
Doctoral Dissertations

Student Theses and Dissertations

2013

Data fusion techniques for nondestructive evaluation and medical image analysis

Soumya De

Follow this and additional works at: https://scholarsmine.mst.edu/doctoral_dissertations



Part of the [Electrical and Computer Engineering Commons](#)

Department: Electrical and Computer Engineering

Recommended Citation

De, Soumya, "Data fusion techniques for nondestructive evaluation and medical image analysis" (2013).
Doctoral Dissertations. 2428.

https://scholarsmine.mst.edu/doctoral_dissertations/2428

This thesis is brought to you by Scholars' Mine, a service of the Missouri S&T Library and Learning Resources. This work is protected by U. S. Copyright Law. Unauthorized use including reproduction for redistribution requires the permission of the copyright holder. For more information, please contact scholarsmine@mst.edu.

DATA FUSION TECHNIQUES FOR NONDESTRUCTIVE EVALUATION AND
MEDICAL IMAGE ANALYSIS

by

SOUMYA DE

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

ELECTRICAL ENGINEERING

2013

Approved by:

R.J. Stanley, Advisor

R. H. Moss

W. V. Stoecker

H. Xiao

A. Adekpedjou

© 2013

Soumya De

All Rights Reserved

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 four articles that have been published/ to be published as follows:

Paper I, Pages 7-47, Soumya De, K. Gupta, R. J. Stanley, R. Zoughi, K. Doering, D. V. Aken, G. Steffes, M. O’Keefe, and D. Palmer, “A Comprehensive Structural Analysis Process for Failure Assessment in Aircraft Lap-Joint Mimics using Intra-Modal Fusion of Eddy Current Data,” *Research in Nondestructive Evaluation*, vol. 23, pp. 146-170, 2012.

Paper II, Pages 48-89, Soumya De, K. Gupta, R. J. Stanley, M. T. Ghasr, R. Zoughi, K. Doering, D. V. Aken, G. Steffes, M. O’Keefe, and D. Palmer, “A Comprehensive Multi-Modal NDE Data Fusion Approach for Failure Assessment in Aircraft Lap-Joint Mimics,” resubmitted after major revision, *IEEE Transactions on Instrumentation and Measurement*, 2012.

Paper III, Pages 90-137, Soumya De, R. J. Stanley, B. Cheng, S. Antani, R. Long, G. Thoma, “Automated Biomedical Text Detection Method in Support of Multi-Scale Content-Based Image Retrieval Systems,” submitted, *Computerized Medical Imaging and Graphics*, 2012.

Paper IV, Pages 138-182, Soumya De, R. J. Stanley, C. Lu, R. Long, S. Antani, G. Thoma, “A Data Fusion-based approach for Uterine Cervical Cancer Histology Image Classification,” submitted, *Computerized Medical Imaging and Graphics*, 2012.

ABSTRACT

Data fusion is a technique for combining data obtained from multiple sources for an enhanced detection or decision. Fusion of data can be done at the raw-data level, feature level or decision level. Applications of data fusion include defense (such as battlefield surveillance and autonomous vehicle control), medical diagnosis and structural health monitoring. Techniques for data fusion have been drawn from areas such as statistics, image processing, pattern recognition and computational intelligence. This dissertation includes investigation and development of methods to perform data fusion for nondestructive evaluation (NDE) and medical imaging applications. The general framework for these applications includes region-of-interest (ROI) detection followed by feature extraction and classification of the detected ROI. Image processing methods such as edge detection and projection-based methods were used for ROI detection. The features extracted from the detected ROIs include texture, color, shape/geometry and profile-based correlation. Analysis and classification of the detected ROIs was performed using feature- and decision-level data fusion techniques such as fuzzy-logic, statistical methods and voting algorithms.

ACKNOWLEDGMENTS

I would like to thank my advisor Prof. Joe Stanley for providing me with an opportunity for graduate studies under his guidance. I should thank him for his constant support and the flexibility he allowed me to pursue my interests for the last five and a half years I have been in graduate school. I would also like to thank Prof. Drewniak and Prof. Pommerenke, my unofficial advisors, for their hands-on teaching and helping me develop an experimental and questioning bent of mind. My sincere thanks to my PhD committee members, Prof. Moss, Prof. Xiao, Prof. Adekpedjou and Prof. Stoecker for taking time out of their schedule and agreeing to serve in my committee.

For the great internship experience I had at Cisco, many thanks are due to Doug White and Stephen Searce. Doug and Stephen are two of the best engineers I have ever met and they always helped me stay focused. The discussions we had in the hallways always motivated me to go that extra mile.

Being in graduate school has been tough at times and for helping me tide those times, I would like to thank all the friends I made in Rolla who made my stay here more fulfilling. My fiancée, Tridha has been a constant source of support and for that I would like to thank her separately. Finally I would like to thank my family- my parents for a lifetime dedicated to their children's success, and my sister, the backbone of my family.

TABLE OF CONTENTS

	Page
PUBLICATION DISSERTATION OPTION.....	iii
ABSTRACT.....	iv
ACKNOWLEDGMENTS	v
LIST OF ILLUSTRATIONS.....	xi
LIST OF TABLES.....	xiv
 SECTION	
1. INTRODUCTION	1
1.1. BACKGROUND	1
1.2. PROBLEM DESCRIPTION.....	2
1.3. SUMMARY OF CONTRIBUTIONS.....	3
 PAPER	
I. A COMPREHENSIVE STRUCTURAL ANALYSIS PROCESS FOR FAILURE ASSESSMENT IN AIRCRAFT LAP-JOINT MIMICS USING INTRA-MODAL FUSION OF EDDY CURRENT DATA.....	7
ABSTRACT	7
1. INTRODUCTION	8
2. LAP JOINT MIMIC PREPARATION.....	11
3. NDE DATA COLLECTION FOR TEST PANELS.....	15
4. COMPREHENSIVE STRUCTURAL ANALYSIS PROCESS.....	17
4.1. IMAGE ANALYSIS.....	18
4.1.1. Corrosion Detection.	18
4.1.2. Material Loss Estimation and Model Generation.....	21
4.2. SIMULATED FATIGUE LOADING	23

5. EXPERIMENTS PERFORMED	27
6. EXPERIMENTAL RESULTS.....	30
6.1. DATA FUSION RESULTS.....	30
6.2. MECHANICAL LOADING RESULTS FOR TEST PANELS	33
7. DISCUSSION	39
8. CONCLUSIONS.....	42
ACKNOWLEDGEMENTS	44
REFERENCES	45
 II. A COMPREHENSIVE MULTI-MODAL NDE DATA FUSION APPROACH FOR FAILURE ASSESSMENT IN AIRCRAFT LAP-JOINT MIMICS	 48
ABSTRACT	48
1. INTRODUCTION	49
2. METHODOLOGY	54
2.1. AIRCRAFT LAP-JOINT MIMIC PANEL DESCRIPTION	54
2.2. NDE DATA COLLECTION	54
2.2.1. Eddy Current Data.....	54
2.2.2. Ultrasonics Data.	56
2.2.3. Millimeter Wave Data.	57
2.3. NDE MULTI-MODAL DATA FUSION-BASED IMAGE ANALYSIS ..	59
2.4. TEST PANEL MODEL GENERATION	68
2.5. SIMULATED FATIGUE LOADING OF TEST PANELS	69
2.6. MECHANICAL LOADING OF TEST PANELS	71
3. EXPERIMENTAL RESULTS.....	74
3.1. MULTI-MODAL DATA FUSION EXPERIMENTS	74

3.2. MECHANICAL LOADING RESULTS FOR TEST PANELS	76
3.3. SIMULATED FATIGUE LOADING RESULTS	77
4. CONCLUSIONS.....	83
ACKNOWLEDGMENTS	86
REFERENCES	87
 III. AUTOMATED BIOMEDICAL TEXT DETECTION METHOD IN SUPPORT OF CONTENT-BASED IMAGE RETRIEVAL SYSTEMS	 90
ABSTRACT	90
1. INTRODUCTION	91
2. METHODOLOGY	95
2.1. DATA SET INVESTIGATED	95
2.2. DETECTION OF HORIZONTAL AND VERTICAL TEXT BLOCKS...	95
2.3. FUSED IMAGE GENERATION	99
2.4. SEGMENTATION OF INDIVIDUAL CHARACTERS	100
2.5. CHARACTER NORMALIZATION AND FEATURE EXTRACTION.	103
2.5.1. Geometry Features.	104
2.5.2. Exemplar-based Correlation Features.	109
2.5.3. Profile-based Basis Function Correlation Features.	111
2.5.4. Fourier Descriptors.	113
2.6 GROUND TRUTHING OF EXTRACTED CHARACTERS.....	114
2.7 CLASSIFICATION OF EXTRACTED CHARACTERS	116
3. EXPERIMENTS PERFORMED	118
3.1. EVALUATION OF THE TEXT BLOCK DETECTION STEP	118
3.2. BENCHMARK OCR EVALUATION.....	118

3.3. EVALUATION OF THE METHOD DEVELOPED	118
3.4. FEATURE EVALUATION AND SELECTION	119
4. EXPERIMENTAL RESULTS AND ANALYSIS	120
5. CONCLUSIONS.....	129
REFERENCES	131
 IV. A FUSION-BASED APPROACH FOR UTERINE CERVICAL CANCER HISTOLOGY IMAGE CLASSIFICATION.....	 138
ABSTRACT	138
1. INTRODUCTION	139
2. METHODOLOGY	145
2.1. MEDIAL AXIS DETECTION	146
2.2. VERTICAL IMAGE SEGMENTATION	154
2.3. FEATURE EXTRACTION	157
2.3.1. Texture Features.	158
2.3.2. Intensity Shading Features.	159
2.3.3. Geometry (Triangle) Features.	161
2.3.4. Profile-based Correlation Features.	163
2.4. CLASSIFIER ALGORITHMS	166
3. EXPERIMENTS PERFORMED	168
3.1. CLASSIFICATION OF VERTICAL SEGMENT IMAGES	168
3.2. FUSION OF THE CIN GRADES OF VERTICAL SEGMENTS	168
3.3. CLASSIFICATION OF THE WHOLE EPITHELIUM.....	169
3.4. FEATURE EVALUATION AND SELECTION	170
4. EXPERIMENTAL RESULTS AND ANALYSIS	171

5. CONCLUSIONS.....	177
REFERENCES	179
SECTION	
2. CONCLUSIONS	183
BIBLIOGRAPHY.....	185
VITA	187

LIST OF ILLUSTRATIONS

PAPER I	Page
Figure 2.1: Schematic of aluminum plates for corroded test panel.	12
Figure 2.2: Masked aluminum plates with corroded patches prior to riveting.	14
Figure 2.3: Pictures of painted corroded and pristine panels in riveted regions.	14
Figure 3.1: Reference panel used for calibration purposes.	16
Figure 4.1: Overview of the structural analysis process.	17
Figure 4.2: Defect and De-Meaning Diameter for RX detector.	19
Figure 4.3: ANSYS generated model of test panel.	23
Figure 4.4: Location key for crack initiation sites in test panels.	24
Figure 4.5: Boundary conditions defined for simulated fatigue loading.	25
Figure 5.1: Cyclical fatigue loading setup.	28
Figure 5.2: Data collection and mechanical testing rounds.	29
Figure 6.1: EC Data for test panel 1 prior to the first round of mechanical loading.	31
Figure 6.2: EC Data for test panel 1 after the first round of mechanical loading.	31
Figure 6.3: EC Data for test panel 1 after second round of mechanical loading.	31
Figure 6.4: EC data for pristine panel prior to the first round of mechanical loading.	32
Figure 6.5: EC data for pristine panel after first round of mechanical loading.	32
Figure 6.6: EC data for pristine panel after second round of mechanical loading.	32
Figure 6.7: Developed crack in panel 1 after the second round of mechanical loading ...	34
 PAPER II	
Figure 1.1: Flowchart of the structural analysis process.	52
Figure 2.1: Picture of painted pristine and corroded panels with riveted area shown.	55
Figure 2.2: Reference panel and EC image of corroded test panel.	56
Figure 2.3: Reference panel and UT image of corroded test panel.	57
Figure 2.4: Schematic of the V-band millimeter wave reflectometer.	59
Figure 2.5: Overview of the multi-modal structural analysis process.	60
Figure 2.6: Various modality combinations.	63
Figure 2.7: Sample schematic of reference panel used for material loss estimation.	67
Figure 2.8: 3-D Model generation for a test panel.	69

Figure 2.9: Meshed model showing the nodes under analysis.....	71
Figure 2.10: Cyclical mechanical loading setup.	72
Figure 2.11: Side view schematic of the panel (a) unloaded and (b) loaded.	73
Figure 3.1: EC and Material loss estimation data of test panel 1.	74
Figure 3.2: Data collected and data fusion based material loss estimation result.....	75
Figure 3.3: Data collected and data fusion based material loss estimation.	75
Figure 3.4: Multi-modal data fusion based material loss estimation for panel 1.....	76
Figure 3.5: Location key for crack initiation sites in test panels.	77
Figure 3.6: Fuzzy-logic based method of automatically finding the threshold.	80
PAPER III	
Figure 1.1: Overview of the image analysis steps.	94
Figure 2.1: Text Block detection	99
Figure 2.2: Creation of fused image I_F	100
Figure 2.3: First step of segmenting individual characters.	101
Figure 2.4: Individual character segmentation.....	102
Figure 2.5: Breaking up original character into 9 equal sized windows.....	104
Figure 2.6: Breaking up the normalized character image into two halves.....	107
Figure 2.7: Sample exemplar image for alphabet ‘A’ [58].	110
Figure 2.8: Profiles of character images.	111
Figure 2.9: The WDD functions used.	112
Figure 2.10: Assigning labels to extracted characters	115
Figure 2.11: Assigning labels to extracted characters (numbers)	116
Figure 4.1: Distribution of character labels across the alphabet classes.....	120
Figure 4.2: Comparison of overall accuracy.....	126
Figure 4.3: Sample results obtained.....	127
Figure 4.4: Character overlap during localization	128
PAPER IV	
Figure 1.1: CIN grading label examples.	140
Figure 1.2: Overview of CIN grade classification method developed in this study.	143
Figure 2.1: Original image and pathologist segmented image of epithelium.	145
Figure 2.2: Binary mask rotation (Step 1-4).	147

Figure 2.3: Distance transform image (D) obtained (Step5).....	148
Figure 2.4: Medial axis of the binary mask I^B (Step 6).....	148
Figure 2.5: Determined medial axis v/s desired medial axis.	149
Figure 2.6: Dividing the medial axis into three pieces (Step 7).....	150
Figure 2.7: The center line of the bounding box of the epithelium (Step 8).	150
Figure 2.8: The intersection points were obtained (Step 9).....	151
Figure 2.9: The edge image was computed and divided into three parts.....	152
Figure 2.10: Determined medial axis and final medial axis (Step 14).....	154
Figure 2.11: Schematic showing the vertical segments.	154
Figure 2.12: Vertical segment images creation.....	156
Figure 2.13: Vertical segment images.....	157
Figure 2.14: Representative shadings within a sample epithelium image.	160
Figure 2.15: Triangles formed from the vertical segments.....	162
Figure 2.16: Profile obtained for computing the profile-based features.....	164
Figure 2.17: The WDD functions used.	164
Figure 3.1: Fusion-based approach v/s whole image approach.	170

LIST OF TABLES

PAPER I	Page
Table 6.1 : Crack locations summary after each round of mechanical loading.....	33
Table 6.2 : Summary of Alt-SI values for simulated fatigue loading.....	35
Table 6.3 : Summary of crack locations identified.....	37
Table 6.4 : Accuracy of simulated loading vs. mechanical loading.....	37
PAPER II	
Table 3.1: Summary of crack locations for two rounds of mechanical loading.....	77
Table 3.2: Summary of Alt-SI values for test panel 1.....	78
Table 3.3: Alt-SI values for test panels 1-4 and pristine panel {EC, UT}.....	79
Table 3.4: Accuracy metrics of simulated fatigue loading vs. mechanical loading.....	81
Table 3.5: Summary of accuracy metrics.....	82
PAPER III	
Table 2.1: Feature Description.....	103
Table 4.1: Accuracy metrics for automatic text block detection.....	122
Table 4.2: Ten-fold cross-validation results comparison.....	123
Table 4.3: Comparison of accuracies using only certain features.....	125
Table 4.4: p-values obtained from the MLR for the 112 ($\alpha=0.01$).....	125
PAPER IV	
Table 2.1: Feature Description.....	158
Table 4.1: Result of SVM classifier (features from vertical segments).....	171
Table 4.2: Result of LDA classifier result (features from vertical segments).....	171
Table 4.3: Result of SVM classifier result (features from whole images).....	172
Table 4.4: Result of LDA classifier result (features from whole images).....	172
Table 4.5: Features with corresponding p-values.....	173
Table 4.6: Result of SVM classifier (reduced features & vertical segments).....	174
Table 4.7: Result of LDA classifier (reduced features & vertical segments).....	174

1. INTRODUCTION

1.1. BACKGROUND

Data fusion can be defined as methods by which data from multiple sources can be combined to improve detection and decision accuracy. Additionally, data fusion algorithms are used to obtain complementary information from sensors and/or information sources measuring different physical phenomenon. It has been of interest in areas such as military, remote sensing, finance, medical-diagnosis, structural health monitoring and automated manufacturing [1-5]. Data fusion is a multi-disciplinary research area that draws concepts from fields such as statistics, signal processing, computer science and artificial intelligence.

Types of data fusion include low (raw-data) level fusion, feature level fusion and decision level fusion. Raw data fusion employs methods such as statistical detection and estimation techniques. Feature level fusion is used to fuse descriptive features extracted from multiple sources measuring similar or dissimilar physical phenomenon. This is done by combining the features into a single 1-D vector which could be used for discrimination purposes.

Fusion of preliminary decisions/assessments made using data from single or multiple sources is termed as decision level data fusion [6-8]. Data fusion techniques such as statistical methods, computational intelligence and machine learning, have been used by previous researchers for solving problems such as fault diagnosis, safety of complex systems, plant monitoring and monitoring in biomedicine [9-11].

1.2. PROBLEM DESCRIPTION

In this dissertation, data fusion methods for nondestructive evaluation (NDE) and medical image analysis are investigated. Data fusion is being increasingly utilized for structural analysis by obtaining additional and complementary NDE data. Nondestructive evaluation data fusion typically facilitates structural feature integration and/or feature characterization based on combining information from several measurement modalities. Such modalities may consist of data from multiple sensors and/or the fusion of data processed using different parameters from the same sensor. Data collected from structures using NDE approaches and processed with image/signal processing and computational intelligence techniques can enhance the capability of evaluating those structures [12].

Medical images such as those found in medical publications include graphical images like diagrams, tables, histograms and flowcharts. Since medical images for a given modality (e.g. MRI, Histology or X-Ray) are very similar in nature, existing content-based image retrieval (CBIR) methods based only on the visual features of the whole image, such as texture and shape, may not be sufficient for accurate retrieval of biomedical images [13-16]. In addition to using text (found in image captions and within the article) and visual features, retrieval of characters within images found in biomedical publications is an important task for obtaining complementary information for CBIR and analysis. Therefore, enhanced CBIR and analysis requires fusion of the various categories of information (image text, visual features and image captions).

One of the methods for early cervical cancer diagnosis includes medical image analysis of the squamous epithelium of the cervix. Cervical intraepithelial neoplasia (CIN)

is defined as the pre-malignant condition in which the atypical cells are limited in the epithelium region only [17-18]. It has been observed that CIN shows progressively increasing atypical cells across the spectrum of the epithelium, from top to bottom, resulting in different CIN grades (CIN1, CIN2, CIN3) [17-22].

Recent studies have shown that an epithelium with atypical cells could exhibit different CIN grades in different vertical segments of the epithelium [18]. Therefore for increased diagnostic accuracy, it is important to analyze the different vertical segments of the epithelium and then fuse information from each of these segments to obtain the CIN grade of the whole epithelium.

1.3. SUMMARY OF CONTRIBUTIONS

This dissertation consists of four journal papers, as presented in the publication list. The unique contributions can be summarized as follows:

Nondestructive evaluation of aircraft structures. The investigation on NDE data fusion presented in this dissertation extends the realm and results of previous research in this field. Existing techniques which use data fusion, combine multi-modal NDE data to enhance test structure visualization for analysis. The comprehensive structural analysis process in this dissertation advances traditional/existing data fusion techniques in the context that data fusion is performed and utilized for automated structural analysis. Intra- and inter-modality fusion was performed using RX statistical and Fuzzy-logic based methods to detect corrosion and defects and to quantify the amount and location of the corrosion and defects in structures for automatically developing models of the test structures. The utility of data fusion here is to facilitate better characterization of the

corrosion and defects present within the test structures for generating models representative of the actual structures so that vulnerable locations may be accurately identified with simulated fatigue loading. It also facilitates predictive analysis by automatically generating geometrical models of test structures based on quantized areas of material loss for Finite Element Modeling (FEM)-based analysis. Based on the experimental results obtained, it was observed that the multi-modal data fusion-based FEM models exhibit a better ability to predict failure as compared to the uni-modal models, thereby establishing the potential of multi-modal data fusion for automated prediction of failure.

Content-based image retrieval. For CBIR and analysis, an end-to-end system for automatic detection and recognition of text labels in biomedical publication images is investigated. Previous research in this domain has been primarily focused on analyzing the whole biomedical image as compared to the ROIs (such as text blocks, arrows, etc.). Image analysis and feature-based approaches are used to extract and recognize text blocks within these images. Raw data fusion of the horizontal and vertical text blocks was performed to create a fusion image. Individual characters were segmented from the fusion image, after which features were extracted from each segmented character. The features extracted from the individual characters were then fused into a single vector which was used as the input to a Support Vector Machine (SVM)-based classifier for character recognition. Support vector machine gave superior results as compared the benchmark OCR tool-Tesseract, which suggests the usefulness of data fusion-based method developed in this study. Furthermore, a multinomial logistic regression (MLR)-based statistical method was investigated for feature level data fusion, wherein the feature

vector was used as input to the MLR-based method. Fusion of the various categories of features was tested to select the best set of features for character discrimination by analyzing the p-values obtained from MLR. It was seen that the features selected by the MLR-based method to select the statistically significant features provided the best accuracy, when subjected to the SVM classifier, as compared to using all the features. This demonstrates that data fusion is useful at the feature-level to remove the statistically insignificant features for improving detection accuracy.

Histology image analysis for Cervical Intraepithelial Neoplasia (CIN) diagnosis.

The research presented for CIN diagnosis in this dissertation builds off of the methods developed in [18, 20]. A novel fusion-based approach was developed for medial axis (skeleton) determination in epithelium images. Fusion of distance transform- and bounding box-based method was performed for enhanced medial axis estimation in this study. It was observed during our experiments that the distance transform approach provided a good estimate of the medial axis in the middle portions of the epithelium image, but the bounding box based method provided a better estimate of the medial axis along the left and right edges of the epithelium. Based on the experimental results, the fused medial axis provided a better estimate of the skeleton of the epithelium region as compared to just using the distance transform. This showed the importance of data fusion at the raw-level for enhanced medial axis estimation. On the other hand, expert pathologists usually perform CIN diagnosis by analyzing the epithelium as a whole and providing a CIN grade for the whole epithelium [21-23]. However recent studies have shown that the epithelium might not have a unique CIN grade throughout and might exhibit different CIN grades in different parts of the epithelium [20]. In this dissertation,

a novel fusion-based automated method for CIN grading is proposed, wherein the epithelium is divided into vertical segments each of those segments are classified into one of the CIN grades and then the vertical segment grades were fused to obtain the CIN grade of the whole epithelium. In this manner, fusion of CIN diagnosis from the different parts of the epithelium was performed for an enhanced CIN diagnosis. Experiments showed that the CIN diagnosis accuracy obtained using the vertical segments fusion method was about 15% higher in comparison to the benchmark method [23], which suggests the utility of data fusion for enhanced detection accuracy.

PAPER

I. A COMPREHENSIVE STRUCTURAL ANALYSIS PROCESS FOR FAILURE ASSESSMENT IN AIRCRAFT LAP-JOINT MIMICS USING INTRA-MODAL FUSION OF EDDY CURRENT DATA

Soumya De^a, Kapil Gupta^b, R. Joe Stanley^a, Reza Zoughi^a, Kenneth Doering^c, David Van Aken^c, Gary Steffes^d, Matt O'Keefe^c, Donald Palmer^e

^aDepartment of Electrical and Computer Engineering, Missouri University of Science and Technology (S&T), Rolla, MO, 65409, USA

^bCoherix Inc., Ann Arbor, MI, 66062, USA

^cDepartment of Materials Science & Engineering, Missouri University of Science and Technology (S&T), Rolla, MO, 65409, USA

^dAir Force Research Laboratory (AFRL), Wright Patterson, OH 45433, USA

^eThe Boeing Company, Boeing Phantom Works, St. Louis, MO 63166, USA

ABSTRACT

A comprehensive structural analysis process is presented for quantizing and evaluating characteristics of aircraft lap-joint mimics. The process investigated here consists of NDE data acquisition, defect detection and characterization involving material loss estimation, three-dimensional structural model generation, finite-element modeling to simulate fatigue damage and comparison with actual tensile fatigue-induced structural analysis data (mechanical loading). The structural analysis process is examined using five test panels consisting of stacked and riveted aluminum plates, configured as lap-joints, four of them with different corroded patches at different layers of the lap-joints and one painted pristine panel used as a reference. The test panels were subjected to three rounds of fatiguing with the final round resulting in complete fatigue failure. Eddy current data was obtained from the test panels prior to each round of mechanical loading. Comparing the simulated fatigue loading and the mechanical loading results for identifying susceptible-to-failure areas on the test panels, this comprehensive structural analysis process found the correct location of failure areas at rates as high as 88.9%

1. INTRODUCTION

Corrosion is a significant maintenance issue, especially for aging aircraft, with the U.S. Air Force recently estimating corrosion maintenance costs exceeding \$5 billion per year according to the 2009 Cost of Corrosion Report [1]. Corrosion results in defects such as surface and fatigue cracks, deterioration of fasteners, and skin disbonds [2]. Corrosion is often hidden under paint and cannot be visually detected. Nondestructive evaluation (NDE) techniques have been used to enhance the process for analyzing corrosion under paint and thin laminates [3], including for structural analysis of aging aircraft structures [4]. In addition, the benefit of NDE methods for fatigue analysis has been examined based on cost constraints for an experimental setup developed for predicting the residual life in corroded fuselage lap joints [5].

NDE methods may provide qualitative information (i.e., defect or non-defect) and/or quantitative evaluation (i.e., attributes such as damage dimensions including thickness, spatial extent, etc.). However, limitations of the physical process involved in collecting NDE data can inhibit the detection of certain defects which are actually present in the structure [2]. These limitations may include changes in the operating distance of the NDE source from the structure for data collection (e.g., liftoff change), the data value ranges of the NDE source, the physical space (i.e., square dimensions on the structure's surface that correspond to a single measurement) represented on the structure by those data values collected from the NDE source (e.g. physical dimensions associated with a pixel), the geometry of the structure, or measurement errors from the NDE source

resulting in noisy data. An overall system using NDE data for structure inspection should minimize structure downtime and required inspection resources and account for the limitations of the NDE source(s) [6].

Data fusion is being increasingly utilized for obtaining additional and complementary NDE data, corresponding to structural information, through signal or pixel-based fusion [6]. NDE data fusion typically facilitates structural feature integration and/or feature characterization based on combining information from several measurement modalities, which may consist of data from multiple sensors and/or the fusion of data processed using varied strategies/algorithms/parameters from the same sensor [6]. Data collected from structures using NDE approaches and processed with image and signal processing and computational intelligence techniques can enhance the capability of evaluating those structures.

Eddy current (EC), ultrasound, optical, thermography, and radiography are common modalities for structural nondestructive inspection. Variations of these modalities and their application to structural inspection of discontinuities are presented in [6]. Data fusion techniques using these modalities have been used for visually enhancing structural features [7], improving signal-to-noise ratio, improving signal interpretation [8], enhanced defect detection and characterization [9], and improved data registration [10], among others. Incorporating data fusion methods using NDE data results in improved analysis of structural features for inspection.

However, these methods require subsequent analysis in order to: 1) compare structural features to physical parameters for those features, 2) evaluate the overall structural health including the determination of vulnerabilities in the structure susceptible

to failure, and 3) assess the residual useful life of the structure. In this research, a comprehensive structural analysis process is presented for quantizing and measuring characteristics of aluminum aircraft lap-joint mimics. The process investigated here consists of NDE data acquisition, defect detection and characterization, three-dimensional (3-D) structural model generation, finite-element modeling-based (FEM) simulation of fatigue damage for structural analysis. This research builds upon preliminary works by Gupta *et al.* for automated detection of hidden corrosion [11] and material loss estimation [12] from EC images collected for lap-joint mimics using intra-modal fusion methods. The comprehensive structural analysis process provides an end-to-end method from structure NDE data collection to FEM-based structural modeling, including a novel approach for computer-assisted 3-D model generation directly from the NDE data collected. For verification of the proposed process, comparison of simulated fatigue damage is performed with experimental cyclic loading tests. The following sections of this paper presents the description of: 1) preparation of the lap-joint mimic test panels, 2) NDE data collection for the test panels, 3) the comprehensive structural analysis process, 4) the mechanical loading of the test panels, 5) the detailed algorithms developed for the different steps of the investigated comprehensive structural analysis process, 6) the experimental results of the structural analysis process for simulated fatigue loading and the comparison to the cyclic loading, and 7) finally a discussion of the results.

2. LAP JOINT MIMIC PREPARATION

A lap-joint is a region in an aircraft fuselage where at least two layers (i.e., inner and outer) of skin overlap and are riveted and/or spot-welded together, with the region covered with a thin layer of paint or appliqué. Water vapor, rain, and other agents can be pulled under the paint and into thin gaps present between the metal plates and the rivet holes through capillary action and remain within the lap-joint for extended periods of time, resulting in corrosion. Corrosion can cause metal thinning and stress points resulting in future cracks (under in-service loading). The geometry present in lap-joints complicates the detection of corrosion in these regions since, without removing the paint and the rivets, visual inspection is impossible.

For this research, five aluminum aircraft lap-joint mimics were produced and used as test panels for corrosion and fatiguing analysis. The lap-joint test panel set consisted of four corroded and one pristine panel. Each lap-joint test panel was produced using two AL2024-3 aluminum plates with dimensions of 254 mm (10") by 76.2 mm (3") with a nominal thickness of 1.6 mm (0.063"). Figure 2.1a presents the cross-section schematic with dimensions of a test panel, showing the rivets and areas of corrosion (shown in orange).

The four corroded test panels were constructed as follows. First, the aluminum plates were masked and selectively corroded to produce corrosion patches. The corrosion patches with dimensions for the top and bottom of each aluminum plate used in forming a test panel are given in Figure 2.1b (shown in orange). Note that the aluminum plates were

riveted after the corrosion process was completed, with the rivet heads in the designated locations given in Figure 2.1b.

Figure 2.2 shows the actual corroded patches of the test plates used in forming a corroded test panel. Accelerated corrosion of these eight plates was induced in a salt fog chamber following the ASTM B117 test standard. The ASTM B117 standard, designed in 1939, is a widely used technique for evaluating the relative corrosion of various metals and/or coatings. The salt spray chambers designed according to this standard are automated to maintain a specified environment within the chamber [13].

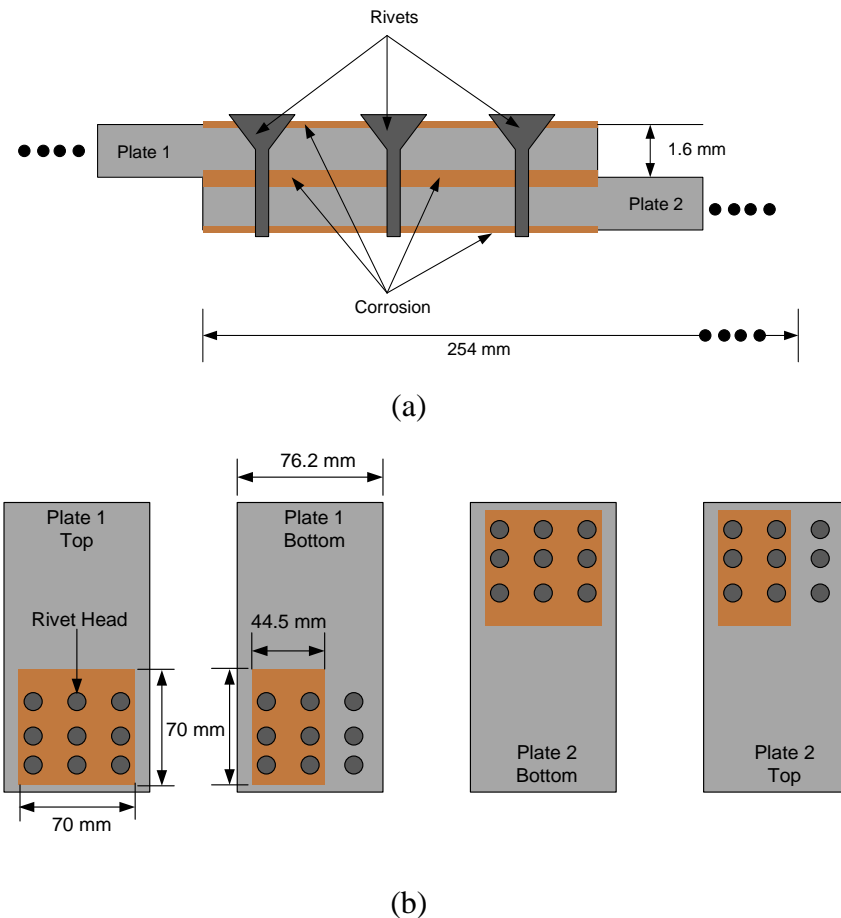


Figure 2.1: Schematic of aluminum plates for corroded test panel.
a) Cross-section of riveted, corroded test panel. b) Schematics with dimensions of the corrosion regions for the top and bottom of each aluminum plate used to form each test panel (all schematics are not-to-scale).

Prior to putting the plates in the chamber, they were rinsed with acetone followed by tap water. Next, the plates were immersed for five minutes in an alkaline cleaning solution, Turco-4215 NC LT (manufactured by Turco, Spain and distributed by Henkel), at a temperature of 55°C and with a pH of 9. Turco-4215 NC LT, which is non-corrosive to Aluminum, is a mild alkaline cleaner used in the aerospace industry. The cleaning procedure occurs at elevated temperatures (above 40 °C), when one of the surface active agents becomes less soluble and thus separates from the metal and floats on the surface, which can be removed by draining [14]. After alkaline cleaning, the plates were rinsed again with de-ionized water. Afterwards, the plates were immersed in a 1 wt% sulfuric acid solution heated at 50°C for five minutes to expose the inter-metallic compounds to facilitate corrosion promotion. Finally, the plates were rinsed with de-ionized water and masked off using electroplaters tape in order to expose the desired corrosion patches, as shown in Figure 2.2. The plates were corroded in the salt chamber between 3-7 days, being evaluated every day to visually determine the presence of pits greater than 1 mm in size and salting of the pits to provide the extent of corrosion. The plates were pulled from the salt spray chamber when the amount of corrosion was determined to be constant based on the localized pits.

Second, the aluminum plates were stacked and riveted in the corresponding rivet head positions from Figure 2.1b, at the Boeing Company (St. Louis, MO), according to aircraft manufacturing specifications. The two primary goals in the sample preparation process were: a) presence of an overlap in at least some corroded areas in the lap-joints, and, b) corrosion present in some of the riveted interfaces.

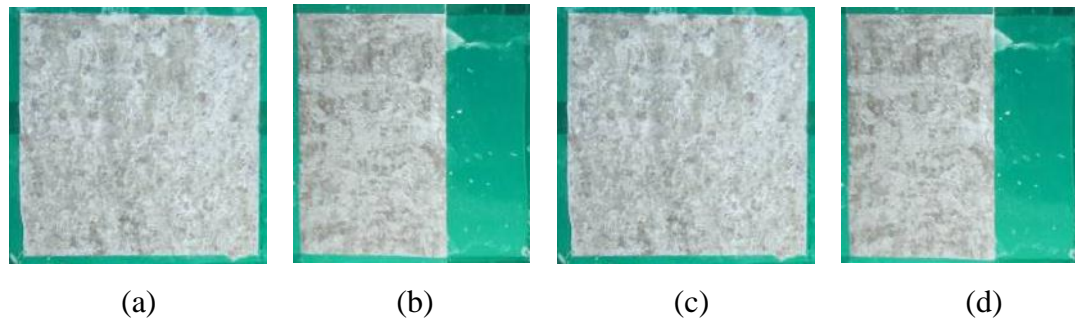


Figure 2.2: Masked aluminum plates with corroded patches prior to riveting.
 (a) Top plate 1. (b) Bottom plate 1. (c) Bottom plate 2. (d) Top plate 2.

Each test panel was assembled to produce a 76.2 mm by 76.2 mm (3" by 3") lap-joint mimic having nine rivets with a spacing of 25.4 mm (1") among them. These riveted panels were then painted, according to specifications used for aircraft painting. Figure 2.3 shows the riveted region of a final corroded and the pristine panel showing top, side, and bottom views.

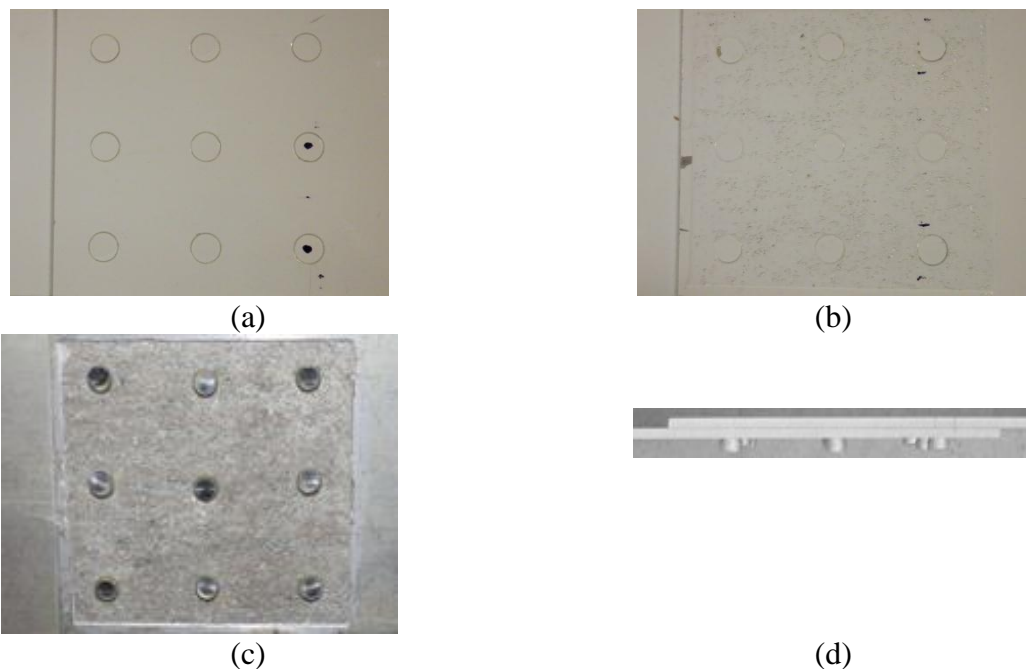


Figure 2.3: Pictures of painted corroded and pristine panels in riveted regions.
 (a) Top corroded, (b) top pristine, (c) bottom corroded, and (d) side.

3. NDE DATA COLLECTION FOR TEST PANELS

For the five constructed lap-joint mimic test panels, metal thinning associated with the presence of corrosion induced in these panels was evaluated using eddy current data collection and evaluation. In order to assess and characterize corrosion, eddy current data was collected from the painted (top) side of the test panels, as would be performed for actual NDE data collection.

The test panels were inspected at frequencies of 1, 2, 4, 8, 10, 20, 50, 100, and 200 kHz, to provide data as a function of depth into the panels, thereby providing information about the layer right beneath the paint and all the way to the bottom layer [15]. Since signals at lower frequencies penetrate more than those at higher frequencies, this range of EC frequencies was selected based on the thickness of the test panels [16] and the effective depth of penetration, denoted as δ (in mm). This phenomenon could be represented as Equation (1):

$$\delta = \frac{660}{\sqrt{C\mu f}} \quad (1)$$

where C is the International Annealed Copper Standard constant for the material's electrical conductivity (between 28.5-32.5 MS/m for 2024-T3 aluminum), μ is the magnetic permeability of the material (for aluminum $\mu = 1 \text{ H/m}$), and f is the frequency of the eddy current inspection (Hz).

The EC data was collected using the Boeing MAUS[®] (Mobile Automated Scanner) raster scanning system, with phase lag adjustment to discern defect/corrosion

depth (see Data Collection block in Figure 4.1). During the data collection process, a reference panel was used with three separate areas of material loss representative of 0%, 10%, 15%, and 20% material loss, as shown in Figure 3.1. Physically, for example, the pixels contained in the 15% corrosion section represent 15% deviation from the pixels contained in the 0% corrosion region. These reference panel images were produced for calibration purposes for corrosion detection and material loss estimation algorithm development.

In general for the aircraft industry, a material loss of less than 10% is considered appropriate for flying without any significant maintenance issues; however, material losses of 10% and above would invoke corrective measures for fixing the corrosion related to thinning [17].

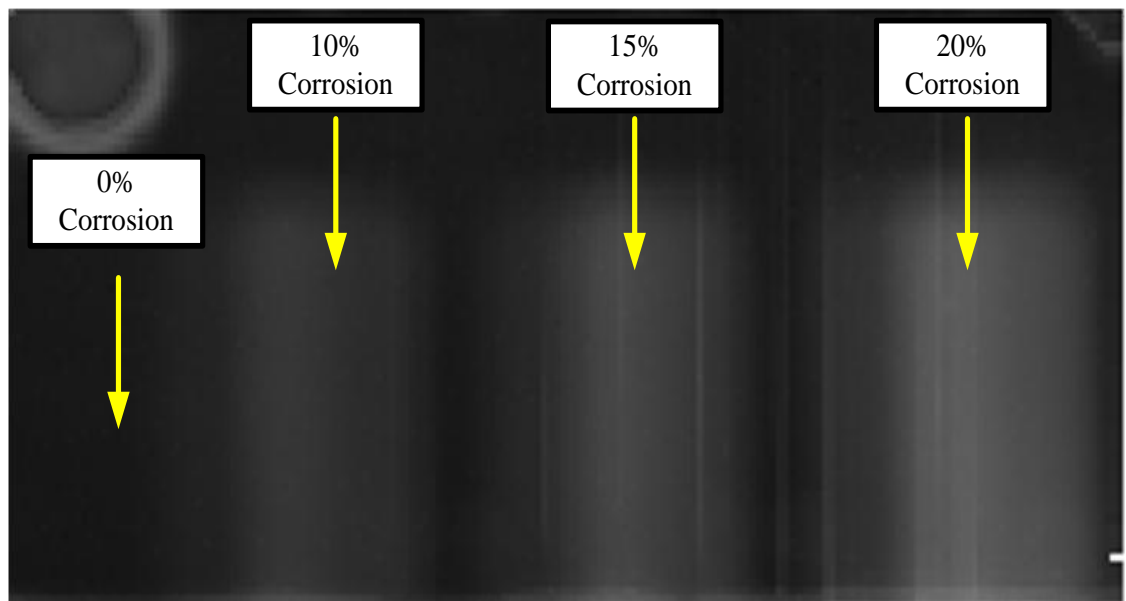


Figure 3.1: Reference panel used for calibration purposes

4. COMPREHENSIVE STRUCTURAL ANALYSIS PROCESS

For each test panel and the EC data collected from that test panel, the comprehensive structural analysis process, presented in Figure 4.1, is applied for defect detection, characterization, modeling, and simulated fatigue loading to determine positions of potential vulnerability on the test panel. This process involves steps for computer-assisted analysis of a test panel, including: a) EC data collection, i.e. image acquisition, b) image analysis based defect detection and characterization and model generation, and c) FEM-based structural model analysis. The following algorithms and methods were used for the different steps of the structural analysis process using the obtained EC images.

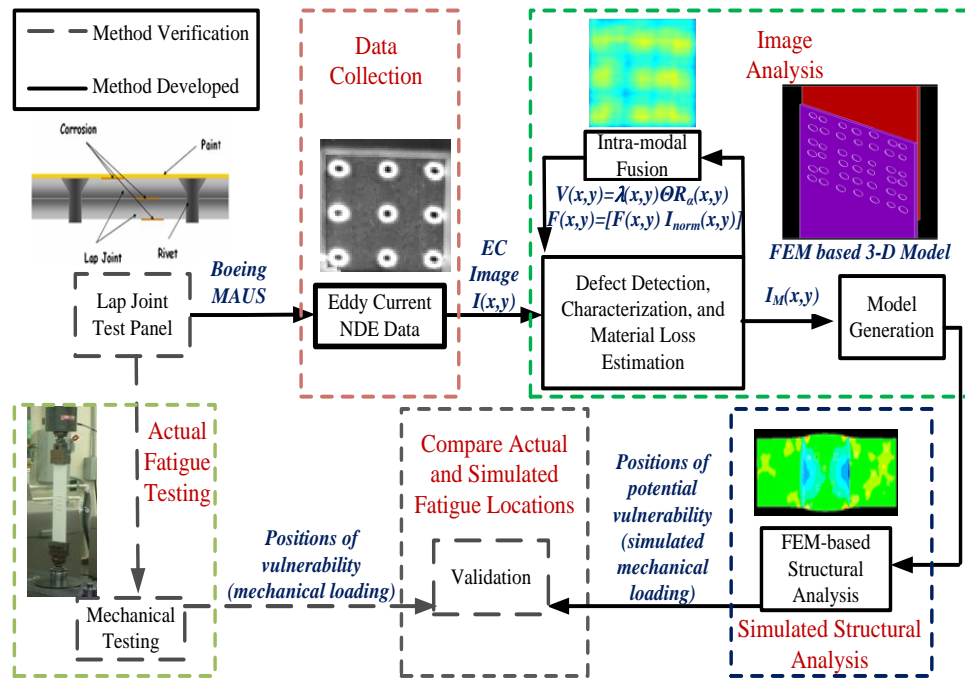


Figure 4.1: Overview of the structural analysis process.

4.1. IMAGE ANALYSIS

4.1.1. Corrosion Detection. For corrosion detection, intra-modal data fusion was performed based on integrating the outputs of two anomaly detectors, RX and data-driven fuzzy logic, applied to the same EC images. The anomaly detectors use localized image analysis to find potential defects. This integrated approach was explored in previous research for corrosion detection [11] and material loss estimation [12], with experimental results showing enhanced corrosion detection and improved material loss estimation based on combining these methods as compared to using them individually.

