Fall 2012

Data fusion by using machine learning and computational intelligence techniques for medical image analysis and classification

Beibei Cheng

Department: Electrical and Computer Engineering

DATA FUSION BY USING MACHINE LEARNING AND COMPUTATIONAL INTELLIGENCE TECHNIQUES FOR MEDICAL IMAGE ANALYSIS AND CLASSIFICATION

by

Beibei Cheng

A DISSERTATION

Presented to the Faculty of the Graduate School of the

MISSOURI UNIVERSITY OF SCIENCE AND TECHNOLOGY

In Partial Fulfillment of the Requirements for the Degree

DOCTOR OF PHILOSOPHY

in

COMPUTER ENGINEERING

2012

Approved by:

R.J. Stanley, Advisor
W. V. Stoecker
R. H. Moss
H. Xiao
Z. Yin

PUBLICATION DISSERTATION OPTION

This dissertation has been prepared in the style utilized by the Missouri University of Science and Technology and consists of the following ten articles that published as follows:

Paper I, Cheng B., Stanley R.J., Stoecker W.V., Stricklin S.M., "Analysis of Clinical and Dermoscopic Features for Basal Cell Carcinoma Neural Network Classification", has been published in Skin Research and Technology, 2012 Jun 22. [Epub ahead of print]

Paper II, Cheng B., Stanley R.J., Stoecker W.V., "Automatic Dirt Trail Analysis in Dermoscopy Images", has been published in Skin Research and Technology, 2012 Jan 11. [Epub ahead of print]

Paper III, Cheng B., Erdos D., Stanley R.J., Stoecker W.V., Hinton K., "Automatic Telangiectasia Analysis in Dermoscopy Images using Adaptive Critic Design", has been published in Skin Research and Technology, 2011 Dec 5. [Epub ahead of print]

Paper IV, Cheng B., Stanley R.J., Soumya D., Antani S., Thoma G.R., "Automatic Detection of Arrow Annotation Overlays in Biomedical Images", has been published in International Journal of Healthcare Information Systems and Informatics (IJHISI), International Journal of Healthcare Information Systems and Informatics, October-December 2011, 6(4), pp.23-41.

Paper V, Cheng B., Erdos D., Stanley R.J., Stoecker W.V., Calcara D., Gomez D., "Automatic Detection of Basal Cell Carcinoma Using Telangiectasia Analysis in Dermoscopy Skin Lesion Images", has been published in Skin Research and Technology, August 2011, 17(3), pp. 278–287.

Paper VI, Cheng B., Antani S., Stanley R.J., Thoma G.R., "Graphical Figure Classification Using Data Fusion for Integrating Text and Image Features", have been submitted to SPIE Electronic Imaging, San Francisco, California, 2013.

Paper VII, Cheng B., Antani S., Stanley R.J., Thoma G.R., "Graphical Image Classification Combining an Evolutionary Algorithm and Binary Particle Swarm Optimization", Proceedings of SPIE Electronic Imaging, San Francisco, California, vol.8297, pp.1-8, 2012.

Paper VIII, Cheng B., Antani S., Stanley R.J., Thoma G.R., "Automatic Segmentation of Subfigure Image Panels for Multimodal Biomedical Document Retrieval", Proceedings of the SPIE Electronic Imaging, San Francisco, California, vol. 7874, pp. Z-1-11, 2011.

Paper IX, Cheng B., Stanley R.J., Szalapski T., Stoecker W.V., "A Hybrid Computational Intelligence Algorithm for Automatic Skin Lesion Segmentation in Dermoscopy

Images", Proceedings of Artificial Neural Networks in Engineering (ANNIE), St. Louis, MO, vol.20, pp.379-386, 2010.

Paper X, Cheng B., Stanley R.J., Antani S., Thoma G.R., "A Novel Computational Intelligence based Approach for Medical Image Artifacts Detection", Proceedings of the 2010 International Conference on Artificial Intelligence and Pattern Recognition, Orlando, FL, pp.113-20, July 2010.

ABSTRACT

Data fusion is the process of integrating information from multiple sources to produce specific, comprehensive, unified data about an entity. Data fusion is categorized as low level, feature level and decision level. This research is focused on both investigating and developing feature- and decision-level data fusion for automated image analysis and classification. The common procedure for solving these problems can be described as: 1) process image for region of interest' detection, 2) extract features from the region of interest and 3) create learning model based on the feature data. Image processing techniques were performed using edge detection, a histogram threshold and a color drop algorithm to determine the region of interest. The extracted features were low-level features, including textual, color and symmetrical features. For image analysis and classification, feature- and decision-level data fusion techniques are investigated for model learning using and integrating computational intelligence and machine learning techniques. These techniques include artificial neural networks, evolutionary algorithms, particle swarm optimization, decision tree, clustering algorithms, fuzzy logic inference, and voting algorithms. This work presents both the investigation and development of data fusion techniques for the application areas of dermoscopy skin lesion discrimination, content-based image retrieval, and graphic image type classification.

ACKNOWLEDGMENTS

First and foremost, I would like to offer my sincerest gratitude to my advisor, Dr. R. J. Stanley. Dr. Stanley has supported me with his knowledge and patience while giving me the freedom to work on my own ideas. Without him, I would not have been able to overcome so many difficulties and accomplish so many publications. One could not ask for a better, friendlier advisor.

I am also extremely indebted to my committee members: Dr. W.V. Stoecker, Dr. R. Moss, Dr. H. Xiao, and Dr. Z. Ying. Dr. Stoecker helped me gain domain knowledge in the dermatology related aspects of my research. Dr. Moss provided a great deal of advice in both my research and my paper written. The suggestions given by Dr. Xiao and Dr. Ying during both my comprehensive exam and final defense were very useful toward improving not only my research but also my dissertation.

I must also acknowledge Dr. S. Antani and Dr. G. Thoma from the U.S. National Library of Medicine: National Institutes of Health. Dr. Antani and Dr. Thoma offered many interesting research topics and gave me the opportunity to get involved with the projects. I also appreciate their time spent in reviewing my paper. This work would not have been possible without their continued help and support. Last but not least, I am thankful to my ever-supportive husband, Dr. R. Wang, and my mother, Ms. L. Zhao. Their love, encouragement, and faith in my abilities kept me going throughout my work.

TABLE OF CONTENTS

PUBLICATION DISSERTATION OPTION iii
ABSTRACT v
ACKNOWLEDGMENTS vi
LIST OF ILLUSTRATIONS xiv
LIST OF TABLES xviii
SECTION
1. INTRODUCTION 1
1.1 BACKGROUND AND PROBLEM STATEMENT 1
1.2 SUMMARY OF CONTRIBUTIONS 2
PAPER 5
I. ANALYSIS OF CLINICAL AND DERMOSCOPIC FEATURES FOR BASAL CELL CARCINOMA NEURAL NETWORK 5
ABSTRACT 5
1. INTRODUCTION 6
2. DESCRIPTION OF FEATURE CATEGORIES 8
3. DISCRIMINATION ALGORITHMS 11
3.1 EVOLVING ARTIFICIAL NEURAL NETWORKS (EANN's) 11
3.2 EVOLVING ARTIFICIAL NEURAL NETWORK ENSEMBLES (EANNE's) 12
4. EXPERIMENTS PROCEDURE 13
5. RESULTS AND DISCUSSIONS 14
6. CONCLUSIONS 16
ACKNOWLEDGEMENTS 16
REFERENCES 17
II. AUTOMATIC DIRT TRAIL ANALYSIS IN DERMOSCOPY IMAGES 19
ABSTRACT 19
1. INTRODUCTION 20
2. METHODOLOGY 21
2.1 DATA SET DESCRIPTION 21

2.2 DIRT TRAIL DETECTION ALGORITHM OVERVIEW 21

3. DIRT TRAIL DETECTION ALGORITHM 22

3.1 GAUSSIAN BANDPASS FILTER 22

3.2 MEDIAN FILTER 24

3.3 IMAGE SUBTRACTION 25

3.4 HISTOGRAM PROCESSING 25

3.5 LOGIC OPERATION 26

3.6 NOISE FILTERING FOR HAIRS AND BUBBLES 26

3.7 NOISE FILTERING FOR ISOLATED NOISE 27

4. CLASSIFIER INPUT FEATURES AND CLASSIFIER METHODOLOGY 28

4.1 FEATURES COMPUTED FOR LESION DISCRIMINATION 28

4.2 CLASSIFIER DESCRIPTION 28

5. EXPERIMENTAL RESULTS 30

5.1 DIRT TRAIL MASK EXAMPLES 30

5.2 FEATURE ANALYSIS USING LOGISTIC REGRESSION 30

5.3 LESION DISCRIMINATION RESULTS 30

6. CONCLUSIONS 33

ACKNOWLEDGMENTS 33

REFERENCES 34

III. AUTOMATIC TELANGIECTASIA ANALYSIS IN DERMOSCOPY IMAGES USING ADAPTIVE CIRTIC DESIGN 35

ABSTRACT 35

1. INTRODUCTION 36

2. PROBLEM DESCRIPTION 41

3. ADAPTIVE CRITIC DESIGN METHODOLOGY 45

4. RESULTS AND DISCUSSION 48

4.1 DIAGNOSTIC ACCURACY 48

4.2 STRUCTURE DETECTION VS. DIAGNOSIS 49

5. CONCLUSIONS AND FUTURE WORK 50

REFERENCES 51

IV. AUTOMATIC DETECTION OF ARROW ANNOTATION OVERLAYS IN BIOMEDICAL IMAGES 53

ABSTRACT 53

1. INTRODUCTION 54

2. ARROW DETECTION PROCESS 57

2.1 OBJECT SEGMENTATION 57

2.2. FEATURE EXTRACTION 58

2.2.1 Region Property Features. 59

2.2.2 Shape Features. 60

2.2.3 Correlation-based Features. 61

2.3 CLASSIFIER ALGORITHMS 62

3. EXPERIMENTS PERFORMED 64

4. DISCUSSION 69

5. CONCLUSIONS AND FUTURE WORK 72

ACKNOWLEDGEMENTS 72

REFERENCES 73

APPENDIX 76

V. AUTOMATIC DETECTION OF BASAL CELL CARCINOMA USING TELANGIECTASIA ANALYSIS IN DERMOSCOPY SKIN LESION IMAGES 80

ABSTRACT 80

1. INTRODUCTION 81

2. METHODS 83

2.1 TELANGIECTASIA DETECTION OVERVIEW 83

2.2 DATA SET USED 83

2.3 PRE-PROCESSING 84

2.4 COLOR-DROP VESSEL DETECTION 84

2.5 NOISE ELIMINATION 86

2.5.1 Brown Network Filtering. 87

2.5.2 Hair Filtering. 88

2.5.3 Bubble Filtering. 88

2.5.4 Blob Filtering for Mask Density. 89
2.5.5 Post-processing: Connecting Vessel Segments. 89
2.5.6 Post-processing: Length and Area Lower Bounds. 90
2.5.7 Post-processing: Independent and Principal Component Analysis, Histogram Pursuit. 91
2.6 VESSEL FEATURE GENERATION 91
2.7 NEURAL NETWORK METHODS FOR BCC DIAGNOSIS 93
3. EXPERIMENTS PERFORMED 94
4. RESULTS 95
5. DISCUSSION 96
5.1 DIAGNOSTIC ACCURACY 96
5.2 EXAMPLES OF ERRORS 96
5.3 RELATION TO CLINICAL PRACTICE 97
5.4 STRUCTURE DETECTION VS. DIAGNOSIS 97
5.5 LIMITATIONS OF STUDY AND SCOPE OF FUTURE RESEARCH 97
REFERENCES 99
VI. GRAPHICAL FIGURE CLASSIFICATION USING DATA FUSION FOR INTEGRATING TEXT AND IMAGE FEATURES 101
ABSTRACT 101
1. INTRODUCTION 102
2. RELATED WORK 104
3. IMAGE FEATURE EXTRACTION AND SELECTION 105
3.1 IMAGE FEATURE EXTRACTION 105
3.2 IMAGE FEATURE SELECTION 105
4. TEXT FEATURE EXTRACTION AND SELECTION 107
4.1 TEXT FEATURE EXTRACTION 107
4.2 TEXT FEATURE SELECTION 107
5. OUTPUT COMBINATION 109
6. EXPERIMENTAL RESULTS 110
7. DISCUSSION AND CONCLUSIONS 113

ACKNOWLEDGEMENTS 113

REFERENCES 114

VII. GRAPHICAL IMAGE CLASSIFICATION COMBINING AN EVOLUTIONARY ALGORITHM AND BINARY PARTICLE SWARM OPTIMIZATION 115

ABSTRACT 115

1. INTRODUCTION 116

2. RELATED WORK 118

3. METHODS 120

3.1 FEATURE EXTRACTION 120

3.2 CLASSIFIER 121

4. OPTIMAL FEATURE SELECTION 123

4.1 EVOLUTIONARY ALGORITHM (EA) 123

4.2 BINARY PARTICLE SWARM OPTIMIZATION (BPSO) 123

4.3 EABPSO 124

5. EXPERIMENTAL SETUP AND RESULTS 125

5.1 DISCUSSION 126

6. CONCLUSIONS 128

ACKNOWLEDGEMENTS 128

REFERENCES 129

VIII. AUTOMATIC SEGMENTATION OF SUBFIGURE IMAGE PANELS FOR MULTIMODAL BIOMEDICAL DOCUMENT RETRIEVAL 131

ABSTRACT 131

1. INTRODUCTION 132

2. PRIOR WORK 134

3. METHODS 135

3.1 IMAGE TYPE IDENTIFICATION 135

3.2 SEGMENTING MULTI-PANEL REGULAR IMAGES 136

3.3 SEGMENTING MULTI-PANEL ILLUSTRATION IMAGES 137

4. EXPERIMENTS 144

5. CONCLUSIONS 146

ACKNOWLEDGEMENTS 146

REFERENCES 147

IX. A HYBRID COMPUTATIONAL INTELLIGENCE ALGORITHM FOR AUTOMATIC SKIN LESION SEGMENTATION IN DERMOSCOPY IMAGES 149

ABSTRACT 149

1. INTRODUCTION 150

2. PREPROCESSING 152

3. EVOLVING VECTOR QUANTIZATION 153

3.1 STANDARD PSO 153

3.2 DIFFERENTIAL EQUATION-BASED PSO 154

3.3 POSTPROCESSING 155

4. RESULTS AND DISCUSSION 157

5. CONCLUSIONS AND FUTURE WORK 158

REFERENCES 159

X. A NOVEL COMPUTATIONAL INTELLIGENCE-BASED APPROACH FOR MEDICAL IMAGE ARTIFACTS DETECTION 160

ABSTRACT 160

1. INTRODUCTION 161

2. FEATURE GENERATION 165

3. ADAPTIVE CRITIC DESIGN METHODOLOGY 170

4. OTHER CLASSIFICATION ALGORITHMS 173

4.1 MLP BACKPROPGATION NEURAL NETWORK 173

4.2. PARTICLE SWARM OPTIMIZATION (PSO) FOR TRAINING OF A MLP NEURAL NETWORK 173

4.3. GENETIC ALGORITHM (GA) FOR TRAINING OF A MLP NEURAL NETWORK 174

5. RESULTS AND DISCUSSION 175

6. CONCLUSIONS AND FUTURE WORK 177

ACKNOWLEDGEMENTS 177

REFERENCES 178

SECTION

2. CONCLUSIONS 180

BIBLIOGRAPHY 181

VITA 182

LIST OF ILLUSTRATIONS

Page

PAPER I

Figure 1. Examples of BCC BASAL structures visible using dermoscopy. 7

PAPER II

Figure 1. Dirt trail examples in dermoscopic skin lesion images, shown by arrows, with dirt trails containing dots and clods of varying sizes. 20
Figure 2. Overview of the dirt trail detection algorithm. 23
Figure 3. RGB plane. 23
Figure 4. Gaussian bandpass filter representation in the spatial frequency domain. 23
Figure 5. Bandpass-filtered images converted to spatial domain for *R*, *G*, and *B* planes.. 24
Figure 6. Median filter output images from *R*, *G*, *B* planes. 25
Figure 7. Output images from scalarized Otsu's method for *R*, *G*, *B* planes. 26
Figure 8. Otsu's output image *A* after logical AND of the threshold color plane images. 26
Figure 9. Image overlay 27
Figure 10. Dirt trail detection mask examples.. 31
Figure 11. ROC curve and AUC (area under curve) for backpropagation neural network. 32

PAPER III

Figure 1. Telangiectasia example.. 37
Figure 2. Overview of vessel detection and discrimination process. 38
Figure 3. Overview of image analysis algorithm. 38
Figure 4. Examples of telangiectasia in the skin image from Figure 1 and the vessel mask generated using the algorithm in Figure 3. 39
Figure 5. Images with varied contrast enhancement (CE) 42
Figure 6.Vessel mask overlays change with varying contrast enhancement. 43
Figure 7. Schematic diagram of ADHDP. 45
Figure 8. Neural network structure of the critic network. 47
Figure 9. ROC curve and AUC (area under curve) for neural networks for all 10 text sets. 48

PAPER IV

Figure 1. Medical image with arrows.. 55
Figure 2. Overview of arrow detection process. 56
Figure 3. Edge detection operator mask. 58
Figure 4. Object segmentation example..... 59
Figure 5. MinPixelNo feature.. 60
Figure 6. Samples for generating WDD features. 62
Figure 7. The WDD functions used to compute arrow features. 62
Figure 8. ROC curves for a representative test set for case 1 for the different classifiers with M=5, N=75 (except for Backpropagation ANN). 66
Figure 9. Arrow objects incorrect classified.. 70
Figure 10. Non-arrow objects incorrect classified. 71

PAPER V

Figure 1. Telangiectasia.. 82
Figure 2. Overview--BCC diagnosis by telangiectasia detection. 83
Figure 3. Direction Mask Used for Pixel Marking. 85
Figure 4. Mask image with different red drops..... 85
Figure 5. Mask Images with Different NumPix Values..... 86
Figure 6. Brown Area Filtering..... 87
Figure 7. Hair filtering.. 88
Figure 8.Bubble filtering..... 89
Figure 9. Blob density filtering. 89
Figure 10. (a) Mask after noise filtering. (b) Mask after dilation, radius 3, and erosion, radius 2. 90
Figure 11. Lower area bound. 90
Figure 12. Receiver operating characteristic curve and area under curve (AUC) results for different feature combinations. 95
Figure 13. BCC misdiagnosed. 96
Figure 14. Benign lesion misdiagnosed.. 98

PAPER VI

Figure 1. Examples of graphical figure with its caption. 103
Figure 2. Overview of EABPSO procedure..... 106

PAPER VII

Figure 1. Graphical image type examples. 117
Figure 2. Examples illustrating problems, discussed in Section 2, in existing approach after preprocessing. 119
Figure 3. Image preprocessing example. 120
Figure 4. Overview of EABPSO procedure. 124
Figure 5. Root Mean Squared Error (RMSE) versus iteration number for each of the feature selection schemes. 126

PAPER VIII

Figure 1. Examples of different types of multi-panel images. 133
Figure 2. Overview of sub-image segmentation process. 135
Figure 3. Image processing algorithm example. 137
Figure 4. Overview of multi-panel images decomposition process. 138
Figure 5. Sobel edge operator processing.. 139
Figure 6. Form bounding boxes. 139
Figure 7. Incorrect bounding box segmentation. 140
Figure 8. Steps for applying Particle Swarm Optimization algorithm. 141
Figure 9. Membership Function. 142
Figure 10. Sample results of PSO clustering with different fitness functions. 143
Figure 11. Input image and its fitness function. 145

PAPER IX

Figure 1. Overview of lesion segmentation process. 151
Figure 2. Preprocessing example. 152
Figure 3. Evolving vector quantization algorithm flow chart. 154
Figure 4. Example of image output within ten iterations 155
Figure 5. Post-processing 156

PAPER X

Figure 1. Medical image examples. 161
Figure 2. Overview of arrow detection and discrimination process. 163
Figure 3. Edge detection. 163
Figure 4. Image processing algorithm example.. 164

Figure 5. MinPixelNo feature. 165
Figure 6. The WDD functions used to compute arrow features. 167
Figure 7. Samples for generating WDD features. 167
Figure 8. Falsely identified arrow. 169
Figure 9. Arrows with different rotation value. 169
Figure 10. Schematic diagram of ADHDP. 171
Figure 11. Basic concept of PSO. 173
Figure 12. ROC curve and AUC (area under curve) for neural networks. 175

LIST OF TABLES

Page

PAPER I

Table 1. Overview of feature groups examined for lesion discrimination. 9

Table 2. Feature combinations for lesion discrimination. 10

Table 3. Average area under the ROC curve discrimination results for different feature and discrimination algorithm combinations. 14

PAPER II

Table 1. Descriptions for Dirt Trail Dermoscopic Features 28

PAPER III

Table 1. Vessel-based features investigated with descriptions 44

PAPER IV

Table 1. Seven cases with their feature combinations and NN architectures. 65

Table 2. Average area under the ROC curve test results and (precision/recall) for different algorithms and feature combinations. 67

PAPER V

Table 1. Vessel-based features investigated with descriptions. 92

Table 2. Feature measures generated from final vessel mask, included in final model, SAS procedure logistic. 93

PAPER VI

Table 1. Top 20 features with its information gain value. 107

Table 2. Top 20 features with its Chi-square value. 108

Table 3. Five classes of fuzzy set union and intersection. 109

Table 4. Performance comparison for image feature set with MLP NN classifier. 111

Table 5. Performance for text feature set with MLP NN classifier. 112

Table 6. Performance for combination by fuzzy logic union and intersection. 112

PAPER VII

Table 1. Extracted Features. 122
Table 2. Performance comparisons for different feature combinations. 127

PAPER VIII

Table 1. Information gain and chi-square value for image features useful in detecting image type. 136
Table 2. Fuzzy rule. 143
Table 3. Result of image type identification. 144

PAPER IX

Table 1. Standard PSO and DEPSO error result with three planes. 157

PAPER X

Table 1. Information gain and chi-square value. 168

1. INTRODUCTION

1.1 BACKGROUND AND PROBLEM STATEMENT

The analysis of medical images is essential in modern medicine. Both new challenges and opportunities arise for different phases of the clinical routine and image retrieval as the amount of not only patient data but also medical publications increase. This work uses state-of-the-art data fusion techniques for the following three aspects:

- Dermoscopy skin lesion discrimination
- Content-based image retrieval
- Graphic image type classification

Data fusion is the process of integrating information from multiple sources to produce specific, comprehensive, unified data about an entity [1]. Data fusion is categorized as low-level, feature-level, and decision-level. Low-level data fusion combines several sources of raw data to produce new raw data. Feature-level data fusion combines the descriptive features extracted from multiple sensors, measuring either a similar or a dissimilar phenomenon into a single feature vector for discrimination purposes.

Decision-level data fusion involves the fusion of a sensor using preliminary decisions, made from individual or multiple sensors and information sources.

Techniques for data fusion include computational intelligence, machine learning, statistical estimation, and more. This research focused on developing both machine learning techniques and computational intelligence for data fusion.

Machine learning is defined as computer program that "is said to learn from experience *E* with respect to some class of tasks *T* and performance measure *P*, if its performance at tasks in *T*, as measured by *P*, improves with experience *E*" [2]. The approaches of machine learning include decision tree learning, artificial neural network, clustering, and so forth. Each approach can be overviewed as follows. A decision tree is a decision support tool that uses a tree-like graph to illustrate the choices available to a decision maker, every possible choice with its estimated outcome being shown as a separate branch of the tree [3]. An artificial neural network can be viewed as a parallel and distributed processing system which consists of a huge number of simple, massively connected processors [4]. Clustering is the task of partitioning data objects (patterns, entities, instances, observances, and units) into a certain number of clusters (groups, subsets, or categories) so that the objects in the same cluster are more similar to each other than to those in other clusters [5].

Computational intelligence is a set of nature-inspired exploration into the adaptive mechanisms that enable intelligent behavior in complex and changing environments which includes evolutionary computation, swarm intelligence and fuzzy systems [6]. The concept of evolutionary computation is survival of the fittest: the weak may die while the elite move to the next level [7]. Swarm intelligence is the study of swarms of social organisms such as a flock of birds or a school of fish [8]. Fuzzy logic is a form of either many-valued logic or probabilistic logic; it deals with reasoning that is approximate rather than fixed and exact [9].

Feature-level fusion involves the extraction of representative features from sensor data. These features combine into a single vector as the input to a classification approach based on neural networks, clustering algorithms, and so on [1]. Current studies include the investigation of feature-level fusion using hand and face biometrics [10], feature-level data fusion for bimodal person recognition [11], feature-level data fusion for land mine detection [12], and so forth. Decision-level data fusion involves fusing sensors using preliminary decisions made from individual or multiple sensors and information sources. Examples of decision-level fusion methods include weighted decision methods (voting techniques) [13] and inference methods [14].

This work presents both the investigation and development of data fusion in not only feature-level and also decision-level fusion for a varied range of applications. Feature-level fusion techniques are explored for the skin cancer diagnosis and content-based image retrieval. These techniques include artificial neural network, evolutionary algorithm, particle swarm optimization, adaptive critic design, decision tree and clustering algorithms. Decision-level fusion approaches including fuzzy logic controller, fuzzy intersection/union, and voting algorithm are developed for graphic image type classification.

1.2 SUMMARY OF CONTRIBUTIONS

This dissertation consists of a number of journal papers and conference papers, as presented in the publication list. My unique contributions regarding solving the proposed three aspects are:

Dermoscopy skin lesion discrimination. Basal cell carcinoma (BCC) is the most commonly diagnosed skin cancer in the USA. An automated image analysis and classification method for dermoscopy skin lesion discrimination was developed by applying a novel classifier (Adaptive Critic Design) in the feature-level data fusion. The representative features were first extracted from sensor data according to dermatologist's suggestion. Those features were then combined into a single vector as the input to Adaptive Critic Design (ACD). ACD is a

biologically-inspired reinforcement learning approach which can combine feature information from multiple sources. ACD gives results superior to those from existing methods such as a backpropagation neural network. This technique can be generally applicable to structure detection that is sensitive to image parameters such as contrast. In addition, it is a useful means for global feature data fusion in order to improve the accuracy of pattern recognition. Furthermore, skin cancer discrimination was enhanced by adding information gathered from the patient and the physician, demonstrating the potential of data fusion in aiding the lesion diagnostic process.

Content-based image retrieval. The detection of medical image artifacts, such as arrows, is important for highlighting both supplemental and context-based information. This information is helpful in improving biomedical information retrieval. Computational intelligence methods, including evolving artificial neural networks (EANNs) and evolving artificial neural networks ensembles (EANNEs), have been investigated in the feature level data fusion for arrow object discrimination. Both EANNs and EANNEs are hybrid artificial intelligence algorithms by combining evolutionary algorithm and particle swarm optimization with artificial neural network. EANNs are evolving into one of the best neural networks available. EANNEs are evolving into neural network ensembles, with the minimum correlation information among the ensembles. Both EANNs and EANNEs outperform the backpropagation NN, indicating the benefit of incorporating multiple neural networks for data fusion. EANNEs outperform EANN, as EANNEs play a better role in fusing the features within a large diversity and variety.

Graphic image type classification. Automatic figure type identification is an important step toward improving both multimodal (text + image) information retrieval and clinical decision support applications. An automated image type classification was developed by applying a data fusion approach to combine information from both text and image sources believed to contain complementary information. First, feature-level fusion techniques were developed for image feature selection in biomedical images for chart classification. These techniques include Particle Swarm Optimization (PSO) and Evolutionary Algorithm (EA) paradigms used for feature selection, with Support Vector Machines (SVM) utilized as the fitness value determination. A hybrid EAPSO algorithm was invented for feature selection by combining the new individual generation functions of both EA and PSO, to attain the global minimum at high speed. Second, a decision-level fusion approach (fuzzy set intersection/union to improve graphic image type classification) was developed. For each biomedical image, both its image features and its caption features were input into artificial neural network classifiers.

The classifiers' outputs were characterized as fuzzy sets to determine the final classification result. This research demonstrates that data fusion techniques are useful in the feature-level data fusion to remove the negative related or unrelated features, while data fusion techniques are useful in the decision level data fusion to combine the resource information.

PAPER

I. ANALYSIS OF CLINICAL AND DERMOSCOPIC FEATURES FOR BASAL CELL CARCINOMA NEURAL NETWORK

Beibei Cheng[a], R. Joe Stanley[a], William V, Stoecker[b], Sherea M. Stricklin[b], Kristen A. Hinton[b], Thanh K. Nguyen[a], Ryan K. Rader[b], Harold S. Rabinovitz[c] , Margaret Oliviero[c] , Randy H. Moss[a]

[a]Department of Electrical and Computer Engineering, Missouri University of Science and Technology (S&T), Rolla, MO, 65409, USA

[b]Stoecker & Associates, Rolla MO, 65401, USA

[c]Skin and Cancer Associates, Plantation, FL, USA

ABSTRACT

Background: Basal cell carcinoma (BCC) is the most commonly diagnosed cancer in the USA. In this research, we examine four different feature categories used for diagnostic decisions, including patient personal profile (patient age, gender, etc.), general exam (lesion size and location), common dermoscopic (blue-gray ovoids, leaf-structure dirt trails, etc.), and specific dermoscopic lesion (white/pink areas, semitranslucency, etc.). Specific dermoscopic features are more restricted versions of the common dermoscopic features.

Methods: Combinations of the four feature categories are analyzed over a data set of 700 lesions, with 350 BCCs and 350 benign lesions, for lesion discrimination using neural network-based techniques, including evolving artificial neural networks (EANNs) and evolving artificial neural network ensembles.

Results: Experiment re sults based on 10-fold cross validation for training and testing the different neural network-based techniques yielded an area under the receiver operating characteristic curve as high as 0.981 when all features were combined. The common dermoscopic lesion features generally yielded higher discrimination results than other individual feature categories.

Conclusions: Experimental results show that combining clinical and image information provides enhanced lesion discrimination capability over either information source separately. This research highlights the potential of data fusion as a model for the diagnostic process.

1. INTRODUCTION

Basal cell carcinoma (BCC), characterized as a slow-growing skin malignancy originating within the basal layer of the epidermis, is the most commonly diagnosed cancer with an estimated 3 million new cases annually in the USA [1]. In the consultation process when a patient exhibits a skin lesion, physicians gather patient information, determine general information about the skin lesion, and may use devices such as a dermatoscope (3Gen LLC, San Juan Capistrano, CA, USA; Heine Optotechnik, Herrsching, Germany) for determining a preliminary diagnosis. Currently, dermatopathology examination of a biopsy is used as the diagnostic gold standard.

In this research, the diagnostic process, here characterized as BCC versus benign lesion discrimination is examined based on information gathered from the patient, physician, and the dermoscopic image of the lesion. There are several BCC structures identifiable using dermoscopy that strongly suggests a BCC diagnosis. These have been incorporated into the BASAL acronym: Blue-gray ovoids and globules, Arborizing telangiectasia, Semitranslucency/Spoke wheel structures, Atraumatic ulcerations, and Leaf-like structures/dirt-trails [2]. Figure 1(a)–(c) presents dermoscopic lesion image examples of these structures.

While there have been numerous studies based on dermoscopic image feature analysis for pigmented lesion discrimination, few studies have specifically addressed BCC versus benign lesion discrimination by a classifier [3-6]. In those studies, BCC lesion discrimination was focused on the detection and analysis of particular dermoscopic features, including telangiectasia [3, 4], leaf-dirt trails [5], and semitranslucency [6]. This research explores the efficacy of fusing clinical and dermoscopic features to enhance skin lesion discrimination capability. Four categories of features were investigated in this research for BCC discrimination: (i) patient personal profile descriptors, (ii) general exam descriptors, (iii) common dermoscopic skin lesion image features associated with BCC, and (iv) specific dermoscopic skin lesion image features used for detecting uncommon BCC presentations. Combinations of these feature categories are examined for skin lesion discrimination using neural network techniques, including standard backpropagation neural networks, evolving artificial neural networks (EANNs), and EANN ensembles (EANNEs).

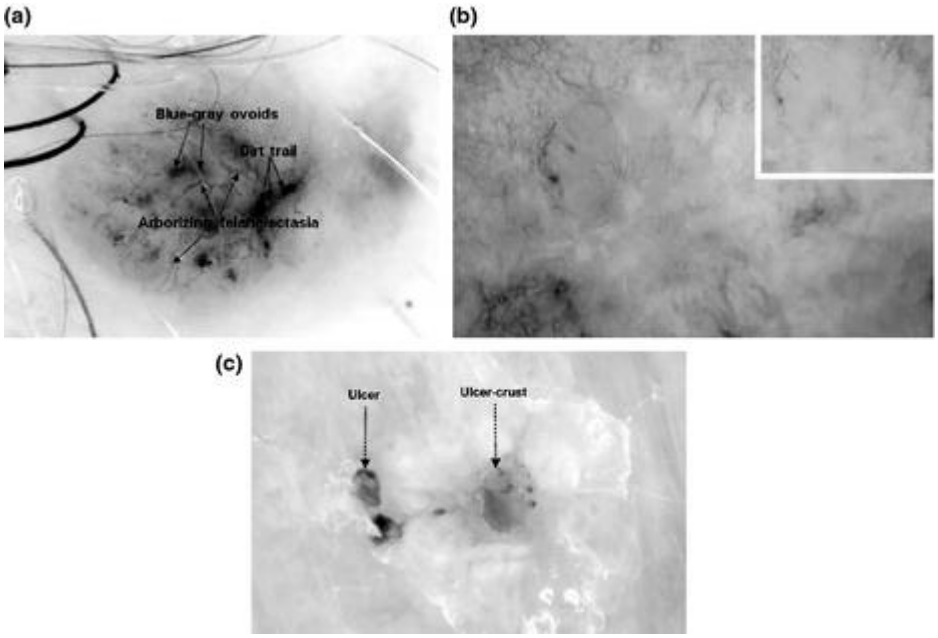


Figure 1. Examples of BCC BASAL structures visible using dermoscopy. All images are contact dermoscopy images except b. (a) Blue-gray ovoids, Arborizing telangiectasia, and dirt trails (rudimentary Leaf-like structures), (b) Semitranslucency, non-contact dermoscopy (inset-contact dermoscopy), and (c) Atraumatic ulceration, with similar ulcer-crust.

The remainder of the paper is presented in the following sections:

1. Description of feature categories
2. Discrimination algorithms
3. Experimental procedure
4. Results and discussion
5. Conclusion.

2. DESCRIPTION OF FEATURE CATEGORIES

Table 1 provides an overview of the four overall feature groups. The personal profile descriptors used are patient age, gender, race/ethnicity, and geographic location. Age was given in a binary format for either older than or younger than (and including) 50 years of age. Race/ethnicity is quantized as either non-Hispanic or Hispanic. The geographic location is based on clinic location, either above (Missouri, Connecticut) or below (Florida) the Tropic of Cancer. The final general exam attribute was lesion location. The third category of features are basic and common BCC dermoscopic descriptors BASAL features including blue-gray ovoids, vessels, pink veil (a more general form of semitranslucency), atraumatic ulcerations, and leaf-dirt trails. The fourth category of features comprises more specific dermoscopic features found to be present in this set of 350 BCCs. For both the common and specific dermoscopic features, a dermatologist (W.V.S.) identified the presence or absence of each feature within the lesion. Two of these specific features have been previously described and have high specificity: arborizing telangiectasia and semitranslucency [2, 7]. We have determined in this research other features with high specificity found in BCC that have not been previously described. These features are developed here to detect unusual BCC presentations such as white areas, pink areas, purple blotches, pale areas, large (majority of lesion) ulcer/crust plus pink regions, and majority white and pink regions. The common and specific dermoscopic features in each feature group are shown in Table 1.

The experimental data examined to evaluate the discrimination capability of the different feature categories includes images acquired from four clinic locations across the USA: The Dermatology Center (Rolla, MO, USA), Drugge & Sheard (Stamford, CT, USA), Skin & Cancer Associates (Plantation, FL, USA) and Columbia Dermatology & Mohs Skin Cancer Surgery. (Columbia, MO, USA). Three hundred and fifty lesions with a BCC diagnosis and 350 benign lesions were used as the test set. Benign lesions consisting of 80 acquired nevi, 71 seborrheic keratoses, 60 actinic keratoses, 51 lentigines, 15 congenital nevi, 10 lichen planus-like keratoses, 9 sebaceous hyperplasia, and 54 lesions of miscellaneous types constituted the competitive set.

Table 1. Overview of feature groups examined for lesion discrimination.

Patient Personal profile descriptors	**General exam descriptors**	**Common BCC features**	**Specific BCC features**
Age	Size	Blue-gray ovoids	Arborizing telangiectasia
Gender	Lesion Location	Pink veil	White/pink areas
Race/Ethnicity	N/A	Vessels	Semitranslucency
Geographic Location	N/A	Leaf-Dirt Trails	Purple Blotches
N/A	N/A	Atraumatic ulcerations	Pale Areas
N/A	N/A	N/A	Majority of Lesion
N/A	N/A	N/A	Ulcer/Crust plus Pink Regions
N/A	N/A	N/A	Majority Pink/White Regions

Table 2 presents the different feature group combinations examined for BCC versus benign lesion discrimination based on the four feature categories. Column 1 gives the feature combination, Column 2 provides the feature categories included in the feature combination, Column 3 lists the total number of features for the feature combination, and Column 4 gives the multilayer perceptrons neural network architecture used for lesion discrimination for the different feature combinations.

Table 2. Feature combinations for lesion discrimination.

Feature Combination	Feature Groups Included	Total Number of Features	Neural Network Architecture
1	Personal profile descriptors	4	5x5x1
2	General exam descriptors	2	3x5x1
3	Common lesion features	5	6x5x1
4	Specific lesion features	6	7x5x1
5	Personal profile descriptors, general exam descriptors	6	7x5x1
6	Personal profile descriptors, common lesion features	9	10x5x1
7	Personal profile descriptors, specific lesion features	10	11x5x1
8	General exam descriptors, common lesion features	7	8x5x1
9	General exam descriptors, specific lesion features	8	9x5x1
10	Common lesion features, specific lesion features	11	12x5x1
11	Personal profile descriptors, general exam descriptors, common lesion features	11	12x5x1
12	Personal profile descriptors, general exam descriptors, specific lesion features	12	13x5x1
13	Personal profile descriptors, common lesion features, specific lesion features	15	16x5x1
14	General exam descriptors, common lesion features, specific lesion features	13	14x5x1
15	Personal profile descriptors, general exam descriptors, common lesion features, specific lesion features	17	18x5x1

3. DISCRIMINATION ALGORITHMS

Using the different combinations of feature categories given in Table 2, EANN's and EANNE's are examined for BCC versus benign lesion discrimination. The implementation for each algorithm is given in this section.

3.1 EVOLVING ARTIFICIAL NEURAL NETWORKS (EANN's)

EANN's refer to a class of artificial neural networks (ANN's) in which both evolution and learning are fundamental forms of adaption. Evolutionary algorithms (EAs) can be used to train the connection weight, design the architecture, and select the input features of the ANN's. In this research, EAs using particle swarm optimization (PSO) [8] and genetic algorithm (GA) [9] for neural network training are investigated.

PSO is the study of swarms of social organisms such as a flock of birds, in which each particle in the swarm moves toward its previous best location (Pbest) and global best location (Gbest) at each time step [8]. The PSO algorithm utilized in this study is presented in detail in [10] and is overviewed as follows. We initialize M particles, where each particle is a D-dimensional vector with each element of the vector representing a connection weight and D being the total number of weights. For example, feature combination 1 from Table 2 has an architecture of $5 \times 5 \times 1$. The total number of weights (D) is 11 (5 + 5 + 1) and each element in this *D*-dimensional vector is the connection weight. Furthermore, the *M* particles represent *M* ANNs.

The connection weights in each ANN are updated when the elements in each particle are trained as follows. The initial value for each element of the vector is randomly set at a value from –0.1 to 0.1. In each training time step, the element's value of each particle is updated toward Pbest and Gbest. Pbest is the particle of the *M* particles that gives the least root mean square error (RMSE) between the current training epoch and the previous training epoch. Gbest is the particle among the *M* particles which generates the minimum RMSE, where the RMSE is calculated based on the difference between the ground truth and the actual ANN's output. The details for the updating process are given in [8]. The same process is repeated for *N* epochs. The final Gbest particle is selected for the final ANN weights for the test vector.

GAs use the "survival of the fittest" concept where the weak may die and elites are able to progress to the next level [9]. The GA approach investigated here is presented in detail in [10]. We initialize *M* parents, where each parent is a *D*-dimensional vector with each element of the vector representing a connection weight, with *D* being the total number of weights. In a

similar fashion defined for the PSO configuration, *M* parents represent *M* ANNs. The initial value for each element of the vector is randomly set from –0.1 to 0.1. In each training step, these *M* parents generate *M* offspring after implementing random selection, uniform crossover and a mutation operation. Then, the next *M* parents are selected based on whether the parents or their offspring minimize the RMSE. The same process is repeated for *N* epochs. From the final parent pool, the parent which minimizes the RMSE is selected for the final ANN weights for the test vectors.

3.2 EVOLVING ARTIFICIAL NEURAL NETWORK ENSEMBLES (EANNE's)

Neural network ensembles provide an approach for using and combining the outputs from several networks, with each ANN having the same inputs and generating its own outputs [11]. The purpose of EANNs is to make use of the whole population of ANNs rather than a single one. Training the network ensembles and determining the final output from the network ensembles are two main components for EANNEs design.

Negative correlation learning [12, 13] is implemented to train neural network ensembles in order to minimize the mutual informal among the networks. We initialize an ensemble with *M* ANNs with the initial weights in each ANN randomly set to a value from –0.1 to 0.1. In each training time step, each ANN in the ensemble is trained for a certain number of epochs using negative correlation learning firstly. Then, *M* offspring ANNs will be created by using selection, crossover, and a mutation operation in GA and replacement of the worst ANNs. The same process is repeated for *N* times.

For the final output determination, there are several criteria such as averaging, winner-taking-all and voting for combining the outputs. Here we simply choose averaging to deliver the final output as shown in Eq. (1).

$$y_{final} = \frac{1}{M}\sum_{m=1}^{M} F_m \tag{1}$$

where *M* is the number of the individual ANNs in the ensemble, F_m is the output for each ANN.

4. EXPERIMENTS PROCEDURE

Fifteen different feature combinations were investigated as inputs to the neural network architectures. The feature combinations and the neural network architectures for the EANN approach trained using the GA algorithm, the EANN approach trained using the PSO algorithm, the EANNE approach, and standard backpropagation ANN are given in Table 2. Each neural network architecture includes the total number of features for the feature combination and one bias as the inputs to each classifier algorithm. Each neural network architecture from Table 2 includes five nodes in a hidden layer and one node in the output layer. The input and output layers use linear transfer functions, and the hidden layer utilizes sigmoid transfer functions. A 10-fold cross validation methodology is used for generating training/test sets for each neural network's architecture [13]. The same training/test sets from the cross-validation process are applied to all feature combinations and classification algorithms presented. The 10-fold cross validation process is repeated five times for each feature combination. Classification results are based on averaging the area under receiver operating characteristic (ROC) curves [14] generated for each of the 10-test sets over the five separate 10-fold cross validation sets. The area under the ROC curve was utilized as the classification measure in this research because the area under the ROC curve does not require selecting a decision boundary or threshold to determine detection accuracy.

5. RESULTS AND DISCUSSIONS

Table 3 presents the average area under the ROC curve results over the five sets of 10-fold cross validation test sets for the different feature combinations and discrimination algorithms examined. BP denotes the standard backpropagation ANN algorithm. EANN-GA and EANN-PSO represent the EANN neural network approaches using the GA and PSO methods for neural network training, respectively.

Table 3. Average area under the ROC curve discrimination results for different feature and discrimination algorithm combinations.

Feature Combination	**Discrimination Algorithm**			
	BP	EANNE	EANN-GA	EANN-PSO
1	0.612	0.613	0.604	0.607
2	0.670	0.704	0.739	0.746
3	0.667	0.823	0.812	0.818
4	0.841	0.774	0.799	0.712
5	0.632	0.794	0.815	0.837
6	0.830	0.924	0.909	0.928
7	0.809	0.890	0.879	0.894
8	0.822	0.927	0.928	0.941
9	0.830	0.909	0.910	0.941
10	0.710	0.853	0.888	0.937
11	0.828	0.934	0.918	0.950
12	0.842	0.913	0.901	0.942
13	0.890	0.970	0.946	0.974
14	0.904	0.973	0.954	0.977
15	0.897	0.972	0.948	0.981

Several conclusions can be obtained from Table 3. First, the EANN algorithm using PSO neural network training typically yielded the highest overall area under the ROC curve results from the four discrimination algorithms investigated. Second, using all features (Feature

combination 15) gave the highest overall discrimination results, with an area under the ROC curve of 0.981 for the EANN-PSO approach. Feature combinations 13–15 yielded comparable results for the EANNE, EANN-GA, and EANN-PSO classification methods, with Feature combination 14 (general exam descriptors, common dermoscopic lesion features, and specific dermoscopic lesion features) providing slightly higher results than using all features (Feature combination 15). This is not a surprising result since the personal profile descriptors (Feature combination 1) yielded the overall lowest BCC discrimination results compared to the other individual feature groups. Third, inspecting the individual feature categories, the common dermoscopic lesion features yielded higher discrimination results than the other feature categories, except for the standard back propagation method. The personal profile descriptors gave consistently lower discrimination results than the other feature categories. Size and location information (general exam descriptors) and specific dermoscopic lesion features gave more discerning discrimination information than personal profile descriptors for all classification algorithms examined.

6. CONCLUSIONS

In this research, BCC versus benign lesion discrimination was explored using four types of clinical and dermoscopic lesion image features: patient personal profile, general exam, common dermoscopic image features associated with BCC, and specific dermoscopic image features with uncommon BCC presentations. Lesion discrimination was performed for different combinations of these features using EANNs and EANNEs architectures. Experiment results showed an area under the ROC curve as high as 0.981 when all features were combined. Common dermoscopic lesions features generally gave higher discrimination results than other individual feature categories. Overall, experimental results highlight that combining clinical and image information enhances lesion discrimination capability over either information source separately, demonstrating the potential of data fusion in aiding the lesion diagnostic process.

ACKNOWLEDGEMENTS

This publication was made possible by Grant Number SBIR R43 CA153927-01 of the National Institutes of Health (NIH). Its contents are solely the responsibility of the authors and do not necessarily represent the official views of the NIH.

REFERENCES

1. Rogers HW, Weinstock MA, Harris AR, Hinckley MR, Feldman SR, Fleischer AB, Coldiron BM. "Incidence estimate of nonmelanoma skin cancer in the United States". Arch Dermatol 2010, 146, pp.283–287.

2. Stoecker WV, Stolz W. "Dermoscopy and the diagnostic challenge of amelanotic and hypomelanotic melanoma". Arch Dermatol, 2008,144, pp. 1120–1127.

