area in AT which is 2 cm away from the calcaneal insertion with length of 4 cm. This is expected to provide valuable diagnosis information for the physicians.

6 Conclusion and Future Work

The study determined the ROI in AT US image automatically by dividing the image into sub-images and targeting the most vulnerable area. Determining the ROI area and size based on the medical reports states that the most vulnerable area in AT is 2 to 6 cm proximal to the calcaneal insertion point. In this study a matrix with 68 different texture features was extracted from the whole samples using the feature extraction methods mentioned previously. Using the KPCA algorithm as a reduction method with the Gaussian kernel function these features are reduced to the minimum number that corresponds to the largest eigenvalues. The summary of various segmentation and classification techniques with their classification accuracy and sensitivity of AT detection has been presented. To enhance the classification process results and select the most suitable classifier for the data, different

classifiers were tested. DT, Discriminant Analysis, Quadratic SVM, k-NN and different ensemble classifiers are used to classify the dataset images through the extracted features. A comparison between these classifiers' results using a confusion matrix and ROC curves was performed. From the experiments conducted, we noticed that the ensemble classifiers especially Bagged trees type achieved a high accuracy than the other classifiers. From Table (2), it is proven that the classification performance of the ensemble classifiers is superior to most of the others. However, the study has been found that the highest accuracy result, 97.21% is achieved by the Bagged tree ensemble classifier.

Future studies will investigate the integration of AT thickness as a feature in the classification process. Automatic thickness measuring technique will be explored in future studies.

REFERENCES

1. Maffulli, N. and L.C. Almekinders, *The Achilles Tendon*. 2007, London: Springer-Verlag London Limited.
2. Holli, K.K., et al., *Texture analysis of MR images of patients with mild traumatic brain injury.* BMC medical imaging, 2010. **10**(1): p. 8.
3. Cheng, H.-D., et al., *Automated breast cancer detection and classification using ultrasound images: A survey.* Pattern Recognition, 2010. **43**(1): p. 299-317.
4. Stoitsis, J., et al., *Computer aided diagnosis based on medical image processing and artificial intelligence methods.* Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment, 2006. **569**(2): p. 591-595.
5. Hojjatoleslami, S. and F. Kruggel, *Segmentation of large brain lesions.* IEEE Transactions on Medical Imaging, 2001. **20**(7): p. 666-669.
6. Kim, S., et al. *Automatic segmentation of leg bones by using active contours.* in *Engineering in Medicine and Biology Society (EMBC), 2014 36th Annual International Conference of the IEEE.* 2014. IEEE.
7. Zhang, K., et al., *Active contours with selective local or global segmentation: a new formulation and level set method.* Image and Vision computing, 2010. **28**(4): p. 668-676.
8. Li, et al., *Integrating Spatial Fuzzy Clustering with Level Set Methods for Automated Medical Image Segmentation.* Comput Biol Med, 2011. **41**(1): p. 1-10.
9. Chan, K. and K. McCarty. *Aspects of the statistical texture analysis of medical ultrasound images*. in *Ultrasound Instrumentation, IEE Colloquium on*. 1990. IET.
10. Basset, O. and J.-L. Mestas, *Texture Analysis of Ultrasonic Images of the Prostate by Means of Co-Occurrence Matrices.* ULTRASONIC IMAGING, 1993.
11. Singh, M., S. Singh, and S. Gupta. *Investigations on ROI selection for liver classification*. in *Electrical and Computer Engineering (CCECE), 2014 IEEE 27th Canadian Conference on*. 2014. IEEE.
12. Osicka, T., M.T. Freedman, and F. Ahmed, *Characterization of pulmonary nodules on computer tomography (CT) scans: The effect of additive white noise on features selection and classification performance.* Medical Imaging, Spie-Int Society Optical Engineering,, 2007. **6512**: p. 51245-51245.
13. Al-Kadi, O.S., *Assessment of texture measures susceptibility to noise in conventional and contrast enhanced computed tomography lung tumour images.* Computerized Medical Imaging and Graphics, 2009. **34**(6): p. 494-503.
14. He, Z., et al. *Dependence of tissue characterization features on region of interest (ROI) size: studies on phantoms and simulations*. in *Ultrasonics Symposium, 2004 IEEE*. 2004. IEEE.
15. ping Tian, D., *A review on image feature extraction and representation techniques.* International Journal of Multimedia and Ubiquitous Engineering, 2013. **8**(4): p. 385-396.
16. Tang, X., *Dominant Run-Length Method for Image Classification*. 1997, Woods Hole Oceanographic Institution.
17. Crofts, G., et al., *Reliability of ultrasound for measurement of selected foot structures.* Gait & posture, 2014. **39**(1): p. 35-39.
18. Fawcett, T., *An introduction to ROC analysis.* Pattern recognition letters, 2006. **27**(8): p. 861-874.