The RX and fuzzy logic methods were used to compute confidence values for anomalies representative of corrosion in an EC image. These methods have been presented in detail in [18] and are only briefly mentioned here. With reference to the output of the Data Collection step in Figure 4.1, let $I(x, y)$ denote the EC image acquired for a test panel, with cropping of the outer 3 rows and 3 columns to address EC data edge effects. Then, the RX confidence value at each pixel (x, y) , denoted as $\lambda(x, y)$ with value range $0 \leq \lambda(x, y) \leq 1$, is defined as:

$$\lambda(x, y) = p^T N p \quad (2)$$

where p is a representative signature of a corrosion patch, N is the unknown background covariance matrix, which is computed from the zero mean image, M . The zero mean image and the background covariance matrix is obtained using Equation 3 and 4 respectively as below,

$$M(x, y) = I(x, y) - \frac{\sum_{(a,b) \in D} I(a, b)}{|D|} \quad (3)$$

$$N = \frac{\sum_{(a,b) \in E} M(a,b)M(a,b)^T}{|E|} \quad (4)$$

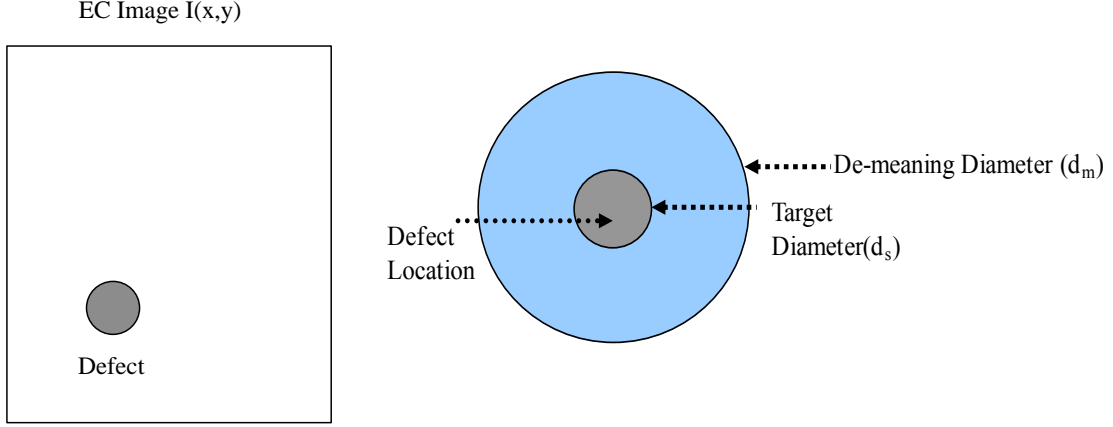


Figure 4.2: Defect and De-Meaning Diameter for RX detector.

where D represents a circular demeaning mask with diameter d_m and cardinality $|D|$ as shown in Figure 4.2. E represents the potential target mask in the form of a circular corrosion patch with diameter d_s and cardinality $|E|$. Hence, the representative corrosion patch, p , marked as the grey area in Figure 4.2, can be represented as Equation 4,

$$p = \frac{\sum_{(a,b) \in E} M(a,b)}{\sqrt{|E|}} \quad (5)$$

The demeaning mask and the target mask are concentric circles, (as shown in Figure 4.2) where $d_m > d_s$, with $d_m = 13$ and $d_s = 7$ determined experimentally in previous research [11]. Physically, the RX statistic $\lambda(x,y)$ provides the confidence value for corrosion detection at each pixel in the image $I(x,y)$ wherein higher values of λ would represent stronger ‘hits’ in detecting corrosion.

For the fuzzy logic method, let G denote the fuzzy set representative of the gray levels associated with the non-corrosion areas (background) in an EC image. The histogram of the EC image I is computed and at each bin (z) of the secondary histogram, the associated membership function $\mu_G(z)$ can be represented mathematically as follows:

$$\mu_G(z) = \begin{cases} (z/t)^{0.5} & \text{for } 0 \leq z < t \\ 1 & \text{for } t \leq z \end{cases} \quad (6)$$

where t is determined empirically to be 95% of the area under the secondary histogram of the EC image I [11]. If $|S_\alpha(x, y)|$ is the number of eight connected neighbors of a particular pixel $I(x, y)$ such that $\mu_G(I(x, y)) \geq \alpha$ and $|S_0(x, y)|$ is the number of eight connected neighbors such that $\mu_G(I(x, y)) \geq 0$ then, a fuzzy clustering anomaly confidence measure, denoted as $R_\alpha(x, y)$ at the pixel location (x, y) is defined as:

$$R_\alpha(x, y) = \frac{|S_0(x, y)|}{|S_\alpha(x, y)|} \quad (7)$$

where $0 \leq R_\alpha(x, y) \leq 1$ and $\alpha = 0.7$ from previous work [11].

In order to reduce false corrosion detection, intra-modal fusion was performed to generate a final corrosion mask for each test panel based on combining the respective RX and the fuzzy ratio confidence maps using an iterative maximum-likelihood weighted-average scheme to generate a final corrosion mask, denoted as $V(x, y)$, with confidence value between 0 and 1, presented in detail in [11], and as shown in Equation 8.

Here, $\sigma_{\lambda(x,y)}$ and $\sigma_{R_\alpha(x,y)}$ denote the standard deviation for the RX and fuzzy ratio confidence maps.

$$V(x, y) = \frac{(\sigma_{\lambda(x,y)})^2 \lambda(x, y) + (\sigma_{R_\alpha(x,y)})^2 R_\alpha(x, y)}{(\sigma_{\lambda(x,y)})^2 + (\sigma_{R_\alpha(x,y)})^2} \quad (8)$$

4.1.2. Material Loss Estimation and Model Generation. At the non-zero confidence values in the final corrosion mask, i.e. $V(x, y) \neq 0$, a material loss estimation map was computed for each position (x, y) in the EC image $I(x, y)$ using interpolation of cluster distance from representative material loss signatures of 0%, 10%, 15% and 20% determined from the reference panel. The clustering technique, presented in detail in [12], can be mathematically represented as follows. For each image pixel, $I(x, y)$ where, $V(x, y) \neq 0$ the feature vector $F(x, y) = [V(x, y) \quad I_{norm}(x, y)]^T$ is formed, where $I_{norm}(x, y) = \frac{(I(x, y) - m_0)}{\sigma_0}$ and m_0 and σ_0 are the mean and standard deviation values of 0% (background/no-corrosion region) material loss from the reference panel. The Euclidean distance, denoted as $I_M(x, y)$ at the output of the Defect Detection, Characterization and Material Loss Estimation block in Figure 4.1, is computed between $F(x, y)$ and median feature vectors determined from the reference panel for the 0% ($F_0(x, y)$), 10% ($F_{10}(x, y)$), 15% ($F_{15}(x, y)$), and 20% ($F_{20}(x, y)$) material loss. Interpolation of the Euclidean distances found is used for estimating the material loss represented for an EC image I . The Image Analysis block in Figure 4.1 represents the steps for the corrosion detection, material loss estimation and the 3-D model generation.

A 3-D geometrical model for the riveted test panels was generated based on a localized non-overlapping window method using material loss estimate map of a given test panel. The method divided the height and width dimensions of the test panel into 10x10 (approximately 5 mm x 5 mm) non-overlapping windows. The average material loss over a non-overlapping window was estimated, and the 3-D center of mass was computed. A circular puck-like region was created with a thickness equal to the average material loss of the window, and area equal to the number of non-zero pixels (indicating corrosion and thus material loss) present in the window. This circular puck-like region was then centered within the non-overlapping window. This method was repeated for all non-overlapping windows in each panel. Such a 3-D model obtained for test panel 1 is shown in Figure 4.3, illustrated using ANSYS®. Note that the puck-like regions shown in Figure 4.3 are actually solid material loss regions, which could not be visually represented using the ANSYS® display tool. Also, note that the 3-D modeling method was developed based on constraints in ANSYS®, which is used for simulated fatigue loading of the test panels. Intra-modal data fusion was performed based on integrating the outputs of two anomaly detectors, RX and data-driven fuzzy logic, applied to the same EC images. The anomaly detectors use localized image analysis to find potential defects. This integrated approach was explored in previous research for corrosion detection [11] and material loss estimation [12], with experimental results showing enhanced corrosion detection and improved material loss estimation based on combining these methods as compared to using them individually.

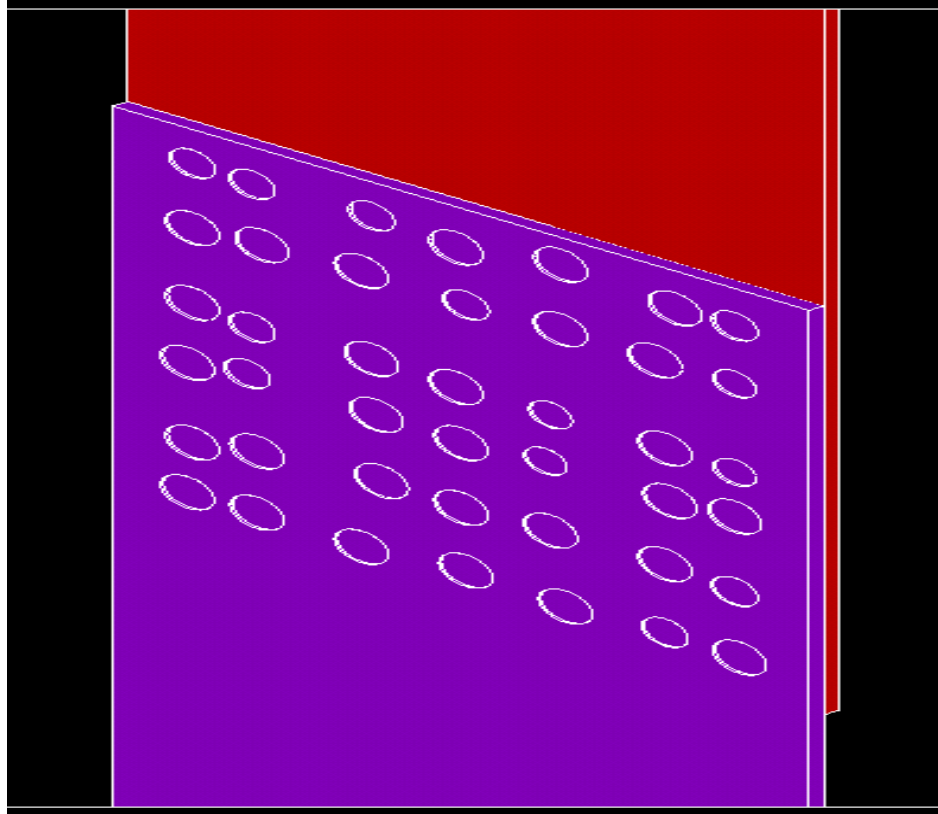


Figure 4.3: ANSYS generated model of test panel.

4.2. SIMULATED FATIGUE LOADING

As part of the structural analysis process, the 3-D model of each test panel was generated to facilitate simulated fatigue loading. The 3-D model was used as the input to ANSYS® to facilitate simulated structural and fatigue analysis of the test panel [19]. The physical location of all of the rivets were calculated and marked on the meshed structure for the test panel from ANSYS®, as shown in Figure 4.4. In the structural modeling of the panels and subsequent simulated fatigue evaluation, the riveted regions were marked and the displacements with respect to each layer were set to zero to simulate the actual riveted structure of the panels. After the rivet regions were marked and their displacement set to zero for all directions at these regions, the boundary conditions applied to the panel,

during simulated loading (see Figure 4.5), were setup to simulate the mechanical loading, as shown in Figure 5.1, where the test panel was subjected to a fixed load on the top and an actuator at the bottom to facilitate the cyclical loading of the panel. As shown in Figure 4.5, the boundary condition applied to the top of the test panel involved restricting the displacement for this particular region to zero (e.g., see circle 1 in Figure 4.5) for all the directions (X, Y and Z). The bottom of the test panel was subjected to a cyclical load to simulate the actuator cyclical loading step. In addition, the sides of the test panel (e.g., see circles 2-5 in Figure 4.5) were restricted to displacement only in the Y direction to account for the movement of the test panel in the $\pm Y$ direction during the simulated fatigue loading process.

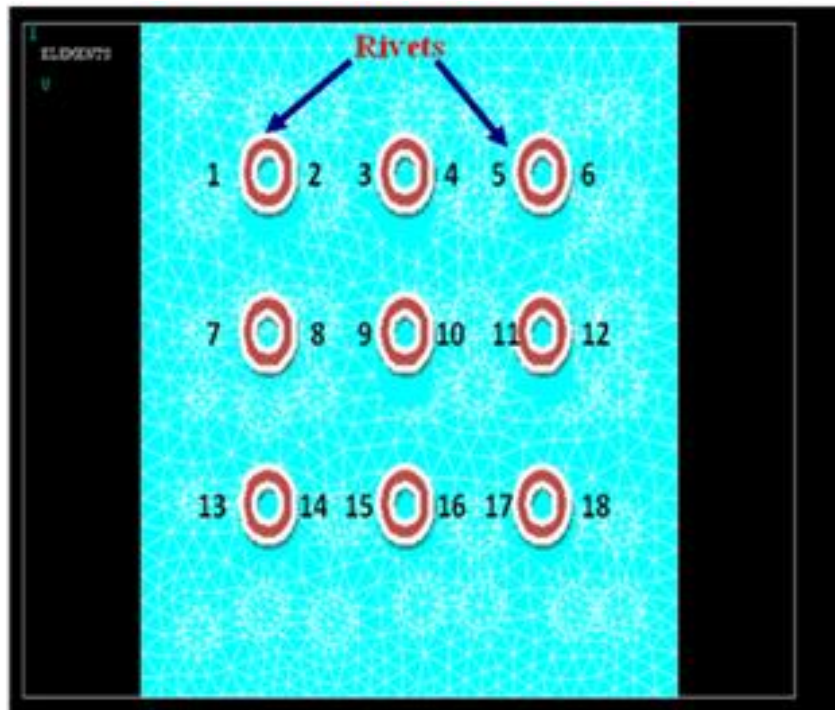


Figure 4.4: Location key for crack initiation sites in test panels.

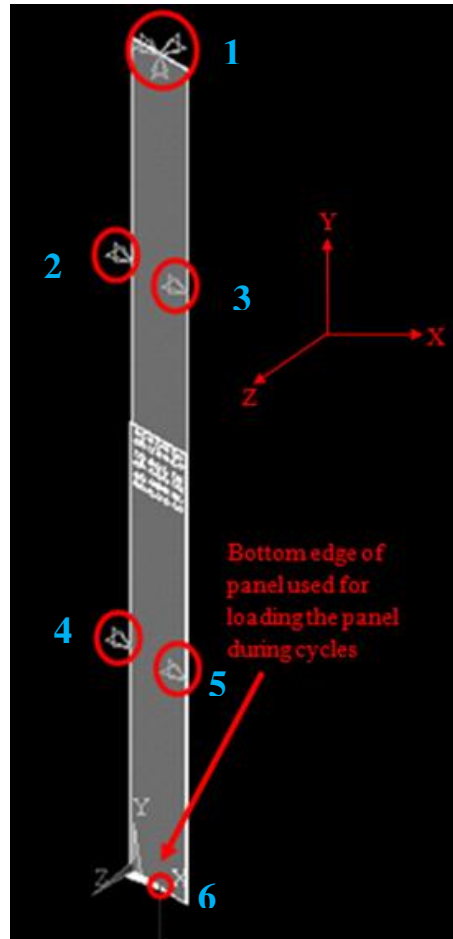


Figure 4.5: Boundary conditions defined for simulated fatigue loading.

Once the loading of the panel had been performed, the Fatigue Analysis module in ANSYS® was used to study the fatigue life of the panel. A standard and systematic approach was used for the fatigue analysis of the test panel's 3-D model [20]. Alternating Stress Intensity (Alt-SI) values for all the nodes of the meshed panel were determined using ANSYS® as a measure of the vulnerability of a specific node/region to failure [21]. The Alt-SI values were thresholded, determined empirically from the mechanical loading data, to determine positions (nodes) on the test panel for potential vulnerability.

These positions of potential vulnerability were compared with the fatigue positions found from mechanical loading of the actual test panel. The Compare Actual and Simulated Fatigue Locations block in Figure 4.1 represents this validation step in the process.

5. EXPERIMENTS PERFORMED

The comprehensive structural analysis process was applied to the five test panels based on three rounds of EC data collection for test panel analysis, including: 1) EC data collection prior to mechanical loading of the test panels, 2) EC data collection after the first round of mechanical loading, and 3) EC data collection after the second round of mechanical loading. After each round of EC data collection, mechanical loading (cyclical loading) of each test panel was performed. From Figure 4.1, for each round of EC data collection, image analysis, simulated structural analysis, and comparison between simulated and mechanical loading was conducted for each test panel.

For comparison with simulated fatigue loading analysis performed for the test panels, the following procedure was used for mechanical loading of the test panels (see Actual Fatigue Testing Block in Figure 4.1). The panels were cyclically loaded between 1.04 kN and 10.4 kN (load ratio, $R = P_{min}/P_{max} = 0.1$) at a frequency of 4 Hz, using the setup shown in Figure 5.1. The calculated stress at the maximum load (10.4 kN) in the unriveted portions of the panels was 82.7 MPa. The first round of loading was intended to initiate small fatigue cracks, wherein, the test panels were cyclically loaded until a crack was noticed (see Table 6.1).

For the round 2 of loading, cracks were allowed to grow to twice their former lengths or if no cracks originated in the first round, cracks were initiated and allowed to grow to a length of ~3 mm. Round 3 of mechanical loading was conducted only on corroded test panels 1, 4 and the pristine panel. In round 3, the panels were tested until a

crack progressed to the panel edge or a crack progressed halfway to an adjacent rivet (a 9 mm crack length for both cases). The panels were photographed using a stereomicroscope to inspect and measure crack lengths after each round. These cracks were meant to simulate in-service flaws. Only test panels 1, 4 and the pristine were subjected to loading in the third round to provide additional data to compare with the comprehensive structural analysis methodology presented (see experimental results).

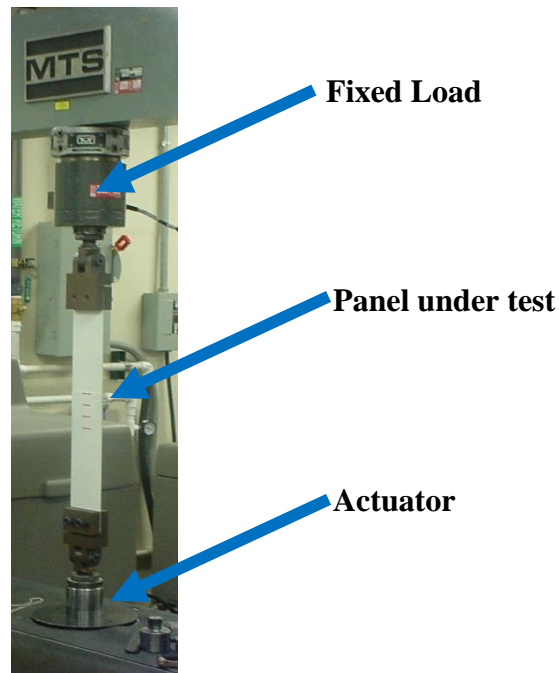


Figure 5.1: Cyclical fatigue loading setup.

Panels 2 and 3 were not subjected to round 3 of loading to allow comparison and analysis of the crack growth progression between round 2 and round 3 of loading between the test panels. The cyclic loop displacement data for the combined three rounds of fatigue loading in conjunction with visual observations to estimate the cyclic life at which crack initiation began (fatigue initiation life) was used to determine the fatigue propagation life [22]. For reference purposes, a schematic outlining the various rounds of

the NDE data collection and mechanical loading process is provided in Figure 5.2 for the different panels. Figure 4.4 gives the position labels, with respect to the rivet locations, for the test panels that were used for specifying crack locations identified from the mechanical (actual fatigue cracks) and simulated fatigue loading (positions of potential vulnerability) processes.

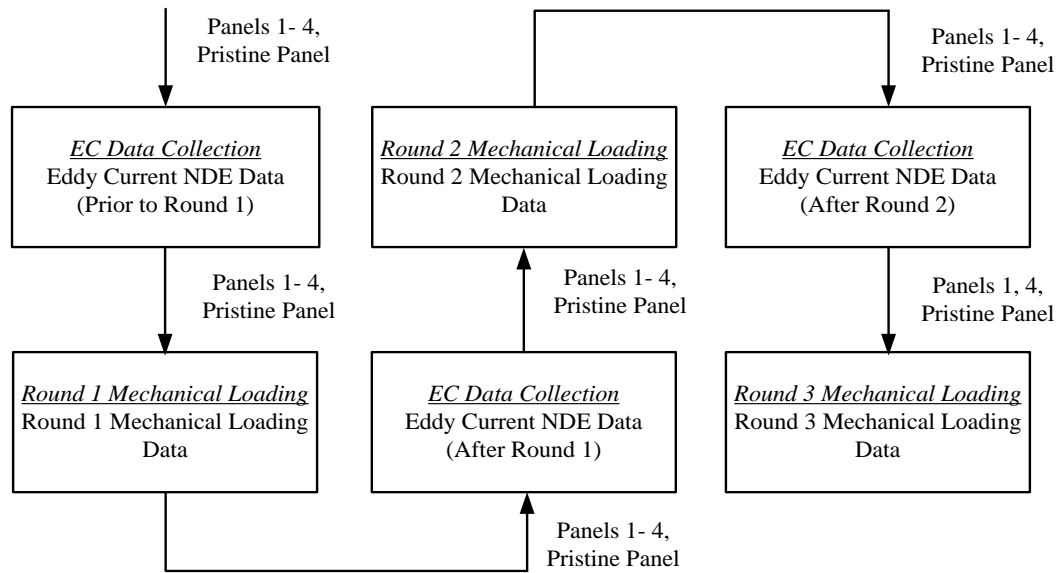


Figure 5.2: Data collection and mechanical testing rounds.

6. EXPERIMENTAL RESULTS

6.1. DATA FUSION RESULTS

Since the depth of penetration of the EC signal for producing the EC image is inversely proportional to the frequency of the EC signal, the lowest frequency EC images (1kHz) were used for data fusion algorithm development. Experimental results for the raw EC scan at 1kHz, corrosion detection confidence maps, and material loss estimation maps with the reference scale for material loss estimation between 0%-20% are presented for corroded test panel 1 prior to mechanical loading (Figure 6.1), after the round 1 of mechanical loading (Figure 6.2), and after round 2 of mechanical loading (Figure 6.3). Note that material loss did not exceed 20% in any test panel based on the corrosion inducement process described in section II.

Figure 4.3 shows the 3-D model used for simulated fatigue loading for panel 1 prior to round 1 of mechanical loading, which was generated from the material loss estimation map given in Figure 6.1c. Similar experimental results are given for the pristine panel prior to round 1 of mechanical loading (Figure 6.4), after round 1 of mechanical loading (Figure 6.5), and after round 2 of mechanical loading (Figure 6.6). The arrows in the images of the test panels, collected after rounds 1 and 2 of mechanical loading (Figures 13, 14, 16, and 17), indicate the location of produced cracks. A summary of crack locations will be presented and discussed later.

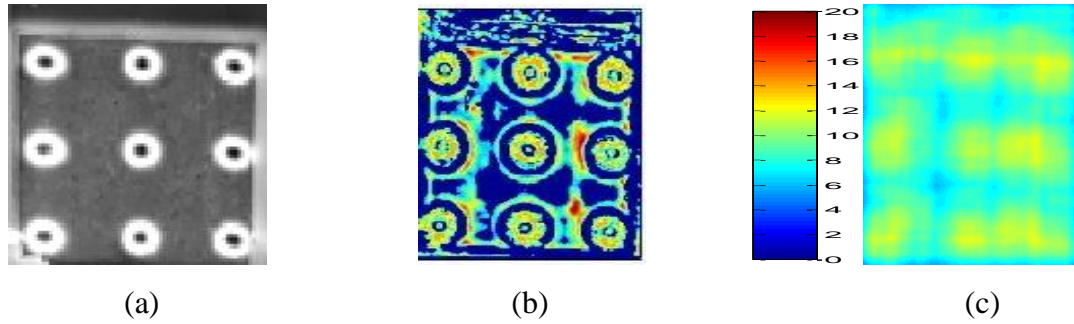


Figure 6.1: EC Data for test panel 1 prior to the first round of mechanical loading.
 (a) EC data for test panel 1, (b) corrosion detection, (c) material loss estimation map.

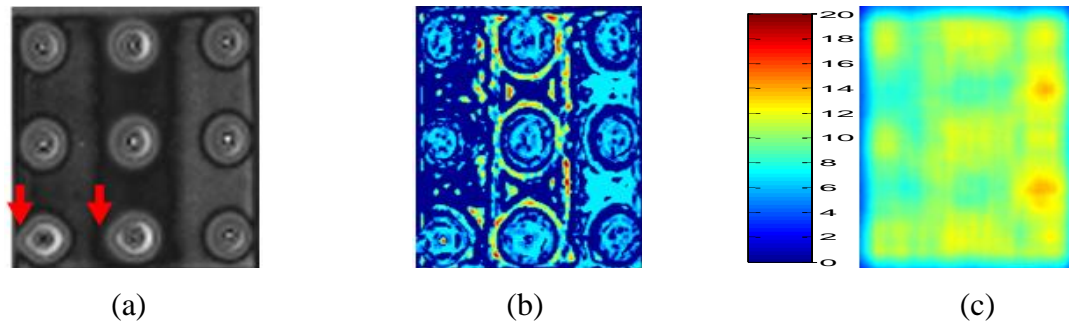


Figure 6.2: EC Data for test panel 1 after the first round of mechanical loading.
 (a) EC data for test panel 1 with locations of cracks shown, (b) corrosion detection confidence map, (c) material loss estimation map with reference scale.

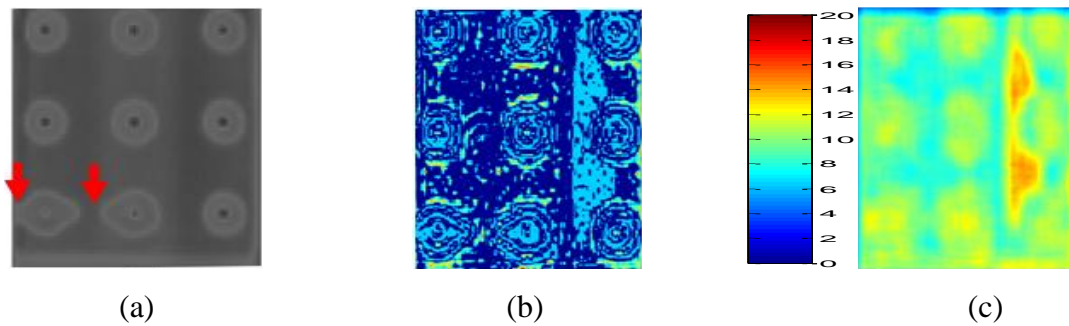


Figure 6.3: EC Data for test panel 1 after second round of mechanical loading.
 (a) EC data for test panel 1 with locations of cracks shown, (b) corrosion detection confidence map, (c) material loss estimation map with reference scale.

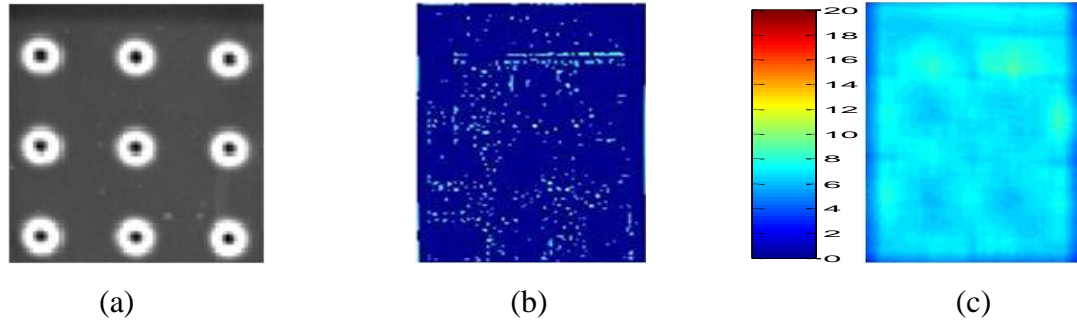


Figure 6.4: EC data for pristine panel prior to the first round of mechanical loading. (a) EC data for the pristine panel, (b) corrosion detection confidence map, (c) material loss estimation map with reference scale.

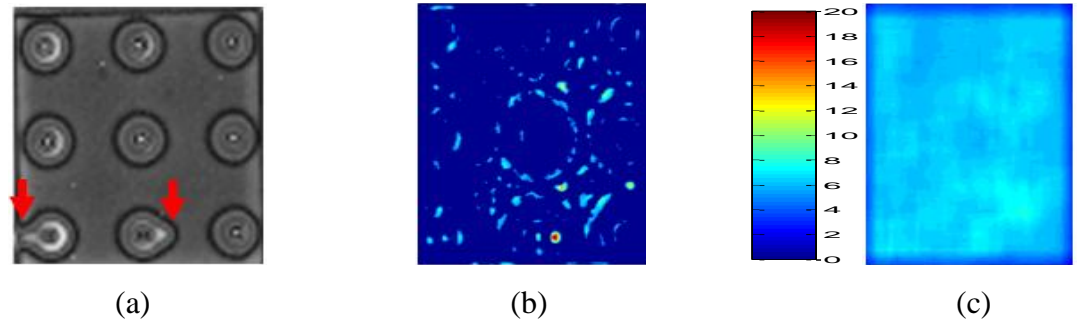


Figure 6.5: EC data for pristine panel after first round of mechanical loading. (a) EC data for the pristine panel with locations of cracks shown, (b) corrosion detection confidence map, (c) material loss estimation map with reference scale.

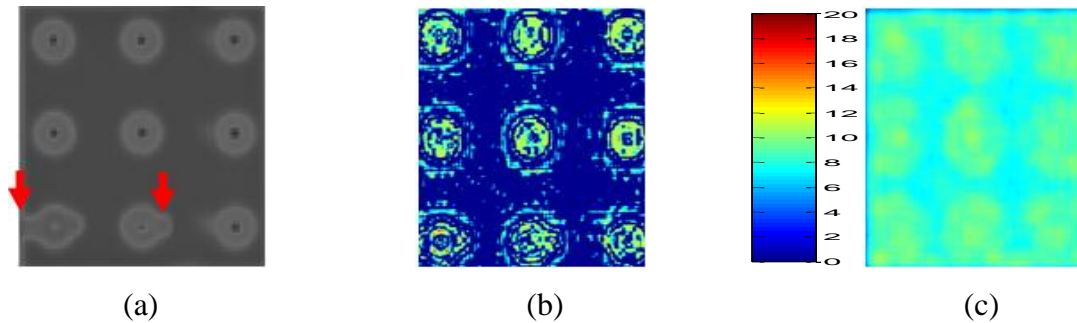


Figure 6.6: EC data for pristine panel after second round of mechanical loading. (a) EC data for the pristine panel with locations of cracks shown, (b) corrosion detection confidence map, (c) material loss estimation map with reference scale.

For this investigation, using the intra-modal data fusion approach, the average material loss over the non-zero corrosion areas in the material loss maps found approximately 13% for the corroded test panel 1.

6.2. MECHANICAL LOADING RESULTS FOR TEST PANELS

The mechanical loading results in the form of fatigue (crack) locations identified after each round of loading are summarized in Table 6.1. Experimentally, it was observed that all cracks developed on the lower row of rivets (locations 13-18), suggesting a higher stress concentration at this row. Photograph of an example crack, obtained using a stereomicroscope is shown in Figure 6.7, which was generated in panel 1 at location 13 (as shown in Figure 4.4) after round 2 of fatigue loading.

Table 6.1 : Crack locations summary after each round of mechanical loading.

Panel	Condition	Crack locations From Round 1	Cracks locations From Round 2	Cracks locations From Round 3
1	Corroded	13, 15	13, 15	13, 14, 15, 16, 18
2	Corroded	13	13, 14	-
3	Corroded	None	14	-
4	Corroded	None	14	13, 14
6	Pristine	13, 16	13, 14, 16	13, 14, 16

– denotes panels not subjected to mechanical loading.

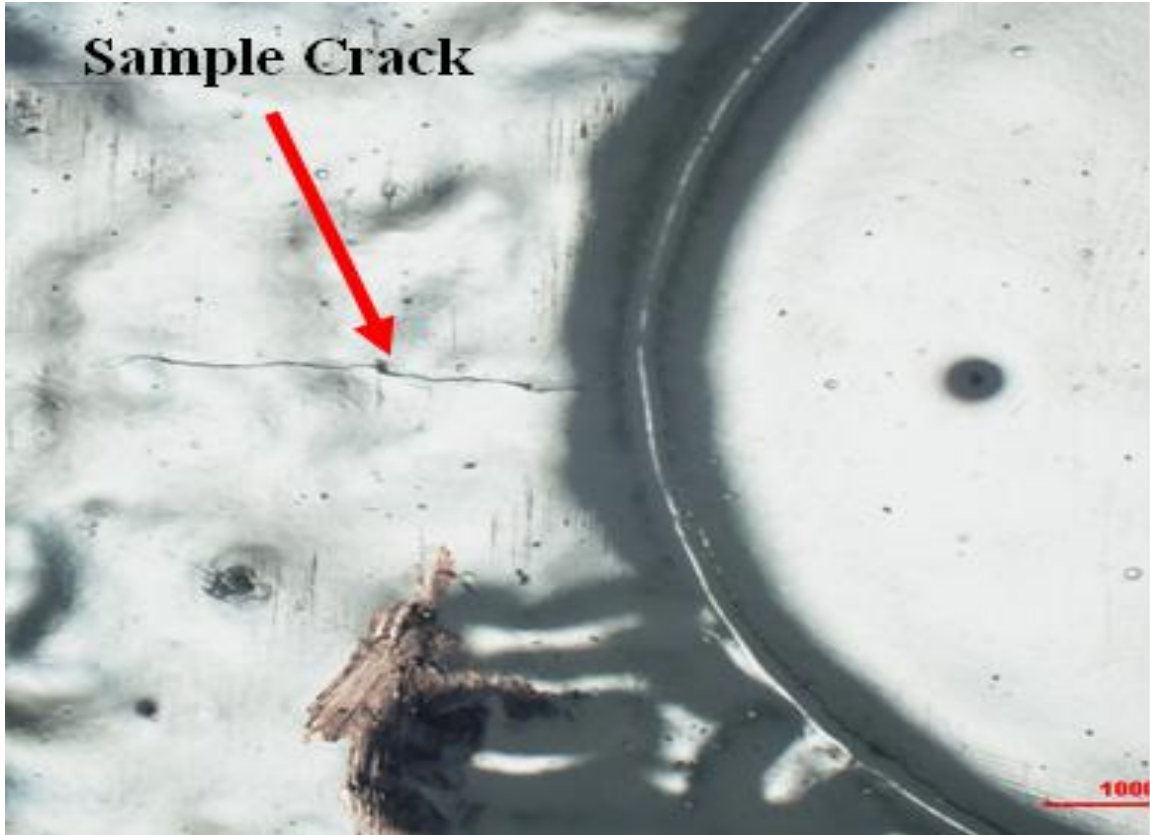


Figure 6.7: Developed crack in panel 1 after the second round of mechanical loading

A summary of the Alt-SI values for test panels 1-4 and the pristine panel, for the three rounds of EC data collection, are presented in Table 6.2. Note that round 1, round 2, and round 3 refer to the different rounds of mechanical loading, as designated in Table 6.1. From Table 6.2, Alt-SI values greater than 70 MPa (shown in **bold** in Table 6.2 as threshold values) were used for comparison of the selected areas given in Figure 6.1 to corresponding areas inspected in the mechanical loading process (given in Table 6.1). Considering test panels 1-4 and the pristine panel, there are a total of 90 critical locations (18 locations per panel). From Table 6.2, if a location (shown as Loc 1-Loc 18 in Table 6.2) in a test panel had an Alt-SI value greater than or equal to 70 MPa, the location was labeled as a crack site.

Table 6.2 : Summary of Alt-SI values for simulated fatigue loading.

		Panel 1 Alt-SI (MPa)	Panel 2 Alt-SI (MPa)	Panel 3 Alt-SI (MPa)	Panel 4 Alt-SI (MPa)	pristine Alt-SI (MPa)
Loc 1	Prior to Round 1	9.1	2.5	1.1	0.5	2.2
	After Round 1	1.7	0	0.4	1.3	1.5
	After Round 2	2	3.1	2.4	1.3	0.5
Loc 2	Prior to Round 1	5.8	0.5	0.6	0.4	0.8
	After Round 1	0.5	0.6	0.5	1.1	1
	After Round 2	1.3	0.8	1	0.8	0.4
Loc 3	Prior to Round 1	4.9	0.3	0.4	0.2	0.6
	After Round 1	0.1	0.1	0.2	0.4	0.7
	After Round 2	0.8	0.2	0.4	0.5	0.3
Loc 4	Prior to Round 1	2.8	0	0.2	0.2	0.4
	After Round 1	0.3	0	0.2	0.7	0.2
	After Round 2	1.2	0.1	0.7	0.4	0.2
Loc 5	Prior to Round 1	2.8	0.2	0.4	0.3	0.6
	After Round 1	0.5	0.8	0.1	0.6	0.6
	After Round 2	0.7	1.5	0.5	0.8	0.3
Loc 6	Prior to Round 1	0.9	2.4	1.5	0.6	3.6
	After Round 1	1	0	0	0	1.1
	After Round 2	3.4	0	4	2.2	0
Loc 7	Prior to Round 1	39.1	1.5	7.1	5.8	8.6
	After Round 1	9.9	0	0.9	7.1	5.9
	After Round 2	7.7	14.9	4.6	8.7	4.9
Loc 8	Prior to Round 1	28.8	2.5	2.7	2.5	4.6
	After Round 1	2.9	3.6	2.4	5.8	3.5
	After Round 2	2.6	8.8	2.7	10.2	1.9
Loc 9	Prior to Round 1	22.4	1.3	2.7	3.2	4.4
	After Round 1	3.2	1.5	1.1	4.7	1.5
	After Round 2	4.9	5.9	4.5	1.4	3.9
Loc 10	Prior to Round 1	12.9	2.6	2.6	2.3	3.1
	After Round 1	1.3	1.6	1	1.6	4.2
	After Round 2	2.6	4.9	3.5	3.3	2.4
Loc 11	Prior to Round 1	15.2	3.6	2.9	2.8	6
	After Round 1	3.1	4.3	3.2	4.9	5.6
	After Round 2	6.3	5.4	3.9	1.3	3.3
Loc 12	Prior to Round 1	7.1	9.7	8.8	0.5	7.6
	After Round 1	3.4	0	0	11.1	5.5
	After Round 2	12.9	8.9	8.7	9	7.5
Loc 13	Prior to Round 1	182.3	92.6	29.6	75.9	84.9
	After Round 1	95.3	0	0	39.6	50.9
	After Round 2	96.6	97.2	65.8	65.7	81.3
Loc 14	Prior to Round 1	99.4	16.9	50.5	82.5	69.5
	After Round 1	4.9	78.6	35.2	27.3	35
	After Round 2	13.2	18	14.8	45.5	6.3
Loc 15	Prior to Round 1	149.7	44.8	43.9	23.8	55.8
	After Round 1	34.7	9.7	6.1	15.5	34.7
	After Round 2	10.3	14.2	49.1	36.5	5.4
Loc 16	Prior to Round 1	119.1	60.9	19.5	83	21.4
	After Round 1	40.9	3.2	0	0	0.1
	After Round 2	33.8	0	56.3	9.3	0
Loc 17	Prior to Round 1	86.1	49.3	108.8	25.2	35.9
	After Round 1	19.6	4.9	11.3	22.1	13.4
	After Round 2	63.2	23.2	64.7	37.4	91.3
Loc 18	Prior to Round 1	125.6	50	104.2	9	108
	After Round 1	70.9	0	0	0	25
	After Round 2	125	0.2	63.2	59.7	89.2

Since during actual mechanical loading the maximum stress observed was 82.7 MPa, a value of 70 MPa, which is >10 MPa lower than the maximum stress observed, was chosen as the threshold value for selecting ‘potential’ vulnerable sites. True positive and true negative crack location detection rates were computed over the four corroded and one pristine test panels. A true positive crack location is correctly labeled if the location satisfies the Alt-SI threshold constraint, and the location in the test panel for the corresponding round of mechanical loading is designated as a crack site, as given in Table 6.1. A true negative location is labeled correctly if the location does not satisfy the Alt-SI threshold constraint, and the location in the test panel for the corresponding round of mechanical loading is not called a crack site, as shown in Table 6.1. False positives are the number of cracks that were erroneously detected across all the panels, and false negatives are the cases when the simulated fatigue loading method failed to predict a crack location that was actually a crack location obtained by the mechanical loading process.

Table 6.3 summarizes the locations of cracks predicted after each round of simulated loading. Table 6.4 summarizes the true positive and true negative detections based on comparing the simulated fatigue loading vulnerable locations to the locations of cracks detected from the corresponding mechanical loading. There are 18 total locations used for true positive and true negative scoring (from Figure 4.4). The number on the left side of the slash indicates the number of vulnerable locations (true positive) and the number of non-vulnerable locations (true negative) found from the simulated fatigue loading process, and the number to the right of the slash gives the correct number of true positive and true negative locations, as found from mechanical loading. The column

labels provide the different EC data corresponding to the times when simulated fatigue loading was performed (see Figure 5.2). The results in the column “Prior to Round 1” refer to the simulated fatigue loading results for the EC data collected prior to the first round of mechanical loading and are compared to the results from round 1 of mechanical loading. The simulated fatigue loading results for “After Round 1” (refers to the simulated fatigue testing results obtained from round 1 of mechanical loading) and “After Round 2” (refers to the simulated fatigue testing results obtained from round 2 of mechanical loading) are compared to the mechanical loading results from rounds 2 and 3, respectively.

Table 6.3 : Summary of crack locations identified.

Panel	Prior to Round 1	After Round 1	After Round 2
1	13, 14, 15, 16, 17, 18	13, 14, 15, 16, 17, 18	13, 14, 15, 16, 17, 18
2	13	13, 14	13, 14
3	17, 18	17, 18	17, 18
4	13, 14, 16	13, 14, 16	13, 14, 16
pristine	13, 18	13, 18	13, 17, 18

Table 6.4 : Accuracy of simulated loading vs. mechanical loading.

	Prior to Round 1 loading		After Round 1 loading		After Round 2 loading	
	True Positive	True Negative	True Positive	True Negative	True Positive	True Negative
Panel 1	2/2	12/16	2/2	12/16	5/5	12/13
Panel 2	1/1	17/17	2/2	16/16	-	-
Panel 3	0/0	16/18	0/1	15/17	-	-
Panel 4	0/0	15/18	1/1	15/17	2/2	15/16
Pristine	1/2	15/16	1/3	14/15	1/3	13/15

– denotes panels not mechanically loaded for Round 3.

The accuracy for each round of simulated fatigue loading was computed by calculating the sum of the number of detected true positives and true negatives divided by the total number of positions for each round of fatigue loading. Before round 1 of

mechanical loading, there were 79 locations correctly identified (4 true positive locations and 75 true negative locations), which yields a correct recognition rate of 87.8% using the structural analysis process. After round 1 and after round 2 of mechanical loading, the corresponding correct recognition rates are 86.7% and 88.9%, respectively. Note that the results after round 2 only include panel 1, panel 4, and the pristine panel because only these panels were mechanically loaded in round 3. Overall, the structural analysis process yielded a consistent correlation for identifying vulnerable locations compared to visual inspection from mechanical loading.

7. DISCUSSION

In this research, the average material loss obtained using the data fusion based approach for test panel 1 was approximately 13%. Based on a report obtained from the Air Force Research Laboratory, where the test panels were physically analyzed for corrosion, the material loss in test panel 1 was measured to be approximately 16-17% [12].

Comparing Figures 6.1a-6.3a with Figures 6.4a-6.6a, the presence of noise in the EC images provide the appearance of material loss, very minor compared to the corroded panels, as observed in Figures 6.1c-6.3c and Figures 6.4c-6.6c. In general, the regions of corrosion in the decision level image for panels 1-4 could be detected and are significantly more than the pristine case resulting in significantly higher material loss as compared to the pristine case.

Evaluating the simulated fatigue loading and mechanical loading results, several observations can be made. Test panel 4 did not produce any cracks in round 1 of the mechanical loading. However, the first round of the simulated fatigue loading process predicted cracks at locations 13, 14 and 16 (from Table 6.2). From rounds 2 and 3 of the mechanical loading, cracks were found in locations 13 and 14 (Table 6.2). Experimentally, it was observed that all cracks developed on the lower row of rivets (locations 13-18), suggesting a higher stress concentration at this row. In some instances, visible cracks were not produced and observed. However, it is possible that small cracks may have initiated at the mating surfaces of each panel and, thus, be hidden from view

and below the EC crack detectability limit. Based on the simulated fatigue evaluation from ANSYS®, it was also found that the critical locations of the panel (locations 13-18) have significantly higher values as compared to the non-critical locations (1-12) from Figure 6.1. Furthermore, mechanical loading results showed no cracks or vulnerable areas in locations 1-12. These results show the predictive capability of this structural analysis approach for evaluation of these test panels.

As shown in Table 6.1, most of the cracks originated in the left edge of the panel (locations 13/14). This could be attributed to the slight bending of the panels towards the left (side towards locations 1, 7, and 13 from Figure 4.4) during the mechanical loading process resulting in higher stress intensities towards the left side of the panels. However, the simulated fatigue loading results showed that stress intensities were relatively balanced with respect to both the left and right sides of the test panels resulting in the tendency of the panels to fail along both sides (locations 13/14 and 17/18). In the mechanical loading results, it was observed that after the round 1 of testing, the pristine panel had cracked, whereas corroded panels 3 and 4 (see Table 6.1) did not, when both panels were subjected to the same number of fatigue cycles. This result highlights the utility of the structural analysis process in detecting potential vulnerable areas in unexpected cases such as a pristine panel. This result also highlights a possible reason for the true negative detections in the pristine panel for the different rounds of simulated fatigue loading.

The individual plates of the test panels received selective corrosion treatments such that only the surfaces around the center column of rivets received corrosion on both mating surfaces. Subsequent riveting may have resulted in unequal distribution of

residual stresses and possibly imparted compressive residual stresses in locations prone to fatigue cracking, such as locations 13 and 14. The ANSYS® FEM technique is limited in its ability to apply such residual stresses in the panels for the fatigue life evaluation.

8. CONCLUSIONS

To summarize, this study involved the design and implementation of five aluminum lap-joint mimics, representative of aircraft lap-joints, and the process of NDE data acquisition, defect detection and characterization involving material loss estimation and three-dimensional structural model generation for finite-element modeling based simulated fatigue analysis and comparison with actual mechanical loading data. The comprehensive structural analysis process presented in this paper (see Figure 4.1) provided the basis for computer-assisted analysis of a test structure based on acquired NDE data. Each step of the comprehensive structural analysis process was considered from the perspectives of how: 1) conventional NDE methods are used to collect data to extract structural information from a test structure and 2) the development and implementation of algorithmic approaches that can be used in each step of the process to characterize, model, and FEM-based simulated fatiguing for direct comparison with actual tensile fatiguing and manual inspection of the test structure (lap-joint mimic).

It should be noted that the algorithms for automated defect detection, defect characterization, material loss estimation, modeling, and FEM-based structural analysis of the test panel are not specifically required for the comprehensive structural analysis process to work. One of the primary considerations for the algorithms referenced and presented for defect detection, defect characterization, and material loss estimation is that there was limited data available for training the algorithms. A reference test panel [12] was used for providing corrosion and different levels of material loss that for algorithms

developed/presented in this paper. The algorithmic approaches referenced [11] and presented in this paper have been experimentally shown to yield results [23] which have correlated reasonably well with actual inspection results of the test panels [12]. Other algorithms/techniques could readily be used here.

