

FUSION-FEATURES AND VISUAL-DICTIONARY IMAGE
RECOGNITION METHODS FOR APPLE CLASSIFICATION
IN SMART MANUFACTURING

AHSIAH BINTI ISMAIL

FACULTY OF COMPUTER SCIENCE AND
INFORMATION TECHNOLOGY
UNIVERSITY OF MALAYA
KUALA LUMPUR

2020

**FUSION-FEATURES AND VISUAL-DICTIONARY
IMAGE RECOGNITION METHODS FOR APPLE
CLASSIFICATION IN SMART MANUFACTURING**

AHSIAH BINTI ISMAIL

**THESIS SUBMITTED IN FULFILMENT OF THE
REQUIREMENTS FOR THE DEGREE OF DOCTOR OF
PHILOSOPHY**

**FACULTY OF COMPUTER SCIENCE AND
INFORMATION TECHNOLOGY
UNIVERSITY OF MALAYA
KUALA LUMPUR**

2020

UNIVERSITY OF MALAYA
ORIGINAL LITERARY WORK DECLARATION

Name of Candidate: Ahsiah Binti Ismail

Matric No: 17026665/2 (WHA150001)

Name of Degree: Doctor of Philosophy

Title of Project Paper/Research Report/Dissertation/Thesis (“this Work”):

Fusion-Features and Visual-Dictionary Image Recognition Methods for Apple Classification in Smart Manufacturing

Field of Study: Image Processing (COMPUTER SCIENCE)

I do solemnly and sincerely declare that:

- (1) I am the sole author/writer of this Work;
- (2) This Work is original;
- (3) Any use of any work in which copyright exists was done by way of fair dealing and for permitted purposes and any excerpt or extract from, or reference to or reproduction of any copyright work has been disclosed expressly and sufficiently and the title of the Work and its authorship have been acknowledged in this Work;
- (4) I do not have any actual knowledge nor do I ought reasonably to know that the making of this work constitutes an infringement of any copyright work;
- (5) I hereby assign all and every rights in the copyright to this Work to the University of Malaya (“UM”), who henceforth shall be owner of the copyright in this Work and that any reproduction or use in any form or by any means whatsoever is prohibited without the written consent of UM having been first had and obtained;
- (6) I am fully aware that if in the course of making this Work I have infringed any copyright whether intentionally or otherwise, I may be subject to legal action or any other action as may be determined by UM.

Candidate’s Signature

Date:

Subscribed and solemnly declared before,

Witness’s Signature

Date:

Name:

Designation:

FUSION-FEATURES AND VISUAL-DICTIONARY IMAGE RECOGNITION METHODS FOR APPLE CLASSIFICATION IN SMART MANUFACTURING

ABSTRACT

Smart manufacturing enables an efficient manufacturing process to optimize production. The optimization is performed through data analytics that requires reliable and informative data as input. Therefore, in this research, two image recognition feature extraction methods namely Curvelet Wavelet-Gray Level Co-occurrence Matrix (CW-GLCM) and Fuzzy-Spatial Pyramid Matching (F-SPM) are proposed to provide reliable inputs for vision-based apple classification in smart manufacturing. Feature extraction is one of the major steps that could influence the efficiency of the manufacturing process. The CW-GLCM method is a feature extraction of fusion-features with Decision Tree classifier, while the F-SPM method uses a visual-dictionary based method to extract features of visual pattern and the output is processed by Support Vector Machine (SVM) classifier. To evaluate the performance of the proposed methods, they are compared with five existing methods, which are Bag of Words (BOW), Spatial Pyramid Matching (SPM), Gray Level Co-occurrence Matrix (GLCM) Texture analysis, Convolutional Neural Network (CNN) and Contrast-Limited Adaptive Histogram Equalization + GLCM + Extreme Learning Machine (CLAHE+GLCM+ELM). Three datasets which are NDDA, NDDAW and DA datasets with a total of 1310 apple images are collected to test the proposed methods. The NDDA and NDDAW datasets are both binary-class of defective and non-defective apple dataset, with NDDAW contains more low-quality region images compared to the NDDA. Conversely, the DA dataset comprised of five different types of defective apples to be used in multi-class tests. The proposed methods are trained and evaluated using 10-fold cross-validation. Their classification accuracy, precision and recall rate are then measured. Training and testing times are also recorded. From the evaluation, the proposed F-SPM method attained 98.15% classification accuracy, 96.30%

precision and 100% recall for NDDA, 91.07% for accuracy, 100% precision and 84.85% recall for NDDAW, 86.33% for accuracy, 91.43% precision and 85.00% recall for DA dataset. The F-SPM method outperformed the existing methods especially for NDDAW and DA datasets. Alternatively, the CW-GLCM method able to obtain 98.15% accuracy, 96.30% precision and 100% recall for NDDA, 89.11% accuracy, 86.79% precision and 91.01% recall for NDDAW, 85.20% of accuracy, 88.33% precision and 85.00% recall for DA dataset. The proposed CW-GLCM also shows the highest percentage (100%) for all measurements (accuracy, precision and recall) and it even outperform others in recognizing the Bruise defect. These results indicate that both proposed methods are reliable and have the potential to be used for vision classification in smart manufacturing.

Keywords: Image recognition; feature extraction; classification; smart manufacturing; data analytics.

**KAEDAH PENGECAMAN IMEJ BERASASKAN CIRI-GABUNGAN (*FUSION-
FEATURES*) DAN KAMUS-VISUAL (*VISUAL-DICTIONARY*) BAGI
PENGELASAN EPAL DI DALAM PENGILANGAN PINTAR**

ABSTRAK

Pengilangan pintar mampu membantu meningkatkan kecekapan dan mengoptimalkan pengeluaran. Pengoptimuman dapat dilakukan melalui data analisis yang memerlukan maklumat data yang boleh dipercayai sebagai input. Oleh itu, dalam kajian ini, dua kaedah pengekstrakan ciri pengecaman imej iaitu *Curvelet Wavelet-Gray Level Co-occurrence Matrix* (CW-GLCM) dan *Fuzzy-Spatial Pyramid Matching* (F-SPM) dicadangkan bagi proses pengelasan epal berasaskan visi mesin di dalam pengilangan pintar. Pengekstrakan ciri adalah salah satu langkah utama yang dapat mempengaruhi kecekapan proses pembuatan. Kaedah CW-GLCM adalah pengekstrakan ciri-gabungan (*fusion-features*) dan menggunakan pengelas Pepohon Keputusan (*Decision Tree*), manakala kaedah F-SPM menggunakan kaedah berasaskan kamus-visual (*visual-dictionary*) untuk mengekstrak ciri-ciri corak visual dan menggunakan pengelas *Support Vector Machine* (SVM). Untuk menilai prestasi kaedah-kaedah yang dicadangkan, kaedah-kaedah tersebut dibandingkan dengan lima kaedah yang sedia ada iaitu *Bag of Words* (BOW), *Spatial Pyramid Matching* (SPM), *Gray Level Co-occurrence Matrix* (GLCM) *Texture analysis*, *Convolutional Neural Network* (CNN) and *Contrast-Limited Adaptive Histogram Equalization + GLCM + Extreme Learning Machine* (CLAHE+GLCM+ELM). Tiga set data iaitu NDDA, NDDAW dan DA set data yang berjumlah 1310 imej epal dikumpulkan bagi menilai kaedah yang dicadangkan. NDDA dan NDDAW adalah set data bagi kelas perduaan imej epal yang berpenyakit atau rosak (*defective*) dan imej epal yang tidak menunjukkan sebarang penyakit atau rosak (*non-defective*), dimana set data bagi NDDAW mengandungi lebih banyak rantau berkualiti rendah berbanding NDDA. Sebaliknya, set data DA terdiri daripada lima jenis imej epal

berpenyakit atau rosak yang akan digunakan dalam ujian multi-kelas. Kaedah yang dicadangkan dilatih dan dinilai menggunakan *10-fold cross-validation*. Kadar ketepatan pengelasan, kejitian dan perolehan kembali (*recall*) bagi kaedah-kaedah tersebut kemudiannya diukur. Jumlah masa yang diambil bagi proses latihan dan ujian juga direkodkan. Berdasarkan penilaian yang dijalankan, kaedah F-SPM mencapai ketepatan pengelasan sebanyak 98.15%, 96.30% kejitian dan 100% perolehan kembali (*recall*) bagi NDDA, 91.07% ketepatan, 100% kejitian dan 84.85% perolehan kembali (*recall*) bagi NDDAW, 86.33% ketepatan, 91.43% kejitian dan 85.00% perolehan kembali (*recall*) bagi set data DA. Kaedah cadangan F-SPM melebihi ketepatan pengelasan semua kaedah yang lain terutamanya bagi set data NDDAW dan DA. Kaedah CW-GLCM pula memperoleh ketepatan sebanyak 98.15%, 96.30% kejitian dan 100% perolehan kembali (*recall*) bagi NDDA, 89.11% ketepatan, 86.79% kejitian dan 91.01% perolehan kembali (*recall*) bagi NDDAW, 85.20% ketepatan, 88.33% kejitian dan 85.00% perolehan kembali (*recall*) bagi set data DA. CW-GLCM yang dicadangkan juga menunjukkan peratusan tertinggi (100%) terhadap semua pengukuran (ketepatan, kejitian dan perolehan kembali (*recall*)) serta dapat mengatasi semua kaedah yang lain dalam mengenal pasti epal lebam (*Bruise*). Keputusan-keputusan ini menunjukkan bahawa kedua-dua kaedah yang dicadangkan boleh dipercayai dan mempunyai potensi untuk digunakan bagi pengelasan visi mesin di dalam pengilangan pintar.

Kata kunci: Pengecaman imej; pengekstrakan ciri; pengelasan; pengilangan pintar; data analisis.

ACKNOWLEDGEMENTS

First and foremost all praise to Allah the Almighty, for giving me the blessing, strength and perseverance to complete this study. It is a great pleasure to express my gratitude to those who also supported and contributed in making this thesis possible.

My deepest and sincere gratitude to my supervisors, Assoc. Prof. Dr. Mohd Yamani Idna Idris, Dr. Mohamad Nizam Ayub and Assoc. Prof. Dr. Por Lip Yee for their unlimited support and provided me great guidance throughout my study. I am truly grateful and very much appreciate their invaluable guidance, patience and encouragement throughout and along this journey. Without all their knowledge and dedicated involvement, this study would never been completed.

I would like to express my deepest sense of gratitude also to my beloved husband, Amirul Amri Ahmad Roslan, who have been ultimate source of motivation with his precious understanding, encouragement and support with his love and patience throughout my study journey. Thanks also to my lovely little son, Ammar Harith for constantly showers me with joy and happiness.

I would like to extend my sincere appreciations to my mother, Supiah Said, sister, Allezawati and all my family and friends for their unconditional love, endless supports, prayers and encouragements throughout my years of study.

Last, but not least, a special acknowledgement also goes to MyBrain15 program of the Ministry of Higher Education, Malaysia for providing financial support during this PhD study. This research was also supported by Postgraduate Research Grant (PPP)–Research (PG040-2015B).

TABLE OF CONTENTS

Abstract	iii
Abstrak	v
Acknowledgements	vii
Table of Contents	viii
List of Figures	xii
List of Tables.....	xv
List of Symbols and Abbreviations.....	xvii
List of Appendices	xix
CHAPTER 1: INTRODUCTION.....	1
1.1 Introduction.....	1
1.2 Problem Statement.....	6
1.3 Research Questions.....	8
1.4 Research Objectives.....	9
1.5 Research Scopes and Limitations	10
1.6 Thesis Contributions	10
1.7 Significance of Research	11
1.8 Thesis Overview	12
CHAPTER 2: LITERATURE REVIEW.....	16
2.1 Smart Manufacturing	16
2.2 Machine Vision.....	18
2.3 Image Recognition Method for Machine Vision Classification	19
2.3.1 Statistical	21
2.3.2 Morphology and Spectral	28

2.3.3	Model-Based.....	35
2.3.4	Deep Learning	45
2.4	Discussion.....	49
2.5	Summary.....	51
CHAPTER 3: RESEARCH METHODOLOGY.....		52
3.1	Introduction.....	52
3.2	Data Collection	54
3.2.1	Binary-Class Datasets.....	57
	3.2.1.1 NDDA	57
	3.2.1.2 NDDAW.....	58
3.2.2	Multi-Class Dataset	59
3.3	Materials and Methods	60
3.3.1	BOW	61
3.3.2	SPM.....	62
3.3.3	CNN.....	62
3.3.4	Texture Analysis.....	63
3.3.5	CLAHE+GLCM+ELM	63
3.4	Proposed Methods	64
3.4.1	Feature Extraction	66
	3.4.1.1 CW-GLCM (Proposed Method I)	66
	3.4.1.2 F-SPM (Proposed Method II).....	67
3.4.2	Classification	68
3.5	Performance Evaluation.....	71
3.5.1	Binary-Class Evaluation.....	74
3.5.2	Multi-Class Evaluation.....	74
3.6	Chapter Summary	75

CHAPTER 4: CW-GLCM METHOD	76
4.1 Introduction.....	77
4.2 CW-GLCM Method.....	78
4.2.1 Feature Extraction	80
4.2.1.1 Curvelet Transform	80
4.2.1.2 Wavelet Transform.....	82
4.2.1.3 GLCM	85
4.2.2 Classification	89
4.3 Experimental Results and Analysis	89
4.3.1 Performance Measure for Fusion-Features	89
4.3.2 Comparison with Existing Methods	94
4.3.3 Analysis of Classification Performance against Low-Quality Region.....	96
4.3.4 Performance Evaluation with Different Image Resolution	100
4.4 Discussion.....	101
4.5 Summary.....	102
CHAPTER 5: F-SPM METHOD	104
5.1 Introduction.....	105
5.2 F-SPM Method	106
5.2.1 Feature Extraction	107
5.2.1.1 Fuzzy Logic Detection	108
5.2.1.2 SIFT Descriptor.....	111
5.2.1.3 Spatial Pyramid Pooling.....	111
5.2.1.4 Clustering	113
5.2.2 Classifier.....	114
5.3 Experimental Result.....	114
5.3.1 Performance Evaluation	115

5.3.2	Performance Measure for Edge Detection	117
5.3.3	Comparison of Different Classifier	118
5.3.4	Effectiveness of the Proposed F-SPM Method	121
5.3.4.1	Binary-Class Classification Performance.....	122
5.3.4.2	Performance on Multi-Class Classification.....	127
5.4	Discussion.....	136
5.5	Summary.....	136
CHAPTER 6: CONCLUSION.....		138
6.1	Conclusion	138
6.2	Future Work Direction.....	140
References		143
List of Publications and Papers Presented		158
Appendix A: <i>K</i> -fold Cross-Validation Evaluation.....		159

LIST OF FIGURES

Figure 1.1: Fruits production recorded between 2005-2016. The graph is obtained from Ministry of Agriculture (USDA, 2017).....	4
Figure 1.2: Illustration for smart manufacturing which employ visual sensors that enable automated inspection, sorting and advanced further analytics to improve productivity (defective (D), non-defective (N), Scab (Sb), Rot (Rt), Cork Spot (Ct), Blotch (Bh) and Bruise (Be)) ¹	5
Figure 1.3: Overview and organization of this thesis. Arrow shows the chapters are combined to achieve the thesis goal.	13
Figure 2.1: Integration of computational intelligent in smart manufacturing lifecycle (J. Wang et al., 2018).	17
Figure 2.2: Image recognition workflow.	20
Figure 2.3: Categories of image recognition method for classification.	21
Figure 2.4: Examples the SIFT keypoint matching of (a) same apple images (b) different apple images.	39
Figure 2.5: Architecture of BOW for image classification.	40
Figure 2.6: Structure of the CNN for apple image.	46
Figure 3.1: Research methodology and design.	53
Figure 3.2: Illustration of the data acquisition process using vision sensors.	56
Figure 3.3: Datasets construction diagram.	57
Figure 3.4: Examples of apple images in the NDDA dataset (a) non-defective (b) defective ¹	58
Figure 3.5: Examples of apple images in the NDDAW dataset (a) non-defective (b) defective. The low-quality regions on the apple skin are indicated by the arrows.	59
Figure 3.6: Examples of defective images in the DA dataset (a) Scab (b) Rot (c) Cork Spot (d) Blotch (e) Bruise.	60
Figure 3.7: Overall scheme of five existing image recognition methods (BOW, SPM, CNN, Texture analysis and CLAHE+GLCM+ELM).	61
Figure 3.8: Proposed CW-GLCM image recognition method.	65

Figure 3.9: Proposed F-SPM image recognition method.....	65
Figure 3.10: Illustration of 10-fold cross-validation.....	73
Figure 4.1: Process flow of the proposed CW-GLCM method for apple classification ²	79
Figure 4.2: Representation of edges in Curvelet transform.	81
Figure 4.3: Representation of edges in Wavelet transform.....	83
Figure 4.4: Image decomposition using analysis filter banks. Note that h is low-pass filter and g is high-pass filer and $\downarrow 2$ is keeping one sample out of two (down sampling) ²	84
Figure 4.5: 2D-DWT wavelet decomposition at two resolution levels (a) the structure diagram (b) apple image via wavelet decomposition at two resolution levels ²	85
Figure 4.6: Directionality of GLCM ²	87
Figure 4.7: Comparative results for the proposed fusion-features using SVM classifier on NDDA dataset (a) precision, recall and accuracy (b) training and testing time.	91
Figure 4.8: Comparative results of fusion-features for different classifier on NDDA dataset (a) precision, recall and accuracy (b) training and testing time.	92
Figure 4.9: Examples of misclassified apple images with a low-quality region. The low- quality regions on apple skin images are pointed by the arrows ²	96
Figure 4.10: Comparative results of precision, recall and accuracy in (a) NDDA dataset (b) NDDAW dataset (c) training time and (d) testing time for BOW (Csurka et al., 2004), SPM (Lazebnik et al., 2006), CNN (dos Santos Ferreira et al., 2017), Texture analysis (Olaniyi et al., 2017), CLAHE+GLCM+ELM (W. Li et al., 2019) and proposed CW-GLCM method.	99
Figure 4.11: Examples misclassification images of the proposed CW-GLCM on (a) NDDA and (b) NDDAW	102
Figure 5.1: (a) Calyx (b) stem end (c) defect ¹	106
Figure 5.2: The proposed F-SPM method for apple classification.	107
Figure 5.3: Fuzzy logic detection (a) original image (b) grayscale (c) Fuzzy logic.	109
Figure 5.4: 3x3 Kernal used for Fuzzy logic detection.	109

Figure 5.5: (a) Grayscale image (b) I_x (c) I_y (d) Fuzzy logic.	110
Figure 5.6: Membership functions of the inputs/outputs of edge.	110
Figure 5.7: Feature histogram of level 2 pyramid.....	112
Figure 5.8: Example of two feature detection techniques for apple image classification (a) feature detector using Dense regular grid (b) feature detector using <i>keypoint</i> (SIFT).	116
Figure 5.9: Comparative results of the proposed visual-dictionary features with different classifier on NDDA dataset (a) precision, recall and accuracy (b) training and testing time.	119
Figure 5.10: Comparative results of precision, recall and accuracy in (a) NDDA dataset (b) NDDAW dataset (c) training time (d) testing time for BOW (Csurka et al., 2004), SPM (Lazebnik et al., 2006), CNN (dos Santos Ferreira et al., 2017), Texture analysis (Olaniyi et al., 2017), CLAHE+GLCM+ELM (W. Li et al., 2019), CW-GLCM (Proposed method I) and F-SPM (Proposed method II).....	125
Figure 5.11: Comparative results of precision, recall and accuracy between the proposed methods with the existing image recognition methods for each individual class (a) Blotch, (b) Bruise, (c) Cork Spot, (d) Scab and (e) Rot on DA dataset.	131
Figure 5.12: Examples of unobvious Bruise defects (a) red skin apple and (b) green skin apple.....	134
Figure 5.13: Comparative results of precision, recall and accuracy for each method on DA dataset (a) average performance (b) training and testing time.....	135

LIST OF TABLES

Table 2.1: Highlighted studies on machine vision classification using statistical approach	26
Table 2.2: Highlighted studies on machine vision classification using morphology and spectral approach.....	33
Table 2.3: Highlighted studies on machine vision classification using model-based approach.	43
Table 2.4: Image classification using CNN method of deep learning approach.....	47
Table 2.5: List of commonly used image recognition methods including their strengths and limitations.	49
Table 3.1: Keywords used for collected images via the Google search engine.	56
Table 3.2: Details of the characteristics in NDDA dataset.	58
Table 3.3: Details of the characteristics in NDDAW apple dataset.....	59
Table 3.4: Details of the characteristics in DA dataset.	60
Table 3.5: Terminology and derivations of the evaluation metrics.	74
Table 3.6: Measures precision, recall and accuracy for multi-class classification based on evaluation metrics for five defect classes C_i : TP_i are true positive for C_i , FP_i are false positive, FN_i are false negative and TN_i are true negative counts respectively.....	75
Table 4.1: Parameters setting of Curvelet, Wavelet and GLCM.	88
Table 4.2: Comparative results (precision, recall, accuracy, training time and testing time) for the proposed fusion-features using SVM classifier on NDDA.....	90
Table 4.3: Comparison of the proposed fusion-features with different classifiers on NDDA dataset.	91
Table 4.4: Comparison of confusion matrix for BOW (Csurka et al., 2004), SPM (Lazebnik et al., 2006), CNN (dos Santos Ferreira et al., 2017), Texture analysis (Olaniyi et al., 2017), CLAHE+GLCM+ELM (W. Li et al., 2019) and the proposed CW-GLCM method on NDDA dataset: Defective (D) and Non-defective (N).	94
Table 4.5: Confusion matrix for BOW (Csurka et al., 2004), SPM (Lazebnik et al., 2006), CNN (dos Santos Ferreira et al., 2017), Texture analysis (Olaniyi et al., 2017),	

CLAHE+GLCM+ELM (W. Li et al., 2019) and proposed CW-GLCM method on NDDAW dataset: Defective (D) and Non-defective (N).	97
Table 4.6: Comparative results (precision, recall, accuracy, training time and testing time) of the proposed method for different resolution images.....	100
Table 5.1: Comparison of SPM method performances between two feature detectors on NDDA dataset.	115
Table 5.2: Comparison of edge detection techniques in modified SPM method on NDDA dataset.	117
Table 5.3: Comparison of the proposed visual-dictionary features with different classifiers on NDDA dataset.....	118
Table 5.4: Comparison of confusion matrix for BOW (Csurka et al., 2004), SPM (Lazebnik et al., 2006), CNN (dos Santos Ferreira et al., 2017), Texture analysis (Olaniyi et al., 2017), CLAHE+GLCM+ELM (W. Li et al., 2019), CW-GLCM (Proposed Method I) and F-SPM (Proposed Method II) on NDDA dataset: Defective (D) and Non-defective (N).....	123
Table 5.5: Comparison confusion matrix for BOW (Csurka et al., 2004), SPM (Lazebnik et al., 2006), CNN (dos Santos Ferreira et al., 2017), Texture analysis (Olaniyi et al., 2017), CLAHE+GLCM+ELM (W. Li et al., 2019), CW-GLCM (Proposed Method I) and F-SPM (Proposed Method II) on NDDAW dataset: Defective (D) and Non-defective (N).....	124
Table 5.6: Comparison results for BOW (Csurka et al., 2004), SPM (Lazebnik et al., 2006), CNN (dos Santos Ferreira et al., 2017), Texture analysis (Olaniyi et al., 2017), CLAHE+GLCM+ELM (W. Li et al., 2019), CW-GLCM (Proposed Method I) and F-SPM method (Proposed Method II) on DA dataset.....	128

LIST OF SYMBOLS AND ABBREVIATIONS

GLCM	:	Gray Level Co-occurrence Matrix
SVM	:	Support Vector Machine
SPM	:	Spatial Pyramid Matching
BOW	:	Bag of Words
CNN	:	Convolutional Neural Network
KNN	:	K-Nearest Neighbors
LBP	:	Local Binary Patterns
HIS	:	Hue, Saturation and Intensity
LDA	:	Linear Discriminant Analysis
RGB	:	Red, Green, Blue
RVM	:	Relevance Vector Machine
ELBP	:	Elliptical Local Binary Pattern
LDP	:	Local Directional Pattern
M-ELBP	:	Mean -ELBP
GLRLM	:	Gray Level Run Length Matrix
PCA	:	Principal Component Analysis
SDA	:	Stepwise Discriminant Analysis
RFE	:	Recursive Feature Elimination
HSV	:	Hue, Saturation and Value
RBPNN	:	Radial Basis Probabilistic Neural Network
PSO	:	Particle Swarm Optimization
RBF	:	Radial Basis Function
CLAHE	:	Contrast-Limited Adaptive Histogram Equalization
ELM	:	Extreme Learning Machine

AR	:	Autoregressive
VMD	:	Variational Mode Decomposition
SIFT	:	Scale Invariant Feature Transform
SURF	:	Speeded up Robust Features
FAST	:	Features from Accelerated Segment Test
DoG	:	Difference of Gaussians

LIST OF APPENDICES

Appendix A: <i>K</i> -fold Cross-Validation Evaluation	159
--	-----

CHAPTER 1: INTRODUCTION

This chapter presents a brief introduction to smart manufacturing, image recognition and motivation behind this research work. Then, problem statement, research questions, research objectives and the research scopes are outlined. The contribution and significance of this research are also explained. Finally, the overview structure of this thesis is presented.

1.1 Introduction

Smart manufacturing employs a high level adaptability of computer control and various technologies into the existing manufacturing process in optimizing the productivity. A huge volume, variety and velocity of data in smart manufacturing or referred to as big data, offers an opportunity not only for managing large amount of information, but also to improved diagnostics and prognostics capabilities (Moyne & Iskandar, 2017; Nagorny, Lima-Monteiro, Barata, & Colombo, 2017). The analytics in the manufacturing process can shift from a reactionary to a predictive practice by improving the existing capabilities such as product defect detection and supporting new capabilities for future planning and prediction (Raghupathi & Raghupathi, 2014; Wan et al., 2017). In delivering high quality predictive solution for future planning, the data quality is the most important big data factor (Moyne & Iskandar, 2017). The effective and accurate method is required to provide reliable information as the input for analytics models to make a better decision (Shin, Woo, & Rachuri, 2014).

Therefore, in this research, the reliability of image recognition method to classify between defective and non-defective apple including types of defect for automated inspection and sorting processes is investigated. The defective and non-defective effect

including types of the defective information can be further used as the input for analytics model for future prediction. The research focuses on the detection of suitable features and feature extraction method that able to increase the classification accuracy of binary-class defective and non-defective apple including multi-class classification types of defects apple. The apple classification is challenging due to the physical and biological influence such as the presence of low-quality regions of bright features or flecks features on the apple skin and the appearances of the defects can be similar to the stem ends or calyxes that are the natural part of the apple. As agriculture product, apples have different colors depending on its cultivar as well as various types of defect and severity, which makes detecting between defective and non-defective challenging.

In fruit production of agriculture industry, apple have the highest production rate and had steadily increase over the year as reported by United States Department of Agriculture (USDA) (USDA, 2017) (see Figure 1.1). The proposed method could be beneficial to help on the production productivity and improving the product quality of apple. This is possible since the proposed method can be implemented in the inspection process of the manufacturing for automatic inspection and sorting processes. Further data analytics can also be performed based on the current production data of the defective and non-defective including types of the defect in apple production to improve the apple growth and processing efficiency as illustrate in Figure 1.2. However, this requires accurate and reliable informative data as input for analytics. The detection and extraction of important features such as low-quality region and the spatial information is crucial to differentiate between defective and non-defective apples, including types of the defect. Failure to detect these features may reduce the classification accuracy.

For this reason, two new image recognition feature extraction methods are proposed. The first proposed method called Curvelet Wavelet-Gray Level Co-occurrence Matrix

(CW-GLCM) is the feature extraction method of fusion-features based on Gray Level Co-occurrence Matrix (GLCM) method and used the Decision Tree as classifier. The second proposed method called Fuzzy-Spatial Pyramid Matching (F-SPM) is the feature extraction of visual-dictionary features inspired from Spatial Pyramid Matching (SPM) method and use the Support Vector Machine (SVM) classifier.

The GLCM method is selected since it can analyze and describe the spatial relationship of a neighboring pixels on the surface structure of the images properties. However, this method is dependent on the images texture information (Maule, Shete, Wani, & Dawange, 2015), which can be unstable on the low-quality region. Thus, the features may not be effectively extracted from the low-quality region images.

Alternatively, the SPM is chosen due to the spatial layout information included in the method that will help in distinguishing between the defect or natural part of stem end or calyx of the apple. However, the limitation of the SPM method is that it generates large numbers of unnecessary and redundant high dimensionality features (Chanti & Caplier, 2018; Lin, Tsai, Chen, & Ke, 2016; Penatti, Silva, Valle, Gouet-Brunet, & Torres, 2014). These irrelevant features may affects the stability of the method and reduce performance on apple classification. Therefore, it is important to propose new image method for defective and non-defective apple even with low-quality region images including types of the defect to suite the problem stated previously.

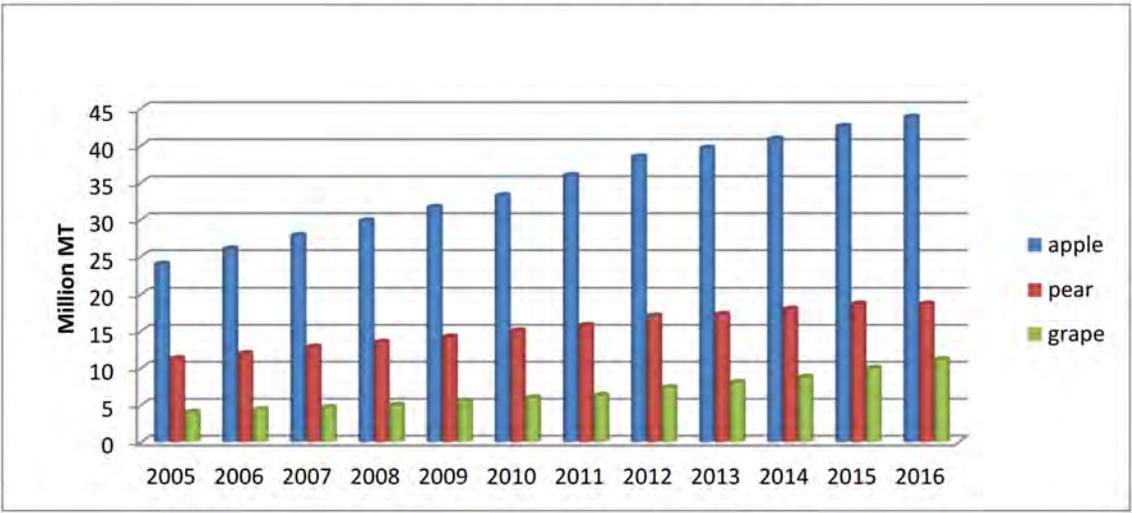


Figure 1.1: Fruits production recorded between 2005-2016. The graph is obtained from Ministry of Agriculture (USDA, 2017).

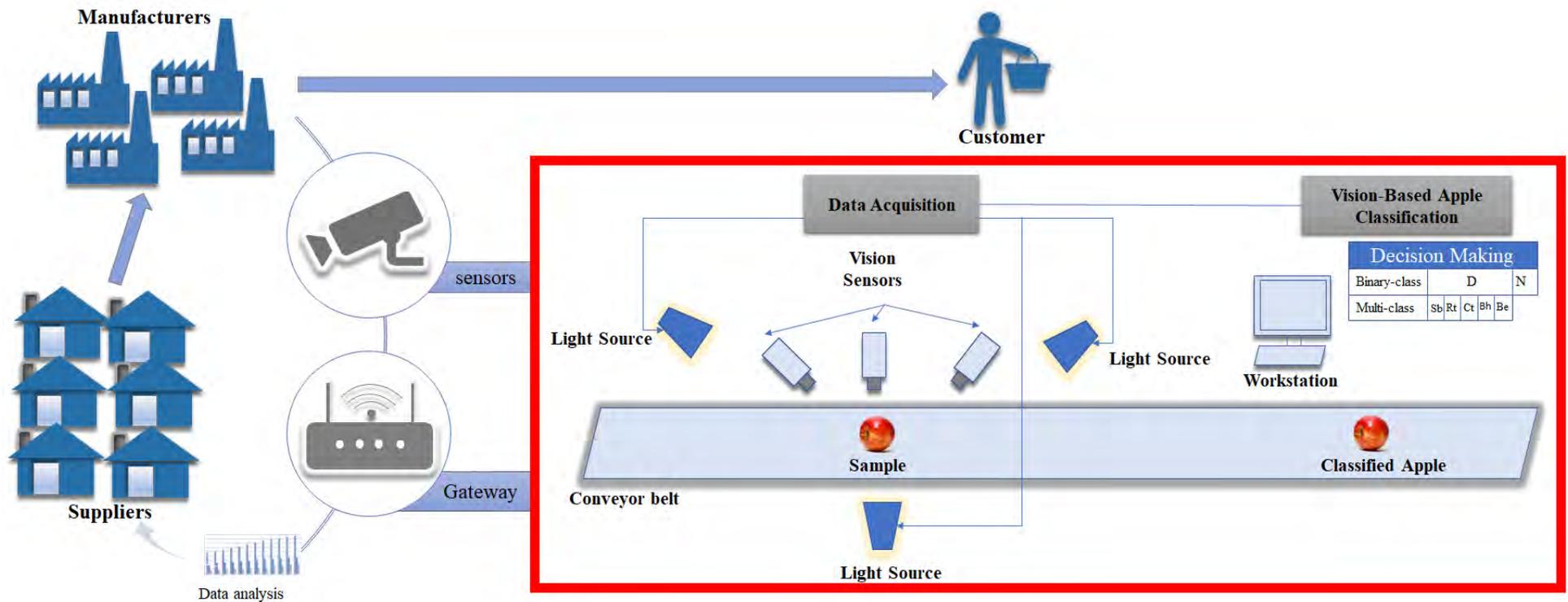


Figure 1.2: Illustration for smart manufacturing which employ visual sensors that enable automated inspection, sorting and advanced further analytics to improve productivity (defective (D), non-defective (N), Scab (Sb), Rot (Rt), Cork Spot (Ct), Blotch (Bh) and Bruise (Be))¹.

¹ Ismail, A., Idris, M. Y. I., Ayub, M. N., & Por, L. Y. (2018). Vision-Based Apple Classification for Smart Manufacturing. *Sensors*, 18, 4353.

1.2 Problem Statement

In image recognition, attaining the capability of human recognition remains a challenge (Ying Zhang, Chen, Huang, & Gao, 2019). It is become more challenging for defective and non-defective apple classification including types of defects. It requires image recognition method that is capable to detect and extract low-quality region features and the spatial information in the images to precisely classify between defective with non-defective apple. However, the existing image recognition methods may not be feasible to detect these kinds of features and subsequently, poses a great research problem.

