

©Copyright 2019

Rouzbeh Davoudi

# A Machine Learning and Computer Vision Framework for Damage Characterization and Structural Behavior Prediction

Rouzbeh Davoudi

A dissertation  
submitted in partial fulfillment of the  
requirements for the degree of

Doctor of Philosophy

University of Washington

2019

Reading Committee:

Gregory R. Miller, Chair

Marc O. Eberhard

Paolo Calvi

Program Authorized to Offer Degree:  
Civil and Environmental Engineering

University of Washington

## **Abstract**

A Machine Learning and Computer Vision Framework for Damage Characterization and Structural Behavior Prediction

Rouzbeh Davoudi

Chair of the Supervisory Committee:  
Professor Gregory R. Miller  
Civil and Environmental Engineering

This work focuses on using computer vision to relate surface (and limited subsurface) damage observations to quantitative damage and load levels in structural components. In particular, image processing and machine learning regression techniques have been used to build predictive models capable of estimating internal loads (e.g., shear and moment) and damage states in RC beams, slabs, and panels based on surface crack pattern images. The predictive models have been generated and tested using image data sets obtained from earlier published studies, which together provide about 1,500 crack pattern images captured from about 170 individual RC beam, slab, and panel tests across a range of load and damage levels.

Working with these existing image data sets, various textural and geometric attributes of surface crack patterns have been defined and evaluated with respect to their effectiveness in building useful estimation models. Relatively simple crack representations have been used, consistent with the varying nature of the images available in the earlier studies, but also with an eye toward potential field applications in which image capture and segmentation quality could be limited. The fundamental state quantification tasks range from the relatively simple (e.g., given an image showing damage, generate from that image a predicted load level in terms of percentage of capacity) to relatively advanced (e.g., estimating principal stresses in a panel).

In addition to investigating the basic feasibility of the approach, these studies also identify and evaluate a range of strategies, algorithms, and parameters that affect the accuracy of the estimations, and these are discussed, as well. In general terms, the results show that the predictive models based on surface crack image data can work well across a wide range of geometries, loadings, concrete strengths, and reinforcement details. Size effects can be accounted for by including specimen physical dimensions in the feature sets used for model training, and fundamental design relations can be used to develop useful non-dimensional prediction parameters.

## TABLE OF CONTENTS

	Page
List of Figures . . . . .	iv
Chapter 1: Introduction . . . . .	1
1.1 Overview and Motivation . . . . .	2
1.2 Goals of Research . . . . .	3
1.3 Organization of the dissertation . . . . .	4
Chapter 2: An exploratory model for relating surface observation to internal damage	9
2.1 Abstract . . . . .	10
2.2 Introduction and Motivation . . . . .	10
2.3 Preliminary Work and X-Ray Image Data Set . . . . .	11
2.4 Image Analysis and Machine Learning Methodology: . . . . .	12
2.5 Finding Dominant features . . . . .	18
2.6 Crack volume estimation based on surface area of cracks . . . . .	19
2.7 Pull-out Work Estimation Based on Bazant's Model . . . . .	19
2.8 Model analysis . . . . .	20
2.9 Performance of predictive algorithms . . . . .	21
2.10 Conclusion and Results . . . . .	22
Chapter 3: Computer Vision-Based Damage Inspection Framework for Shear Critical Beams (With permission from ASCE) . . . . .	39
3.1 abstract . . . . .	40
3.2 Introduction and Motivation . . . . .	40
3.3 Previous Experimental Studies and Image Data . . . . .	45
3.4 Image-Based Estimation Framework . . . . .	46
3.5 Modal Analysis and Cross-Validation . . . . .	53
3.6 Performance of estimation models . . . . .	54

3.7	Results . . . . .	57
3.8	Summary and Conclusions . . . . .	61
Chapter 4: Computer Vision-Based Damage Inspection Framework for RC Beams and Slabs . . . . . 79		
4.1	abstract . . . . .	80
4.2	Introduction and Motivation . . . . .	80
4.3	Prior Work Applying Computer Vision for Damage Inspection . . . . .	83
4.4	Prior Experimental Studies and Image Data Sets . . . . .	84
4.5	Image-Based Classification and Estimation Framework . . . . .	87
4.6	Error Measures and Cross Validation . . . . .	91
4.7	Estimation Results . . . . .	93
4.8	Summary and Conclusions . . . . .	98
Chapter 5: Computer Vision-Based Stress State Quantification for RC Panels . . . 116		
5.1	abstract . . . . .	117
5.2	Introduction . . . . .	117
5.3	Scope and Organization of the chapter . . . . .	121
5.4	Image Data Set . . . . .	121
5.5	Image Processing . . . . .	122
5.6	Principal Stress/Strain Orientation (Theoretical vs Image-Based) . . . . .	125
5.7	Data Subsampling . . . . .	126
5.8	Feature Engineering . . . . .	127
5.9	Crack-Based Estimation Models . . . . .	128
5.10	Model Analysis and Performance of Estimation Models . . . . .	129
5.11	Estimation Results . . . . .	130
5.12	Summary and Conclusions . . . . .	134
Chapter 6: Conclusion and Suggested Future Work . . . . . 149		
6.1	Conclusion and contributions . . . . .	150
6.2	Limitations . . . . .	151
6.3	Future work in building a data repository . . . . .	152
6.4	Future work in damage quantification . . . . .	152
6.5	Future work in combining mechanics-based with data-driven approaches . . . 152	

Bibliography ..... 154

## LIST OF FIGURES

Figure Number	Page
1.1 Typical damage assessment by the inspection team . . . . .	3
1.2 (a) Typical cracked bridge girder. The photo captured at Exit 2C, I5 N, Seattle. (b) Developed structural cracks in the girder . . . . .	4
2.1 Pull-out test setup [72] used to obtain the CT specimen images analyzed in the present study . . . . .	24
2.2 Typical Pull Out Test Load-displacement Response [72]. CT imaging points are indicated by "Step n" labels on the figure. Strength loss for last two steps was shown . . . . .	25
2.3 (a) 3D CT image data for each specimen/load step combination, captured from a total of eight specimens using one or two load steps per specimen. (b) Extended surface crack image data for each specimen/load step instance, leading to 40 total surface crack image sets separated into training and test data sets. (c) sub-sampling of surface crack images using two methods: 1) Four crack regions extracted via a randomly translated fixed sample length; 2) Four crack regions extracted using non-overlapping contiguous intervals . . . . .	27
2.4 (a) Typical cylinder splitting failure mode [72] (b) Raw X-ray image. (c) Segmented image using threshold technique. (d) Automated process to find and extract cracks in each 2D segmented image. (e) 3D segmentation (f) Crack surface density distribution (g) Crack volume density distribution (h) Longitudinal surface crack 1 from surface observation. (i) Feature set matrix extracted from longitudinal surface cracks (j) True class set matrix representing the values of the dependent variable of the feature set . . . . .	28
2.5 (a) Raw X-ray image. (b) Segmentation using gray threshholding technique. (c) de-nosing process through LmakeHmoonRC's method, this method improves segmentation by smoothing the image (e) Filling the holes left from de-noised image. . . . .	29
2.6 Used procedure to predict total fracture energy for the specimen (i.e. specimen i, load step jj). $U_1$ , and $U_{p1}$ are true and predicted fracture energy corresponding to crack plane 1 receptively. Modified LOOCV was used to define training and test data sets as explained in Figure 2.3 . . . . .	29



2.7	Comparison of volume density predictive models using Gaussian regression technique and Standard deviation feature with crack volume estimation model (Baseline). Data set was generated using entire crack length (No sampling) .	30
2.8	Comparison of volume density predictive models using Gaussian regression technique and Standard deviation feature with linear fit model. Data set was generated using sampling method 1 . . . . .	31
2.9	Comparison of volume density predictive models using Gaussian regression technique and Standard deviation feature with linear fit model. Data set was generated using sampling method 2 . . . . .	32
2.10	Comparison of fracture energy predictive models using SOM regression technique and Area feature with fracture energy estimation model (Baseline) . .	33
2.11	Feature importance weight using relief method for (a) Volume density, (b) Fracture energy density . . . . .	34
2.12	Performance of Volume Density predictive models using Gaussian Process regression technique and Standard Deviation feature on primary data set. The data set was generated using entire crack length (No sampling) . . . . .	35
2.13	Performance of Volume Density predictive models using Gaussian Process regression technique and Standard Deviation feature on data set 1 generated using sampling method 1 . . . . .	36
2.14	Performance of Volume Density predictive models using Gaussian Process regression technique and Standard Deviation feature on data set 1 generated using sampling method 2 . . . . .	37
2.15	Performance of Fracture Energy predictive models using SOM regression technique and Area feature on different data sets . . . . .	38
3.1	(a) Typical four point load test setup. (b) Specimen detail. (c) Shear diagram. (d) Moment diagram. . . . .	64
3.2	Image data segmentation for two different types of data. Type I: Typical raw image (a) segmented as shown in (b). Type II: pre-segmented crack pattern images in high resolution (c) or low resolution (d) form. . . . .	65
3.3	(a) Typical raw image, $i$ , with dimensions $w_i \times h_i$ . (b) The largest image in the data set with dimensions $w_{\max} \times h_{\max}$ . (c) Image standardization based on simple embedding in the largest image: the typical image, $i$ , is centered in an image frame with the size of $w_{\max} \times h_{\max}$ , with extra non-crack pixels added as needed. (d) Image standardization based on re-scaling to a common image size: image height and width are scaled to consistent pixel dimensions (i.e., $w_0 \times h_0$ ). . . . .	65

3.4	(a) Typical raw scaled segmented photo, with segmented cracks having a varying width. (b) The segmented image in (a) is skeletonized using a morphological process so that all cracks have uniform thickness. . . . .	65
3.5	Schematic showing axes used to compute moments of area for the crack images.	67
3.6	(a) Distance transform of a typical segmented image (see Figure 5); (b) Watershed transform of the distance transform; (c) The watershed regions used to calculate average distance between cracks; (d) Distance values in the distance transform image along watershed ridge lines. . . . .	69
3.7	(a) Deep Neural Network topology. (b) Typical Neuron in a hidden layer. . . . .	70
3.8	Training and test data sets definitions for performing 10-fold cross validation at the image level. (a) Total 558 images obtained from the different load stages used for performing 10-fold cross validation, gray and white boxes represent the testing fold and training folds, respectively, for one cross validation cycle. Each fold is randomly selected from 10 percent of the entire data set; (b) 10-fold cross validation: the randomly selected images in each fold are predicted using the rest of the images, cycling through all ten folds. . . . .	71
3.9	Training and test data set definitions for performing 10-fold cross validation at the specimen level. (a) Representation of all 84 specimens used for performing 10-fold cross validation. Each testing fold is derived from a random partitioning of the total data set into 10 subsets, with the selection for Testing Fold 1 illustrated above; (b) Each randomly selected testing fold is cycled through in the role of a test set with the remaining folds used as training data. . . . .	72
3.10	Performance of the failure ratio estimation models on Data Set 1 generated using image standardization methods 1 (embedding) and 2 (rescaling) using different feature sets. . . . .	73
3.11	Comparison of the failure ratio estimation models using Gaussian technique on Data Set 1 generated using scaling method 1 and 2. . . . .	74
3.12	Comparison of SVM models to estimate failure ratio using 10-fold cross validation at image and specimen level. . . . .	75
3.13	Comparison of the SVM regression models trained with different feature sets to estimate failure ratio using 10-fold cross validation at specimen level. . . . .	76
3.14	Comparison of the different regression models trained with the same feature set to estimate failure ratio using 10-fold cross validation at specimen level. . . . .	77
3.15	Performance of the models to estimate internal shear force directly using the feature set and indirectly through the estimated scaled shear force (figure b). . . . .	78

4.1	Conceptual image-based inspection framework. Existing laboratory tests are used to train an estimation model that correlates the surface image of the structure to known load and damage levels. The predictive model could then in principle be used in field applications such that captured photos could lead to statistical damage estimates. . . . .	82
4.2	Overview of number of specimens in each test category, including labeling scheme for data subset combinations. For the IDs, ‘B’ refers to beam specimens, ‘S’ refers to slabs, and ‘H’ identifies haunched beams, while ‘3’ and ‘4’ correspond to 3-point and 4-point loading, respectively, and ‘U’ means uniform loading. . . . .	86
4.3	Typical crack pattern images in the specimens having different shear reinforcement details. Figures a, b, c and d show crack pattern images at ultimate load stage for specimens YB2000-4, YB2000-6, YB2000-9 and YB2000-0 respectively [17]; and [83]. All structural details except the distance between shear reinforcements are similar in the beams. $d/s$ is 3.2, 1.4 and 0.7 for YB2000-4, YB2000-6 and YB2000-9 respectively. Specimen YB2000-0 does not have shear reinforcement. . . . .	88
4.4	Typical crack pattern images in the specimens having different shear span to depth ratio ( $a/h$ ). Figures a and b show crack pattern images corresponding to the ultimate load stage for specimens BS3-0.5-Vs and BS3-2-Vs respectively [34]. $a/h$ is 0.5 and 2 for BS3-0.5-Vs and BS3-2-Vs respectively. . . . .	89
4.5	Schematics of the machine-learning framework components and pipelines used to develop estimation and classification models. . . . .	102
4.6	ROC plot showing the shear reinforcement (Data Set 2)/no shear reinforcement (Data Set 1) categorization performance of the Multilayer Perceptron model trained using all crack pattern attributes and rounded $a/h$ . . . . .	103
4.7	SVM model failure ratio estimation accuracy for shear reinforced specimens (Data Set 2) comparing the effects of using different feature sets. . . . .	104
4.8	Performance of SVM models in estimating failure ratio for shear reinforced specimens (Data Set 2) with increasing geometric specimen information included in the feature sets. . . . .	105
4.9	Error metrics for SVM model failure ratio estimation for cases in Figure 4.8. . . . .	106
4.10	Comparison of estimation performance of SVM models using dominant features ( $I_P$ , $B$ , $N_c$ , $a/h$ , $h$ ) to estimate failure ratios for all unreinforced specimens (Data Set 1), all reinforced specimens (Data Set 2), and a subset of unreinforced specimens (Data Set 5). . . . .	107

4.11	Performance of SVM models to estimate failure ratio for combined shear reinforced and unreinforced specimens (Data Sets 3 and 4) using features $I_P$ , $B$ , $N_c$ , $a/h$ , and $h$ . . . . .	108
4.12	Performance of direct SVM/Gaussian models in estimating absolute shear force on different data sets using the feature set $I_P$ , $B$ , $N_c$ , $a/h$ , $h$ , and $b$ . . .	109
4.13	Performance of SVM/Gaussian models in estimating absolute moment on different data sets using the feature set $I_P$ , $B$ , $N_c$ , $a/h$ , $h$ , and $b$ . Classification and regression are combined in (d). . . . .	110
4.14	Estimation performance of Gaussian Process models using dominant features ( $I_P$ , $B$ , $N_c$ , $a/h$ , $h$ , $b$ ) to estimate normalized shear forces ( $V/(2\sqrt{f'_c} bd)$ and $V/(2\sqrt{f'_c} bd + A_s f_y d/s)$ ) and scaled moment ( $M/(bh^2)$ ) for reinforced specimens (Data Set 2). Note the axes scales vary between subfigures. . . . .	111
4.15	Performance of the Gaussian Process models for estimating internal shear for shear-reinforced specimens (Data Set 2) directly using the feature set $I_P$ , $B$ , $a/h$ , $h$ , $N_c$ , and $b$ , and indirectly through the estimated scaled moment, $M/(bh^2)$ , using the same feature set. . . . .	112
4.16	Performance of SVM models with and without pre-classification to directly estimate absolute moment on all specimens (Data Set 3) and all 3-point bending specimens (Data Set 4) using the feature set $I_P$ , $B$ , $N_c$ , $a/h$ , $h$ , and $b$ . . . . .	113
4.17	Error metrics for SVM moment estimations shown in Figure 4.16 for all specimens (Data Set 3) and all 3-point bending specimens (Data Set 4). . . . .	114
4.18	Performance of ultimate load calculation by ACI procedure, image-based internal shear force level estimation, and ultimate load stage estimations using combined classification and regression model for 3-point bending specimens (Data Set 4) using the feature set $I_P$ , $B$ , $N_c$ , $a/h$ , $h$ , and $b$ . The ACI plot reflects the removal of an outlier unreinforced slab specimen with dimensions outside the applicability of the ACI formulas. This specimen was not removed from the image-based data set. . . . .	115
5.1	(a) Typical cracked bridge girder. (b) Developed structural cracks in the girder	118
5.2	Typical shear panel test setup. . . . .	120
5.3	(a) Typical crack pattern; (b) Distance transform of a typical segmented image; (c) Watershed transform of the distance transform; (d) The watershed regions used to calculate average distance between cracks; (e) Distance values in the distance transform image along watershed ridge lines. . . . .	124

5.4	Angle calculation of crack object. (a) Typical crack pattern image. $\theta_p$ is the theoretical calculated principal angle corresponding to the plane of maximum tension, $\sigma_1$ . $\theta_c$ is the averaged crack angle based on image analysis. (b) Individual crack segment orientation, $\theta_i$ , and length, $l_i$ calculated using MATLAB regionprops function; (c) Crack objects outside the range of one standard deviation away from the mean orientation were considered outliers, and removed from overall average angle calculation; (d) final weighted average calculation.	125
5.5	Comparison between $\theta_p$ calculated based on the nominal state principal orientation, and overall orientation of cracks calculated using image processing for representative fiber reinforced load stage images. Loading increases left to right for each row of images, which results in increased induced critical stress, and thus rotating principal orientations. . . . .	137
5.6	Comparison between $\theta_p$ calculated based on the nominal state principal orientation, and overall orientation of cracks calculated using image processing for representative conventionally reinforced concrete load stage images. Loading increases left to right for each row of images, which results in increased induced critical stress, and thus rotating principal orientations. . . . .	138
5.7	Resampling approaches used to generate more image blocks . . . . .	139
5.8	(a) Typical concrete panel subjected to monotonic loading (obtained from Luo 2014). $\sigma_{tot}$ and $\tau_{xy}$ are the total normal and shear stress acting at the side of the panel. (b) Segmented crack patterns. $\sigma_{xx}$ and $\tau_{xy}$ are the normal and shear stress in the concrete. $\sigma_{xx}$ is calculated based on the compatibility condition between steel and concrete. The location and angle of the image blocks were randomly selected with orientation ( $0^\circ \leq \theta_s \leq 180^\circ$ ) to generate new subsampled image blocks; (c) Mohr's circle representing the stress state of the panel and randomly selected image blocks. The randomly selected image blocks have the same maximum shear stress ( $\tau_{max}$ ) and principal stresses ( $\sigma_1$ and $\sigma_2$ ) (d) Pipeline used to estimate the failure ratio and stress state for a typical resampled image block, marked with a star. . . . .	140
5.9	Effects of the sub-image block size on different cracks attributes. Sampling technique III was used to sample ten sub-image blocks having different orientations from a typical fiber crack pattern image . . . . .	141
5.10	Effects of the image block size on Normalized Mean Absolute Error (NMAE) for different cracks attributes. Sampling technique III was used to sample ten sub-image blocks having different orientations from all fiber specimens . . . .	142
5.11	Image-based average fracture plane orientation, $\theta_c$ , correlation with theoretical principal orientation, $\theta_p$ , based on nominal applied stress state. . . . .	143

5.12	Index of Agreement convergence performance for fracture ratio estimation as a function of reference sampling block size for SVM models using dominant features. Plotted curves correspond to different training and test sampling strategies as indicated, with size ( $a_s$ ) and/or orientation ( $0^\circ \leq \theta_s \leq 90^\circ$ ). $10\times$ resampling used per load stage for all cases other than the full ( $a_s = 100\%$ ) panel cases. . . . .	144
5.13	Performance of SVM models to estimate fracture ratio trained by dominant features on different training and test data sets . . . . .	145
5.14	Comparison of SVM models trained by dominant feature set to estimate fracture ratio on entire and resampled data sets . . . . .	145
5.15	Performance of Gaussian process models to estimate $\tau_{\max}/\sqrt{f_t}$ , $\tau/\sqrt{f_t}$ and $\sigma_2/\sqrt{f_t}$ for monotonic data set using best-fit feature sets. All cases uses $L_c$ , and ave crack distance and homogeneity. In addition, cases (a) and (c) include aspect ratio, while case (b) includes $I_p$ and $N_c$ . . . . .	146
5.16	Performance of Gaussian process and linear models to estimate $\sigma_{xx}/f_t$ and $\tau_{xy}/f_t$ trained with dominant feature set (homogeneity, $L_c$ , $I_p$ , $N_c$ , total major axis, average distance between cracks) on monotonic SFRC data set . . . . .	147
5.17	Performance of SVM models to estimate $\sigma_{x'x'}/f_t$ and $\tau_{x'y'}/f_t$ trained with all crack features including overall crack orientation, $\theta_c$ , and sampling block size, $a_s$ , on sub-sampled RC and SFRC data sets. Case I sampling strategy, with size of ( $30\% \leq a_s \leq 100\%$ ) and orientation range of ( $0^\circ \leq \theta_s \leq 180^\circ$ ) was used. . . . .	148

## ACKNOWLEDGMENTS

First and foremost I would like to express my sincere gratitude to my advisor, Dr. Miller, for his endless patience, motivation, and immense knowledge. I am deeply indebted to him since without his continuous support, guidance and constant feedback this PhD would not have been achievable. He spent endless hours editing my research papers and giving me insightful suggestions that I will always remember. I could not have imagined having a better advisor and mentor during my academic journey. There are not enough words to thank him for all the lessons I have learned in the past 4 years.

I would also like to express my special appreciation and thanks to Dr. Kutz, he has been a tremendous mentor for me. I would like to thank him for encouraging my research and for helping me to grow as a researcher. His advice on both research as well as on my career have been invaluable. I have learned so much from him during this academic journey. Most importantly, he is my role model for being an excellent researcher. It has been a rare pleasure to work with someone so inspiring and supportive. Working with him was indeed a turning point in my career that I will never forget.

Besides my advisor, I would like to thank the rest of my thesis committee: Dr. Eberhard and Dr. Calvi, for their insightful comments, and invaluable feedback. Their extensive experience and expertise in areas of structural mechanics, and non destructive testing helped me a lot to widen my research from various perspectives. Without their exceptional guidance, this PhD would not complete.

A special thanks to my family. Words can not really express how grateful I am to my mother, father, brother and sister for all of the sacrifices that they have made on my behalf. Their prayer for me was what sustained me thus far.

Chapter 1  
**INTRODUCTION**



## 1.1 Overview and Motivation

Infrastructure inspection plays a key role in public safety in regards to both long-term damage accumulation and post extreme-event scenarios. Given its importance, there has been increasing interest in automating damage inspection tasks using computer vision and data driven techniques in the last ten years. To date such automated inspection approaches have primarily focused on identifying and categorizing surface damage, but this at best allows for relatively coarse-grained, qualitative damage classification. The purpose of the research presented here has been to investigate how machine learning and image processing techniques can be used to enable *quantitative* damage/deterioration state assessments based on visual observations. This required the assembly of appropriate image and load level data from available sources, which then enabled the application of computer vision technologies and the subsequent evaluation of its effectiveness.

As a representative example, there are approximately 580,000 major bridges in the US that must be inspected on a two-year cycle. The inspection process is primarily visual and detailed photographs are captured from all of the vulnerable components of the structure (see, e.g., Figure 1.1). If it is assumed these 580,000 bridges have been inspected consistently for 40 years, and approximately 20 photos have been captured for each bridge every other year during this period, a data set with approximately 232 million photos of the structural components could be available for automated analysis. This kind of data set offers many opportunities for machine learning and partially automated inspection support.

On a more focused scale, cracking in bridge decks and girders are common damage mechanisms in RC bridges during their lifetime. A photo of a typical cracked bridge outrigger beam captured in Seattle is shown in the Figure 1.2. Such cracks in girders can raise concerns about the safety and longevity of the affected structure. Existing damage inspection approaches cannot provide reliable quantified information about the level of the damage and safety associated with such observable damage. Mechanisms for finding correlations between such photos and the damage level of the components have the potential to enable significant



Figure 1.1: Typical damage assessment by the inspection team

improvements in inspection utility.

## **1.2 Goals of Research**

This research project is focused on rethinking how computer vision and data driven techniques can be used to improve damage assessment based on surface damage and crack patterns in concrete structures. The specific goals are as follows:

- an exploration into developing predictive models for relating external/surface observations to useful internal information.
- presenting computer vision models to estimate quantitative damage and load levels based on crack patterns for various reinforced concrete components.
- collecting more image data sets for extending the framework for inspecting different structural components such as bridge girders.



(a)



(b)

Figure 1.2: (a) Typical cracked bridge girder. The photo captured at Exit 2C, I5 N, Seattle. (b) Developed structural cracks in the girder

- investigating the feasibility of linking the predictive models developed for different components such as beams, slabs and concrete panels together to generate a comprehensive damage inspection model that could be used for other RC components.

### **1.3 Organization of the dissertation**

All research work to complete the dissertation is explained in Chapters 2-5, which provide full technical details and results of the presented computer vision and machine learning framework to quantify damage and load states in structural components such as beams, slabs, and concrete panels. Exploratory work for developing predictive models for relating outside observations to internal damage states is provided in Chapter 2. The application of

the proposed method in this chapter could be used in diagnostic problems in which extracting internal information requires significantly higher cost and/or special facilities compared to making external observations. Chapter 2 is designed to provide the reader with a better understanding of how data-driven approaches could be used as an alternative way to leverage damage assessment tasks.

Chapters 3-5 explain the presented framework for damage classification and quantification for beams, slabs, and reinforced concrete panels. The majority of these chapters are devoted to explaining different applications and the limitations of the presented framework for damage inspection tasks. The following subsections outline the presented computer vision framework used to develop estimation/prediction models to be calibrated and tested with the collected experimental data for different cases. The data set used for training predictive models for each case is explained in this chapter. However, the primary results, full literature review, and technical details for all cases are presented in Chapters 2-5 themselves.

### *1.3.1 Chapter 2 (An exploratory model for relating surface observation to internal damage)*

Most field inspections are based on external observations, but key damage phenomena often occur under the surface of the structural components. The goal of this exploratory work is to show the feasibility of building models relating surface damage to internal damage state. For this purpose, full internal and external 3D and 2D image data sets captured from cracked rebar de-bond specimens collected from previous studies were used for training the models. The primary objective is to estimate internal crack volume as a simple proxy for internal damage based on external crack image data. This work also investigates the effects of using different machine learning techniques along with different feature sets extracted from crack images to estimate the internal crack volumes. The results indicate that trained models can estimate internal crack volume based solely on external observations, although for this particular data set the performance of simple linear models was found to be comparable to more advanced machine learning models.

### 1.3.2 Chapter 3 (*Computer Vision-Based Damage Inspection Framework for Shear Critical Beams*)

Chapter 3 focuses on developing models for estimating quantified load and damage levels for beams without shear reinforcement based solely on surface crack patterns and geometries of the components. The predictive models were trained and tested using image data sets obtained from ten different earlier published studies, which provided 563 crack pattern images captured from 86 specimens, including RC beams and slabs without shear reinforcement. These existing experimental tests covering a wide range of geometries, loadings, concrete strengths, and reinforcement details were conducted for studying the effects of different parameters such as aggregate size and concrete strength on ultimate shear strength of the RC beams. The crack pattern images used for training the models were captured at various load and damage levels, and so the resulting framework is capable of translating simple photographic data into estimated levels of damage calibrated relative to the component's actual capacity. Various textural and geometrical attributes of surface crack patterns have been defined and used for extracting meaningful features from the images to train the predictive models. An image transformation step combined with a morphological process is also used for normalizing and registering the crack pattern images in a standard format to take account of the different specimen and image dimensions (i.e., image resolutions). Results show that the presented framework is capable of estimating internal load and damage levels based only on surface crack patterns and basic geometries of the specimens.

### 1.3.3 Chapter 4 (*Computer Vision-Based Damage Inspection Framework for RC Beams and Slabs*)

Chapter 4 presents an investigation of the performance of the inspection framework presented in the previous part for an extended data set that includes beams and slabs with shear reinforcement. Unlike the previous study in which collected specimens were designed to be shear critical, the focus of this study is on more conventionally reinforced specimens designed

to have flexural-shear or flexural failures, and which are more representative of the beams and slabs likely to be found in existing buildings. To this end, the new data set was collected from seven earlier published studies, which provided 304 crack pattern images captured from 43 RC beams and slabs (with web reinforcement) across a range of load and damage levels. It was found that knowledge of structural properties such as shear reinforcement ratio and shear index beyond basic external geometries are not necessary for damage estimation. However, adding such information can improve the performance of the models. Results indicate that the system is capable of quantifying internal load and damage levels using crack pattern images both across and within different classes of reinforcement. The main distinction in this work relative to the two previous studies include the larger data set used (both in terms of the amount of data and breadth of its sources), the much more challenging range of geometries, structural details and different behaviors (i.e., shear, flexural-shear and flexural behaviors). In addition to damage and load-level quantification using the presented framework, Chapter 4 also investigates the feasibility of developing classification models to categorize the beam type (i.e., beam with and without shear reinforcement) based solely on external damage observations in which the web reinforcement is not visible. In general terms, the presented framework has the potential to be used for quantified damage estimation and beam type classification for RC beams and slabs across a wide range of geometries, loadings, concrete strength, longitudinal and shear reinforcement ratio.

#### *1.3.4 Chapter 5 (Computer Vision-Based Stress State Quantification for RC Panels)*

The models in the previous sections can only be used for damage assessment of reinforced concrete beams and slabs undergoing similar loading to that used in the calibration testing (i.e., other components such as columns and shear walls cannot be handled by models trained using beams and slabs). The goal of Chapter 5 is to investigate a more general approach to developing crack-based models for damage assessment of a variety of RC and SRFC components. In particular, Chapter 5 investigates the feasibility of relating local crack patterns to quantified stress states. For this purpose, a similar computer vision framework

was used to build predictive models generated and tested using image data sets obtained from five different earlier published studies at University of Toronto, which provided 189 crack pattern images captured from from 33 concrete and SRFC shear panel specimens. These specimens tested with the objectives of obtaining data for building constitutive relationships for monotonic and cyclic behavior of fiber and SRFC materials, and developing crack models using non-linear regression techniques for the calculation of crack widths and crack spacing based on the loading. The objective of the current study is to rethink how such correlations between loading conditions and crack patterns can be used for generating a reverse solution. In particular, the model is intended to use these correlations to estimate the stress states based on the crack pattern images. The results show that the stress state estimation models can work well across a wide range loadings (i.e., reverse cyclic and monotonic) and materials (concrete and SRFC). It was also found that trained models can estimate principal stresses and maximum shear stress for different subsampled stress blocks, with accuracy varying according to sub-image sampling strategies. This direct correlation between crack patterns and stress states could be used in broader context for developing more advance predictive models for damage assessment of different concrete and SRFC members.

### *1.3.5 Chapter 6 (Conclusion)*

Chapter 6 summarizes the key conclusions regarding the computer vision framework presented in the previous chapters, and outlines possible directions for future studies in automated vision-based inspection field. This chapter also discuss possibilities, uncertainties, and limitations of such data-driven damage inspection approaches.

## Chapter 2

# AN EXPLORATORY MODEL FOR RELATING SURFACE OBSERVATION TO INTERNAL DAMAGE



## **2.1 Abstract**

This chapter investigates the feasibility of developing predictive models relating external observation of damaged infrastructure components to internal damage states and structural behavior via computed tomography (CT) scanning and machine learning techniques. The primary focus of this particular study was to investigate the accuracy of various machine learning techniques in estimating internal crack volumes, internal fracture energy, strength loss and bar slippage in reinforced concrete members based on external crack image data. Crack volume, fracture energy and strength loss are conceptually straightforward to quantify and were chosen to serve as simple metrics of structural damage states. Bar slippage is also an important quantifiable structural behavior that can significantly influence structural performance of RC members. Full internal and external 3D and 2D image data sets captured from cracked rebar debond specimens were used as training data, and test cases were selected in which only surface imaging was used as model input. Various textural and geometric attributes were defined to train the predictive models. The results show that predictive models based on combined internal and external image training data outperform predictions based solely on external observations. It was also found that the use of feature selection techniques along with multiple nonlinear regression approaches can significantly improve estimation accuracy. These results indicate that systematic development of X-ray based internal damage image data sets combined with appropriate machine learning techniques could be used in broader contexts to enhance fundamentally the value of both routine and post-extreme-event field infrastructure inspections for which surface observations alone are available.

## **2.2 Introduction and Motivation**

Infrastructure inspection plays a key role in public safety in regards to both long-term damage accumulation and post extreme event scenarios. Most field inspections are based on visual/external observation, but key damage phenomena often occur under the surface of the

structural components. Internal inspection capability can be expensive and time-consuming to implement and most advanced specialized equipments such as X-ray based CT scanning ultimately is limited to lab use. Therefore, the goal of this chapter is to investigate the feasibility of leveraging lab-based CT scanning to enhance damage/deterioration state assessments based on visual observations. The approach presented here could enable field inspectors to reliably estimate sub-surface structural damage simply by surface imaging. The broader concept of the proposed method could be used in any diagnostic problems in which extracting internal information requires significantly higher cost and/or special facilities compared to making external observations. Various damage metrics have been suggested in earlier studies such as fracture energy, fracture strain and stress. In the current case study, the crack volume was considered as a principle internal observation representing structural damage. Other observations including fracture energy, strength loss and bar slippage were also considered to explore damage state and structural behavior of the specimens.

To that end, predictive models based on a combination of internal and external image training data were used to carry out volume, fracture energy, slip and strength loss predictions on new external image data. For this purpose, predictive models using combinations of Image Processing and Wavelet toolboxes and WEKA machine learning package [56] in MATLAB were developed , to predict crack volumes based on surface imaging. The models can perform image segmentation, crack extraction, feature extraction, volume and surface area calculations and non linear regression analysis in a fully automated fashion.

### ***2.3 Preliminary Work and X-Ray Image Data Set***

As mentioned before, X-ray images of bond test specimens have been collected from previous research projects [72]. The high resolution images in previous studies were captured mainly for the purpose of studying bond failure mechanisms. A typical single-ended pull-out tests setup is illustrated in Figure 3.1. Total bonded area was imaged to study behavior of the bond failure. Table 2.1 provides key structural parameters and imaging details for each image data set used in this study. Imaged specimens can be divided into two different categories.

In the first one, five specimens were imaged at four load steps during the pull-out test. A typical load-displacement response for a specimen is illustrated in Figure 2.2. The specimen was subjected to a splitting-type failure as shown in Figure 2.4 (a) and, the black dots in Figure 2.2 (step1-step4) indicate points at which CT imaging was carried out. The X-ray image data sets were collected from the last two imaging steps (step 3 and 4) in which cracks were fully developed. Therefore, these five specimens provide ten x-ray image sets. In the second case, three specimens were imaged following the testing that creates three more image sets. Overall, thirteen sets of X-ray images were collected from 8 specimens as shown in Table 2.1. In order to increase the number of data sets; sampling techniques were employed to generate more data sets. As shown in Figure 2.3, each specimen might have a different numbers of crack "shields" (i.e., vertical 3D crack volumes bounded by crack surfaces). Two distinct sampling methods were used to construct training and test data sets from 2D surface images of crack shields or longitudinal cracks. In the first one, the training data set was randomly generated from longitudinal cracks. Different numbers of samples and lengths of generated cracks were considered to evaluate the consistency of the method. It is noted that there is no overlap between the test data set and the generated training data since the latter is generated from other longitudinal cracks. The sampling process is illustrated in Figure 2.3. In the second case, the longitudinal crack was divided into a number of sub-cracks such that all data sets do not overlap. Figure 2.3 shows the method 2 sampling approach. As noted, the training and test. data sets do not overlap. Also, test sub-cracks would be selected randomly between one of the 4 sub-cracks as shown in Figure 2.3.

#### **2.4 Image Analysis and Machine Learning Methodology:**

Image analysis and machine learning were performed using a series of predictive models. Four major steps are employed to perform image analysis: (i) 2D segmentation, (ii) 3D segmentation volume and surface area calculations, (iii) feature extraction and (IV) machine learning based regression techniques. The initial and major process in any computer vision system is the segmentation of the image. Image segmentation is the process in which raw

images are transformed to binary images (pixel intensity of 0 or 1). It should be noted that all of the raw images were scaled to their actual specimen size before segmentation process so that following analyses results can be properly comparable. The 2D image segmentation process was carried out using a thresholding technique isolating objects by converting grayscale images into binary images. The constant range of [55,95] was chosen to set gray scale level. In order to de-noise the image and remove non-crack objects from the segmented image, hybrid morphological reconstruction technique was employed to improve segmentation process by Lu et. al. 2012 [51]. The proposed technique performs an automated robust nuclei segmentation through reducing intensity variation withing nuclei regions in histopathological images [51]. In the current study, the hybrid morphological technique was utilized to make the segmented image homogeneous by removing noise and aggregates. The left holes from the morphological de-nosing step were located and filled with white pixels using image analysis toolbox in MATLAB. It should be noted that, filling the holes after primary segmentation without using the hybrid morphological de-noising step could over-simplify the crack structure and result in losing some parts of the crack. The segmentation process presented here could accurately perform image segmentation in most of the images. The segmentation process was illustrated in Figure 2.5 (a) to (d). An automated process was developed to locate and extract cracks from the segmented images as shown in Figure 2.4 (d). The algorithm is able to identify rebar's position and its diameter in order to extract the rebar from the segmented image. The algorithm could also track cracks in 2D images set and extract them to reconstruct 3D crack planes as illustrated in Figure 2.4 (d) and (e). The height of 3D segmented images was also scaled to their specimens' size to ensure consistency between specimens.

The volume and internal surface area of crack planes were defined based on contained number of voxels and surface area of crack shields. The volume and surface area also needs to be scaled by specimens' dimension to obviate dimensions difference problem in the specimens. Figure 2.4 (f) and (g) indicate crack volume density and surface area density distributions in one of the specimens. These distributions will be used later to define real volume density in

re-sampled cracks. As mentioned before, internal damage and structural behavior perditions are based on external observations. Figure 2.4 (h) indicates a longitudinal surface crack corresponding to crack plane 1. A feature set needs to be defined based on features extracted from binary surface cracks to train predictive models. The feature set matrix is indicated schematically in Figure 2.4 (i). Each row of the matrix corresponds to one crack plane and columns of the matrix would be features set. In the feature extraction step, two types of attributes were considered: (i) feature set obtained by wavelet transform and SVD [12] (ii) Geometrical and Textural features. As shown in Figure 2.4 (j), a quantitative class set matrix including crack volume density, fracture energy, slip and strength loss was also defined to be used for comparison between predicted and true values. Crack volume density of crack planes was straightforward to quantify and was defined by volume of the crack plane scald by volume of the specimen.

Fracture energy,  $G_f$ , is the amount of energy required to open unite area of crack [7]. In the current study, total cracking energy in the the specimen was considered as a fracture energy and distributed between crack planes based on surface area of the crack planes in the specimen. In fact, the crack planes having a larger surface area, was subjected to a bigger portion of fracture energy. The total energy in the pull-out test is equal to the work done by external load which can be obtained by integrating under load displacement curve. Figure 2.2 shows a typical load-displacement response for a specimen. The total pull-out test energy can also be represented in terms of change in strain energy of concrete and rebar, and concrete fracture energy as illustrated in equation 2.1. Similar procedures were performed by Marthin 2006 [55] . Since part of the specimen was imaged under CT scan machine as shown in Figure 3.1, cracking energy should be adjusted based on unseen and total heights of the specimen. It was assumed cracking pattern in unseen part of specimen is similar to imaged part.

$$\Delta W = \Delta U_{Strain} + \Delta U_{Cracking,Imagedheight} \frac{Specimen'sheight}{Imagedheight} \quad (2.1)$$

where  $\Delta U_{Strain}$  is change in elastic strain energy of concrete and steel materials. The

strain energy can be calculated using compatibility and equilibrium conditions between bar and concrete materials [55]. Ideally, the total pull-out work subtracted by elastic strain energy of steel and concrete materials should result in concrete fracture energy as shown in equation 2.2. However, the left side of the equation is 100 percent bigger than the right side. Other un-known friction mechanisms between bar and concrete, energy absorbed by spirals or noise can explain the energy inequality. Results in Tyler's study also showed 100 percent noise in fracture energy calculations.

$$\Delta W - \Delta U_{Strain} \stackrel{?}{=} G_f \Delta A_{Imagedheight} \frac{Specimen's\ height}{Imagedheight} \quad (2.2)$$

where  $\Delta A$  is the change in crack surface area in imaged height of the specimen.  $G_f$  is the concrete fracture energy density obtained using Bazant's model [7]. This chapter considers the total pull-out work subtracted by elastic strain energy of steel and concrete materials as crack fracture energy. The total crack fracture energy was divided between crack planes based on their internal surface areas as illustrated by equation 2.3. This assumption was made to investigate possible corroboration between surface observations and fracture energy in crack planes.

$$U_i = \Delta U_{Cracking,Imagedheight} \frac{A_i}{\sum A_i} \quad (2.3)$$

where  $U_i$  is the crack fracture energy corresponds to crack i.  $A_i$  is internal surface area of crack i.

Slippage of the rebar or slip, is deformation in the rebar due to separation of concrete and rebar in RC components which can significantly influence structural behavior of the members. In the current study, slip was identified by finding the difference between the total deformation of the specimen and the elastic elongation of the bar. The total slip in the specimen was divided between crack planes based on ratio of surface area of cracks in each specimen.

Strength loss ratio for last two load steps was defined as ratio of strength loss I and II

(Figure 2.2) over ultimate strength of the specimen. Similar to slip, strength loss ratio was also divided between crack planes based on ratio of surface area of cracks in each specimen.

#### 2.4.1 (I) feature set obtained by common image transforms and SVD

The common transform methods such as Hough, Radon, Discrete cosine, Discrete sin and Wavelet transforms convert an image from spatial domain to other dimensions in order to extract or highlight some special features or patterns in the image. Such transforms have been widely used in the literature for pattern recognition, image classification and filtering tasks. The wavelet transform approach provides an effective method to represent multi scale information. In particular, 2D cracks can be projected onto multiple constructed bases. High degrees of freedom (bases) can be transformed to lower dimensional bases using singular value decomposition method. In fact, Singular Value Decomposition (SVD) is a useful technique to find the dominant modes of each data set. The SVD pulls out the diagonal matrix in decreasing order along with two unitary matrices  $U$  and  $V$ . This method is guaranteed to exist for any matrix. In fact, the SVD makes it possible for every matrix to be diagonal if the proper bases for the domain and range are used. Matrix  $A$  can be represented with its bases as follows in equation 2.4 [12] and [46].

$$A = \sum_{j=1}^r \sigma_j u_j v_j^* \quad (2.4)$$

The objective is to define a procedure by which we can define a wavelet based feature to train supervised algorithm for classification or regression analyses. In order to perform regression analysis, test data set must be transformed to  $V$  bases obtained from training data set decomposition. The algorithm can be summarized the following key steps [12] and [46]:

- Step 1: Apply images decomposition on segmented training data set
- Step 2: Find the principal components associated with crack data set
- Step 3: Find unitary  $U$  and  $V$  matrices

- Step 4: Project test data set on V bases of the training data set

#### 2.4.2 (II) Geometrical and Textural features

The second approach involves the conversion of crack image data sets into quantitative attributes. The quantitative features which are mostly geometric attributes corresponding to each crack were extracted from binary images. The feature extraction was performed in MATLAB using image processing toolbox. These extracted attributes are as follows: compactness, aspect ratio, thresh-out, entropy, contrast, correlation, energy, homegenity, local variance, area, perimeter, euler number, standard deviation, central moment, moment of inertia and polar moment of inertia. Some of these features are explained here.

compactness: is perimeter divided by area of the crack.

aspect ratio: is ratio of length over width of the crack.

Thresh-out: is threshold value of gray-scale crack image returned by "edge" function in Image processing toolbox.

entropy: is a scalar value representing entropy of gray-scale crack images.

contrast: is a scalar value representing intensity contrast between pixels over whole image.

correlation: is a value identifying correlation between a pixel and its neighbor over the entire image.

Energy: is sum of squared elements in gray level co occurrence matrix

Homegenity: is scalar value between 0 and 1. representing distribution of elements in gray level co occurrence matrix.

local variance: is average of local standard deviation of the 3-by-3 neighborhood around the each pixel in the image.

euler number: is a scalar value which is calculated by total number of objects in the image subtracted by total number of holes in those objects.



Standard deviation: is the standard deviation of one-dimensional reshaped crack images. This scalar feature represents area distribution through length of the crack.

Central moment: is a weighted moment of the image pixels through each crack image as shown in equation 2.5.

$$M_{pq} = \sum_{i=0}^m \sum_{i=0}^n (x - \bar{x})^p (y - \bar{y})^q \quad (2.5)$$

where  $m$  and  $n$  are the length of row and column. The  $\bar{x}$  and  $\bar{y}$  are the components of the centroid. In this study,  $p$  and  $q$  parameters are considered to be equal to 2.

moment of inertia: is the moment of inertia of entire crack about it's centroid.

Polar moment of inertia: is summation of moment of inertia in both x and y directions about crack's centroid.

New crack volumes were predicted using machine learning based regression techniques. Here, the WEKA toolbox which is an open source machine learning package has been employed in MATLAB [38]. This toolbox provides a number of supervised machine learning techniques such as SVM, J48 decision tree and Random forest algorithms. The aforementioned algorithms along with the regression techniques were used to predict Volume. Regression analysis can transform large amounts of data into meaningful numerical patterns such that missed or new data set can be estimated.

## 2.5 Finding Dominant features

The Relief algorithm was employed for features selection. In this algorithm,  $K$  nearest neighbors method was used to computes ranks and weights of attributes for primary input features matrix. This method can pull out the dominant features for both numerical and non-numerical response vectors. In other words, both classification and regression problems can be addressed using this method. It was found that area and standard deviation can play a major role in both volume and fracture energy predictions for the data sets generated using both sampling methods. In slip prediction, predicted fracture energy was used as new

feature set. The Relied algorithm identifies predicted fracture energy and aspect ratio as the most effective features for slip and strength loss predilections respectively.

Feature selection using this technique for volume density, fracture energy, slip and strength loss predictions are compared as shown in Figure 2.11. Negative importance weighs indicate an adverse effect on volume prediction. Effects of excluding these attributes from the feature set will be discussed later.

## **2.6 Crack volume estimation based on surface area of cracks**

In order to demonstrate the viability of the proposed predictive algorithm, the predicted crack volume using relevant algorithms was compared with the estimated crack volumes determined from external data alone which can be considered as a baseline. For the baseline estimation, It was assumed each shield of the crack has a constant rectangular cross section along the length of the crack. Therefore, the volume of each longitudinal crack can be calculated by multiplying the surface crack area by a constant value which can be expressed by each specimen's radius. In other words, a linear relation between the crack volume and the surface area was considered. The estimated volumes based on the surface crack area is shown via a correlation plot in Figure 2.12 (a). Each data point represents the estimated crack volume for one crack image in the test data set. Histogram frequency bar on the true crack volume axis indicates that the cracks having low volumes have relatively higher frequency compared to others.

## **2.7 Pull-out Work Estimation Based on Bazant's Model**

Similar to crack volume estimation, a baseline case for fracture energy estimation was defined to compare performance of predictive models. The fracture energy density,  $G_f$ , is the amount of energy required to make a unite area of crack. If  $\Delta A$  and  $G_f$  are the change in crack area and fracture energy density in the concrete respectively, The total cracking energy can be defined by multiplication of  $\Delta A$  and  $G_f$ . Bazant (2002) [7] proposed an expression to predict  $G_f$  based on concrete properties (equation 2.6). The value of fracture energy density,  $G_f$ ,

was found 0.46 using Bazant's model for all the specimens as they have the same mechanical properties. It's noted that two specimens (Martin 5A-6 and YAXLON sample B in table 2.1) were excluded in fracture energy prediction study as load-displacement data for those specimens were not available. Similar to internal volume estimation, each shield of the crack can be assumed to have a constant rectangular cross section along the length of the crack. Therefore, the internal surface area of crack planes can be estimated by multiplication of specimen's radius and crack length on the surface of the specimen.

$$G_f = 2.5\alpha_0 \left( \frac{f'_c}{0.051} \right)^{0.46} \left( 1 + \frac{d_a}{11.27} \right)^{0.22} \left( \frac{w}{c} \right)^{-0.30} N/m \quad (2.6)$$

where  $f'_c$  is the compressive strength of the concrete in Mpa,  $\alpha_0$  is an aggregate shape factor which is 1.44 for regular aggregate,  $d_a$  is the maximum aggregate size in mm, and  $w/c$  is the water-cement ratio of the concrete.