However, the comprehensive structural analysis process attempts to address practical considerations in presenting an original computer-assisted testing method from NDE data collection through FEM-based structural analysis that would reflect actual analysis of a test structure. Conventional NDE methods provide data for a test structure which can be used to visualize the presence of defects such as corrosion. Current NDE methods and analysis techniques fall short in being able to automatically use NDE data to estimate material loss, quantize the regions of material loss, and generate a model of a test structure for FEM-based structural analysis. The comprehensive structural analysis process provides for these automated steps and gives areas of potential structural vulnerability in the test structure as the output of the process. The positions of potential vulnerability in the test structure can be validated by direct comparison to actual inspection results, which was done in this paper with accuracy as high as 88.9%.

Future research would involve extending the single NDE modality approach used in this investigation to a multi-modal data fusion approach to better represent surface and subsurface information in defect detection and characterization and material loss estimation for the ultimate goal of generating more accurate 3-D models to be used for simulated fatigue loading.

ACKNOWLEDGEMENTS

This work was supported in part by the Air Force Research Laboratory (AFRL) under contract number FA8650-04-C-5704 in conjunction with the Center for Aerospace Manufacturing Technologies (CAMT).

REFERENCES

1. E. Herzberg. The annual cost of corrosion for DoD. Presented at the Department of Defense Corrosion Conference, Washington, DC (2009)
2. M. Brassard, A. Chahbaz, and A. Pelletier. Combined NDT inspection techniques for corrosion detection of aircraft structures. In 15th World Congress on Nondestructive Testing. Roma, Italy (October 2000)
3. P. J. Shull. Nondestructive Evaluation: Theory Techniques and Applications. Marcel Dekker Incorporated, New York, NY (2002)
4. D.S. Forsyth and B.A. Lepine. Development and verification of NDI for corrosion detection and quantification in airframe structures. AIP Conference Proceedings 615:1787 (2002)
5. N.C. Bellinger, J.P. Komorowski, and T.J. Benak. Residual life predictions of corroded fuselage lap joints. International Journal of Fatigue 23:349-356 (2001)
6. Z. Liu, D. S. Forsyth, J.P. Komorowski, K. Hanasaki, and T. Kirubarajan. Survey: State of the Art in NDE Data Fusion Techniques. IEEE Trans. Instrum. Meas. 56:2435-2451 (2007)
7. Z. Liu, K. Tsukada, and K. Hanasaki. One-dimensional eddy current multifrequency data fusion: A multiresolution analysis approach. Insight 40(4):286–289 (1998)
8. X.E. Gros, P. Strachan, and D.W. Lowden. Theory and implementation of NDT data fusion. Research in Nondestructive Evaluation 6:227-236 (1995)
9. X.E. Gros, Z. Liu, K. Tsukada, and K. Hanasaki. Experimenting with pixel level NDT data fusion techniques. IEEE Trans. Instrum. Meas. 49(5):1083–1090 (2000)
10. K. Sun, S.S. Udpa, L. Udpa, T. Xue, and W. Lord. Registration issues in the fusion of eddy current and ultrasonic NDE data using Q-transforms in Review of Progress in QNDE, vol. 15, pp. 813-820, D. O. Thompson and D. E. Chimenti, eds. Plenum, New York (1996)

11. K. Gupta, R. J. Stanley, M. T. Ghasr, S. Kharkovsky, R. Zoughi, and G. Steffes. Fusion of Multimodal NDE Data for Improved Corrosion Detection. In 3rd International Conference on Electromagnetic Near-Field Characterization and Imaging, pp. 407-412. St. Louis, MO (2007)
12. S. De, K. Gupta, R.J. Stanley, G. Steffes, D. Palmer, and R. Zoughi. A data fusion based approach for evaluation of material loss in corroded aluminum panels. In ITSC '09. 12th International IEEE Conference on Intelligent Transportation Systems, pp. 1-6. St. Louis, MO (2009)
13. T. Doppke, A. Bryant. The Salt Spray Test: Past, Present, and Future. In Proc. of the 2nd Automotive Corrosion Prevention Conference, pp 57-72 (1983).
14. Henkel Surface Technologies. TURCO 4215 NC-LT. Technical Process Bulletin (2009).
[[https://tds.us.henkel.com//NA/UT/HNAUTTDS.nsf/web/51A97021F65F4DE3852571ED005BBF7B/\\$File/TURCO%204215%20NC-LT-EN.pdf](https://tds.us.henkel.com//NA/UT/HNAUTTDS.nsf/web/51A97021F65F4DE3852571ED005BBF7B/$File/TURCO%204215%20NC-LT-EN.pdf)]
15. Z. Liu, K. Tsukada, K. Hanasaki and M. Kurisu. Two-dimensional eddy current signal enhancement via multifrequency data fusion. Research in Nondestructive Evaluation 11:165-177 (1999)
16. E.M. Franklin. Eddy-current inspection. Materials Evaluation 40:1008-1010 (1982)
17. D. L. DuQuesnay, P.R. Underhill, and H.J. Britt. Fatigue crack growth from corrosion damage in 7075-T6511 aluminium alloy under aircraft loading. International Journal of Fatigue 25:371-377 (2003)
18. S. Agarwal, P. Sriram, P.P. Palit, and O.R. Mitchell. Algorithms for IR imagery based airborne landmine and minefield detection. Detection and Remediation Technologies for Mines and Minelike Targets VI, Proc. SPIE, vol. 4394, pp. 284-295 (2001)
19. ANSYS Inc. ANSYS Structural Analysis Guide (2007)
20. D. Roylance, 3.11 Mechanics of Materials, Fall 1999. (Massachusetts Institute of Technology: MIT OpenCourseWare), <http://ocw.mit.edu> (Accessed 17 Sep, 2010). License: Creative Commons BY-NC-SA.
21. C. Nyquist, K. Haghighi. ANSYS Tutorial: Performing a Fatigue Analysis. Purdue University (2007).

22. S. Faanes. Inclined cracks in fretting fatigue. *Engineering Fracture Mechanics* 25(1):71-82 (1995)
23. K. Gupta, M.T. Ghasr, S. Kharkivskiy, R. Zoughi, R.J. Stanley, A. Padwal, M. OKeefe, D. Palmer, J. Blackshire, G. Steffes, and N. Wood. Fusion of Microwave and Eddy Current Data for a Multi-Modal Approach in Evaluating Corrosion Under Paint and in Lap Joints. *AIP Conference Proceedings* 894:611-618 (2007).

II. A COMPREHENSIVE MULTI-MODAL NDE DATA FUSION APPROACH FOR FAILURE ASSESSMENT IN AIRCRAFT LAP-JOINT MIMICS

Soumya De^a, Kapil Gupta^b, R. Joe Stanley^a, Mohammad T. Ghasr^a, Reza Zoughi^a, Kenneth Doering^c, David Van Aken^c, Gary Steffes^d, Matt O’Keefe^c, Donald Palmer^e

^aDepartment of Electrical and Computer Engineering, Missouri University of Science and Technology (S&T), Rolla, MO, 65409, USA

^bCoherix Inc., Ann Arbor, MI, 66062, USA

^cDepartment of Materials Science & Engineering, Missouri University of Science and Technology (S&T), Rolla, MO, 65409, USA

^dAir Force Research Laboratory (AFRL), Wright Patterson, OH 45433, USA

^eThe Boeing Company, Boeing Phantom Works, St. Louis, MO 63166, USA

ABSTRACT

In this research, a comprehensive multi-modal structural analysis process is presented that includes intra- and inter-modal NDE data fusion based on eddy current (EC), millimeter wave (MW), and ultrasonic (UT) data obtained from five lap-joint mimic test panels. The process includes defect detection, defect characterization and Finite Element Modeling-based simulated fatigue loading for structural analysis. The multi-modal structural analysis process is evaluated using four test panels with corroded patches at different layers of the lap-joints and one painted pristine panel used as a reference. Different NDE modality combinations are examined for test panel modeling, including: 1) EC, 2) UT, 3) MW, 4) EC and UT, 5) EC and MW, and 6) EC, UT, and MW. Experiments are performed to compare the simulated fatigue loading and the mechanical loading results to find susceptible-to-failure areas in the test panels. Experimental results showed that the EC and UT modality combination yielded a correct vulnerable (crack) location recognition rate of 98.8%, an improvement of 14.7% over any individual modality, demonstrating the potential for multi-modal data fusion for characterizing corrosion and defects.

1. INTRODUCTION

Corrosion-induced aircraft structural degradation is the subject of constant attention and concern for those involved in maintaining operational integrity and safety of aging aircraft fleets (i.e., commercial, cargo and military). Hidden corrosion under paint or in between various lap-joint layers tends to facilitate early fatigue crack initiation and enhanced crack growth when the aircraft experiences in-service loadings [1, 2]. To this end, the U.S. Air Force has proposed a “predict and manage” scheme for corrosion management as a deviation from the current “find-it and fix-it” approach [3].

To characterize and evaluate corrosion in aging aircraft structures many nondestructive evaluation (NDE) methods have been explored in the past and are currently being implemented [4-11]. In general, NDE methods may provide qualitative information (e.g., corroded or not) and/or quantitative evaluation (e.g., corrosion thickness/severity, spatial extent, etc.) about a structure. Also, each NDE method used for corrosion detection has its own limitations and advantageous features as outlined in detail in [8]. Some of the more prominent methods for aircraft structural integrity analysis, and more specifically for corrosion evaluation, include eddy current (EC), ultrasonic (UT), optical, thermography, radiography and millimeter wave (MW) methods [8].

EC inspection systems are small, portable and inexpensive, and are capable of providing effective metal loss information. Both pulsed and conventional EC NDE methods have been used for corrosion detection in aging aircrafts [5-7]. Conventional EC methods can be used to provide structural metal loss/thinning information, while pulsed

EC methods provide material loss depth information as well [10]. Ultrasonic methods are effective inspection tools for a number of NDE applications and employ high frequency sound waves (with relatively short wavelengths in the material). Ultrasonic signal amplitude, frequency or time of flight (ToF) may be used to detect and evaluate a number of defect characteristics. Ultrasonic methods can be limited in its capability to distinguish features in subsequent structural layers underneath the surface layer. This is due to the lack of an effective sound coupling media between layers. In contrast, EC is not limited by these constraints, has a relatively high signal-to-noise ratio, and is able to provide depth information as well [4,10].

Millimeter wave NDE methods offer several advantages for detection and evaluation of corrosion under paint and thick composite laminate coatings [9,12,13]. Millimeter wave signals can penetrate inside of low-loss dielectric materials, and are also sensitive to changes associated with dielectric properties and boundary interfaces, which make them very attractive for detecting the presence of a corrosion layer under paint. Near-field MW NDE techniques, employing open-ended rectangular waveguide probes, have been successfully used to detect the presence of corrosion and corrosion precursor pitting under paint and primer in both steel and aluminum substrates [9,12-17].

However, MW signals do not penetrate inside of electrically conducting materials such as metals and can therefore only provide information (i.e., presence, thickness, etc.) about surface flaws such as pitting and corrosion under paint or dielectric coatings. Consequently, as it relates to the objectives of this investigation, other methods must complement MW methods for detecting corrosion in between the various layers of the lap joints.

Results of previous research have clearly demonstrated that an effective way to overcome the limitations of each NDE method is to properly integrate the results of several of these methods using multi-modal data fusion techniques. In this way the complementary information across the various NDE modalities can be effectively combined to provide significantly more information than any one method individually, resulting in substantial increase in measurement reliability. Using complementary information, data fusion techniques have the potential to provide a much better and more comprehensive understanding of the physical process, which in this case is corrosion induced fatigue damage, as compared to using one sensor data or using multiple sensors in an independent manner [4, 8, 10, 18]. Data fusion provides a mathematical framework to enhance the data analysis process based on data collected from a single modality and processed using multiple methods (intra-modality fusion) and/or multiple modalities processed using one or more techniques (inter-modality fusion). The context for data fusion in this investigation is to “measure” (i.e., assess and predict) corrosion-induced crack locations in aircraft lap joints. Pixel level multi-modal NDE data fusion techniques have been applied to aerospace structures using probabilistic and wavelet methods [19, 20], and simple evidence theory as in [21] for enhanced detection of defects. NDE data fusion works fusing EC and UT measurement data obtained from aerospace structures have also been performed in [19, 22-25].

This investigation extends the realm and results of previous research (outlined above) to provide a semi-automated end-to-end system, which accommodates different combinations of NDE data collection sources, for failure assessment of aircraft structures. Existing techniques which use data fusion, combine multi-modal NDE data to enhance

test structure visualization for analysis. The comprehensive structural analysis process developed in this research advances the traditional/existing data fusion techniques in the context that data fusion is performed and utilized for automated structural analysis. It also facilitates predictive analysis by automatically generating a geometrical model of a test panel based on quantized areas of material loss for FEM-based simulated fatigue loading. This in turn could be used for comparison to mechanical loading (actual fatigue loading) to identify areas of vulnerability. In this investigation, a novel end-to-end system has been developed that has the potential to integrate data obtained from multiple NDE modalities for aircraft structures in the form of data fusion. Subsequently, it uses this fused data for structural analysis and evaluation using a Finite Element Modeling (FEM) based tool (ANSYS). The flowchart of this multi-modal structural analysis process is shown in Figure 1.1 [11].

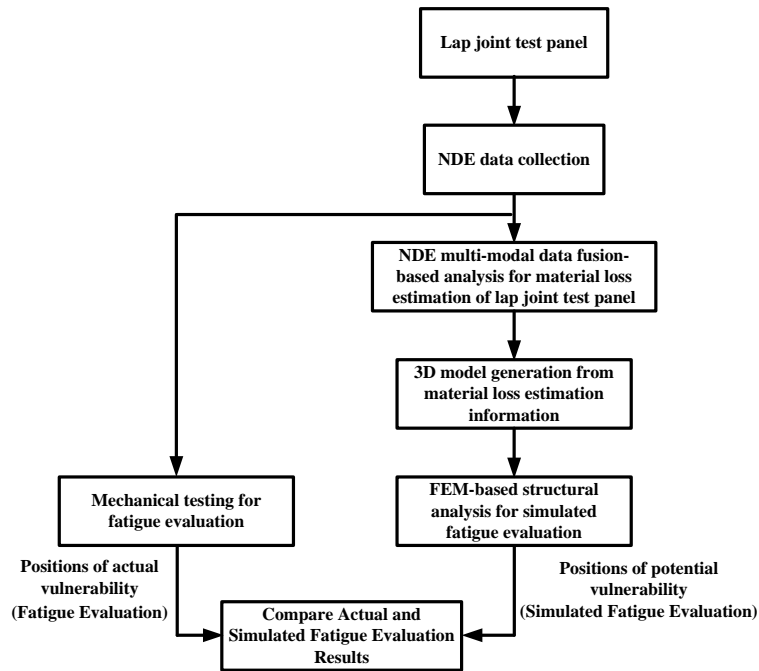


Figure 1.1: Flowchart of the structural analysis process.

As shown in Figure 1.1, the comprehensive structural analysis process consists of the following seven steps after the lap-joint mimic test panel has been prepared:

- 1) Collection of NDE data for the test panel,
- 2) Using NDE data to perform data fusion-based corrosion detection and material loss estimation,
- 3) Generating a 3-D model of the test panel from the material loss estimation information,
- 4) Simulating fatigue loading of the test panel for structural analysis using FEM-based evaluation of the 3-D model for a designated number of cycles,
- 5) Performing mechanical loading of the test panel for a designated number of cycles,
- 6) Comparing simulated fatigue loading and mechanical loading results of test panels to identify areas of vulnerability,
- 7) Repeating steps 2-6 after comparing mechanical and simulated fatigue loading results for validation of the process.

The comprehensive structural analysis method provided the basis for computer-assisted analysis of a test panel based on acquired NDE data. Each step of the comprehensive structural analysis process was considered with a view to: 1) understand how conventional NDE methods could be used to collect data to extract structural information from a test structure, and 2) develop and implement algorithmic approaches that can be used in each step of the process to characterize, model, and simulate the fatiguing process for direct comparison with mechanical loading.

2. METHODOLOGY

In this section, the details of each of the seven steps outlined above (also see Figure 1.1) are presented.

2.1. AIRCRAFT LAP-JOINT MIMIC PANEL DESCRIPTION

A set of five aluminum lap-joint mimic test panels, four of which were “corroded” and one “pristine”, were used. These panels were constructed from ten Al 2024-T3 aluminum plates with dimensions of 254 mm (10") by 76.2 mm (3") and with a nominal thickness of 1.6 mm (0.063"). Each of the five lap-joint mimics was created from two stacked and riveted aluminum plates. Prior to stacking and riveting, the aluminum plates of the corroded test panels were subjected to accelerated corrosion in a salt fog chamber following the ASTM B117 test standard, which is a widely used technique for evaluating relative corrosion in various metals and/or coatings. The salt spray chambers designed according to this standard are automated to maintain a specified environment within the chamber [26]. Figure 2.1 shows one of each set of the painted pristine and corroded panels, with the close-up views showing the riveted regions in each of the panels.

2.2. NDE DATA COLLECTION

2.2.1. Eddy Current Data. To evaluate the properties of corrosion/defects in the test panels, EC measurements were performed on each of the panels for two rounds of

data collection at the Air Force Research Laboratory, OH, where images of the test panels were obtained from the painted side only corresponding to the only accessible side on an aircraft. The EC data collection was performed using the Boeing MAUS[®] (Mobile Automated Scanner) raster scanning system, with phase lag adjustment to ascertain corrosion depth. The imaged area covered the corroded regions containing the rivets, as shown in Figure 2.1.

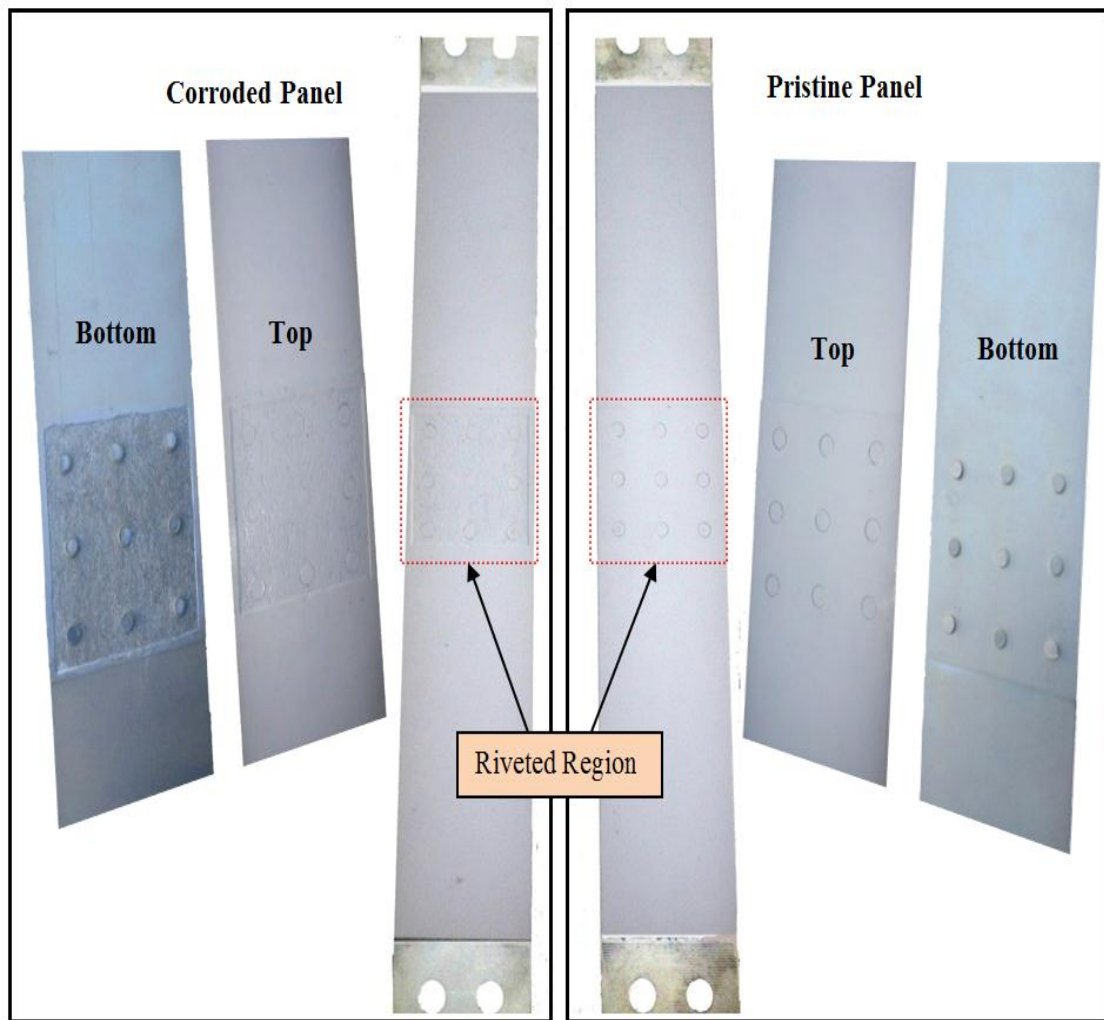


Figure 2.1: Picture of painted pristine and corroded panels with riveted area shown.

Figure 2.2a shows EC image of a reference panel with known amounts of metal thinning corresponding to the shown corrosion levels. They were primarily used for

calibration purposes for corrosion detection and material loss estimation algorithm development. In Figure 2.2a, physically, for example, the pixels contained in the 15% corrosion section represent 15% deviation from the pixels contained in the 0% corrosion region. Figure 2.2b shows the raster image of one of the corroded test panels.

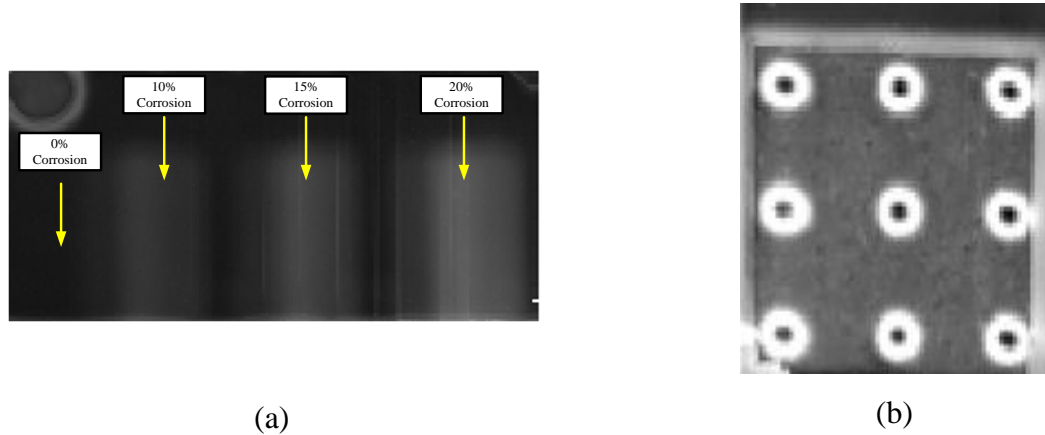


Figure 2.2: Reference panel and EC image of corroded test panel.
 (a) Reference panel used for calibration purposes, with areas of various levels of metal thinning; and (b) EC image of a corroded test panel.

2.2.2. Ultrasonics Data. Ultrasonics time-of-flight NDE data was also collected at the Air Force Research Laboratory, OH, in a similar fashion as those for EC. Measurements were conducted at a frequency of 15 MHz and raster scanned images were produced [27]. The Ultrasonic time-of-flight between the front wall and back wall signal responses were used to determine the remaining material in the panels. Since we know the velocity of sound in aluminum, one can calculate the remaining material (non-corroded material) in the test panel. We also know the thickness of the plates prior to their salt fog exposure, and this was verified by the reference standard, representing 0%, 10%, 15%, and 20% material loss, as shown in Figure 2.3a (similar to the EC reference

panel image). Figure 2.3b shows a UT image of the riveted region, for the same test panel whose EC data was shown in Figure 2.2b.

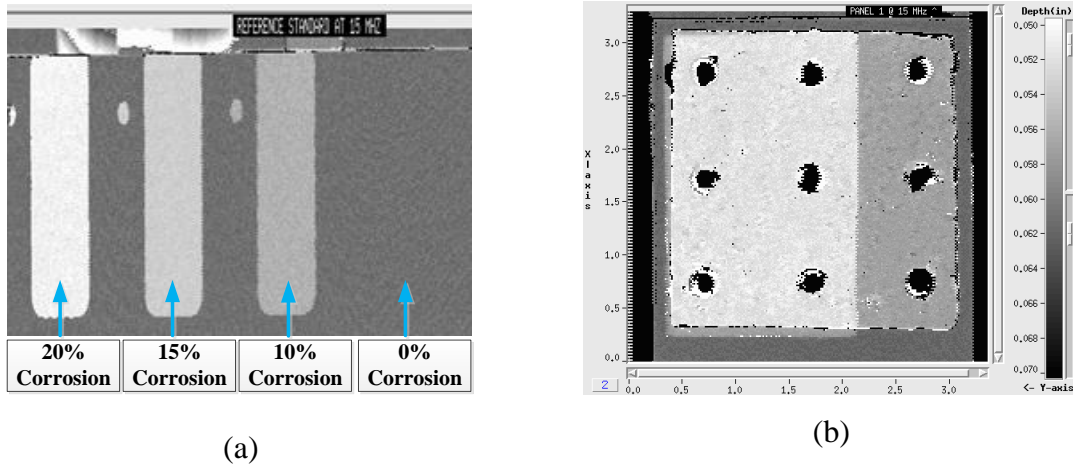


Figure 2.3: Reference panel and UT image of corroded test panel.
 (a) Reference panel used for calibration purposes with areas of various levels of removed metal; and (b) UT image of corroded test panel taken at 15 MHz.

2.2.3. Millimeter Wave Data. The MW images were produced at the Applied Microwave Nondestructive Testing Laboratory, Missouri S&T. A MW reflectometer, operating at 67 GHz (V-band) consisting of a continuous-wave (CW) MW source (Gunn oscillator), a magic-tee (i.e., power splitter and combiner), and a diode detector was used, as shown in Figure 2.4. Through extensive measurements and considering trade-offs such as detection capability vs. spatial resolution, V-band (50-75 GHz) was shown to be a suitable frequency range for this investigation. As shown in Figure 2.4, a portion of the incident signal irradiates the sample through an open-ended rectangular waveguide probe with dimensions of 3.75 mm by 1.87 mm. The reflected signal is routed to the detector input. In addition, a portion of the incident signal is also routed to the detector input via a short circuit (e.g., reference signal) at the remaining port of the magic-tee. In this way the reflected and reference signals combine at the input to the detector, that behaves like a

mixer with a detector DC output voltage proportional to the magnitude and phase of the reflected signals.

Given the small amount of expected corrosion, the phase of the reflected signal is mostly sensitive to the presence of corrosion. Subsequently, the lap joint panels were scanned by this reflectometer at certain standoff distance, producing MW images of the panels. These images in general are quite suitable for quantitative analysis, such as estimation of corrosion thickness, and data fusion with other modalities. Images were produced at 67 GHz for corroded and pristine panels corresponding to scanned areas of 80 mm by 55 mm. These images were subsequently analyzed to determine the presence of corrosion and to evaluate corrosion thickness. It was found that the top 12 to 15 mm of each image consisted of non-corroded area, so that this information may be used as reference. Since the measurements were affected by the presence of sharp edges of the samples, using two orthogonal polarizations yielded a closer estimate of material loss. The technique for measuring the material loss (due to corrosion) is outlined in [15] in which a first order approximate model for the probe, not taking into account multiple reflections, was developed. In this investigation, the parameters of the probe were directly measured using a performance network analyzer. Furthermore, the actual detector's input-output curve (from the datasheet) was used instead of assuming a square-law region of operation for more accurate model compared to [15]. As a result a relatively accurate model was obtained that converts a reflection coefficient at the aperture (5) (see Figure 2.4) of the probe to a voltage output at the detector. The theoretical reflection coefficient at the aperture of the probe was calculated using the model for open-ended waveguide radiating a multilayer structure (e.g., paint-corrosion-

conductor) [16]. The dielectric constants of similar paint and corrosion layers were previously measured in [30]. Ultimately a curve was constructed relating the amount of corrosion/metal loss to the voltage output of the probe.

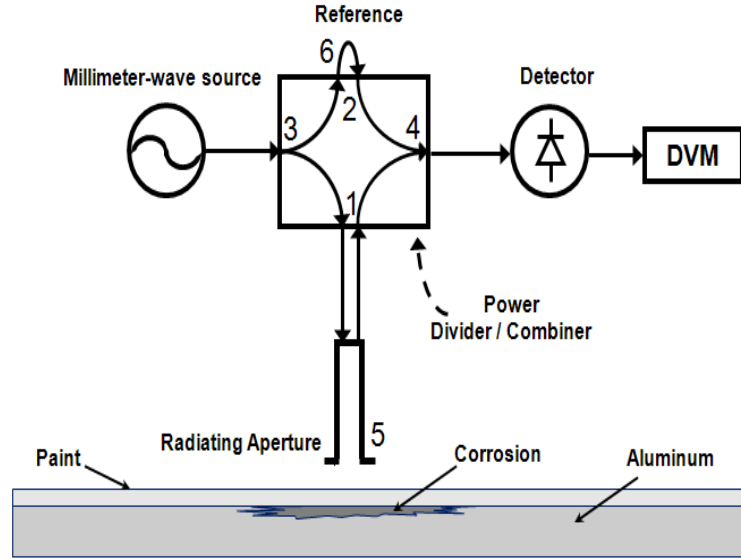


Figure 2.4: Schematic of the V-band millimeter wave reflectometer.

2.3. NDE MULTI-MODAL DATA FUSION-BASED IMAGE ANALYSIS

Figure 2.5 presents the multi-modal comprehensive structural analysis process used in this study to evaluate test panel vulnerabilities, extending the previously investigated process involving only uni-modal EC NDE data, the flowchart of which is shown in Figure 1.1 earlier [11].

With reference to Figure 2.5, let $X^{m_1}, X^{m_2}, \dots, X^{m_n}$ denote the registered images obtained for each modality for a test panel, where, m_1, m_2, \dots, m_n denote the n different modalities. Registering the different modality images for a test panel was performed by manually selecting the rivet positions, and an affine transformation was applied to generate images of standard size for the comprehensive structural analysis process.

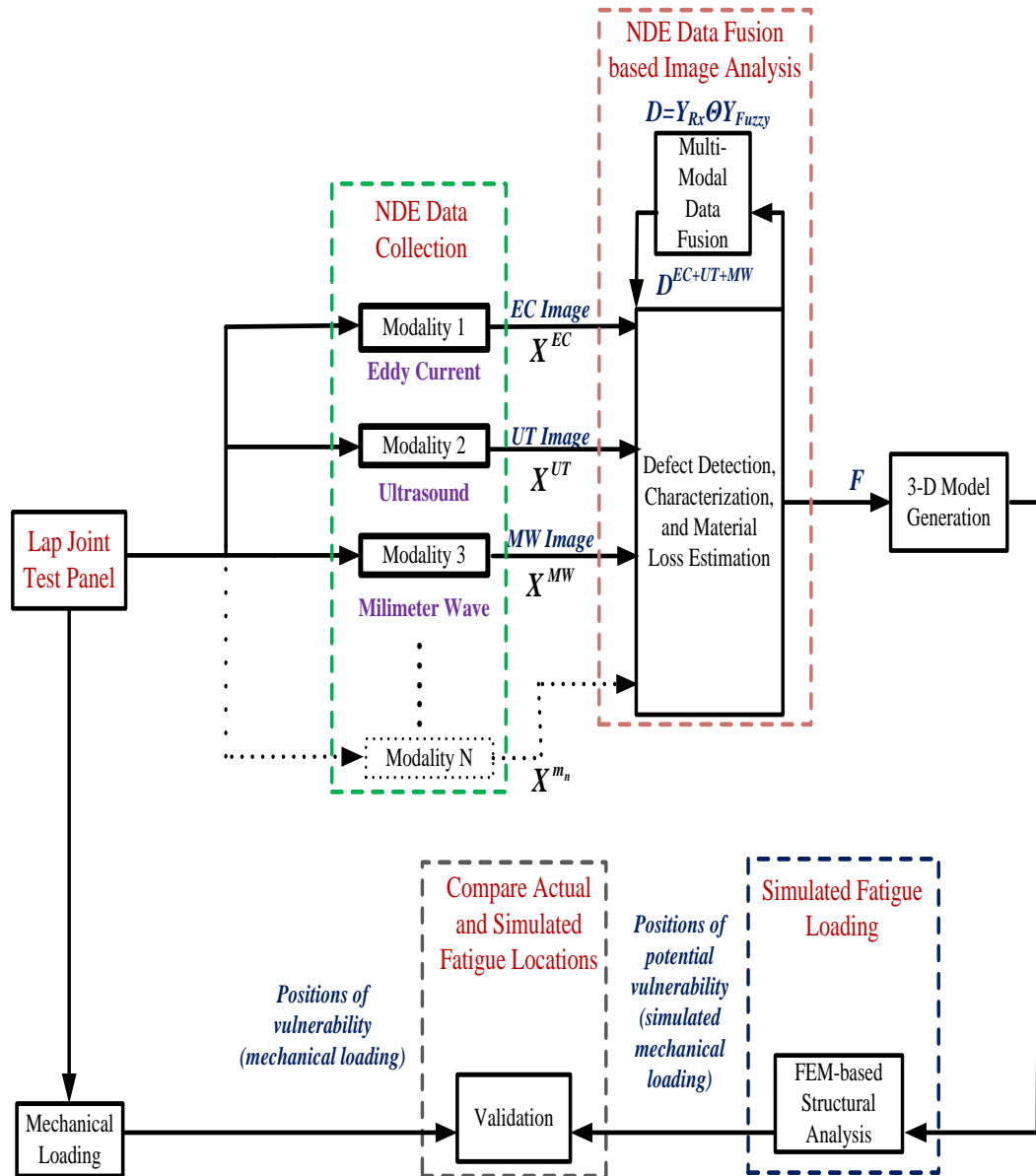


Figure 2.5: Overview of the multi-modal structural analysis process.

The next step in the process is corrosion detection. In previous study, an Rx statistical method (Rx Detector) and a data-driven fuzzy logic-based (Fuzzy Detector) data fusion method was developed for corrosion detection [27-29]. The RX confidence value at each

pixel (x, y) of the registered modality image, X^{m_n} denoted as $\lambda(x, y)$ with value range $0 \leq \lambda(x, y) \leq 1$, is defined as:

$$\lambda(x, y) = p^T N p \quad (1)$$

where p is a representative signature of a corrosion patch, N is the unknown background covariance matrix, which is computed from the zero mean image, M . Physically, the RX statistic $\lambda(x, y)$ measures the confidence value for corrosion detection at each pixel in the image $I(x, y)$ where, a higher value of λ would represent stronger ‘hits’ in detecting corrosion.

In the fuzzy logic-based clustering method, let G denote the fuzzy set which represents the gray levels associated with the non-corrosion areas in the registered NDE input image X^{m_n} . Then the histogram of this image is computed, and at each bin (z) of the secondary histogram, the associated membership function $\mu_G(z)$ can be expressed as follows:

$$\mu_G(z) = \begin{cases} (z/t)^{0.5} \\ 1 \end{cases} \quad (2)$$

where t is determined empirically as 95% of the area under the secondary histogram of the image [11]. If $|S_\alpha(x, y)|$ is the number of eight connected neighbors of a particular pixel (x, y) such that $\mu_G(z)(X^{m_n}(x, y)) \geq \alpha$ and $|S_0(x, y)|$ is the number of eight

connected neighbors such that $\mu_G(z)(X^{m_n}(x, y)) \geq 0$ then, a fuzzy clustering anomaly confidence measure, denoted as $R_\alpha(x, y)$ at the pixel location (x, y) is defined as:

$$R_\alpha(x, y) = \frac{|S_0(x, y)|}{|S_\alpha(x, y)|} \quad (3)$$

where $0 \leq R_\alpha(x, y) \leq 1$ and $\alpha = 0.7$ from previous work [11]. For enhancing corrosion detection and reducing the rate of false corrosion detection, intra-modal fusion of the RX and Fuzzy clustering anomaly detectors was performed to generate a final corrosion mask for each test panel. The final corrosion mask, denoted as $V(x, y)$, with confidence value between 0 and 1, was created using a maximum-likelihood weighted-average method as shown in Equation 4. $\sigma_{\lambda(x, y)}$ and $\sigma_{R_\alpha(x, y)}$ denote the standard deviation for the RX and fuzzy ratio confidence maps.

$$V(x, y) = \frac{(\sigma_{\lambda(x, y)})^2 \lambda(x, y) + (\sigma_{R_\alpha(x, y)})^2 R_\alpha(x, y)}{(\sigma_{\lambda(x, y)})^2 + (\sigma_{R_\alpha(x, y)})^2} \quad (4)$$

In [11], the data fusion method involved obtaining the decision-level fusion image (representing detected corrosion) from EC images only. In this study, a two-step data fusion approach was developed capable of combining NDE data from multiple modalities to create a final decision-level fusion image (see Figure 2.6).

First, the multi-modal fusion between the results of the individual anomaly detectors (Rx and Fuzzy Detectors) was determined to create the fused detector level image. Second, the decision-level multi-modal fusion of the Rx and Fuzzy Detector images to produce the multi-modal decision-level fusion image.

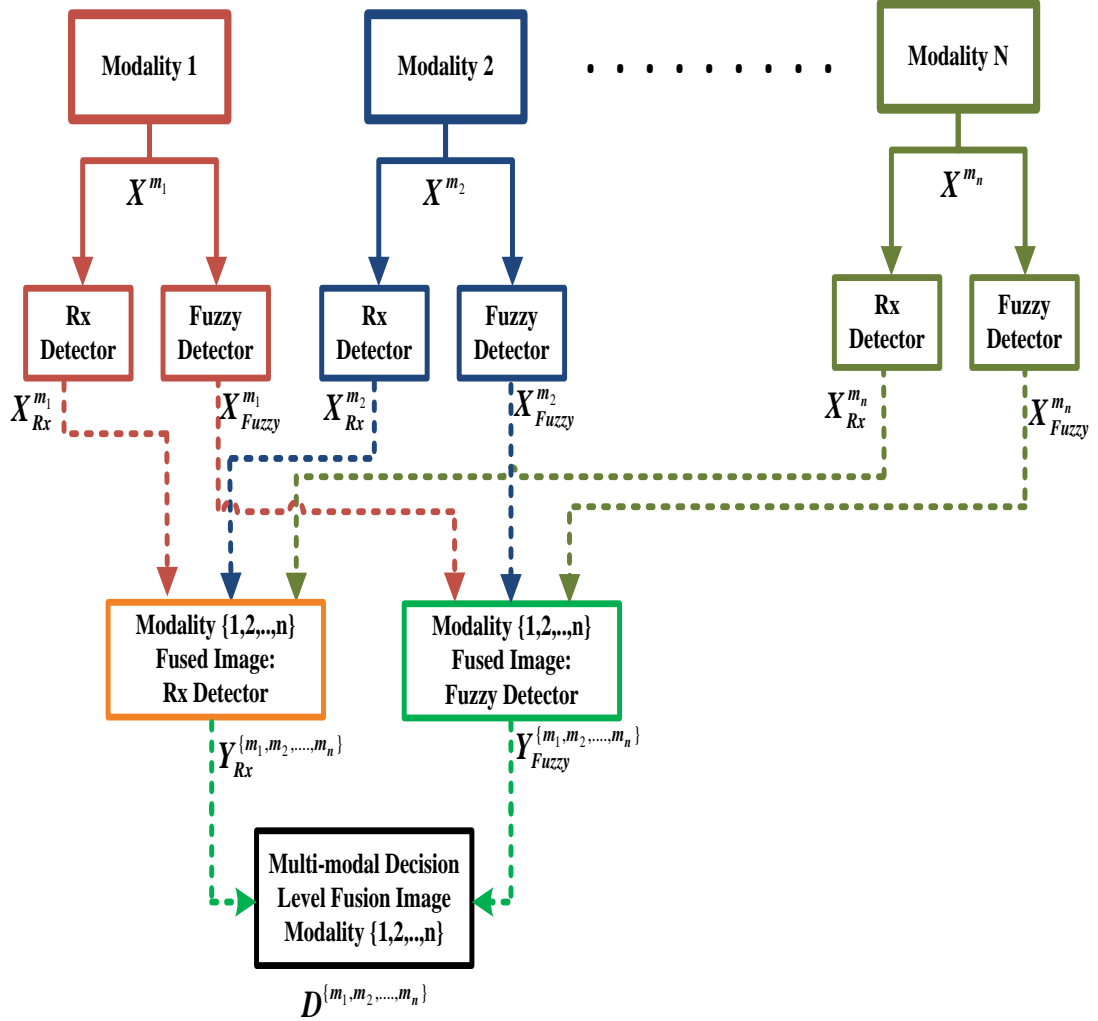


Figure 2.6: Various modality combinations.

The Rx and Fuzzy detector images, as applied to the original modal images, are denoted as $X_{Rx}^{m_1}, X_{Rx}^{m_2}, \dots, X_{Rx}^{m_n}$ and $X_{Fuzzy}^{m_1}, X_{Fuzzy}^{m_2}, \dots, X_{Fuzzy}^{m_n}$, respectively. The next step involved the fusion of the detector images using a maximum-likelihood based method [8], which is given in Equations (5) and (6), where $Y_{Rx}^{\{m_1, m_2, \dots, m_n\}}$ and $Y_{Fuzzy}^{\{m_1, m_2, \dots, m_n\}}$ denote the fused Rx detector and Fuzzy detector images for the modality combination $\{m_1, m_2, \dots, m_n\}$, respectively, while $\sigma_{Rx}^{m_1}, \sigma_{Rx}^{m_2}, \dots, \sigma_{Rx}^{m_n}$ and $\sigma_{Fuzzy}^{m_1}, \sigma_{Fuzzy}^{m_2}, \dots, \sigma_{Fuzzy}^{m_n}$

denote the corresponding Rx and Fuzzy detector standard deviations from application to the original images, respectively.

$$Y_{Rx}^{\{m_1, m_2, \dots, m_n\}} = \frac{(\sigma_{Rx}^{m_1})^2 X_{Rx}^{m_1} + (\sigma_{Rx}^{m_2})^2 X_{Rx}^{m_2} + \dots + (\sigma_{Rx}^{m_n})^2 X_{Rx}^{m_n}}{(\sigma_{Rx}^{m_1})^2 + (\sigma_{Rx}^{m_2})^2 + \dots + (\sigma_{Rx}^{m_n})^2}, \quad (5)$$

$$Y_{Fuzzy}^{\{m_1, m_2, \dots, m_n\}} = \frac{(\sigma_{Fuzzy}^{m_1})^2 X_{Fuzzy}^{m_1} + (\sigma_{Fuzzy}^{m_2})^2 X_{Fuzzy}^{m_2} + \dots + (\sigma_{Fuzzy}^{m_n})^2 X_{Fuzzy}^{m_n}}{(\sigma_{Fuzzy}^{m_1})^2 + (\sigma_{Fuzzy}^{m_2})^2 + \dots + (\sigma_{Fuzzy}^{m_n})^2} \quad (6)$$

Then, $\sigma_{Rx}^{\{m_1, m_2, \dots, m_n\}}$ and $\sigma_{Fuzzy}^{\{m_1, m_2, \dots, m_n\}}$ represent the standard deviation for the fused Rx and fused Fuzzy Detector images, respectively. Hence, the multi-modal decision-level fusion image is represented as $D^{\{m_1, m_2, \dots, m_n\}}$, and obtained as shown in Equation (7).

$$D^{m_1+m_2+\dots+m_n} = \frac{(\sigma_{Rx}^{m_1+m_2+\dots+m_n})^2 Y_{Rx}^{m_1+m_2+\dots+m_n} + (\sigma_{Fuzzy}^{m_1+m_2+\dots+m_n})^2 Y_{Fuzzy}^{m_1+m_2+\dots+m_n}}{(\sigma_{Rx}^{m_1+m_2+\dots+m_n})^2 + (\sigma_{Fuzzy}^{m_1+m_2+\dots+m_n})^2} \quad (7)$$

If X^{EC} , X^{UT} and X^{MW} denote the images that are obtained using the three different modalities, then X_{Rx}^{EC} , X_{Rx}^{UT} and X_{Rx}^{MW} . Let X_{Rx}^{EC} , X_{Rx}^{UT} and X_{Rx}^{MW} represent the Rx Detector images of the EC, UT and MW modalities, respectively. Similarly, X_{Fuzzy}^{EC} , X_{Fuzzy}^{UT} and X_{Fuzzy}^{MW} represent the corresponding Fuzzy Detector images. Then, the fused Rx Detector images are represented as $Y_{Rx}^{\{EC, UT\}}$, $Y_{Rx}^{\{EC, MW\}}$ and $Y_{Rx}^{\{EC, UT, MW\}}$ (Equations (8)-(10)) for the $\{EC, UT\}$, $\{EC, MW\}$ and $\{EC, UT, MW\}$ cases.

$$Y_{Rx}^{\{EC, UT\}} = \frac{(\sigma_{Rx}^{EC})^2 X_{Rx}^{EC} + (\sigma_{Rx}^{UT})^2 X_{Rx}^{UT}}{(\sigma_{Rx}^{EC})^2 + (\sigma_{Rx}^{UT})^2} \quad (8)$$

$$Y_{Rx}^{\{EC,MW\}} = \frac{(\sigma_{Rx}^{EC})^2 X_{Rx}^{EC} + (\sigma_{Rx}^{MW})^2 X_{Rx}^{MW}}{(\sigma_{Rx}^{EC})^2 + (\sigma_{Rx}^{MW})^2} \quad (9)$$

$$Y_{Rx}^{\{EC,UT,MW\}} = \frac{(\sigma_{Rx}^{EC})^2 X_{Rx}^{EC} + (\sigma_{Rx}^{UT})^2 X_{Rx}^{UT} + (\sigma_{Rx}^{MW})^2 X_{Rx}^{MW}}{(\sigma_{Rx}^{EC})^2 + (\sigma_{Rx}^{UT})^2 + (\sigma_{Rx}^{MW})^2} \quad (10)$$

Similarly, the fused Fuzzy Detector images are represented as $Y_{Fuzzy}^{\{EC,UT\}}$, $Y_{Fuzzy}^{\{EC,MW\}}$ and

$Y_{Fuzzy}^{\{EC,UT,MW\}}$ (Equations (11)-(13)).

$$Y_{Fuzzy}^{\{EC,UT\}} = \frac{(\sigma_{Fuzzy}^{EC})^2 X_{Fuzzy}^{EC} + (\sigma_{Fuzzy}^{UT})^2 X_{Fuzzy}^{UT}}{(\sigma_{Fuzzy}^{EC})^2 + (\sigma_{Fuzzy}^{UT})^2} \quad (11)$$

$$Y_{Fuzzy}^{\{EC,MW\}} = \frac{(\sigma_{Fuzzy}^{EC})^2 X_{Fuzzy}^{EC} + (\sigma_{Fuzzy}^{MW})^2 X_{Fuzzy}^{MW}}{(\sigma_{Fuzzy}^{EC})^2 + (\sigma_{Fuzzy}^{MW})^2} \quad (12)$$

$$Y_{Fuzzy}^{\{EC,UT,MW\}} = \frac{(\sigma_{Fuzzy}^{EC})^2 X_{Fuzzy}^{EC} + (\sigma_{Fuzzy}^{UT})^2 X_{Fuzzy}^{UT} + (\sigma_{Fuzzy}^{MW})^2 X_{Fuzzy}^{MW}}{(\sigma_{Fuzzy}^{EC})^2 + (\sigma_{Fuzzy}^{UT})^2 + (\sigma_{Fuzzy}^{MW})^2} \quad (13)$$

The decision-level fusion images were obtained utilizing the detector level fusion images using the same maximum likelihood scheme defined earlier [31].