3. Cheng B, Erdos D, Stanley RJ, Stoecker WV, Calcara D, Gomez D. "Automatic detection of basal cell carcinoma using telangiectasia analysis in dermoscopy skin lesion images". Skin Res Technol 2011, 17, pp.278.

4. Cheng B, Stanley RJ, Stoecker WV, Hinton K. "Automatic telangiectasia analysis in dermoscopy images using adaptive critic design". Skin Res Technol, 2011, pp.1–8.

5. Cheng B, Stanley RJ, Stoecker WV, Osterwise CTP, Stricklin SM, Hinton KA, Moss RH, Oliviero M, Rabinovitz HS. "Automatic dirt trail analysis in dermoscopy images". Skin Res Technol, Jan 2012.

6. Stoecker WV, Gupta K, Shrestha B et al. "Detection of basal cell carcinoma using color and histogram measures of semitranslucent areas". Skin Res Technol, 2009,15, pp.283–287.

7. Menzies SW, Altamura D, Menzies SW *et al.* "Dermatoscopy of basal cell carcinoma: morphologic variability of global and local features and accuracy of diagnosis". J Am Acad Dermatol, 2010, 62, pp. 67–75.

8. Kennedy J, Eberhart R. "Particle swarm optimization". Proceedings of the IEEE International Conference on Neural Networks, Piscataway, NJ, 1995, pp. 1942-1948.

9. Holland JH. "Adaptation in Natural and Artificial Systems", University of Michigan Press, Ann Arbor, 1975.

10. Cheng B, Stanley RJ, De S, Antani S, Thoma GR. "Automatic detection of arrow annotation overlays in biomedical images". Int J Healthcare Inform Syst Informatics. 2011, 6, pp.23–41.

11. Yao X, Islam MM. "Evolving artificial neural network ensembles". Computational Intelligence Magazine, 2008, 3(1), pp.31-42.

12. Liu L, Yao X. "Negatively correlated neural networks can produce best ensembles", Aust. J. Intell. Inf. Proc. Syst., 1998, 4(3), pp.176–185.

13. Kohavi R. "A study of cross-validation and bootstrap for accuracy estimation and model selection", Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence, 1995, 14, pp. 1137–1143.

14. Fogarty J, Baker RS, and Hudson SE. “Case studies in the use of ROC curve analysis for sensor-based estimates in human computer interaction”. Proceedings of Graphics interface, Canadian Human-Computer Communications Society, School of Computer Science, University of Waterloo, Waterloo, Ontario Victoria, British Columbia, 2005, 112, pp. 129-136.

II. AUTOMATIC DIRT TRAIL ANALYSIS IN DERMOSCOPY IMAGES

Beibei Cheng[a], R. Joe Stanley[a], William V, Stoecker[b], Christopher T.P. Osterwise[a], Sherea M. Stricklin[b], Kristen A. Hinton[b], Randy H. Moss[a], Margaret Oliviero[c], Harold S. Rabinovitz[c]

[a]Department of Electrical and Computer Engineering, Missouri University of Science and Technology (S&T), Rolla, MO, 65409, USA

[\b]Stoecker & Associates, Rolla MO, 65401, USA

[c]Skin and Cancer Associates, Plantation, FL, USA

ABSTRACT

Background: Basal cell carcinoma (BCC) is the most common cancer in the US. Dermatoscopes are devices used by physicians to facilitate the early detection of these cancers based on the identification of skin lesion structures often specific to BCCs. One new lesion structure, referred to as dirt trails, has the appearance of dark gray, brown or black dots and clods of varying sizes distributed in elongated clusters with indistinct borders, often appearing as curvilinear trails.

Methods: In this research, we explore a dirt trail detection and analysis algorithm for extracting, measuring, and characterizing dirt trails based on size, distribution, and color in dermoscopic skin lesion images. These dirt trails are then used to automatically discriminate BCC from benign skin lesions.

Results: For an experimental data set of 35 BCC images with dirt trails and 79 benign lesion images, a neural network-based classifier achieved a 0.902 area under a receiver operating characteristic curve using a leave-one-out approach.

Conclusion: Results obtained from this study show that automatic detection of dirt trails in dermoscopic images of BCC is feasible. This is important because of the large number of these skin cancers seen every year and the challenge of discovering these earlier with instrumentation.

Key words: basal cell carcinoma – image analysis – dirt trails –neural network – dermoscopy

1. INTRODUCTION

Basal cell carcinoma (BCC), a slowly growing skin malignancy, is the most common cancer, with an estimate of 3 million new cases annually in the US [1]. To allow early detection of these skin cancers, physicians employ a device called a dermatoscope (3Gen LLC, San Juan Capistrano, CA; Heine Optotechnik, Herrsching, Germany). Classic basal cell carcinoma structures, visible and measurable with the dermatoscope, have been summarized by the BASAL acronym: Blue-gray ovoids and globules, Arborizing telangiectasia, Semitranslucency / Spoke wheel structures, Atraumatic ulcerations, and Leaf-like structures [2]. One newly described feature is brown to black dots/globules, which were found in 132 cases (21.7%) of a series of 609 basal cell carcinomas [3]. The distribution of these tiny pigmented structures has not been previously characterized. In one series of 351 basal cell carcinomas in a previous study, we found 46 cases (13.1%) of 351 BCCs to have the appearance of dark gray, brown or black dots and clods of varying sizes distributed in elongated clusters with indistinct borders, often appearing as curvilinear trails. These clustered objects may be characterized as dirt trails. Figure 1 gives two examples of dirt trails present in dermoscopy skin lesion images. In this research, we explore a dirt trail detection and analysis algorithm for extracting, measuring, and characterizing dirt trails based on size, distribution, and color in dermoscopic skin lesion images. These dirt trails are then used to automatically discriminate basal cell carcinoma from benign skin lesions. The following sections of the paper include: II. Methodology, III. Experimental Results, IV. Conclusion, and References.

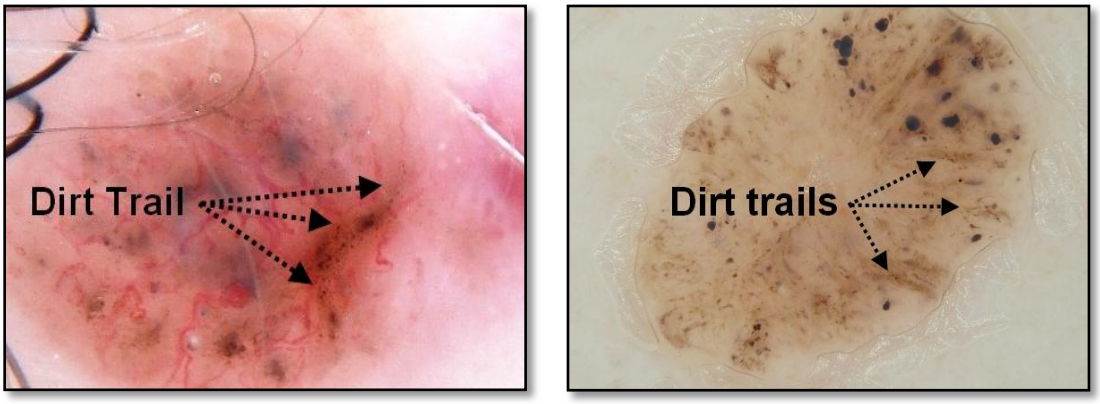


Figure 1. Dirt trail examples in dermoscopic skin lesion images, shown by arrows, with dirt trails containing dots and clods of varying sizes.

2. METHODOLOGY

2.1 DATA SET DESCRIPTION

The image set used for this study includes contact, non-polarized dermoscopic images of 35 basal cell carcinomas (BCCs) with dirt trail areas and 79 non-BCC benign lesions collected from two clinics: The Dermatology Center, Rolla, Missouri; and Skin and Cancer Associates, Plantation, Florida. All lesions with any dermoscopy features of malignancy and any benign lesions for which there was any uncertainty were biopsied. All images were contact, non-polarized dermoscopy images taken with a Sony DSC-W70 7.2 megapixel digital camera with a 3Gen DermLite Fluid dermoscopy attachment (3Gen LLC, San Juan Capistrano, CA). The 79-image competitive set to which the BCC set was compared consisted of a variety of lesions encountered in the clinic during the same period in which the BCC images were acquired, included 38 actinic keratoses (pre-cancers), 15 acquired nevocellular nevi (benign moles), 9 cases of sebaceous hyperplasia, and 15 cases of other benign diagnoses.

2.2 DIRT TRAIL DETECTION ALGORITHM OVERVIEW

An overview of the dirt trail detection algorithm is given in Figure 2. First, for the dermoscopy image data set, the individual red, green and blue (R, G, B) color planes were used for skin lesion analysis. Next, a Gaussian bandpass filter was applied to each color plane. To find the small dirt trail dots and clods, the bandpass-filtered image was next subtracted from each 3x3 median-filtered image. A scalarized Otsu threshold was then applied, followed by hair and bubble noise filtering. Features were determined for the resultant objects remaining in the mask. A backpropagation neural network using a leave-one-out method was applied to these features to determine presence or absence of dirt trails. The dirt trail detection algorithm, feature extraction and lesion discrimination are presented in detail in sections II.3 and II.4, respectively.

3. DIRT TRAIL DETECTION ALGORITHM

3.1 GAUSSIAN BANDPASS FILTER

The first step in the dirt trail detection algorithm is to obtain the individual *R*, *G*, and *B* color planes for the skin lesion image (Figure 3). The second step is to apply Gaussian bandpass filtering to the individual color planes. A Gaussian lowpass filter, denoted as *H*, defined for each pixel position (*u*,*v*) is constructed and is given in Eqs. 1 and 2[4]:

$$D(u,v) = \left[\left(u - \frac{M}{2}\right)^2 + \left(u - \frac{N}{2}\right)^2\right]^{\frac{1}{2}} \tag{1}$$

$$H(u,v) = e^{-\frac{D(u,v)^2}{2{D_0}^2}} \tag{2}$$

where *M* and *N* represent the width and height of the image and (*M/2*, *N/2*) is the center of the frequency rectangle. *H(u,v)* is the resulting lowpass filtered image based on the value for D_0. Two different Gaussian lowpass filters were determined empirically to find dirt trail objects by setting D_0 to be 40 and 100. The difference between these lowpass filters, denoted as *W(u,v)*, represents the bandpass filter, as given in Eq. (3):

$$W(u,v) = e^{-\frac{D(u,v)^2}{2(100^2)}} - e^{-\frac{D(u,v)^2}{2(40^2)}} \tag{3}$$

Figure 4 shows the representation for the bandpass filter *W*. *W* is applied to the spatial frequency domain representation from the discrete Fourier transform for each of the color plane images for the skin lesion (see Figure 3). The Gaussian filtered images for the *R*, *G*, and *B* planes are determined based on the frequency domain and converted to the spatial domain. The resulting bandpass images for the *R*, *G*, and *B* planes are denoted as W_R, W_G, and W_B respectively. Figure 5 presents examples of the bandpass filter process for each color plane, with the original color plane image on the left side and the filtered image on the right side.

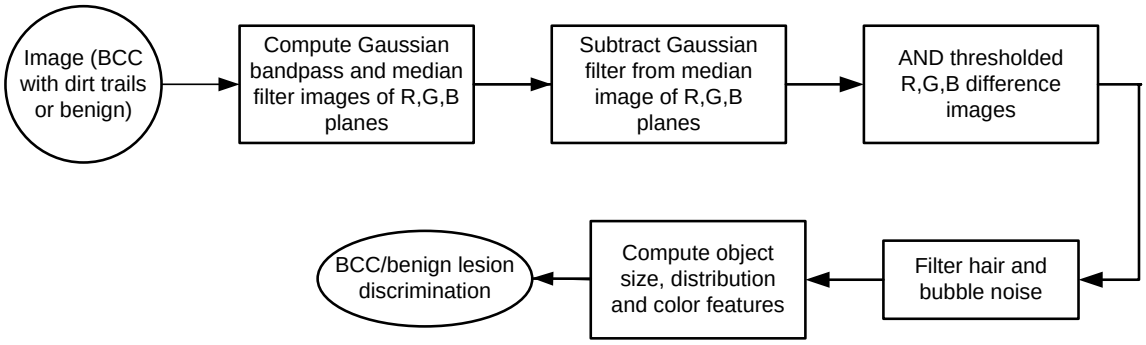
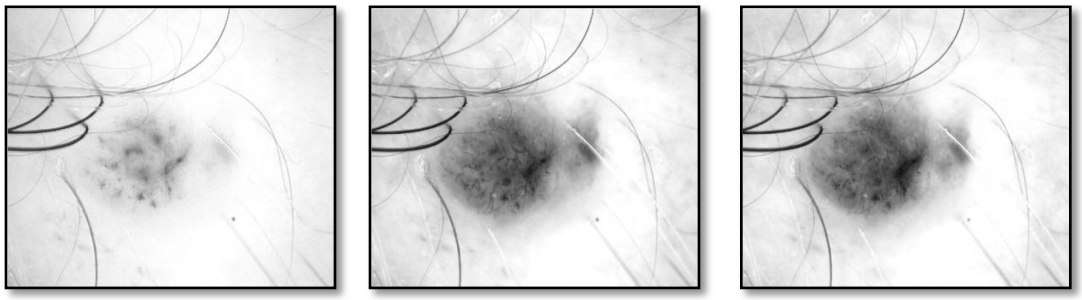


Figure 2. Overview of the dirt trail detection algorithm.



(a) (b) (c)

Figure 3. RGB plane. (a) Red plane. (b) Green plane. (c) Blue plane.

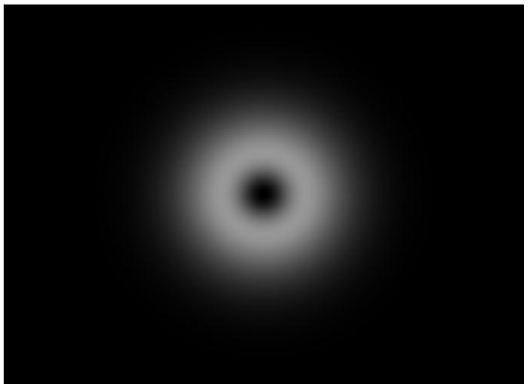
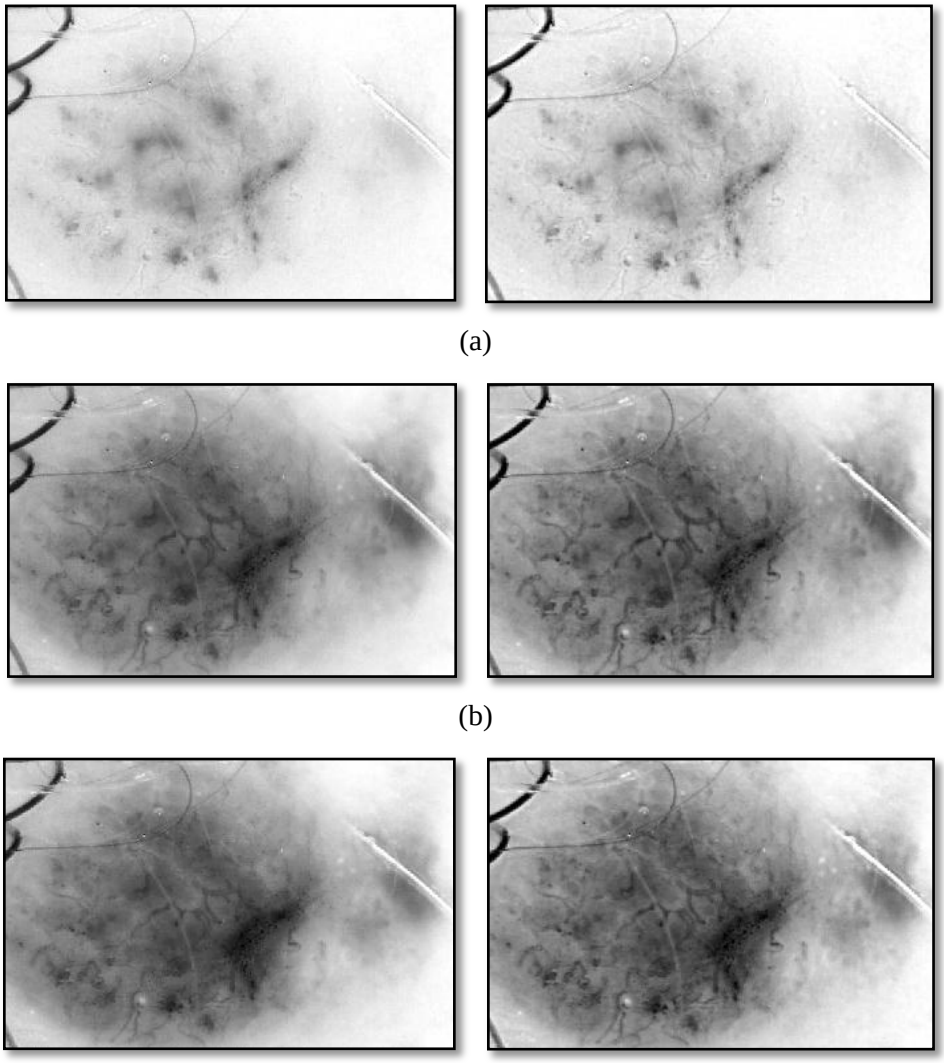


Figure 4. Gaussian bandpass filter representation in the spatial frequency domain.

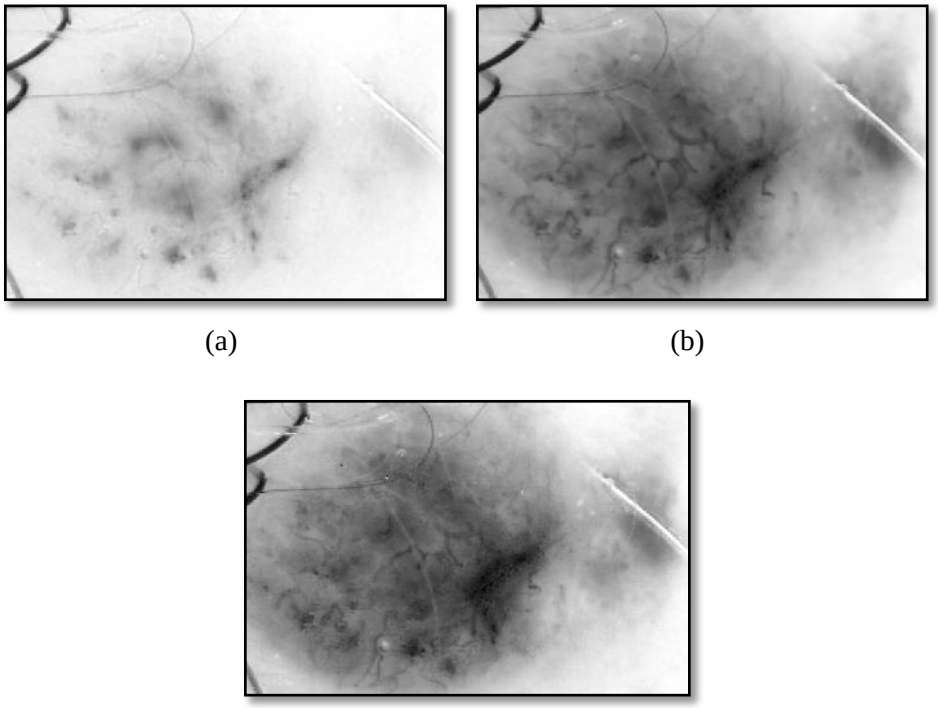


(a) (b) (c)

Figure 5. Bandpass-filtered images converted to spatial domain for *R*, *G*, and *B* planes. (a) Red plane. (b) Green plane. (c) Blue plane. The original color plane images are on the left, and the filtered images W_R, W_G, and W_B are on the right.

3.2 MEDIAN FILTER

Since the dirt trail resembles small salt-and-pepper noise, a 3x3 median filter is applied to each original color plane image, with median filter results shown in Figure 6 for the individual color plane images from Figure 3. Let F_R,F_G, and F_B denote the median-filtered images for the *R*, *G*, and *B* color planes, respectively.



(a) (b)

(c)

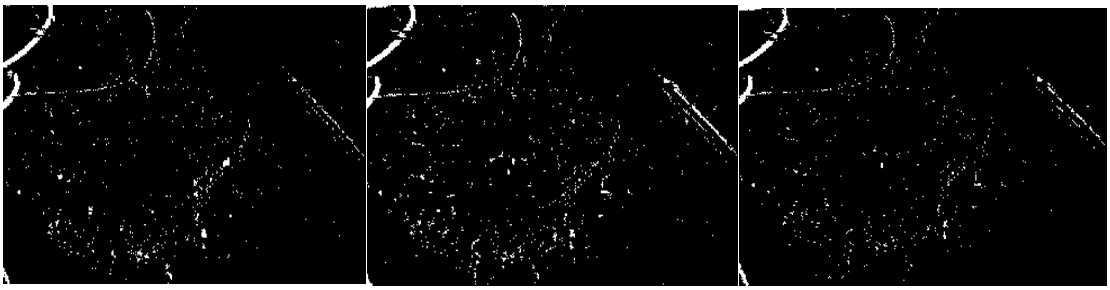
Figure 6. Median filter output images from *R*, *G*, *B* planes. (a) F_R, (b) F_G, (c) F_B.

3.3 IMAGE SUBTRACTION

The small dirt trail objects are found by computing the difference between the bandpass filtered image and the median filtered image for each color plane. Let S_R,S_G, and $S_B$represent the difference images for the *R*, *G*, and *B* color planes, respectively, with $S_R = F_R - W_R$, S_G and S_B are similarly defined. This corresponds to subtracting the corresponding color plane images, Figure 5, from the median filtered images, Figure 6.

3.4 HISTOGRAM PROCESSING

Using the difference images S_R,S_G, and S_B for the pixels inside lesion border, the Otsu's method is implemented for these pixels to find the histogram threshold [5], with the threshold multiplied by a scalar of 2, determined empirically from the experimental data set, in order to increase the sensitivity of dirt trail detection. Let T_R, T_G, and $T_B$denote the threshold images for the *R*, *G*, and *B* color planes, respectively. These are shown in Figure 7.



(a) (b) (c)

Figure 7. Output images from scalarized Otsu's method for *R*, *G*, *B* planes.
(a) T_R, (b)T_G, (c)T_B.

3.5 LOGIC OPERATION

In order to extract dirt trail-like objects, the threshold images for the different color planes are merged using a logical AND operation, as given in equation 4 and denoted as *A*. Figure 8 gives an AND image example for the threshold images from Figure 7.

$$A = T_R \cap T_G \cap T_B \quad (4)$$

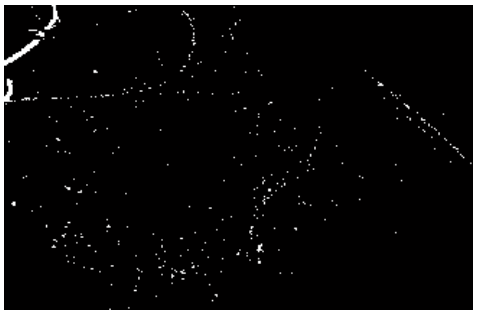


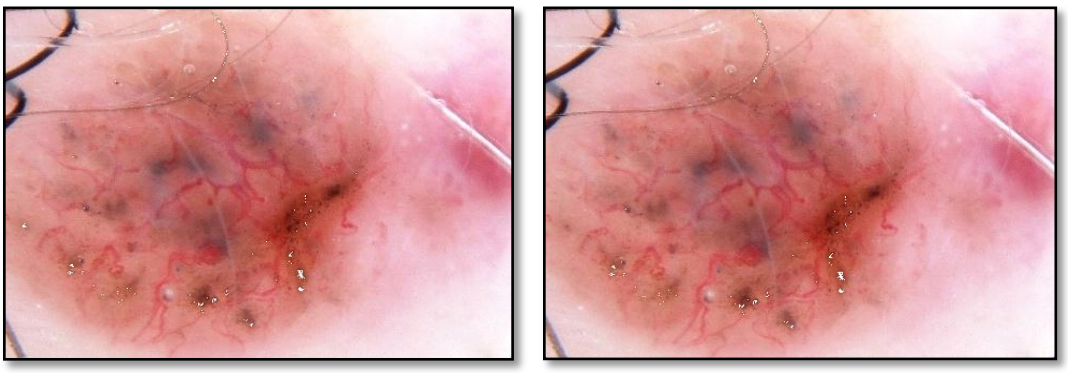
Figure 8. Otsu's output image *A* after logical AND of the threshold color plane images.

3.6 NOISE FILTERING FOR HAIRS AND BUBBLES

The logical AND of the threshold images provides an initial mask for dirt trail-like objects. A mathematical morphological-based approach was applied in order to filter hairs and bubbles commonly observed in dermoscopy skin lesion images [6]. Let *B* represent the resulting hair and bubble objects detected from *A*. Then, the resultant mask is given by $R = A - B$.

3.7 NOISE FILTERING FOR ISOLATED NOISE

A secondary noise filter was applied to R in order to remove isolated objects I. This step was motivated by the fact that dirt trails consist of a cluster of objects (black dots and clods). Each object in R was given a blob label. All objects within an empirically determined radius of 300 pixels of the object's centroid were counted. If the number of objects within this radius was less than 10, the isolated noise object I was removed from R, to create the final dirt trail mask $K = R - I$. Let K represent the final dirt trail mask after performing the clustering operation. An example image is given in Figure 9, with overlays on the original color image in (a) showing the mask R after hair and bubble removal and (b) the dirt trail mask K after hair and bubble and isolated object removal.



(a) (b)

Figure 9. Image overlay. (a) Image overlay R, after hair and bubble removal. (b) Dirt trail image overlay K after isolated object removal.

4. CLASSIFIER INPUT FEATURES AND CLASSIFIER METHODOLOGY

4.1 FEATURES COMPUTED FOR LESION DISCRIMINATION

Using the final dirt trail mask K from the previous section and the RGB skin lesion image, nine features were calculated from each dirt trail mask for each image for lesion discrimination, as shown in Table 1. These nine features were selected from known dirt trail features—they are dark elongated areas within a relatively bright, pink area of the BCC.

Table 1. Descriptions for Dirt Trail Dermoscopic Features

Feature	**Description of features for objects within lesion**	**Meaning**	**p-val, Wald Chi-sq ***
OBN	Number of objects	BCC have more dirt trails	0.3301* n.s.
AREA_AV	Average area of objects	BCC have clods in trails	0.4335 n.s.
MAX_EC	Maximum eccentricity of objects: ratio of elliptical axes	Dirt trail objects are longer	0.1902 n.s.
AVE_EC	Average eccentricity of objects: ratio of elliptical axes	Dirt trail objects are longer	> 0.5 n.s. (feature not selected)
RED_AV	Average red value of objects	BCC trails are darker	0.0689 n.s.
GREEN_AV	Average green value of objects	BCC trails are darker	0.0108
BLUE_AV	Average blue value of objects	BCC trails are darker	0.0085
REL_BLUE	Ratio of BLUE_AV to average blue surrounding object	BCC surrounds are pink	0.2826 n.s.
GR/BLUE	Ratio GREEN_AV/ BLUE_AV	BCC objects are greener (brighter)	0.1232 n.s.

*p-values from SAS model for all variables in logistic regression model with $p < 0.5$, maximum likelihood estimate, Wald Chi-sq.

4.2 CLASSIFIER DESCRIPTION

The dirt trail detection algorithm and lesion features presented in the previous section were examined for BCC versus benign lesion discrimination in the data set of 35 BCC images with dirt trails and 79 benign lesion images described above. BCC versus benign skin lesion discrimination was done based on a standard backpropagation neural network classifier. For the neural network classifier, an architecture of 11x5x1 was used: 10 features and a bias for the input layer, 5 nodes in a hidden layer, and one output was employed. Linear transfer functions were used for the input and output layers, and sigmoid transfer functions were utilized in the

hidden layer. A leave-one-out methodology was used for training and testing the neural network with each network being trained for a maximum of 200 epochs (on-line weight updating) or RMSE < 0.001. Discrimination results were generated and examined using a receiver operating characteristic (ROC) curve, varying the neural network threshold and computing the sensitivity vs. 1-specificity (true positive and false negative rates) at each threshold value.

5. EXPERIMENTAL RESULTS

5.1 DIRT TRAIL MASK EXAMPLES

Figure 10 presents the dirt trail detection masks for a BCC lesion and a benign lesion. As observed in Figure 10, the detection algorithm finds dirt trail regions with some false positive areas. Features are computed based on the dirt trail masks found from the BCC and benign lesion data set. Ideally, the dirt trail mask is empty for the benign lesions. We found that 51 of the 79 benign lesions contained false positive objects similar to dirt trails. Accordingly, features are calculated based on the generated dirt trails mask determined for each lesion image.

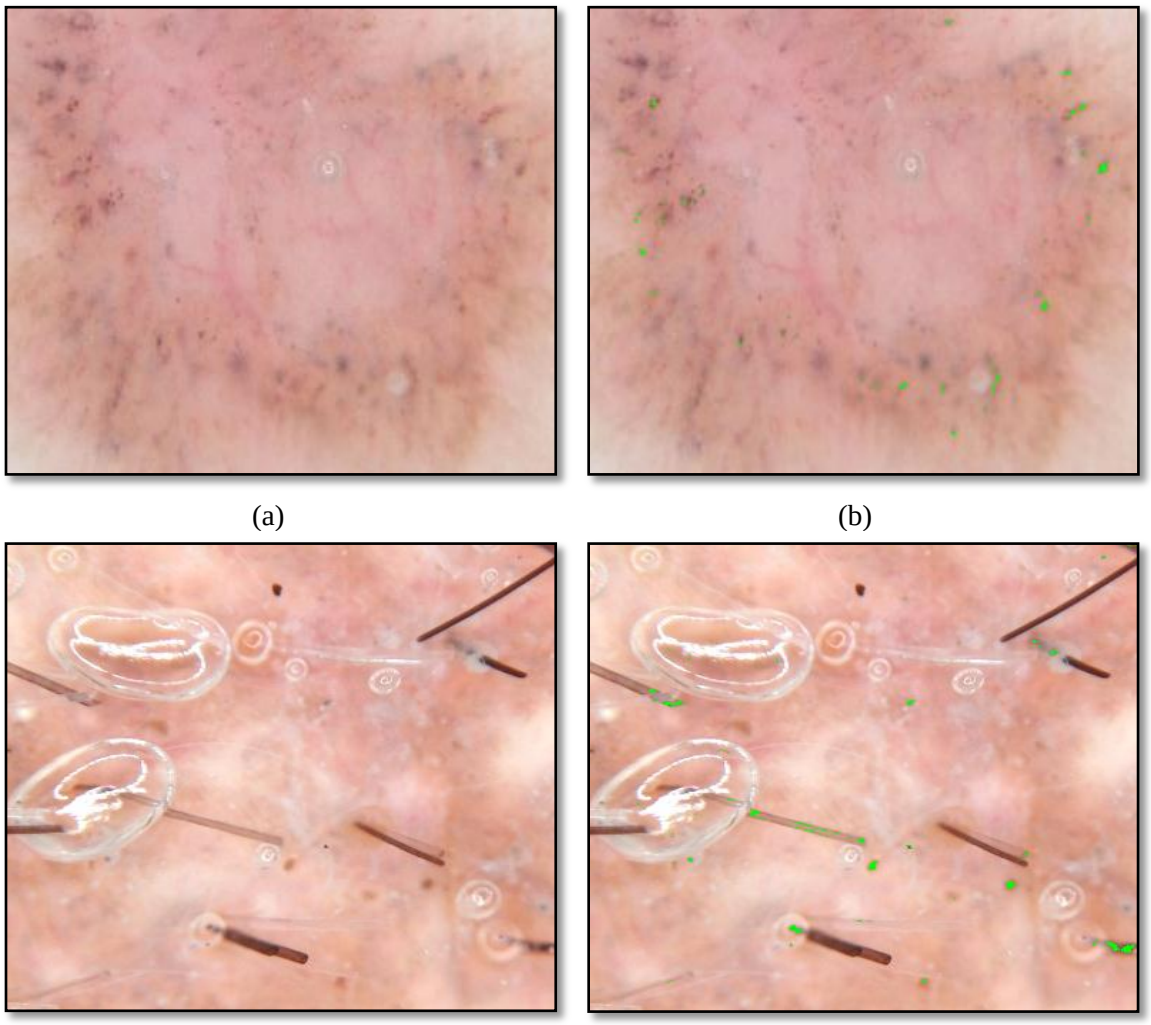
5.2 FEATURE ANALYSIS USING LOGISTIC REGRESSION

Statistical analysis using Wald Chi-Square maximum likelihood estimates [7] was performed on the features computed for the BCC and benign lesions for objects found in the experimental test set of lesions in order to evaluate the discrimination capability of those features. Table 1 presents the statistical results for these features for the nine features analyzed.

The features are shown in the order selected by logistic regression, using SAS software (SAS Institute, Cary, NC), with the features OBN, the number of objects, selected first. Eight of the nine features were included in the model, with only one feature, AVE_EC, the average eccentricity of the objects, not selected. The other eight features were all included in the final model. Only two features reached significance using the Wald Chi-Square maximum likelihood estimates--GREEN_AV and BLUE_AV. The blue and green planes distinguish the dirt trails from their surrounds better than the red plane, Figure 3.

5.3 LESION DISCRIMINATION RESULTS

Figure 11 shows the plot of the ROC curve and the area under the ROC curve, denoted as AUC, for the neural network results based on the above features using on-line neural network training and leave-one-out training and testing.



(a) (b)

(c) (d)

Figure 10. Dirt trail detection mask examples. (a) Dirt trail images. (b) Dirt trail image overlays. (c) Benign images. (d) Benign image overlays.

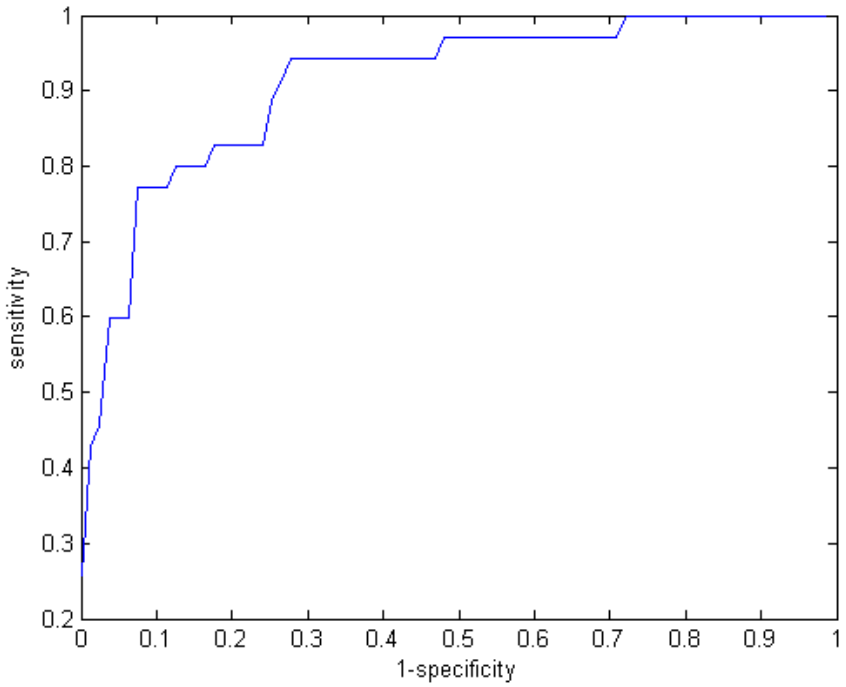


Figure 11. ROC curve and AUC (area under curve) for backpropagation neural network. AUC=0.902.

6. CONCLUSIONS

Results obtained from this study show that automatic detection of dirt trails in dermoscopic images of BCC is feasible. This is important because of the large number of these skin cancers seen every year and the challenge of discovering these earlier with instrumentation. Filtering techniques developed in this study include the combined use of band-pass Gaussian filtering and median filtering. In this manner, the primary filter is a spatial screen for the small objects and the secondary filter removes midrange frequencies. Dot-size noise can easily mimic the objects seen in dirt trails. Since dot-size objects are very common in benign lesions [8,9], the neural network classifier results of *AUC =0.902* are higher than might be expected in this noisy environment. The features in the individual objects therefore have meaningful information for detecting dirt trails, allowing discrimination of BCC from benign skin lesions.

Future studies could incorporate assessments of cluster elongation and global distribution analysis using lesion deciles [10]. It is likely that parameters describing paracentral location of dirt trail objects will be useful in achieving better discrimination results. In addition, further optimization of Gaussian bandpass filter parameters may improve classification results.

ACKNOWLEDGMENTS

This publication was made possible by Grant Number SBIR R43 CA153927-01 of the National Institutes of Health (NIH). Its contents are solely the responsibility of the authors and do not necessarily represent the official views of NIH.

REFERENCES

1. H.W. Rogers, M.A.Weinstock, A.R. Harris, M.R. Hinckley, S.R. Feldman, A.B. Fleischer, B.M. Coldiron, "Incidence estimate of nonmelanoma skin cancer in the United States, 2006", Arch. Dermatol., 2010, 146, pp. 283-7.

2. W.V. Stoecker, W. Stolz. "Dermoscopy and the diagnostic challenge of amelanotic and hypomelanotic melanoma," Arch. Dermatol., 2008, 144, pp. 1120-7.

3. D. Altamura, S.W. Menzies, G. Argenziano, I. Zalaudek, H.P. Soyer, F. Sera, M. Avramidis, K. DeAmbrosis, M.C. Fargnoli, K. Peris, "Dermatoscopy of basal cell carcinoma: morphologic variability of global and local features and accuracy of diagnosis," J. Am. Acad. Dermatol. 2010, 62, pp. 67-75.

4. R.C. González, R.E. Woods, Digital Image Processing, Pearson/Prentice Hall, Upper Saddle River, NJ: Prentice Hall, 2008.

5. N. Otsu, "A Threshold Selection Method from Gray-Level Histograms," IEEE Transactions on Systems, Man, and Cybernetics, 1979, 9, pp. 62-66.

6. H. Wang, R.H. Moss, X. Chen, R.J. Stanley, W.V. Stoecker, E. Celebi, J.M. Malters, J.M. Grichnik, A.A. Marghoob, H.S. Rabinovitz, S.W. Menzies, T.M. Szalapski, "Modified watershed technique and post-processing for segmentation of skin lesions in dermoscopy images," Computerized Medical Imaging and Graphics, Special Dermatology Issue, 2011, 35(2), pp. 116-120.

7. G.W. Corder, D.I. Foreman, Nonparametric Statistics for Non-Statisticians: A Step-by-Step Approach Wiley, 2009.

8. J. Xu, K. Gupta, W.V. Stoecker, Y. Krishnamurthy, H.S. Rabinovitz, A. Bangert, D. Calcara, M. Oliviero, J.M. Malters, R. Drugge, R.J. Stanley, R.H. Moss, M.E. Celebi . "Analysis of globule types in malignant melanoma," Arch Dermatol., Nov 2009, 11, pp.1245-51.

9. T.G. Salopek, A.W. Kopf, C.M. Stefanato, K. Vossaert, M. Silverman, S. Yadav, "Differentiation of atypical moles (dysplastic nevi) from early melanomas by dermoscopy," Dermatol Clin, 2001, 19, pp. 337-345.

10. A. Dalal, R.H. Moss, R.J. Stanley, W.V. Stoecker, K. Gupta, D.A. Calcara, J. Xu, B. Shrestha, R. Drugge, J.M. Malters, L.A. Perry, "Concentric decile segmentation of white and hypopigmented areas in dermoscopy images of skin lesions allows discrimination of malignant melanoma," Comput Med Imaging Graph., 2011, 35, pp. 148-54.

III. AUTOMATIC TELANGIECTASIA ANALYSIS IN DERMOSCOPY IMAGES USING ADAPTIVE CIRTIC DESIGN

B. Cheng[a], R.J. Stanley[a], W. V. Stoecker[b] and K. Hinton[b]

[a]Department of Electrical and Computer Engineering, Missouri University of Science and Technology, Rolla, MO

[b]Stoecker and Associates, Rolla, MO

ABSTRACT

Background: Telangiectasia, tiny skin vessels, are important dermoscopy structures used to discriminate basal cell carcinoma (BCC) from benign skin lesions. This research builds off of previously developed image analysis techniques identified vessels automatically to discriminate benign lesions from BCCs.

Methods: A biologically-inspired reinforcement learning approach is investigated in an Adaptive Critic Design framework to apply Action Dependent Heuristic Dynamic Programming (ADHDP) for discrimination based on computed features using different skin lesion contrast variations to promote the discrimination process. Lesion discrimination results for ADHDP are compared with multi-layer perception (MLP) backpropagation artificial neural networks.

Results: This study uses a data set of 498 dermoscopy skin lesion images of 263 BCCs and 226 competitive benign images as the input sets. This data set is extended from previous research [12]. Experimental results yielded a diagnostic accuracy as high as 84.6% using the ADHDP approach, providing an 8.03% improvement over a standard MLP method.

Conclusions: We have chosen BCC detection rather than vessel detection as the endpoint. Although vessel detection is inherently easier, BCC detection has potential direct clinical application. Small BCCs are detectable early by dermoscopy and potentially detectable by the automated methods described in this research.

Key words: Image processing, telangiectasia, backpropagation, adaptive critic design, ADHDP

1. INTRODUCTION

Telangiectasia are dilated blood vessels of varying diameter within the superficial dermis. These vessels are observed with a number of diseases such as rosacea and are a prominent feature of basal cell carcinoma (BCC), the most common skin cancer [1]. They are commonly present as a background skin feature in fair-skinned persons, especially in sun-exposed areas in older persons. Telangiectasia can best be visualized with dermoscopy, using either a glass plate with fluid interface (contact non-polarized dermoscopy) or cross-polarized lighting (non-contact polarized dermoscopy), together with 10-power magnification. In BCC, the classical form of telangiectasia is termed "arborizing telangiectasia," with a thick central trunk vessel with narrow radiating branch vessels [2, 3]. Figure 1 shows a BCC with arborizing telangiectasia and fine telangiectasia, in addition to background telangiectasia resulting from sun damage. The goal of the image analysis is to differentiate the types of telangiectasia shown in Figure 1 in order to discriminate BCC from non-malignant lesions.

Numerous image processing approaches from various medical domains [4-11] have been reviewed. These techniques from other (non-cutaneous) domains rely on monochromatic methods and are adapted for larger vessels, inappropriate for telangiectasia. Therefore a pixel-based color drop algorithm was developed for optimal image processing in [12].

Figure 2 presents the overall skin lesion discrimination process, including image processing and classification, that is explored in this research. From Figure 2, the image analysis techniques investigated to detect and characterize telangectasias extend previous research [12]. An overview of image analysis algorithms employed is shown in Figure 3. Image processing techniques are used to produce vessel masks, which generate feature sets used as inputs to different classifiers to make the benign/malignant decision.

The data set consisted of 263 basal cell carcinoma (BCC) dermoscopy images with visible telangiectasia obtained from two clinics: Skin and Cancer Associates, Plantation, FL, and the Dermatology Center, Rolla, MO. Two hundred twenty-six benign dermoscopy images were selected from benign lesions in the same study for image processing algorithm development, including: 55 dysplastic nevi (45 with mild atypia and 10 with moderate atypia), 50 seborrheic keratoses, 63 benign acquired nevi, 17 lentigines, 13 actinic keratoses, 1 hemangioma, 8 lichen-planus like keratoses, 4 dermatofibroma, 2 warts, and 13 single examples of various benign diagnoses were identified in the same dermoscopy study and used as the competitive test set. All images were taken with the HR II DermLite (3Gen, Dana Point, CA) using ultrasonic gel for the fluid interface. Lesions were included in the study if by clinical or

dermoscopy evaluation they were malignant, potentially malignant, changing, or of interest to physician and/or patient. Lesions that were biopsied for cosmetic reasons were not included in the study. The data set was of moderate diagnostic difficulty and was felt to represent basal cell carcinomas and benign lesions encountered in the clinic. The Phelps County Regional Medical Center Institutional Review Board, Rolla Missouri, approved this research and each subject or subject's parent or guardian signed a consent form for this research. All basal cell carcinomas in the study had histopathology examined by a dermatologist or a dermatopathologist.

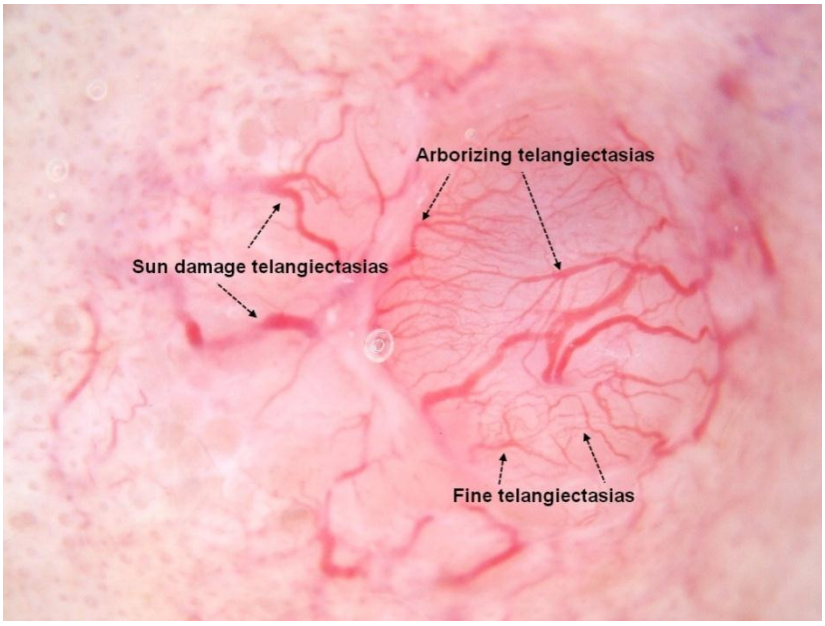


Figure 1. Telangiectasia example. Arborizing telangiectasia (trunk and branches) are the classical telangiectasia seen in a contact, non-polarized dermsocopy image of basal cell carcinoma (BCC). Fine telangiectasia are more numerous than arborizing telangiectasia in BCC. Note that sun damage telangiectasia are wider, shorter, have less sharp edges, have a greater variation in width, and are more sparse than BCC telangiectasia.