## 2.8 Model analysis

As it was mentioned earlier, one hundred and sixty 2D sub-sampled surface cracks were generated from forty image data sets. In order to assess performance of the predictive models, the models analysed using 20-fold cross-validation. Predictive models based on sub-sampled training data sets outperform volume predictions for each sub-sample surface crack as an input test data. Each data output represents the average of predicted crack volumes obtained from 20-fold cross-validation for each surface crack image in the test data set. It's noted that, training and test data sets were randomly generated in each cross validation while they do not overlap as shown in Figure 2.3. As it was explained earlier, two different feature sets (I) wavelet transform and SVD and (II) Geometrical features were utilized to train the predictive models. Different combinations of non-linear machine learning-based regression techniques using aforementioned features were applied on sub-sampling data sets generated from method 1 and 2 to outperform volume prediction.

## 2.9 Performance of predictive algorithms

Various goodness-of-fit statistics were considered to compare the predictive techniques. The regression accuracy metrics describe how well the predicted crack volumes fit true data. The regression metrics used include Mean-Absolute Error (MAE), Index of Agreement (IA), Correlation Coefficient (R) and Explained variance Score [56] [50].

Mean-Absolute Error (MAE): is defined as the average of absolute difference between predicted and true variables as shown in equation 2.7.

$$MAE = \frac{1}{n_{samples}} \sum_{i=0}^{n-1} |y_j - y'_j| \quad (2.7)$$

where,  $y'_j$  and  $y_j$  are the estimated output and the corresponding true value respectively.

Index of Agreement(IA): IA is a non-dimensional and bounded measure. IA values closer to 1 show better agreement between predicted and actual data sets.

Correlation Coefficient (R): is a measure of the strength and direction of a linear relationship between two variables. A correlation of 1 indicates a perfect one-to-one linear relationship.

Explained variance Score: is defined based on the ratio of residual variance to variance of true data set as illustrated in equation 3.7. Similar to IA, Scores closer to 1 indicate a better correlation between predicted and true values.

$$Explained\ Variance = 1 - \frac{Var(y_j - y'_j)}{Var(y)} \quad (2.8)$$

where,  $y'_j$  and  $y_j$  are the estimated output and the corresponding true value respectively.

In regression analysis, the low R-squared value could show a robust trend while there might be a noisy scatter around the regression line. The Correlation coefficient (R) can also be influenced by extreme values in non-normally distributed data ([56] [50]). In the current study, the data set includes some outlier values as shown in Figure 2.12 (a) that necessitates using IA than R-squared to compare the trend between models. The performances of the predictive algorithms trained by different feature sets are illustrated visually by scatter correlation plots in Figure 2.12 and Figure 2.15. A 45 degree line in the plots indicates an ideal line in which predicted crack volumes would match true volumes.

Figure 3.11 illustrates the prediction accuracy for predictive models applied on data set 1 compared to the estimated crack volume considered as a baseline case. Figure 2.10 illustrates the prediction accuracy for predictive models applied on data set 2 compared to the estimated crack volume considered as a baseline case.

## **2.10 Conclusion and Results**

As it was mentioned before, predictive models were built using various machine learning based regression technique including Gaussian process, J48, SVMs, SOM, Multilayer perceptron, Decision trees, Logistic regressions, Nearest neighbors, linear regression and Random forest along with distinctive features such as image transform based features, geometric and textural features. It was found that image transform based features such as Wavelet, Hough and Radon along with SVD/PCA techniques cannot perform predictions well compared to nonlinear regression models using geometric features. As mentioned earlier, such image transform techniques are more suitable to be used for pattern recognition and classification problems than regression problems. In the current study, the longitudinal cracks do not seem to have different patterns such that they can be differentiated or classified. Singular value decomposition analysis also shows that no strong dominant modes can be identified from the image transforms features.

Results in Figures 2.13 and 2.14 show that volume predictive models using Gaussian process regression techniques and standard deviation feature outperform predictions better

than baseline model for different data sets generated from different sampling approaches. It was found that, Gaussian process and standard deviation feature gives a better performance than other techniques. The volume predictive models deliver same performance for different data sets as indicated in Figure 3.11 that reveals a statistically robust relationship between standard deviation of the surface cracks and volume density.

As explained earlier, fracture energy based on the Bazant's model under estimate cracking energy in the pull-out test due to some other possible failure mechanisms such as friction between bar and concrete and energy absorption in the confinement. SOM regression technique with area feature predicts fracture energy better than the baseline case for different data sets except three outlier points corresponding to one of the specimens (Figure 2.15 ). It was seen that the corresponding specimen has a fairly larger resilience after the ultimate load compared to the other specimens. Other phenomena such as cracking in the unseen part of the specimen (i.e. was not imaged in the CT scan machine) might be the reason for the larger pull-out work in the global behavior of the specimen. Performance of the fracture energy predictive model for different data sets is shown in Figure 2.10. Similar performance of the model indicates a strong relation between area feature and fracture energy in concrete members. Results in volume density and fracture energy predictions illustrates that predictive models for primary data set (No crack sampling) indicates a little better performance compared to the other data sets generated by the re sampling techniques (Figure 2.8 and 2.9). In fact, some crack features could be missed through re sampling.

Both internal volume density and fracture energy as quantified internal damage states can be predicted using area and standard deviation features that proves the concept of relating internal behavior to external observations.

Different nonlinear regression techniques along with predicted energy density as a feature set were employed to predict slip. However, the results do not indicate a strong relation between the external observations and slip. Figure 2.10 illustrates the performance of different models. It was found that, Gaussian process using predicted fracture energy as a feature can outperform prediction better than other methods. As mentioned earlier, the

total slip on the specimen was divided between the crack planes based on their surface areas. However, mentioned questionable assumption does not seem to be completely correct since the specimens were tested under a displacement based controlled test with the similar final displacements. In other words, some cracks in the data set were subjected to the same slip while they have a different external features. Although the current data set does not seem to be an appropriate case for slip prediction based on the surface observations, results showed a correlation between slip and predicted fracture energy.

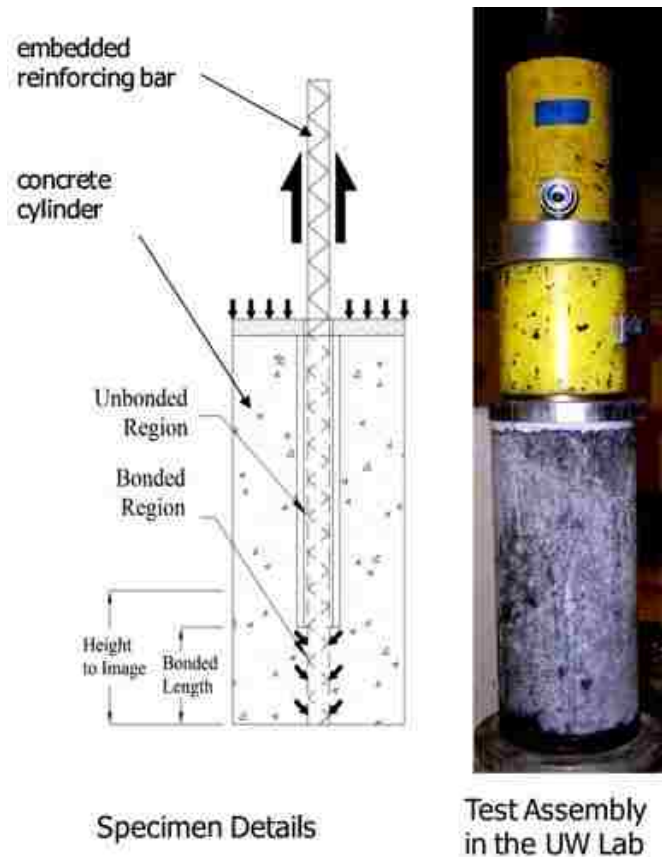


Figure 2.1: Pull-out test setup [72] used to obtain the CT specimen images analyzed in the present study

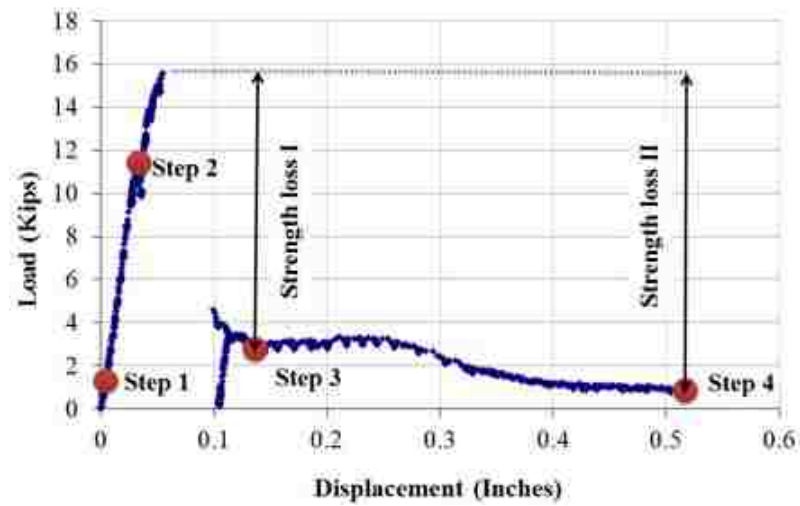


Figure 2.2: Typical Pull Out Test Load-displacement Response [72]. CT imaging points are indicated by "Step n" labels on the figure. Strength loss for last two steps was shown



Test Name	Number of load steps used	Imaging step	Diameter (in)	Height (in)	Approx. Imaged Length (in)	Image Resolution
Martin 5A-6	1	following testing	6	12	4	0.3 mm x 0.3 mm x 1 mm
Martin 5A-9	1	following testing	6	12	4	0.3 mm x 0.3 mm x 1 mm
YAXLON SampleB.	1	following testing	8	16	3	0.2 mm x 0.2 mm x 1 mm
Martin 5B-P1.	2	under load	6	12	2	0.3 mm x 0.3 mm x 1 mm
Martin 5B-P3F.	2	under load	6	12	4	0.3 mm x 0.3 mm x 1 mm
Martin 5B-P3W1A.	2	under load	6	12	4	0.3 mm x 0.3 mm x 1 mm
Martin 5B-P3W1B.	2	under load	6	12	4	0.3 mm x 0.3 mm x 1 mm
Martin 5B-P3W2.	2	under load	6	12	4	0.3 mm x 0.3 mm x 1 mm

Table 2.1: X-ray imaging details for each specimen

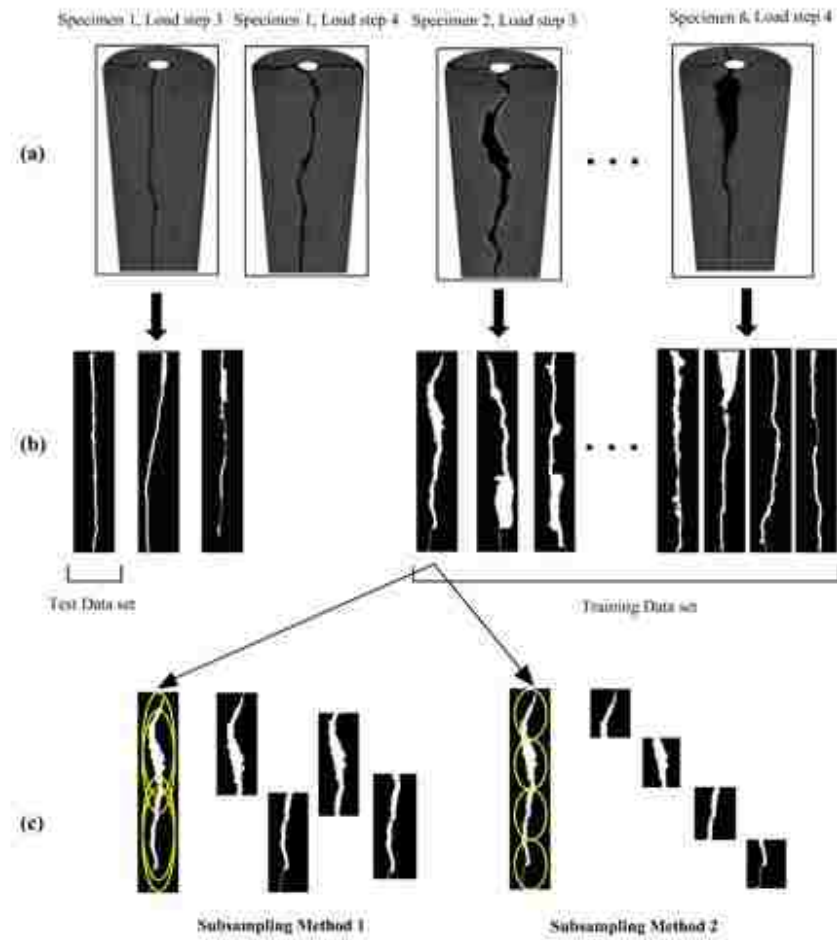
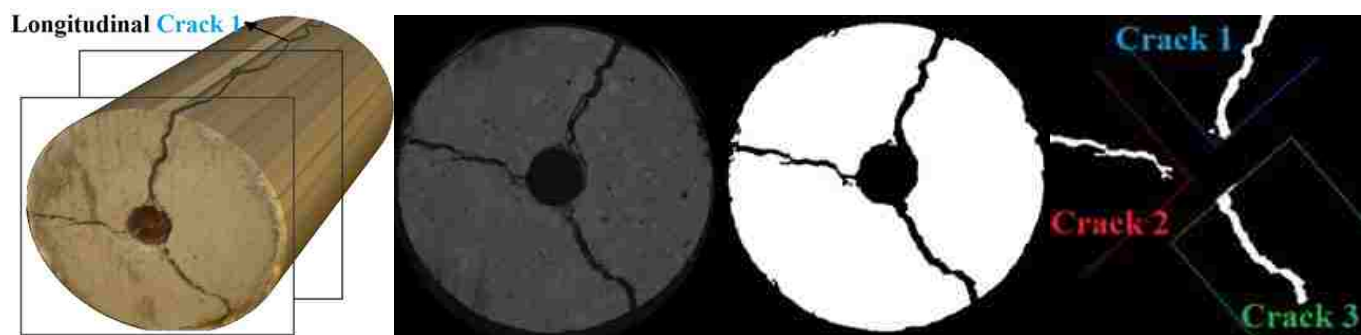
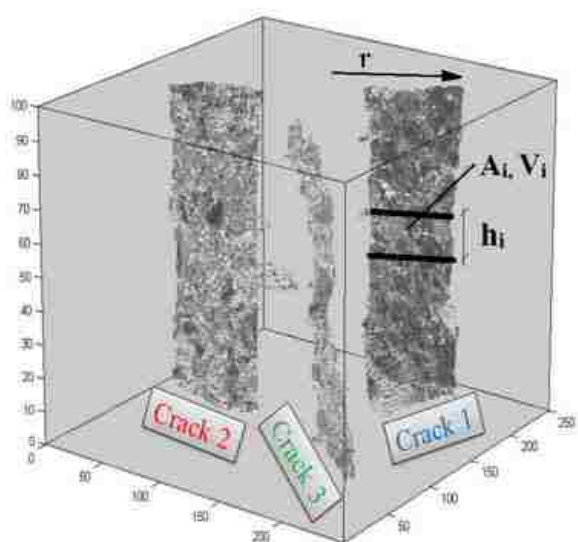


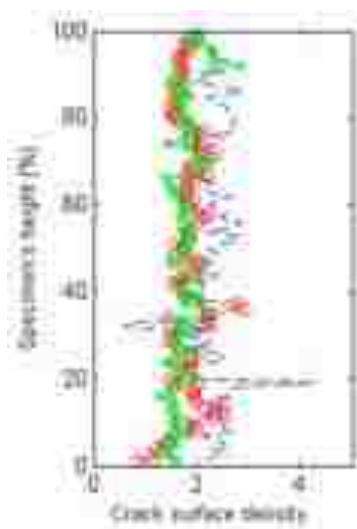
Figure 2.3: (a) 3D CT image data for each specimen/load step combination, captured from a total of eight specimens using one or two load steps per specimen. (b) Extended surface crack image data for each specimen/load step instance, leading to 40 total surface crack image sets separated into training and test data sets. (c) sub-sampling of surface crack images using two methods: 1) Four crack regions extracted via a randomly translated fixed sample length; 2) Four crack regions extracted using non-overlapping contiguous intervals



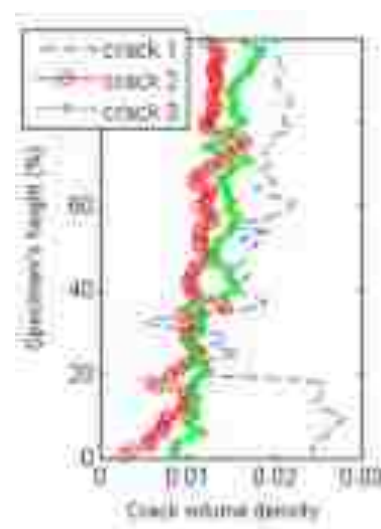
(a) Typical CT scan slice (b) Raw X-ray image (c) 2D segmentation (d) Crack extraction



(e) 3D segmentation



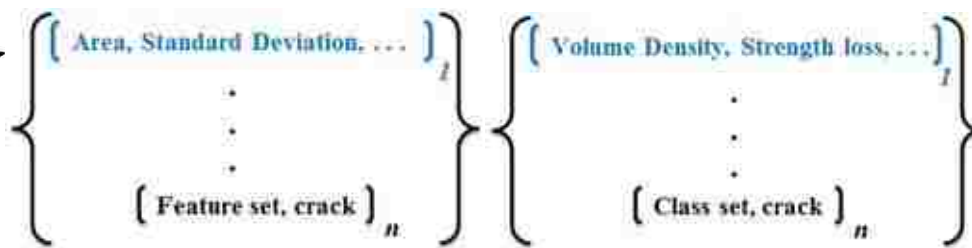
(f) Crack surface density distribution ( $\frac{A_i}{rh_i}$ )



(g) Crack volume density distribution ( $\frac{V_i}{\pi r^2 h_i}$ )



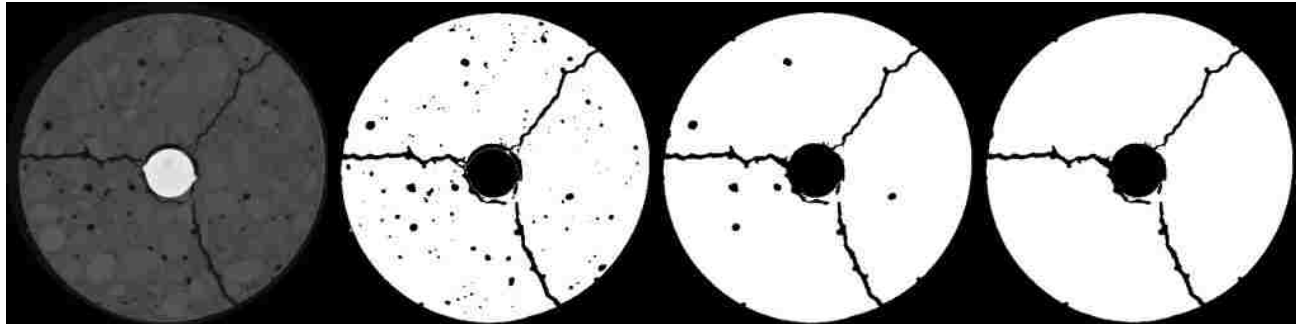
(h) Longitudinal Crack 1



(i) Feature extraction

(j) Defining true class set

Figure 2.4: (a) Typical cylinder splitting failure mode [72] (b) Raw X-ray image. (c) Segmented image using threshold technique. (d) Automated process to find and extract cracks in each 2D segmented image. (e) 3D segmentation (f) Crack surface density distribution (g) Crack volume density distribution (h) Longitudinal surface crack 1 from surface observation. (i) Feature set matrix extracted from longitudinal surface cracks (j) True class set matrix representing the values of the dependent variable of the feature set



(a) Raw X-ray image (b) 2D segmentation using gray thresholding (c) De-noising process (d) Filling the holes

Figure 2.5: (a) Raw X-ray image. (b) Segmentation using gray thresholding technique. (c) de-noising process through LmakeHomoonRC's method, this method improves segmentation by smoothing the image (e) Filling the holes left from de-noised image.

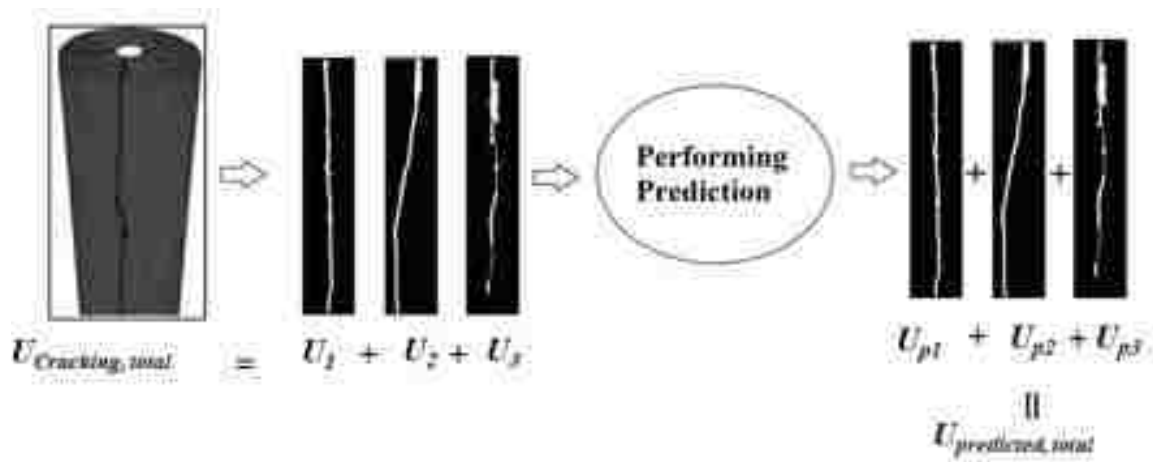


Figure 2.6: Used procedure to predict total fracture energy for the specimen (i.e. specimen  $i$ , load step  $jj$ ).  $U_1$ , and  $U_{p1}$  are true and predicted fracture energy corresponding to crack plane 1 respectively. Modified LOOCV was used to define training and test data sets as explained in Figure 2.3

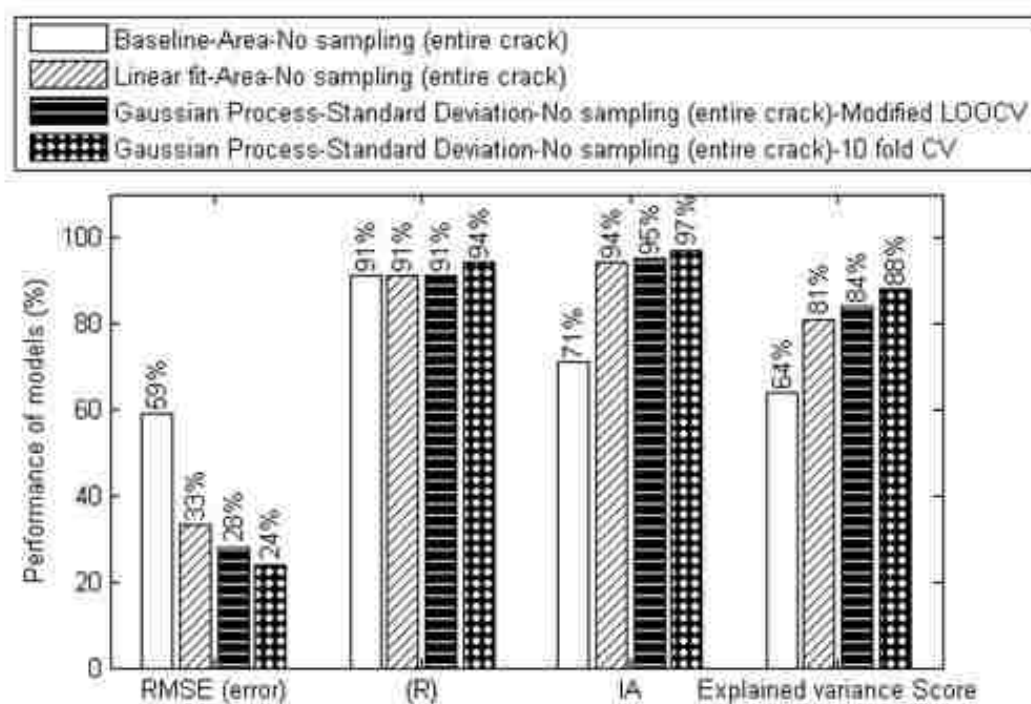


Figure 2.7: Comparison of volume density predictive models using Gaussian regression technique and Standard deviation feature with crack volume estimation model (Baseline). Data set was generated using entire crack length (No sampling)

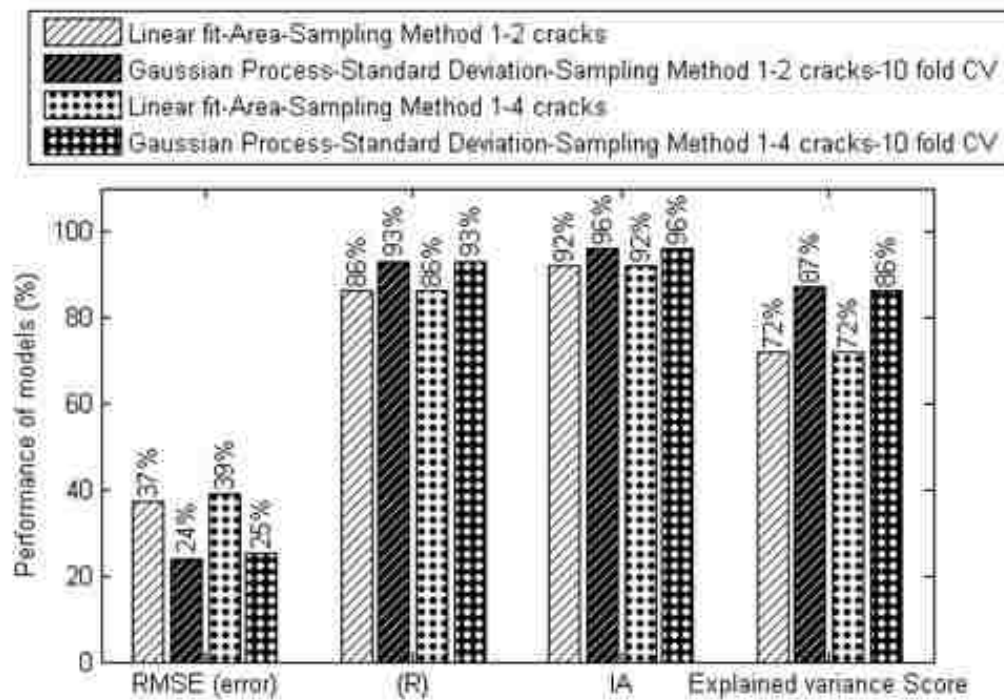


Figure 2.8: Comparison of volume density predictive models using Gaussian regression technique and Standard deviation feature with linear fit model. Data set was generated using sampling method 1



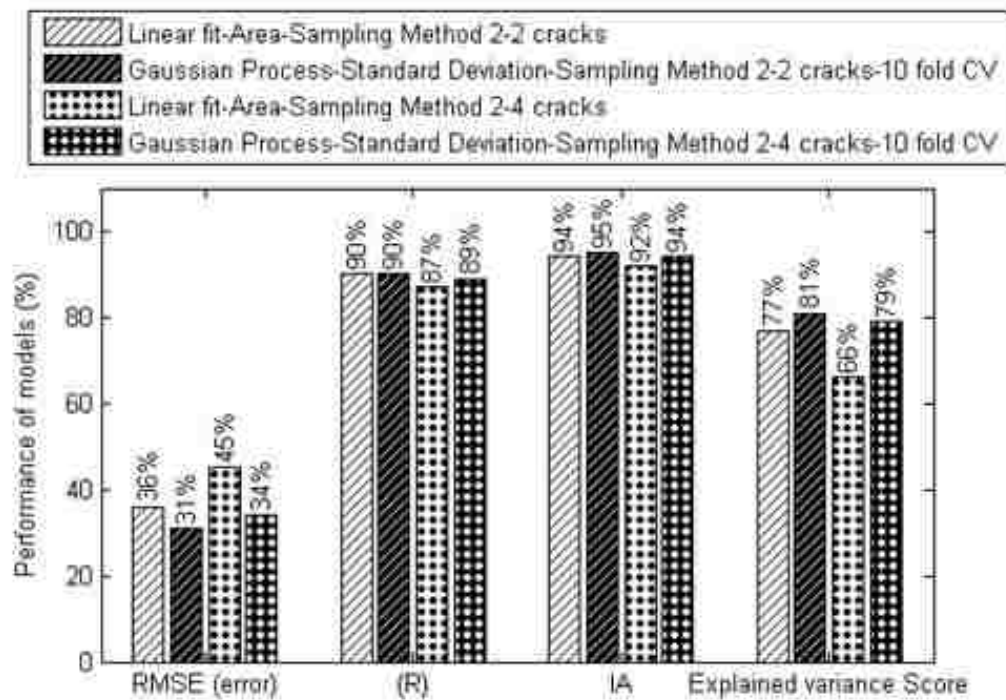


Figure 2.9: Comparison of volume density predictive models using Gaussian regression technique and Standard deviation feature with linear fit model. Data set was generated using sampling method 2

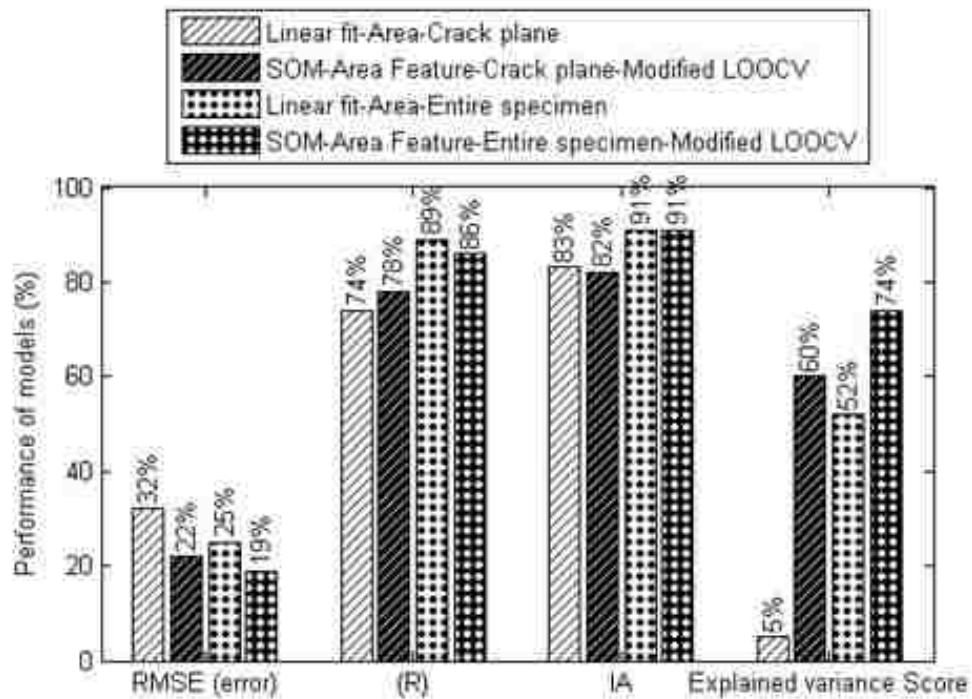
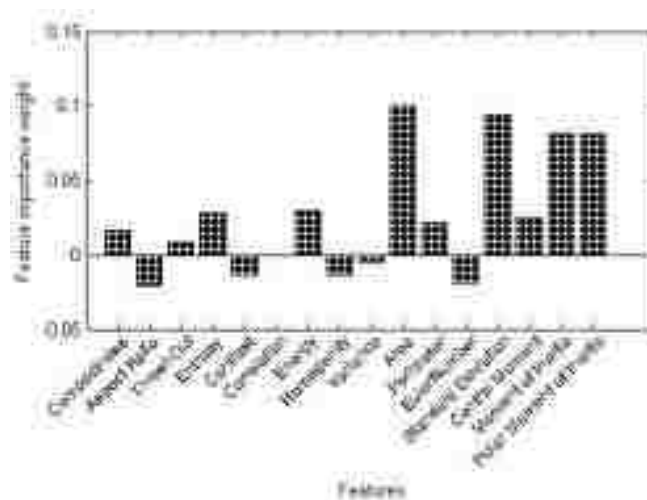
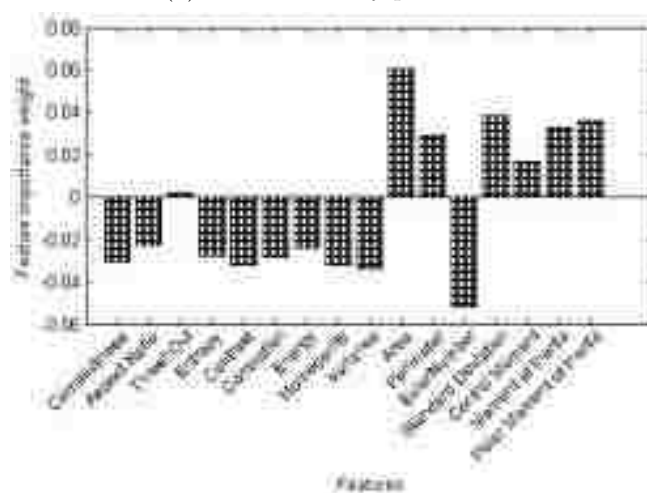


Figure 2.10: Comparison of fracture energy predictive models using SOM regression technique and Area feature with fracture energy estimation model (Baseline)



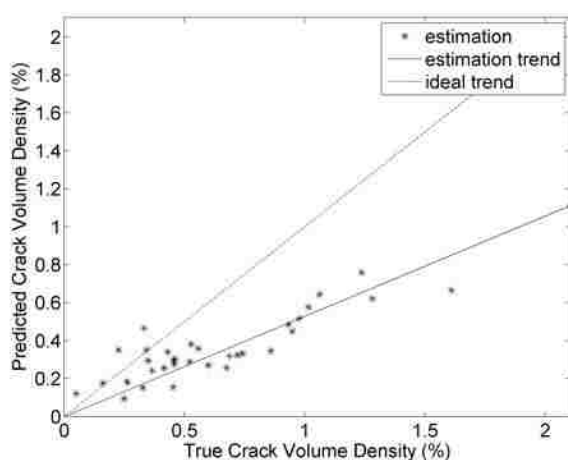


(a) Volume density prediction

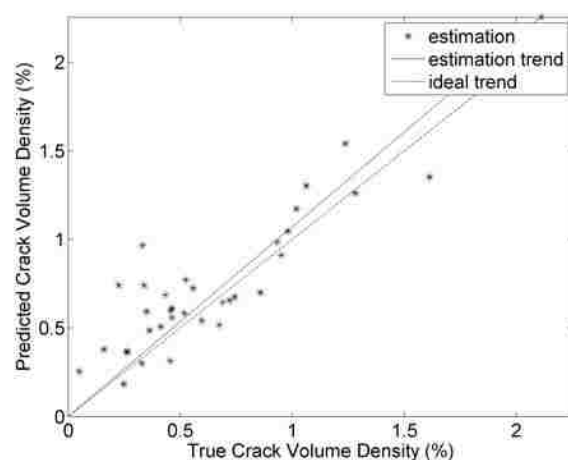


(b) Fracture energy prediction

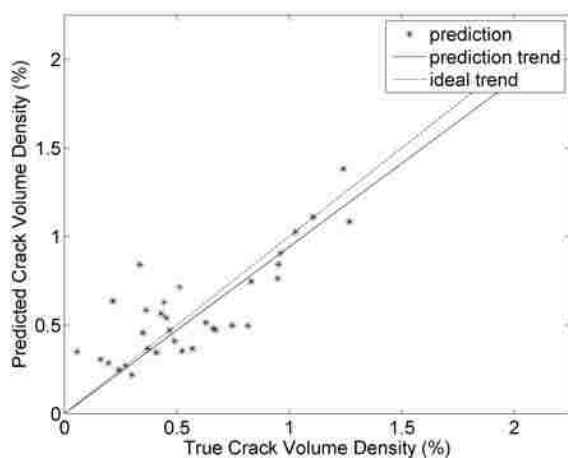
Figure 2.11: Feature importance weight using relief method for (a) Volume density, (b) Fracture energy density



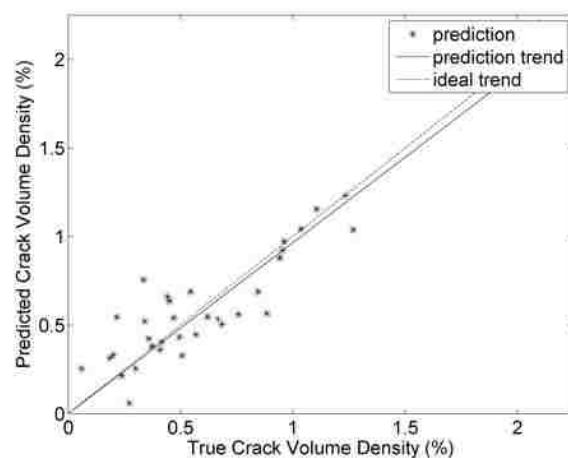
(a) Baseline-Area-No sampling (entire crack)



(b) Linear fit-Area-No sampling (entire crack)

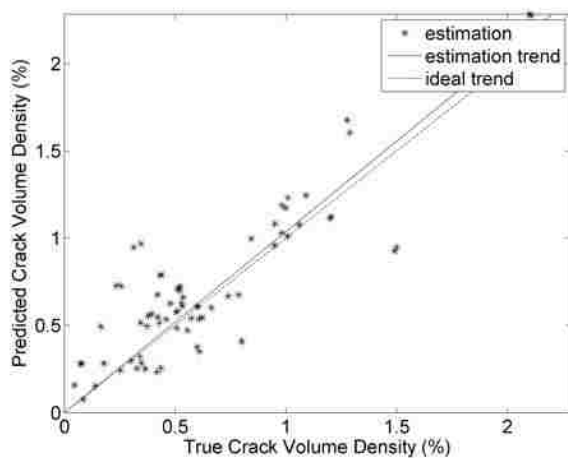


(c) Gaussian Process-Standard Deviation Feature-No sampling (entire crack)-Modified LOOCV

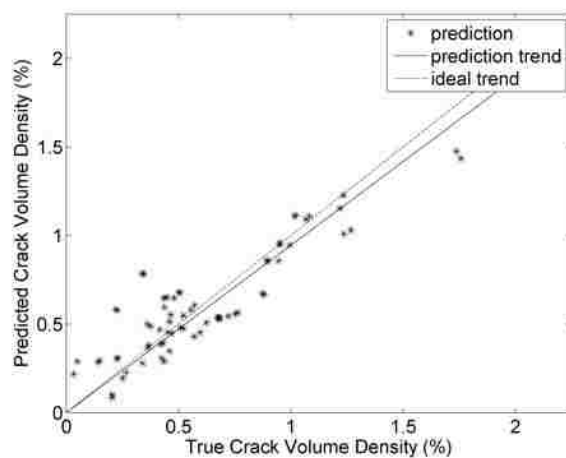


(d) Gaussian Process-Standard Deviation Feature-No sampling (entire crack)-10 Fold CV

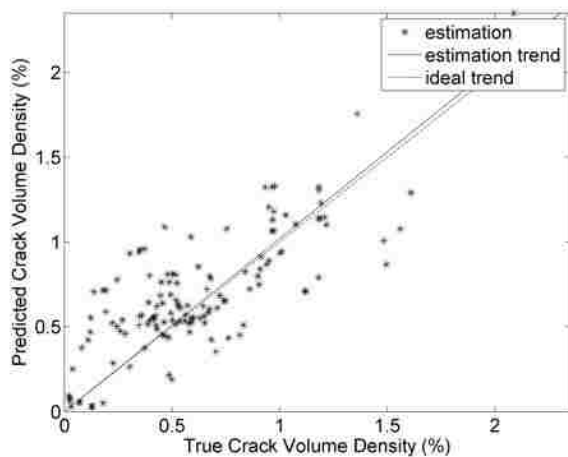
Figure 2.12: Performance of Volume Density predictive models using Gaussian Process regression technique and Standard Deviation feature on primary data set. The data set was generated using entire crack length (No sampling)



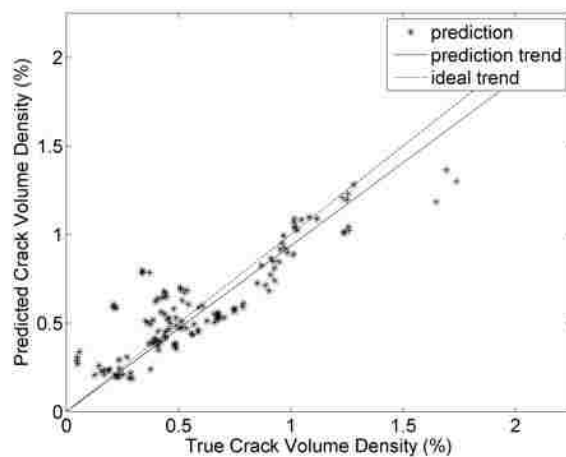
(a) Linear fit-Area-Method 1- 2 cracks



(b) Gaussian Process-Standard Deviation Feature-Method 1- 2 cracks-10 Fold CV

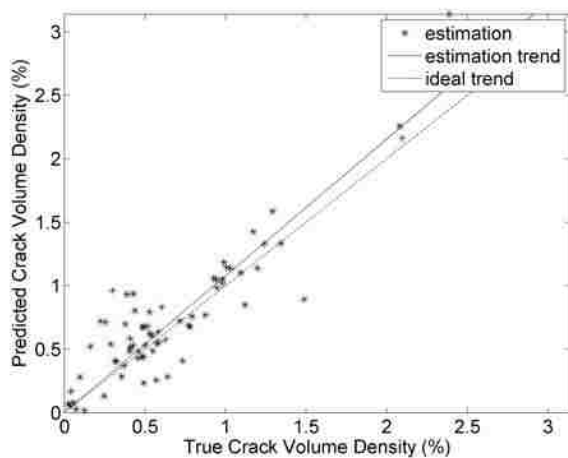


(c) Linear fit-Area-Method 1- 4 cracks

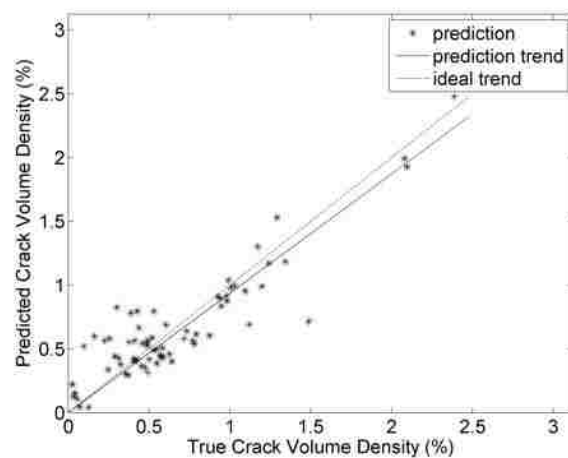


(d) Gaussian Process-Standard Deviation Feature-Method 1- 4 cracks-10 Fold CV

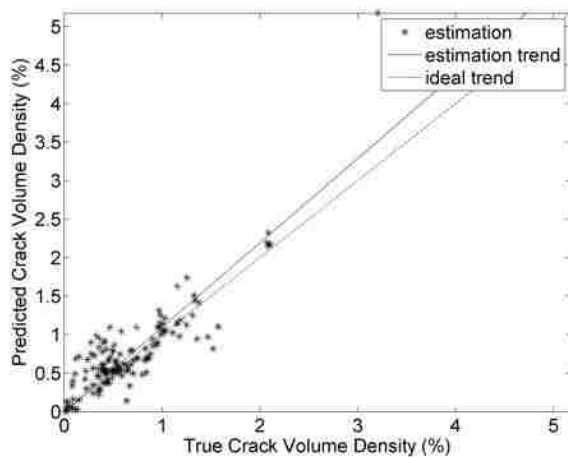
Figure 2.13: Performance of Volume Density predictive models using Gaussian Process regression technique and Standard Deviation feature on data set 1 generated using sampling method 1



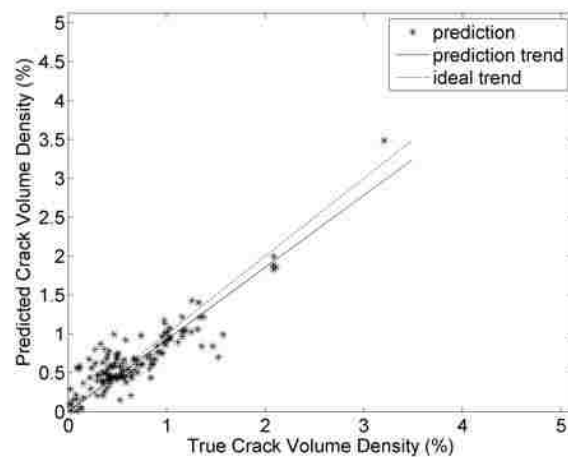
(a) Linear fit-Area-Method 2- 2 cracks



(b) Gaussian Process-Standard Deviation Feature-Method 2- 2 cracks-10 Fold CV

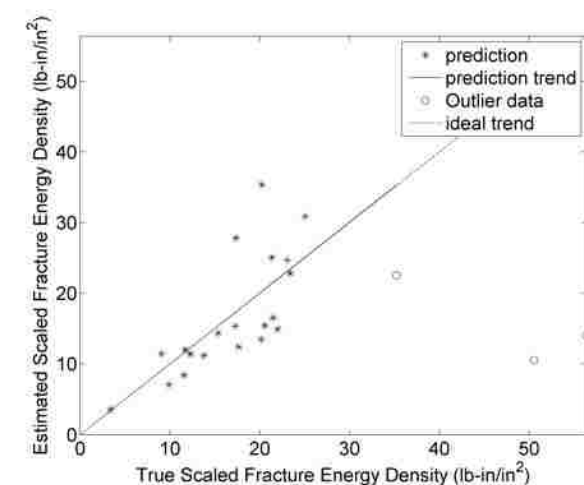


(c) Linear fit-Area-Method 2- 4 cracks

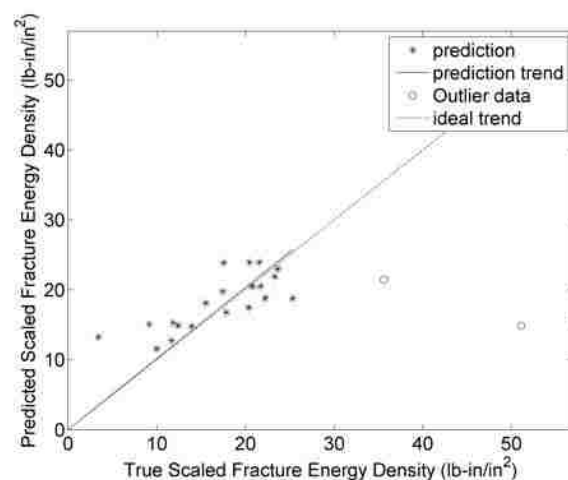


(d) Gaussian Process-Standard Deviation Feature-Method 2- 4 cracks-10 Fold CV

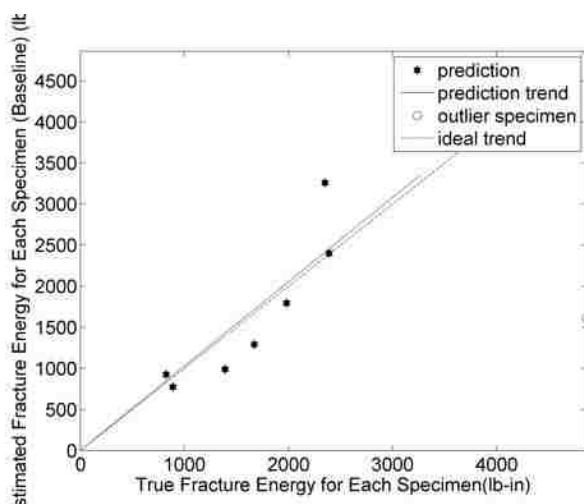
Figure 2.14: Performance of Volume Density predictive models using Gaussian Process regression technique and Standard Deviation feature on data set 1 generated using sampling method 2



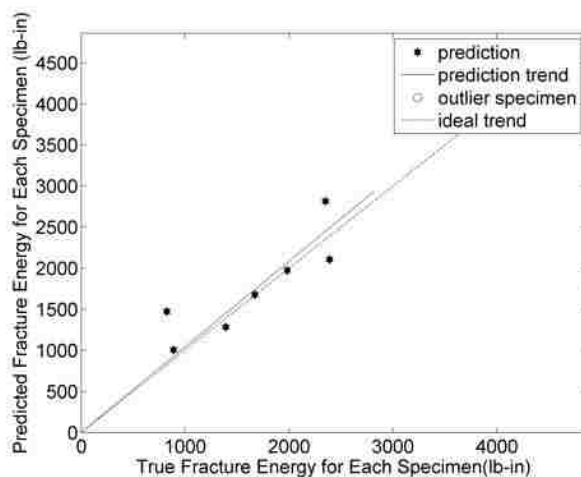
(a) Linear fit-Area-Crack plane



(b) SOM-Area Feature-Crack plane-Modified LOOCV



(c) Linear fit-Area-Entire specimen



(d) SOM-Area Feature-Entire specimen-Modified LOOCV

Figure 2.15: Performance of Fracture Energy predictive models using SOM regression technique and Area feature on different data sets

Chapter 3

**COMPUTER VISION-BASED DAMAGE INSPECTION  
FRAMEWORK FOR SHEAR CRITICAL BEAMS (WITH  
PERMISSION FROM ASCE)**

### **3.1 abstract**

This chapter presents the results of using machine vision to relate surface observations to quantitative load level estimates in structural components. In particular, image processing and machine learning regression techniques were used to build estimation models capable of quantifying internal load levels (i.e., shear and moment) in shear-critical RC beams and slabs based on surface crack pattern images. The estimation models were generated and tested using image data sets obtained from ten different earlier published studies, which provided 558 crack pattern images captured from 84 shear-critical RC beams and slabs across a range of load and damage levels. Working with these existing image data sets, various textural and geometric attributes of surface crack patterns were defined and evaluated with respect to their effectiveness in building useful estimation models. Multiple statistical error measures and cross-validation methods are used to quantify predictive accuracy, and several training/test methodologies were considered relative to potential field application scenarios. The results show that the estimation models based on surface crack image data can work well across a wide range of geometries, loadings, concrete strengths, and reinforcement details. Size effects can be accounted for by including specimen physical dimensions in the feature sets used for model training, and fundamental design relations can be used to develop useful non-dimensional prediction parameters.

