Automated Periodontal Diseases Classification System

Aliaa A. A. Youssif

Department of Computer Science,
Faculty of Computers and
Information
Helwan University
Cairo, Egypt

Abeer Saad Gawish

Department of Oral Medicine and
Periodontology, Faculty of Dental
Medicine (Girls' Branch)
Al-Azhar University
Cairo, Egypt

Mohammed Elsaid Moussa

Department of Computer Science,
Faculty of Computers and
Information
Helwan University
Cairo, Egypt

Abstract— **This paper presents an efficient and innovative system for automated classification of periodontal diseases, The strength of our technique lies in the fact that it incorporates knowledge from the patients' clinical data, along with the features automatically extracted from the Haematoxylin and Eosin (H&E) stained microscopic images. Our system uses image processing techniques based on color deconvolution, morphological operations, and watershed transforms for epithelium & connective tissue segmentation, nuclear segmentation, and extraction of the microscopic immunohistochemical features for the nuclei, dilated blood vessels & collagen fibers. Also, Feedforward Backpropagation Artificial Neural Networks are used for the classification process. We report 100% classification accuracy in correctly identifying the different periodontal diseases observed in our 30 samples dataset.**

Keywords-Biomedical image processing; epithelium segmentation; feature extraction; nuclear segmentation; periodontal diseases classification.

I. INTRODUCTION

Periodontitis is a chronic inflammatory disease of vascularized supporting tissues of the teeth [1], Periodontal disease occurs when inflammation or infection affect the gingiva and extend to the periodontal apparatus [2], The 1999 classification system for periodontal diseases and conditions listed seven major categories of periodontal diseases [3,4]: Gingivitis, Chronic periodontitis, Aggressive periodontitis, Periodontitis as a manifestation of systemic disease, Necrotizing ulcerative gingivitis/periodontitis, Abscesses of the periodontium, Combined periodontic-endodontic lesions; The latter 4 are associated with systemic diseases [5,6,7,8,9]; Hence, this work preliminary focuses on identifying the different types of periodontal diseases by using computer-assisted microscopy system for automated classification of periodontal diseases to increase the accuracy and reduce the workload in classifying and diagnosis of the different categories of periodontal diseases, which help in designing the treatment plan used.

The paper is organized as follows: Section II explains the materials and methods that we followed to get our dataset, Section III describes the implementation details of our proposed system, Section IV shows experimental results and discussions about these results, and Section V concludes the paper and introduces the future work that can be done.

II. MATERIALS AND METHODS

A. Study Cases

The study was conducted on 32 patients attending between February 2009 and March 2011 to the Oral Medicine, Periodontology, Oral Diagnosis and Radiology Department, Faculty of Dental Medicine Girls' Al-Azhar University.

Patients were suffering from gingival inflammation, which may extend to include the periodontium (different types of periodontitis) or gingival overgrowth (due to different etiological factors).

Patients were excluded from the present study if they were smokers, pregnant or post-menopausal women. All selected patients had not undergone any periodontal therapy for at least six months.

B. Collected Clinical Data

The following clinical parameters were collected and recorded on six sites at each tooth; all linear measurements were recorded to the nearest 0.5 mm using William graduated periodontal probe: Plaque index (PI) [10] Pocket Depth (PD) and Clinical attachment level (CAL) was measured from the CEJ to the apical part of the sulcus, All included patients were indicated for surgical flaps as a line of their treatment.

C. Histopathological Sample Preparation

After the surgical procedures; the excised tissue samples were immersed in 10% formalin and decalcified in multiple baths of 10% trichloroacetic acid. The blocks were immersed in paraffin, and semi-serial 4 μm histologic sections were stained with Haematoxylin and Eosin (H&E).

D. Image Capture

Representative sections were photographed using a Leitz DMRD Microscope (Leica, Wetzlar, Germany) with 20 X objective UPLanFl (resolution 0.67 μm) at a size of 1600 X 1200 pixels (interpixel distance 0.62 μm) using a JVC KY-55B 3-CCD colour camera attached to a 24 bit RGB frame grabber (Imaging Technologies IT4PCI, Bedford,

MA, U.S.A.) controlled by Optimas v. 6.51 (Media Cybernetics, Silver Spring, MD, U.S.A.)

E. Periodontal Diseases Groups

The 32 patients were classified into 4 diagnostic groups according to their periodontal status; 16 patients were classified into Gingival Enlargement group, 7 patients into Chronic Gingivitis group, another 7 patients into Chronic Periodontitis group, and the last 2 patients into Aggressive Periodontitis group.

We excluded the patients which were classified into the Aggressive Periodontitis group from our classification experiments; because of the very small number of study cases that we have for this group; which was not sufficient for our neural networks based classification system to work with. Although we included these 2 cases in our preprocessing and feature extraction experiments, to make sure that our proposed system is generic enough to handle all of the different types of diseases.

We also included another study case for a healthy person to our preprocessing and feature extraction experiments.

III. PROPOSED SYSTEM

Our proposed system can be divided into several well-defined stages as illustrated in Figure 1. We first get the H&E stained slide, do some preprocessing steps to it to remove background and undesired objects, segment it to the tissue's main components epithelium and connective tissue, then extract features from both parts, and then use these features along with the clinical data of the patient to start the classification process.
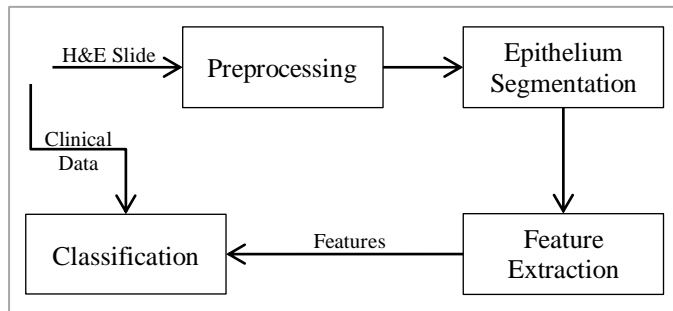

Figure 1 Schematic structure of the proposed system

A. Pre-processing

Mainly images consist of 2 parts: Tissue (Epithelium + Connective Tissue) and Background. In this preprocessing phase, we need to remove the background areas from the image, allowing the further processing to be done only on the tissue part.