Extending Equation (7) for the EC, UT and MW modalities, the decision-level fusion images for the multi-modal case is mathematically expressed by $D^{\{EC,UT\}}$, $D^{\{EC,MW\}}$ and $D^{\{EC,UT,MW\}}$ as shown in Equations (14)-(16), for the $\{EC,UT\}$, $\{EC,MW\}$ and, $\{EC,UT,MW\}$ modality combinations, respectively.

$$D^{\{EC,UT\}} = \frac{(\sigma_{Rx}^{\{EC,UT\}})^2 Y_{Rx}^{\{EC,UT\}} + (\sigma_{Fuzzy}^{\{EC,UT\}})^2 Y_{Fuzzy}^{\{EC,UT\}}}{(\sigma_{Rx}^{\{EC,UT\}})^2 + (\sigma_{Fuzzy}^{\{EC,UT\}})^2} \quad (14)$$

$$D^{\{EC,MW\}} = \frac{(\sigma_{Rx}^{\{EC,MW\}})^2 Y_{Rx}^{\{EC,MW\}} + (\sigma_{Fuzzy}^{\{EC,MW\}})^2 Y_{Fuzzy}^{\{EC,MW\}}}{(\sigma_{Rx}^{\{EC,MW\}})^2 + (\sigma_{Fuzzy}^{\{EC,MW\}})^2} \quad (15)$$

$$D^{\{EC,UT,MW\}} = \frac{(\sigma_{Rx}^{\{EC,UT,MW\}})^2 Y_{Rx}^{\{EC,UT,MW\}} + (\sigma_{Fuzzy}^{\{EC,UT,MW\}})^2 Y_{Fuzzy}^{\{EC,UT,MW\}}}{(\sigma_{Rx}^{\{EC,UT,MW\}})^2 + (\sigma_{Fuzzy}^{\{EC,UT,MW\}})^2} \quad (16)$$

Once the decision-level fusion image is obtained for corrosion detection, the next step in the overall process is to compute the material loss estimation image, denoted as $F^{\{m_1, m_2, \dots, m_n\}}$ (see Figure 2.5), which gives information about the material loss in the test panel and that is used to create the 3-D models for FEM-based simulated fatigue testing. At the non-zero confidence values in the final decision-level fusion image, i.e. $D^{\{m_1, m_2, \dots, m_n\}}(x, y) \neq 0$, the material loss estimation map was created for each position (x, y) , using interpolation of cluster distance from representative material loss signatures of 0%, 10%, 15% and 20% determined from the reference panel. The clustering technique can be represented as follows [27].

For each image pixel, $D^{\{m_1, m_2, \dots, m_n\}}(x, y)$ where $D^{\{m_1, m_2, \dots, m_n\}}(x, y) \neq 0$, the feature

vector $F(x, y) = \begin{bmatrix} D^{\{m_1, m_2, \dots, m_n\}}(x, y) & D_{norm}^{\{m_1, m_2, \dots, m_n\}}(x, y) \end{bmatrix}^T$ is created, where:

$$D_{norm}^{\{m_1, m_2, \dots, m_n\}}(x, y) = \frac{(D^{\{m_1, m_2, \dots, m_n\}}(x, y) - m_0)}{\sigma_0}, \quad I_{norm}(x, y) = \frac{(I(x, y) - m_0)}{\sigma_0}, \quad \text{and } m_0 \text{ and } \sigma_0 \text{ are}$$

the mean and standard deviation values of 0% (background/no-corrosion region) material loss from the reference panel (Figure 2.2(a) for EC and Figure 2.3(a) for UT NDE data; sample reference panel schematic is shown in Figure 2.7). The Euclidean distance, denoted as $D_M^{\{m_1, m_2, \dots, m_n\}}(x, y)$ at the output of the Defect Detection, Characterization and Material Loss Estimation block in Figure 2.5, is computed between $F(x, y)$ and median feature vectors determined from the reference panel for the 0% ($D_0^{\{m_1, m_2, \dots, m_n\}}(x, y)$), 10% ($D_{10}^{\{m_1, m_2, \dots, m_n\}}(x, y)$), 15% ($D_{15}^{\{m_1, m_2, \dots, m_n\}}(x, y)$), and 20% ($D_{20}^{\{m_1, m_2, \dots, m_n\}}(x, y)$) material loss estimates. Finally, interpolation of the Euclidean distances is used for creating the material loss estimation map for the fused model $\{m_1, m_2, \dots, m_n\}$.

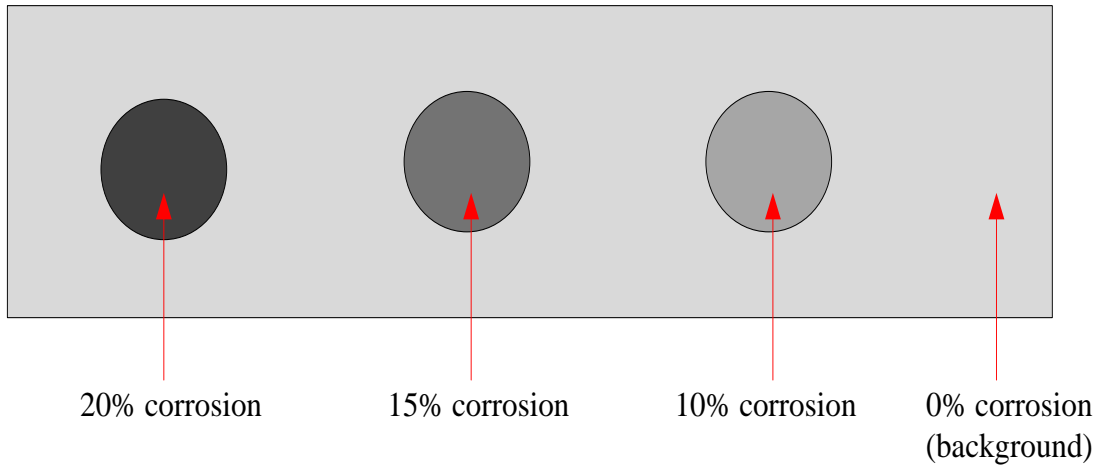


Figure 2.7: Sample schematic of reference panel used for material loss estimation.

For the different modality combinations used in this research, $\{EC, UT\}$, $\{EC, MW\}$ and, $\{EC, UT, MW\}$, the material loss estimation produces the final material

loss estimation images, $F^{\{EC,UT\}}$, $F^{\{EC,MW\}}$ and $F^{\{EC,UT,MW\}}$ respectively. Experimental results for the data fusion based experiments have been presented in Section 3.1 of the paper.

2.4. TEST PANEL MODEL GENERATION

Using the material loss estimation information obtained in the previous section, a 3-D geometrical model representation for the riveted test panels was generated. A localized non-overlapping windowing method was used for the 3-D model generation. This involved dividing the height and width dimensions of the test panel into 10x10 (approximately 5 mm x 5 mm) non-overlapping windows. The average material loss over every non-overlapping window was estimated, and the center of mass was computed. Finally, a circular puck-like region was created with a thickness equal to the average material loss of the 10x10 window, and area equal to the number of non-zero pixels (indicating corrosion) present in the same window.

The circular puck-like region was then centered (using the center of mass information) within the non-overlapping window. An example of a 3-D model obtained for test panel 1 for the $\{EC,UT,MW\}$ modality combination is shown in Figure 2.8, illustrated using ANSYS[®]. It is worthwhile to mention here that the puck-like regions shown in Figure 2.8 are actually solid material loss regions, which could not be visually represented using the ANSYS[®] display tool. Also, note that the 3-D modeling method was developed based on constraints in ANSYS[®], which is used for simulated fatigue loading of the test panels [11].

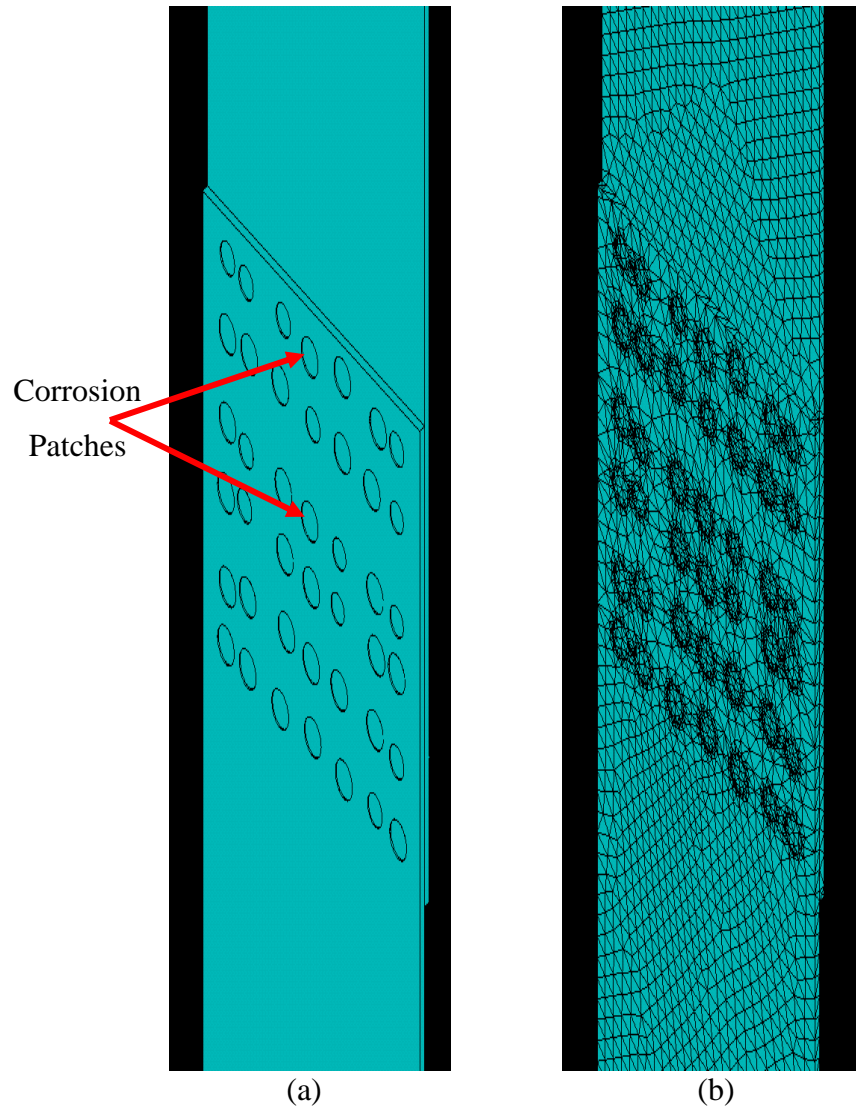


Figure 2.8: 3-D Model generation for a test panel.

(a) Showing corroded patches obtained using localized non-overlapping window method, displayed using ANSYS® for test panel 1 for the {EC, UT, MW} modality combination; and (b) Snapshot of the meshed model.

2.5. SIMULATED FATIGUE LOADING OF TEST PANELS

The ANSYS® Mechanical APDL tool was used to perform the simulated fatigue loading based on the 3-D model of the panel. The 3-D model was subjected to boundary conditions so as to simulate the mechanical loading (details of mechanical loading is

presented in the next section). A fixed load was applied to the top end of the panel while a cyclical load was applied to the bottom. The loading of the panels was done in conjunction with the Fatigue Analysis module in ANSYS® to perform the simulated fatigue evaluation of the test panels.

A standard method, based on analyzing cyclic loading of structures, was used for performing the simulated fatigue loading [32]. To measure the vulnerability of a specific node/region within the 3-D model of the test panels, the Alternating Stress Intensity (Alt-SI) values for all the nodes of the meshed panel were determined using ANSYS® Fatigue Analysis tool. The Alt-SI values at the nodes of the finite elements essentially are representative of the degree of stress at the nodes. A higher Alt-SI value at a node indicates that that particular node/area is more vulnerable to damage as compared to nodes with lower Alt-SI values.

Previous research has shown that most of the damage/failure in these test panels occurred at the sides of the rivets, horizontally [11]. Based on this, only the Alt-SI values of the nodes around the rivets were used for comparison to the mechanical loading data. Figure 2.9a shows the 3-D FEM based model of the test panel while Figure 2.9(b) shows the riveted region of the test panel with the nodes (circled in blue) whose Alt-SI values were used for comparison with mechanical loading results. Experimental results obtained from the simulated fatigue loading and its comparison to mechanical loading is presented in Section 3.3.

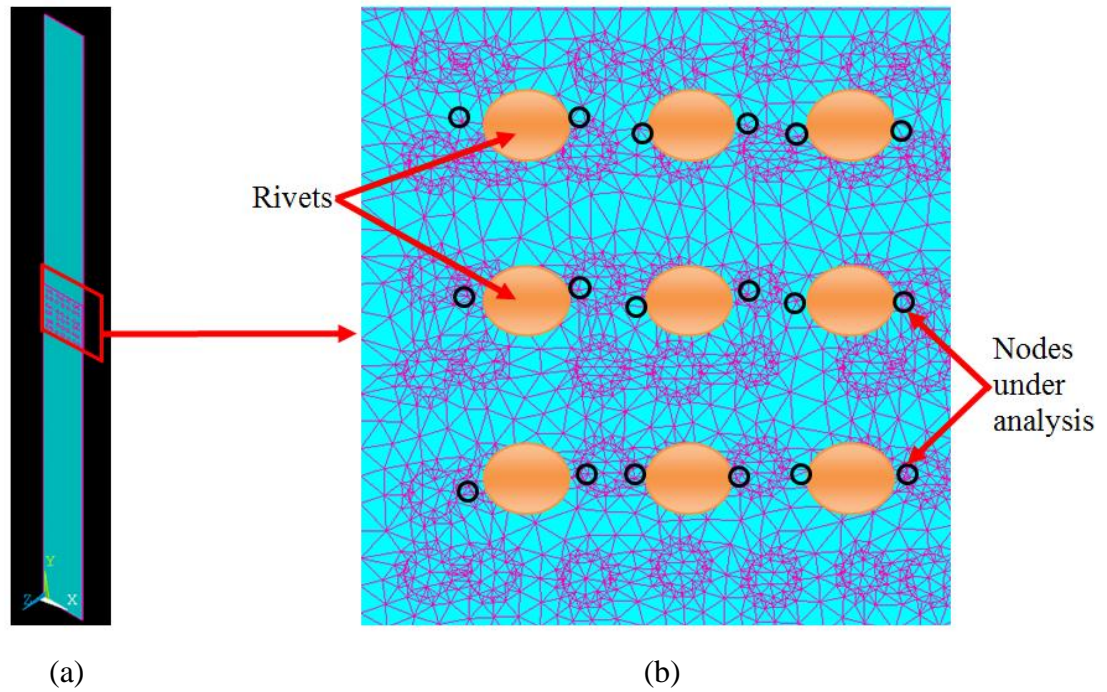


Figure 2.9: Meshed model showing the nodes under analysis.
 (a) Entire 3-D model for the {EC,UT,MW} modality combination; (b) Zoomed-in view of the riveted region with the nodes whose Alt-SI values are recorded shown in circles.

2.6. MECHANICAL LOADING OF TEST PANELS

The test panels were subjected to two rounds of mechanical loading. NDE data was collected before each round of mechanical loading for visual identification of cracks in the test panels to compare with the vulnerable locations found with simulated fatigue loading (obtained using NDE data) with each round of mechanical loading (actual fatigue testing) data.

For the mechanical loading, the test panels were cyclically loaded between 1.04 kN and 10.4 kN (load ratio, $R = P_{min}/P_{max} = 0.1$) at a frequency of 4 Hz for approximately 30,000 cycles across the two rounds of loading. The mechanical loading setup is shown in Figure 2.10.

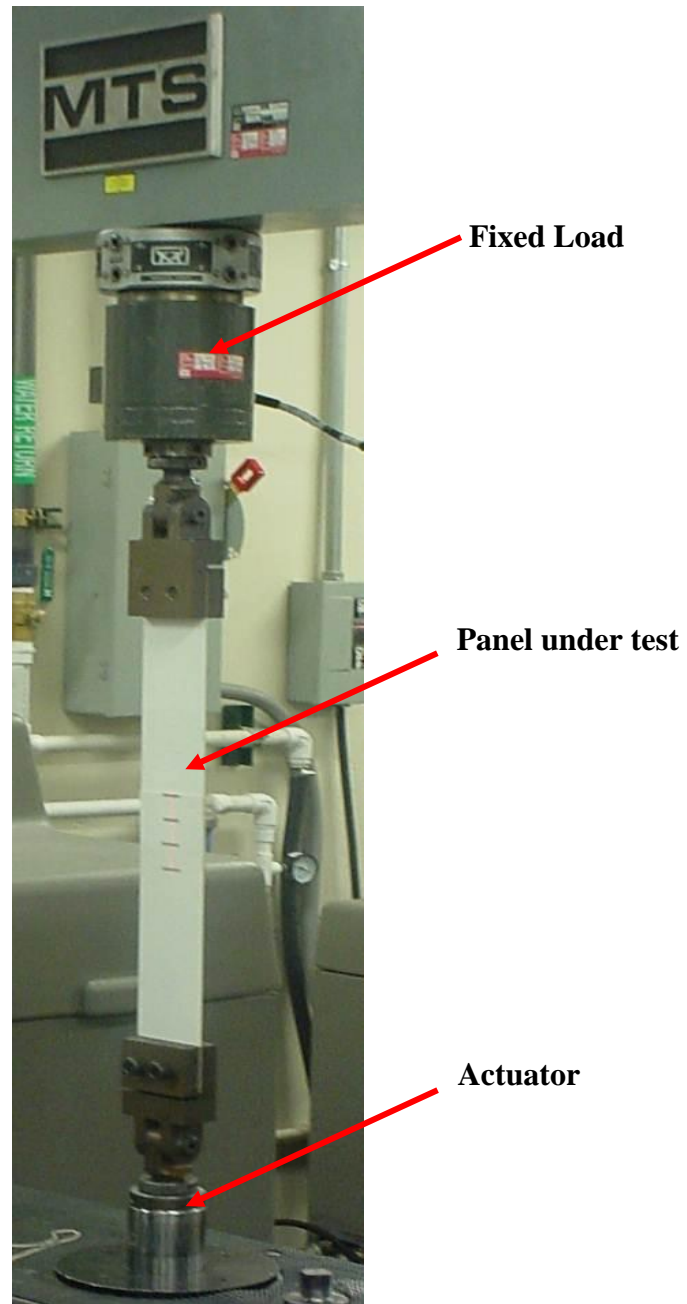


Figure 2.10: Cyclical mechanical loading setup.

In-situ strain measurement of the test panel was attempted using a MTS LX 300 laser extensometer. However, the test did not produce meaningful data as the test panels got deflected around the front rivet portion upon application of the load. Figure 2.11 shows this phenomenon.

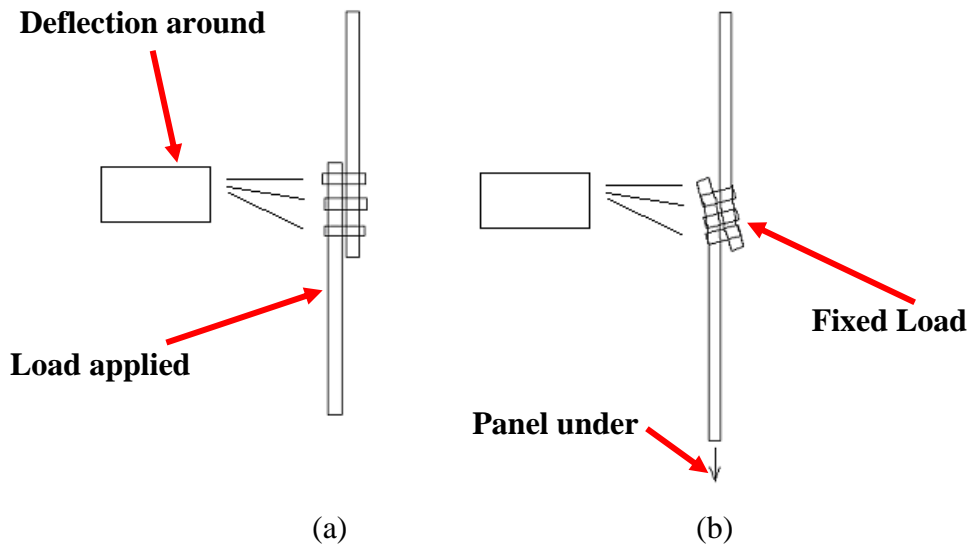


Figure 2.11: Side view schematic of the panel (a) unloaded and (b) loaded.

Since the strain measurements could not be carried out, the panels were visually inspected after each round of mechanical loading and also photographed using a stereomicroscope to inspect for cracks after each round of mechanical loading. These cracks were meant to simulate in-service flaws. The first round of mechanical loading was intended to initiate small fatigue cracks, wherein, the test panels were cyclically loaded until a crack was noticed (see Table 3.1). For the second round of mechanical loading, cracks were allowed to grow to twice their former lengths or if no cracks originated in the first round, cracks were initiated and allowed to grow to a length of ~3 mm. The experimental results for the mechanical loading are presented in Section 3. 2.

3. EXPERIMENTAL RESULTS

3.1. MULTI-MODAL DATA FUSION EXPERIMENTS

The NDE data collected from the three modalities were subjected to the RX and Fuzzy detectors and then to the decision-level fusion process, the details of which have been explained in Section 2.3. Experimental results for the raw EC, UT and MW images, with their respective material loss estimation images are shown in Figures 3.1-3.3, with the reference scale for material loss estimation between 0% and 20% are presented for corroded test panel 1. Figure 3.1 presents EC NDE measurement data before round 1 of mechanical loading [11], Figure 3.2 presents the UT data and Figure 3.3 shows the MW data. The material loss estimation images using the data fusion method for the individual modalities were performed so that the results could be compared to the multi-modal data fusion-based material loss estimation method.

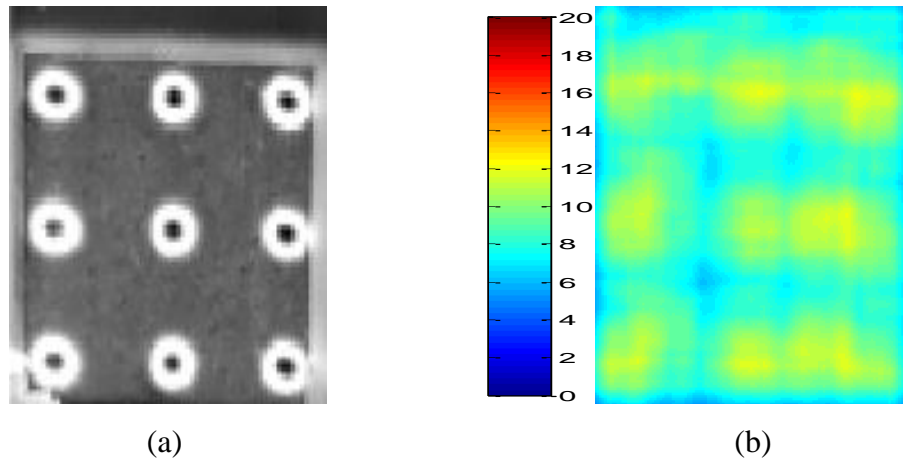


Figure 3.1: EC and Material loss estimation data of test panel 1. Data collected and data fusion based material loss estimation algorithm results for test panel 1 before round 1 of mechanical loading. (a) EC data for test panel 1, (b) material loss estimation map with reference scale.

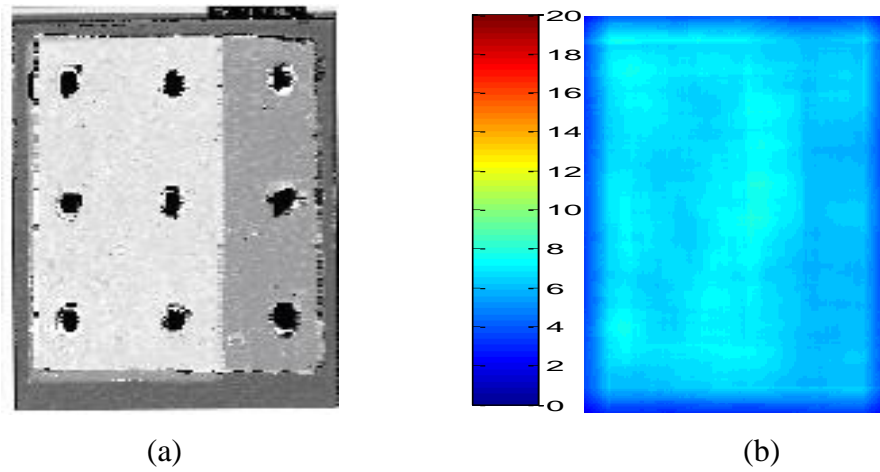


Figure 3.2: Data collected and data fusion based material loss estimation result. Results for test panel before round 1 of mechanical loading. (a) UT data for test panel 1, (b) material loss estimation map with reference scale.

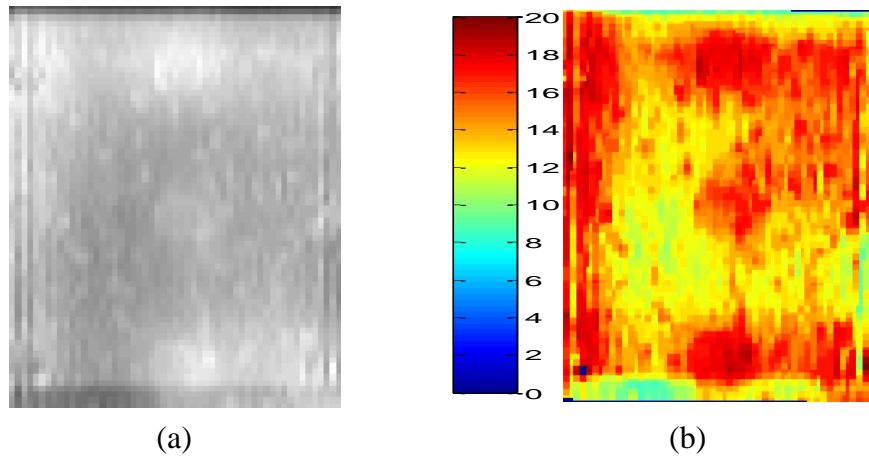


Figure 3.3: Data collected and data fusion based material loss estimation. Results for test panel 1 before round 1 of mechanical loading. (a) MW data for test panel 1, (b) material loss estimation map with reference scale.

Once the material loss estimation results for all three modalities were obtained, the multi-modal data fusion algorithm explained in Section 2.3 was used to obtain the material loss estimation images.

For test panel 1, Figure 3.4a shows the EC and UT fused results, Figure 3.4b shows the EC and MW fused results while the EC, UT and MW fused results for the NDE data obtained before the first round of mechanical loading are shown in Figure 3.4c.

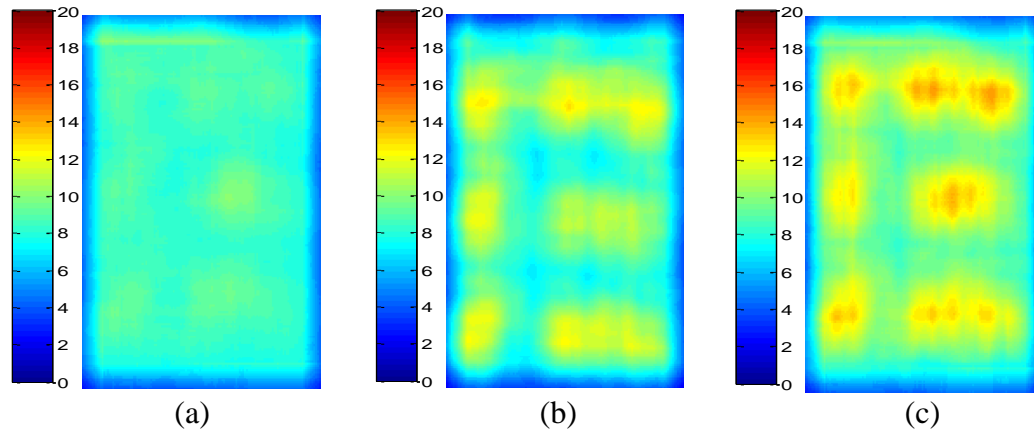


Figure 3.4: Multi-modal data fusion based material loss estimation for panel 1. NDE data obtained before round 1 of mechanical loading. (a) EC and UT fused result; (b) EC and MW fused result; and (c) EC, UT and MW fused result.

3.2. MECHANICAL LOADING RESULTS FOR TEST PANELS

During the mechanical loading of the panels we observed that all of the cracks developed on the lower row of rivets, suggesting a higher stress concentration at this row. Table 3.1 presents a summary of the mechanical loading crack locations across the two rounds of mechanical loading [11]. These results are used for comparison purposes with the crack locations predicted by the data fusion based simulated fatigue loading process using FEM structural modeling and analysis.

Table 3.1: Summary of crack locations for two rounds of mechanical loading.

Panel #	Crack locations found after Round 1 of mechanical loading	Cracks locations found after Round 2 of mechanical loading
1	13, 15	13, 15
2	13	13, 14
3	None	14
4	None	14
6	13, 16	13, 14, 16

3.3. SIMULATED FATIGUE LOADING RESULTS

For simulated fatigue loading, the Alt-SI (stress intensity) values were analyzed for the critical and non-critical locations of the test panel. Since all the cracks found from mechanical loading were in the lowest row of the rivets, these panel locations were designated as the critical locations of the panel (locations 13-18), as shown in from Figure 3.5. The other regions of the panel, viz. locations 1-12 are designated as the non-critical panel locations. Table 3.2 provides a summary of the Alt-SI values at the critical and non-critical locations (nodes) of the test panel 1 after the simulated fatigue loading.

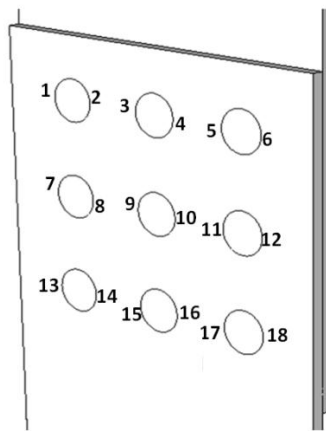


Figure 3.5: Location key for crack initiation sites in test panels.

Table 3.2: Summary of Alt-SI values for test panel 1.

	Before round 1 of actual loading			Before round 2 of actual loading		
	{EC,UT} (MPa)	{EC,MW} (MPa)	{EC,UT,MW} (MPa)	{EC,UT} (MPa)	{EC,MW} (MPa)	{EC,UT,MW} (MPa)
Location 1	3.2	2.7	2.2	5.5	4.7	8.1
Location 2	0.2	1.3	0.8	1.0	0.6	0.3
Location 3	0.7	1.0	0.6	1.1	0.5	0.2
Location 4	0.5	1.3	0.4	0.7	0.5	0.2
Location 5	0.2	1.3	0.6	1.0	0.3	0.0
Location 6	3.9	2.2	3.5	5.5	9.7	6.4
Location 7	28.7	28.0	8.5	6.8	2.2	2.5
Location 8	1.1	0.8	4.5	3.7	3.0	1.2
Location 9	2.1	6.5	4.3	3.8	5.0	1.7
Location 10	0.4	1.0	3.0	3.6	4.8	0.8
Location 11	2.7	1.4	5.9	2.8	1.5	1.2
Location 12	4.8	5.0	4.0	21.0	5.1	0.8
Location 13	134.9	109.3	83.6	55.6	84.1	76.9
Location 14	72.9	63.2	70.5	108.1	67.5	45.0
Location 15	57.5	75.9	55.0	70.7	64.6	51.9
Location 16	33.2	92.7	21.1	46.5	54.5	47.4
Location 17	9.2	63.5	35.3	55.4	47.9	36.0
Location 18	40.0	34.7	37.4	45.4	54.8	43.3

Furthermore, the crack locations identified during the mechanical loading (Table 3.1) exhibited no cracks in the test panels at locations 1-12. All the detected cracks were between locations 13-18.

Table 3.3 gives the Alt-SI values for the different locations from Figure 3.5 for test panel 1 for different modality combinations. Similar data tables were generated for test panels 1-4 and the pristine panel. Table 3.3 presents the Alt-SI values for test panels 1-4 and the pristine panel for the $\{EC,UT\}$ modality combination.

Table 3.3: Alt-SI values for test panels 1-4 and pristine panel {EC, UT}.

	Before round 1 of actual loading (in MPa)					Before round 2 of actual loading (in MPa)				
	Test Panel 1	Test Panel 2	Test Panel 3	Test Panel 4	Pristine Panel	Test Panel 1	Test Panel 2	Test Panel 3	Test Panel 4	Pristine Panel
Location 1	3.2	0.8	3.4	1.3	1.5	5.5	2.3	9.1	10.0	12.7
Location 2	0.2	0.5	4.6	2.3	1.0	1.0	0.4	0.7	0.5	2.0
Location 3	0.7	0.4	4.2	3.3	0.7	1.1	0.4	0.5	0.5	1.7
Location 4	0.5	0.5	6.0	1.7	0.2	0.7	0.3	0.4	0.7	1.2
Location 5	0.2	0.6	5.5	2.3	0.5	1.0	0.1	0.1	0.2	0.3
Location 6	3.9	2.0	2.4	18.8	1.1	5.5	8.0	1.4	9.1	1.6
Location 7	28.7	4.8	6.8	33.4	5.8	6.8	11.5	9.7	24.3	3.0
Location 8	1.1	3.3	3.5	8.8	3.5	3.7	2.3	1.6	3.9	4.4
Location 9	2.1	2.0	10.9	9.2	1.4	3.8	2.4	2.0	4.0	4.3
Location 10	0.4	0.4	0.9	7.4	4.1	3.6	11.4	3.1	4.9	4.0
Location 11	2.7	1.6	2.6	6.3	5.5	2.8	1.7	2.2	3.0	4.5
Location 12	4.8	1.7	3.9	6.4	5.4	21.0	5.1	3.6	6.1	34.1
Location 13	134.9	44.0	48.7	56.2	50.1	55.6	73.0	32.9	49.4	92.8
Location 14	72.9	35.6	33.3	55.4	1.0	108.1	58.2	110.9	56.7	60.8
Location 15	57.5	39.3	53.6	51.4	34.2	70.7	47.0	56.6	49.0	49.1
Location 16	33.2	26.2	31.3	55.8	34.5	46.5	43.1	39.7	49.3	67.1
Location 17	9.2	26.1	36.5	53.2	13.2	55.4	42.1	50.1	53.9	46.2
Location 18	40.0	53.0	49.6	51.1	24.6	45.4	56.3	53.2	61.7	59.5

A fuzzy-set approach was used to find the cutoff value for automatic detection of the cracks from the Alt-SI values of the simulated fatigue loading of the panels. A fuzzy set \mathbf{P} provides the degree of vulnerability in a structure (corroded and pristine panels) at a specified position in terms of Alt-SI. Let $\mu_p(\mathbf{V})$ represent the membership value in \mathbf{P} for an Alt-SI value, \mathbf{V} , determined in a structure at a specified position. $\mu_p(\mathbf{V})$ is defined in Equation 17:

$$\mu_p(\mathbf{V}) = \begin{cases} \frac{\mathbf{V}}{70}, & \text{if Alt - SI} < 70 \text{ MPa} \\ 1, & \text{otherwise} \end{cases} \quad (17)$$

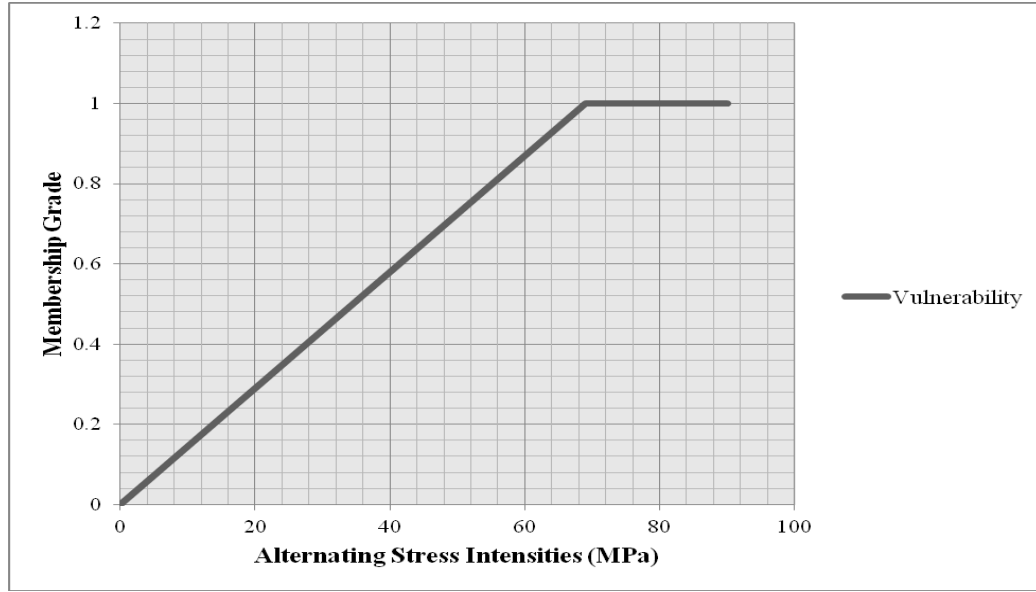


Figure 3.6: Fuzzy-logic based method of automatically finding the threshold.

Figure 3.6 shows the membership function $\mu_p(V)$. Alpha-cuts (α) on the fuzzy set P were examined for determining test panel vulnerability positions, where all positions with $\mu_p(V) \geq \alpha$ are designated as vulnerable for $0 \leq \alpha \leq 1$ as shown in Equation 18.

$$\text{Vulnerability} = \begin{cases} 1, & \text{if } \mu_p(V) \geq \alpha \\ 0, & \text{otherwise} \end{cases} \text{ for } 0 \leq \alpha \leq 1 \quad (18)$$

An alpha-cut value of 0.45 ($\alpha = 0.45$) produced the best results, the details of which for each test panel and each modality combination for all the two rounds of simulated fatigue loading are summarized in Table 3.4. True Positive (TP) and True Negative (TP) crack location detection rates were computed over the four corroded and one pristine test panels. A true positive crack location is correctly labeled if the location satisfies the Alt-SI threshold constraint, and the location in the test panel for the corresponding round of

mechanical loading is designated as a crack site, as given in Table 3.1. A true negative location is labeled correctly if the location does not satisfy the Alt-SI threshold constraint, and the location in the test panel for the corresponding round of mechanical loading is not called a crack site, as shown in Table 3.1. False positives are the number of cracks that were erroneously detected across all panels, and false negatives refer to cases when the simulated fatigue loading method failed to predict a crack location that was actually a crack location obtained by the mechanical loading process. For each round of loading, and for each modality combination, 18 total number of true positive and true negative locations resulted. In Table 3.4, the number on the left side of the slash indicates the number of vulnerable locations (true positive) and the number of non-vulnerable locations (true negative) found from the simulated fatigue loading process, and the number to the right of the slash gives the true positive and true negative locations as found from mechanical loading.

Table 3.4: Accuracy metrics of simulated fatigue loading vs. mechanical loading.

	Before round 1 of mechanical loading						Before round 2 of mechanical loading					
	True Positive			True Negative			True Positive			True Negative		
	{EC,UT}	{EC,MW}	{EC,UT,MW}	{EC,UT}	{EC,MW}	{EC,UT,MW}	{EC,UT}	{EC,MW}	{EC,UT,MW}	{EC,UT}	{EC,MW}	{EC,UT,MW}
Panel 1	2/2	2/2	2/2	14/16	12/16	13/16	2/2	2/2	2/2	12/16	12/16	12/16
Panel 2	1/1	1/1	1/1	17/17	17/17	17/17	2/2	2/2	2/2	16/16	16/16	16/16
Panel 3	0/0	0/0	0/0	18/18	18/18	18/18	1/1	1/1	1/1	17/17	17/17	17/17
Panel 4	0/0	0/0	0/0	18/18	18/18	18/18	1/1	1/1	1/1	17/17	17/17	17/17
pristine	2/2	2/2	1/2	16/16	16/16	16/16	3/3	2/3	1/3	15/15	15/15	15/15
Total	5/5	5/5	4/5	83/85	81/85	82/85	9/9	8/9	7/9	77/81	77/81	77/81

The accuracy for each round of simulated fatigue loading is computed by calculating the sum of true positives and true negatives detected divided by the total number of positions for each round of the mechanical loading. Table 3.5 shows the summarized results obtained using this method. Inspecting the Overall Accuracies (OA) in Table 3.5, all multi-modal NDE data combinations yielded higher Overall Accuracies (OA) than the individual modalities.

For the multi-modal fused data, the correct prediction rate was as high as 98.8% for the $\{EC, UT\}$ case before round 1 of mechanical loading. It should be noted that since the MW data was obtained for one round only, Table 3.5 shows the MW data same for before and after round 1 of actual loading. In general, the data fusion process yielded a consistent correlation for vulnerability location compared to visual inspection from mechanical loading.

Table 3.5: Summary of accuracy metrics.

	{EC} (%)			{UT} (%)			{MW} (%)			{EC,UT} (%)			{EC,MW} (%)			{EC,UT,MW} (%)		
	TP	TN	OA	TP	TN	OA	TP	TN	OA	TP	TN	OA	TP	TN	OA	TP	TN	OA
Before round 1 of actual loading	80.0	88.2	84.1	80.0	81.1	80.6	75.0	73.2	74.1	100.0	97.6	98.8	100.0	95.3	97.7	80.0	96.5	88.2
Before round 2 of actual loading	66.7	88.9	77.8	75.0	78.1	76.6	75.0	73.2	74.1	100.0	95.1	97.5	88.9	95.1	92.0	77.8	95.1	86.5

4. CONCLUSIONS

For the comprehensive structural analysis process explored in this research, different NDE modality combinations were examined for test panel modeling, including: 1) EC, 2) UT, 3) MW, 4) EC and UT, 5) EC and MW, and 6) EC, UT, and MW. From Table 5, experimental results showed that using multi-modal NDE data for test panel modeling improved vulnerability discrimination over individual modality NDE data for test panel modeling. Eddy Current-based models yielded the highest single-modality vulnerability recognition rate for the test panels examined. Different NDE modalities provide different perspectives of the test panel that can be used to enhance the decision making process. However, each modality has relative strengths. Eddy Current and UT were collected at relatively low frequencies. With the material properties of Aluminum, these modalities provide depth information about the corrosion and defects present in the different layers of the test panels. Millimeter Wave NDE data was collected at relatively high frequencies and provide surface information about the presence of corrosion and its severity (e.g., thickness) in the test panels.

Experimental results showed that the EC and UT modality combination yielded the highest correct vulnerable (crack) location recognition rate compared to cracks identified with mechanical loading. Experimental results also showed that the MW modality gave the lowest correct vulnerable location recognition rate compared to mechanical loading. This modality, as applied, was more sensitive to the presence and severity of corrosion and corrosion pitting than detecting tiny cracks around rivets. These

experimental results appear to indicate that: 1) there is complementary information in the different NDE modalities that can be used to enhance vulnerability recognition and assessment in the experimental test panels and, 2) more NDE sources for data fusion to characterize corrosion and defects does not necessarily translate into better vulnerable location recognition. The utility of data fusion here is to facilitate better characterization of the corrosion and defects present within the test panels for generating models representative of the actual test panels so that vulnerable locations may be accurately identified with simulated fatigue loading. The severe corrosion regions within the test panels scatter the NDE signals in a similar fashion as a small closed crack. Thus, detection of a crack in a severely corroded region may not be possible, and this may have contributed to the MW method not indicating crack locations but indicating severe corrosion. Experiments were performed to determine vulnerable regions in the test panels using the comprehensive structural analysis process for different NDE modality combinations to characterize corrosion and defects present to generate geometrical models for simulated fatigue loading. Characterizing corrosion and defects throughout the test panel (i.e. depth information) is important for vulnerability determination, as the vulnerability recognition results show for the fused EC and UT NDE modalities.

The comprehensive structural analysis process facilitates the collection of multi-modal NDE data (EC, UT and MW data were collected in this study) from a test structure. Intra- and inter-modality fusion were performed to detect corrosion and defects and to quantify the amount and location of the corrosion and defects in the test panel, i.e. material loss estimation, for automatically developing a model of the test panel. The purpose of data fusion, in this research, is to improve the detection capability of corrosion

and other defects and anomalies in a test panel based on collected NDE data. The model of the test panel is analyzed using an FEM-based tool (ANSYS® in this study) for simulated fatigue loading to determine where the test panel is vulnerable to failure. As part of this comprehensive process, a fuzzy logic-based algorithm automatically determines the test point locations based on the Alt-SI values obtained from FEM-based modeling where the test panel is vulnerable based on the FEM simulated stress intensities. Traditional/existing techniques which use data fusion combine multi-modal NDE data to enhance test structure visualization for analysis. This comprehensive process advances traditional/existing data fusion techniques in the context that data fusion is performed and utilized for automated structural analysis. It also facilitates predictive analysis by automatically generating a geometrical model of a test panel based on quantized areas of material loss for FEM-based simulated fatigue loading which can be used for comparison to mechanical loading (as was done in this study) and for simulation of the test panel to failure (as was done in this study) to identify areas of vulnerability.

Future research will explore different types of data fusion approaches to combine NDE modality data to characterize the test structures to attempt to integrate unique information from each modality. In order to address the limitations of FEM-based methods in their ability to predict fatigue, more specialized numerical modeling techniques need to be investigated which would allow the generation of a more consistently structured mesh or a consistently refined mesh in which it is possible to acquire stress data from the same location between meshing increments, which would allow for more consistent prediction of stresses.

ACKNOWLEDGMENTS

This work was supported in part by the Air Force Research Laboratory (AFRL) under contract no. FA8650-04-C-5704 in conjunction with the Center for Aerospace Manufacturing Technologies (CAMT).