### **3.2 Introduction and Motivation**

Infrastructure inspection plays a key role in public safety in regards to both long-term damage accumulation and post extreme event scenarios. As part of the rapid developments in data-driven technologies that are transforming many fields in engineering and science, machine learning and computer vision techniques are increasingly capable of reliably diagnosing and classifying patterns in image data, which has clear applications in inspection contexts. Such automated inspection approaches have been shown to be effective for identifying and characterizing surface damage, but work to date has generally focused on qualitative dam-

age classification. The purpose of the current study is to focus on the potential for machine learning and image processing techniques to enable quantitative damage and load level assessments based on captured visual data (e.g., digital photographs taken with a phone). In particular, image processing and machine learning regression techniques have been used to build models capable of estimating internal load levels in shear-critical RC beams and slabs using images of surface crack patterns. The primary application context motivating this work is field inspection scenarios in which captured images of surface damage could be used for making decisions concerning repair, replacement, additional monitoring, or evacuation depending on the time scales and events in question.

As will be described in more detail later, the estimation models in this study were generated and tested using image data sets obtained from previously published experimental studies with shear-critical specimens covering a wide range of geometries, loadings, concrete strengths, and reinforcement details, and with image and related data captured at various load and damage levels. Working with these existing image data sets, various textural and geometric attributes of surface crack patterns were defined and evaluated with respect to their effectiveness in building useful estimation models. Relatively simple crack representations were used, consistent with the varying nature of the images available in the earlier studies, but also with an eye toward potential field applications in which image capture and segmentation quality could be limited. The fundamental state quantifications included both absolute and relative measures. The relative measures focused on relating a given image to a corresponding load level estimate in terms of percentage of capacity, while the absolute measures were used to estimate moment and shear values directly.

In addition to investigating the basic feasibility of the approach, this study also identified and evaluated a range of strategies, algorithms, and parameters that affect the accuracy of the estimations. Multiple statistical error measures and cross validation approaches are used to quantify predictive accuracy, and several training/test methodologies are considered relative to potential field application realities.



### 3.2.1 Relation to Prior Work

A number of previous studies have focused on machine vision applications in infrastructure contexts. For reinforced concrete in particular, there has been an emphasis on the fundamental problem of segmentation (i.e., automatically extracting cracks from images). Representative work in this area can be found in [41], [44] and [40] with one of the primary challenges being the development of robust algorithms capable of reliable extraction with realistically noisy field images. The focus of the present study is on post-segmentation analysis, and so any of the various available techniques could be used as needed as a front-end for the estimation models presented later.

Moving from segmentation to machine learning, several studies have developed mechanisms for categorizing damage according to machine heuristics ([81], [27], [86], [87], [30], [28], and [35]). This categorization can range from simple binary damage/no-damage identification to more general classifications of discrete levels of damage. The primary distinction in the present work is the emphasis on quantitative estimation rather than more qualitative categorization or classification. Of course, continuous quantified estimates could be used for categorization by introducing discrete threshold values to partition data into coarse bins, so these are not completely independent, but back-end categorization of this kind is different than using machine learning categorization (versus regression) algorithms up front.

Stepping beyond basic descriptive categorization, there also have been limited attempts to use more fine grained, regression-based computer vision approaches to generate quantitative damage assessments such as peak experienced drift ([60]; [47]) in reinforced concrete columns. In [60] surface damage observations including cracking and spalling were associated with prescribed descriptive damage states for RC columns. This study also sought to indirectly quantify experienced peak drift in RC columns through data analyses of associated damage states. The study in [47] used images captured during cyclic column testing to predict the maximum drift to which the columns had been subjected based on the crack and spalling patterns photographed during the testing. This study is most closely related

to the work in this study in that the goal was to use purely image-based data to generate quantitative estimates of structural loading. The main distinctions in the work in the present study relative to these earlier studies include the much larger data set used (both in terms of the amount of data and the breadth of its sources), the focus on beams and slabs rather than columns, the wider range of quantities predicted, the larger set of image features used, and the more in-depth consideration of error measures and parameter sensitivities.

Machine learning also has been used to relate data other than images to quantitative estimates of structural system behavior. For example, [21] and [54] used neural network algorithms to estimate design strength for concrete beams based on knowledge of fundamental material and geometric design properties. Although there are some similarities between these earlier studies and the current work in terms of techniques used, estimating design strength based on design parameters is fundamentally different from relating imaged external damage to load levels. Also, most the predictive models presented in this study assume that there are no available data concerning design details such as concrete compressive strength ( $f'_c$ ), area of longitudinal reinforcing steel ( $A_s$ ), reinforcing arrangement, etc. This is intended to be consistent with likely field inspection circumstances. However, results are also presented for cases in which partial design data are assumed known, the idea being to illustrate the effect such added data could have on estimation accuracy.

### *3.2.2 Organization of the chapter*

The overall goal motivating this work is to be able to use captured image data to estimate corresponding levels of loading expressed numerically either as a failure ratio (e.g., current shear divided by shear capacity) or absolute resultant value (e.g., current shear or moment). To achieve this the primary challenges include: (i) assembling a database of images and load levels captured during experimental testing to use for model training and validation; (ii) developing suitable methods of image preprocessing and standardization to support machine learning; (iii) identifying effective machine learning algorithms capable of quantitative estimation of structural loading from image data; (iv) selecting evaluation metrics to com-

pare approaches and characterize reliability of the results obtained; and (v) characterizing sensitivities related to the multiple layers of algorithmic and parameter selections inherent in the overall process. The scope of this study includes addressing each of these challenges in the context of shear-critical, conventionally-reinforced concrete beams and one-way slabs, focusing primarily on inspection applications for which no internal properties of the structural members are known.

In terms of the scope of the results obtained, it is important to note that in many computer vision applications (e.g., facial recognition) there is a definitively correct answer for a given image, and so the goal is 100% accuracy. However, in the context of structural load estimation there is no single deterministic value associated with a given image due to the inherent variability of the relevant structural materials and systems considered. This makes it difficult to identify a unique upper bound target accuracy metric—the errors in estimation can at best match the reproducibility of the underlying damage phenomena themselves. Because this is the first published study considering image-based quantitative load level estimation for a beam and slab data set of this size and complexity, the fundamental purpose here is to generate baseline results capable of demonstrating the basic feasibility of the approach, and to characterize the accuracy and variability of the associated load-level estimations as a function of particular modeling algorithms and parameters. The data set will be made available for further improvements by others, which could include both the expansion of the image data set itself and the application of different computer vision approaches.

The presentation to follow begins with a summary of the ten previously published experimental studies from which the image and load-level data used in this study were obtained, and then summarizes the data sets themselves. The next section outlines the image-processing and machine learning framework used to develop the families of estimation models to be calibrated, tested, and cross-validated with the collected experimental data. The primary results are presented next, illustrating the effects of varying algorithmic and sampling approaches, and characterizing the accuracy of the estimations achieved for various quantities of practical interest. A suite of error measures are used for purposes of comparison, leading

to a set of conclusions relevant to the feasibility of using the approach in field inspection contexts.

### ***3.3 Previous Experimental Studies and Image Data***

The image data sets used in this study were collected from previously published monotonic point load and uniform loading bending tests on conventional RC beams and slabs without shear reinforcement. The results of these tests have been presented in ten earlier publications: [82]; [16]; [71]; [70]; [58]; [9]; [62]; [59], [57] and [66]. These largely independent experimental tests were carried out for the purpose of studying various effects of concrete strength, beam size, longitudinal reinforcement ratio, and aggregate size on the ultimate shear strength of concrete beams and slabs. A typical test setup with shear and moment diagrams is illustrated in Figure 3.1 [9]. Table 3.1 summarizes the ten test series from which the data used here were obtained, and lists study references, key structural parameters, test type, and number of images used in each data set. As can be seen from the last row of Table 3.1, a total of 558 images were obtained from 84 different specimens, with the collected data covering orders of magnitude ranges of structural and geometric parameters, and with a variety of load patterns. This provides a challenging breadth of behavior for the image-based estimation models developed here, but also ultimately enhances the robustness of the results.

Although there was wide variability of the principal parameters in these previous tests, they were nonetheless all designed to have shear-dominant failures. In this sense, the results of this present study are limited in their generality, because the classes of failure considered are limited to those which are shear critical. More general data sets are being considered for future publications, but the more focused class of data considered here provides a good test case for evaluation of the overall approach and for more detailed consideration of alternative algorithms and methods for estimation.

The prior studies were not undertaken for the purposes considered here, but they all have the key common feature of having captured sequences of images linked to recorded load levels. Because the images were captured at multiple load levels for each specimen, it is possible to

look for patterns and correlations between the images and the loading. Given that even a cursory visual scan through the images reveals a clear increase in cracking with increasing load level for any given specimen, it is not surprising that basic correlations exist, so the key question in this study is the degree to which these broad correlations can be refined to enable useful quantitative load estimations. To be useful these estimations should be valid across as wide a range of load and geometric parameters as possible, and for purposes of the present study they ideally should be based solely on externally visible data of the kind likely to be available in a field inspection scenario (i.e., parameters like concrete compressive strength, rebar size, etc. should not be part of the feature set used to build the estimation models). The computer vision framework discussed in the next section is intended to provide the means to develop and evaluate such image-based estimation models.

### **3.4 Image-Based Estimation Framework**

As mentioned earlier, the machine vision estimation framework used here is based on computer vision, which combines image processing and machine learning. In the context of the present study, the image processing primarily revolves around extracting crack patterns and scaling of images to allow for consistent analysis, while the machine learning focuses on identifying image features and regression algorithms that could enable successful linking of images to corresponding estimated load levels. For the most part existing techniques and algorithms have been used for both the image processing and machine learning, and so the emphasis in this study is on identifying effective combinations of algorithms within the application scenario constraints, and characterizing the resulting accuracy of the estimations.

The four major steps employed to perform the estimations in this study align with those commonly used in computer vision applications<sup>1</sup> [47]: (i) image segmentation; (ii) image transformation; (iii) feature extraction; and (iv) machine learning (with regression in this case). Segmentation is the process of converting a color or gray scale image to a binary image

---

<sup>1</sup>Convolution Neural Network approaches represent notable exceptions to this four-step process, but these typically require larger data sets to be effective.

in which pixels are categorized as either inside or outside cracked regions, and are represented by 0 and 1 indices, respectively. For the data used in the current study the collected data sets have a wide range of actual and pixel dimensions, and so the image transformation step focused on converting the segmented images to a standard baseline format such that crack pattern features were preserved. Various textural and geometrical features were defined and then computed (extracted) using the binary crack pattern images, essentially providing a means of projecting images into shared parametric spaces. Machine-learning-based regression techniques can then be used to relate extracted features to load and damage states. The details of each of these processing steps are outlined in the following subsections.

### *3.4.1 Image Segmentation*

As mentioned earlier, the image data sets were obtained from ten different references. As shown in Figure 3.2 there were two broad classes of crack pattern images contained in the data sets. The images collected from [9] were cropped from laboratory observation videos in which cracks were manually marked during the experimental test (Figure 3.2(a)), and these raw images were then segmented into black-and-white images. Figure 3.2(b) is representative of the crack pattern images for the nine other data sets, which were published in an essentially pre-segmented form in the associated references by means of comprehensive drawn crack patterns. Thus, the raw images obtained from [9] were the only ones for which segmentation was required, and this task was simplified because of the markings applied during the experimental test. This major simplification of the segmentation task streamlined the work for this study, but it is worth noting there are more general segmentation approaches that could be used for extracting cracks from the concrete surfaces in field contexts (see, e.g. [41], [44], and [40]). For the single set of images for which segmentation was necessary, it was found adequate to use Canny's method [15], which is a common edge detection technique.

One side effect of using pre-segmented crack pattern images is the loss of accurate crack width measures. In effect, cracks are depicted as lines of constant unquantified width, which

clearly results in data loss that could be important in some contexts. However, the simplicity of this representation has attractions on several fronts, and so the geometric features to be used later are all based on crack extent and crack patterns without reference to crack width. With this in mind, the thresholding parameters and aspect ratio filtering used for the Canny's method segmentation mentioned above were adjusted to generate results consistent with the hand-drawn crack patterns in the other references. Figure 3.2(a) shows a typical segmentation result, and the details of this process are described more fully in the next subsection.

### 3.4.2 Image Transformation

In order to take account of the different specimen and image dimensions, an image transformation step was used for normalizing and registering the crack pattern images in a standard format. In particular, the images were all converted to a consistent form such that comparable features could be extracted from the images using an equivalent image space (i.e., constant dimension binary matrices). There are two basic ways to accomplish this. In the first approach, additional background pixels can be added horizontally and vertically such that all images in the entire data sets have equal pixel dimensions matching the biggest overall image as shown in Figure 3.3 (c). As can be seen,  $w_i$  and  $w_{max}$  correspond to the initial width of the image, and the adjusted width of the image with respect to a bigger image respectively. In this approach the initial image size will be in proportion to the specimen dimensions. For instance, a 4-m deep beam should have an image size four times larger than a 1-m deep beam before adding the additional black pixels to make the size of all images equal.

In the second approach, the input images are scaled in both the horizontal and vertical directions such that all images in the data set have consistent dimensions. Any convenient dimensions can be used to do the scaling. As can be see in Figure 3.3 (d),  $w_i$  and  $w_0$  correspond to the initial width of the image and the scaled width of the image with respect to a bigger image. Although the crack pattern structure might be slightly changed as shown

in Figure 3.3 (d), dimensionless crack features can be extracted from the crack pattern images scaled using this second approach.

As discussed above, a morphological process also was applied after the scaling step to renormalize the thickness of the cracks to essentially a single pixel. Figure 3.4 illustrates this transformations for a case in which segmentation was the initial step. It is worth noting the results presented later in this study were not found to be highly sensitive to the details of these transformations.

### 3.4.3 Feature Extraction

In the context of standard computer vision, feature extraction is the process by which image data can be projected into lower dimensional spaces for purposes of comparison and analysis. This is accomplished by computing numerical quantities associated with the binary segmented images using various geometric and textural properties of the pixel patterns. For the present study a combination of geometric and textural quantities was used to develop feature sets. Some groups of attributes were based on the standard geometric and textural analysis functions built into image processing toolbox in MATLAB [56], which automatically are applied at the level of individual cracks, and which were then averaged or totaled across multiple cracks. Another group of attributes was also computed in this study to directly represent the overall distribution of all cracks.

Altogether the computed features can be organized into four categories: (i) per-crack averaged quantities; (ii) per-crack total quantities; (iii) overall image properties; and (iv) custom multi-crack features. The features in the first category (per-crack averaged quantities) are computed by averaging individual crack quantities across all crack objects in the image. The individual crack quantities are computed using the MATLAB function `regionprops` and are listed in the first part of Table 3.2. The features in the second category (per-crack total quantities) provide a representation of the total quantities summed over all crack objects in the image. These features are also based similarly on the MATLAB `regionprops` function and are as listed in the second part of Table 3.2. The third category



of features, listed in Table 3.3 and also based on built-in MATLAB functions, are used to characterize the overall texture and related image properties. The image features in the final category (custom multi-crack features) were defined and implemented as a part of the current study, and are described in more detail in the next section.

#### 3.4.4 Multi-Crack features

It is not difficult to see that the overall attributes of the crack patterns likely would be important factors in relating image features to load level. To this end, some additional multi-crack attributes were used in this study to directly represent the amount and distribution of cracking at each load stage. These overall features are described below:

- Moments of area of the pixels about the centroidal  $x$  and  $y$  axes,  $I_{xx}$ ,  $I_{yy}$ , and  $I_{xy}$ :

$$I_{xx} = \sum_{i=1}^n (x_i - \bar{x})^2 \quad (3.1)$$

$$I_{yy} = \sum_{j=1}^m (y_j - \bar{y})^2 \quad (3.2)$$

$$I_{xy} = \sum_{i=1}^n \sum_{j=1}^m (x_i - \bar{x})(y_j - \bar{y}) \quad (3.3)$$

in which  $x_i$  and  $y_i$  are the pixel coordinates of the non-zero pixels in the image,  $\bar{x}$  and  $\bar{y}$  are the components of the pixel centroid as shown in Figure 3.5, and  $n$  and  $m$  are the image dimensions in the  $x$  and  $y$  directions, respectively.

- Polar moment of area,  $I_p = I_{xx} + I_{yy}$ .
- Total and average distance between cracks. Figure 5.3 outlines the process for calculating the total and average distance between cracks. First, the distance transform of the binary segmented image is computed using MATLAB's `bwdist` function in the image processing toolbox. For each pixel in the segmented image, the distance transform computes the distance between that pixel and the nearest nonzero pixel. The

calculated distance values are assigned to each pixel in the distance transform matrix. Figure 5.3(a) shows an example colormap representation of the distance transform for a representative segmented image (the cracks themselves have been superposed in white). The color bar indicates the computed distance values. Then, the watershed transform in MATLAB's image processing toolbox is used to convert the distance transform image to a label matrix that identifies the "watershed ridge lines" surrounding each crack object. This is illustrated in Figure 5.3(b) in which each region surrounding a crack object is identified with a certain color. Figure 5.3(c) clarifies the procedure by showing watershed ridge lines along with crack objects in the same figure. The watershed ridge lines identify the furthest points between cracks. The total distance between cracks can be computed by adding distance values at pixel locations along the watershed ridge lines in the distance transform matrix. Similarly, the average distance between cracks is the total distance between cracks divided by the total number of pixels corresponding to watershed ridge lines. Figure 5.3(d) indicates the distance values in the distance transform image at location of watershed ridge lines.

In addition to these aforementioned geometrical and textural image features, externally visible overall structural geometric properties such as span-to-depth ratio,  $a/d$ , and depth,  $d$ , can be added to the feature set to take account of the size effect in the estimation models.

The set of features outlined here is neither exhaustive nor unique, and in many cases one would not expect strong correlations to emerge between load levels and some of the chosen trial features. However, the features chosen are wide ranging in their coverage of the image characteristics, and it is the job of the machine learning algorithms described in the next section to identify features with strong correlations (i.e., dominant features) to the quantities of interest. In this regard it generally is best to err on the side of inclusion rather than exclusion of features.

### 3.4.5 Machine Learning and Data Clustering

In the present context machine learning involves the identification and tuning of correlations between image features and structural quantities of interest, in essence training the underlying models to recognize damage characteristics in terms of image features. For this study various machine learning based regression techniques were employed to accomplish this training of the predictive models. The WEKA toolbox was used, which is an open source machine learning package implemented in MATLAB [38]. This toolbox provides several supervised machine learning techniques such as Gaussian process, SVM, J48 decision tree and random forest algorithms.

Regression analysis can transform large amounts of data into meaningful numerical patterns such that missing or new data set can be estimated. Combining the supervised machine learning techniques and regression methods enables the development of quantitative predictive models.

In addition, the Tensor Flow toolbox in MATLAB was used to perform Deep Neural Network (DNN) modeling [63]. The DNN approach is an extension of simple neural network modeling inspired by the structure of the brain to find relations in highly nonlinear data sets. DNN is based on networks of interconnected processing units (called neurons) that can be trained to predict and classify output from a given input. The neurons are connected with each other across different layers with numerical values called weight,  $w_{ij}$ , as shown in Figure 3.7(a). Each neuron computes a numerical value,  $Z_j$ , based on input weights and values from its previous layer. The calculated  $Z_j$  value then is scaled using a predefined activation function to be transmitted to other neurons as inputs as shown in Figure 3.7(b). During this process, a bias value,  $b_j$ , which is not related to input weights can also be added to compute  $Z_j$  in order to improve the performance of the network. The Tensor Flow neural network toolbox is a multi-layer back-propagation artificial neural network and the DNN model used here was built using the following configuration strategies and parameters:

- Data scaling: the input (i.e., feature sets) values were centered to have mean of zero

and scaled to have standard deviation of 1.0. Final output values were scaled back to the original data range.

- Network topology: different combinations of layers and neurons were used to build the predictive models. A network with 2 hidden layers, the first with 20 neurons and the second with 7 neurons, was found to have good performance for both failure ratio and scaled shear strength predictions.
- Activation function: a hyperbolic tangent function was used.
- Learning rate: a value of 0.2 was used to generate the results presented below.
- Batch size: the maximum size of the training data set was used to set the batch size.
- The Stochastic Gradient Descent (SGD) optimizer was used.

At present there are no rigorous procedures to optimize parameter choices for neural networks such as those given above, so the above parameter value choices were based on basic testing and should be viewed as illustrative rather than definitive.

### ***3.5 Modal Analysis and Cross-Validation***

A key step to evaluate machine learning models is to define independent training and test data sets. Multi-fold cross validation and leave-one-out cross validation (LOO-CV) have been common approaches to evaluate machine learning models. Previous studies demonstrated that performing 10-fold cross validation usually produces reliable results in practice [45]. It was also shown that performing LOO-CV can be justified when a small number of samples are available since it keeps a relatively large percentage of the data active at all times. However, the results of LOO-CV can change if the data set is later extended. LOO-CV can be particularly effective for model selection [45] , and this approach was used for some of the preliminary results presented later.

As mentioned earlier, the image data set used in the present study consists of 558 crack pattern images obtained from 84 specimens. Figure 3.8 illustrates 10-fold cross validation as would normally be applied at the image level. The overall set of images is partitioned randomly into 10 approximately equally sized subsets (folds), and then sequentially each fold is used as a test data set while the other 9 folds are used for training. Overall results are based on averaging across the outcomes of the 10 cross-validation iterations.

In looking more closely at Figure 3.8 it is worth noting that the image-level approach introduces what could be thought of as an unfair bias to the results relative to the field application scenarios envisioned as a long-term outcome of this approach. Consider for example sample S1 in Figure 3.8(a). The image/load level combination marked in gray would initially be in the test set for the given fold, but the training set would have all the other images from that same specimen, thus bracketing the test image with the damage states before and after. In normal inspection contexts one would not have prior and subsequent damage images when viewing a current state<sup>2</sup>, and so including these images from the same specimen skews the results to appear better than would be the case in reality.

Therefore, for the purposes of this study a modified 10-fold cross validation applied at the specimen level was used to generate more independent training and test data sets as shown in Figure 3.9. The training and test data sets are randomly generated from the specimens such that the test fold and training fold images do not share any common specimens. This style of cross validation on the specimen level more fairly represents the situation relevant for field inspection scenarios. The difference in performance between image-level and specimen-level cross-validation will be demonstrated in the results presented below.

### **3.6 Performance of estimation models**

As noted earlier there are many model variations possible within the same overall framework: image processing, image features, machine learning algorithms, algorithm parameters, and

---

<sup>2</sup>With continuous monitoring one might have prior damage images, but still one would not have the corresponding load levels.

sampling techniques can be combined freely to obtain related but distinct estimation models. This section presents and compares results for a representative set of such model variations both to provide a sense of the effect of different modeling choices, and to illustrate the estimation accuracy possible using the overall approach. Given the breadth of combinations possible the results presented here are at best illustrative rather than definitive, but this is consistent with the purpose of this study which is to demonstrate what is possible without claiming to have identified what is demonstratively optimal.

Because the specimen data assembled for this study were not uniformly distributed across many of the key dimensions considered, any single error measure is limited in its conclusiveness, especially in an absolute sense. Therefore, multiple standard quality-of-fit measures were used to compare the relative performance of the estimation model variations via various accuracy metrics describing how well the estimations fit the true data. The specific regression metrics used are described in more detail below [56] [50]:

**Normalized root-mean-square error (NRMSE)** : normalized prediction error representing the difference between predicted and true variables defined as follows:

$$NRMSE = \frac{\sqrt{\frac{\sum_{i=1}^n (y_i - y'_i)^2}{n}}}{y_{max} - y_{min}} \quad (3.4)$$

in which  $y'_i$  and  $y_i$  are the estimated output and the corresponding true values, respectively, and  $n$  is the total number of samples.

**Confidence interval (CI)** : estimate of an interval likely to contain future predictions with a certain confidence level if the prediction is repeated a large number of times. For example, the 95% confidence level means the user can be 95% certain future predictions would lie in the interval associated with the 95 % confidence level. Tighter confidence intervals in relation to the fit line indicate improved quality of the predictive model.

**Correlation Coefficient (R)** : measures the strength and direction of a linear relation-

ship between two variables. A correlation of 1.0 indicates a perfect one-to-one linear relationship.  $R$  is defined as:

$$R = 1 - \frac{\sum_{i=1}^n (y_i - \bar{y}_i)(y'_i - \bar{y}'_i)}{\sqrt{\sum_{i=1}^n (y_i - \bar{y}_i)^2 \sum_{j=1}^n (y'_j - \bar{y}'_j)^2}} \quad (3.5)$$

in which  $y'_i$  and  $y_i$  are the estimated output and the corresponding true values,  $\bar{y}'_i$  and  $\bar{y}_i$  are the averages of estimated and true values, respectively, and  $n$  is the total number of samples.

**Index of Agreement (IA)** : a non-dimensional and bounded measure proposed in [79].

IA values closer to 1.0 indicate better agreement between predicted and actual data sets. Unlike the correlation-based metrics such as the correlation coefficient which are sensitive to outliers, IA is less sensitive to extreme values [50]. IA is defined as:

$$IA = 1 - \frac{\sum_{j=1}^n (y_i - y'_i)^2}{\sum_{j=1}^n (|y_i - \bar{y}_i| + |y'_i - \bar{y}'_i|)^2} \quad (3.6)$$

in which the same notation is used as in equation (3.5) for  $R$  above.

**Explained Variance Score** : the ratio of residual variance to the variance of the true data

set. Similar to IA, scores closer to 1.0 indicate a better correlation between predicted and true values. This quantity can be computed as:

$$\text{Explained Variance} = 1 - \frac{\text{Var}(y - y')}{\text{Var}(y)} = 1 - \frac{\sum_{j=1}^n (y_i - y'_i - \bar{y}_i + \bar{y}'_i)^2}{\sum_{j=1}^n (y_i - \bar{y}_i)^2} \quad (3.7)$$

in which the same notation is again used.

In general the various error measures were consistent with each other in terms of degree of variation and ordering of method quality, showing similar trends throughout. This will be seen in the results presented below.

### 3.7 Results

The results presented in this section are intended to show the effects of varying major modeling components and to demonstrate the overall accuracy of the method for estimating failure ratios and current shear and moment values. The major modeling components considered include the scaling technique used to standardize the image formats (see Figure 3.3), the sampling strategy (image versus specimen level), the machine learning algorithms, the image feature sets, and the externally observable specimen dimensions ( $a/d$  and  $d$ ). Some additional results are presented to illustrate how predictions can be improved in some circumstances if internal properties are known, such as  $f'_c$ . In general results are presented for representative subsets of the full set of analyses performed, with an emphasis on the most accurate approaches. For example, results are only presented for Gaussian Process, SVM, and DNN machine learning algorithms, which outperformed techniques more suited to classification, such as J48, random forest, etc.

As mentioned earlier, LOO-CV is a particular case of K-fold cross validation in which the number of folds,  $K$ , is equal to the number of samples ( $n$ ), meaning for each fold the test data set would be a single observation, and the remaining observations are put together to form the training set. The process cycles through each individual observation, so it is more computationally expensive than something like 10-fold CV. Because in LOO-CV the training data set is being generated from a larger portion of the samples compared to 10-fold CV, this can improve the performance of the predictive models. However, this improved performance can be partially associated with overfitting due to the similarity of training sets, and it is known that LOO-CV can produce higher variance results than 10-fold CV when new data are considered [45]. The current study uses LOO-CV initially on a subset of the data to investigate the effect of different feature sets and image scaling techniques, and then uses 10-fold CV for the more comprehensive consideration of the full data sets.

The majority of the results to be presented focus on estimation/prediction of how close a specimen is to failure based on captured images. This is quantified in terms of the Because



these were all shear critical specimens, strictly speaking  $V_{\text{failure}}$  is the shear that caused the failure, while  $M_{\text{failure}}$  is the corresponding moment at failure, but for the simple loading configurations used in these tests the quantities are coupled such that failure ratios are the same. As mentioned earlier, the practical use of such ratios would be to be able to estimate for a given state of damage how close to true capacity the member was or is when subjected to the damaging loading (e.g.,  $V/V_{\text{failure}} = 0.6$  means a given load took the member to 60% of its capacity).

Figures 3.10 and 3.11 compare the load-level estimation performance of models generated using the two image transformation/standardization approaches described earlier, and evaluated using LOO-CV on Data Set 1 [9]. Each subfigure shows predicted (estimated) load level ratios for each load-stage image plotted against the true load level ratio known from the experimental data. Perfect prediction would lie along the 45° line. In this and in the following figures the estimation trend line and 95% confidence interval lines are shown to help indicate the prediction accuracy. Error measures are also provided (see Figure 3.11 in this case) to aid in comparing methods and algorithms.

The geometric feature sets used in Figures 3.10 and 3.11 represent a small subset of the full set of features presented earlier in this study. The polar moment of area of the image pixels,  $I_p$ , total crack length, and total number of cracks were found through testing to be dominant features for this case, and so the estimations are based on these reduced feature sets. The effect of feature sets will be considered in more detail below.

The images in Data Set 1 were used for this part of the study because unlike the other data sets, the starting point was raw captured images that were subjected to the full image preprocessing pipeline. The Gaussian process technique was found to be more effective to perform the estimations for this particular data set among the other aforementioned techniques. In comparison with other machine learning techniques, the Gaussian process technique has the advantage of learning the kernel and regularizing the average noise parameter [36]. The results shown throughout this study used Gaussian parameters of 10% noise and the Pearson universal kernel [74]. As indicated in Figure 3.11(a)–(b), the Gaus-

sian model generated using the rescaling image standardization approach (method 2) shows a better performance compared to the simple image embedding approach (method 1) for a given crack feature set. However, the performance of Gaussian models trained with the crack features and additional geometric properties is similar for both standardization approaches as shown in Figure 3.11(c)–(d). This indicates that adding structural geometric properties such as span-to-depth ratio,  $a/d$ , and depth,  $d$ , to the feature set can significantly improve the performance of the Gaussian process models for both standardization approaches. This is not surprising given that both standardization approaches remove true physical dimension information. Because method 2 consistently led to better results as shown here and confirmed in numerous other cases, this was the preferred approach, and it is used in all subsequent results presented here.

As mentioned earlier, the collected data sets provide a challenging breadth of geometry and image dimensions. For example, the overall range of specimen width and length are 12.2 cm to 300 cm and 108 cm to 2380 cm as shown in Table 3.1. This also favors the rescaling approach for standardization, which was found to be more robust with respect to working with dimensionless images.

Figure 3.12 presents results comparing image-level versus specimen-level sampling for the full set of images using 10-fold CV as discussed earlier. The machine learning algorithm in this case is SVM and all image crack features are used (this does not include specimen geometric dimensions  $a/d$  and  $d$ ). Figure 3.12 shows the overall superior performance of the image-level sampling models to estimate the failure ratio compared to specimen-level sampling. This is consistent with the fact that image-level sampling enables bracketing of a given image with similar images from the same load sequence from the same specimen, which enables better fitting to be achieved. However, the results of failure ratio estimation based on specimen-level sampling is a better indicator of the accuracy achievable in potential field application scenarios in which the test and training specimen sets would be disjoint.

Figure 3.13 compares the effects of using different feature sets and structural geometric properties on the performance of the models for the full set of images using 10-fold CV at

specimen level. As indicated in Figure 3.13(a), (b), and (d), adding structural geometric properties such as span-to-depth ratio,  $a/d$ , and depth,  $d$ , to the feature set can improve the performance of the SVM models. This is an expected result given that size effect is a key factor in determining the shear behavior of the concrete beams. As shown in Figure 3.13(b), (c), and (d), the performance of the SVM model using all crack attributes and geometric properties (b) is similar to the SVM model using the dominant features. These features are total major axis,  $I_p$ , and number of cracks along with structural geometric properties ( $a/d$ , and  $d$ ) which were selected based on the dominant features found earlier in model selection part for Data Set 1.

Figure 3.14 presents results comparing different machine learning techniques such as Gaussian process, SVM, and DNN. In order to compare different algorithms consistently, the dominant feature set found in the previous results are used for all models. As indicated in Figure 3.14, the DNN model showed slightly better performance compared to SVM, with each outperforming the Gaussian process model in this case. It should be noted the topology of the network and other parameters in a deep neural network model need be tuned to obtain the best performance for different inputs/outputs (i.e., different feature sets and output variables) which makes this technique more complex and open-ended to use compared to traditional machine learning techniques such as SVM. In addition, DNN is more computationally expensive compared to traditional machine learning techniques.

The results presented so far focused on estimating failure ratios based solely on the external observations (crack pattern features and geometric properties of the beam). The same framework can also be used to estimate current shear and moment values in the beams. Figure 3.15(a) indicates the internal shear estimation result using the same feature set found to be dominant for failure ratio prediction. As can be seen, the DNN model failed to estimate internal shear force based solely on the external observations. It makes sense that external damage alone is difficult to relate to absolute load values without some additional measure of strength. In the present case knowing some internal properties such as  $f'_c$  can enable

shear load estimations if scaled shear values (e.g.,  $\frac{V}{2\sqrt{f'_c} bd}$ ) are predicted as shown in the Figure 3.15(b). Therefore, the internal shear values can be estimated indirectly through predicted scaled shear values (i.e., shear is determined by multiplying the scaled shear by  $2\sqrt{f'_c} bd$ ) as illustrated in the Figure 3.15(c). Similar to internal shear force, internal moment can also be predicted with comparable accuracy if  $f'_c$  is known (i.e., moment is determined by multiplying the shear in the beam by the shear span for the simple loading configurations considered here).

### 3.8 Summary and Conclusions

A computer vision based inspection framework for concrete members has been developed and tested using image data assembled from previously published experimental studies with specimens covering a wide range of geometries, loadings, concrete strengths, and reinforcement details. The images were captured at various load and damage levels, and so the resulting framework is capable of translating simple photographic data into estimated levels of damage calibrated relative to the component's actual capacity. Estimating absolute rather than relative load levels was more difficult, with knowledge of  $f'_c$  being needed to obtain consistent results. For the applications envisioned here, the utility of absolute load level estimation would be limited compared to relative load levels (i.e., relative to actual capacities), so this is not considered a critical issue.

The key aspects of the approach presented here include:

- Definition of a simple and uniform procedure to build new image data sets from available data sources such as existing video clips or crack pattern images for future studies.
- Definition of a simple, yet effective image transformation process to take account of the different specimen and image dimensions.
- Use of simplified crack representations (i.e., only crack patterns without any information about crack width) for damage and load levels estimation.

- Identification of effective and dominant geometric and textural crack features for different prediction tasks.
- Characterization of different machine learning techniques for prediction tasks.
- Accommodation for the size effect phenomenon via inclusion of  $a/d$  and  $d$  in model feature sets.
- Imposing specimen-based cross validation for evaluating the performance of the models to correspond to potential field scenarios.

As in any early study of this type, there also are a number of caveats and limitations for the work presented here, including:

- The data set used in this study was relatively small by data analytics standards, and as is common in concrete testing the specimen size distribution was skewed towards small and medium specimen dimensions.
- The features used to characterize the image data and the learning algorithms used to build the estimation models were shown to be effective, but this does not mean they are necessarily optimal.
- The estimation models were built using data corresponding to shear-critical simply supported beams and slabs only. Extension to other classes of structural systems would require appropriately expanded training data sets.
- The image standardization process used in this work removed damage features such as crack width and spalling, which are known to be important in assessing damage. Expanded access to raw image data collected during testing would make it possible to increase the generality of the standardized image content, which might improve overall estimation accuracy.

Despite these limitations, the results show that the approach is capable of estimating internal damage levels with quantifiable accuracy and reliability based purely on externally observable data. This indicates the procedure has the potential for use in a variety of field inspection scenarios in which low-cost and rapid inspection to assess the structural condition of a visibly damaged component would be useful.

In the interest of ensuring reproducibility, the authors are happy to share the data sets and code used in this work. The full data sets and code will be uploaded to: [github.com/Rouzbehe/Image-Based-Load-Estimation/](https://github.com/Rouzbehe/Image-Based-Load-Estimation/) for academic use.

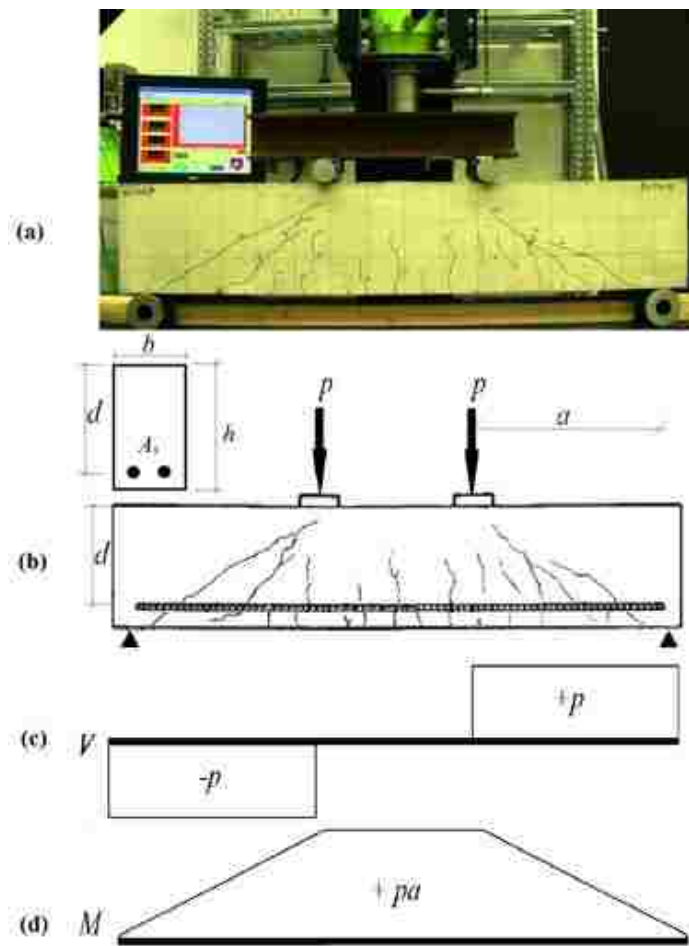


Figure 3.1: (a) Typical four point load test setup. (b) Specimen detail. (c) Shear diagram. (d) Moment diagram.

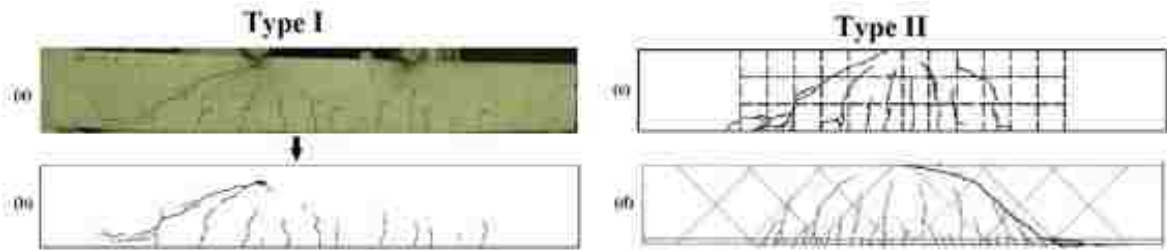


Figure 3.2: Image data segmentation for two different types of data. Type I: Typical raw image (a) segmented as shown in (b). Type II: pre-segmented crack pattern images in high resolution (c) or low resolution (d) form.

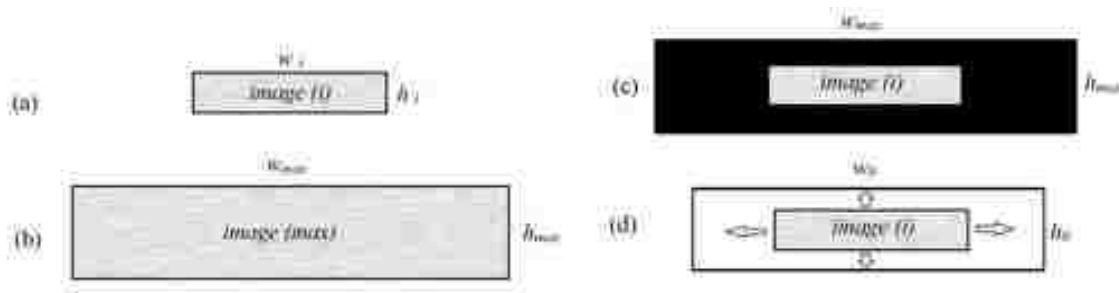


Figure 3.3: (a) Typical raw image,  $i$ , with dimensions  $w_i \times h_i$ . (b) The largest image in the data set with dimensions  $w_{max} \times h_{max}$ . (c) Image standardization based on simple embedding in the largest image: the typical image,  $i$ , is centered in an image frame with the size of  $w_{max} \times h_{max}$ , with extra non-crack pixels added as needed. (d) Image standardization based on re-scaling to a common image size: image height and width are scaled to consistent pixel dimensions (i.e.,  $w_0 \times h_0$ ).

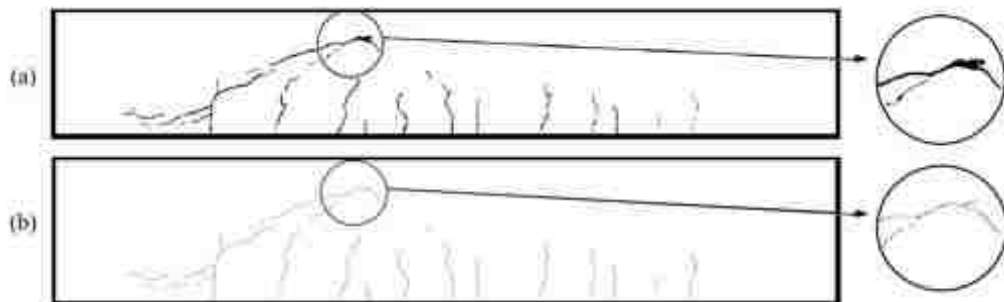


Figure 3.4: (a) Typical raw scaled segmented photo, with segmented cracks having a varying width. (b) The segmented image in (a) is skeletonized using a morphological process so that all cracks have uniform thickness.



Table 3.1: Summary table of prior experimental study data sets. See Figure 1 for definitions of the tabulated geometric parameters.

Reference	No. Specimens	No. images	Test type	Width (cm)	Length (cm)	$a/d$	$d$ (cm)	$\rho$ (%)	$f'_c$ (MPa)
[82]	1	4	3-pt load, deep beam	30	1150	2.9	189	0.7	31.8
[16]	2	12	3-pt load, deep beam	30	1079.5–1121.9	2.8–2.9	192	0.4–1.5	26.2–28.3
[71]	8	52	3-pt load, deep beam	22.9–57.15	198–495.3	2.3	26.7–66.8	0.55–0.85	18.6–32.4
[70]	30	198	3-pt load, beam and slab	12.2–300.5	108–810	2.79–3.4	28–145	0.3–1.33	29.1–77.3
[58]	8	88	3-pt load, beam	20.3–61	139.7–493	2.97–3	23.1–81.2	1.2–1.3	64.8–74.8
[9]	15	42	4-pt load, beam	20	125	1.1–4.81	8.1–30.74	1.3–1.55	31.4
[62]	6	42	uniform, beam	30	358.1–599.4	—	92.5	0.98	39–64
[59]	3	28	3-pt load, haunched beam	19.8	180–299.7	3–5.0	30	1.57–3.1	35.4–59.1
[57]	10	82	3-pt load, beam	22.9–57.4	197.9–609.7	2.3–2.9	26.9–67	0.63–0.98	22.8–33.8
[66]	1	10	3-pt load, deep beam	24.9	2380	3.1	384	0.7	40.0
Overall	84	558		12.2–300.5	108–2380	1.1–5.0	8.1–384	0.3–3.1	18.6–77.3

Table 3.2: Summary of individual crack features computed from crack images.

Individual Crack Features, Averaged	
Major axis length	Major axis of crack elliptical bounding envelope
Aspect ratio	Ratio of major axis length over minor axis length of the elliptical crack envelope
Perimeter	Outline envelope of the crack
Compactness	Perimeter divided by crack area (crack length and crack area are essentially equivalent in the present case due to the morphological transformations described earlier)
Individual Crack Features, Totaled	
Total major axis	Sum of individual crack major axis lengths
Total crack length	Sum of individual crack lengths
Total perimeter	Sum of individual crack perimeters

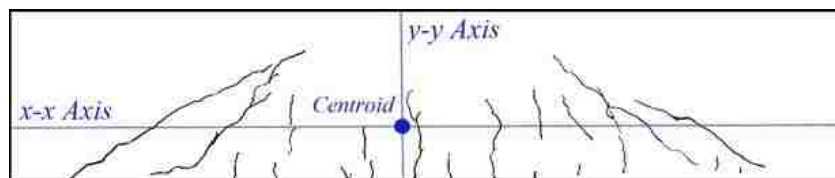
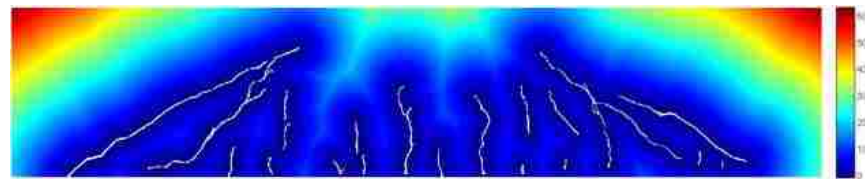


Figure 3.5: Schematic showing axes used to compute moments of area for the crack images.

Table 3.3: Summary of overall image features computed using built-in MATLAB functions.