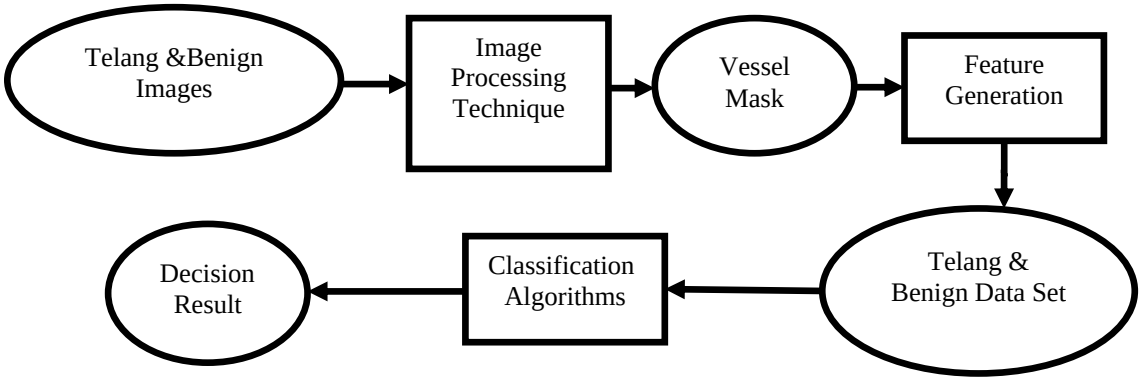


Figure 2. Overview of vessel detection and discrimination process.

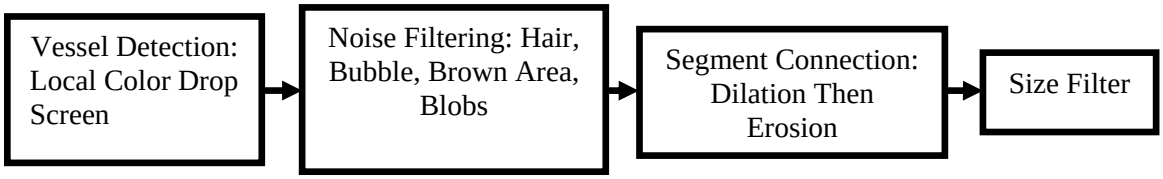
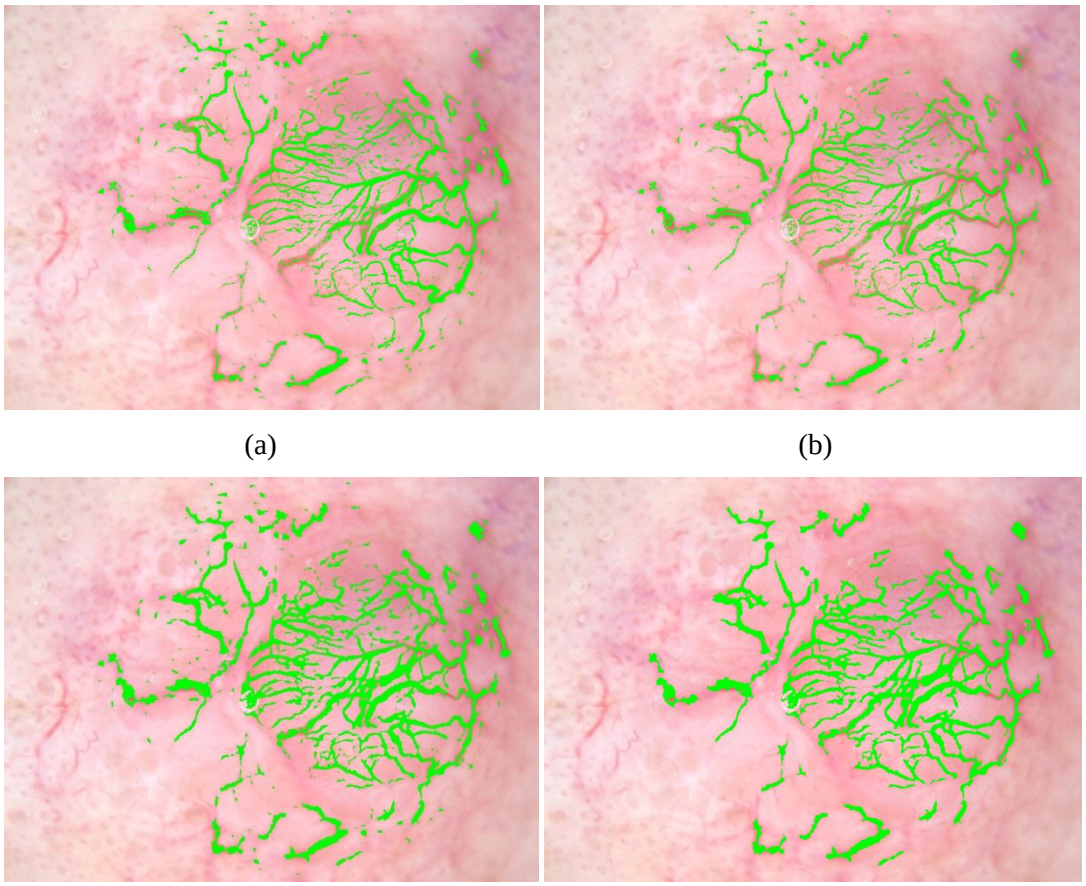


Figure 3. Overview of image analysis algorithm.

From Figure 3, image processing techniques including a pixel-based color drop algorithm, noise filtering, segment connection and size filtering, previously described in detail [12]. The color drop algorithm exploits known specific sharp color gradients found predominantly at vessel edges. It iterates through every pixel inside the lesion, labeling a new designated center pixel, then moves outward a set number of pixels from this given center in 8 directions, and requires a minimum color drop, optimizing gradients for each color channel [12]. Noise sources such as hair and bubbles are subsequently eliminated by specific morphologic and size filters [12]. Vessel segments are connected using a closing operation and size filtering limits noise. Figure 4 presents an example of the image processing steps for the original image in Figure 1 used to generate the vessel mask for feature calculations.

Contrast enhancement is needed as a pre-processing step to detect the low-contrast vessels. In this research, seven different contrast enhancement variations are applied to the image data set, generating seven data sets for vessel mask determination and feature and discrimination analysis. These seven contrast enhancement masks provide multiple perspectives

of the skin lesions for feature extraction and data fusion analysis using reinforcement learning techniques for lesion discrimination.



(a) (b)

(c) (d)

Figure 4. Examples of telangiectasia in the skin image from Figure 1 and the vessel mask generated using the algorithm in Figure 3. (a) Vessel mask after color drop algorithm. (b) Vessel mask after noise filtering. (c) Vessel mask after segmentation. (d) Vessel mask after size filter.

Reinforcement learning (RL) is the problem faced by an agent that must learn behavior through trial-and-error interactions with a dynamic environment. It is a computational approach to learning whereby an agent tries to maximize the total amount of reward it receives when interacting with a complex, uncertain environment [13]. RL has been developed in various

applications such as neuro-computing [14], multi-resolution object recognition [15] and image artifacts detection [16].

The adaptive critic design (ACD) technique provides a workstation for implementing RL. ACD approximates the neuro-dynamic programming by using an action and a critic network, respectively [17]. An RL-based ACD network is applied to the seven features data sets. One ACD algorithm, direct neural dynamic programming (NDP), is implemented as the classifier. In direct NDP, two generic (neural) function approximators are used to represent both the value function (for control performance evaluation) and the action function (for control law generation – policy function). Furthermore, the state information is used directly in learning the control law where the controller parameters are the weight parameters in the action function implemented by a neural network. Direct NDP is a model independent approach to action dependent heuristic programming (ADHDP) [18]. This research introduces the application of ACD to skin lesion discrimination. The application incorporates feature analysis of the different skin lesion contrast variations to promote the discrimination process.

The remaining sections of this paper include: 1) Problem Description, 2) Adaptive Critic Design Methodology, 3) Benchmark Classification Technique, 4) Results and Discussion, and 5) Conclusion and Future Work.

2. PROBLEM DESCRIPTION

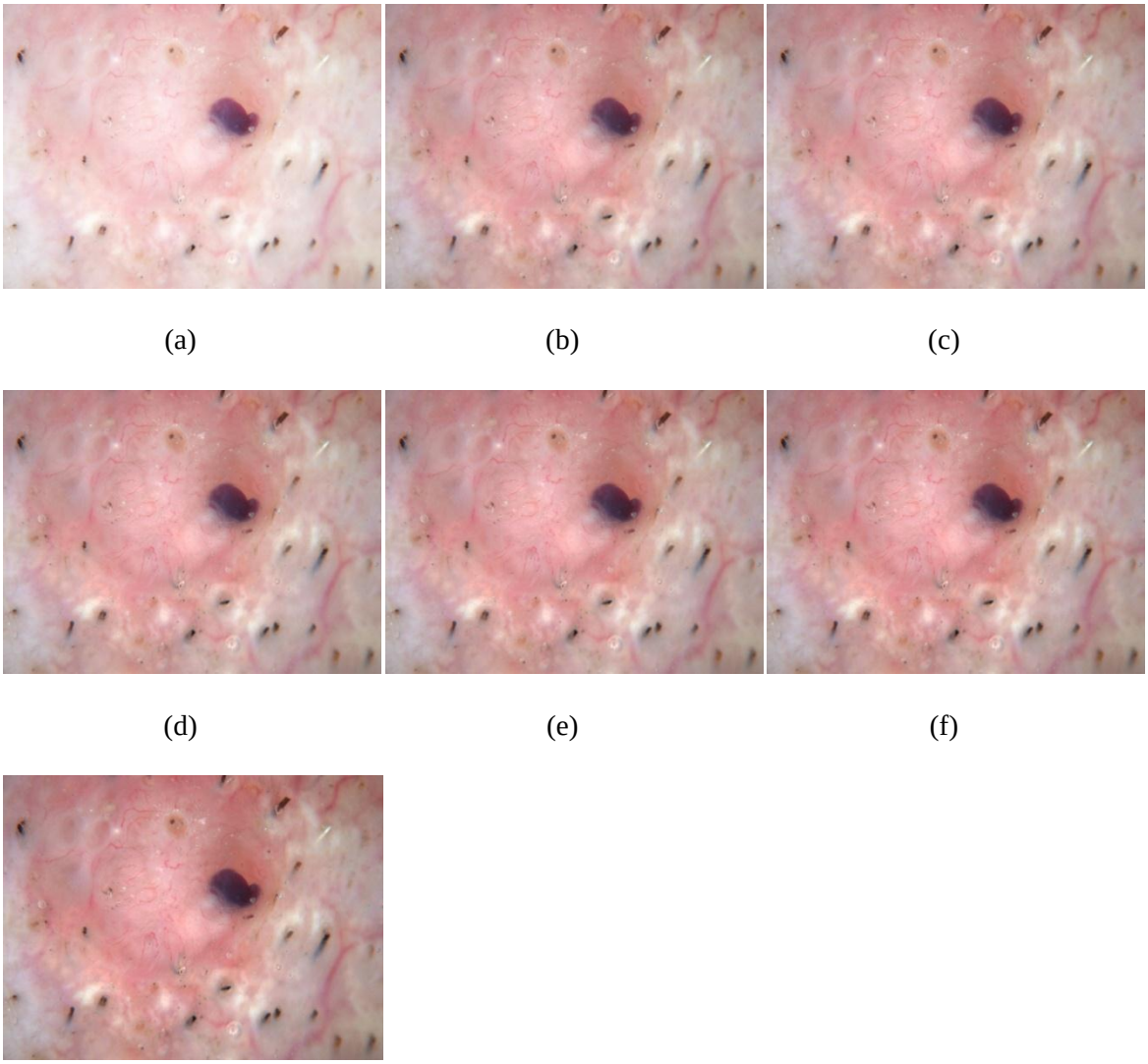
Figure 5 shows a lesion image with telangiectasia and its output with contrast enhancement by using a gain and an offset to stretch the dynamic range of the image [19]. For a scene with dynamic range between I_{max} and I_{min}, and a display medium with dynamic range d_{max}, the contrast enhancement (CE) is given for each pixel (x,y) as

$$R_i(x,y) = \frac{d_{max}}{I_{max} - I_{min}} (I_i(x,y) - I_{min}) \quad (1)$$

where, d_{max} is 255; I_{max} is the maximum gray value; I_{min} is the minimum gray value, varied over 6 different CE values: $0.05I_{max}$, $0.1I_{max}$, $0.15I_{max}$, $0.2I_{max}$, $0.25I_{max}$, and $0.3I_{max}$. Image processing with I_{min} greater than $0.3I_{max}$ generates too many false vessels after implementing image processing techniques.

Based on varying I_{min}, seven total images of the lesion are generated, six CE images and the original lesion image. The telangiectasia detection algorithm is applied to the different contrast enhanced images for each lesion. The seven vessel mask overlays are called state 1 to state 7 (Figure 6). As shown in Figure 6, more vessels can be detected with increasing the contrast. However, with increased contrast, more false vessels are found as well. An ACD-based telangiectasia-discrimination classifier is then applied to those seven data sets. The input to this algorithm is the features generated from the vessel mask.

The features utilized to provide numerical descriptors of the vessel-like objects in the vessel mask for each image are given in Table 1. The physical motivation and the Chi-Square statistical analysis for selecting these features are presented in [12].

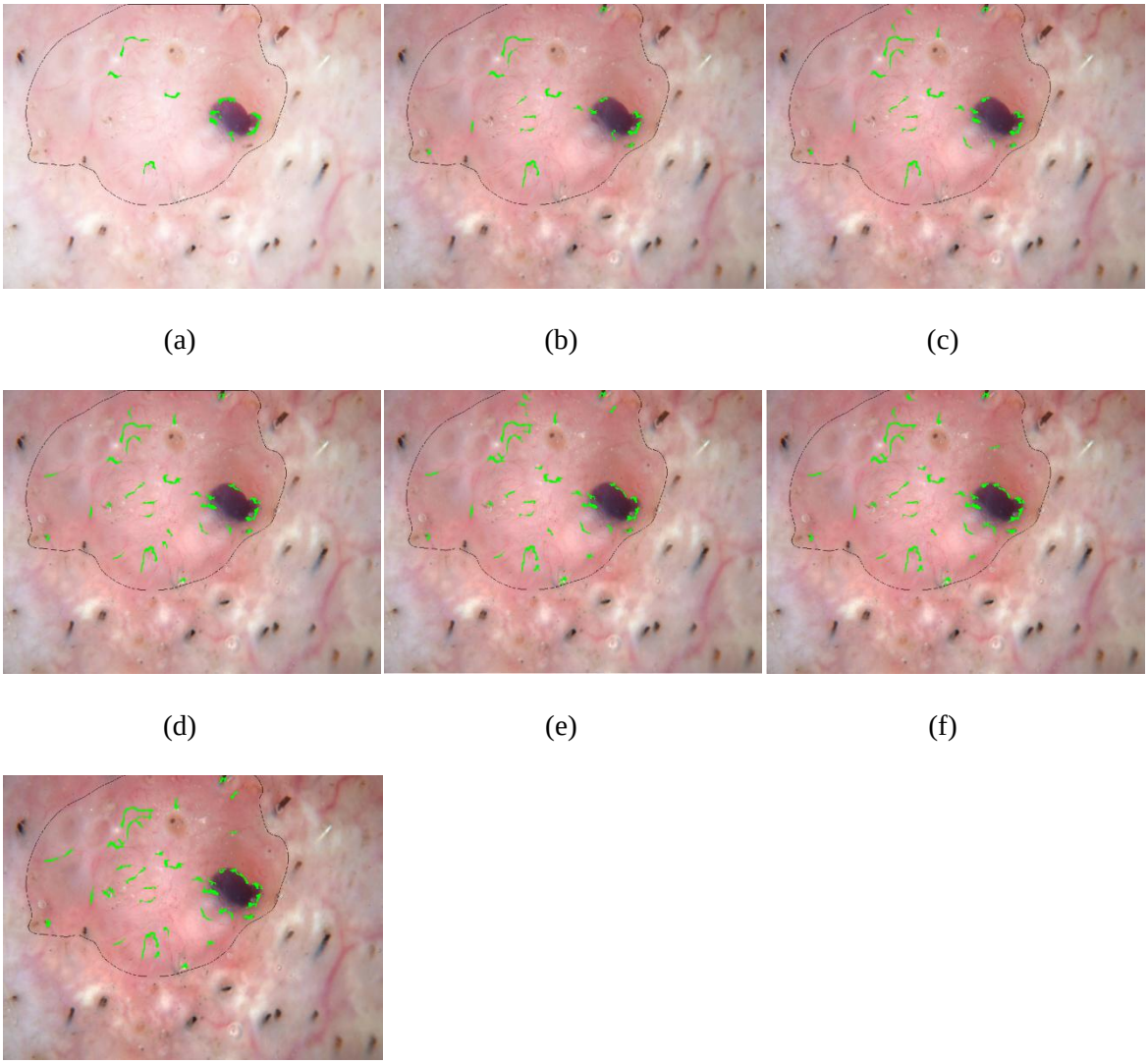


(a) (b) (c)

(d) (e) (f)

(g)

Figure 5. Images with varied contrast enhancement (CE). (a) Original image (b) I_{min}/I_{max} =0.05 (c) I_{min}/I_{max} =0.10 (d) I_{min}/I_{max}=0.15 (e) I_{min}/I_{max}=0.20 (f) I_{min}/I_{max}=0.25 (g) I_{min}/I_{max}=0.3



(a) (b) (c)

(d) (e) (f)

(g)

Figure 6.Vessel mask overlays change with varying contrast enhancement.

(a) State 1. (b) State 2.

(c) State 3. (d) State 4. (e) State 5. (f) State 6. (g) State 7.

Table 1. Vessel-based features investigated with descriptions

No.	Measure (all normalized except those marked*)	Description (measures marked # taken after skeletonizing)	Meaning
1-4	Object area after 1,2,9,10 erosions	Erode the vessel mask with circular structuring element of radius from 1 to 10 and record the remaining mask area for each erosion and divide by square root of the lesion area	BCC object areas smaller after given number erosions
5-8	Object number after 1, 2, 6, 7 erosions*	Erode the final vessel mask with circular structuring element of radius from 1 to 10 and record the remaining object number after each erosion	BCC object fewer after given number erosions
9	Maximum object length	Maximum length for all vessels/square root lesion area#	BCC vessels are longer
10	Average object width	Average width for all vessels/square root lesion area	BCC vessels are narrower
11	Standard deviation object width*	Standard deviation for all vessel widths	BCC vessel widths are more uniform
12-13	Maximum/ Average eccentricity*	Maximum/Average ratio of distance between the foci of the ellipse enclosing the vessels and its major axis length	BCC vessels are straighter

3. ADAPTIVE CRITIC DESIGN METHODOLOGY

This model employs reinforcement learning (RL) through direct neural dynamic programming. The term "direct" is influenced by the adaptive control literature where "direct adaptive control" means no plant model, and thus no plant parameter is estimated. Instead, certain plant information is used directly to find appropriate and convergent control laws and control parameters.

Figure 7 shows the model of ADHDP used in this study, which is based on the model in [18]. In the current problem setting, let the discounted total reward-to-go *R(t)* at time *t* be given by

$$R(t) = r(t+1) + \alpha r(t+1) + \cdots = \sum_{k=1}^{\infty} \alpha^{k-1} r(t+k) \tag{2}$$

where the function of $r(t)$ is the reinforcement value at time t, and α is a discount factor between 0 and 1. In the context of this research, t refers to a contrast enhanced image in the sequence of 7 increasing CE images for a given skin lesion (Figure 6).

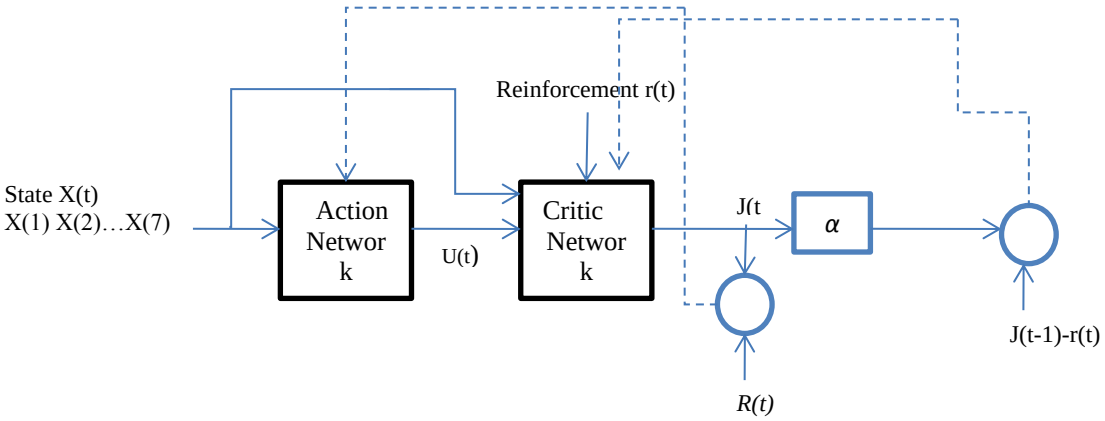


Figure 7. Schematic diagram of ADHDP.

The critic network (Figure 7) is used to provide an output *J(t)*, which is an approximation for *R(t)*, the weighted total future reward-to-go. The reward function *R(t)* at time

t is given by Eq.(1). We define the prediction error, and consequently the Bellman error, for the critic element as

$$e_c(t) = \alpha J(t) - [J(t-1) - r(t)] \tag{3}$$

and the objective function to be minimized in the critic network is

$$E_c(t) = \frac{1}{2} {e_c}^2(t) \tag{4}$$

The principle in adapting the action network is to backpropagate the error between the desired ultimate performance objective, denoted by *R**, and the approximate function *J* from the critic network. Since r_s has been defined as the reinforcement signal for "success," *R** is set to $r_s/(1-\alpha)$ in the direct NDP design paradigm and in subsequent case studies. In this paper, r_s is set to zero for simplification.

In the action network, the state measurements are used as inputs to create a control as the output of the network. In turn, the action network can be implemented by either a linear or a nonlinear network, depending on the complexity of the problem. The weight update in the action network can be formulated as follows.

$$e_a(t) = J(t) - R^* \tag{5}$$

An artificial neural network is chosen for implementation of the action and critic networks. The structure of the neural networks for both the action and critic networks are implemented as a multi-layer feed forward (MLP) three layer neural network consisting of an input layer, a hidden layer and an output layer. The hidden layer neurons have a sigmoid transfer function while other layers have linear transfer functions. For the action network, the architecture is 14x5x1, with 13 features and a bias in the input layer, five nodes in the hidden layer and 1 output layer. For the critic network, the architecture is 15x5x1, with 13 features, a bias and the output from the action network in the input layer, five nodes in the hidden layer and 1 output layer (Figure 8).

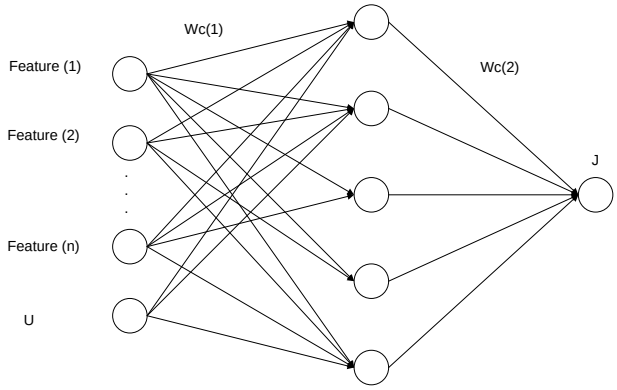


Figure 8. Neural network structure of the critic network.

A ten-fold cross-validation methodology is used for training/test set generation for the neural network. All neural networks are trained up to 1000 epochs, using online stochastic delta learning. In this case, the next input pattern is selected randomly from the training set, to prevent any bias that may occur due to the sequences in which patterns occur in the training set. For each training feature, seven different data states (original image feature data set and its six contrast enhancement feature data sets) are applied as the input one by one for both action network and critic network to update the weights. If the difference between the action network output U(t) and the target is less than 0.5, the reinforcement signal r(t) takes the reward "0", otherwise, r(t) takes the punishment "-1". The learning rates for both critic and action network are set to be 0.001. The discount factor α is set to be 0.1. The test set is the original image feature data set only.

With the target value for the BCC data set to 1 and the benign data set to 0, action network outputs after testing are between -1 and 1. Receiver operating characteristic (ROC) curves are generated for classification results based on the neural network output obtained for all ten test sets. The ROC curve is represented by plotting the fraction of true positives out of the BCC lesions (TPR = true positive rate) versus the fraction of false negatives from the benign lesions (FNR = false negative rate) [20].

A multilayer perception backpropagation neural network is investigated for lesion discrimination [21]. Sigmoid transfer functions are used in the hidden layers, and a linear transfer function is used in the input and output layer, the neural network architecture is 14x5x1 (all features as input). Neural network training and ROC curve generation proceed as above.

4. RESULTS AND DISCUSSION

4.1 DIAGNOSTIC ACCURACY

In this research, ACD is applied for BCC classification. The output is compared with MLP backpropagation neural network, using the same 14x5x1 neural network architecture (13 features and a bias for the input layer), training and test strategy, and ROC curve generation for all neural networks.

In addition to comparing the discrimination results for all 10 test sets from the different classifiers using AUC (Figure 9), we used a common true positive rate from the ROC curves to determine a true negative rate. Specifically, we defined the true negative rate to be the largest true negative value while the true positive rate is constant and at least 85%. With the overall true positive constant at 85.2%, the true negative rates for MLP backpropagation NN and Direct NDP are 64.2% and 80.6%, respectively.

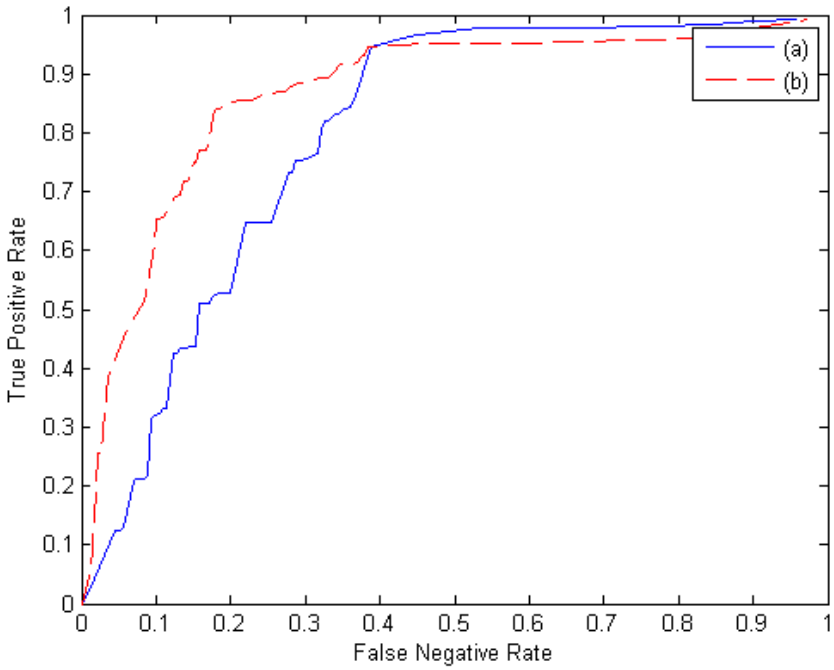


Figure 9. ROC curve and AUC (area under curve) for neural networks for all 10 text sets. (a)MLP back propagation NN. AUC=0.766. (b) Direct NDP. AUC=0.846.

Based on the AUC and the true positive and true negative comparisons, we found that adaptive critic design achieved the best results among the different classifiers investigated. Specifically, from Figure 9, direct NDP yielded higher AUC discrimination results (0.846), compared with the standard backpropagation NN (0.766). The direct NDP results show an improvement of 8.03% for AUC and 16.5% in true negative rate (based on the constant true positive rate of 85.2%) over the benchmark standard backpropagation NN method and features investigated in [12]. In direct NDP, the critic network is used as the performance evaluation; the reward of reinforcement learning is helping to update its weight and the output J(t). Therefore, the action network could update its weight based on total seven state feature data instead of only one state feature for other techniques. The global feature data improves the accuracy discriminating BCC from benign lesions.

4.2 STRUCTURE DETECTION VS. DIAGNOSIS

We have chosen BCC detection rather than vessel detection as the endpoint. Although vessel detection is inherently easier, BCC detection has potential direct clinical application. Small BCC are detectable early by dermoscopy, and potentially detectable by the automated methods described here.

5. CONCLUSIONS AND FUTURE WORK

This paper introduces the ACD design to image recognition using ADHDP algorithm. ADHDP result gives results superior to those from existing methods. This technique may be generally applicable to structure detection that is sensitive to image parameters such as contrast, which was explored here. The ACD technique provides a means to automatically tune the system for optimal results.

Future work will involve exploring other adaptive critic design algorithms such as action dependent dual heuristic dynamic programming (ADDHP) or action dependent globalized dual heuristic dynamic programming (ADGHP) for solving these problems. Moreover, the PSO-based ADHDP, ADDHP and ADGHP can also be implemented as a comparison. Last, the convenience sample from our clinics provided results that need confirmation. It is planned to pursue telangiectasia detection methods on a larger set of lesions.

REFERENCES

1. Bolognia JL, Jorizzo JL, Rapini RP, eds. “Dermatology”. New York: Mosby, 2003, pp.1653–1654.

2. Johr RH, Soyer P, Argenziano G, Hofman-Wellenhof R, Scalvenzi M. “Dermoscopy the essentials”. Edinburgh: Elsevier, 2004, pp.107.

3. Stolz W, Braun-Falco O, Bilek P, Landthaler M, Burgdorf WHC, Cognetta A. “Color atlas of dermatoscopy”, 2nd edn. Berlin: Blackwell Science, 2002, pp.33.

4. Thackray BD, Nelson AC. “Semiautomatic segmentation of vascular network images using a rotating structuring element (ROSE) with mathematical morphology and dual feature thresholding”. IEEE Trans Med Imaging, 1993, 12, pp.385–392.

5. Stansfield SA. “ANGY: a rule-based expert system for automatic segmentation of coronary vessels from digital subtracted angiograms”. IEEE Trans Pattern Anal Mach Intell, 1986, 8, pp.188–199.

6. Fei Y, Park JW. “Automatic segmentation of liver blood vessels using level set method”. Proceedings of international conference on audio, language and image processing, Shanghai, China, 2008, pp.1718–172.

7. Matsopoulos GK, Asvestas PA, Delibasis KK, Mouravliansky NA, Zeyen TG. “Detection of glaucomatous change based on vessel shape analysis”. Comput Med Imaging Graph, 2008, 32, pp.183–192.

8. Klingensmith JD, Vince DG. “Bspline methods for interactive segmentation and modeling of lumen and bessel surfaces in three-dimensional intravascular ultrasound”. Comput Med Imaging Graph, 2002, 26, pp.429–438.

9. Saad AA, Shapiro LG. “Shape decomposition approach for ultrasound color Doppler image segmentation In”, Proceedings of the 18th international conference on pattern recognition, Hong Kong, 2006, 4, pp.691–694.

10. Staal J, Abramoff MD, Niemeijer M, Viergever MA, van Ginneken B. “Ridge-based vessel segmentation in color images of the retina”. IEEE Trans Med Imaging, 2004, 23, pp.501–509.

11. Bhuiyan A, Nath B, Chua J, Kotagiri R. “Blood vessel segmentation from color retinal images using unsupervised texture classification”. In: Proceedings of the IEEE international conference on image processing, San Antanio, TX, 2000, 5, pp.521–52.

12. Cheng B, Erdos D, Stanley RJ, Stoecker WV, Calcara DA, Gómez DD. “Automatic detection of basal cell carcinoma using telangiectasia analysis in dermoscopy skin lesion images”. Skin Research and Technology, 2011, 17, pp.278-287.

13. Sutton RS, Barto AG. “Reinforcement Learning: An Introduction”. Cambridge: MIP Press, 1998.

14. Sutton RS, Barto AG, Williams RJ. “Reinforcement learning is direct adaptive optimal control”. In: Proceedings of 1991 IEEE American Control Conference, Boston, MA, 1991, pp.2143-2146.

15. Iftekharuddin KM, Li Y. “A biologically-inspired computational model for transformation invariant target recognition”. In: Proceedings of the IEEE International Joint Conference on Neural Networks, Hong Kong, 2008, pp.1049-1056.

16. Cheng B, Stanley RJ, Antani S, Thoma GR. “A Novel Computational Intelligence based Approach for Medical Image Artifacts Detection”. In: Proceedings of the 2010 International Conference on Artificial Intelligence and Pattern Recognition, Orlando, FL, 2010, pp.113-20.

17. Prokhorov DV, Wunsch D. “Adaptive critic designs”. IEEE Transactions on Neural Networks, 1997, 8(5), pp. 997-1007.

18. Venayagamoorthy GK, Harley RG, Wunsch D. “Applications of Approximate Dynamic Programming in Power Systems Control”. Handbook of Learning and Approximate Dynamic Programming. Hoboken, NJ: Wiley-IEEE Press, 2004, pp.479-515.

19. Rahman Z, Woodell GA, Jobson DJ. “A comparison of the multiscale retinex with other image enhancement techniques”. In: Proceedings of IS&T’s 50th Annual Conference: A Celebration of All Imaging 50, Cambridge, Massachusetts, 1997, pp.426-431.

20. Fogarty J, Baker RS, Hudson SE. “Case studies in the use of ROC curve analysis for sensor-based estimates in human computer interaction”. In: Proceedings of Graphics interface, Waterloo, Ontario, 2005, 112, pp.129-136.

21. Haykin S. “Neural Networks: A Comprehensive Foundation”. Upper Saddle River, NJ: Prentice Hall, 1998.

IV. AUTOMATIC DETECTION OF ARROW ANNOTATION OVERLAYS IN BIOMEDICAL IMAGES

Beibei Cheng[a], R. Joe Stanley[a], Soumya De[a], Sameer Antani[b], George R. Thoma[b]

[a]Department of Electrical and Computer Engineering

Missouri University of Science and Technology

Rolla, MO 65409-0040 USA

[b]Lister Hill National Center for Biomedical Communications

National Library of Medicine, National Institutes of Health, DHHS

Bethesda, MD 20894 USA

ABSTRACT

Images in biomedical articles are often referenced for clinical decision support, educational purposes, and medical research. Authors-marked annotations such as text labels and symbols overlaid on these images are used to highlight regions of interest which are then referenced in the caption text or figure citations in the articles. Detecting and recognizing such symbols is valuable for improving biomedical information retrieval. In this research, image processing and computational intelligence methods are integrated for object segmentation and discrimination and applied to the problem of detecting arrows on these images. Evolving Artificial Neural Networks (EANNs) and Evolving Artificial Neural Network Ensembles (EANNEs) computational intelligence-based algorithms are developed to recognize overlays, specifically arrows, in medical images. For these discrimination techniques, EANNs use particle swarm optimization and genetic algorithm for artificial neural network (ANN) training, and EANNEs utilize the number of ANNs generated in an ensemble and negative correlation learning for neural network training based on averaging and Linear Vector Quantization (LVQ) winner-take-all approaches. Experiments performed on medical images from the imageCLEFmed'08 data set, including 395 images with one or more arrows and 288 images with no arrows, yielded area under the receiver operating characteristic curve and precision/recall results as high as 0.988 and 0.928/0.973, respectively, using the EANNEs method with the winner-take-all approach.

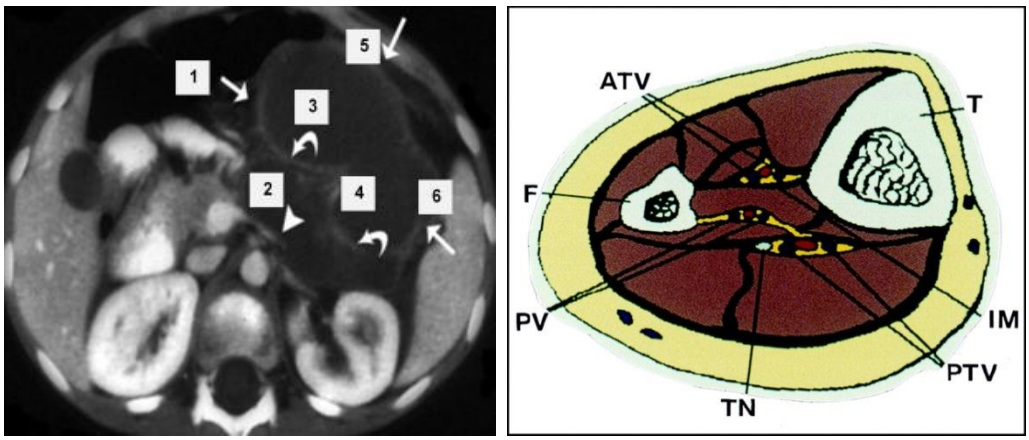
1. INTRODUCTION

Authors of biomedical publications use images to illustrate medical concepts and highlight special cases. These images often convey essential information and can be very valuable for improved clinical decision support (CDS) and education. Biomedical information retrieval has, so far, been largely text-based and limited mostly to bibliographic information. To be of greater value, it is desirable to retrieve images from biomedical publications. However, they need to be first annotated with respect to their usefulness for CDS to help determine relevance to a clinical query or to queries for special cases important in educational settings (Demner-Fushman, 2007, 2008, 2009).

Image retrieval can be achieved using the following methods: (i) traditional text-based approaches that index figure captions, (ii) image retrieval approaches that index the visual content of the images, and (iii) an intelligent combination of the above. To enhance text-based retrieval, content-based image retrieval (CBIR) has been explored to retrieve information from images in the biomedical field (Demner-Fushman, 2007). However, the approaches have not taken advantage of specific image regions of interest (ROIs) highlighted by the author using overlaid symbols, such as arrows and other text labels, and identifying them in the caption text. Further, it has been shown that whole image retrieval without attention to specific regions of interest marked by annotations, such as arrows (Figure 1), is not as promising as retrieval of text, primarily due to "semantic gap" introduced by less relevant image regions (Deserno, 2009). It is commonly understood in the field that low level features such as color, texture, and shape used in CBIR are insufficient to represent medical concepts or meaningful diagnostic information in the images effectively unless they can be applied to the key image regions such as those identified by the author, as in the case of images from biomedical articles. To improve the relevance quality of conventional retrieval approaches, we have proposed an approach using hybrid (text and image) features (Antani, 2008; You, 2009, 2010). Information retrieval techniques are used to identify key textual features in the title, abstract, figure caption, and figure citation ("mention") in the article. Structured vocabularies, such as the National Library of Medicine's Unified Medical Language System (UMLS®) are used as well to identify the biomedical concepts in these (Demner-Fushman, 2009; You, 2009). Unlike conventional CBIR schemes that extract features from the entire image, our approach uses a combination of features: those computed from specific image region of interests (ROIs) in addition to the ones obtained from the entire image. The ROIs are detected by localizing and recognizing image annotations

such as arrows overlaid by authors. Annotations and ROIs in retrieved images can be identified by the annotation recognizer and then used to re-rank the results of the recognizer.

There are some techniques that have been implemented to find arrows in previous research. Sparce pixel vectorization has been explored to detect arrowheads (Dov, 1999). In addition, arrow sign identification has been investigated for robot navigation using a camera-based method (Park, 2008). Compared to the existing approaches, the arrow symbols seen in the medical images experimental data set used in this research have a more complex shape. As shown in Figure 1(a), arrows in these medical images do not necessarily have to be straight (see arrow 3, arrow 4 in Figure 1(a)) and the shape of the arrows can be significantly different (see arrow 2 in Figure 1(a)) as well. Furthermore, the example image in Figure 1(b) shows objects such as characters and symbols which can be of similar size to arrows, providing potential false arrow detections. Therefore, a general and robust arrow detection algorithm is needed for discrimination from other medical image artifacts.



(a) (b)

Figure 1. Medical image with arrows. (a) Arrow example (Adapted from (Caskey, 1999)). (b) Non-arrow example (Adapted from (Fraser, 1999)).

This work extends previous research for a computational intelligence-based approach for medical image symbol (arrow) analysis (Cheng, 2010). In this research, a data set of 683 medical images annotated by modality (radiological, photo, etc.), was selected from the

imageCLEFmed'08 (http://www.imageclef.org) data set, including 395 images with one or more arrows and 288 images with no arrows (Müller, 2010). An overview of the arrow detection analysis process for medical images is shown in Figure 2. Since arrow, text and symbol objects are white or black, they can be segmented using image analysis techniques. After generating the binary image containing only text-like and symbol-like objects, feature sets are used as input to classifiers so that we can discriminate the arrows from noise and other types of medical symbols (Cheng, 2010). The various steps in the flowchart presented in Figure 2, are explained in Section 2. Section 3 gives the experimental results, Section 4 provides the discussion, and Section 5 presents conclusions and future work.

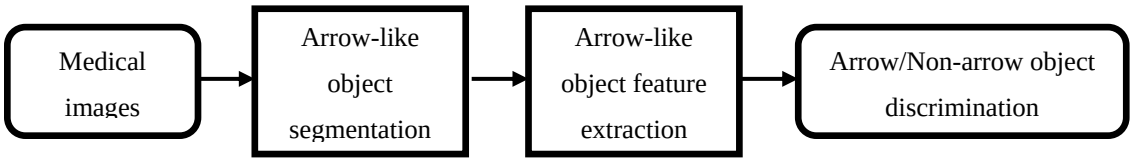


Figure 2. Overview of arrow detection process.

2. ARROW DETECTION PROCESS

As previously stated, medical images can contain arrow, text, and symbol objects. From these images, arrows typically are white or black objects as distinguished color from the background. These arrow objects can be segmented based on grayscale thresholding and edge detection, which is the basis for the segmentation algorithm presented. Thresholding and edge detection provide complementary information for finding arrow-like objects within the image with similar gray levels and potentially varying contrast with the surrounding background. Arrow-like object analysis and pruning are performed using extracted features with computational intelligence techniques. In the following sections of the paper, the algorithmic details are given for the methods used in the different blocks of arrow detection process flowchart from Figure 2.

2.1 OBJECT SEGMENTATION

From the medical images, the initial step is to segment arrow-like objects using a combination of thresholding and edge detection techniques. The object segmentation algorithm consists of the following steps:

1) Convert RGB images into luminance grayscale images.
2) Use Otsu's method (Otsu, 1979) to generate a preliminary object mask for arrow-like objects.
3) Remove objects that are considered small (pixel number of the object area is less than 60) from the preliminary mask in Step 2.
4) Generate an edge image of arrow-like objects using a gray drop method, extending the algorithm developed in (Cheng, 2011). If the absolute gray value of the center pixel (C) minus the gray value of NW, N, NE, W, E, SW, S, SE (see Figure 3) is greater than the gray drop, determined experimentally as 30, this pixel will be marked in the edge image. Figure 3 shows the edge detection operator mask.
5) Compute the bounding boxes of the objects in the masks from Steps 3 and 4.
6) Compare bounding box sizes for corresponding objects from the masks in Steps 3 and 4 and retain the objects with the same bounding box size. Let R_G denote the mask image determined from the grayscale image.

7) Repeat Steps 1-6 for inverted grayscale images since arrow, text and symbol objects may also be black. Let R_I denote the mask image determined from the inverted grayscale image.
8) Compute the final arrow-object mask as the OR image of R_G and R_I, denoted as $R_{G+I} = R_G + R_I$.

Figure 4 presents an image example of the image processing steps for the original image to generate the binary mask for feature calculations. Note that Figure 4 (i) and (k) are empty images because there are no arrow-like objects resulting from these steps in the image process steps to find arrows.

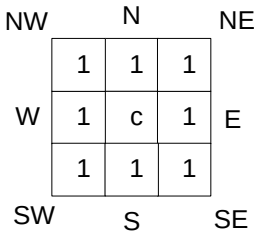


Figure 3. Edge detection operator mask.

2.2. FEATURE EXTRACTION

After completing the image processing steps for arrow-like object segmentation, features are extracted from each object. In (Park, 2008), features including extent and solidity were selected for arrow discrimination and line segment features were utilized to estimate the orientation of arrow objects (You, 2010). For the (You, 2010) study, arrow orientation, not detection, was explored. A typical arrow has a head region with varying stem types. Variations of arrow heads and arrow stems can be observed in Figure 1 (a). In order to address the complexity of the size and shape variations of the arrows in the medical images for the experimental data set, features in three categories are examined, including region property features, shape features and correlation-based features. To this end, multiple features are computed for each object in the mask *RG+I* (see Step 8 in Section 2.1) to facilitate arrow/non-arrow discrimination by using the object mask image computed as shown in section 2.1. The feature set descriptions are given as follows.

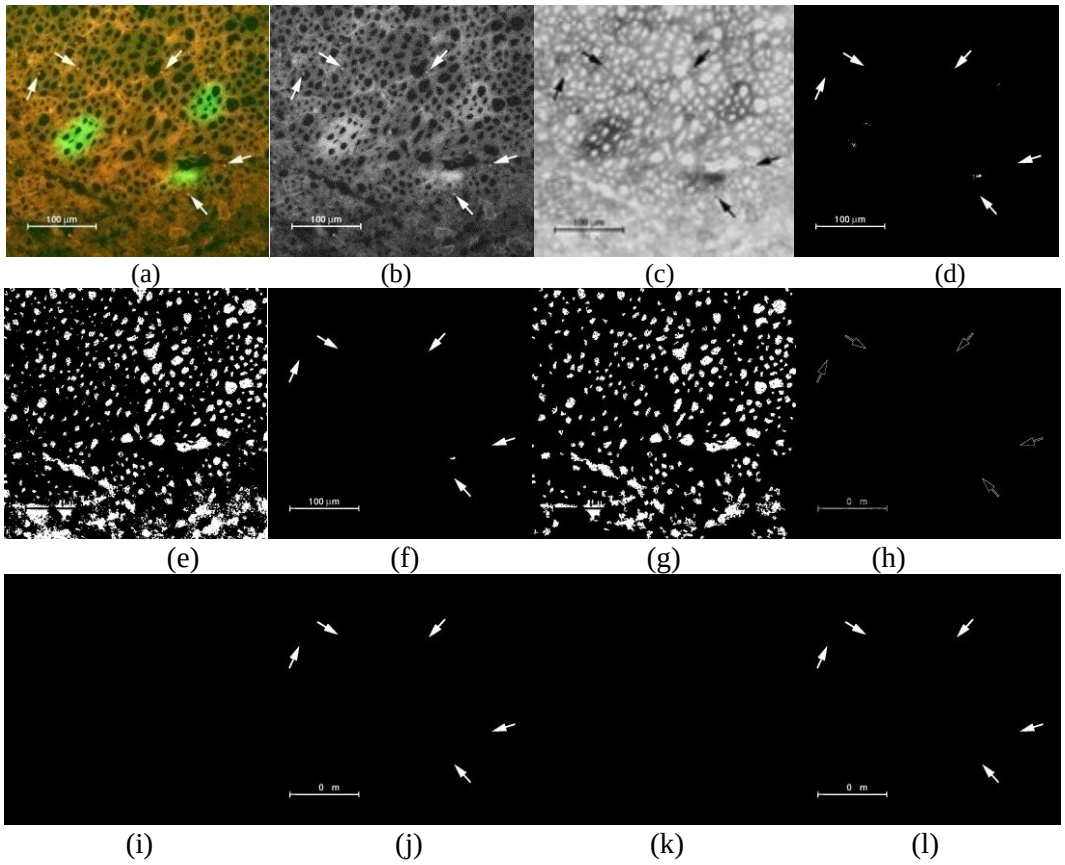


Figure 4. Object segmentation example. (a) Original image (Adapted from (Li, 2003)). (b) Grayscale image. (c) Inverted grayscale image. (d) Gray threshold image for (b). (e) Inverted gray threshold image for (c). (f) Gray threshold image after noise removal. (g) Inverted gray threshold image after noise removal. (h) Gray edge image for (b). (i) Inverted gray edge image for (c). (j) Gray image comparing (f) to (h) with the bounding box size. (k) Inverted gray image comparing (g) to (i) with the bounding box size. (l) Final OR-image of (j) and (k).