REFERENCES

1. M. Brassard, A. Chahbaz, A. Pelletier, "Combined NDT inspection techniques for corrosion detection of aircraft structures," in *15th World Congress on Nondestructive Testing* Roma, Italy, 2000.
2. G. H. Koch, "Corrosion in Aluminum Alloy 2024-T3 Lap Joints," in *CORROSION* 99, Houston, TX, 1999.
3. C. L. Brooks, K. Honeycutt, S. Prost-Domasky, "Case studies for life assessments with age degradation," in *Fourth Joint DoD/FAA/NASA Conference on Aging Aircraft*, St. Louis, MO, 2000.
4. S. Young-Won and S. S. Udpa, "A new morphological algorithm for fusing ultrasonic and eddy current images," in *Ultrasonics Symposium, 1996. Proceedings., 1996 IEEE*, 1996, pp. 649-652 vol.1.
5. B. A. Lepine, B. P. Wallace, D. S. Forsyth, and A. Wyglinski, "Pulsed eddy current method developments for hidden corrosion detection in aircraft structures," *Canadian Society for Non-Destructive Testing Journal*, vol. 20, pp. 6-14, 1999.
6. M. A. Robers and R. Scottini, "Pulsed Eddy Current in Corrosion Detection," presented at the 8th European Conference on Nondestructive Testing, Barcelona (Spain), 2002.
7. J. Skramstad, R. Smith, and D. Harrison, "Enhanced detection of deep corrosion using transient eddy currents," in *7th Joint DoD/FAA/NASA Conference on Aging Aircraft*, New Orleans (USA), 2003.
8. D. Forsyth, J. Komorowski, "Fusion of Multimodal NDI Images for Aircraft Corrosion Detection and Quantification," in *Multi-Sensor Image Fusion and Its Applications*, ed: CRC Press, 2005, pp. 375-404.
9. M. T. Ghasr, S. Kharkovsky, R. Zoughi, and R. Austin, "Comparison of near-field millimeter-wave probes for detecting corrosion precursor pitting under paint," *Instrumentation and Measurement, IEEE Transactions on*, vol. 54, pp. 1497-1504, 2005.
10. Z. Liu, D. S. Forsyth, J. P. Komorowski, K. Hanasaki, T. Kirubarajan, "Survey: State of the Art in NDE Data Fusion Techniques," *Instrumentation and Measurement, IEEE Transactions on*, vol. 56, pp. 2435-2451, 2007.

11. S. De, Kapil Gupta, R. Joe Stanley, Reza Zoughi, Kenneth Doering, David Van Aken, Gary Steffes, Matt O'Keefe, and D. Palmer, "A Comprehensive Structural Analysis Process for Failure Assessment in Aircraft Lap-Joint Mimics using Intra-Modal Fusion of Eddy Current Data," *Research in Nondestructive Evaluation*, vol. 23, pp. 146-170, 2012.
12. N. Qaddoumi, Shroyer, A., and Zoughi, R., "Microwave Detection of Rust Under Paint and Composite Laminates" *Research in Nondestructive Evaluation* vol. 9, pp. 201-212 1997.
13. R. Zoughi, *Microwave Non-Destructive Testing and Evaluation*: Kluwer Academic Publishers, 2000.
14. M. Ghasr, B. Carroll, S. Kharkovsky, R. Zoughi, and R. Austin, "Size Evaluation of Corrosion Precursor Pitting Using Near-Field Millimeter Wave Nondestructive Testing Methods," *AIP Conference Proceedings*, vol. 760, pp. 547-553, 2005.
15. M.T. Ghasr, R. Zoughi, and G. Steffes, "Estimation of Corrosion Thickness under Paint using a Simple Millimeter Wave Probe," in *3rd International Conference on Electromagnetic Near-Field Characterization & Imaging (ICONIC 2007)*, St.Louis, USA, 2007, pp. 83-88.
16. M. T. Ghasr, D. Simms, and R. Zoughi, "Multimodal Solution for a Waveguide Radiating into Multilayered Structures - Dielectric Property and Thickness Evaluation," *Instrumentation and Measurement, IEEE Transactions on*, vol. 58, pp. 1505-1513, 2009.
17. J. S. Cargill, J.A. Pecina, S.M. Shepard, J.D. Weir, R. Zoughi, and A. V. Bray, "Nondestructive Testing for Corrosion Under Paint," *Materials Evaluation*, vol. 63, pp. 102-109, 2005.
18. V. Kaftandjian, Y. M. Zhu, O. Dupuis, D. Babot, "The combined use of the evidence theory and fuzzy logic for improving multimodal nondestructive testing systems," *Instrumentation and Measurement, IEEE Transactions on*, vol. 54, pp. 1968-1977, 2005.
19. X. E. Gros, L. Zheng, K. Tsukada, K. Hanasaki, "Experimenting with pixel-level NDT data fusion techniques," *Instrumentation and Measurement, IEEE Transactions on*, vol. 49, pp. 1083-1090, 2000.
20. B. Matuszewski, L. Shark, M. Varley, "Region-based wavelet fusion of ultrasonic, radiographic and shearography non-destructive testing images," in *15th World Conference on Nondestructive Testing*, Roma, Italy, 2000.
21. N. Francois, "A New Advanced Multitechnique Data Fusion Algorithm for NDT," in *15th World Congress on Nondestructive Testing* Roma, Italy, 2000.

22. X. E. Gros, Strachan, P., Lowden, D., and Edwards, I., "NDT data fusion," in *6th European Conference on Non-Destructive Testing*, 1994, pp. 355–360.
23. X. E. Gros, *NDT Data Fusion*: Halsted Press, 1996.
24. K. Sun, S. Udpa, L. Udpa, T. Xue, and W. Lord, "Registration issues in the fusion of eddy current and ultrasound NDE data using Q-transforms," *Review of Progress in QNDE*, vol. 15, pp. 813–820.
25. L. Zheng, D. S. Forsyth, J. P. Komorowski, K. Hanasaki, and T. Kirubarajan, "Survey: State of the Art in NDE Data Fusion Techniques," *Instrumentation and Measurement, IEEE Transactions on*, vol. 56, pp. 2435-2451, 2007.
26. T. Doppke and A. Bryant., "The Salt Spray Test: Past, Present, and Future," in *2nd Automotive Corrosion Prevention Conference*, 1983, pp. 57-72
27. S. De, K. Gupta, R. J. Stanley, G. Steffes, D. Palmer, R. Zoughi, "A data fusion based approach for evaluation of material loss in corroded aluminum panels," in *ITSC '09. 12th International IEEE Conference on Intelligent Transportation Systems*, St. Louis, MO, 2009, pp. 1-6.
28. K. Gupta, M. T. Ghasr, S. Kharkovsky, R. Zoughi, R. J. Stanley, A. Padwal, M. O'Keefe, D. Palmer, J. Blackshire, G. Steffes, and N. Wood, "Fusion of Microwave and Eddy Current Data for a Multi-Modal Approach in Evaluating Corrosion Under Paint and in Lap Joints," *AIP Conference Proceedings*, vol. 894, pp. 611-618, 2007.
29. K. Gupta, R. J. Stanley, M. T. Ghasr, S. Kharkovsky, R. Zoughi, G. Steffes, "Fusion of Multimodal NDE Data for Improved Corrosion Detection," in *3rd International Conference on Electromagnetic Near-Field Characterization and Imaging*, St. Louis, MO, 2007, pp. 407-412.
30. D. Hughes, N. Wang, T. Case, K. Donnell, R. Zoughi, R. Austin, and M. Novack, "Microwave nondestructive detection of corrosion under thin paint and primer in aluminum panels," *Special Issue of Subsurface Sensing Technologies and Applications: on Advances and Applications in Microwave and Millimeter Wave Nondestructive Evaluation* vol. 2, pp. 435-451, 2001.
31. D. Hall, S. A. H. McMullen, *Mathematical Techniques in Multisensor Data Fusion*, Artech House, Boston, MA, 2004.
32. C. Nyquist, K. Haghighi. "ANSYS Tutorial: Performing a Fatigue Analysis," Purdue University, 2007.

III. AUTOMATED BIOMEDICAL TEXT DETECTION METHOD IN SUPPORT OF CONTENT-BASED IMAGE RETRIEVAL SYSTEMS

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

^aDepartment of Electrical and Computer Engineering, Missouri University of Science and
Technology, Rolla, MO

^bLister Hill National Center for Biomedical Communications
National Library of Medicine, National Institutes of Health, DHHS
Bethesda, MD

ABSTRACT

In this research, automatic detection and recognition of text labels in biomedical publication images is investigated. The broader objective is to correlate regions of interest with text labels within a biomedical image to descriptive texts and derived biomedical concepts that may appear in the form of image captions or discussions. This paper presents image analysis and feature-based approaches to extract specific regions of interest within images in biomedical publications and to segment and recognize characters in those regions. A data set of 6515 characters were extracted and analyzed from text labels in 200 images contained in articles from the *British Journal of Oral and Maxillofacial Surgery*. Character recognition experiments using 185 geometry, exemplar-based correlation, Fourier descriptors, and profile-based basis function correlation features yielded a correct recognition rate as high as 94.27% with a support vector machine classifier, compared to a 75.90% correct recognition rate with a benchmark OCR technique.

1. INTRODUCTION

Essential information is often conveyed through images in biomedical publications. Images such as diagrams, tables, histograms and flowcharts are typically rich in content, summarizing the important results/methods presented in an article. Such images, when used in conjunction with either the image caption text or the image citations (mention) in the publications, can enhance the performance of Clinical Decision Support (CDS) systems [1-3]. In previous studies, the retrieval of biomedical information for CDS has been primarily text-based, limited mainly to bibliographic information. To that end, traditional Content-Based Image Retrieval (CBIR) provides automated indexing and retrieval of large image collections. Biomedical images for a given modality (e.g. MRI, Histology or X-Ray) are however very similar in nature. Therefore, existing CBIR techniques based only on the visual features (texture/shape) of images are not sufficient for accurate retrieval of biomedical images [4-7]. In addition to text (image captions/citations) and visual features, retrieving characters from biomedical images is necessary to obtain complementary information for enhanced CBIR.

As part of CBIR, regions of interest (ROIs) within biomedical images are those which contain illustrations such as arrows/symbols/text-labels. Commonly used methods for CBIR, however, do not utilize these ROIs. The semantic gap in biomedical image analysis can be reduced by characterizing the ROIs, as compared to only analyzing the image as a single entity [2-3]. Lehmann et al. proposed that three additional semantic abstraction levels are required of CBIR systems to understand complex medical

knowledge [8]. These include low-level medical information to understand the imaging modality, mid-level information obtained from ROIs, and high-level information obtained from the spatial relationships of ROIs [8-11].

Images in biomedical articles are generally of two types: regular images and graphical images. Regular images include MRIs, CT-Scans, X-rays, photographs and so forth. Graphical images, such as diagrams, statistical charts, flowcharts, and tables represent images that are created to either illustrate biomedical concepts or allow for biomedical data analysis. In previous studies involving CBIR, classification of graphical images into its various modalities have been successfully performed [12-14]. The information present within these graphical images must be extracted, however, to support both multimodal (image + text) biomedical information retrieval and CDS [15-16]. The study presented in this paper is focused on enhancing the retrieval of textual information from biomedical images.

As previously stated, authors often include several forms of annotations with their images. These annotations include but are not limited to text, text labels (e.g., A, B, and C), pointers (e.g., arrows and arrowheads) and symbols (e.g., asterisk). Such annotations are used to identify a ROI in the image. In previous CBIR-based research, arrow detection has been found to be successful in several types of biomedical images [16-20]. The integration of semantic annotation and information visualization was performed by Herold et.al to analyze fluorescence micrographs of tissue samples for CBIR applications [19]. Previous studies have analyzed biomedical images with text-like characteristics for both the extraction and recognition of textual characters [20-25].

Image features such as color, shape and texture have been previously investigated for analyzing the modality of biomedical images [26-36]. Techniques such as wavelet decomposition [37, 38], hierarchical methods [39], and multi-modal detection [26-27, 40] were used to analyze the graphical content in biomedical images. Hybrid feature extraction methods were examined to improve the information derived from the low-level features [26-28]. The framework of this study focuses on extracting characters/text from biomedical images, as compared to analyzing the whole image. Therefore, geometry and correlation-based features were developed in conjunction with texture features for automatic character recognition [14, 41]. Unlike conventional CBIR schemes that extract features from the entire image, in this research, features were extracted only from specific ROIs. Such ROIs were analyzed by first localizing and then recognizing text regions typical of annotations within biomedical images.

Previous research to both extract and classify text regions from biomedical images include methods/tools such as the BioText search engine [20], Yale Image Finder (YIF) [22-23], TextFinder [21] and FigText [28]. The BioText allows users to both search and browse articles for image captions [20]. For FigText, biomedical text extraction is conducted through not only image processing but also text recognition using a commercially available OCR [28]. More recent works in text recognition from biomedical images include the YIF. The YIF uses an iterative text detection algorithm followed by text recognition using a commercial OCR [22-23].

This research was built on previous methods by using an automated text detection and recognition technique. The method developed in this study was used to extract, parse and classify individual characters from text blocks found in biomedical images. The

proposed approach was compared to a benchmark OCR technique for character recognition. The flowchart in Figure 1.1 summarizes the main steps developed in this study.

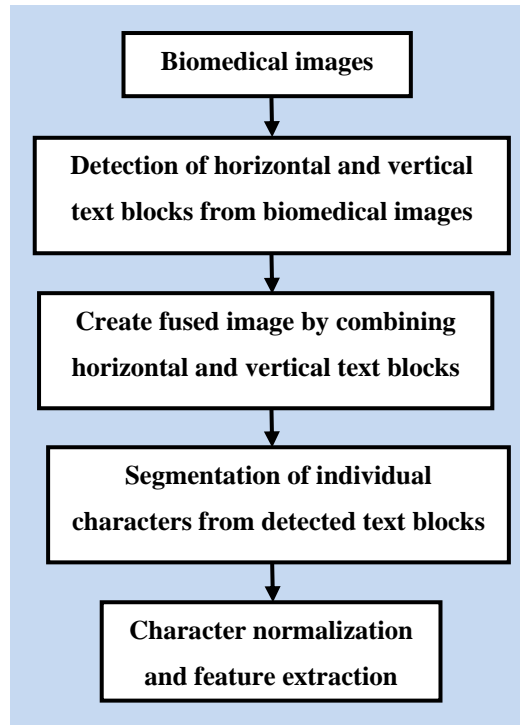


Figure 1.1: Overview of the image analysis steps.

The remainder of the paper is organized as follows. Section 2 presents the methodology used in this research. Section 3 presents the experiments performed while Section 4 contains both the experimental results and analysis. Section 5 summarizes the study presented in this paper.

2. METHODOLOGY

This section presents the methods investigated for automated detection and recognition of characters within biomedical images. Image analysis techniques explored for character detection include projection-based automated text block detection, followed by segmentation of the text blocks into individual characters (as shown in Figure 1.1). Features were extracted from these individual character images and used as inputs for an SVM-based classifier. The following sections contain details regarding each step in the overall text detection and recognition algorithm developed in this study.

2.1. DATA SET INVESTIGATED

For this research, a set of annotated images were examined from 2004-2005 issues of the *British Journal of Oral and Maxillofacial Surgery* [15]. More specifically, automatic text extraction, text analysis, and character recognition were performed on 200 annotated images. These images contained 2708 text blocks with 6515 characters. The experimental data set consisted of manual class labeling of characters in the alphabetic classes a-z and A-Z for character classification.

2.2. DETECTION OF HORIZONTAL AND VERTICAL TEXT BLOCKS

The first step in the development of the automated text detection and recognition system involved identifying the text found in biomedical images. Several studies have been performed to detect text within images. Ohya et al. presented an algorithm for text

detection [42]. This algorithm detected character components according to differences in gray-level and correlated the components with standard character forms. A limitation to this approach is that it could not address changes in either the color or the orientation of the text blocks present in the image. Zhong et al. introduced a RGB color histogram-based method for finding texts blocks in color images [43]. Text detection has also been achieved using both neural network-based approaches [44-45] and a support vector machine (SVM)-based method [46] to detect text within video.

A projection-based text region detection method was used in this study to extend the approach developed by Xu et al. [22-23]. The steps of the algorithm implemented in this study are presented as follows. Given an input binary image I (see Figure 2.1a):

Step 1: The image is searched for both lines and panel boundaries. Both are generally used for layout purposes and do not contain any diagnostic information. These lines and panel boundaries are removed from the images prior to text detection to reduce false text block detection. A summing filter was used in this step based on the implementation discussed in [22]. The summing filter found contiguous non-zero (“on”) pixels in both the horizontal and vertical directions of the input image I , thereby indicating either lines or panel boundaries [22]. The lines and panel boundaries detected in this step were removed from the input image I to obtain the image I_{LP} .

Step 2: Horizontal projections for I_{LP} were computed to find text regions. If I_{LP} has dimensions $D_R \times D_C$, where D_R is the number of rows and D_C the number of columns, then the horizontal projections could be denoted as $B_H(1, 2, \dots, D_C)$. If $I_{LP}(i, j)$

represented each pixel of I_{LP} , then the horizontal projection B_H computed could be defined as:

$$B_H(j) = \sum_{i=1}^{D_C} I_{LP}(i, j) \quad (1)$$

Step 3: The horizontal projection (B_H) was thresholded to obtain the columns of I_{LP} that corresponded to potential vertical text blocks present in the input image I . The threshold $\tau_{projection}$ was calculated adaptively as: $\tau_{projection} = \min(B_H) + 0.275 * (\max(B_H) - \min(B_H))$ [47]. The factor 0.275 was obtained experimentally. The threshold was applied to obtain the thresholded vertical projection as:

$$B_H(j) = \begin{cases} B_H(j), & \text{if } B_H(j) > \tau_{projection} \\ 0, & \text{if } B_H(j) < \tau_{projection} \end{cases} \quad (2)$$

for $j = 1, 2, \dots, D_C$.

In this manner, all the non-zero values of B_H (which actually denote the sum of columns of I_{LP} , shown in Equation 1) corresponded to the potential vertical text block columns in I_{LP} . The vertical segmented image I_{LP}^V was obtained using these thresholded projection values as follows:

$$I_{LP}^V(i, j) = \begin{cases} I_{LP}(i, j), & \text{if } B_H(j) > 0 \\ 0, & \text{otherwise} \end{cases} \quad (3)$$

for $i = 1, 2, \dots, D_R$ and $j = 1, 2, \dots, D_C$.

Next, the vertical segmented image I_{LP}^V was subjected to a connected-component analysis to obtain the regions corresponding to the actual vertical text blocks. The connected-components were thresholded based on the major axis length of the ellipse that contained the connected-component region. Connected component regions with major axis lengths ranging from 15-100 (determined empirically) were selected as candidates for creating the vertical text blocks. These were denoted as $I_{VS_1}, I_{VS_2}, \dots, I_{VS_\delta}$ for the ‘ δ ’ vertical text blocks.

Step 4: Vertical projections were obtained from I_{LP} and denoted as $B_V(1, 2, \dots, D_R)$. If $I_{LP}(i, j)$ represented each pixel of I_{LP} , then the vertical projection B_V could be defined as:

$$B_V(i) = \sum_{j=1}^{D_R} I_{LP}(i, j) \quad (4)$$

Step 5: Similar to Step 3, the vertical projections were thresholded to obtain the horizontal segmented image I_{LP}^H followed by a connected component analysis to obtain the horizontal text blocks denoted as $I_{HS_1}, I_{HS_2}, \dots, I_{HS_\beta}$ for the ‘ β ’ horizontal text blocks.

Thus, the outputs of the automatic text block detection step include both the vertical text blocks ($I_{VS_1}, I_{VS_2}, \dots, I_{VS_\delta}$) and the horizontal text blocks ($I_{HS_1}, I_{HS_2}, \dots, I_{HS_\beta}$). A sample input image is given in Figure 2.1a, with the horizontal and vertical segmented blocks given in Figure 2.1b.

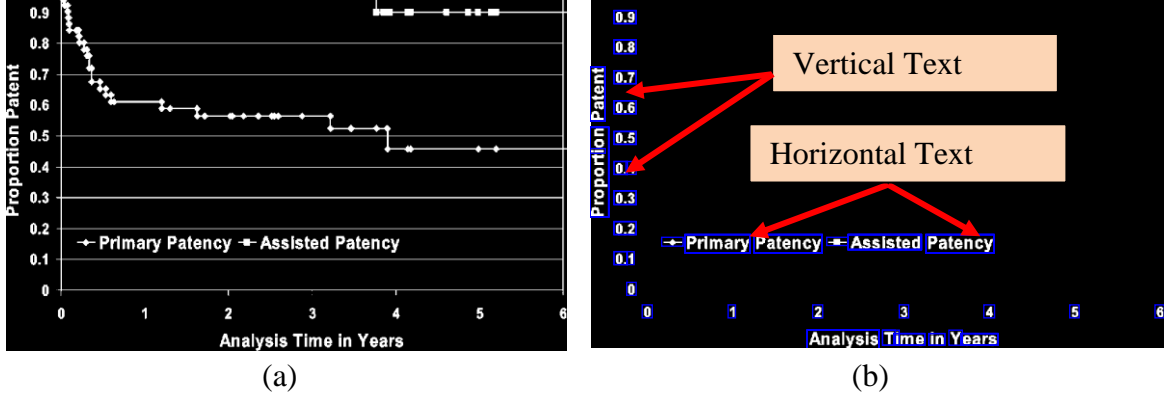


Figure 2.1: Text Block detection

(a) Input image, (b) Resultant image from the text detection step, showing the horizontal and vertical text blocks detected.

2.3. FUSED IMAGE GENERATION

After the horizontal and vertical text blocks were detected, the next step was to combine them into a fused image, I_F (see Figure 2.2b). The fusion image was created such that all of the text blocks were aligned one below the other (see Figure 2.2b) rather than scattered over the entire image (see Figure 2.2a). This was done to reduce the space required to store the image by eliminating all the non-text areas. It also facilitated faster processing during the segmentation of individual characters within the text blocks.

To obtain I_F , first the horizontal text blocks were placed one below the other with a 5- pixel-gap (5 rows of zero pixels between each text block). Second, the vertical text blocks were rotated 90° clockwise and aligned below the horizontal text blocks with the same 5-pixel-gap between each block. The fused image created in this manner is shown in Figure 2.2b and was represented as:

$$I_F = (I_{VS_1}, I_{VS_2}, \dots, I_{VS_\delta}) \cup (I_{HS_1}, I_{HS_2}, \dots, I_{HS_\beta}) \quad (5)$$



Figure 2.2: Creation of fused image I_F .

(a) Text block detected image, (b) Horizontal and vertical text blocks fused image.

2.4. SEGMENTATION OF INDIVIDUAL CHARACTERS

After creating I_F , all the characters within I_F were extracted into separate images and saved for further analysis. To perform this, the first step involved using the 5-pixel gap between every text block to segment out the text blocks. For illustrative purposes, the 5-pixel gap between each text block is given in red in Figure 2.3b. As shown, these red lines provide segmenting lines between each block of text. The final segmented image is given in Figure 2.3c. Let $I_{S_1}, I_{S_2}, \dots, I_{S_n}$ denote each of these segments for the ' n ' different segments. Note that here ' n ' refers to the total number of text blocks (segments) in the fusion image I_F and can be obtained as $n = \delta + \beta$, where δ is the number of vertical text blocks and β is the number of horizontal text blocks as stated previously.

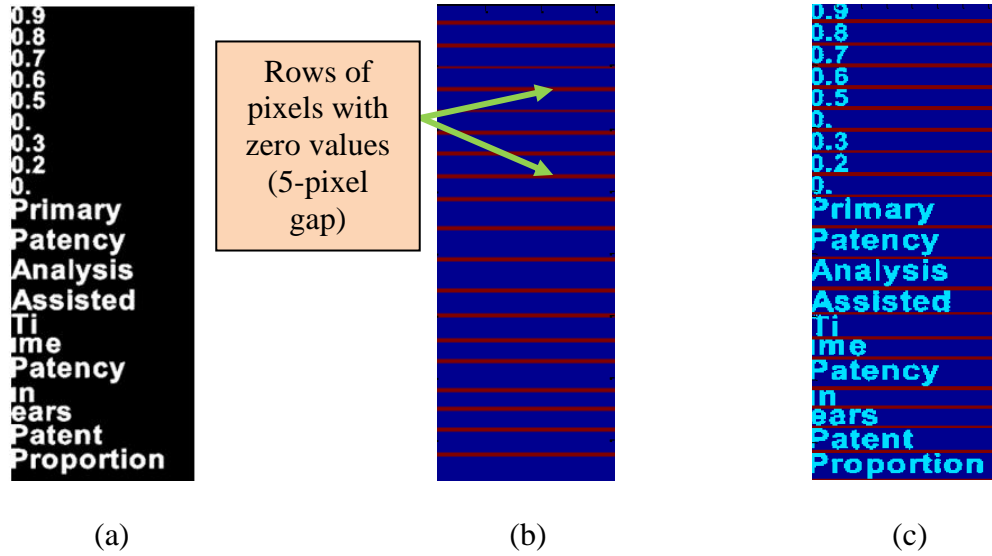


Figure 2.3: First step of segmenting individual characters.
 (a) Fused (Horizontal and Vertical text) image, (b) Red lines representing the separating line between each row of text block in Figure 2.3a, (c) The fused image being broken up into segmented rows.

The second step was to segment each character within the segmented text blocks $(I_{S_1}, I_{S_2}, \dots, I_{S_n})$. For this step, each segmented text block I_{S_n} was scanned to find zero-pixel columns (see Figure 2.4b). These zero-pixel columns (columns with all row pixels having zero-value) acted as segmenting lines between each character in the text block. As shown in Figure 2.4b, the region between two subsequent zero-pixel columns represented a character and was extracted (segmented). For the first text block I_{S_1} , each segmented character image was denoted as $I_{C_1S_1}, I_{C_2S_1}, \dots, I_{C_mS_1}$ for the ‘ m ’ characters within I_{S_1} , and so forth. To generalize, each segmented character image was represented as $I_{C_mS_n}$, where ‘ n ’ was the segment (text block) number and ‘ m ’ represented the m^{th} character in the n^{th} segment. For example, in Figures 2.4a and 2.4b, the 16th text block ($I_{S_{16}}$) is shown to be segmented into its seven characters, $I_{C_1S_{16}}, I_{C_2S_{16}}, \dots, I_{C_7S_{16}}$. Figure 2.4c shows the text blocks with segmented individual character images.

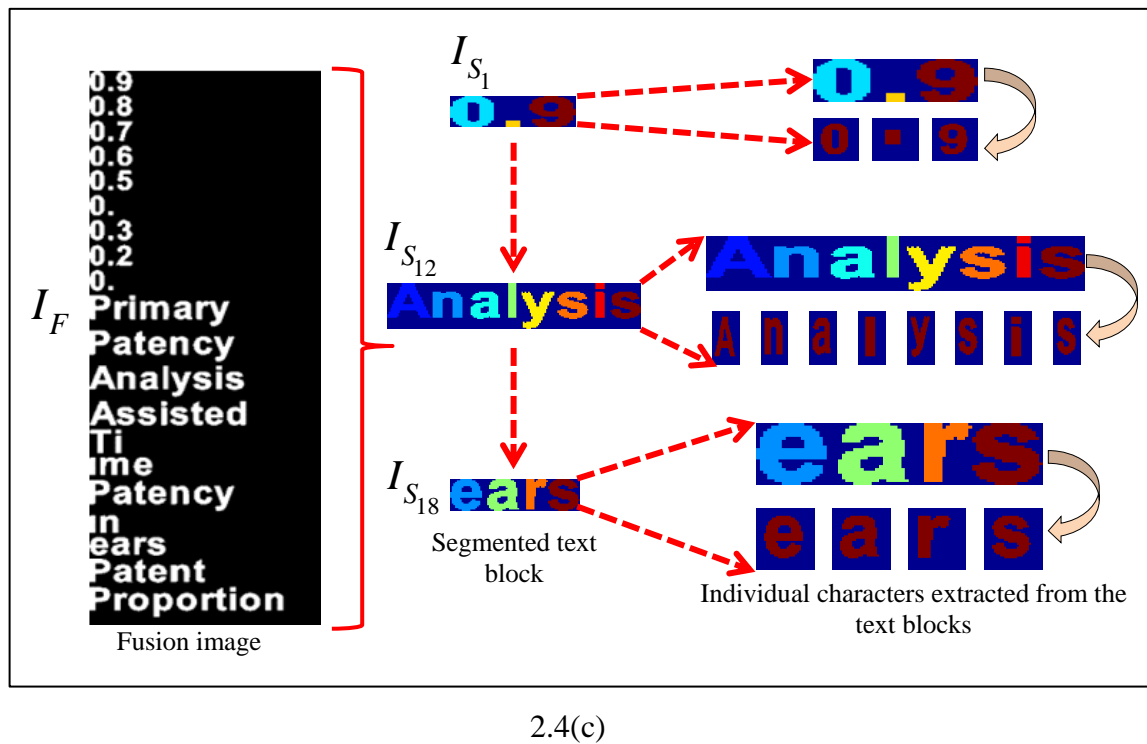
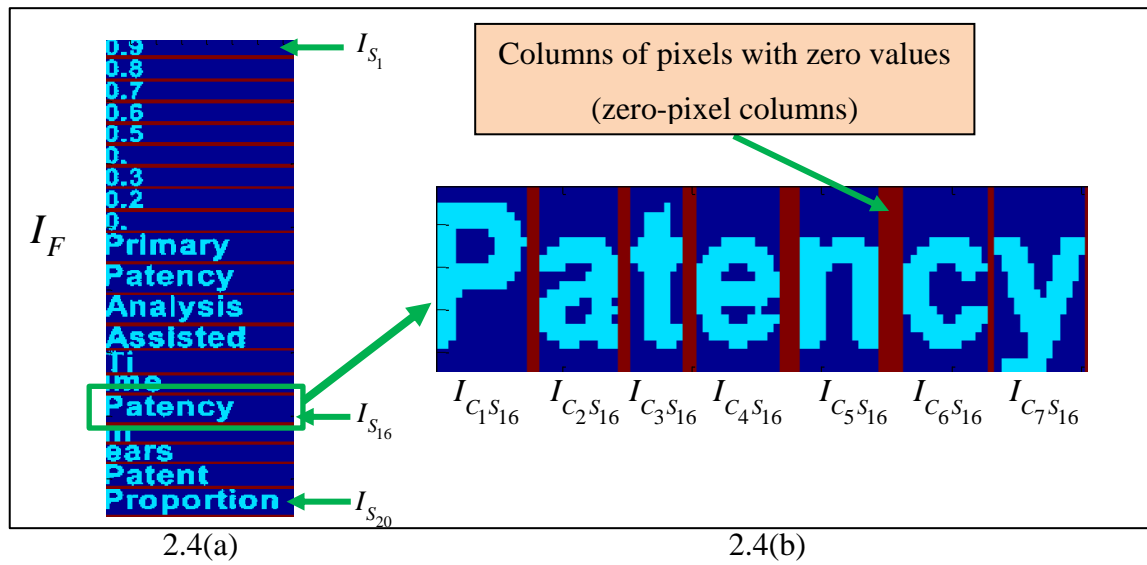


Figure 2.4: Individual character segmentation.

(a)-(c): Overview of extracting individual characters from the fused text block image.

2.5. CHARACTER NORMALIZATION AND FEATURE EXTRACTION

After all the characters were extracted from I_F into separate character images $I_{C_m S_n}$, these character images were normalized to a size of 42x24 pixels. For the characters extracted from the first text block I_{S_1} , the normalized character images were denoted as $I_{C_1 S_1}^Z, I_{C_2 S_1}^Z, \dots, I_{C_m S_1}^Z$ and so forth. A total of 185 features including geometry-based, region-based, exemplar-based, Fourier descriptors and profile-based correlation were extracted for each normalized character image $I_{C_m S_n}^Z$. An overview of the features with associated variable names is given in Table 2.1.

Table 2.1: Feature Description

Feature set	Variable Name	Measure
Geometry Features (annotated as G#, where # is the feature number)	G1-G9	Number of horizontal lines in window
	G10-G18	Number of vertical lines in window
	G19-G27	Number of right diagonal lines in window
	G28-G36	Number of left diagonal lines in window
	G37-G45	Length of horizontal lines in window
	G46-G54	Length of vertical lines in window
	G55-G63	Length of left diagonal lines in window
	G64-G72	Length of right diagonal lines in window
	G73-G81	Non-zero/unoccupied area in window
	G82	Euler number of whole image
	G83	Top/Bottom ratio of non-zero pixels in each half of image
	G84	Area
	G85	Centroid
	G86	Major axis length
	G87	Minor axis length
	G88	Eccentricity
	G89	Orientation
	G90	Convex area
	G91	Filled area
	G92	Equivalent Diameter
Exemplar-based correlation feature (annotated as E#, where # is the feature number)	E1- E26	Correlation to characters 'A'-'Z' (capital alphabets)
	E27-E52	Correlation to characters 'a'-'z' (small alphabets)
Fourier Descriptors (annotated as F#, where # is the feature number)	F1-F14	Frequency domain features
Profile-based basis function correlation features (annotated as W#, where # is the feature number)	W1-W24	Weighted density distribution

2.5.1. Geometry Features. One of the principal ways to differentiate one character from another is to analyze either the geometry or the shape of a character. Geometrical features are quite traditional in nature for character recognition; they have been used successfully by prior studies [48-51]. Geometrical features investigated in this study include structural features (G1-G83) and shape features (G84-G95). To compute the structural features, the character image $I_{C_m S_n}^Z$ was partitioned into nine equal sized windows, each of size 14x8 pixels. The number, length, and types of lines found within each of the nine windows (see Figure 2.5) were used for feature calculations [52].

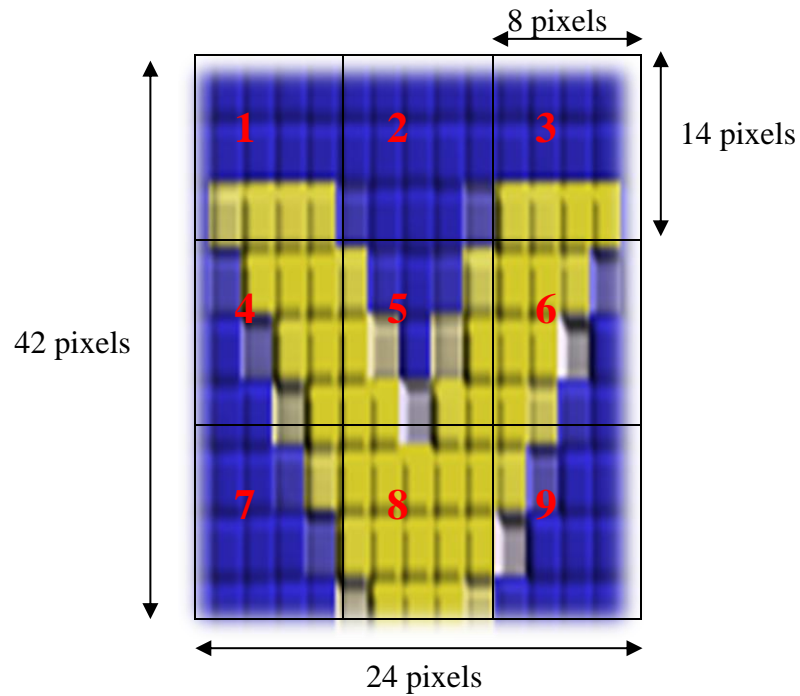


Figure 2.5: Breaking up original character into 9 equal sized windows

Nine structural features were computed for each of the above windows (1-9). Details of the geometry features used in this study can be found in [34] and are overviewed here for reference purposes.

The structural features extracted were:

- *Number of horizontal lines (G1-G9)*: For each window, the number of horizontal lines represented the number of rows with non-zero pixels for all columns.
- *Number of vertical lines (G10-G18)*: For each window, the number of vertical lines represented the number of columns with non-zero pixels for all rows.
- *Number of right diagonal lines (G19-G27)*: For each window, the number of right diagonal lines represented the number of diagonal lines that started on the left bottom and ended on the right top, with non-zero pixels for all diagonal pixels.
- *Number of left diagonal lines (G28-G36)*: For each window, number of left diagonal lines represented the number of diagonal lines that started on the left top and ended on the right bottom, with non-zero pixels for all diagonal pixels.
- *Normalized number of all horizontal lines (G37-G45)* represented the number of horizontal lines, normalized over the non-zero area of the window.
- *Normalized number of all vertical lines (G46-G54)* represented the number of vertical lines, normalized over the non-zero area of the window.
- *Normalized number of all left diagonal lines (G55-G63)* represented the number of left diagonal lines, normalized over the non-zero area of the window.
- *Normalized number of all right diagonal lines (G64-G72)* represented the number of right diagonal lines, normalized over the non-zero area of the window.
- *Normalized non-zero area of each window (G73- G81)*: For each window normalized non-zero area represented the number of non-zero pixels in the window normalized over the area of the window.

Eighty-one features (nine features computed for each window) were obtained using the above mentioned windowing scheme. Apart from the above features, two additional structural features were computed: *Euler number* (G82), and the *Ratio of the top half of the character to the bottom half of the character* (G83). Euler number is defined as the difference of the number of objects and holes in an object. The ratio of the top and bottom halves (Γ) was obtained by first partitioning the normalized image into two halves (shown in Figure 2.6), and then finding the non-zero pixels for the top half (Γ_{top}) and bottom half (Γ_{bottom}). The number of non-zero pixels for the top half (Γ_{top}) was represented as Equation 6, while Equation 7 represented the number of non-zero pixels for the bottom half (Γ_{bottom}). Equation 8 defined the ratio of the top half to the bottom half.

If $I_{C_m S_n}^Z(i, j)$ represents each pixel of the normalized character image, $I_{C_m S_n}^Z$, then $i=(1,2,...,R)$ are the rows and $j=(1,2,...,C)$ are the columns of the normalized character image. For this study, we fixed $R=42$ and $C=24$, since all the normalized character images were of size 42x24 pixels.

$$\Gamma_{top} = \sum_{i=1}^{R/2} \sum_{j=1}^C I_{C_m S_n}^Z(i, j), \text{ where } I_{C_m S_n}^Z(i, j) \neq 0 \quad (6)$$

$$\Gamma_{bottom} = \sum_{i=(R/2)+1}^R \sum_{j=1}^C I_{C_m S_n}^Z(i, j), \text{ where } I_{C_m S_n}^Z(i, j) \neq 0 \quad (7)$$

$$\Gamma = \frac{\Gamma_{top}}{\Gamma_{bottom}} \quad (8)$$

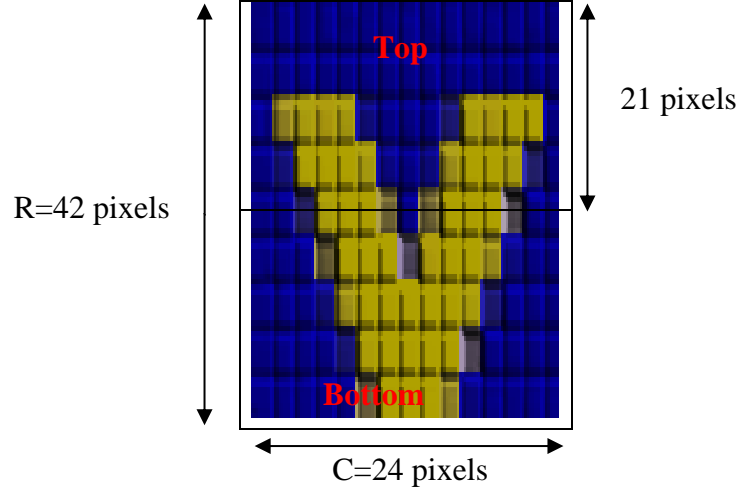


Figure 2.6: Breaking up the normalized character image into two halves.

The next set of geometry features extracted included the shape features. These features analyzed the region occupied by the character. Shape features have been used in prior studies for the recognition of Chinese characters [53], Devnagari characters, [54] and English characters [55-56]. The following shape features were extracted for each normalized character image $I_{C_m S_n}^Z$.

- *Area* (G84) represented the area of the non-zero pixels in $I_{C_m S_n}^Z$. It can be represented as:

$$Area = \sum_{i=1}^R \sum_{j=1}^C I_{C_m S_n}^Z(i, j), \quad \text{where } I_{C_m S_n}^Z(i, j) \neq 0 \quad (9)$$

- *Centroid* (G85) represented the center of mass of $I_{C_m S_n}^Z$ and can be represented as:

$$Centroid = \left\{ \frac{\sum_{i=1}^R (i)}{R}, \frac{\sum_{j=1}^C (j)}{C} \right\}, \quad \text{where } I_{C_m S_n}^Z(i, j) \neq 0 \quad (10)$$

- *MajorAxisLength* (G86) represented the length of the major axis of the ellipse with the same normalized second central moments as $I_{C_m S_n}^Z$.
- *Minor Axis Length* (G87) represented the length of the minor axis of the ellipse with the same normalized second central moments as $I_{C_m S_n}^Z$;
- *Eccentricity* (G88) represented the eccentricity of the ellipse with the same second-moments as the normalized character image $I_{C_m S_n}^Z$.
- *Orientation* (G89) represented the angle (-90° to $+90^\circ$) between the x-axis and the major axis of the ellipse with the same second-moments as the normalized character image $I_{C_m S_n}^Z$.
- *ConvexArea* (G90) was the number of pixels in the Convex Image. Convex Image was the image that specified the Convex Hull. Convex Hull was defined as the smallest convex polygon that could contain the normalized character image $I_{C_m S_n}^Z$;
- *FilledArea* (G91) represented the number of pixels specifying the Filled Image of the normalized character image. A Filled Image was one which had all of the holes inside it filled.
- *EquivalentDiameter* (G92) specified the diameter of the circle with the same area as the normalized character image $I_{C_m S_n}^Z$.
- *Solidity* (G93) specified the proportion of the pixels in the Convex Hull that were also within the region of the normalized character image. It could be computed as:

$$Solidity = \frac{Area}{ConvexArea} \quad (11)$$

- *Extent* (G94) specified the ratio of the non-zero pixels within the normalized character image $I_{C_m S_n}^Z$ to the area of bounding box of the same normalized character image. It was computed as the area of the normalized character image divided by the area of the bounding box that contained the characters within the same image (*BoundingBoxArea*). It was represented as:

$$Extent = \frac{Area}{BoundingBoxArea} \quad (12)$$

- *Perimeter* (G95) was the distance along the boundary of the continuous “on” (contiguous non-zero) pixels in the normalized character image $I_{C_m S_n}^Z$.

2.5.2. Exemplar-based Correlation Features. Exemplar images corresponding to each alphabet of the English language (A-Z and a-z) were used to compute the exemplar-based correlation features. Prior work in text recognition has shown that both template-matching and distance-based methods provide excellent results for automatic text classification [55, 57]. The nearest-neighbor-based distance metric by itself provided an accuracy of 99% for a very large dataset as demonstrated by Smith et al. [57]. In this research, a template-matching approach was used by computing the correlation between the exemplar images of English alphabets with the normalized character image $I_{C_m S_n}^Z$. These exemplar images were found online through the freely available Matlab[®] Exchange files database [58], which were comparable to the characters present in the biomedical images used in this study. The exemplar images contained fifty-two images of

the characters in the English character set (A-Z and a-z). Figure 2.7 illustrates a sample exemplar image for the alphabet ‘A’.



Figure 2.7: Sample exemplar image for alphabet ‘A’ [58].

During feature extraction, the correlation coefficient was computed between each normalized character image ($I_{C_m S_n}^Z$) and the exemplar images of the English alphabets (A-Z and a-z). If E_A represented the exemplar image for ‘A’ (Figure 2.7), then the correlation coefficient for the character image ($I_{C_m S_n}^Z$) with respect to ‘A’ could be represented as ρ_A and computed as:

$$\rho_A = \frac{\sum_{i=1}^R \sum_{j=1}^C (I_{C_m S_n}^Z(i, j) - \mu_{C_m S_n}^Z) (E_A(i, j) - \mu_A)}{\sqrt{\left(\sum_{i=1}^R \sum_{j=1}^C (I_{C_m S_n}^Z(i, j) - \mu_{C_m S_n}^Z)^2 \right)} \sqrt{\left(\sum_{i=1}^R \sum_{j=1}^C (E_A(i, j) - \mu_A)^2 \right)}} \quad (13)$$

where $I_{C_m S_n}^Z(i, j) \neq 0$ and $E_A(i, j) \neq 0$. Also, $\mu_{C_m S_n}^Z$ and μ_A were the mean of $I_{C_m S_n}^Z$ and E_A , respectively. Since both $I_{C_m S_n}^Z$ and E_A were binary images, the mean values represented the number of non-zero pixels in the image divided by the area of the binary image. Note that the exemplar images were of the same size as the normalized character image size (42x24 pixels). In this way, each normalized character’s correlation with all of the fifty-two alphabets (A-Z and a-z) were computed and used as features. Each character

would thus have fifty-two exemplar-based correlation features corresponding to the fifty-two correlation values, $\rho_a, \rho_b, \dots, \rho_z$ and $\rho_A, \rho_B, \dots, \rho_Z$.

2.5.3. Profile-based Basis Function Correlation Features. In previous research, profile-based Weighted-Density Distribution (WDD) basis functions were used on 1-D profiles of metal detector signals for landmine discrimination [59]. These WDD basis functions were also applied for dermatology skin lesion discrimination based on a 1-D histogram representation of skin lesions [60]. In a related study, these functions were applied to 1-D projection profiles of arrow-like objects to extract shape information for object discrimination in CBIR [16]. Because the basis functions provided shape-related information to determine correlation-based, size-variant, and spatially distributed features from 1-D profiles for object discrimination, they were used in the current study as well. The profile of each normalized character image $I_{C_m S_n}^Z$ was obtained by projecting the character image upon planes at 90° and 0° to create the profiles at 90° (Figure 2.8b) and 0° (Figure 2.8c) respectively. The profile was denoted as P_ϕ for projection angle ϕ .

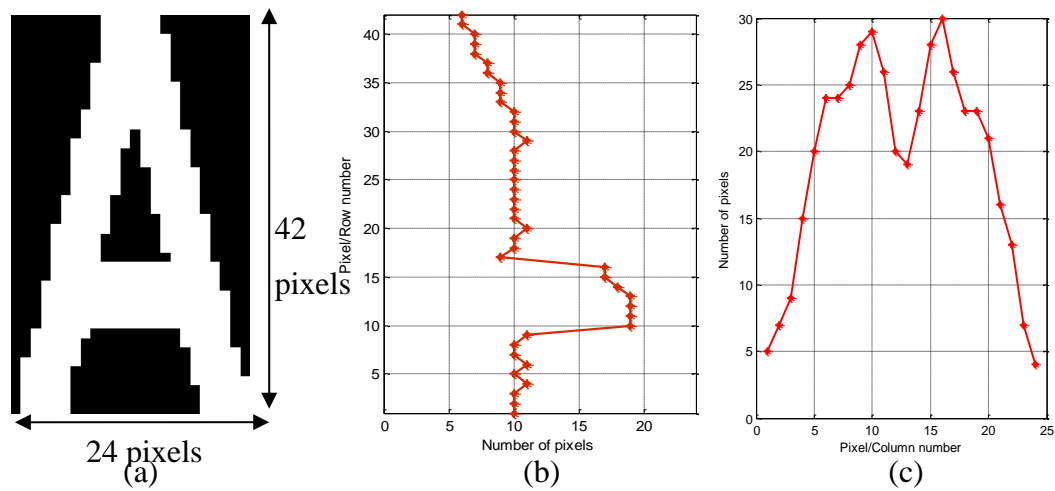


Figure 2.8: Profiles of character images.

- (a) Sample normalized character image, (b) Profile of character at 90° projection, (c) Profile of character at 0° projection.