Thresh-out	threshold value of the crack image returned by the edge function in the image processing toolbox. This scalar indicates the relative proportion of the cracks to non-crack objects
Entropy	scalar value representing entropy of crack images
Contrast	scalar value characterizing the texture of the image using the intensity of the contrast between pixels over the entire image
Correlation	textural attribute which identifies correlation between neighboring pixels over the entire image
Energy	textural attribute calculated by sum of squared elements in the gray-level co-occurrence matrix
Homogeneity	textural attribute between 0 and 1 representing the distribution of elements in the gray-level co-occurrence matrix
Local variance	average local standard deviation of the $3 \times 3$ neighborhood around each pixel in the image
Euler number	scalar value corresponding to the total number of objects in the image minus the total number of holes in those objects
2D standard deviation	standard deviation of matrix elements calculated by the std2 function in MATLAB using the pixel values within the image



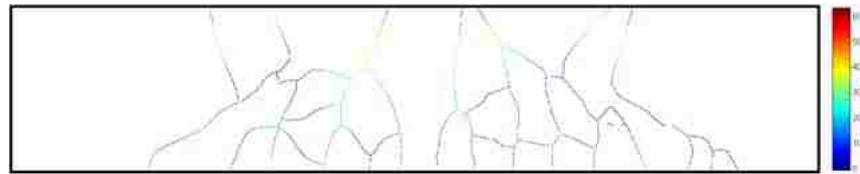
(a) Distance transform of the segmented image.



(b) Watershed transform of the distance transform.

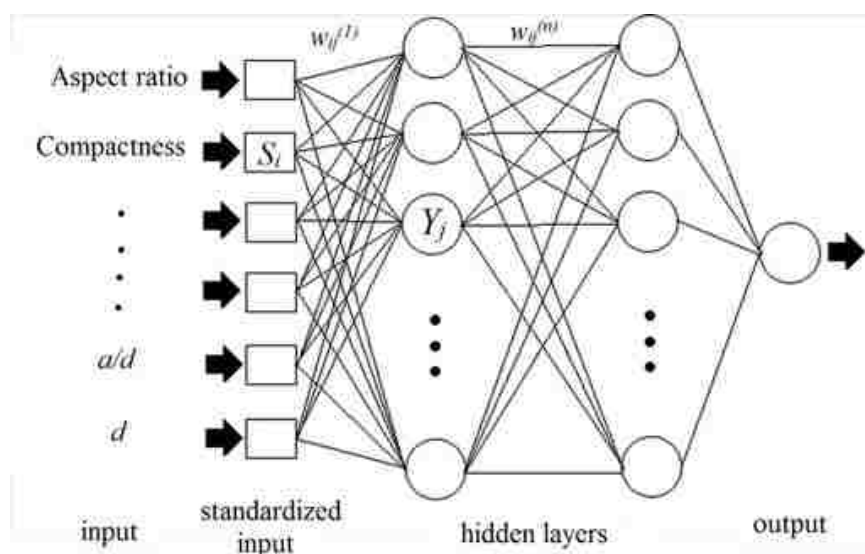


(c) The watershed regions and cracks.

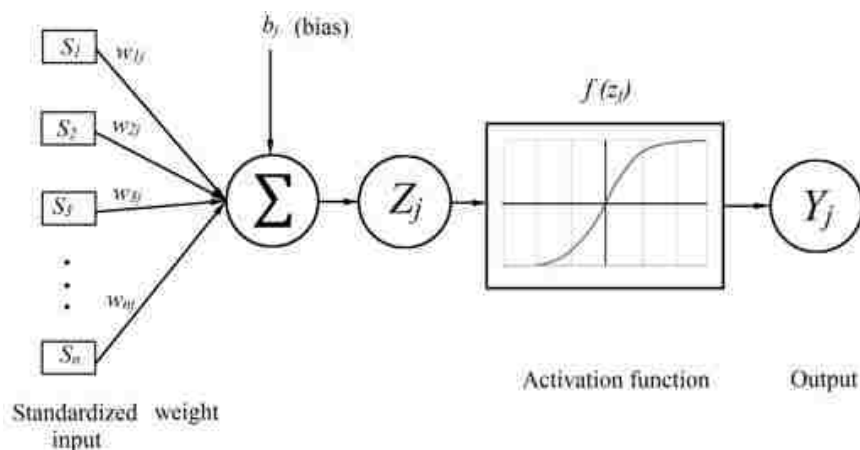


(d) Distance values at “watershed ridge lines”.

Figure 3.6: (a) Distance transform of a typical segmented image (see Figure 5); (b) Watershed transform of the distance transform; (c) The watershed regions used to calculate average distance between cracks; (d) Distance values in the distance transform image along watershed ridge lines.



(a) Deep Neural Network topology

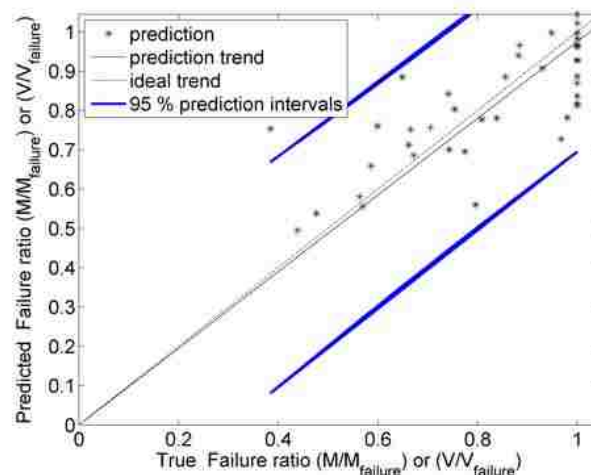
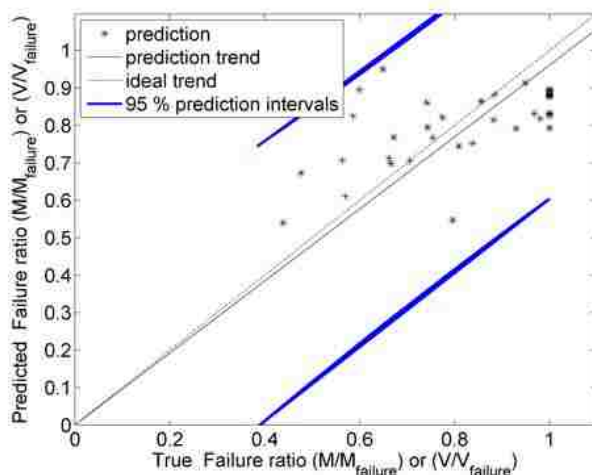


(b) Typical Neuron in a hidden layer

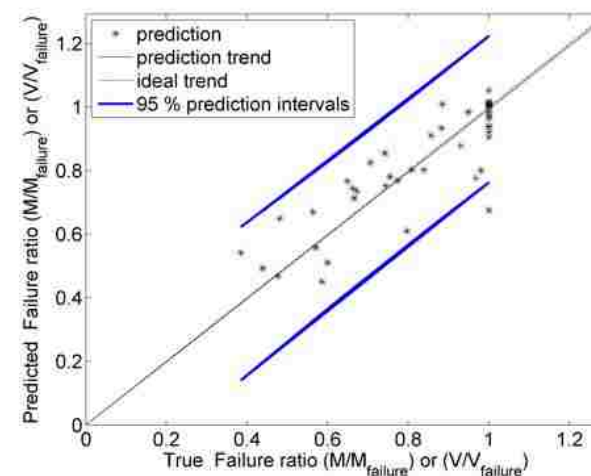
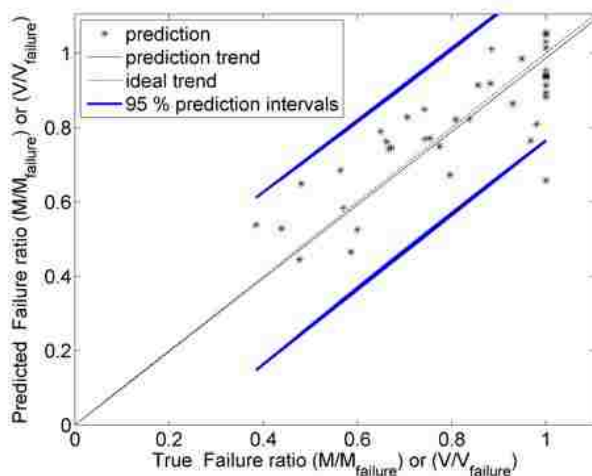
Figure 3.7: (a) Deep Neural Network topology. (b) Typical Neuron in a hidden layer.







(a) Image transformation method 1-Gaussian Process-  $I_p$ - Total crack length  
 (b) Image transformation method 2-Gaussian Process-  $I_p$ - Total crack length



(c) Image transformation method 1-Gaussian Process-  $I_p$ - Number of cracks- $a/d-d$   
 (d) Image transformation method 2-Gaussian Process-  $I_p$ - Number of cracks- $a/d-d$

Figure 3.10: Performance of the failure ratio estimation models on Data Set 1 generated using image standardization methods 1 (embedding) and 2 (rescaling) using different feature sets.



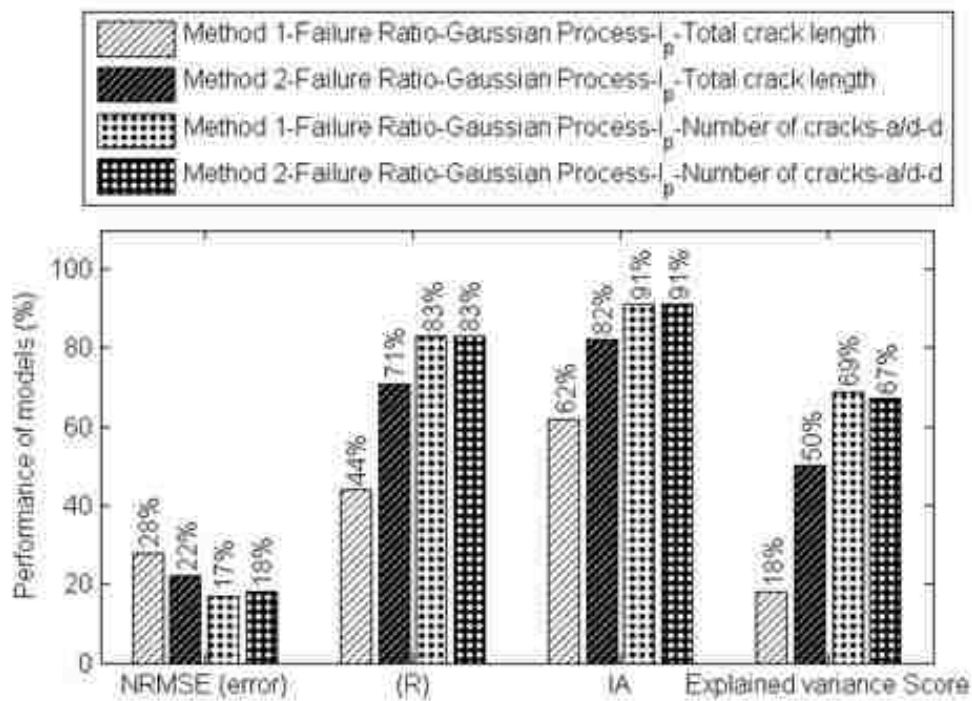
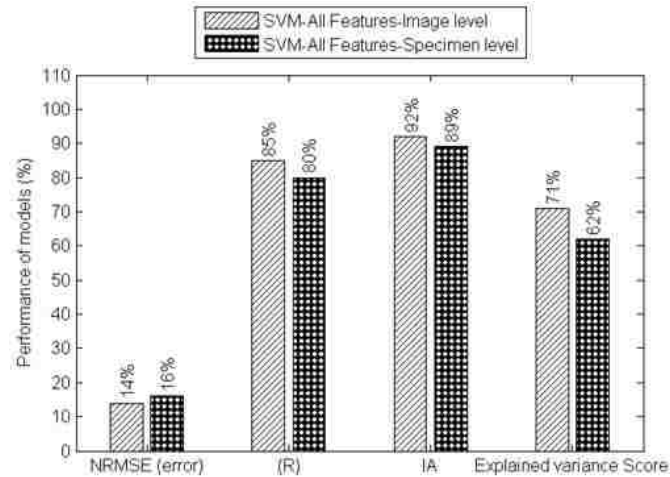
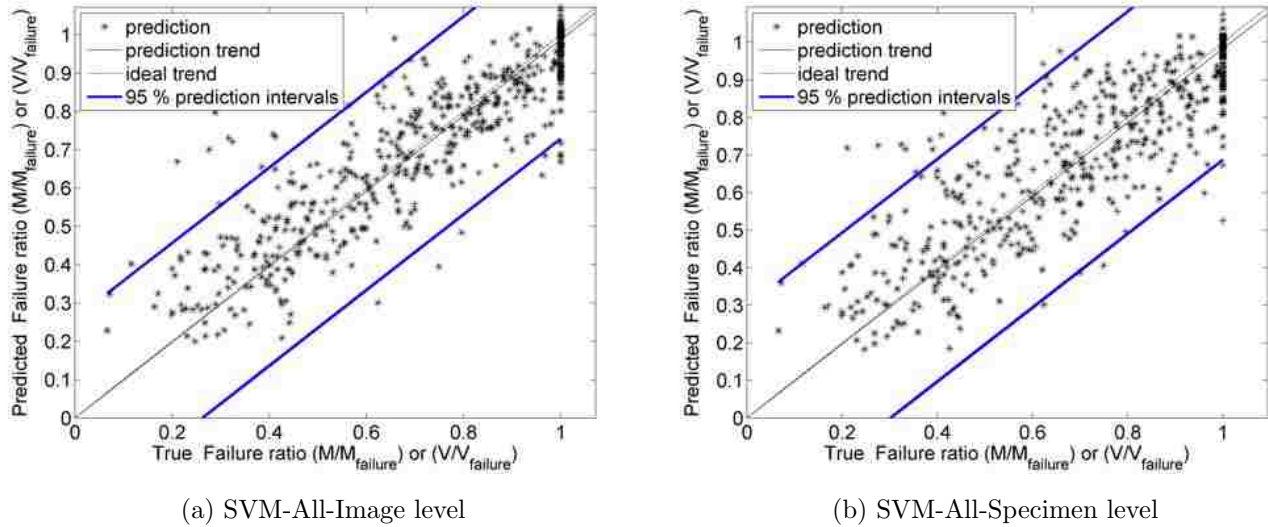
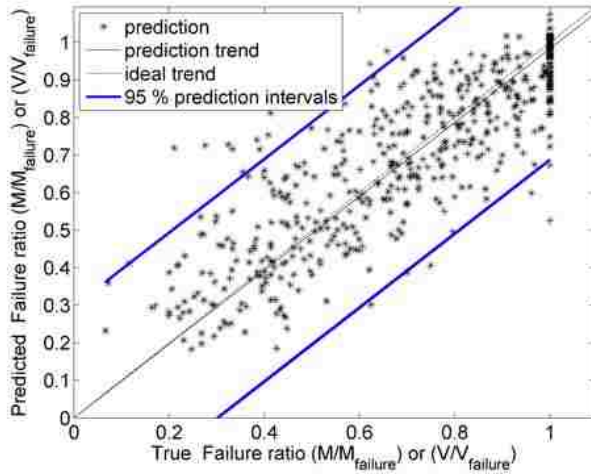


Figure 3.11: Comparison of the failure ratio estimation models using Gaussian technique on Data Set 1 generated using scaling method 1 and 2.

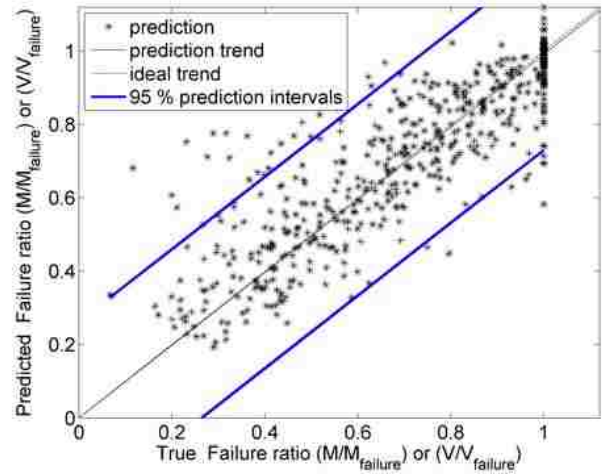


(c) Performance of the models

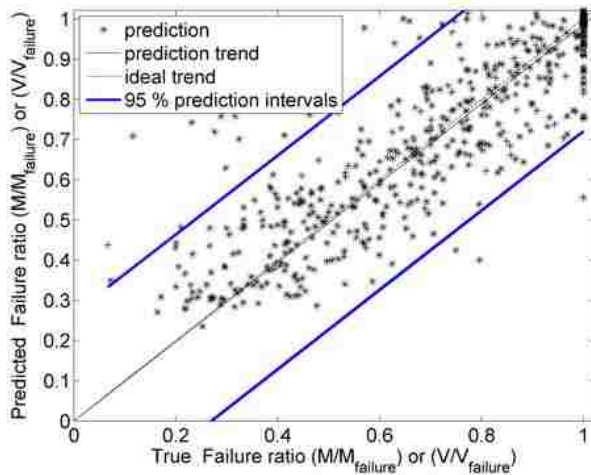
Figure 3.12: Comparison of SVM models to estimate failure ratio using 10-fold cross validation at image and specimen level.



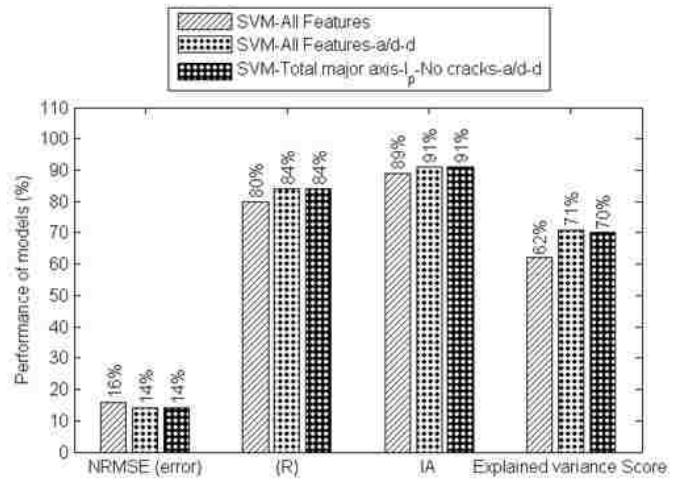
(a) SVM-All



(b) SVM-All-a/d-d

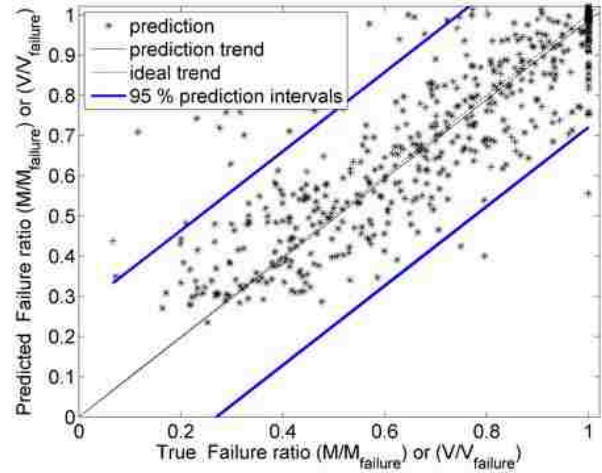
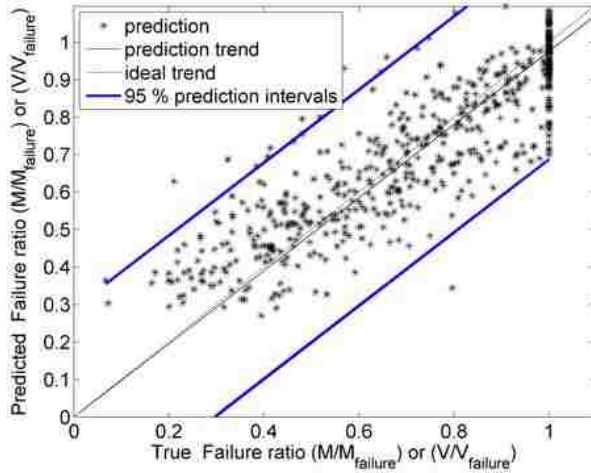


(c) SVM-Total major axis- $I_p$ -Number of cracks-a/d-d

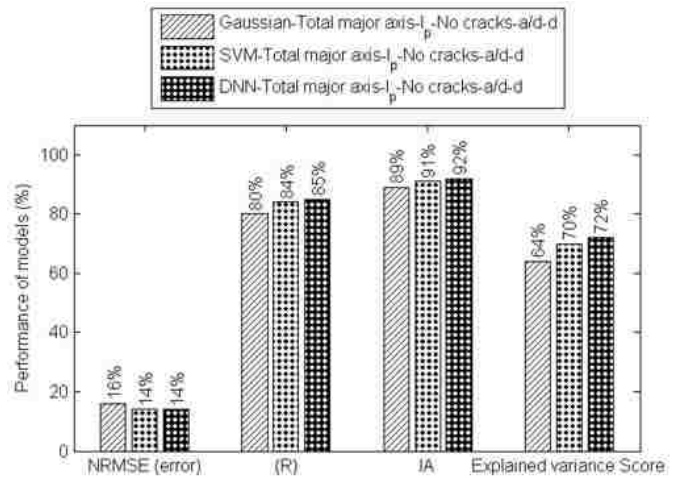
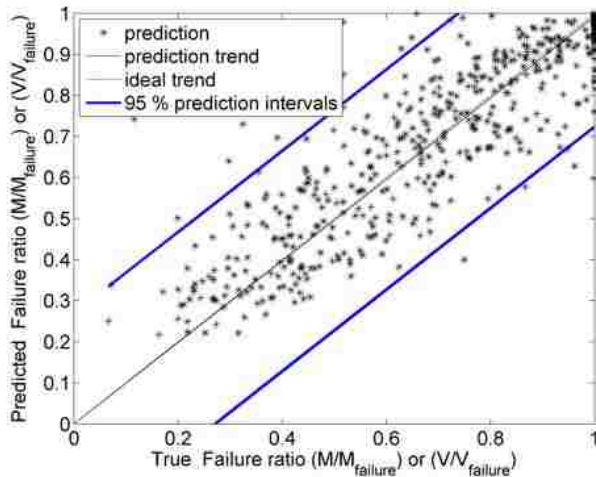


(d) Performance of the models

Figure 3.13: Comparison of the SVM regression models trained with different feature sets to estimate failure ratio using 10-fold cross validation at specimen level.



(a) Gaussian-Total major axis- $I_p$ -Number of cracks- $a/d-d$  (b) SVM-Total major axis- $I_p$ -Number of cracks- $a/d-d$

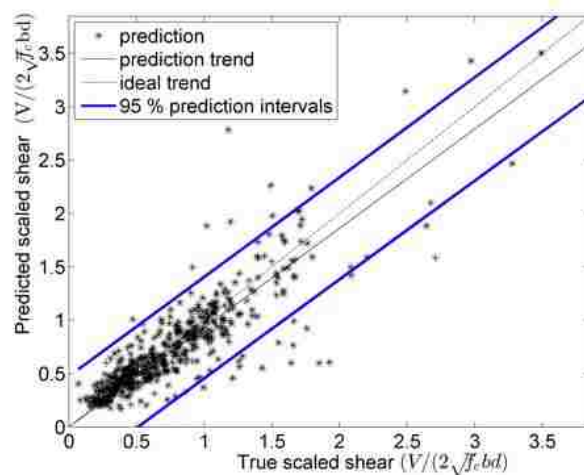
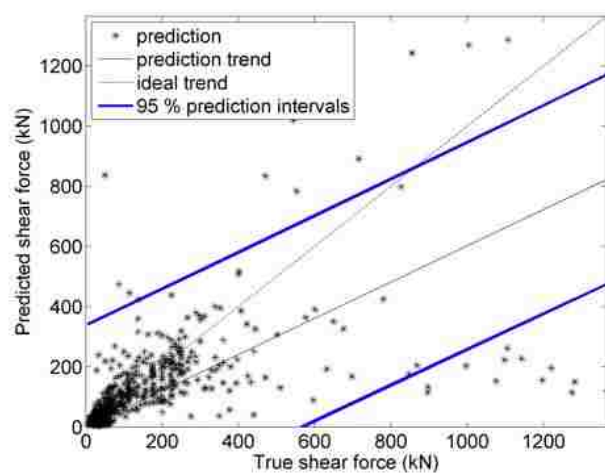


(c) DNN-Total major axis- $I_p$ -Number of cracks- $a/d-d$

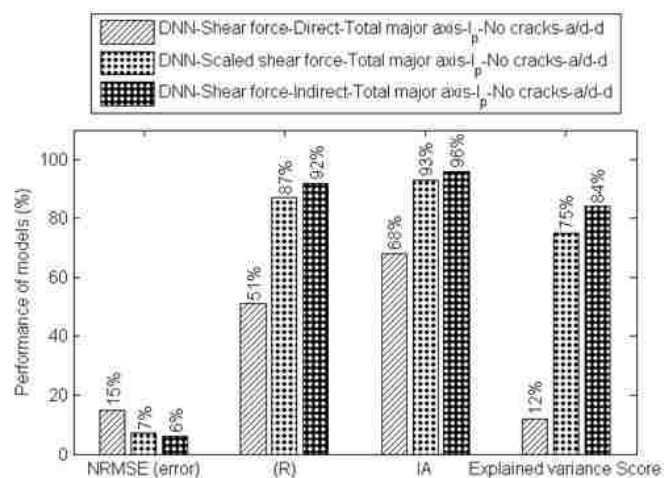
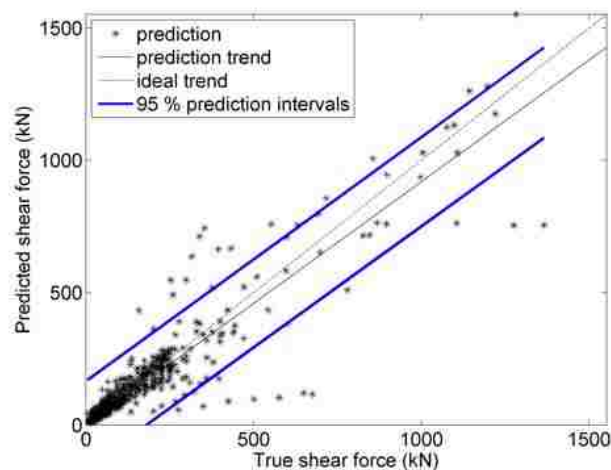
(d) Performance of the models

Figure 3.14: Comparison of the different regression models trained with the same feature set to estimate failure ratio using 10-fold cross validation at specimen level.





(a) DNN-Direct-Total major axis- $I_p$ -Number of cracks- $a/d-d$  (b) DNN-Total major axis- $I_p$ -Number of cracks- $a/d-d$



(c) DNN-Indirect-Total major axis- $I_p$ -Number of cracks- $a/d-d$

(d) Performance of the models

Figure 3.15: Performance of the models to estimate internal shear force directly using the feature set and indirectly through the estimated scaled shear force (figure b).

Chapter 4

**COMPUTER VISION-BASED DAMAGE INSPECTION  
FRAMEWORK FOR RC BEAMS AND SLABS**

#### **4.1 abstract**

We show that computer-vision-based inspection can relate surface observations to quantitative damage and load level estimates in common reinforced concrete beams and slabs subjected to monotonic loading. This work is related to an earlier study focused on shear-critical beams and slabs (i.e., specimens lacking shear reinforcement), but here an expanded image database has been assembled to include specimens with both flexural and shear reinforcement such as would be found in practice. Using this expanded dataset, a supervised machine learning algorithm builds cross-validated predictive models capable of estimating internal loads (i.e., shear and moment) and damage levels based on surface crack pattern images. The expanded data set contains a total of 127 specimens and 862 images captured in past studies across a range of load and damage levels. Textural and geometric attributes of surface crack patterns were used for feature engineering and tuning of predictive models. The expanded data set enables comparison of the estimation accuracy for shear-critical and shear-reinforced beams and slabs considered separately and in combined form. This includes the capability to categorize whether shear reinforcement is present or not. Estimation models based on surface observations for shear-reinforced elements are found to be comparable to those for shear-critical beams and slabs, with variability observed due to loading type range, member geometries, whether categorization is combined with regression, and the image feature sets used.

#### **4.2 Introduction and Motivation**

Infrastructure inspection is an important task in various contexts, including structural condition monitoring, post-disaster damage assessment, and commissioning of new systems. The outcomes of such inspections include decisions regarding repairing or replacing damaged systems, predicting remaining useful life, evacuating building occupants, or opening new facilities to full public use. Research focused on augmenting existing inspection procedures with computer-vision-based tools has been underway in different contexts for many years. The work in this paper is part of series of studies focused specifically on characterizing the feasibility of quantitative estimation of load and damage levels based on images of surface crack patterns—that is, determining the degree to which computer vision techniques can successfully relate surface observations to quantitative load level estimates in structural components.

For any image-based method of this type to work, there must be visible (but not necessarily unique) correlations between observed behavior and the underlying mechanics of the structural systems in question. This is similar to the expectation for any instrumented system, be it an accelerometer on a bridge or a strain gauge on a beam. For cracks in concrete beams and slabs subjected to standard loadings, the characteristic crack patterns observed under increasing load are indeed reflective of the stresses and strains induced in the system, with apparent distinctions between, for example, shear and flexural failures, and with clear

correlations between load levels and degree of cracking. The existence of such qualitative correlations between mechanical behavior and observable damage leads to the possibility this could be extended to the level of quantitative correlation using computer vision technology. As in any data-driven evaluation approach with major empirical components, sufficiently comprehensive data sets would be needed to bracket the parameter ranges arising in targeted applications, and extrapolation beyond the targeted applications would not be viable. Similarly, correlation accuracy and reliability could be improved by including additional structural information beyond that which could be measured directly in situ. Within these limitations, the results presented in this paper demonstrate that potentially useful quantitative correlations exist between crack pattern images and load and damage levels for monotonically-loaded, simply-supported beams and one-way slabs.

Figure 4.1 illustrates conceptually how such capabilities potentially could be used in a typical field inspection scenario requiring damage and condition assessment. For the conceptual application shown, a cell-phone is used to capture an image of a damaged structural component, and the image is then processed such that the crack pattern is highlighted and extracted (segmented). The segmented crack pattern images are evaluated using a damage estimation model that has been trained and cross-validated on a large data set drawn from experimental studies in which images have been captured at known load levels for similar structural systems or components. The captured image thus could be converted into an estimate of the load or damage level (expressed in terms of percentage capacity and absolute shear and moment values in the figure, including confidence intervals), which could then guide further decision-making. The key question in making such a system functional concerns the degree to which purely external imaging of structural members can be used to make quantitative estimates of the load levels associated with visible damage, and that is the overall context motivating the present work.

This present study adapts the computer-vision-based inspection approach from an earlier study focused on reinforced concrete beams and slabs without shear reinforcement [26], to the case of beams and slabs with standard shear reinforcement. As in the earlier study, the estimation scenario is based on surface crack pattern images such as one would be able to obtain in field applications like the conceptual scenario in Figure 4.1. In particular, only externally visible data are assumed available (i.e., quantities like shear reinforcement spacing, concrete compressive strength, etc., are taken as unknown). For the present paper additional image data sets were collected from seven previously published experimental laboratory studies, which provided an additional 304 crack pattern images captured from 43 RC beams and slab specimens across a range of load and damage levels. These additional data augmented the 558 images from 84 specimens used in the prior study, giving a combined data set suitable for the study of prediction across a wide range of failure modes. As before, the images captured correspond to known load levels (supervised learning) recorded



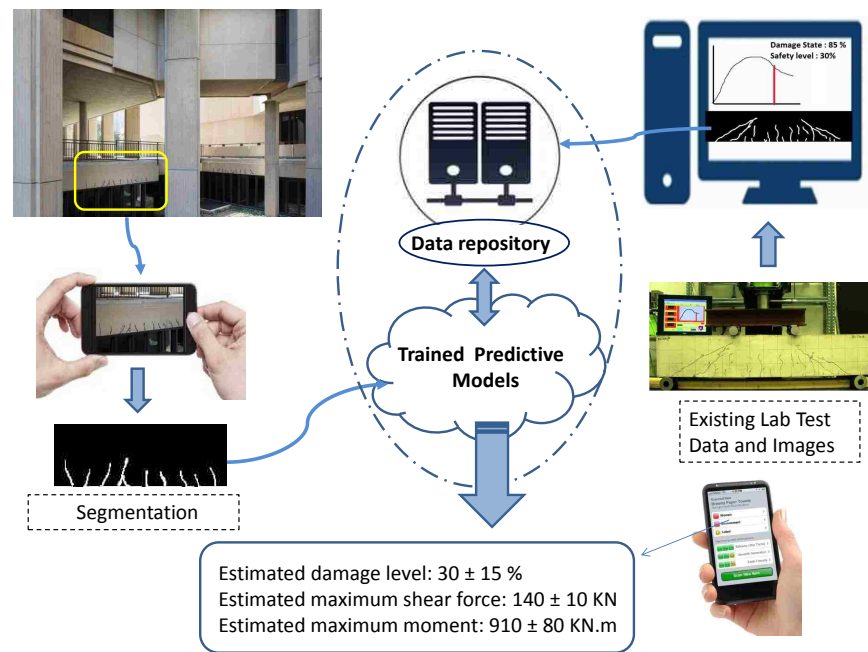


Figure 4.1: Conceptual image-based inspection framework. Existing laboratory tests are used to train an estimation model that correlates the surface image of the structure to known load and damage levels. The predictive model could then in principle be used in field applications such that captured photos could lead to statistical damage estimates.

during the testing, and the ultimate capacity of each specimen was also determined during the tests, which makes it possible to correlate image features to load levels, both absolute and relative to ultimate capacity (e.g., current moment over failure moment,  $M/M_{\text{failure}}$ ).

Our previous study [26] demonstrated the basic feasibility of estimating quantitative damage and load levels based on surface image data, and considered fundamental issues arising in the development and evaluation of such models. This included studying the effects of different image standardization techniques and machine learning algorithms on estimation accuracy, identifying appropriate cross-validation approaches consistent with field application scenarios, and developing effective image feature sets. In addition, a suite of error measures was assembled to provide a consistent means for comparing estimation accuracy between different models. The results of this previous study have thus guided the present work in terms of selection of algorithms, image feature sets, image standardization, cross-validation techniques, and error assessment metrics.

The inclusion of shear reinforcement in beams and slabs results in flexural-shear and flexural failure modes, which gives rise to different crack patterns than those observed in shear critical cases. It is thus

necessary to develop additional trained and calibrated models for data sets containing these classes of failures, but the overall framework is similar to that used before: images are transformed into a standard format based on uniform-width, segmented binary crack images; geometric image features are extracted for each image and fed into a machine learning process combining systematic identification of training and testing sets to support cross validation; and the estimation accuracy of the models is analyzed.

Because the full data set used in this study includes both shear-reinforced and shear-critical (members with no shear reinforcing) specimens, this makes it possible to address several related questions. First, the relative accuracy of estimation for each of these classes of specimens taken separately can be compared. Second, the two classes of specimen data can be used to investigate how well specimens can be classified into their corresponding shear reinforcement categories based on crack pattern image analysis. Finally, assuming differential estimation quality within each specimen category, it makes sense to study whether one can obtain more accurate estimates for the combined overall data set by first classifying, and then using regression methods on the distinct data classes.

The paper is organized as follows: first, a summary of related prior work is reviewed, after which the data sets used in the current study are presented and grouped according to characteristics of the underlying tests and specimens, with the primary delineation being the presence or absence of shear reinforcement. The next section briefly reviews the main ingredients of the machine vision framework used for the model development and evaluation. The remainder of the paper focuses on results obtained for different combinations of data groupings, considering estimation accuracy for relative and absolute load levels with an eye toward field application scenarios.

### ***4.3 Prior Work Applying Computer Vision for Damage Inspection***

There are many methods that have been applied to the detection of damage in structures using machine vision techniques. Digital Image correlation (DIC) is one popular damage detection method capable of measuring deformations in structural components ([27], [61]). This method is based on comparisons between a series of images captured in a constant camera placement condition. Augmented Reality (AR) is another approach that has been explored for post disaster damage detection ([43], [23]). Here, a computer-generated graphic can be superimposed on an undamaged view of the structure to detect permanent changes such as inter-story drift. There are also studies focused on analyzing the dynamic response of structures based on camera calibration. In one recent study, [84] determined displacement time history, natural frequency, and damping ratio based on camera based monitoring of the signal structures.

For reinforced concrete in particular, there has been an emphasis on binary damage/no-damage identification based on observation of concrete surfaces, and more general classifications of qualitative damage types

in structural components. Given its close ties to the broader field of machine vision, this remains a rapidly developing field, and representative work in this area can be found in [81], [27], [86], [87], [35], [30], and [28]. The techniques for identifying and isolating cracks, which forms the basis of our feature engineering, is a key component in the development of the type of system illustrated in Figure 4.1. For the specific work presented here, in all but one case the crack images had already been extracted as part of the reporting of the experimental work. Thus, the potentially complicating factor of assessing a crack isolation algorithm has been removed from the scope of the present work. It is also worth noting that from the perspective of damage quantification, in most instances the image data sets used in developing the crack identification/extraction techniques presented in references like those listed above cannot be adapted or extended to generate quantitative damage estimation models. This is because any such machine learning approach must have training data combining large numbers of images with corresponding known load levels and ultimate failure loads. Such combined data have rarely been captured (or would be capturable in the case of field-captured images) for purposes of these crack detection publications. Thus, existing image-only data would not be adequate for quantitative model development.

In terms of prior publications focused on quantitative damage characterization, [26] provides a discussion of the recent work most closely related to what is presented here. Additional works include those focused on machine-learning-based damage classification (see, e.g., [81], [27], [86], [87], [30], [28], and [35]) and those going beyond classification to more fine-grained quantification (e.g., [60] and [47]). Relative to these earlier studies, the primary distinguishing characteristics of the present work include the large data set used collected across multiple experimental studies performed at different labs, the broad range of beam and slab geometries and structural parameters considered, the focus on predicting key quantities which can be directly used for practical damage inspection tasks (i.e., internal load level and quantitative damage state), and the scaled-up application of error measures and cross validation.

#### ***4.4 Prior Experimental Studies and Image Data Sets***

This section presents a basic overview of the experimental data gathered from prior published studies, including both the shear-critical cases used in [26] and the expanded data used in this current study. The image data set in [26] was obtained from ten different studies, which provided 558 crack pattern images captured from 84 shear-critical RC beams and slabs (i.e., RC beams and slabs without shear reinforcement). For the current study, an additional image database has been assembled based on specimens containing standard flexural and shear reinforcement (stirrups), which distinguishes it from the earlier shear-unreinforced data. This new image data set was collected from previously published monotonic bending tests on RC beams and slabs with shear reinforcement. The results of these tests have been presented in seven earlier

publications: [52]; [10]; [64]; [5]; [17]; [83]; and [34]. These experimental tests were conducted for the purpose of studying various effects of structural and material details such as specimen's dimension, reinforcement details and concrete strength on the structural behavior of the specimen under different monotonic loading conditions.

Table 4.1 summarizes the specimen and test configuration data from the full set of experimental studies, including references to the studies themselves, ranges of key structural parameters, test type, and number of images used in each data set. These largely independent experimental tests were carried out for the purpose of studying various effects of transverse and longitudinal reinforcement ratio, concrete strength, and beam size on the ultimate shear strength of concrete beams and slabs, and so they are being repurposed in the present work. The first part of Table 4.1 contains the newly assembled data corresponding to the shear-reinforced specimen and tests, while the second part of the table contains data from the shear-critical tests considered in [26] (here the external depth  $h$  of the sections tested is used instead of the previously used effective depth,  $d$ , as will be discussed later). As can be seen from Table 4.1, a total of 304 images obtained from 43 specimens were added in this study, with the overall collected data covering orders of magnitude ranges of structural and geometric parameters, and with a variety of load patterns. Most the studies across the entire data set involved beams subjected to 3-point loading (notated B3 in the table), but there also were slab tests involving 3-point bending (S3 in the table) and uniform loading (SU), haunched beams with 3-point loading (H3), and beams with 4-point loading (B4). In addition to  $d$  and  $h$ , the physical quantities listed in Table 4.1 have their usual definitions:  $b$  is the cross-section width;  $L$  is the overall length of the specimen;  $a$  is the effective shear span;  $s$  is the shear reinforcement spacing;  $\rho$  is the reinforcement ratio given by  $A_s/bd$  in which  $A_s$  is the area of the flexural steel; the shear index is  $\rho_v f_y$  in which  $f_y$  is the steel yield strength and  $\rho_v = A_{sv}/bs$  with  $A_{sv}$  being the area of the shear reinforcement; and the concrete compressive strength is  $f'_c$ .

Taken together, the combined image data sets represent 17 different published studies, with a total of 862 crack pattern images captured from 127 RC beam and slab specimens. Figure 4.2 provides an overview of the various data sets, and also introduces a naming scheme to aid in comparing across and within the two primary data set groupings (shear reinforced and shear unreinforced). The complete set of shear unreinforced specimens (which was also used in [26]) is labeled as Data Set 1, the complete set of shear reinforced specimens form Data Set 2, the full combination of reinforced and unreinforced specimens is Data Set 3, Data Set 4 refers to the partial combination of Data Sets 1 and 2 in which only the 3-point bending test data common to both are included, and Data Set 5 corresponds to a subset of the shear unreinforced data with consistent overlap with the reinforced data. In particular, Data Set 5 is composed of the 3-point loading subset of shear unreinforced specimens with the additional removal of the 4-m deep specimens, since nothing close to this

scale was included in the shear reinforced data. This “Data Set  $n$ ” naming scheme will be used henceforth in this paper.

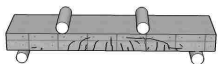
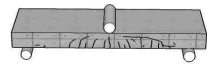
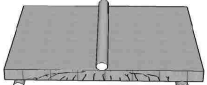
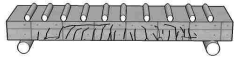

		Data Set 3	
Specimen/Load type	ID	No shear reinforcement (Data Set 1)	Shear reinforced (Data Set 2)
	B4	15	-
Data Set 4	 	52 8	37 6
	SU	6	-
	H3	3	-

Figure 4.2: Overview of number of specimens in each test category, including labeling scheme for data subset combinations. For the IDs, ‘B’ refers to beam specimens, ‘S’ refers to slabs, and ‘H’ identifies haunched beams, while ‘3’ and ‘4’ correspond to 3-point and 4-point loading, respectively, and ‘U’ means uniform loading.

The image data collected in all these data sets were captured at multiple load levels for each specimen. This makes it possible to investigate the correlations between the images and the loading, and in the previous study [26], it was shown there are sufficiently strong correlations between crack pattern features and damage levels in shear critical RC beams and slabs that one can build quantitative load estimation models with accuracy comparable to standard design formulas. The key question in this paper is the degree to which such correlations exist across and within combined data sets involving a range of shear-critical and flexure-critical behavior. Figure 4.3 illustrates the effect varying the distance between stirrups has on the crack pattern at maximum load in beams having otherwise identical geometry and properties. As can be seen, the most local cracking occurs in beam (a), which has the closest shear stirrup spacing (i.e., highest shear reinforcing ratio),

while beam (d), which has no shear reinforcing (shear reinforcing ratio of zero), has comparatively few, smaller cracks. These visually-evident qualitative differences indicate that image-based machine learning could be capable both of making quantitative load-level estimates and qualitative classification of failure modes.

Related to the qualitative observation of crack pattern variations as a function of shear reinforcement ratio, Figure 4.4 compares the crack patterns at maximum load for two specimens in which the shear reinforcement ratio was held constant, while the shear span to depth ratio ( $a/h$ ) was varied from 0.5 to 2.0. As can be seen in Figure 4.4, more flexural cracks developed in the specimen having the higher shear-span-to-depth ratio. In the specimen depicted in Figure 4.4(b), which has a low shear span-to-depth ratio, compression struts are formed that cause compression failure in this specimen as shown in Figure 4.4(a). Again, this simple anecdotal example illustrates the basic idea of linking images to failure modes and ultimately load levels, while also highlighting the multifaceted interaction of the structural parameters and the resulting accumulated damage.

Given that basic correlations exist between the crack patterns, load capacity, and structural parameters like shear reinforcement ratio, the goal of this paper is to use machine learning to refine these broad correlations to enable useful quantitative load-level estimations such as would be needed for applications like those illustrated earlier in Figure 4.1. Ideally, these damage and load level estimations should be valid across as wide a range of load and geometric parameters as possible, and they also should be based solely on externally visible data of the kind likely to be available in a field inspection scenario (i.e., parameters like concrete compressive strength, rebar size, distance between stirrups, etc. should not be part of the feature set used to build the estimation models). The computer vision framework discussed in the next section is intended to provide the means to develop and evaluate such image-based estimation models.

#### **4.5 Image-Based Classification and Estimation Framework**

The machine vision framework underlying the estimation models combines image processing, computer vision, and machine learning as illustrated in Figure 4.5. Figure 4.5(a) shows schematically how the data sets used in this paper were obtained, with images captured for tested specimens at different load levels collected into data sets based on specimen and load features. Figure 4.5(b) shows the basic image processing steps used to construct feature-set/load-level pairs by converting image data to feature sets. Figure 4.5(c) depicts the features and load-level pairs for the various data sets being used to train and validate machine learning regression models, while Figure 4.5(d) shows a similar process for classification. Figure 4.5(e) illustrates the combination of classification and regression, with a trained classification model initially splitting a data set into shear reinforced and unreinforced subsets (i.e., Data Set 1' and Data Set 2' in the figure), with

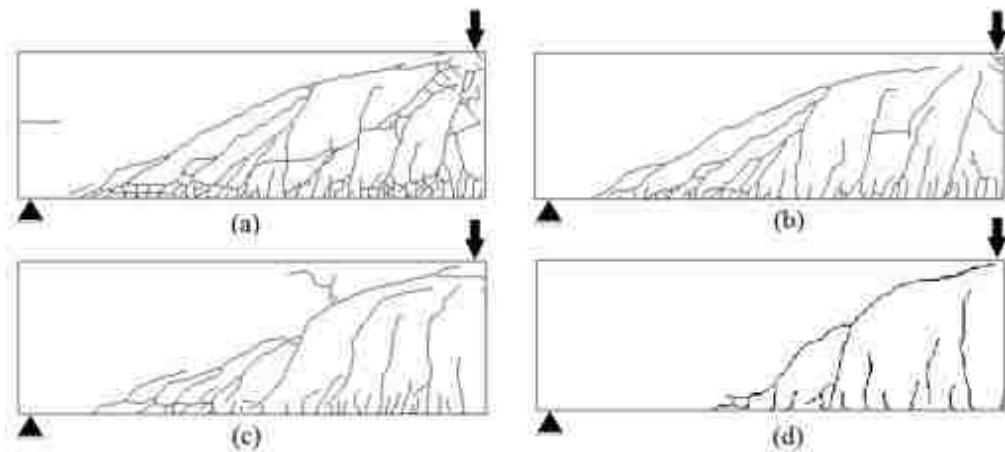


Figure 4.3: Typical crack pattern images in the specimens having different shear reinforcement details. Figures a, b, c and d show crack pattern images at ultimate load stage for specimens YB2000-4, YB2000-6, YB2000-9 and YB2000-0 respectively [17]; and [83]. All structural details except the distance between shear reinforcements are similar in the beams.  $d/s$  is 3.2, 1.4 and 0.7 for YB2000-4, YB2000-6 and YB2000-9 respectively. Specimen YB2000-0 does not have shear reinforcement.

specialized regression models subsequently applied to the classified data sets.

Considering the steps in Figure 4.5(b) in more detail, segmentation refers to the process of converting an image into binary form by isolating pixels corresponding to cracked regions. There is an ever-growing literature covering techniques for robust crack extraction from raw images, but for this particular study crack pattern line drawings already had been constructed as part of the experimental test reporting for all but one of the specimen sets considered ([9]). This made segmentation quite simple. For the few cases in which more involved segmentation was needed, a simple implementation of Canny's method [15] was found to be adequate in extracting crack patterns comparable to those in the line drawing depictions.

Continuing with the steps in Figure 4.5(b), the images were then transformed to a standard size using simple linear scaling. Also, as discussed in [26], because the crack line drawings from the earlier studies did not include details like crack widths, all crack images were subjected to an additional morphological transformation forcing all cracks to have a consistent single pixel width. If a full set of raw images were available for all the tests in question, more detailed crack representations could be used, possibly leading to improved estimation accuracy.

