An automatic intelligent system for diagnosis and confirmation of Johne's disease

Inad Aljarrah , Anas Toma and
Mohammad Al-Rousan

Computer Engineering Department,
Jordan University of Science and Technology,
Irbid 22110, Jordan
Email: inad@just.edu.jo
Email: anas.toma@student.kit.edu
Email: alrousan@just.edu.jo

Abstract: Johne's disease is one of the most widespread bacterial diseases of domestic animals. It causes yearly losses of billions of dollars worldwide. In this paper an automatic intelligent computer-aided system is proposed for the diagnosis of Johne's disease, the system uses image analysis and computer vision techniques to extract features from two different microscopic images, then those features are classified using neural networks and K-nearest neighbour (K-NN) techniques to diagnose Johne's disease. The proposed system employs histopathological examination to extract 192 different texture features. The features are then reduced into only 8 features and classified using artificial neural networks (ANN). The acid fast stain test is used to confirm the positive cases. The construction and testing of both models are carried out using a total of 294 microscopic images, 194 images for the histopathological examination test which produces an overall accuracy of 98.33%. The other 100 images are used for the acid fast stain test, and it achieves an accuracy of 96.97%.

Reference to this paper should be made as follows: Aljarrah, I., Toma, A. and Al-Rousan, M. (2015) 'An automatic intelligent system for diagnosis and confirmation of Johne's disease', *Int. J. Intelligent Systems Technologies and Applications*, Vol. 14, No. 2, pp.128–144.

Biographical notes: Inad Aljarrah has been an Assistant Professor in the Department of Computer Engineering at the Jordan University of Science and Technology since 2006. He received his BS in Electrical Engineering from the Jordan University of Science and Technology in 1999. He received his Master's and PhD in Electrical and Computer Engineering from the Ohio University, Athens, Ohio, USA in the years 2002 and 2006, respectively. His research interests are computer vision, image processing and analysis, artificial intelligent systems. Currently, he is the Acting Chair of the Computer Engineering Department at the Jordan University of Science and Technology.

Anas Toma is a PhD student at the Karlsruhe Institute of Technology (KIT), Germany. His main research interests are computer vision, high-performance and energy-efficient scheduling, algorithm design and analysis, and real-time systems. He received his BSc in Computer Engineering from the An-Najah National University, Palestine, in 2005. He worked for one year as Teaching and Research Assistant at the An-Najah University. In 2008, he received his MSc in Computer Engineering from the Jordan University of Science and Technology, Jordan. From 2008 to 2011, he worked as a Lecturer at the An-Najah University. Then, he joined KIT and started his doctoral research in October 2011.

Mohammad Al-Rousan is a Professor at the Network Engineering and Security Department in the College of Computer and Information Technology at the Jordan University of Science and Technology. He received his BS in Computer Engineering from the King Saud University, 1986, MSc in Electrical Engineering from the University of Missouri/Columbia, 1992, and his PhD in Electrical Engineering from the Brigham Young University, 1996. His research interests include the areas of sensor networks, artificial intelligence, web-based systems, programming languages, nano technology, cloud computing. He served as the Dean of the College of Computer and Information Technology between 2008 and 2012 at the Jordan University of Science and Technology.

1 Introduction

Diagnosis in medicine is often performed using visual recognition of abnormal conditions which are either directly viewed by the physician or scanned into an image from X-rays, magnetic resonance imaging (MRI), or microscope (Lew, 2001). Computer algorithms and analytical methods have endeavoured to maximise the anatomical and physiological information derived by physical principles for different imaging modalities (Dikshit et al., 2005). It has become an important field of clinical and research laboratories. Clinical applications include diagnosis, treatment planning, and surgical guidance. In research laboratories, medical imaging is a powerful tool for investigating normal/abnormal physiological processes and for monitoring experimental treatment effects. Medical imaging depends mainly on computer aided techniques, it is a correlative field that includes physics, mathematics, biology, image processing, pattern recognition and artificial intelligence. Pathologist diagnosis diseases by means of observing the pathological changes in the infected tissue compared with the healthy one. There are many features that can be extracted from medical images after a series of steps of enhancements and filtering. The features depend on the image contents (tissue components) so they present a mathematical descriptors for that images which can be used to diagnose different diseases using one of the classification techniques such as Neural Networks. Many intelligent systems have been employed on diagnosis of human diseases, but none has been devoted to Johne's disease. The main objective of this research is to develop a novel computer-aided system using techniques from texture analysis, colour segmentation, pattern recognition, and machine learning to

provide the diagnostic category of Johne's disease. The system consists of three main stages. The first one includes the analysis of two microscopic images from sheep intestine tissue using different image processing techniques to extract features. Secondly, the discriminative features from the previous stage which directly affect disease diagnosis are selected by different data mining algorithms. Finally, diagnosis of Johne's disease by a decision support system based on neural network and K-nearest neighbour classifiers is carried out (Toma, 2008).