2.2.1 Region Property Features. The first set of features is based on the region properties and is computed using the Matlab® function regionprops (Hanselman, 2004). The regionprops features represent the visualization of the objects based on:

- ***MajorAxisLength***: length (in pixels) of the major axis of the ellipse that has the same normalized second central moments as the region.
- ***MinorAxislength***: length (in pixels) of the minor axis of the ellipse that has the same normalized second central moments as the region.

- ***Axis Ratio***: ratio of *MajorAxisLength* to *MinorAxislength*.
- ***Normalized area***: area of the region divided by the whole image.
- ***Solidity:*** area of the region divided by the convex hull area.
- ***EulerNumber:*** equal to the number of objects in the region minus the number of holes in those objects.
- ***EquiDiam***: the diameter of a circle with the same area as the region.
- ***Extent:*** ratio of area to bounding box area.

2.2.2 Shape Features. The second set of features computed for the arrow-like objects are shape features. These features include:

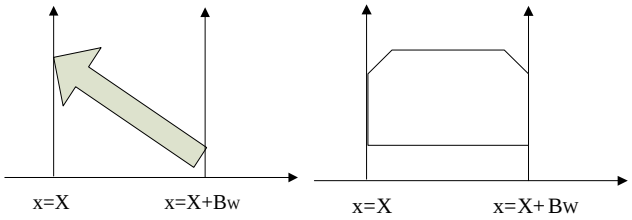
- ***AvgSkelDist:*** average width of object. It can be expressed in the following equation:

$$AvgSkelDist = \sum_{s=1}^{S} D_s / S \quad (1)$$

The skeleton of the arrow-like object was determined using the morphological skeleton algorithm (Serra, 1982). S is the total number of the pixels inside the skeleton object. D_s is the distance from pixel inside the skeleton object to the nearest pixel in the boundary of the object.

- ***MinPixelNo:*** the minimum number of intersection areas for the object and the two lines as shown in Figure 5. The function of these two lines are $x=X$ and $x=X+B_W$ (X is the left column of the bounding box; B_W is the width of bounding box), which is shown in Equation 2. The value of *MinPixelNo* for arrow (Figure 5(a)) is usually smaller than the value for noise (Figure 5(b)) due to the shape of the arrowhead.

$$MinPixelNo = \min(Area_{and(x-X,Object)}, Area_{and(x-X+B_W,Object)}) \quad (2)$$



(a) (b)

Figure 5. MinPixelNo feature. (a) Arrow. (b) Noise.

2.2.3 Correlation-based Features. The third set of features is based on computing shape profiles of binary arrow-like objects and correlating those profiles with basic functions. A one dimensional shape profile of each arrow-like object is found as follows. The bounding box of the arrow-like object is determined. Let $\boldsymbol{B_W}$ and $\boldsymbol{B_H}$ denote the bounding box height and width, respectively. The profile at each sample, $\boldsymbol{L(i)}$, is defined as Equation 3.

$$L(i) = \sum_{j=1}^{B_W} R_{G+I}\,(i,j) \tag{3}$$

for $i = 1, \ldots, B_H$. An example of the samples used for profile calculation is given in Figure 6. Let $L = \{L(1), L(2), \ldots, L(B_H)\}$ be the sequence of profile values. Correlation-based features are extracted by correlating the shape profile of the arrow-like object with weighted density distribution (WDD) functions (Piper, 1989), shown in Figure 7. Let W_1 denote the WDD function in Figure 7(a), W_2 denote the WDD function in Figure 7(b), and so on. In previous research, WDD functions have been explored in previous research for: 1) landmine discrimination based on 1-dimensional profiles of metal detector signals (Stanley, 2002) and 2) dermatology skin lesion discrimination based on a 1-dimensional histogram representation of skin lesions (Stanley, 2008). In both previous research applications, WDD functions provided shape-related information in the determination of correlation-based, size-variant, spatially distributed features from 1-dimensional profiles for object discrimination. In this research, the WDD functions have been applied to 1-dimensional projections profiles of arrow-like objects to extract shape information such as symmetry of the objects for object discrimination. These WDD functions account for the degree of change in the spatial distribution encapsulated in a 1-dimensional profile as well as the symmetry of those values for arrow discrimination. The twelve correlation-based features are computed as follows.

Six WDD features $(f_1, \ldots, f_6)$ are computed using the profile L according to the following expression:

$$f_k = \sum_{i=1}^{B_H} L(i) W_k(i) \tag{4}$$

for $k = 1,2, \ldots,6$. Six additional features $(f_7, \ldots, f_{12})$ are computed by correlating the six WDD functions with the sequence of absolute differences between samples value as follows:

$$f_k = \sum_{i=1}^{B_H} |L(i) - L(i-1)| W_k(i) \quad (5)$$

for $k = 1, \ldots, 6$ and $L(0) = 0$.

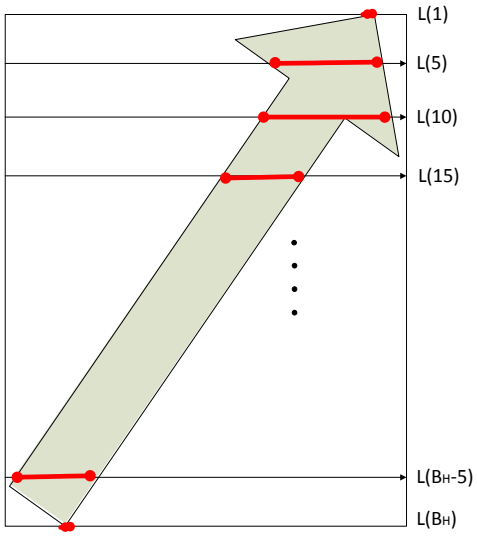


Figure 6. Samples for generating WDD features. (B_H is the height of bounding box)

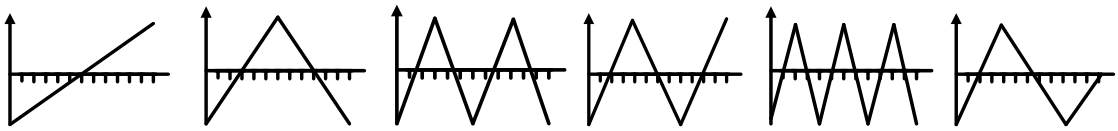


Figure 7. The WDD functions used to compute arrow features (Adapted from (Stanley, 2008)).

2.3 CLASSIFIER ALGORITHMS

Using the features computed for each arrow-like object within the object masks for all images in the experimental data set, Evolving Artificial Neural Networks (EANNs) and Evolving Artificial Neural Network Ensembles (EANNEs) are examined for arrow/non-arrow discrimination. A brief overview for each algorithm is presented in this section, while the implementation details are presented in the Appendix.

EANNs refer to a class of artificial neural networks (ANNs) in which evolution is another fundamental form of adaptation, in addition to learning (Yao, 1993). Evolutionary algorithms (EAs) are used to perform various tasks, such as connection weight training, architecture design, learning rule adaptation, input feature selection, connection weight initialization and rule extraction from ANNs. In our implementation, the Particle Swarm Optimization (PSO) (Kennedy, 1995) method and the Genetic Algorithm (GA) (Holland, 1975) method are explored for connection weight training. Both the PSO and GA use the same scheme for candidates' representation, where each candidate is a D-dimensional vector with each element of the vector representing a connection weight and D being the total number of weights. The population is initialized in the sense that each element in a vector is randomly setting a value from -0.1 to +0.1. The fitness values for PSO and GA are set to be the root mean square error (RMSE) given the connection weights. The RMSE is calculated based on the difference between the grand truth and the actual ANN's output. The algorithms for training the connection weights for ANNs in PSO and GA are explained in detail in the Appendix section.

Evolving Artificial Neural network ensembles (EANNEs) provide a method for utilizing and combining the outputs of several networks (Yao, 2008). Each ANN has the same inputs and generates its own outputs for decision making. The ensemble method is based on the premise that a population contains at least as much information as any single individual. There are two components for ensemble design: 1) a method of training the networks to encourage the diversity of behaviors and 2) a mechanism to decide the final output based on the outputs from the individual networks. For the first component, a cooperative ensemble learning system (CELS) is used for training individual networks. CELS is used to create negatively correlated ANNs using a correlation penalty term in the error function of each individual network so that the mutual information among the networks in the ensemble can be minimized based on the Liu and Yao approach (Liu, 1999). For the second components, since the outputs of the ANNs are floating-point numbers, averaging and winner-taking-all for combining/aggregating the outputs were examined. The algorithms for each component are shown in the Appendix.

3. EXPERIMENTS PERFORMED

The experimental data set consisted of 683 medical image annotated by modality (radiological, photo, etc.) selected from the imageCLEF08 data set (Müller, 2010), including 395 images with one or more arrows and 288 images with no arrows. These images were manually assigned as arrow/no-arrow images for creating the ground-truth database. The object segmentation for arrow-like object segmentation, feature extraction from those objects, and arrow/no arrow discrimination algorithms presented in Section 2 were applied to the image set. Using the object segmentation algorithm from Section 2.1, 724 arrow objects and 1450 text/noise objects were segmented from those images. The 22 input features computed from each arrow-like object include 8 region property features, 2 shape features, and 12 correlation-based shape profile features. Standard backpropagation ANNs, EANNs, and EANNEs algorithms with variations were investigated for arrow/non-arrow discrimination. In order for an object to be scored correctly as an arrow object, the object had to be labeled by the discrimination algorithm as an arrow object, the object had to be completely segmented, and the object had to be ground truthed as an arrow object.

Seven different feature combinations are investigated as inputs to the multilayer perceptrons(MLP) neural network architectures, with the neural network architectures summarized in the following cases: 1) 9x5x1 consisting of an input layer of 8 region property features and a bias with linear neurons, a hidden layer of 5 neurons with sigmoid transfer functions, and an output layer of one output with a linear neuron; 2) 3x5x1 consisting of an input layer of 2 shape features and a bias with linear neurons; 3) 13x5x1 consisting of an input layer of 12 correlation-based features and a bias with linear neurons; 4) 11x5x1 consisting an input layer of combined 8 region property features with 2 shape features and a bias with linear neurons; 5) 21x5x1 consisting of an input layer of combined 8 region property features with 12 correlation-based features and a bias with linear neurons; 6) 15x5x1 consisting of an input layer of combined 2 region property features with 12 shape features and a bias with linear neurons; 7) 23x5x1 consisting of an input layer of combining all three feature groups and a bias with linear neurons. These architectures are summarized in Table 1. A ten-fold cross validation methodology is used for generating training/test sets for each neural network's architecture (Kohavi, 1995). The same training/test sets from the cross-validation process are applied to all feature combinations and classification algorithms presented. Classification results are based on averaging the area under Receiver Operating Characteristic (ROC) curves (Fogarty, 2005) generated for each of the ten test sets. The area under the ROC curve was given as the

evaluation measure because it does not require selecting a decision boundary or threshold to determine detection accuracy. In addition, experimental results are reported using precision and recall (Bar-Ilan, 1998). Here, precision is defined as the number of arrow objects correctly called arrow objects (true positive classifications) divided by the total number of objects called arrow objects, and recall is defined as the true positive classifications divided by the sum of the true positive classifications and the number of objects which were not classified as arrow objects but should have been, i.e. the total number of arrows in the evaluation set of images.

Table 1. Seven cases with their feature combinations and NN architectures.

Case No.	Feature Combination	NN Architecture
1	Region property features	9x5x1
2	Shape features	3x5x1
3	Correlation-based features	13x5x1
4	Region property& Shape features	11x5x1
5	Region property & Correlation-based features	21x5x1
6	Shape & Correlation-based features	15x5x1
7	Region property& Shape & Correlation-based features	23x5x1

Figure 8 presents the ROC curve results for a representative test set for case 1 for the different classifiers with M=5 and N=75 (except for the backpropagation ANN algorithm). Table 2 shows the area under ROC curve results and precision and recall (given in parentheses) for the seven different input features combinations (Case 1 to Case 7) for the EANN, EANNE, and standard backpropagation ANN classifiers investigated. For the EANN and EANNE classifiers, Table 2 includes the results for different population size (M) and the maximum number of generations (N), and the standard online backpropagation ANNs were trained over 2000 epochs. Area under the ROC curve and precision and recall are presented based on averaging the results over the ten test sets from the ten-fold cross validation process. Precision and recall are presented based on specifying precision as a constant of at least 90% (based on the ROC curve) and computing the corresponding recall.

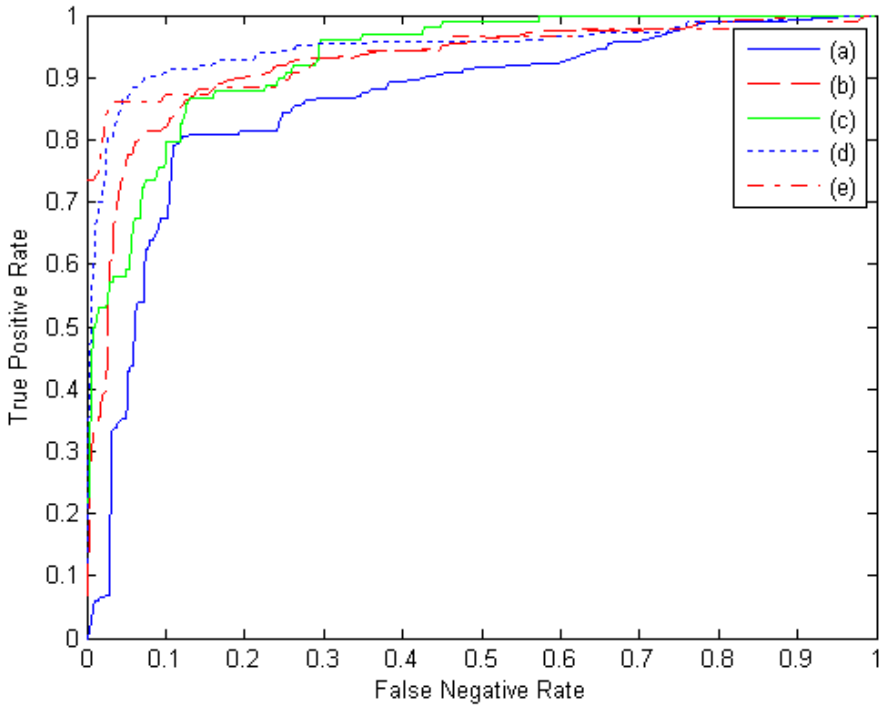


Figure 8. ROC curves for a representative test set for case 1 for the different classifiers with M=5, N=75 (except for Backpropagation ANN). (a) EANN GA algorithm. (b) EANN PSO algorithm. (c) EANNE Average algorithm. (d) EANNE LVQ algorithm. (e) Backpropagation ANN.

Table 2. Average area under the ROC curve test results and (precision/recall) for different algorithms and feature combinations. For the EANN and EANNE algorithms M refers to the population size, and N is the maximum number of generations.

Case 1		M=5,N=75	M=10,N=75	M=10,N=100	M=15,N=100
EANN	GA	0.862 (0.900/ 0.652)	0.867 (0.900/ 0.667)	0.865 (0.900/ 0.666)	0.884 (0.904/0.635)
	PSO	0.927 (0.900/0.835)	0.964 (0.900/ 0.922)	0.960 (0.905/ 0.942)	0.965 (0.906/0.941)
EANNE	Average	0.930 (0.900/0.783)	0.958 (0.900/0.865)	0.975 (0.906/0.972)	0.979 (0.900/0.970)
	LVQ	0.944 (0.901/ 0.838)	0.965 (0.903/0.890)	0.977 (0.906/0.972)	0.980 (0.906/0.973)
Backpropagation ANN		0.914 (0.900/0.795)			
Case 2					
EANN	GA	0.892 (0.900/0.550)	0.901 (0.900/0.540)	0.904 (0.900 0.686)	0.904 (0.900/0.686)
	PSO	0.876 (0.900/0.526)	0.917 (0.905/ 0.637)	0.924 (0.909/0.609)	0.914 (0.900/0.625)
EANNE	Average	0.884 (0.904/0.635)	0.884 (0.904/0.635)	0.893 (0.900/0.550)	0.902 (0.903/0.550)
	LVQ	0.884 (0.907/0.620)	0.883 (0.903/0.620)	0.896 (0.900/0.551)	0.905 (0.900/0.554)
Backpropagation ANN		0.905 (0.944/0.640)			
Case 3					
EANN	GA	0.792 (0.900/0.503)	0.855 (0.900/0.652)	0.830 (0.913/0.533)	0.821 (0.900/0.553)
	PSO	0.892 (0.900/0.782)	0.921 (0.901/0.820)	0.944 (0.900/0.864)	0.947 (0.905/0.850)
EANNE	Average	0.924 (0.902/0.861)	0.935 (0.900/0.860)	0.947 (0.900/0.855)	0.950 (0.900/0.861)
	LVQ	0.923 (0.902/0.861)	0.940 (0.900/0.861)	0.952 (0.900/0.863)	0.958 (0.900/0.865)
Backpropagation ANN		0.895 (0.944/0.681)			
Case 4					
EANN	GA	0.837 (0.913/0.540)	0.844 (0.923/0.553)	0.884 (0.904/0.635)	0.912 (0.900/0.624)
	PSO	0.954 (0.901/0.838)	0.961 (0.902/0.838)	0.974 (0.906/0.956)	0.978 (0.906/0.958)

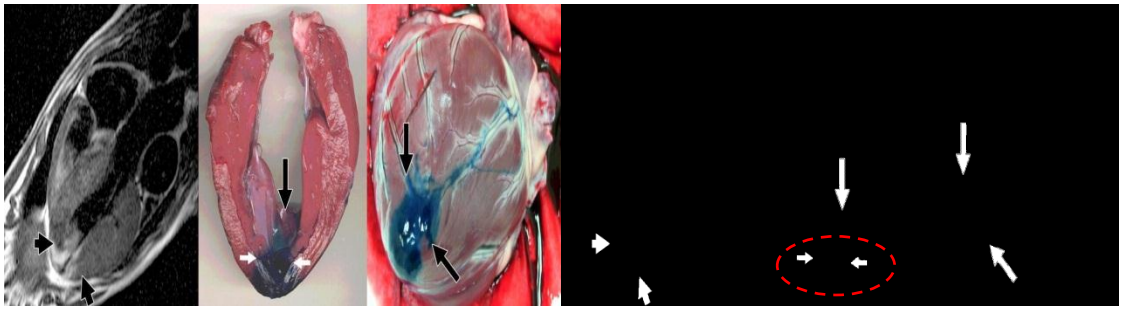
Table 2. Average area under the ROC curve test results and (precision/recall) for different algorithms and feature combinations. For the EANN and EANNE algorithms M refers to the population size, and N is the maximum number of generations (con't).

EANNE	Average	0.967 (0.900/0.950)	0.974 (0.900/0.960)	0.977 (0.906/0.958)	0.977 (0.906/0.958)
	LVQ	0.970 (0.900/0.958)	0.978 (0.906/0.958)	0.982 (0.903/0.892)	**0.988 (0.928/0.973)**
Backpropagation ANN		0.922 (0.917/0.830)			
Case 5					
EANN	GA	0.843 (0.923/0.553)	0.832 (0.932/0.516)	0.827 (0.932/0.514)	0.859 (0.900/0.655)
	PSO	0.942 (0.905/0.958)	0.974 (0.906/0.962)	0.963 (0.906/0.954)	0.985 (0.913/0.970)
EANNE	Average	0.956 (0.900/0.863)	0.976 (0.906/0.958)	0.984 (0.913/0.965)	0.980 (0.903/0.960)
	LVQ	0.956 (0.900/0.863)	0.977 (0.906/0.958)	0.985 (0.913/0.970)	0.986 (0.913/0.970)
Backpropagation ANN		0.909 (0.917/0.731)			
Case 6					
EANN	GA	0.828 (0.932/0.514)	0.831 (0.913/0.533)	0.896 (0.900/0.550)	0.873 (0.912/0.666)
	PSO	0.861 (0.900/ 0.667)	0.948 (0.906/0.958)	0.951 (0.905/0.940)	0.956 (0.900/0.963)
EANNE	Average	0.965 (0.906/0.954)	0.974 (0.906/0.958)	0.956 (0.900/0.863)	0.956 (0.900/0.963)
	LVQ	0.968 (0.906/0.955)	0.975 (0.906/0.958)	0.962 (0.906/0.954)	0.957 (0.900/0.963)
Backpropagation ANN		0.925 (0.917/ 0.865)			
Case 7					
EANN	GA	0.858 (0.900/0.667)	0.867 (0.906/0.667)	0.872 (0.912/0.667)	0.895 (0.900/0.550)
	PSO	0.924 (0.900/0.865)	0.961 (0.905/0.960)	0.976 (0.906/0.959)	0.985 (0.913/0.970)
EANNE	Average	0.978 (0.906/0.958)	0.975 (0.906/0.958)	0.974 (0.906/0.944)	0.976 (0.906/0.959)
	LVQ	0.975 (0.906/0.958)	0.979 (0.906/0.958)	0.983 (0.903/0.630)	0.983 (0.903/0.960)
Backpropagation ANN		0.958 (0.917/0.856)			

4. DISCUSSION

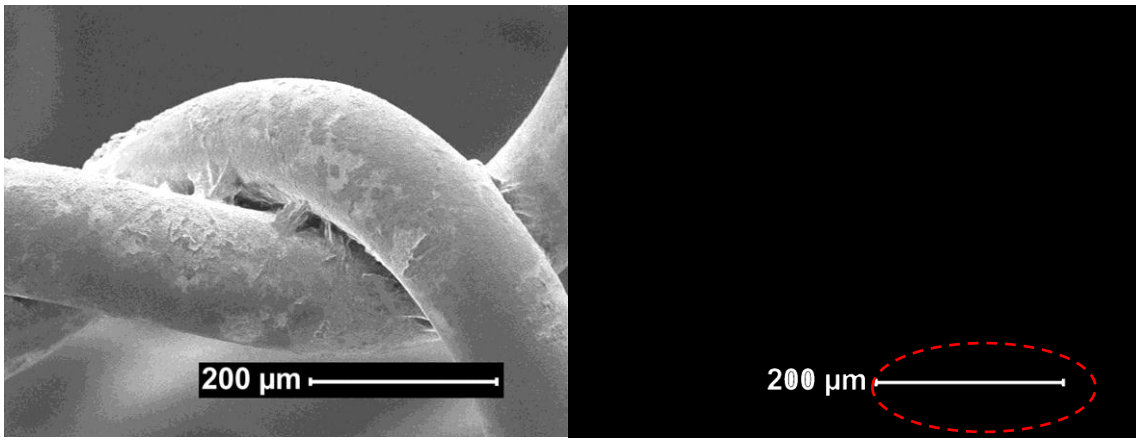
Inspecting Figure 8 and Table 2, several observations can be made. First, because the arrows in the experimental data set are uniform in grayscale, there were no cases of partially segmented arrow objects. Second, according to Figure 8, the classifier accuracy ranking from highest to lowest based on area under the ROC curve is EANNE LVQ algorithm, EANNE Average algorithm, EANN PSO algorithm, EANN GA algorithm, and backpropagation ANN. Third, the area under the ROC curve and precision/recall results are not directly related. The area under the ROC curve provides a measure of overall arrow discrimination capability over different classifier output thresholds. Precision and recall gives a measure of relevance to the objects labeled as arrows. Having a high area under the ROC curve does not always result in high precision/recall, as can be observed in Table 2. Having a high area under the ROC curve and high precision/recall demonstrates arrow objects can be successfully discriminated from non-arrow objects and that arrow objects are correctly found and not omitted in the selection/segmentation process within the medical images. The highest overall discrimination rates based on area under the ROC curve and precision/recall are 0.988 and 0.928/0.973, respectively, for the region property and shape features (case 4) using the EANNE with LVQ (winner-take-all) approach. Other feature combinations including all features (case 7) and the region property and correlation-based features (case 5) yielded similar results using the EANNE with LVQ approach. Fourth, the PSO algorithm consistently gives higher results, area under the ROC curve and precision/recall, than the GA approach for the EANN algorithm for the different feature combinations. Fifth, the EANN with PSO for weight updating and the EANNE methods consistently outperformed the standard backpropagation neural network benchmark approach for all feature combinations, highlighting the benefit of incorporating multiple neural networks in the training process. This is supported with the general observation that the area under ROC curve and precision/recall results are higher with more neural networks integrated into the training process, M = 10, 15 versus M = 5 for the EANN and EANNE classifiers. Sixth, for the EANNE approach, the LVQ winner-take-all method for integrating multiple neural networks gave consistently higher classification results than the averaging method for N=100 (last 2 columns of Table 2) for all feature combinations. Overall, the EANNE discrimination algorithms slightly outperform the EANN methods for the same input feature combinations. This experimental result highlights the robustness of the EANNE algorithms to the variations in the types of features explored for arrow discrimination as well as the size and shape variations of the arrows present in the experimental data set (Cheng, 2010).

The experimental results show that arrow discrimination can be performed at a high success rate using the arrow segmentation, feature extraction, and computational intelligence methods presented in this paper. The arrow-like object segmentation algorithm found all arrows with numerous false positive arrow-like objects, hence, the need for feature and discrimination analysis. The approach presented demonstrates the utility of integrating imaging and computational intelligence methods for object segmentation. The objects circled by the red in Figure 9 and Figure 10 provide some examples with incorrect classification. Since arrows are typically narrow and long, small arrows with large width may be incorrectly classified as noise, as shown in Figure 9. In addition, narrow and long noise objects may be mistaken for arrow objects, as shown in Figure 10.



(a) (b)

Figure 9. Arrow objects incorrect classified (Adapted from (Saeed, 2004)). (a) Original image. (b) Binary object mask image.



(a) (b)

Figure 10. Non-arrow objects incorrect classified (Adapted from (Schürmann, 2004)). (a) Original image. (b) Binary object mask image.

5. CONCLUSIONS AND FUTURE WORK

This paper presents a process for integrating image and feature analysis and computational intelligence-based techniques for arrow discrimination in annotated medical images. The arrow discrimination results show the potential for merging imaging and computational intelligence methods for accurate arrow discrimination and segmentation based on object pruning, i.e. labeling objects of interest. Experimental results yielded area under the ROC curve and precision/recall as high as 0.988 and 0.928/0.973, respectively, using the EANNE approach with winner-take-all LVQ approach. Future work will involve integrating the detection of medical annotations into an overall approach for fusing data such as key words, modality of medical image and figure captions to improve the relevance of the search results for medical publication querying. Future work will involve determining and incorporating arrow orientation information to assist in the assessment process of this symbol in medical images.

ACKNOWLEDGEMENTS

This work was supported by the National Library of Medicine, NLM under contract number 276200800413P and the Intramural Research Program of the National Institutes of Health (NIH), NLM, and Lister Hill National Center for Biomedical Communications (LHNCBC).

We thank the organizers of ImageCLEFmed 2008 (Müller, 2010) for providing the images for our research, some of which have been reproduced in this article. See http://www.imageclef.org for details.

We appreciate the contributions and efforts of the anonymous IJHISI reviewers for this paper.

REFERENCES

1. Antani, S., Demner-Fushman, D., Li, J., Srinivasan, B.V., Thoma, G.R., "Exploring use of images in clinical articles for decision support in Evidence-Based Medicine". In Proceedings of SPIE-IS&T Electronic Imaging, San Jose, CA, 2008, 6815, pp. Q(1-10).

2. Bar-Ilan, J., "On the overlap, the precision and estimated recall of search engines: A case study of the query "Erdos" ". Scientometrics, 1998, 42(2), pp.207-208.

3. Caskey, C.I., Berg, W.A., Hamper, U.M., Sheth, S., Chang, B.W., Anderson, N.D., "Imaging spectrum of extracapsular silicone: correlation of US, MR imaging", mammographic, and histopathologic findings, Radiographics, 1999, 19, pp.S39-S51.

4. Cheng, B., Stanley, R.J., Antani, S., Thoma G.R., "A Novel Computational Intelligence-based Approach For Medical Image Artifacts Detection", In International Conference on Artificial Intelligence and Pattern Recognition, Orlando, FL, 2010, pp.113-20.

5. Cheng, B., Erdos, D., Stanley, R.J., Stoecker, W.V., Calcara, D., Gomez, D., "Automatic Detection of Basal Cell Carcinoma Using Telangiectasia Analysis in Dermoscopy Skin Lesion Images". Skin Research and Technology, 2011, 17(3), pp.278–287.

6. Demner-Fushman, D., Antani, S.K., Thoma, G.R., "Automatically Finding Images for Clinical Decision Support". In Workshop on Data Mining in Medicine. 7th IEEE Intl Conf on Data Mining, 2011, pp. 139-144.

7. Demner-Fushman, D., Antani, S.K., Simpson, M., Thoma, G.R., "Combining Medical Domain Ontological Knowledge and Low-level Image Features for Multimedia Indexing". In 2nd International "Language Resources for Content-Based Image Retrieval" Workshop (OntoImage 2008), part of 6th Language Resources and Evaluation Conference, 2008.

8. Demner-Fushman, D., Antani, S., Simpson, M., Thoma, G.R., "Annotation and Retrieval of Clinically Relevant Images", International Journal of Medical Informatics: Special Issue on Mining of Clinical and Biomedical Text and Data, 2009, 78(12), pp.e59-e674.

9. Deserno, T.M., Antani, S., Long, R., "Ontology of Gaps in Content-Based Image Retrieval", Journal of Digital Imaging, 2009, 22(2), pp.202-15.

10. Dov, D., Liu, W., "Automated CAD Conversion with the Machine Drawing Understanding System: Concepts, Algorithms, and Performance", IEEE Transaction on Systems, Man, and Cybernetics-Part A: Systems and Humans, 1999, 29(4), pp.411-416.

11. Dov, D., Liu, W., "The sparse pixel vectorization algorithm and its performance evaluation", IEEE Trans on Pattern Analysis and Machine Intelligence, 1999, 21, pp.202–215.

12. Fogarty, J., Baker, R.S., Hudson, S.E., "Case studies in the use of ROC curve analysis for sensor-based estimates in human computer interaction", In Proceedings of Graphics interface, Canadian Human-Computer Communications Society, School of Computer

Science, University of Waterloo, Waterloo, Ontario Victoria, British Columbia, 2005, pp.129-136.

13. Fraser, J.D., Anderson, D.R., "Deep venous thrombosis: recent advances and optimal investigation with US". Radiology, 1999, 211, pp. 9-24.

14. Hanselman, D.C., Littlefield, B.L., "Mastering MATLAB 7", Upper Saddle River, NJ: Prentice Hall, 2004.

15. Holland, J.H., "Adaptation in Natural and Artificial Systems", Ann Arbor, MI: University of Michigan Press, 1975.

16. Kennedy, J., Eberhart, R., "Particle swarm optimization", In Proceedings of the IEEE International Conference on Neural Networks, Piscataway, NJ, 1995, pp.1942-1948.

17. Kohavi, R., "A study of cross-validation and bootstrap for accuracy estimation and model selection", In Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence, San Francisco, CA,1995, 2, pp.1137–1143.

18. Kohonen, T., "Self-Organizing Maps", 3rd edition. New York: Springer-Verlag, 1995.

19. Li, T., Tachibana, K., Kuroki, M., Mesahide, K., "Gene transfer with echo-enhanced contrast agents: comparison between albunex, optison, and levovist in mice—initial results", Radiology, 2003, 229, pp. 423-428.

20. Liu, Y., Yao, X., "Ensemble learning via negative correlation", Neural Networks, 1999, 12, pp. 1399–1404.

21. Müller, H., Clough, P., Deselaers, T., Caputo, B. (Eds)., "Experimental Evaluation in Visual Information Retrieval", The Information Retrieval Series, Berlin: Springer, 2010. 32.

22. Otsu, N., "A threshold selection method from gray-level histograms", IEEE Transactions on Systems, Man and Cybernetics, 1979, 9(1), 62-66.

23. Park, J., Rasheed. W., Beak, J., "Robot Navigation Using Camera by Identifying Arrow Signs", Grid and Pervasive Computing Workshops, GPC Workshops '08, The 3rd International Conference on Grid and Pervasive Computing, Kunming, 2008, pp.382-386.

24. Piper, J., Granum, E., "On fully automatic feature measurement for banded chromosome classification", Cytometry, 1989, 10, 242-255.

25. Saeed, M., Lee, R., Martin, A., Weber, O., Krombach, G.A., Schalla, S., Lee, M. Saloner, D., Higgins, C.B., "Transendocardial delivery of extracellular myocardial markers by using combination x-ray/MR fluoroscopic guidance: feasibility study in dogs", Radiology, 2004, 231, pp.689-696.

26. Serra, J., "Image Analysis and Mathematical Morphology", London: Academic Press, 1982, 1.

27. Schürmann, K., Lahann, J., Niggermann, P., Klosterhalfen, B., Meyer, J., Kulisch, A., Klee, D., Günther, R.W., Vorwerk, D., "Biologic response to polymer-coated stents: in vitro analysis and results in an Iliac artery sheep model", Radiology, 2004, 230, pp.151-162.

28. Stanley, R.J., Stoecker, W.V., & Moss, R.H., "A basis function feature-based approach for skin lesion discrimination in dermatology dermoscopy images", Skin Research and Technology, 2008, 14(4), pp.425-435.

29. Stanley, R.J., Gader, P., Ho, D., "Feature and decision level sensor fusion of electromagnetic induction and ground penetrating radar sensors for landmine detection with hand-held units", Information Fusion, 2002, 3(3), pp.215-223.

30. Yao, X., "A review of evolutionary artificial neural networks", International Journal of Intelligent Systems, 1993, 8(4), pp.539–567.

31. Yao, X., Islam, M.M., "Evolving artificial neural network ensembles", Computational Intelligence Magazine, 2008, 3(1), 31-42.

32. You, D., Apostolova, E., Antani, S.K., Demner-Fushman, D., Thoma, G.R., "Figure content analysis for improved biomedical article retrieval", In SPIE-IS&T Electronic Imaging, San Jose, CA, 2008, 7247, pp. v(1-10).

33. You, D., Antani, S.K., Demner-Fushman, D., Rahman, M.M., Govindaraju, V, Thoma, G.R., "Biomedical article retrieval using multimodal features and image annotations in region-based CBIR", In SPIE-IS&T Electronic Imaging, San Jose, CA, 2010, 7534, pp.v(1-12).

APPENDIX

A.1 EVOLVING ARTIFICIAL NEURAL NETWORKS

A.1.1 Evolving Artificial Neural Networks trained by Particle Swarm Optimization. Particle Swarm Optimization (PSO) (Kennedy, 1995) is the study of swarms of social organisms such as flock of birds, which each particle in the swarm moves toward its previous best location (*Pbest*) and global best location (*Gbest*) defined below at each time step. To train the connection weights in the ANNs, each candidate is a particle. *Pbest* is the particle of the *M* particles that gives the least RMSE between the current epoch of ANN's training and the previous epoch with ANN's training. *Gbest* is the particle among the *M* particles which generates the minimum RMSE. The velocity to update the particle is presented in Equation 6. The position vector of the particles is changed as shown in Equation 7. The same process is used for obtaining the next set of particles, which is repeated by *N* epochs.

The velocity of the particles is given as follows:

$$V_{md}(n+1) = wV_{md}(n) + c_1 rand_1\left(Pbest_{md} - X_{md}(n)\right) + c_2 rand_2\left(Pbest_{md} - X_{md}(n)\right) \quad (6)$$

The position vector of the particles is changed as follows:

$$X_{md}(n+1) = X_{md}(n) + V_{md}(n+1) \quad (7)$$

where *n* is the current iteration (time step) *(* $1 \le n \le N$ *)*, *m* is the current particle *(*$1 \le m \le M$*)*, *d* is the weight element (*1*≤*d*≤ number of weights), $V_{md}(n)$ is the particle's current velocity, $V_{md}(n+1)$ is the particle's new velocity, $X_{md}(n)$ is the particle's current position, $X_{md}(n+1)$ is the particle's new position, $rand_1$ and $rand_2$ are the random values selected from *0* to *1*, *w* is the inertia weight chosen as *0.7*, c_1 is the cognitive acceleration constant of *1.5*, and c_2 is the social acceleration constant of *1.5*.

A.1.2 Evolving Artificial Neural Networks trained by Genetic Algorithm. Genetic Algorithm (GA) (Holland, 1975) provides optimization by using selection, crossover, mutation and elitism operators. The implementation used in this research consists of generating M offspring, i.e. particles from a pool of M sets of initial weights comprising a parent pool.

The offspring are generated as follows: 1) randomly select two parents (sets of weights)

from the parent pool of M sets of initial weights; 2) initialize offspring weights as the first parent weights and apply a randomly generated binary mask with the size of the weights matrix for the crossover process in order to recombine selected parents to get offspring. For this randomly generated binary mask, if a random value is selected which is less than 0.5, then the bit of binary mask is 0, otherwise, it is 1; 3) apply a mutation process for weight updating. The mutation process consists of selecting a weight and then adding a random value to it. A weight is chosen if a random value is less than 0.5, then another random value between -1 to +1 is added to the weight value.

MLP training is performed using the parent pool of weights and the offspring pool of weights based on the ANN architecture above. The next parent pool is chosen based on whether the parent is used for initialization or its offspring minimizes the RMSE error. The same process is used for obtaining the next set of offspring, which is repeated by N epochs. From the final parent pool, the parent which minimizes the RMSE error over the training feature vectors is selected for the final ANN weights for the test vectors.

A.2 EVOLVING ARTIFICIAL NEURAL NETWORK ENSEMBLES

A.2.1 Training the Networks. A learning paradigm named negative correlation learning (NCL) is used for training neural network ensembles. The idea of negative correlation learning is to introduce a correlation penalty term into the error function of each individual network so that the mutual information among the networks in the ensemble can be minimized (Liu, 1999).

The steps of genetic algorithm neural network ensembles for training are given as follows. First, generate an initial population of *M* ANNs, and set the iteration number *n* to be 1, the random initial weights are distributed uniformly inside a small range. Second, train each ANN in the initial population on the training set for a certain number of epochs using negative correlation learning. Third, calculate the fitness of *M* ANNs in the population. Fourth, create *M* offspring ANNs by using selection, crossover, and mutation. Fifth, replace the worst *M* ANNs in the current population with *M* offspring ANNs, and train the whole population using negative correlation learning for another epoch. Sixth, stop the evolution process if the maximum number of iterations (*N*) has been reached. Otherwise, $n = n + 1$ and go to step 2.

A.2.2 Final output decision. Once a population of networks has been trained, a mechanism is needed to decide the final output based on the outputs from individual networks. Different methods are considered such as averaging, majority voting, and winner-taking-all. Since the outputs of the ANNs are floating-point numbers, we explored averaging and winner-taking-all for combining/aggregating the outputs. For averaging, the output (y_{avg}) is simply expressed as follows:

$$y_{avg} = \frac{1}{M} \sum_{m=1}^{M} F_m \tag{8}$$

where, *M* is the number of the individual ANNs in the ensemble.

For winner-taking-all approach, the output of the network with the strongest activation is chosen. A learning vector quantization network (LVQ) is trained after training the neural network ensembles (Kohonen, 1995). There are two layers in LVQ-competitive layer and output layer. The net output of the first layer of the LVQ is given by W, expressed as:

$$L_1 = argmin_m ||x - w_{1m}|| \tag{9}$$

where x is the same input vector as the input to the ensemble of ANNs, $m=1,2,\ldots,M$ and w_{1m} is the weight of the m^{th} neuron in the first layer.

The network output of the second layer of the LVQ is given by:

$$L_2 = w_2 L_1 \tag{10}$$

where w_2 is the weight in the second layer. The second layer of the LVQ network is used to combine subclasses into a single class. The columns of w_2 represent the subclasses (*M*) and the rows of the matrix represent the classes (*C*). w_2 has a single *1* in each column, with others elements set to zero. The value *1* in each row indicates which class (the row number) the appropriate subclass (the column number) should be combined into. We set *C* equal to *M*.

The training target (*T*) is given as:

$$T = \begin{cases} argmin(F_m), \ if \ class \ 0 \\ argmax(F_m), \ if \ class \ 1 \end{cases} \tag{11}$$

We adjust the synaptic weight vectors of all neurons by using the update formula on the nth training epoch:

$$w_{1m}(n+1) = w_{1m}(n) + \eta(n)(x - w_{1m}(n)) \quad (12)$$

where, $\eta(n)$ is the learning rate which is set to be *0.2* if L_2 is the same as the *T*, otherwise, it is *-0.2*.

Therefore, we have a trained LVQ network with the same input as EANNs and the winner neural network number as the output (*1, 2,…, or M*) indicated by L_2.

V. AUTOMATIC DETECTION OF BASAL CELL CARCINOMA USING TELANGIECTASIA ANALYSIS IN DERMOSCOPY SKIN LESION IMAGES

Beibei Cheng[a], David Erdos[a], R. Joe Stanley[a],William V. Stoecker[b], Randy H. Moss[a], David A. Calcara[b], David D. Gómez[c]

[a]Department of Electrical and Computer Engineering, University of Missouri-Rolla, Rolla, MO

[b]Stoecker and Associates, Rolla, MO

[c]Computational Imaging Lab, Department of Information and Communication Technologies, Universitat Pompeu Fabra, Barcelona, Spain

ABSTRACT

Telangiectasia, dilated vessels of varying diameter near the surface of the skin of varying diameter, are critical dermoscopy structures used in the detection of basal cell carcinoma. Distinguishing these vessels from other telangiectasia, that are commonly found in sun-damaged skin, is challenging. In this research, image analysis techniques are developed to identify vessels found in basal cell carcinoma automatically. The primary screen for vessels uses an optimized local color drop technique. A noise filter is developed to eliminate false positive structures, primarily bubbles, hair, and blotch and ulcer edges. From the telangiectasia mask containing candidate vessel-like structures, shape, size and normalized count features are computed to facilitate the discrimination of benign skin lesions from basal cell carcinomas with telangiectasia using a neural network classifier.

1. INTRODUCTION

Telangiectasias (telangiectases) are dilated blood vessels of varying diameter within the superficial dermis. They are common in fair-skinned persons, especially in sun-exposed areas in older persons. These vessels are also observed with a number of diseases, including rosacea, congenital lesions such as port-wine stains, scleroderma, inherited disorders such as ataxia-telangiectasia and hereditary hemorrhagic telangiectasia, and with prolonged use of oral or topical corticosteroids [1]. Basal cell carcinoma (BCC), the most common skin cancer, frequently displays telangiectasia. Although often visible without magnification, these blood vessels can best be visualized with dermoscopy, using either a glass plate with fluid interface (contact non-polarized dermoscopy) or cross-polarized lighting, together with 10-power magnification. In BCC, the classical form of telangiectasia is termed "arborizing telangiectasia," with a thick central trunk vessel with narrow radiating branch vessels. [2, 3] Figure 1 presents an example of a dermoscopy skin lesion image with arborizing telangiectasia and fine telangiectasia. Figure 1 also shows sun-damaged skin that has telangiectasia, found in many non-malignant lesions, often seen in sun-damaged skin. The challenge here is to differentiate the wider, shorter, often less distinct telangiectasia seen commonly from the longer, finer vessels of BCC in order to discriminate BCC from non-malignant lesions.

There have been numerous imaging-based approaches for vessel segmentation in a variety of medical domains. Approaches investigated for vessel extraction and analysis include: 1) mathematical morphology and grayscale histogram thresholding of edge pixels for semi-automatic extraction of vessels in vascular network images [4], 2) a Rule-Based Expert System for segmentation of coronary vessels from digital angiograms which uses edge and region information along with domain specific knowledge for segmenting, grouping and performing shape analysis [5], 3) morphological and modified Otsu threshold techniques for initialization a level set method with for tracking interfaces and shapes in the segmentation process for automatic segmentation of liver blood vessels using [6], 4) image analysis, self-organizing map-based registration and neural network classification techniques for vessel shape analysis to detect glaucomatous change [7], 5)B-spline methods for semi-automated segmentation and modeling of lumen and vessel surfaces in three-dimensional intravascular ultrasound [8], 6) shape decomposition for segmenting blood vessels in ultrasound color Doppler images based on shape decomposition [9], 7) ridge- and feature-basedvessel segmentation in two-dimensional color images of the retina [10], 8) texture-based segmentation of blood vessels in retinal images using unsupervised clustering techniques which addresses local contrast variations for detecting

minor vessels [11], among others.

In this research, we explore a vessel detection algorithm for extracting and characterizing telangiectasias in dermoscopy skin lesion images for discriminating basal cell carcinoma from benign skin lesions. The proposed technique incorporates a color drop-based method for finding vessel-type objects with morphological and other image analysis methods to eliminate vessel-type mimics in common lesion features.

The remaining sections of this paper include: overview, description of data sets used, pre-processing, first pass vessel detection by local drops, image noise filtering technique, post-processing, vessel features used for skin lesion discrimination, and neural network methods for BCC diagnosis, experiments performed, results, discussion, and conclusions.

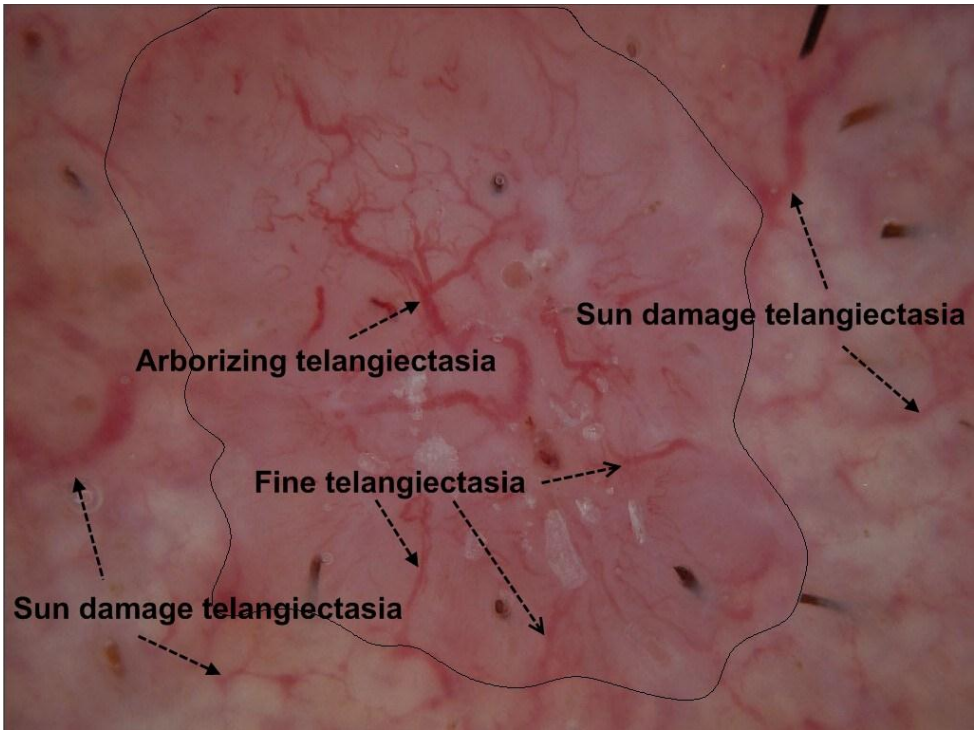


Figure 1. Telangiectasia. Arborizing telangiectasia (trunk and branches) are the classical telangiectasia seen in a contact, non-polarized dermsocopy image of basal cell carcinoma (BCC). Fine telangiectasia are more numerous than arborizing telangiectasia in BCC. Note that sun damage telangiectasia are often rudimentary and tend to be wider, shorter, have less sharp edges, have a greater variation in width, and are less numerous per area than BCC telangiectasia.