The feature extraction step in Figure 4.5(b) is the final step of the image processing. The basic idea of feature extraction is to project the image into a lower dimensional feature space based on numerical

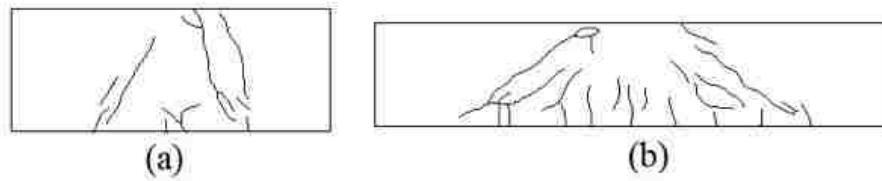


Figure 4.4: Typical crack pattern images in the specimens having different shear span to depth ratio ( $a/h$ ). Figures a and b show crack pattern images corresponding to the ultimate load stage for specimens BS3-0.5-Vs and BS3-2-Vs respectively [34].  $a/h$  is 0.5 and 2 for BS3-0.5-Vs and BS3-2-Vs respectively.

geometric properties such as pixel areas, bounding ellipse aspect ratios, and so on. It is also possible and useful to augment the pure image-based features with externally visible basic geometric properties of the test specimens such as the section depth,  $h$ , width,  $b$ , and span,  $a$ . The full set of 22 crack image features used in this study matches that developed and presented in detail in [26]. These crack image features include both individual crack features averaged across all crack segments present (e.g., average crack length), and summed quantities computed by adding together the individual contributions of each crack segment (e.g., total crack length). No claim can be made regarding optimality, but similar to the results from [26], subsets of features were found to give results with near equivalent accuracy to larger feature sets, indicating successful identification of dominant features. In the results to be presented later in this paper, the specific features used will be reported for each case and will include particular combinations of the total number of cracks ( $N_c$ ), polar moment of inertia of crack pixels ( $I_p$ ), and total (summed) major axis of crack segments ( $B$ ), as well as the full set of 22 image features (All).

The outcome of Figure 4.5(b) is a set of image and specimen features (which can be thought of as feature vectors) and corresponding load levels that were recorded when the images were captured. The load levels themselves can be used both in absolute form (i.e., recorded values in relevant units) and in a relative form designated the failure ratio, defined as the current load level normalized by the failure load level (e.g.,  $V/V_{\text{failure}}$ ,  $M/M_{\text{failure}}$ ). This particular quantity could be especially useful in field applications in which a key question might be how close a structural member is to its capacity (most likely presented as a percentage as in Figure 4.1). Load levels also can be either moment or shear values depending on the context, and converting between these quantities based on statics and known specimen properties is possible. For the regression-based estimation modeling, both the features and load levels serve as the data input as shown in Figure 4.5(c). In particular, the load levels represent the dependent values to be estimated from the independent variables (input features), which are crack features plus basic geometric properties such as



shear span to depth ratio,  $a/h$ , and depth,  $h$ . For classification (Figure 4.5(d)), only the feature set data are needed, along with the known specimen category. The objective of the classification model is to predict the beam type based on the same input features mentioned earlier. The input features were standardized using the z-score to make each feature in the feature set have a mean of zero and variance of unity [56] before feeding into the classifiers. The combination of classification followed by regression shown in Figure 4.5(e) uses both features and load levels in sequential fashion. Combining classifiers and regression models has been done for the purpose of improving regression models on the combined data set (Data Set 3).

There are many algorithms available for machine learning applications involving image data, with technology continuing to evolve on this front, and so no claim is made here of obtaining the best possible results. However, the load-level results in this paper have been generated using techniques found to be effective through extensive testing in [26], so they are representative of the performance one can expect with current feature-based algorithms. The specific algorithms used here for load-level estimation focus on Gaussian Process [68] and Support Vector Machine (SVM) [37] approaches as provided in the MATLAB-based WEKA toolbox ([38], [49]). Additional algorithms were used in [26], most notably Artificial Neural Networks [63], but with comparable accuracy. For the categorization tasks, the algorithms used include the WEKA-based J48 decision tree [67], random forest algorithms [11], logistic [48], and multilayer perceptron [39]. The processing tools and programming codes used in this study are available online in [dataset] [24], and at the following URL: <http://github.com/Rouzbehe/FullRC-Beam-Slab-Image-Based-Load-Estimation/>. Because the focus of this paper is more concerned with feasibility than optimality, default parameter settings suggested in the WEKA library [39] were mostly used to train the estimation and classifier models. The main settings for these parameters are summarized for each machine learning technique below.

- Multilayer perceptron: One hidden layer with the number of nodes equal to the sum of the number of features and classes divided by two was used to build the network. Momentum and learning rates of 0.2 and 0.3 were used, respectively, for the back propagation algorithm. The number of epochs was set to 500 for training and a Sigmoid function was used as the activation function for the model [39].
- Tree: The REPTree technique is a fast decision tree learner. The number of folds for reduced error pruning and the minimum number of instances per leaf were set to 10 and 2 respectively.
- J48: setting parameters were chosen based on recommendation of the WEKA library [39]. This includes confidence threshold factor and minimum number of instances per leaf values of 0.25 and 2, respectively.
- Logistic Regression: a modified version of the logistic model is implemented in the WEKA library that can handle instance weights [39]. The ridge value in the log-likelihood function was set to  $1.0E - 8$ ,

and the maximum number of iterations was set such that iteration continues until convergence occurs.

- Random Forests: 100 trees with unlimited depth were used to build the model. Consistent with the suggestion of the WEKA library [39], the number of randomly selected features was chosen using  $\log_2(N + 1)$ , in which  $N$  is the number of predictors or features [11].
- SVM: the sequential minimal optimization (SMO) technique, which is a solver for quadratic programming (QP) problems, was used for training the SVM model. The Pearson VII function-based universal kernel, which is known as Puk, was used as supplied in the WEKA library [39]. Omega and Sigma parameters were set to 1 in the Puk kernel.
- Gaussian Process: the version of the Gaussian model implemented in WEKA applies normalization to the target attribute as well as to the other attributes by default [39]. The level of Gaussian noise with respect to the transformed target was chosen as 10%. The radial basis function (RBF) kernel is a common kernel in machine learning problems and was used in this paper. Gamma and cache size, which should be a prime number, were set to 1 and 250,007, respectively, for the RBF kernel.

In some cases different parameter values were explored in trial and error fashion, but the above parameters were found to give suitable results for the purposes of this basic feasibility study.

#### **4.6 Error Measures and Cross Validation**

The inclusion of both classification and regression models leads to a need for two different kinds of performance metrics. In the case of classification, there is an unambiguous correct answer for each classification question: a given specimen either had shear reinforcement or it did not (this is not the same as classifying a failure as shear critical or flexure critical, which would not have the same crisp boundary between classes). In this kind of unambiguous classification context it is possible in principle to obtain 100% accuracy, and so there is a natural target level of optimal performance against which error measures can be evaluated. With respect to load-level estimation, the story is quite different. In this case there is no known correct answer for a given observation: ostensibly identical reinforced concrete specimens will exhibit different strengths and fracture details, and so estimation models cannot be more accurate than allowed by the inherent variability in the systems themselves. There is thus no optimal accuracy target as in the classification case, and so the observed error metrics are useful primarily for relative comparison among models, which could include design formula models as considered in [26].

To compare the performance of classifiers in distinguishing between shear-reinforced and shear-unreinforced specimens, the overall accuracy of the classifiers (i.e., number of correct predictions over total number of observations) was considered as the main performance metric. In addition, separate accuracy for each class

was calculated to evaluate the classifier accuracy: number of specimens assigned to a given class divided by the total number of specimens in that class. This is particularly important because of the unequal numbers of each specimen type in the combined data set (Data Set 3). In particular, 65% of the images corresponded to shear-unreinforced specimens (Data Set 1) and 35% corresponded to shear-reinforced specimens (Data Set 2). In addition to simple normalized numerical metrics, the use of receiver operating characteristic (ROC) curves is another common technique to represent the overall accuracy [65] of a classifier and the tradeoff between sensitivity (true positive rate) and specificity (1–false positive rate) at different categorization threshold values. Accuracy in this approach is determined by measuring the area under the ROC curve, which has as its axes the false positive rate (abscissa) and the true positive rate (ordinate), with the maximum possible area of 1.0 represents perfect classification performance.

In regards to regression estimation error metrics, it is important to note that because the data assembled for this paper were generated from experiments originally performed for other purposes, the overall distributions of structural parameter ranges are not uniformly distributed. This can cause reliability issues for any single statistical error measure. Therefore, a suite of goodness-of-fit statistics were used to compare the performance of the estimation models, and similar to what was observed in [26], the various error/fit metrics were found to be consistent with each other in terms of ordering and general trends. The regression metrics used include: 95% confidence interval lines, normalized root-mean-square error (NRMSE), index of agreement (IA) [50], correlation coefficient (R), and Explained Variance Score. With the exception of IA, these standard metrics are built into MATLAB's statistical toolbox [56].

Defining independent learning and test data sets is of paramount importance in evaluating the performance of machine learning models. This paper uses 10-fold cross validation, and this cross validation is applied at the specimen level rather than at the image level. It was shown in [26] that while this leads to less accurate estimation, it more fairly represents the situation relevant for field inspection scenarios. In particular, under normal circumstances images bracketing a given damage state would not be available for a field specimen, because the field specimen would not have a pre- and post-observation record of damage images and load levels such as is the case in a lab-based test set. By sampling at the specimen level it is not possible for training images for a particular specimen to also appear in the test image set. To illustrate the effect of image-based versus specimen-based cross-validation, Table 4.2 compares summary error measures for the two techniques for Data Set 2 (shear reinforced specimens) and a representative feature set/algorithm combination. The error measures show that image-based cross-validation leads to higher accuracy than would be obtainable in field scenarios represented by the specimen-based results, consistent with the shear-critical results presented in [26].

## 4.7 Estimation Results

This section presents results demonstrating the estimation accuracy of the machine vision models in regards to both load-level predictions and reinforcement-level categorization (i.e., determining whether a specimen had shear reinforcement or not). It is helpful to again refer to Figure 4.5(c)-(e) for an overview of the organization of the results to be presented. In particular, results will include classification accuracy as indicated in Figure 4.5(d), load-level accuracy as in Figure 4.5(c), and estimation accuracy combining classification followed by load-level regression as in Figure 4.5(e). Results in the case of load-level estimation will be presented using a combination of estimation vs. true value scatter plots (including 95% confidence intervals) and the various summary error measures described earlier. Categorization accuracy will be presented using a combination of ROC curves and standard summary numerical prediction statistics. Results generated using different subsets of the overall data set are used to highlight particular sensitivities of the algorithms used relative to specimen variations likely to arise in practical field applications.

The biasing effect of image-based sampling in the case of load-level estimations has already been discussed. A different biasing artifact arose in the consideration of specimen categorization. When the specific specimen geometric parameters  $a/h$  or  $h$  were included in the feature sets used to categorize specimens as being either shear reinforced or shear unreinforced, the categorization accuracy was boosted unrealistically high because specific geometric parameter values (e.g.,  $a/h = 2.91$ ) were idiosyncratic to particular test series such that they effectively acted like labels for these test series. Thus, the classifier could identify which test series a given image belonged to independent of any of the image features, and the test series itself was a simple proxy for whether the test included reinforcement or not, leading to near 100% accuracy. To remove this artificial biasing of the results,  $h$  values were not used in the categorizations, while rounded  $a/h$  values were used such that the physical variation in the values were preserved, but the labeling effect was removed. Testing showed that using crack pattern features combined with  $a/h$  values rounded to the nearest 0.5 avoided artificial, non-physical bias in training the classifiers. (Note that this same effect does not arise in the regression cases, since knowing which test series a specimen belongs to does not provide significant information concerning specific load and damage level correlations. This was also verified through testing.)

Tables 4.3 and 4.4 present categorization accuracy across multiple algorithms for two different feature sets: (i) all crack image features (Table 4.3); and (ii) all crack image features plus rounded  $a/h$  values (Tables 4.4). In these tables “Total Accuracy” refers to the number of correct predictions overall divided by the total number of observations, while “Class 1” and “Class 2” accuracies designate the number of correct identifications of specimens without and with shear reinforcement, respectively, divided by the number of specimens in the corresponding class. “AUC” refers to the area under the curve of the ROC curve corresponding to the given algorithm. As seen in the performance metrics the performance of the various

classifiers is roughly similar, but adding size effect data such as  $a/h$  improves the classification accuracy for all but the Random Forest Class 1 case. As shown earlier, the crack patterns are different in shallow and slender beams due to differences in the shear transfer mechanisms, and the size effect feature ( $a/h$ ) can play a key role in identifying slenderness of the beam, leading to the observed improvement in classification accuracy. The highest total accuracy of 84% was obtained using the Multilayer Perceptron model trained with crack features and rounded  $a/h$  as shown in Table 4.4. It should be noted that the baseline (random guess) accuracy of classification without using machine learning would be 66% since the combined data set (Data Set 3) is made of two unequally sized data sets (66% of the specimens from Data Set 1 and 34% from Data Set 2). The ROC plot representing the accuracy of the Multilayer Perceptron model is shown in Figure 4.6. As indicated in the figure the area under the curve value for the Multilayer Perceptron classifier is 90%, which reflects the robustness of the classification.

Figure 4.7 is the first of a series of similar plots showing estimation performance and the effects of varying modeling parameters, feature sets, and data sets. Only a subset of results from the various possible combinations are shown, with an emphasis on the effects of different combinations of structural features such as reinforcing, loading, and basic configurations. Both relative failure ratios and estimated load levels are considered, including scaled and unscaled moment and shear values. Figure 4.7 itself compares the effects of using different feature sets and size effect features ( $a/h$  and  $h$ ) on the performance of the SVM models for Data Set 2 (shear reinforced) using 10-fold cross validation at the specimen level. The figure shows the ability of the dominant feature subset to produce results similar in accuracy to the full feature set. This is consistent with the results observed in the previous study [26], and the reduced feature set thus can be used in these kinds of estimations.

Figure 4.8 compares the effects of combining dominant image features with different geometric parameters ( $a/h$  and  $h$ ) and internal structural properties such as the depth to distance between stirrups ratio,  $d/s$ , and shear index on the performance of the SVM models for Data Set 2 (shear reinforced) using 10-fold specimen-level cross validation. Although adding  $d/s$  and shear index to the dominant crack features can improve the performance of the SVM model as shown in Figure 4.8, the failure ratio estimation based on dominant crack features and basic geometric properties ( $a/h$  and  $h$ ) is a better indicator of the accuracy achievable in potential field application scenarios in which  $d/s$ , and shear index would not be known from surface observations. The range of margin of error for the 95% confidence intervals in Figure 4.8, and the error measures in Figure 4.9 also show that the performance of the SVM model solely using basic geometric properties ( $a/h$  and  $h$ ) is quite close in accuracy compared to the non-realistic case in which internal information is provided. It is worth noting that [26] used the effective section depth,  $d$ , rather than the external section depth,  $h$ , for its analyses, but extensive testing showed no significant difference in estimation accuracy using either  $h$  or  $d$  for

the data sets in question. The preference for using  $h$  is that it is an externally visible parameter, while  $d$  is not.

Figure 4.10 compares performance of the SVM models to estimate failure ratio load levels for shear unreinforced (Data Set 1) and shear reinforced (Data Set 2) specimens using dominant features and 10-fold specimen-level cross validation. Based on the results for the reinforced and unreinforced complete data sets shown in parts (a) and (b) of the figure, it appears that shear unreinforced cases are less predictable than shear reinforced, but closer examination shows the performance of the SVM model on Data Set 2 is better than Data Set 1 primarily because there is a narrower range of specimen variability in Data Set 2. As explained earlier in the discussions of Table 4.1 and Figure 4.2, Data Set 1 (shear unreinforced) includes a broader parameter range and more varied loading conditions compared to Data Set 2 (shear reinforced). This can be confirmed by considering Figure 4.10(c), showing model estimation results for Data Set 5, which contains a subset of the shear unreinforced specimens having similar loading and geometry ranges as Data Set 2 (i.e., 3-point bending only, and no 4-m deep beams). Given similar parameter ranges, the estimates are actually better for the shear-unreinforced specimens overall, which is also evident in the numerical error metrics in Figure 4.10(d). Looking more closely at Figure 4.10(b) and (c) it can be seen that there is less scatter for the reinforced specimens (Figure 4.10(b)) for failure ratios less than about 0.40 and more scatter above 0.60, while the unreinforced specimens (Figure 4.10(c)) have more consistent scatter throughout. Because of the more ductile failure in the shear-reinforced specimens, there is a less crisply identified failure load, and so crack pattern images near failure tend to be mapped to a narrower band of load-level values than in a less flat failure curve, leading to less delineation between loads and crack patterns, and thus more scatter in the model estimations.

Figure 4.11 compares the performance of SVM models in estimating failure ratios for cases combining all shear reinforced and unreinforced specimens (Data Set 3) and all 3-point bending specimens (shear reinforced and unreinforced) with the 4-m deep specimens removed (Data Set 4) using dominant crack features. These results can be considered representative of the kind of accuracy that might be obtained in field applications with an unknown mix of shear-critical and flexure-critical beams and slabs. The performance of the model for Data Set 4 is slightly better than for Data Set 3, but both these cases are noticeably less accurate than when the reinforced and unreinforced specimens were handled separately as in Figure 4.10. This suggests that using classification prior to regression could lead to improved results for the combined case. This is indeed the case if one can get 100% correct classification, but when using actual classifiers such as those described above, testing demonstrated virtually no benefit to pre-classifying prior to regression.

In addition to estimating load levels in terms of failure ratios, it is also possible to estimate absolute load levels. Figure 4.12 compares the performance of SVM and Gaussian models when estimating absolute shear

force for: (a) shear-unreinforced specimens (Data Set 1); (b) shear-reinforced specimens (Data Set 2); and (c) all specimens combined (Data Set 3), using dominant crack features and externally observable geometric properties such as  $a/h$ ,  $h$  and  $b$ . As is evident in this figure, the estimation accuracy for the shear unreinforced specimens is relatively poor, while the reinforced-specimen results are more comparable to those observed in the failure ratio estimation results (although direct comparison of error metrics across these analyses can be misleading due to the data range and distribution differences). In the case of direct moment estimation as shown in Figure 4.13, the same phenomenon can be seen: poor performance for the shear-unreinforced specimens, and much better accuracy for the shear-reinforced specimens. It is also evident in this case that there is a strong influence from a relatively small number of outliers associated with the 4-m deep beam specimen.

There are several possible means to improve the estimation results for absolute load levels, including using normalizing parameters other than the specimen failure load, which would not be known in field applications. Figure 4.14 shows the performance of the Gaussian Process model to estimate scaled moment and scaled shear for shear-reinforced specimens (Data Set 2) using different scaling factors. The shear is scaled in two ways in parts (a) and (b) of the figure, respectively: (i) using the nominal concrete shear capacity,  $V/(2\sqrt{f'_c} bd)$ ; and (ii) using the nominal total shear capacity accounting for the shear reinforcing,  $V/(2\sqrt{f'_c} bd + A_s f_y d/s)$ . These scaled shear estimations use internal properties such as  $f'_c$  that would not be known from surface observations in field applications, and therefore are not directly applicable to the broader goals of this work. However, the figure demonstrates the degree to which there is a correlation between identified dominant crack features and normalized shear values. More usefully, the scaled moment  $M/bh^2$  can also be estimated using dominant crack features and observable geometric properties such as  $a/h$ ,  $h$ , and  $b$  as shown in Figure 4.14(c). These results indicate a strong correlation between  $M/bh^2$  and dominant crack features as shown in Figure 4.14(d).  $M/bh^2$  is proportional to the nominal bending stress  $Mc/I$  for a rectangular section beam, and this particular normalization was observed to have stronger correlations than other potential normalizing factors such as nominal bending capacity.

The ability to estimate scaled moments can also be used to indirectly estimate shear levels. Figure 4.15 compares absolute shear estimation using both direct and indirect approaches for shear-reinforced specimens (Data Set 2). The indirect absolute shear values in Figure 4.15(b) were calculated using estimated values of  $M/bh^2$  as shown in Figure 4.14(c) by multiplication with  $bh^2/a$  (for 3-point bending,  $M = Va$ ). Figure 4.15(a) is the same as Figure 4.12(b), repeated here for convenient comparison. As is evident in the figures and in the error metrics in Figure 4.15(c), using  $M/bh^2$  can noticeably improve absolute shear estimation for Data Set 2.

Given the difference in absolute load-level estimation accuracy between the shear-reinforced and unrein-



forced data sets for moment and shear, it is worthwhile to again consider the use of a special combination of classification and regression. In particular, Figure 4.16(a) shows the absolute direct moment estimation results for all specimens, and the poor performance for the shear-unreinforced data as seen in Figure 4.13(a) observed earlier drags down the accuracy of the estimation for the combined specimen set. However, the absolute shear predictions for the shear-unreinforced specimens (Data Set 1) were more accurate than the moment estimations (see Figure 4.12(a)), and thus could be used to indirectly calculate better moment estimates via  $M = Va$ . Thus, by pre-classifying the combined data into predicted unreinforced and reinforced categories (see Figure 4.7(e)), using direct moment calculations for images labeled as reinforced and indirect moment calculations for images labeled as unreinforced, and then generating combined predictions leads to the results shown in Figure 4.16(b). In this case the pre-classification (using the Multilayer Perceptron classifier described earlier) combined with category-specific regression leads to better overall results. Clearly a significant part of the improvement is related to the reining in of the outlier effects, but Figure 4.16(c) and (d) show the same kind of analysis for the case in which only 3-point bending specimens are included and the 4-m deep specimen is removed. This takes away the effect of the major outliers, but it can still be seen that the results are improved (notice especially the trend line) in part (d) of the figure. The error metrics for this final set of cases are presented in Figure 4.17.

Although results are not presented here, a similar approach could be used in the case of absolute shear estimation, leveraging the fact that shear-unreinforced specimen shears can be estimated well directly, while shear-reinforced specimen shears can be best estimated via  $V = M/(bh^2a)$  (see Figure 4.15). Classification followed by custom regression was again found to lead to improved results compared to straight regression, although with less effect than in the moment case.

In order to provide one additional comparison reference for the overall accuracy of the computer vision framework presented in this study, the absolute shear estimation model performance was compared to ultimate load predictions based on the standard ACI procedure [22] for design strength determination. Using the ACI procedure, the nominal moment capacity and shear strength of the specimen were calculated to find the governing design load controlling the behavior of the beam (i.e., based on shear or flexural failure). Figure 4.18(a) shows the performance of model predictions based on the ACI design formula. Figure 4.18(b) shows shear estimations using a combined classification and regression shear estimation model for all 3-point bending specimens (Data Set 4). As shown earlier, pre-classification combined with category-specific regression results in better overall shear estimations. In this case, the Multilayer Perceptron classifier was used to classify the beam type, and the SVM model trained with the dominant features was employed to perform the estimations. The shear-unreinforced specimen shears were estimated directly, while shear-reinforced specimen shears were estimated via  $V = M/(bh^2a)$ . Figure 4.18(c) extracts the subset of estimation points from



Figure 4.18(b) that correspond to the final loading stage, which is similar to the ultimate load prediction in Figure 4.18(a). Note that these subplots are not showing identical quantities: the first is an ultimate strength prediction; the second contains a series of load estimations for damage observations throughout a loading history; and (c) is an indirect image-based strength prediction which ultimately would have little practical utility (the member in question would have failed already, so damage estimation would be moot). Thus, the purpose of Figure 4.18 is to provide a means for making broad comparisons between estimation methods with very different purposes. In this light, it can be seen that the image-based predictions are similar to the ACI-based prediction in terms of overall prediction accuracy. This demonstrates that for conditions similar to those applied here the machine learning approach can be capable of estimating absolute load levels based solely on external crack patterns and externally visible geometric properties with accuracy similar to common design procedures.

#### 4.8 Summary and Conclusions

This work has considered the application of computer vision for quantitative assessment of damage and load levels based on surface crack patterns. Building on previous work focused on shear critical (i.e., shear unreinforced) beams and slabs [26], the computer vision framework, which generates corresponding supervised machine learning predictive models, was extended to include beams and slabs with standard shear reinforcement. Unlike mechanics-based approaches, which can be generalized using first principles, data-driven methods require additional data to reliably extend their applicability beyond the calibration data range, and so additional test data for shear-reinforced beams and slabs were identified and incorporated into the modeling framework. This enabled the study of the effectiveness of the computer vision approach to estimate load and damage levels across and within several key categories of reinforced concrete flexural members subjected to monotonic loading, and to investigate the computer vision framework's ability to identify the presence or absence of shear reinforcement based on externally observed damage.

The primary conclusions to be drawn from the present work are listed below:

- Categorization of beams and slabs according to the presence or absence of shear reinforcing using the techniques presented here can be achieved with accuracy rates on the order of 85–90% (relative to a baseline of 66%).
- Load-level estimations for failure ratios were comparable across all specimen combinations, with the lowest NRMSE (12%) corresponding to the subset of shear unreinforced specimens subjected to 3-point bending and with the deepest (4-m) outlier geometries excluded. For the shear-reinforced specimens (which were all 3-point bending and had no outlier deep sections) the best observed NRMSE was 15%, and for all specimens together the best NRMSE observed was 17%.

- Absolute shear and moment estimations had wider variations than failure ratio estimations. The best results were obtained for the shear-reinforced specimens, and good results could be obtained for any specimen set by using appropriate direct or indirect calculation strategies.
- Failure ratio estimations likely would be more useful in practical applications than absolute load estimations because failure ratios provide a direct estimate of the percentage of capacity the member has reached.
- Load-level estimations for shear critical members tended to be more accurate than those for shear reinforced members for all the error measures considered. Although the increased ductility inherent in shear reinforced members might be expected to lead to more predictable behavior, the increased ductility provided a less crisp delineation of failure, and so this weakened the strength of the correlations between load levels and damage.
- Pre-categorization of specimens using machine learning combined with sub-category regression provided consistent but small improvements over straight regression.
- Using alternative span-to-depth measures ( $a/d$  or  $a/h$ ) leads to similar estimation accuracies.
- Knowing additional structural properties such as  $f'_c$  and shear reinforcement spacing can lead to improved estimations.
- For categorization in particular, it is important to be careful to keep experimental artifacts from acting as implicit labels, leading to misleadingly high apparent accuracy.

Because the technologies and algorithms associated with computer vision are continually improving, the results and conclusions presented here can best be thought of as setting a baseline for what is currently possible. The levels of accuracy in estimation achieved with the framework used in this paper are promising with respect to potential field applications, but there is also a significant gap between this feasibility study based on lab studies and practical deployment.

The limitations of the results presented here are similar to those identified in [26], and include the limited size of the data sets relative to those common in machine learning contexts, the relatively simple crack representations available in the reported prior work, and the specific focus on simply-supported beams and slabs undergoing monotonic loading. It would not make sense to apply the specific models trained here outside the controlled context in which they were developed. However, none of these limitations arise from inherent problems within the overall framework, and in principle each could be addressed with sufficient further testing and/or data gathering.

In the spirit of aiding reproducibility and further development in this field, the collected data and processing tools used in this study are available online in [dataset] [24], and at the following URL: <http://github.com/Rouzbehe/FullR-Beam-Slab-Image-Based-Load-Estimation/>.

Table 4.1: Summary table of prior experimental study data sets. See Fig. 1 for definitions of the tabulated geometric parameters.

Reference	No. Specimens	No. images	Test type	Width, $b$ (cm)	Length, $L$ (cm)	$a/h$	$h$ (cm)	$d/s$	Shear Index	$\rho$ (%)	$f'_c$ (MP)
<b>Shear Reinforced Specimens</b>											
[52]	12	97	B3, S3	25–117.2	208–370.8	3–3.65	28.19–49.8	1.62–2.15	0.35–1.97	1.01–1.74	34.8–4
[10]	11	86	B3	20	259	3.6	35	14	0.57–1.29	2.28–2.99	49.9–
[64]	7	48	B3	29.9–100	179.8–600	2.9–3	26.9–92.45	1.54–2.75	0.193–0.4	0.76–1.4	38–4
[5]	4	29	B3	29.9	600	2.9	92.5	1.54–3.08	0.4	1.01	21–8
[17]	2	11	B3	29.9	1077–1082	2.81–2.91	184.4–192.5	1.37–1.42	0.3	0.36–1.52	26.2–2
[83]	3	21	B3	29.9	1081	2.86	189	0.7–3.2	0.33–0.37	0.74	31.8–
[34]	4	12	B3	9.9	74.9–146.8	0.5–2.00	24.13	3.4	31.6	0.32	22.2–3
Overall	43	304		9.9–117.2	74.9–1082	0.5–3.65	24.13–192.5	0.7–3.4	0.193–31.6	0.32–2.99	21–8
<b>Specimens without Shear Reinforcement (as used in [26])</b>											
[83]	1	4	B3	30	1150	2.9	189	-	-	0.7	31.8
[17]	2	12	B3	30	1079.5–1121.9	2.8–2.9	192	-	-	0.4–1.5	26.2–2
[71]	8	52	B3	22.9–57.15	198–495.3	2.3	26.7–66.8	-	-	0.55–0.85	18.6–3
[70]	30	198	B3, S3	12.2–300.5	108–810	2.79–3.4	28–145	-	-	0.3–1.33	29.1–7
[58]	8	88	B3	20.3–61	139.7–493	2.97–3	23.1–81.2	-	-	1.2–1.3	64.8–7
[9]	15	42	B4	20	125	1.1–4.81	8.1–30.74	-	-	1.3–1.55	31.4
[62]	6	42	SU	30	358.1–599.4	—	92.5	-	-	0.98	39–6
[59]	3	28	H3	19.8	180–299.7	3–5.0	30	-	-	1.57–3.1	35.4–5
[57]	10	82	B3	22.9–57.4	197.9–609.7	2.3–2.9	26.9–67	-	-	0.3–0.98	22.8–3
[66]	1	10	B3	24.9	2380	3.1	384	-	-	0.7	40.0
Overall	84	558		12.2–300.5	108–2380	1.1–5.0	8.1–384	-	-	0.3–3.1	18.6–7

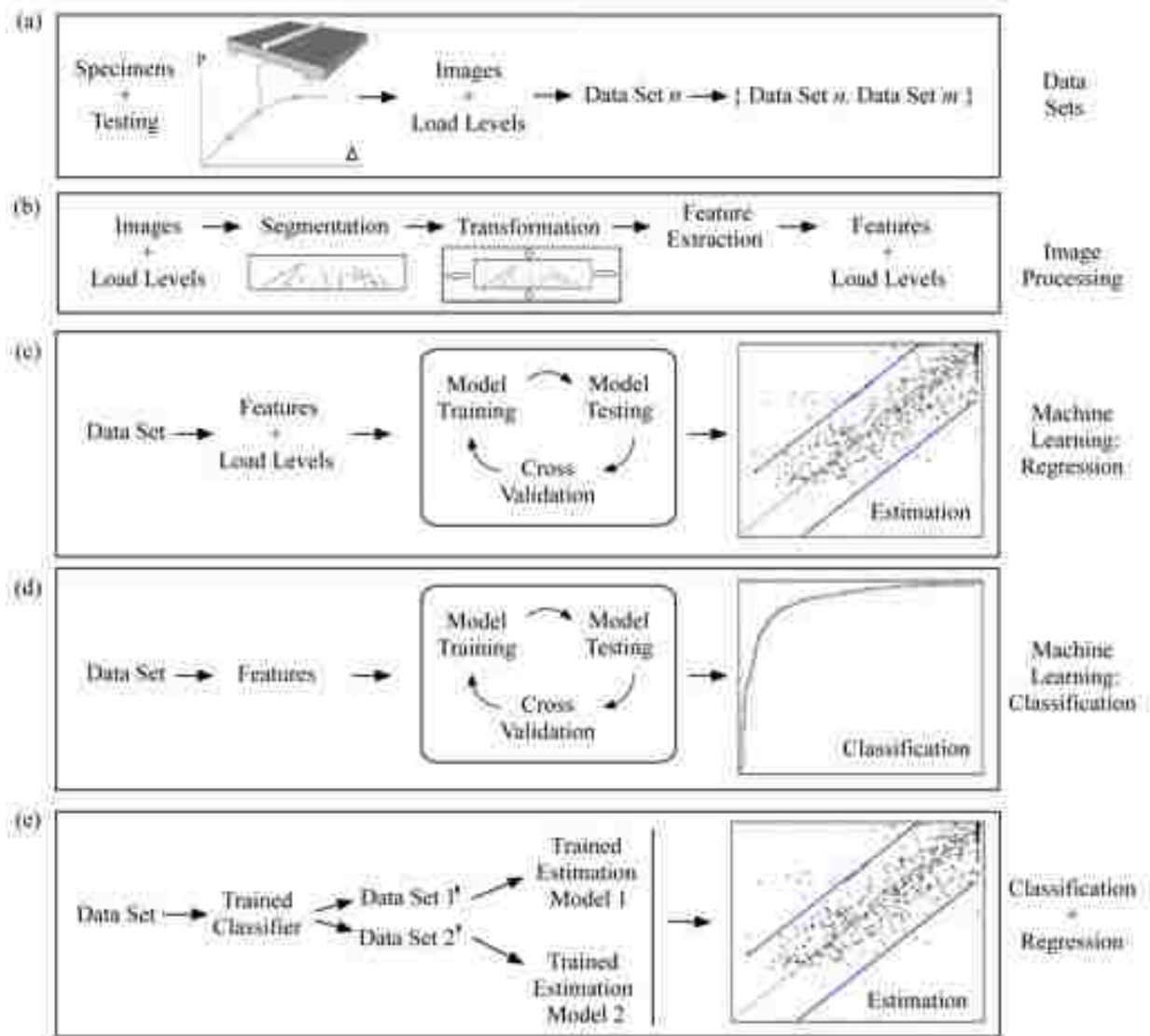


Figure 4.5: Schematics of the machine-learning framework components and pipelines used to develop estimation and classification models.

Table 4.2: SVM model failure ratio estimation performance using all crack features and  $a/h$  and  $h$ , comparing image and specimen level cross validations.

Metric	Image level	Specimen level
NRMSE (error)	13%	15%
R	91%	87%
Index of Agreement	95%	93%
Explained Variance Score	83%	76%

Table 4.3: Shear reinforcement (Data Set 2)/no shear reinforcement (Data Set 1) classification performance using all crack pattern attributes.

	J48	Logistic	Random Forest	Tree	Multilayer Perceptron
Total Accuracy	74%	80%	80%	74%	80%
Class 1 Accuracy	80%	90%	89%	78%	84%
Class 2 Accuracy	63%	62%	65%	67%	73%
AUC (ROC area)	72%	86%	87%	77%	86%

Table 4.4: Shear reinforcement (Data Set 2)/no shear reinforcement (Data Set 1) classification performance using all crack pattern attributes and rounded  $a/h$ .

	J48	Logistic	Random Forest	Tree	Multilayer Perceptron
Total Accuracy	81%	81%	81%	76%	84%
Class 1 Accuracy	88%	91%	88%	79%	85%
Class 2 Accuracy	69%	64%	70%	73%	83%
AUC (ROC area)	89%	87%	90%	83%	90%

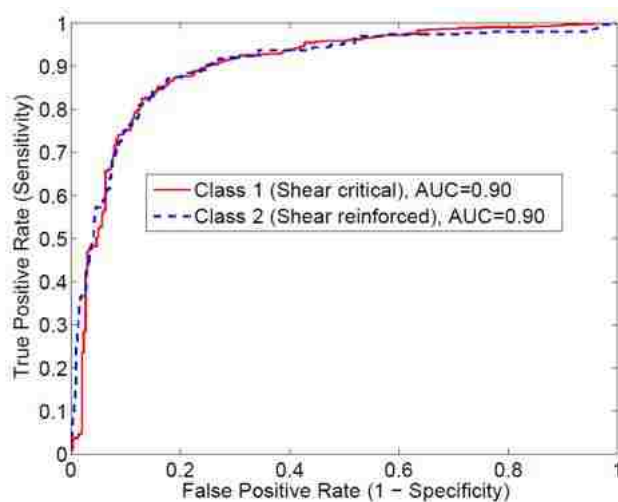
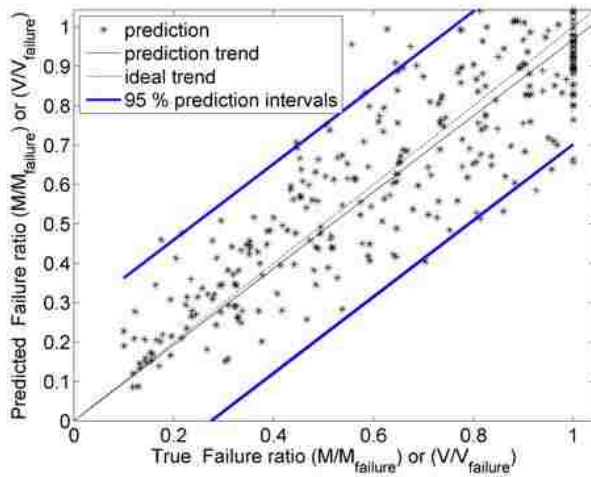
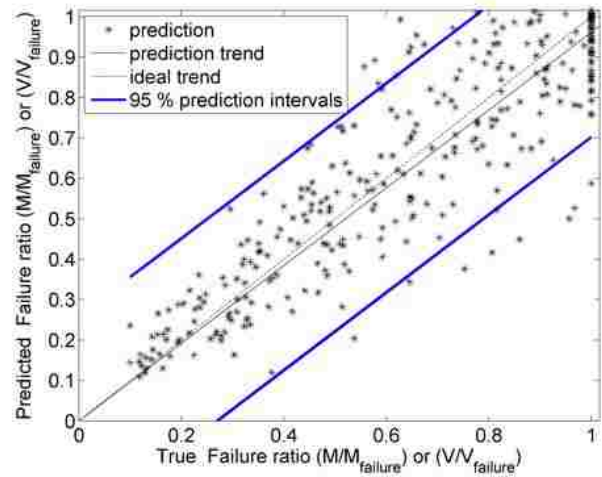


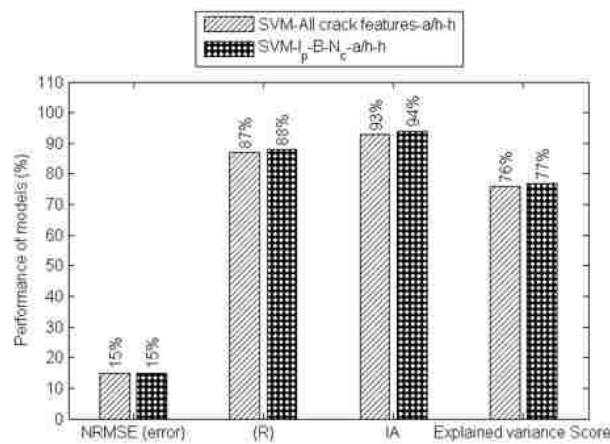
Figure 4.6: ROC plot showing the shear reinforcement (Data Set 2)/no shear reinforcement (Data Set 1) categorization performance of the Multilayer Perceptron model trained using all crack pattern attributes and rounded  $a/h$ .



(a) Estimation using all crack features,  $a/h$ , and  $h$ .



(b) Estimation using  $I_p$ ,  $B$ ,  $N_c$ ,  $a/h$ , and  $h$ .



(c) Estimation error metrics.

Figure 4.7: SVM model failure ratio estimation accuracy for shear reinforced specimens (Data Set 2) comparing the effects of using different feature sets.



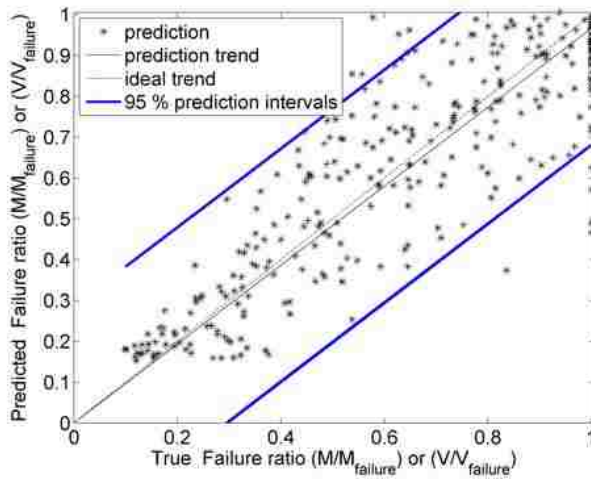
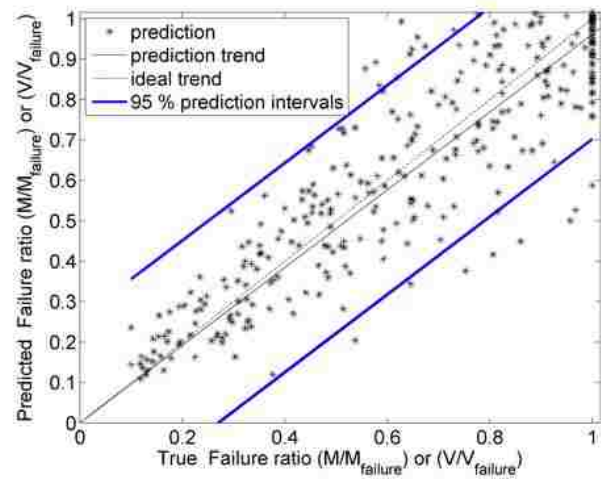
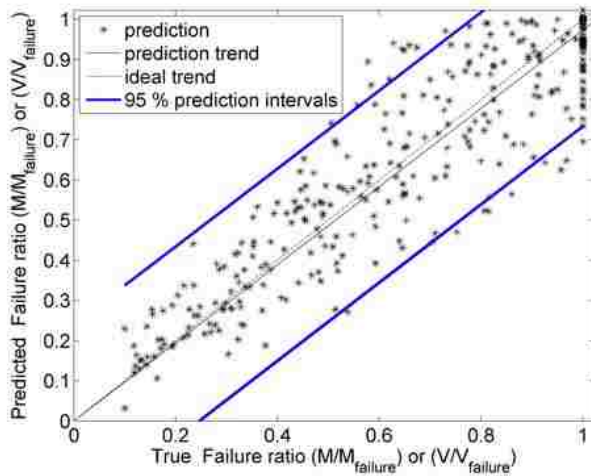
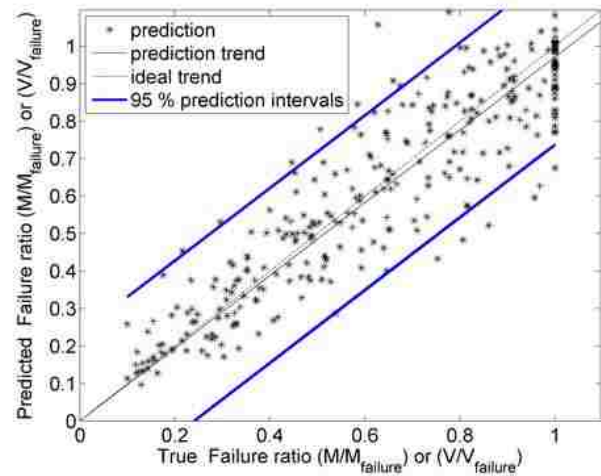
(a) Features:  $I_P$ ,  $B$ ,  $N_c$ (b) Features:  $I_P$ ,  $B$ ,  $N_c$ ,  $a/h$ ,  $h$ (c) Features:  $I_P$ ,  $B$ ,  $N_c$ ,  $a/h$ ,  $h$ ,  $d/s$ (d) Features:  $I_P$ ,  $B$ ,  $N_c$ ,  $a/h$ ,  $h$ ,  $d/s$ , shear index

Figure 4.8: Performance of SVM models in estimating failure ratio for shear reinforced specimens (Data Set 2) with increasing geometric specimen information included in the feature sets.



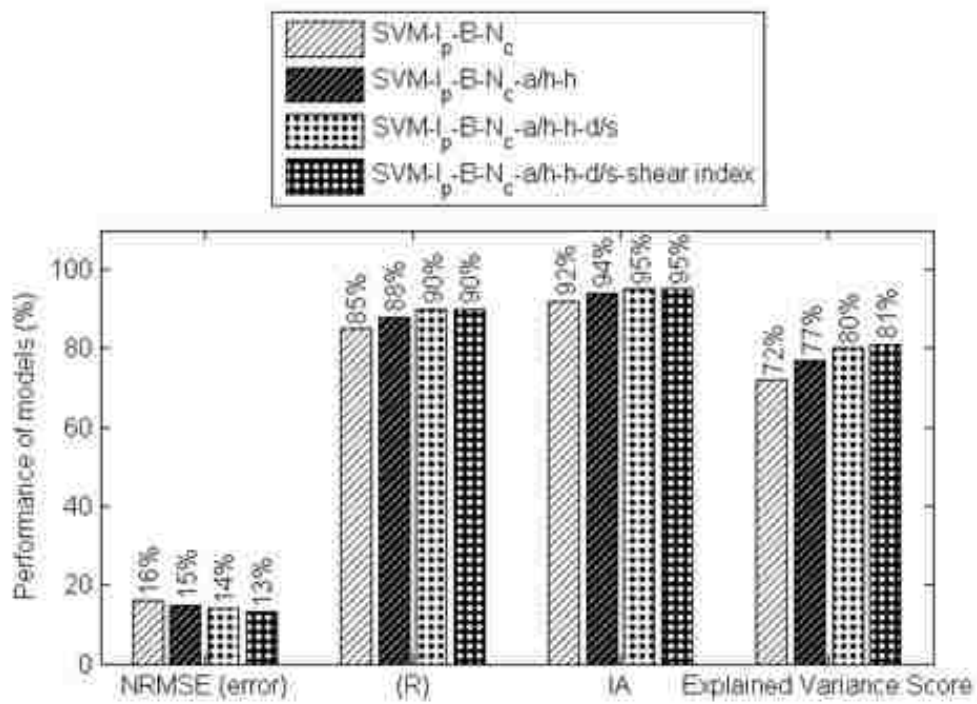
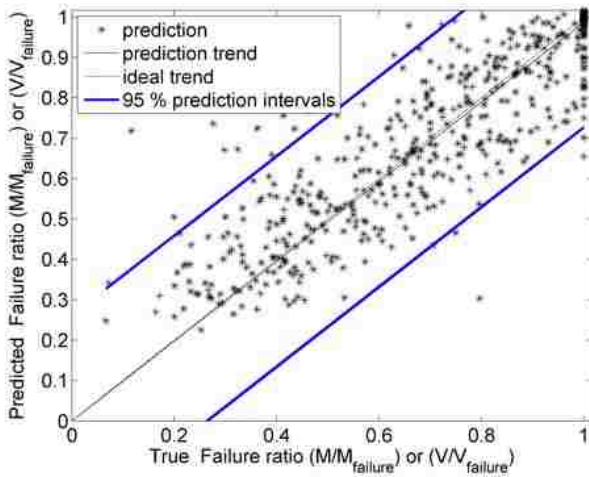
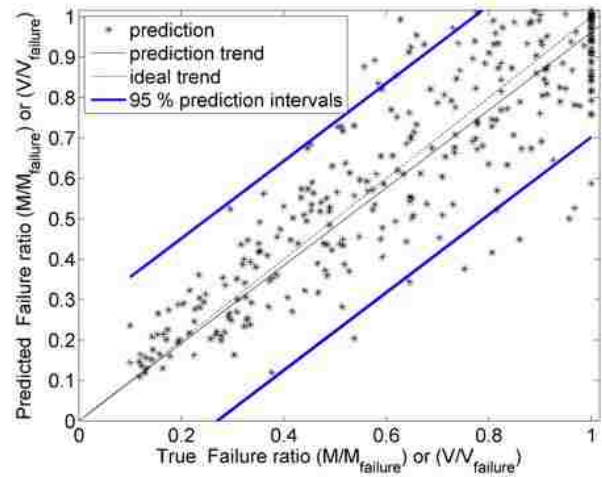


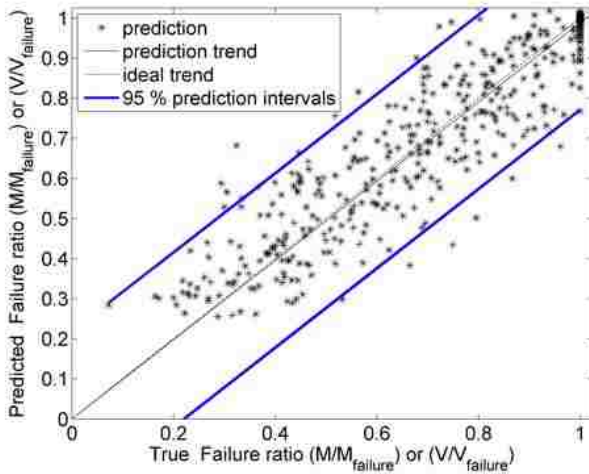
Figure 4.9: Error metrics for SVM model failure ratio estimation for cases in Figure 4.8.