2 Johne's disease

Johne's disease (Paratuberculosis) is caused by the bacterium Mycobacterium avium subspecies paratuberculosis (MAP). Paratuberculosis affects domestic and wild animals on the five continents (Ayele, 2001; Rolf et al., 1996). This disease induces emaciation, decrease in milk and meat production, diarrhea and death. The yearly losses due to the disease were estimated to exceed $1.5 billion in the US. Moreover, since Johne's disease is one of the most widespread bacterial diseases of domestic animals, its impact on the world economy is enormous (Cocito et al., 1994; Gupta et al., 2012). Crohn's disease is a chronic inflammation of the intestine in humans which presents some clinical and pathological similarities to the lymphocytic/paucimicrobial form of Paratuberculosis in animals. Mycobacterium avium subspecies paratuberculosis has been detected in approximately 30% of pathological samples from patients with Crohn's disease (Sechi et al., 2001; Commission, 2000). Johne's disease can be detected by three different diagnostic methods; bacterial culture, histopathology, or PCR technology. In our work, we will be using the histopathology diagnostic method only in sheep, where the disease diagnosis can be done by the observation of the pathological changes on the tissue compared with the healthy one. Two microscopic images taken from intestine tissue (ileocecum) are used to diagnose Johne's disease. The first one is tested by histopathological examination (called test1 in our system), by observing the cells pattern in an intestine tissue. The affected tissue has epithelioed cells as shown in Figure 1(b). The second test is the acid fast stain test which is used for confirmation. The normal tissue has a blue background, Figure 2(a). On the other hand, a bacterium of red colour appears in the infected tissue as shown in Figure 2(b).

Figure 1 Normal vs. infected tissues for histopathological examination: (a) normal tissue and (b) infected tissue (see online version for colours)
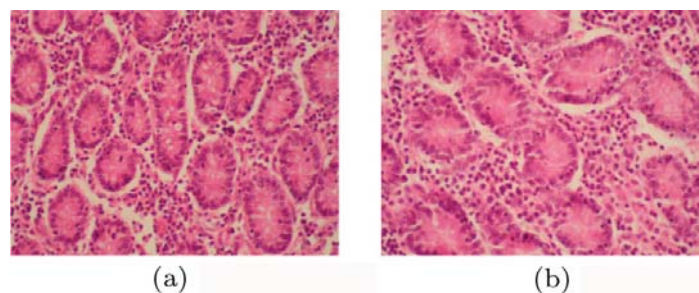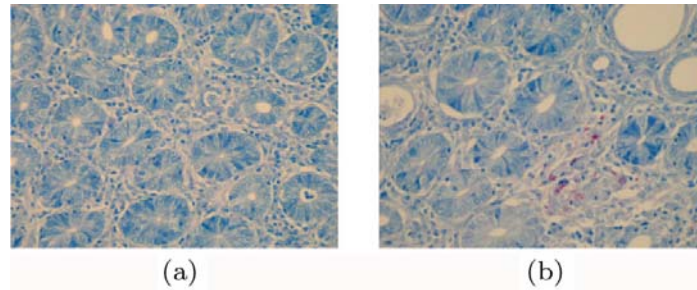

(a)   (b)

Figure 2 Normal vs. infected tissues using acid fast stain test: (a) normal tissue with blue background and (b) image shows red colour bacterium (see online version for colours)

(a)　　　　　　　　　(b)

3 Image acquisition

Images were captured using the Nikon imaging system. It consists of a microscope provided with a coloured digital camera, which is linked to a computer via an image capturing equipment. A total of 294 coloured images where captured at $40\times$ magnification factor, with the help of the College of Veterinary Medicine at Jordan University of Science and Technology. The size of the image is 640×480. More than one image were taken from the same slide, each image is considered a separate sample.

4 System architecture

The proposed approach is implemented using Matlab. It begins by reading an image from the user, then a preprocessing step to enhance the image in order to get more accurate classification results is performed. Texture analysis techniques are used to extract features, which are then fed to the artificial neural networks (ANN) to classify the input image. If the diagnosis is positive then a confirmation step is carried out by asking the user to input acid fast stain test image.

The system uses two different images of each sample to perform two tests. The first test is histopathological examination (test1) that is used mainly for diagnosis of Johne's disease. In this test giant and epithelioid cells are observed in the infected tissue, five texture analysis techniques are used in our system to recognise it. The second test (test2) is the acid fast stain test which is used for confirmation, in this test the red colour bacteria is detected using colour segmentation.
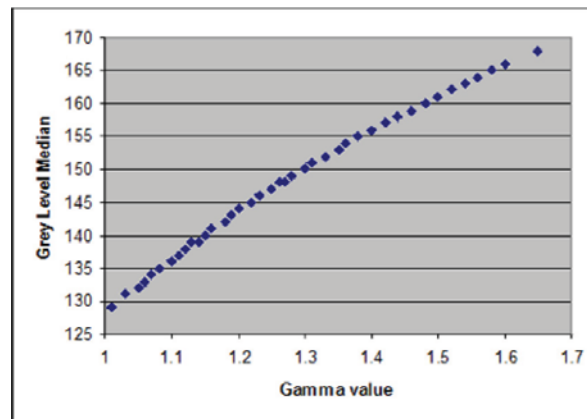
5 Histopathological examination

Two main problems were encountered in this test. Firstly, the amount of the staining over the tissue during slides preparation phase results in a non uniform illumination between images. This affects the grey level values when converting the coloured image into grey. Also, changes in microscope light intensity increase the difficulty of processing the images. The second problem happens through the trimming operation during preparation of the slides. The shape of the cells is different from image to another according to the cutting section (vertical, horizontal or any degree). Those two problems cannot be avoided during

the preparation process, but we overcame the above problems using image processing techniques that will be discussed later. Test1 starts with image preprocessing to solve the intensity problem, then extracts a total of 192 texture features to recognise the infected tissues and overcome the differences in the crossing sections. Feature selection step is used to select the reduced discriminative features to simplify the classification process.