The detection performance of the image recognition method behaves differently with respect to the type of object and the image's information complexity. The existing image recognition methods reported in the literature focus on the use of the texture, keypoint and visual-dictionary features. The texture based image recognition method such as GLCM can analyze and describes the spatial relationship of neighboring pixels on the surface structure of the image. The GLCM method is an effective method to recognize the object with texture information but ineffective when different objects exhibit a similar texture and low-quality regions (Fahrurozi, Madenda, & Kerami, 2016; Maule et al., 2015).

The keypoint features extensively used in image recognition due to their various advantages. It is initially designed for matching between two images (Fan, Jin, Wang, Zhang, & Li, 2019). However, the ability of solely dependent on keypoint feature is limited to the classification of the same object from the same images. This is due to difficulties to match the keypoint between the same object for a different image. Other than that, exclusively dependent on keypoint feature unable to define a shape or images due to non-uniform distribution. Thus, the keypoint feature requires a complementary

method to encode the information of the neighborhood keypoint for the classification task.

The visual-dictionary based image recognition method utilizes keypoint patches or Dense regular grid patches to extract visual patterns (visual words) from the images. Examples of the well-known visual-dictionary based method are Bag of Words (BOW) and SPM. However, the visual-dictionary based method generates a large number of unnecessary and redundant high dimensionality features (Chanti & Caplier, 2018; Q. Li, Peng, Li, & Ren, 2017; Lin et al., 2016; Penatti et al., 2014; Yan, Xu, Xu, Lin, & Li, 2012). These irrelevant features can reduce the stability and performance of the method.

From the issues mentioned above, the features utilized in the existing image recognition methods are limited in the classification of apple images which contains low-quality region and requires spatial information. There are two major concerns of the existing image recognition methods for apple classification. These concerns can be summarized as follows:

- i. Texture based image recognition methods such as GLCM have difficulty to distinguish different object with quite similar texture and images with low-quality region.
- ii. Visual-dictionary based methods (i.e. BOW and SPM) generates large numbers of unnecessary and redundant high dimensionality features, therefore these irrelevant features can reduce the stability and performance of the method.

Therefore, this thesis proposed two feature extraction methods for image recognition to extract features from the low-quality region and spatial information that reduce irrelevant features so that only significant features are selected for further classification. These methods are proposed for binary-class classification of defective and

non-defective apples including low-quality region images. In this research, apart from investigating between the binary-class defective and non-defective apple, the second problem concentrate on investigating the defective apple and extended to multi-class type of defects. The multi-class classification between types of defects allows the recognition of the specific defective type. This include various severity of defects whether they are obvious or unobvious defect.

1.3 Research Questions

The research questions for this thesis include:

- i. What are the characteristics of the apple images?
- ii. What does the effect of low-quality region and spatial information on apple images?
- iii. How to extract low-quality region on apple images?
- iv. How to extract spatial information on apple images?
- v. How the proposed method improves the performance of other existing image recognition methods in extracting low-quality region on apple images?
- vi. How the proposed method include spatial information and selects only significant features for further classification to reduce unnecessary features for apple classification?
- vii. How the proposed method improves the classification performance of the existing image recognition methods?

1.4 Research Objectives

The main aim of this research is to propose feature extraction image recognition methods to detect and classify defective and non-defective apples even in the presence of the low-quality region images including types of defects. This aim is achieved with the following objectives:

- i. To investigate the suitability of the existing image recognition methods for apple classification.
- ii. To propose a set of fusion-features that able to effectively classify between defective and non-defective apple images including images with low-quality region.
- iii. To propose visual-dictionary features that eliminate unnecessary high dimensionality features, where the elimination of the unnecessary features can increase the accuracy performance of the apple classification methods including their types of defects.
- iv. To evaluate the classification performance of the proposed image recognition methods.

In the first objective, the main focus is to investigate the strength and limitations of the existing image detection and recognition methods that can be applied to apple classification. On the other hand, the second objective is looking into methods for detecting and extracting suitable features for classifying images with low-quality region. It is believed that the introduction of highly informative fusion-features able to enhance the feature detection on low-quality image region, which will lead to the increase of classification accuracy between defective and non-defective apple. In the third objective, a method to extract only significant features from spatial layout information is devised.

For this objective, unnecessary and redundant high dimensionality features from the spatial layout information (extracted using SPM method) is reduced or eliminated. These irrelevant features can reduce the accuracy performance in apple classification. Therefore, a visual-dictionary feature extraction method is proposed to reduce unnecessary features produced by the SPM method. The selection of only reliable and significant features is expected to increase the accuracy performance in apple classification including their types of defects. In the fourth objective, the performance of the proposed image recognition methods on apple classification will be evaluated and analyzed. The performances of the proposed image recognition methods are also compared with the existing image recognition methods.

1.5 Research Scopes and Limitations

The scope and limitation of this research are outlined as follows:

- i. This research focuses on the image recognition method for apple classification.
- ii. The datasets used in the experiments is limited to apple images.

1.6 Thesis Contributions

This research proposed two image recognition methods of CW-GLCM and F-SPM. These methods are proposed to detect and classify binary-class classification of defective and non-defective apples including low-quality region images and multi-class classification between types of defects. The CW-GLCM is a fusion-features based on the GLCM Texture analysis method whereas the F-SPM is a visual-dictionary feature based on the SPM method. The performances of CW-GLCM and F-SPM are evaluated and

compared with the existing image recognition methods on three datasets. Additional contributions of this thesis are outlined as follows:

- i. The limitations of the existing image recognition methods are identified from the conducted literature review.
- ii. The limitations of the existing image recognition methods on apple classification are established through performance evaluation and analysis.
- iii. A new image recognition method based on the GLCM Texture analysis method is implemented using a proposed fusion-features for the detection on the low-quality region of apple images.
- iv. A new image recognition method of the visual-dictionary features is proposed by incorporating the spatial layout information of the SPM. The method concentrates on reduces unnecessary SPM features through Fuzzy logic detection to include only significant features for classification and improving the detection of the defective apple.
- v. Finally, the directions of the future research are presented.

1.7 Significance of Research

This research proposed new image recognition methods for binary-class classification of defective and non-defective apple images including low-quality region image and multi-class classification between types of defects. The proposed methods are tested on three newly created apple datasets, namely NDDA, NDDAW and DA. The effectiveness and reliability of the proposed methods in the binary-class and multi-class classifications of the apple images will benefit the manufacturing industry in optimizing the productivity. The proposed method can be used for automatic inspection and sorting

processes. Other than that, further data analytics can also be performed based on the current production data of the defective and non-defective apple including the types of defective in apple production. The data analytics identifies and learns the patterns for future planning and prediction, which would help in improving the apple growth and processing efficiency.

1.8 Thesis Overview

This thesis is divided into six chapters as summarised in Figure 1.3. The summary of each chapter is briefly described as follows: (The thesis interpolates materials from the article's title "Vision-Based Apple Classification for Smart Manufacturing" and "Investigation of Fusion Features for Apple Classification in Smart Manufacturing" written by Ismail et al., (Ismail, Idris, Ayub, & Por, 2018; Ismail, Idris, Ayub, & Yee, 2019); the first author is the author of this thesis).

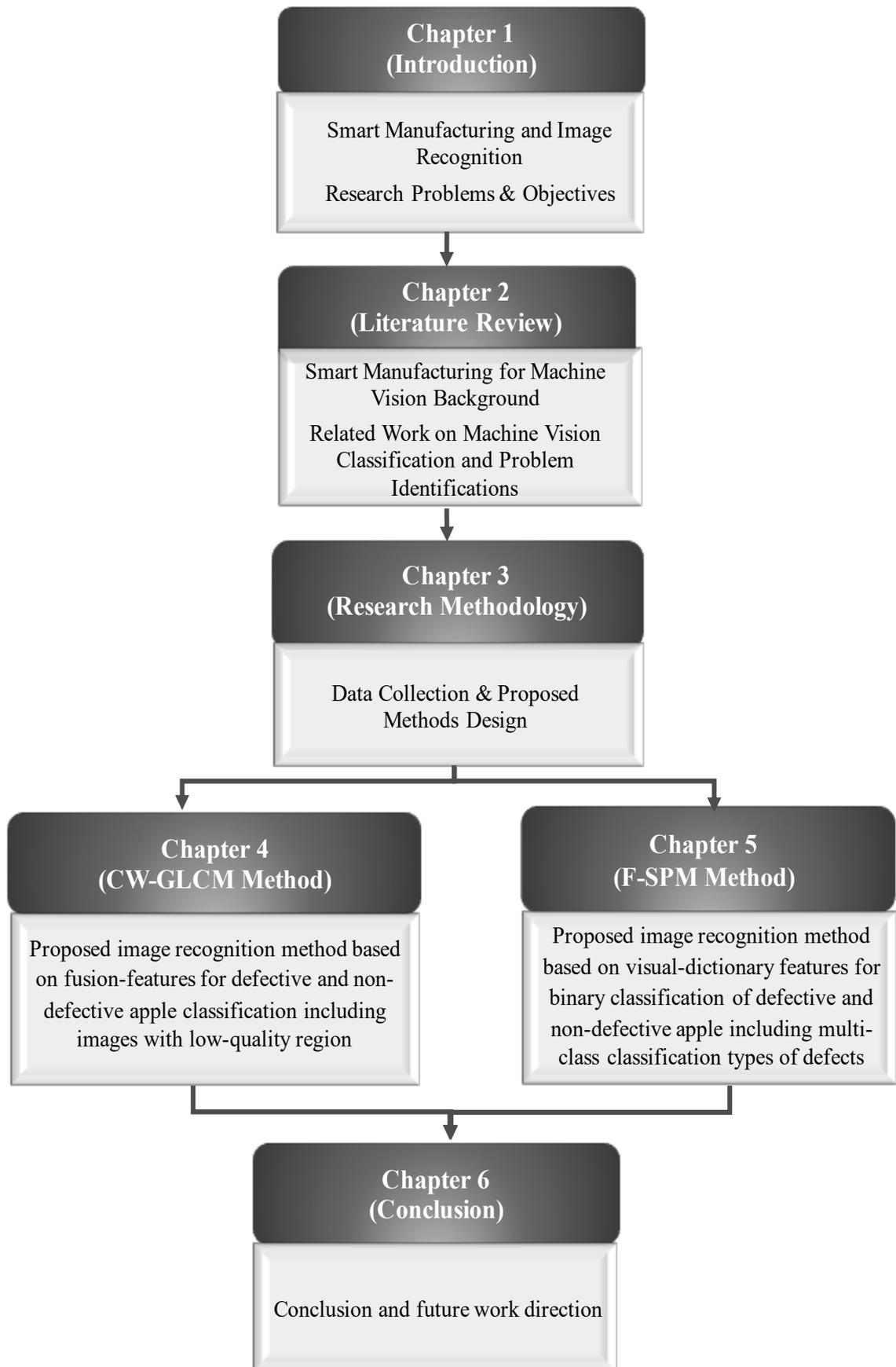


Figure 1.3: Overview and organization of this thesis. Arrow shows the chapters are combined to achieve the thesis goal.

Chapter 1 introduces the image recognition technologies in smart manufacturing. The related problems of image recognition in the apple classification are clearly defined. Then, the research questions, objectives, research scopes and the motivation of this research are presented.

Chapter 2 presents the overview of the smart manufacturing. Then, the related image recognition methods for machine vision classification in the literature are reviewed and analyzed for their strength and limitation. The related image recognition methods include statistical, morphology and spectral, model-based and deep learning approaches.

Chapter 3 describes the overall methodology in this research. The chapter contains plans, structure and strategy of investigation, data acquisition and explanation of the proposed methods.

Chapter 4 explains the proposed CW-GLCM method which based on the GLCM method incorporating Curvelet and Wavelet transform. The CW-GLCM method is demonstrated in binary-class classification of defective and non-defective apple images including images with low-quality region.

Chapter 5 presents the second proposed method of F-SPM. The F-SPM method improves the drawbacks of the proposed CW-GLCM method in detecting defective apples by incorporating spatial layout information of SPM. The F-SPM method also reduces the unnecessary SPM features through Fuzzy logic detection to include only significant features for further classification. The effectiveness of the F-SPM method in the binary-class classification of defective and non-defective apple images is also extended to recognizing the types of defective apples in the multi-class classification. The multi-class classification between types of defects allows the recognition of the specific defective type when the testing images belong to the certain defective category.

Chapter 6 summarizes the findings of the proposed CW-GLCM and F-SPM methods.
Then, the direction for future research is suggested.

CHAPTER 2: LITERATURE REVIEW

This chapter introduces the smart manufacturing and machine vision using image recognition method for classification. First, the overview of smart manufacturing and machine vision classification are presented. Then, the image recognition methods covering the statistical, morphology and spectral, model-based and deep learning approaches are reviewed. The review discusses the strength and limitation of these methods in the defective and non-defective classification. Furthermore, the potential of these methods in the classification of the defective and non-defective apple images including the types of defects are presented. Finally, brief discussions of the methods utilized in the proposed methods are provided.

2.1 Smart Manufacturing

The smart manufacturing representing the manufacturing revolution which integrates high level adaptability of computer control and various technologies into the existing manufacturing process. The integration offers the manufacturing intelligence that can increase flexibility, quality of the production process and optimize productivity (Rüßmann et al., 2015; J. Wang, Ma, Zhang, Gao, & Wu, 2018). The statistics had reported 82% of the companies that employ smart manufacturing experienced increase in efficiency and 45% of the companies experienced increase in customer satisfaction (J. Wang et al., 2018). There are two important concepts in smart manufacturing which are data and automation (L. Li, Ota, & Dong, 2018). The data can be acquired via sensors at any level of the manufacturing process. These sensory data can be learned and analyzed using computational intelligence. The computational intelligence is an important part of the manufacturing intelligence to enable accurate decision making for automation in the manufacturing process. The computational intelligence has been widely investigated at

different lifecycle stages of the manufacturing industry covering the concept development, design, evaluation, production, operation and sustainment as shown in Figure 2.1.

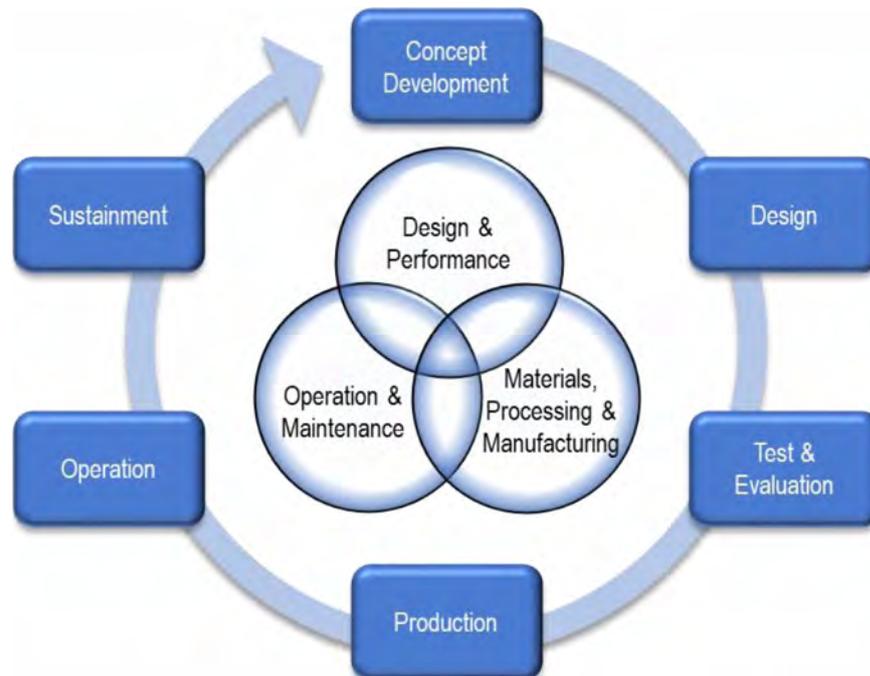


Figure 2.1: Integration of computational intelligent in smart manufacturing lifecycle (J. Wang et al., 2018).

This research emphasizes the importance of computational intelligence in smart manufacturing. Particularly, the implementation of the machine vision through vision-based classification in the manufacturing lifecycle stages. The vision sensor data collected across the manufacturing level can be used for vision-based classification to enable automatic inspection, sorting processes and further data analytics based on the current production data. The analytics process will identify and learn the patterns for future planning and prediction to improve production and processing efficiency.

2.2 Machine Vision

The technological advances in sensors, machines and internet system have driven interest for machine vision system in many industries such as agriculture, food processing industry, medical devices technology, environmental monitoring, security surveillance and others (Da Xu, He, & Li, 2014; Mehdizadeh, Minaei, Hancock, & Torshizi, 2014; Neves, Mehdizadeh, Tschärke, de Alencar Nääs, & Banhazi, 2015; Semeniuta, Dransfeld, Martinsen, & Falkman, 2018; B. Zhang et al., 2014). The implementation of the machine vision system in smart manufacturing can potentially optimize productivity by exploiting visual sensing technologies with the evaluation of big data (Ebrahimi, Khoshtaghaza, Minaei, & Jamshidi, 2017; Koch, Georgieva, Kasireddy, Akinci, & Fieguth, 2015). The visual sensing technologies acquire the data at any level of the manufacturing process that can be transmitted, shared and exchanged to improve the product quality and the production productivity. While the evolution of big data which refers to a huge volume, variety and velocity of data in smart manufacturing offers an opportunity not only for managing large amount of information, but also offers improvement on the diagnostics and prognostics capabilities (Moyne & Iskandar, 2017; Nagorny et al., 2017). The analytics in the manufacturing process can shift from a reactionary to a predictive practice by improving the existing capabilities such as product defect detection and supporting new capabilities for future planning and prediction (Raghupathi & Raghupathi, 2014; Wan et al., 2017).

In general, the machine vision for the defect detection and classification use vision sensors to acquire images and computer vision solution to detect and classify the desired features (Ebrahimi et al., 2017; Semeniuta et al., 2018). The computer vision extracts important features from the images to facilitate the computer understanding and interpretation of the contents in the images (K. K. Patel, Kar, Jha, & Khan, 2012; B. Zhang

et al., 2018). Basically, the image processing method use for machine vision to automate the defect detection and classification are built upon image recognition method.

2.3 Image Recognition Method for Machine Vision Classification

This section presents the image recognition method for machine vision classification of defective and non-defective as well as other classification tasks. In any classification tasks, identifying relevant and important features is an essential first step to accurately represent the image information. Then, the extracted features are learned using the classifier in the classification process to identify the object (Koch et al., 2015). The defective classification presents a unique challenge. In contrast to other classification task, there are some criterions that makes defective classification is more challenging. In defective classification, the types of defects and its severity influence the outcome of the classification. Thus, a further strategy is required to precisely detect the defective region and classify between the defective and non-defective.

The general workflow of the image recognition method as presented in Figure 2.2 consists of two main phases, namely feature extraction and feature classification. These phases represent the image in a way that the computer can understand and perform the classification task on those representations. The feature extraction phase extracts relevant or important features from an image and use it for recognizing the object (X. Chen, Kopsaftopoulos, Wu, Ren, & Chang, 2019). The features are the measurable characteristic of the object such as color, shape, edge, texture, keypoint and visual-dictionary features. The similarity between objects can also be determined through these features.

The classification phase assigns the extracted features to probability classes based on their similarity then be classified into its classes using a classifier (Antonucci & Corani, 2017; Zhao, Zheng, Xu, & Wu, 2019). The classifiers compare the input features with the stored pattern to find the best matching class.

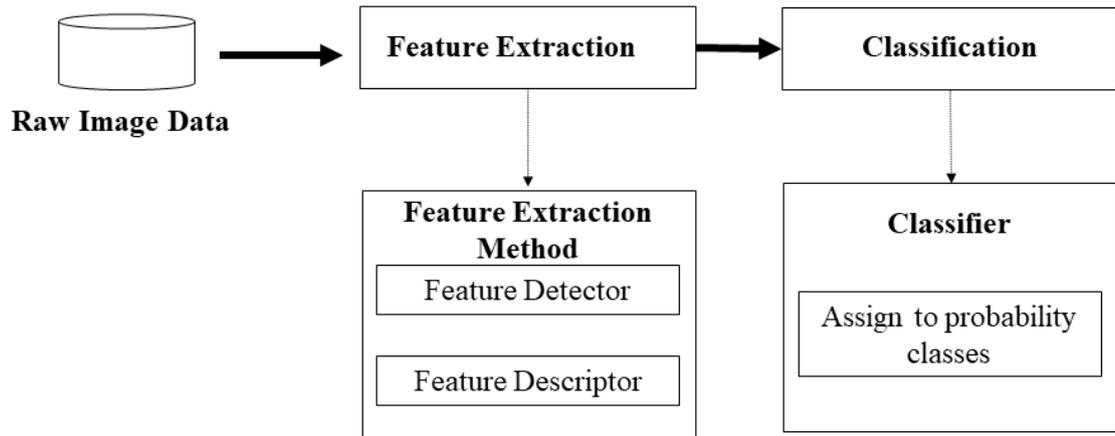


Figure 2.2: Image recognition workflow.

This research focuses on the effective feature extraction method and suitable feature selection in classifying between defective and non-defective apple images including low-quality region images and recognizing the types of defects. The image recognition method for classification can be categorized into statistical, morphology and spectral, model-based and deep learning approaches as depicted in Figure 2.3. Each category and related classification that suits this research are discussed in the following subsections.

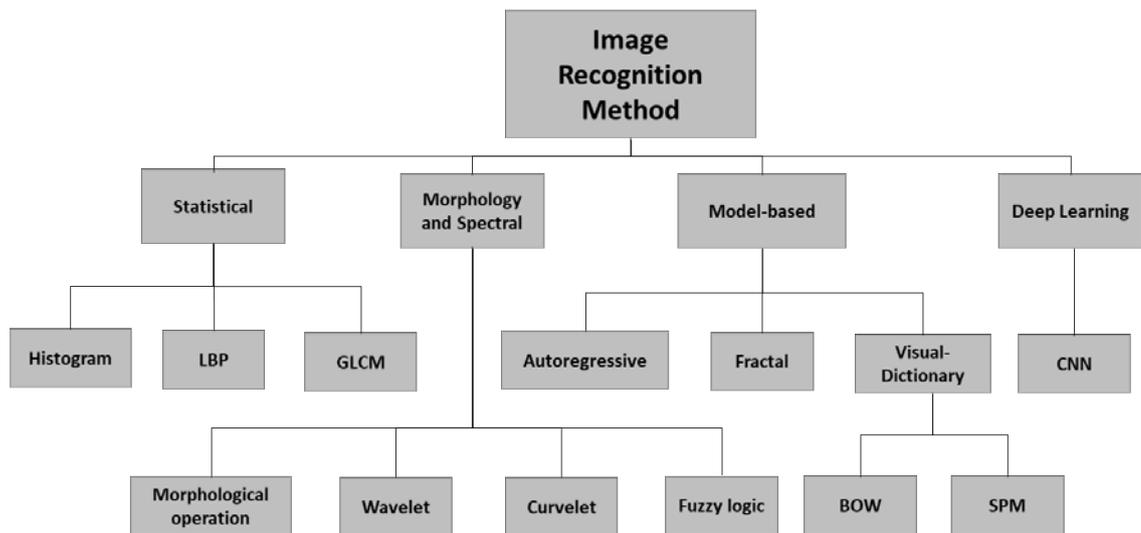


Figure 2.3: Categories of image recognition method for classification.

2.3.1 Statistical

The statistical approach image recognition method is widely adopted in many classification areas especially in the defect detection and classification of the agricultural product (L. Li et al., 2018). The statistical method is preferred because its capabilities to measure the spatial distribution of the pixel values in the image. Various statistical approaches were proposed to extract features from the image ranging from first order statistics to higher order statistics such as Histogram, Local Binary Patterns (LBP) and GLCM.

Histogram method is a simple and low-level statistical based image recognition method that invariant to translation, rotation, and insensitive to the exact spatial distribution of the color pixels. The histogram method is widely used to capture the color features of the image in HIS (hue, saturation, intensity) color space. The histogram represents the distribution of the color in a region of an image by counting the pixels with a given set of color values.

Kazmi et al. utilized Histogram with Linear Discriminant Analysis (LDA) classifier for weed detection and classification in sugar beet fields (Kazmi, Garcia-Ruiz, Nielsen, Rasmussen, & Andersen, 2015). The work addresses the weed detection in sugar beet field under natural, outdoor conditions using color features. The color features are defined based on the color space in the image (Singha & Hemachandran, 2012). Kazmi et al. reported high classification accuracy for their work in between 94.05% to 97.83%. However, this work is limited in the consideration of only color features. Solely dependent on color features can be challenging to classify between defective and non-defective apple as the apple varies depending on its cultivar and maturity level.

Therefore, B. Zhang et al. (B. Zhang et al., 2015) proposed automatic lightness correction with GLCM, RGB (red, green, blue) and HIS methods to extract color, texture and other statistical features from Fuji apple. The extracted features are classified as defective and non-defective Fuji apple using weighted Relevance Vector Machine (RVM) classifier. This work is intended to improve the detection of the defect region in Fuji apple. For selecting relevant features, I-RELIEF algorithm is used. However, the I-RELIEF algorithm introduces blind selection problem and increases the complexity in the method.

Mokhtar et al. (Mokhtar et al., 2015) introduced the defective and non-defective classification of tomato leaves based on the texture features. The texture features are properties representing the surface and structure of an image (ping Tian, 2013). Mokhtar et al. used the GLCM method to extract texture features and the SVM classifier with different kernel functions for classification. In the GLCM method, the texture features are extracted from a co-occurrence matrix based on the selection of the GLCM features to be observed depending on the texture data encountered (Fahrurozi et al., 2016).

The combination of Local Binary Pattern (LBP) with the GLCM method to extract features is presented in (Chowdhury, Verma, & Stockwell, 2015). The LBP (Ojala, Pietikäinen, & Harwood, 1996) is a texture descriptor where the texture in the image is described using histogram of label obtained from thresholding the neighborhood pixels with the centre pixel (Al-Hammadi, Muhammad, Hussain, & Bebis, 2013; Silva, Bouwmans, & Frélicot, 2015). Although the LBP operator was initially intended for texture descriptor, various modifications have been proposed to extend its applicability in various recognition fields (Papakostas, Koulouriotis, Karakasis, & Tourassis, 2013). However, the capability of the LBP operator is limited in capturing dominant features of large scale structures (Abdulrahman, Gwadabe, Abdu, & Eleyan, 2014; Ojala, Pietikäinen, & Mäenpää, 2002).

Though there are many LBP and its variants (Silva et al., 2015) such as Moment-Based Local Binary Patterns, Classic LBP, Elliptical Local Binary Pattern (ELBP), Uniform ELBP, Local Directional Pattern (LDP), Mean-ELP (M-ELBP) and others have shown the potential in many applications, the LBP method captures the pattern in the image based on the circular pattern of the neighborhood. This can limit the performance of the LBP and its variants to detect the low-quality region on the apple skin for defective and non-defective apple classification. To detect the low-quality region on apple images, a more specific spatial directional method is required to extract texture at different directions and orientation covering the low-quality region on the apple skin.

The GLCM is an effective method to extract texture information (Mondal, Kole, & Roy, 2017; Sthevanie & Ramadhani, 2018) where it can describe the relationship of the neighboring pixels in the image. Thus, Kurtulmuş et al. (Kurtulmuş & Ünal, 2015) incorporated GLCM with Gray Level Run Length Matrix (GLRLM) on the LBP method to improve its capabilities in capturing the color texture features. The GLRLM was

proposed by Galloway (Galloway, 1974) to describe texture features. To extract color texture features, the images are converted to HSI and luminance-chrominance in blue-chrominance in red (YCbCr) color models. The YCbCr space is chosen in their work because of its coordinate system that sensitive to the depth value (Oni, Ojo, Alabi, Adebayo, & Amoran, 2018). Then, sets of feature models are optimized with Principal Component Analysis (PCA), Stepwise Discriminant Analysis (SDA) and Recursive Feature Elimination (RFE) feature selections. However, the PCA can completely overlap the data from different classes (Cong & Duan, 2015), consequently, reduces the classification accuracy. Similarly, Capizzi et al. (Capizzi, LO SCIUTO, Napoli, Tramontana, & WOŹNIAK, 2016) also includes the color features in addition to the texture features to improve the detection of the orange dataset. The color information of the images is represented by hue, saturation and value (HSV) space whereas texture by the GLCM. These features are classified using Radial Basis Probabilistic Neural Network (RBPNN) classifier. Despite promising results in the classification of the defective oranges, the method can be ineffective in the presence of low-quality image region.

Moallem et al. (Moallem, Serajoddin, & Pourghassem, 2017) proposed statistical, textural and geometric features for golden delicious apple classification using the SVM classifier. The GLCM method is used to extract textural features, which are contrast, correlation, energy, homogeneity and entropy. The method achieved convincing accuracy (89.20%–92.50%) for grading the golden delicious apple but the accuracy decreases when the defective region is close to the stem ends area.

The selection of the textural features is further extended in (Olaniyi, Adekunle, Odekuoye, & Khashman, 2017). Olaniyi et al. proposed a Texture analysis method based on eight features from first order statistic and second order statistic, which is, the GLCM. The first order features are mean, variance and standard deviation, whereas the second

order are contrast, correlation, energy, homogeneity and entropy. The method was tested using three classifiers of Radial Basis Function (RBF) model, SVM and Backpropagation Neural Network. The method achieved the classification accuracy of 96.25% up to 100% and improved the performance in (Moallem et al., 2017) by utilizing the first order statistic features. However, utilizing the texture based method alone limits the capability of the method to distinguish between different object with similar texture representation and images with low-quality region (Fahrurozi et al., 2016; Y. Li, Wang, Tian, & Ding, 2015).

Recently, W.Li et al. (W. Li et al., 2019) proposed Contrast-Limited Adaptive Histogram Equalization+GLCM+Extreme Learning Machine (CLAHE+GLCM+ELM) method using Contrast-Limited Adaptive Histogram Equalization (CLAHE) and GLCM with Extreme Learning Machine (ELM) classifier to address the limitations of the GLCM. The CLAHE is used to depress the noises and improve the local contrast while the ELM classifier is used to reduce the time complexity. However, their method unable to perform well in terms of sensitivity, specificity and accuracy (W. Li et al., 2019).

A summary of the highlighted studies in this section with their classification area, feature extraction method, feature exploited, classifier and success rate are listed in Table 2.1.

Table 2.1: Highlighted studies on machine vision classification using statistical approach.

Author (s)	Classification Area	Feature Extraction Method	Features Exploited	Classifier	Success Rate
(Kazmi et al., 2015)	Sugar beet weed classification	Histogram	Color	LDA	The classification accuracy achieved between 94.05% to 97.83%
(B. Zhang et al., 2015)	Defective and non-defective Fuji apple	RGB, HIS and GLCM	Color, statistical and texture	RVM	The overall classification accuracy achieved is 95.63%
(Mokhtar et al., 2015)	Defective and non-defective classification of tomato leaves	GLCM	Texture	SVM	The highest classification accuracy achieved is up to 99.83%
(Chowdhury et al., 2015)	Classification of Vegetation	GLCM and LBP	Texture	SVM, Feed Forward Back-Propagation Neural Network (FF-BPNN) and K-Nearest Neighbor (KNN)	The classification accuracy achieved in between 70.00% to 91.82%
(Kurtulmuş & Ünal, 2015)	Classification of seed	HSI and YCbCr GLCM, GLRLM and LBP	Color and texture	KNN, Stochastic Gradient Descent (SGD) and SVM	The best classification accuracy scores of 99.24% was achieved for the SVM classifier

(Capizzi et al., 2016)	Defective and non-defective orange	HSV color histogram and GLCM	Color and texture	Radial Basis Probabilistic Neural Network (RBPNN)	The classification accuracy achieved is up to 88.00%
(Moallem et al., 2017)	Defective and non-defective golden delicious apple	GLCM	Statistical, texture and geometric	SVM, Multi-Layer Perceptron (MLP), Neural Network and KNN	The highest classification accuracy achieved by SVM in between 89.20% to 92.50%
(Olaniyi et al., 2017)	Defective and non-defective banana classification	Texture analysis GLCM	Texture	Radial Basis Function (RBFN), SVM and Backpropagation Neural Network	The classification accuracy achieved in between 96.25% to 100%
(W. Li et al., 2019)	Healthy and non-healthy gum	CLAHE and GLCM	Texture	Extreme Learning Machine (ELM)	The classification accuracy achieved is 74.00%

2.3.2 Morphology and Spectral

Morphology approach is the mathematical morphology feature extraction method based on preliminary geometry information of the object (Solomon & Breckon, 2011). A morphological approach examines the image set using a small cluster known as configuration element. It used the basic mathematical morphology operations of expansion, erosion, opening and closing. Spectral approach removes the basic structure in the images and then generalizes the basic structure with spatial layout rules (Hanbay, Talu, & Özgüven, 2016). In the defective and non-defective apple classification, the spatial domain information is important to identify the location of the defect. This will help to distinguish between defective and non-defective region. The studies conducted on the defective classification that based on morphology and spectral approaches are mainly used Morphological operation, Wavelet, Curvelet and Fuzzy logic.

Chung et al. (Chung et al., 2016) proposed HSV and Morphological operator with SVM classifier for the classification of defective and non-defective rice seedlings. First, thresholding in HSV color space component is implemented to remove the background from the rice seedlings. Subsequently, the Morphological operations are applied to the identified seedlings based on the shape features. However, the method relies on the preprocessing step of the boundary detection and segmentation to remove the background from the rice seedlings and capture the shape of the seeding.

To improve this limitation, Mondal et al. (Mondal et al., 2017) proposed a Morphological operation with Pearson Correlation Coefficient method. The work finds dominating texture feature for defective and non-defective classification including the severity of the defect. The dominating texture feature is selected according to the specified criterion of the correlation, in which high, moderate, low and very low. Thus, the effectiveness of the method depends on the threshold value. Ganganagowder et al.

(Ganganagowder & Kamath, 2017) improved the detection through morphological features of size and shape with color and texture features using the combination of CIEL*a*b*, Morphological, GLCM and Correlation-based Feature Selection (CFS) method. In their work, 35 features of morphological, color and texture are extracted for the classification. However, the large number of features tends to overfitting, thus, reduce the classification model interpretability (Dougherty, Hua, & Sima, 2009; Guyon & Elisseeff, 2003; Kurtulmuş & Ünal, 2015; Saeys, Inza, & Larrañaga, 2007).

Therefore, a smaller number of features is considered in (J. Zhang, Wang, Yuan, Chen, & Wu, 2017). The work extracts six wavelet features using continuous Wavelet analysis for defective and non-defective classification. Similarly, Zolfaghari et al. (Zolfaghari, Noor, Rezazadeh Mehrjou, Marhaban, & Mariun, 2018) proposed a combination of five most significant features extracted using Wavelet transform for defective and non-defective classification including the defective severity. The Wavelet transform projects an image into spaces in multiple scales (resolutions) which enhance the characteristics and composition for efficient representation. The Wavelet is a spectral approach that capable to provide spatial resolution and frequency (P. Patel & Bhandari, 2019). The information in spatial and frequency domains are important for defect detection (Hanbay et al., 2016). The frequency domain information can identifies the presence of the defect while the spatial domain information can identifies its location. The Wavelet transform ensures a good frequency resolution at lower frequencies and good spatial resolution at higher frequencies (P. Patel & Bhandari, 2019; Wu, Liu, & Jiang, 2014). Using Wavelet in their work, the spatially localized details of the defect region can be obtained. The defect region can be described as the discontinuities in intensity from one pixel to another. The difference in the intensity of the pixels forms edges describing boundaries or outline of the defect region on the surface object (Jain & Kaur, 2017). The Wavelet transform

can isolate a singularity point i.e., the discontinuity well across an edge pixel (Ganesan & Sajiv, 2017; Shukla & Changlani, 2013). However, its limitation lies in the curved region.