2. METHODS

2.1 TELANGIECTASIA DETECTION OVERVIEW

An overview of methods used is shown in Figure 2. Following pre-processing of images, color drops are used for vessel localization and characterization. Noise filtering, segment connection and size filtering are used to produce vessel masks. Various features sets are used as input to the neural network to generate ROC curves.

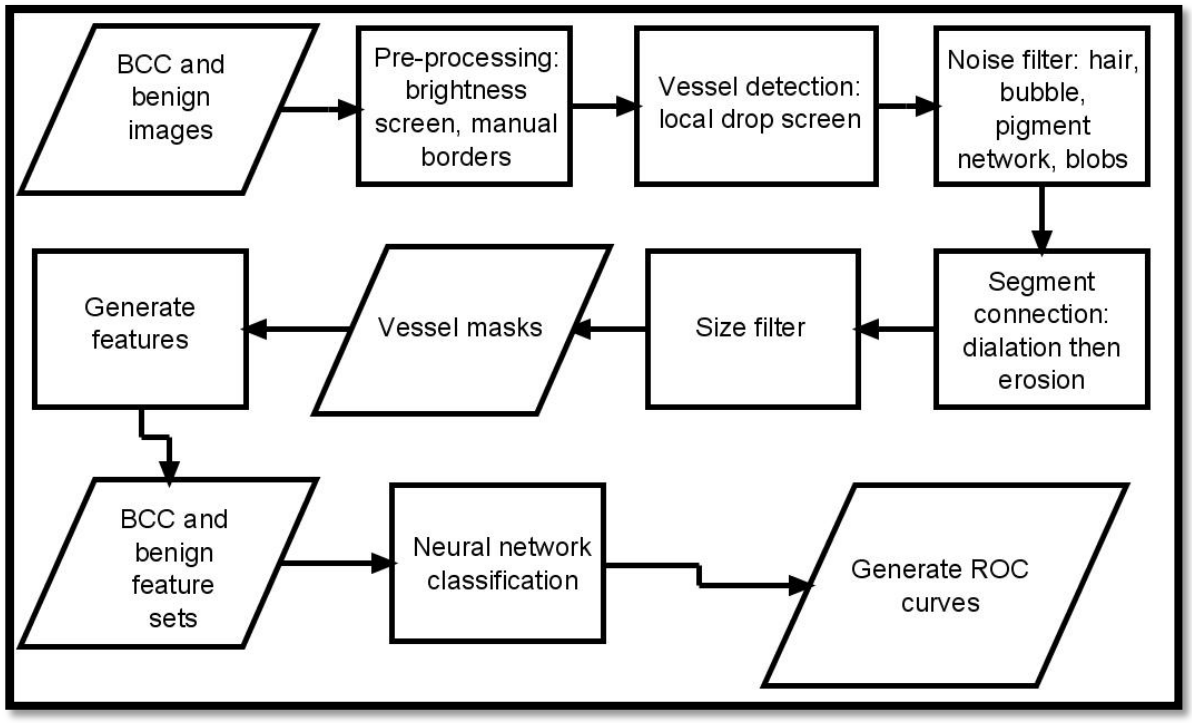


Figure 2. Overview--BCC diagnosis by telangiectasia detection.

2.2 DATA SET USED

Fifty-nine basal cell carcinoma dermoscopy images with visible telangiectasia were selected at random from 297 basal cell carcinomas encountered in 2 clinics: Skin and Cancer Associates, Plantation, FL, and the Dermatology Center, Rolla, MO. 116 benign dermoscopy images were selected from benign lesions in the same study: 31 dysplastic nevi (23 with mild atypia and 8 with moderate atypia), 21 seborrheic keratoses, 13 congenital nevi, 8 nevocellular nevi, 8 intradermal nevi, 7 lentigines, 6 compound nevi, 3 junctional nevi, 2 acral nevi, 2 actinic

keratoses, 2 hemangiomas, 2 lichen-planus like keratoses, 2 warts, and 9 single examples of various benign diagnoses were identified in the same dermoscopy study and used as the competitive test set. For the test set, 152 benign dermoscopy images were used from the same study: 34 seborrheic keratoses, 32 dysplastic nevi (28 with mild atypia and 4 with moderate atypia), 19 congenital nevi, 13 lentigines, 11 nevocellular nevi, 8 actinic keratoses, 5 compound nevi, 4 blue nevi, 4 lichen-planus like keratoses, 3 intradermal nevi, 2 Grover's disease, 2 histiocytomas, 2 lichen simplex chronicus, 2 junctional nevi, and 11 single examples of various benign diagnoses. (NIH SBIR R44 CA-101639-02A2). All images were taken with the HR II DermLite (3Gen, Dana Point, CA) using ultrasonic gel for the fluid interface. Lesions were included in the study if by clinical or dermoscopy evaluation they were malignant, potentially malignant, changing, or of interest to physician and/or patient. Lesions that were biopsied for cosmetic reasons were not included in the study. The data set was of moderate diagnostic difficulty and was felt to represent basal cell carcinomas and benign lesions encountered in the clinic. The Phelps County Regional Medical Center Institutional Review Board, Rolla Missouri, approved this research and each subject or subject's parent or guardian signed a consent form for this research. All basal cell carcinomas in the study had histopathology examined by a dermatologist or a dermatopathologist.

2.3 PRE-PROCESSING

The images obtained are not all of equal contrast. The dark vessels cannot be found in low-contrast images. All images failing a minimum contrast threshold, measured by standard deviation, underwent a uniform contrast enhancement and brightness decrease. Borders are then found manually (dc) and corrected by a dermatologist (wvs), by selecting border points, creating a closed curve with a send-order spline, and a subsequent binary lesion border mask [12].

2.4 COLOR-DROP VESSEL DETECTION

To a human, a vessel looks red compared to the surrounding skin. Using machine vision, it is seen that pixels inside the vessel have green and blue color drops from the surrounding pixels while the red pixels are at most, only slightly brighter.

The algorithm is based upon the premise that the vessels are narrow. Iterating through every pixel inside the lesion, labeling a new designated center pixel as c (Figure 3), it then moves outward a set number of pixels (NumPix) from a given center in eight directions, and demands a minimum drop, different for each color-- red, green and blue, respectively optimized at -2, 4, and 12. A candidate pixel is marked if for any satisfied direction pair, the drop

requirements are satisfied in at least two directions. As determined experimentally, 135 degrees is large enough to compare the surrounding pixel with the center pixel; 45 degrees and 90 degrees yield too much noise, while 180 degrees may miss some actual vessel pixels. The goal at this stage is to find all possible vessel pixels, regardless of noise found.

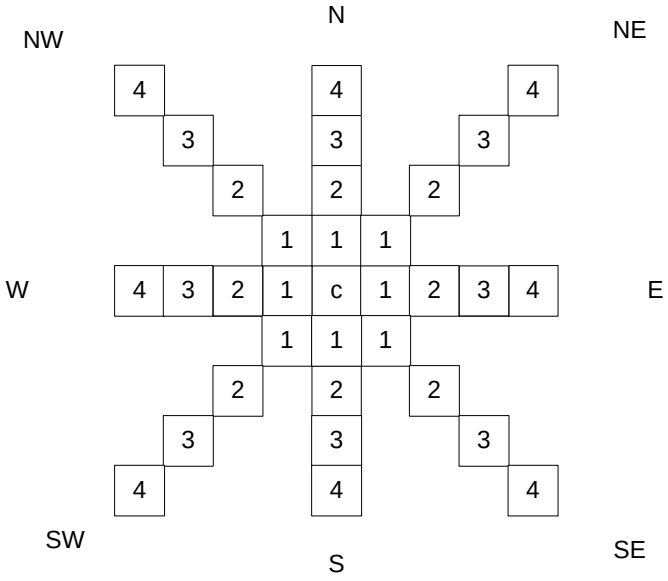
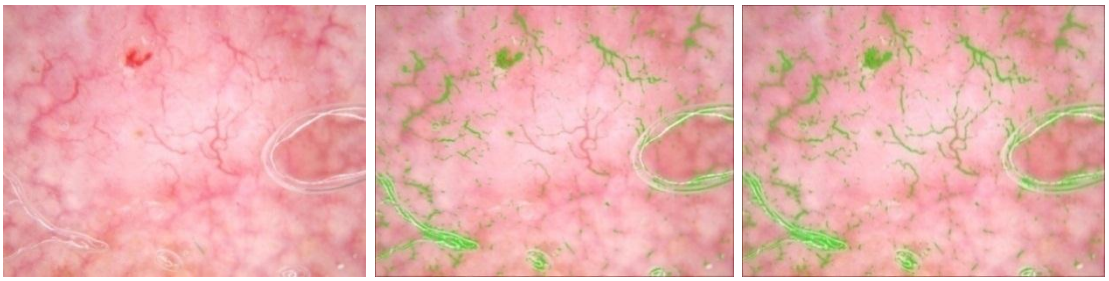


Figure 3. Direction Mask Used for Pixel Marking.

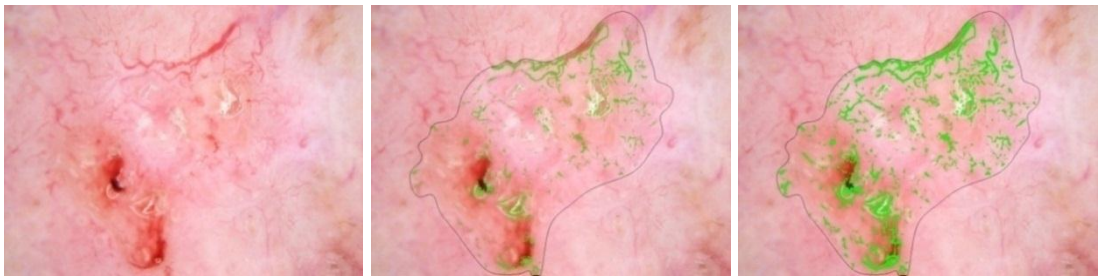
Figure 4 presents an example of the output images based on different red color drops with 2 in (b) and -2 in (c). With decreasing the red drop from 2 to -2, some vessels missed in (b) can be detected.



(a) (b) (c)

Figure 4. Mask image with different red drops. (a) Original image. (b) Mask image with red drop of 2. (c) Mask image with red drop of -2.

Vessels have different widths. Wider vessels are missed, i.e. no drop is detected, unless the outward pixel search parameter (NumPix) is large enough to include both vessel and surround. Figure 5 shows NumPix = 4 is not sufficiently high to detect the widest vessels that NumPix = 7 can detect.



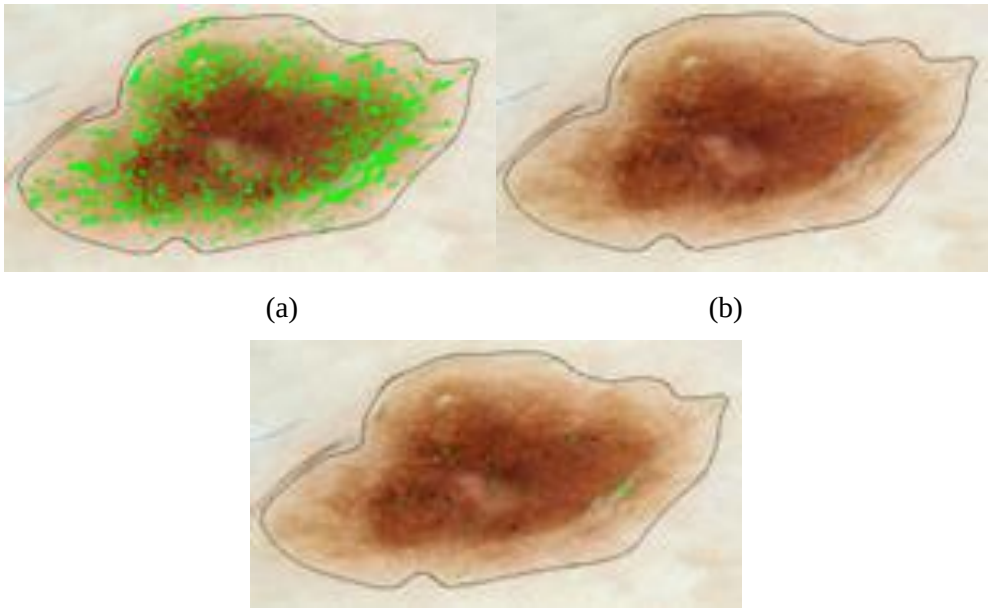
(a) (b) (c)

Figure 5. Mask Images with Different NumPix Values. (a) Original Image. (b) NumPix=4. (c) NumPix=7.

2.5 NOISE ELIMINATION

After implementing the above algorithm, noise sources such as brown network, blobs, hair, bubbles and brown network were marked. A noise filtering technique is used to mitigate these noise sources. Figure 6 (a) shows brown areas that are labeled as vessels using the vessel detection technique. The following sections present noise filtering approaches used to address the different types of noise sources. All of these techniques are applied on a pixel-by-pixel basis to pixels included in the vessel mask from the vessel detection algorithm.

2.5.1 Brown Network Filtering. Figure 6 (a) shows the mask result after applying above vessel detection algorithm, with most noise due to brown areas. Let RGB represent the red, green and blue values, respectively, at a pixel location within the lesion or pixels within brown network structures, G > B, since brown is a degraded orange, which has G > B. The opposite is true for vessels. We find that this condition, G>B+5, could be used to filter the brown area as shown in Figure 6 (b). Figure 6 (c) shows the mask result using the filter condition G>B+20 for comparison purposes. Based on empirical analysis of the dataset, as can be observed comparing Figure 6 (b) and Figure 6 (c), the filter condition G>B+5 provided optimal removal of brown areas compared to other G and B combinations.



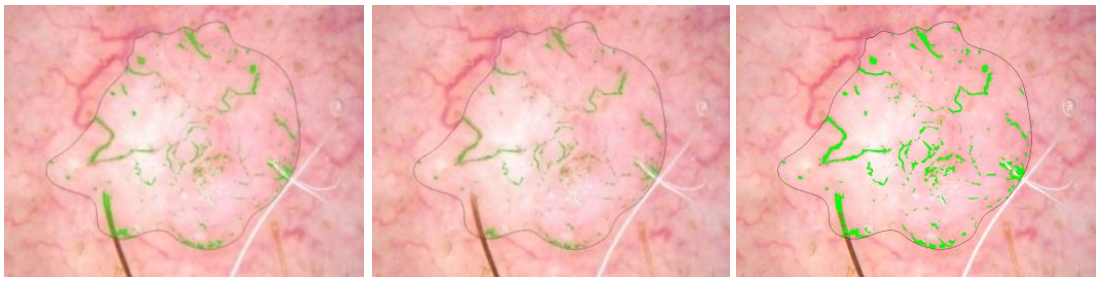
(a) (b)

(c)

Figure 6. Brown Area Filtering. (a) Brown Areas Labeled as Vessels. (b) Brown Areas after Removal from Vessel Mask by G>B+5 Filter. (c) Brown Areas after Removal from Vessel Mask by G>B+20 Filter. Filter shown in (b) is optimal.

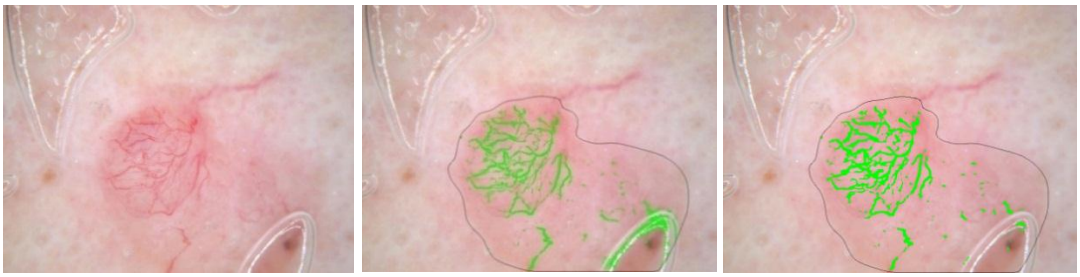
2.5.2 Hair Filtering. Hairs are noise found by the vessel algorithm, as shown in Figure 7(a). Hair is best distinguished from vessels by means of the variance of R/G, which is higher at hair boundaries than at blood vessel boundaries. This is due to the very high relative drop in red moving from surround to inside the hair, compared to an essentially zero drop in red moving from surround to inside the vessel, as noted above in the color drop discussion in section 2d. Note that other features of hair- lower luminance, sharp double boundaries, and long and narrow shape, are all present in blood vessels. The optimal hair filter, compared to other filters investigated, shown in Figure 7(b) employed a 5x5 sliding window to eliminate all pixels whose centered 5x5 surrounding areas had variance of the R/G > 0.01, the highest ratio which eliminates almost all hairs and preserves vessels. Figure 7 (c) shows the unmarked hair part if setting the Red/Green ratio>0.02.

2.5.3 Bubble Filtering. From the skin lesion dataset, it was observed that there are dark and light red vessels. For the light red vessels, the red color value is typically much larger than the green value at a given vessel pixel location, with the relationship generalized as $G/R < 0.6$ based on empirical analysis of the dataset. For the dark red vessels, the red color value is also typically larger than the green value at a given vessel pixel location, with the relationship generalized at $G/R < 0.7$ based on empirical analysis of the dataset. Using $G/R > 0.6$ for light red values ranging from 0 to 130 and $G/R > 0.7$ for dark red values ranging from 0 to relative red were used to reduce the labeling of bubble type regions as vessels. Figure 8 presents a skin lesion image example showing the original image with bubbles in (a), bubble regions labeled as vessels using the vessel detection algorithm in (b) and the reduction of bubble regions labeled as vessels in (c).



(a) (b) (c)

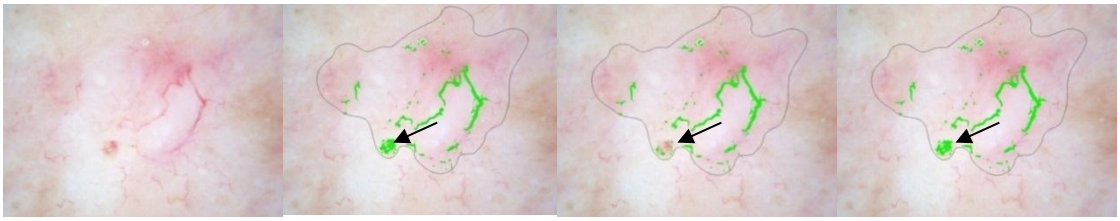
Figure 7. Hair filtering. (a) Hair labeled as vessels. (b) Hair Areas after Removal by R/G >0.01 Filter (c) Hair Areas after Removal R/G> 0.02 Filter. Filter shown in (b) is optimal.



(a) (b) (c)

Figure 8.Bubble filtering. (a) Original image. (b) Bubble labeled as vessels. (c) Unmarked bubble using the algorithm presented above.

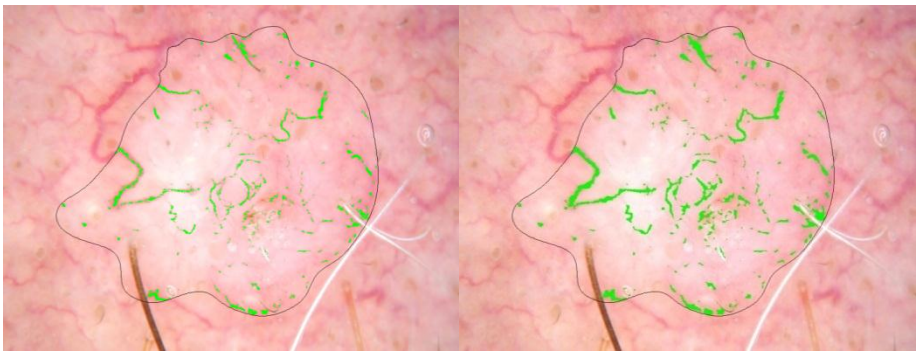
2.5.4 Blob Filtering for Mask Density. Some blobs such as the ulcer in Figure 9 satisfy color drops and remain in the vessel mask. A 41x41 square mask centered on all mask pixels is used for a density screen. A density screen of 70% or more of the pixels inside this square is used, and all centered pixels with high density are unmarked (arrow in Figure 10(c)). Figure 10(d) shows the comparison image with setting the density 80%



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

Figure 9. Blob density filtering. (a) Original image. (b) Big blob labeled as vessel. (c) Unmarked big blob. (d) Big blob unmarked with 80% density.

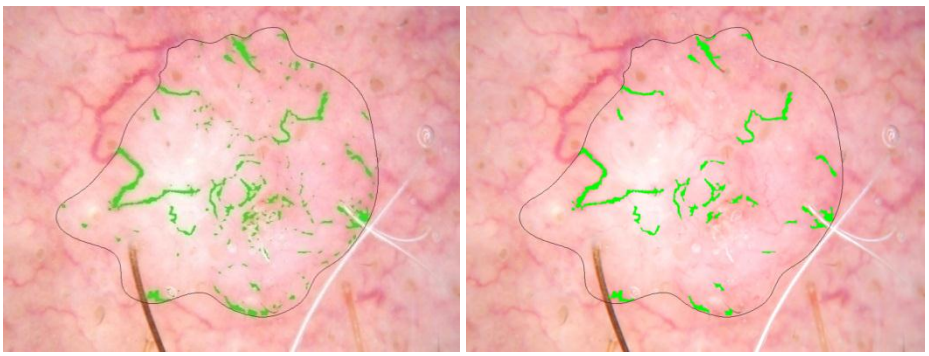
2.5.5 Post-processing: Connecting Vessel Segments. Image dilation and erosion are used to connect vessel segments. After noise filtering, vessel segments become disconnected. The disconnected segments are joined by first dilating and then eroding with a circular structuring element. Connecting segments is accomplished using a dilating element of greater radius than the eroding element. The optimum radii are found to be 3 for dilation and 2 for erosion, as shown in Figure 10.



(a) (b)

Figure 10. (a) Mask after noise filtering. (b) Mask after dilation, radius 3, and erosion, radius 2.

2.5.6 Post-processing: Length and Area Lower Bounds. The vessel mask after performing all of the previous noise removal steps was still noisy. The vessel mask image is skeletonized using Matlab® function bwmorph (BW, 'skel') to ,remove short objects that are not long enough to be considered linear vessels (< 30 pixels long) or too small (< 40 pixels area). The result of the area lower bound is shown in Figure 11.



(a) (b)

Figure 11. Lower area bound. (a) Mask with noise. (b) Mask after area lower bound.

2.5.7 Post-processing: Independent and Principal Component Analysis, Histogram Pursuit. The final vessel masks on some images do not detect all the vessels and do not detect all segments of detected vessels. Fainter vessels and fainter vessels segments are least likely to be detected. Because the vessel masks at this stage were reliable, an attempt was made to expand the

amount of vessel area found without finding many false vessels. Three techniques were tried. Although independent component analysis [13] proved better than principal component analysis [14], almost no new vessel areas were found with either method. Similar results were found with independent histogram pursuit [15].

2.6 VESSEL FEATURE GENERATION

Vessel masks are computed on the skin lesion image dataset using the algorithm presented in section 2.a-2.h above. The final vessel masks on some benign images may have telangiectatic vessels. Table 1 presents the vessel-based features investigated in order to discriminate BCC from benign skin lesions images. The first 10 features are general vessel descriptors (1-10). For this feature group, the first six features were selected to best represent the narrower, longer, and more numerous vessels in BCC. The seventh feature, standard deviation of object (vessel) width, represents the more uniform width of BCC vessels. The next three eccentricity features represent the straighter vessels in BCC (Figure 1). The second feature set, denoted as Object Area Descriptors, includes 10 features computed by morphologically eroding the vessel mask with a circular structuring element of radius from 1 to 10 and finding the remaining total mask area after each erosion and dividing the remaining mask area by the lesion area. The third feature set, denoted as Object Number Descriptors, includes 10 features determined by morphologically eroding the final vessel mask with a circular structuring element of radius from 1 to 10 and record the remaining object number after each erosion. The Object Area and Object Number Descriptors represent different approaches to quantify variations in vessel width. Although the classic dermoscopy feature for BCC is arborizing telangiectasia, with varying vessel widths, observation of numerous vessels in BCC and benign images shows a greater variation in benign vessel widths, as seen in Figure 1. Upon computing the features for the entire dataset, multivariate analysis was performed using the SAS procedure logistic stepwise logistic procedure to determine for feature analysis and selection. The measures included in the final model are shown in Table 2 in the order selected using the feature labels from Table 1. The most significant measures in the final model measured the thinness and straightness of the vessels. The SAS stepwise selection by logistic procedure resulted in the

following variables in the final model. Chi-square significance in the model, which tests against the null hypothesis that the predictors' regression coefficient is not equal to zero in the model, is shown in the rightmost column of Table 2[16].

Table 1. Vessel-based features investigated with descriptions.

Feature Number	**Measure (all normalized for area except those marked*)**	**Description (measures marked # taken after skeletonizing)**	**Meaning**
1	Object number	Total number of vessels/lesion area	Basal cell carcinoma (BCC) have more vessels
2	Maximum object length	Maximum length for all vessels/square root lesion area	BCC vessels are longer
3	Average object length	Average length for all vessels/square root lesion area	BCC vessels are longer
4	Maximum object area	Maximum area for all vessels/lesion area	BCC vessels are larger
5	Average object area	Average area for all vessels/lesion area	BCC vessels are larger
6	Average object width	Average width for all vessels/square root lesion area	BCC vessels are narrower
7	Standard deviation object width*	Standard deviation for all vessel widths	BCC vessel widths are more uniform
8	Maximum eccentricity*	Maximum ratio of distance between the foci of the ellipse enclosing the vessels and its major axis length	BCC vessels are straighter
9	Average eccentricity*	Average ratio of the distance between the foci of the ellipse enclosing the vessels and its major axis length	BCC vessels are straighter
10	Minimum eccentricity*	Minimum ratio of the distance between the foci of the ellipse enclosing the vessels and its major axis length	BCC vessels are straighter
Object number descriptors (11–20)	Object number after 1–10 erosions* (10 features)	Erode the final vessel mask with circular structuring element of radius from 1 to 10 and record the remaining object number after each erosion	BCC object fewer after given number erosions
Object area descriptors (21–30)	Object area after 1–10 erosions (10 features)	Erode the vessel mask with circular structuring element of radius from 1 to 10 and record the remaining mask area for each erosion and divide by square root of the lesion area	BCC object areas smaller after given number erosions

Table 2. Feature measures generated from final vessel mask, included in final model, SAS procedure logistic.

Order of feature selection	Feature number	Chi-Square Significance
1	30	<.0001
2	9	<.0001
3	2	0.0210
4	8	0.1836
5	6	0.0602
6	7	0.0101
7	11	0.0816
8	12	0.0444
9	16	0.1274
10	29	0.1156
11	21	0.0576
12	22	0.0030
13	17	0.2492

2.7 NEURAL NETWORK METHODS FOR BCC DIAGNOSIS

In order to evaluate the effectiveness of the vessel detection process, lesion discrimination is performed using the two classes Basal Cell Carcinoma (BCC) and benign lesions. Standard backpropagation neural network architectures were explored for lesion discrimination. All neural network architectures were implemented in Matlab®,using sigmoid transfer functions are in the input and hidden layers and a linear transfer function in the output layer.

3. EXPERIMENTS PERFORMED

For BCC discrimination from benign lesions over the experimental data set, four different combinations of features computed from the telangectasia masks were investigated for inputs to the different neural network architectures. First, all 30 features listed in Table 1 were examined, including the general, object number descriptors and object area descriptors. An architecture of

31x5x1 was used for the standard backpropagation neural network implementation, with 30 features and a bias in the input layer, five nodes in the hidden layer and one output. Second, 20 the general and object number descriptors from Table 1 were examined, providing 20 input features. An architecture of 21x5x1 was utilized for the neural network implementation. Third, the general and object area descriptors were explored, providing 20 features. An architecture of 21x5x1 were used for the neural network implementation. Fourth, descriptors selected from logistic procedure given in Table 2, providing 13 features. An architecture of 14x5x1 was utilized for the neural network implementation. Because of the limited number of BCC cases, a leave-one-out training/test methodology is used for all neural network architectures. The neural networks are trained up to 15 epochs or until root-mean- square error was <0.001, using online (stochastic/delta) learning, with the weights adjusted after each pattern presentation. In this case, the next input pattern is selected randomly from the training set, to prevent any bias that may occur due to the sequences in which patterns occur in the training set. With the target value for the telangiectases data set set to 1 and the benign data set set to 0, network outputs after testing are between -1 and 1. ROC curves are generated for classification results based on the neural network outputs obtained for the leave-one-out cases. The ROC curve is a plot of the sensitivity for a binary classifier system as its discrimination threshold is varied. In other words, the ROC curve provides the corresponding true positive and false negative rates at each discrimination threshold.

4. RESULTS

Figure 12 shows the plot of ROC curves and areas under the ROC curves, denoted as AUC, for the neural network results based on the four different feature combinations using on-line neural network training and leave-one-out training and testing. For the different ROC curves presented, the vertical axis shows the true positive rate, and the horizontal axis gives the false negative rate.

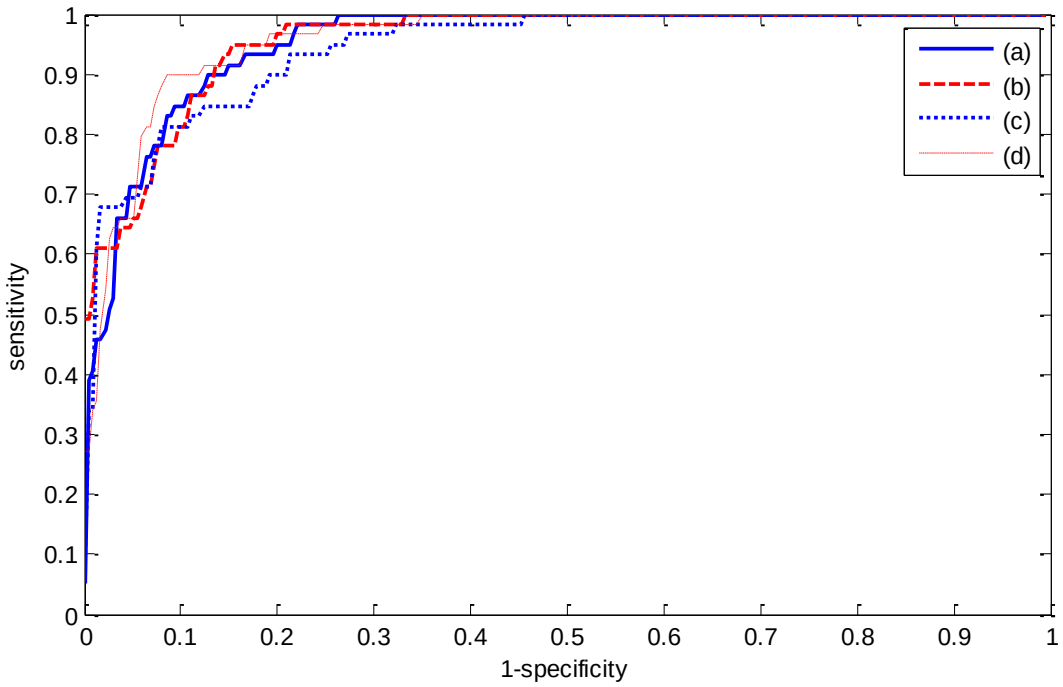


Figure 12. Receiver operating characteristic curve and area under curve (AUC) results for different feature combinations. (a) All 30 features from Table 1 with AUC=0.9548. (b) General descriptors and object area descriptors with AUC=0.9670. (c) General descriptors and the object number descriptors with AUC=0.9482. (d) Reduced feature set selected using SAS Procedure logistic with AUC=0.9547.

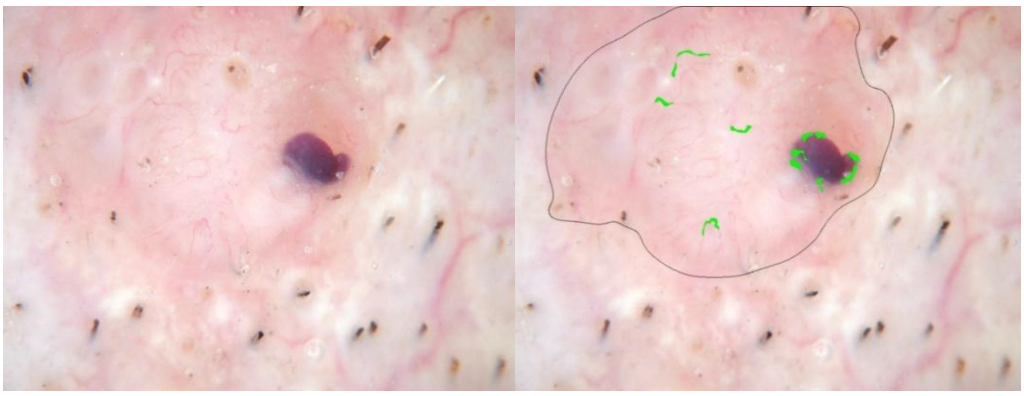
5. DISCUSSION

5.1 DIAGNOSTIC ACCURACY

Based on AUC analysis from Figure 12, all feature combinations achieved at 94.8% diagnostic accuracy. The diagnostic results based on all 30 features are 95.5%. Using the general descriptors and object area descriptors adds approximately 1% to this diagnostic accuracy (96.7%), and using the general descriptors and object number descriptors subtracts approximately 1% from the diagnostic accuracy (94.8%). The reduced feature set based on SAS analysis yields AUC results (95.5%) similar to all 30 features but slightly <20 features consisting of the general descriptors and the object area descriptors.

5.2 EXAMPLES OF ERRORS

Figure 13 (a) shows one of the BCC images which is detected as a benign lesion falsely. The detected vessels are not long enough to be considered BCC telangiectasia. Figure 13 (b) shows one of the benign lesion images which is discriminated as BCC falsely. The noise around the bubbles and the noise around the hair could not be removed after applying the noise filters. Those areas are long and big enough to be considered as telangiectasia.



(a) (b)

Figure 13. BCC misdiagnosed. (a) Original image. (b) Telangiectasia mask.

5.3 RELATION TO CLINICAL PRACTICE

Although textbooks describe telangiectasia visualized in basal cell carcinoma as arborizing, we found that arborizing telangiectasia are actually present in only a small number of basal cell carcinomas. Our attempt to capture this feature by eroding the telangiectasia by variably-sized structural elements and measuring areas left actually resulted in somewhat decreased classifier accuracy (Figure 12). In current clinical practice, it is possible to find smaller basal cell carcinomas by dermoscopy than by clinical inspection. These smaller basal cell carcinomas in our experience (wvs and dc) lack arborizing telangiectasia. Although almost all of these small BCCs have telangiectasia, they can be short and thin. Normalization of lengths and areas helps to detect these smaller BCCs.

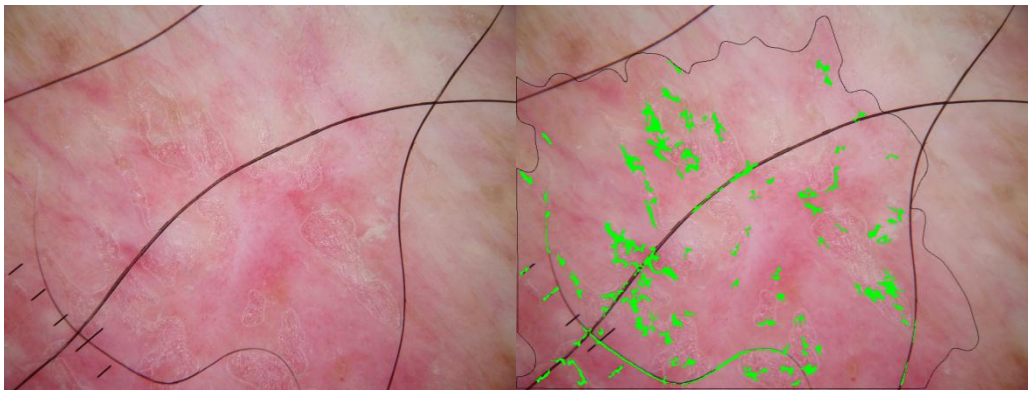
5.4 STRUCTURE DETECTION VS. DIAGNOSIS

We have chosen BCC detection rather than vessel detection as the endpoint. Although vessel detection is inherently easier, BCC detection has potential direct clinical application. Small BCC are detectable early by dermoscopy, and potentially detectable by the automated methods described here.

5.5 LIMITATIONS OF STUDY AND SCOPE OF FUTURE RESEARCH

Errors such as those shown in Figures 13 and 14 are due to 1) the inability of image processing to always eliminate noise, in this bubbles and hair which have sharp drops which mimic telangiectasia, 2) incomplete manifestation of telangiectasia in early lesions, and 3) presence of telangiectasia in many benign lesions. It is likely that the diagnostic accuracy can be improved by increased ability to find dim bubbles, such as those in Figure 14.

Lesion images unselected and were acquired as time permitted in the clinic. It is possible that the results obtained may not be generalized. It is planned to pursue telangiectasia detection methods on a larger set of lesions.



(a) (b)

Figure 14. Benign lesion misdiagnosed. (a) Original image. (b) Telangiectasia mask.

REFERENCES

1. Stolz W, Braun-Falco O, Bilek P, Landthaler M, Burgdorf WHC, Cognetta A. "Color Atlas of Dermatoscopy", 2nd Edition, Berlin: Blackwell Science, 2002, p. 33.

2. Johr RH, Soyer P, Argenziano G, Hofman-Wellenhof R, Scalvenzi M. "Dermoscopy the essentials". Edinburgh: Elsevier, 2004, pp.107.

3. Bolognia JL, Jorizzo JL, Rapini RP, ed., "Dermatology", New York: Mosby, 2003, pp. 1653-1654, 2406.

4. Thackray BD, Nelson AC. "Semi-automatic segmentation of vascular network images using a rotating structuring element (ROSE) with mathematical morphology and dual feature thresholding". IEEE Trans Med Imag 1993, 12(3), pp.385-392.

5. Stansfield SA, "ANGY: A rule-based expert system for automatic segmentation of coronary vessels from digital subtracted angiograms". IEEE Trans Patt Anal Mach Intell 1986, PAMI-8(2), pp.188-199.

6. Fei Y, Park JW, "Automatic segmentation of liver blood vessels using level set method". Proceedings of International Conference on Audio, Language and Image Processing, Shanghai, China, July 7-9, 2008, pp. 1718-1720.

7. Matsopoulos GK, Asvestas PA, Delibasis KK, Mouravliansky NA, Zeyen TG. "Detection of glaucomatous change based on vessel shape analysis", Comput Med Imag Graph, 2008, 32(3), pp.183-192.

8. Klingensmith JD, Vince DG, "B-spline methods for interactive segmentation and modeling of lumen and Bessel surfaces in three-dimensional intravascular ultrasound", Comput Med Imag Graph 2002, 26(6), pp.429-438.

9. Saad AA, Shapiro LG, "Shape decomposition approach for ultrasound color Doppler image segmentation", Proceedings of the 18th International Conference on Pattern Recognition, Hong Kong, September 2006, 4, pp. 691-694.

10. Staal J, Abramoff MD, Niemeijer M, Viergever MA, van Ginneken B. "Ridge-based vessel segmentation in color images of the retina", IEEE Trans Med Imag, 2004, 23(4), pp.501-509.

11. Bhuiyan A, Nath B, Chua J, Kotagiri R. "Blood vessel segmentation from color retinal images using unsupervised texture classification", Proceedings of the IEEE International Conference on Image Processing, San Antonio, TX, November 2007, 5, pp. v521-524.

12. Celebi ME, Kingravi HA, Iyatomi H, Aslandogan YA, Stoecker WV, Moss RH, et. al. "Border detection in dermoscopy images using statistical region merging", Skin Res Technol. 2008, 14(3), pp.347-53.

13. Stone JV. "A Brief Introduction to Independent Component Analysis in Encyclopedia of Statistics in Behavioral Science", Everitt BS, Howell DC, eds. Chichester, John Wiley & Sons, 2005. pp. 907-912.
14. Jolliffe IT. "Principal Component Analysis", Series: Springer Series in Statistics, 2nd ed., New York: Springer, 2002.

15. Gómez DD, Butakoff C, Ersbøll BK, Stoecker W. "Independent histogram pursuit for segmentation of skin lesions". IEEE Trans Biomed Eng. 2008, 55(1), pp.157-61.

16. Schervish MJ, "P Values: What They Are and What They Are Not", The American Statistician 1996, 50(3), pp.203–206.

VI. GRAPHICAL FIGURE CLASSIFICATION USING DATA FUSION FOR INTEGRATING TEXT AND IMAGE FEATURES

Beibei Cheng[a], Sameer Antani[b], R. Joe Stanley[a], George R. Thoma[b]

[a]Department of Electrical and Computer Engineering, Missouri University of Science and Technology, Rolla, Missouri

[b]U.S. National Library of Medicine, National Institutes of Health, Bethesda, Maryland

ABSTRACT

This paper describes an image+text learning approach to automatically identify three graphical figure types commonly found in biomedical literature, namely, diagrams, statistical figures and flow charts. The main purpose of this work is to improve multimodal (image+text) information retrieval for figures in biomedical journal articles. In this contribution, we describe a data fusion approach to combine information from both text and image sources, believed to contain complementary information. Text information about the image is extracted from the figure caption. The data fusion process includes a hybrid of EA and BPSO (EABPSO) method applied to find an optimal subset of extracted image features with Chi-square statistic and information gain utilized to select the optimal subset of extracted text features. Then, an optimal subset of image features and an optimal subset of text features are input to Multi-Layer Perceptron Neural Network classifiers, respectively, whose outputs are characterized as fuzzy sets to determine the final classification result. Evaluation performed on 1707 figure images extracted from a test subset of BioMedCentral® journals extracted from PubMed Central ® repository provided by the U.S. National Library of Medicine yielded classification accuracy as high as 96.1%.

Keywords: Image Processing, Feature Selection, Data Fusion, Binary Particle Swarm Optimization (BPSO), Evolutionary Algorithm (EA), Multi-Layer Perceptron Neural Network (MLP-NN), Fuzzy Set Union, Fuzzy Set Intersection

1. INTRODUCTION

A variety of biomedical images needed for instructional purposes or in support of clinical decisions are often found in biomedical articles, but are not easily accessible to retrieval tools. Broadly, the images found as figures in the articles can be classified into two categories: regular and graphical images, respectively. Regular images are those that are acquired through an imaging device and include MRI, CT, X-ray, photographs, etc. Graphical images (also described as, graphics) are those that are created by authors to illustrate biomedical processes or content or biomedical data analyses. Graphical images can be further classified into four classes: diagrams, statistical figures, flow-charts, and tables. It is necessary to annotate these images to support multimodal (image + text) medical information retrieval and clinical decision support systems. Image type identification is a key step toward such automatic annotation for figures extracted from scientific publications. The task of separating regular images from graphics is reported in [1] and the task of separating diagrams, statistical figures, flow-charts, and tables is reported in [2]. In this research, we classify diagrams, statistical figures, flow-charts since tables are often represented in XHTML form in online articles today. The classification is based on a larger amount of data set and combined information from both the images and their captions. For this purpose, 1707 graphical figures are selected from a test subset of BioMedCentral ® journals available in the Open Access dataset from the PubMed Central® repository of the U.S. National Library of Medicine, part of the National Institutes of Health. The caption for each figure is automatically extracted from NXML coded full-text journal articles using a rule-based script. Figure 1 provides the examples for each type of figure with its caption. Since both figure and its caption bring complete meaning to the image and they are complementary information for the reader, the objective of this project is to develop an image+text learning approach to identify the three graphics types, namely, graph, flow chart, and diagram.

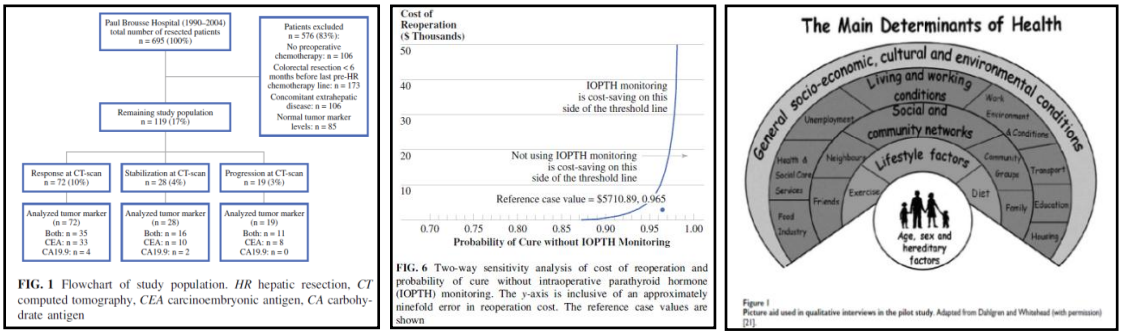


FIG. 1 Flowchart of study population. *HR* hepatic resection, *CT* computed tomography, *CEA* carcinoembryonic antigen, *CA* carbohydrate antigen

FIG. 6 Two-way sensitivity analysis of cost of reoperation and probability of cure without intraoperative parathyroid hormone (IOPTH) monitoring. The y-axis is inclusive of an approximately ninefold error in reoperation cost. The reference case values are shown

Figure 1
Picture aid used in qualitative interviews in the pilot study. Adapted from Dahlgren and Whitehead (with permission) [21].

(a) (b) (c)

Figure 1. Examples of graphical figure with its caption. (a) Flow chart, (b) Graph, and (d) Diagram.