5.1 Preprocessing

The preprocessing step is an important step to enhance the images in order to get more accurate classification results. It includes converting the coloured images into grey model and the illumination correction. We start by converting the coloured input images into grey, then illumination correction is done using gamma correction (Lew, 2001; Gonzalez and Woods, 2008) to overcome the problem of the difference in the intensity values mentioned previously. Gamma value depends on the median of image histogram, we shifted the median of all images histograms to 128 (grey levels from 0 to 256), Figure 3 shows the relation between the grey level median of the image and gamma value for 40 samples. The previous relation can be used for further preprocessing operation to get the unknown gamma value depending on the median of image histogram. Shifting to a 128 grey level median value makes the illumination of all images nearly equal.

Figure 3 Relation between grey level median of the image and the gamma value (see online version for colours)

5.2 Feature extraction (texture analysis)

In this step, five texture analysis techniques are used to extract 192 different features, texture analysis was used because changes in the structure of the infected images were observed. A large number of features are necessary because of the difference in the cells pattern and shape among images of the same class, the main reason is the cutting section problem mentioned previously. The texture features extracted from the input images in our proposed system are histogram, wavelet, co-occurrence matrix, run-length matrix and Fourier spectrum features.

5.2.1 Histogram features

The histogram shape provides much information as to the character of the image, different useful parameters (image features) can be worked out from the histogram to describe the first-order statistical properties of the image. Eight descriptors were extracted from the images histogram. Mean, variance, standard deviation, smoothness, skewness, kurtosis, energy and entropy are the extracted histogram features (Materka and Strzelecki, 1998; Antani et al., 2002). We observed that there are difference in the histogram shape between the healthy and the infected tissue images, but that difference did not classify all images, so other analysis techniques were used to cover all images patterns.

5.2.2 Wavelet features

The discrete wavelet transform (DWT) is a linear transformation that operates on a data vector whose length is an integer power of two, transforming it into a numerically different vector of the same length (Kocioek et al., 2001; Mallat, 1989). The wavelet transform features were extracted to analyse the frequency contents of the images within different scales. Haar wavelet transform is used with 3 scales to produce 10 different channels, and then the energy of each channel is calculated to get 10 different features (Castellano et al., 2004). The energy of each channel can be evaluated by calculating the mean magnitude of its wavelet coefficients.

5.2.3 Co-occurrence matrix features

The co-occurrence matrix is a technique that allows for the extraction of statistical information from the image regarding the distribution of pairs of pixels (Haralick, 1979). The co-occurrence matrix technique is used to construct 20 co-occurrence matrices for five distances between image pixels ($d = 1, 2, 3, 4$ and 5), and for horizontal, vertical, 45° and 135° directions. From each matrix, six features were extracted, energy, inertia, entropy, homogeneity (Inverse difference), maximum probability and absolute value. As a result of using this analysis, 120 co-occurrence matrix based features are used to describe the input images. Because of that large number of feature, some sort of feature selection method must be used to select the most relevant features (Chen et al., 2005).

5.2.4 Run-length matrix features

The run-length matrix is a way of searching the image, across a given direction, for runs of pixels having the same grey-level value. The grey level run-length approach characterises coarse textures as having many pixels in a constant grey one run and fine textures as having few pixels in a constant grey tone run. A total of 44 run-length matrix based features were extracted. Four run-length matrices were constructed from each image using horizontal, vertical, 45° and 135° directions. The 11 features calculated from each run-length matrix are short run emphasis (SRE), long run emphasis (LRE), grey-level non uniformity (GLN), run length non uniformity (RLN), run percentage respectively (RP) (Xiaoou, 1998), low grey-level run emphasis (LGRE), high grey-level run emphasis (HGRE) (Chu et al., 1990), short run low grey-level emphasis (SRLGE), short run high grey-level emphasis (SRHGE), long run low grey-level emphasis (LRLGE) and long run high grey-level emphasis (LRHGE) (Dasarathy and Holder, 1991).

5.2.5 Fourier spectrum features

The spectral features are based on the Fourier spectrum, they can be used to discriminate between periodic and non periodic texture patterns, and quantifying differences between periodic patterns, mainly by observing the high-energy bursts in the spectrum. Features extracted using Fourier spectrum analysis includes the highest value and its location, mean, variance, and the distance between the mean and the highest value of both $S(r)$ and $S(\theta)$ functions.

5.3 Feature selection

Due to the large number of features extracted (a total of 192), it is difficult and complicated to use all of them in a classification system. The accuracy of the classification system may be affected because there are irrelevant features to class labels. Also, a large storing space is needed for lazy classifiers, and a long training time is spent for eager classifier because of many redundant features. Principal component analysis (PCA) (Han, 2005) determines the highest correlated feature with the class label vector. Also, Fisher's discriminant ratio (FDR) (Theodoridis and Koutroumbas, 2006) is a separability measure between two classes (positive and negative). Both techniques were used to reduce the number of features from 192 to only 8 discriminant features.