We implemented 2 different algorithms to achieve this task; the first one is fully-automated, while the second one is semi-automated because it requires some clicks from the user to specify the seed points in the background area. The pseudo codes for fully-automated and semi-automated algorithms are shown below in Pseudocode I & Pseudocode II respectively. 0shows one sample result of the 2 algorithms. Also, sensitivity and specificity results of these algorithms are displayed in the Results and Discussion section.

Pseudocode I   FULLY-AUTOMATED PREPROCESSING ALGORITHM

1. Get the "a" channel from L*a*b* color space of the original image, $Image_1$.
2. Smoothing of $Image_1$ with a 5 · 5 average filter to preserve only large detail [11], $Image_2$.
3. Convert $Image_2$ to black & white with 0.2 threshold value, and then invert it, $Mask_1$.
4. Morphologically open $Mask_1$ to remove all black objects that have fewer than 1000 pixels, $Mask_2$.
5. Morphologically open $Mask_2$ to remove all white objects that have fewer than 1000 pixels, $Mask_3$.
6. Perform a morphological dilatation with a 10 · 10 window size on the $Mask_3$, $Mask_4$.
7. Fill Holes of $Mask_4$, $Mask_5$.
8. Perform a morphological erosion with a 10 · 10 window size on the $Mask_5$, $Mask_6$.
9. Remove areas from $Mask_6$ that are not on border [12], $Mask_7$.
10. Morphologically open $Mask_7$ to remove all objects that have more than 10000 pixels, $Mask_8$.
11. Apply final smoothing for $Mask_8$, $Mask_9$.

Pseudocode II   SEMI-AUTOMATED PREPROCESSING ALGORITHM

1. Get the "Green" channel of the original image, $Image_1$.
2. Get Seed Point from the pathologist, inside the background region, then set the initial threshold to the intensity value of this seed point.
3. Based on the initial threshold, get all the pixels that its intensity in a range of 50 values, $Mask_1$, i.e.
   abs(newPixelValue - initalThreshold) <= 50
4. Perform image reconstruction on the $Mask_1$ to get only the pixels that are connected to the seed point, $Mask_2$.
5. Perform a morphological erosion with a 1 · 1 window size on the $Mask_2$, $Mask_3$.
6. Morphologically open $Mask_3$ to remove all white objects that have fewer than 1000 pixels, $Mask_4$.
7. Perform a morphological dilation with a 2 · 2 window size on the $Mask_4$, $Mask_5$.
8. Morphologically open $Mask_5$ to remove all black objects that have fewer than 1000 pixels, $Mask_6$.
9. Perform a morphological erosion with a 1 · 1 window size on the $Mask_6$, $Mask_7$.
10. Repeat steps 2 to 10 for each separate background region.
11. Edge shifting for the $Mask_7$ by performing morphological erosion with a disk-shaped structuring element [13], [14] with radius 2 the $Mask_7$, $Mask_8$.

For very small background regions, we skipped steps from 6 to 10 for these regions only. This happened only 3 times over our 33 images.
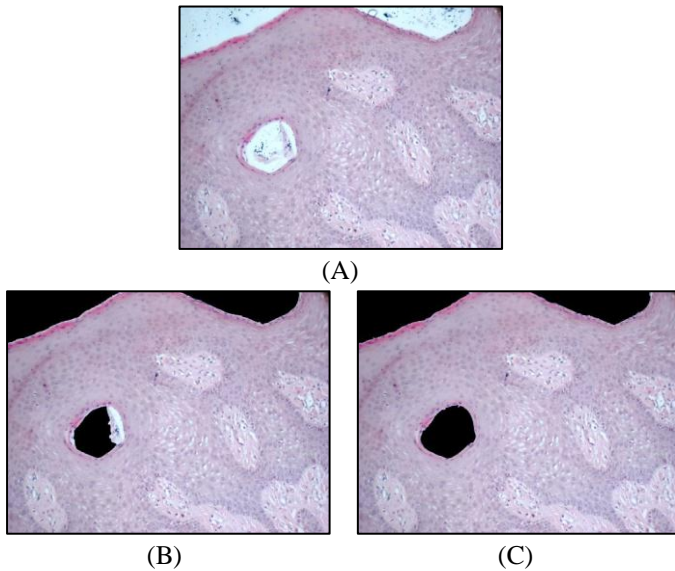

(A)

(B)   (C)

Figure 2 Pre-processing sample (A) The original image. (B) After applying fully-automated algorithm. (C) After applying semi-automated algorithm.

B. Epithelium Segmentation

In this phase of our system, we need to segment the tissue into Epithelium and Connective Tissue; Epithelium segmentation was targeted in some studies [15,16,17]. As stated in [15] there are a lot of challenges doing this task, and they are mainly caused by staining artifacts, lighting acquisition conditions, and undesired touching objects. Some studies [15] used the saturation channel in the HSV color space, others [16] used multiple gray level automated thresholding over the image's green channel, while others [17] used region growing algorithm [18] over the gray channel of the image.

We found that 73% of our samples give better results when using the HSV Saturation channel, while the remaining 27% give better results when using the RGB green channel. We implemented an automated way for this segmentation process; Pseudocode III describes our epithelium segmentation algorithm.

We also implemented a post-processing step, to move some parts from epithelium to connective tissue or vice versa manually, because our proposed algorithm didn't achieve highly accurate results for all study cases, because the brightness value of the infiltrate was similar to that of the epithelium in images with high inflammatory infiltrate [16]. Information about sensitivity and specificity of our algorithm is found in the Results and Discussion section. Figure 3 shows one sample result of this algorithm.