2. RELATED WORK

Several approaches to chart type identification have been proposed with varying degrees of success. Zhou et al proposed a method of graphics type identification based on the Hough transform [3]. Huang et al presented a classification method that uses the revised diverse density algorithm [4]. Liu et al developed an approach using vectorized graphical information extraction from an image [5]. There are some common drawbacks in these existing approaches. First, these approaches are developed to identify a particular graphic, viz., statistical charts, such as pie, bar and line charts. Second, only a small number of charts have been examined, which make the performance results published for these algorithms inapplicable for the wide variety of graphical image types. Finally, none of these algorithms integrate text and image information to do the classification while the caption usually contains important information about the image. i.e. the keywords "Flowchart", "y-axis" and "picture" in the caption of Figure 1(a), 1(b) and 1(c) demonstrate each image's type.

Approaches to combine text and image information have been actively researched in the recent years. Shatkay et al described an approach for document categorization by using both subfigure identifier terms and text terms [6]. Caicedo et al presented a method for detecting relevant images for the query topic by combining visual features and text data using latent semantic kernels [7]. In this research, we implement a feature-based learning approach to classify flow chart, graph and diagram based on the following steps: 1), image features are extracted from each individual image and a hybrid of Evolutionary Algorithm and Binary Particle Swarm Optimization (EABPSO) method is investigated to find an optimal subset of extracted image features; 2), text features are extracted from each image's caption and Chi-square statistic and information gain are used to obtain an optimal subset of extracted text features; and 3), fuzzy logic union/intersection functions are used to fuse both image and text classification outputs.

The remaining sections of this paper include the following: the image feature extraction and selection, the text feature extraction and selection, and fuzzy logic union/intersection functions to combine classifiers' outputs, experiments results, discussion and conclusions.

3. IMAGE FEATURE EXTRACTION AND SELECTION

3.1 IMAGE FEATURE EXTRACTION

The approach for image feature extraction is provided in [2]. Those extracted seventy-nine features can be grouped into five sets, viz., textural features, shape features, weighted density distribution (WDD) features, Hough features and hole object features. Besides Hough features are used in previous research [3], textural features measure smoothness, coarseness, and regularity of an image; shape features extract edge (object shape) characteristics in the image; WDD features measure the image's symmetry; hole object features compare the objects in an image.

3.2 IMAGE FEATURE SELECTION

Three different feature selection algorithms have been examined in [2] to obtain the optimal feature subset, which are Evolutionary Algorithm (EA), Binary Particle Swarm Optimization (BPSO) and a combined Evolutionary Algorithm and Binary Particle Swarm Optimization (EABPSO), the result in [2] proves the hybrid one generate the best output. Therefore, we apply the hybrid of EA and BPSO (EABPSO) method to find an optimal subset of extracted image features and an optimal subset of extracted text features. The scheme, where each individual in a population is an N-dimensional binary vector with each element of the vector representing a feature and N being the total number of features, is used for candidate (feature subset) representation,. For each element of the binary vector, '1' means that the corresponding feature is selected. The initial population is randomly initialized in the sense that each element in a vector is randomly picked as 0 or 1. The fitness value for EABPSO is set to the accuracy of the classifier applied to the selected feature set.

Figure 2 shows an overview of the EABPSO procedure. As can be seen, EA and BPSO both work with the same initial population. To solve an M-dimensional problem, 2M individuals are randomly generated in the sense that each element in an individual is randomly picked from 0 or 1. These individuals may be considered analogous to chromosomes in the case of EA, or as particles in the case of BPSO. The 2M individuals are sorted by fitness, and the top M individuals are fed into EA to create M new offspring by crossover and mutation operations, as described in [2]. The new offspring are used as the input to BPSO to compute M new particles. The new parent and new particles are combined and sorted in preparation for repeating the entire run.

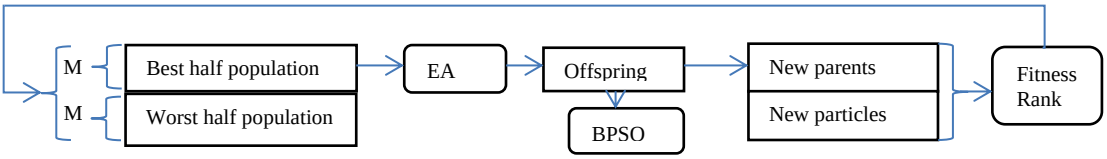


Figure 2. Overview of EABPSO procedure.

4. TEXT FEATURE EXTRACTION AND SELECTION

4.1 TEXT FEATURE EXTRACTION

The caption corresponding to each image is collected in a separate data set. Each word extracted from the caption becomes a feature and its numeric value is its occurrence frequency. Then the raw data is cleaned by removing frequently occurring words and reducing the words to their etymological stems. The cleaned word vector is analyzed to have the most relevant features selected for the modeling stage as shown in the following section.

4.2 TEXT FEATURE SELECTION

After converting the caption list into a cleaned and reduced feature list, we obtain approximately a 1080 word vector. It is unnecessary to implement all of these features as the post-processing input since certain features in the initial data are either ineligible to be used as independent variables for the data mining tasks or are irrelevant to the project and it is time consuming to process with the entire word vector. The top two hundred features which contain more important information than others would be chosen as the input features based on different criteria such as information gain and chi-square. These two different techniques to select the most relevant features for classification are described as follows.

Information gain: information gain [8] is frequently employed as a term-goodness criterion in the field of machine learning. It measures the number of bits of information obtained for category prediction by knowing the presence or absence of a term in a caption. The top twenty features with its information gain value as ranked by the information gain filter in Weka are listed in Table 1.

Table 1. Top 20 features with its information gain value.

flow 0.22442	stud 0.11965	chart 0.09159	tr 0.07891
signif 0.07661	diagram 0.06933	group 0.06722	differ 0.0579
valu 0.05728	flowchart 0.05546	particip 0.04419	surviv 0.04094
tim 0.03842	plot 0.03726	lin 0.03707	scor 0.03593
titl 0.03476	curv 0.03472	schem 0.0347	error 0.03243

Chi-square statistic: Chi-square[9] is one of the most widely used theoretical probability distributions in inferential statistics. A chi-square statistic is used to investigate whether distributions of categorical variables differ from one another. These top twenty ranked features are shown in Table 2.

Table 2. Top 20 features with its Chi-square value.

flow 563.76962	stud 315.90134	chart 229.37274	tr 209.13612
diagram 168.19293	signif 141.32304	group 134.91097	flowchart 132.43568
particip 119.29557	differ 114.44652	valu 112.49262	schem 96.12087
draw 90.75539	titl 79.46252	random 78.28262	proc 75.5469
select 73.88812	tim 73.84643	recruitm 73.19221	map 69.98771

The list of features obtained from each of these methods is combined to create a list of most selected features. Two hundred features are chosen to create the final feature list.

5. OUTPUT COMBINATION

We integrated both image and text features for the purpose of image classification. After getting the output from the classifier of the optimal image feature subset and the output from the classifier of the optimal text feature subset, the final output is generated by combining these two results. Fuzzy logic controller (FLC) can be used to deal with this problem since this problem has nonlinear and imprecise measurement information and FLC allows nonlinear input/output relationship to be expressed by a set of qualitative if then rules. However, it is difficult to set the optimal cut-off value to produce a degree of membership for each fuzzy set as well as the fuzzy rules. Thus some classes of fuzzy set unions and intersections are implemented here to fuse two outputs from the classifiers. Five proposed classes of fuzzy set union along with the corresponding class of fuzzy set intersections are given in Table 3.

Table 3. Five classes of fuzzy set union and intersection.

Reference	Fuzzy Unions	Fuzzy Intersections	Range of Parameter
Schweizer& Sklar[10]	$1-\max[0,(1-a)^{-p}+(1-b)^{-p}-1]^{1/p}$	$\max(0,a^{-p}+b^{-p}-1)^{-1/p}$	$p\in(-\infty,\infty)$
Frank[11]	$1-\log_s[1+\frac{(s^{1-a}-1)(s^{1-b}-1)}{s-1}]$	$\log_s[1+\frac{(s^{a}-1)(s^{b}-1)}{s-1}$	$s\in(0,\infty)$
Yager[12]	$\min[1,\ (a^w+b^w)^{1/w}]$	$1-\min[1,\ ((1-a)^w+(1-b)^w)^{1/w}]$	$w\in(0,\infty)$
Dubois & Prade[13]	$\frac{a+b-ab-\min(a,b,1-\alpha)}{\max(1-a,1-b,\alpha)}$	$\frac{ab}{\max(a,b,\alpha)}$	$\alpha\in(0,1)$
Dombi[14]	$\frac{1}{1+[(\frac{1}{a}-1)^{-\lambda}+(\frac{1}{b}-1)^{-\lambda}]^{-1/\lambda}}$	$\frac{1}{1+[(\frac{1}{a}-1)^{\lambda}+(\frac{1}{b}-1)^{\lambda}]^{1/\lambda}}$	$\lambda\in(0,\infty)$

6. EXPERIMENTAL RESULTS

The experimental data set consists of 1707 medical images annotated by type including 324 diagrams, 905 graphs, and 478 flow charts, extracted from a test subset of BioMedCentral® journals selected from the Open Access subset of NLM's PubMed Central repository.

79 image features are extracted from these images and EAPSO is applied to find an optimal subset of extracted features that are then classified using MLP Neural Network. For training the MLP Neural Network, we have its architecture as (X+1)x10x3 consisting of an input layer of X selected features obtained from EAPSO and a bias with linear neurons, a hidden layer of 10 neurons with sigmoid transfer functions, and an output layer of three output with a linear neuron. The classification output can be written as below:

$$ImagePrediction = Argmax_{i=1,2,3}(ImageNN_i) \quad (1)$$

Where $ImageNN_i$ is the i output of neural network. A five-fold cross validation is used to set up the training and testing data sets. The data set is divided into five parts where 4/5th is used for training and the rest is used for testing. This procedure is repeated five times. Therefore, for each time, the training set is 1366 and the representative test set is 341 images. The accuracy of the classifier presented in the next section is based on averaging the accuracy of the five test sets. We can measure recall, precision, and accuracy as summarized in Table 4. Precision and recall for each image type are defined as:

$$Precision = \frac{tp}{tp+fp}; Recall = \frac{tp}{tp+fn} \quad (2)$$

Where, *tp* is the number of true positives, i.e., the number of flowchart identified as flow chart, *fp* is the number of false positive, i.e., the number of flowchart identified as non-flowchart, *fn* is the number of false negative, i.e., the number of non-flowchart identified as flowchart. Recall in this context is also referred to as the True Positive Rate.

From Table 4, Neural Network delivers 1)93.2% with the entire image feature set and 2)94.4% with the optimal image feature subset. The number of selected features is 45.

1080 text features are initially extracted from each image caption and we obtain 200 top ranked features from them based on the information gain and Chi-square criteria. Similar as the image feature classification training, these 200 features are used to train a MLP neural network, with the architecture 201x10x3 consisting of an input layer of 200 features and a bias with linear neurons, a hidden layer of 10 neurons with sigmoid transfer functions, and an output layer of three output with a linear neuron. The result is given in Table 5 based on a five-fold cross validation.

Then we use fuzzy logic union/intersection function (Table 3) to combine the image classification output and the text classification output. Take the fuzzy intersection function from Dubois & Prade for example. The combined output can be written as:

$$Prediction = Argmax_{i=1,2,3} \frac{ImageNN_i \cdot TextNN_i}{\max(ImageNN_i, TextNN_i, \alpha)} \quad (3)$$

where $\alpha \in (0,1)$, $\mathrm{ImageNN_i}$ is the i output of neural network trained by image features, $\mathrm{TextNN_i}$ is the i output of neural network trained by text features. Table 6 provides the final prediction accuracy with the chosen parameter.

Table 4. Performance comparison for image feature set with MLP NN classifier.

	Entire Set		Optimal Subset(45)	
	Precision	Recall	Precision	Recall
FlowChart	0.955	0.937	0.954	0.950
Graphic	0.947	0.944	0.953	0.958
Diagram	0.86	0.892	0.903	0.895
Accuracy	93.2%		**94.4%**	

Table 5. Performance for text feature set with MLP NN classifier.

FlowChart	0.868	0.866
Graphic	0.888	0.935
Diagram	0.859	0.735
Accuracy	**87.8%**	

Table 6. Performance for combination by fuzzy logic union and intersection.

Reference	Fuzzy Unions	Parameter	Fuzzy Intersections	Parameter
Schweizer & Sklar	94.4%	p=-1	95.5%	p=-0.7
Frank	96.0%	s =2	95.4%	s=4
Yager	**96.1%**	w=1	95.4%	w=6
Dubois & Prade	96.0%	α =0.1	95.4%	α =0.1
Dombi	96.0%	λ =0.1	95.4%	λ =10

7. DISCUSSION AND CONCLUSIONS

As seen in Table 4, the higher image feature classification accuracy of 94.4% is achieved with the feature subset selected by EAPSO than the entire feature set, which shows that EAPSO removes the negative related or unrelated features. Besides, as shown in Table 6, the accuracy of every combination by fuzzy logic union and intersection is at least the same as the accuracy of either image feature or text feature. The best overall classification accuracy is 96.1% by Yager with the parameter w to be 1.

This research presents a first exploration of the image type classification by combining both image information and its caption information, extended from the previous research [2]. This research shows data fusion techniques are useful with the early data fusion to remove the negative related or unrelated features and the late data fusion to combine the resource information. In addition, this research proves that the caption data has much to offer in support of image type classification. The experimental results demonstrates that integration of various image processing techniques, feature extraction techniques, and data fusion techniques for information combination as proposed in this paper can achieve high classification accuracy. Future research will focus on character recognition within these image types and statistical chart interpretation.

ACKNOWLEDGEMENTS

This work was supported by NLM under contract number 276200800413P and the Intramural Research Program of the National Institutes of Health (NIH), NLM, and Lister Hill National Center for Biomedical Communications (LHNCBC).

REFERENCES

1. B. Cheng, S. Antani, R.J. Stanley, G.R. Thoma, "Automatic segmentation of subfigure image panels for multimodal biomedical document retrieval", Proceedings of the SPIE Electronic Image, San Francisco, California, 2011, 7874, pp. Z-1-11.

2. B. Cheng, S. Antani, R.J. Stanley, G.R. Thoma, "Graphical image classification using a hybrid of evolutionary algorithm and binary particle swarm optimization", Proceedings of SPIE Electronic Image, San Francisco, California, Jan 2012.

3. Y. P. Zhou, C. L. Tan, Hough technique for bar charts detection and recognition in document images. International Conference on Image Processing, 2000, pp. 494-497.

4. W. Huang, Chart Image Classification Using Multiple-instance Learning. IEEE workshop on Applications of Computer Vision, 2007, pp. 27.

5. R. Liu., W. Huang. , C. L. Tan, Extraction of Vectorized Graphical Information from Scientific Chart Images, Document Analysis and Recognitio, ICDAR, Ninth International Conference on, 2007, 1, pp.521-525.

6. H.Shatkay, N. Chen, D. Blostein, Integrating image data into biomedical text categorization, Bioinformatics, 2006, 22 (14), pp.e446-e453.

7. J. C. Caicedo, J. G. Moreno, E. A. Ni~no, and F. A. Gonzalez. Combining visual features and text data for medical image retrieval using latent semantic kernels. In Proceedings of the ACM MIR'10 International Conference on Multimedia Information Retrieval. ACM, 2010.

8. H. Liu, and R. Setiono, Chi2: Feature selection and discretization of numeric attributes, Proc. IEEE 7th International Conference on Tools with Artificial Intelligence, pp.338-391, 1995.

9. T.M. Mitchell, Machine Learning, the Mc-Graw-Hill Companies, Inc, 1997.

10. B. Schweizer, A. Sklar, Associative functions and statistical triangle inequalities, Publications Mathematicae Debrecen, 10, pp.69~81.

11. M.J. Frank, On the simultaneous associativity of F(x,y) and x+y-F(x,y), Aequationes Math, 19, pp.149~226.

12. R.R. Yager, On a general class of fuzzy connectives, Fuzzy Sets and Systems, 4, pp.235~242.

13. D. Dubois and H. Prade, Fuzzy logics and the generalized modus ponens revisited, Cybernetics and Systems, 15, pp.293~331.

14. Dombi. J, A general class of fuzzy operators, the De Morgan class of fuzzy operators and fuzziness measures included by fuzzy operators, Fuzzy Sets and Systems, 8, pp. 149~163.

VII. GRAPHICAL IMAGE CLASSIFICATION COMBINING AN EVOLUTIONARY ALGORITHM AND BINARY PARTICLE SWARM OPTIMIZATION

Beibei Cheng[a], Renzhong Wang[a], Sameer Antani[b], R. Joe Stanley[a], George R. Thoma[b]

[a]Department of Electrical and Computer Engineering, Missouri University of Science and Technology, Rolla, Missouri

[b]U.S. National Library of Medicine, National Institutes of Health, Bethesda, Maryland

ABSTRACT

Biomedical journal articles contain a variety of image types that can be broadly classified into two categories: regular images, and graphical images. Graphical images can be further classified into four classes: diagrams, statistical figures, flow charts, and tables. Automatic figure type identification is an important step toward improved multimodal (text + image) information retrieval and clinical decision support applications. This paper describes a feature-based learning approach to automatically identify these four graphical figure types. We apply Evolutionary Algorithm (EA), Binary Particle Swarm Optimization (BPSO) and a hybrid of EA and BPSO (EABPSO) methods to select an optimal subset of extracted image features that are then classified using a Support Vector Machine (SVM) classifier. Evaluation performed on 1038 figure images extracted from ten BioMedCentral® journals with the features selected by EABPSO yielded classification accuracy as high as 87.5%.

Keywords: image processing, feature selection, Binary Particle Swarm Optimization (BPSO), Evolutionary Algorithm (EA), Support Vector Machine (SVM), graphical image

1. INTRODUCTION

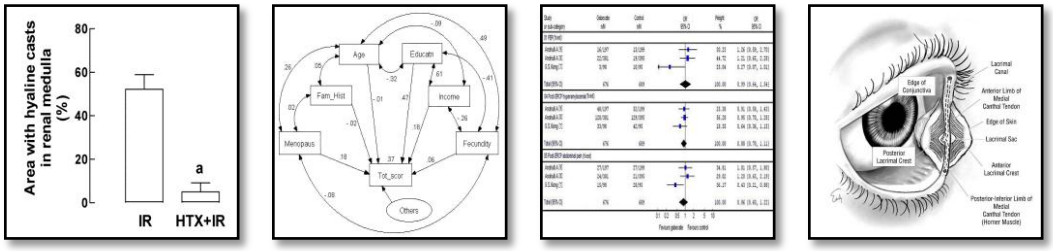
A variety of biomedical images needed for instructional purposes or in support of clinical decisions are often found in biomedical articles, but are not easily accessible to retrieval tools. Broadly, the images found as figures in the articles can be classified into two categories: regular and graphical images, respectively. Regular images are those that are acquired through an imaging device and include MRI, CT, X-ray, photographs, etc. Graphical images (henceforth, graphics) are those that are created by authors to illustrate biomedical processes or content or biomedical data analyses. These images can be further classified into four classes: diagrams, statistical figures, flow-charts, and tables. Although tables are often represented in XHTML form in online articles today, older issues still provide them as images. It is necessary to annotate these images to support multimodal (image + text) medical information retrieval and clinical decision support systems. Graphical figure type identification is a key step toward such automatic annotation for figures extracted from scientific publications. The task of separating regular images from graphics is also a goal of the project and has been reported earlier [1].

Graphics used in medical articles often appear in a variety of formats such as tables, graphs, flow charts, and diagrams as illustrated in Figure 1. For this paper, we use 1038 graphical images selected from ten BioMedCentral journals (Cancer, Cardio, Urology, Gastroenterology, Musculoskeletal Disorders, Nephrology, Ophthalmology, Pulmonary Medicine, Surgery and Dermatology) available in the Open Access dataset from the PubMedCentral® repository of the National Library of Medicine, part of the U.S. National Institutes of Health. The objective of this project is to develop a feature-based learning approach to identify the four graphics types, viz., table, graph, flow chart, and diagram.

This article describes our feature-based learning approach applying Evolutionary Algorithm (EA), Binary Particle Swarm Optimization (BPSO) and a hybrid of EA and BPSO (EABPSO) methods to an optimal subset of extracted image features which are then classified using a Support Vector Machine (SVM) [2] classifier. The proposed approach can address complex, hybrid, and composite graphics, which existing approaches [3, 4, 5] fail to identify satisfactorily. For example, the widely used Hough transform [3] can identify chart types such as pie charts and bar charts by detecting arc and line components inside the image based on the fact that arcs only appear in pie charts while vertical lines with similar length often exist in bar charts.

We also address other graphics types such as tables and flow charts with more complicated compositions. For example, the X-axis in a graph does not have to be present

(Figure 1(a)); a flow chart can contain curved lines (Figure.1(b)); some blocks in a table may be missing (Figure 1(c)); and a diagram may have pictures within it (Figure.1(d)). Classifying such figures can be more challenging. To address these challenges, multiple features, including textural features [6, 7], region property features [8, 9], weighted density distribution (WDD) features [8, 10], Hough features [3] and hole object features, associated with the chart types are extracted. A hybrid of Evolutionary Algorithm [11] and binary Particle Swarm Optimization [12] (EABPSO) is employed to remove irrelevant features using the SVM classifier's output as feedback for evaluating the merits of a generated feature subset.



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

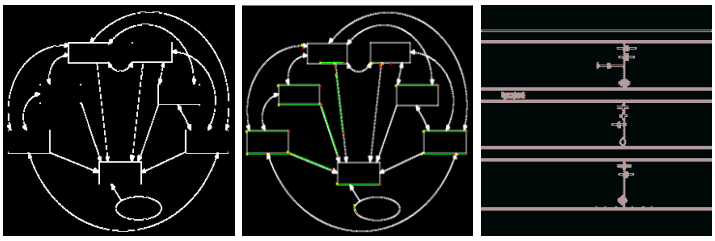
Figure 1. Graphical image type examples. (a) Graph, (b) Flow chart, (c) Table, and (d) Diagram.

2. RELATED WORK

Several approaches to chart type identification have been proposed with varying degrees of success. Zhou et al proposed a method of graphics type identification based on the Hough transform [3]. Huang et al presented a classification method that uses the revised diverse density algorithm [4]. Liu et al developed an approach using vectorized graphical information extraction from an image [5]. There are two common drawbacks in these existing approaches. First, these approaches are developed to identify a particular graphic, viz., statistical charts, such as pie, bar and line charts. Second, only a small number of charts have been examined, which make the performance results published for these algorithms inapplicable for the wide variety of graphical image types.

Techniques from existing research are implemented to detect flow charts. Take Figure 1(b) as an example. Text is first removed from an image and an edge map is obtained through edge detection (Figure 2(a)); Next, straight lines are detected by using the Hough transform (Figure 2(b)). There are three problems with this approach. First, it cannot detect all the lines if there are a large number of line segments of various lengths in a chart. Second, even if all of the lines and arcs are precisely detected by the Hough transform or by the Line net global vectorization [13], it is still difficult to determine the chart type due to complex composition. For example, both flow charts and diagrams can contain lines and arcs. Third, removing text from an image may also remove important information in the chart. For example, only lines are left after preprocessing the table (Figure 2(c)) using this approach.

To solve the first and second problems, in addition to lines and arcs, more features that are associated with the shape of a chart need to be considered, for example, the region property features and weighted density distribution features. For the third problem, textural features are extracted before removing the text. However, feature extraction may yield some features that are not relevant to classification. Incorporation of these unrelated features may have an adverse effect on the classifier's performance. Feature selection [14], the process of selecting the best feature subset that contains the least number of features, contributes most to accuracy and efficiency. Both EA and PSO are stochastic search procedures and are generally suitable for solving this problem. However, EA has a slower convergence rate so that it usually takes longer to reach the global optimum while PSO has a higher convergence speed and is easily trapped in local optimum, as shown in Section 5. Therefore, a hybrid of EA and PSO is developed to combine the advantages of both and is described in Section 4.

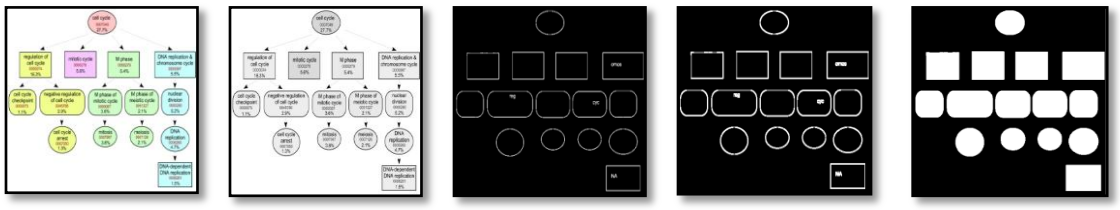


(a) (b) (c)

Figure 2. Examples illustrating problems, discussed in Section 2, in existing approach after preprocessing. (a) Flow chart. (b) Flow chart with marked straight lines. (c) Table image.

3. METHODS

In our approach, an image is preprocessed in four steps with various features extracted at each step. This process is illustrated in Figure 3. The first step is to convert the original image (Figure 3(a)) to a gray scale image (Figure 3(b)) and extract textural features [6, 7]. Second, the gray scale image is binarized using Otsu's method [15] and then small objects with short length are removed so that it contains only frame-like objects (Figure 3(c)). Shape features, weighted density distribution features can be extracted. Third, the Sobel edge filter [16] is applied to the binary image and Hough features are computed (Figure 3(d)). Finally, the hole inside the binary image is filled [17] and features from these holes are extracted (Figure 3(e)).



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

Figure 3. Image preprocessing example. (a) Original image. (b) Gray image. (c) Binary image with chart frame. (d) Binary image with chart frame edge. (e) Binary image with hole objects.

The features thus obtained are then evaluated and subsets are selected by applying feature selection algorithms, i.e., EABPSO. The results from the classifier are used as feedback for assessing the merits of the candidate feature subset. This process proceeds iteratively until an optimized feature subset is obtained. The optimized feature subset is used as the input to the classifier, which determines the type of the chart.

3.1 FEATURE EXTRACTION

Relevant features extracted from the gray scale image and the three object mask images can be grouped into five sets, viz., textural features, shape features, weighted density distribution (WDD) features, Hough features and hole object features. Besides Hough features are used in previous research, textural features measure smoothness, coarseness, and regularity [7] of an image; Shape features extract edge (object shape) characteristics in the image; WDD

features measure the image's symmetry; hole object features compare the objects in an image. Table 1 summarizes these features and further explanations of these are given as follows.

A. Textural features: The textural features are items 1 to 6 in Table 1. However, these features use only histograms, which carry no information regarding the relative positions of pixels with respect to each other. To solve this problem, a co-occurrence matrix [7] is used which considers pixel position. Additional textural features from these co-occurrence matrices are labeled 7 to 30 in Table 1.

B. Shape features: These features are based on region properties and are obtained by applying the MATLAB® function regionprops [18] to the entire image. Ten region property features extracted in this project are listed in Table 1 labeled 31~40.

C. Weighted Density Distribution (WDD) features: The third set of 24 WDD features (rows 41 to 64 in Table 1) is obtained by computing a one dimensional shape profile of each object in the binary chart frame and correlating those profiles with WDD functions [19].

D. Hough features: After the straight lines inside an image are identified by the Hough transform algorithm, five Hough features are generated as shown in Table 1 rows 65 to 73.

E. Hole object features: This set of features is generated from the hole objects inside an image, as illustrated in Figure 4(d). The hole objects are obtained by applying the hole filling algorithm [17]. Five hole object features are generated for the sample image in this paper. They are shown in Table 1 rows 74 to 79.

3.2 CLASSIFIER

SVM classifier is chosen since it delivers a deterministic solution. The key features of SVMs are the use of kernels, the absence of local minima, the sparseness of solution, and the capacity control achieved by optimizing the margin.

Table 1. Extracted Features.

Feature set	Label	Measure	Description
Textural features	1	Mean of histogram	The first moment of gray image
	2	Variance of histogram	The second moment of gray image
	3	Skewness of histogram	The third moment of gray image
	4	Flatness of histogram	The fourth moment of gray image
	5	Maximum of histogram	Uniformity of gray image
	6	Entropy of histogram	Average entropy of gray image
	7~10	Contrast	The intensity contrast of correlation matrices
	11~14	Correlation	The correlation of correlation matrices
	15~19	Uniformity	The uniformity of correlation matrices
	20~23	Closeness	The homogeneity of correlation matrices
	24~27	Strongest response	The maximum probability of correlation matrices
	28~30	Randomness	The average entropy of correlation matrices
Shape features	31	MajorAxisLength	Length (in pixels) of the major axis of the ellipse that has the same normalized second central moments as the region.
	32	MinorAxislength	Length (in pixels) of the minor axis of the ellipse that has the same normalized second central moments as the region .
	33	Axis ratio	Ratio of MajorAxisLength to MinorAxislength.
	34	Normalized area	Area of the region divided by the whole image.
	35	Solidity	Area of the region divided by the convex hull area.
	36	EulerNumber	The number of objects in the region minus the number of holes in those objects.
	37	EquiDiam	The diameter of a circle with the same area as the region.
	38	Extent	Ratio of area to bounding box area
	39	Horizontal MinPixelNo	The minimum number of intersection area for the object and its bounding box horizontally.
	40	Vertical MinPixelNo	The minimum number of intersection area for the object and its bounding box vertically.
WDD features	41~64	WDD	Correlation of binary chart frame and WDD function.
Hough features	65	Line number	Number of straight lines
	66	Longest line length	Longest line's length
	67	Longest line slope	Longest line's slope
	68	2nd Longest line length	Second longest line's length
	69	2nd Longest line slope	Second longest line's slope
	70	Line slope	Average value of lines' slope
	71	Line length	Average value of lines' length
	72	Variance of line slope	Variance of lines' slope
	73	Variance of line length	Variance of lines' length
Hole Object features	74	Hole number	Number of hole objects
	75	Largest hole area	Area of largest hole object
	76	Hole area	Average hole objects' area,
	77	Area variance	Variance of hole objects' area,
	78	Area ratio	Average ratio of hole object's area to its bounding box's area
	79	Area ratio variance	Variance of ratio of hole object's area to its bounding box's area.

4. OPTIMAL FEATURE SELECTION

After the 79 features listed in Table 1 are extracted from the images, Evolutionary Algorithm (EA), Binary Particle Swarm Optimization (BPSO) and a combined Evolutionary Algorithm and Binary Particle Swarm Optimization (EABPSO) are then applied to obtain the optimal feature subset. They all use the same scheme for candidate (feature subset) representation, where each individual in a population is an N-dimensional binary vector with each element of the vector representing a feature and N being the total number of features. For each element of the binary vector, '1' means that the corresponding feature is selected. The initial population is randomly initialized in the sense that each element in a vector is randomly picked as 0 or 1. The fitness values for EA, BPSO and EABPSO are set to the accuracy of the SVM classifier applied to the selected feature set. The algorithms for generating the candidate feature subsets in EA, BPSO and EABPSO are described below.

4.1 EVOLUTIONARY ALGORITHM (EA)

The offspring of the Evolutionary Algorithm are generated as follows: 1) randomly select two parents from the parent pool of M initial candidates; 2) generate two offspring by applying a uniform [20] crossover operator; 3) offspring are then altered by performing a mutation operation. A random parameter ranging from 0 to 1 is generated for each bit of the candidate vector, which will flip once the parameter is greater than a predefined threshold. The next parent pool is selected based on whether the parents or their offspring maximize the classification accuracy. The same process is used for obtaining the next generation of offspring and this process is repeated for N epochs. From the final parent pool, the parent which maximizes the classification accuracy is selected as the final result. Since EA evaluates many points simultaneously in the search space it is more likely to find the global solution but at the cost of higher computation time.

4.2 BINARY PARTICLE SWARM OPTIMIZATION (BPSO)

In implementing BPSO, the velocity and position of a candidate are computed using Eq.1 and Eq. 2 respectively. A sigmoid transformation of the velocity component is applied to keep the velocity values constrained in the range (0, 1). However, the BPSO algorithm can be easily trapped into a local minimum and may lead to premature convergence. It has been observed that when BPSO reaches a local optimal solution, all particles tend to gather around it making it difficult to find a global optimum.

$$V_{mn}(t+1) = wV_{mn}(t) + c_1 rand_1\left(Pbest_{mn} - X_{mn}(t)\right) + c_2 rand_2\left(Gbest_{mn} - X_{mn}(t)\right) \quad (1)$$

$$X_{mn}(t+1) = \begin{cases} 1 & \text{if } (rand < \frac{1}{1+e^{-V_{mn}(t+1)}}) \\ 0 & \text{else} \end{cases} \quad (2)$$

where t is the iteration index (time step), m is the current particle ($1 \leq m \leq M$) in a population of M, n is the attribute element ($1 \leq n \leq N$), $V_{mn}(t)$ is the particle's current velocity, $V_{mn}(t+1)$ is the particle's new velocity, $X_{mn}(t)$ is the particle's current position, and $X_{mn}(t+1)$ is the particle's new position, $Gbest_{mn}$ is the global best position, $Pbest_{mn}$ is the previous best position, c_1 and c_2 are the random value from 0 to 1, w is the learning weight, selected from 0 to 1.

4.3 EABPSO

To address the individual shortcomings of these two algorithms, we design EABPSO combining the feature evolution idea of EA and BPSO into a hybrid solution appropriate for discrete (binary) problems. EABPSO is an improvement over prior hybrid evolutionary algorithms [21, 22] that solved continuous problems.

Figure 4 shows an overview of the EABPSO procedure. As can be seen, EA and BPSO both work with the same initial population. To solve an M-dimensional problem, 2M individuals are randomly generated in the sense that each element in an individual is randomly picked from 0 or 1. These individuals may be considered analogous to chromosomes in the case of EA, or as particles in the case of BPSO. The 2M individuals are sorted by fitness, and the top M individuals are fed into EA to create M new offspring by crossover and mutation operations, as described in section 4.1. The new offspring are used as the input to BPSO to compute M new particles as described in section 4.2. The new parent and new particles are combined and sorted in preparation for repeating the entire run.

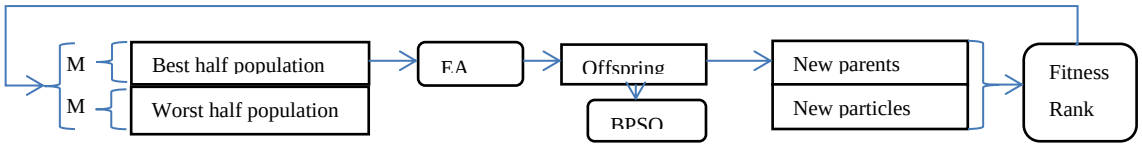


Figure 4. Overview of EABPSO procedure.

5. EXPERIMENTAL SETUP AND RESULTS

The experimental data set consists of 1038 medical images annotated by type including 306 diagrams, 329 graphs, 154 tables, and 249 flow charts, which are selected from the ten BioMedCentral journals mentioned in Section 1. 79 features are extracted from these images and optimal features selection step is applied. In the BPSO algorithm, the inertia weight w is empirically set to 0.8, the cognitive acceleration constant c_1, 1, and the social acceleration constant c_2, 1. In EA, the uniform crossover operator evaluates each bit in the parent strings for exchange with a probability of 0.5. The predefined mutation threshold is set as 0.8. EABPSO shares the same parameters with BPSO and EA. In addition, the dimension size (N) is the same as the number of features. For training the SVM, Platt's sequential minimal optimization algorithm [23] was implemented. It globally replaces all missing values and transforms nominal attributes into binary ones. It also normalizes all attributes and uses the polynomial kernel by default. A three-fold cross validation is used to set up the training and testing data sets. The data set is divided into three parts where 2/3rd is used for training and the rest is used for testing. This procedure is repeated three times. Therefore, for each time, the training set is 692 and the representative test set is 346 images. The accuracy of the classifier presented in the next section is based on averaging the accuracy of the three test sets.

Seven different schemes for feature subset selections are used. They are as follows: (i) **Case 1**: EA feature selection with uniform crossover operator; (ii) **Case 2**: BPSO feature selection; (iii) **Case 3**: EABPSO; (iv) **Case 4**: voting algorithm based on the selected feature set. The voting algorithm selects features based on the frequency of their occurrence in the three feature selection algorithms; (v) **Case 5**: Chi-square statistic [24]; (vi) **Case 6**: information gain [25] also used in order to compare classifier performance against EA, BPSO, EABPSO; and, finally, (vii) **Case 7**: uses all features as the input.

In Figure 5, the root mean square error (RMSE) performance measures (1-Accuracy) for cases 1, 2 and 3 are shown as the training progresses for one run. The population size (M) is 30 and the total training epoch (T) is 100. Table 2 shows the final accuracy of the SVM for feature subset from case 1 to case 7. We choose the particle size (M) to be 20 and 30, the total training epoch (T) to be 50, 100 and 150. Within the different combinations of the particle size (M) and the total training epoch (T), for case 1 to case 3, the best and averaged accuracy for the ten runs are listed in Table 2. The number of features in the subset is listed after the best accuracy. The accuracy of cases 4, 5 and 6 is also listed in Table 2 based on the feature combination that gives the best accuracy for case 1, 2 and 3.

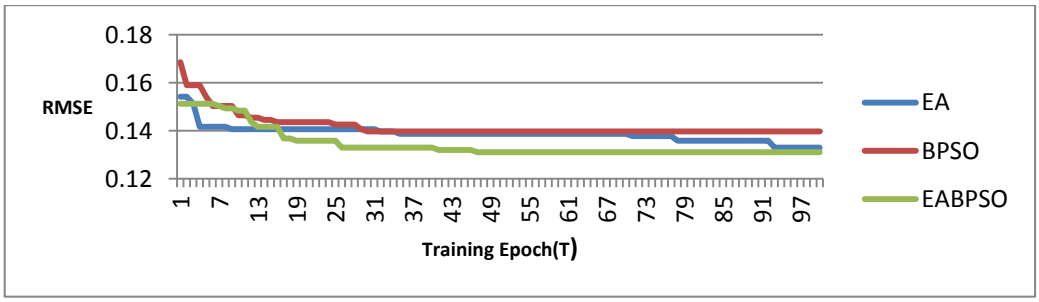


Figure 5. Root Mean Squared Error (RMSE) versus iteration number for each of the feature selection schemes.

As can be seen in Figure 5, for one hundred epochs, the accuracy ranking from high to low is 1) EABPSO, 2) EA, and 3) BPSO. BPSO stops converging at epoch 31; EA keeps converging until epoch 94; EABPSO achieves the global minimum at epoch 48. Figure 5 shows that BPSO algorithm gets trapped at the local minimum although it has a very fast convergence speed. EA does a good job to reach the lowest RMSE value but it takes a long time. Since EABPSO combines the evolutionary ideas of both BPSO and EA where BPSO helps to enhance the offspring created by EA in order to generate fitter feature combinations (elites) in each epoch, EABPSO has the ability to attain the best RMSE at a higher speed. Also, with N features, the computational complexity is on the order of O(N!) if every possible combination of features is explored. The PSO/EA and EABPSO reduce the complexity to O(N).

5.1 DISCUSSION

As seen in Table 2, the best overall classification accuracy of 87.5% was achieved with feature subset selected by BPSO (case 2) and EABPSO (case 3) for population size 20 and training epoch 100/150. For six different combinations of population size (M) and training epoch (T), EABPSO has the best accuracy. The average overall classification accuracy, 87.0% was achieved by both EA and EABPSO. This shows that EABPSO keeps delivering good and consistent results. Further, with the same particle size (M =30) as the epoch increases from 100 to 150 the average accuracy increases in EA (case 1) while staying the same in both BPSO (case 2) and EABPSO (case 3). This is due to longer convergence time for EA also shown in Figure 5. For six different combinations of population size (M) and training epoch (T), the number of features for the best subset obtained by EABPSO (case 3) stays around 45, while EA and BPSO have a larger range. Also, the advantage of using the voting algorithm for feature selection is

seen through the consistently good overall performance of case 4 since the voting algorithm's input features are the best selected features of cases 1, 2, 3. Justification for the usefulness of BPSO, EA and EABPSO is found through the observation that the highest classifier accuracies are achieved from cases 1-4.

Table 2. Performance comparisons for different feature combinations. (M = population size; T = maximum #iterations, Performance/XX: Here XX = the number of features used in computing the performance). Best Performance in **bold**.

Case No.	**Accuracy**	**M=20, T=50**	**M=20, T=100**	**M=20, T=150**	**M=30, T=50**	**M=30, T=100**	**M=30, T=150**
1. EA	(Average)	0.863	0.867	0.867	0.866	0.868	0.870
	(Best)	0.866/49	0.870/38	0.868/46	0.869/51	0.871/45	**0.872/49**
2.BPSO	(Average)	0.864	0.869	0.869	0.867	0.867	0.867
	(Best)	0.866/48	**0.875/51**	**0.875/51**	0.870/46	0.870/52	0.870/50
3.EABPSO	(Average)	0.870	0.870	0.870	0.868	0.869	0.869
	(Best)	**0.872/48**	**0.875/45**	**0.875/46**	**0.871/48**	**0.872/43**	**0.872/43**
4.Voting algorithm	Accuracy	0.870	0.872	0.872	0.869	0.870	0.870
5.Chi square	Accuracy	0.846	0.851	0.851	0.846	0.837	0.837
6.Information gain	Accuracy	0.846	0.850	0.850	0.846	0.836	0.836
7.All features	Accuracy	0.863					

6. CONCLUSIONS

This paper proposes a framework for graphical image type identification based on image feature analysis and computational intelligence techniques [26, 27]. Several feature extraction techniques are applied to the preprocessing of the images. Multiple features associated with the chart types are then extracted. EA and binary PSO are employed to find the optimal subset of features since both are stochastic search procedures and are generally suitable for solving the optimization problem. PSO has a higher convergence speed but easily trapped in local optimum while EA usually takes longer time to reach the global optimum although it has a mutation operator that can keep it out of local minimum. Thus EABPSO is proposed to combine the new individual generation functions of both EA and PSO, to attain the global minimum at high speed. The experimental results demonstrate that integration of various image processing techniques, feature extraction techniques, and computational intelligence methods for optimal feature selection as proposed in this paper can achieve high classification accuracy.

ACKNOWLEDGEMENTS

This work was supported by NLM under contract number 276200800413P and the Intramural Research Program of the National Institutes of Health (NIH), NLM, and Lister Hill National Center for Biomedical Communications (LHNCBC).

REFERENCES

1. M. Simpson, M.M. Rahman, S. Singhal, D. Demner-Fushman, S. Antani, G. Thoma. Text- and Content-based Approaches to Image Modality Detection and Retrieval for the ImageCLEF 2010 Medical Retrieval Track. CLEF 2010 LABs and Workshops Notebook Papers. ISBN 978-88-904810-2-4. ISSN 2038-4963. September 2010.

2. D. Meyer, F. Leisch, K. Hornik, The support vector machine under test. Neurocomputing, vol. 55, pp. 169-186, 2003.

3. Y. P. Zhou, C. L. Tan, Hough technique for bar charts detection and recognition in document images. International Conference on Image Processing, pp. 494-497, 2000.

4. W. Huang, Chart Image Classification Using Multiple-instance Learning. IEEE workshop on Applications of Computer Vision, pp. 27, 2007.

5. R. Liu, W. Huang , C. L. Tan, Extraction of Vectorized Graphical Information from Scientific Chart Images, Document Analysis and Recognitio, ICDAR, Ninth International Conference on, vol.1, pp.521-525, 2007.

6. R. C. González, R. E. Woods, Digital Image Processing, Pearson/Prentice Hall, 2008.

7. R.M.Haralick, K. Shanmugan, I. Dinstein, Textural Features for Image Classification, IEEE Transactions on Systems, Man, and Cybernetics, vol. SMC-3, pp. 610-621, 1973.

8. B. Cheng, R. J. Stanley, S. Antani, G. Thoma, A Novel Computational Intelligence based Approach for Medical Image Artifacts Detection, Proceedings of the 2010 International Conference on Artificial Intelligence and Pattern Recognition. Orlando, FL, pp.113-120, 2010.

9. B. Cheng, S. Antani, R. J. Stanley, Automatic segmentation of subfigure image panels for multimodal biomedical document retrieval, Proceedings of the SPIE Electronic Imaging, vol. 7874, pp. Z-1-11, 2011.

10. R.J. Stanley, J. Keller, P. Gader, C.W. Caldwell, Data-driven homologue matching for chromosome identification. IEEE Trans Med Imaging, vol. 17, no.3, pp, 451-462, 1998.

11. J. H. Holland, Adaptation in Natural and Artificial Systems. University of Michigan Press, Ann Arbor, 1975.

12. J. Kennedy, R. Eberhart, Particle swarm optimization. Proceedings of the IEEE International Conference on Neural Networks, Piscataway, NJ, pp. 1942-1948, 1995.

13. J. Song, F. Su, J. Chen, C. L. Tai, and S. Cai, Line net global vectorization: an algorithm and its performance analysis. IEEE Conference on Computer Vision and Pattern Recognition, South Carolina, pp. 383-388, 2000.

14. H.C. Peng, F. Long, C. Ding, Feature selection based on mutual information: criteria of max-dependency, max-relevance, and min-redundancy. IEEE Trans Pattern Anal and Machine Intel, 27(8), pp. 1226–1238, 2005.

15. M. Sezgin, B. Sankur, Survey over image thresholding techniques and quantitative performance evaluation, Journal of Electronic Imaging, vol. 13, no. 1, pp. 146-165, 2003.

16. I. Sobel, G. Feldman, A 3x3 Isotropic Gradient Operator for Image Processing. Presented at a talk at the Stanford Artificial Project in 1968, unpublished but often cited, orig. in Pattern Classification and Scene Analysis, Duda R and Hart P, John Wiley and Sons, pp. 271-272, 1973.

17. Soille, P., Morphological Image Analysis: Principles and Applications, Springer-Verlag, 1999, pp. 173-174.

18. D. C. Hanselman, B. L. Littlefield, Mastering MATLAB 7. Prentice Hall, 2004.

19. J. Piper, Granum E. On fully automatic feature measurement for banded chromosome classification. Cytometry, vol.10, pp.242-255, 1989.

20. G. Sywerda, Uniform crossover in genetic algorithms. Proceedings of the third international conference on genetic algorithms, pp.2-9, 1989.

21. Y. Kao, E.Zahara, A hybrid genetic algorithm and particle swarm optimization for multimodal functions, Elsevier, Applied Soft Computing, vol: 8(2), pp. 849-857, 2008.

22. C. Juang, A hybrid of genetic algorithm and particle swarm optimization for recurrent network design, Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on , 34(2), pp. 997- 1006, 2004.

23. J. Platt, Machines using Sequential Minimal Optimization. In B. Schoelkopf and C. Burges and A. Smola, editors, Advances in Kernel Methods - Support Vector Learning, 1998.

24. H. Liu, and R. Setiono, Chi2: Feature selection and discretization of numeric attributes, Proc. IEEE 7th International Conference on Tools with Artificial Intelligence, pp.338-391, 1995.