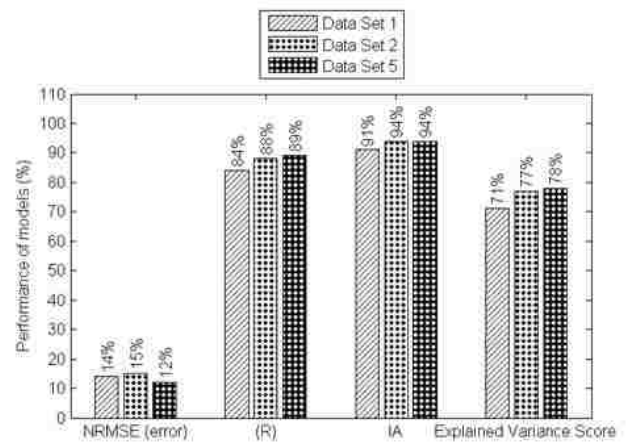
(a) Unreinforced specimens (Data Set 1).



(b) Reinforced specimens (Data Set 2).

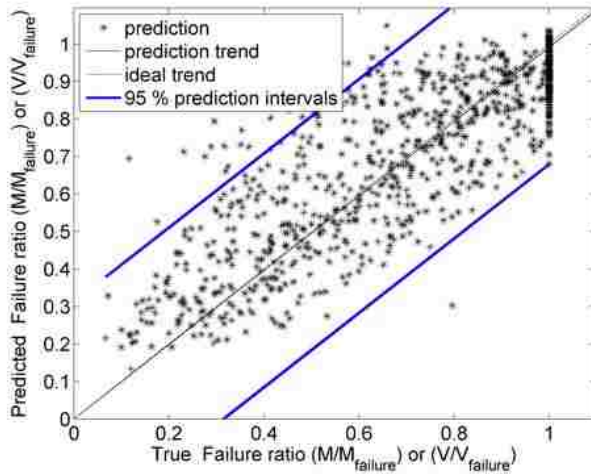


(c) Subset of unreinforced specimens (Data Set 5).

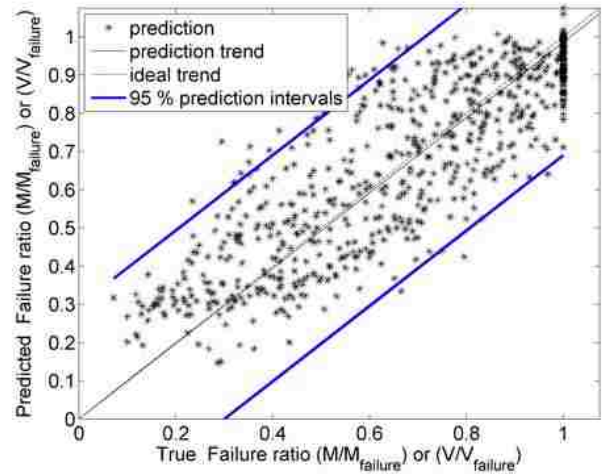


(d) Estimation error metrics

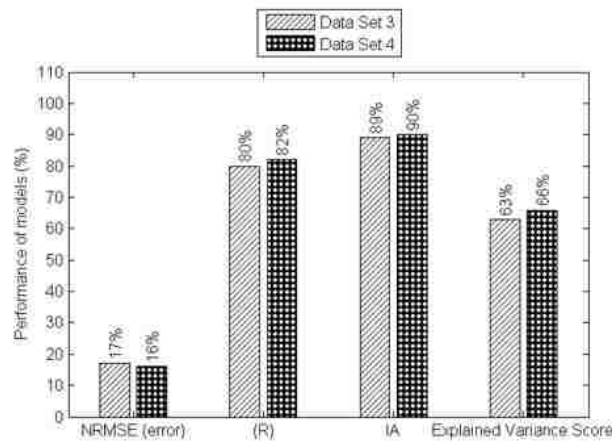
Figure 4.10: Comparison of estimation performance of SVM models using dominant features ( $I_P$ ,  $B$ ,  $N_c$ ,  $a/h$ ,  $h$ ) to estimate failure ratios for all unreinforced specimens (Data Set 1), all reinforced specimens (Data Set 2), and a subset of unreinforced specimens (Data Set 5).



(a) All specimens (Data Set 3).

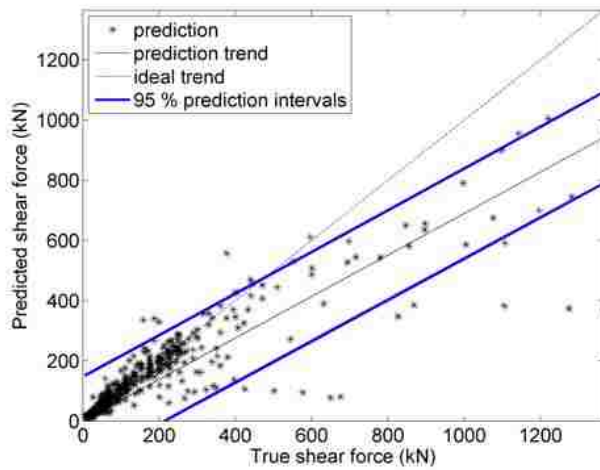


(b) All 3-point load specimens (Data Set 4).

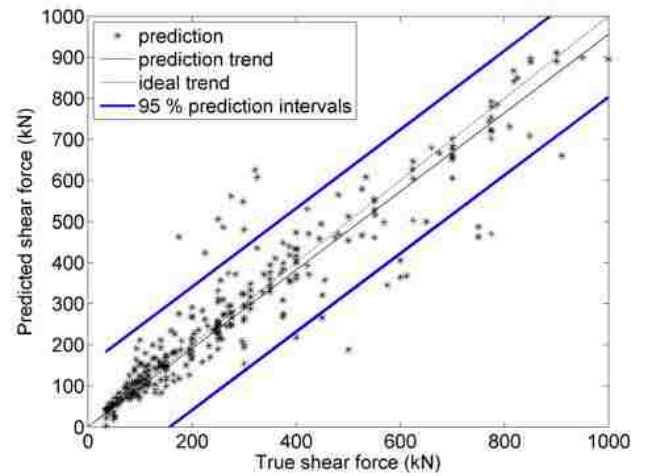


(c) Estimation error metrics.

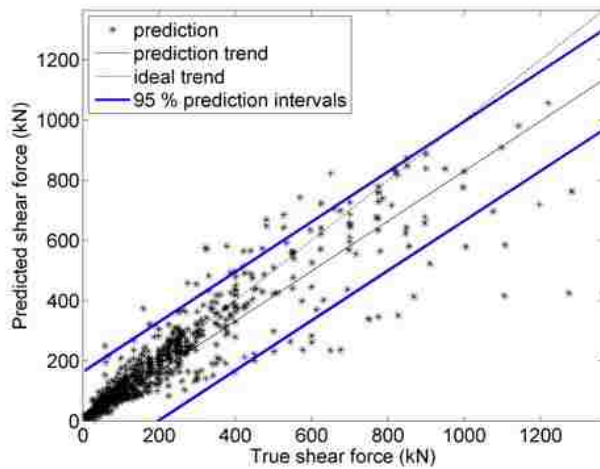
Figure 4.11: Performance of SVM models to estimate failure ratio for combined shear reinforced and unreinforced specimens (Data Sets 3 and 4) using features  $I_p$ ,  $B$ ,  $N_c$ ,  $a/h$ , and  $h$ .



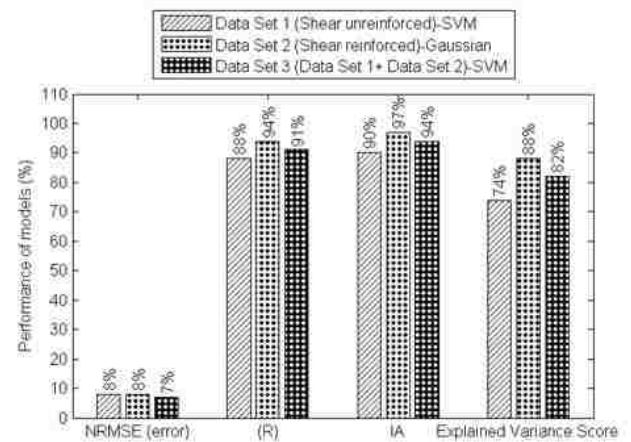
(a) Shear unreinforced (Data Set 1), SVM.



(b) Shear reinforced (Data Set 2), Gaussian Process.

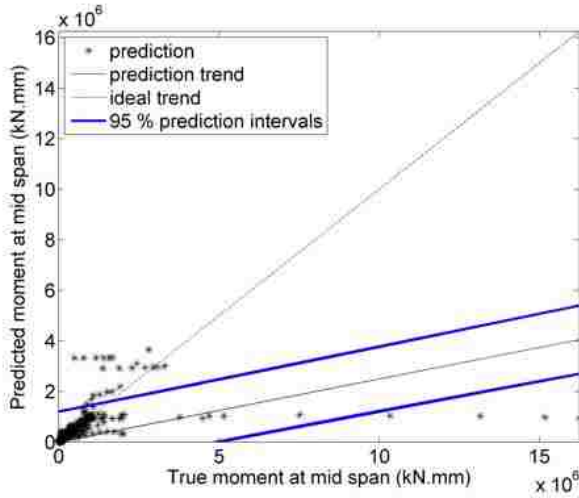


(c) All specimens (Data Set 3), SVM.

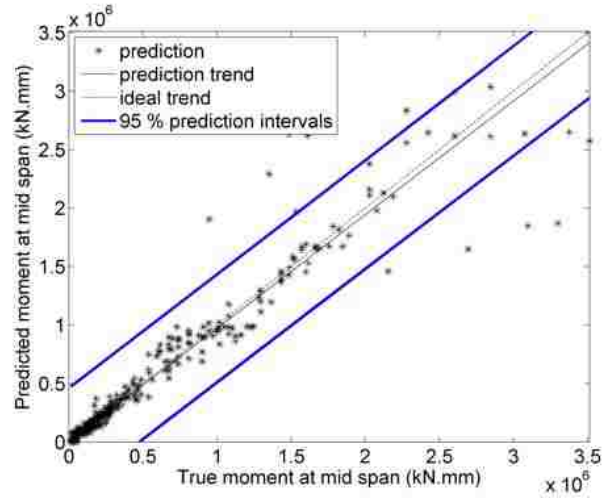


(d) Estimation error metrics.

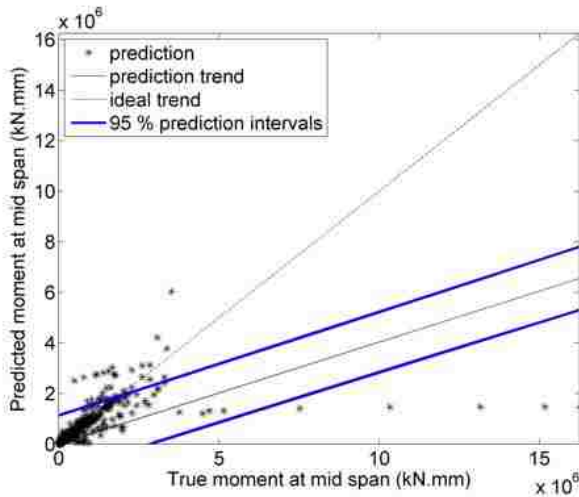
Figure 4.12: Performance of direct SVM/Gaussian models in estimating absolute shear force on different data sets using the feature set  $I_p$ ,  $B$ ,  $N_c$ ,  $a/h$ ,  $h$ , and  $b$ .



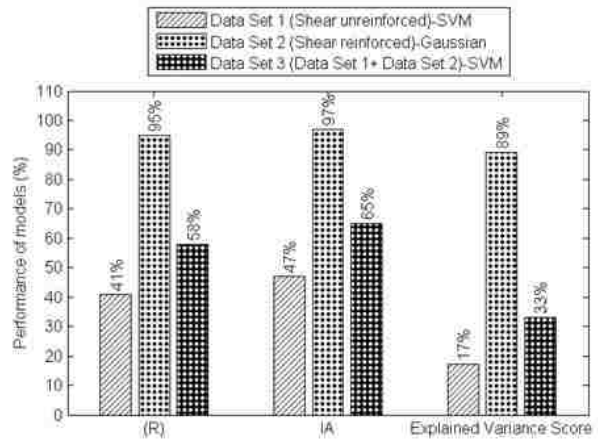
(a) Shear unreinforced (Data Set 1), SVM



(b) Shear reinforced (Data Set 2), Gaussian Process.



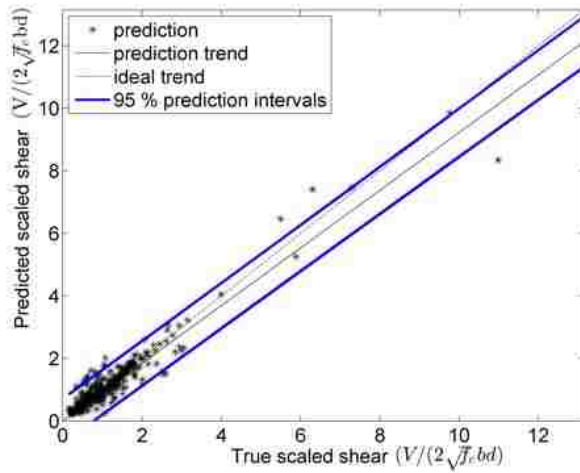
(c) All specimens (Data Set 3), SVM.



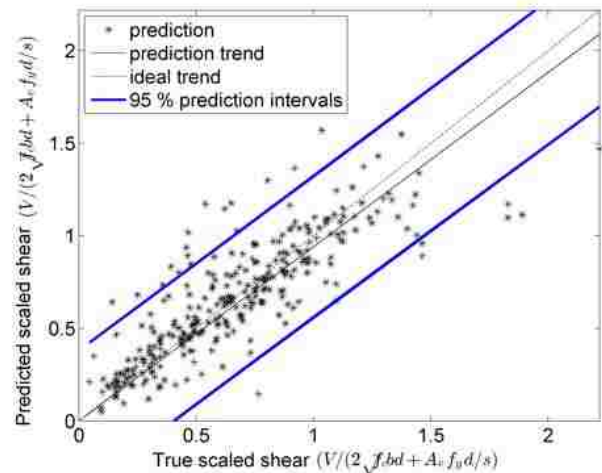
(d) Error estimation metrics.

Figure 4.13: Performance of SVM/Gaussian models in estimating absolute moment on different data sets using the feature set  $I_P$ ,  $B$ ,  $N_c$ ,  $a/h$ ,  $h$ , and  $b$ . Classification and regression are combined in (d).

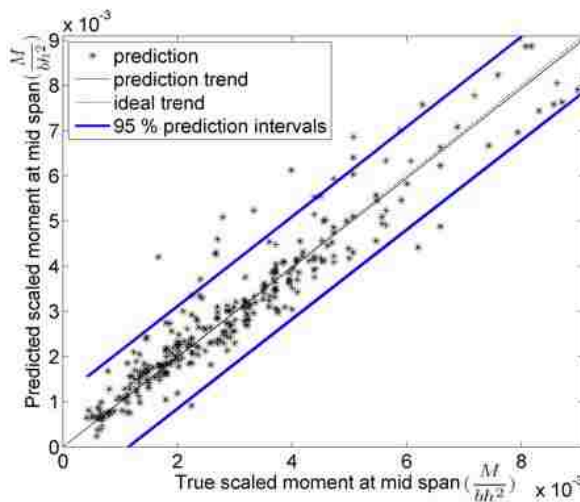




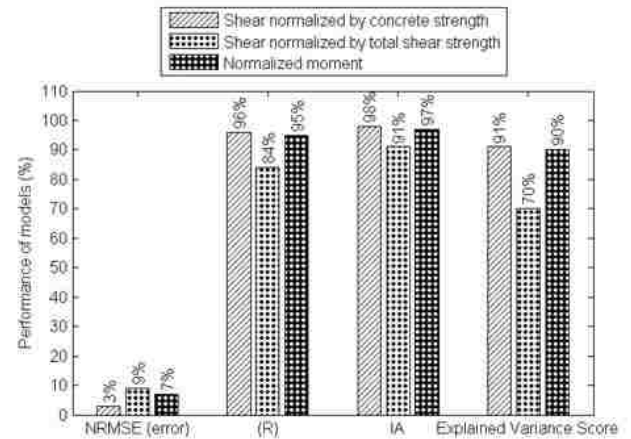
(a) Shear normalized by concrete strength.



(b) Shear normalized by total shear strength.

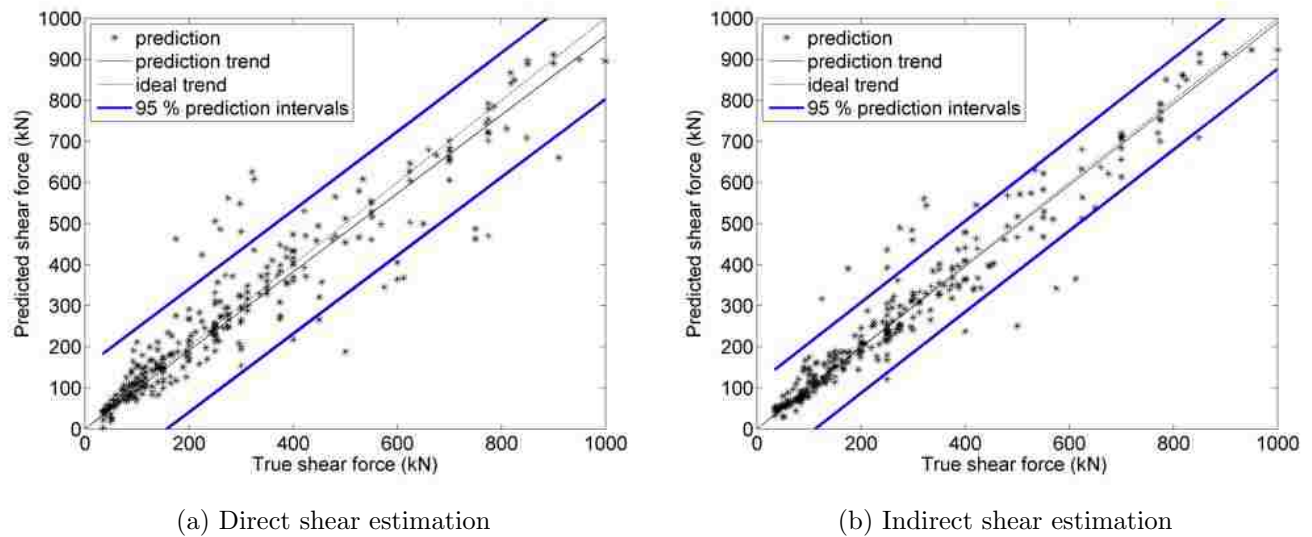


(c) Normalized moment



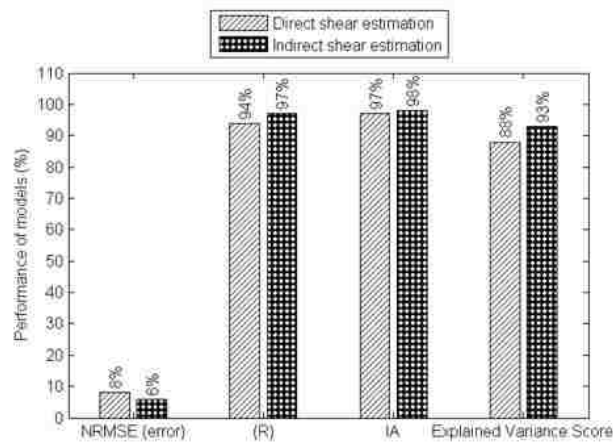
(d) Estimation error metrics.

Figure 4.14: Estimation performance of Gaussian Process models using dominant features ( $I_P$ ,  $B$ ,  $N_c$ ,  $a/h$ ,  $h$ ,  $b$ ) to estimate normalized shear forces ( $V/(2\sqrt{f'_c} bd)$  and  $V/(2\sqrt{f'_c} bd + A_s f_y d/s)$ ) and scaled moment ( $M/(bh^2)$ ) for reinforced specimens (Data Set 2). Note the axes scales vary between subfigures.



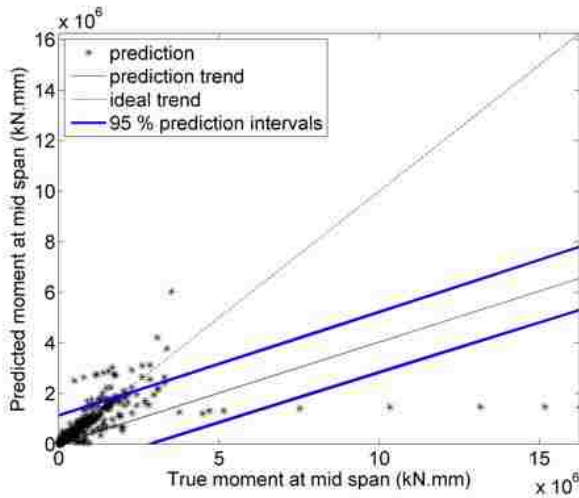
(a) Direct shear estimation

(b) Indirect shear estimation

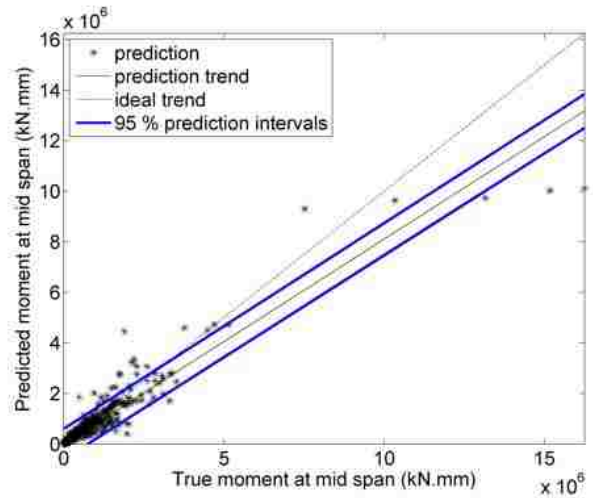


(c) Estimation error metrics.

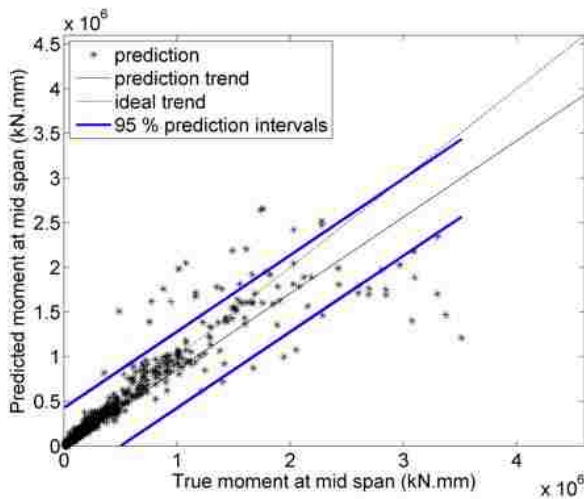
Figure 4.15: Performance of the Gaussian Process models for estimating internal shear for shear-reinforced specimens (Data Set 2) directly using the feature set  $I_P$ ,  $B$ ,  $a/h$ ,  $h$ ,  $N_c$ , and  $b$ , and indirectly through the estimated scaled moment,  $M/(bh^2)$ , using the same feature set.



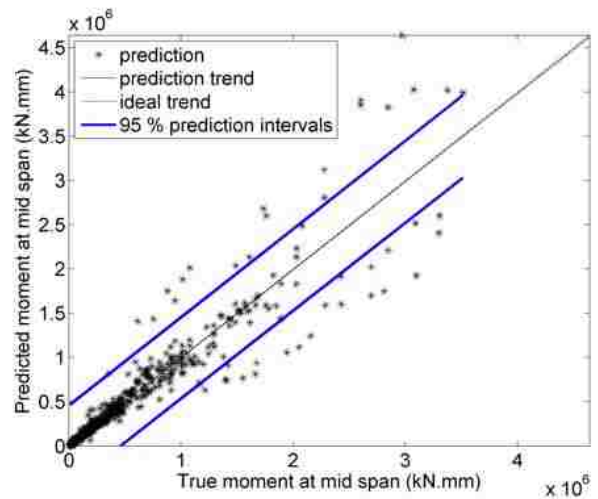
(a) Data Set 3, regression only.



(b) Data Set 3, classification followed by regression.



(c) Data Set 4, regression only.



(d) Data Set 4, Classification followed by regression.

Figure 4.16: Performance of SVM models with and without pre-classification to directly estimate absolute moment on all specimens (Data Set 3) and all 3-point bending specimens (Data Set 4) using the feature set  $I_P$ ,  $B$ ,  $N_c$ ,  $a/h$ ,  $h$ , and  $b$ .



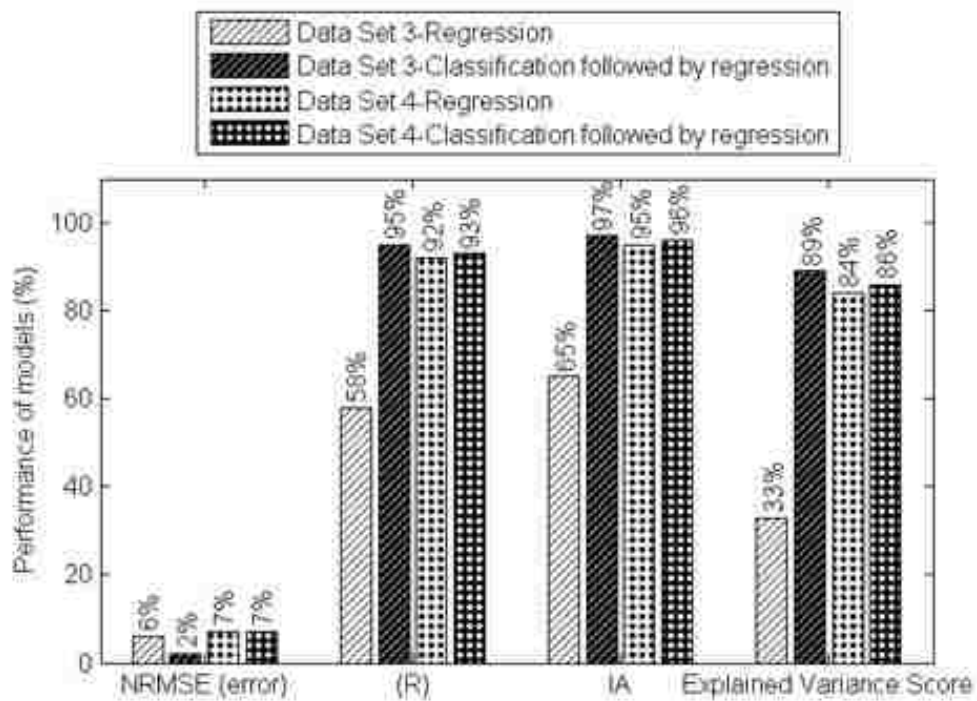
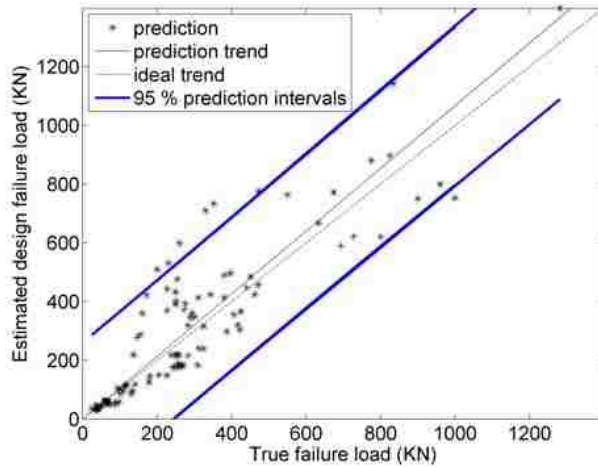
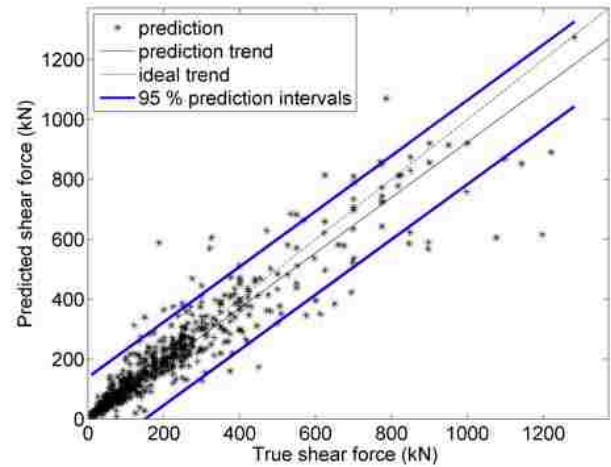


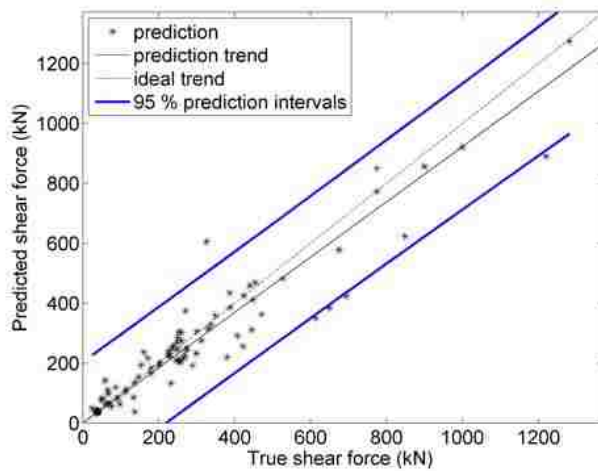
Figure 4.17: Error metrics for SVM moment estimations shown in Figure 4.16 for all specimens (Data Set 3) and all 3-point bending specimens (Data Set 4).



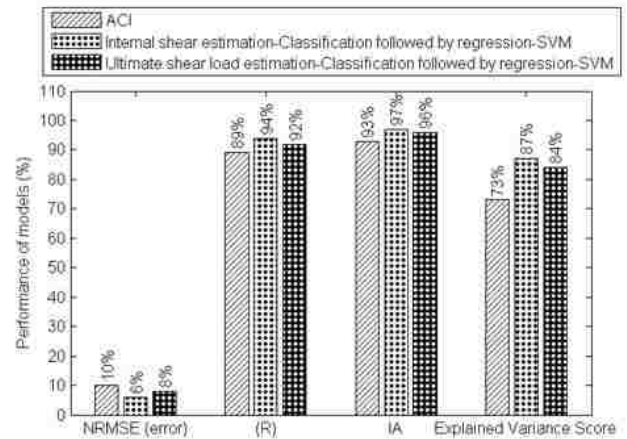
(a) ACI.



(b) Internal shear estimation, Classification followed by regression, SVM.



(c) Internal load estimation at ultimate (last) load stage, Classification followed by regression, SVM.



(d) Estimation error metrics.

Figure 4.18: Performance of ultimate load calculation by ACI procedure, image-based internal shear force level estimation, and ultimate load stage estimations using combined classification and regression model for 3-point bending specimens (Data Set 4) using the feature set  $I_P$ ,  $B$ ,  $N_c$ ,  $a/h$ ,  $h$ , and  $b$ . The ACI plot reflects the removal of an outlier unreinforced slab specimen with dimensions outside the applicability of the ACI formulas. This specimen was not removed from the image-based data set.

Chapter 5

**COMPUTER VISION-BASED STRESS STATE  
QUANTIFICATION FOR RC PANELS**

## 5.1 *abstract*

This chapter presents a computer vision damage assessment that relates surface crack patterns to damage levels and stress state characteristics in conventionally reinforced concrete and steel fiber-reinforced concrete (SFRC) panels. Previous studies have focused on crack patterns for specific structural element types such as beams and columns, but this study considers stress states in a more general and arbitrary framework. In particular, image data from previously published panel test specimens subjected to nominally constant stress have been collected to develop image-based estimation models capable of quantifying damage levels and stress components for full-panel crack patterns, and to investigate sub-image sampling strategies to approximate full-panel results using partial-panel images. The objective here is to show that the analog of representative volume elements can be extended to image-based analysis contexts. The image data sets used in this study have been obtained from five different published studies, which provided 189 crack pattern images captured from 33 concrete and SFRC shear panel specimens. Given the limited size of the data set, a feature-based computer-vision approach has been used, with various geometric attributes of surface crack patterns used to train the estimation models. Within the limits of the data available, the preliminary results presented here indicate that quantifiable correlations exist such that stress state and damage level estimation models are valid across a range of loadings (i.e., reverse cyclic and monotonic) and materials (reinforced concrete and steel fiber-reinforced concrete), and that with appropriate sampling techniques it is possible for sub-sampled images to yield estimations similar to full-panel results. These localized correlations between crack patterns and stress states potentially could be used in broader contexts for damage assessment of more general reinforced concrete and SFRC members.

## 5.2 *Introduction*

We investigate the use of image-based machine learning models capable of estimating load levels and stress states based on crack pattern data gathered from experimental panel tests originally used to successfully characterize and quantify fundamental reinforced concrete behavior for purposes of design and analysis. The motivation for this present work is to augment inspection and damage characterization in contexts in which the kinds of fundamental information needed for design and analysis are not necessarily known, thus leading to an inverse problem of sorts: determining the state of a structure based solely on external observation. Earlier similar studies [47], [26], [25] and [24] have focused on crack patterns in specific structural elements such as beams and columns, relating overall crack patterns to damage and load level measures for particular combinations of loading and boundary conditions. In this study, published results from conventionally reinforced concrete (RC) and steel fiber-reinforced concrete (SFRC) panel experiments designed to test representative volume elements under nominally constant stress are used to explore how well a similar

concept of representative image elements can be used to estimate stress and damage levels based on localized observation, potentially supporting more general applications.

Figure 5.1 provides a simplified conceptual overview of a possible future application of such an approach. As shown in the figure, observed cracking over a given region could potentially be related to estimated local stress state quantities. There are two key assumptions in the conceptual application shown in Figure 5.1: (i) at some appropriate scale crack patterns can be related to nominal states of stress; and (ii) at such scales it is possible to use representative sub-images of a structural component to characterize local aspects of the stress state and damage levels. To test these assumptions in the specific context of panel test data, we use full-panel images to set a baseline for estimation of stress state and damage level quantities, and different methods of using subsampled, partial-panel images are investigated with respect to producing similar results to the full-panel images. General subsampling approaches include both translation and rotation of sub-image blocks within the overall full-panel image. Because the stress state in a given test panel is nominally constant, any subsampled image should in principle result in a matching estimation, but in reality this will depend on many factors as demonstrated later in this study.

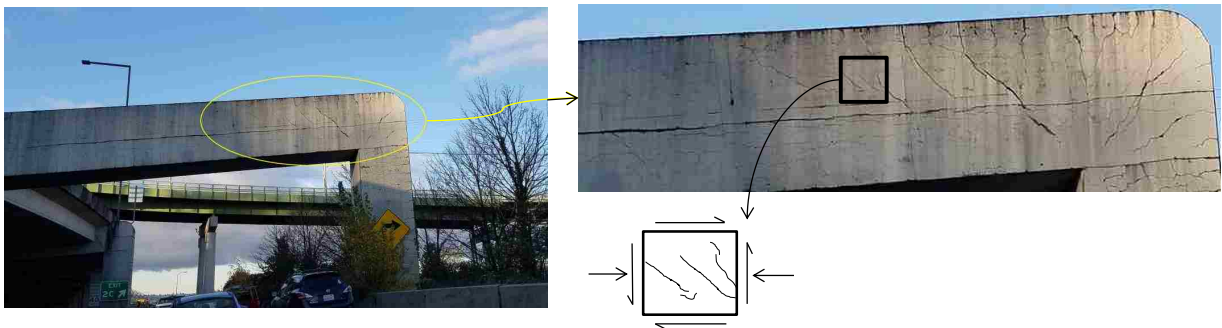


Figure 5.1: (a) Typical cracked bridge girder. (b) Developed structural cracks in the girder

### 5.2.1 Panel-Based Testing and Models

The panel test data used in this study were originally captured in the 1980s following the development of the first Panel Element Tester [78] and the consequent formulation of the Modified Compression Field Theory (MCFT) [77]. The governing constitutive models of the MCFT were developed empirically from a series of reinforced concrete panel elements tested at the University of Toronto using the Panel Element Tester (shown in Figure 5.2). In particular, the results of the panel tests led to the development of compression softening and tension stiffening relationships that were incorporated into the constitutive compressive response for cracked

reinforced concrete. More recently the Disturbed Stress Field Model (DSFM) ([75], [76]) was implemented to explicitly incorporate deformations due to crack slip in the element compatibility relationships. The MCFT has been integrated in the CSA A23.3 [8], AASHTO LRFD [1] and fib Model [32] Codes. The MCFT and the DSFM have also been implemented into a number of software packages to provide engineers with the ability to predict and/or assess the response of beams and other structural elements subjected to shear, moments and axial loads, including providing estimates of crack widths and deflections. It can be seen overall that the panel-based approach to studying and characterizing concrete response for purposes of applications in more general contexts has proven successful.

The tools derived from the panel-based approach have been demonstrated to be effective in conducting detailed analyses of existing reinforced concrete and fiber reinforced concrete structures and capturing complex aspects of the response. However, from an inspection perspective it can be difficult to relate crack patterns observed in the field to the output of these models. In fact, only very recently have attempts been made to develop mechanical models capable of estimating the stress state and reserve capacity of a structure from field measurements of crack widths, crack slips, and crack orientations ([13], [14]). While promising, such mechanical models can require information about loading conditions and structural details such as concrete strength and reinforcement ratio that can make them unsuitable for particular field inspection scenarios in which such information may not be available. This leads to the complementary approach in this study, which is to leverage panel test data for potential field inspection applications based solely on empirical correlations between crack patterns and load states determined using computer vision and machine learning techniques.

### *5.2.2 Relation to Prior Computer Vision Work*

There are now broad research efforts focused on the detection, classification, segmentation, and quantification of damage in structures using computer vision. Most recent work has involved the use of deep neural networks with various combinations of convolutional and other coupled layers (see, e.g., [19], [31], [3], [20], [33]), which is becoming the dominant approach in general image recognition and segmentation tasks across many domains. These techniques can enable the effective use of millions of fitting parameters, and with suitable training data they are capable of levels of performance comparable to human observers. These techniques also implicitly include feature extraction, which removes the need for ad hoc feature engineering. However, these techniques normally require large training sets to achieve high accuracy, and while it is possible to augment image data sets for purposes of classification, it can be problematic to use similar data augmentation techniques for physical testing results as considered in this study. Nonetheless, for purposes of identifying damage such as cracks, and for robustly segmenting crack images down to the pixel level, deep learning methods are currently



Figure 5.2: Typical shear panel test setup.

the most effective option [4], and the starting point for the work in this study is the assumption that cracks have been detected and segmented to the pixel level. This sets the stage for the next assumed task, which is to relate these segmented images of damage to quantitative estimates of the corresponding load levels and stress states.

In this respect the current work is most closely related to other recent studies designed to quantify load levels and damage states in RC beams, columns, and slabs based on surface crack patterns obtained by assembling data from earlier published experimental studies ([47], [26], [25]). Similar to these related studies, simplified crack patterns are used for damage estimation, specific geometric and textural crack features are identified and employed for damage quantification, and machine-learning models are trained to enable quantified internal shear and moment load estimation. While these previous research efforts focused on quantifying damage and load levels in specific structural components such as beams, slabs, and columns, this use of panel test data has the potential for damage and stress state estimation at a more general level. In particular, in the spirit of the successful application of panel testing to the development of widely applicable physics-based models, the current study seeks to investigate the correlation between local crack patterns and damage and stress states in RC and SFRC panel specimens designed to reproduce stress states known to arise in general applications. The quasi-continuum, nominally constant stress field induced in the panel



tests theoretically results in uniformly distributed crack patterns, which in principle makes the panel tests suitable for investigating the correlation between local crack patterns and global stress states.

This work also shares some similarity with studies that have applied computer vision techniques to concrete cylinder strength estimation based on microstructure images ([6], [85], [42]). The similarity lies in the basic approach of using generic, local image data to infer global behavior characteristics in a quantitative manner by means of correlating captured images and experimental test results. However, these studies have used much different image scales (e.g., microscale) consistent with the goal of estimating in-situ material properties rather than progressive damage assessment or load level estimation. From this perspective, the approach in this study considers intermediate scales between full structural members (or subsystems) and regions comparable in scale to those inherent in the damage mechanisms themselves (e.g., crack lengths, crack spacings), with an emphasis on damage status and load level estimates.

### ***5.3 Scope and Organization of the chapter***

The overall objective of this study is to investigate the correlations between observable surface crack attributes and corresponding generic stress states for reinforced concrete and steel fiber-reinforced concrete panel specimens at different scales. This includes identification of important image features, application of suitable machine learning algorithms, generation of estimation results for both load levels and stress components, and application of error metrics and scatter plots to assess and compare results. Both full and subsampled images are used to determine the degree to which the concept of a nominally constant stress state can be reproduced by estimates based on machine learning algorithms for varying sub-image sizes, locations, and orientations.

The chapter begins with an overview of the data set assembled from prior published panel test studies, and this is followed by several sections considering the identification and development of different scalable crack features and sampling techniques to be used later in the machine learning modeling, and to explore the possible correlations between crack features and theoretical stress transformation principles independent of machine learning considerations. This sets the stage for an overview of the machine learning implementation itself, followed by in-depth presentation and discussion of the estimation model performance results for both full-panel and subsampled image cases. The chapter closes with a summary of the results and limitations of the study, including suggestions for further work.

### ***5.4 Image Data Set***

The image data set was assembled from published monotonic and reverse cyclic tests on reinforced concrete and fiber reinforced concrete panels as described earlier. The full results of the tests used here have been



presented in five published dissertations from the University of Toronto: [80], [69], [18], [73], and [53]. These experimental tests were conducted for the purpose of investigating different effects of concrete strength, fiber properties, and reinforcement details on cracking behavior, and shear strength of panels subjected to different monotonic and reverse cyclic stress state conditions. Table 5.1 summarizes the specimen and test configuration data from the full set of experimental studies, including the range of tensile strength, compressive strength, and number of images used for each material type.

There are a number of limitations to be noted for the dataset collected in this study. The panel tests were mainly designed to model mechanical behavior of concrete and fiber materials in particular structural components such as bridge girders. Thus, the loading conditions in the panels are limited to certain combinations of axial and shear loadings similar to those arising in different structural components. For instance, all monotonic SFRC panels were subjected only to pure shear loading to model the mechanical behavior of SFRC material in structural components such as bridge girder webs, in which shear loadings are dominant. Although the shear loading produces secondary axial stresses in both horizontal and vertical directions in the panels, combinations of different primary axial and shear loadings can result in different mechanical and cracking behavior in the panels, which should be considered for future studies. Two RC panels in the collected data set were subjected to combined compression and shear loading with loading ratios of 4.0 and 2.0. Observed spalling and cracking patterns in these specimens were different from the other RC panels subjected to the pure shear loading.

Most of the images in the collected dataset were not available in raw captured image form, and cracks were marked manually in those images during the testing process. The crack line drawings in the image data did not include details like crack widths and spalling, which are known to be important in damage estimation. Although more than 200 RC and SFRC panel tests have been conducted in previous studies, this study collected images from the 33 panel tests that had full coupled load and image data, leading to a total of 189 image and load-history pairs. While this is not a large data set by normal data analytics standards, it is adequate for the focused preliminary analysis presented here.

## 5.5 Image Processing

In the context of the present study, image processing is used to analyze raw load stage images and extract quantified crack attributes. Three major steps are employed to perform this image processing:

(i) **segmentation**: image segmentation is the process by which objects of interest are extracted from general images by creating binary images (pixel intensity of 0 or 1) in which the nonzero pixels are those pixels contained in the object representation in the original image. The objects of interest in the current context are cracks, and so segmentation results in black-and-white images containing crack and non-crack

Material type	No. Specimens	No. images	No. monotonic spec	No. cyclic spec	$f_t$ (MPa)	$f'_c$ (MPa)
Concrete	11	60	9	2	2.04-4.35	38.5-90
Fiber reinforced concrete	22	129	15	7	3.32-4.99	43-79
Overall	33	189	24	9	2.04-4.99	38.5-90

Table 5.1: Summary table of prior experimental study data sets.

pixels. In the majority of cases in this study the crack objects in the collected images were marked manually during the experimental tests, and so were effectively pre-segmented. As a result, it was not necessary to apply advanced segmentation techniques. For those cases in which segmentation was necessary, simple edge detection ([15]) and manual cleaning of the resulting images were adequate for the purposes of this study, the focus of which is post-segmentation analysis.

(ii) **morphological operation:** Because so much of the image data set was only available in the form of simple line drawings, for consistency all the image data were converted to the lowest common denominator equivalent of simple line drawings. To this end, an additional morphological operation was performed to remove crack width, and force all crack objects to have a width of one pixel in the images regardless of the image source or scale [26].

(iii) **feature extraction:** this step involves representing crack properties in segmented binary images using quantified crack attributes to allow for consistent analysis in lower dimensional spaces than those of full images. The primary set of 22 geometric and textural crack image features considered in this study matches that used in [25] and [26] with minor modifications. Some features (such as total crack length and polar moment of inertia of crack pixels) were scaled based on the size of the image to allow for consistent feature extraction for different resampled image blocks. In addition to these primary features, two additional features were also included for this particular study: (i) average distance between cracks; and (ii) average orientation of crack segments. The computation of these additional features is described next.

Figure 5.3 outlines the process for calculating the total and average distance between cracks. First, the distance transform of the binary segmented image is computed using MATLAB's `bwdist` function in the image processing toolbox [56]. For each pixel in the segmented image, the distance transform computes the distance between that pixel and the nearest nonzero pixel. The calculated distance values are assigned to each pixel in the distance transform matrix. Figure 5.3(a) shows an example colormap representation of the distance transform for a representative segmented image (the cracks themselves have been superposed in

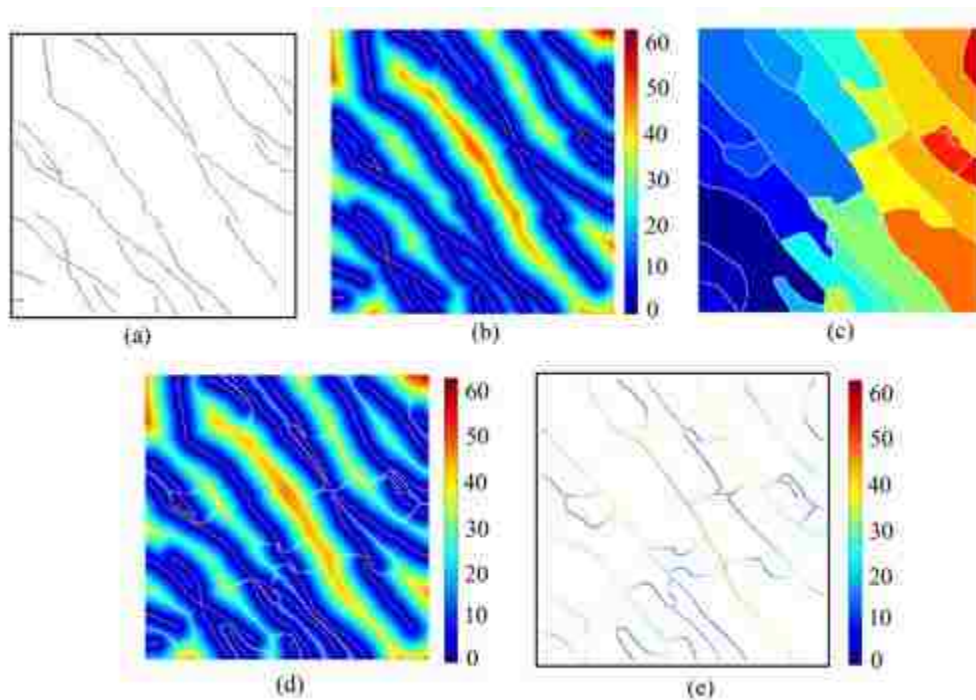


Figure 5.3: (a) Typical crack pattern; (b) Distance transform of a typical segmented image; (c) Watershed transform of the distance transform; (d) The watershed regions used to calculate average distance between cracks; (e) Distance values in the distance transform image along watershed ridge lines.

white). The color bar indicates the computed distance values. Then, the watershed transform in MATLAB's image processing toolbox is used to convert the distance transform image to a label matrix that identifies the "watershed ridge lines" surrounding each crack object. This is illustrated in Figure 5.3(b) in which each region surrounding a crack object is identified with a certain color. Figure 5.3(c) clarifies the procedure by showing watershed ridge lines along with crack objects in the same figure. The watershed ridge lines identify the farthest points between cracks. The total distance between cracks can be computed by adding distance values at pixel locations along the watershed ridge lines in the distance transform matrix. Similarly, the average distance between cracks is the total distance between cracks divided by the total number of pixels corresponding to watershed ridge lines. Figure 5.3(d) indicates the distance values in the distance transform image at location of watershed ridge lines.