5.4 Learning and classification

Classification (also called pattern recognition) is a form of data analysis that can be used to extract models describing important data classes. First of all, the data with known class labels should be divided into two sets, training set and testing set. The training set is used to construct the classification system. Whereas the testing set is used to evaluate the system. Artificial neural networks (ANN) (Jain et al., 1996; Rich and Knight, 1991) were used to classify the data in test1 into two classes, healthy and infected with Johne's disease, ANN was used as a classifier because it is a highly tolerance to noisy data and can closely approximate any function. A two layer feed forward neural network with back propagation is designed and trained for this purpose.

The input layer consists of 8 inputs which represents the number of reduced input features extracted using texture analysis, while there are 5 neurons in the single hidden layer. Because we have only two classes, one neuron is used in the output layer, the zero output denote healthy tissue and one output for infected one. The reduced texture features were normalised to the range (0–1) before starting training and testing. 194 tuples of features and its classes were used in training and testing using holdout method. The data were randomly partitioned into two independent sets; training set which includes 2/3 of the data for ANN construction, and the remaining data is called test set and used for accuracy estimation. The ANN was trained until an acceptable accuracy was achieved, and then it was used for further classification to diagnose Johne's disease.
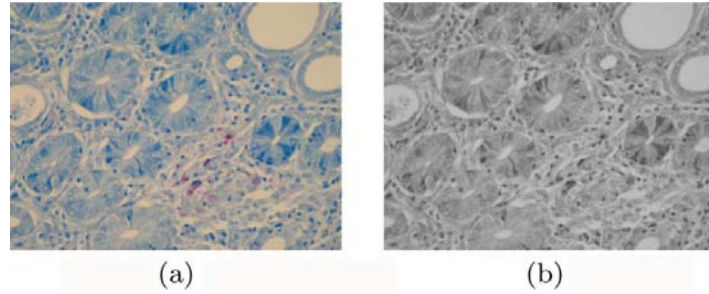
6 Acid fast stain test

The main steps of test2 include colour segmentation to detect the red colour bacteria, and removing the noise using morphological operations. Finally, features are extracted to detect the disease by the help of the classification system.

6.1 Segmentation

Image segmentation is the process of separating the image into regions of similar properties. Simply, it means to check each individual pixel to see whether it belongs to an object of interest or not. A pixel has the value one if it belongs to the object; otherwise it is zero. Using one of segmentation algorithms the operation will produce a binary image. In this step, the goal is to isolate the regions of interest (ROI) which represent the bacteria from each image. Converting the coloured image into grey scale makes it difficult for segmentation process to isolate the ROI (bacteria). Figure 4(a) shows an image of infected tissue with clear red-coloured bacteria, but they are invisible in Figure 4(b) which is the grey scale of Figure 4(a).

Figure 4 Conversion from colour to grey scale: (a) infected tissue and (b) healthy tissue (see online version for colours)
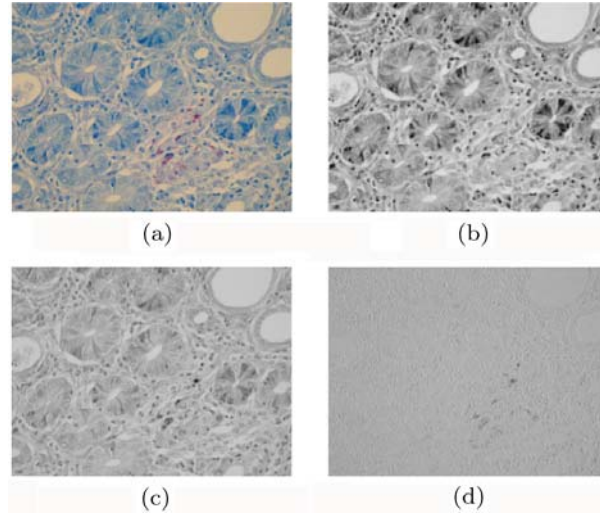

(a)   (b)

So, we start by studying the red, green and blue components of a sample of infected images. Figure 5 shows an infected image (a) and its red (b), green (c) and blue (d) components. The bacteria that appear in the input image (a) are nearly obvious in green and blue components while the red component do not give any indication for the infected region. Red (R), green (G) and blue (B) values were taken for 38 samples from different images, 19 of them represent positive results (+ve, bacteria colour) and the other 19 for negative results (–ve, healthy tissue). Correlation analysis shows that the green and blue components are highly correlated with the class label therefore, the ratio between (G + B) and (R + G + B) is used in the conversion formula that will be discussed later.

Figure 6 shows that the two classes are more separable in green and blue components than in the red component. The following conversion formula was used to convert the coloured images into grey ones, Figure 7(b).

$$\text{Grey} = \frac{(\alpha \times G + \beta \times B) \times C}{R + G + B}, \tag{1}$$

where $\alpha = 0.79$ and $\beta = 0.21$, there values were calculated using try-and-error method to increase the discrimination between the two classes. C is a constant that is equal to 510 and is used for scaling.