C. Feature Extraction

We extracted a set of features from our Haematoxylin and Eosin (H&E) stained slides, that we thought it will be helpful in the classification process later. Below are the extracted features along with the algorithms we used to extract it:

Pseudocode III   EPITHELIUM SEGMENTATION ALGORITHM

1. Get the "Saturation" channel from the HSV color space of the tissue image [a], $Image_1$, which retrieved from the preprocessing phase, $Channel_1$.
2. Adjust [b] intensity values in the grayscale $Channel_1$, to map it to new values in such that 1% of data is saturated at low and high intensities of the $Channel_1$. This increases the contrast of the output channel, $Channel_2$. [15].
3. Apply $10 \cdot 10$ median filter [c] for removing the noise in the $Channel_2$ without harming edges, $Channel_3$. [19]
4. Convert $Channel_3$ to black & white by applying global thresholding through Otsu's method [20], $Mask_1$. [d]
5. Fill Holes of $Mask_1$, $Mask_2$.
6. Perform a morphological erosion with a $10 \cdot 10$ window size on the $Mask_2$, $Mask_3$.
7. Morphologically open $Mask_3$ to remove all white objects that have fewer than 1000 pixels, $Mask_4$.
8. Perform a morphological dilation with a $20 \cdot 20$ window size on the $Mask_4$, $Mask_5$.
9. Morphologically open $Mask_5$ to remove all black objects that have fewer than 1000 pixels, $Mask_6$.
10. Perform a morphological erosion with a $10 \cdot 10$ window size on the $Mask_6$, $Mask_7$.
11. Edge shifting for the mask by performing a morphological erosion with a strel $2 \cdot 2$ window size on the $Mask_7$, $Mask_8$. [13], [14].

[a] For 9 samples of our 33 samples, we used RGB Green channel instead of HSV Saturation channel, because we found it giving better results.

[b] For 1 sample, we replaced image adjustment step with histogram equalization.

[c] For 6 samples, we used a $20 \cdot 20$ window size for median filtering instead of the $10 \cdot 10$ window size, and for another 1 sample we used a $9 \cdot 9$ window size.

[d] For 12 samples, we applied histogram equalization after step 4.

1) Epithelium Percentage

After segmentation of the tissue image into its components (Epithelium and Connective Tissue) in the previous phase, we calculated the percentage of the Epithelium part.

(Note that the background regions are already excluded)

2) Nuclei in Epithelium

We segmented the nuclei that found in the epithelium part, and used its count and percentage from the whole epithelium in our features list, Nuclei segmentation was targeted in some studies such as [11], [21], [22] which were using nuclear localization achieved using the colour deconvolution algorithm developed by [23] to obtain the optical density of the Haematoxylin stain alone plus spatial partition of the epithelial compartment representing the exclusive area of influence of each nucleus profile, [24] which is based on the multiscale Laplacian-of-Gaussian (LoG) filter, [25] which depends on ellipse fitting and the watershed transform, [26] which relies on some morphology-based techniques.
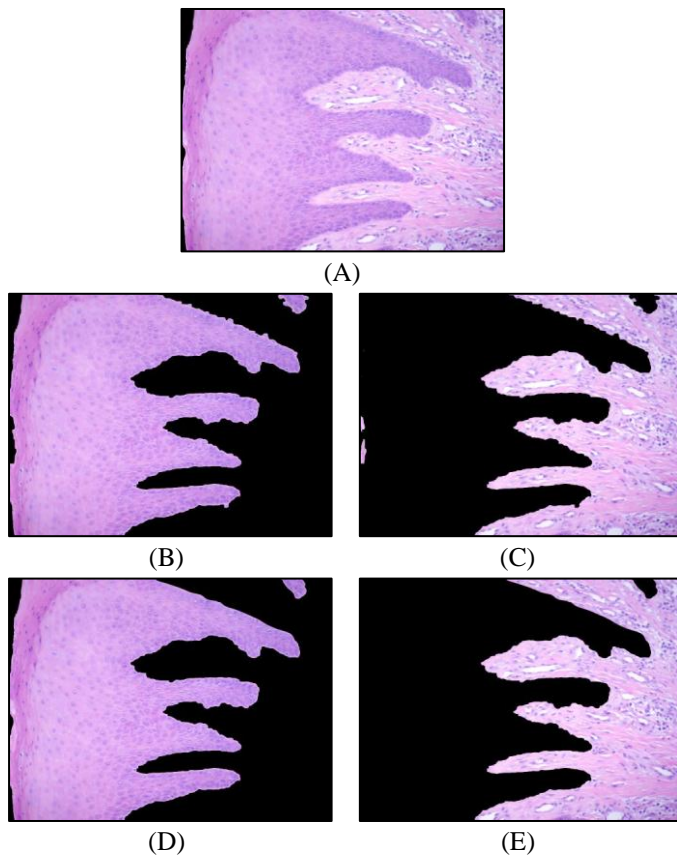
(A)

(B)

(C)

(D)

(E)

Figure 3 Epithelium Segmentation (A) Original tissue image after pre-processing phase. (B) Automatically segmented epithelium. (C) Automatically segmented connective tissue. (D) Epithelium after post-processing. (E) Connective-Tissue after post-processing.

In our proposed system, we used steps in Pseudocode IV to segment the nuclei in the epithelium part of the image; Figure 4 shows two sample results of this algorithm.

3) Nuclei in Connective Tissue

We segmented the nuclei that found in the connective tissue component, and used its count and percentage from the whole connective tissue in our features list.

We do this by getting 2 masks; the first one represents the light areas in the connective tissue (blood vessels) while the second one represents the dark areas (nuclei). Then we need to filter the nuclei mask to exclude the ones that are related to the dilated blood vessels. Our steps for doing this segmentation part are described in Pseudocode V. Figure 5 shows one sample result of this algorithm.

4) Dilated Blood Vessels

We segmented the dilated blood vessels that found in the connective tissue part, and used its count and percentage from the whole connective tissue in our features list.

We do this with almost the same steps described in the Pseudocode V that were used in the connective tissue nuclei segmentation. We get the first mask with exactly the same steps, while excluding steps e, f and j from the ones required to get the second mask. And to get the final mask, we filter parts in $Mask_1$ (dilated blood vessels candidates) to only the

ones that intersect at least 3 times with parts of $Mask_2$ (nuclei); finally, we count and get the percentage of all objects in $Mask_3$. 0Figure 6 shows one sample result of this algorithm.