RC and SFRC specimens are vulnerable to cracking under tensile stress due to the mechanical behavior of concrete as a material. Orientation of cracks is an important feature that in panel test contexts can be interpreted as a representation of overall principal stress and strain orientations in the specimens. To enable

the use of crack orientation as a feature, overall average orientations of cracks were calculated based on a weighted average of the crack objects as shown in Figure 5.4. In particular, the orientations of crack objects are computed using MATLAB's regionproperties function in the image processing toolbox. The crack objects with orientations falling in the interval of two standard deviation from the initial mean crack angle were considered for calculating the final average orientation of cracks ( $\theta_c$ ) as indicated in Figure 5.4.  $\theta_c$  is defined based on a weighted average formula as follows:

$$\theta_c = \frac{\sum_{-s}^s l_i \theta_i}{\sum_{-s}^s l_i} \quad (5.1)$$

in which  $l_i$  and  $\theta_i$  are the length and the orientation of crack objects in the image, and  $s$  is the standard deviation of the crack angles.

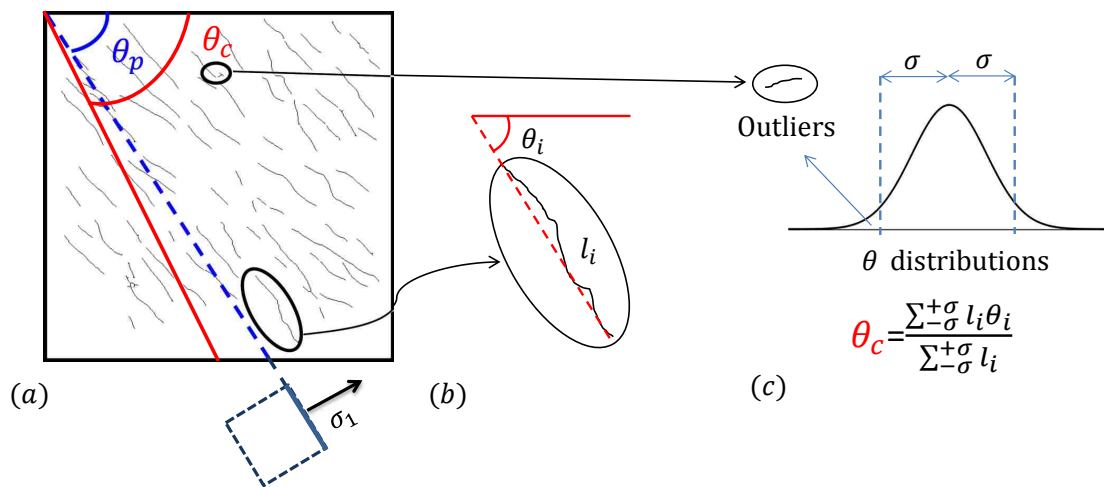


Figure 5.4: Angle calculation of crack object. (a) Typical crack pattern image.  $\theta_p$  is the theoretical calculated principal angle corresponding to the plane of maximum tension,  $\sigma_1$ .  $\theta_c$  is the averaged crack angle based on image analysis. (b) Individual crack segment orientation,  $\theta_i$ , and length,  $l_i$  calculated using MATLAB regionprops function; (c) Crack objects outside the range of one standard deviation away from the mean orientation were considered outliers, and removed from overall average angle calculation; (d) final weighted average calculation.

## 5.6 Principal Stress/Strain Orientation (Theoretical vs Image-Based)

This section compares the adjusted average orientation of cracks based on image processing with the theoretical principal stress and strain planes for the concrete and fiber reinforced concrete panels in the data set. As explained in the previous section, the empirical overall orientation of cracks was calculated based on a

weighted average of crack objects falling in the interval of two standard deviation from the initial average angle of the crack objects. The theoretical principal stress and strain orientations were calculated directly using stress and strain transformation equations, with known target stress and strain values generated from the panel tests.

Figures 5.5 and 5.6 show overall orientation of cracks based on theoretical principal stress and principal strain plane orientations compared to image-based average crack orientation for typical monotonically loaded RC and SFRC load stage images. The applied stress states along with orientation of reinforcement are also shown for each panel. As can be seen in the figures, principal orientations rotate from left to right for each row of load stage images as the stress level increases, with mismatches in the principal stress and strain orientations. The applied stress level leads to increased induced critical stress, and eventually failure in the panels. As mentioned earlier, the SFRC panel specimens used here were subjected to pure shear loading while the RC panels were tested under both pure shear and more general stress states. As can be seen in the figures, there is good agreement between image-based and theoretical principal orientations in RC and SFRC panels subjected to pure shear loading. However, image-based orientations deviate from theoretical orientations for some of the load stages in combined axial and shear loading conditions as shown in Figure 5.6(b) and (c). The cracks appear to develop in different directions due to the combined stress state condition as loading increases left to right for each row of images as shown in Figure 5.6(b) and (c). Although the SFRC panels were not subjected to combined stress states similar to the RC panels shown here, there is generally better agreement between image-based and theoretical principal orientations in the SFRC panels compared to the RC panels. The fiber material used in the SFRC panels improves the tensile strength of the concrete/fiber composite, which results in more evenly distributed cracks in the SFRC panels compared to the RC panels, as can be seen in Figures 5.5 and 5.6. It was also found that there is better agreement between image-based and theoretical principal orientations in SFRC and RC panels as more cracks form, which makes sense given the averaging approach used to quantify the image-based orientation.

### **5.7 Data Subsampling**

In order to generate different stress states and understand the consistency of subsampling images in a nominally constant stress state field, different sampling techniques were employed to generate more data observations. As shown in Figure 5.7, four distinct sampling methods were used to resample sub-image stress blocks from the concrete panels. In the first resampling approach, horizontal, fixed-size image blocks were randomly resampled from the panels. In the second approach, horizontal image blocks having different dimensions were randomly resampled. In the third case, fixed-size image blocks having different orientations were resampled from the entire panel. In the fourth and final case, image blocks having random sizes and

orientations were resampled from the panels.

Different numbers of samples and subsampled block sizes were considered to evaluate the consistency of the models in different estimation tasks, which will be discussed in the machine learning section below. It is worth noting that there can be some overlap between resampled image blocks for each load stage, but this does not affect the performance of the estimation models since specimen-level cross validation [26] is used to generate the training data, involving test data set from all resampled blocks. Data subsampling was also used to generate sampled blocks with different sizes and orientations to identify effective normalized features, as explained in more detail in the next section.

## 5.8 Feature Engineering

Because deep learning has not been used here, it is necessary to identify effective crack features with minimum dependencies on size and orientation of sampled image blocks. The candidate normalized features should have two main characteristics: (i) convergence as sub-image size increases; and (ii) efficiency for use in machine learning models. As mentioned earlier, the full set of 22 geometric and textural crack features found to be useful for machine learning analyses in [25] and [26] were also used in the present study. Depending on the machine learning model in use, it often was preferable to use a subset of dominant features rather than the full set of features. Some of these features, such as homogeneity, aspect ratio, average distance between cracks, and overall crack orientation are already normalized, and are not dependent on the size and orientation of the subsampled image blocks. However, some other features such as total crack length and polar second moment of area of crack pixels should be scaled based on the size of the subsampled image block to allow for consistent features across the panel. To this end, the total crack length and polar second moment of area of cracks were scaled based on  $a^2$  and  $a^4$  respectively, in which  $a$  is the size of the subsampled image block.

Figure 5.9 illustrates the effects of the sampling size and orientation on different crack features including aspect ratio, average distance between cracks, total crack length, and polar second moment of area of cracks. For each sub-image block size, 10 image blocks were sampled from a typical SFRC crack pattern image across the entire panel based on sampling technique III (random orientations, fixed size). As can be seen in this figure, all crack features converge as sub-image block size increases from 0 to 100% of the panel size. It is also evident that as the block size increases, the crack features become similar to the features extracted from the entire panel. In order to evaluate the accuracy of the normalized features, resampled block features should be compared with the features extracted from the entire panel without sampling. To this end, normalized mean absolute error (NMAE) was calculated for all RC and SFRC sampled blocks for a range of 0-100% of the panel size as shown in Figure 5.10. These error plots represent the overall behavior of the identified

normalized crack features, and the tradeoff between accuracy and block size for different crack features.

Orientation of cracks is another useful feature which should not be dependent on size and orientation of the resampled sub-image blocks. Figure 5.11(a) compares the orientation of cracks based on image processing with theoretical principal planes for subsampled SFRC blocks using sampling technique II. This comparison is shown with a prediction versus true value scatter plot in which crack orientation based on theoretical principal planes was considered as the ground truth. As is evident in this figure, the overall crack orientation range in the test data is small, which influences the overall estimation accuracy. In order to broaden this range, sampling technique IV was used to generate rotated blocks with different angles. Figure 5.11(b) shows the comparison between the image-based and theoretical principal planes approaches to estimate the orientation of cracks for subsampled SFRC blocks using technique IV. Although sampling technique IV resulted in a better estimated fit, the margin of error corresponding to 95% prediction intervals are almost the same for both sampling techniques, which shows the consistency of the image-based model in estimating the overall crack orientation. For the RC data set, similar results can also be obtained, which shows the consistency of the presented approach in estimating overall crack orientation using image processing.

## 5.9 Crack-Based Estimation Models

With coupled image and load data, features sets, and sub-sampling strategies in hand, machine learning models can be developed to quantify correlations between image data and load levels and stress state characteristics for both full-panel and subsampled images. The primary load level metric used here is the failure ratio, which refers to the level of loading relative to the panel's ultimate load capacity (e.g., failure ratio = 0.7 means the panel has reached 70% of its actual load capacity). In monotonic load cases the failure ratio also can serve as a damage metric, indicating how close the current loading is to capacity.

The crack-based model is intended to estimate the transformed stress components based on the local crack patterns for sampled image blocks with varying orientations. For model training, stress state transformation and Mohr's circle concepts can be used to calculate theoretical  $\sigma_{x'x'}$  and  $\tau_{x'y'}$  values based on the known nominal stress states applied on the panels and the known sub-image orientation as shown in Figure 5.8(b) and (c). Using training data enables the machine learning model to correlate crack pattern features (including relative crack orientation and image orientation) and load levels and stress invariants. Once the model has been trained, Figure 5.8(d) illustrates the pipeline used to estimate the failure ratio and the stress state for a typical test resampled image block marked with a star in the relevant sub-figures. During estimation the image block is the only input fed into the pipeline, at which point the relevant features are extracted, and the estimate made based on these feature values. It should be noted that at this point the sampling orientation,  $\theta_s$ , is not considered in the feature set to ensure there is no reliance on data that would not be known in



field applications (i.e., the background stress state and its “natural” orientation would be unknown, so while  $\theta_c$  can be calculated based on the image itself,  $\theta_s$  would be unknown). As shown, the output of the trained estimation model can include the failure ratio (FR in the figure), stress state invariants such as the maximum in-plane shear stress,  $\tau_{\max}$  and the minor principal stress,  $\sigma_2$  (the major principal stress is assumed to be linked to the effective cracking strength of the concrete), and the normal and shear components of stress aligned with the image block,  $\sigma_{x'x'}$  and  $\tau_{x'y'}$ . Model performance can be evaluated by comparing model estimates with known values for data taken from test sets independent from the training data.

### 5.10 Model Analysis and Performance of Estimation Models

In order to evaluate the performance of the crack-based models in quantifying load levels and stress-state characteristics, various goodness-of-fit statistics and cross validation methods commonly used for regression analyses were considered. The regression accuracy metrics quantify in summary fashion how well the estimated damage and stress components values fit observed data. Since the overall distributions of dependent values in the data used in the regression analyses here are not uniformly distributed, multiple error measures having different sensitivities to outliers were used. The regression metrics and visualization approach considered in this study are the same as those used in [25] and [26]: (i) Normalized root-mean-square error (NRMSE) represents the normalized difference between estimated and true variables; (ii) Index of Agreement (IA) is a non-dimensional and bounded error measure ranging between 0 (no correlation) and 1 (perfect correlation) which considers proportional differences in the true and estimated means and variances; (iii) Correlation Coefficient (R) measures the strength and direction of a linear relationship between two variables; and (iv) Explained Variance Score is the ratio of residual variance to the variance of the true data set. In addition to these scalar error measures, regression estimation results also are presented using scatter plots providing visualization of estimated values along with ideal line, fit-through-origin trend line, and confidence intervals corresponding to 95% confidence level.

In order to assess the performance of the estimation models, 10-fold cross-validation was used to define independent training and test data sets. Similar to earlier studies [25] and [26], specimen-based cross-validation was used to evaluate the performance of the models. Specimen-based cross-validation corresponds to a way of distributing training and test data sets such that training images (load stage images) for a particular specimen do not also appear in the test image set. This differs from image-level cross validation in which training and test images can correspond to the same specimen, albeit at different load stages. Specimen-level cross validation leads to reduced accuracy compared to image-level sampling, but machine learning results based on specimen-level sampling are more realistic relative to the practical field scenarios envisioned here [26]. Importantly, specimen-level sampling allows for the evaluation of the extrapolation



performance as it does not exclusively do an interpolation task.

### 5.11 Estimation Results

This section presents an evaluation of the accuracy of the crack-based estimation models in quantifying load levels and stress state components for both RC and SFRC panel specimen data sets. In particular, the effects of varying major modeling components are investigated, including loading conditions (monotonic and reverse cyclic), resampling approaches (translated and/or rotated subsampled block), machine learning algorithms, and crack feature sets. The estimation models were built using various machine-learning-based regression techniques including Gaussian process [68], J48 [67], and Support Vector Machine (SVM) [37] along with distinct features discussed in the feature engineering section. Results will be presented for the best-performing algorithms for each case, but in most instances the variation in results for different algorithms were modest (see [26] for algorithm comparisons in related contexts). For both monotonic and cyclically loaded specimens, failure ratio (i.e., current load level divided by ultimate strength of the panel) is used as a normalized value to quantify damage states in RC and SFRC panels.

The first subsection of the results describes the effects of using different resampling approaches on failure ratio estimation model accuracy for combined monotonic and cyclic RC and SFRC panels for full-panel and subsampled image data. The second subsection presents the results of crack-based model estimation of stress state components for both resampled sub-image blocks and full panel images. Stress component estimation focuses on monotonic loading only, with RC and SFRC cases handled separately. This is due to the fundamental difference in cyclic loading stress state interpretation relative to monotonic loading, and the low correlations between SFRC and RC behavior coupled with unbalanced data sizes.

#### 5.11.1 Effects of sampling on estimation model convergence

Different sampling approaches were used to generate sampled image blocks to understand the consistency of subsampling images in panels which are in a nominally constant stress field. In theory, all resampled blocks with different ranges of size and orientation from a given panel should represent the same load level or stress state, which should correlate with the full-panel case. As explained in the Feature Engineering section, normalized features were defined to reduce the effects of sampling size on the crack features. The goal of this section is to investigate how machine learning models trained with those normalized features perform for varying sub-image block sizes in different training and test scenarios.

Figure 5.12 presents the performance of SVM failure ratio estimation models trained and tested using different subsampled and/or full-panel image data to investigate the effects of sample block size, sampling and orientation strategies, and training/test set identification approaches. Index of Agreement is used in

Figure 5.12 as the primary error metric to compare the performance of models in different cases, but the results would be similar for any of the other error metrics used in this study. In Figure 5.12, a total of six different cases are considered in which training and test data sets were generated from resampled (sub-image) blocks and/or full panel images using varying relative sample block sizes,  $a_s$ , and orientations,  $\theta_s$ . The load stage images were resampled 10 times with a sample orientation range of  $0^\circ$  to  $90^\circ$  ( $0^\circ \leq \theta_s \leq 90^\circ$ ) for all cases other than the full panel ( $a_s = 100\%$ ) cases for which sample orientation can only be equal to 0 ( $\theta_s = 0$ ). The relative sample block size has been randomly selected in the range from the reference sample block size,  $S_R$ , to the full-panel size ( $S_R \leq a_s \leq 100\%$ ) for the cases other than the fixed sized sample block size cases (i.e.,  $a_s = S_R$  and  $a_s = 100\%$ ).

As can be seen in Figure 5.12, the performance of the estimation models for all cases converges as expected to the full-panel result (Case 0) as the range of the reference sample block size approaches 100%. The sampling block size is a key factor determining the performance of the models, but sampling strategy is also important. As is evident in the figure, models trained using data resampled over a range of sizes and orientations show better convergence performance compared to models trained solely using entire panels. As mentioned in the Feature Engineering section, sample block size directly influences the crack features that can be detected. Since no “full-panel” case exists for structural components in the field, varying relative sample block size in the test set (case I and II) appears to offer the best approach to develop models in potential field scenarios. Comparing the performance of the failure ratio models for case I and case II in Figure 5.12 illustrates that including different subsampled image block sizes could potentially improve the performance of estimation models by helping the models learn the error relevant to sample block size and orientation.

Index of Agreement provides one perspective for comparing the performance of models for the different cases in Figure 5.12. Another perspective is provided in Figure 5.13, which illustrates the full performance of SVM failure ratio estimation models using scatter plots for some of the key cases in Figure 5.12, while Figure 5.14 summarizes all error measures for the cases shown in Figure 5.13(a)-(d). In Figures 5.13 and 5.14, the benchmark performance of failure ratio estimation models is again defined as Case 0, in which the entire panel is used for generating both training and test data sets. Figure 5.13(b) shows the worst-case performance of estimation models in Figure 5.12 trained with full panels and tested based on a resampled block size of  $a_s = 10\%$  (this corresponds to the left-most point of the Case III curve in Figure 5.12). As shown in Figure 5.14, the Explained Variance Score for this case was found to be 0 which indicates a very weak correlation between estimated and true values. The range of the estimated failure ratios corresponding to the ultimate load stage (i.e., true failure ratio=100%) is as wide as the entire estimation range, which indicates the estimated failure ratios are essentially randomly distributed without any correlations, as is

evident in Figure 5.13(b). In effect, sample image sizes that are too small can easily land between cracks and thus have no data to work with.

Figures 5.13(c) and (d) show the results of SVM models for the first two points on the case IV curve in Figure 5.12 trained with reference sample block sizes of  $a_s = 10\%$  and  $a_s = 20\%$  respectively, and tested on full-panel images. As is evident in these figures, model performance can be significantly improved as minimum sample block size,  $S_R$ , increases from 10% to 20%. The range of the estimated failure ratios is noticeably larger in Figure 5.13(d) compared to Figure 5.13(c), which shows the superior performance of Case IV with  $a_s = 20\%$  compared to the Case IV with  $a_s = 10\%$ . This is also evident in the error measure results in Figure 5.14. For these kinds of panel tests, a relative block size of 20% appears to be appropriate for a minimum sample block size to reduce the effect of sampling size on the performance of estimation models. The estimation results based on the case I sampling strategy is a better indicator of the accuracy achievable in potential field application scenarios in which full-panel size and stress block size are not known for a given structural component. For the same reason, the Case I sampling strategy is used to evaluate the performance of the stress state estimation models, which will be explained in more detail in the next section.

### 5.11.2 Stress state estimation results

The results presented so far have focused on estimating failure ratios for resampled image blocks or full-panel images. The same framework can be used to estimate stress state components based on crack patterns. As noted earlier, stress state transformation and Mohr's circle concepts can be used to calculate  $\sigma_2$ ,  $\tau_{\max}$ ,  $\sigma_{x'x'}$  and  $\tau_{x'y'}$  based on the nominal stress states applied on the panels. All resampled image blocks with different sizes and orientations for a given panel image correspond to the same nominal  $\sigma_2$  and  $\tau_{\max}$  as illustrated in the Mohr's circle in Figure 5.8. Similarly,  $\sigma_{x'x'}$  and  $\tau_{x'y'}$  for each resampled image block can be calculated based on stress state transformation rules. The goal of the crack-based model presented here is to show that normalized stress components can be estimated based solely on crack patterns using new resampled image blocks. As shown in Figure 5.8, different crack pattern features including overall orientation of cracks were used to train the stress component estimation models.

Knowing some internal properties such as the concrete tensile strength,  $f_t$ , can enable stress state quantification if the scaled stress state is estimated. It was found that maximum in-plane shear stress,  $\tau_{\max}$ , and minor principal stress,  $\sigma_2$ , can be estimated with higher accuracy if they are normalized based on  $\sqrt{f_t}$ . Figure 5.15 presents results of Gaussian process models to estimate  $\tau_{\max}/\sqrt{f_t}$ ,  $\tau_{xy}/\sqrt{f_t}$ , and  $\sigma_2/\sqrt{f_t}$  for full-panel images for the monotonic data set. As mentioned earlier, combining reverse cyclic and monotonic data sets does not make sense in this case, as stress state transformation and Mohr's circle concepts for monotonic and cyclic loadings are not the same. Figure 5.15(b) shows the presented image-based model can

estimate scaled shear stress for combined RC and SFRC data sets with relatively good performance. The image-based model also can estimate maximum scaled shear and principal stresses, which indicates a strong correlation between the overall stress state of the panels and the crack pattern. As can be seen in this figure, the overall distributions of scaled shear stress and maximum scaled shear stress are different because the monotonic data used here were subjected to different combinations of normal and shear loadings. However, the maximum scaled shear and minor principal stress estimations lead to similar estimation trends and scatter plots, which indicates these two quantities are directly correlated to each other as expected from the perspective of Mohr's circle and stress transformation concepts.

Figure 5.16 presents results of Gaussian process and linear models to estimate simultaneously  $\sigma_{xx}/f_t$  and  $\tau_{xy}/f_t$  using the monotonic fiber data set (full-panel), and trained with dominant features. Pseudo 3-D scatter plots are used to show 3-way correlations between estimated and true stress components on 2-D plots. The  $x$  and  $y$  axes in the scatter plots show the estimated versus true values for one of the stress components, while the color bars on the right side of the plots show the absolute error for another stress component. For instance, Figure 5.16(a) simultaneously shows estimated scaled shear stress ( $\tau_{xy}/f_t$ ), and the absolute error for scaled normal stress ( $\sigma_{xx}/f_t$ ) estimation. Index of Agreement and Explained Variance error measures are also provided for each plot to evaluate the performance of the models. The results show the overall good performance of the models in estimating stress states in the SFRC panels based on dominant features. For RC panels, similar models can be used to estimate the stress components. However, it was found that the performance of the model in estimating the stress components for the SFRC data set is better than the RC data set because the RC data set includes specimens with combined normal and shear stress while the SFRC data set includes only pure shear loading panels. Combined normal and shear loading conditions result in different stress states and crack patterns than pure shear loading cases, which makes it more challenging for the models to learn the differences based on the crack features used here, especially with fewer specimens subjected to combined normal and shear stress. Other crack features, such as crack width, seem to be more important in such combined loading conditions for RC specimens in which the crack width and spalling are key factors relating the damage to the stress state.

As discussed earlier, image data were subsampled to generate different stress components for different image blocks. Results in the Feature Engineering section indicated normalized crack features can be affected by the size of the resampled blocks. However, as shown in the failure ratio estimation results above, the estimation models can potentially learn the effects of sample block size and orientation if different subsampled blocks are included in the training set. Thus, both training and test data sets were resampled with different sizes and orientations. The generated stress components are directly correlated with the crack density and overall orientation of cracks in the resampled image blocks. In order to estimate stress components, all crack

pattern features, including the dominant features used before and overall orientation of cracks were used to train the models. Figure 5.17 presents pseudo 3-D scatter plot results of SVM models to estimate  $\sigma_{x'x'}/f_t$  and  $\tau_{x'y'}/f_t$  on the sub-sampled SFRC and RC data sets. As illustrated in Figure 5.8, each data point in the plots indicates estimated stress components ( $\sigma_{x'x'}/f_t$  and  $\tau_{x'y'}/f_t$ ) on a rotated sampled image block. The full-panel images were resampled 10 times to generate different stress components for sub-image blocks with a sample block size range of 30% to 100%, and sample orientation range of  $0^\circ$  to  $180^\circ$ .

RC specimens with combined compression and shear loading were excluded from the monotonic RC data set used here because those specimens showed a completely different cracking behavior due to the large amount of compressive normal stress applied during the test, and there was insufficient test data to adequately train for these cases. The same interpretations can be made here about the effect of sampling orientation range on performance of the models as mentioned above. However, Figure 5.17 indicates the overall superior performance of the crack-based models to estimate the stress components in the SFRC data set compared to the RC data set due to a key difference in cracking behavior in concrete and the fiber material. Adding fibers to RC specimens improves the effective tensile strength of the members, which results in more uniformly distributed crack patterns. This allows for more consistent feature extraction in resampled image blocks, which also improves the performance of estimation models. Results in this figure show that the crack-based model can estimate stress components for resampled SFRC and RC image blocks with any rotations. This indicates the potential capability of the model trained with different resampled blocks to estimate different stress components based only on local crack patterns in a given structural component.

### 5.12 Summary and Conclusions

The work in this study has focused on two related themes involving computer vision applied to images of damaged concrete: (i) baseline load-level and stress state estimation for concrete panel test specimens; and (ii) convergence behavior of subsampled images taken from nominally constant stress fields. The baseline load-level results provide initial correlation accuracy levels achievable using crack field characteristics to estimate corresponding load level and stress state components based on full images of test panel specimens. These full-panel results were also used as target values for estimations based on sub-sampled images, with a focus on identifying sampling strategies and feature sets leading to rapid convergence (i.e., close matching to full-panel results with relatively small sub-images). The overall purpose has been to lay the groundwork for potential future field applications in which only partial observation of damaged infrastructure is possible or convenient, requiring global estimation based on local data.

This work is intended to be preliminary but foundational. It is preliminary in that there are many limitations to the scope of what has been done, but it is foundational in that it demonstrates basic feasibility

of local-to-global damage estimation based on panel testing with and without partial image sampling. Within this restricted scope the principal results and conclusions of this study can be summarized as follows:

- Correlations between cracked panel images and recorded load levels can be used to train machine learning models to make quantitative estimations of relative load levels based on simplified segmented crack images with accuracies comparable to other structural subsystems.
- Panel-specific image features such as average crack orientation and mean distance between cracks are useful in improved estimation accuracy, and average orientation data also can be used to estimate other characteristics of the nominal stress state, such as principal orientations and Mohr's circle constructions.
- The crack patterns associated with the fiber-reinforced concrete specimens were more amenable to the image-based analysis used here, largely because of the more even distribution of relatively large numbers of cracks compared to conventionally reinforced specimens, in which damage tended to be concentrated in smaller numbers of relatively large cracks. Crack segmentation that preserves crack widths and includes spalling likely would improve estimation for conventionally reinforced cases.
- For the data set considered here, image subsampling using varying sub-image sizes and orientations led to estimation accuracies similar to full-panel results above a threshold minimum sample size of about 20% the overall panel size. Such a threshold sample size would likely vary for different data sets depending on crack pattern characteristics.
- Applying models trained with full-panel data to sub-image data leads to poor estimations relative to using sub-image trained models on either full or sub-panel data.
- Most results show only moderate sensitivity to machine learning algorithm choices and parameter settings.

There also are a number of important caveats, limitations, and opportunities for future extensions for the work presented here:

- While in some ways the test series data gathered for this study were quite general (e.g., fiber-reinforced and conventionally reinforced, monotonic and cyclic loading, pure shear and shear/normal loading, etc.), because the previously published studies were undertaken for completely different purposes, the amount of data covering these various cases was often sparse and not distributed evenly. Any data-driven technique requires adequate amounts of data to become reliable, and so a great deal of further testing would be needed to make any of the preliminary observations made here applicable for more general use. In addition to generating more data for the kinds of cases considered here, such

further testing would require investigation of other key parameters such as loading patterns, panel dimensions, concrete types, and reinforcement characteristics.

- The normalized features used here to characterize the crack patterns, and the learning algorithms used to build the estimation models were shown to be effective, but this does not mean they are necessarily optimal. In particular, manually engineered image feature sets of the type used here are currently being replaced by convolution-based deep learning models, and so it is likely with sufficiently large data sets in the future, estimation accuracy could be improved.
- The image standardization process used in this work sacrificed damage features such as crack width and spalling, which are known to be important in assessing damage, especially for RC specimens subjected to combined normal and shear loadings. Expanded access to raw image data collected during testing combined with more refined segmentation would make it possible to increase the generality of the standardized image content, which likely would improve overall estimation accuracy and model applicability.
- Physics-based numerical models could potentially provide a means to generate larger data sets, as well. Such models would need to be capable of reliably generating crack patterns and crack characteristics similar to those observed experimentally, with machine learning outcome metrics being part of the criteria defining this similarity.

It is also worth noting a couple general issues that extend beyond the specifics of this particular study. Machine-learning models in their current form do not tend to fail gracefully, and so care must be taken after "training data" in limiting applications to circumstances closely matching the training data. It is well known that machine learning models have great difficulty with generalization. Similarly, cracks are not ideal markers for load levels in general contexts, as they are not necessarily distributed smoothly spatially or in time, they can be caused by various phenomena other than applied loading, they are much less useful in non-brittle materials, and in some cases they can become difficult to see following load removal. Thus, while crack image analysis of the kind presented here could lead to useful engineering applications when applied within appropriate boundaries, combining computer vision with other sensing modalities should be pursued, as well.

In the spirit of aiding reproducibility and further development in this field, the collected data and processing tools used in this study are available online in [dataset] [24], and at the following URL:

[github.com/Rouzbehe/RC-Panel-Data-Driven-Image-Based-Damage-Stress-Estimation/](https://github.com/Rouzbehe/RC-Panel-Data-Driven-Image-Based-Damage-Stress-Estimation/).



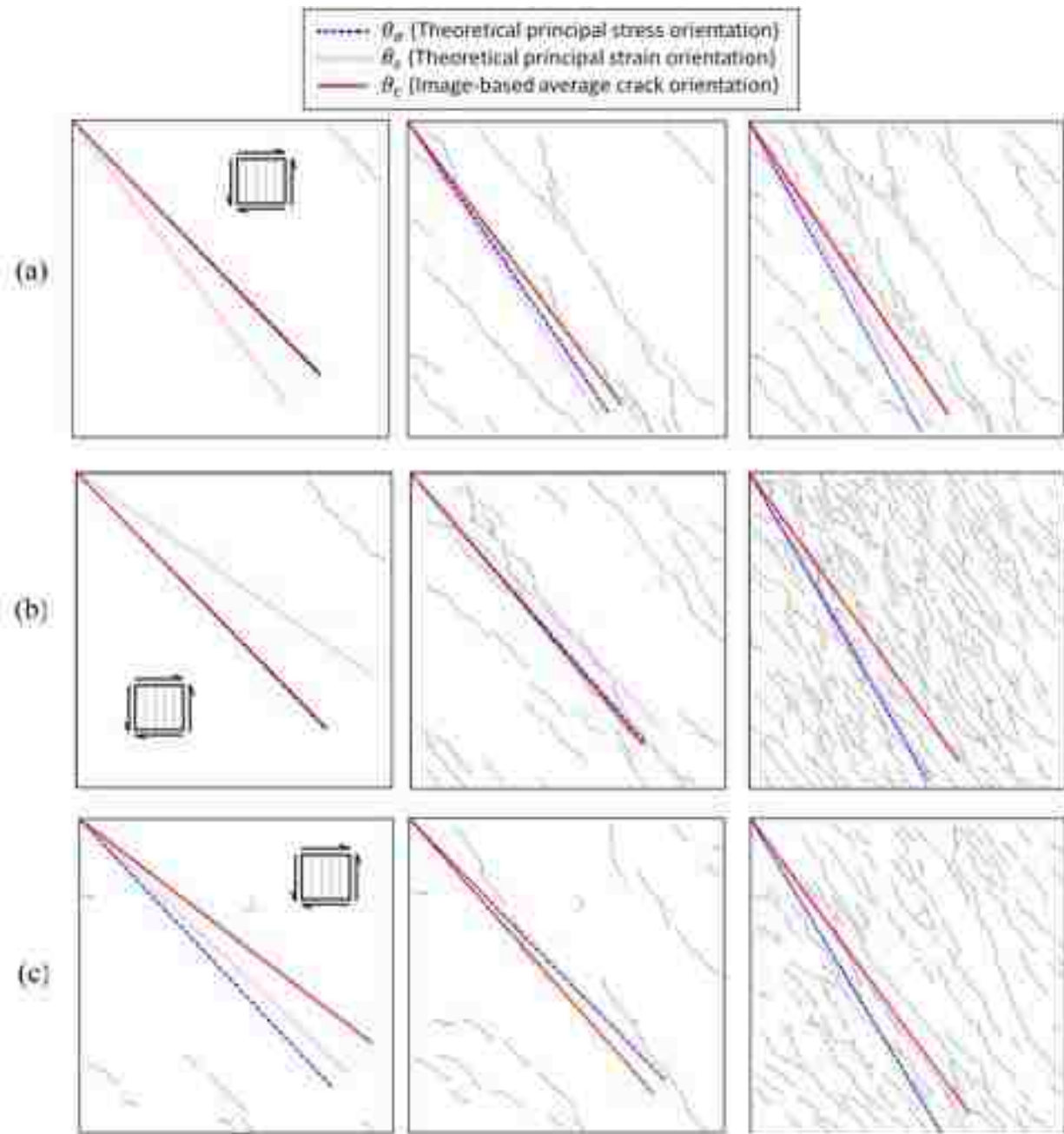


Figure 5.5: Comparison between  $\theta_p$  calculated based on the nominal state principal orientation, and overall orientation of cracks calculated using image processing for representative fiber reinforced load stage images. Loading increases left to right for each row of images, which results in increased induced critical stress, and thus rotating principal orientations.



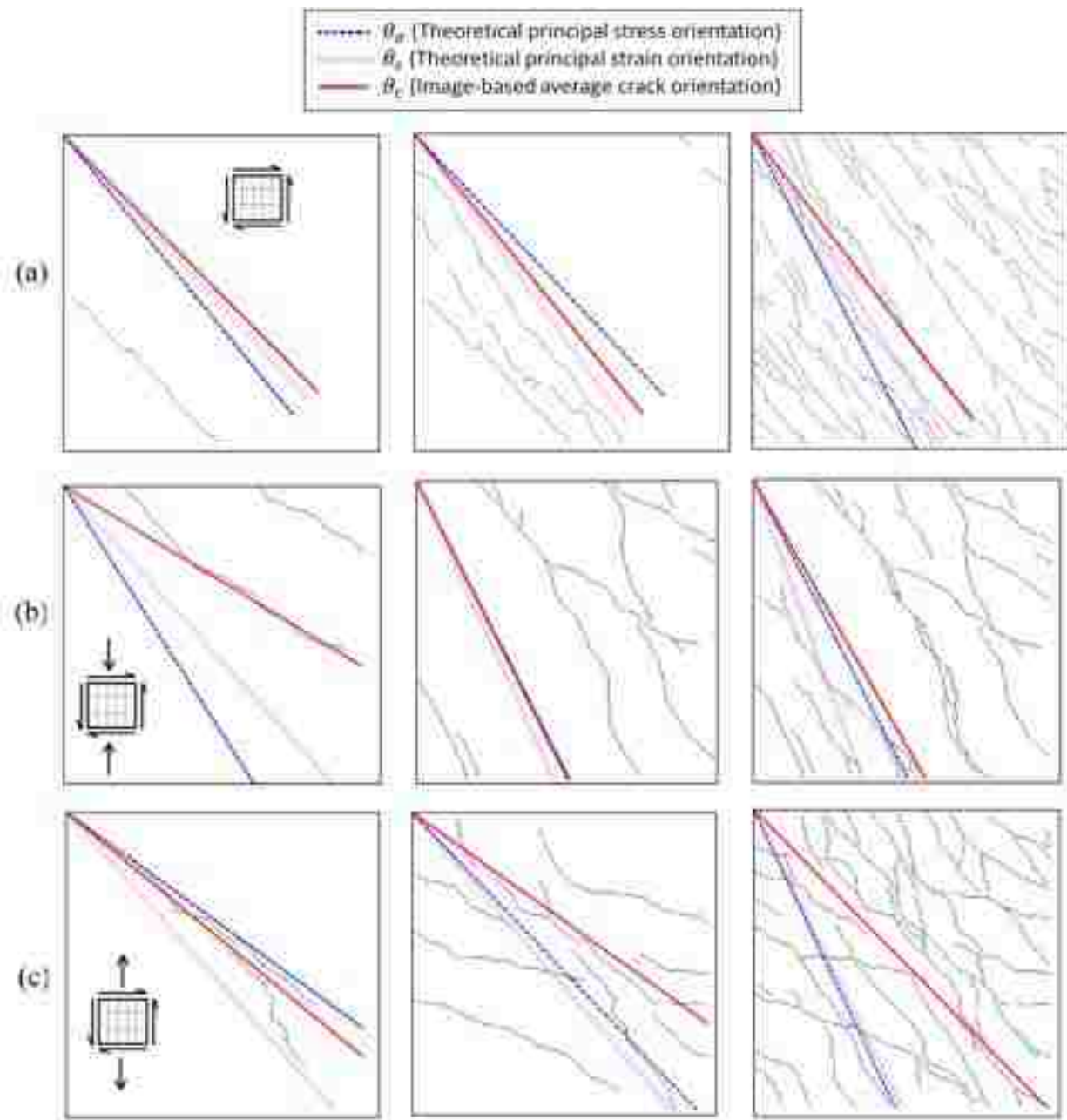


Figure 5.6: Comparison between  $\theta_p$  calculated based on the nominal state principal orientation, and overall orientation of cracks calculated using image processing for representative conventionally reinforced concrete load stage images. Loading increases left to right for each row of images, which results in increased induced critical stress, and thus rotating principal orientations.

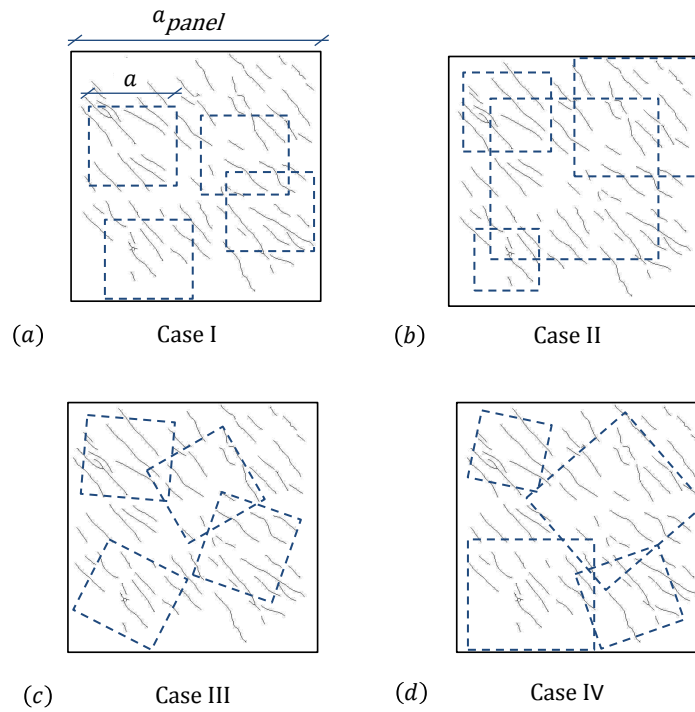


Figure 5.7: Resampling approaches used to generate more image blocks

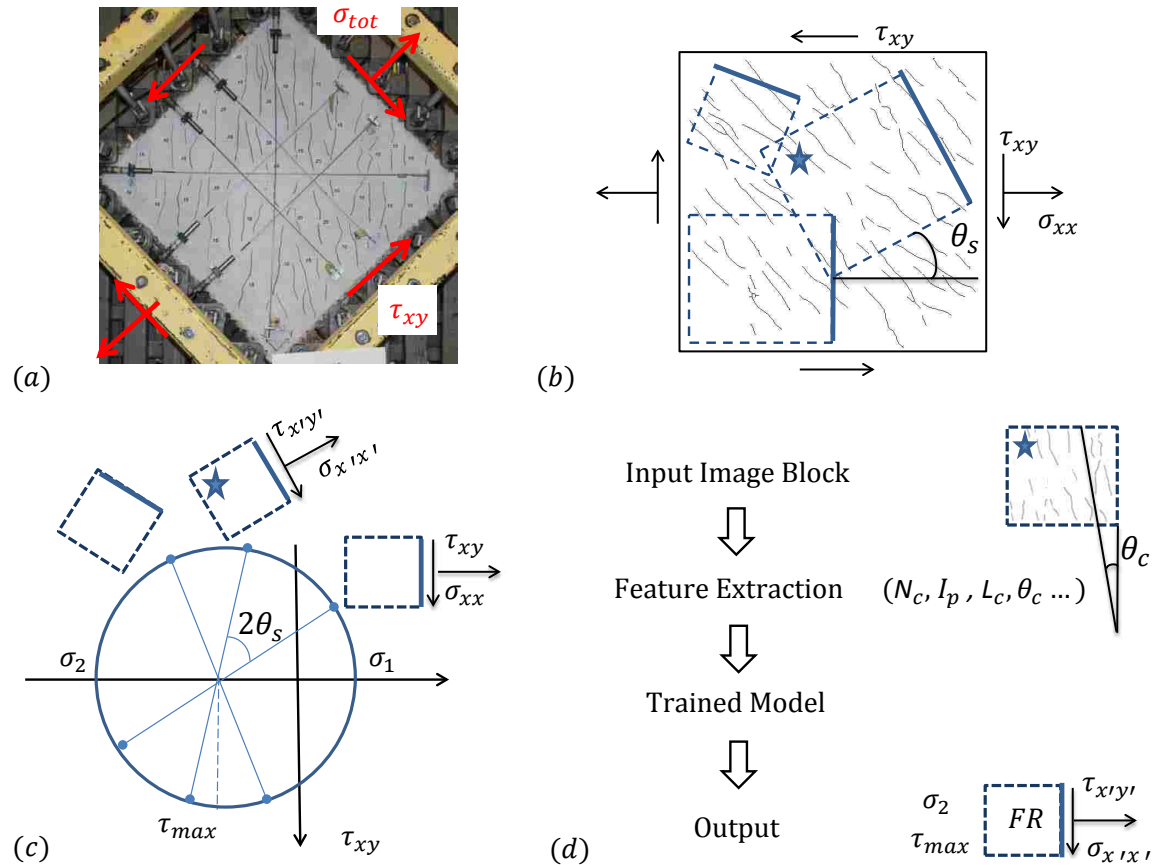


Figure 5.8: (a) Typical concrete panel subjected to monotonic loading (obtained from Luo 2014).  $\sigma_{tot}$  and  $\tau_{xy}$  are the total normal and shear stress acting at the side of the panel. (b) Segmented crack patterns.  $\sigma_{xx}$  and  $\tau_{xy}$  are the normal and shear stress in the concrete.  $\sigma_{xx}$  is calculated based on the compatibility condition between steel and concrete. The location and angle of the image blocks were randomly selected with orientation ( $0^\circ \leq \theta_s \leq 180^\circ$ ) to generate new subsampled image blocks; (c) Mohr's circle representing the stress state of the panel and randomly selected image blocks. The randomly selected image blocks have the same maximum shear stress ( $\tau_{max}$ ) and principal stresses ( $\sigma_1$  and  $\sigma_2$ ) (d) Pipeline used to estimate the failure ratio and stress state for a typical resampled image block, marked with a star.