For a normalized character image $I_{C_m S_n}^Z$, the profile for $\phi=90^\circ$ projection (denoted as P_{90}) was obtained as Equation 14. Similar profile data (denoted as P_0) was obtained for the $\phi=0^\circ$ projection too.

$$P_{90}(i) = \sum_{j=1}^C I_{C_m S_n}^Z(i, j) \quad (14)$$

for $i = (1, 2, \dots, R)$. Here, $P_{90} = \{P_{90}(1), P_{90}(2), \dots, P_{90}(R)\}$ were the sequences of profile values, represented as a vector. Correlation-based features were extracted by correlating this profile of the character (P_ϕ) with the WDD basis functions [62], shown in Figure 2.9. Let λ_1 denote the WDD function in Figure 2.9(a), λ_2 denote the WDD function in Figure 2.9(b), and so forth.

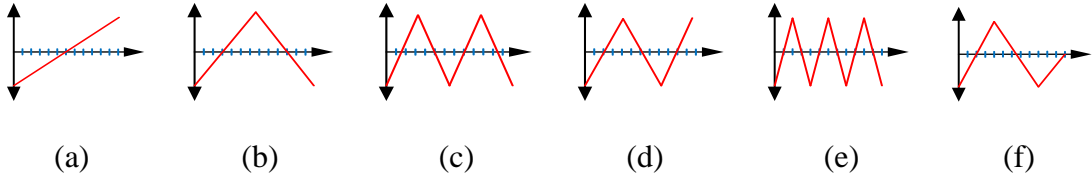


Figure 2.9: The WDD functions used.

The first six WDD features W_1, W_2, \dots, W_6 are computed using the profile vector P_ϕ according to the expression:

$$W_k = \sum_{i=1}^R P_\phi(i) D_k(i) \quad (15)$$

for $k = 1, 2, \dots, 6$, representing the six WDD functions in Figure 2.9. Six additional features W_7, W_8, \dots, W_{12} are computed by correlating the WDD functions with the sequence of absolute differences between the profile values as shown in Equation 16.

$$W_{k+6} = \sum_{i=1}^R |P_{\phi}(i) - P_{\phi}(i-1)| D_k(i) \quad (16)$$

In this manner, a total of 24 WDD features are obtained for the two different projection angles: a) $\phi=90^\circ$ projection (W1-W12), b) $\phi=0^\circ$ projection (W13-W24).

2.5.4. Fourier Descriptors. Previous research on text recognition used Fourier transforms extensively for automatic classification of text. Fourier descriptors provide means by which the shape of a character can be characterized in terms of its spatial frequency content. These descriptors are also invariant to translation, rotation and scale and hence provide a good estimation of characters for automatic text detection [62-65]. For this research, based on the algorithm in [64], the boundary coordinates of the normalized character image ($I_{C_m S_n}^Z$) were obtained first using a Sobel edge detector and can be represented as $\{x(g), y(g)\}$. For $g=1, 2, \dots, L$ edge points, $x(g)$ denoted the x-coordinates and $y(g)$ denoted the y-coordinates. Second, the boundary coordinates were separately (for the x- and y-coordinates) transformed to the frequency domain using a 1-D Fast Fourier Transform. The Fourier coefficients $a(u)$ and $b(u)$ thus obtained can be represented as:

$$a(u) = \frac{1}{L} \sum_{g=1}^L x(g) e^{-jn\omega_0 g} \quad (17)$$

$$b(u) = \frac{1}{L} \sum_{g=1}^L y(g) e^{-jn\omega_0 g} \quad (18)$$

where, $\omega_0 = 2\pi / L$. Third, the frequency domain components were transformed to give the required invariance as shown in Equation 19 [64].

$$r(u) = \sqrt{|a(u)|^2 + |b(u)|^2} \quad (19)$$

Discarding $r(0)$ (since it is 1), $r(u)$ has been shown to be invariant to translation and rotation [67]. Finally, $s(u)$ was computed as shown in Equation 20, where $r(u)$ is normalized to $r(1)$, thus removing dependence of $r(u)$ on the size of the character [64].

$$s(u) = r(u) / r(1) \quad (20)$$

Here the first 15 coefficients of $s(u)$, with $u = 1, 2, \dots, 15$ could be used as image features and termed as Fourier features. However, since the first coefficient is always 1, we use only the 2nd-15th coefficients (F1-F14).

Finally, the feature vector (185x1) for each normalized character image $I_{C_m S_n}^Z$ was created as:

$$F_{C_m S_n} = [G_1, \dots, G_{95}, E_1, \dots, E_{52}, F_1, \dots, F_{14}, W_1, \dots, W_{24}]_{C_m S_n}^T \quad (21)$$

2.6 GROUND TRUTHING OF EXTRACTED CHARACTERS

Using the individual characters segmented with the methods presented in the previous sections, the next task was to assign text labels to each of the characters. The ground truthing (assigning labels) of the extracted characters was needed to assist the classifier for automated recognition of the characters. For this step, a ground truthing tool was developed in Matlab[®]. In the tool, the extracted characters were displayed one by one (see Figure 2.10), and the user was asked to assign a label to the displayed character. After label assignment, the character blocks were stored in a database along with the labels. It should be noted that for this study, only English alphabets (A-Z and a-z) were

used as labels during ground truthing. Thus when the tool asks the user: ‘Which letter is this?’ (Figure 2.10) the user must specify which English alphabet is being represented by the character displayed.

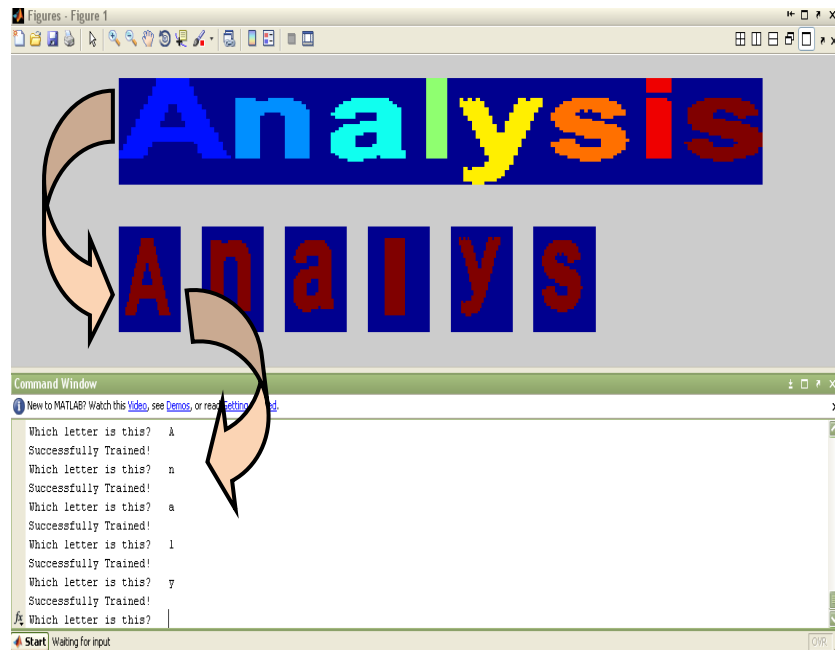


Figure 2.10: Assigning labels to extracted characters

If a numeric character or a symbol appeared during ground truthing (see Figure 2.11), and even if the user assigned a label to this number or symbol, the system did not save this character or its label. For example, in Figure 2.11, three characters are provided: ‘0’, ‘.’ and ‘9’. As shown, the user assigned ‘0’ to the first extracted character; however the ground truthing tool did not save this label in the database. In this manner, all the non-alphabetic (numbers and symbols) characters were excluded from the experimental data set for character recognition. This also limited the automated character recognition developed in this research to recognize alphabetic characters only (A-Z and a-z). If a numeric character or any symbol shows up during ground truthing (see Figure 2.11), and

if the user assigns this character a number or a symbol, the system even then does not save this character and the following message is displayed: ‘The letter was not trained’. This is a check to make sure that the system does not confuse English alphabets with numbers or symbols during the recognition step.



Figure 2.11: Assigning labels to extracted characters (numbers)

2.7 CLASSIFICATION OF EXTRACTED CHARACTERS

For this study, a Support Vector Machine (SVM) - based classifier was investigated based on the implementation in [68-69]. In general, the goal of SVM-based classifiers is to find an optimal hyperplane, which acts as a decision function to classify data in high dimensions. For classes that are not linearly separable, the SVM optimizes to find a hyperplane that maximizes the class separation while minimizing a quantity that is directly proportional to the misclassified instances. In particular, the design of a SVM-based classifier relates to the minimization of a cost function (kernel) [70]. Since for this

study we have a significantly large number of features (185), the linear kernel for the SVM classifier was used [68]. The penalty parameter to be used in the SVM classifier for the linear kernel was obtained using the ten-fold cross validation method explained in the experimental results section [68].

3. EXPERIMENTS PERFORMED

Four sets of experiments were performed in this study. This section outlines the four sets of experiments as follows.

3.1. EVALUATION OF THE TEXT BLOCK DETECTION STEP

The detected text blocks were evaluated in this set of experiments. The 200 biomedical images used in this study had 2708 text blocks that were manually located and recorded by the authors. For performance evaluation, text blocks detected automatically were compared to the manually recorded text blocks to evaluate text block detection accuracy.

3.2. BENCHMARK OCR EVALUATION

For the second set of experiments, the characters automatically extracted using the individual character segmentation algorithm were evaluated with the benchmark OCR tool-Tesseract [71]. The 6515 extracted characters were subjected to a ten-fold cross validation test with Tesseract. This was done to compare the performance of the benchmark OCR with the automated text detection method developed in this research.

3.3. EVALUATION OF THE METHOD DEVELOPED

The third set of experiments was done to test the performance of the automatic text recognition method developed in this research. For test images, after text block

detection and individual character segmentation, features were extracted from each normalized character image and used as inputs to the SVM-based classifier to perform the automatic text recognition. The results from the SVM-based classifier were compared to the manual labels assigned to the characters for performance evaluation.

3.4. FEATURE EVALUATION AND SELECTION

For this set of experiments, the goal was to find a statistically significant set of features as compared to all the features. This was mainly done to reduce the number of features (dimensions) and to find the set of features which had the best discriminative abilities for automatic character recognition. A multinomial logistic regression-based approach was used to find the best set of features.

4. EXPERIMENTAL RESULTS AND ANALYSIS

As previously stated, the automatic text detection and recognition algorithm was applied for character recognition on a data set of 200 biomedical images from which 6515 alphabetic characters were extracted. A total of 185 features were obtained from each of these 6515 characters. Figure 4.1 presents the distribution of the labels of the characters obtained using the ground truthing tool. Since in this research only lowercase and uppercase English alphabetic characters were extracted and evaluated, non-alphabetic characters such as symbols and numbers could not be automatically recognized. If such non-alphabetic characters are encountered during character recognition, they would be assigned to the most resembling English alphabet. For example, the number '0' could be assigned to the alphabet 'O' (uppercase) or 'o' (lowercase), while the number '1' could be assigned to 'I' (uppercase) or 'I' (lowercase) (see Figure 4.3b for examples).

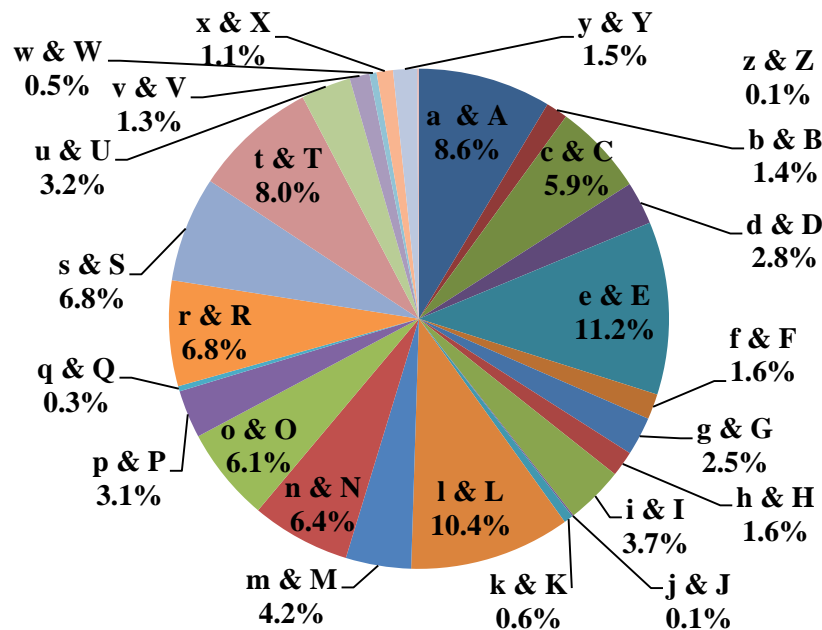


Figure 4.1: Distribution of character labels across the alphabet classes

For the first set of experiments, the text blocks automatically detected ($I_{s_1}, I_{s_2}, \dots, I_{s_n}$) using the text detection method were compared to the text blocks manually recorded by the authors. The performance evaluation of the automatic text block detection was done in the following way: A true positive (tp) defined a case when a text block detected automatically matched one of the text blocks manually recorded for the same image; A false positive (fp) defines a case when an automatically detected text block does not match with any of the manually recorded text blocks for the same image; False negatives (fn) are cases when the text blocks which were manually recorded were not detected automatically. In Table 4.1, the tp , fp and fn values shown are normalized to the number of actual text blocks manually recorded, as $TP = \frac{tp}{2708}$, $FP = \frac{fp}{2708}$ and $FN = \frac{fn}{2708}$. In this manner, all the text blocks detected automatically from the 200 images were compared to the 2708 manually recorded text blocks. The Precision and Recall values were also computed and can be mathematically expressed as Equations 22-23. Precision and Recall gives a measure of relevance of the text blocks detected automatically to the text blocks manually recorded [72]. Table 4.1 shows the summarized results of the performance of the automatic text block detection step for this research.

$$Precision = \frac{TP}{TP + FP} \quad (22)$$

$$Recall = \frac{TP}{TP + FN} \quad (23)$$

Table 4.1: Accuracy metrics for automatic text block detection

True Positive (TP)	False Positive (FP)	Precision	Recall
0.9632	0.0200	0.9800	0.9900

From Table 4.1, the automated text detection algorithm developed in this study has a high correct text block detection rate (high TP) with a relatively low false positive rate. Also, high precision/recall values demonstrate that text blocks could be successfully discriminated from non-text blocks and that the text blocks were correctly identified (and not omitted) in the selection/segmentation process within the biomedical images.

The second set of experiments was done to evaluate the benchmark OCR tool-Tesseract for the set of characters extracted. For this experiment, the 6515 characters automatically extracted were divided into ten independent character sets to perform the ten-fold cross validation test. The character images $I_{C_m S_n}^Z$ extracted using the automatic text detection step acted as inputs to the OCR tool-Tesseract. To evaluate the accuracy of Tesseract, the following scoring method was used: In a round-robin manner, nine out of the ten character sets were used for training and the remaining one set was used for testing. For each character test set, the accuracy was computed as Equation 24.

$$\text{Accuracy per character test set} = \frac{\text{Number of correctly recognized characters per character test set}}{\text{Number of characters in character test set}} \quad (24)$$

The results for the evaluation of Tesseract are presented in Table 4.2.

For the third set of experiments the same ten character sets (as used for the previous set of experiments, Section 4.2) and the same scoring method as Equation 24 were used. In this manner, a ten-fold cross validation was performed to evaluate the automatic text recognition method developed in this research. However, for this case, the

inputs to the SVM-based classifier were the feature vectors $F_{C_m S_n} = [G_1, \dots, G_{95}, E_1, \dots, E_{52}, F_1, \dots, F_{14}, W_1, \dots, W_{24}]_{C_m S_n}^T$ obtained from the feature extraction step. Table 4.2 shows the ten-fold cross validation results for the Tesseract-based character recognition in comparison to the automated text detection and recognition method developed in this research.

Table 4.2: Ten-fold cross-validation results comparison.

Ten Fold Cross-Validation Test Sets	Accuracy (%) (OCR tool-Tesseract)	Accuracy (%) (Automated Text Detection developed in this research)
1	73.24	91.35
2	74.31	91.52
3	70.93	90.65
4	71.78	92.22
5	78.23	92.93
6	73.37	91.59
7	76.38	91.52
8	83.27	93.01
9	74.23	92.54
10	83.24	92.38
Average Accuracy	75.90	91.97

As can be seen in Table 4.2, the automated text detection technique developed in this research provided higher accuracy as compared to the OCR tool Tesseract. The average accuracy for Tesseract was 75.90% as compared to 91.97% using the automated text detection method developed in this research. In this way, the method developed in this study was able to outperform the benchmark OCR-Tesseract by 16.07%.

The next set of experiments was performed for feature evaluation and feature reduction. For the automated text detection method developed in this study, Table 4.2 showed the results when all the features were used for ten-fold cross validation. In order to reduce the dimension (number of features) of the data set, a MLR-based approach was

tested using the statistical analysis software, SAS[®]. Multinomial logistic regression is used for modeling nominal outcome variables, in which the log odds of the outcomes are modeled as a linear combination of the predictor variables. As a metric for best feature selection, the p-values from the MLR method were analyzed to evaluate which features provided statistically significant information as compared to the other features. The p-values are used to determine whether or not the null hypothesis that a particular predictor's regression coefficient is zero, given that the rest of the predictors in the model can be rejected. In general, if the p-value is less than alpha (α) (usually 0.05 or 0.01), then the null hypothesis can be rejected and the parameter estimate is considered to be statistically significant at that alpha level [70, 73-74]. The p-values for all the features were obtained in this manner.

For feature selection, we used a thresholding method, wherein we used different values of α to select only those features whose p-values were less than α . For this study, we used α values of 0.05, 0.04, 0.03, 0.02, 0.01, 0.001 and 0.0015. Using these values of α and thus selecting only certain features, the ten-fold cross validation was performed on the reduced set of features to evaluate which set of features provided the best accuracy for character recognition. The summary of the ten-fold cross validation results using reduced features were compared to using all the features in Table 4.3. As shown, the best-case accuracy (94.27% shown in bold red in Table 4.3) was obtained for an alpha value of 0.01, which employed 112 features as compared to all the 185 features. In general, the SVM-classifier accuracies for automated character recognition were quite consistent across the ten-fold cross validation approach.

Table 4.3: Comparison of accuracies using only certain features.

Number of features Test Set #	83 ($\alpha=0.0015$)	97 ($\alpha=0.001$)	112 ($\alpha=0.01$)	119 ($\alpha=0.02$)	126 ($\alpha=0.03$)	129 ($\alpha=0.04$)	132 ($\alpha=0.05$)	185 (all features)
1	91.82	91.90	91.67	91.12	91.04	91.27	91.27	91.35
2	92.14	92.62	93.01	91.91	91.83	91.59	91.91	91.52
3	91.28	91.20	91.59	91.04	91.28	90.49	91.04	90.65
4	91.83	92.38	92.62	92.14	93.87	92.69	92.54	92.22
5	92.07	92.77	92.93	92.93	91.36	92.46	93.17	92.93
6	92.46	92.93	92.93	91.75	92.07	91.83	91.59	91.59
7	92.22	92.69	92.77	91.67	93.01	91.67	91.20	91.52
8	93.95	93.87	94.27	93.40	92.14	93.09	92.93	93.01
9	91.44	92.38	92.54	91.91	91.91	92.38	92.77	92.54
10	92.22	92.30	92.38	92.30	92.62	92.38	92.54	92.38
Average Accuracy	92.14	92.51	92.67	92.02	92.11	91.99	92.10	91.97

The p-values of the features corresponding to the best-case accuracy are shown in Table 4.4 below.

Table 4.4: p-values obtained from the MLR for the 112 ($\alpha=0.01$).

Feature	p-value	Feature	p-value	Feature	p-value	Feature	p-value	Feature	p-value	Feature	p-value	Feature	p-value
G1	0.0001	G38	0.0001	G70	0.0085	E1	0.0001	E23	0.0001	E46	0.0001	F14	0.0001
G2	0.004	G45	0.0001	G72	0.0001	E2	0.0001	E24	0.0001	E47	0.0001	W2	0.0001
G8	0.0001	G46	0.0001	G78	0.0018	E5	0.0001	E25	0.0001	E48	0.0002	W4	0.0001
G9	0.0034	G47	0.0001	G82	0.0051	E6	0.0001	E26	0.0001	E49	0.0002	W5	0.0001
G13	0.0001	G50	0.0001	G83	0.0001	E7	0.0001	E30	0.0001	E51	0.0001	W8	0.0001
G14	0.0001	G54	0.0021	G84	0.0001	E9	0.0067	E31	0.0021	F1	0.0001	W9	0.0001
G15	0.0001	G55	0.0006	G85	0.0001	E10	0.0001	E32	0.0006	F2	0.0001	W10	0.0001
G17	0.0007	G57	0.0036	G86	0.0001	E13	0.0002	E33	0.0036	F4	0.0001	W11	0.0001
G18	0.0029	G58	0.0001	G87	0.0001	E14	0.0001	E34	0.0001	F5	0.0001	W13	0.0001
G22	0.0074	G59	0.0002	G88	0.0001	E15	0.0001	E37	0.0002	F6	0.0001	W14	0.0001
G25	0.0043	G61	0.003	G90	0.0001	E16	0.0004	E39	0.003	F7	0.0001	W15	0.0001
G28	0.0001	G64	0.0001	G91	0.0001	E17	0.0001	E40	0.0001	F9	0.0001	W16	0.0001
G29	0.0001	G65	0.0006	G92	0.0001	E19	0.0001	E41	0.0006	F10	0.0001	W17	0.0001
G31	0.0001	G66	0.0001	G93	0.0001	E20	0.0001	E42	0.0001	F11	0.0008	W21	0.0001
G35	0.0001	G67	0.0001	G94	0.0001	E21	0.0001	E43	0.0001	F12	0.0001	W22	0.0001
G36	0.0001	G69	0.0009	G95	0.0001	E22	0.0001	E45	0.0009	F13	0.0001	W24	0.0001

Figure 4.2 shows how the average accuracies vary upon using the different number of features. One can conclude that having a lot of features does not necessarily help in the automated character recognition from biomedical images. It is important to use statistically significant features for the classification step in comparison to using all the features.

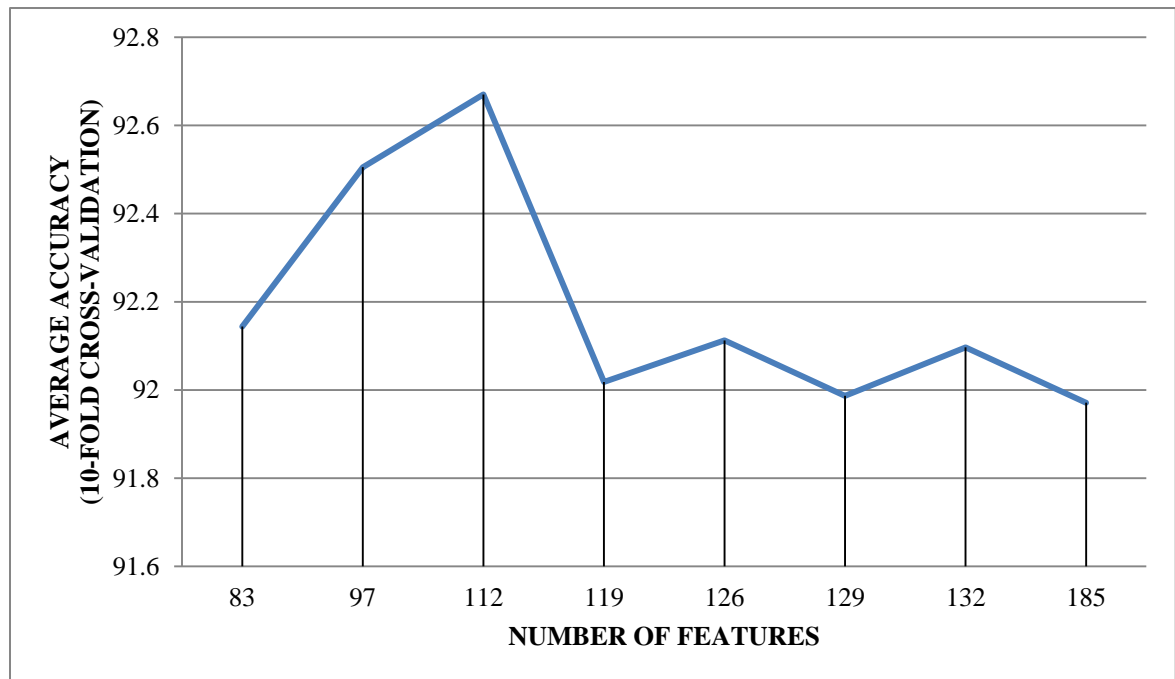


Figure 4.2: Comparison of overall accuracy.

From the feature evaluation, one could state that all the categories of features provided low p-values, indicating statistically significant information present within each category of features and also validating the basis of our feature selection.

Figure 4.3(a)-(b) shows the sample output (text file) of the overall system, which includes the detected text along with the position information of the detected text blocks. The text file has the following information in order: 1) characters recognized from each

Even though the automated text detection technique developed here has a very good accuracy, there are limitations with this approach. The detection of the individual characters from the text blocks seem to work quite well except for cases when two characters are very close to each other (see Figure 4.4). Since the input image itself had a partial overlap of the characters (letters 'r' and 'y' in Figure 4.4), the system detects these two letters as a single character instead of separating them. Also, the system is not able to recognize non-alphabetic characters such as numbers and symbols. Whenever such characters are encountered, they are classified as the nearest resembling alphabetic character.

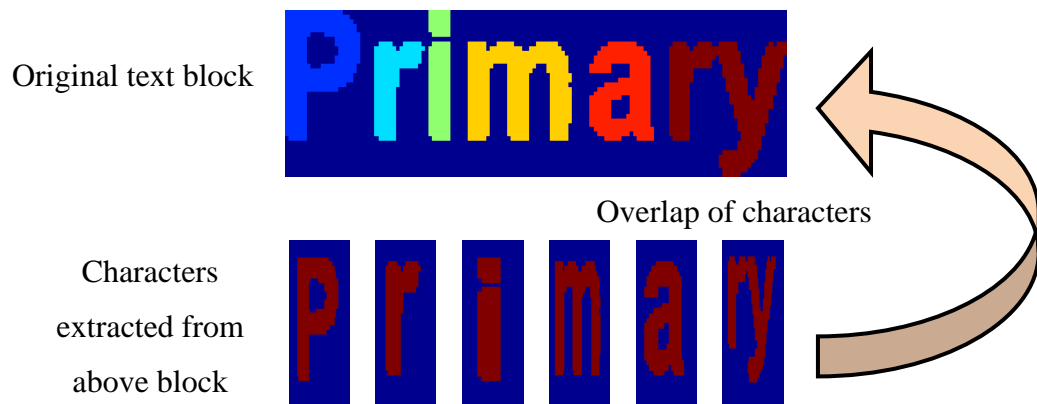


Figure 4.4: Character overlap during localization

5. CONCLUSIONS

In this research, an ROI-based analysis of image features is investigated for biomedical image text detection and recognition. This method can automatically detect text blocks from biomedical images using a histogram-based processing. Individual character segmentation is performed on the detected text blocks followed by character normalization and feature extraction. Different categories of features are investigated such as geometry-based, exemplar-based, correlation-based and frequency-domain-based Fourier descriptors. The features are used as an input to the SVM classifier for automated character recognition.

Since the automated text detection and recognition method presented in this study is specifically developed for biomedical images, the accuracy of the text detection and recognition is very high as compared to the commercially available OCR tool Tesseract. One of the primary reasons for the high degree of accuracy developed in this study is due to the features used. Tesseract uses simple features that are based on polygonal approximation of the periphery of the character [70]. However for complex images such as the ones used in this study, more descriptive features such as correlation-based and region-based need to be used for a better character recognition. Based on the study presented in this paper, one can conclude that using features that analyze the character in a more comprehensive manner as compared to just analyzing the boundary of the character is more important for character recognition.

Future research would be directed towards merging the automated text detection method (developed in this research) and automatic arrow detection methods (developed

in previous research, [12]) into a single composite system that would be able to characterize a biomedical image completely. Future work will involve integrating the detection of medical annotations into an overall approach for fusing data such as image text (developed in this research), key words, modality of the biomedical image and image captions to improve the relevance of the search results for medical publication querying.

REFERENCES

1. Demner-Fushman, D., Antani, S.K., Simpson, M., Thoma, G.R. (2008). 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.
2. Demner-Fushman, D., Antani, S., Simpson, M., Thoma, G.R. (2009). Annotation and Retrieval of Clinically Relevant Images, *International Journal of Medical Informatics: Special Issue on Mining of Clinical and Biomedical Text and Data*, 78(12), pp.e59-e674.
3. Deserno, T.M., Antani, S., & Long, R. (2009). Ontology of Gaps in Content-Based Image Retrieval. *Journal of Digital Imaging*, 22(2), 202-15.
4. Eakins, J.P., Graham, M.E. (2000). Content-based image retrieval, Tech. Rep. JTAP-039, *JISC Technology Application Program*, Newcastle upon Tyne.
5. Pfund, T., Marchand-Maillet, S. (2002). Dynamic multimedia annotation tool, in: G. Beretta, R. Schettini (Eds.), *Proceedings of the SPIE Photonics West Conference on Internet Imaging III*, vol. 4672, San Jose, CA, USA, pp. 216—224.
6. Müller, H., Michoux, N., Bandon, D., Geissbuhler, A. (2004). A review of content-based image retrieval systems in medical applications—clinical benefits and future directions, *International Journal of Medical Informatics*, Volume 73, Issue 1, Pages 1-23, ISSN 1386-5056, 10.1016/j.ijmedinf.2003.11.024.
7. Tang, L., Hanka, R., Ip, H. (1999). A review of intelligent content-based indexing and browsing of medical images. *Health Informatics Journal*, 5(1):40-49.
8. Lehmann, T.M., Güld, M.O., Thies, C., Fischer, B., Spitzer, K., Keysers, D., Ney, H., Kohnen, M., Schubert, H., Wein, B.B. (2004). Content-based image retrieval in medical applications. *Meth Informat Med*; 43(4):354–61.
9. Lehmann T.M., Wein, B., Dahmen, J., Bredno, J., Vogelsang, F., Kohnen, M. (2000) Content-based image retrieval in medical applications: A novel multistep approach. *Proceedings SPIE*; 3972: 312-20.
10. Lehmann T.M., Schubert, H., Keysers, D., Kohnen, M., Wein, B.B. (2003). The IRMA code for unique classification of medical images. *Proceedings SPIE*; 5033: 440-51.

11. Lehmann, T.M., Güld, M.O., Deselaers, T., Keysers, D., Schubert, H., Spitzer, K., et al. (2005). Automatic Categorization of Medical Images for Content-based Retrieval and Data Mining. *Computerized Medical Imaging and Graphics*; 29(2):143–155.
12. Rahman, M.M., Desai, B.C., Bhattacharya, P. (2008). Medical image retrieval with probabilistic multi-class support vector machine classifiers and adaptive similarity fusion. *Computerized Medical Imaging and Graphics*; 32(2): 95-108.
13. Pourghassem, H., Ghassemian, H. (2008). Content-based medical image classification using a new hierarchical merging scheme. *Computerized Medical Imaging and Graphics*; 32 (8): 651-661.
14. Stanley, R. J., De, S., Demner-Fushman, D., Antani, S., Thoma, G. R. (2011). An image feature-based approach to automatically find images for application to clinical decision support. *Computerized Medical Imaging and Graphics*; 35(5):365-72.
15. Demner-Fushman, D., Antani, S.K., Thoma, G.R. (2007). Automatically Finding Images for Clinical Decision Support. In Workshop on Data Mining in Medicine. 7th *IEEE Intl Conf on Data Mining*, pp. 139-144.
16. Cheng B., Stanley R.J., De, S., Antani S., Thoma G.R. (2011). “Automatic Detection of Arrow Annotation Overlays in Biomedical Images”, *International Journal of Healthcare Information Systems and Informatics (IJHISI)*, October-December 2011, 6(4), pp.23-41.
17. Dov, D., & Liu, W. (1999). 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*, 29(4), 411-416.
18. Park, J., Rasheed. W., Beak, J. (2008). 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*, 382-386.
19. Herold, J., Zhou, L., Abouna, S., Pelengaris, S., Epstein, D., et al. (2010). Integrating semantic annotation and information visualization for the analysis of multichannel fluorescence micrographs from pancreatic tissue. *Computerized Medical Imaging and Graphics* 34: 446–452.
20. Hearst, M. et al. (2007). BioText Search Engine: beyond abstract search. *Bioinformatics*, 23, 2196–2197.
21. Wu, V., Manmatha, R., Riseman, E. (1999). TextFinder: An automatic system to detect and recognize text in images, *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 21, no. 11.

22. Xu, Songhua, and Michael Krauthammer. (2010). A new pivoting and iterative text detection algorithm for biomedical images. *Journal of Biomedical Informatics* 43(6):924-31.
23. Xu, S., McCusker, J., Krauthammer, M. (2008). Yale Image Finder (YIF): a new search engine for retrieving biomedical images. *Bioinformatics*, 24:1968-1970.
24. You, D., Apostolova, E., Antani, S.K., Demner-Fushman, D., & Thoma, G.R. (2009). Figure content analysis for improved biomedical article retrieval. In *SPIE-IS&T Electronic Imaging: Vol. 7247*. San Jose, CA (72470V(1-10)).
25. You, D., Antani, S.K., Demner-Fushman, D., Rahman, M.M., Govindaraju, V., & Thoma, G.R. (2010). Biomedical article retrieval using multimodal features and image annotations in region-based CBIR. In *SPIE-IS&T Electronic Imaging: Vol. 7534*. San Jose, CA (75340V(1-12)).
26. Deserno, T.M., Antani S., Long, L.R. (2009). Content-based image retrieval for scientific literature access. *Methods of Information in Medicine*;48(4):371-80.
27. Antani, S., Demner-Fushman, D., Li, J., Srinivasan, B.V., & Thoma, G.R. (2008). Exploring use of images in clinical articles for decision support in Evidence-Based Medicine. In *Proceedings of SPIE-IS&T Electronic Imaging: Vol. 6815*. San Jose, CA (68150Q(1-10)).
28. Kim, D., Yu, H. (2011). Figure Text Extraction in Biomedical Literature. *PLoS ONE* 6(1): e15338. doi:10.1371/journal.pone.0015338.
29. Hiremath, P.S., Pujari, J. (2007). Content based image retrieval using color texture and shape features. Proc. 15th *International Conference on Advanced Computing and Communications*;780-784.
30. Jalaja, K., Bhagvati, C., Deekshatulu, B.L., Pujari, A.K. (2005). Texture element feature characterizations for CBIR. Proc. *IEEE International Geoscience and Remote Sensing Symposium*;2:733-736.
31. Hua, Y., Xiao-Ping, Z. (2004). Texture image retrieval based on a Gaussian mixture model and similarity measure using a Kullback divergence. Proc. *IEEE International Conference on Multimedia and Expo*;3:1867-1870.
32. Patino-Escarcina, R.E., Ferreira Costa, J.A. (2008). The semantic clustering of images and its relation with low level color features. Proc. *IEEE International Conference on Semantic Computing*;74-79.
33. Deng, Y., Manjunath, B.S., Kenney, C., Moore, M.S., Shin. (2007). H. An efficient color representation for image retrieval. *IEEE Trans Image Proc.* ;10(1):140-147.

34. Soffer, A. (1997). Image categorization using texture features. *Proc. Fourth International Conference on Document Analysis and Recognition*; 1:233-237.
35. Tao, D., Li, X., Yuan, Y., Yu, N., Liu, Z., Tang, X. (2002). A set of novel textural features based on 3D co-occurrence matrix for content-based image retrieval. *Proc. Fifth International Conference on Information Fusion*; 2:1403-1407.
36. Sergyan, S. (2008). Color histogram features based image classification in content-based image retrieval systems. *Proc. 6th International Symposium on Applied Machine Intelligence and Informatics*; 221-224.
37. Huihui, H., Wei, H., Zhigang, L., Weirong, C. ; Qingquan, Q. (2005). Content-based color image retrieval via lifting scheme. *Proc. International Symposium on Autonomous Decentralized Systems*; 378-383.
38. Vadivel, A.; Majumdar, A.K.; Sural, S. (2004). Characteristics of weighted feature vector in content based image retrieval applications. *Proc. IEEE Conference on Intelligent Sensing and Information Processing*; 127-132.
39. Patino-Escarcina, R.E.; Costa, J.A.F. (2007). Content based image retrieval using a descriptors hierarchy. *Proc. 7th International Conference on Hybrid Intelligent Systems*; 228-233.
40. Frigui, H., Caudill, J., Chamseddine B., Adballah, A. (2008). Fusion of multi-modal features for efficient content-based image retrieval. *Proc. IEEE International Conference on Fuzzy Systems* ;1992-1998.
41. Stanley, R.J., Stoecker, W.V., Moss, R.H. (2008). A basis function feature-based approach for skin lesion discrimination in dermatology dermoscopy images. *Skin Research and Technology*; 14(4):425-435.
42. Ohya, J., Shio, A., Akamatsu, S. (1994). Recognizing characters in scene images. *IEEE Trans Pattern Anal Mach Intell*; 16(2):214–20. doi:10.1109/34.273729.
43. Zhong, Y., Karu, K., Jain, A.K. (1995). Locating text in complex color images. *Proc Third Int Conf Document Anal Recognit*; 1:146–9. doi:10.1109/ICDAR.1995.598963.
44. Jung, K. (2001). Neural network-based text location in color images. *Pattern Recognition Letters*; 22(14):1503–15. doi:10.1016/S0167-8655(01)00096-4.
45. Li, H., Doermann, D., Kia, O. (2001). Automatic text detection and tracking in digital video. *IEEE Trans Image Process*; 9(1):147–56.
46. Kim, K.I., Jung, K., Park, S.H., Kim, H.J. (2001) Support vector machine-based text detection in digital video. *Pattern Recognition*; 34(2):527–9.

47. Xu, S., Krauthammer, M. (2011). Boosting text extraction from biomedical images using text region detection. *Biomedical Sciences and Engineering Conference (BSEC), 2011* , pp.1-4.
48. Bhooshan, S., Kumar, V., Singh, A.K., Negi, T. Miglani, J. (2009). Character Recognition Using Geometrical Features of Alphabet: A Novel Technique. *Communication Software and Networks. ICCSN '09. International Conference on*, vol., no., pp.119-125.
49. Chhabra, A.K., An, Z., Balick, D., Cerf, G., Loris, K., Sheppard, P., Smith, R., Wittner, B. (1993). High-order statistically derived combinations of geometric features for handprinted character recognition. *Document Analysis and Recognition, Proceedings of the Second International Conference on*, vol., no., pp.397-401.
50. Xiaodong, Y., Duoqian, M. (2009). Rough Geometry and Its Applications in Character Recognition. *Transactions on Rough Sets X* :136-156.
51. Halawani, S.M., Albidewi, I.A. (2010). Recognition of Hand Printed Characters Based on Simple Geometric Features. *J. Comput. Sci.*, 6: 1518-1523
52. Blumenstein, M., Verma B. K., Basli, H. (2003). A Novel Feature Extraction Technique for the Recognition of Segmented Handwritten Characters. *7th International Conference on Document Analysis and Recognition*, Eddinburgh, Scotland: pp.137-141.
53. Pai, T., Shyu, K., Chen, L., Tai, G. (1995). Mathematical morphology-based shape feature analysis for chinese character recognition systems. *Visual Communications and Image Processing*. 0001; 922-930.
54. Mukherji, P., Rege, P. (2009). Shape Feature and Fuzzy Logic Based Offline Devnagari Handwritten Optical Character Recognition. *Journal of Pattern Recognition Research* 4 (2009) 52-68.
55. Lu, S., Linlin, L., Chew, L.T. (2008). Document Image Retrieval through Word Shape Coding. *Pattern Analysis and Machine Intelligence, IEEE Transactions on* , vol.30, no.11, pp.1913-1918.
56. Amit, Y., Geman, D., Wilder, K. (1997). Joint induction of shape features and tree classifiers. *Pattern Analysis and Machine Intelligence, IEEE Transactions on* , vol.19, no.11, pp.1300-1305.
57. Smith, S.J., Bourgoin, M.O., Sims, K., Voorhees, H.L. (1994). Handwritten character classification using nearest neighbor in large databases. *Pattern Analysis and Machine Intelligence, IEEE Transactions on* , vol.16, no.9, pp.915-919.

58. Hansen, J. (2002). A Matlab Project in Optical Character Recognition (OCR). Course Project, University of Rhode Island.
59. Stanley, R. J., Gader, P., Ho, D. (2002). Feature and decision level sensor fusion of electromagnetic induction and ground penetrating radar sensors for landmine detection with hand-held units. *Information Fusion*; 3: 215–223.
60. 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.
61. Everingham, M. R., Thomas, B. T., Troscianko, T. (1998). Head- mounted mobility aid for low vision using scene classification techniques. *International Journal of Virtual Reality*, 3(4), 3–12.
62. Piper, J., Granum, E. (1989). On fully automatic feature measurement for banded chromosome classification. *Cytometry*, 10, 242-255.
63. Wang, S.S., Chen, P.C., Lin, W.G. (1994). Invariant pattern recognition by moment fourier descriptor, *Pattern Recognition*, Volume 27, Issue 12, Pages 1735-1742.
64. Shridhar, M., Badreldin, A. (1984) High accuracy character recognition algorithm using Fourier and topological descriptors. *Pattern Recognition*, 17, 5, pp. 515-524.
65. Everingham, M. R., Thomas, B. T., Troscianko, T. (1998). Head- mounted mobility aid for low vision using scene classification techniques. *International Journal of Virtual Reality*, 3(4), 3–12.
66. Morns, I.P., Satnam S.D. (1997). Character recognition using Fourier descriptors and a new form of dynamic semisupervised neural network, *Microelectronics Journal*, Volume 28, Issue 1, Pages 73-84.
67. Badreldin, A., Wong, A.K.C., Prasad, T., Ismail, M. (1980). Shape descriptors for N-Dimensional curves and trajectories, *IEEE Proc. on Cybernetics and Societ.* pp. 713-717.
68. Fan, R.E., Chen, P.H. and Lin, C.J. (2005). Working set selection using second order information for training SVM. *Journal of Machine Learning Research* 6, 1889-1918.
69. Chang, C.C. and Lin, C.J. (2001). LIBSVM: a library for support vector machines. Software available at <http://www.csie.ntu.edu.tw/~cjlin/libsvm>
70. Pal, M. (2012). Multinomial logistic regression-based feature selection for hyperspectral data, *International Journal of Applied Earth Observation and Geoinformation*, Volume 14, Issue 1, February, Pages 214-220.

71. Smith, R. (2007). An Overview of the Tesseract OCR Engine. ICDAR, vol. 2, pp.629-633, *Ninth International Conference on Document Analysis and Recognition (ICDAR 2007)*.
72. Bar-Ilan, J. (1998). On the overlap, the precision and estimated recall of search engines: A case study of the query “Erdos”. *Scientometrics*, 42(2), pp.207-208.
73. Hosmer, D. and Lemeshow, S. (2000). *Applied Logistic Regression (Second Edition)*. New York: John Wiley & Sons, Inc.
74. Agresti, A. (1996). *An Introduction to Categorical Data Analysis*. New York: John Wiley & Sons, Inc.

IV. A FUSION-BASED APPROACH FOR UTERINE CERVICAL CANCER HISTOLOGY IMAGE CLASSIFICATION

Soumya De^a, R. Joe Stanley^a, Cheng Lu^a, Rodney Long^b, Sameer Antani^b,
George Thoma^b

^aDepartment of Electrical and Computer Engineering
Missouri University of Science and Technology
Rolla, MO 65409-0040 USA

^bLister Hill National Center for Biomedical Communications
National Library of Medicine, National Institutes of Health, DHHS
Bethesda, MD 20894 USA

ABSTRACT

One method of cervical cancer screening uses visual testing, by analyzing the color change of cervix tissues when exposed to acetic acid. Cervicography is a technique that augments this visual screening by recording a film image of the acetic acid-treated cervix. In this research, sixty-two cervicography images of the cervix were examined. The images were analyzed for classifying the squamous epithelium into varying grades of cervical intraepithelial neoplasia (CIN), including Normal, CIN1, CIN2 and CIN3. The steps for image-based epithelium classification include a novel distance transform and bounding box-based medial axis determination algorithm, which was used to partition the epithelium vertically into ten segments. Features were computed for each vertical segment for individual segment classification. A voting scheme for fusing the CIN grades of the vertical segments was used for image-based epithelium classification. Using a Linear Discriminant Analysis-based classifier with a leave-one-image-out approach yielded an overall classification rate as high as 95.1% for Normal versus CIN classes. Classification of the epithelium within a one label windowed variation of the correct image label yielded a 100.0% correct recognition rate.

1. INTRODUCTION

Annually, there are 400,000 new cases of invasive cervical cancer out of which 15,000 occur in the U.S. alone [1]. Cervical cancer is the second most common cancer affecting women worldwide and the most common in developing countries. Cervical cancer can be cured in almost all patients, though, if detected early, and treated. However, cervical cancer incidence and mortality remain high in resource-poor regions, where high-quality screening programs often cannot be maintained because of inherent complexity. An alternative cervical cancer screening uses analysis of visual testing based on color change of cervix tissues when exposed to acetic acid. Cervicography is a technique that augments this visual screening by recording a film image of the acetic acid-treated cervix, and has been widely used over the last few decades [2]. The National Cancer Institute (NCI) has collected a vast amount of visual information, 100,000 cervigrams (35 mm digital color slides), screening thousands of women by this technique. The long-term objective of the proposed project is to facilitate the development of a unique web-based database of digitized cervix images. This is needed for investigating the role of human papillomavirus (HPV) in the development of cervical cancer and its intraepithelial precursor lesions in women [3].

One of the methods for cervical cancer diagnosis includes using the cervigrams to analyze the squamous epithelium region of the cervix. As stated in [3], cervical intraepithelial neoplasia (CIN) could be defined as a pre-malignant condition for cervical cancer in which the atypical cells are limited to the epithelium only. Cervical

intraepithelial neoplasia is divided into grades such as Normal, CIN1, CIN2 and CIN3 [3-5]. Normal refers to the absence of CIN, CIN1 corresponds to mild dysplasia while CIN2 and CIN3 have similar annotations for moderate dysplasia and severe dysplasia respectively. The pathologists provide diagnostic decisions related to CIN and its various grades based on the visual analysis of the cervigrams [3-7]. Figure 1.1 presents image examples of the different CIN grades, as determined by an expert pathologist. In general, a decision on the CIN grade can be made by analyzing the epithelium, which contains the atypical cells (Figure 1.1).