The limitation of the Wavelet in the curved region is addressed in the Curvelet transform (Candes, Demanet, Donoho, & Ying, 2006). The Curvelet is capable of capturing directional edges of curves, corners and profiles (Agarwal & Bedi, 2015; J. Luo, Song, Xiu, Geng, & Dong, 2014). Furthermore, the Curvelet offers a high dimension of Wavelet transform for richer information in both spatial and spectral domains (Hagargi & Shubhangi, 2018; Tunio, Memon, Khuhawar, & Abro, 2019). In comparison to other transforms, the Curvelet is effective and accurate at capturing edges and other singularities along the curves (Acharya et al., 2016). Therefore, the Curvelet transform can generally be used as filtering, image enhancement and detection of edges in image recognition. The filtering is a process to reduce the quantity of the data (pixel) while keep the essential structural assets of an image by reducing the noise and filters out the futile information in the images. Noise is the corrupted random variations in intensity values of the image. The common types of the noise in the images are impulse noise, Gaussian noise, salt and pepper noise.

Riyadi et al. employed Discrete Curvelet Transform (DCT) (Riyadi, Azizah, Damarjati, & Hariadi, 2018) as image enhancement to improve the quality and clarity of the images. Then, the statistical features of mean, standard deviation, entropy and energy are extracted for defective and non-defective mangosteen classification. Conversely, Tunio et al. (Tunio et al., 2019) used k -mean filtering with Curvelet transform to extract Curvelet features for defective classification. The Curvelet transform is a multivariable function that map into space spanned by curvelets in multiple scales and orientations. This gives an efficient representation of curved singularities and smoothest away from discontinuities across the curves. The Curvelet transform captures the curve segments in

an image of varying coefficients strength over all discrete orientations. The strong coefficients represent the meaningful edges while the weak coefficients describe the very blurred boundaries. Then, the prominent edges or curved elements are packed into a small number of coefficients in the Curvelet domain.

Other works as presented in (Mohammed & Alhamdani, 2019) and (Prabuwono et al., 2019) use other methods such as Gabor filter and Prewitt edge detection to capture edge points and edge lines for classification. Edge feature is among the important geometric primitive features for the classification task (Vijayakumar & Durai, 2017). There are variety of edge detection methods such as Sobel, Canny and Fuzzy logic. These methods respond differently to the same gray levels image due to the image's information complexity (Anas, 2016). Therefore, require the researchers to utilize different method to reach the performance needed.

In image classification, most researchers considered the edge detection method of Sobel and Canny in their work due to its accuracy and not the computation time (Vijayakumar & Durai, 2017). The Sobel method detects the edges by performing a 2-D spatial gradient quantity on an image to highlight the regions with high spatial frequency. However, it less capable in the spatial domain (Vijayakumar & Durai, 2017). The Canny method detects the edges based on adjustable parameters including the size of the Gaussian filter and thresholds. These parameters can affect the effectiveness and the computation time of the method (Jain & Kaur, 2017). The upper and lower threshold values in Canny method are set to extract strong edge pixels and weaker edge pixels, respectively. However, both Sobel and Canny methods have limitations in removing noise in the images (J. Song, Wang, & Li, 2019). Noise in the images is inevitable, thus, its important to remove the noise from the input image to achieve more relevant information in extracting the edges in the image.

A more accurate and powerful edge detection method that been used in many areas of digital image processing including image recognition is Fuzzy logic (Aborisade, 2010; Anas, 2016; C. Liu, Shirowzhan, Sepasgozar, & Kaboli, 2019; Vijayakumar & Durai, 2017). In contrast to Sobel and other edge detection methods, the Fuzzy logic method improves the quality of the edges (Aborisade, 2010; Kaur, Chawla, Khiva, & Ansari, 2018). The Fuzzy logic method is very good at handling edge detection problem decision making of partially true and partially false values between completely true and completely false values (J. Song et al., 2019). This method is suitable to determine the boundaries of a specific region for further image analysis (Aborisade, 2010). Generally, the Fuzzy logic method detects the edges in the image by dividing the gray level into three values. These values are the entropy maxima that describe the image and used to build the membership function in the Fuzzy logic. The membership function of the Fuzzy logic operates based on the membership degrees that assigned in the fuzzy rules according to the image information (Jain & Kaur, 2017; J. Song et al., 2019). This concept offers high flexibility to accurately discriminate between neighboring pixels at the edge lines or curves. However, the edges are defined as a sudden change in the gray level of the adjacent pixels. Thus, the performance of all edge detection methods may be affected by many factors related to image properties such as sensitive to noise, lighting, blurred images and dynamic background (Anas, 2016; Lakshmi & Sankaranarayanan, 2010; J. Song et al., 2019).

Despite various morphological and spectral approaches available for the classification task, these approaches require a high degree of periodicity of the defect structure to provide image geometry to obtain better results (Hanbay et al., 2016; J. Song et al., 2019). A summary of the highlighted studies in this section along with their classification area, feature extraction method, features exploited, classifier and success rate are listed in Table 2.2.

Table 2.2: Highlighted studies on machine vision classification using morphology and spectral approach.

Author (s)	Classification Area	Feature Extraction Method	Features Exploited	Classifier	Success Rate
(Chung et al., 2016)	Defective and non-defective classification of rice seedlings	HSV and Morphological operator	Color and shape	SVM	The overall classification accuracy achieved is 87.90%
(Mondal et al., 2017)	Defective classification on plant leaves	Morphological operation with Pearson Correlation Coefficient	Texture	Naive Bayes	The classification accuracy achieved in between 82.67% to 95.00%
(Ganganagowder & Kamath, 2017)	Food product classification	CIEL*a*b* color space, Morphological, GLCM and Correlation-based Feature Selection (CFS)	Color, size, shape and texture	Multilayer Perceptron (MLP), SVM, Random Forest (RF), Simple Logistic (SLOG) and Sequential Minimal Optimization (SMO)	The classification accuracy achieved in between 70.00% to 73.00%
(J. Zhang et al., 2017)	Defective and non-defective classification of crop	Wavelet	wavelet features	Fisher's Linear Discriminant Analysis (FLDA)	The classification accuracy achieved is 77.00%
(Zolfaghari et al., 2018)	Defective and non-defective rotor bar classification	Discrete Wavelet Transforms and statistical	statistical features	Feed-Forward Neural Network (FFNN)	The average classification accuracy achieved is 98.80%

(Riyadi et al., 2018)	Defective and non-defective mangosteen classification	DCT and statistical	Mean, standard deviation, entropy and energy	LDA	The classification accuracy achieved in between 83.75% to 92.50%
(Tunio et al., 2019)	Defective leaves classification	Curvelet	Curvelet features	SVM	The classification accuracy achieved is 93.50%
(Mohammed & Alhamdani, 2019)	Defective and non-defective fabric classification	Gabor filters	Statistical and geometry	Fuzzy Back Propagation Neural Network (FBPNN)	The classification accuracy achieved is 91.43%
(Prabuwono et al., 2019)	Classification of bottle caps	Mathematical Morphology and Prewitt	Edge	Fuzzy logic	The classification accuracy achieved in between 96.66% to 97.91%

2.3.3 Model-Based

Model-based approach is based on the construction of the image recognition method that adopts a combination of feature extraction methods to extract features. A combination of the feature extraction methods is constructed by considering the condition of the classification task. The recognition is then performed from the model built using the classifier (Z. X. Zhang, Tan, Huang, & Wang, 2012). The model-based approach can offer a powerful recognition in terms of detection, parameter estimation, interpretation, learning and generalization (Affonso, Sassi, & Barreiros, 2015). In defective classification, the most commonly used model-based approaches are Fractal model, Autoregressive model and Visual-dictionary based model.

The Fractals model (Mandelbrot, 1983) used the concept of geometric primitives that are self-similar and irregular in nature. A fractal is infinitely complex patterns created by repeating a process in an ongoing feedback loop. The different between fractals and geometric is the way they are scale. Fractals exhibit similar patterns at increasingly small scales which is called “self-similarity”. In (Al-Kadi, 2015), the Fractal model was proposed with wavelet decomposition in medical device technology for brain tumour classification. The fractal dimension is used for guiding the sub-band tree-structure decomposition to extract textural fractal characteristic for detecting the brain tumours. The texture measure is chosen in the work due to its scale invariance, capable to investigate self-similarity and its roughness surface estimation. These characteristics can be used to detect variance in structure orientation and the size of the cell for brain tumour classification. The limitation of (Al-Kadi, 2015) is the lacunarity of fractal features is ignored. In the Fractal model, the most important measurements are the fractal dimension and lacunarity. The fractal dimension is the measure of complexity and irregularity while lacunarity is the structural variation or inhomogeneity measures. Therefore, exclusively

depends on the fractal dimension limits the ability to quantify all the characteristic of a fractal object (Z. Ma, Zhou, Hepburn, & Cowan, 2016).

To address this limitation, Z. Ma et al. (Z. Ma et al., 2016) suggested the fractal dimension and lacunarity for defective types partial discharge classification within the rotating machine. The fractal dimension is used to describe the surface roughness of the polar defective pattern while the lacunarity is used to quantify the density of the defective pattern. In their work, three types of defective partial discharge geometries were built and tested to validate the methods. However, the Fractal model method unable to perform efficiently on several defective patterns because the similarity of the fractal dimension although the fractals are completely different (Karbauskaitė & Dzemyda, 2016).

Alternatively, Autoregressive (AR) model offers more stable performance for pattern variations. The AR model is invariant to rotation, low complexity and fast computation (Hanbay et al., 2016). The AR model used a linear equation system solution to express the degree of linear dependence between different pixels of an image. Han et al. (Han & Jiang, 2016) use the AR model with Variational Mode Decomposition (VMD) for defective and non-defective classification of rolling bearing. The VMD decomposed vibration signals and a series of stationary component signals. For each component mode, the AR model is established to extract the feature vector from the AR model parameter. The VMD is used because it handles the mixing mode effectively. However, the problem with mixing mode arises when the number of the decomposed modes is too small or too large that will cost computational time. Additionally, their method is limited for the linear model to detect the defect. The defect condition signals are not always in linear behaviour and certain defect characteristics, may contain nonlinear and non-stationary behaviours. These lead to difficulty to detect the defect effectively, thus reduces the classification accuracy.

Consequently, J.Ma et al. (J. Ma, Xu, Huang, & Huang, 2017) introduced a hybrid model based on the AR model to resolve the nonlinearity issue for defective bearing detection. The method combines the general expression for linear and nonlinear autoregressive (GNAR) model and a Generalized Autoregressive Conditional Heteroscedasticity (GARCH) model. The main advantage of this method is its capability in both linear and nonlinear signals. However, the AR model parameters are very sensitive to the change of the condition such as noise and lighting. Thus, can be challenging to correctly detect the defect.

To overcome this limitation, a visual-dictionary based model is used for defective and non-defective classification. The BOW is a well-known method in the visual-dictionary based model and widely used for various classification task (Chanti & Caplier, 2018). The BOW method describes the images by counting the occurrence frequency of each visual pattern (visual words) and used it as a feature to train the classifier. The *keypoint* approach was introduced in the BOW method by Csurka et al. (Csurka, Dance, Fan, Willamowski, & Bray, 2004) to extract the visual patterns (visual words) from the images.

The widely used *keypoint* methods are Harris corner detection (Harris & Stephens, 1988), Scale Invariant Feature Transform (SIFT) (Lowe, 2004), Speeded up Robust Features (SURF) (Bay, Ess, Tuytelaars, & Van Gool, 2008; Bay, Tuytelaars, & Van Gool, 2006) and Features from Accelerated Segment Test (FAST) (Rosten, Porter, & Drummond, 2010). The Harris corner detection extracts keypoints on corners and edges based on local auto-correlation function. Its performance is consistent in natural images, robust in matching, good stability and repeatability (M. Idris, Arof, Tamil, Noor, & Razak, 2009; Warif et al., 2016). However, it sensitive to the scale changes.

SIFT detector and descriptor (Lowe, 2004) provide unique features with high repeatability and accuracy. This makes SIFT robust to affine distortion, translation,

illumination changes, scale changes and rotation changes (Agaian et al., 2016; Lowe, 2004). However, SIFT descriptor is computationally high cost (Lee, Cho, & Park, 2015). Therefore, SURF (Bay et al., 2008) was proposed to reduce the computational cost of computing gradients for encoding descriptors. SURF is faster than SIFT without degrading the quality and more robust to noise (M. Y. I. Idris et al., 2019; Panchal, Panchal, & Shah, 2013). To further reduce the computational time of the earlier methods, Rosten et al. (Rosten et al., 2010) introduced FAST in which, faster than both SURF and SIFT but sensitive to the scale changes (Loncomilla, Ruiz-del-Solar, & Martínez, 2016).

Among the local descriptor methods, SIFT is the most robust with the best performance in the presence of the geometrical changes (Agaian et al., 2016; Mikolajczyk & Schmid, 2005; Sachdeva et al., 2017). SIFT descriptor combines the interest region detector of Difference of Gaussians (DoG) and a corresponding feature descriptor. SIFT extracts local extrema of DoG interest point from the scale space as keypoints. Then, the local neighboring information around the keypoints or patches are utilized for computing the descriptors. SIFT and other *keypoint* methods demonstrate high performance in image detection but unable to define a shape or images due to non-uniform distribution (Warif et al., 2016). The main limitation of the *keypoint* method is the patches of the keypoints can have a similar descriptors for completely different contexts in the images. A similar keypoint patches can also be represented by different descriptors due to noise or distortion (Sachdeva et al., 2017). This is because the *keypoint* method was initially designed for matching between two images (Fan et al., 2019).

The matching between two images using keypoints includes three stages, namely detecting the keypoint, calculating the descriptor and matching the keypoint with their associated descriptor. Examples of the matching between apple images are illustrated in Figure 2.4. The examples show the keypoints matching between the same apple images

can result in a successful matching of 100% as depicted in Figure 2.4 (a) but failed to match any of the keypoints when two different apple images are involved as depicted in Figure 2.4 (b). The issue of keypoints matching between different images of the same object requires a complementary technique to encode and pool the neighborhood information of the keypoint features for the classification task.

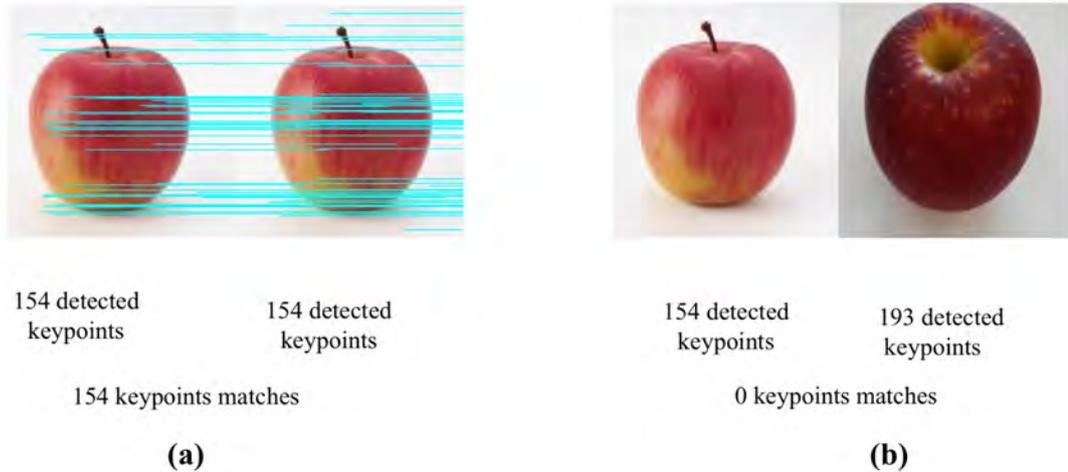


Figure 2.4: Examples the SIFT keypoint matching of (a) same apple images (b) different apple images.

The mentioned limitation of the *keypoint* method can be addressed with BOW method which is easy to implement, robust to object occlusion, image clutter, non-rigid deformation and also robust to viewpoint changes (Y. Liu, Zhang, Zhang, & Liu, 2016; Kejriwal, Kumar, & Shibata, 2016; H.-L. Luo, Wei, & Lai, 2011). The implementation of the BOW involves extracting the visual pattern (visual words) patches from the images using *keypoint* method. The patches are coded in the detected point to compute the descriptor and extract the features from an image. Then, these huge features extracted are clustered and build frequency histogram. The images are represented by counting the number of occurrences patches of each visual words in the images and used it as a feature to train the classifier as illustrated in Figure 2.5.

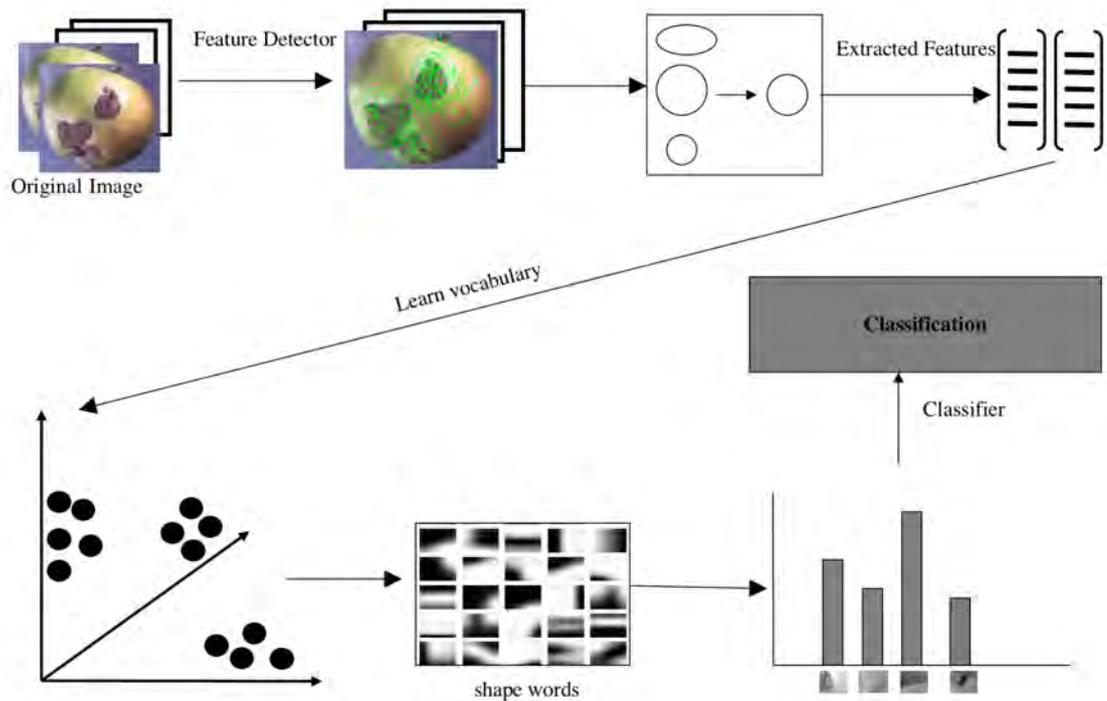


Figure 2.5: Architecture of BOW for image classification.

Nasirahmadi et al. (Nasirahmadi & Ashtiani, 2017) used five *keypoint* detectors along with SIFT descriptor in BOW method which also known as Bag-of-features (BOF) to find the best detector for defective and non-defective almond classification. The method achieved convincing classification accuracy (79.00% – 91.00%) for the combination of the Harris-Laplace detector with SIFT descriptor and SVM classifier. However, the BOW method disregards the spatial layout information in the visual words that lead to the missing spatial arrangement features on the image composition (Kejriwal et al., 2016; Lin et al., 2016). Consequently, reduce the performance of the BOW method.

Therefore, the spatial layout information is included in the SPM method (Lazebnik, Schmid, & Ponce, 2006) to improve image representation for image classification. The spatial layout information is important to discriminate between objects because different objects can have the same visual appearance but in different spatial arrangement

(Aldavert, Rusinol, Toledo, & Lladós, 2015). The SPM method works by partitioning the images into sub-regions on dense regular grids and compute the histogram of local features found in each grid. This method recorded a significant performance improvement over the BOW method but generates a large number of unnecessary and redundant high dimensionality features (Chanti & Caplier, 2018; Q. Li et al., 2017; Lin et al., 2016; Penatti et al., 2014; Yan et al., 2012). These irrelevant features can reduce the stability and performance of the method.

To eliminate the redundancies and select the representative keypoints, Lin et al. (Lin et al., 2016) and Q. Li et al. (Q. Li et al., 2017) proposed a keypoint selection technique to resolve this issue. Similarly, Xie et al. (Xie et al., 2018) proposed a new spatial partitioning scheme to avoid feature redundancy by modifying the pyramid matching kernel. The method proposed new pyramid matching kernel and spatial partition scheme instead of calculating the distribution of local visual descriptor histogram on each region. However, this modification of the spatial pyramid requires computational complexity to calculate the mid-level features.

There are various studies attempted to embed the spatial orders but the Spatial Pyramid Matching representation (Lazebnik et al., 2006) is the most popular and effective in encoding the spatial distribution for the classification task. The capability of the SPM method was successfully demonstrated in both scene and object recognition (B.-D. Liu et al., 2019). Recently, X. Wei et al. (X. Wei et al., 2019) investigated the performance of the BOW and SPM methods for defective classification of railway fastener. In their work, the BOW method is applied to the fastener classification by constructing dictionary construction of bag of words of the fastener. The experiment is also carried out for the SPM method that includes the spatial location information of the fastener image using Spatial Pyramid Matching representation (Lazebnik et al., 2006) for defective

classification of fastener. In the SPM method, the Dense-SIFT (Fei-Fei & Perona, 2005) feature extraction is used instead of SIFT because SIFT less sensitive to illumination changes and lacks global features. These SIFT issues can lead to misclassification and reduce the classification accuracy. In the study, the classification accuracy of the SPM method (99.26%) successfully improved the classification accuracy of the BOW method (96.30%). This proved that the spatial layout information using the Spatial Pyramid Matching representation in the SPM method (Lazebnik et al., 2006) can improve the detection of the defect. A summary of the highlighted studies in this section along with their classification area, feature extraction method, features exploited, classifier and success rate are listed in Table 2.3.

Table 2.3: Highlighted studies on machine vision classification using model-based approach.

Author (s)	Classification Area	Feature Extraction Method	Features Exploited	Classifier	Success Rate
(Al-Kadi, 2015)	Classification of brain tumour	Fractal and Wavelet	Fractal dimension	SVM, Naïve Bayes and KNN	The classification accuracy achieved by SVM classifier is 98.76%, Naïve Bayes 92.90% and KNN 79.70%
(Z. Ma et al., 2016)	Classification of defective partial discharge in rotating machine	Fractal	Fractal dimension and lacunarity	Centour Score Algorithms	The classification accuracy achieved for average centour score in between 47.75% to 60.25%
(Han & Jiang, 2016)	Defective and non-defective classification of rolling bearing	VMD and AR	Feature vectors from AR model parameters	Random Forest (RF) and SVM	The classification accuracy achieved in between 81.97% to 98.63%
(J. Ma et al., 2017)	Defective rolling bearing classification	Combination GNAR model and GARCH model	Statistical data	KNN	Mean classification accuracy achieved is 88.43% and the maximum accuracy achieved up to 93.33%

(Nasirahmadi & Ashtiani, 2017)	Defective and non-defective classification of almond	BOW	Visual-dictionary	KNN and SVM	Classification accuracy achieved for KNN classifier in between 67.00% to 79.00% and SVM in between 79.00% to 91.00%
(X. Wei et al., 2019)	Defective classification of railway fastener	BOW SPM	Visual-dictionary	SVM	Classification accuracy achieved for BOW is 96.30% and SPM 99.26%

2.3.4 Deep Learning

The recent development of deep learning approach has received considerable attention in computer vision. The deep learning approach is unique as it learns the important features automatically over training without the need for feature engineering through instructed algorithm (Kamilaris & Prenafeta-Boldú, 2018; H.-B. Zhang et al., 2019). The features or representation for the detection, distinguishing and classifying are automatically discovered by feeding the raw data or pixel values into the deep learning method. This approach is useful in learning and discovering complex structure data. The deep learning approach that commonly used for analyzing visual images is Convolutional Neural Network (CNN) (Kahraman, Karas, & Akay, 2018; Zhao et al., 2019).

In CNN method, the convolutional layers act as a feature extractor to extract important features from the input and the pooling layers reduce the dimensionality of the input image. Various convolutions are performed on the network at some layers, which create different representations of the learning dataset. Starting from the larger layers that consist general image representation, then become more specific at deeper layers. Multiple features from the lower-level into more discriminative features are encoded in the convolutional layers in the spatial-context aware. It may be understood as bank of filters that transform input images into another form which highlight the specific pattern. In many cases, the fully connected layers are placed near the model output which act as a classifier (Kamilaris & Prenafeta-Boldú, 2018). The structure of the CNN for apple images is given in Figure 2.6.

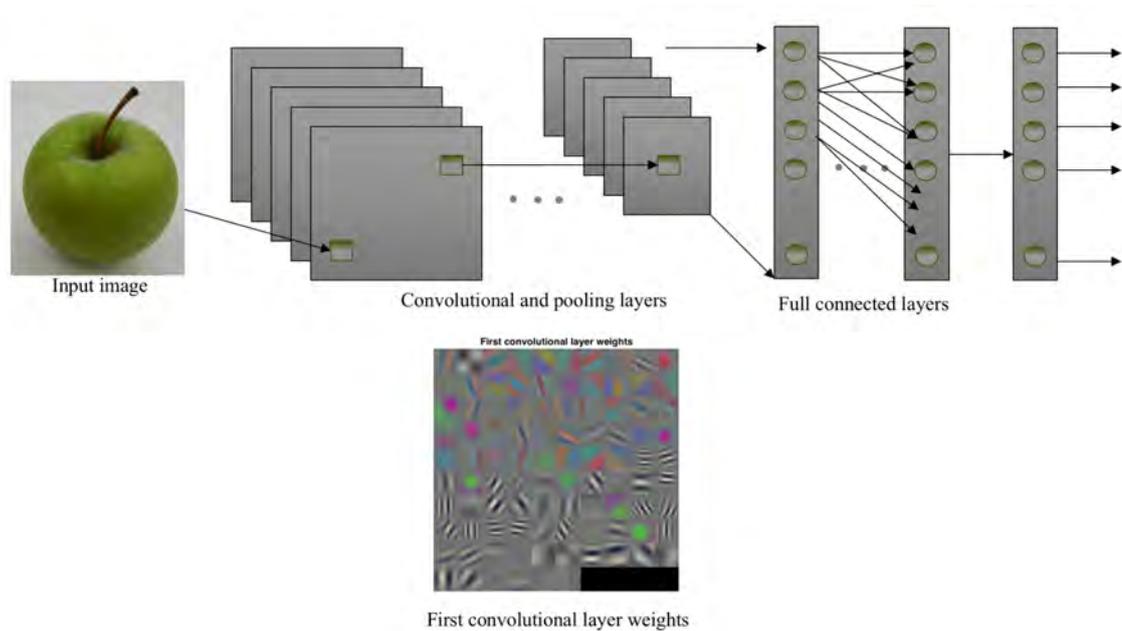


Figure 2.6: Structure of the CNN for apple image.

The CNN method has been implemented in various fields of the classification task including defective and non-defective classification as presented in Table 2.4. The high performance of the CNN method is consistently observed on different classifiers reported in (dos Santos Ferreira, Freitas, da Silva, Pistori, & Folhes, 2017) with accuracy ranged between 98.00% to 99.50%. However, there are several limitations in the structure of the CNN method. One of the major issues is the requirement for the fixed-size input image (He, Zhang, Ren, & Sun, 2015; Qayyum et al., 2017). If the images with arbitrary sizes are applied, the CNN method will resize the input images to a fixed size via cropping or warping the images (Donahue et al., 2014; Girshick, Donahue, Darrell, & Malik, 2014; He et al., 2015; Krizhevsky, Sutskever, & Hinton, 2012). To address this issue, He et al. (He et al., 2015) proposed a network structure (SPP-net) method that generates a fixed-length representation regardless of the image size or scale (X. Cheng, Zhang, Chen, Wu, & Yue, 2017; He et al., 2015; Krizhevsky et al., 2012).

Another issue in the CNN method is the difficulties to train the neural network when the network depth in the structure increases (X. Cheng et al., 2017; He, Zhang, Ren, & Sun, 2016). To improve training for a deeper network, a residual learning framework (He et al., 2016) was proposed that reformulates the network layers as learning residual functions. Other than that, the CNN features lose a lot of detailed information in the images due to multiple levels of abstraction through various convolutions in the method (Y. Zhang et al., 2019). This makes defective and non-defective apple classification task including the types of defects become more challenging. Apple images also consist of low-quality regions such as bright features or flecks features on the apple skin which can be difficult to detect with the lost information on CNN features. The detection on the low-quality region is important to differentiate between defective and non-defective apples. Failure to detect these features may reduce the classification accuracy.

Table 2.4: Image classification using CNN method of deep learning approach.

Author (s)	Classification Area	Classifier	Success Rate
(Sladojevic, Arsenovic, Anderla, Culibrk, & Stefanovic, 2016)	Defective and non-defective classification of leaf	Softmax	The overall classification accuracy achieved is 96.30%
(Pinto, Furukawa, Fukai, & Tamura, 2017)	Defective classification of coffee	Softmax	The classification accuracy achieved in between 67.50% to 98.75%
(Lu, Yi, Zeng, Liu, & Zhang, 2017)	Defective classification of rice crop	Softmax	The classification accuracy achieved is 95.48%
(dos Santos Ferreira et al., 2017)	Classification of weed in soybean crop	ConvNets, Adaboost, RF and SVM	The classification accuracy achieved in between 98.00% to 99.50%
(X. Cheng et al., 2017)	Classification of pests	SVM, BP Neural Network and Deep Convolutional Neural Network	The classification accuracy achieved for SVM is 44.00%, BP Neural Network 42.67% and Deep Convolutional Neural Network achieved in between 86.67% to 98.67%

(Z. Wang, Hu, & Zhai, 2018)	Defective classification of blueberry	ResNet, ResNeXt, Sequential Minimal Optimization (SMO), Linear Regression (LR), RF, Bagging and Multilayer Perceptron (MLP)	The classification accuracy achieved for ResNet is 88.44%, ResNeXt 87.84%, SMO 80.82%, LR 76.06%, RF 73.14%, Bagging 71.13% and MLP 78.27%
(Yan Zhang, Cui, Liu, & Yu, 2018)	Defective classification of tire	CNN Caffe framework	The classification accuracy achieved is 96.51%
(D. Li, Cong, & Guo, 2019)	Defective classification of sewer	Deep learning multi-class	The classification accuracy achieved is 64.80%

Other main issues of the deep learning approach is it computationally expensive. The deep learning approach consists of various parameters that required large dataset such as ImageNet dataset to automatically learn from these large number of training images to avoid over-fitting in order to achieve the desired result. However, there is limited dataset available publicly in the domain of the agriculture industry for researchers to work with. Moreover, the available datasets do not completely describe the targeted problem, thus required researchers to develop their own sets of images (Kamilaris & Prenafeta-Boldú, 2018; X. Song et al., 2016). The optimization issues might arise when using a small dataset for pre-trained models of deep learning approach that is significantly different in models' complexity and hardware restrictions (Chaturvedi, Ragusa, Gastaldo, Zunino, & Cambria, 2018; X. Cheng et al., 2017; He et al., 2015; Kamilaris & Prenafeta-Boldú, 2018; Krizhevsky et al., 2012; Y. Zhang et al., 2019). Thus, there is a need for effective and accurate image recognition method that works on a small dataset with limited computing power consumption especially for apple classification that consists of low-quality region on the apple skin.

2.4 Discussion

Researchers have proposed many image recognition methods for machine vision classification in industry, especially for defect detection and classification. From the highlighted studies in this chapter, the defective classification is mainly performed using GLCM, BOW, SPM or CNN image recognition methods. The GLCM performs the detection through texture features representing the surface and structure of the image while the BOW and SPM use visual-dictionary features for image classification. On the other hand, the CNN method use deep CNN features. Each method performs differently due to the features employed for the detection and the image's information complexity. The strength and limitations of these methods are summarized in Table 2.5.

Table 2.5: List of commonly used image recognition methods including their strengths and limitations.

Method	Strengths	Limitations
GLCM	<ul style="list-style-type: none"> - Effective to recognize the object with texture information - Extracting spatial relationships with different statistical computation - Produces features that describe the relationship of the neighboring pixels in the image 	<ul style="list-style-type: none"> - Dependent on the rotation and scaling - Have difficulty to distinguish different object with quite similar texture representation and images with low-quality region
BOW	<ul style="list-style-type: none"> - Easy to implement - Robust to several parameters such as occlusion, clutter, non-rigid deformation and viewpoint changes 	<ul style="list-style-type: none"> - High computational in vector quantization step - Disregards the spatial layout information
SPM	<ul style="list-style-type: none"> - Include spatial information to improve the image representation and better distinguish objects 	<ul style="list-style-type: none"> - Generates large numbers of unnecessary and redundant high dimensionality features
CNN	<ul style="list-style-type: none"> - Learns the important features automatically over training without the need for feature engineering through instructed algorithm (abstracted features are learned by stack convolutional and sampling layer) 	<ul style="list-style-type: none"> - Requirement for the fixed-size input image - Lose detailed information in the images due to multiple levels of abstraction through various convolutions - High computational complexity for model training

-
- Computationally expensive and consists of many parameters that required large dataset
-

The critical foundations in image recognition are the images used for the classification task (Cui, 2019). In this research, three dataset of apple images are used to test the proposed methods for defective and non-defective classification including types of defects. The challenges in apple classification due to physical and biological influences include the presence of the low-quality regions of bright or flecks features on the apple skin and the similarity of the appearance between defects and stems ends or calyxes which are the natural part of the apple. The detection of the low-quality region can be challenging due to low contrast and unclear boundary. On the other hand, the discrimination between the defect and the natural part of the apple requires the information on its location to precisely classify between the defective and non-defective region.

After considering the strength and applicability of the image recognition methods on apple images as presented in the previous sections, the GLCM and SPM methods are the most suitable methods for classifying defective and non-defective apple including the types of defects. The GLCM capable of analyzing and describing the spatial relationship of neighboring pixels on the surface structure of the image properties. While the SPM method includes the spatial layout information that is important to discriminate between defect with natural parts of the non-defective apple.