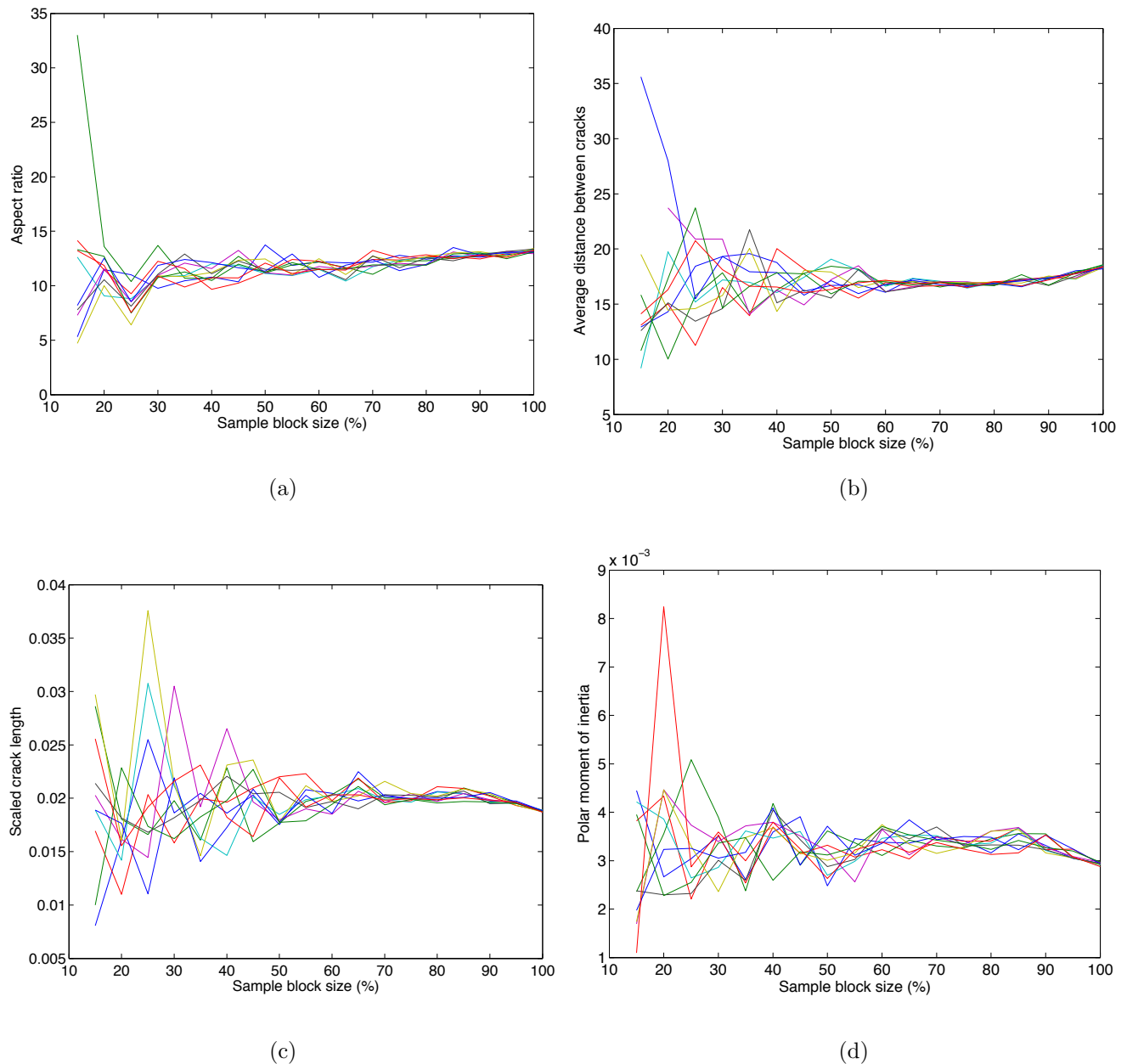
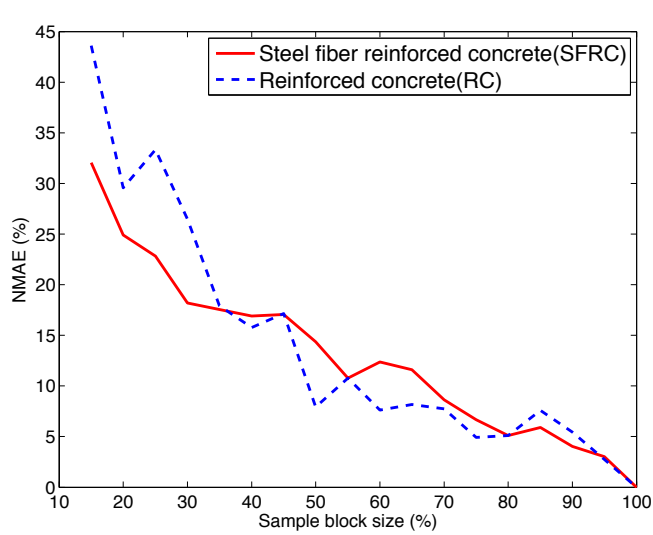
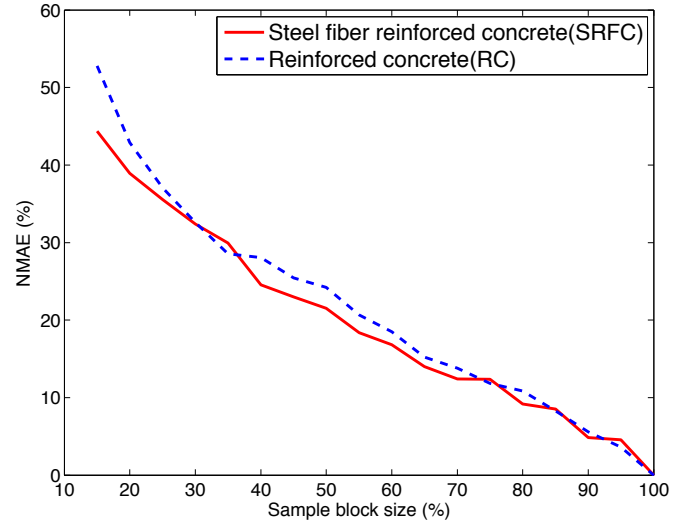


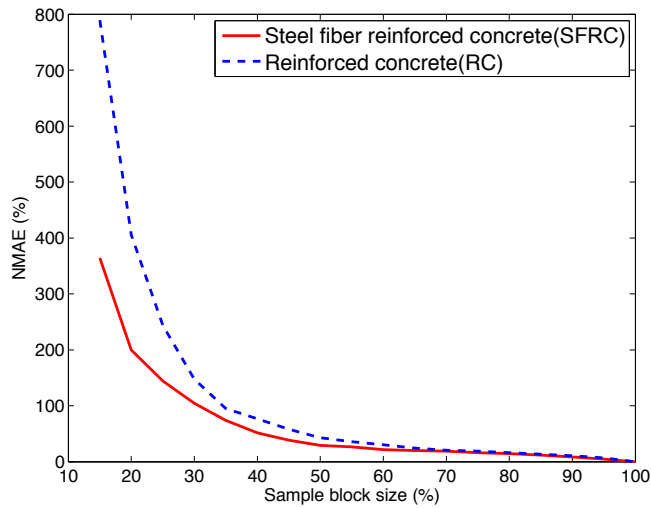
Figure 5.9: Effects of the sub-image block size on different cracks attributes. Sampling technique III was used to sample ten sub-image blocks having different orientations from a typical fiber crack pattern image



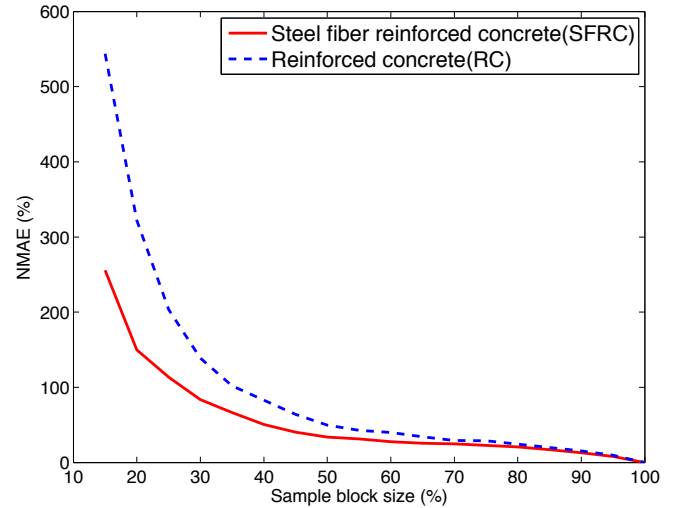
(a) Aspect Ratio



(b) Average distance between cracks

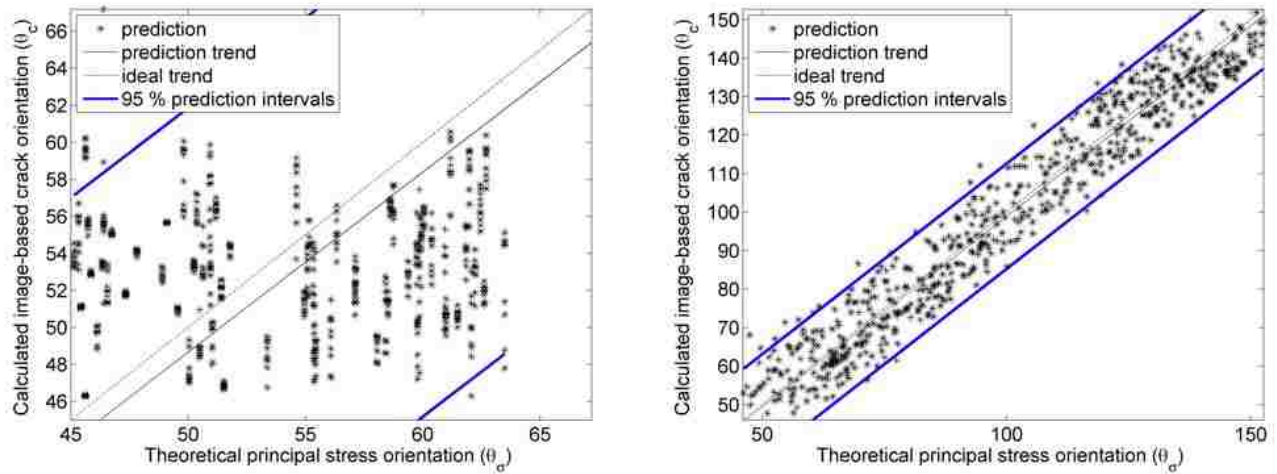


(c) Scaled crack length



(d) Scaled polar moment of inertia

Figure 5.10: Effects of the image block size on Normalized Mean Absolute Error (NMAE) for different cracks attributes. Sampling technique III was used to sample ten sub-image blocks having different orientations from all fiber specimens



(a) Case 2:  $a_s$  : 30%–100%; fixed orientation;  $10\times$  resampling  
 (b) Case 4:  $a_s$  : 30%–100%;  $\theta_s$ :  $0^\circ$ – $90^\circ$ ;  $10\times$  resampling

Figure 5.11: Image-based average fracture plane orientation,  $\theta_c$ , correlation with theoretical principal orientation,  $\theta_p$ , based on nominal applied stress state.

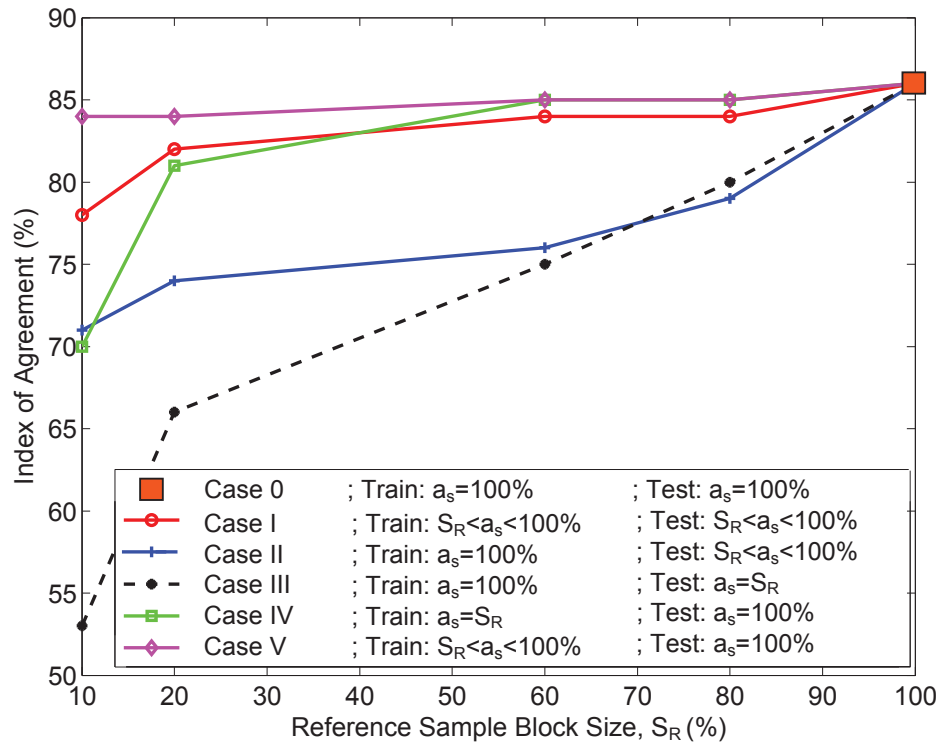
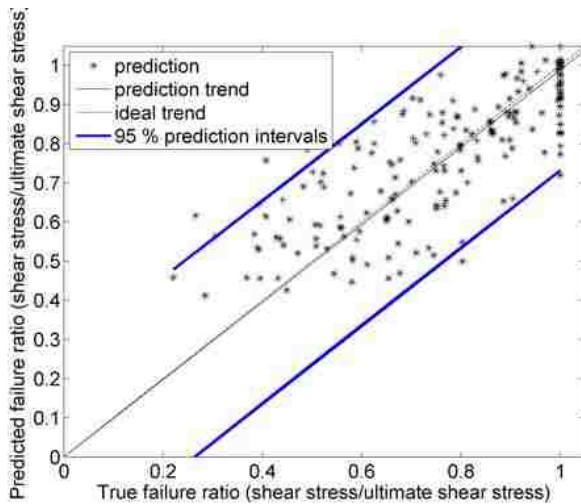
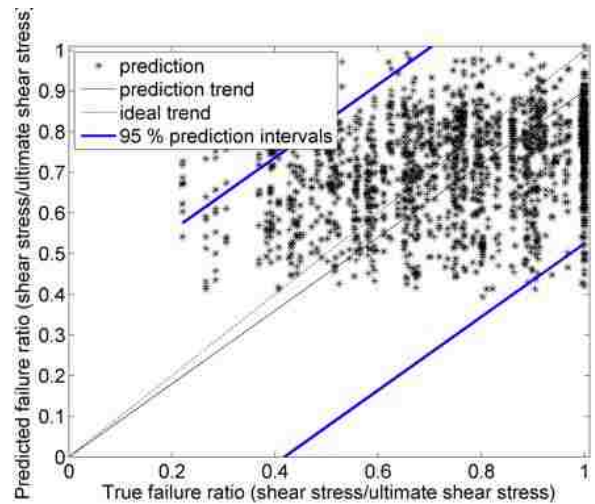


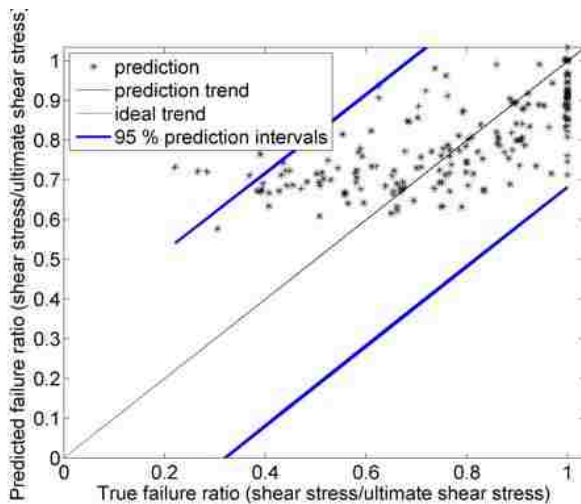
Figure 5.12: Index of Agreement convergence performance for fracture ratio estimation as a function of reference sampling block size for SVM models using dominant features. Plotted curves correspond to different training and test sampling strategies as indicated, with size ( $a_s$ ) and/or orientation ( $0^\circ \leq \theta_s \leq 90^\circ$ ).  $10\times$  resampling used per load stage for all cases other than the full ( $a_s = 100\%$ ) panel cases.



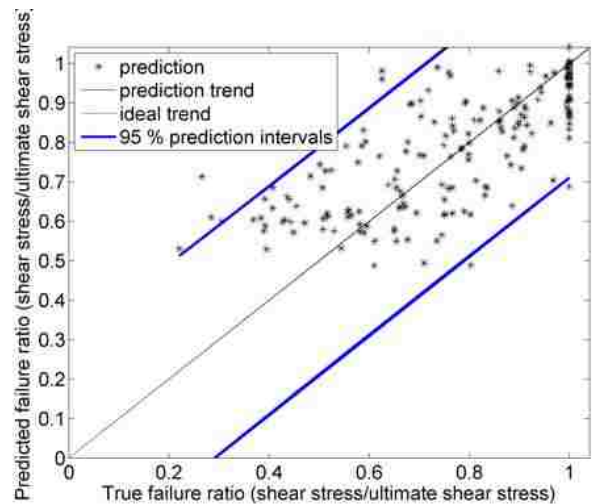
(a) Train  $a_s$  : 100%; Test  $a_s$  : 100% (Case 0)



(b) Train  $a_s$  : 100%; Test  $a_s$  : 10% 10× resampling (Case III)



(c) Train  $a_s$  : 10% 10× resampling; Test  $a_s$  : 100% (Case IV)

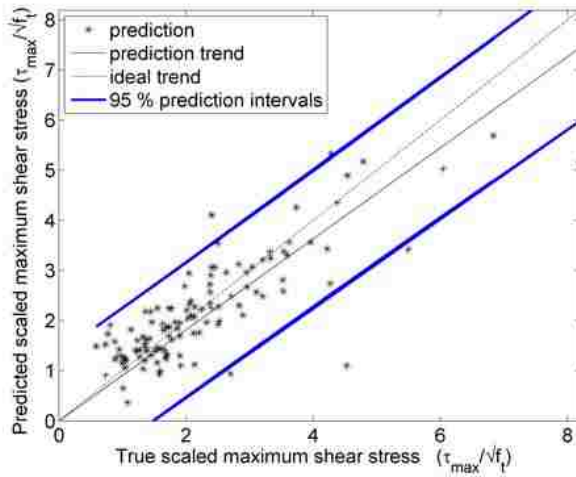


(d) Train  $a_s$  : 20% 10× resampling; Test  $a_s$  : 100% (Case IV)

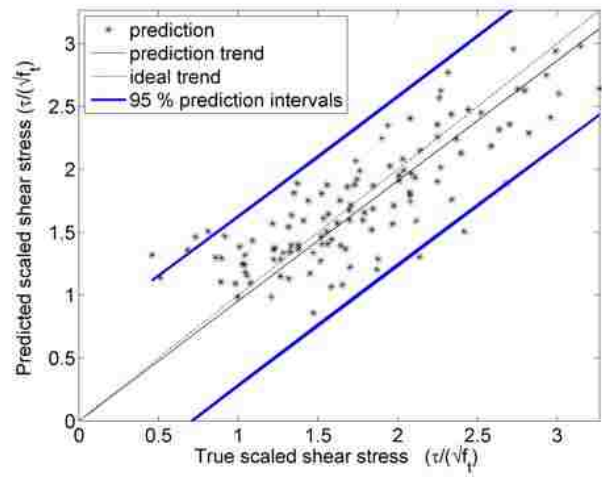
Figure 5.13: Performance of SVM models to estimate fracture ratio trained by dominant features on different training and test data sets

Figure 5.14: Comparison of SVM models trained by dominant feature set to estimate fracture ratio on entire and resampled data sets

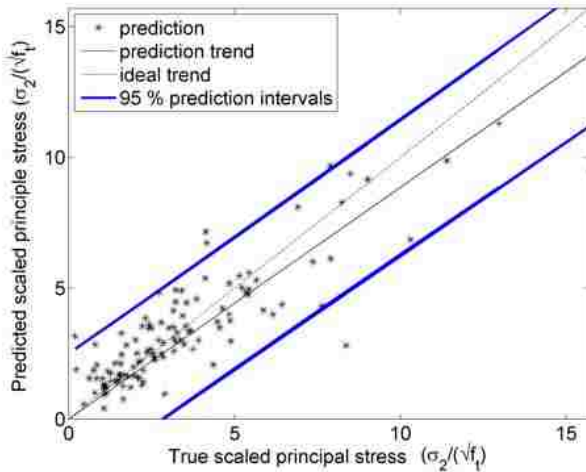




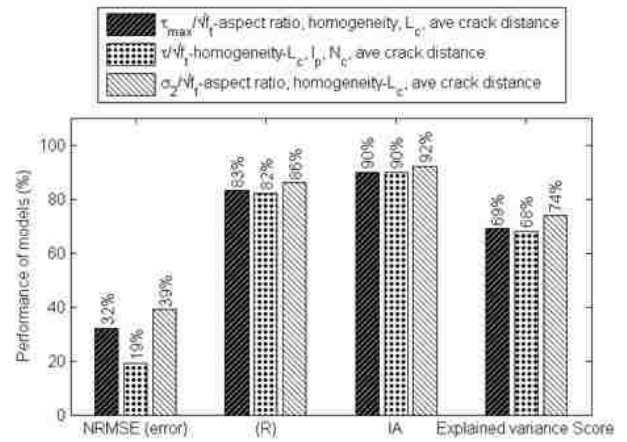
(a) Maximum scaled in plane shear stress



(b) Scaled shear stress



(c) Minor principal stress



(d) Error measures

Figure 5.15: Performance of Gaussian process models to estimate  $\tau_{max}/\sqrt{f_t}$ ,  $\tau/\sqrt{f_t}$  and  $\sigma_2/\sqrt{f_t}$  for monotonic data set using best-fit feature sets. All cases uses  $L_c$ , and ave crack distance and homogeneity. In addition, cases (a) and (c) include aspect ratio, while case (b) includes  $I_p$  and  $N_c$

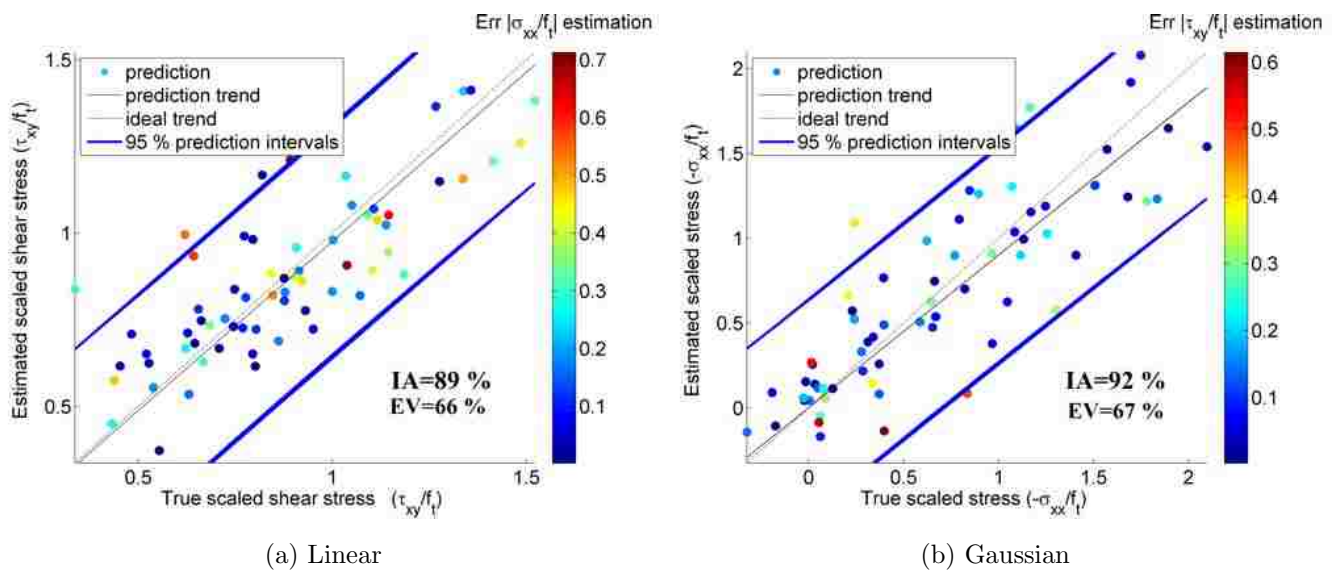
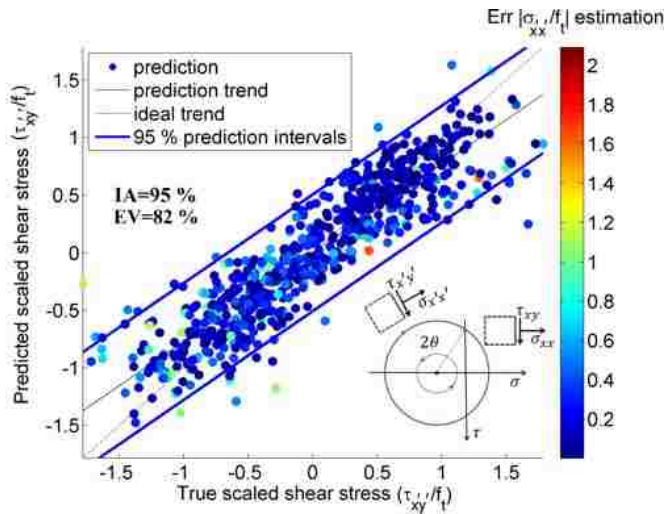
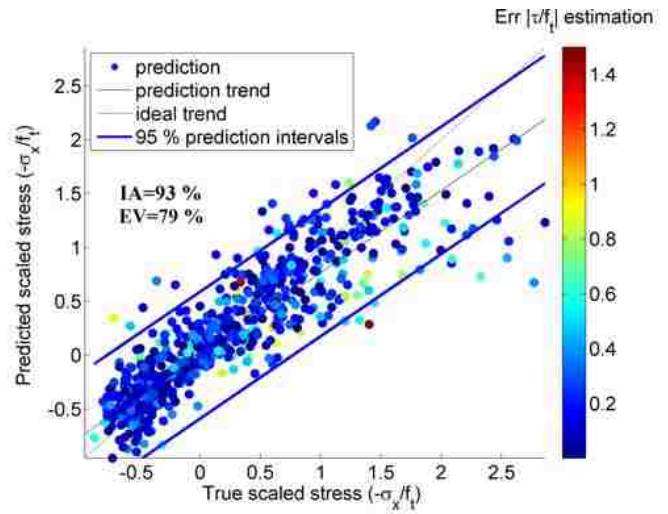


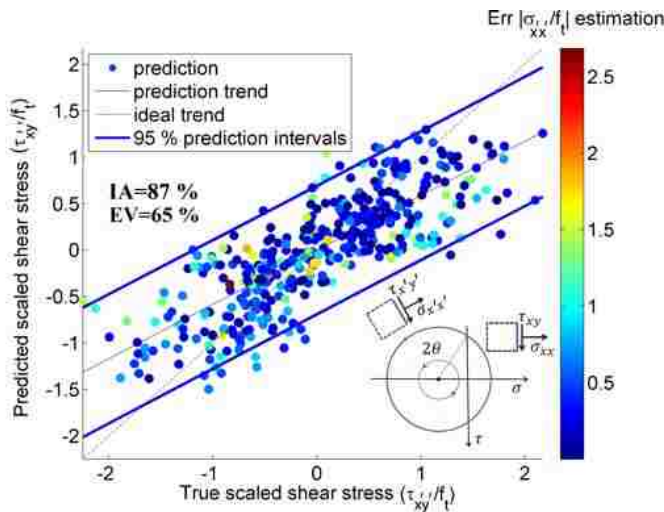
Figure 5.16: Performance of Gaussian process and linear models to estimate  $\sigma_{xx}/f_t$  and  $\tau_{xy}/f_t$  trained with dominant feature set (homogeneity,  $L_c$ ,  $I_p$ ,  $N_c$ , total major axis, average distance between cracks) on monotonic SFRC data set



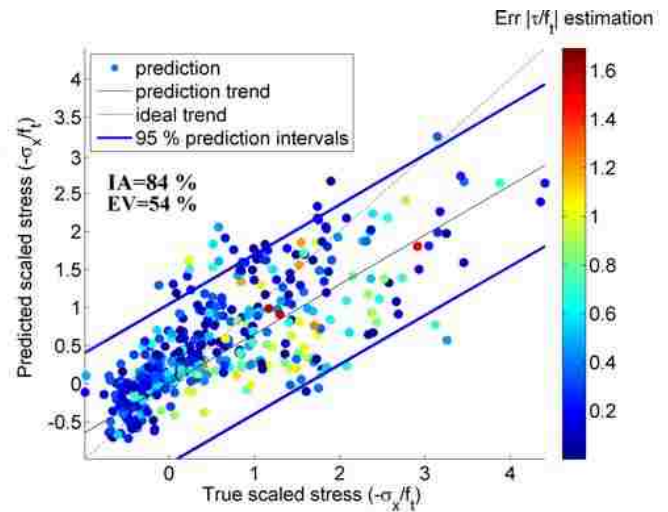
(a) Scaled shear stress on rotated block (SFRC)



(b) Scaled normal stress on rotated block (SFRC)



(c) Scaled shear stress on rotated block (RC)



(d) Scaled normal stress on rotated block (RC)

Figure 5.17: Performance of SVM models to estimate  $\sigma_{x'x'}/f_t$  and  $\tau_{x'y'}/f_t$  trained with all crack features including overall crack orientation,  $\theta_c$ , and sampling block size,  $a_s$ , on sub-sampled RC and SFRC data sets. Case I sampling strategy, with size of ( $30\% \leq a_s \leq 100\%$ ) and orientation range of ( $0^\circ \leq \theta_s \leq 180^\circ$ ) was used.

Chapter 6

**CONCLUSION AND SUGGESTED FUTURE WORK**

## 6.1 Conclusion and contributions

A computer vision based inspection framework for concrete members has been developed and tested using image data assembled from previously published experimental studies with specimens covering a wide range of geometries, loadings, concrete strengths, and reinforcement details. The images were captured at various load and damage levels, and so the resulting framework is capable of translating simple photographic data into estimated levels of damage calibrated relative to the component's actual capacity. Unlike mechanics-based approaches, which can be generalized using first principles, data-driven methods require additional data to reliably extend their applicability beyond the calibration data range, and so additional test data can significantly leverage the prediction capability of computer vision based models. As shown in previous chapters, incorporating new data such as additional shear-reinforced beams and slabs could extend the capability of predictive models both in classification and regression tasks. This enabled the study of the effectiveness of the computer vision approach to estimate load and damage levels across and within several key categories of reinforced concrete members subjected to monotonic loading, and to investigate the computer vision framework's ability to classify the damage type (i.e., identifying the presence or absence of shear reinforcement based on externally observed damage). The present study provides several contributions:

- Assembly of an image/load-level database capable of enabling the development of data-driven estimation models.
- Defining a simple and uniform procedure to build new image data sets from available data sources such as existing video clips or crack pattern images for future studies.
- Defining a simple, yet effective image transformation process to take account of the different specimen and image dimensions.
- Use of simplified crack representation (i.e., only crack patterns without any information about crack width) for damage and load level estimation
- Identification of effective and dominant geometric and textural crack features for different prediction tasks
- Suggesting different simple and advanced machine learning techniques for prediction tasks
- Accommodation for size effect phenomenon in concrete structures
- Defining a new modified cross validation approach for evaluating the performance of the models to resemble potential field scenarios.

As discussed in previous chapters, knowledge of structural properties such as reinforcement ratio and concrete strength beyond basic external geometry was assumed to not be available for damage level estimation. However, adding such information can improve the performance of the models. Results indicate that quantifying internal load and damage levels with practically useful confidence levels and estimation errors is feasible. It therefore could be possible to apply these procedures in future field inspection scenarios. Potential application opportunities can be identified as follows:

- low-cost and rapid preliminary inspection becomes feasible for anyone with an interest in assessing the structural health of a visibly damaged component.
- structural health monitoring companies, building and infrastructure officials, and first responders can be more effective in inspecting buildings and infrastructure components and systems via results obtained using easily captured images.
- non-specialists could be mobilized in disaster scenarios such as major earthquakes and storms to gather rapid preliminary inspection data useful for guiding necessary actions to help reduce widespread human and economic losses.
- autonomous continuous camera monitoring of critical structural systems could also be used for quantitative damage detection over the life of a structure.

## 6.2 *Limitations*

Because the technologies and algorithms associated with computer vision are continually improving, the results and conclusions presented in this research can best be thought of as setting a baseline for what is currently possible. The levels of accuracy in estimation achieved with the framework used in this work are promising with respect to potential field applications, but there is also a significant gap between this feasibility study based on lab studies and practical deployment.

The limitations of the results presented here include the limited size of the data sets relative to those common in machine learning contexts, the relatively simple crack representations available in the reported prior work, and the specific focus on simply-supported beams and slabs undergoing monotonic loading. It would not make sense to apply the specific models trained here outside the controlled context in which they were developed. However, none of these limitations arise from inherent problems within the overall framework, and in principle each could be addressed with sufficient further testing and/or data gathering.

### **6.3 Future work in building a data repository**

As mentioned earlier, there is no serious effort focused on quantified damage state estimation using machine learning and big data approaches. This project suggests creating a data repository that enables further developments in machine vision based inspection techniques. We already found thousands of image data for beams, slabs, columns, shear walls and concrete panels. But, a major challenge would be collecting more data and encouraging different researchers around the world to contribute to this project by adding new data. Researchers in civil engineering field who have conducted many experimental lab tests don't seem to be very interested in using such innovative data analytic and big data approaches for inspecting the structures. The main step would be to design an online website, and develop a proof of concept framework that encourages other researchers to join this research effort. Another challenge would be to design of a simple and comprehensive procedure to clarify what types of data and images can be added in the data repository.

### **6.4 Future work in damage quantification**

Most of the latest research efforts in the area of computer vision damage assessment are focused on categorizing the damage type using machine learning or deep learning techniques [2] [29]. A key aspect of these approaches is the removal of the need for feature engineering. There are opportunities to build further predictive models similar to image-based models presented in this dissertation building on this related contemporary work. Higher resolution image histories could make it possible to explore further possible correlations between observable defects/damages with quantities like deformation, load, and damage states.

Conducting new experimental tests should be considered in future works to expand the raw image data in order to investigate the possibility of improving image-based models. Most of the available image data are not available in raw captured image form, and for much of the data used here cracks were marked manually during loading tests. The resulting crack line drawings in the image data did not include details like crack widths and spalling, which are known to be important in damage estimation. Expanded access to raw image data collected during testing would also make it possible to increase the generality of the standardized image content, which could improve overall estimation accuracy.

### **6.5 Future work in combining mechanics-based with data-driven approaches**

The performance of machine learning models are significantly dependent on the size, range, and quality of the data. Incorporating mechanics-based models with pure data-driven models could be one of the ways to address this problem. Mechanics-based models can be used to generate more data, and also to complement

the results of data-driven approaches. This combination of traditional mechanic-based models with machine learning models can potentially improve the usefulness of existing computer vision frameworks.



## BIBLIOGRAPHY

- [1] LRFD AASHTO. Aashto lrfd bridge design specifications. *Transportation (Amst). American Association of State Highway and Transportation Officials, Inc.: Washington, DC*, 2007.
- [2] Hars D.K. Alipour, M. and G.R. Miller. Robust pixel-level crack detection using deep fully convolutional neural networks. *Journal of Computing in Civil Engineering*, 2019.
- [3] Mohamad Alipour and Devin K. Harris. Adapting deep learning models for robust infrastructure crack detection in different materials. *Journal of Computing in Civil Engineering*, 2018. in review.
- [4] Mohamad Alipour, Devin K. Harris, and Gregory R. Miller. Robust pixel-level crack detection using deep fully convolutional neural networks. *Journal of Computing in Civil Engineering*, 2018. in review.
- [5] Dino Angelakos. The influence of concrete strength and longitudinal reinforcement ratio on the shear strength of large-size reinforced concrete beams with, and without transverse reinforcement. Master's thesis, University of Toronto, Canada, 1999.
- [6] Celalettin Başıyigit, Bekir Çomak, Şemsettin Kılınçarslan, and İsmail Serkan Üncü. Assessment of concrete compressive strength by image processing technique. *Construction and Building Materials*, 37:526 – 532, 2012. Non Destructive Techniques for Assessment of Concrete.
- [7] Zdeněk P Bažant. Concrete fracture models: testing and practice. *Engineering fracture mechanics*, 69(2):165–205, 2002.
- [8] Evan C Bentz and Michael P Collins. Development of the 2004 canadian standards association (csa) a23. 3 shear provisions for reinforced concrete. *Canadian Journal of Civil Engineering*, 33(5):521–534, 2006.
- [9] S. R. Birgisson. Shear resistance of reinforced concrete beams without stirrups. Master's thesis, Reykjavik University, Iceland, 2011.
- [10] A. C. Bohigas. *Shear design of reinforced high-strength concrete beams*. PhD thesis, UNIVERSITAT POLITCNICA DE CATALUNYA, Barcelona, Spain, 2002.
- [11] Leo Breiman. Random forests. *Machine learning*, 45(1):5–32, 2001.
- [12] Steven L Brunton and J Nathan Kutz. *Data-driven Science and Engineering: Machine Learning, Dynamical Systems, and Control*. Cambridge University Press, 2019.
- [13] Paolo M Calvi, Evan C Bentz, and Michael P Collins. Model for assessment of cracked reinforced concrete membrane elements subjected to shear and axial loads. *ACI Structural Journal*, 115(2):501–509, 2018.
- [14] Paolo M Calvi, Giorgio T Proestos, and David M Ruggiero. Towards the development of direct crack-based assessment of structures. *ACI, Special Publication*, 328:9–1, 2018.
- [15] J. Canny. A computational approach to edge detection. *Institute of Electrical and Electronics Engineers (IEEE) Transactions on Pattern Analysis and Machine Intelligence*, PAMI-8(6):679–698, Nov 1986.
- [16] S. Cao. Size effect and the influence of longitudinal reinforcement on the shear response of large reinforced concrete members. Master's thesis, University of Toronto, Canada, 2001.

- [17] Shen Cao. Size effect and the influence of longitudinal reinforcement on the shear response of large reinforced concrete members. Master's thesis, University of Toronto, Canada, 2001.
- [18] David Joseph Carnovale. *Behaviour and analysis of steel and macro-synthetic fibre reinforced concrete subjected to reversed cyclic loading: a pilot investigation*. PhD thesis, University of Toronto, 2013.
- [19] Young-Jin Cha, Wooram Choi, and Oral Büyüköztürk. Deep learning-based crack damage detection using convolutional neural networks. *Computer-Aided Civil and Infrastructure Engineering*, 32(5):361–378, 2017.
- [20] F. Chen and M. R. Jahanshahi. Nb-cnn: Deep learning-based crack detection using convolutional neural network and naive bayes data fusion. *IEEE Transactions on Industrial Electronics*, 65(5):4392–4400, May 2018.
- [21] A. Cladera and A.R. Mar. Shear design procedure for reinforced normal and high-strength concrete beams using artificial neural networks. part i: beams without stirrups. *Engineering Structures*, 26(7):917 – 926, 2004.
- [22] ACI Committee, American Concrete Institute, and International Organization for Standardization. Building code requirements for structural concrete (aci 318-08) and commentary. American Concrete Institute, 2008.
- [23] Fei Dai, Ming Lu, and Vineet R. Kamat. Analytical approach to augmenting site photos with 3d graphics of underground infrastructure in construction engineering applications. *Journal of Computing in Civil Engineering*, 25(1), 2010.
- [24] Rouzbeh Davoudi, Gregory R. Miller, and Paolo M Calvi. Data and code for image-based damage and stress estimation in rc and sfrc panels. 2018.
- [25] Rouzbeh Davoudi, Gregory R. Miller, and J. Nathan Kutz. Data-driven vision-based inspection for reinforced concrete beams and slabs: Quantitative damage and load estimation. *Automation in Construction*, 96:292 – 309, 2018.
- [26] Rouzbeh Davoudi, Gregory R. Miller, and J. Nathan Kutz. Structural load estimation using machine vision and surface crack patterns for shear-critical rc beams and slabs. *Journal of Computing in Civil Engineering*, 32(4):04018024, 2018.
- [27] J. F Destrebecq, E. Toussaint, and E. Ferrier. Analysis of cracks and deformations in a full scale reinforced concrete beam using a digital image correlation technique. *Experimental Mechanics*, 51(6):879–890, 2011.
- [28] Arvin Ebrahimkhanlou, , and Salvatore Salamone. A probabilistic model for visual inspection of concrete shear walls. In *SPIE 10168, Sensors and Smart Structures Technologies for Civil, Mechanical, and Aerospace Systems*, ed. Jerome P. Lynch, 2017.
- [29] Arvin Ebrahimkhanlou, Brennan Dubuc, and Salvatore Salamone. A generalizable deep learning framework for localizing and characterizing acoustic emission sources in riveted metallic panels. *Mechanical Systems and Signal Processing*, 130:248 – 272, 2019.
- [30] Arvin Ebrahimkhanlou, Alireza Farhidzadeh, and Salvatore Salamone. Multifractal analysis of crack patterns in reinforced concrete shear walls. *Structural Health Monitoring*, 15(1):81–92, 2016.
- [31] Chen Feng, Ming-Yu Liu, Chieh-Chi Kao, and Teng-Yok Lee. *Deep Active Learning for Civil Infrastructure Defect Detection and Classification*. 2017.
- [32] New Model Code fib special activity group, Luc Taerwe, and Stijn Matthys. *fib model code for concrete structures 2010*. Ernst & Sohn, Wiley, 2013.

- [33] Yuqing Gao and Khalid M. Mosalam. Deep transfer learning for image-based structural damage recognition. *Computer-Aided Civil and Infrastructure Engineering*, 33(9):748–768, 2018.
- [34] Y. H. Gedik. *Experimental and numerical study on shear failure mechanism of RC deep beams*. PhD thesis, Nagoya University, Nagoya, JAPAN, 2011.
- [35] Stephanie German, Ioannis Brilakis, and Reginald DesRoches. Comprehensive property retrieval and measurement of concrete spalling using machine vision for post-earthquake safety assessments. In *proceedings of SPIE Vol*, volume 8345, pages 83454S–1, 2012.
- [36] Zoubin Ghahramani. A tutorial on gaussian processes (or why i don't use svms). 2011.
- [37] Steve R Gunn. Support vector machines for classification and regression. Technical report, interactive software invocation system, 1998.
- [38] Mark Hall, Eibe Frank, Geoffrey Holmes, Bernhard Pfahringer, Peter Reutemann, and Ian H. Witten. The weka data mining software: An update. *SIGKDD Explorations Newsletter*, 11(1):10–18, November 2009.
- [39] Mark Hall, Eibe Frank, Geoffrey Holmes, Bernhard Pfahringer, Peter Reutemann, and Ian H. Witten. The weka data mining software: An update. *SIGKDD Explor. Newsl.*, 11(1):10–18, November 2009.
- [40] A. Ito, Y. Aoki, and S. Hashimoto. Accurate extraction and measurement of fine cracks from concrete block surface image. In *Institute of Electrical and Electronics Engineers (IEEE) 2002 28th Annual Conference of the Industrial Electronics Society. IECON 02*, volume 3, pages 2202–2207 vol.3, Nov 2002.
- [41] M. R. Jahanshahi, J. S. Kelly, S. F. Masri, and G. S. Sukhatme. A survey and evaluation of promising approaches for automatic image-based defect detection of bridge structures. *Structure and Infrastructure Engineering*, 5, 2009.
- [42] Y. Jang, Y.H. Ahn, and Ha Young Kim. Estimating the compressive strength of concrete using deep convolutional neural networks with digital microscope images. *Journal of Computing in Civil Engineering*, 2018. in press.
- [43] Vineet R. Kamat and Sherif El-Tawil. Evaluation of augmented reality for rapid assessment of earthquake-induced building damage. *Journal of Computing in Civil Engineering*, 21(5):303–310, 2007.
- [44] Christian Koch, Kristina Georgieva, Varun Kasireddy, Burcu Akinci, and Paul Fieguth. A review on computer vision based defect detection and condition assessment of concrete and asphalt civil infrastructure. *Advanced Engineering Informatics*, 29(2):196 – 210, 2015. Infrastructure Computer Vision.
- [45] Ron Kohavi. A study of cross-validation and bootstrap for accuracy estimation and model selection. In *Proceedings of the 14th International Joint Conference on Artificial Intelligence - Volume 2, IJCAI'95*, pages 1137–1143, San Francisco, CA, USA, 1995. Morgan Kaufmann Publishers Inc.
- [46] J Nathan Kutz. *Data-driven modeling & scientific computation: methods for complex systems & big data*. Oxford University Press, 2013.
- [47] David Lattanzi, Gregory R. Miller, Marc O. Eberhard, and Olafur S. Haraldsson. Bridge column maximum drift estimation via computer vision. *Journal of Computing in Civil Engineering*, 30(4), 2015.
- [48] S. le Cessie and J.C. van Houwelingen. Ridge estimators in logistic regression. *Applied Statistics*, 41(1):191–201, 1992.
- [49] Sunghoon Lee. Using weka in matlab, 2015.

- [50] David R. Legates and Gregory J. McCabe. Evaluating the use of goodness-of-fit measures in hydrologic and hydroclimatic model validation. *Water Resources Research*, 35(1):233–241, 1999.
- [51] Cheng Lu, Muhammad Mahmood, Naresh Jha, and Mrinal Mandal. A robust automatic nuclei segmentation technique for quantitative histopathological image analysis. *Analytical and Quantitative Cytology and Histology*, 34:296–308, 2012.
- [52] Adam Scott Lubell. *Shear in wide reinforced concrete members*. PhD thesis, University of Toronto, Canada, 2006.
- [53] Wei Jun Wei Luo. *Behaviour and Analysis of Steel Fibre-reinforced Concrete under Reversed Cyclic Loading*. PhD thesis, University of Toronto, 2014.
- [54] M.Y. Mansour, M. Dicleli, J.Y. Lee, and J. Zhang. Predicting the shear strength of reinforced concrete beams using artificial neural networks. *Engineering Structures*, 26(6):781 – 799, 2004.
- [55] Joshua S. Martin. An experimental investigation of bond in reinforced concrete. Master’s thesis, University of Washington, 2006.
- [56] The Mathworks, Inc., Natick, Massachusetts. *MATLAB version 8.5.0.197613 (R2013b)*, 2013.
- [57] K. A. McCain. The effect of scale on the resistance of reinforced concrete beams to shear. Master’s thesis, Purdue University, West Lafayette, Indiana, 2012.
- [58] Matthew Murray and Santiago Pujol. An investigation of the unit shear strength of geometrically scaled reinforced concrete beams. Master’s thesis, Purdue University, West Lafayette, Indiana, 2017.
- [59] Vu Hong Nghiep. Shear design of straight and haunched concrete beams without stirrups. Master’s thesis, Schriftenreihe des Instituts fr Massivbau der TUHH, 2011.
- [60] Stephanie G. Paal, Jong-Su Jeon, Ioannis Brilakis, and Reginald DesRoches. Automated damage index estimation of reinforced concrete columns for post-earthquake evaluations. *Journal of Structural Engineering*, 141(9):04014228, 2015.
- [61] Jason Peddle, Adam Goudreau, Emily Carlson, and Erin Santini-Bell. Bridge displacement measurement through digital image correlation. *Bridge Structures*, 7(4):165–173, 2011.
- [62] S. M. J. Perkins. Shear behaviour of deep reinforced concrete members subjected to uniform load. Master’s thesis, University of Toronto, Canada, 2011.
- [63] Raymond Phan. A matlab implementation of the tensorflow neural network playground, 2015.
- [64] Bogdan A. Podgorniak-Stanik. The influence of concrete strength, distribution of longitudinal reinforcement, amount of transverse reinforcement and member size on shear strength of reinforced concrete members. Master’s thesis, University of Toronto, Canada, 1998.
- [65] Foster Provost and Tom Fawcett. Analysis and visualization of classifier performance: Comparison under imprecise class and cost distributions. In *Proceedings of the Third International Conference on Knowledge Discovery and Data Mining*, KDD’97, pages 43–48. AAAI Press, 1997.
- [66] P. T. Quach. Understanding and safely predicting the shear response of large-scale reinforced concrete structures. Master’s thesis, University of Toronto, Canada, 2016.
- [67] J. R. Quinlan. Induction of decision trees. *Mach. Learn.*, 1(1):81–106, March 1986.
- [68] Carl Edward Rasmussen and Christopher K. I. Williams. *Gaussian Processes for Machine Learning (Adaptive Computation and Machine Learning)*. The MIT Press, 2005.
- [69] David Michael Volpe Ruggiero. *The Behaviour of Reinforced Concrete Subjected to Reversed Cyclic Shear*. PhD thesis, University of Toronto, 2015.

- [70] E.G. Sherwood. *One-way Shear Behaviour of Large, Lightly-reinforced Concrete Beams and Slabs*. PhD thesis, 2008.
- [71] L. H. Sneed. Influence of member depth on the shear strength of concrete beams. Master's thesis, Purdue University, West Lafayette, Indiana, 2007.
- [72] Tyler S. Sprague. X-ray tomography for evaluation of damage in concrete bond. Master's thesis, University of Washington, 2006.
- [73] Jimmy Susetyo. *Fibre reinforcement for shrinkage crack control in prestressed, precast segmental bridges*. PhD thesis, University of Toronto, 2009.
- [74] Bülent Üstün, Willem J Melssen, and Lutgarde MC Buydens. Facilitating the application of support vector regression by using a universal pearson vii function based kernel. *Chemometrics and Intelligent Laboratory Systems*, 81(1):29–40, 2006.
- [75] FJ Vecchio. Disturbed stress field model for reinforced concrete: formulation. *Journal of structural engineering*, 126(9):1070–1077, 2000.
- [76] FJ Vecchio. Disturbed stress field model for reinforced concrete: implementation. *Journal of Structural Engineering*, 127(1):12–20, 2001.
- [77] Frank J Vecchio and Michael P Collins. The modified compression-field theory for reinforced concrete elements subjected to shear. *ACI J.*, 83(2):219–231, 1986.
- [78] Frank Joseph Vecchio. *Shear Rig Design*. PhD thesis, University of Toronto, 1979.
- [79] Cort J. Willmott, Steven G. Ackleson, Robert E. Davis, Johannes J. Feddema, Katherine M. Klink, David R. Legates, James O'Donnell, and Clinton M. Rowe. Statistics for the evaluation and comparison of models. *Journal of Geophysical Research: Oceans*, 90(C5):8995–9005, 1985.
- [80] Liping Xie. *The influence of axial load and prestress on the shear strength of web-shear critical reinforced concrete elements*. PhD thesis, University of Toronto, 2009.
- [81] T. Yamaguchi and S. Hashimoto. Fast crack detection method for large-size concrete surface images using percolation-based image processing. *Machine Vision and Applications*, 21(5):797–809, 2010.
- [82] Y. Yoshida. Shear reinforcement for large lightly reinforced concrete members. Master's thesis, University of Toronto, Canada, 2000.
- [83] Yoichi Yoshida. Shear reinforcement for large lightly reinforced concrete members. Master's thesis, University of Toronto, Canada, 2000.
- [84] Yasha Zeinali, Yamin Li, Dinesh Rajan, and Brett Story. Accurate structural dynamic response monitoring of multiple structures using one ccd camera and a novel targets configuration. In *11th International Workshop of Structural Health Monitoring (IWSHM) 2017, Stanford University*, pages 3107–3114, 2017.
- [85] Liangliang Zhang, Xinya Yue, Lin Wang, and Bo Yang. Estimating cement compressive strength from microstructural images using gep with probabilistic polarized similarity weight tournament selection. In *Proceedings of the Genetic and Evolutionary Computation Conference, GECCO '18*, pages 1403–1410, New York, NY, USA, 2018. ACM.
- [86] Zhenhua Zhu, Ioannis Brilakis, and Gustavo Parra-Montesinos. Real time concrete damage visual assessment for first responders. In *American Society of Civil Engineers (ASCE) Construction Research Congress*, pages 1204–1213. ASCE Reston, Va., 2009.
- [87] Zhenhua Zhu, Stephanie German, and Ioannis Brilakis. Visual retrieval of concrete crack properties for automated post-earthquake structural safety evaluation. *Automation in Construction*, 20(7):874–883, 2011.