25. T.M. Mitchell, Machine Learning, the Mc-Graw-Hill Companies, Inc, 1997.

26. B. Cheng, R. J. Stanley, T. Szalapski, W. V. Stoecker, A Hybrid Computational Intelligence Algorithm for Automatic Skin Lesion Segmentation in Dermoscopy Images, Proceedings of Artificial Neural Networks in Engineering (ANNIE), pp. 379-386, 2010.

27. B. Cheng, R.J. Stanley, W.V. Stoecker, Automatic Telangiectasia Analysis in Dermoscopy Images Using Adaptive Critic Design, Skin Research and Technology, in Press, 2011.

VIII. AUTOMATIC SEGMENTATION OF SUBFIGURE IMAGE PANELS FOR MULTIMODAL BIOMEDICAL DOCUMENT RETRIEVAL

Beibei Cheng[a], Sameer Antani[b], R. Joe Stanley[a], George R. Thoma[b]

[a]Department of Electrical and Computer Engineering, Missouri University of Science and Technology, Rolla, MO 65409

[b]U.S. National Library of Medicine, National Institutes of Health, Bethesda, MD 20894

ABSTRACT

Biomedical images are often referenced for clinical decision support (CDS), educational purposes, and research. The task of automatically finding the images in a scientific article that are most useful for the purpose of determining relevance to a clinical situation is traditionally done using text and is quite challenging. We propose to improve this by associating image features from the entire image and from relevant regions of interest with biomedical concepts described in the figure caption or discussion in the article. However, images used in scientific article figures are often composed of multiple panels where each sub-figure (panel) is referenced in the caption using alphanumeric labels, e.g. Figure 1(a), 2(c), etc. It is necessary to separate individual panels from a multi-panel figure as a first step toward automatic annotation of images.

In this work we present methods that add make robust our previous efforts reported here. Specifically, we address the limitation in segmenting figures that do not exhibit explicit inter-panel boundaries, e.g. illustrations, graphs, and charts. We present a novel hybrid clustering algorithm based on particle swarm optimization (PSO) with fuzzy logic controller (FLC) to locate related figure components in such images.

Results from our evaluation are very promising with 93.64% panel detection accuracy for regular (non-illustration) figure images and 92.1% accuracy for illustration images. A computational complexity analysis also shows that PSO is an optimal approach with relatively low computation time. The accuracy of separating these two type images is 98.11% and is achieved using decision tree.

Keywords: Biomedical image analysis, biomedical article retrieval, content-based image retrieval, figure panels, image segmentation

1. INTRODUCTION

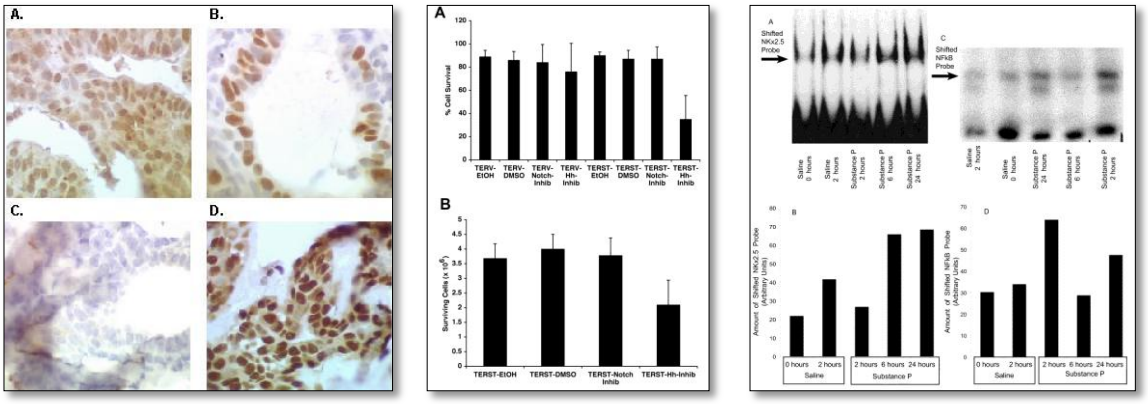
Biomedical images are frequently used in publications to illustrate the medical concepts or to highlight special cases. They are invaluable in establishing diagnosis, acquiring technical skills, and implementing best practices in many areas of medicine. Conventional approaches for biomedical journal article retrieval have been text-based with little attention devoted to the use of images in the articles. Text-based retrieval uses text information automatically extracted from title, abstract, figure captions, and discussions (mention). It provides fairly good results; however, the relevance quality sometimes is not satisfactory. Content-based image retrieval (CBIR) also has been applied to biomedical image retrieval [1]. However, the retrieval performance is far behind the text-based retrieval due to several gaps [2]. Low level features such as color, textual, and shape used in CBIR are insufficient to represent medical concepts or meaningful diagnostic information in the images effectively.

To improve the relevance quality of conventional retrieval approaches, we have proposed an approach using hybrid (text and image) features [3-8]. Information retrieval (IR) techniques are used to identify key textual features in the title, abstract, figure caption, and figure citation (mention) in the article. Structured vocabularies, such as the National Library of Medicine's Unified Medical Language System (UMLS ®) are used as well to identify the biomedical concepts in these [3,7]. Unlike conventional CBIR that uses image features from the entire image, our proposed approach uses a combination of features computed over the entire image and those computed from specific image regions of interest (ROIs). In computing image features, however, it is necessary that the image should be unimodal and individual. Often, in biomedical articles author put related images from different modalities (CT, MR, or Ultrasound, for example) as different subfigures in a single figure, or put slices from a CT or MR study as subfigures. For CBIR to be effective it is necessary to separate individual subfigure panels.

This article presents our efforts to improve our prior work [4, 5] on subfigure localization and segmentation toward improved biomedical article retrieval. In [4], the feasibility of automatically classifying images by usefulness (utility) in finding evidence was explored using supervised machine learning and achieved 84.3% accuracy using image captions for modality and 76.6% accuracy combining captions and image data for utility from articles over 2 years from a clinical journal. However, the figures images in this study had to be manually segmented into individual panels. In this work we present methods that add make robust our previous efforts [5] that, though successful, were limited in their scope and were unable to meet the challenges of segmenting figure illustrations, graphs, and charts. For the

latter, we present a novel hybrid clustering algorithms which are particle swarm optimization (PSO) with fuzzy logic controller (FLC) method to locate related figure components. Results from preliminary evaluation are very promising with 93.64% accuracy for regular (non-illustration) figure images and 92.1% accuracy for illustration images. The correctness of separating these two type images are 98.11% by using decision tree. More intensive tests are in progress to evaluate impact of automatic figure panel segmentation and use of ROI in image annotation and retrieval.

This article is organized as follows. Section 2 describes prior work on the topic. Section 3 describes the methods. Section 4 presents the experiments and we conclude with Section 5.



(a) (b) (c)

Figure 1. Examples of different types of multi-panel images. (a) Regular image. (b) Graph/chart image. (c) Mixed image.

2. PRIOR WORK

In our prior work a heuristic two-phase algorithm was developed and applied for detection and decomposition of multi-panel images. Information from figure caption was used to obtain an estimate of number of panels [5]. The algorithm looked for strong white or black lines or a sharp transition between image panels. If any panels were found then the image was segmented along identified boundaries and recursively applied to segmented panels until no further segmentations are found. Detection and decomposition of multi-panel images was tested on 516 figure images extracted from 2 years (2004 – 2005) issues of the British Journal of Oral and Maxillofacial Surgery. In this set, 427 images were single panel images and 89 were multi-panel. Overall 409 or 95.78% of the single panels and 84 or 94.38% of the multi-panel images were correctly identified. In case of multi-panel images, 6 of 84 had been correctly identified as multi-panel images having a disagreement with the caption analysis. This disagreement was usually minor (±1 panel). These images are deemed as correct detection of a multi-panel image for purposes of that evaluation. Overall result combining these scores was 95.54% detection and decomposition accuracy. The method typically failed on cases where (i) inter-panel boundary width assumption exceeded our thresholds or (ii) there was a lack of a sharp transition between panels. A further limitation of the method is that it was heuristic and could not adapt to variations in the figure layouts.

3. METHODS

For this research, we use 2111 medical images annotated by modality (radiological, photo, etc.) selected from 6 BioMedCentral journals (Cancer, AFPS, Urology, Surgery and Cardiovascular Ultrasound, Dermatology). The set comprises 1052 regular images (as shown in Figure 1(a)) and 1059 illustration images (as shown in Figure 1(b)). An overview of the algorithm investigated is shown in Figure 2.

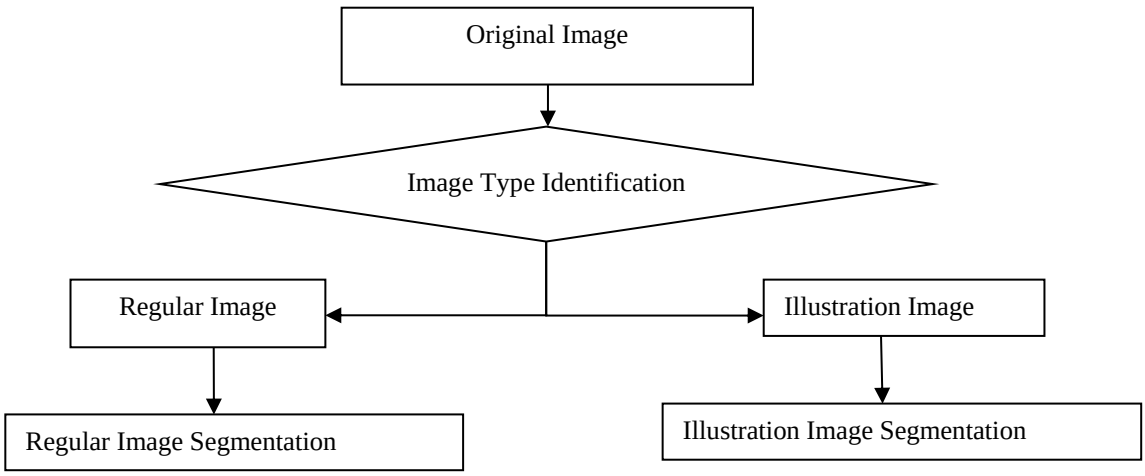


Figure 2. Overview of sub-image segmentation process.

3.1 IMAGE TYPE IDENTIFICATION

As different methods are used to detect panels on multi-panel regular or illustration image, it is necessary to distinguish its type first. We generate 11 features for this purpose. These characteristic features with their explanations, information gain [9] values and chi-square [10] values are shown in Table 1. Several discrimination methods were explored and we found that decision trees provided the highest accuracy of 98.11%.

Table 1. Information gain and chi-square value for image features useful in detecting image type.

Features	Inform-gain value	Chi-square value
Mean of red color value	0.859	1872.868
Mean of green color value	0.929	1999.198
Mean of blue color value	0.924	1985.9172
Standard deviation of red color value	0.344	859.0623
Standard deviation of green color value	0.33	833.523
Standard deviation of blue color value	0.311	779.8279
Mean of skewness of red color value	0.666	1557.3166
Mean of skewness of green color value	0.804	1782.0291
Mean of skewness of blue color value	0.803	1776.4037
Mean of number of coherent pixels	0.53	1219.7187
Mean of number of incoherent pixels	0.515	1211.3774

3.2 SEGMENTING MULTI-PANEL REGULAR IMAGES

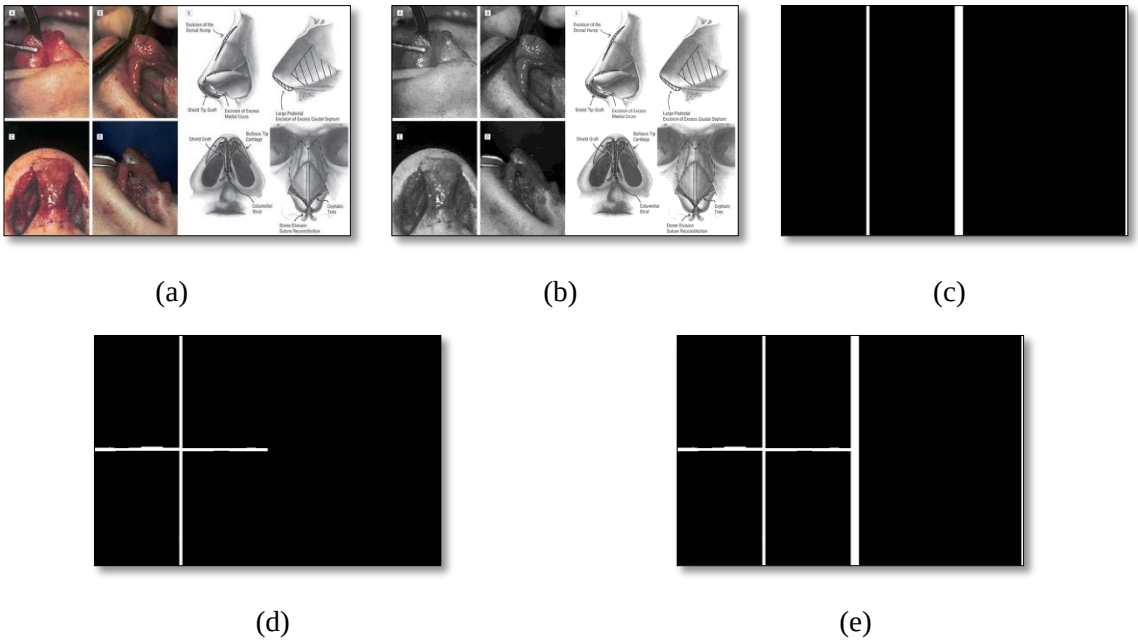
A two-phase algorithm is developed and applied for decomposition of regular images. Such images tend to have strong inter-panel boundaries. First phase of the method detected these, as follows:

1. Convert RGB images into gray images.
2. Calculate the variance of each vertical line and horizontal line across the image. The line with low variance is marked as the boundary. Call this image_1
3. Get the boundary edge with gray horizontal and vertical dynamic range between the minimum and maximum gray drop values of 10 and 25. A boundary edge is defined as a set of pixels whose horizontal neighbor difference is less than the minimum gray drop value while its vertical gray drop value is greater than the maximum gray drop value, or vice versa.
4. A logical OR image outputs from steps 2) and 3). Call this image_2

Figure 3 presents an image example of the image processing steps for the original image to generate the boundary output. From image_1 and image_2, we compute the number of subfigure panels and assign them to variables: panel_1 and panel_2, respectively. In addition, an estimate of number of panels is obtained using Natural Language Processing (NLP) techniques from the figure caption [4], named panel_3. Next, four features are generated, which are

(i)panel_3 /panel_1, (ii)panel_3 /panel_2, (iii) standard deviation of the subpanels size from image_1, and (iv)the standard deviation of the subpanels size from image_2.

The second phase is using above features as the input to train and test the neural network. The performance is computed by comparing the values of panel_1, panel_2, and panel_3 with a manually determined real_panel_number. If real_panel_number=panel_1, then the output is set to 1; else if real_panel_number=panel_2, then the output is set to 2; else if real_panel_number= panel_3, then the output is set to 3; else, it is set to 4.



(a) (b) (c)

(d) (e)

Figure 3. Image processing algorithm example. (a) Original image. (b) Gray image. (c) Output by using low variance line finding technique. (d) Output by using gray drop technique. (e) Output by Or ((c), (d)).

3.3 SEGMENTING MULTI-PANEL ILLUSTRATION IMAGES

Unlike regular images illustration images do not have a clear inter-panel boundary, but the separation is obvious to the human observer. This makes segmenting the panels more difficult. Some cues regarding illustration placement and size were used to develop an algorithm for segmenting the drawn groups. An overview of the algorithm investigated is shown in Figure 4. Each flowchart component is discussed in turn.

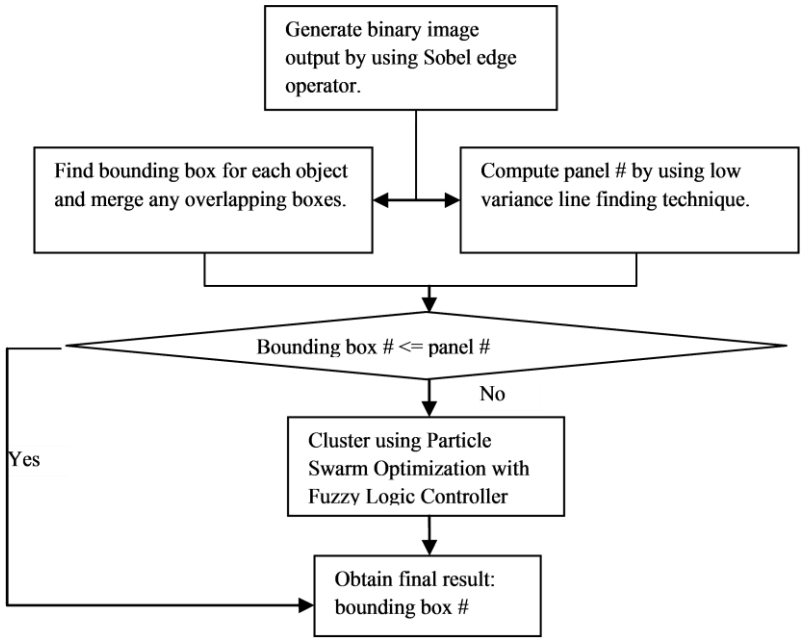


Figure 4. Overview of multi-panel images decomposition process.

Sobel edge operator: Edges in the image are computed in the vertical and horizontal direction using the Sobel edge detector [11] with kernel values shown in Equations (1) and (2).

$$dx = \begin{matrix} -0.05 & 0 & 0.05 \\ -0.1 & 0 & 0.1 \\ -0.05 & 0 & 0.05 \end{matrix} \quad (1)$$

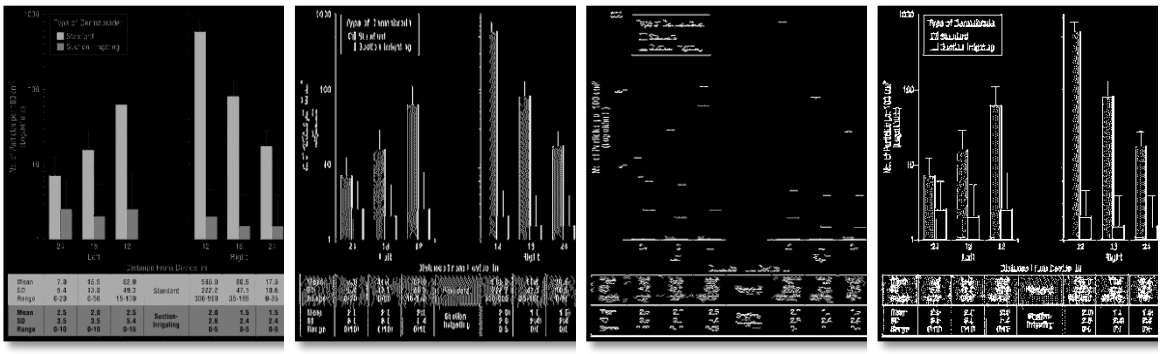
$$dy = dx' \quad (2)$$

The input image is convoluted with dx and dy and then Otsu's filter [12] is applied to the output. The outputs are logically OR-ed to generate the image for use in the next step. Example images from this step are shown in Figure 5.

Form bounding boxes: The image output from above step is processed to fill holes inside each foreground object. Small noise objects are removed by morphological processing. Bounding box can be drawn around each object and any overlapping bounding boxes are merged. Small bounding boxes may be removed. Example images from this step are shown in Figure 6

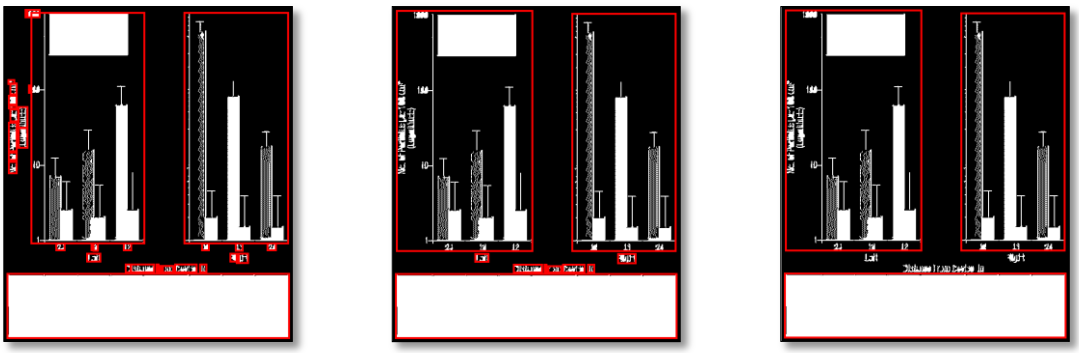
Particle Swarm Optimization (PSO) with Fuzzy Logic Controller (FLC) clustering: As shown in Figure 7, some images may not result in bounding boxes surrounding each subfigure. To overcome this difficulty, we compare the number of bounding boxes with the

number of subpanel identified in the text caption using low variance line finding technique. If the number of bounding boxes is less equal to the number of subpanels then it is deemed correct. However if it is greater than the number of subpanels, an over-segmentation may be assumed and the particle swarm optimization (PSO) [13] algorithm is used to correct any error. Using shape, size, and position constraints, we devise a fitness function for the PSO algorithm (shown in Figure 8) where subfigure objects of similar shape and size that are near each other are considered as belonging to one region.



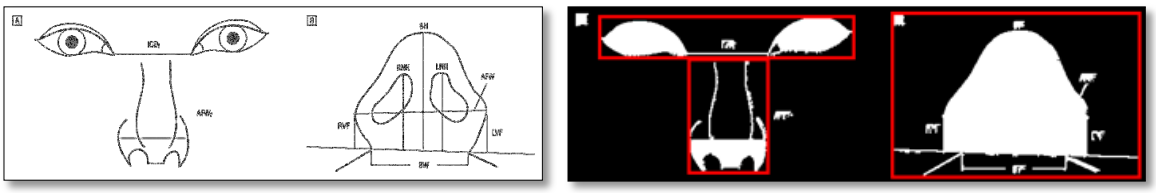
(a) (b) (c) (d)

Figure 5. Sobel edge operator processing. (a) Original image. (b) Image output after convoluting the input with ***dx*** and using Ostu's method. (c) Image output after convoluting the input with ***dy*** and using Ostu's method. (d) Shows the image after logically OR-ingimages in (b) and (c).



(a) (b) (c)

Figure 6. Form bounding boxes. (a) Bounding box for each object. (b) Bounding box with overlap removed. (c) Bounding box after removing small noise bounding boxes.



(a) (b)

Figure 7. Incorrect bounding box segmentation. (a) Original image. (b) Bounding box result.

We select five features for computing the fitness function of size and fourteen features for shape. These characteristic features and their explanations are shown as follows, the first five are size features and others are shape features:

Width: The width of bounding box.

Height: The height of bounding box.

MajorAxisLength: length (in pixels) of the major axis of the ellipse that has the same normalized second central moments as the region.

MinorAxislength: length (in pixels) of the minor axis of the ellipse that has the same normalized second central moments as the region.

Axis Ratio: ratio of MajorAxisLength with MinorAxislength.

Solidity: area of the region divided by the area within the Convex Hull.

Extent: ratio of area to bounding box area.

Weighted Density Distribution features: Twelve features are extracted by correlating the shape samples of the object inside the bounding box with weighted density distribution functions (WDD) [14, 15]. The samples are computed by dividing the height of the bounding box by 20. This value is empirically determined from prior work [15]. The length of the section of the object at each of these 20 horizontal segments is used as a shape sample for convolution and computation of the WDD features.

For objects, each particle in the PSO is an n-dimensional vector where the dynamic range of each dimension n and objects with same value are assigned to the same cluster. The velocity is constrained to 2. The fitness function of size and shape are defined as follows:

$$fitness_{size} = var(fitness_{size})\ \ fitness_{shape} = var(fitness_{shape}) \quad (3)$$

Since both fitness values have their individual effect on the result, we build a fuzzy logic controller (FLC) system to get the final fitness value with these two as the inputs.

$$fitness = FLC(fitness_{size}, fitness_{shape}) \quad (4)$$

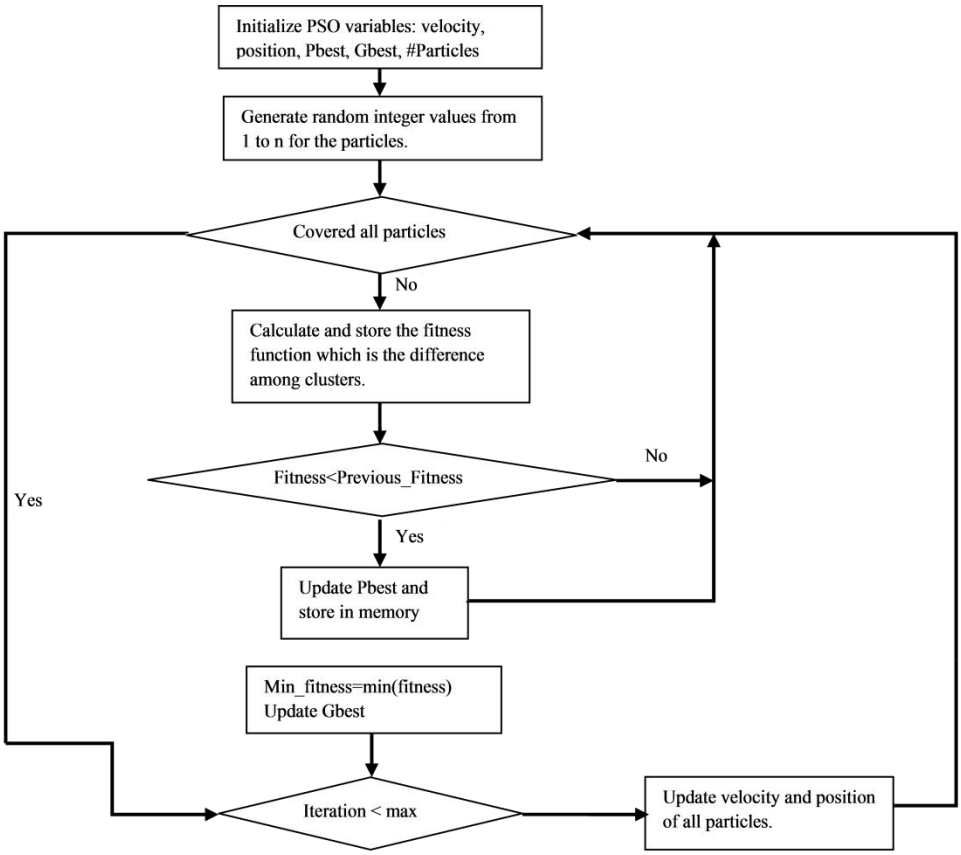


Figure 8. Steps for applying Particle Swarm Optimization algorithm.

Fuzzy logic is a form of multi-value logic derived from fuzzy set theory to deal with reasoning that is approximate rather than precise [16]. Fuzzy control is a control method based on fuzzy logic [17]. The fuzzy logic controller consists of three main modules described below: fuzzification, the inference engine, and defuzzification.

Fuzzification: The input membership functions take the inputs to the controller and produce a degree of membership for each fuzzy set. This value is usually designated by the symbol μ. Figure 9 shows the membership function takes as inputs the fitness value and assigns to that a degree of membership for each fuzzy set in the graph, where "s" represents "small", "m" represents "medium", "l" is large.

Inference engine: Once the degrees of membership for each fuzzy set have been determined for a particular input, they are presented to the inference engine that determines which rules should be evaluated. Table 2 gives the fuzzy rule result.

Defuzzification: Once the degrees of membership of the outputs have been found via the inference engine, the defuzzification process takes these values and translates them into an output dispatch signal. If multiple rules have been asserted, the center of mass of the weighted outputs is used. The output membership function is shown in Figure 9.

Figure 10 shows sample results of PSO clustering with different fitness functions.

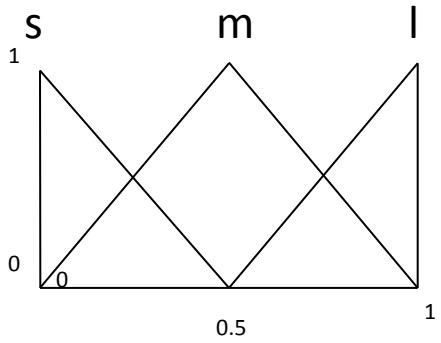
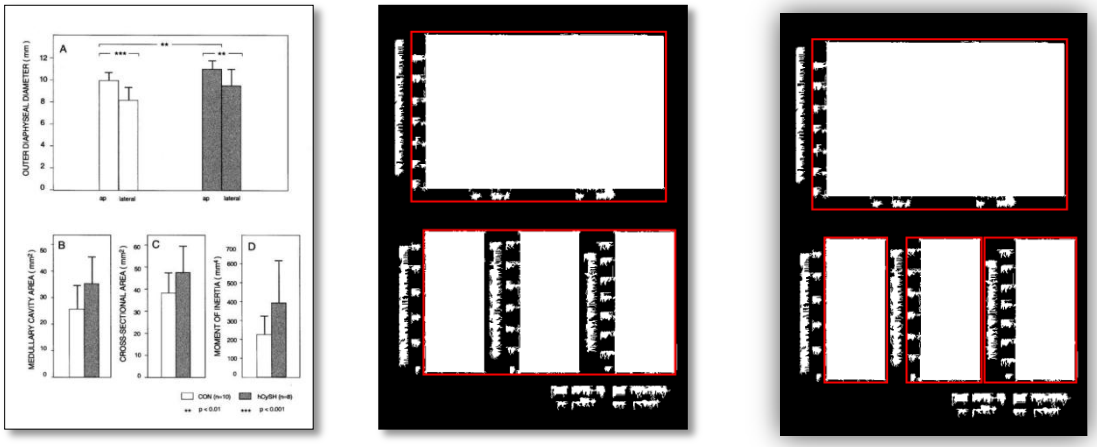


Figure 9. Membership Function.

Table 2. Fuzzy rule.

	$fitness_{size}$	$fitness_{shape}$	$fitness_{total}$
1	s	s	s
2	s	m	s
3	s	l	s
4	m	s	s
5	l	s	s
6	m	m	m
7	m	l	l
8	l	m	l
9	l	l	l



(a) (b) (c)

Figure 10. Sample results of PSO clustering with different fitness functions. (a) Original image. (b) $Fitness = fitness_{size} + fitness_{shape}$. (c) $Fitness = FLC\ (fitness_{size}, fitness_{shape})$.

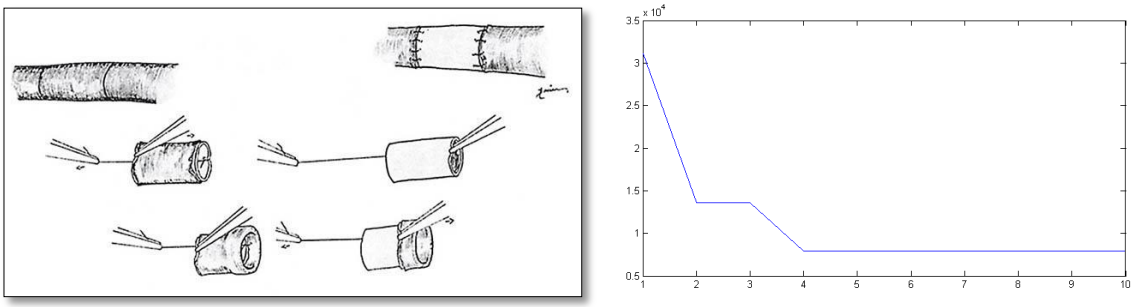
4. EXPERIMENTS

Table 3 presents the true positive rate and true negative rate of image type identification for 2111 images by using three different algorithms with ten-fold cross validation [21]. Decision tree provides the highest accuracy.

Table 3. Result of image type identification.

Algorithm	**True Positive Rate**	**True Negative Rate**
MLP Backpropgation Neural Network [18]	96.26%	97.17%
Support Vector Machine [19]	96.20%	96.13%
Decision Tree [20]	97.62%	98.58%

The method for separating subfigure panels in regular images is evaluated by generating the ROC curve [22] using a neural network with the architecture 5x5x1 (5 neurons including four features plus one bias in the input layer, 5 neurons in the hidden layer and one output neuron) and ten-fold cross validation. For illustration images we measure the performance by comparing the estimated number of panels with the manually generated ground truth. And we set the number of particles as 30 and number of iterations as 10. For Figure 11(a), the fitness value at every iteration is shown in Figure 11(b). Note that the fitness value reaches the optimal value at the fourth iteration. We obtain very promising results with 93.64% accuracy for regular (non-illustration) figure images and 92.1% accuracy for illustration images. Compared with our previous effort, this approach to locate subfigure panels is more robust. The fitness function of PSO includes not only the size but also the shape of each object, and is consequently able to identify similar shaped objects of different sizes. In addition, with n subpanels, the computational complexity will be O(n!) if we try every possible combination for subpanels. The PSO helps to reduce the complexity to be O(n).



(a) (b)

Figure 11. Input image and its fitness function. (a) Input image. (b)Fitness value function.

5. CONCLUSIONS

Figure image segmentation is an important and necessary first step in annotating images for improved information retrieval for clinical decision support. This step helps subsequent image annotation and CBIR methods perform optimally. For accurate sub-figure image segmentation we first need to detect the image type. Regular images usually provide a strong inter-panel boundary which is used to detect the sub-figure panels. Finding subfigure panels in illustration images is more challenging. For success we use a combination of top-down and bottom-up approaches. The low variance line finding technique is like segmenting image in a top-down method, while, finding foreground objects and bounding boxes is akin to segmenting images in a bottom-up fashion. Steps from this method will be used as a initial step in indexing and retrieval of images and articles for from PubMedCentral at the National Library of Medicine in its iMedline project.

ACKNOWLEDGEMENTS

This research was supported by the Intramural Research Program of the Lister Hill National Center for Biomedical Communications, an R&D division of the National Library of Medicine, at the National Institutes of Health, U.S. Department of Health and Human Services. We thank Dr. Dina Demner-Fushnanm, Dr. Matthew Simpson, and Dr. Zhiyun Xue of the National Library of Medicine for their contributions in this effort.

REFERENCES

1. Antani S, Long R, Thoma GR. Content-Based Image Retrieval for Large Biomedical Image Archives. Medinfo,San Francisco, CA, 2004, pp.829-33.

2. Deserno TM, Antani S, Long R. Ontology of gaps in content-based image retrieval. Journal of Digital Imaging, 2009, vol. 22, no.2, pp.202-15.

3. Demner-Fushman D, Antani S, Simpson M, Thoma GR. Annotation and Retrieval of Clinically Relevant Images. International Journal of Medical Informatics: Special Issue on Mining of Clinical and Biomedical Text and Data, 2009, vol.78, no.12, e59-e67.

4. Demner-Fushman D, Antani SK, Thoma GR. Automatically Finding Images for Clinical Decision Support. Proceedings of Workshop on Data Mining in Medicine, 7th IEEE Intl Conf on Data Mining, Omaha, NE, 2007, pp.139-44.

5. Antani S, Demner-Fushman D, Li J, Srinivasan BV, Thoma GR. Exploring use of images in clinical articles for decision support in Evidence-Based Medicine. Proc. SPIE-IS&T Electronic Imaging, San Jose, CA, 2008, 6815:68150Q(1-10).

6. Demner-Fushman D, Antani SK, Simpson M, Thoma GR. Combining Medical Domain Ontological Knowledge and Low-level Image Features for Multimedia Indexing. Proc. 2nd International "Language Resources for Content-Based Image Retrieval" Workshop (OntoImage 2008), part of 6th Language Resources and Evaluation Conference (LREC 2008), 2008, CDROM Proceedings.

7. You D, Antani SK, Demner-Fushman D,Rahman MM, Govindaraju V, Thoma GR. Biomedical article retrieval using multimodal features and image annotations in region-based CBIR. Proc. SPIE-IS&T Electronic Imaging, San Jose, CA, 2010, 7534:75340V(1-12).

8. You D, Apostolova E, Antani SK, Demner-Fushman D, Thoma GR. Figure content analysis for improved biomedical article retrieval. Proc. SPIE-IS&T Electronic Imaging, San Jose, CA, 2009, 7247:72470V(1-10).

9. Mitchell TM. Machine Learning, the Mc-Graw-Hill Companies, Inc, 1997.

10. Mood A, Franklin AG, Duane CB. Introduction to the Theory of Statistics, McGraw-Hill Companies, Inc, 1974.

11. Sobel I, Feldman G. A 3x3 Isotropic Gradient Operator for Image Processing. Presented at a talk at the Stanford Artificial Project in 1968, unpublished but often cited, orig. in Pattern Classfication and Scene Analysis, Duda R and Hart P, John Wiley and Sons, '73, pp. 271-272.

12. Sezgin M, Sankur B.Survey Over Image Thresholding Techniques and Quantitative Performance Evaluation. Journal of Electronic Imaging, 2003, vol. 13, no. 1, pp. 146-165.

13. Kennedy J, Eberhart R. Particle Swarm Optimization. Proceedings of the IEEE International Conference on Neural Networks, Piscataway, NJ, 1995, pp. 1942-1948.

14. Piper J, Granum E. On Fully Automatic Feature Measurement for Banded Chromosome Classification. Cytometry, 1989, vol. 10, no. 3, pp. 242-255.

15. Stanley RJ, Stoecker WV, Moss RH, Rabinovitz HS, Cognetta AB, Argenziano G, and Soyer HP. A Basis Function Feature-based Approach for Skin Lesion Discrimination in Dermatology Dermoscopy Images. Skin Research and Technology, 2008, vol. 14, no. 4, pp. 425-435.

16. Von Altrock C. Fuzzy logic and Neuro Fuzzy applications explained. Upper Saddle River, NJ: Prentice Hall PTR, 1995.

17. Driankov D, Hellendoorn H, Reinfrank,M. An introduction to fuzzy control, second edn, Springer-Verlag, Berlin, 1996.

18. Haykin S. Neural Networks: A Comprehensive Foundation, Prentice Hall, 1998.

19. Meyer D, Leisch F, Hornik k. The support vector machine under test, Neurocomputing, 2003, vol. 55, pp. 169-186.

20. Breiman L, Friedman JH, Olshen RA, Stone CJ. Classification and regression trees, Monterey, CA: Wadsworth & Brooks/Cole Advanced Books & Software, 1984.

21. Fogarty J, Baker RS, and Hudson SE. Case Studies in the Use of ROC Curve Analysis for Sensor-based Estimates in Human Computer Interaction. Proceedings of Graphics interface, Canadian Human-Computer Communications Society, School of Computer Science, University of Waterloo, Waterloo, Ontario Victoria, British Columbia, 2005, vol. 112, pp. 129-136.

22. Kohavi R. A Study of Cross-validation and Bootstrap for Accuracy Estimation and Model selection. Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence, 1995, vol.14, pp. 1137–1143.

IX. A HYBRID COMPUTATIONAL INTELLIGENCE ALGORITHM FOR AUTOMATIC SKIN LESION SEGMENTATION IN DERMOSCOPY IMAGES

Beibei Cheng, R. Joe Stanley, William V, Stoecker, Thomas Szalapski,
Ganesh K. Venayagamoorthy, Hanzheng Wang

ABSTRACT

In this paper, an unsupervised approach based on evolving vector quantization (EVQ) is presented for enhancing dermatology images for skin lesion segmentation. Vector quantization (VQ) as a famous compression technique has been widely used in image signal compression and speech signal compression. The EVQ algorithm extends the Linde, Buzo, and Gray (LBG) vector quantization method with particle swarm optimization to cluster the pixels inside the image based on merging similar gray value pixels. The proposed enhancement technique is evaluated using 100 dermoscopy skin lesion images for skin lesion segmentation. The EVQ algorithm is applied to the individual color planes red, green, and blue, respectively. Segmentation results using these three planes are compared and scored based on manual borders obtained from three dermatologists. In addition, differential equation-based particle swarm optimization is implemented and their results are compared with the standard PSO.

1. INTRODUCTION

Malignant melanoma is the most deadly form of skin cancer (Jemal et al, 2009). The incidence of melanoma cases, especially melanoma in situ, is increasing rapidly (Jemal et al, 2009). This adds to the importance of further developing early detection methods such as those based on digital image analysis. Automatic lesion segmentation is a fundamental step in computer-based image analysis of pigmented skin lesions. Numerous methods have been proposed for lesion segmentation in dermoscopy images including: 1) a partial differential equations (PDE)-based system (Chung and Sapiro, 2000); 2) an independent histogram pursuit (IHP) algorithm (Gomez et al,2008); and a watershed-based algorithm (Wang et al, 2010). In this research, a novel evolving vector quantization (EVQ) is presented for image enhancement for lesion segmentation. The EVQ algorithm extends the the Linde, Buzo, and Gray (LBG) vector quantization method with particle swarm optimization (PSO).

LBG vector quantization (Linde et al, 1980), is an iterative algorithm which alternatively solves the two optimality criteria: 1) nearest neighbor condition, the encoding region should consists of the vectors that are closer than any of the other codevectors, and: 2) centroid condition, the codevector should be the average of all the training vectors that are in the encoding region. The basic steps of the LBG algorithm are:

1. Determine the number of codewords, N, or the size of the codebook.
2. Select N codewords at random, and let that be the initial codebook. The initial codewords can be randomly chosen from the set of input vectors.
3. Using the Euclidean distance measure clusterize the vectors around each codeword.
4. Compute the new set of codewords.
5. Repeat steps 2 and 3 until the either the codewords don't change or the change in the codewords is small.

PSO is a form of evolutionary computation technique developed by Kennedy and Eberhart (1995). In the swarm intelligence algorithm, each particle has random velocity and memory that keeps track of the previous best position and corresponding fitness. The previous best value of the particle position is called the ‘pbest’. It has another value called ‘gbest’, which is the best value of all the ‘pbest’ positions in the swarm. The basic concept of PSO is that each particle in the swarm move toward its pbest and gbest locations at each time step. In this research, 100 skin lesion images from dermoscopy research are used as the benchmark. An overview of this project investigated is shown in Figure 1. The remaining sections of this paper

include: 1) pre-processing, 2) Evolving Vector Quantization methodology, 3) post-processing, 4) results and discussion, and 5) conclusions and future work.

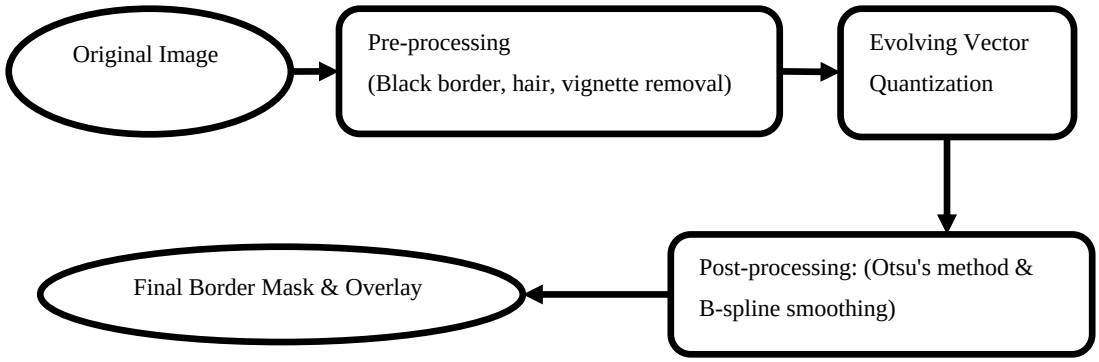
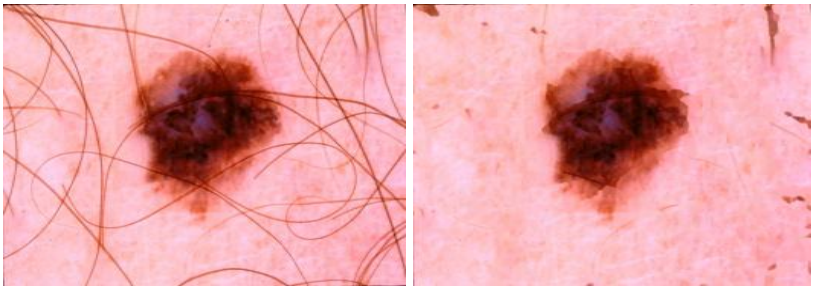


Figure 1. Overview of lesion segmentation process.

2. PREPROCESSING

A few noise sources such as hair, black border and vignetting affect lesion segmentation. The preprocessing approach investigated is as follows: Firstly, a hair removal method with the morphological closing operator is introduced to remove hairs if they exist. An example is given in Figure 2. Secondly, the vignetting removal method is detected in the red plane (because it is least affected by the lesion) and normalized for each of the other two planes. Thirdly, the black border is cropped from the innermost point of each black rim (Wang et al, 2010). Also, a compensation step that involves resizing the image to256x256 is applied before filtering noise.



(a) (b)

Figure 2. Preprocessing example. (a) Original image. (b) Image after hair removal.

3. EVOLVING VECTOR QUANTIZATION

For the VQ of dimension M and size N, each particle in PSO is a (MxN) dimensional vector, representing a possible codebook. The range of each dimension is the dynamic range of the pixel value, which is 256. The velocity is constrained to 25 (10% of the pixel value range). The fitness function is defined as:

$$fitness = \sum_{m=1}^{256} \sum_{n=1}^{256} abs(sig(m,n) - re\_sig(m,n)) \quad (1)$$

where sig(m,n) represents the input image and re_sig(m,n) represents the reconstructed image after quantization. The PSO procedures for VQ are summarized in Figure 3. During each iteration, the particles fly in the solution space to search for better fitness and they could converge to an optimal solution representing an optimal VQ codebook. After each iteration, the input image becomes the reconstructed image which has the best fitness value. Two different PSO version tried in this work are described in Section 3.1 and 3.2.

3.1 STANDARD PSO (Valle et al,2008)

The following equations are used to update particle velocity and position:

$$V_{i,new} = w \cdot V_{i,new} + c_1 \cdot rand1 \cdot \left(pbest_i\_sig_{i,old}\right) + c_2 \cdot rand2 \cdot \left(gbest\_sig_{i,old}\right) \quad (2)$$

$$sig_{i,new} = sig_{i,old} + V_{i,new} \quad (3)$$

where $1 \leq i \leq$ particle number, $rand1$ and $rand2$ are the random values between 0 and 1. In this research, w = 0.1, c_1 = 1 and c_2 = 1. There are four cases examined. In case 1, population size is 20, iteration number is 10; in case 2, population size is 20, iteration number is 5; in case 3, population size is 10, iteration number is 10; in case 4, population size is 10, iteration number is 5. For example, Figure 4 gives the original image and output image of each iteration using standard PSO in case 3.