8

Figure 5 Infected tissue image and its red green and blue components: (a) the input image; (b) red components of image; (c) green components of image and (d) blue components of image (see online version for colours)

Then, grey-level slicing (Gonzalez and Woods, 2008) is performed to the grey pictures to clarify the ROI as shown in Figure 7(c), and then thresholding was used to isolate them using thresholding value $T = 154$, Figure 7(d). The output image is a binary image in which the bacteria are represented with the white pixels.
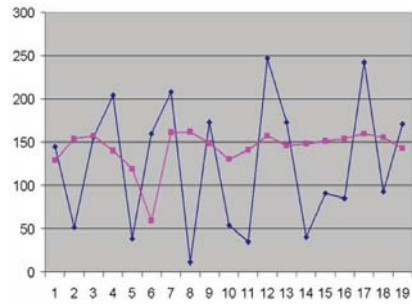
6.2   Morphological operations

After the image has been segmented, we applied some post processing steps like filing the holes of extracted objects, smoothing them, and removing the tiny noisy objects that are believed to be smaller than any bacterium body. Post processing steps is done using morphological operations (Materka and Strzelecki, 1999; Jähne, 2005) which is based on mathematical morphology and rely only on the relative ordering of pixel values, not on their numerical values, they are especially suited to the processing of binary images, Figure 7(e).
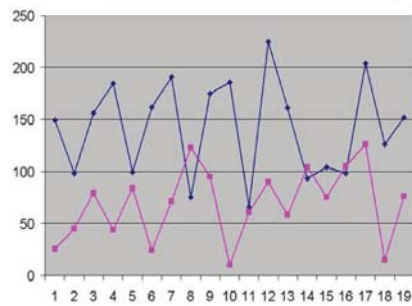
6.3   Feature extraction

Features are important to recognise if input images are healthy or infected. The previously discussed analysis procedures produced binary images with two main forms. In the first form (infected tissue) the binary image contains white coloured objects which represent the bacteria. In the second form (healthy tissue) there is no white objects which indicates there is no bacteria and the tissue is healthy, but in some cases there are some small noisy objects that cannot be removed by the morphological operations discussed previously. Two features were extracted which are the count of white pixels and the number of objects in the image. These two features are used to distinguish between bacteria and noisy objects. In the image with noisy objects the number of white pixels is close to the number of objects, while in the infected images the number of white pixels is much larger than the number of objects.
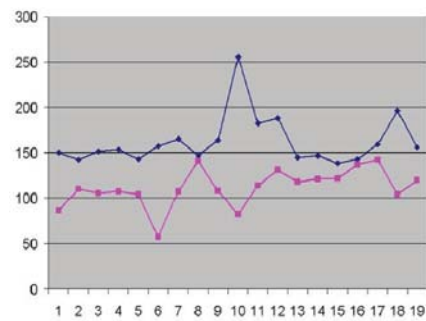
9

Figure 6 Positive (red graph) and negative (blue graph) classes representations: (a) red components of image; (b) green components of image and (c) blue components of image (see online version for colours)
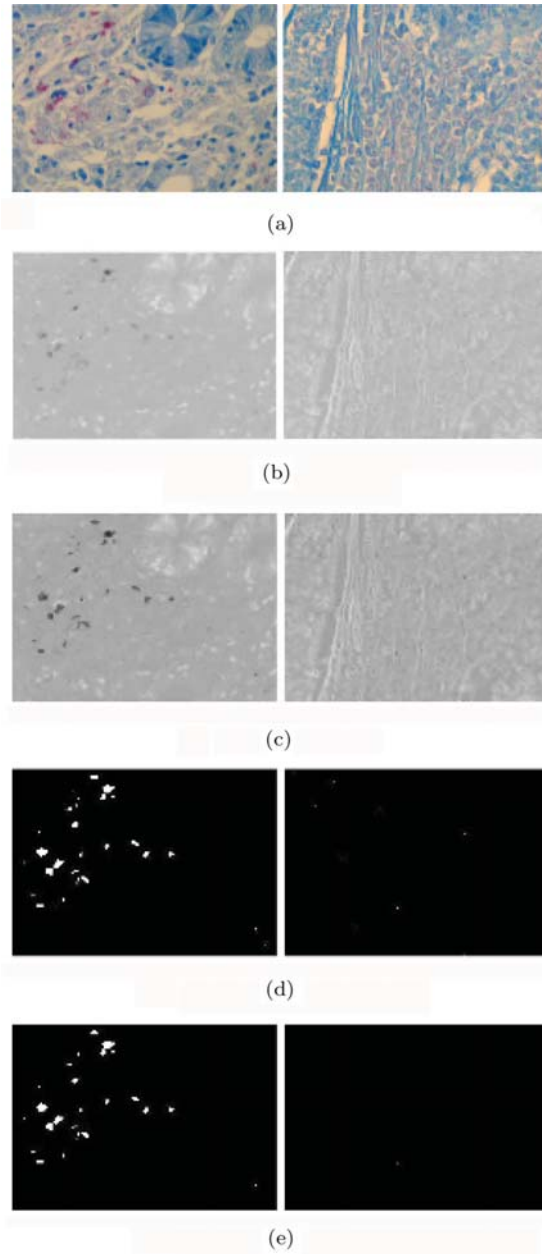

(a)

(b)

(c)

6.4   Classification

In test2 classification, k-nearest neighbour algorithm (k-NN) with k= 3 was used to classify the input images into healthy or infected with Johne's disease. k-NN is a type of instance-based learning, or lazy learning that can be used for classifying objects based on closest k training examples in the feature space. Choosing k-NN classifier was due to the nature of training data, we have only two feature vectors, so there is no need to large storing space. Also, the data is more separable than in the test1. The 100 tuples input data were

normalised to (0–1), and divided into training and testing sets using holdout method. The data were randomly partitioned into two independent sets; training set which includes 2/3 of the data for model construction, and the remaining data is called test set and used for system evaluation.