However, there are several limitations of the GLCM and SPM methods. The GLCM method is dependent on the images texture information which ineffective to extract features from the low-quality region images. On the other hand, the SPM method

generates a large number of unnecessary and redundant high dimensionality features. These irrelevant features can reduce the stability and performance of the method. To address these limitations on apple classification, the first proposed method aims to improve the ability of the GLCM method in detecting features on the low-quality region of the apple image. The second proposed method focuses on the selection of significant features to reduce unnecessary SPM features for classification. The binary-class classification of defective and non-defective apple images is also extended to the multi-class classification between types of defects. The details for each method are presented in the following chapters.

2.5 Summary

In this chapter, the overview of the image recognition methods on feature extraction and classification were presented. The strength and limitations of the image recognition methods were also analyzed in terms of detection and recognition. Finally, the challenges and potential improvement of the presented image recognition methods on apple classification were also discussed.

CHAPTER 3: RESEARCH METHODOLOGY

This chapter describes the details of the research methodology to address the research problems and achieve the outlined objectives. The details include plans, structure and strategy of investigation, data acquisition, existing methods and proposed methods. Lastly, the proposed methods are compared with the existing image recognition methods to discuss the final results. The links between chapter 4 and chapter 5 are also explained.

3.1 Introduction

The effectiveness of any image classification is strongly dependent on the features extracted from the images as highlighted in the reviewed literature of Chapter 2. The existing image recognition methods are mainly focused on general defective and non-defective image recognition problems. Thus, the detection and extraction of the features to recognize and classify the image with low-quality region become essential. Particularly for classification of non-defective and defective apple including the types of defects. Therefore, this research focuses on the detection of suitable features to increase the accuracy in small sample binary-class classification of defective and non-defective apple images including the low-quality region images and multi-class classification between types of defects. Figure 3.1 illustrates the research design or methodology summarizing all the processes to achieve the research objectives.

The first step in the research methodology is conducting a comprehensive literature review to identify and formulates the research problems and objectives. In the second step, the required apple datasets are created. In the next step, five existing image recognition methods, which are BOW (Csurka et al., 2004), SPM (Lazebnik et al., 2006), CNN (dos Santos Ferreira et al., 2017), GLCM Texture analysis (Olaniyi et al., 2017) and

CLAHE+GLCM+ELM (W. Li et al., 2019) are replicated to evaluate and compare the performance of each method on apple classification. These five methods are chosen due to their popularity and stability to represent the visual-dictionary based method, deep-learning and texture based method. To evaluate the performances of the image recognition methods on apple datasets, 10-fold cross-validation is implemented. Based on the performances and analysis of the existing image recognition methods, two new image recognition methods are proposed to improve the accuracy for binary-class classification of defective and non-defective apple including the low-quality region images and multi-class classification between types of defects. Finally, the performances of two proposed methods are evaluated using 10-fold cross-validation on apple datasets. Then, these results are compared with five existing image recognition methods.

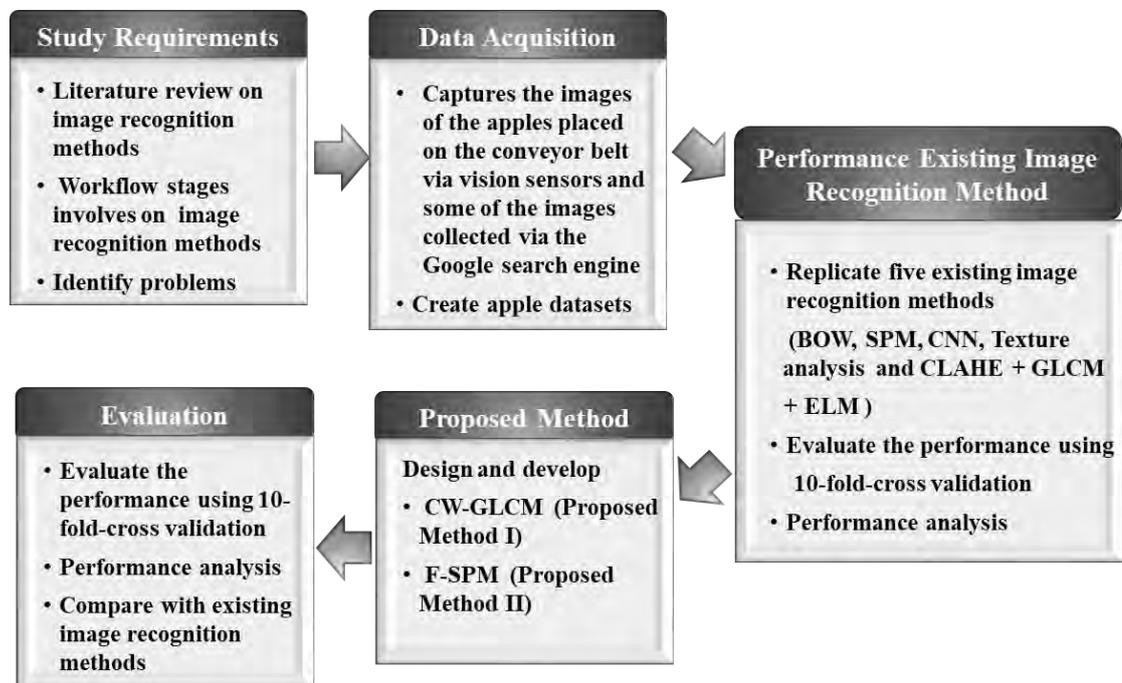


Figure 3.1: Research methodology and design.

3.2 Data Collection

This research focuses on the binary-class classification of defective and non-defective apple including images with low-quality region and multi-class classification between types of defects. The evaluation of the binary-class and multi-class classifications are performed on three new apple image datasets, namely NDDA, NDDAW and DA. These datasets are created due to the shortage of public agriculture image datasets on defective and low-quality region (Kamilaris & Prenafeta-Boldú, 2018; X. Song et al., 2016). The NDDA and NDDAW are both binary-class datasets containing defective and non-defective apple images, with NDDAW contains more images with low-quality region compared to the NDDA. Conversely, the DA dataset contains five types of defective apple images, which are Scab, Rot, Cork Spot, Blotch and Bruise. These categories are verified by the agriculture practitioners from Malaysian Agricultural Research and Development Institute (MARDI) and Pahang Agriculture Department.

The apple images in the datasets are collected using vision sensors and via the Google search engine with several keywords as listed in Table 3.1. The apple images from the Google search engine are included due to the difficulties in obtaining various types of defective apples. Majority of the apple images from the Google search engine are centered, cleaned and occupied most of the image without or very few cluttered environments. This is similar to the apple images captured via the vision sensor where the apples are placed on the conveyor belt as illustrated in Figure 3.2. To reduce the shadow and glare effects, the light sources are placed near to the vision sensor and uniformly distributed. The resolution of all the images captured using vision sensor are set at 900×700 pixels. However, the resolution of the images obtained via the Google search engine are varied from 205×220 pixels to 552×512 pixels. Therefore, the images obtained from the Google search engine are rescaled to 900×700 pixels following the resolution of the images captured via vision sensor. The image rescaling have no effect

on the classification performance but would effect the computational time. This is due to the image recognition methods used in this thesis (i.e. visual-dictionary, deep learning and texture based method) are invariance to scale changes (Arya, Singh, Kumar, & Mandoria, 2018; Chan et al., 2015; Csurka et al., 2004; Graham, 2014; Hashemi, 2019; W. Li, Dong, Xiao, & Zhou, 2016; Sachdeva et al., 2017).

The first dataset created in this research is the NDDA dataset, a binary-class dataset containing defective and non-defective apple images. The NDDA dataset is created to evaluate the capability of the image recognition method to detect defective and non-defective apples. Based on the performance analysis of the existing image recognition methods on NDDA dataset, this research creates another new dataset of binary-class defective and non-defective apple images called NDDAW. This dataset is used to further evaluate the effectiveness of the proposed method against low-quality region. The NDDAW dataset consists more images with low-quality region intended address the limitation of classifying low-quality region images.

The third dataset namely, DA dataset is a new multi-class dataset consisting five types of defective apple. This dataset aims to evaluate the multi-class classification performances of the image recognition methods. The construction diagram of the datasets is illustrated in Figure 3.3 and the details of each dataset are described in the following subsections.

Table 3.1: Keywords used for collected images via the Google search engine.

Keywords
fresh+apple
healthy+apple
apple+disease
damage+apple
defect+apple
low+grade+apple+fruit
Scab+apple
Rot+apple
Rotten+apple
Cork+Spot+apple
Blotch+apple
Bruise+apple

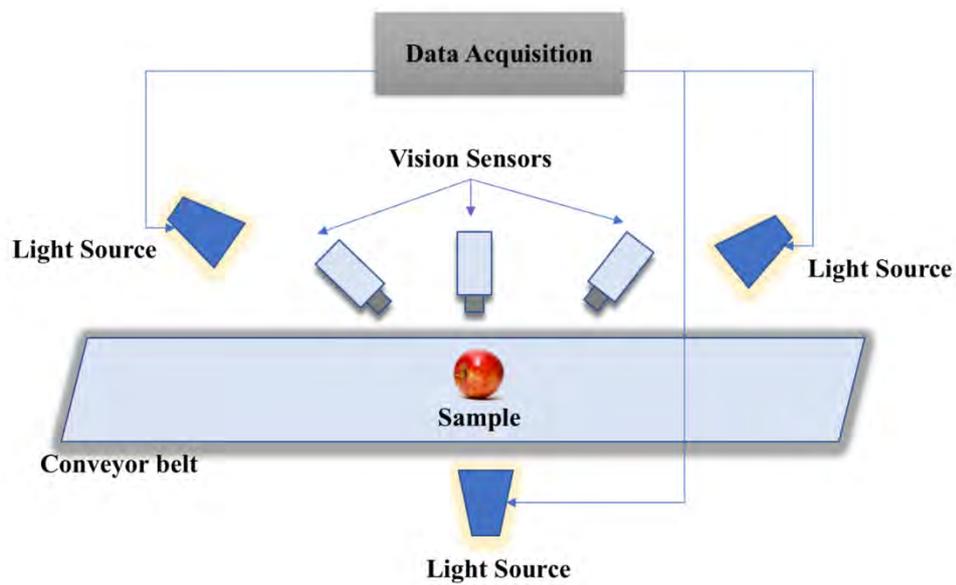


Figure 3.2: Illustration of the data acquisition process using vision sensors².

² Ismail, A., Idris, M. Y. I., Ayub, M. N., & Yee, L. (2019). Investigation of Fusion Features for Apple Classification in Smart Manufacturing. *Symmetry*, 11(10), 1194.

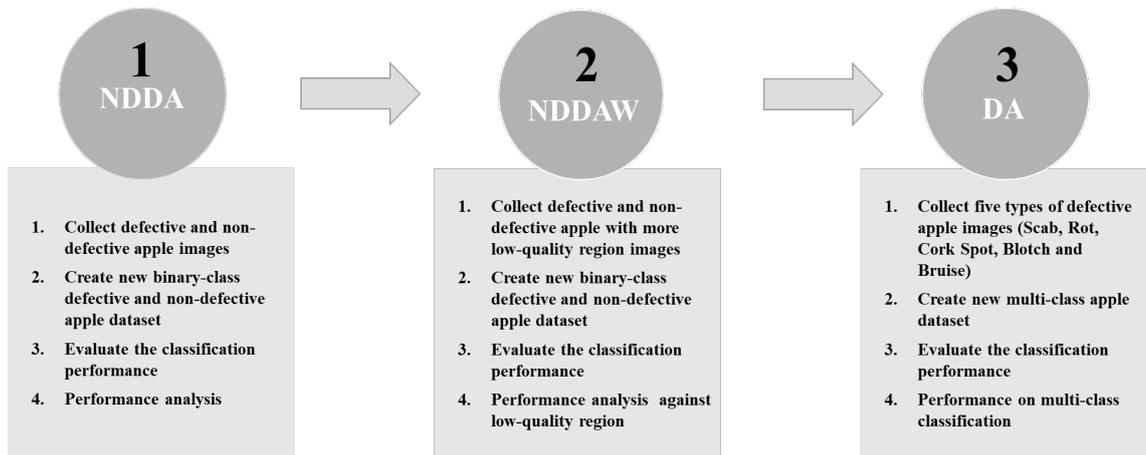


Figure 3.3: Datasets construction diagram.

3.2.1 Binary-Class Datasets

The binary-class datasets of the NDDA and NDDAW datasets contain 1110 apple images from defective and non-defective categories. The NDDA dataset is created to evaluate the capability to detect various defective and non-defective apple types, while the NDDAW is created particularly to include more apple images with low-quality region on its skin. The details of each dataset are described in the following subsection.

3.2.1.1 NDDA

This dataset consists of 550 apple images; 275 of the images are non-defective apples and 275 of the images are defective apples. Also, the natural parts of the apple are visible on 370 images for the stem end and 61 images for the calyx. Other than that, the low-quality region on the apple skin can be found on 76 images. The properties of this dataset are listed in Table 3.2. The apple images for the non-defective category were collected from five apple cultivars; Red Delicious, Gala, Fuji, Honeycrisp and Granny Smith. While the defective category was collected from five groups of defects which are Scab, Rot, Cork Spot, Blotch and Bruise. The defective category includes the external defects

on the apple skin with varying severity. These defects are visible to the naked eyes and found at different location, region and size. Examples of apple images in the dataset are shown in Figure 3.4.

Table 3.2: Details of the characteristics in NDDA dataset.

Non-Defective		Defective	
Cultivars	Total	Types	Total
Red Delicious	53	Scab	71
Gala	57	Rot	79
Fuji	58	Cork Spot	61
Honeycrisp	56	Blotch	32
Granny Smith	51	Bruise	32
Total	275	Total	275

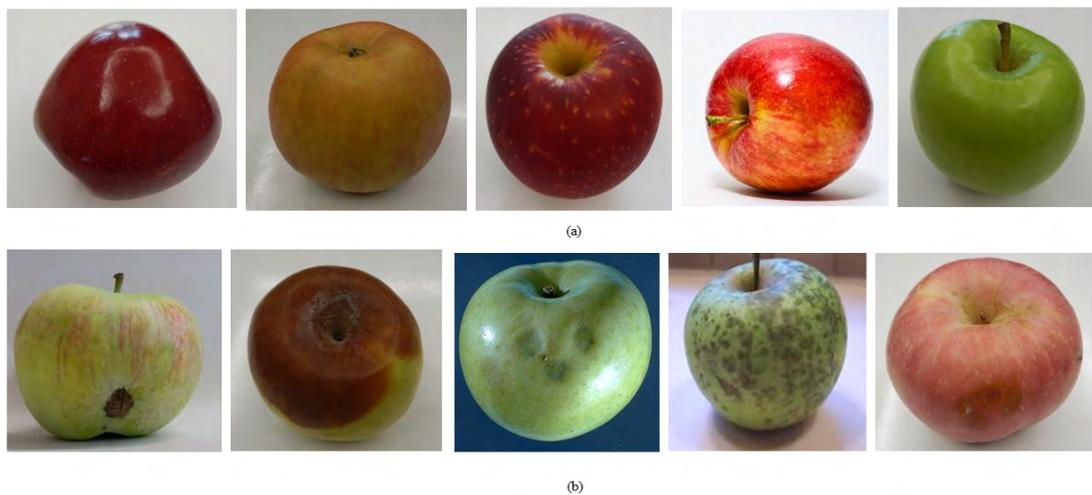


Figure 3.4: Examples of apple images in the NDDA dataset (a) non-defective (b) defective¹.

3.2.1.2 NDDAW

The NDDAW dataset contains a total of 560 apple images. These images are categorized into non-defective and defective categories wherein each category is composed of 280 images. The overall setup of this dataset follows the NDDA dataset. However, the main difference between the NDDAW and NDDA datasets is that the NDDAW dataset have more apple images with low-quality region on its skin (159

images) which intended to address the limitation of classifying low-quality image region. In the dataset, the stem end is visible on 248 apple images whereas calyx on 130 apple images. The properties and samples of the NDDAW dataset are shown in Table 3.3 and Figure 3.5, respectively.

Table 3.3: Details of the characteristics in NDDAW apple dataset.

Non-Defective		Defective	
Cultivars	Total	Types	Total
Red Delicious	44	Scab	42
Gala	16	Rot	55
Fuji	26	Cork Spot	27
Honeycrisp	59	Blotch	7
Granny Smith	135	Bruise	149
Total	280	Total	280

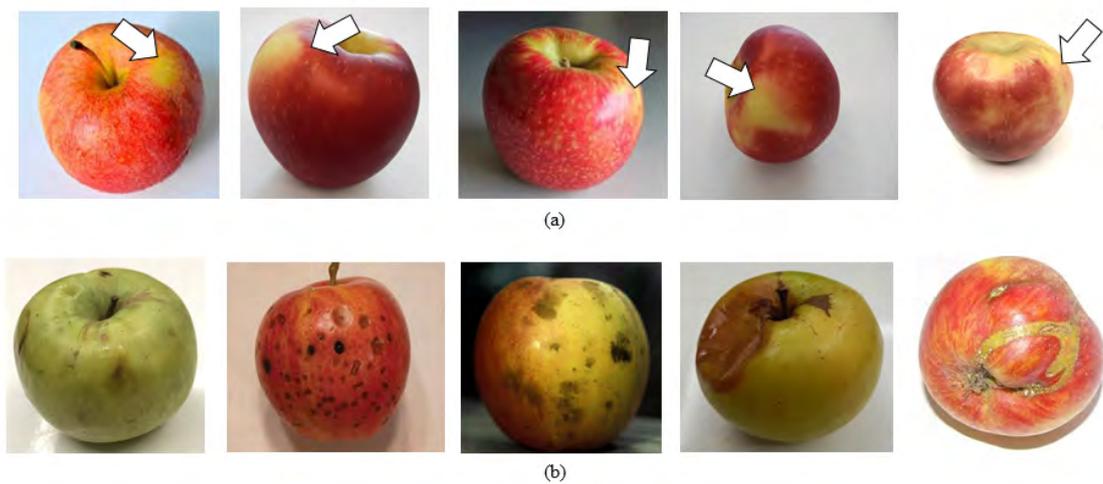


Figure 3.5: Examples of apple images in the NDDAW dataset (a) non-defective (b) defective. The low-quality regions on the apple skin are indicated by the arrows.

3.2.2 Multi-Class Dataset

The DA multi-class dataset is created to further evaluate the performance of the proposed methods for multi-class classification. This dataset contains five types of defective apples. The types of defective apples are Scab, Rot, Cork Spot, Blotch and Bruise. Each of the type consists of 40 images. The natural parts of the apple such as stem

end (75 images) and calyx (53 images) are also visible in this dataset. The defective category includes variations of obvious and unobvious defects with different types, severity, region and size. The properties and sample of DA dataset are shown in Table 3.4 and Figure 3.6, respectively.

Table 3.4: Details of the characteristics in DA dataset.

Defective	
Types	Total
Scab	40
Rot	40
Cork Spot	40
Blotch	40
Bruise	40
Total	200

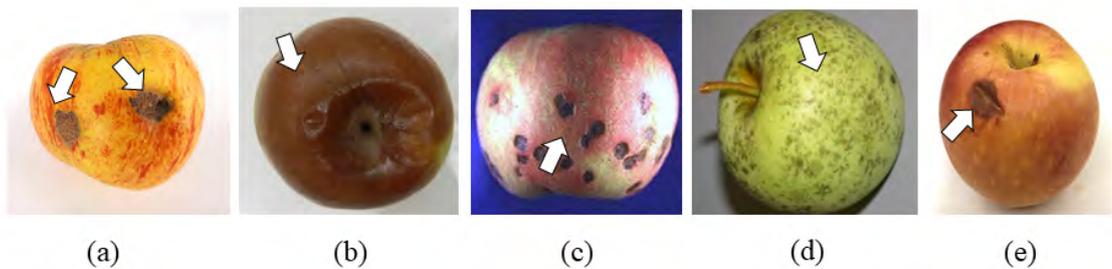


Figure 3.6: Examples of defective images in the DA dataset (a) Scab (b) Rot (c) Cork Spot (d) Blotch (e) Bruise.

3.3 Materials and Methods

This research replicates five existing image recognition methods for comparison purpose in the evaluation. The selected existing image recognition methods are BOW (Csurka et al., 2004), SPM (Lazebnik et al., 2006), CNN (dos Santos Ferreira et al., 2017), Texture analysis (Olaniyi et al., 2017) and CLAHE+GLCM+ELM (W. Li et al., 2019). These methods are chosen due to their popularity and stability in visual-dictionary based method, deep learning and texture based method. The existing methods are tested on the created NDDA, NDDAW and DA datasets. The overall scheme of the five replicated

existing methods are shown in Figure 3.7. Each method is briefly described in the following subsections.

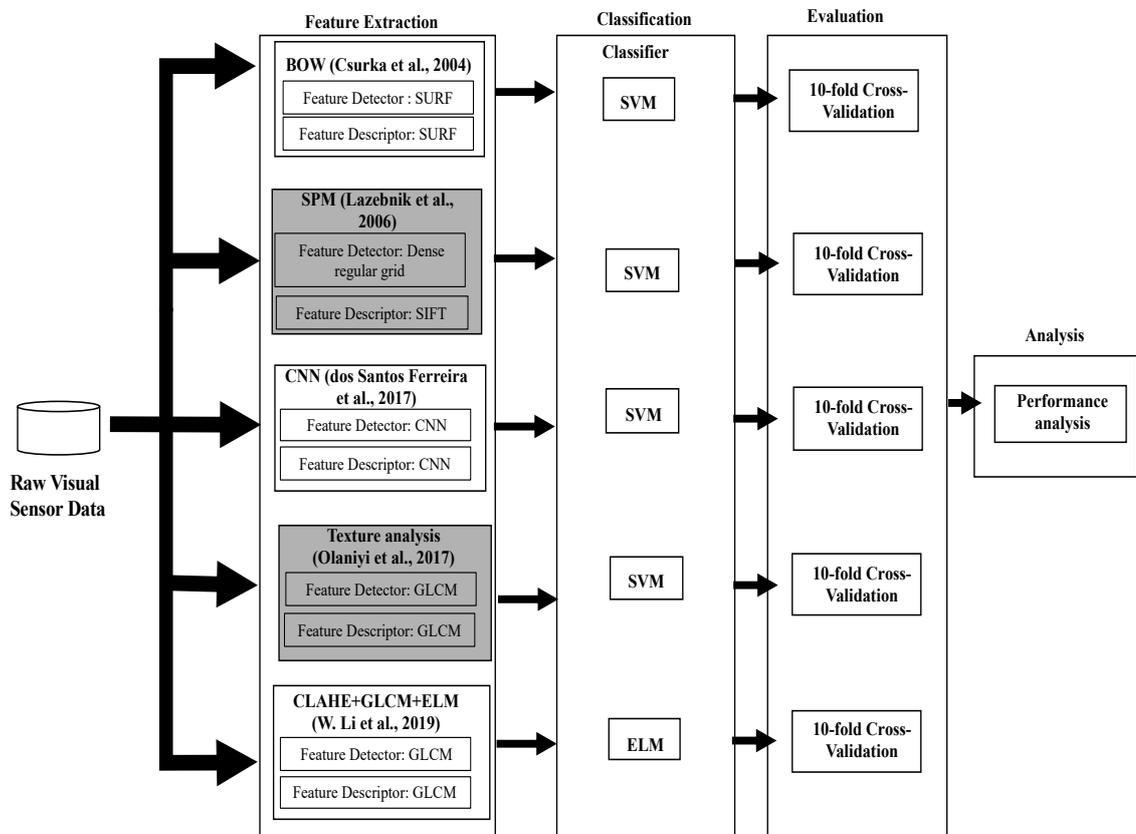


Figure 3.7: Overall scheme of five existing image recognition methods (BOW, SPM, CNN, Texture analysis and CLAHE+GLCM+ELM).

3.3.1 BOW

The BOW (Csurka et al., 2004) method use keypoint to extract visual pattern (visual words) patches from the images. In the BOW method, the images are described by visual patterns (visual words). The first step in the method is to detect keypoint features from the images. Then, the patches are coded in the detected point, compute the descriptor and extract the features from an image. Next, these huge extracted features are clustered into “similar looking” regions using *k*-means algorithm in the clustering step of the BOW

method. Once the descriptors are assigned to clusters forming feature vector, the classification is performed using a classifier to predict its probability classes.

3.3.2 SPM

The SPM (Lazebnik et al., 2006) method includes the spatial layout information to improve image representation in the BOW method for image classification. The SPM method partitions the image into increasingly fine sub-regions and computes the histograms of local features found inside each sub-region. In the feature extraction stage, the Dense regular grid with SIFT descriptor is used to extract feature vectors from all over the image on a regular grid. Each grid represents a patch of 16×16 pixels and overlaps by 8 pixels with its neighbors. The patch is further divided into 4×4 cells wherein the gradient orientation of 8-bins histogram is computed for each cell. Then, the features extracted from the entire cell grid are concatenated forming the feature vector. Next, the extracted feature vectors are clustered into a visual word vocabulary to create a codebook using the *k*-means clustering algorithm. The features are then fed to classifier for the classification task.

3.3.3 CNN

The CNN deep learning method is composed of a method that automatically discovers the features. The CNN (dos Santos Ferreira et al., 2017) method extracts the features by feeding the raw data of each image into the CNN deep learning method. Then, it automatically discovers the features. The CNN structure consists of input layer, convolutional and pooling layers, full connected layers, output layer and classification. The convolutional layers act as a feature extractor to extract important features from input

and the pooling layers reduce the dimensionality of the input image. Various convolutions are performed on the network layers, which create different representations of the learning dataset. Starting from the larger layer that consists of general image representation, then become more specific at the deeper layers. Multiple features from the lower-level into more discriminative features are encoded in the convolutional layers in spatial-context aware. It also acts as a bank of filters that transform input images into another form which highlights specific pattern in the images. Then, the extracted deep features are sent to classifiers for the classification task.

3.3.4 Texture Analysis

Texture analysis (Olaniyi et al., 2017) based on the GLCM method extracts textural features from the image. The gray comatrix function of GLCM is used to characterize the texture in the image by determining the number of a pixel of a specific gray level intensity value. This will create a GLCM matrix. Then, eight features are extracted from the gray level co-occurrence matrixes, namely contrast, correlation, energy, homogeneity, entropy, mean, variance and standard deviation. Next, the features are normalized and fed to the classifier for the classification task.

3.3.5 CLAHE+GLCM+ELM

The CLAHE+GLCM+ELM (W. Li et al., 2019) method is the combination of CLAHE, GLCM and ELM. The CLAHE is an image enhancement method which enhances the edges and improves the local contrast in the images. Then, the GLCM method extracts the texture features from the images. In the classification phase, the ELM

is used as a classifier to reduce the time complexity of the method since its a fast learning classifier.

3.4 Proposed Methods

Based on the limitations identified from the performance analysis of the existing image recognition methods for apple classification, two image recognition feature extraction methods called CW-GLCM and F-SPM are proposed. These proposed methods are aimed to increase the accuracy of binary-class classification of defective and non-defective apple images including images with low-quality region and multi-class classification between types of defects. Specifically, the CW-GLCM is proposed to improve the ability of the GLCM Texture analysis method in detecting features on the low-quality region of the apple image for binary-class classification while the F-SPM concentrates on improving the drawbacks of the proposed CW-GLCM method by including the spatial layout information of the SPM. The proposed F-SPM method also reduces the unnecessary SPM features through Fuzzy logic detection to include only significant features for further classification. The capability of the F-SPM is also extended to the multi-class classification. Each of the proposed method consists of two main phases. These phases are feature extraction and feature classification as shown in Figure 3.8 and Figure 3.9, which will be discussed in the following subsections.

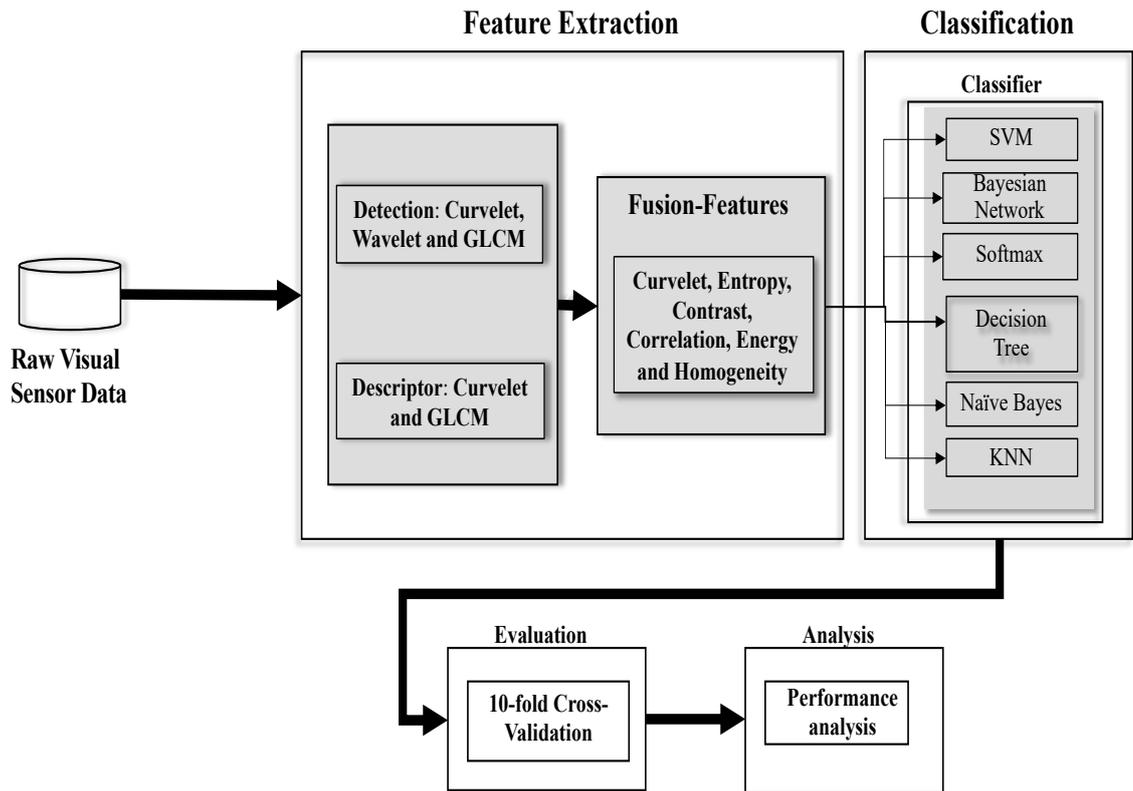


Figure 3.8: Proposed CW-GLCM image recognition method.

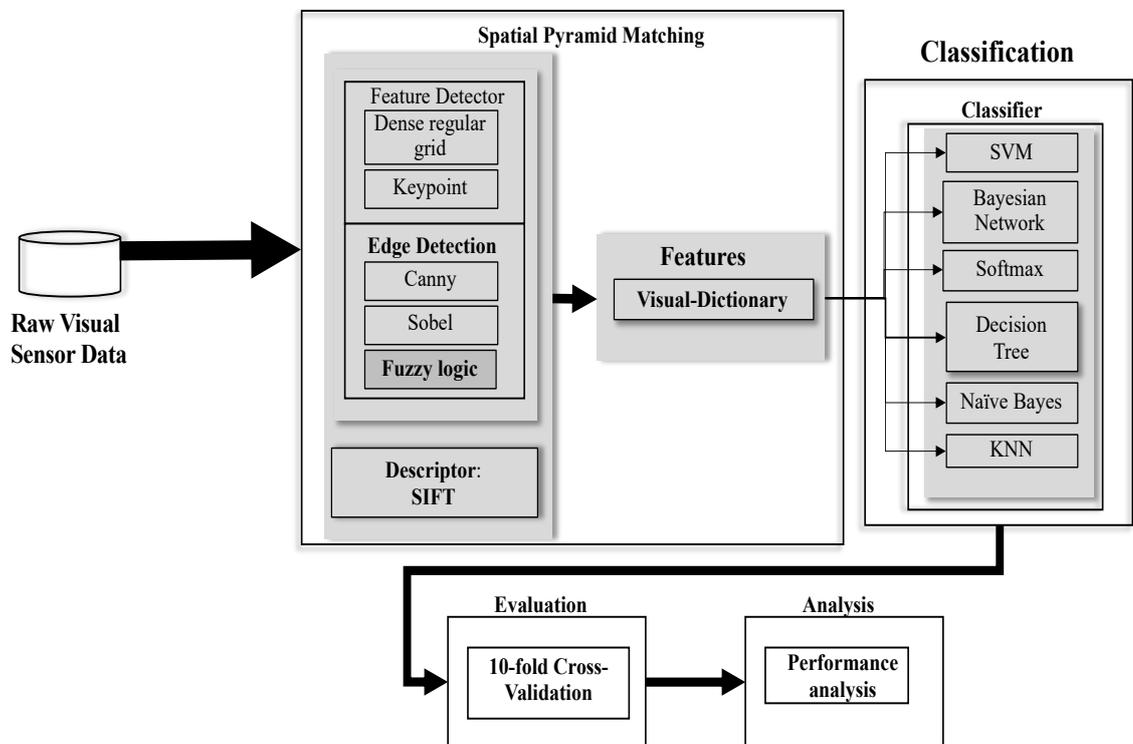


Figure 3.9: Proposed F-SPM image recognition method.

3.4.1 Feature Extraction

This section briefly describes the feature extraction phase of the proposed methods. The feature extraction phase extracts important elements of an object while the feature classification phase classifies the object into classes based on their similarity. Therefore, it is important to define and select useful features to recognize the object in the feature extraction phase (X. Chen et al., 2019). The proposed CW-GLCM method is developed to improve the detection of the GLCM Texture analysis method on low-quality region images. The F-SPM method, on the other hand, is developed to select significant features to reduce unnecessary high dimensionality features of the SPM and improve the detection in apple classification. The description of the two methods are presented as follows:

3.4.1.1 CW-GLCM (Proposed Method I)

The CW-GLCM is a feature extraction of fusion-features that inspired by GLCM Texture analysis method (Olaniyi et al., 2017). The GLCM Texture analysis method is chosen since it achieved high accuracy of 96.25%-100% in defective and non-defective classification as shown in (Olaniyi et al., 2017). However, the GLCM method presents limited capability in detecting images with low-quality region. For this reason, Wavelet and Curvelet transform is introduced in the proposed method. The Wavelet transform is used to improve the quality of the texture at the low-quality region in the GLCM method. However, its limitation lies in the curved region areas. To address this limitation, Curvelet transform is added to the proposed method because of its ability in capturing the directional edges of curves, corners and profiles. The Curvelet transform also provides richer information in both spatial and spectral domains. In the CW-GLCM method, the features from the Curvelet transform are fused with five GLCM features which are

entropy, contrast, correlation, homogeneity and energy that extracted based on the Wavelet coefficients to produce a highly informative fusion-features. The details of the proposed CW-GLCM method and its specific contribution is further described in Chapter 4.