Pseudocode IV   EPITHELIUM NUCLEI SEGMENTATION ALGORITHM

1. Get the optical density of the Haematoxylin as an 8 bit grayscale image, by colour deconvolution [23] of the epithelium part, retrieved from the epithelium segmentation phase, $Channel_1$. [a]
2. Normalize intensity values in the grayscale $Channel_1$, to change the range of pixel intensity values to be between 0 and 255, $Channel_2$. [b]
3. Convert $Channel_2$ to black & white by computing the extended-minima transform, which is the regional minima of the H-minima transform [27], with a value of 30, $Mask_1$.
4. Fill Holes of $Mask_1$, $Mask_2$.
5. Remove areas from $Mask_2$ that are not on border, $Mask_3$.
6. Morphologically open $Mask_3$ to remove all objects that have less than 15 pixels, $Mask_4$.
7. Morphologically open $Mask_4$ to remove all objects that have more than 800 pixels, $Mask_5$.
8. Convert each connected object to the ellipse that fit it, $Mask_6$.
9. Finally we count and get the percentage of all objects in $Mask_6$.

[a] For 9 samples of our 33 samples, we used RGB Red channel instead of Haematoxylin channel, because we found it giving better results.

[b] For 3 samples, we also apply image adjustment after step 2.
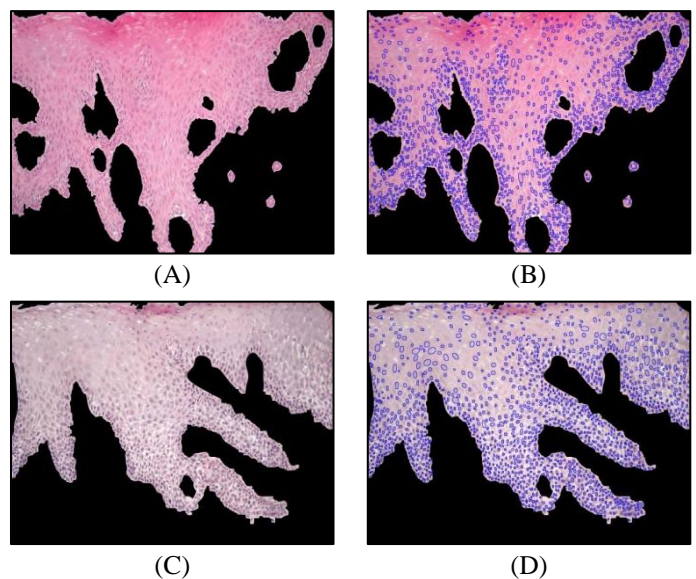

(A)

(B)

(C)

(D)

Figure 4 Nuclei in Epithelium. (A), (C) Two samples of epithelium images. (B), (D) After applying epithelium nuclei segmentation algorithm over (A) and (C) respectively. [Nuclei are surrounded by blue ellipses].

Pseudocode V   CONNECTIVE-TISSUE NUCLEI SEGMENTATION ALGORITHM

1. Get $Mask_1$
   a. Get the "Green" channel of the connective tissue part, retrieved from the epithelium segmentation phase, $Channel_1$.
   b. Enhance the contrast of the grayscale channel $Channel_{1\_1}$ by transforming the values using contrast-limited adaptive histogram equalization (CLAHE) [28], $Channel_{1\_1}$.
   c. Apply 3 • 3 median filter for removing the noise [19] in the $Channel_{1\_1}$ without harming edges, $Channel_{1\_2}$.
   d. Convert $Channel_{1\_2}$ to black & white with 0.8 threshold value, $Mask_{1\_1}$.
   e. Morphologically open $Mask_{1\_1}$ to remove all objects that have less than 80 pixels, $Mask_1$.
2. Get $Mask_2$
   a. Get the optical density of the Haematoxylin as an 8 bit grayscale image, by colour deconvolution [23] of the connective tissue part, retrieved from the epithelium segmentation phase, $Channel_2$.
   b. Adjust intensity values in the grayscale $Channel_2$, to map it to new values in such that 1% of data is saturated at low and high intensities of the $Channel_2$. This increases the contrast of the output channel, $Channel_{2\_1}$. [15].
   c. Apply 2-D adaptive filtering with a 5 · 5 window size over $Channel_{2\_1}$, $Channel_{2\_2}$.
   d. Convert $Channel_{2\_2}$ to black & white by computing the extended-minima transform, which is the regional minima of the H-minima transform [27], with a value of 30, $Mask_{2\_1}$.
   e. Morphologically open $Mask_{2\_1}$ to remove all objects that have less than 15 pixels, $Mask_{2\_2}$.
   f. Morphologically open $Mask_{2\_2}$ to remove all objects that have more than 800 pixels, $Mask_{2\_3}$.
   g. Apply watershed algorithm for $Mask_{2\_3}$, $Mask_{2\_w}$.
   h. Perform a morphological dilation with a 10 • 10 window size on the $Mask_{2\_3}$, $Mask_{2\_4}$.
   i. Apply AND operator between $Mask_{2\_4}$ and $Mask_{2\_w}$, $Mask_{2\_5}$.
   j. Remove areas from $Mask_{2\_5}$ that are not on border, $Mask_2$.
3. Get Final Mask
   a. Filter parts in $Mask_2$ (Nuclei) to only ones that not intersect with parts of $Mask_1$ (Light Parts) that intersects with at least 3 times with parts of $Mask_2$, $Mask_{3\_1}$.
   b. Apply image reconstruction algorithm between $Mask_{3\_1}$ and $Mask_{2\_3}$, $Mask_{3\_2}$.
   c. Convert each connected object in $Mask_{3\_2}$ to the ellipse that fit it, $Mask_3$.
4. Finally we count and get the percentage of all objects in $Mask_3$.
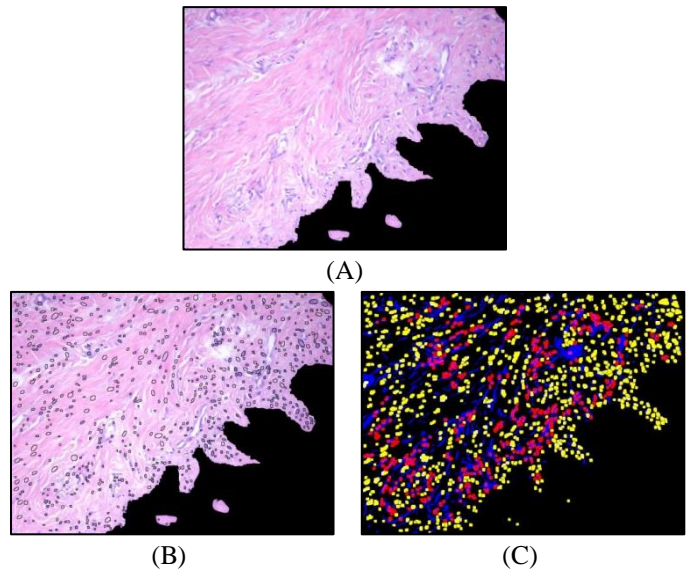

(A)

(B)   (C)

Figure 5 Nuclei in Connective Tissue. (A) Sample of connective tissue image. (B) Nuclei in connective tissue surrounded by black ellipses. (C) $Mask_1$ is in blue, $Mask_2$ is in red and yellow, and Final Mask is in yellow.