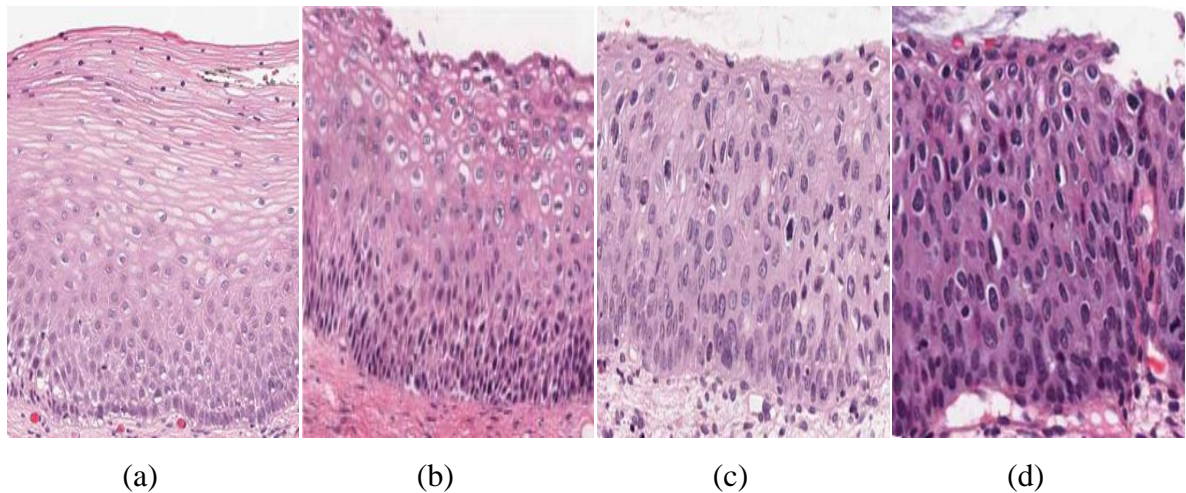


Figure 1.1: CIN grading label examples.
(a) Normal, (b) CIN 1, (c) CIN 2, (d) CIN 3.

Cervical intraepithelial neoplasia has been observed to show progressively increasing atypical cells across the spectrum, from top to bottom [4-7]. For example as shown in Figure 1.1, atypical cells are seen mostly in the bottom third of the epithelium for CIN 1 (Figure 1.1b). For CIN2, the atypical cells lie in the bottom two thirds of the epithelium (Figure 1.1c), and full length of the epithelium for CIN 3 (Figure 1.1d). When these atypical cells extend beyond the epithelium region and start to enter into the

surrounding tissues and organs, it can lead to cancer [5]. In addition to analyzing the progressively increasing atypical cells from top to bottom of the epithelium, identification of nuclear atypia is also important [3]. Nuclear atypia is characteristic of nuclear enlargement, thereby resulting in different shapes and sizes of the nuclei present within the epithelium region. Since CIN grading by pathologists is visual-based, diagnostic grading repeatability suffers from inter- and intra-pathologist variation [6-8].

In previous studies, computer-assisted methods (digital pathology) using cervigrams have been investigated to augment the pathologists' decision (traditional pathology) for CIN diagnosis [9-15]. Keenan et al. introduced an automated system for CIN degree classification based on an input squamous epithelium region image [10]. The nuclei centers were determined and used as node points for a Delaunay triangulation mesh. The epithelium region was divided into three equal horizontal compartments and morphological features were extracted from the triangles within the compartments [10]. Guillaud et al. extracted texture features from the epithelium region to estimate the absolute intensity and density levels of the nucleus. Morphological features were also extracted to estimate the nuclear shape, size and boundary irregularities [11-13]. Miranda et al. determined the nuclei in the epithelium using a Watershed segmentation method followed by Delaunay triangulation to facilitate CIN analysis. This method is uniquely assigned CIN grade labels based on triangles formed using the Delaunay triangulation method, instead of making a CIN grade decision on the whole epithelium image [14].

In another study, CIN classification was performed by analyzing histological features and using a Bayesian belief network classifier. Histological features such as degree of nuclear pleomorphism and number and levels of mitotic figures in the

epithelium were examined. These features were linked to a decision node, wherein the decision on CIN grade was made based on conditional probability of the features. However these features were subjective and hence it was difficult to use them in a consistent and reproducible manner for CIN grade classification [15]. One of the recent works applied a decision tree classifier upon morphological features extracted from the nuclei for CIN grade classification [16]. The nuclei were segmented from the epithelium using a K-means clustering and graph-cut segmentation method. The approach used a decision tree for CIN grade classification with empirically determined rules. Wang et al. developed a comprehensive method for full-size cervigram analysis for CIN grade classification. The method included multi-resolution texture analysis for automated epithelium segmentation, iterative skeletonization to approximate the epithelium medial axis (skeleton), morphological feature extraction from blocks along orthogonal lines to the skeleton, and CIN grade classification of each block was done by a SVM-based classifier using the features extracted from the blocks [17-18]. In a recent pathology study by Marel et al., the epithelium was analyzed in a heterogeneous manner, wherein different regions of the epithelium were shown to exhibit different CIN grades. It was proposed that a particular epithelium with atypical cells could exhibit different CIN grades in different regions of the epithelium along the medial axis, as compared to a homogeneous CIN grade of the epithelium [19]. The research presented in this paper builds off of the CIN diagnosis methods proposed by Wang et al. and Marel et al. [5, 19]. A novel distance transform and bounding box-based technique for epithelium medial axis determination is presented in this paper. The medial axis was used as the basis to partition the epithelium into vertical segments. The vertical segments were each assigned a CIN

grade based on the extracted features. A novel fusion-based CIN grade classification for the epithelium region is also presented in this study. The fusion-based CIN grade was based on the individual vertical segment CIN grade classifications.

For individual segment analysis, several categories of features were investigated, including intensity shading-, texture-, correlation- and geometry-based, which were used as inputs to a classifier. Support Vector Machine- (SVM), and Linear Discriminant Analysis (LDA)-based classifiers were investigated to classify the individual segments into the varying grades of CIN. Finally, a voting scheme was used to fuse the class labels of the vertical segments to assign a CIN grade to the whole epithelium. Figure 1.2 shows the flowchart of the overall method developed in this study for CIN grade classification of the epithelium.

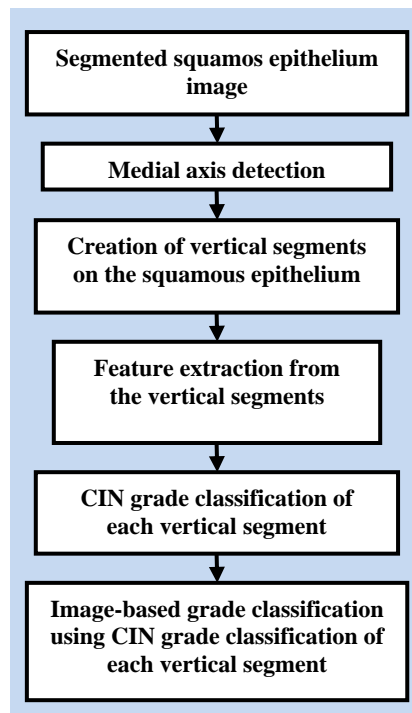
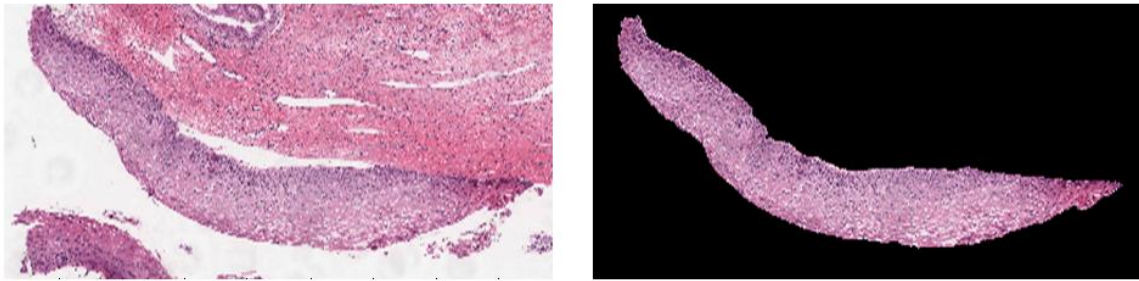


Figure 1.2: Overview of CIN grade classification method developed in this study.

The rest of the paper is organized as follows. Section 2 presents the methodology used in this research. Section 3 presents the experiments performed. Section 4 contains the experimental results and discussion while Section 5 provides the conclusions from this study.

2. METHODOLOGY

The goal of this research is to classify the squamous epithelium regions from the cervigrams into different grades of CIN. For this study, the squamous epithelium region had been manually segmented by expert pathologists in collaboration with the National Library of Medicine (NLM). Figure 2.1 shows a sample image and the manual pathologist segmented epithelium region. The segmented epithelium image is shown in Figure 2.1b and denoted as I (in RGB format). Let the luminance image of I be denoted as L .



(a) Original image

(b) Pathologist segmented image I

Figure 2.1: Original image and pathologist segmented image of epithelium.

Classification of the segmented epithelium images into the various CIN grades was performed using a five-step approach, as outlined in the flowchart in Figure 1.2, including:

Step 1: Find the medial axis of the segmented epithelium region (medial axis detection);

Step 2: Divide the segmented image into ten vertical segments, which were orthogonal to the medial axis (vertical segments generation);

Step 3: Extract features from each of the vertical segments (feature extraction);

Step 4: Classify each of these segments into one of the CIN grades (classification);

Step 5: Fuse the CIN grades from each vertical segment to obtain the CIN grade of the whole epithelium for image-based classification.

The following sections present each step in detail.

2.1. MEDIAL AXIS DETECTION

A distance transform and bounding box-based method was used to find the medial axis from the epithelium image I . The following steps detail the medial axis detection algorithm developed in this study:

Step 1: The binary mask of the segmented epithelium region shown in Figure 2.1b was created, and denoted as I^B . Note that the binary mask was cropped to contain only the region of the epithelium. The binary mask thus created is shown in Figure 2.2a.

Step 2: The bounding box of the binary mask I^B was determined and is shown in Figure 2.2b. Next, the Extent of I^B was found and denoted as ε_{before_Rot} . Extent is defined as the ratio of the non-zero pixels within the bounding box to the total number of pixels within the bounding box.

Step 3: The orientation of the binary mask I^B was determined as the angle (-90° to 90°) between the x-axis and the major axis of the ellipse that has the same second-moments as the binary mask. Figure 2.2b shows the orientation angle obtained using this manner and denoted as θ^B . The binary mask was rotated by the negative of the orientation angle, $-\theta^B$ and denoted as $I_{temp_rot}^B$. Next, the bounding box for $I_{temp_rot}^B$ was obtained and the Extent of $I_{temp_rot}^B$ was computed and denoted as ε_{after_Rot} . The bounding box for $I_{temp_rot}^B$ is shown in Figure 2.2d.

Step 4: The binary mask for the next steps (\mathbf{I}_{rot}^B) was selected as follows: if $\varepsilon_{after_Rot} > \varepsilon_{before_Rot}$, then $\mathbf{I}_{rot}^B = \mathbf{I}_{temp_rot}^B$. If $\varepsilon_{after_Rot} < \varepsilon_{before_Rot}$, then $\mathbf{I}_{rot}^B = \mathbf{I}^B$. It was experimentally observed that using a binary mask with a larger *Extent* provided better estimates of the medial axis. For the example shown in this paper, $\varepsilon_{after_Rot} = 0.42$ and $\varepsilon_{before_Rot} = 0.28$. Hence, $\mathbf{I}_{rot}^B = \mathbf{I}_{temp_rot}^B$ and is shown in Figure 2.2d.

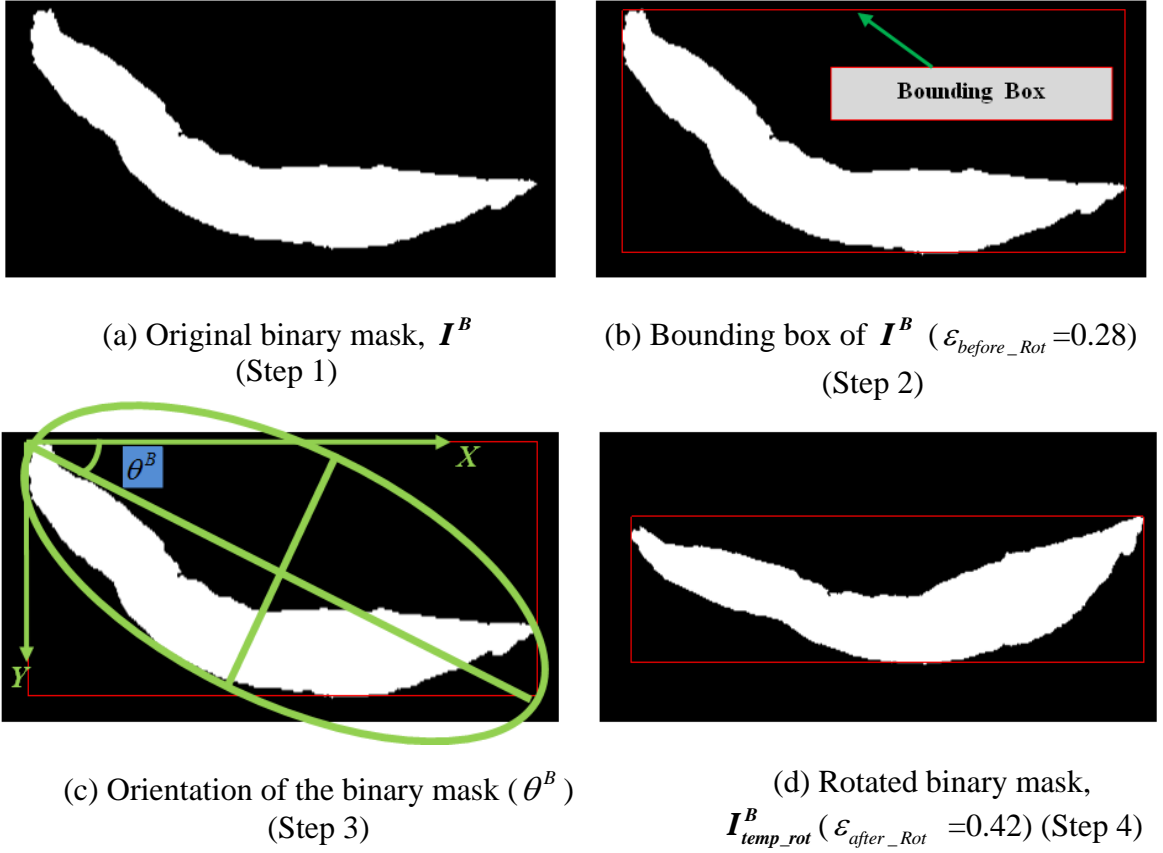


Figure 2.2: Binary mask rotation (Step 1-4).

Step 5: The distance transform was applied to the inverted, rotated binary mask, denoted as $\overline{\mathbf{I}_{rot}^B}$. The Euclidean distance was computed between each pixel, $\overline{\mathbf{I}_{rot}^B}(x, y)$ within $\overline{\mathbf{I}_{rot}^B}$ and the nearest non-zero pixel in $\overline{\mathbf{I}_{rot}^B}$, using the Matlab-based implementation in [21].

The distance transform applied to the mask $\overline{I_{rot}^B}$ resulted in the distance transform image, D as shown in Figure 2.3.

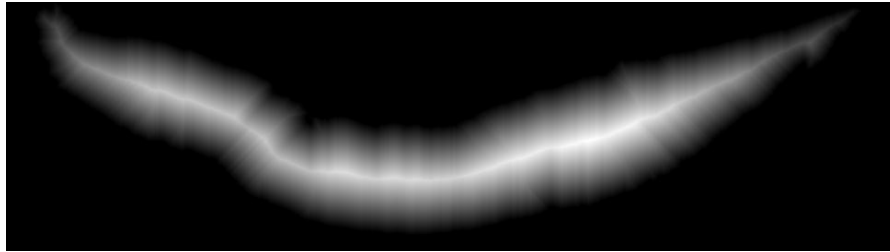


Figure 2.3: Distance transform image (D) obtained (Step5).

Step 6: Each column of the distance transform image D was scanned for the highest intensity pixel. The coordinates corresponding to the highest intensity pixels obtained in this manner were termed as the medial axis coordinates and denoted as the ‘determined’ medial axis λ^D . Thus λ^D is a set of points representing the coordinates of the determined medial axis and is shown in Figure 2.4 (magenta color), overlaid on the binary mask I_{rot}^B .

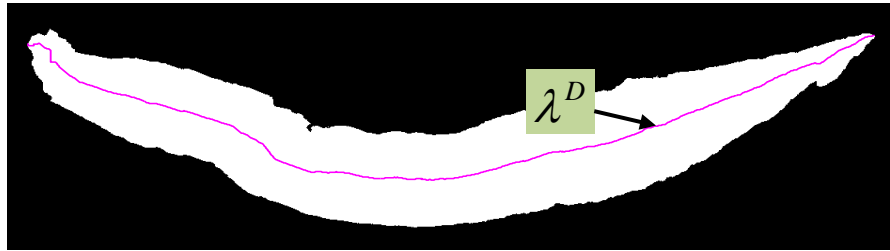


Figure 2.4: Medial axis of the binary mask I^B (Step 6).

The determined medial axis λ^D showed medial artifacts for cases when the epithelium images had a somewhat rectangular-like shape, as shown in Figure 2.5. The determined medial axis using the distance transform-based approach is shown in blue, with the manually marked desired medial axis shown in red. These artifacts occur in the left and right sections of the epithelium as shown in Figure 2.5. This is because for the left and

right sections, the maximally separated pair of end points (the max for each column) in D is along the corners, as compared to the edges. Therefore the medial axis tends to bend towards the corners as shown in Figure 2.5.

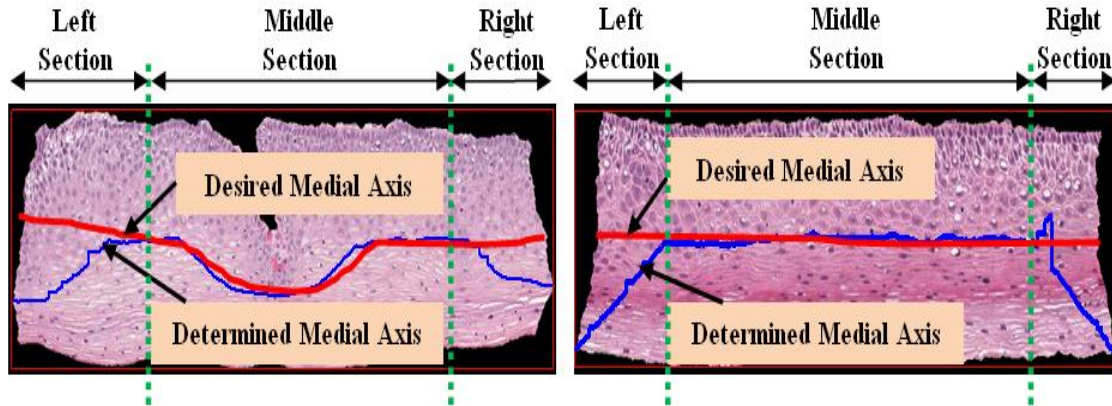


Figure 2.5: Determined medial axis v/s desired medial axis.

Step 7: From Figure 2.5, it can be observed that the determined medial axis is in approximate agreement with the desired medial axis in the middle/interior section of the medial axis. However, the determined medial axis showed deviations from the desired medial axis in the leftmost 20% and the rightmost 20% of the epithelium. In order to compensate for end section variations, the determined medial axis λ^D was divided into three pieces, such that $\lambda^D = \{\lambda_{left}^D, \lambda_{middle}^D, \lambda_{right}^D\}$ as shown in Figure 2.6. Here, λ_{left}^D (the piece shown in red) represented the leftmost 20% of the medial axis, λ_{middle}^D (the piece shown in black) represented the middle 60% of the medial axis and λ_{right}^D (the piece shown in blue) represented the rightmost 20% of the medial axis.

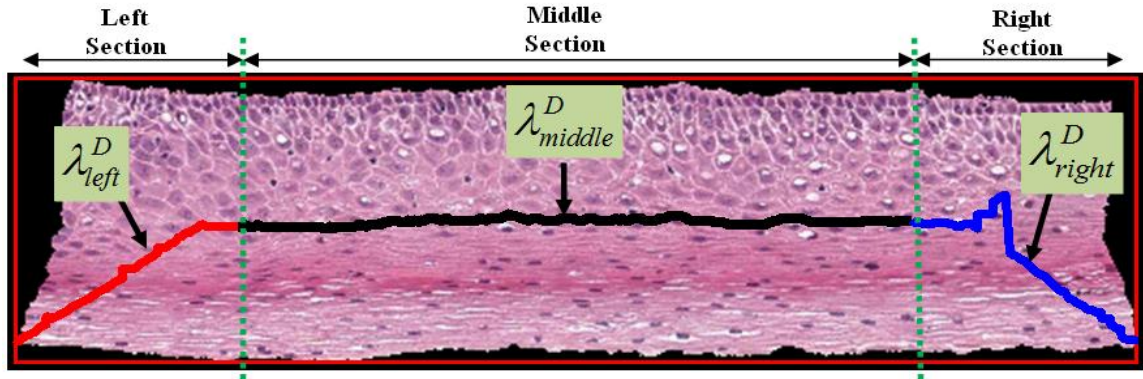


Figure 2.6: Dividing the medial axis into three pieces (Step 7).

Step 8: The center line ψ_{center} of the bounding-box for the epithelium was found and is shown in cyan color in Figure 2.7. ψ_{center} contains the coordinates of the line that divided the bounding box into two halves, top and bottom.

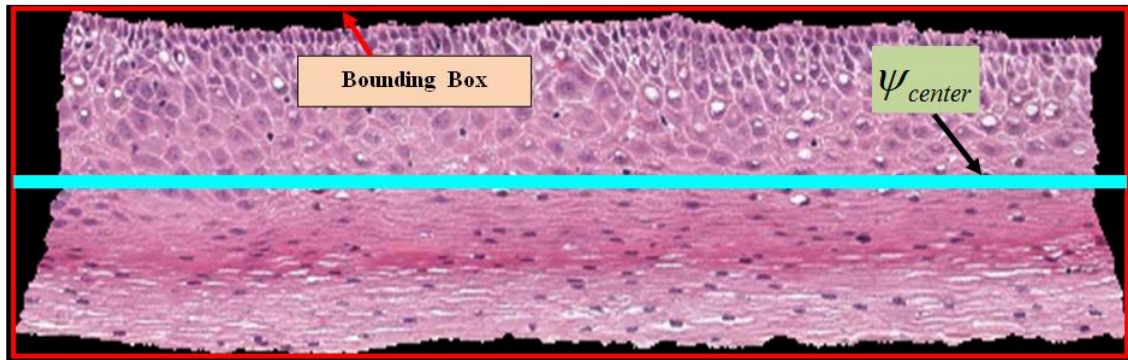


Figure 2.7: The center line of the bounding box of the epithelium (Step 8).

Step 9: Next, the intersection points, as shown in Figure 2.8, were determined. An intersection point was defined as the point where the leftmost (or the rightmost) medial axis piece intersected with the center line (see Figure 2.8). The left intersection point was obtained by computing the Euclidean distances between each point (coordinate) of λ_{left}^D and ψ_{center} . The point on λ_{left}^D which had the minimum Euclidean distance to ψ_{center} was

used as the left intersection point and denoted as $\rho_{left}(x, y)$. If more than two points were found, the point geometrically closer to the middle section (see Figure 2.8) was used as the left intersection point. In a similar manner, the right intersection point was determined by obtaining the intersection between the rightmost medial axis piece λ_{right}^D and the center line ψ_{center} . The right intersection point was denoted as $\rho_{right}(x, y)$. These intersection points were used as the points around which the leftmost and the rightmost medial axis pieces were adjusted to create the final medial axis. The intersection points are illustrated in Figure 2.8 as yellow dots.

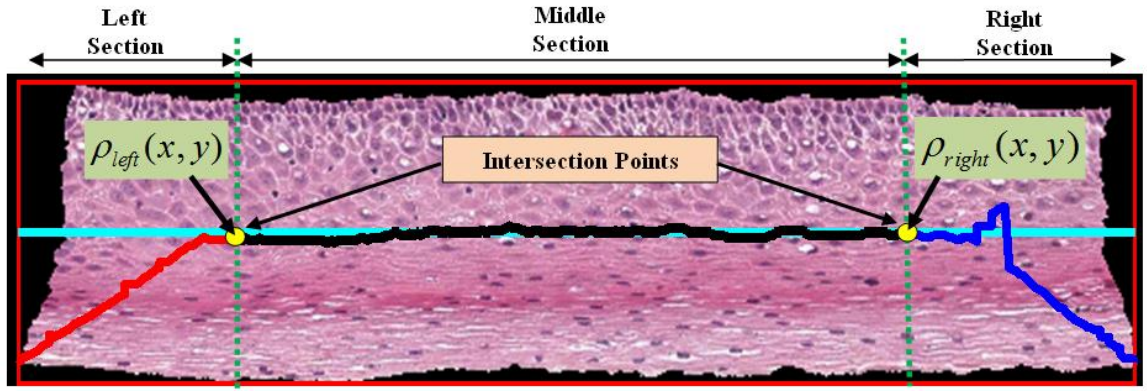


Figure 2.8: The intersection points were obtained (Step 9).

Step 10: Next, the edge image of the epithelium region was determined. The edge image was obtained by applying a Sobel edge detector on the rotated mask I_{rot}^B and denoted as I^E . The edge image I^E was created to determine the outer endpoints of the medial axis approximations for the leftmost 20% and rightmost 20% regions. The edge image I^E was divided into three parts, as shown in Figure 2.9, including: leftmost 20% (denoted as I_{left}^E), middle 60% (denoted as I_{middle}^E), and rightmost 20% (denoted as I_{right}^E). After dividing the edge image into three parts, the median values of edge coordinates of the leftmost edge

part I_{left}^E and rightmost edge piece I_{right}^E were obtained. These median values were denoted as the left median point $\mu_{left}(x, y)$ and the right median point $\mu_{right}(x, y)$, for I_{left}^E and I_{right}^E , respectively. The median points are shown as yellow dots in Figure 2.9. It was observed experimentally that the median points $\mu_{left}(x, y)$ and $\mu_{right}(x, y)$ provide a good estimate of the endpoints of the medial axis piece λ_{left}^D and λ_{right}^D , respectively.

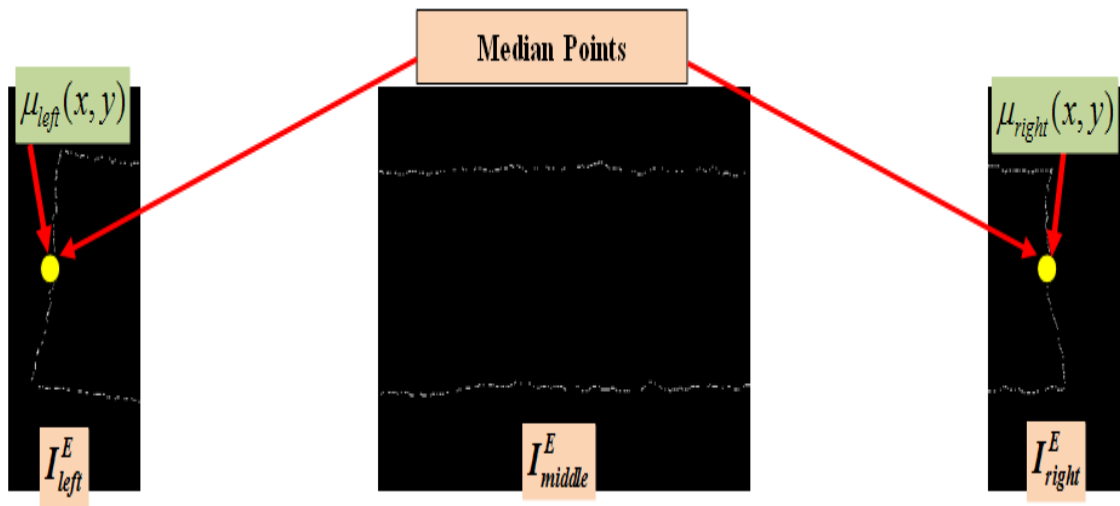


Figure 2.9: The edge image was computed and divided into three parts. The endpoints of the medial axis were found using the median of the edge points for the left and right parts (Step 10).

Step 11: The orientation of the leftmost and rightmost medial axis pieces were estimated to rotate those pieces of the medial axis determined. The orientation is defined as the angle by which the leftmost and the rightmost medial axis pieces (λ_{left}^D and λ_{right}^D respectively) are rotated/adjusted to obtain the final medial axis. The determined medial axis pieces were adjusted so as to obtain medial axis pieces that were in close proximity to the desired medial axis. Using the left intersection point $\rho_{left}(x, y)$ obtained from Step 9 and the left

median point $\mu_{left}(x, y)$ from Step 10, the orientation of the leftmost medial axis piece λ_{left}^D was estimated, and denoted as θ_{left}^D . In a similar manner, the orientation of the rightmost medial axis piece λ_{right}^D was also determined and denoted as θ_{right}^D . Note that the orientations of the leftmost and rightmost medial axis pieces obtained in this step were used to adjust the medial axis pieces as described in Step 12.

Step 12: In this step, the leftmost medial axis piece λ_{left}^D was rotated by the negative of the orientation $-\theta_{left}^D$ to obtain the ‘adjusted’ leftmost medial axis piece λ_{left}^A (note that the superscript ‘A’ refers to the ‘adjusted’ medial axis piece). Similarly the rightmost medial axis piece λ_{right}^D was rotated by the negative of the orientation $-\theta_{right}^D$ to obtain the ‘adjusted’ rightmost medial axis piece λ_{right}^A .

Step 13: The final medial axis coordinates λ^F were obtained from the determined medial axis coordinates $\lambda^D = \{\lambda_{left}^D, \lambda_{middle}^D, \lambda_{right}^D\}$ by replacing λ_{left}^D and λ_{right}^D with λ_{left}^A and λ_{right}^A respectively. The final medial axis was represented as $\lambda^F = \{\lambda_{left}^A, \lambda_{middle}^D, \lambda_{right}^A\}$. Note that the middle medial axis piece λ_{middle}^D is the same as that of the determined medial axis without any adjustment/rotation.

Step 14: Finally, λ^F was smoothed using a moving average filter with a span of 150 to remove some of the inundations in λ^F . Figure 2.10 shows the determined medial axis (in blue), the final medial axis before smoothing (in green) and the final medial axis after smoothing (in red), overlaid on the epithelium image.

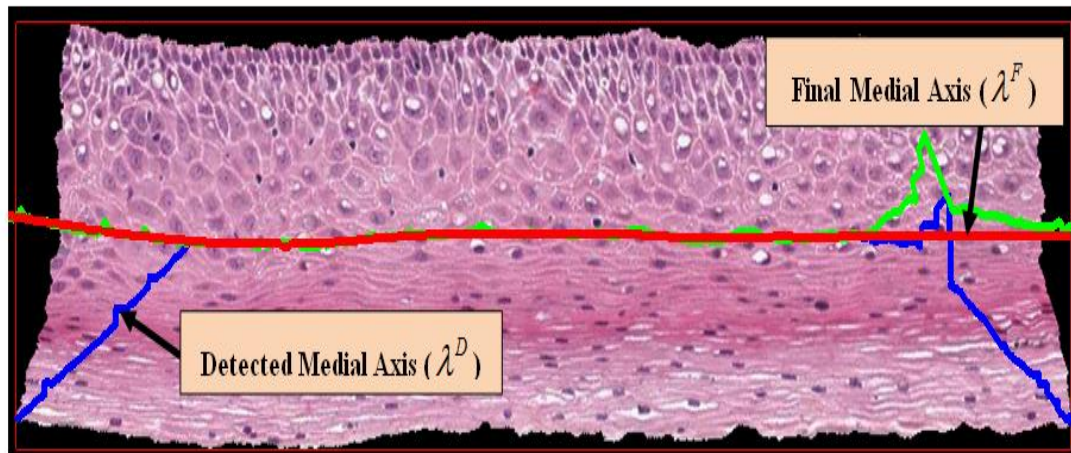


Figure 2.10: Determined medial axis and final medial axis (Step 14).

2.2. VERTICAL IMAGE SEGMENTATION

Once the final medial axis was determined, the next step was to use the medial axis to create vertical segments on the epithelium. Figure 2.11 presents an illustration of the vertical segments to be partitioned within the epithelium.

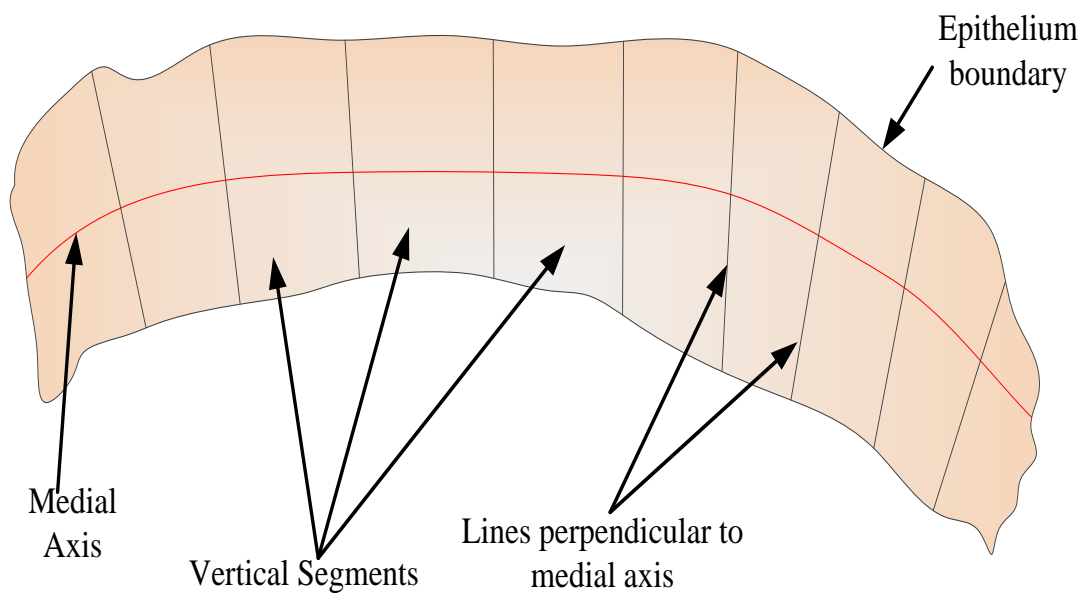


Figure 2.11: Schematic showing the vertical segments.

For the vertical segments creation, the final medial axis λ^F was partitioned into ten equal length line pieces such that $\lambda^F = \{\lambda^{v_1}, \lambda^{v_2}, \dots, \lambda^{v_{10}}\}$ (the superscript ‘ v ’ denotes the medial axis being partitioned into ten pieces, which would be used for ‘vertical’ image segments creation). The first line piece λ^{v_1} for one of the epithelium images is shown in Figure 2.12a. A three-step approach was used for vertical image segments generation:

Step 1: Each medial axis line piece $\lambda^{v_1}, \lambda^{v_2}, \dots, \lambda^{v_{10}}$ was fitted to a linear approximation using a least-squares regression algorithm [22]. An example of the line fitting algorithm is shown in Figure 2.12b. Let $\lambda^{r_1}, \lambda^{r_2}, \dots, \lambda^{r_{10}}$ denote the fitted medial axis line pieces obtained in this manner (note that the superscript ‘ r ’ denotes regression-fit line). This step was performed to obtain straight line pieces to facilitate orthogonal line generation for creating the vertical image segments.

Step 2: Next, the mid-points of each of the fitted medial axis line pieces $\lambda^{r_1}, \lambda^{r_2}, \dots, \lambda^{r_{10}}$ were obtained and used as points upon which perpendicular lines were drawn as shown in Figure 2.12c. The perpendicular lines were extended by 200 pixels in the top and bottom directions so as to include all the pixels within the epithelium region.

Step 3: For each fitted medial axis line piece $\lambda^{r_1}, \lambda^{r_2}, \dots, \lambda^{r_{10}}$, a bounding box was created by joining the endpoints of the fitted medial axis of each line piece and the endpoints of the perpendicular lines (obtained from Step 2) of each piece (illustrated in Figure 2.12d). In this manner, for each epithelium image I , ten bounding boxes (regions) were created.

These bounding boxes were used to crop or partition the epithelium image I (in RGB format) into ten sections, referred to as vertical segment images and denoted as

$I_S^1, I_S^2, \dots, I_S^{10}$ (in RGB format). Note that depending on the linear approximation of the ten different partitions for the medial axis, there is the potential that the bounding boxes containing the vertical segment images may overlap and there may be gaps. Features were extracted from each of the vertical segmented images and used for CIN grade classification.

Figure 2.13 shows the bounding box and the resulting ten vertical image segments obtained using this method for one of the sample epithelium images.

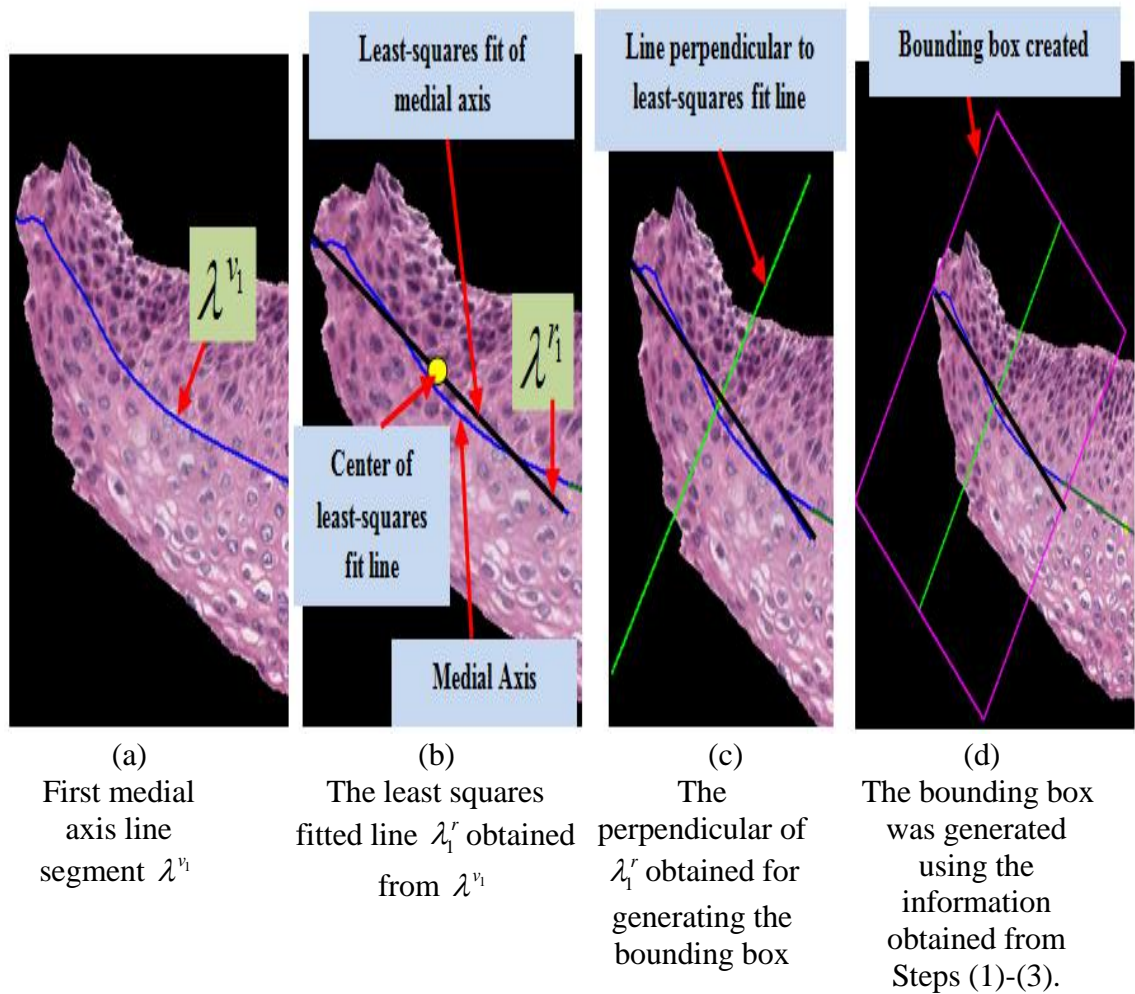


Figure 2.12: Vertical segment images creation.

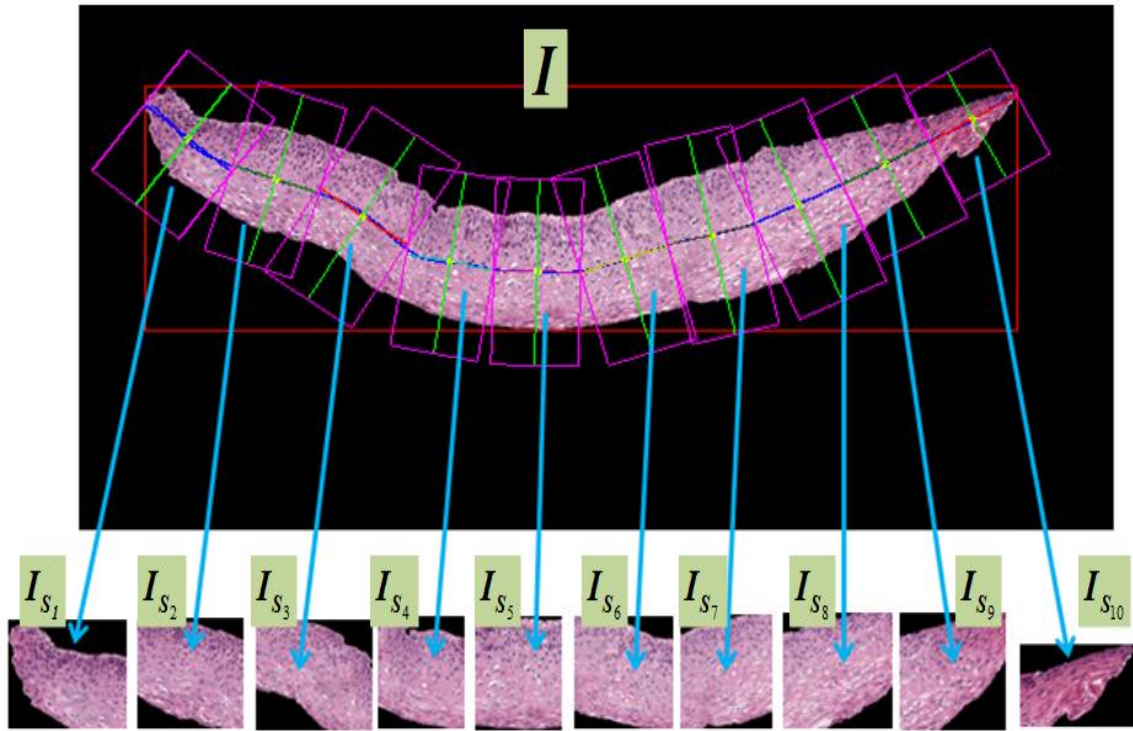


Figure 2.13: Vertical segment images.

2.3. FEATURE EXTRACTION

For each vertical segment image, $I_s^1, I_s^2, \dots, I_s^{10}$, four different types of features were computed, including: 1) texture features, 2) intensity shading features, 3) geometry (triangle) features, and, 4) profile-based correlation features. Table 2.1 shows the feature labels and a brief description of each feature while the following sections elaborate the features. Prior to feature extraction, the RGB images of the vertical segments ($I_s^1, I_s^2, \dots, I_s^{10}$) were converted to luminance grayscale images. The luminance images for the vertical segments were denoted as $L_s^1, L_s^2, \dots, L_s^{10}$ and were used for feature extraction. A summary of the features extracted from each vertical segment image are shown in Table 2.1 and are presented in the following sections. If I_s^n denotes the RGB image of each

vertical segment, where $n=1,2,...,10$ for the ten vertical segment images, then the luminance image of each vertical segment can be denoted as L_s^n . Features were extracted for each vertical segment L_s^n .

Table 2.1: Feature Description

Feature set	Label	Measure	Description
Texture features	F1	Contrast of segment	Returns a measure of the intensity contrast between a pixel and its neighbor over the whole image.
	F2	Energy of segment	Measures the entropy (squared sum of pixel values in the segment)
	F3	Correlation of segment	Returns a measure of how correlated a pixel is to its neighbor over the whole image.
	F4	Homogeneity of a segment	Returns a value that measures the closeness of the distribution of pixels in the segment to the segment diagonal.
	F5-F6	Contrast of GLCM	Measure of the contrast of the GLCM matrix obtained from the segment.
	F7-F8	Correlation of GLCM	Returns a value that measures the closeness of the distribution of elements in the GLCM to the GLCM diagonal.
	F9-F10	Energy of GLCM	Returns the sum of squared elements in the GLCM.
Intensity Shading features	F11	Percentage Light	Percentage of region that has the light pixels.
	F12	Percentage Medium	Percentage of region that has the medium pixels.
	F13	Percentage Dark	Percentage of region that has the dark pixels.
Triangle features	F14	Average area of triangles	This is the average area of the triangles formed by using Delaunay triangulation on the nuclei detected.
	F15	Std deviation of area of the triangles	This is the standard deviation of the area of the triangles formed by using Delaunay triangulation on the nuclei detected.
	F16	Average edge length	This is the mean of the length of the edges of the triangles formed.
	F17	Std deviation of edge length	Standard deviation of the length of the edges of the triangles formed.
Profile-based Correlation features	F18-F65	Weighted density distribution	Correlation of profile of the segment and WDD function.

2.3.1. Texture Features. Texture features are usually extracted using structural, spectral or statistical methods [23]. Previous studies on CIN grade classification used texture features such as fractal dimension [12] and the measures based on the gray-level

co-occurrence matrix (GLCM) [5]. The texture features used in this study include contrast (F1), energy (F2), correlation (F3) and uniformity (F4) of the segmented region, combined with the same statistics (contrast, energy and correlation) obtained from the GLCM of the segment (F5-F10, see Table 2.1) [24]. For calculation of the GLCM matrix, the immediate neighborhood pixels which lay within a radius of 1 were selected, producing the six GLCM related texture measures [5]. This means that the gray level co-occurrence was computed for each pixel $L_s^n(i, j)$ of the vertical image segments over the 3x3 neighborhood of $L_s^n(i, j)$ [25].

2.3.2. Intensity Shading Features. Previous studies for CIN grade classification primarily used texture and morphological features [5, 12, 22]. Previous studies for CIN grade classification primarily used features based on texture differences and the morphology of the epithelium region [5, 12, 23]. However, one of the important characteristics that pathologists use for CIN grade diagnosis includes the identification of hyperchromasia (dark staining) within the epithelium [3]. To that end, dark, light and medium intensity shading descriptors were explored in this research to characterize the intensity distribution within the different CIN grades. The luminance features characterized the dark structures such as the nuclei, light structures such as the light areas within the epithelium and the medium shading representative of the cytoplasm.

For extracting the luminance features, three different regions were marked as light, medium and dark shading within one of the sample images (CIN3 in this case) as shown in Figure 2.14. These three regions were randomly selected based on visual inspection of the image in Figure 2.14. Let us denote the three different regions as

marked in Figure 2.14 as *pixelsLight* , *pixelsMedium* and *pixelsDark* for the light, medium and dark shadings, respectively. The cluster centers representative of these three regions were obtained from these small patches and used to compute features for the remaining sixty-one images. The cluster centers were obtained by taking the mean value of the pixels within the light, medium and dark regions and denoted as *clustLight* , *clustMedium* and *clustDark* .