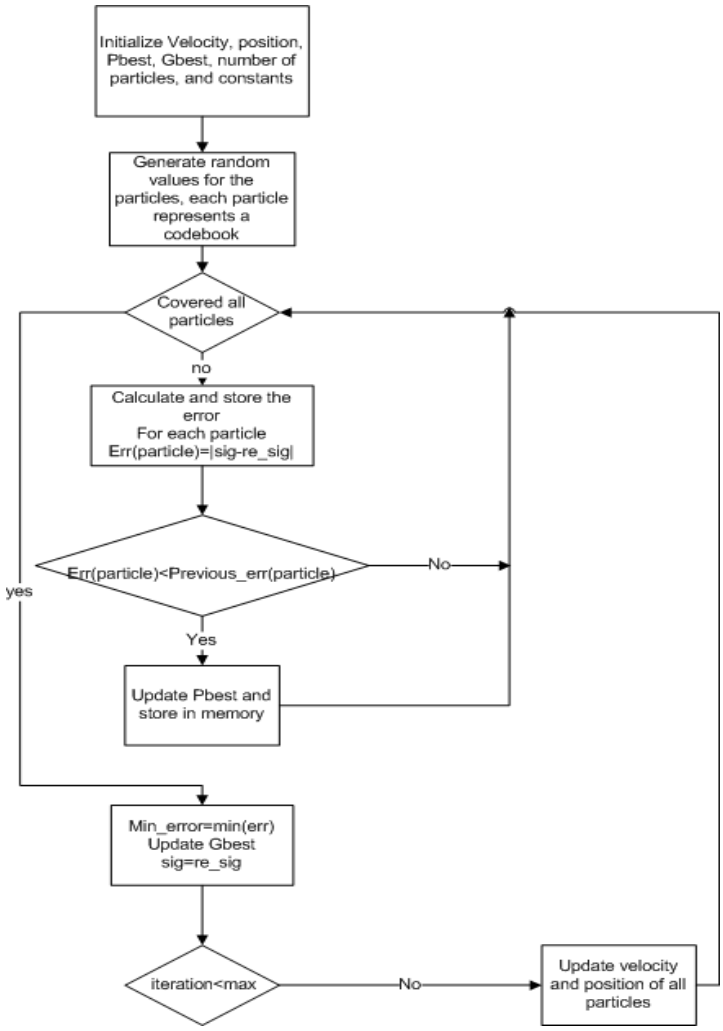


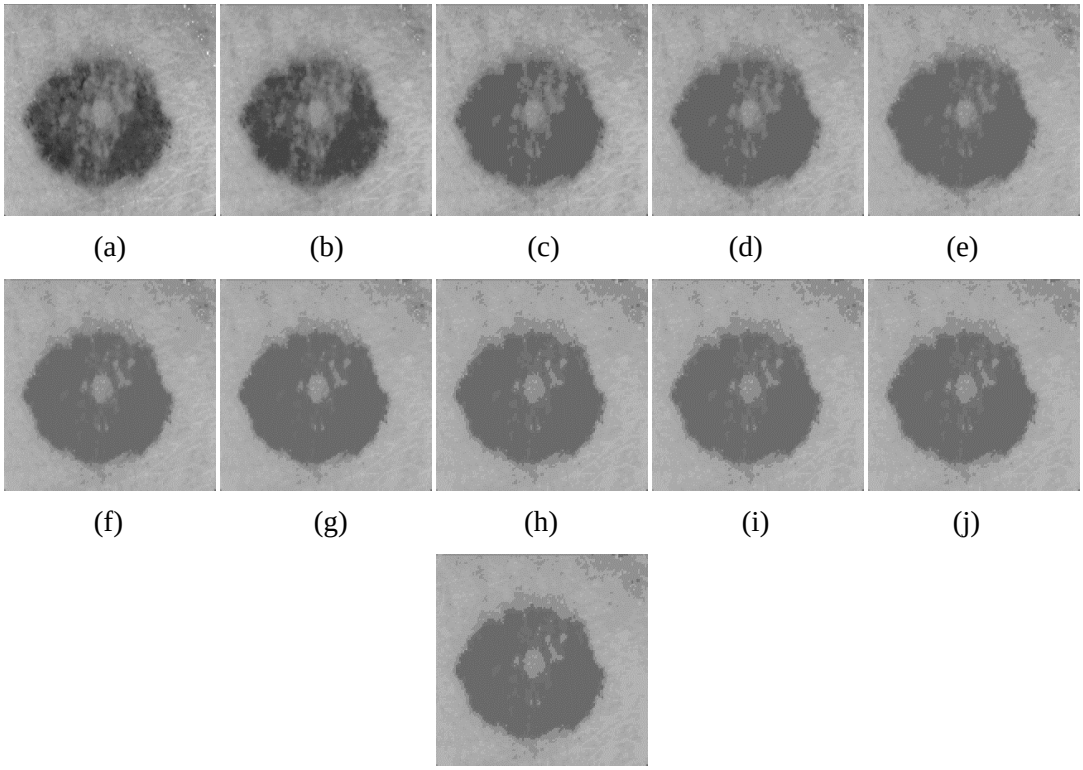
Figure 3. Evolving vector quantization algorithm flow chart.

3.2 DIFFERENTIAL EQUATION-BASED PSO (Zhang and Xie, 2003)

The PSO algorithm is used for determining the best codebook for representing the compression coefficients for the image. The differential equation-based PSO (Zhang and Xie, 2003) is investigated to see if an improvement in the compression coefficients can be obtained based on the fitness equation (1). In DEPSO, the differential equation operator can be applied to improve the pbest of the PSO particles based on randomly choosing four different pbest particles from the trained PSO algorithm. Let Δ_d be represented as follows. Then $\text{pbest}_{i,new}$ gives the updated row entry for the codebook, $\text{pbest}_{i,new}$ is given as follows.

$$\Delta_d = [(pbest_1 - pbest_2) + (pbest_3 - pbest_4)]/2 \quad (4)$$

$$\text{pbest}_{\text{i,new}} = \begin{cases} \text{pbest}_{\text{i,old}} + \Delta_d \; if \; fitness(\text{pbest}_{\text{i,old}}) \leq \text{fitness}(\text{pbest}_{\text{i,new}}) \\ \text{pbest}_{\text{i,old}} \quad if \; fitness(\text{pbest}_{\text{i,old}}) > \text{fitness}(\text{pbest}_{\text{i,new}}) \end{cases} \quad (5)$$



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

(f) (g) (h) (i) (j)

(k)

Figure 4. Example of image output within ten iterations**.** (a) Original image. (b) ~ (k) output image within 1~10 iteration.

The same four cases as section 3.1 are examined as well. Upon completing the PSO and DEPSO algorithms for performing EVQ, the codebook is determined for the input image, yielding the transformed or enhanced image.

3.3 POSTPROCESSING

Otsu's method (Sezgin and Sankur, 2003) is implemented with the enhanced image using the algorithm from the previous steps, which chooses the threshold to minimize the intra-

class variance of the black and white pixels. This converts the gray images into binary images. In order to mitigate missing lesion pixels, a relaxed threshold was used, which is the first gray value higher than the value from Otsu's method. Then, the second-order B-Spline closed curve fitting (Wang et al, 2010) is applied to the border of the EVQ segmentation. The final border overlay will then be generated (Figure 5).

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

Figure.5. Post-processing. (a) original image. (b) output image after EVQ. (c) output image after Otsu's method. (d) image after B-Spline smoothing. (e) image overlay: dermatologist border (blue) EVQ (green).

4. RESULTS AND DISCUSSION

The presented enhancement and threshold methods for lesion segmentation were applied to 100 dermoscopy skin lesion images. These images have been used in previous studies (Wang et al, 2010). The segmented lesions are evaluated based on comparison to an average border obtained from borders manually determined by three dermatologists. The error of the segmented lesion from the average dermatologist border is found as follows. Let A denote the automatically segmented lesion and xor represent the exclusive-OR operation. The percentage border error E is given by $E = (A \text{ xor } M_{lesion})/M_{lesion} \cdot 100$, where M_{lesion} is the average dermatologist border. The average error over 100 lesions is presented in Table 1 for the PSO and the DEPSO for the four different cases described in section 3.

Table 1. Standard PSO and DEPSO error result with three planes.

		Blue	Green	Red
Standard PSO	Case1	28.71	33.39	49.36
	Case2	21.87	20.36	40.85
	Case3	20.37	36.88	39.98
	Case4	19.84	25.04	39.50
DEPSO	Case1	25.93	28.96	45.86
	Case2	24.69	20.27	37.69
	Case3	19.66	31.59	40.64
	Case4	18.67	24.68	38.49

This table shows that blue plane always gives the best result, confirming results of Celebi et.al (2007). For standard PSO, case 4 in blue plane has the lowest error value within the smallest population size and the iteration. DEPSO generally has slightly better output than standard PSO.

5. CONCLUSIONS AND FUTURE WORK

The EVQ algorithm can be considered as a clustering algorithm which extends the Linde, Buzo, and Gray (LBG) Vector Quantization method with Particle Swarm Optimization in order to cluster the pixels inside the image based on merging similar gray value pixels. After that, the enhanced image could be easily segmented.

This project uses Otsu's method to finish the image segmentation. In future work, some more intelligent image segmentation methodologies will be developed within EVQ as the preprocessing technique. Furthermore, supervised EVQ could be explored to improve the fitness function.

REFERENCES

1. Chung, D. and Sapiro G., 2000, "Segmenting skin lesions with partial-differential-equations-based image processing algorithms", Medical Imaging, IEEE Transactions on, vol. 19(7), pp. 763-767.

2. Celebi, E.M., Aslandogan, Y.A., Stoecker, W.V., Iyatomi, H., Oka, H., and Chen, X., 2007, "Unsupervised border detection in dermoscopy images", Skin Res Technol. vol. 13(4), pp. 565-462.

3. Gomez, D.D., Butakoff, C., Ersboll, B.K., and Stoecker, W., 2008, "Independent Histogram Pursuit for Segmentation of Skin Lesions", vol. 55(1), pp. 157-161.

4. Jemal, A., Siegel, R., Ward, E., Hao, Y., Xu, T.J., and Thun, M. J., 2009, "Cancer Statistics 2009", CA: A Cancer Journal for Clinicians, vol. 59(1), pp. 225-249.

5. Kennedy, J., and Eberhart, R., 1995, "Particle swarm optimization", Proceedings of the IEEE International Conference on Neural Networks, Piscataway, NJ, pp. 1942-1948.

6. Linde, Y., Buzo, A., and Gray, R.M., 1980, "An algorithm for vector quantizer design," IEEE Transactions on Communications, vol. 28(1), pp. 84–95.

7. Sezgin, M., and Sankur, B., 2003, "Survey over image thresholding techniques and quantitative performance evaluation", Journal of Electronic Imaging, , vol. 13(1), pp. 146-165.

8. Valle, Y., Venayagamoorthy, G.K., Mohagheghi, S., Hernandez, J.C., Harley, R.G., 2008, Particle Swarm Optimization: Basic Concepts, Variants and Applications in Power Systems", IEEE Transactions on Evolutionary Computation, vol. 12(2), pp. 171 – 195.

9. Wang, H., Chen, X., Moss, R.H., Stanley, R.J., Stoecker, W.V., Celebi, M. E., Szalapski, T. M., Malters, J. M., Grichnik, J. M., Marghoob, A. A., Rabinovitz, H. S., Menzies, S.W., 2010, "Watershed segmentation of dermoscopy images using a watershed technique", online pre-publication, Skin Research Technology.

10. Zhang, W. and Xie, X., 2003, "DEPSO: hybrid particle swarm with differential evolution operator", Systems, Man and Cybernetics, IEEE International Conference on, vol. 4, pp. 3816-3821.

X. A NOVEL COMPUTATIONAL INTELLIGENCE-BASED APPROACH FOR MEDICAL IMAGE ARTIFACTS DETECTION

Beibei Cheng[a], Renzhong Wang[a], Sameer Antani[b], R. Joe Stanley[a], George R. Thoma[b]

[a]Department of Electrical and Computer Engineering, Missouri University of Science and Technology, Rolla, Missouri

[b]U.S. National Library of Medicine, National Institutes of Health, Bethesda, Maryland

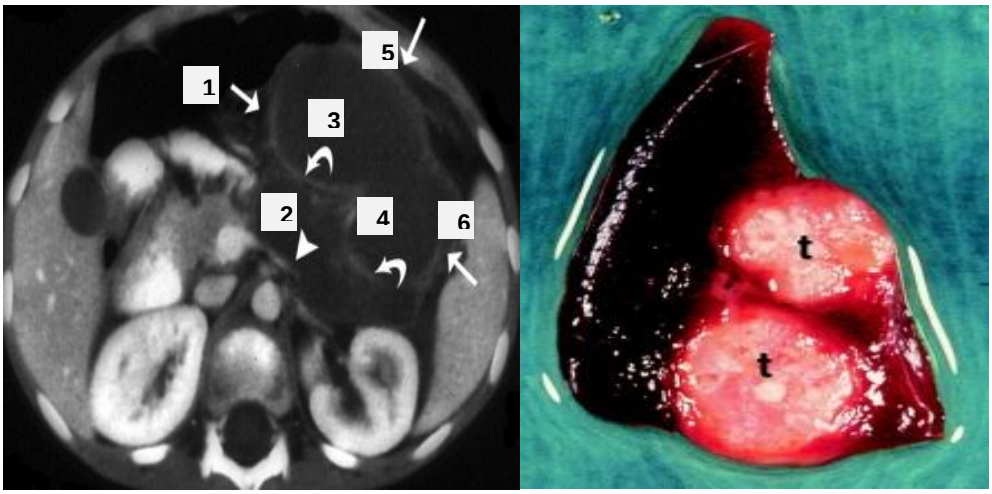
ABSTRACT

In this research, a novel computational intelligence-based algorithm to detect artifacts, specifically arrows, in medical images is presented. Image analyses techniques are developed to find the symbols and text automatically. Features are computed from the shape of arrow for the discrimination of arrows from other artifacts. We investigate a biologically-inspired reinforcement learning (RL) approach in an adaptive critic design (ACD) framework to apply Action Dependent Heuristic Dynamic Programming (ADHDP) for arrow discrimination based on the computed features. Experimental results for ADHDP are compared with feed forward multi-layer perception (MLP) back-propagation artificial neural networks (BP-ANN), particle swarm optimization (PSO) for training of a MLP neural network, genetic algorithm (GA) for training of a MLP neural network, k-nearest neighbor (KNN), and support vector machine (SVM).

1. INTRODUCTION

The detection of medical image artifacts such as arrows is important to highlighting supplemental and context-based information which may be helpful in understanding medical images. It is necessary to have an accurate algorithm in discriminating arrows from other characters and symbols. There are several methodologies that have been implemented to find arrows in previous research. Laurent Wendling and Salvatore Tabbone [1] proposed a method in recognizing arrows based on the aggregation of geometric criteria using the choquet interal; J.E. den Hartogtz and T.K. ten Katet [2] gave a solution of finding arrows in utility maps using a neural network; Jongan Park, Waqas Rasheed, and Junguk Beak [3] proposed a way of identifying arrow signs for Robot Navigation using a camera-based method.

Extending techniques from previous research, the arrow symbols present in the medical images in our experimental data set have variety in shape and size. Arrows do not necessarily have to be straight (arrow 3, arrow 4) and the shape of the arrows can change (arrow 2) as you can see in Figure 1(a). Furthermore, a noise example can include characters and symbols, which may be of similar size to arrows as shown in Figure 1 (b). Therefore, a general and robust arrow detection algorithm is needed for discrimination from other medical image artifacts.



(a) (b)

Figure 1. Medical image examples. (a) Arrow image example. (b) Noise image example.

Since both text and symbol objects are white or black, they can be segmented by some image analysis techniques. After generating the binary image containing only text-like and symbol-like objects, features sets are used as input to classifiers. An overview of the algorithm investigated is shown in Figure 2. This study uses 144 medical images annotated by modality (radiological, photo, etc.) selected from 2004-2005 issues of the British Journal of Oral and Maxillofacial Surgery, including 79 images with one or more arrows and 65 images with no arrows. The image analysis techniques are follows:

1. Convert RGB images into gray images and inverted gray images.
2. Use Otsu's method [4], which chooses the threshold to minimize the intra-class variance of the black and white pixels to convert gray images into binary images.
3. Remove objects that are considered small and objects that are considered short.
4. Get the edge of object with gray drop. The value of gray drop is 30. If the absolute value of center pixel minus NW, N, NE, W, E, SW, S, SE is greater than the gray drop, this pixel will be marked. (Figure 3)
5. Compare the edge image getting after step 4) with the image getting after step 3), keep the objects with the same bounding box size.
6. Or image result with inverted image result.

Figure 4 presents an image example of the image processing steps for the original image to generate the binary mask for feature calculations.

The remaining sections of this paper include: 1) Feature Generation, 2) Adaptive Critic Designs Methodology, 3) Other Classification Methodologies, 4) Results and Discussion, 5) Acknowledgement, and 6) Conclusions and Future Work.

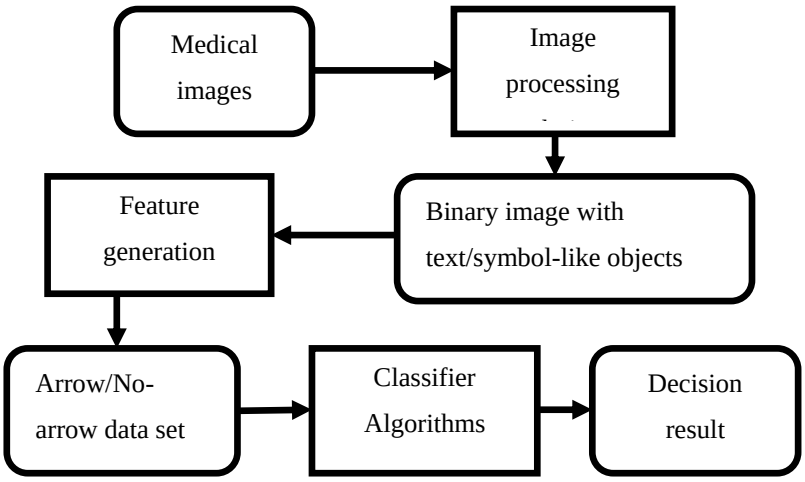


Figure 2. Overview of arrow detection and discrimination process.

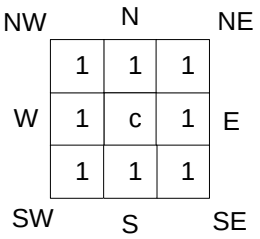


Figure 3. Edge detection.

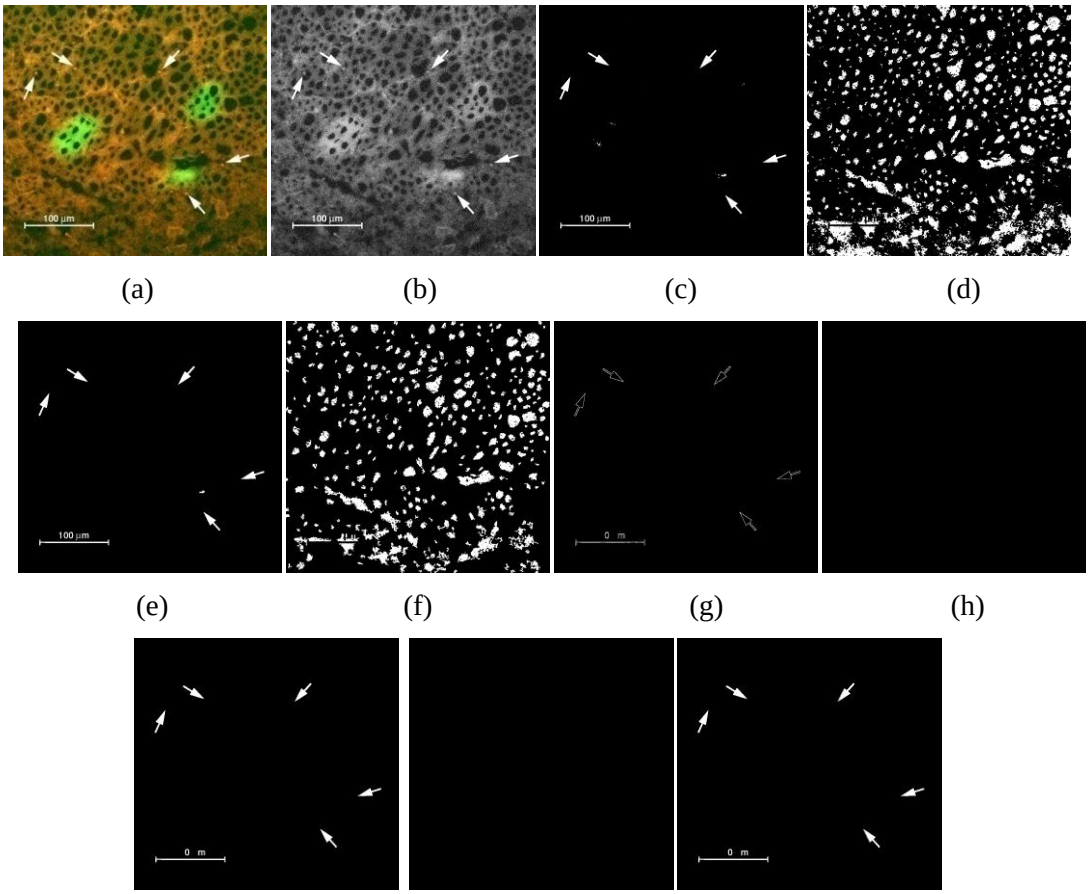


Figure 4. Image processing algorithm example. (a) Original image. (b) Grey image. (c) Image by using threshold. (d) Inverted image by using threshold. (e) Image after noise removal. (f) Inverted image after noise removal. (g) Image by using edge detection. (h) Invert image by using edge detection. (i) Image by comparing (e) to (g) with the bounding box size. (j) Image by comparing (f) to (h) with the bounding box size. (k) Final image by or (i) and (j).

2. FEATURE GENERATION

After generating the binary image containing only text-like and symbol-like objects by using above techniques, certain features are selected that will help in the identification of arrows from the rest of objects. This selection process is very important, because the identification stage depends heavily upon this process. We select twenty-two features that will help us in distinguishing the arrows. These characteristic features and their explanations are shown as follows:

MajorAxisLength: length (in pixels) of the major axis of the ellipse that has the same normalized second central moments as the region.

MinorAxislength: length (in pixels) of the minor axis of the ellipse that has the same normalized second central moments as the region.

Axis ratio: ratio of MajorAxisLength with MinorAxislength.

Normalized area: area of the region divided by the whole image.

Solidity: area of the region divided by the Convexhull Area.

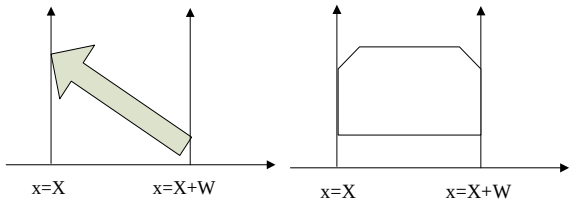
EulerNumber: equal to the number of objects in the region minus the number of holes in those objects.

EquiDiam: the diameter of a circle with the same area as the region.

Extent: ratio of area to bounding box area.

AvgSkelDist: average width of object.

MinPixelNo: for two line x=X and x=X+W (X is the top left point of bounding box horizontal value; W is the width of bounding box), the minimum number of pixels of intersection for region and each line as shown in Figure 5.



(a) (b)

Figure 5. MinPixelNo feature**.** (a)Arrow. (b) Noise.

Weighted Density Distribution features: other arrow features are extracted by correlating the shape samples of the arrow with weighted density distribution functions (WDD) [5-6]. Let $E = \{E(s_1), E(s_2), \dots, E(s_m)\}$ be the sequential of shape samples collected, where m is the number of samples collected at a constant rate and $E(s_i) \in R_n$, where $1 \leq i \leq m$. Figure 6 shows the WDD functions used in the experiments. Twelve WDD-based features are computed. Each of the WDD function is decomposed into 12 discrete points for WDD feature calculations. Let W_1 denote the WDD function in Figure 6(a), W_2 denote the WDD function in Figure 6(b) and so on. The gray horizontal position marker for each WDD function shown in Figure 6 points to the reference position for which WDD features is computed. The reference position corresponds to the current sample s. Six WDD features ($f_1(s), \dots, f_6(s)$) corresponding to the measurement at sample s are computed according to the following expression:

$$f_k(s) = \sum_{i=s-10}^{s+1} E(i) W_k\,(i - s + 11) \tag{1}$$

For $k = 1,2, \dots,6$. Six additional features ($f_7(s), \dots, f_{12}(s)$) are computed by correlating the six WDD functions with the sequence of absolute differences between samples value as follows

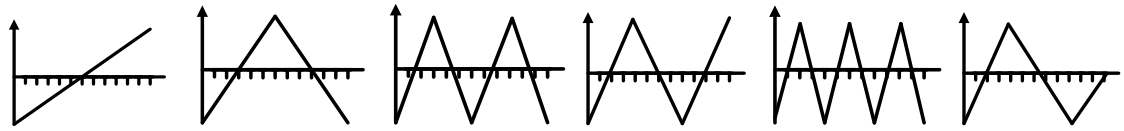
$$f_k(s) = \sum_{i=s-10}^{s+1} |E(i) - E(i-1)| W_k\,(i - s + 11) \tag{2}$$

In this research, the shape samples of arrow are computed in the following way: divide the height of the bounding box of arrow by twenty and get line $y=Y_1$ $y=Y_2 \dots$, $y=Y_{20}$. The 20 samples is the length of line $y=Y_1$ to $y=Y_{20}$ intersected with this arrow, as you can see in Figure 7, the red lines. After all, twelve WDD features could be generated according to these samples.

Therefore, with the 144 medical images as the input, after image processing and feature extraction, there are 154 arrow objects and 276 text/noise objects generated. They are manually catalogued with class 1 (arrow objects) and class 0 (other objects). To evaluate these features, the attribute selection criteria are information gain [7] and chi-square [8] by Weka®. Table 1 shows information gain value and chi-square value for each feature.

The orientation of each object can impact feature generation. For example, the feature MinPixelNo is supposed to be smaller for arrow objects than for others because of the accurate point of arrow. But since the various shapes and sizes of arrows, as can be seen in Figure 8, this value can still be quite large. This arrow is falsely identified as noise by the MLP backpropgation neural network.

To improve the outcome, rotating the region is an effective solution. Figure 9 shows the arrow image output with different rotation values (0 degree, 15 degree, 30 degree, 45 degree, 60 degree, 75 degree, 90 degree). Therefore, seven data sets will be generated after applying feature extraction to these seven image sets.



(a) (b) (c) (d) (e) (f)

Figure 6. The WDD functions used to compute arrow features [6].

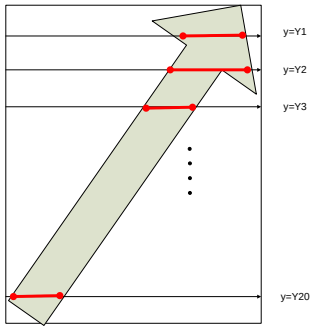


Figure 7. Samples for generating WDD features.

Table 1. Information gain and chi-square value.

Feature_Name	Info_gain_value	Chi_square_value
MajorAxisLength	0.5875	287.0769
MinorAxislength	0.1924	106.3261
Axis Ratio	0.4380	217.0534
Normalized area	0.5052	248.0459
Solidity	0.1038	58.169
eulerNumber	0.1587	73.3162
EquiDiam	0.5628	278.9818
Extent	0.2260	112.2445
AvgSkelDist	0.4566	230.6459
MinPixelNo	0.1555	75.0847
f1	0.1837	109.0502
f2	0.2504	141.9929
f3	0.2768	155.7666
f4	0.1658	95.7217
f5	0.0457	27.9648
f6	0.4004	213.7316
f7	0.3361	187.9629
f8	0.1205	71.3723
f9	0.1350	81.8844
f10	0.0538	32.9075
f11	0.1837	25.5232
f12	0.2768	159.0519

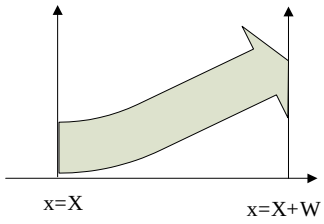


Figure 8. Falsely identified arrow.

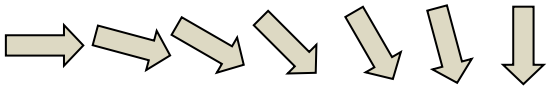


Figure 9. Arrows with different rotation value.

An ACD-based arrow discrimination methodology is then applied in the presence of these seven data sets. So the decision of arrow/no-arrow is made by the global data set but not the single data set. The input for the neural network is the features generated from those seven different orientations based on rotation.

3. ADAPTIVE CRITIC DESIGN METHODOLOGY

Reinforcement learning is the problem faced by an agent that must learn behavior through trial-and-error interactions with a dynamic environment. It is a computational approach to learning whereby an agent tries to maximize the total amount of reward it receives when interacting with a complex, uncertain environment [9]. The RL has been developed in various applications such as neuro-computing [10], and multi-resolution object recognition [11].

The adaptive critic design provides a workstation for implementing RL. An ACD approximates the neuro-dynamic programming by using an action and a critic network, respectively [12]. This model employs reinforcement learning (RL) through direct neural dynamic programming (Direct NDP) [13]. The term "direct" is influenced by the adaptive control literature where "direct adaptive control" means no plant model, and thus no plant parameter estimation takes place but instead certain plant information is used directly to find appropriate and convergent control laws and control parameters, which is required in this research. Direct NDP is a model independent approach to action dependent heuristic programming (ADHDP).

Figure 10 shows the model of ADHDP used in this study, which is based on the model in [12]. In the current problem setting, let the discounted total reward to go R(t) at time t be given by

$$R(t) = r(t+1) + \alpha r(t+1) + \cdots = \sum_{k=1}^{\infty} \alpha^{k-1} r(t+k) \tag{3}$$

Where the function of r(t) is the reinforcement value at time t, and α is a discount factor between 0 and 1.

The critic network is used to provide an output J(t), which is an approximation for R(t), the weighted total future reward to go. The reward function R(t) at time t is given by Eq. (3).

We define the prediction error, and consequently the Bellman error, for the critic element as:

$$e_c(t) = \alpha J(t) - [J(t-1) - r(t)] \tag{4}$$

and the objective function to be minimized in the critic network is

$$E_c(t) = \frac{1}{2} {e_c}^2(t) \tag{5}$$

In the action network, the weight update in the action network can be formulated as follows.

$$e_a(t) = J(t) - R^* \tag{6}$$

The principle in adapting the action network is to backpropagate the error between the desired ultimate performance objective, denoted by R*, and the approximate function J from the critic network. Since $r_s$has been defined as the reinforcement signal for "success," R* is set to $r_s/(1-\alpha)$ has in the direct NDP design paradigm and in subsequent case studies. In this paper, r_s is set to be zero for simplification.

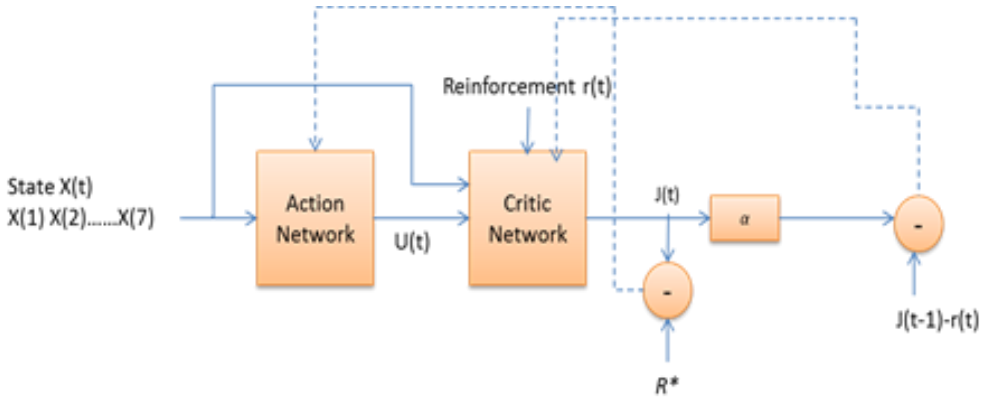


Figure 10. Schematic diagram of ADHDP.

An artificial neural network is chosen for implementation of the action and critic networks. The structure of the neural networks for both the action and critic networks are implemented as a multi-layer feed forward (MLP) neural network. It consists of the input layer, the hidden layer and the output layer. The hidden layer neurons have a sigmoid transfer function while other layers have linear neurons. For the action network, the architecture is 23x5x1, with twenty-two features and a bias in the input layer, five nodes in the hidden layer and one output layer. For the critic network, the architecture is 24x5x1, with twenty-two features, a bias and the

output from action network in the input layer, five nodes in the hidden layer and one output layer.

A ten-fold cross validation methodology is used for training/test set generation for the neural network [14]. The neural networks are trained up to 1000 epochs, using online (Stochastic/Delta) learning. In this case, the next input pattern is selected randomly from the training set, to prevent any bias that may occur due to the sequences in which patterns occur in the training set. For each training feature, 7 different data states (original image feature data set and its six different rotated orientation feature data sets) are applied as the input one by one for both the action network and the critic network to update the weights. If the difference between action network output u(t) and the target is less than 0.5, the reinforcement signal r(t) takes the reward "0", otherwise, r(t) takes the punishment "-1". The learning rates for both critic and action network are set to be 0.001. The discount factor α is set to be 0.1. The test set is the original image feature data set only.

With the target value for the arrow data set to 1 and the no-arrow data set to 0, action network outputs after testing are between -1 and 1. Receiver operating characteristic (ROC) curves are generated for classification results based on the neural network outputs obtained for the ten-fold cross cases [15]. The ROC curve is a plot of the sensitivity for a binary classifier system as its discrimination threshold is varied. The ROC curve represents equivalently the fraction of true positives versus false negatives rate.

4. OTHER CLASSIFICATION ALGORITHMS

4.1 MLP BACKPROPGATION NEURAL NETWORK

A multilayer perception backpropagation neural network is investigated for arrow discrimination [16]. Sigmoid transfer functions are used in the hidden layers, and a linear transfer function is used in the input and output layer, the neural network architecture is 23x5x1. The neural networks are trained up to 1000 epochs, using online (Stochastic/Delta) learning, ROC curves are generated by the ten-fold cross strategy.

4.2. PARTICLE SWARM OPTIMIZATION (PSO) FOR TRAINING OF A MLP NEURAL NETWORK

In swarm intelligence algorithm [17], each particle has random velocity and memory that keeps track of previous best position and corresponding fitness. The previous best value of the particle position is called the ‘pbest’. It has another value called ‘gbest’, which is the best value of all the ‘pbest’ positions in the swarm. The basic concept of PSO is that each particle in the swarm move toward its pbest and gbest locations at each time step. The basic concept of PSO is that each particle in the swarm move toward its pbest and gbest locations at each time step. (Figure 11)

In this research, PSO is used to train a MLP neural network, which has the same architectures and training epochs as the last one. To train this neural network, all neuron weights are together to contribute one of the particles. Each particle is updated toward the global best position, which will minimize the difference between the neural network output and target value. Detail of this algorithm is described in [18].

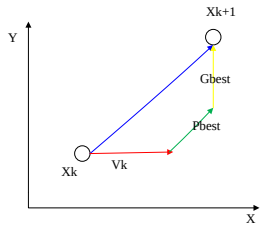


Figure 11. Basic concept of PSO.

4.3. GENETIC ALGORITHM (GA) FOR TRAINING OF A MLP NEURAL NETWORK

Genetic Algorithm is a kind of optimization by using selection, crossover, mutation and elitism operators [19]. This MLP neural network still has the same architectures and training epochs as usual. In the training procedure, all neuron weights are put together as the parents firstly, after applying the selection, crossover and mutation operators, offspring could be generated. The next offspring is chosen based on whether parent or its offspring minimizes the difference between the neural network output and target value.

5. RESULTS AND DISCUSSION

Adaptive Critic Designs is applied for the arrow classification. The output is compared with MLP backpropgation neural network, PSO for training of a MLP neural network, GA for training of a MLP neural network. Ten-fold cross testing strategy is implemented for all of them. ADHDP and other three neural networks are all build on the same twenty-two input features, use the same number of training epochs (1000). Since the output of the neural network is not an accurate class number, the ROC curve is generated as the comparison criteria. AUC represent the area under the ROC curve. Figure 12 shows the ROC curves and AUC result for these four algorithms.

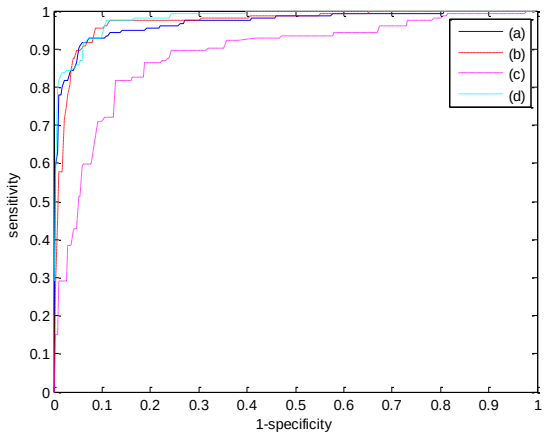


Figure 12. ROC curve and AUC (area under curve) for neural networks. (a) MLP backpropgation NN. AUC=0.9672. (b) PSO based NN. AUC=0.9681. (c) GA based NN. AUC=0.8839. (d) Direct NDP. AUC=0.9790.

In addition to comparing the discrimination results from the different classifiers using AUC, as given in Figure 12, we used a highest true positive rate and highest true negative rate with minimum difference between them from the ROC curves. Therefore, the true positive rate and true negative rate for (a), (b), (c), (d) are 91.53% and 91.56%, 91.53% and 92.77%, 87.66% and 82.97%, 92.86% and 92.03%.

Based on the AUC and the true positive and true negative comparison, we found that adaptive critic design achieved the best results among the different classifiers investigated. In

direct NDP, the critic network is used as the performance evaluation; the reward of reinforcement learning is helping to update its weight and the output J(t). Therefore, the action network could update its weight based on total seven state feature data instead of only one state feature for other techniques. The global feature data improves the accuracy of discriminating arrow or no-arrow.

Furthermore, K-nearest neighbor (KNN) [20] and Support Vector Machine (SVM) [21] are also applied to the same data set. In KNN algorithm, generally Euclidean distance function is used to calculate distance between two points. K is set to be fifteen. In SVM algorithm, polynomial is used as kernel function value to train the data. The true positive rate and true negative rate for KNN and SVM are 93.50% and 88.76%, 90.90% and 93.03%, respectively.

6. CONCLUSIONS AND FUTURE WORK

This paper introduces the ACD design to image recognition using ADHDP algorithm. The result is very promising. ADHDP demonstrates greater strength and superiority than existing methods. In addition, the features extracted as the input of different classifiers are approved to be significantly useful.

The future work will include finding the exact orientation of each arrow. In addition, other adaptive critic design algorithms such as action dependent dual heuristic dynamic programming (ADDHP) or action dependent globalized dual heuristic dynamic programming (ADGHP) could be implemented as comparison.

ACKNOWLEDGEMENTS

This work was supported by NLM under contract number 276200800413P and the Intramural Research Program of the National Institutes of Health (NIH), NLM, and Lister Hill National Center for Biomedical Communications (LHNCBC).

REFERENCES

1. L. Wendling, S. Tabbone, “Recognition of arrows in line drawings based on the aggregation of geometric criteria using the Choquet integral”, Document Analysis and Recognition, Proceedings of Seventh International Conference on, 2003, vol. 1, pp. 299-303.

2. J.E. den Hartogtz, T.K. ten Katet, “Finding arrows in utility maps using a neural network”, Proceedings of the 12th IAPR International Conference on Pattern Recognition, Conference B: Computer Vision & Image Processing, 1994, vol. 2, pp. 190-194.

3. J. Park, W. Rasheed, and J. Beak, “Robot Navigation Using Camera by Identifying Arrow Signs”, Grid and Pervasive Computing Workshops, GPC Workshops '08, The 3rd International Conference on Grid and Pervasive Computing, Kunming, May 2008, pp. 382-386.

4. M. Sezgin, B. Sankur, "Survey over image thresholding techniques and quantitative performance evaluation", Journal of Electronic Imaging, 2003, vol. 13, no. 1, pp. 146-165.

5. J. Piper, E. Granum, “On fully automatic feature measurement for banded chromosome classification”, Cytometry, 1989, vol. 10, no. 3, pp. 242-255.

6. R.J. Stanley, W.V. Stoecker, R.H. Moss, H.S. Rabinovitz, A.B. Cognetta, G. Argenziano, and H.P. Soyer, “A basis function feature-based approach for skin lesion discrimination in dermatology dermoscopy images”, Skin Research and Technology, 2008, vol. 14, no. 4, pp. 425-435.

7. T.M. Mitchell, Machine Learning, the Mc-Graw-Hill Companies, Inc, 1997.

8. A. Mood, A.G. Franklin, and C.B. Duane, Introduction to the Theory of Statistics, McGraw-Hill Companies, Inc, 1974.

9. R.S. Sutton, A.G. Barto, Reinforcement Learning: An Introduction, MIT Press, 1998.

10. R.S. Sutton, A.G. Barto, and R.J. Williams, “reinforcement learning is direct adaptive optimal control”, Proceedings of IEEE American Control Conference, Boston, 1991, pp. 2143- 2146.

11. K.M. Iftekharuddin, Y. Li, “A biologically-inspired computational model for transformation invariant target recognition”, IEEE International Joint Conference on Neural Networks, Hong Kong, 2008, pp.1049 – 1056.

12. D.V. Prokhorov, D.C. Wunsch, “Adaptive Critic Designs”, IEEE Transactions on Neural Networks, 1997, vol. 8, no. 5, pp. 997-1007.

13. J. Si, A.G. Barto, W.B Powell, D.C. Wunsch, Handbook of Learning and Approximate Dynamic Programming, Wiley-IEEE Press, 2004.

14. R. Kohavi, "A study of cross-validation and bootstrap for accuracy estimation and model selection", Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence, 1995, vol.14, pp. 1137–1143.

15. J. Fogarty, R.S. Baker, and S.E. Hudson, “Case studies in the use of ROC curve analysis for sensor-based estimates in human computer interaction”, Proceedings of Graphics interface, Canadian Human-Computer Communications Society, School of Computer Science, University of Waterloo, Waterloo, Ontario Victoria, British Columbia, 2005, vol. 112, pp. 129-136.

16. S. Haykin, Neural Networks: A Comprehensive Foundation, Prentice Hall, 1998.

17. J. Kennedy, R. Eberhart, “Particle swarm optimization”, Proceedings of the IEEE International Conference on Neural Networks, Piscataway, NJ, 1995, pp. 1942-1948.

18. G. Guidse, G. K. Venayagamoorthy, “Comparison of Particle Swarm Optimization and Backpropagation as Training Algorithms for Neural Networks”, Swarm Intelligence Symposium, 2003, pp. 110-117.

19. J.H. Holland, Adaptation in Natural and Artificial Systems, University of Michigan Press, Ann Arbor, 1975.

20. T.M. Cover, P.E. Hart, "Nearest neighbor pattern classification," IEEE Transactions on Information Theory, 1967, vol. 13, no. 1, pp. 21-27.

21. D. Meyer, F. Leisch, and K. Hornik, “The support vector machine under test”, Neurocomputing, 2003, vol. 55, pp. 169-186.

SECTION
2. CONCLUSIONS

This dissertation proposes novel data fusion techniques, using both machine learning and computational intelligence algorithms, to improve image classification accuracy. Both feature- level and decision-level data fusion techniques were investigated and developed for a varied range of applications. Feature level fusion techniques were explored for not only skin cancer diagnosis but also arrow detection, by applying artificial neural network, evolutionary algorithm, particle swarm optimization, decision tree and clustering algorithms; decision-level fusion approaches included both fuzzy logic, and a voting algorithm for a graphic image type classification.

This research suggests feature-level data fusion techniques are useful for the removal of negative related/unrelated features and make the classification decision. While the decision level data fusion techniques are useful to combine the resource information and improve the classification result. Experimental results indicate that integrating various image processing techniques, feature-level data fusion techniques, and decision-level data fusion techniques, as proposed in this dissertation, can achieve high classification accuracy.

The contributions of this work are summarized in five journal papers and five conference papers, which are included in this dissertation.

BIBLIOGRAPHY

1. Hall D.L., Llinas J., “An introduction to multisensor data fusion”, Proceedings of the IEEE, 85(1), pp. 6-23, 1997.
2. Mitchell T., “Machine Learning”, McGraw Hill. ISBN 0-07-042807-7, p.2, 1997.
3. Quinlan J. R., “Induction of Decision Trees”, Kluwer Academic Publishers, Machine Learning, pp.81-106, 1986.
4. Jain A. K., Mao J., “Artificial neural networks: A tutorial”, IEEE Computer 29, pp.31–44, 1996.
5. Xu R., Wunsch D. I. I., “Survey of clustering algorithms”, IEEE Transactions on Neural Networks, 16(3), pp. 645–678, 2005.
6. Englebrecht A.P., “Computational Intelligence: An Introduction”, John Wiley & Sons, Inc., New York, 2002.
7. Banzhaf W., Nordin P., Keller R. E., Francone F. D., “Genetic Programming: An Introduction”, Morgan Kaufmann, San Mateo, CA, 1998.
8. Kennedy J., Eberhart R., “Particle swarm optimization”, Proceedings of the IEEE International Conference on Neural Networks, Piscataway, NJ, pp. 1942-1948, 1995.
9. Zadeh L.A., "Fuzzy sets", Information and Control, 8 (3), pp. 338–353, 1965.
10. Ross A. A., Govindarajan R., “Feature level fusion of hand and face biometrics”, Proceedings of SPIE Conference on Biometric Technology for Human Identification II, pp.196–204, March 2005.
11. Chibelushi C. C., Mason J. S. D., Deravi F., “Feature-level data fusion for bimodal person recognition,” Proceedings of 6th IEE Int. Conf. Image Processing and its Applications, pp. 399–403, 1997.
12. Gunatilaka A.H., Baertlein B.A., "Feature-level and decision-level fusion of noncoincidently sampled sensors for land mine detection," Pattern Analysis and Machine Intelligence, IEEE Transactions on, 23(6), pp.577-589, Jun 2001.
13. Parhami B., “Voting algorithms”, IEEE Trans. Reliability, 43 (4), pp.617–629, 1994.
14. Prabhakar S., Jain A. K., “Decision-level fusion in fingerprint verification”, Pattern Recognition, 35 (4), pp. 861-874, 2002.

VITA

Beibei Cheng received her BS in electronic engineering from the University of Electronic Science and Technology of China (UESTC), 2005. She earned an MS in computer engineering from Missouri University of Science and Technology (Missouri S&T), USA in May 2009. She began her PhD study at the Department of Electrical and Computer Engineering at Missouri S&T in June 2009. Her research interests include image processing, data fusion, machine learning, and computational intelligence. She received her PhD in Computer Engineering from Missouri University of Science and Technology in December 2012.