Figure 7 Test2 main processing steps: (a) image of infected tissue to the left and healthy one to the right; (b) conversion to grey; (c) grey-level slicing; (d) morphological operations and (e) The output binary images (see online version for colours)
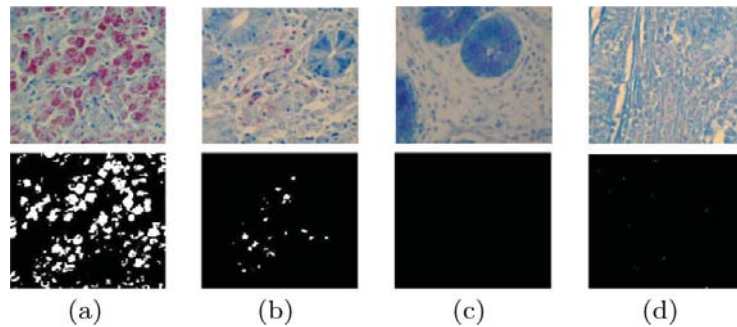

(a)

(b)

(c)

(d)

(e)

7 Results and discussion

7.1 Colour segmentation

The colour region segmentation was used in test2 to isolate the bacteria based on the green and blue colour vectors. Our target is to have an obvious and smooth segmentation, which represent the bodies of the bacteria.
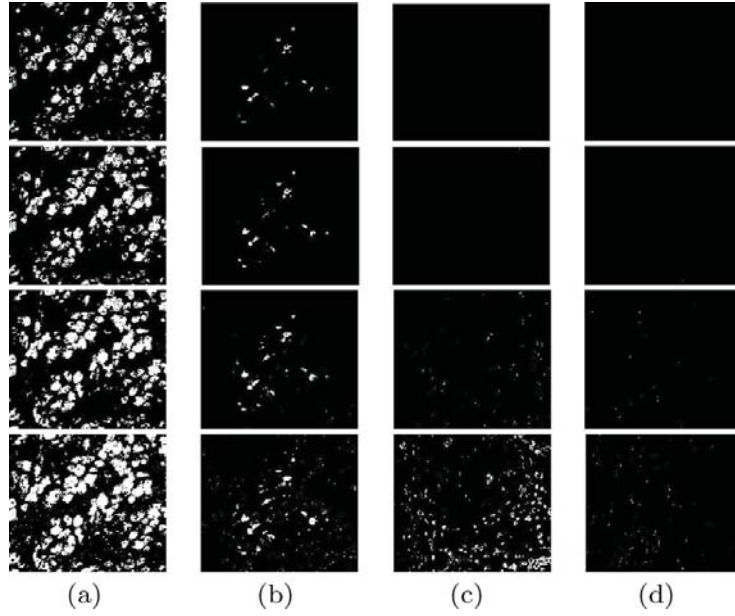
In this section, we will compare our method with the other colour segmentation techniques. A set of four different images were used for this purpose as shown in the upper row of Figure 8, and the lower row represents there output after segmentation process, (a) and (b) are positive images, (c) and (d) are negatives. The ordinary colour segmentation techniques (8) are based on the RGB colour vectors, the segmentation process depends on a threshold value T that represents the distance between the average of samples taken from ROI and a selected point. Figure 9 shows the segmentation results of the four original samples from Figure 8, thresholding values T are chosen to be: 40, 50, 60 and 70 arranged from top to bottom for each image. Images in column (a) correspond to the coloured input image in Figure 8(a), images in column (b) correspond to the coloured input image in Figure 8(b), and so on. 19 RGB values were taken from the bacteria colour, the average vector was calculated which is equal to (143 67 111) for the red, green and blue components respectively. Images segmented using T = 40 and 50 are smooth and clean for all images. Also, it is perfect for the negative images (column c and d) because there is no any white objects, but in positive images (column a and b) it could not capture all the bacterium body. Using the values $T = 60$ and 70 is more effective for capturing and representing the bacteria bodies but it is very sensitive to noise in both positive and negative images. We observe that the bacteria are clear in positive images but with a large amount of noise. On the other hand, it is not acceptable results for the negative images because of the white objects and noise that should not appear.

Figure 8 Samples of images and there segmentation output using our proposed system. The input images (a), (b), (c), and (d) are in the upper row. The segmentation outputs are in the lower row (see online version for colours)

The proposed system solves the previous problems, because the segmentation allows smooth and continuous regions to be isolated with relatively little amount of noise as shown in Figure 8.

Figure 9 The effect of thresholding value T on segmentation. Images (a), (b), (c), and (d) are the segmentation output for images (a), (b), (c), and (d) from Figure 8 respectively. From top to bottom T values are (40, 50, 60, and 70)

(a)   (b)   (c)   (d)

7.2 Histopathological examination test classification

As discussed previously, a total of 192 texture features were extracted for the histopathological examination, but it is too complicated to use all the 192 features as an input for the classification system. Two techniques were used for dimensionality reduction, in order to choose the most correlated and discriminative features. Also, those methods remove the redundant and irrelevant features. The first measure is the FDR, which can be described as following:

$$\text{FDR} = \frac{(\mu_1 - \mu_2)^2}{\sigma_1^2 + \sigma_2^2}. \tag{2}$$

Fisher's Discriminant Ratio is a separability measure between two classes using one-dimensional feature vector. The other measure is the correlation which is calculated between each feature and the class labels attribute, the correlation is an indication for the strength and direction of the linear relationship between each feature attribute and the class label attribute. Table 1 shows the absolute correlation and the FDR values for the highest 15 features. The two used measures give the same order for the selected features. A rank order is given for each feature depending on its order.