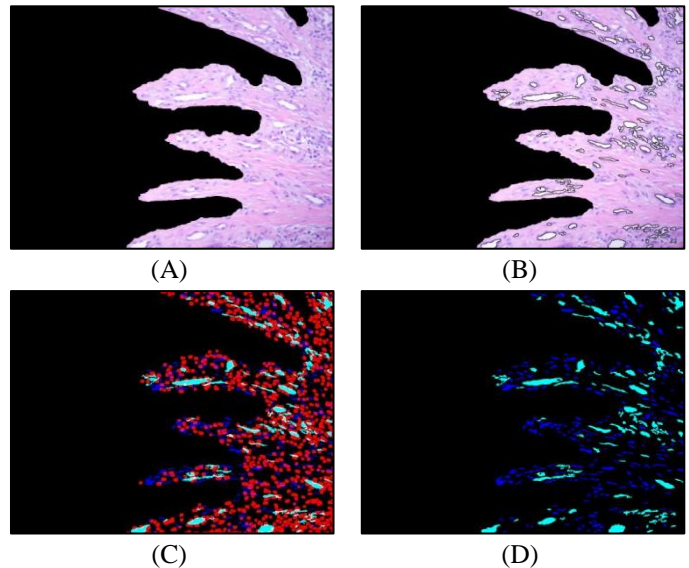(A)   (B)

(C)   (D)

Figure 6 Dilated Blood Vessels. (A) Sample of connective tissue image. (B) Dilated Blood Vessels in connective tissue surrounded by black ellipses. (C) Candidate blood vessels are in blue and light blue. Red areas are nuclei. (D) Candidate blood vessels are in blue and light blue. Final Mask (passed blood vessels) is in light blue.

5) Collagen Fiber

We segmented the collagen fiber that found in the connective tissue part, and used its percentage from the whole connective tissue in our features list.

We do this by getting 2 masks, the first one represents the light areas in the connective tissue, and the second one

represents the dark areas, then we need to filter the nuclei mask to exclude the ones that are related to the dilated blood vessels. Our steps for doing this segmentation part are described in Pseudocode VI. 0Figure 7 shows one sample result of this algorithm.

Pseudocode VI	COLLAGEN FIBER SEGMENTATION ALGORITHM

1. Get $Mask_1$
   a. Get the "Green" channel of the connective tissue part, retrieved from the epithelium segmentation phase, $Channel_1$.
   b. Enhance the contrast of the grayscale channel $Channel_{1\_1}$ by transforming the values using contrast-limited adaptive histogram equalization (CLAHE) [28], $Channel_{1\_1}$.
   c. Apply 3 • 3 median filter for removing the noise [19] in the $Channel_{1\_1}$ without harming edges, $Channel_{1\_2}$.
   d. Convert $Channel_{1\_2}$ to black & white with 0.75 threshold value, $Mask_1$.
2. Get $Mask_2$
   a. Get the optical density of the Haematoxylin as an 8 bit grayscale image, by colour deconvolution [23] of the connective tissue part, retrieved from the epithelium segmentation phase, $Channel_2$.
   b. Adjust intensity values in the grayscale $Channel_2$, to map it to new values in such that 1% of data is saturated at low and high intensities of the $Channel_2$. This increases the contrast of the output channel, $Channel_{2\_1}$. [15].
   c. Apply 2-D adaptive filtering with a 5 · 5 window size over $Channel_{2\_1}$, $Channel_{2\_2}$.
   d. Convert $Channel_{2\_2}$ to black & white by applying global thresholding algorithm using the value got from Otsu's method [20] multiplied by 1.1, and then invert it, $Mask_2$.
3. Get Final Mask by excluding $Mask_1$ and $Mask_2$ from the original connective tissue area, $Mask_3$.
4. Finally we get the percentage of all objects in $Mask_3$.

D. Classification

After extracting all the features, we used Feedforward Backpropagation Artificial Neural Networks in the classification process.

We used 11 inputs; 3 clinical data (Plaque Index, Pocket Depth, and Attachment Level) and 8 extracted features (Epithelium Percentage, Epithelium Nuclei Count & Percentage, Connective Tissue Nuclei Count & Percentage, Dilated Blood Vessels Count & Percentage, and Collagen Fiber Percentage), 10 hidden neural networks, and 3 outputs representing our 3 diseases groups (Gingival Enlargement, Chronic Gingivitis, and Chronic Periodontitis); the used neural network model is shown in Figure 8.
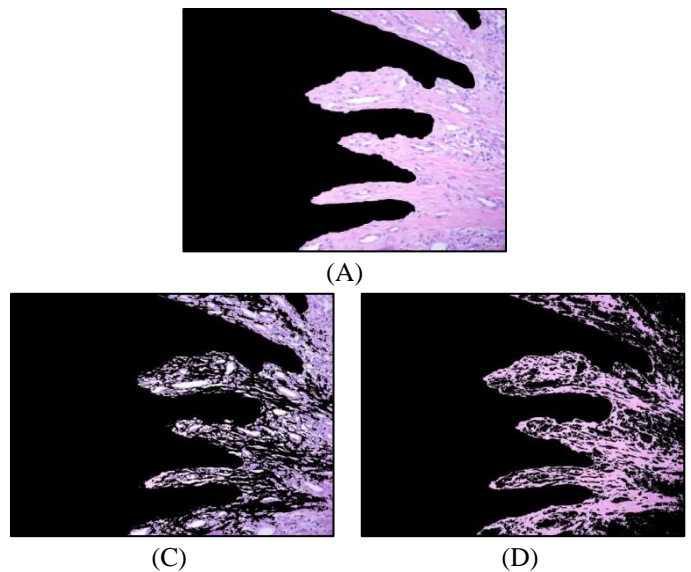

(A)

(C)   (D)

Figure 7 Collagen Fiber (A) Sample of connective tissue image. (B) Mask1 + Mask2. (C) Final Mask (Collagen Fiber).