3.4.1.2 F-SPM (Proposed Method II)

The second proposed method called F-SPM improves the classification accuracy of the proposed CW-GLCM method in detecting defective apples. The F-SPM method is a visual-dictionary feature extraction that includes spatial orders information of the SPM method (Lazebnik et al., 2006) to improve image representation for classification. The SPM method is chosen to encode the spatial distribution in the proposed method because it achieved excellent performance in the classification task and defective classification as reported in (X. Wei et al., 2019). The spatial layout information is important in the apple classification to discriminate between the defect and the natural parts of the stem end and calyx. To ensure the proposed F-SPM method is stable and reliable for apple classification, the existing SPM method is modified to reduce unnecessary SPM features through Fuzzy logic detection to include only significant features for further classification. The existing SPM method (Lazebnik et al., 2006) extracts feature vectors from all over the image using Dense regular grid. This generates a large number of unnecessary and redundant high dimensionality features (Chanti & Caplier, 2018; Lin et al., 2016; Penatti et al., 2014). These irrelevant features can reduce the accuracy performance in apple classification. The Fuzzy logic detection is introduced in the proposed method to improve the detection and reduce the unnecessary SPM features by selecting only the significant features to resolve the limitation on the SPM method. The

details of the proposed F-SPM method and contribution are further described in Chapter 5.

3.4.2 Classification

The classification phase classifies the extracted features into the relevant categories using classifier. Thus, an accurate image recognition method requires not only a good feature extraction method but also a good classifier (R. Chen, 2018). There are several properties of a good classifier such as generalization and compactness. The generalization is the ability of the classifier to learn the pattern based on the input data and accurately predict the unseen data. The compactness is related to simplicity and ability of the classifier model to make fast predictions.

In the learning structure of a classifier, the dataset is split into training set and validation set. The training set is used to learn the classifier and the validation set is reserved for evaluation (Kumar, Selvam, & Kumar, 2018; Singh, Thakur, & Sharma, 2016). There are two problems potentially occur during the training phase, which are high bias and high variance. The high bias is the overfitting situation where the classifier successfully learns the training set but fails to generalize. Thus, produces a high misclassification on a validation set. The high variance is the underfitting situation where the classifier fails to learn the training data because the model is too simple. Therefore, in the apple classification, complementing the proposed feature extraction with a suitable classifier is important to achieve optimal performance. One of the major challenges is to determine the suitable classifier that able to achieve better classification accuracy (H.-D. Cheng et al., 2006).

The suitable classifier to ensure the optimal performance for each proposed feature is identified by testing six classifiers in the classification phase. These classifiers are KNN, Bayesian Network, Naïve Bayes, Decision Tree, Softmax and SVM. They are selected based on their specific advantage.

The KNN classifier is among widely used classifier for image recognition task because of its simplicity and easy to implement (Avila, 2013; Duda, Hart, & Stork, 2001; Jadhav & Channe, 2016; Syaliman, Nababan, & Sitompul, 2018). It is an instance-based learning classifier where the hypotheses are constructed directly from the training instances. KNN classifies an object by referring to the feature similarity based on its nearest neighbors in the training instances (Singh et al., 2016). The object is assigned to the class based on the majority vote among its k nearest neighbors, where the k value determine the class and typically is a small positive integer. In finding the nearest neighbors, the Euclidean distance technique is normally used.

In contrast, the Bayesian Network is the classifier that based on bias-variance trade-off network structure. The network structure models will allow the Bayesian Network to precisely capture the fine details in the data (Petitjean, Buntine, Webb, & Zaidi, 2018). Bayesian Network able to interpret problem using a structural relationship among predictors. In Bayesian Network, no free parameters to be set and the training data are learned without having to hold the data in the main memory (Petitjean et al., 2018; Singh et al., 2016). It used a graphical model to represents probabilities distribution for set of variable via a directed acyclic graph (Chaturvedi et al., 2018). The acyclic graph of Bayesian Network structure consists of nodes that represents variables which can be observed or measured. The variable can take many distinctives value within a node, in which each of it has its special probability.

The Naïve Bayes is also an efficient and effective machine learning classifier (Domingos; Duda, Hart, & Stork, 2012; Karandikar, McLeay, Turner, & Schmitz, 2015) that derived from Bayesian Network theory. It is a simple probability that can be implemented with a linear complexity efficiently. This learning algorithm is based on Bayes theorem where it consist only one parent and several child nodes (Kumar et al., 2018; Langarizadeh & Moghbeli, 2016). It applies a naïve assumption that assumes features in a class are completely independent. This simplistic assumption enables efficient calculation and resulting highly scalable classifier.

Another simple and fast classifier that can achieve accurate result in most cases is the Decision Tree classifier (Jadhav & Channe, 2016). Furthermore, it also works well with the noisy data (Jadhav & Channe, 2016). The Decision Tree composed of the decision rules that predict an outcome based on optimal feature cutoff values that split independent variables into different groups recursively in a hierarchical manner (W. Wei, Polap, Li, Woźniak, & Liu, 2018). The tree structure made of root node, internal node and leaf nodes. The tree design starts with a root node which contains all set of input samples. When each time a rule is applied, the set is divided into two child nodes (internal node). If the internal node contains instances from a single class, the decision is readily made and it became a leaf; otherwise, the splitting process is continued. The analogy of applying the rules imposed on the internal nodes through certain attribute is called tree edges or branches and a class decision is called leaves (Kumar et al., 2018). The longer or shorter the tree structure is depending on the selection of priority features. The structure of tree grows until the terminal nodes or leaves obtained, which determine the class probability.

The Softmax classifier is one of the most commonly-used logistic regressions classifier for multi-class classification (Le & Mikolov, 2014; Pellegrini, 2015) especially in deep

learning method (Girshick et al., 2014; Qi, Wang, & Liu, 2017). The Softmax classifier provides flexible approach in classification process. The learning framework of the Softmax classifier offer flexibility where it is trained on a reference dataset using spectral and spatial information. The classification process in Softmax classifier used the logical regression method, which judge the input and then outputs the result as single label class (Qi et al., 2017).

Finally, the SVM classifier is selected because of it is well established technique in many image recognition tasks and its high accuracy performance (Attamimi, Araki, Nakamura, & Nagai, 2013; Auria & Moro, 2008; Danades, Pratama, Anggraini, & Anggriani, 2016; Ozkan, Ergin, Isik, & Isikli, 2015; Syaliman et al., 2018). In the SVM classifier, the Kernel function is used to model a higher dimensional data by adding the additional dimensions to draw the higher dimensional data. The used of kernel function in the SVM classifier also helps in reducing the time complexity of high dimensional data as it capable to compute faster.

All the mentioned classifiers are evaluated with the proposed feature extraction and a test decision of the best performing classifier on the proposed feature extraction determine the suitable classifier for the proposed method.

3.5 Performance Evaluation

This section describes the performance evaluation and the evaluation metrics used in this research. To demonstrate the reliability of the proposed methods, a series of comprehensive experiments are conducted in VLSI laboratory Faculty of Computer Science and Information Technology, University of Malaya, Kuala Lumpur, Malaysia. All the experiments are conducted using MATLAB R2017b on a computer with the

following specification: Windows 10 Pro and an Intel Core i7-4770 CPU (3.4GHz) processor with 8.00 GB RAM.

In the experiments, the training and testing sets are specified using K -fold cross-validation technique. The value of K strongly depends on the quantity of the data and suggested between 5 to 10 folds (Hastie, Tibshirani, Friedman, & Franklin, 2005; Roberts et al., 2017). Based on the experiment conducted on the NDDA dataset that evaluates the number of folds in connection with the classification accuracy (see Appendix A), the K parameter of 10 yields the highest classification accuracy. The experiment shows that the classification accuracy increased with the increment of folds. This is because the number of training images is also increased when the number of folds increased, thus increases the classification accuracy. Therefore, the 10-fold cross-validation is employed in this research. Also, the 10-fold cross-validation is mostly employed in the work related to this domain for evaluating the performance (Chui & Lytras, 2019; de Haan et al., 2017; George & Zwiggelaar, 2019; Roberts et al., 2017). The 10-fold cross-validation randomly partitioned the dataset into ten folders, in which each folder has virtually the same number of class distribution. Nine of the folders are used for training and one folder for validation. This process is repeated ten times until each folder is used exactly once as a validation set. Finally, the results from the ten experiments are averaged. Figure 3.10 illustrates the K -fold cross-validation, where $k = 10$.

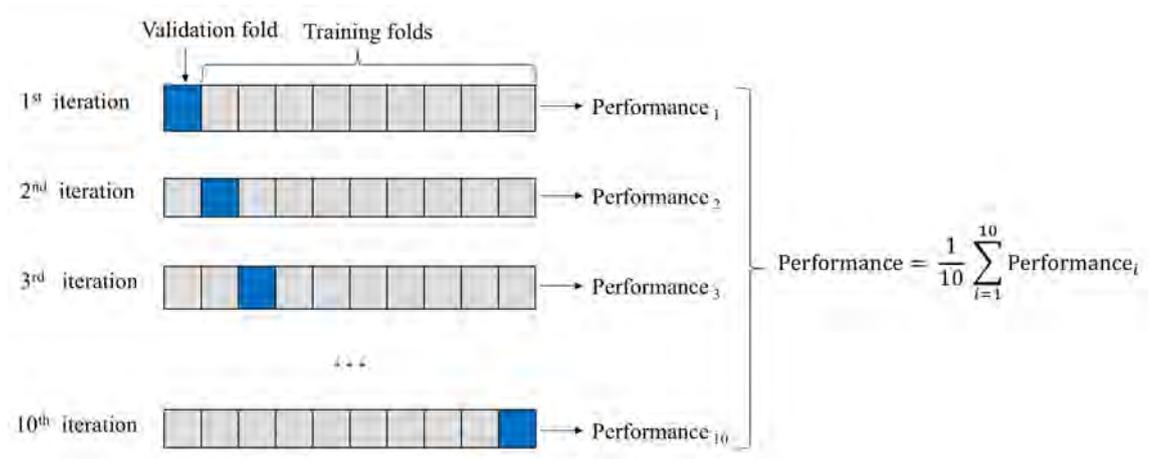


Figure 3.10: Illustration of 10-fold cross-validation.

The classification performances in the experiments are measured in terms of precision, recall and accuracy. These measurements are calculated from TP (true positive), TN (true negative), FP (false positive) and FN (false negative). The TP, TN, FP and FN identify the amount of correctly or incorrectly classified instances to the positive or the negative classes. The accuracy score is considered in this research to describe the correct rate of a classifier. The accuracy score often used as a statistic performance score measure when dealing with classifier. Several different score which are precision and recall also considered and calculated. The precision is measured to depict a prediction of the possible true positive value which is the proportion ratio of the positive instances that are correctly classified. Finally, the recall is included to represent the proportion of the negative instances that are correctly classified by the classifier. Other than that, the computational time is also recorded and reported. The details of the evaluation for binary-class and multi-class classifications are described in the following subsections.

3.5.1 Binary-Class Evaluation

The positive and negative classes can clearly define in the binary-class classification (Sokolova & Lapalme, 2009). The measurement of the precision, recall and accuracy can be computed as follows:

$$\text{Precision} = \text{TP}/(\text{TP} + \text{FP}), \quad (1)$$

$$\text{Recall} = \text{TP}/(\text{TP} + \text{FN}), \quad (2)$$

$$\text{Accuracy} = (\text{TN} + \text{TP}) / (\text{TP} + \text{FN} + \text{FP} + \text{TN}), \quad (3)$$

where the terminology and derivations of TP, TN, FP and FN are given in Table 3.5.

Table 3.5: Terminology and derivations of the evaluation metrics.

Terminology	Derivations
TP (true positive)	Defective apple is correctly classified as defective apple
TN (true negative)	Non-defective apple correctly classified as non-defective apple
FP (false positive)	Defective apple incorrectly classified as non-defective apple
FN (false negative)	Non-defective apple is incorrectly classified as defective apple

3.5.2 Multi-Class Evaluation

The assessment of the multi-class classification performance is defined by $\text{TP}_i, \text{FN}_i, \text{TN}_i, \text{FP}_i$ of the individual class C_i . The $\text{Precision}_i, \text{Recall}_i, \text{Accuracy}_i$ are calculated from the counts for C_i (Sokolova & Lapalme, 2009). The overall classification are calculated from the summation of the cumulative $\text{TP}_i, \text{FN}_i, \text{TN}_i, \text{FP}_i$ and then the average performance of precision, recall and accuracy are calculated as shown in Table 3.6.

Table 3.6: Measures precision, recall and accuracy for multi-class classification based on evaluation metrics for five defect classes C_i : TP_i are true positive for C_i , FP_i are false positive, FN_i are false negative and TN_i are true negative counts respectively.

Measure	Formula	Evaluation Focus
Precision	$\frac{\sum_{i=1}^5 \frac{TP_i}{TP_i + FP_i}}{5}$	An average per-class agreement of the data class labels with those of a classifier
Recall	$\frac{\sum_{i=1}^5 \frac{TP_i}{TP_i + FN_i}}{5}$	An average per-class effectiveness of a classifier to identify class labels
Accuracy	$\frac{\sum_{i=1}^5 \frac{TP_i + TN_i}{TP_i + FN_i + FP_i + TN_i}}{5}$	The average per-class effectiveness of a classifier

3.6 Chapter Summary

This chapter explained the research methodology conducted to answer the research questions and meet the objectives. The design and implementation of the proposed CW-GLCM and F-SPM were briefly presented. The details of the design and performance for each proposed method will be elaborated in chapter 4 and chapter 5, respectively.

CHAPTER 4: CW-GLCM METHOD

This chapter presents the first proposed method, namely CW-GLCM. The CW-GLCM is proposed to improve the ability of the GLCM method in detecting features on the low-quality region of the apple image for binary-class classification of defective and non-defective apple images. The proposed method inspired by the Texture analysis of GLCM method (Olaniyi et al., 2017) since it presented high accuracy for defective and non-defective classification. However, the limitation of the method is it dependent on the texture information features provided in the images. Utilizing the GLCM method alone limits the capability of the method to distinguish object with quite similar texture and images with low-quality region (Fahrurozi et al., 2016; Y. Li et al., 2015). In apple classification, the accuracy of the method degrades due to the presence of low-quality image region features on the apple skin. Thus, the proposed CW-GLCM method improve and enhance the detection of features on the low-quality image region by incorporate the Curvelet and Wavelet transform with GLCM method. The CW-GLCM method fused the features of Curvelet and five GLCMs features based on Wavelet coefficient to produce highly informative fusion-features that able to effectively classify between defective and non-defective apple images including images with low quality region in which, the key contribution of this research. To evaluate the performance of the proposed CW-GLCM method, two datasets binary-class of defective and non-defective apple images, which are NDDA and NDDAW are considered. These binary-class datasets contain 1110 apple images from defective and non-defective categories. The NDDA dataset is used to evaluate the capability to detect various defective and non-defective apple types, while the NDDAW particularly to evaluate the effectiveness of the proposed method against low-quality region. The chapters are divided into six main sections: the first section (section 4.1) briefly introduced the method. The second section (section 4.2) described the process flow of the proposed CW-GLCM method. The third section (section 4.3)

presented the experimental results and analysis; followed by the discussion in fourth section (Section 4.4). Whereas the last section (Section 4.5) summarized the whole chapter.

4.1 Introduction

Texture features is one of the suitable features for defective and non-defective classification as it can represent the surface and structure of the image. The GLCM is an effective method to extract texture information (Mondal et al., 2017; Sthevanie & Ramadhani, 2018) where it can describe the relationship of the neighboring pixels in the image. However, the method is dependent on the images texture information which ineffective to extract features from the low-quality region images. In apple classification, the detection and extraction of features on low-quality region is important to differentiate between defective and non-defective apples. Failure to detect these features may reduce the classification accuracy.

For this reason, a new image recognition method of CW-GLCM method to effectively classify defective and non-defective apple images including images with low quality region is proposed. The CW-GLCM is a feature extraction of fusion-features based on the GLCM method. The proposed method improve the detection on the existing GLCM Texture analysis (Olaniyi et al., 2017) method in detecting on low-quality region images. In GLCM Texture analysis, the features were extracted from a co-occurrence matrix based on the selection of GLCM features. In contrast, the proposed CW-GLCM method incorporate the Curvelet and Wavelet transform with GLCM method to enhance the apple images especially on low-quality region by improving their texture information. The proposed CW-GLCM method fused the features of Curvelet and five GLCMs features

extracted based on Wavelet coefficient forming a set of fusion-features to improve the detection.

4.2 CW-GLCM Method

This section describes the phases involved in the proposed CW-GLCM method. The proposed CW-GLCM method focuses to improve the detection on low-quality region image. The proposed method introduces the Curvelet and Wavelet transform in the modified GLCM method to improve the detection. The Wavelet transform is used in the GLCM method to improve the quality of the texture on the low-quality region images. Due to Wavelet transform limitation that lies in the curved region areas, the Curvelet transform are also used in the proposed method to effectively deals with a low-quality region area since it has a better ability in capturing the directional edges of curves, corners and profiles (Agarwal & Bedi, 2015; J. Luo et al., 2014). The Curvelet transform also provides richer information in both spatial and spectral domains (Hagargi & Shubhangi, 2018). These will enhance the apple images especially on low-quality region by improving their texture information.

The proposed CW-GLCM method consist of two main phases, feature extraction and feature classification as shown in Figure 4.1. The feature extraction phase concentrate on the selection of fusion-features that able to increase the classification accuracy for defective and non-defective apple including low-quality region images. For this reason, the Curvelet and Wavelet transform is introduced to enhance the detection of features on the low-quality image region by improving their texture information. In the feature extraction phase, the images are subjected to the Curvelet transform to obtain the Curvelet features. The images are also subjected to the Wavelet transform in order to obtain the Wavelet coefficient. From these Wavelet coefficients, five GLCMs features which is

entropy, contrast, correlation, homogeneity and energy are extracted. In this phase, there are six different features in total which is Curvelet, entropy, contrast, correlation, homogeneity and energy that are fused together forming a set of fusion-features. The fusion-features obtained in the feature extraction phase are then transferred to the feature classification phase. In the classification phase, six classifiers are utilized to select the most suitable classifier for the proposed fusion-features in classifying defective or non-defective apple images. With the use of the proposed fusion-features in feature extraction phase, the classification is expected to be more accurate than solely dependent on GLCMs features especially with the presence of low-quality regions in the apple images. The output from the classifier can be used for the data analytics and visualization to identify the patterns and learn for future decision making and actions. The proposed CW-GLCM method that comprises of two phases, will be discussed in the following subsections.

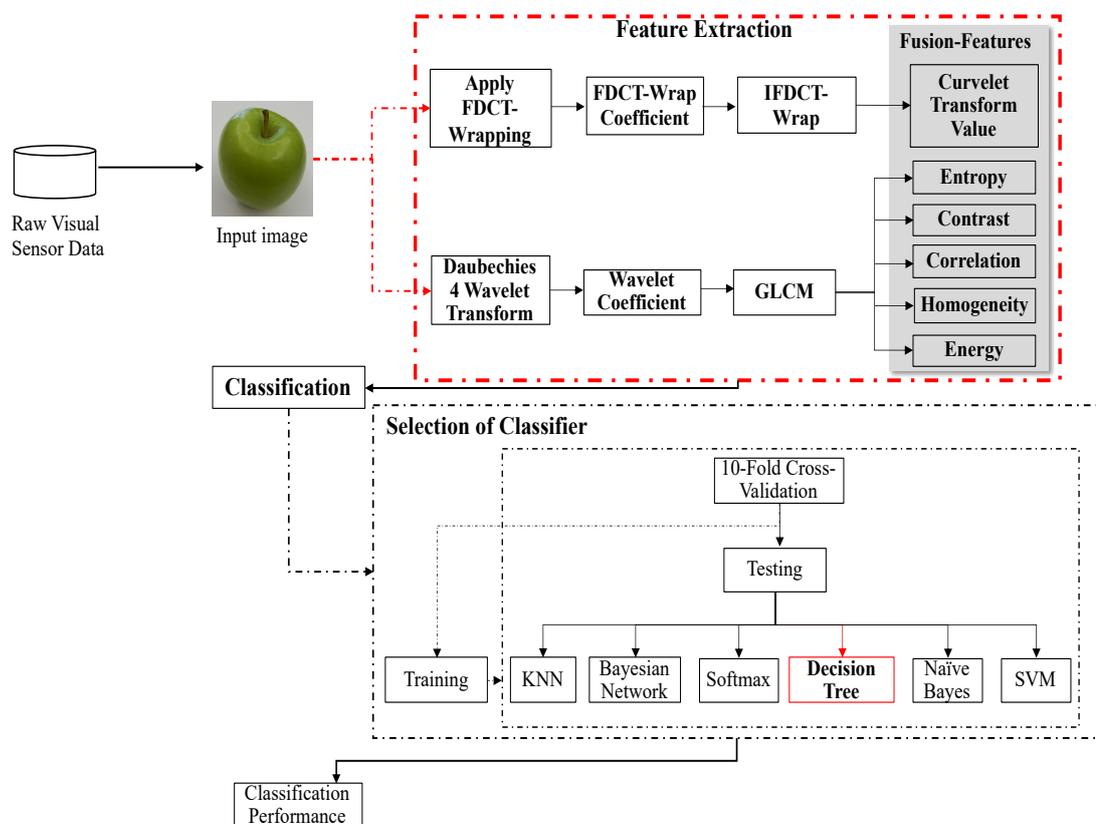


Figure 4.1: Process flow of the proposed CW-GLCM method for apple classification².

4.2.1 Feature Extraction

The phase of the proposed method feature extraction comprises of three major methods, which are Curvelet, Wavelet and GLCM method. The proposed method combines the Curvelet features with five GLCMs features extracted based on the Wavelet coefficient. To retain the image information, the image normalization step is skipped in the proposed method to deal with the low-quality region on the apple skin images. This is to avoid misclassification between defective and non-defective apple.

4.2.1.1 Curvelet Transform

The main reason that the proposed method fuses the Curvelet features is to detect the low-quality apple images region for curves, corners and profiles. As compared with other transforms, the Curvelet is effective and accurate at capturing the edges and other singularities along the curves (Acharya et al., 2016). The Curvelet features will provide more information on low-quality regions in the apple images. In the proposed method, the Curvelet transform based on wrapping of specially selected Fourier samples (FDCT-Wrap) is used because it is the fastest and well-adapted Curvelet transform algorithm to represent edges (Abdullah, Hazem, & Reham, 2017; Candes et al., 2006; curvelet.org). The FDCT-Wrap is applied to enhance the image contrast of the low-quality region. The two consecutive regions between the low-quality regions that has a different pixel value with the nearby region are likely to form “edges” as illustrated in Figure 4.2. This edges are formed based on the variation of pixel values allowing the FDCT-Wrap in the proposed method to detect this edges information. In order to obtain the dominant features, the LL sub-bands filter is applied to set the intensity elements of the FDCT-Wrap coefficient. Then, the inverse transformation is performed on the extracted features from the FDCT-Wrap coefficient to produce the Curvelet transform value.

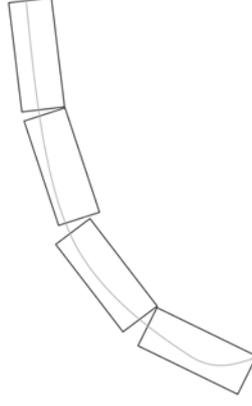


Figure 4.2: Representation of edges in Curvelet transform.

The steps of FDCT-wrap are as follows:

Step 1. Input the image;

Step 2. Apply 2D Fast Fourier Transform (2DFFT) on the image; produce a set of Fourier sample $\hat{f}[n_1, n_2]$;

Step 3. Resample a set of Fourier sample $\hat{f}[n_1, n_2]$ at each pair of scale j and angle direction l in frequency domain. The scale is from finest to coarsest scale with the angle direction start from the top-left corner increases clockwise. This will produce the new sampling function as expressed in (4).

$$\hat{f}[n_1, n_2 - n_1 \tan \theta_l], \quad (n_1, n_2) \in \{(n_1, n_2), n_{1,0} \leq n_1 \leq n_{1,0} + L_{1,j}, n_{2,0} \leq n_2 < n_{2,0} + L_{2,j}\}, \quad (4)$$

where $n_{1,0}$ and $n_{2,0}$ are the initial position of window function $\tilde{u}_{j,l}[n_1, n_2]$, $L_{1,j}$ and $L_{2,j}$ are parameter of length 2^j and width $2^{j/2}$ components of window function support interval. The window function formula is defined in (5).

$$\tilde{u}_{j,l}[n_1, n_2] = w_1(w_1, w_2) v_j \left(s_{\theta} \cdot \frac{(2^{\lfloor j/2 \rfloor} w_2)}{w_1} \right),$$

$$w_1(w_1, w_2) = \sqrt{\phi_{j+1}^2(w^2) - \phi_j^2(w^2)},$$

$$\phi_j(w_1, w_2) = \phi(2^{-j} w_1) \phi(2^{-j} w_2), \quad (5)$$

$$s_{\theta_l} = \begin{bmatrix} 1 & 0 \\ -\tan \theta_l & 1 \end{bmatrix},$$

$$\tan \theta_l = l \times 2^{\lfloor -j/2 \rfloor}, l = -2^{\lfloor -j/2 \rfloor}, \dots, 2^{\lfloor -j/2 \rfloor} - 1$$

where w_1 is a vertical axis but located near the horizontal axis of w_2 ;

Step 4. The new sampling function of $\hat{f}[n_1, n_2 - n_1 \tan \theta_l]$ are multiplied with the window function $\tilde{u}_{j,l}[n_1, n_2]$:

$$\widetilde{f}_{j,l}[n_1, n_2] = \hat{f}[n_1, n_2 - n_1 \tan \theta_l] \tilde{u}_{j,l}[n_1, n_2]; \quad (6)$$

Step 5. Then, the inverse 2DFFT is applied to each of $\widetilde{f}_{j,l}$ obtained in the previous step to produce Curvelet transform value.

Finally, feature vector of the Curvelet transform value are extracted and then fused them with the GLCMs features calculate based on the Wavelet transform in the following step to accomplish defective and non-defective apple classification.

4.2.1.2 Wavelet Transform

To improve the texture information extracted from the GLCM method, the proposed method also modifies the existing GLCM method by extracting the GLCMs features based on the Daubechies 4 Wavelet coefficient. Daubechies 4 Wavelet is chosen as it is suitable for texture classification due to their relations to multiresolution. The Wavelet

coefficient will enhance the visibility of the low-quality region in the apple images, especially on the apple skin by capturing the directional edges in different resolution levels preserving the low and high frequency information. This leads the proposed method to extract better texture information from the apple images. The edges representation in Wavelet transform are shown in Figure 4.3.

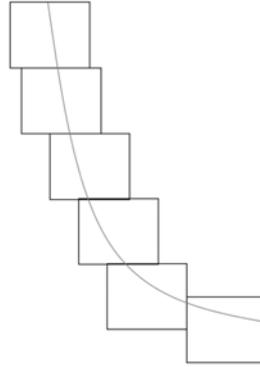


Figure 4.3: Representation of edges in Wavelet transform.

The Wavelet transform are calculated using the wavelet function as follows:

$$W[f(s, \tau)] = \langle f, \psi_{s,\tau}^k \rangle \quad (7)$$

where $s \in Z$ is parameter of scale resolution level, τ is translation, $k \in \{h, v, d\}$ is the orientation and $\psi_{s,\tau}^k = (\frac{1}{\sqrt{s}}) \psi^k(\frac{x-\tau}{s})$ is a wavelet family. The orientation h, v and d parameter represent horizontal, vertical and diagonal direction. The wavelet decomposition is achieved when the value of $s = 2^j$ and $\tau = 2^j \cdot n, j, n \in Z$.

The wavelet and scaling family are constructed using wavelet function $\psi(x)$ and scaling function $\varphi(x)$ as expressed in (8).

$$\psi_{j,n}^k(x) = \frac{1}{\sqrt{2^j}} \psi^k\left(\frac{x-2^j \cdot n}{2^j}\right) \quad \text{and} \quad \varphi_{j,n}^k(x) = \frac{1}{\sqrt{2^j}} \varphi^k\left(\frac{x-2^j \cdot n}{2^j}\right) \quad (8)$$

The two dimensional Wavelet transform are constructed based on the combination of high-pass and low-pass digital filter banks and down-samplers. As the images is in 2D signal, separable function Discrete Wavelet Transform is used in the configuration of DWT structure. The rows and columns of the images are subjected to the 1D Wavelet Transform separately to produce the 2D-DWT. The output of the decomposed images in 2D orthogonal wavelet representation resulting four orthogonal sub-bands component which are Low-Low (LL), Low-High (LH), High-Low (HL) and High-High (HH) as presented in Figure 4.4. The results shown in the figure is for one level decomposition. Every stage of DWT requires high-pass and low-pass digital filter with two down sampling (Sarala & Sivanantham, 2014). This process is further continued and decomposed to another four sub-band components, forming two-level decomposition. The wavelet decomposition at two resolution levels as illustrates in Figure 4.5.

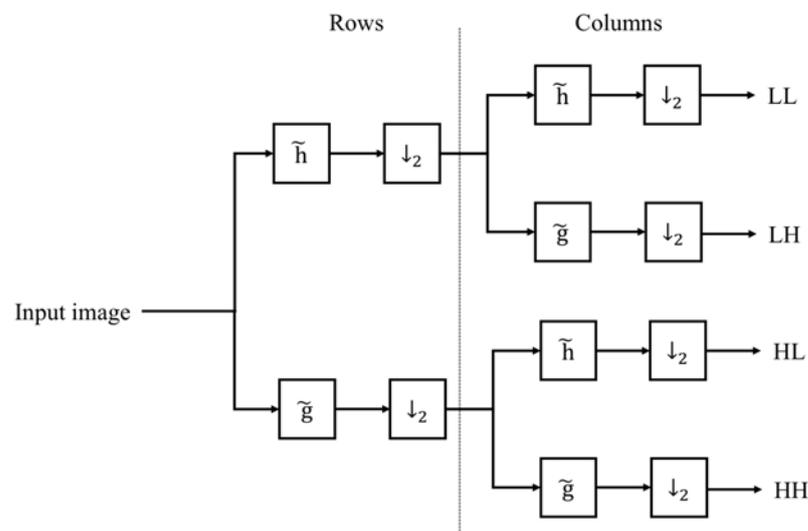


Figure 4.4: Image decomposition using analysis filter banks. Note that \tilde{h} is low-pass filter and \tilde{g} is high-pass filter and \downarrow_2 is keeping one sample out of two (down sampling)².

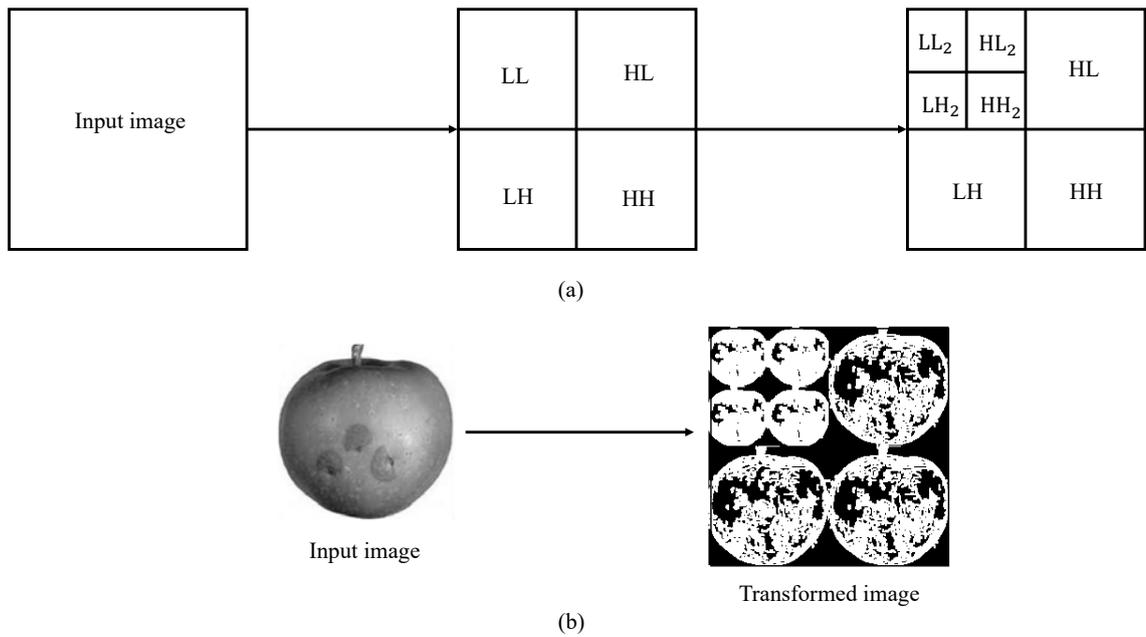


Figure 4.5: 2D-DWT wavelet decomposition at two resolution levels (a) the structure diagram (b) apple image via wavelet decomposition at two resolution levels².

4.2.1.3 GLCM

The GLCMs texture features are extracted from GLCM based on the computed Wavelet coefficient from the prior process in section 4.2.1.2. The GLCMs features are included in the proposed method to estimate the apple images texture properties. Instead of calculating the GLCMs features from GLCM coefficient, the proposed method modifies the original GLCM implementation by using Wavelet coefficient to calculate five of the GLCMs feature. The features are entropy, contrast, correlation, homogeneity and energy. The entropy is a measure of levels disorderliness and randomness in the images. It is the most dominant statistical features and widely used to measure variations between pixel intensities (Acharya et al., 2016). This is important to symbolize texture that appear in the apple images. The contrast measures the variation values and intensity contrast of the neighboring pixel in the gray level. The correlation features are also selected since it measures the correlated pixels to the neighbors over the whole image and

determined the linear dependencies of the gray levels. Homogeneity features are important to measure the uniform region in the images according to its gray level difference and the energy returns the sum value of the squared elements in the GLCM. To extract the texture features from GLCM, the matrix must be symmetric (Girisha, Chandrashekhar, & Kurian, 2013). In order to get a symmetric matrix, the GLCM is transposed and added to the original GLCM. From the symmetrical GLCM, the texture features are extracted. To compute the GLCM, the spatial relationship between two pixels is establish. The first one is the reference pixel which is pixel-of-interest and the other pixel is a neighbor pixel. This process forming the GLCM that contains different combination of pixel gray values. The number of gray levels (G) is ranging from 0 to $G - 1$. The GLCM is highly dependent on two parameters which are distance between the pixel pair (D) and their angular relationship (θ). In the proposed method, the GLCM is computed based on the predefine distance of one pixel ($D = 1$) and θ are quantized in four parameter directions which are 0° , 45° , 90° and 135° . The directionality of GLCM used in the proposed method illustrated as in Figure 4.6. This forms four co-occurrences matrix. The GLCM are calculated in the corresponding matrix by taking the absolute value of each resolution level of the Wavelet coefficient matrix obtained using 2D-DWT in the prior section. Then, each of the GLCMs texture features are computed from the details of the Wavelet coefficient matrix for various resolution level in the corresponding matrix.