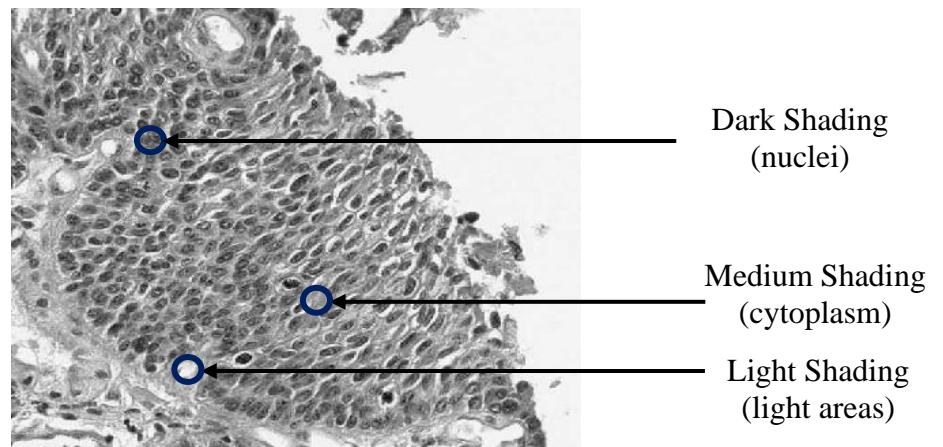


Figure 2.14: Representative shadings within a sample epithelium image.

Next, Euclidean distances in the spatial domain were computed between the cluster centers and each pixel of the vertical segment L_s^n . Based on the Euclidean distances, every pixel $L_s^n(i, j)$ in L_s^n was assigned to the nearest cluster center (*clustLight* , *clustMedium* or *clustDark*), and labeled as *classLight* , *classMedium* or *classDark* . Finally the intensity shading features were computed as Equations 1-3 and denoted as *percentLight* (F11), *percentMedium* (F12) and *percentDark* (F13). If *numLight* , *numMedium* and *numDark* represent the number of pixels that were assigned to *classLight* , *classMedium* and *classDark* respectively, then:

$$percentLight = \frac{numLight}{numLight + numMedium + numDark} \quad (1)$$

$$percentMedium = \frac{numMedium}{numLight + numMedium + numDark} \quad (2)$$

$$percentDark = \frac{numDark}{numLight + numMedium + numDark} \quad (3)$$

2.3.3. Geometry (Triangle) Features. Previous research has shown that the triangles formed by joining the centers of the nuclei detected, using a thresholding technique, provide structural information of the squamous epithelium [5, 10]. The triangles were formed using the Delaunay Triangulation (DT) approach. For a set of points, the DT is a type of triangulation in which the set of points are the vertices of the triangles formed using DT. Delaunay Triangulation exhibits the property that no point lies within the circles that are formed by joining the vertices of the triangles [10]. In other words, as shown in Figure 2.15c, all the triangles formed using DT are unique and do not contain any points within the triangles. The DT uses the coordinates of the points as inputs and provides the vertices of the triangles formed as outputs.

In the current work, first we used a Circular Hough Transform (CHT)-based circle detection method to detect the nuclei (black rounded objects in the epithelium region) from the vertical segmented images. The outputs of the CHT included: 1) Centers of the detected circular objects, and, 2) Distribution of radii (in pixels) of the detected circular objects, found within the vertical segments. For this study, we used the radius in the range $[minR, maxR]$, where $minR = 5$ pixels and $maxR = 15$ pixels, to threshold the circular objects that were representative of the nuclei sizes in the epithelium image

dataset. This radius range was obtained from empirical analysis of the data set for images from each of the CIN grades. The CHT implementation presented in [26] was used in this paper. Once the circles (representing the nuclei) were been detected using CHT, the centers of the detected circles were used as vertices to create triangles using the Matlab-based implementation of the Delaunay Triangulation (DT) method [27].

Figure 2.15 shows an example for locating the nuclei using the CHT and then applying the DT algorithm for triangulation was performed (Figure 2.15c). The features that were obtained from the triangles include: average area of the triangles (F14), standard deviation of the area of the triangles (F15), average distance between the edges of the triangles (F16) and standard deviation of the distance between the edges of the triangles (F17) [5,10].

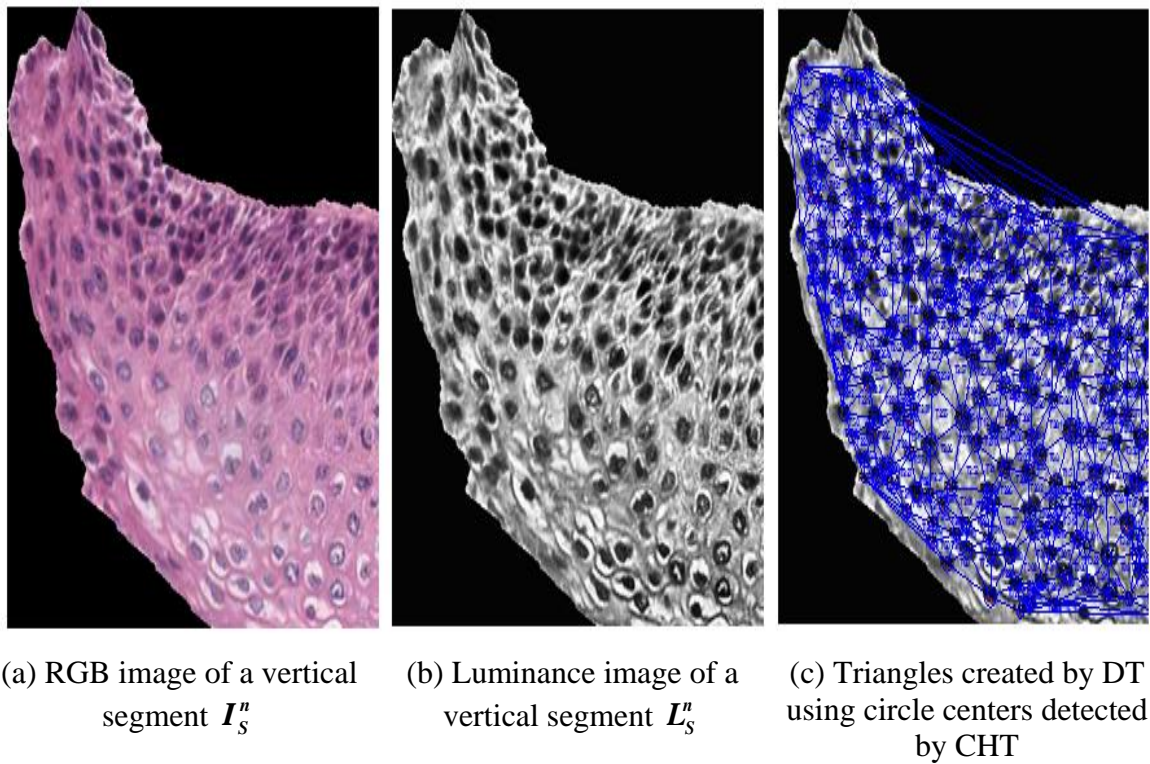


Figure 2.15: Triangles formed from the vertical segments.

2.3.4. Profile-based Correlation Features. The fourth set of features was based on computing profiles of the vertical segments and correlating the profiles with basis functions. In previous research, profile-based correlation features were computed for dermatology skin lesion discrimination based on correlating the luminance histogram for a skin lesion image with a set of Weighted Density Distribution (WDD) basis functions [28]. In this research, the method as used in [28] was applied to the 1-D profiles of the luminance images of the vertical segments L_s^n .

If R denotes the number of rows and C the number of columns for L_s^n , then the profile value of each row, $P_s(i)$ is defined as the average luminance value for each row of pixels inside the epithelium (pixels shown by the red lines in Figure 2.16) for each vertical segment luminance image L_s^n as given in Equation 4.

$$P_s(i) = \frac{\sum_{j=1}^C L_s^n(i, j)}{numPix(i)} \quad (4)$$

for $i = 1, 2, \dots, R$. Here $numPix(i)$ denotes the number of pixels that are inside the epithelium for the i^{th} row in the vertical segment luminance image L_s^n . Let $P_s = \{P_s(1), P_s(2), \dots, P_s(R)\}$ be the sequence of profile values obtained in this manner. Correlation-based features were extracted by correlating the profile (P_s) with the WDD functions shown in Figure 2.17.

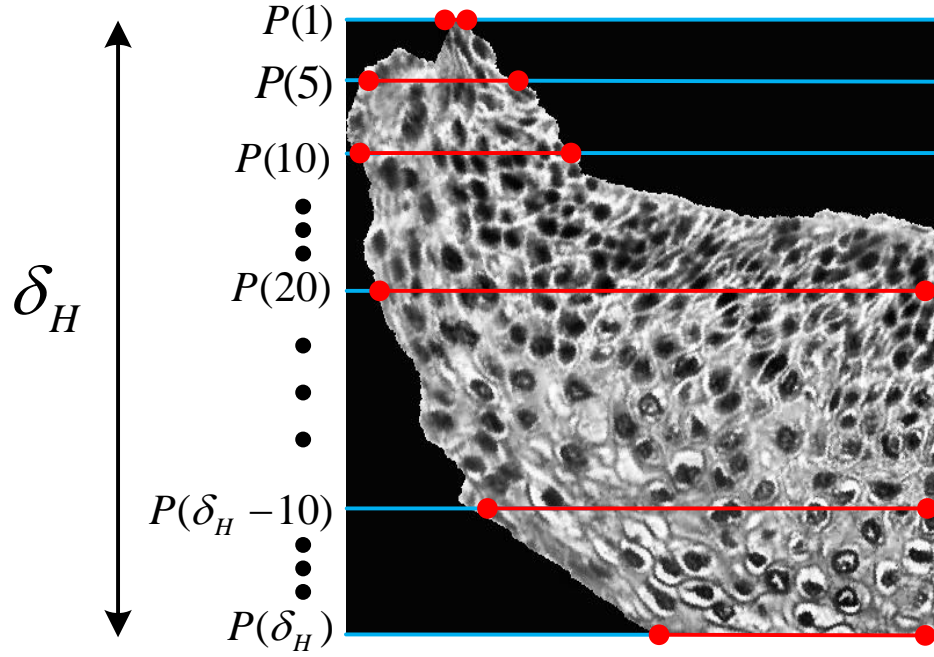


Figure 2.16: Profile obtained for computing the profile-based features.

Let T_1 denote the WDD function in Figure 2.17(a), T_2 denote the WDD function in Figure 2.17(b), and so forth.

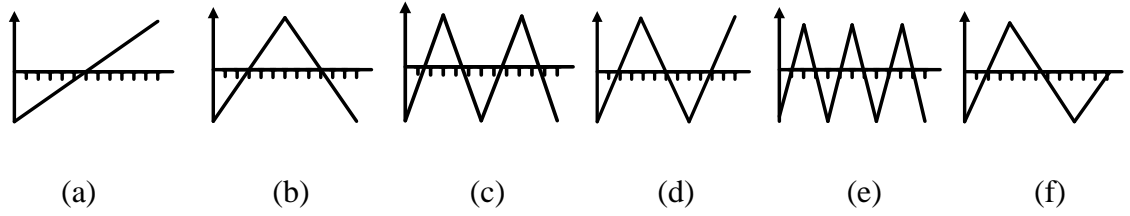


Figure 2.17: The WDD functions used.

The twelve profile-based correlation features were computed as follows.

For the profile P_s , six features q_1, q_2, \dots, q_6 were computed according to Equation 5.

$$q_k = \sum_{i=1}^{\delta_H} P_s(i) T_k(i) \quad (5)$$

for $k = 1, 2, \dots, 6$. Six additional features q_7, q_8, \dots, q_{12} were computed by correlating the six WDD functions with the sequence of absolute differences between profile sample values as Equation 6.

$$q_{k+6} = \sum_{i=1}^{\delta_H} |P_S(i) - P_S(i-1)| T_k(i) \quad (6)$$

for $k = 1, 2, \dots, 6$, and $P_S(i-1) = 0$ for $i = 1$. A total of forty-eight profile-based correlation features were obtained using the above method for the following four variations of the profile samples (P_S), obtained from the vertical segment under analysis:

- a) whole profile, such that $P_S = \{P_S(1), P_S(2), \dots, P_S(\delta_H)\}$ (F18-F29);
- b) top $1/3^{\text{rd}}$ of the profile such that $P_S = \{P_S(1), P_S(2), \dots, P_S(\delta_H/3)\}$ (F30-F41);
- c) middle $1/3^{\text{rd}}$ of the profile, such that $P_S = \{P_S(\delta_H/3 + 1), P_S(\delta_H/3 + 2), \dots, P_S(2\delta_H/3)\}$ (F42-F53); and,
- d) bottom $1/3^{\text{rd}}$ of the profile, such that $P_S = \{P_S(2\delta_H/3 + 1), P_S(2\delta_H/3 + 2), \dots, P_S(\delta_H)\}$ (F54-F65).

As shown above, each $1/3^{\text{rd}}$ of the profile (top to bottom) was analyzed separately along with the whole profile to extract profile-based correlation features for each $1/3^{\text{rd}}$ of the vertical segment images L_s^n . This was done because previous studies on CIN grade classification have shown that the epithelium shows structural variation (number of nuclei, and, nuclei-to-cytoplasm ratio) along each $1/3^{\text{rd}}$ (top to bottom) for the different CIN grades [3]. Finally, the feature vector (65x1) for each vertical segment image L_s^n was created as:

$$U_{L_s^n} = [F1, F2, F3, \dots, F65]^T \quad (7)$$

2.4. CLASSIFIER ALGORITHMS

Sixty-two cervigrams with expert pathologist manual segmentations of the squamous epithelium and labeled CIN grades (Normal, CIN1, CIN2 or CIN3) were obtained from NLM and examined in this study. Manual CIN grade labeling of the vertical segments within each image were done by the authors. Image-based epithelium analysis was performed based on classifying the ten individual vertical segments and using a voting scheme for final image-based classification. For each vertical segment, sixty-five features were extracted. In the classification experiments, the extracted features were used as inputs to the classifier. Support Vector Machines (SVM) and Linear Discriminant Analysis (LDA) were explored for individual vertical segment classification.

The SVM-based classifier presented in the implementation in [29] was used in this study. In general, SVM-based classifiers try to find an optimal hyperplane, which acts as a decision function to classify data in high dimensions. For linearly inseparable classes, the SVM optimizes to find a hyperplane that maximizes the class separation while minimizing a quantity (known as the penalty parameter), which is directly proportional to the number of misclassified instances. In this research, a linear kernel was used for the SVM implementation. The penalty parameter to be used in the SVM classifier for the linear kernel was obtained using the leave-one-image-out method, as explained in the experimental results section [30].

The other classification method that was investigated in this study included the LDA-based classifier [31]. The LDA method optimizes a linear transformation operator which depends on the ratio of the inter-class variance to the intra-class variance.

Optimization of this linear transformation operator was done by maximizing the inter-class variance normalized over intra-class variance, thereby guaranteeing maximum class separability [32-33]. The Matlab[®]-based implementation of the method presented in [31] was used for this study. For classification, the following four steps were performed:

Step 1: Train the classification algorithm (SVM or LDA) using a leave-one-image-out approach, whereby, the classifier is trained based on the individual vertical segment feature vectors $U_{I_s}^r$ for all but the left out epithelium image (used as the test image).

Step 2: Classify each vertical segment of the left out test image into one of the CIN grades using the SVM or LDA classifier.

Step 3: Assign the test epithelium image to the class (Normal, CIN1, CIN2, CIN3) using a voting scheme. The CIN grade of the test epithelium image was assigned based on whichever class is most frequently assigned to each of the vertical segments for the image (most frequently occurring class assignment for the ten vertical segments in Step 2). If there is a tie with the most frequently occurring class assignment among the vertical segments, then the epithelium image is assigned to the higher class. For example, if there is a tie between CIN2 and CIN3, then the image would be labeled as CIN3.

Step 4: Repeat steps 1-3 for all the epithelium images in the experimental data set.

Note that the epithelium image shown in Figure 2.14 was eliminated from the testing step. Hence for classification, sixty-one out of the sixty-two images were used. This was done because the cluster centers for the intensity shading features were computed from the image shown in Figure 2.14. Therefore in order to remove any bias during the classification process, this image was excluded for the classification experiments.

3. EXPERIMENTS PERFORMED

There were four sets of experiments that were performed in this study and are presented in this section.

3.1. CLASSIFICATION OF VERTICAL SEGMENT IMAGES

For the first set of experiments, all the features extracted from the vertical segment images were used as inputs to train the SVM/LDA classifier. Using the leave-one-image out approach (as presented in Step 1 of Section 2.4), the ten vertical segment images of the test image was assigned a CIN grade.

3.2. FUSION OF THE CIN GRADES OF VERTICAL SEGMENTS

In the second set of experiments, the CIN grades of the vertical segment images obtained from Section 3.1 were fused to obtain the CIN grade of the test epithelium image (see Figure 3.1). Fusion of the CIN grades of the vertical segment images was done using a voting scheme as presented in Step 2 of Section 2.4. For scoring the epithelium image classifications, three approaches were examined:

Approach 1 (Exact Class Label): The first approach is exact classification, meaning that if the class label automatically assigned to the test image is the same as the expert class

label, then the image is considered to be correctly labeled. Otherwise, the image is considered to be incorrectly labeled.

Approach 2 (Windowed Class Label): The second scoring approach is a windowed classification scheme. Using this approach, if the predicted CIN grade level for the epithelium image is only one grade off as compared to the actual CIN grade, we considered it as correct prediction. For example, if CIN1 was predicted as Normal or CIN 2, the result would be considered correct. If CIN1 was predicted as CIN3, the result would be considered incorrect.

Approach 3 (Normal vs. CIN): For the third approach, we considered the prediction incorrect when a Normal stage was predicted as any CIN stage and vice-versa.

3.3. CLASSIFICATION OF THE WHOLE EPITHELIUM

For the third set of experiments, features were extracted from the whole epithelium image, without creating the individual vertical segment images (see Figure 3.1). Features extracted from the whole image were used as inputs to the SVM/LDA classifier using the same leave-one-image out approach. This set of experiments was investigated to compare the performance of the fusion-based epithelium classification (Section 3.2) as compared to classifying the epithelium image as a whole. The same scoring approaches as presented in Section 3.2 were used to evaluate the performance of the whole epithelium classification.

Figure 3.1 overviews the classification methods using the fusion-based approach (Section 3.2) and the whole image approach (Section 3.3).

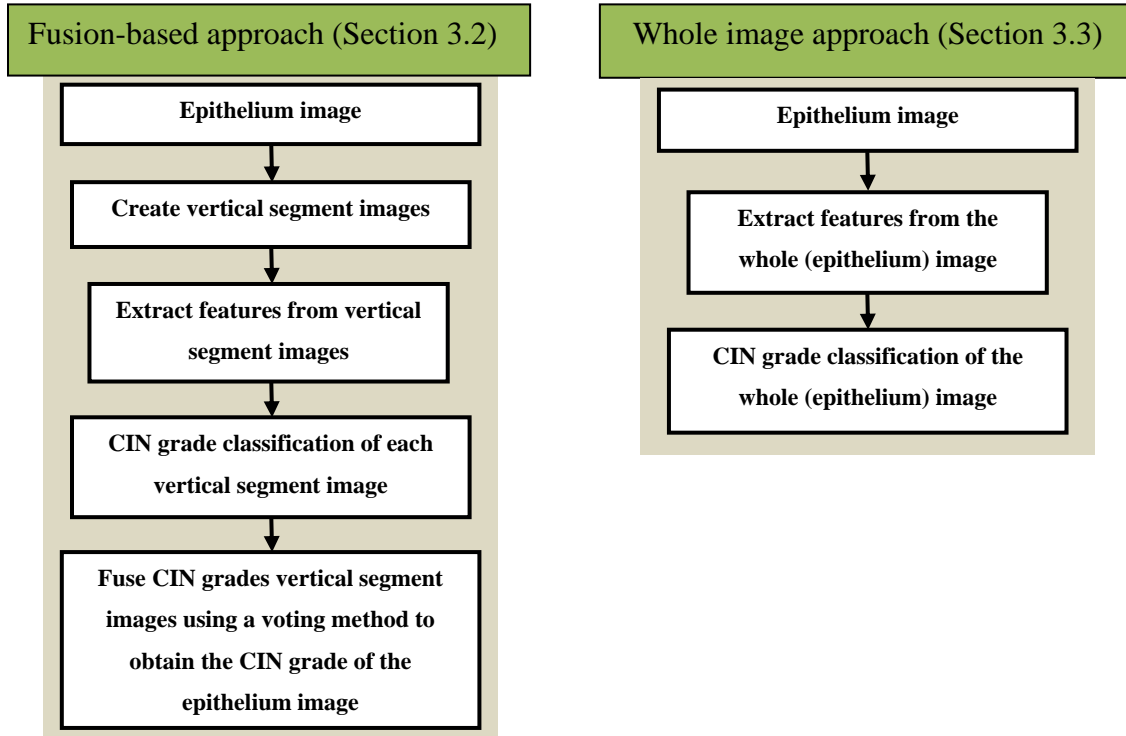


Figure 3.1: Fusion-based approach v/s whole image approach.

3.4. FEATURE EVALUATION AND SELECTION

For feature evaluation and selection, a SAS®-based implementation of Multinomial Logistic Regression (MLR) was used [33-35]. In general, MLR is used for modeling nominal outcome variables, wherein the log odds of the outcomes are modeled as a linear combination of the predictor variables. The p-values obtained from the SAS-based MLR output were used as a metric for feature selection. For feature selection, the p-values are used to check the null hypothesis that a particular predictor's regression coefficient is zero, provided the rest of the predictors in the model are rejected. Previous research has shown that if a feature has a p-value less than alpha (α), then the null hypothesis could be rejected and the feature is considered to be statistically significant [34-36].

4. EXPERIMENTAL RESULTS AND ANALYSIS

As previously stated, the vertical segment image classifications (CIN grading) were obtained using the SVM/LDA classifier using a leave one image out approach. The vertical segment classifications were fused using a voting scheme to obtain the CIN grade of the epithelium image. The performance evaluation of the epithelium image classifications using the fusion approach was done using the three approaches as presented in Section 3.2. Tables 4.1 and 4.2 give the confusion matrix for the epithelium image classification results obtained using the fusion-based approach, for the SVM and LDA classifiers respectively.

Table 4.1: Result of **SVM** classifier (features from vertical segments).

	Normal	CIN1	CIN2	CIN3
Normal	11	0	1	0
CIN1	4	5	4	0
CIN2	1	7	9	10
CIN3	2	0	0	9

Table 4.2: Result of **LDA** classifier result (features from vertical segments).

	Normal	CIN1	CIN2	CIN3
Normal	12	2	0	0
CIN1	4	7	4	0
CIN2	0	3	8	5
CIN3	0	0	2	14

From Table 4.1 and 4.2, we can see that the exact class label scoring scheme shows an accuracy of 55.7% using the SVM classifier and 67.2% using the LDA classifier. The Normal v/s CIN scoring approach yielded an accuracy of 90.2% for both

the SVM and LDA classifiers. For the windowed class scoring approach, SVM showed an accuracy of 96.7% while LDA showed an accuracy of 100.0%. Tables 4.1 and 4.2 present the results when the vertical segment classifications were fused to obtain the CIN grade of each epithelium image. However, in order to analyze the advantage of using the fusion-based approach for epithelium classification, we extracted features from the whole epithelium image instead of partitioning it into vertical segments, and the classified the image using the same SVM/LDA based classifiers as stated previously. Table 4.3 and 4.4 shows the whole image classification results.

Table 4.3: Result of **SVM** classifier result (features from whole images).

	Normal	CIN1	CIN2	CIN3
Normal	9	2	1	2
CIN1	0	5	3	1
CIN2	5	5	5	7
CIN3	2	1	5	8

Table 4.4: Result of **LDA** classifier result (features from whole images).

	Normal	CIN1	CIN2	CIN3
Normal	9	4	1	0
CIN1	5	2	5	1
CIN2	2	3	6	9
CIN3	0	4	2	8

From Tables 4.3 and 4.4, the exact class label scoring scheme provided an accuracy of 44.3% using the SVM classifier and 40.9% using the LDA classifier. The Normal v/s CIN scoring approach exhibited an accuracy of 80.3% for both the SVM and LDA classifiers. For the windowed class scoring approach, SVM's accuracy was 80.3% while LDA showed an accuracy of 86.9%. Comparing the results from Tables 4.1 and 4.2 to Tables 4.3 and 4.4, for the exact class label scoring, the fusion-based approach showed

an improvement of 11.4% in accuracy as compared to the whole image-based approach for the SVM classifier and an improvement of 26.3% for the LDA classifier. For Normal v/s CIN scoring, accuracies for both the SVM and LDA classifiers increased by 9.9% using the fusion-based approach as compared to the whole image-based approach. The accuracy using the fusion-based approach increased by 16.4% for SVM and 13.1% for LDA when the performance was evaluated using the windowed scoring method. The results presented in Tables 4.1-4.4 appear to indicate that the fusion-based method for CIN grade classification of the epithelium images provides a better estimate of the epithelium CIN grade as compared to analyzing the epithelium as a whole, thus validating the fusion-based approach developed in this study.

For feature evaluation and selection experiments, all the sixty-five features extracted from the individual vertical segments were used as inputs to the SAS-based MLR. For this study, we used $\alpha = 0.05$ to obtain the statistically significant features. As shown in Table 4.5, forty-two features out of the overall sixty-five features were obtained, whose p-values were less than α . Only the statistically significant features, as obtained from the feature evaluation and selection are shown in Table 4.5.

Table 4.5: Features with corresponding p-values.

Feature	P-value	Feature	P-value	Feature	P-value	Feature	P-value	Feature	P-value	Feature	P-value
F1	0.0001	F11	0.0001	F22	0.0002	F30	0.0001	F38	0.0002	F56	0.0001
F2	0.0001	F13	0.0001	F23	0.0001	F31	0.0001	F39	0.0067	F57	0.0001
F4	0.0001	F14	0.0021	F24	0.0001	F32	0.0027	F40	0.0001	F60	0.0021
F5	0.0010	F16	0.0039	F25	0.0007	F33	0.0034	F42	0.0007	F61	0.0022
F7	0.0023	F18	0.0047	F26	0.0001	F34	0.0001	F48	0.0001	F62	0.0001
F9	0.0001	F19	0.0019	F28	0.0036	F35	0.0039	F54	0.0045	F64	0.0001
F10	0.0001	F20	0.0033	F29	0.0001	F37	0.0001	F55	0.0047	F65	0.0011

Based on the statistically significant features as obtained in Table 4.5, the next set of experiments were conducted to evaluate the classification accuracies using the reduced features as compared to using all the sixty-five features. Note that the reduced feature classifications were done only for the fusion-based epithelium analysis. The classification algorithms (SVM/LDA) were applied to the reduced features obtained from the vertical segments, followed by fusing of the vertical segment classifications to obtain the CIN grade of the epithelium. The reduced feature classifications obtained in this manner is summarized in the confusion matrices in Table 4.6 and Table 4.7 for the SVM- and LDA-based methods respectively.

Table 4.6: Result of **SVM** classifier (reduced features & vertical segments).

	Normal	CIN1	CIN2	CIN3
Normal	11	0	0	0
CIN1	3	6	3	0
CIN2	2	6	10	11
CIN3	0	0	1	8

Table 4.7: Result of **LDA** classifier (reduced features & vertical segments).

	Normal	CIN1	CIN2	CIN3
Normal	13	0	0	0
CIN1	3	10	3	0
CIN2	0	2	10	4
CIN3	0	0	1	15

From Table 4.6, the following correct recognition rates were obtained for the reduced features using the SVM-based classifier: windowed class label of 96.7%, Normal vs. CIN of 91.8%, and exact label of 57.4%. From the confusion matrix shown in Table 4.7, the following correct recognition rates using the reduced features for the LDA-based

classifier were obtained: windowed class label of 100.0%, Normal vs. CIN of 95.1%, and exact label of 78.7%. Here one can see that the windowed classification scheme produced consistently high correct recognition rates for both the SVM- and LDA-based classifiers. Normal vs. CIN correct recognition rates were high using both classifiers (SVM: 91.8% vs. LDA: 95.1%). For the exact class labels, the LDA-based classifier had a correct recognition rate of 78.7% as compared to SVM's 57.4%. Based on the results in Table 4.6 and Table 4.7, we can observe that the reduced feature analysis of the epithelium images improves the accuracy of CIN grade classification. For the SVM classifier, an improvement of 1.7% in accuracy was obtained for CIN grade classification using the fusion-based approach, while the LDA classifier showed an improvement of 11.5% in accuracy using the fusion-based approach for exact CIN labeling scoring approach. It could also be observed in both Tables 4.6 and 4.7 that most confusion occurred between the CIN1 and CIN2 grades.

Overall, the classification results demonstrate that the windowed approach and the Normal vs. CIN scoring approach provide excellent accuracies of prediction for the data set investigated in this study. The windowed classification results presented in this study are important considering the intra- and inter-expert pathologist variability for CIN grade assignment of the epithelium [6-8]. The high classification accuracy obtained using the windowed approach clearly demonstrates that the nearby classes (Normal/ CIN1 or CIN1/CIN2 or CIN2/CIN3) are very similar and therefore difficult to discriminate. It also explains the intra- and inter- pathologist variation in labeling of these images.

As compared to the windowed (Approach 2) and Normal vs. CIN (Approach 3) classification schemes, the exact class labels (Approach 1) had lower prediction accuracy.

However, the 78.7% accuracy of exact class label prediction using the reduced features is 16.4% higher than the existing benchmark results for automated CIN diagnosis (62.3%) as presented by Keenan et al. in [10] and 10.7% higher than the accuracy of the method used by Guillaud et al. (68%) in [13]. This suggests that the method developed in this study using the vertical segment analysis and fusion of the vertical segment CIN grades for obtaining the image-based CIN classification is an improvement over the existing methods for automated CIN diagnosis. Even though our method outperformed the existing methods for CIN grade classifications, based on our observation, there are possible contributors to the variations and relatively low classification results for the exact class label scoring approach. These factors include similarity between images of nearby CIN grades such as CIN1 and Normal, CIN1 and CIN2. As shown in the confusion matrix, most misclassifications occurred between CIN1 and CIN2, suggesting that these are very similar classes. Also, the investigator designated individual vertical segment class labels have some variations and inaccuracies as well since these were not done by an expert pathologist.

5. CONCLUSIONS

In this study, a framework for automated CIN grade classification of segmented epithelium regions has been developed. The method developed in this research includes a proposed distance transform and bounding box- based medial axis detection algorithm followed by partitioning the epithelium region into vertical segments using the medial axis. Within the current framework, epithelium region classification was performed based on: 1) Expert-labeling of the epithelium images as Normal, CIN1, CIN2, and CIN3, and, 2) Investigator-labeling of the created vertical segments into the same four CIN grades. The vertical segments were classified using a SVM or LDA classifier, based on the investigator-labeled training data of the segments. Finally, a novel fusion-based epithelium image classification is proposed wherein a voting scheme was used to fuse the vertical segment classifications to obtain the image-based classification of the segmented epithelium.

We found that the LDA-based classifier outperformed the SVM-classifier in all the scoring approaches used. Using the LDA classifier upon the reduced set of features and based on a windowed classification scheme for epithelium region classification, correct prediction rates as high as 100% were obtained. Normal vs. CIN classification rates was as high as 95.1% while the rates for exact class labels were as high as 78.7% using the reduced set of features. Future research would involve using computational intelligence-based methods for medial axis detection and CIN grade classification. However for incorporating such methods, it is important to include more training images from the cervigrams to obtain a comprehensive data set for different CIN grades. Since

the bottleneck for automated CIN grade classification is the computationally intensive and time consuming automatic epithelium segmentation, any algorithm developed for medial axis detection and CIN grade classification needs to be computationally very fast in order to facilitate a reasonable time for automated CIN diagnosis [5, 37].

REFERENCES

1. Parkin, D.M., Bray, F.I., Devesa, S.S. (2000). Cancer burden in the year 2000: The global picture. *Eur J Cancer* 2001; 37 Suppl 8:S4-66.
2. Jeronimo, J., Schiffman, M. (2004). A tool for collection of region based data from uterine cervix images for correlation of visual and clinical variables related to cervical neoplasia. *Computer-Based Medical Systems*, Proceedings. 17th IEEE Symposium on, vol., no., pp. 558- 562, 24-25.
3. Kumar, V., Abbas, A., Fausto, N., Aster, J. (2009). Chapter 22: The Female Genital Tract. *Robbins & Cotran Pathologic Basis of Disease*, pp 1017-1021.
4. He, L., Long, L.R, Antani, S., Thoma, G.R. (2011). Computer Assisted Diagnosis in Histopathology. *Sequence and Genome Analysis: Methods and Applications*, Z. Zhao, eds., iConcept Press, pp. 271-287.
5. Wang, Y., Crookes, D., Eldin, O.S., Wang, S., Hamilton, P., Diamond, J. (2009). Assisted Diagnosis of Cervical Intraepithelial Neoplasia (CIN). *Selected Topics in Signal Processing, IEEE Journal of*, vol.3, no.1, pp.112-121.
6. McCluggage, W.G. et al. (1998). Inter- and intra-observer variation in the histopathological reporting of cervical squamous intraepithelial lesions using a modified bethesda grading system. *BJOG: An Int. J. Obstet. and Gynecol.*, vol. 105, no. 2, pp. 206–210.
7. Ismail, S.M. et al. (1990). Reporting cervical intra-epithelial neoplasia (CIN): Intra- and interpathologist variation and factors associated with disagreement. *Histopathology*, vol. 16, no. 4, pp. 371–376.
8. Molloy, C. et al. (2002). Evaluation of colposcopically directed cervical biopsies yielding a histologic diagnosis of CIN 1, 2. *J. Lower Genital Tract Dis.*, vol. 6, no. 2, pp. 80–83.
9. Soenksen, D. (2006). The road to digital pathology. *Proc. Pathology Vision, Aperio Digital Pathology Solution Conf.*, Leeds, U.K.
10. Keenan, S.J., Diamond, J., McCluggage, W.G., Bharucha, H., Thompson, D., Bartels, P. H., Hamilton P.W. (2000). An automated machine vision system for the histological grading of cervical intraepithelial neoplasia (CIN). *Journal of Pathology*, 192(3), 351–362.
11. MacAulay, C., Palcic, B. (1990). An edge relocation segmentation algorithm. *Analyt. Quantit. Cytol. Histol.* 12(3), 165– 171.

12. Guillaud, M., Cox, D., Malpica, A., Staerkel, G., Maticic, J., Niekirk, D. V., Adler-Storthz, K., Poulin, N., Follen, M., MacAulay, C. (2004). Quantitative histopathological analysis of cervical intra-epithelial neoplasia sections: methodological issues. *Cellular Oncology*, 26, 31–43.
13. Guillaud, M., Adler-Storthz, K., Malpica, A., Staerkel, G., Maticic, J., Niekirk, D. V., Cox, D., Poulin, N., Follen, M., MacAulay, C. (2005). Subvisual chromatin changes in cervical epithelium measured by texture image analysis and correlated with HPV. *Gynecologic Oncology*, 99, S16–S23.
14. Barboni Miranda, G.H., Garcia Soares, E., Barrera, J., Cezar Felipe, J. (2012). Method to support diagnosis of cervical intraepithelial neoplasia (CIN) based on structural analysis of histological images. *Computer-Based Medical Systems (CBMS), 2012 25th International Symposium on*, vol., no., pp.1-6, 20-22.
15. Price, G.J., McCluggage, W. G., Morrison, M. L., McClean, G., Venkatraman, L., Diamond, J., Bharucha, H., Montironi, R., Bartels, P. H., Thompson, D., Hamilton, P.W. (2003). Computerized diagnostic decision support system for the classification of preinvasive cervical squamous lesions. *Human Pathology*, 34(11), 1193–1203.
16. Rahmadwati, R., Naghdy, G., Ros, M., & Todd, C. (2012). Computer aided decision support system for cervical cancer classification. In *SPIE Optical Engineering + Applications* (pp. 849919-849919). International Society for Optics and Photonics.
17. Wang, Y. et al. (2007). A color-based approach for automated segmentation in tumor tissue classification. *Proc. 29th Annu. Conf. IEEE Engineering in Medicine and Biology Society 2007 (EMBS 2007)*.
18. Wang, Y., Turner, R., Crookes, D., Diamond, J., Hamilton, P. (2007). Investigation of Methodologies for the Segmentation of Squamous Epithelium from Cervical Histological Virtual Slides. *International Machine Vision and Image Processing Conference IMVIP 2007*. Maynooth, Ireland, pp.83-90, 5-7.
19. Marel, J.V.D., Quint, W. G.V., Schiffman, M., van-de-Sandt, M. M., Zuna, R. E., Terence-Dunn, S., Smith, K., Mathews, C. A., Gold, M. A., Walker, J., Wentzensen, N. (2012). Molecular mapping of high-grade cervical intraepithelial neoplasia shows etiological dominance of HPV16. *Int. J. Cancer*, 131: E946–E953. doi: 10.1002/ijc.27532.
20. Paglieroni, D. (1992). Distance Transforms: Properties and Machine Vision Applications. *Computer Vision, Graphics, and Image Processing: Graphical Models and Image Processing*, Vol. 54, No. 1, pp. 57-58.

21. Maurer, C.R., Rensheng, Q., Raghavan, V. (2003). A linear time algorithm for computing exact Euclidean distance transforms of binary images in arbitrary dimensions. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol.25, no.2, pp. 265- 270.
22. Rao, C.R., Toutenburg, H., Fieger, A., Heumann, C., Nittner, T., Scheid, S. (1999). Linear Models: Least Squares and Alternatives. *Springer Series in Statistics*.
23. Gonzalez, R., Woods, R. (2002). *Digital Image Processing*, 2nd ed. Englewood Cliffs, NJ: Prentice-Hall, 2002.
24. Soh, L.K., Tsatsoulis, C. (1999). Texture analysis of SAR sea ice imagery using gray level co-occurrence matrices. *Geoscience and Remote Sensing, IEEE Transactions on* , vol.37, no.2, pp.780-795.
25. Stanley, R. J., De, S., Demner-Fushman, D., Antani, S., Thoma, G. R. (2011). An image feature-based approach to automatically find images for application to clinical decision support. *Computerized Medical Imaging and Graphics*; 35(5):365-72.
26. Borovicka, J. (2003). Circle Detection using Hough Transforms. *Course Project: COMS30121- Image Processing and Computer Vision*.
27. Preparata, F. R., Shamos, M. I. (1985). *Computational Geometry: An Introduction*. New York: Springer-Verlag.
28. Stanley, R.J., Stoecker, W.V., Moss, R.H. (2008). A basis function feature-based approach for skin lesion discrimination in dermatology dermoscopy images. *Skin Research and Technology*, 14(4), pp.425-435.
29. Chang, C.C., Lin, C.J. (2001). LIBSVM: a library for support vector machines. *Software available at* <http://www.csie.ntu.edu.tw/~cjlin/libsvm>
30. Fan, R.E., Chen, P.H., Lin, C.J. (2005). Working set selection using second order information for training SVM. *Journal of Machine Learning Research* 6, 1889-1918.
31. Krzanowski, W. J. (1988). Principles of Multivariate Analysis: A User's Perspective. New York: Oxford University Press.
32. Johnson, R.A, Wichern, D.W. (1988). Applied multivariate statistical analysis. Prentice Hall.
33. Li, T., Zhu, S., Ogihara, M. (2006). Using discriminant analysis for multi-class classification: an experimental investigation. *Knowl Inf Syst* 10(4): 453–472.

34. Hosmer, D., Lemeshow, S. (2000). Applied Logistic Regression (Second Edition). New York: John Wiley & Sons, Inc.
35. Agresti, A. (1996). An Introduction to Categorical Data Analysis. New York: John Wiley & Sons, Inc.
36. Pal, M. (2012). Multinomial logistic regression-based feature selection for hyperspectral data. *International Journal of Applied Earth Observation and Geoinformation*, Volume 14, Issue 1, February, Pages 214-220.
37. Xue, Z., Long, R.L., Antani, S., Neve, L., Zhu, Y., Thoma, G.R. (2010). A unified set of analysis tools for uterine cervix image segmentation. *Computerized Medical Imaging and Graphics*, Volume 34, Issue 8, Pages 593-604.

SECTION 2. CONCLUSIONS

This dissertation proposes data fusion techniques for NDE data, biomedical images with text, and histology cervical cancer images. Data fusion was applied to these areas to extend the state-of-the-art methods for:

- 1) **structural defect detection and characterization** using intra- and inter-modal fusion of NDE data,
- 2) **automated biomedical text detection and recognition** using fusion of horizontal and vertical text, and,
- 3) **cervical cancer diagnosis** by fusion of diagnosis from different parts of the epithelium to obtain an image-based diagnosis.

To summarize, the research performed showed that fusion of data at the raw-, feature- and decision-level is important for enhanced detection and characterization. Raw-data fusion using geometrical methods were used to obtain a better estimation of the medial axis for histology image classification. Feature level fusion using statistical methods were used for biomedical images for enhanced character recognition.

Multi-modal decision level fusion of NDE data using RX-statistic and fuzzy logic were used to create geometrical models of aircraft structures for FEM and analysis. These multi-modal fused models showed enhanced ability to predict failure as compared to uni-modal models. Additionally, decision level fusion was performed using a voting scheme for the vertical segmented images in histology image classification.

The research presented in this dissertation concludes that: 1) feature-level data fusion techniques are useful for dimension reduction, which in turn can increase the prediction accuracy, 2) decision level data fusion techniques could be employed for enhanced classification. Experimental results from this dissertation show that as compared to conventional image processing techniques, data fusion methods can provide enhanced image analysis and classification by using feature- and decision-level fusion techniques.

BIBLIOGRAPHY

1. Hall D.L., Llinas J. (1997) "An introduction to multisensor data fusion", Proceedings of the IEEE, 85(1), pp. 6-23, 1997.
2. Punska O., "Bayesian Approaches to Multi-Sensor Data Fusion", MS Thesis, Department of Engineering, University of Cambridge, 1999.
3. Manyika J., Whyte, H.D., "Data Fusion and Sensor Management: A Decentralized Information-Theoretic Approach", Ellis Horwood, New-York, London, 1994.
4. Hall D., "Mathematical Techniques in Multisensor Data Fusion", Boston, MA: Artech House, 1992.
5. Klein L.A., "Sensor and Data Fusion Concepts and Applications", SPIEOpt. Engineering Press, Tutorial Texts, vol. 14, 1993.
6. Bayes T., "An essay towards solving a problem in the doctrine of chances", Biometrika, vol. 45, pp. 293-315, 1958.
7. Berger J.O., Statistical Decisions (second edition), Springer-Verlag, Berlin, 1985.
8. Basseville M., Nikiforov, I.V., "Detection of Abrupt Changes: Theory and Application", Prentice Hall, Information and system science series, 1993.
9. Blasch E., Kadar I., Salerno J., Kokar M.M., Das S., Powell G.M., Corkill D.D., Ruspini E.H., "Issues and challenges of knowledge representation and reasoning methods in situation assessment (Level 2 Fusion)", J. of Advances in Information Fusion, Dec. 2006.
10. Mitchell H.B., Multi-Sensor Data Fusion- An Introduction, Springer, 2007.
11. Mitchell H.B., Image Fusion- Theories, Techniques and Applications (first edition), Springer, 2010.
12. Liu Z. et al., "Survey: State of the Art in NDE Data Fusion Techniques", IEEE Trans. Instrum. Meas. 56:2435-2451, 2007.
13. Eakins, J.P., Graham, M.E., "Content-based image retrieval", Tech. Rep. JTAP-039, JISC Technology Application Program, Newcastle upon Tyne, 2000.
14. Pfund, T., Marchand-Maillet, S., "Dynamic multimedia annotation tool", in: G. Beretta, R. Schettini (Eds.), Proceedings of the SPIE Photonics West Conference on Internet Imaging III, vol. 4672, San Jose, CA, USA, pp. 216—224, 2002.

15. Müller, H., Michoux, N., Bandon, D., Geissbuhler, A., “A review of content-based image retrieval systems in medical applications—clinical benefits and future directions”, *International Journal of Medical Informatics*, Volume 73, Issue 1, Pages 1-23, 2004.
16. Tang, L., Hanka, R., Ip, H., “A review of intelligent content-based indexing and browsing of medical images”, *Health Informatics Journal*, 5(1):40-49, 1999.
17. Kumar, V., Abbas, A., Fausto, N., Aster, J., “Chapter 22: The Female Genital Tract. Robbins & Cotran Pathologic Basis of Disease”, pp 1017-1021, 2009.
18. Marel, J.V.D., Quint, W. G.V., Schiffman, M., van-de-Sandt, M. M., Zuna, R. E., Terence-Dunn, S., Smith, K., Mathews, C. A., Gold, M. A., Walker, J., Wentzensen, N, “Molecular mapping of high-grade cervical intraepithelial neoplasia shows etiological dominance of HPV16”, *Int. J. Cancer*, 131: E946–E953. doi: 10.1002/ijc.27532.
19. He, L., Long, L.R, Antani, S., Thoma, G.R., “Computer Assisted Diagnosis in Histopathology”, *Sequence and Genome Analysis: Methods and Applications*, Z. Zhao, eds., iConcept Press, pp. 271-287, 2011.
20. Wang, Y., Crookes, D., Eldin, O.S., Wang, S., Hamilton, P., Diamond, J., “Assisted Diagnosis of Cervical Intraepithelial Neoplasia (CIN)”, *Selected Topics in Signal Processing*, *IEEE Journal of*, vol.3, no.1, pp.112-121, 2009.
21. W. G. McCluggage et al., “Inter- and intra-observer variation in the histopathological reporting of cervical squamous intraepithelial lesions using a modified bethesda grading system”, *BJOG: An Int. J. Obstet. and Gynecol.*, vol. 105, no. 2, pp. 206–210, 1998.
22. Ismail S.M., et al., “Reporting cervical intra-epithelial neoplasia (CIN): Intra- and interpathologist variation and factors associated with disagreement,” *Histopathology*, vol. 16, no. 4, pp. 371–376, 1990.
23. Keenan, S.J., Diamond, J., McCluggage, W.G., Bharucha, H., Thompson, D., Bartels, P. H., Hamilton P.W., “ An automated machine vision system for the histological grading of cervical intraepithelial neoplasia (CIN)”, *Journal of Pathology*, 192(3), 351–362.

VITA

Soumya De was born in Asansol in the state of West Bengal in India. After completing schooling at the St. Partick's Higher Secondary School, Asansol (1991-2001) and the DAV Model School, Durgapur (2001-2003), he went to Nagpur University, Nagpur for his B.E. in Electronics Engineering (2007). Thereafter he went to the Missouri University of Science and Technology (Missouri S&T) at Rolla, where he completed his MS in Electrical Engineering (2008) and continued his PhD in the same program. During his program of study at Missouri S&T, he was a research assistant with the Image Processing laboratory and the Electromagnetic Compatibility Laboratory, where he worked in the areas of image processing methods for nondestructive evaluation, biomedical engineering and signal integrity applications. He worked as an intern with the Image Science and Machine Vision group at the Oak Ridge National Laboratory during the summers of 2009 and 2010. He also worked with the High Speed Design team at Cisco as an intern in 2011 and 2012. He received his PhD in Electrical Engineering in May 2013 at Missouri S&T. He joined Cisco as a Signal Integrity Engineer in December 2012 after completion of his PhD.