We divided our samples randomly into 3 groups, 20 samples for the training set which were presented to the neural network during training, and the network was adjusted according to its error, 5 samples for the validation set which were used to measure network generalization, and to halt training when generalization stops improving, and the last 5 samples for the testing set which had no effect on training so providing an independent measure of network performance during and after training.

We wrote some algorithm to make sure that each data set contains a relative number of study cases for each disease group. We used scaled conjugate gradient backpropagation function for network training. Also, we used mean squared error as our performance function.
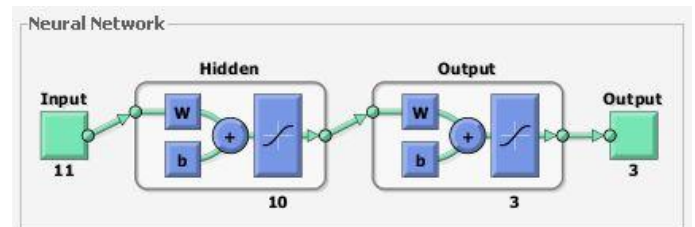

Figure 8 Neural Network Model

IV. RESULTS AND DISCUSSION

For background removal preprocessing phase, we applied both fully-automated and semi-automated preprocessing algorithms described in Pseudocode I and Pseudocode II respectively. We considered the semi-automated results as the ground truth, and compared the results of the fully-automated algorithm to it; over our 33 samples, we achieved a sensitivity of 68.42% and specificity of 98.56%. 6 samples of our samples actually contain no background at all, and when removing their results from our statistics, we achieved a sensitivity of 83.33% and specificity of 98.66% over our remaining 27 samples.

For epithelium segmentation phase, we applied the algorithm described in Pseudocode III over our date, and since the results of this step has to be accurate since it is used in all later phases, we also implemented a manual way for moving parts from the epithelium part to the connective tissue part or vice versa. We considered the results after manual processing the ground truth, and found that our algorithm achieved a sensitivity of 84.99% and specificity of 88.40% over our 33 samples dataset, below are some statistics about most notable results:

- 16 Samples (48.5%) give Sensitivity >= 95%, 15 Samples (45.5%) give Specificity >= 95%, and 11 Samples (33.3%) give Sensitivity & Specificity >= 95%.

- 16 Samples (48.5%) give Sensitivity >= 90%, 20 Samples (60.6%) give Specificity >= 90%, and 13 Samples (39.4%) give Sensitivity & Specificity >= 90%.

- 5 Samples (15.2%) give Sensitivity <= 60%, 2 Samples (6.1%) give Specificity <= 60%, and 2 Samples (6.1%) give Sensitivity & Specificity <= 60%.

For nuclei in epithelium segmentation phase, we applied the algorithm described in Pseudocode IV over our data; we found that optical density of the Haematoxylin stain retrieved through colour deconvolution [23] is the best channel for nuclei segmentation in 24 samples (73%) while the RGB Red channel is the best for the remaining 9 samples (27%). Information regarding average (count, area, and percentage) of nuclei found in the epithelium for each disease group can be found in Table I.

TABLE I.   EPITHELIUM NUCLEI SEGMENTATION STATISTICS.

Group	No. of Samples	Avg. Count	Avg. Area	Avg. Percentage
Normal	1	826	117887	10.31
Gingival Enlargement	16	916	108013	09.46
Chronic Gingivitis	7	1359	131323	12.41
Chronic Periodontitis	7	1020	125925	11.96
Aggressive Periodontitis	2	1284	125068	10.44

For nuclei in connective tissue segmentation phase, we applied the algorithm described in Pseudocode IV over our data; Information regarding average (count, area, and percentage) of nuclei found in connective tissue for each disease group can be found in Table II.

TABLE II.   CONNECTIVE TISSUE NUCLEI SEGMENTATION STATISTICS.

Group	No. of Samples	Avg. Count	Avg. Area	Avg. Percentage
Normal	1	413	27722	03.65
Gingival Enlargement	16	409	22835	03.07
Chronic Gingivitis	7	438	26042	03.06
Chronic Periodontitis	7	510	30459	03.45
Aggressive Periodontitis	2	444	19670	02.92

For dilated blood vessels segmentation phase, we found the following statistics that are described in Table III.

TABLE III.   DILATED BLOOD VESSELS SEGMENTATION STATISTICS.

Group	No. of Samples	Avg. Count	Avg. Area	Avg. Percentage
Normal	1	92	64005	08.29
Gingival Enlargement	16	94	75001	10.57
Chronic Gingivitis	7	77	68666	08.90
Chronic Periodontitis	7	82	65443	08.53
Aggressive Periodontitis	2	117	84809	12.61

For collagen fiber segmentation phase, we applied the algorithm described in Pseudocode VI over our data; Information regarding average (area, and percentage) of collagen fibers for each disease group can be found in Table IV.

TABLE IV.   COLLAGEN FIBER SEGMENTATION STATISTICS.