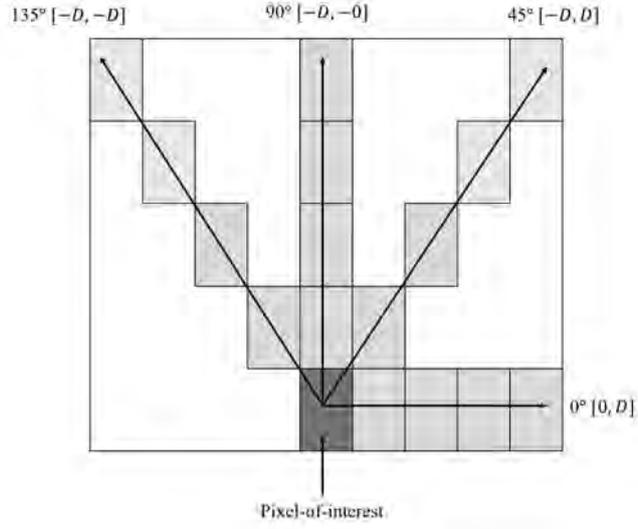


Figure 4.6: Directionality of GLCM².

The procedure of the extraction of GLCMs features based on the Wavelet coefficient are as follows:

Step 1. Input the original image;

Step 2. Compute GLCM from the Wavelet coefficient and calculate based on five GLCMs. The formulations for each of the GLCMs features are computed as follows:

$$\text{Entropy} = \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} P(i, j) * \log(P(i, j)), \quad (9)$$

$$\text{Contrast} = \sum_{i=0}^{G-1} n^2 \left\{ \sum_{i=1}^G \sum_{j=1}^G P(i, j) \right\}, |i - j| = n, \quad (10)$$

$$\text{Correlation} = \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} \frac{\{i * j\} * P(i, j) - \{\mu_x * \mu_y\}}{\sigma_x * \sigma_y}, \quad (11)$$

$$\text{Homogeneity} = \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} \left\{ \frac{P(i, j)}{1 + |i - j|} \right\}, \quad (12)$$

$$\text{Energy} = \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} P(i, j)^2, \quad (13)$$

where P is a pixel, i is row, j is column, n is line of neighborhood, G represents the number of gray levels used, $\mu_x, \mu_y, \sigma_x, \sigma_y$ are the mean and standard deviation value obtained from P_x and P_y respectively. The P_x and P_y are the results obtained after summing the rows $P(i, j)$;

Step 3. Acquire texture features according to (9), (10), (11), (12) and (13).

Finally, the Curvelet features obtained from the prior process in section 4.2.1.1 are fused in the texture features obtained in step 3 to produce highly informative fusion-features. The parameter settings of Curvelet, Wavelet and GLCM used in the proposed method are summarized in Table 4.1.

Table 4.1: Parameters setting of Curvelet, Wavelet and GLCM.

Algorithm	Parameters	Settings
Curvelet	scale (j)	3
	number of angles at the second coarsest level (l)	16
	fine	2
Wavelet	level of decomposition (s)	2
	number of minimum scale	7
	number of vanishing moments	4
	gray levels (G)	0 to $G - 1$
	total number of directions	3
GLCM	angle (θ)	$0^\circ, 45^\circ, 90^\circ$ and 135°
	distance pixel (D)	1
	texture features	entropy, contrast, correlation, homogeneity and energy

4.2.2 Classification

In the classification phase, the fusion-features extracted in the previous phase are classified into the defective and non-defective apples using classifier. Six classifiers which are KNN, Bayesian Network, Softmax, Decision Tree, Naïve Bayes and SVM have been utilized comparatively to select the most suitable classifier for the proposed fusion-features. Their performances are evaluated and compared.

4.3 Experimental Results and Analysis

This section presents the experimental results of the proposed CW-GLCM method on binary-class of NDDA and NDDAW datasets. To demonstrate the reliability performance of the proposed CW-GLCM method, a series set of comprehensive experiments was conducted. In apple classification, it is crucial to detect on the low-quality region in the apple images to enable the classifier to differentiate and classify between defective and non-defective. The results shown in this section prove the success of the proposed CW-GLCM method in detecting features on the low-quality region of the apple image and significantly improve the ability of GLCM method.

4.3.1 Performance Measure for Fusion-Features

As highlighted in Section 4.2, the fusion-features are the combination of Curvelet features and five GLCMs features based on the Wavelet coefficient. This forming a set of fusion-features which consist of six features. They are the Curvelet features, entropy, contrast, correlation, homogeneity and energy. In searching for the best fusion-features, the fusion-features are compared with the Curvelet features and each of the GLCMs features calculated from GLCM coefficient. Their performances are evaluated and

compared with the proposed fusion-features on NDDA dataset using SVM classifier as shown in Table 4.2. Based on the table, the results show that the proposed fusion-features outperformed others with 88.89% precision, 85.71% recall and 87.04% accuracy. Although the fusion-features require the longest time for training and testing, the results proved that the Curvelet and Wavelet transform can improve the detection of the GLCM texture features.

Table 4.2: Comparative results (precision, recall, accuracy, training time and testing time) for the proposed fusion-features using SVM classifier on NDDA.

Features	Precision (%)	Recall (%)	Accuracy (%)	Training Time (s)	Testing Time (s)
Curvelet	77.78	75.00	75.93	317.34	2.97
Contrast	58.57	54.67	55.00	103.29	1.51
Correlation	53.57	52.26	52.32	107.91	0.72
Energy	67.14	58.02	59.29	101.09	0.32
Homogeneity	56.07	55.28	55.36	100.73	0.37
Entropy	78.21	57.94	60.71	101.35	0.32
Fusion-Features (Proposed Features)	88.89	85.71	87.04	364.51	3.12

A graphical comparison performance between the proposed fusion-features with contrast, correlation, energy, homogeneity, entropy and Curvelet using SVM classifier on NDDA dataset are presented in Figure 4.7. Based on the Figure 4.7 (a), the proposed fusion-features is shown to be able to obtain the highest percentage for all measurement of precision, recall and accuracy compared to other features with a minimum value of 85.71% on recall. However, it requires the longest time for the training and testing as shown in Figure 4.7 (b). This is due to the reason that the fusion-features incorporate three major methods which is the Curvelet, Wavelet and the GLCM method. In addition, the normalization step is skipped in the feature extraction phase to retain the information of a low-quality region on the apple skin images. This will increase the time complexity and reduce the speed. Although the fusion-features show high computational time, from these results it can be seen that the fusion-features of Curvelet features with five GLCMs

features calculated based on the Wavelet coefficient outperformed others in term of the precision, recall and accuracy.

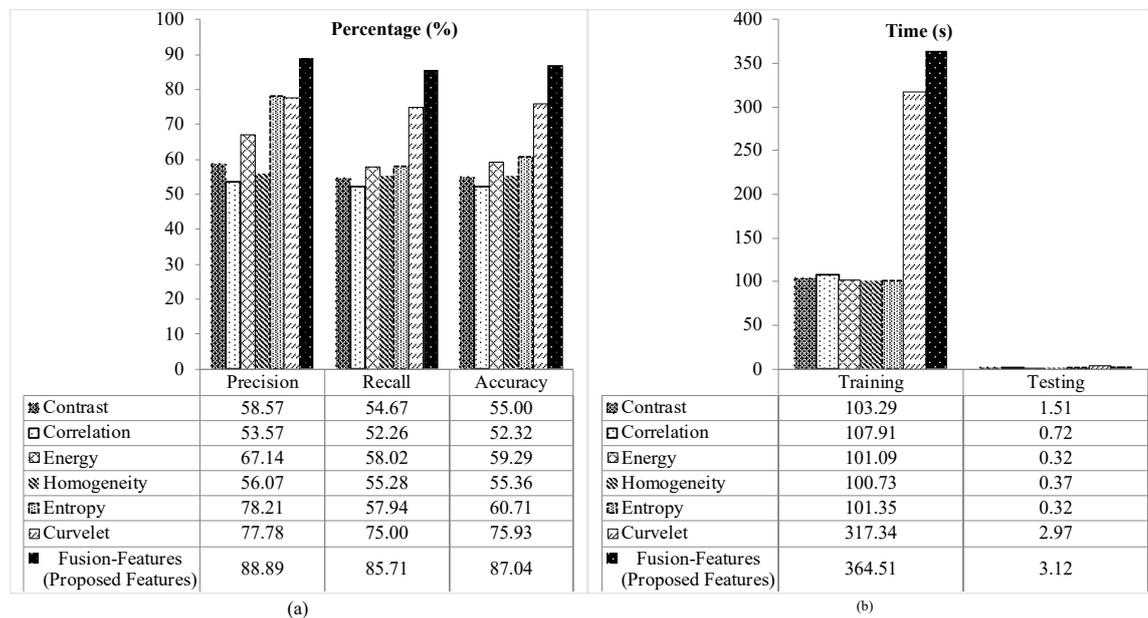


Figure 4.7: Comparative results for the proposed fusion-features using SVM classifier on NDDA dataset (a) precision, recall and accuracy (b) training and testing time.

To obtain the most suitable classifier, the fusion-features utilize six classifiers and are tested on NDDA dataset. The selections of classifiers are KNN, Bayesian Network, Softmax, Decision Tree, Naïve Bayes and SVM. The results for each classifier are presented in Table 4.3.

Table 4.3: Comparison of the proposed fusion-features with different classifiers on NDDA dataset.

Classifier	Fusion-Features (Proposed Features)				
	Precision (%)	Recall (%)	Accuracy (%)	Training Time (s)	Testing Time (s)
KNN	70.37	73.08	72.73	386.19	23.55
Bayesian Network	70.37	76.00	74.07	385.97	0.57
Softmax	79.31	71.88	73.21	372.24	1.10
Decision Tree	96.30	100.00	98.15	344.17	0.25
Naïve Bayes	59.26	76.19	70.37	390.18	0.98
SVM	88.89	85.71	87.04	364.51	3.12

From the Table 4.3, the fusion-features with Decision Tree classifier give the best performance for all for the measurement of precision, recall and accuracy including the computational time. The comparative results for different classifier are presented in Figure 4.8.

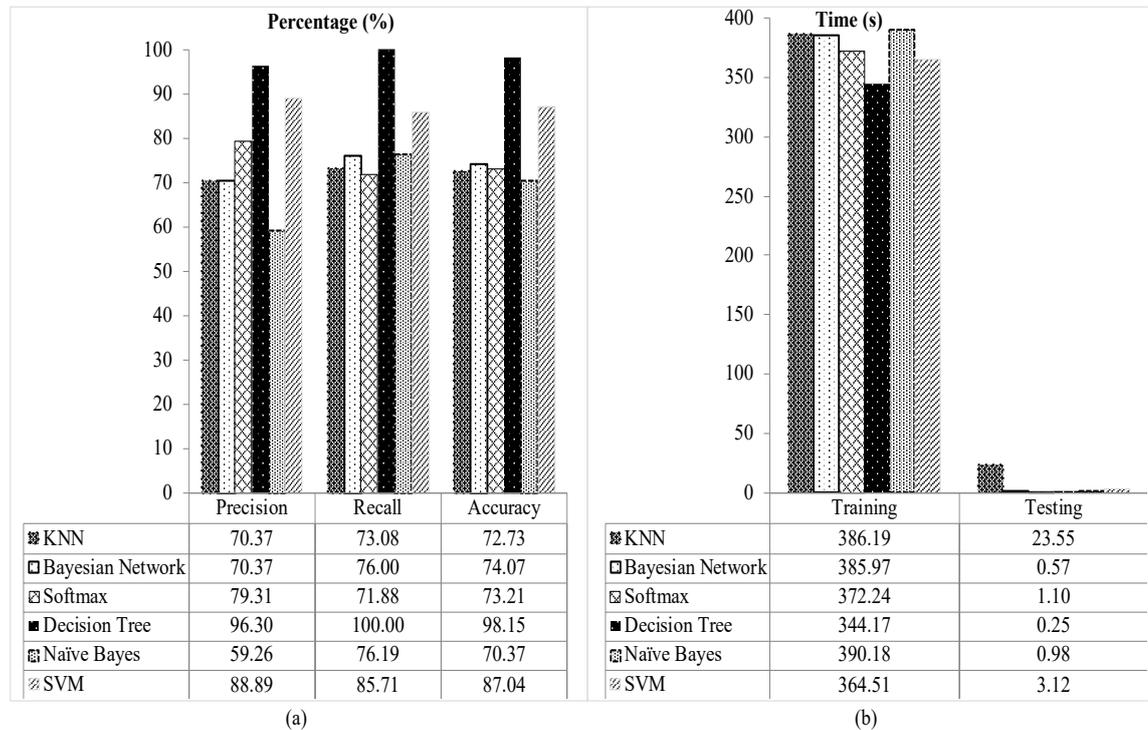


Figure 4.8: Comparative results of fusion-features for different classifier on NDDA dataset (a) precision, recall and accuracy (b) training and testing time.

Based on Figure 4.8, the Decision Tree classifier outperformed others with 96.30% of precision, recall 100% and accuracy 98.15%. In contrast, Naïve Bayes classifier shows the lowest performance for precision (59.26%) and accuracy (70.37%). This is because Naïve Bayes classifier make a very strong assumption that all variables are mutually correlated and contribute towards classification. Due to this assumption, it degrades the classification performance (Jadhav & Channe, 2016). The lowest recall is observed in Softmax classifier with 71.88%. The performance decrease in term of recall rate in the Softmax classifier is due to overfitting from high-variance structure (Pellegrini, 2015).

In terms of computational time among the classifiers, Naïve Bayes takes the longest time for training (390.18 s), whereas KNN for testing (23.55 s). The Naïve Bayes classifier is based on probabilistic that requires the knowledge of prior probability distribution of the class and also data to be classified. This increased the training time in Naïve Bayes classifier. Conversely, the KNN classifier is computationally intensive as it stores all the training data and compares the extracted features on the test images with each training data for classification (Syaliman et al., 2018). In contrast, the Decision Tree classifier is the fastest classifier during the training (344.17 s) and testing (0.25 s).

The results also show that the Decision Tree classifier able to achieve the highest performance for all measurements. This is due to the reason that different ranges of features in the proposed fusion-features does not affect the Decision Tree. The Decision Tree is a non-parametric classifier where it can deal with linearly inseparable data and capable to handle variety of data either nominal, numeric or textual even with noisy data, redundant attributes or missing values (Jadhav & Channe, 2016; Singh et al., 2016). While in the SVM and other classifiers, each of the data instances is represented in the form of real numbers vectors. This transformation may affect the classification performance. Since the data normalization step is skipped in the feature extraction stage, the Decision Tree classifier is found to be more suitable to classify the proposed fusion-features since the Decision Tree classifier good at handling variety data (Jadhav & Channe, 2016). Therefore, in the proposed method, the Decision Tree classifier is chosen.

4.3.2 Comparison with Existing Methods

To further evaluate the performance of the proposed method, five existing methods for image recognition are compared with the proposed method. They are BOW (Csurka et al., 2004), SPM (Lazebnik et al., 2006), CNN (dos Santos Ferreira et al., 2017), Texture analysis (Olaniyi et al., 2017) and CLAHE+GLCM+ELM (W. Li et al., 2019). The average results for 10-fold cross-validation of each method are presented in Table 4.4.

Table 4.4: Comparison of confusion matrix for BOW (Csurka et al., 2004), SPM (Lazebnik et al., 2006), CNN (dos Santos Ferreira et al., 2017), Texture analysis (Olaniyi et al., 2017), CLAHE+GLCM+ELM (W. Li et al., 2019) and the proposed CW-GLCM method on NDDA dataset: Defective (D) and Non-defective (N).

Class	Methods											
	BOW		SPM		CNN		Texture analysis		CLAHE+GLCM+ELM		CW-GLCM (Proposed Method)	
	D	N	D	N	D	N	D	N	D	N	D	N
D	24	3	26	1	25	2	20	7	22	5	26	1
N	4	23	0	27	1	26	4	23	11	16	0	27
Precision (%)	88.89		96.30		92.59		74.07		81.48		96.30	
Recall (%)	85.71		100.00		96.15		83.33		66.67		100.00	
Accuracy (%)	87.04		98.15		94.44		79.63		70.37		98.15	
Training time (s)	402.94		150.97		149.13		135.12		1323.58		344.17	
Testing Time (s)	3.92		0.13		0.08		1.38		0.02		0.25	

Following the 10-fold cross-validation experiment on NDDA dataset, a total number of 275 images from the defective class and 275 images from the non-defective class are divided into ten equal parts. Each part consists of 28 or 27 images of the defective and non-defective classes. Nine parts are used for training and one for testing. This process is repeated ten times until each of the folders is used exactly once as a validation set. Then, the average value for all ten experiment are taken. In Table 4.4, the classification performance is led by the proposed CW-GLCM and SPM method with 98.15% classification accuracy. Both methods correctly classified all 27 images of non-defective apples followed by CNN (26 images), BOW and Texture analysis (23 images), while the lowest goes to CLAHE+GLCM+ELM (16 images). For the defective images, the

proposed CW-GLCM and SPM method correctly classified 26 out of 27 images of defective apples while CNN, BOW, CLAHE+GLCM+ELM and Texture analysis correctly classified 25, 24, 22 and 20 respectively.

Overall, the proposed CW-GLCM and SPM method outperformed others in NDDA dataset with the classification accuracy of 98.15%. This is followed by CNN with 94.44%, BOW 87.04%, Texture analysis 79.63% and CLAHE+GLCM+ELM 70.37% as presented in Table 4.4. The proposed CW-GLCM and SPM method also outperformed others with 96.30% precision and 100% recall rate. Among the methods, CLAHE+GLCM+ELM take the longest time for training (1323.58 s) and the fastest during testing (0.02 s). The CLAHE+GLCM+ELM method required the longest time for training due to the computationally extensive of CLAHE approach in the method. The CLAHE approach are usually used for image enhancement in off-line application (Reza, 2004). In contrast to the training time, the CLAHE+GLCM+ELM able to classify the dataset faster compared to other methods because of the extremely fast learning speed of the ELM classifier used in the method (W. Li et al., 2019). This is followed by the CNN method with 0.08 s, SPM 0.13 s, proposed CW-GLCM 0.25 s, Texture analysis 1.38 s and BOW 3.92 s. The BOW requires the longest time for testing the NDDA dataset because of the high computational cost in vector quantization step in BOW method. In contrast to BOW, the CNN method able to classify faster because of the input images to the CNN method were rescaled from the original of 900×700 pixels to 227×227 pixels. This is due to the CNN requirement of having fixed-size input images (He et al., 2015; Krizhevsky et al., 2012). If the arbitrary sizes of the images are applied, the CNN method will fit the images input to its fixed size via either cropping or warping the images (Donahue et al., 2014; Girshick et al., 2014; He et al., 2015; Krizhevsky et al., 2012). Although the proposed CW-GLCM requires a longer time to classify the dataset than the SPM method, the results are still acceptable as it takes only less than 0.23 s longer than the CLAHE+GLCM+ELM, which is the fastest

method during testing to successfully classify all the non-defective apple images. Other than that, the proposed CW-GLCM only misclassified one out of 27 defective apples. These results indicate that the proposed method able to effectively classify between defective and non-defective apple including the apple with a low-quality region on its skin. The examples of the low-quality region on the non-defective apple images can be found in bright-skinned apple and apple with yellow-white flecks as shown in Figure 4.9. In other methods, these types of apple may be misclassified as defective.

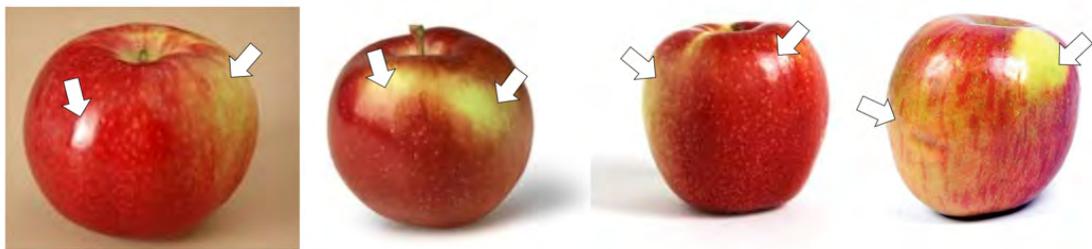


Figure 4.9: Examples of misclassified apple images with a low-quality region. The low-quality regions on apple skin images are pointed by the arrows².

4.3.3 Analysis of Classification Performance against Low-Quality Region

Based on the analysis of the proposed method on the NDDA dataset, the results were further explored. The proposed method is tested with NDDAW dataset in which the dataset was created particularly to include more low-quality apple image region. This dataset consists of 159 apple images with a low-quality region on its skin. The comparison average results for 10-fold cross-validation with other methods are presented in Table 4.5.

Table 4.5: Confusion matrix for BOW (Csurka et al., 2004), SPM (Lazebnik et al., 2006), CNN (dos Santos Ferreira et al., 2017), Texture analysis (Olaniyi et al., 2017), CLAHE+GLCM+ELM (W. Li et al., 2019) and proposed CW-GLCM method on NDDAW dataset: Defective (D) and Non-defective (N).

Class	Methods											
	BOW		SPM		CNN		Texture analysis		CLAHE+GLCM+ELM		CW-GLCM (Proposed Method)	
	D	N	D	N	D	N	D	N	D	N	D	N
D	21	7	20	8	22	6	20	8	14	14	24	4
N	5	23	9	19	6	22	13	15	11	17	2	26
Precision (%)	75.00		71.43		78.57		71.43		50.00		86.79	
Recall (%)	80.65		68.97		78.57		60.61		56.00		91.01	
Accuracy (%)	78.50		69.64		78.57		62.50		53.36		89.11	
Training time (s)	275.61		149.92		146.89		131.32		2075.93		386.63	
Testing Time (s)	3.84		0.47		0.34		1.41		0.02		0.31	

Following the 10-fold cross-validation for NDDAW dataset, a total number of 280 images from the defective class and 280 images from the non-defective class are divided into ten equal parts. Each part consists of 28 images of the defective and non-defective classes. Nine parts are used for training and one for testing. This process is repeated ten times until each of the folders is used exactly once as a validation set. Then, the average value for all ten experiment are taken. From Table 4.5, the classification performance is led by the proposed CW-GLCM method with 89.11% classification accuracy. This is followed by CNN 78.57%, BOW 78.50%, SPM 69.64%, Texture analysis 62.50% and CLAHE+GLCM+ELM 53.36%. The proposed CW-GLCM correctly classified 26 images out of 28 non-defective apples while CNN (22 images), BOW (23 images), SPM (19 images), Texture analysis (16 images) and CLAHE+GLCM+ELM (17 images). For the defective images, the proposed CW-GLCM method correctly classified 24 images of defective apples followed by CNN (22 images), BOW (21 images), SPM and Texture analysis (20 images) while CLAHE+GLCM+ELM (14 images). In this dataset, the proposed CW-GLCM outperformed others whereas the CLAHE+GLCM+ELM method recorded the lowest classification accuracy followed by Texture analysis method.

Overall, the CLAHE+GLCM+ELM method recorded the lowest classification accuracy performance in both datasets tested. This is due to the drawbacks of the CLAHE approach that sometimes may produce unwanted gray level artifact and creates an equal density in all the histogram bins during the image enhancement process (Hassan, Kasim, Jafery, & Shah, 2017). Furthermore, the ELM classifier is a single-hidden-layer, feed-forward neural networks that in the learning procedure of ELM classifier, one may require to tackle few major issues which is free parameters setting, convergence speed and overfitting (Chaturvedi et al., 2018). These reasons may reduce the classification accuracy in the methods. In contrast, the obvious accuracy performance difference can be observed from the SPM method between NDDA and NDDAW dataset. Although the SPM method achieved high percentage for the measurement of precision, recall and accuracy in NDDA dataset, it presents lower performance in NDDAW dataset. This is because the NDDAW dataset contains more apple images with low-quality region compared to NDDA dataset. The result shows that the SPM method is less sensitive in detecting features in the low-quality region. On the other hand, the proposed CW-GLCM method achieved the highest classification accuracy, precision and recall in both datasets. This indicates that the proposed method is more robust in detecting features on low-quality region. Figure 4.10 (a) and Figure 4.10 (b) depicts the performance of precision, recall and accuracy for NDDA and NDDAW. The training and testing time tested on each dataset are presented in Figure 4.10 (c) and Figure 4.10 (d).

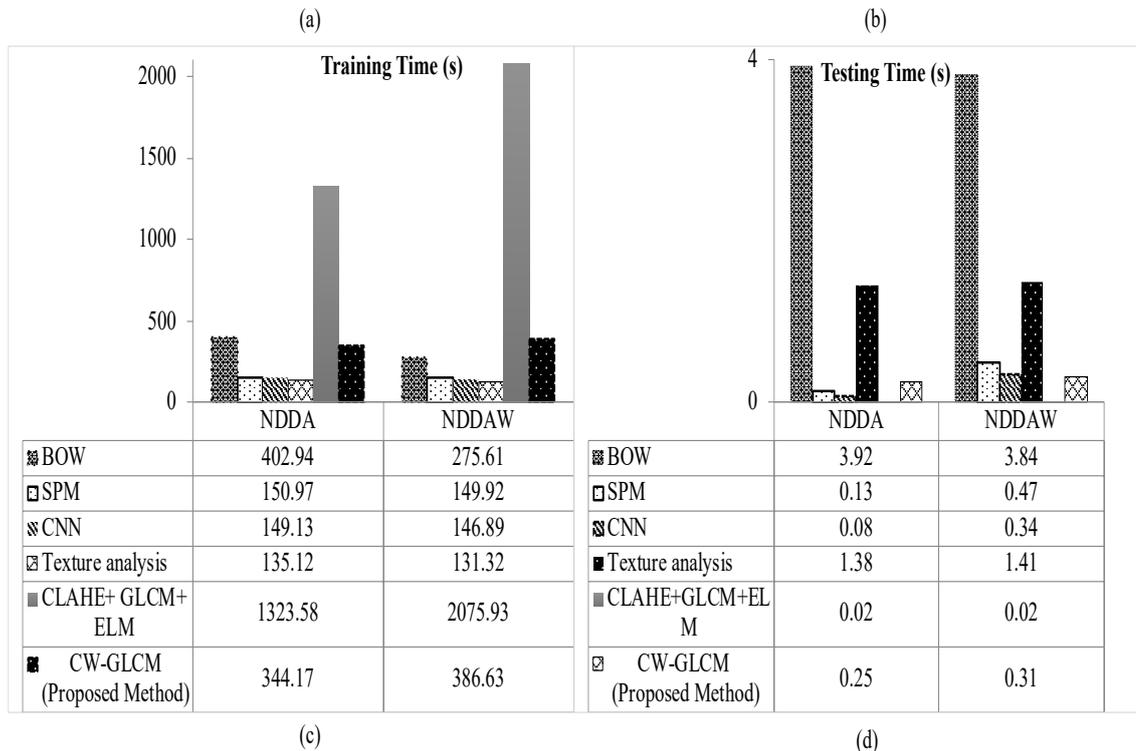
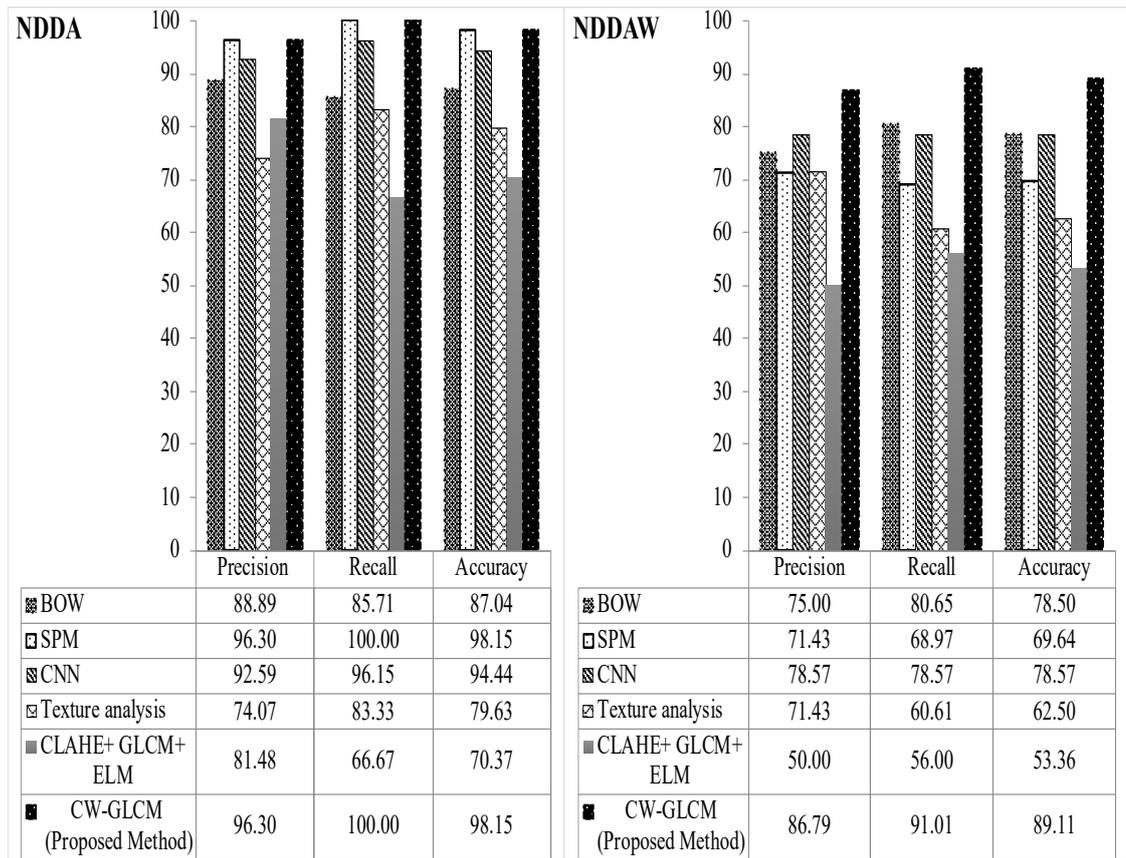


Figure 4.10: Comparative results of precision, recall and accuracy in (a) NDDA dataset (b) NDDAW dataset (c) training time and (d) testing time for BOW (Csurka et al., 2004), SPM (Lazebnik et al., 2006), CNN (dos Santos Ferreira et al., 2017), Texture analysis (Olaniyi et al., 2017), CLAHE+GLCM+ELM (W. Li et al., 2019) and proposed CW-GLCM method.

4.3.4 Performance Evaluation with Different Image Resolution

To evaluate the efficiency of the proposed method against different image resolution, the test with three different image resolution (i.e. original, small and large) is conducted. The small and large images are created by rescaling them with two parameters, 0.5 and 1.5. The comparative results of their precision, recall, accuracy, training and testing time are presented in Table 4.6.

Table 4.6: Comparative results (precision, recall, accuracy, training time and testing time) of the proposed method for different resolution images.

Rescale	Resolution (Pixel)	Precision (%)	Recall (%)	Accuracy (%)	Training Time (s)	Testing Time (s)
0.5	450 × 350	96.30	100.00	98.15	102.76	0.12
Original	900 × 700	96.30	100.00	98.15	344.15	0.25
1.5	1350 × 1050	96.30	100.00	98.15	550.74	0.31

From the results, it can be seen that the performance of the proposed CW-GLCM is not affected by the resolution change. The only difference observed is in the computational time. The time taken for training and testing is getting higher as the number of pixels increased. The results proved that the image resolution does not influence the precision, recall and accuracy of the proposed CW-GLCM. Although the training and testing time are increased with the increment of image resolution, the 10-fold cross-validation experiment on the dataset shows that the proposed method are able to process 27 images within 0.25 s during testing on the original resolution (900 × 700 pixel). These results indicate that the proposed CW-GLCM method can be used in real-time systems.

4.4 Discussion

Overall, the proposed CW-GLCM method outperformed others in detecting important features on the low-quality apple image region. The proposed method performance exceeds 86.79% for all the performance measures in both datasets tested. In contrast, a lower precision, recall and accuracy are observed in the other five methods on NDDAW dataset with the maximum recall of 80.65% in BOW method. The classification accuracy of BOW in NDDAW dataset is 78.50%, SPM 69.64%, CNN 78.57%, Texture analysis 62.50% and CLAHE+GLCM+ELM 53.36%. The lower classification accuracy of the BOW, SPM, Texture analysis and CLAHE+GLCM+ELM methods is influenced by the presence of low-quality region on the apple images in NDDAW dataset. While the reasons that reduce the classification accuracy of the CNN method is due to the small sample dataset utilized in the experiment. The CNN deep learning method requires a large number of images for training in order to obtain a desired classification accuracy result (X. Cheng et al., 2017; He et al., 2015; Krizhevsky et al., 2012; Z. Xiao et al., 2019).

In contrast, the proposed CW-GLCM are able to achieve more than 86.79% for precision, 91.01% recall and 89.11% accuracy for both datasets tested. This indicates that the introduction of Curvelet features and Wavelet Coefficient in the GLCM method can improve the results even with low quality region images. This is possible since the Curvelet and Wavelet transform able to enhance the apple images, especially on the low-quality region. However, the misclassification can still be observed on the defective apple that had been misclassified as non-defective apple. The example of the false positive classification in which defective apple incorrectly classified as non-defective apple is shown in Figure 4.11. This defect region may be misclassified as stem ends or calyxes which are the natural parts of the apple that located at the top and bottom of the apple. This is due to similarities exist between these features.

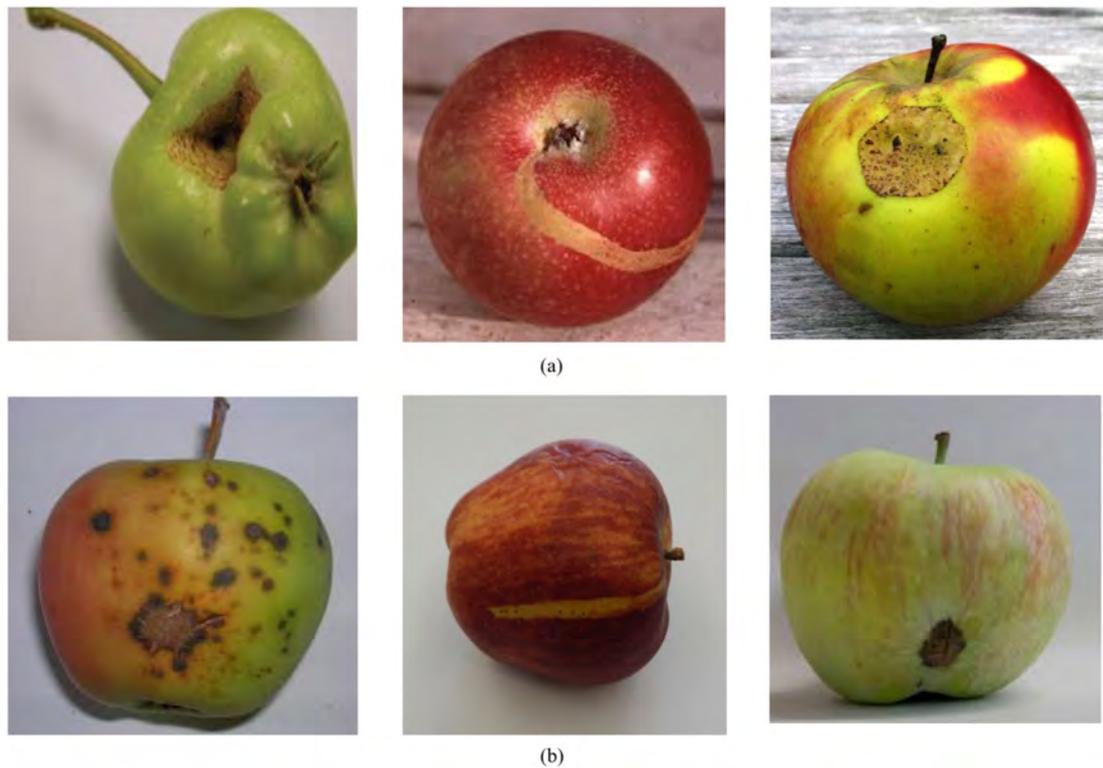


Figure 4.11: Examples misclassification images of the proposed CW-GLCM on (a) NDDA and (b) NDDAW².