Table 1 Absolute correlation and FDR for 15 texture features

Absolute correlation	FDR	Feature no.	Rank
0.72723	2.1216	8	1
0.71302	1.9746	164	2
0.70581	1.8947	175	3
0.70543	1.8915	153	4
0.70161	1.849	5	5
0.70094	1.8459	142	6
0.69116	1.8032	2	7
0.69071	1.7973	185	8
0.67961	1.7081	3	9
0.67862	1.6592	6	10
0.67154	1.5808	7	11
0.66641	1.5583	45	12
0.66587	1.555	39	13
0.66551	1.5535	33	14
0.66481	1.5493	27	15

The next step is to find the best subset features of dimension N to be used in the classification process, we considered the accuracy of the system as a measure to determine the number of features that gives the most accurate results. Figure 10 represents the relationship between the number of features used in the classification system and the accuracy of the results. It can be noticed that the best accuracy was achieved using 8 features as described in Table 2. From Table 2, we observe that the histogram, run-length and Fourier spectrum features can be used in the histopathological examination. Average contrast (μ), uniformity (or energy E) and the kurtosis (μ_4) are the effective features from the histogram analysis which can be extracted using the following formulas:

$$\mu = \sum_{i=0}^{L-1} i \times p(i) \tag{3}$$

$$E = \sum_{i=0}^{L-1} [p(i)]^2 \tag{4}$$

$$\mu_4 = \sigma^{-4} \sum_{i=0}^{L-1} (i - \mu)^4 \times p(i) \tag{5}$$

while only the run length non-uniformity (RLN) in horizontal, vertical, $45°$ and $135°$ directions was selected from the run-length features.

$$RLN = \sum_{j=1}^{N} \left(\sum_{i=1}^{M} p(i,j) \right)^2. \tag{6}$$

Also, the mean of $S(r)$ from the Fourier spectrum features was a discriminant feature.

$$S(r) = \sum_{\theta=0}^{\Pi} S_\theta(r). \tag{7}$$

The other features, wavelets and co-occurrence matrix features were discarded because they are irrelevant to the class label. This reduction process helps in reducing the dimensionality of the features from 192 into only 8 discriminative features.

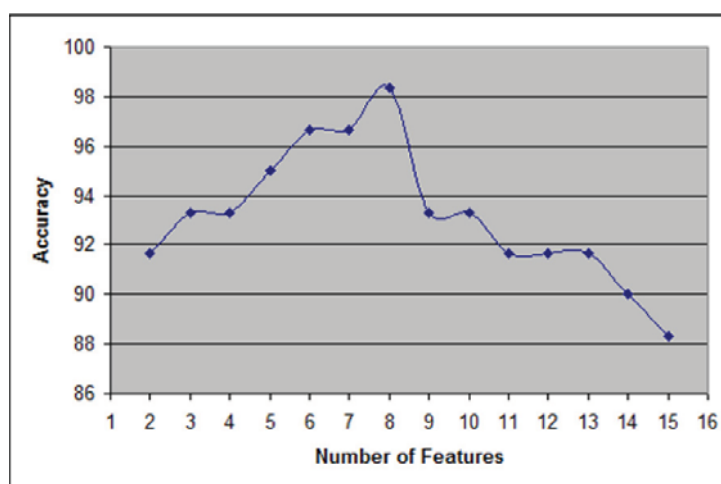**Figure 10** Relation between the accuracy and the number of features (see online version for colours)

Table 2 The best 8 features used for the histopathological examination

No.	Texture feature	Rank order
Histogram features		
2	Average contrast	7
5	Uniformity	5
8	Kurtosis	1
Run-length features		
142	RLN ($0°$ direction)	6
153	RLN ($45°$ direction)	4
164	RLN ($90°$ direction)	2
175	RLN ($135°$ direction)	3
Fourier spectrum features		
185	Mean of $S(r)$	8

Artificial neural network (Jain et al., 1996; Rich and Knight, 1991) was used for the classification step for two main reasons. First, there is an overlapping between negative and positive classes, the ANN can closely approximate any function and perform nonlinear and complicated classification. The other reason is its ability for High tolerance to noisy data. Using hold out method (Han, 2005), the over all 194 images were divided randomly into two sets, 134 images (2/3 of the data) for the training set which used for ANN training and construction, the remainder 60 (1/3 of the data) images were used as testing set for system evaluation. The 194 images include 92 infected (positive) and 102 healthy (negative) images. For testing set, 27 images are positive and 33 are negative.

And to endorse the selection of ANN in our implementation, several classifiers were tried, especially those that performed well in previous works on medical diagnosis systems based on image processing such as neural networks, support vector machines, decision trees, and Naive Bays. WEKA tool was used to run the experiments and the results are tabulated in Table 3. The results show that ANN achieves the highest classification accuracy.