Group	No. of Samples	Avg. Area	Avg. Percentage
Normal	1	347325	45.72
Gingival Enlargement	16	423448	58.54
Chronic Gingivitis	7	383753	51.01
Chronic Periodontitis	7	404528	51.03
Aggressive Periodontitis	2	373231	56.24

For our classification results; Figure 9 shows our neural network performance, Figure 10 0shows the confusion matrix for the training dataset, validation dataset, test dataset, & the whole dataset, while 0Figure 11 shows Receiver Operating Characteristic ROC for them.
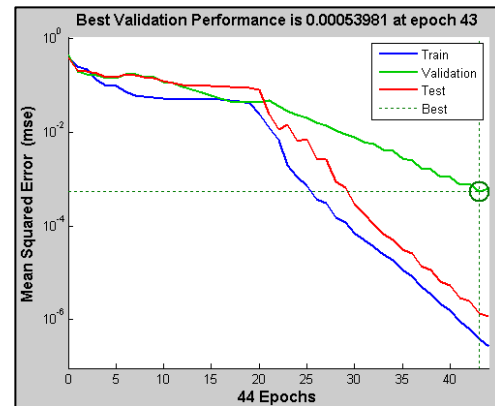

Figure 9 Neural Network Training Performance

I.   CONCLUSION AND FUTURE SCOPE

An automated system has been developed for classification of periodontal diseases using H&E stained microscopic images of the tissues around the affected teeth along with clinical data. The epithelium percentage of the whole tissue, count & percentage of nuclei in epithelium & connective tissue, dilated blood vessels count & percentage, and collagen fiber percentage are used as features during the classification process which is done using Feedforward Backpropagation Artificial Neural Networks. It was found that using these mixed features together achieve a more accurate classification results than using only clinical data or H&E stained images' extracted features.
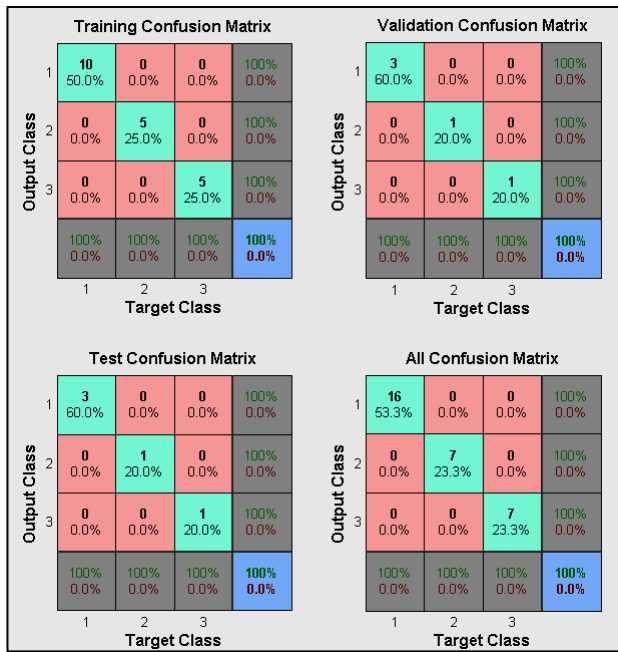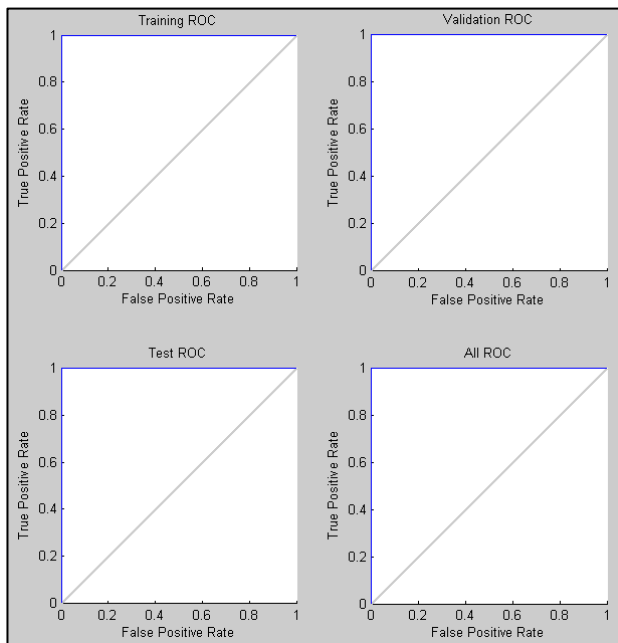
Figure 10 Neural Network Training Confusion Matrix

Figure 11 Neural Network Training Receiver Operating Characteristic ROC

We suggest the following as a future work for this study:

1) Get a larger dataset of diseases, to provide better training and testing of the proposed system, also get test cases in the Aggressive Periodontitis group to include them in the training and testing phases of the program to make it as generic as possible for all periodontal diseases.

2) Enhance the fully-automated background removal algorithm found in Pseudocode I, to achieve similar results to the semi-automated one found in Pseudocode II. Also, enhance the epithelium segmentation algorithm

found in Pseudocode III, to achieve higher Specificity and sensitivity results.

3) Try to find more generic algorithms' parameters, to decrease the number of study cases that we had to alter its parameters to get better results; as shown in the footnotes of Pseudocode III and Pseudocode IV.

REFERENCES

[1] D. V, Prapulla, P. B. Sujatha, and A. R. Pradeep, "Gingival Crevicular Fluid VEGF Levels in Periodontal Health and Disease," *Journal of Periodontology*, vol. 78, no. 9, pp. 1783-1787, September 2007.

[2] C. B. Wiebe and E. E. Putnins, "The periodontal disease classification system of the American Academy of Periodontology--an update," *J Can Dent Assoc.*, vol. 66, no. 11, pp. 594-597, December 2000.

[3] G. C. Armitage, "Development of a classification system for periodontal diseases and conditions," *Annals of Periodontology*, vol. 4, pp. 1-6, December 1999.

[4] (2012, January) Wikipedia, the free encyclopedia. [Online]. http://en.wikipedia.org/wiki/Periodontitis

[5] L. J. Brown and H. Löe, "Prevalence, extent, severity and progression of periodontal disease," *Periodontol 2000*, vol. 2, pp. 57-71, June 1993.

[6] G. C. Armitage, "Periodontal diseases: diagnosis," *Annals of Periodontology*, vol. 1, no. 1, pp. 37-215, 1996.

[7] P. N. Papapanou, "Periodontal Diseases: Epidemiology," *Annals of Periodontology*, vol. 1, no. 1, pp. 1-36, November 1996.

[8] Löe H., Anerud A., Boysen H., and Morrison E., "Natural history of periodontal disease in man. Rapid, moderate and no loss of attachment in Sri Lankan laborers 14 to 46 years of age," *J Clin Periodontol*, vol. 13, no. 5, pp. 431-445, May 1986.