4.5 Summary

This chapter discusses and present the contribution of the CW-GLCM method in detecting on low-quality region in the images. The proposed methods incorporate the Curvelet and Wavelet Transform with the GLCM method to improve its ability in detecting on the low-quality region in the apple images. In apple classification, it is crucial to detect these features to enable the classifier to differentiate and classify between defective and non-defective. Comparative experiments are performed between the proposed CW-GLCM method with other five existing methods namely BOW (Csurka et al., 2004), SPM (Lazebnik et al., 2006), CNN (dos Santos Ferreira et al., 2017), Texture analysis (Olaniyi et al., 2017) and CLAHE+GLCM+ELM (W. Li et al., 2019).

Experimental results show that the proposed CW-GLCM and SPM method attained higher classification accuracy in NDDA dataset with the same precision (96.30%), recall (100%) and accuracy (98.15%). However, lower classification accuracy is observed in SPM method when tested with NDDAW dataset with 71.43% precision, 68.97% recall and 69.64% accuracy. In contrast, the proposed CW-GLCM able to achieve 86.79% precision, 91.01% recall and 89.11% accuracy. In comparison with other methods, the proposed method presents the highest precision, recall and accuracy results in both datasets tested. Though the proposed CW-GLCM method is superior to other existing methods, more robust and achieves more accurate classification in both datasets tested, the method was shown to be less effective in detecting defective apples as shown in section 4.4. This is due to similarities exist between the detect region features with the stem ends or calyxes which are the natural parts of the apple.

CHAPTER 5: F-SPM METHOD

This chapter presents the second proposed method, namely F-SPM. The F-SPM concentrates on improving the drawbacks of the proposed CW-GLCM method in detecting defective apples of binary-class classification between defective and non-defective apple images. The capability of the F-SPM is also extended to the multi-class classification between types of defects. The F-SPM method is based on the SPM method (Lazebnik et al., 2006) which includes the spatial layout information. The spatial layout information is important to discriminate the defect with the natural parts of the non-defective apple. However, the existing SPM method generates a large number of unnecessary and redundant high dimensionality features (Chanti & Caplier, 2018; Lin et al., 2016; Penatti et al., 2014). These irrelevant features can reduce the stability and performance of the method. Therefore, the F-SPM method focuses on reducing unnecessary SPM features by detecting significant features of visual-dictionary through Fuzzy logic. To evaluate the performance of the F-SPM method, three datasets of apple images are considered. Two of the datasets are for evaluating the binary-class classification of defective and non-defective apple images, which are NDDA and NDDAW. The proposed methods of F-SPM and CW-GLCM are also evaluated on DA dataset for multi-class classification between types of defects. This test allows the recognition of the specific defective type on the apple images. The above works are organized as follows: the first section (section 5.1) explains the importance of the spatial layout information in apple image. The proposed F-SPM method is described in section 5.2. The performance evaluation and discussion are presented in Section 5.3 and Section 5.4. A summary of this chapter is provided in Section 5.5.

5.1 Introduction

The spatial layout information is a very important element to improve the classification between the defective and non-defective apple images. Specifically, the stem ends or calyxes that are the natural parts of the apple may be misclassified as defects due to similar visual patterns (L. Jiang, Zhu, Cheng, Luo, & Tao, 2009; B. Zhang et al., 2014) as shown in Figure 5.1. The stem ends and calyxes that are the natural parts of the apple are consistently located at the top and bottom of the apple with a consistent size. In opposite, the defects are located at various positions on the apple with varying sizes. Therefore, the spatial layout information can be the key in resolving this issue because the stem end, calyx or defects can yield a similar visual pattern but in a different spatial arrangement.

The feature with spatial layout information is included in the SPM method (Lazebnik et al., 2006). The SPM is a visual-dictionary based method that adds spatial information in the unstructured BOW model to improve image representation and better at distinguishing objects. For this reason, a new image recognition method of F-SPM for apple classification is proposed based on the SPM method to include spatial layout information. However, the SPM method generates a large number of unnecessary and redundant high dimensionality features (Penatti et al., 2014). These irrelevant features can reduce the stability and performance of the method. Therefore, the proposed F-SPM method introduces Fuzzy logic detection to select only reliable and significant features for apple classification. This is to increase the accuracy performance in apple classification. The Fuzzy logic detection detects the edges or contours to highlight the high-frequency components. The detection is performed by comparing the intensity between the neighboring pixels. The spatial layout information is obtained by subdividing the images into finer regions according to spatial pyramid representation of SPM. Then, the bag of features is computed within each of the sub-regions.

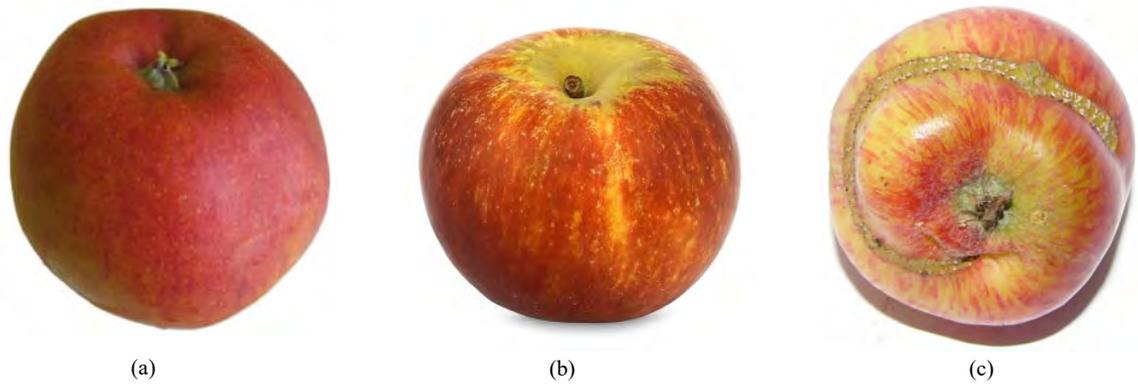


Figure 5.1: (a) Calyx (b) stem end (c) defect¹.

5.2 F-SPM Method

This section describes the stages involved in the proposed F-SPM method. The proposed F-SPM method focuses on addressing the SPM method that generates a large number of unnecessary and redundant high dimensionality features that affect the stability and performance of the method. In the SPM method, the feature vectors are extracted from the image using Dense regular grid by repeatedly subdividing the image into regular grids. Then, histograms of local features are computed for each regular grid. This generates a large number of feature vectors with high dimensionality where some of them are irrelevant for the classification of apple images. Therefore, the proposed F-SPM method is introduced to address this issue by selecting only significant features for classification. The proposed F-SPM method consists of two main phases, namely, feature extraction and feature classification as outlined in Figure 5.2.

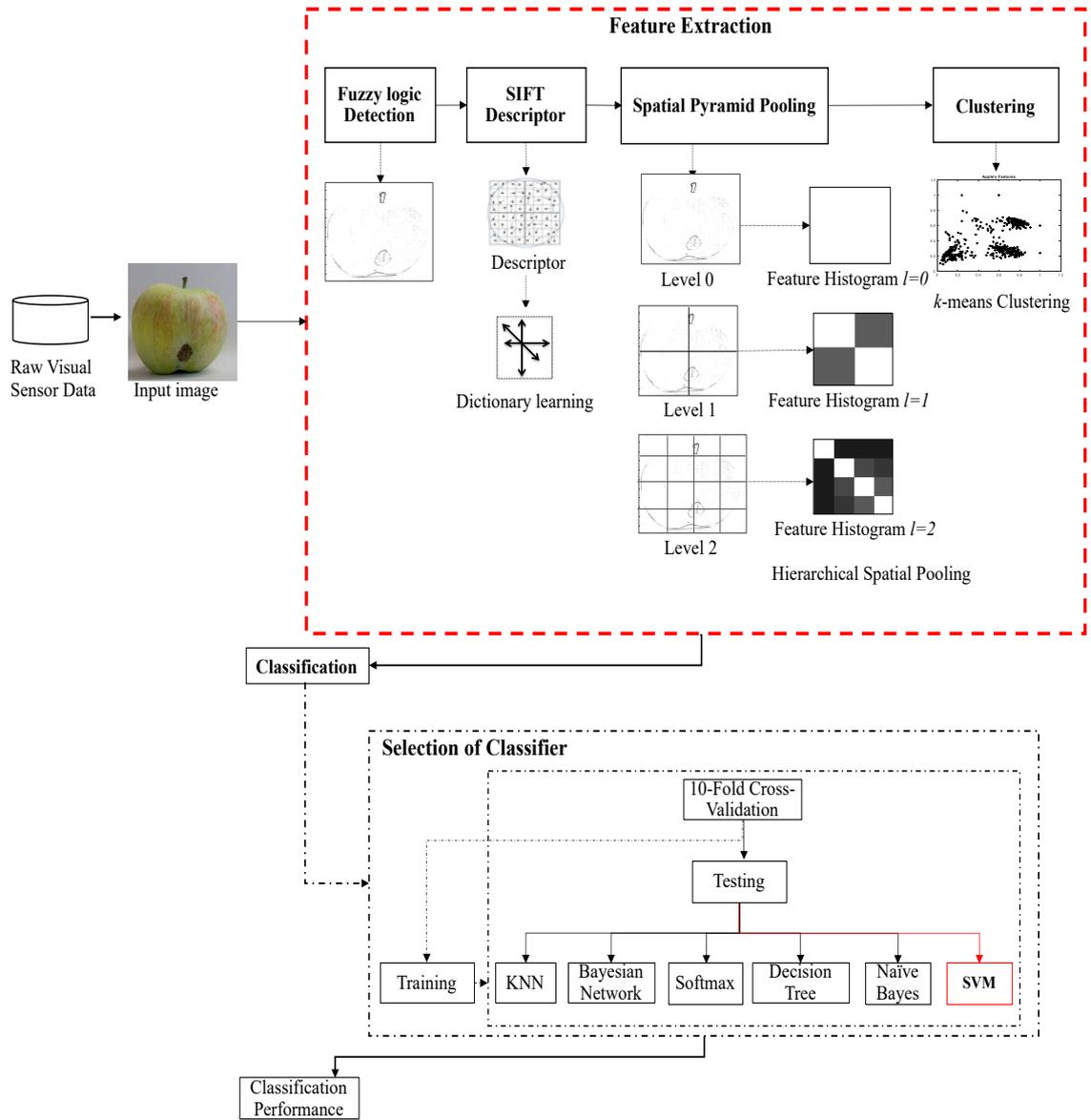


Figure 5.2: The proposed F-SPM method for apple classification.

5.2.1 Feature Extraction

The feature extraction phase selects the significant features from apple image in which, the key contribution of this research. The feature extraction phase is composed of three major methods. These methods are SPM, SIFT and Fuzzy logic. The SPM method is employed in the proposed F-SPM method to include the spatial layout information of the features while the SIFT method computes the descriptors. The Fuzzy logic detection is introduced in the proposed F-SPM method to reduce irrelevant SPM features and only

selects the significant features for classification. The proposed F-SPM method also modifies the existing SPM method by computing the SIFT descriptor at each location of the detected edginess from Fuzzy logic. Also, the normalization step in the SIFT descriptor is skipped to deal with the low-quality region image. The feature extraction phase of the proposed F-SPM method consists of four main stages; Fuzzy logic detection, SIFT descriptor, spatial pyramid pooling and clustering. These stages are discussed in the following subsections.

5.2.1.1 Fuzzy Logic Detection

The Fuzzy logic is introduced in the proposed F-SPM methods to select significant features of high-frequency components in the image. The Fuzzy logic detects the abrupt or sudden changes of any characteristic including the alteration in the texture, color, shades or light absorption in the images at the pixel level (Haq, Anwar, Shah, Khan, & Shah, 2015). The abrupt changes can indicate the defect region on the apple skin when “edges” or “contours” are formed by comparing the intensity between the neighboring pixels. However, in the apple image, the small intensity difference between two neighboring pixels might represents unobvious defect or a low-quality region such as bright or flecks and shading effect. For this reason, the proposed F-SPM method used the Fuzzy logic instead of other edge detection because of the flexibility in defining the degree of membership function for a pixel either belongs to an edge or a uniform non-defective region (Wright & Marwala, 2008). This is because in apple images, the uniform regions of the apple skin are not crisply defined, small intensity differences between two neighboring pixels do not always represent an edge of the defective region. The Fuzzy logic detection relies on the image gradient to locate the breaks in the uniform region as shown in Figure 5.3. The membership function in Fuzzy logic captures a curve of any

shape that maps the input space variable to a number between 0 and 1 that represent the degree of a specific input variable belongs to an edge or a uniform region. The Gaussian membership function is used and the value of the edginess strength for each pixel in the apple image is calculated using three (3) 3x3 linear spatial filters of low-pass, high-pass and edge enhancement filters. These filters are performed through a spatial convolution process with scaling factor 255 along the x -axis and y -axis. Nine convolution coefficients (convolution mask) are defined and labeled to form the 3x3 kernel convolution as shown in Figure 5.4. The associated membership functions with the input and output of Fuzzy logic are shown in Figure 5.5 and Figure 5.6 respectively.

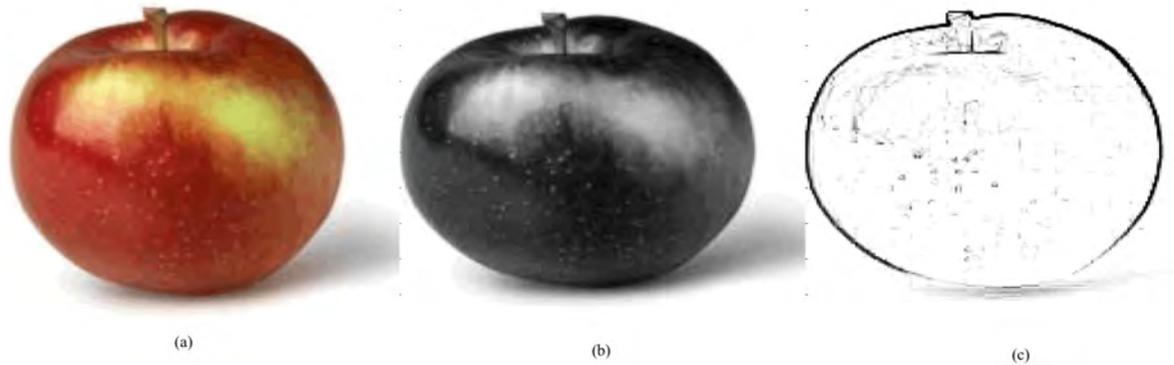


Figure 5.3: Fuzzy logic detection (a) original image (b) grayscale (c) Fuzzy logic.

$$h_{LP} = \begin{bmatrix} \frac{1}{9} & \frac{1}{9} & \frac{1}{9} \\ \frac{1}{9} & \frac{1}{9} & \frac{1}{9} \\ \frac{1}{9} & \frac{1}{9} & \frac{1}{9} \end{bmatrix}, \quad h_{HP} = \begin{bmatrix} -1 & -1 & -1 \\ -1 & 9 & -1 \\ -1 & -1 & -1 \end{bmatrix}$$

$$h_x = \begin{bmatrix} -1 & 0 & +1 \\ -2 & 0 & +2 \\ -1 & 0 & +1 \end{bmatrix}, \quad h_y = \begin{bmatrix} +1 & +2 & +1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix}$$

Figure 5.4: 3x3 Kernel used for Fuzzy logic detection.

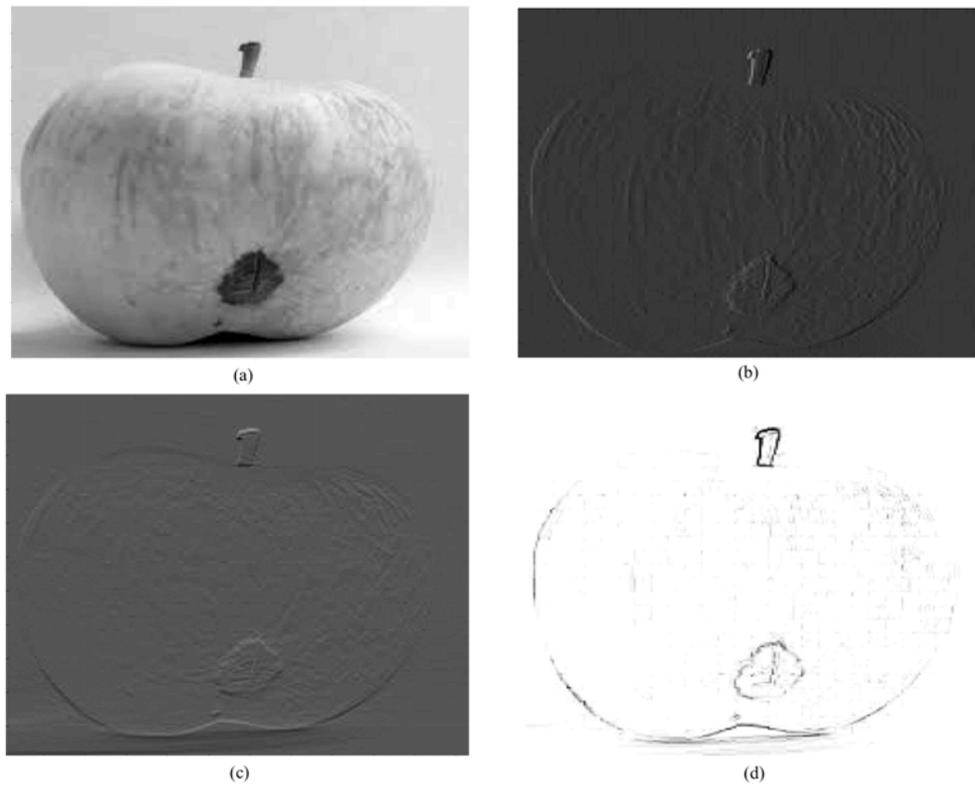


Figure 5.5: (a) Grayscale image (b) I_x (c) I_y (d) Fuzzy logic.

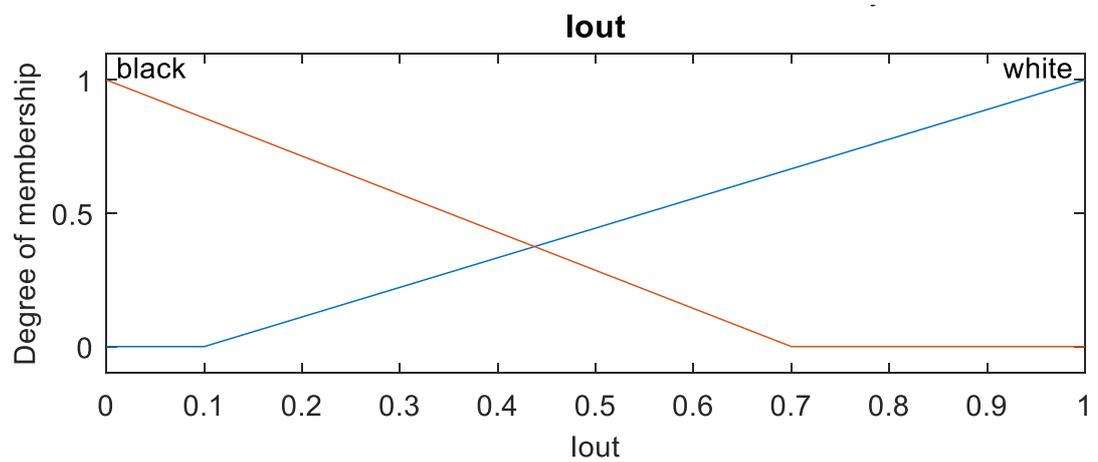


Figure 5.6: Membership functions of the inputs/outputs of edge.

5.2.1.2 SIFT Descriptor

In the second phase, the SIFT descriptor is computed for the detected edges. The SIFT descriptor is chosen due to high repeatability and accuracy which improves the chances of the correct match in the large features database. The SIFT descriptor is computed from the patches of 16×16 pixels and overlap by 8 pixels with its neighbors on each edginess detected by Fuzzy logic. The patch is subdivided into 4×4 cells wherein the gradient orientation of 8-bin histogram is computed for each cell. The 8-bin histograms from all cells in a patch are then combined and represented by a vector. This vector contains a total of 128-elements. The SIFT descriptor for an image I can be obtained through feature extraction as expressed in (14).

$$\Phi(I): I \rightarrow \{(\phi_1, x_1, y_1), (\phi_2, x_2, y_2) \dots (\phi_{nI}, x_{nI}, y_{nI})\}, \quad (14)$$

where, ϕ_i is the local image descriptor and (x_i, y_i) is a pixel location of the centre feature of I .

The computed SIFT descriptors are robust to affine distortion, illumination changes, invariant to scale, orientation and rotation changes (Lee et al., 2015; Lowe, 2004; Warif, Wahab, Idris, Salleh, & Othman, 2017). In the proposed F-SPM method, the normalization step in the SIFT descriptor is skipped to deal with the low-quality region when the patch has a weak gradient magnitude. This is to avoid misclassification between defective and non-defective apple.

5.2.1.3 Spatial Pyramid Pooling

The pyramid in the proposed F-SPM method is built using spatial pooling in two-dimensional image space. Spatial pyramid is the multi-level recursive image decomposition. The image is divided into a sequence of grid according to the pyramid level. This approach gives the advantage in maintaining continuity with the “visual

vocabulary” paradigm. The proposed F-SPM method utilized the spatial pooling with two pyramid levels to partition the pyramid level. The utilization of these two pyramid levels ($L = 2$) are accordance to the experimental setup from the previous studies (Harada, Ushiku, Yamashita, & Kuniyoshi, 2011; Lazebnik et al., 2006; Yan et al., 2012; Yan, Xu, Xu, Lin, & Li, 2015). The settings of the existing SPM (Lazebnik et al., 2006) method used $L = 2$ with vocabulary size of $M = 200$ and only tested on small resolution images of about 300×250 pixels. However, the datasets in this research contain high-resolution images of 900×700 pixels. Therefore, the proposed F-SPM method utilized two pyramid levels with the settings of $L = 2, M = 500$ and the patch size of 1500. The schematic illustration for 2 level pyramid representations is shown in Figure 5.7.

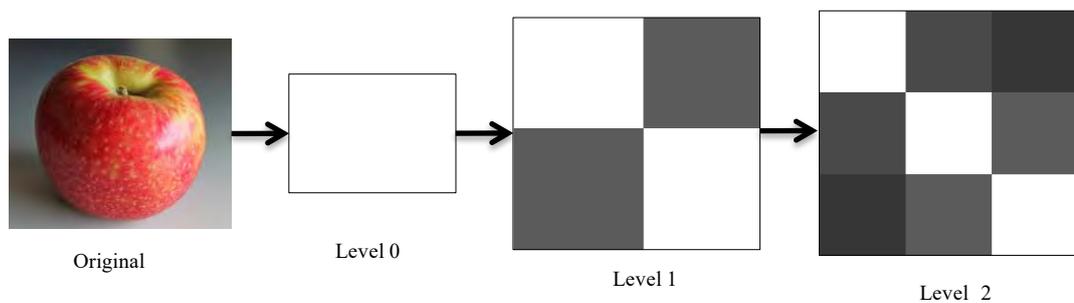


Figure 5.7: Feature histogram of level 2 pyramid.

As depicted in Figure 5.7, the decomposition of the image consists of a single cell at level 0. The image representation at level 0 is equal to the standard BOW method. The image is further divided into four quadrants at level 1 and then nine grid cells at level 2. The feature extracted from the entire grid cell is concatenated forming the feature vector. Each spatial grid cell consists of a feature vector and produces one feature histogram. Thus, each feature vector is the collection of the feature histograms that being computed over grid cells. Specifically, all the feature vectors are quantized into M discrete types. Each of the m channels gives two sets of two-dimensional vectors which is X_m and Y_m .

The X_m and Y_m represent the features of type m coordinates found in the respective images. Then, the final kernel sums the separate channel kernels as in Equation (15).

$$K_L(X, Y) = \sum_{m=1}^M K_L(X_m, Y_m), \quad (15)$$

5.2.1.4 Clustering

The extracted feature vectors from the prior stage are encoded using the vector quantization technique. This is followed by the creation of the visual codebook using k -means algorithm. The k -means algorithm is a clustering technique (Kanungo et al., 2002; Rahim et al., 2018). The clustering process of the k -means algorithm is performed according to distances measured between the feature to each centre of the k clusters. Initially, the position for the centre of the k clusters are randomly selected. Then, the features are assigned to the respective cluster based on their minimum distance to the initial centre of the cluster. The mean value calculated from the current members of each cluster yields as the new centre. These processes are repeated for 100 iterations to form a codebook. The steps above can be summarized as follows:

Step 1: Assign a value to k , the total clusters.

Step 2: Randomly select the initial centre of k

Step 3: Calculate the centre or mean value of k

Step 4: Determine the distance between each pixel from centre

Step 5: Assign the cluster to the pixels based on the minimum distance to a particular centre of the cluster. Re-estimate the position of the centre by calculating the mean value from the current members.

Then, the features are coded into the codebook using coding function defined in (16).

$$g((\phi_i, x_i, y_i)) = (\hat{g}(\phi_i), x_i, y_i), \quad (16)$$

This function converts the extracted feature vectors into a coded version of visual-dictionary features in the codebook associated with the original spatial information.

5.2.2 Classifier

The classification phase classifies the visual-dictionary features into binary-class of defective and non-defective apple. The classification also includes the multi-class between types of defects. For these classifications, six classifiers, namely the KNN, Bayesian Network, Softmax, Decision Tree, Naïve Bayes and SVM are employed to find the most suitable classifier for the proposed visual-dictionary features. Their performances are evaluated and a test decision of the best performing classifier on the proposed visual-dictionary features determine the suitable classifier for the proposed method.

5.3 Experimental Result

This section presents the experimental results of the proposed F-SPM method on NDDA, NDDAW and DA datasets. These datasets are used in the experiments to evaluate the capability and effectiveness of the F-SPM method on different aspects of the classification performances. The NDDA dataset evaluates the detection of various defective and non-defective apple images. The NDDAW dataset consists of 159 images to evaluate the classification capability on apple images with low-quality region on its

skin. The DA dataset is composed of five types of defective apples to evaluate the multi-class classification between types of defects.

The results are analyzed to select the most significant features for apple classification and address the SPM method limitation of generating a large number of unnecessary and redundant high dimensionality features. The results shown in this section prove the success of the proposed F-SPM method in selecting the significant features for binary-class classification of defective and non-defective apple images including low-quality region images and multi-class classification between types of defects.

5.3.1 Performance Evaluation

This section compares the performances of the SPM method between two feature detectors, which are Dense regular grid and keypoint detection. The Dense regular grid is originally employed in the SPM method while for keypoint detection, the SIFT detector is employed due to high repeatability and accuracy (Lee et al., 2015). The comparison of the SPM method between two feature detectors are performed on the NDDA dataset to detect various defective and non-defective apple types. The evaluation results are summarized in Table 5.1.

Table 5.1: Comparison of SPM method performances between two feature detectors on NDDA dataset.

Feature Detector	Precision (%)	Recall (%)	Accuracy (%)	Training Time (s)	Testing Time (s)
Dense regular grid	100.00	96.30	98.15	150.97	0.13
<i>keypoint (SIFT)</i>	92.59	100.00	96.30	833.78	0.54

The overall performance of the SPM method using Dense regular grid shows higher accuracy and faster than the keypoint detection for classification of defective and non-

defective apple images. The Dense regular grid recorded the highest precision (98.15%) and accuracy (100%) while its recall rate is 3.7% less than the keypoint detection. In terms of computational time, the Dense regular grid is also the fastest during training (150.97 s) and testing (0.13 s). In opposite, a lower precision (92.59%) and accuracy (96.30%) with longer training (833.78 s) and testing time (0.54 s) are observed in keypoint detection. This high computational result is expected since the keypoint detection is high computational time (Lee et al., 2015). The keypoint detection requires a longer time to detect the salient keypoints in each of the image but many of the detected keypoints are insignificant for the recognition (Lin et al., 2016). Thus, reduces the classification accuracy of the method. The higher accuracy is observed on the Dense regular grid as the method extracts features from all over the image on a regular grid. This direct approach allows the uniform region in apple images been effectively captured compared to the keypoint detection. The example of these two feature detection technique for apple image classification as shown in Figure 5.8 to clearly illustrate the difference between the two techniques. Figure 5.8 (a) show the example of feature detector using Dense regular grid, whereas the example of *keypoint* (SIFT) feature detector is shown in Figure 5.8 (b).

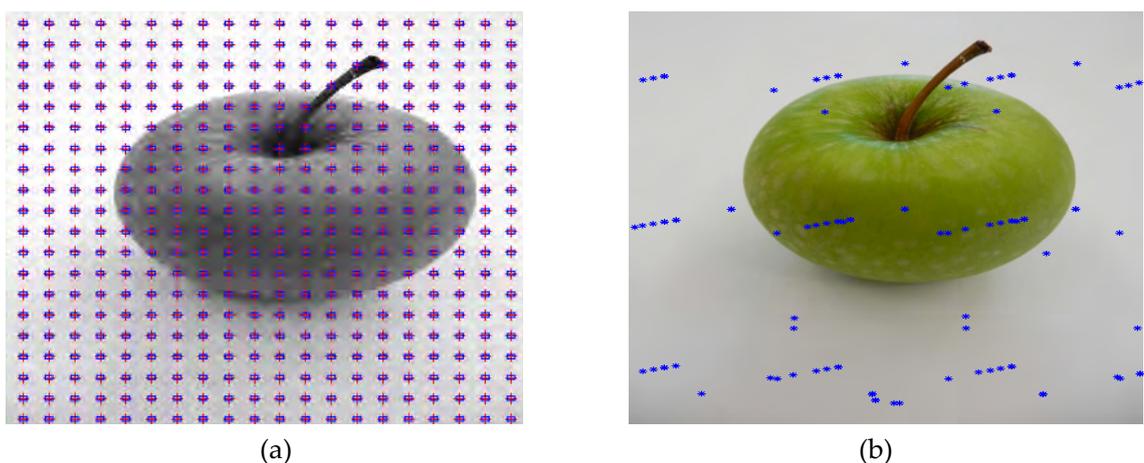


Figure 5.8: Example of two feature detection techniques for apple image classification (a) feature detector using Dense regular grid (b) feature detector using *keypoint* (SIFT).

The findings presented in this section justifies the need for a reliable method to detect significant features and reduce the unnecessary features of the SPM method without decreasing its classification accuracy.

5.3.2 Performance Measure for Edge Detection

As highlighted in section 5.2, the F-SPM method focuses on reducing the unnecessary SPM features through selection of significant features for further classification. The selection of the significant features begins with edges or contours detection to highlight high-frequency components of the defective region and neglect the unnecessary features in the background region of the apple image. The detection of the edges or contours is performed by comparing the intensity between the neighboring pixels. To search the best edge detection technique for the proposed method, the Fuzzy logic is compared with two established edge detection techniques of Canny and Sobel. Their performances in the modified SPM method were tested on NDDA dataset using the SVM classifier.

Table 5.2: Comparison of edge detection techniques in modified SPM method on NDDA dataset.

Edge Detection	Precision (%)	Recall (%)	Accuracy (%)	Training Time (s)	Testing Time (s)
Canny	89.29	100.00	94.55	700.95	0.20
Sobel	88.89	100.00	94.44	753.70	0.44
Fuzzy logic (Proposed)	96.30	100.00	98.15	658.07	0.46

The Fuzzy logic detection obtained the highest precision (96.30%) and accuracy (98.15%) with the fastest training time (658.07 s) compared to Canny and Sobel techniques as shown in Table 5.2. These performances demonstrate the capability of the Fuzzy logic detection to effectively capture the abrupt or sudden changes of texture, color, shades or light absorption in the images characteristic at the pixel level (Haq et al., 2015).

The Fuzzy logic detection relies on the image gradient to locate the breaks and flexible in defining the membership of the pixel either belong to an edge or a uniform region (Wright & Marwala, 2008). This is an important aspect in apple classification because the small intensity difference between two neighboring pixels can indicate an edge of the defective region as well as a low-quality region such as bright or flecks on the apple skin. On the other hand, the ability of Canny and Sobel techniques are limited in dealing with noises in the image which lead to low classification accuracy (J. Song et al., 2019). The presented results show that the Fuzzy logic detection can reduce the unnecessary SPM features without degrading its classification accuracy. Thus, the Fuzzy logic detection is employed in the design of the proposed F-SPM method to extract visual-dictionary features.

5.3.3 Comparison of Different Classifier

To select the most suitable classifier for the proposed visual-dictionary features, six classifiers were tested on NDDA dataset. These classifiers are KNN, Bayesian Network, Softmax, Decision Tree, Naïve Bayes and SVM. The results for each classifier are presented in Table 5.3 and depicted in Figure 5.9.

Table 5.3: Comparison of the proposed visual-dictionary features with different classifiers on NDDA dataset.