Table 3 Accuracy measures of applying several classifiers to the John's disease diagnosis

Classifier	Accuracy
Neural networks	0.9833
Support vector machines	0.9333
Decision trees	0.9666
Naive Bays	0.95

8 Conclusions

In this paper, a system to diagnose Johne's disease is implemented. Texture analysis combined with ANN proved its effectiveness to distinguish between healthy and infected tissues of the histopathological examination. An accuracy of 98.33% is accomplished. Since declaring a positive case is followed by expensive measures, the acid fast stain test was used for confirmation of positive cases. The test isolated the bacteria objects successfully, hence it was easier to extract features with maximum separability between positive and negative classes, this test achieved an accuracy of 96.97%.

References

Antani, S., Kasturi, R. and Jain, R. (2002) 'A survey on the use of pattern recognition methods for abstraction, indexing and retrieval of images and video', *Pattern Recognition*, Vol. 35, No. 4, pp.945–965.

Ayele, Y. (2001) 'The transmission and impact of paratuberculosis infection in domestic and wild ruminants', *Veterinary Medicine*, Vol. 46, No. 7, pp.204–205.

Castellano, G., Bonilha, L., Li, L. M. and Cendes, F. (2004) 'Texture analysis of medical images', *Clin Radiol*, Vol. 59, No. 12, pp.1061–1069.

Chen, C.H., Pau, P.F. and Wang, P.S.P. (2005) *Handbook of Pattern Recognition and Computer Vision*, World Scientific, Singapore.

Chu, A., Sehgal, C.M. and Greenleaf, J.F. (1990) 'Use of gray value distribution of run lengths for texture analysis', *Pattern Recogn. Lett.*, Vol. 11, pp.415–420.

Cocito, C., Gilot, P., Coene, M. and Kesel, M. (1994) 'Paratuberculosis', *Clinical Microbiology*, Vol. 7, No. 3, pp.328–345.

Commission, E. (2000) *Possible Links Between Crohns Disease and Paratuberculosis*, Technical Report.

Dasarathy, B.V. and Holder, E.B. (1991) 'Image characterizations based on joint gray level-run length distributions', *Pattern Recogn. Lett.*, Vol. 12, pp.497–502.

Dikshit, A., Wu, D., Wu, C. and Zhao, W. (2005) 'An online interactive simulation system for medical imaging education', *Comput Med Imaging Graph*, Vol. 29, No. 6, pp.395–404.

Gonzalez, R.C. and Woods, R.E. (2008) *Digital Image Processing*, 3rd ed., Prentice-Hall, Inc., Upper Saddle River, NJ, USA.

Gupta, A., Rani, S.M., Agrawal, P. and Gupta, P.K. (2012) 'Sero-prevalence of paratuberculosis (johne's disease) in cattle population of south-western Bangalore using elisa kit', *Open Journal of Veterinary Medicine*, Vol. 2, No. 4, pp.196–200.

Han, J. (2005), *Data Mining: Concepts and Techniques*, Morgan Kaufmann Publishers Inc., San Francisco, CA, USA.

Haralick, R. (1979) 'Statistical and structural approaches to texture', *Proceedings of the IEEE*, Vol. 67, No. 5, pp.786–804.

Jähne, B. (2005) *Digital Image Processing*, 6th ed., Springer, Berlin.

Jain, A.K., Mao, J. and Mohiuddin, K. (1996) 'Artificial neural networks: a tutorial', *IEEE Computer*, Vol. 29, pp.31–44.

Kocioek, M., Materka, A., Strzelecki, M. and Szczypiski, P.D. (2001) 'Wavelet transform derived features for digital image texture analysis', *International Conference on Signals and Electronic Systems*, Vol. 18, No. 21, pp.163–168.

Lew, M.S. (2001) *Principles of Visual Information Retrieval*, Springer-Verlag, London, UK.

Mallat, S.G. (1989) 'Multifrequency channel decompositions of images and wavelet models', *IEEE Transactions on Acoustics Speech and Signal Processing*, Vol. 37, pp.2091–2110.

Materka, A. and Strzelecki, M. (1998) 'Texture analysis methods: a review', *Methods*, Brussels, pp.1–33.

Materka, A. and Strzelecki, M. (1999) 'Scanner resolution and noise influence on texture parameters of magnetic resonance phantom images', *International Conference – Computers in Medicine*, Lodz, Poland, pp.101–107.

Rich, E. and Knight, K. (1991) *Artificial Intelligence*, 2nd ed., McGraw-Hill, Singapore.

Rolf, B., Samira, B., Reinhard, W., Theodor, S. and Hermann, W. (1996) 'Molecular characterization of mycobacterium paratuberculosis isolates from sheep, goats, and cattle by hybridization with a DNA probe to insertion element is900', *Journal of Clinical Microbiology*, Vol. 34, No. 7, pp.1617–1621.

Sechi, L.A., Mura, M., Tanda, F., Lissia, A., Solinas, A., Fadda, G. and Zanetti, S. (2001) 'Identification of mycobacterium avium subsp. paratuberculosis in biopsy specimens from patients with crohn's disease identified by in situ hybridization', *Journal of Clinical Microbiology*, Vol. 39, No. 12, pp.4514–4517.

Theodoridis, S. and Koutroumbas, K. (2006) *Pattern Recognition*, 3rd ed., Academic Press, Inc., Orlando, FL, USA.

Toma, A.S.M. (2008) *An Automatic Intelligent System for Diagnosis of Johne's Disease*, Jordan University of Science and Technology, Jordan.

Xiaoou, T. (1998) 'Texture information in run-length matrices', *IEEE Transactions on Image Processing*, Vol. 7, No. 11, pp.1602–1609.