[9] R. Attström and U. van der Velden, "Consensus report (epidemiology)," in *Proceedings of the 1st European Workshop on Periodontics, 1993*, London, 1994, pp. 120-126.

[10] H. LOE and J. SILNESS, "Periodontal disease in pregnancy. I. Prevalence and severity," *Acta Odontol Scadinavia*, vol. 21, pp. 533-551, December 1963.

[11] G. Landini and I. E. Othman, "Estimation of tissue layer level by sequential morphological," *Journal of Microscopy*, vol. 209, pt 2, pp. 118– 125, February 2003.

[12] P. Soille, *Morphological Image Analysis: Principles and Applications*, 1st ed.: Springer-Verlag, 1999, pp. 164-165.

[13] R. Adams, "Radial Decomposition of Discs and Spheres," *Computer Vision, Graphics, and Image Processing: Graphical Models and Image Processing*, vol. 55, no. 5, pp. 325–332, September 1993.

[14] R. Jones and P. Soille, "Periodic lines: Definition, cascades, and application to granulometrie," *Pattern*

Recognition Letters, vol. 17, pp. 1057–1063, 1996.

[15] M.M. Sami, M. Saito, H. Kikuchi, and T. Saku, "A computer-aided distinction of borderline grades of oral cancer," in *16th IEEE International Conference on Image Processing (ICIP)*, Cairo, 2009, pp. 4205 - 4208.

[16] R. Abu Eida and G. Landini, "Quantification of the Global and Local Complexity of the Epithelial-Connective Tissue Interface of Normal, Dysplastic, and Neoplastic Oral Mucosae Using Digital Imaging," *Pathology - Research and Practice*, vol. 199, no. 7, pp. 475-482, 2003.

[17] M. Pal et al., "Quantitative dimensions of histopathological attributes and status of GSTM1–GSTT1 in oral submucous fibrosis," *Tissue and Cell*, vol. 40, no. 6, pp. 425-435, December 2008.

[18] J. Maeda, S. Novianto, S. Saga, Y. Suzuki, and V. V. Anh, "Rough and accurate segmentation of natural images using fuzzy region-growing algorithm," in *IEEE Proceedings Image Processing. ICIP 99*, vol. 3, Kobe, 1999, pp. 227-231.

[19] R. C. Gonzalez and R. E. Woods, *Digital Image Processing*, 3rd ed.: Prentice Hall, 2008.

[20] N. Otsu, "A Threshold Selection Method from Gray-Level Histograms," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 9, no. 1, pp. 62-66, 1979.

[21] G. Landini and I. E. Othman, "Quantification of Local Architecture Changes Associated with Neoplastic Progression in Oral Epithelium using Graph Theory," in *Fractals in Biology and Medicine*.: Birkhäuser Basel, 2005, vol. 4, pp. 193-201.

[22] G. Landini, "Quantitative analysis of the epithelial lining architecture in radicular cysts and odontogenic keratocysts," *Head & Face Medicine*, vol. 2, 2006.

[23] A. C. Ruifrok and D. A. Johnston, "Quantification of histological staining by color deconvolution," *Anal Quant Cytol Histol*, vol. 23, pp. 291-299, 2001.

[24] Y. Al-Kofahi, W. Lassoued, and W. Lee, "Improved Automatic Detection and Segmentation of Cell Nuclei in Histopathology Images," *IEEE transactions on biomedical engineering*, vol. 57, no. 4, pp. 841-852, April 2010.

[25] M. Park et al., "Automatic cell segmentation in microscopic color images using ellipse fitting and watershed," in *The 201 0 IEEEIICME International Conference on Complex Medical Engineering*, Gold Coast, Australia, 2010, pp. 69-74.

[26] S. DiCataldo, E. Ficarra, A. Acquaviva, and E. Macii, "Achieving the way for automated segmentation of nuclei in cancer tissue images through morphology-based approach - A quantitative evaluation," *Computerized Medical Imaging and Graphics*, vol. 34, pp. 453–461, 2010.

[27] P. Soille, *Morphological Image Analysis: Principles and Applications*, 1st ed.: Springer-Verlag, 1999, pp. 170-171.

[28] K. Zuiderveld, "Contrast Limited Adaptive Histogram Equalization," in *Graphic Gems IV*. San Diego: Academic Press Professional, 1994, pp. 474–485.

AUTHORS PROFILE

Aliaa A. A. Youssif, professor of computer science, Faculty of Computers and Information, Helwan University. Cairo, Egypt. She received her B.Sc and MSc. degree in telecommunications and electronics engineering from Helwan University. Prof. A. Youssif received the PhD degree in computer science from Helwan University in 2000. She was a visiting professor at George Washington University (Washington DC, USA) in 2005. She was also a visiting professor at Cardiff University in UK (2008). She is currently vice dean for the faculty of Computers and Information, Helwan University. Her fields of interest include pattern recognition, AI researches, and medical imaging. She published more than 40 papers in different fields. She can be reached at aliaay@helwan.edu.eg

Abeer Saad Gawish, professor of Oral Medicine, Periodontology, Oral Diagnosis and Radiology, Faculty of Dental Medicine (Girls'), Al-Azhar University. Cairo, Egypt. She received her B.Sc degree in Oral and Dental Medicine from Alexandria University. Prof. A. Gawish received the MSc. And the PhD degree Oral and Dental Medicine from Cairo University in 2000. She was a visiting professor at Pan Arab Association of Oral &Maxillofacial Surgeon KSA 2010. She is currently vice dean for the faculty of Oral and Dental Medicine, Sinai University. Her fields of interest include periodontal medicine, genetics in periodontal medicine, implantology. She published more than 37 papers in different fields. She can be reached at abeergawish@yahoo.com

Mohammed Elsaid Moussa received his BSc. degree in computer science from Faculty of Computers and Information, Helwan University, Cairo, Egypt in 2005. He is currently working as a teacher assistant at the same faculty, and a master's degree student under the supervision of prof. Aliaa A. A.Youssif and prof. Abeer Saad Gawish. His areas of interests include image processing, medical imaging, cloud computing, and programming languages. He can be reached at engmohammedelsaid@hotmail.com