Classifier	Precision (%)	Recall (%)	Accuracy (%)	Training Time (s)	Testing Time (s)
KNN	96.30	74.29	81.41	674.81	37.44
Bayesian Network	88.89	96.00	92.59	674.82	0.01
Softmax	100.00	90.00	94.44	674.43	2.31
Decision Tree	88.89	100.00	94.44	675.52	0.72
Naïve Bayes	92.59	100.00	96.30	675.74	0.11
SVM	96.30	100.00	98.15	658.07	0.46

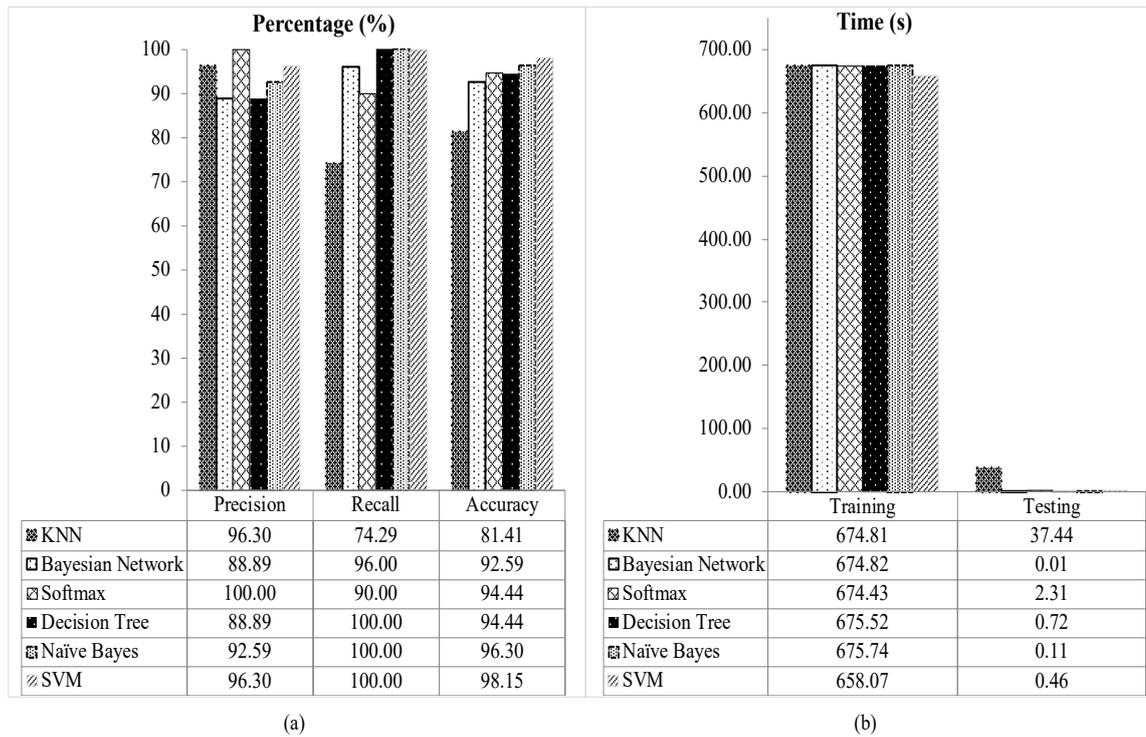


Figure 5.9: Comparative results of the proposed visual-dictionary features with different classifier on NDDA dataset (a) precision, recall and accuracy (b) training and testing time.

From the evaluation, the combination of the proposed visual-dictionary features with SVM classifier achieved the best performance for recall (100%) and accuracy (98.15%) including the fastest training time (658.07 s). This is because the SVM classifier is more suitable in classifying a small sample data with high feature vectors dimensionality (M. Xiao, Jiang, Li, Xie, & Yi, 2017; J. H. Zhang, Meng, Wang, & Hou, 2014) compared to the others. Furthermore, the high dimensionality of the feature vectors has a low impact on the effectiveness of the SVM classifier (Penatti et al., 2014) For this reason, the SVM classifier is a popular classifier for visual-dictionary based method that generates high dimensionality of feature vectors (Lin et al., 2016). The SVM classifier is a complex algorithm but its complexity remains unaffected by high dimensionality features. Other than that, the SVM classifier has a good generalization ability without depending on the size of the data (Petitjean et al., 2018). This prevents the classifier from overfitting on a

small sample data. Additionally, the SVM classifier does not require the prior probability distribution of the classes and data to be classified. Thus, contributes to the success of the classification on the proposed visual-dictionary feature in this experiment.

In contrast, the KNN classifier obtained the lowest performance for accuracy (81.41%) and recall (74.29%). The KNN classifier determines the classes by calculating distance between the data based on the simple vote majority system (Duneja & Puyalnithi, 2017). However, the distance of the data from different classes can be similar (Syaliman et al., 2018), thus, increases the misclassification. Also, the performance of the KNN classifier is adversely affected by noise, irrelevant features, unstable in high dimensionality features and small sample data (Hinneburg, Aggarwal, & Keim, 2000; Singh et al., 2016). The lowest precision is observed in Bayesian Network and Decision Tree classifier with both obtained 88.89%. Although the Decision Tree is one of the successful classifiers for most cases, the performance of this classifier is affected when dealing with high dimensionality data (Do, Lenca, Lallich, & Pham, 2010). The high dimensionality data tends to overfit each of the node, which affecting the classification performance (Singh et al., 2016). The low precision in the Bayesian Network is due to its less sensitive to noise (Druzdzal & Onisko, 2008; Onisko & Druzdzal, 2003; Onisko & Druzdzal, 2013). Furthermore, the performance of the Bayesian Network is affected by high dimensional data in this experiment. The Bayesian Network is ineffective in handling high dimensionality data (Singh et al., 2016). A large network of the Bayesian Network did not feasible in time and space (Petitjean et al., 2018). Another issue that affecting the precision of the Bayesian Network is its network structure which is based on their bias-variance trade-off that is more suitable to be used for a large dataset. The large dataset can provide sufficient examples for the Bayesian Network classifier to learn with precision and precisely capture the fine details in the data (Petitjean et al., 2018). However, the dataset with a limited

number of images (550 images) was utilized in this experiment. This reduces the precision of the Bayesian Network classifier.

In terms of computational time, the SVM classifier is the fastest classifier during training (658.07 s) whereas the Bayesian Network is the fastest during testing (0.01 s). The SVM classifier is the fastest during training because it does not require the prior probability distribution of the classes and data to be classified. This also due to the data scaling in the SVM package that avoids the attribute with the greater numeric range dominating a smaller one (Hsu, Chang, & Lin, 2003); which reduces the time complexity. Among the classifiers, the Naïve Bayes takes the longest time for training (675.74 s), whereas the KNN for testing (37.44 s). The Naïve Bayes classifier is based on probabilistic that requires the knowledge of prior probability distribution of the class and data to be classified. This increases the training time in Naïve Bayes classifier. The KNN classifier is computationally intensive as it stores all the training data and compares the extracted features on the test images with each training data for classification (Syaliman et al., 2018). Finally, the experimental results show that the SVM classifier is the most suitable and effective classifier for the proposed F-SPM method.

5.3.4 Effectiveness of the Proposed F-SPM Method

The proposed F-SPM method is evaluated on NDDA, NDDAW and DA datasets for binary-class and multi-class classifications. The NDDA and NDDAW datasets are intended for the evaluation of the binary-class classification. These datasets include a total of 1110 defective and non-defective apple images. However, the NDDAW dataset contains more low-quality region images compared to the NDDA dataset. For the evaluation of the multi-class classification, the DA dataset is employed containing a total of 200 apple images with five types of defects.

The performances of the F-SPM method on NDDA, NDDAW and DA datasets are compared with five existing methods. These methods are BOW (Csurka et al., 2004), SPM (Lazebnik et al., 2006), CNN (dos Santos Ferreira et al., 2017), Texture analysis (Olaniyi et al., 2017) and CLAHE+GLCM+ELM (W. Li et al., 2019). Also, the proposed CW-GLCM method from chapter 4 which improved the GLCM Texture analysis is included in the comparison. The performances of all methods are compared in terms of precision, recall, accuracy and computational time using 10-fold cross-validation.

5.3.4.1 Binary-Class Classification Performance

The evaluation of the binary-class classification is divided into two experiments. The first experiment of the binary-class classification is performed on the NDDA dataset. In the first experiment, the CW-GLCM, F-SPM and SPM methods recorded similar performances in precision (96.30%), recall (100%) and accuracy (98.15%) as presented in Table 5.4. The accuracy of the CW-GLCM, F-SPM and SPM methods are also the highest compared to other methods. This is followed by the accuracy of the CNN (94.44%), BOW (87.04%), Texture analysis (79.63%) and CLAHE+GLCM+ELM (70.37%).

Table 5.4: Comparison of confusion matrix for BOW (Csurka et al., 2004), SPM (Lazebnik et al., 2006), CNN (dos Santos Ferreira et al., 2017), Texture analysis (Olaniyi et al., 2017), CLAHE+GLCM+ELM (W. Li et al., 2019), CW-GLCM (Proposed Method I) and F-SPM (Proposed Method II) on NDDA dataset: Defective (D) and Non-defective (N).

Class	Methods													
	BOW		SPM		CNN		Texture analysis		CLAHE + GLCM+ ELM		CW-GLCM (Proposed Method I)		F-SPM (Proposed Method II)	
	D	N	D	N	D	N	D	N	D	N	D	N	D	N
D	24	3	26	1	25	2	20	7	22	5	26	1	26	1
N	4	23	0	27	1	26	4	23	11	16	0	27	0	27
Precision (%)	88.89		96.30		92.59		74.07		81.48		96.30		96.30	
Recall (%)	85.71		100.00		96.15		83.33		66.67		100.00		100.00	
Accuracy (%)	87.04		98.15		94.44		79.63		70.37		98.15		98.15	
Training Time (s)	402.94		150.97		149.13		135.12		1323.58		344.17		658.07	
Testing Time (s)	3.92		0.13		0.08		1.38		0.02		0.25		0.46	

The second experiment focuses on the challenge of the low-quality region images in the binary-class classification. The experiment is performed on the NDDAW dataset using a 10-fold cross-validation approach. To implement the 10-fold cross-validation, 280 images from the defective class and 280 images from the non-defective class are divided into ten equal parts. Each part consists of 28 images from the defective and non-defective classes. Nine parts are used for training and one for testing. This process is repeated ten times until each part is used exactly once as a validation set. Then, the average value for all ten experiments is taken and presented in Table 5.5.

Table 5.5: Comparison confusion matrix for BOW (Csurka et al., 2004), SPM (Lazebnik et al., 2006), CNN (dos Santos Ferreira et al., 2017), Texture analysis (Olaniyi et al., 2017), CLAHE+GLCM+ELM (W. Li et al., 2019), CW-GLCM (Proposed Method I) and F-SPM (Proposed Method II) on NDDAW dataset: Defective (D) and Non-defective (N).

	Methods													
	BOW		SPM		CNN		Texture analysis		CLAHE + GLCM +ELM		CW-GLCM (Proposed Method I)		F-SPM (Proposed Method II)	
Class	D	N	D	N	D	N	D	N	D	N	D	N	D	N
D	21	7	20	8	22	6	20	8	14	14	24	4	28	0
N	5	23	9	19	6	22	13	15	11	17	2	26	5	23
Precision (%)	75.00		71.43		78.57		71.43		50.00		86.79		100.00	
Recall (%)	80.65		68.97		78.57		60.61		56.00		91.01		84.85	
Accuracy (%)	78.50		69.64		78.57		62.50		53.36		89.11		91.07	
Training time (s)	275.61		149.92		146.89		131.32		2075.93		386.63		675.23	
Testing Time (s)	3.84		0.47		0.34		1.41		0.02		0.31		0.48	

In this experiment, the classification accuracy of the proposed F-SPM method outperformed the others whereas the CLAHE+GLCM+ELM method recorded the lowest classification accuracy. Specifically, the proposed F-SPM method obtained an accuracy of 91.07% followed by the proposed CW-GLCM (89.11%), CNN (78.57%), BOW (78.50%), SPM (69.64%), Texture analysis (62.50%) and CLAHE+GLCM+ELM (53.36%). The F-SPM method correctly classified all 28 defective apple images followed by the proposed CW-GLCM (24 images), CNN (22 images), BOW (21 images), SPM (20 images), Texture analysis (20 images) and CLAHE+GLCM+ELM (14 images). For non-defective, the F-SPM method correctly classified 23 images out of 28 non-defective apple images while the CW-GLCM (26 images), CNN (22 images), BOW (23 images), SPM (19 images), Texture analysis (16 images) and CLAHE+GLCM+ELM (17 images). Figure 5.10 (a) and Figure 5.10 (b) visualize the comparison of the precision, recall and accuracy for each binary-class dataset of NDDA and NDDAW, respectively. The time recorded during training and testing for each dataset are shown in Figure 5.10 (c) and Figure 5.10 (d).

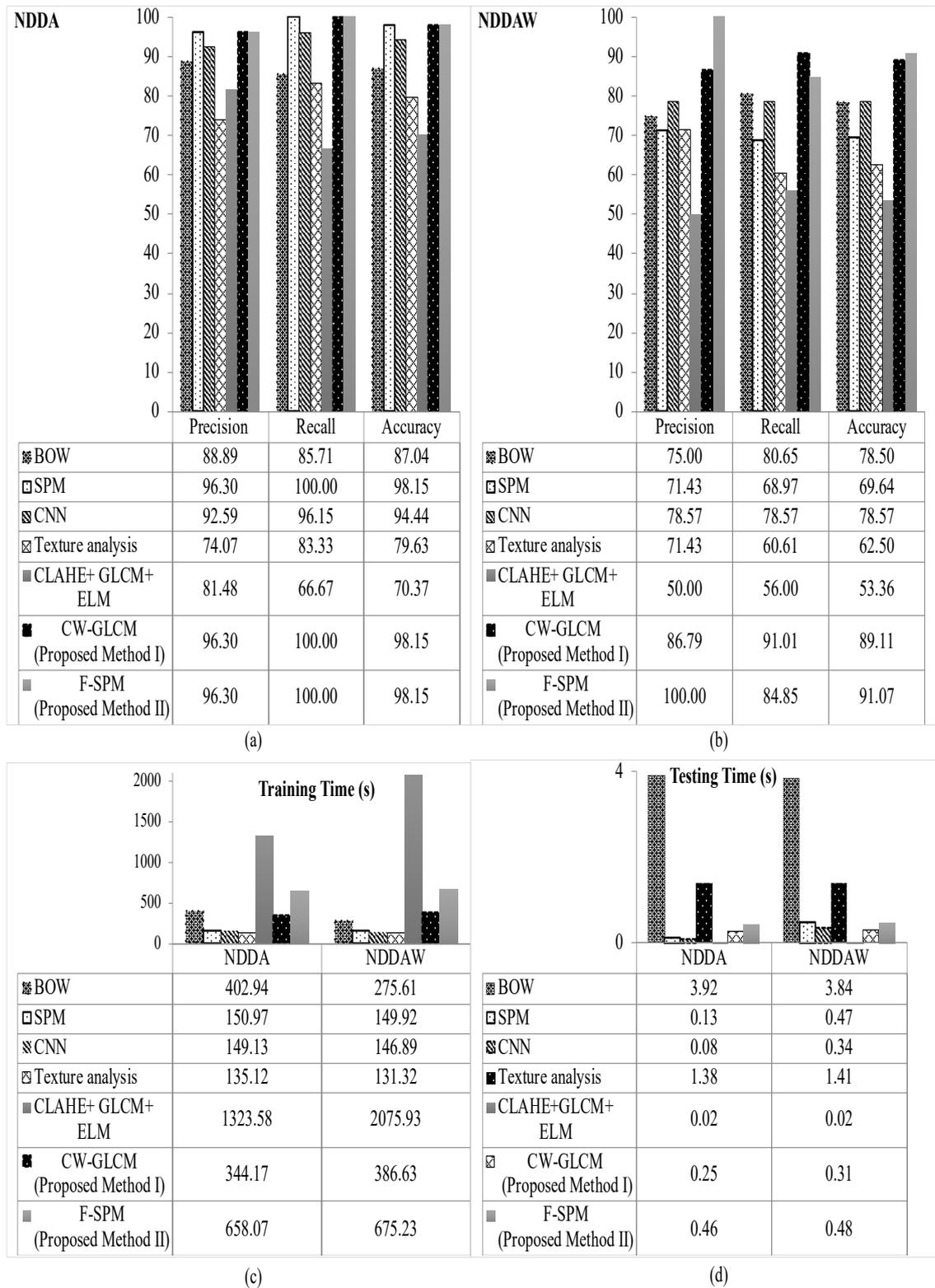


Figure 5.10: Comparative results of precision, recall and accuracy in (a) NDDA dataset (b) NDDAW dataset (c) training time (d) testing time for BOW (Csurka et al., 2004), SPM (Lazebnik et al., 2006), CNN (dos Santos Ferreira et al., 2017), Texture analysis (Olaniyi et al., 2017), CLAHE+GLCM+ELM (W. Li et al., 2019), CW-GLCM (Proposed method I) and F-SPM (Proposed method II).

From Figure 5.10 (a), all the methods achieved the classification accuracy above 70.37% on the NDDA dataset. The proposed methods (CW-GLCM and F-SPM) and the SPM method outperformed the others on NDDA dataset with a similar classification accuracy of 98.15%. This is followed by the CNN with 94.44%, BOW (87.04%), Texture analysis (79.63%) and CLAHE+GLCM+ELM (70.37%). Although the SPM method works effectively on the NDDA dataset, its performance on the NDDAW dataset decreases due to low-quality regions images. Contrarily, the proposed F-SPM method maintained its high performance on the NDDAW dataset and achieved the highest classification accuracy of 91.07% among the methods. This is followed by the proposed CW-GLCM (89.11%), CNN (78.57%), BOW (78.50%), SPM (69.64%), Texture analysis (62.50%) and CLAHE+GLCM+ELM (53.36%).

The high performance of the proposed F-SPM method on the NDDAW dataset demonstrates the ability of the Fuzzy logic detection to select significant features even in the low-quality region images. This is possible since the Fuzzy logic detection capable of capturing the abrupt or sudden change of any alteration in texture, color, shades or light absorption characteristic in the images at the pixel level. On the other hand, the proposed CW-GLCM method achieved the highest recall (91.01%) on the NDDAW dataset as depicted in Figure 5.10 (b). However, the misclassification can still be observed on the defective apple that had been misclassified as non-defective apple as highlighted in section 4.4. The misclassification occurs due to similarity between the defective region and the stem ends or calyxes which are the natural parts of the apple. In the proposed F-SPM method, the spatial layout information and Fuzzy logic detection of the significant features are included which helps in distinguishing between the defect and the stem end or calyx. This is demonstrated in the performance of the proposed F-SPM method on the NDDAW dataset where all 28 defective images from the 10-fold cross-validation are successfully classified as summarized in Table 5.5.

In terms of training and testing time, the CLAHE+GLCM+ELM recorded the longest training time and the fastest testing time on both datasets. The CLAHE+GLCM+ELM method required a long time to train due to the computationally extensive of the CLAHE approach but fast testing time due to extremely fast learning speed of the ELM classifier used in the method (W. Li et al., 2019). For the proposed methods, the training and testing time of the CW-GLCM are slightly faster than the F-SPM on both datasets. The F-SPM method processed all 28 images on NDDAW dataset within 0.48 s during testing which constitute less than 0.02 s per image. This indicates that the CW-GLCM and F-SPM methods have the potential to be implemented in real-time system.

5.3.4.2 Performance on Multi-Class Classification

The previous section evaluates the performance of the proposed F-SPM method for binary-class of defective and non-defective apple including low-quality region images. In this section, the performances of the proposed methods is further evaluate on the multi-class classification. The multi-class classification allows the recognition of the specific type of defect. The experiments are performed on the DA dataset consisting five types of defects. The defects are Blotch, Bruise, Cork Spot, Scab and Rot. The similar experimental setups including the 10-fold cross-validation as in the binary-class classification are repeated in this experiment. The experimental results from the 10-fold cross-validation for all methods are compared in Table 5.6. The graphical presentation for the precision, recall and accuracy in individual class of Blotch, Bruise, Cork Spot, Scab and Rot are also presented in Figure 5.11.

Table 5.6: Comparison results for BOW (Csurka et al., 2004), SPM (Lazebnik et al., 2006), CNN (dos Santos Ferreira et al., 2017), Texture analysis (Olaniyi et al., 2017), CLAHE+GLCM+ELM (W. Li et al., 2019), CW-GLCM (Proposed Method I) and F-SPM method (Proposed Method II) on DA dataset.

Methods					
BOW					
Class	Blotch	Bruise	Cork Spot	Scab	Rot
TP _i (%)	50.00	100.00	50.00	75.00	25.00
FP _i (%)	50.00	75.00	25.00	50.00	0.00
FN _i (%)	50.00	0.00	50.00	25.00	75.00
TN _i (%)	100.00	50.00	75.00	25.00	50.00
Precision _i (%)	50.00	57.14	66.67	60.00	100.00
Recall _i (%)	50.00	100.00	50.00	75.00	25.00
Accuracy _i (%)	60.00	66.67	62.50	57.14	50.00
Average Precision (%)	66.76				
Average Recall (%)	60.00				
Average Accuracy (%)	59.26				
Training Time (s)	143.19				
Testing Time (s)	1.33				
SPM					
Class	Blotch	Bruise	Cork Spot	Scab	Rot
TP _i (%)	75.00	75.00	75.00	100.00	75.00
FP _i (%)	50.00	0.00	0.00	50.00	0.00
FN _i (%)	25.00	25.00	25.00	0.00	25.00
TN _i (%)	75.00	75.00	100.00	75.00	75.00
Precision _i (%)	60.00	100.00	100.00	66.67	100.00
Recall _i (%)	75.00	75.00	75.00	100.00	75.00
Accuracy _i (%)	66.67	85.71	87.50	77.78	85.71
Average Precision (%)	85.33				
Average Recall (%)	80.00				
Average Accuracy (%)	80.67				
Training Time (s)	223.04				
Testing Time (s)	1.27				
CNN					
Class	Blotch	Bruise	Cork Spot	Scab	Rot
TP _i (%)	50.00	75.00	50.00	25.00	100.00
FP _i (%)	25.00	25.00	75.00	50.00	25.00
FN _i (%)	50.00	25.00	50.00	75.00	0.00
TN _i (%)	75.00	50.00	25.00	100.00	50.00
Precision _i (%)	66.67	75.00	40.00	33.33	80.00
Recall _i (%)	50.00	75.00	50.00	25.00	100.00
Accuracy _i (%)	62.50	71.43	37.50	50.00	85.71
Average Precision (%)	59.00				
Average Recall (%)	60.00				
Average Accuracy (%)	61.43				
Training Time (s)	49.94				
Testing Time (s)	0.27				

Texture analysis					
Class	Blotch	Bruise	Cork Spot	Scab	Rot
TP _i (%)	50.00	100.00	0.00	0.00	0.00
FP _i (%)	75.00	150.00	75.00	25.00	25.00
FN _i (%)	50.00	0.00	100.00	100.00	100.00
TN _i (%)	100.00	0.00	0.00	0.00	50.00
Precision _i (%)	40.00	40.00	0.00	0.00	0.00
Recall _i (%)	50.00	100.00	0.00	0.00	0.00
Accuracy _i (%)	54.55	40.00	0.00	0.00	28.57
Average Precision (%)			16.00		
Average Recall (%)			30.00		
Average Accuracy (%)			24.62		
Training Time (s)			20.16		
Testing Time (s)			0.87		
CLAHE+GLCM+ELM					
Class	Blotch	Bruise	Cork Spot	Scab	Rot
TP _I (%)	100.00	50.00	75.00	25.00	75.00
FP _I (%)	25.00	50.00	75.00	0.00	25.00
FN _I (%)	0.00	50.00	25.00	75.00	25.00
TN _I (%)	50.00	75.00	25.00	75.00	100.00
Precision _i (%)	80.00	50.00	50.00	100.00	75.00
Recall _i (%)	100.00	50.00	75.00	25.00	75.00
Accuracy _i (%)	85.71	55.56	50.00	57.14	77.78
Average Precision (%)			71.00		
Average Recall (%)			65.00		
Average Accuracy (%)			65.24		
Training Time (s)			681.80		
Testing Time (s)			0.32		
CW-GLCM (Proposed Method I)					
Class	Blotch	Bruise	Cork Spot	Scab	Rot
TP _i (%)	75.00	100.00	100.00	75.00	75.00
FP _i (%)	0.00	0.00	50.00	25.00	0.00
FN _i (%)	25.00	0.00	0.00	25.00	25.00
TN _i (%)	100.00	100.00	75.00	75.00	75.00
Precision _i (%)	100.00	100.00	66.67	75.00	100.00
Recall _i (%)	75.00	100.00	100.00	75.00	75.00
Accuracy _i (%)	87.50	100.00	77.78	75.00	85.71
Average Precision (%)			88.33		
Average Recall (%)			85.00		
Average Accuracy (%)			85.20		
Training Time (s)			188.27		
Testing Time (s)			0.15		
F-SPM (Proposed Method II)					
Class	Blotch	Bruise	Cork Spot	Scab	Rot
TP _i (%)	100.00	75.00	75.00	75.00	100.00
FP _i (%)	0.00	0.00	0.00	0.00	75.00
FN _i (%)	0.00	25.00	25.00	25.00	0.00
TN _i (%)	75.00	75.00	75.00	100.00	100.00

Precision _i (%)	100.00	100.00	100.00	100.00	57.14
Recall _i (%)	100.00	75.00	75.00	75.00	100.00
Accuracy _i (%)	100.00	85.71	85.71	87.50	72.73
Average Precision (%)			91.43		
Average Recall (%)			85.00		
Average Accuracy (%)			86.33		
Training Time (s)			360.31		
Testing Time (s)			0.61		

Overall, the proposed F-SPM method outperformed others in multi-class classification with an average accuracy of 86.33%. This is followed by the accuracy of the proposed CW-GLCM (85.20%), SPM (80.67%), CLAHE+GLCM+ELM (65.24%), CNN (61.43%), BOW (59.26%) and Texture analysis (24.62%).

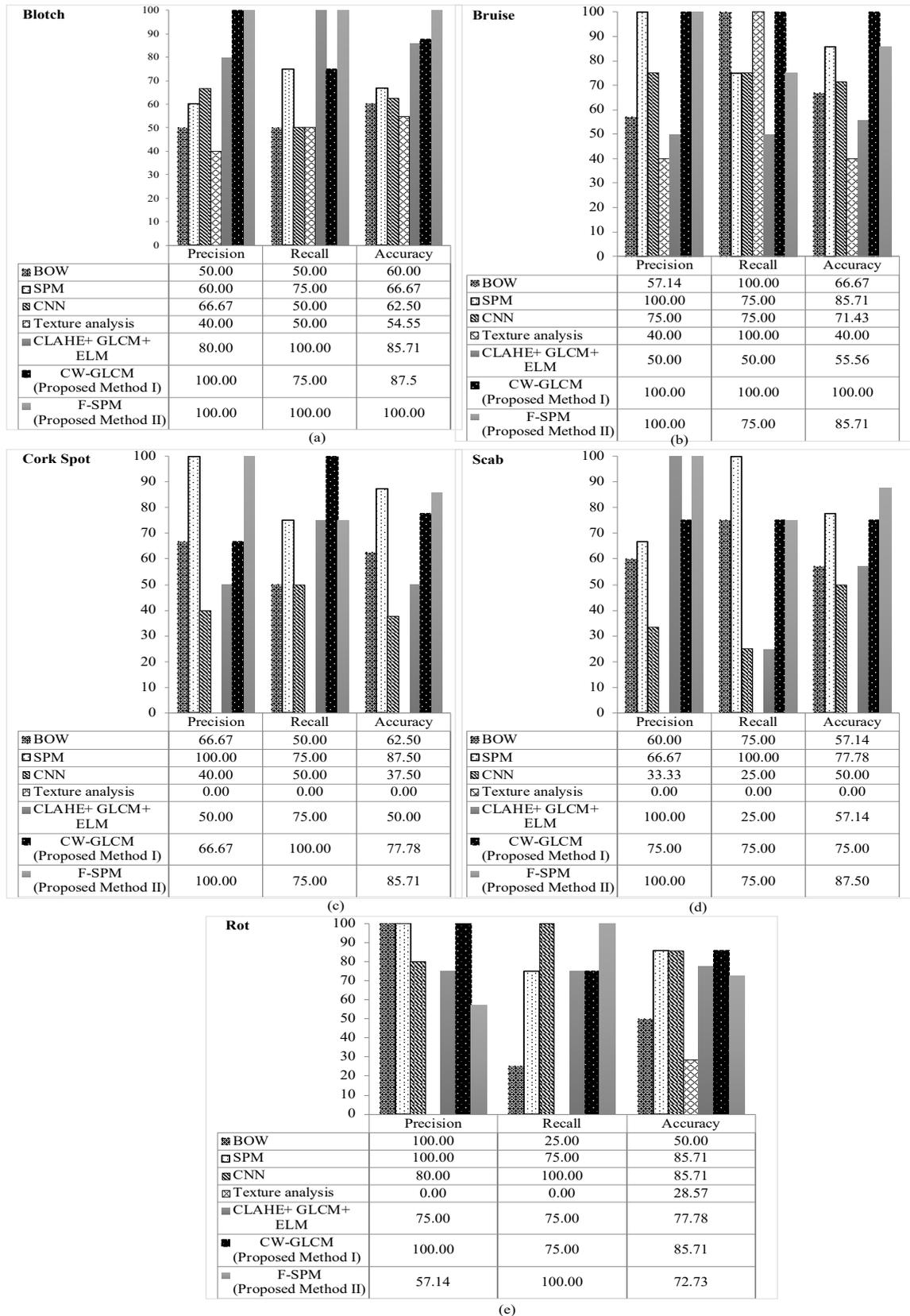


Figure 5.11: Comparative results of precision, recall and accuracy between the proposed methods with the existing image recognition methods for each individual class (a) Blotch, (b) Bruise, (c) Cork Spot, (d) Scab and (e) Rot on DA dataset.

For individual class defect, the proposed F-SPM method outperformed others in classifying the Blotch defect and scored a 100% for all the performance measures of precision, recall and accuracy. The proposed F-SPM method also outperformed others in classifying the Scab defect with an accuracy of 87.50%. These high performances are enabled by the Fuzzy logic detection that successfully captures the presence of the Blotch and Scab defective characteristic on the apple skin. The Fuzzy logic detection offers flexible membership function when defining the degree to which a pixel belongs to either on an edge of the Botch defective region or a uniform region of the non-defective. Furthermore, the spatial layout information in the proposed F-SPM method helps in distinguishing a Scab defect with stem end or calyx. These lead to the success of the defect detection on the apple skin images. On the other hand, the proposed CW-GLCM method outperformed others in the classification of the Bruise defect with a score of 100% for all the performance measures (precision, recall and accuracy). The proposed CW-GLCM method also able to obtain high classification accuracy in classifying the Rot defect with an accuracy of 85.71%. The main reason for the low classification accuracy of the Bruise defect in other methods is because the similarity in appearance between Bruise and Rot defects. However, this similarity has minimal impact on the performances of the proposed CW-GLCM method as the normalization step is skipped in the method. This will retain the image information for each type of the defects and consequently minimizing the misclassification between these defects.

Although score of 100% for all precision, recall and accuracy is observed in classifying the Bruise defect, these high classification results are not affected by overfitting. The overfitting situation might happen when the learning classifier tightly fits the training data given and it could be inaccurate when predicting the untrained data. It produces highly accurate output on the training data, but low accuracy when predicting samples that are not part of the training set.

In this research, the training and testing sets are specified using K -fold cross-validation technique in all the experiments. The K -fold cross-validation technique is one of a technique to prevent overfitting situation (Anguita, Ghelardoni, Ghio, Oneto, & Ridella, 2012). The K -fold cross-validation randomly partitioned a dataset into k number folders, in which each folder has virtually the same number of class distribution. $k-1$ of the folders are used for training and one folder for validation. This process is repeated k times until each folder is used exactly once as a validation set. Thus, there is no overlap between training sets and testing data in the experiments conducted.

The score of 100% in precision, recall and accuracy for CW-CLCM method is only observed on the individual Bruise defect classification for the DA dataset. There is also a possibility these results may be decreased if different input images are used. The reasons for the proposed CW-GLCM method able to obtain high performance measure for Bruise defect lies in the introduction of Curvelet and Wavelet transform in the proposed CW-GLCM. The proposed method able to enhance the texture information on unobvious Bruise defect region in DA dataset. This leads to a 100% score for precision, recall and accuracy on Bruise defect classification. Examples of the images of unobvious Bruise on the apple skin defect in DA dataset are shown in Figure 5.12. The unobvious Bruise defect region appears as a darker pigmented area compared to the healthy region as pointed by the arrow.



Figure 5.12: Examples of unobvious Bruise defects (a) red skin apple and (b) green skin apple.

Despite obtaining 100% score in Bruise defect classification, the average classification performance of the proposed CW-GLCM on five defect classes of DA dataset on the other hand, only able to obtain 88.33% for precision, 85.00% recall and 85.20% accuracy as presented in Table 5.6 and also summarized in Figure 5.13 (a).

In the classification of Cork Spot defect, the SPM method outperformed others with a classification accuracy of 87.50% as shown in Figure 5.11 (c). However, the average classification accuracy of the SPM method on five defect classes remain lower than both proposed methods as summarized in Figure 5.13 (a). The proposed F-SPM obtained the best average performances in precision, recall and accuracy compared to others with a minimum value of 85.00% on recall. In contrast, the Texture analysis method recorded the lowest average performances with precision (16.00%), recall (30.00%) and accuracy (24.62%). The Texture analysis method is highly depended on the texture information that can be ineffective when extracting the feature from the less texture of the defective region. Consequently, lead to the failed classification of the defect classes.

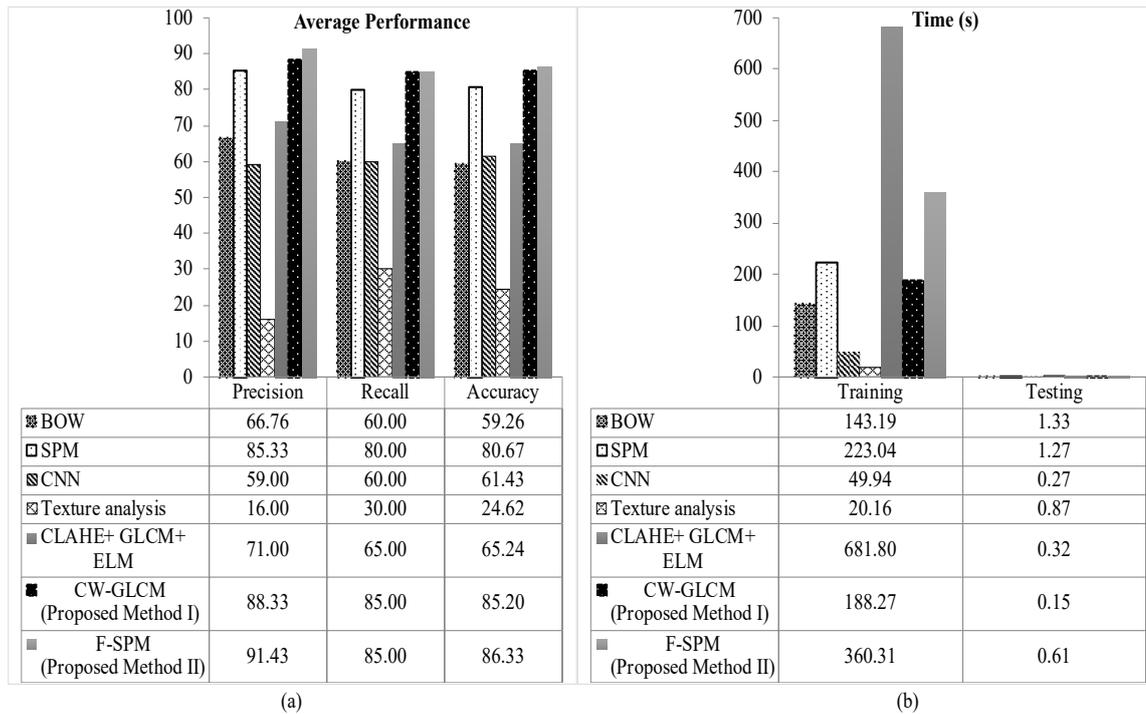


Figure 5.13: Comparative results of precision, recall and accuracy for each method on DA dataset (a) average performance (b) training and testing time.

In terms of the computational time as shown in Figure 5.13 (b), the Texture analysis is the fastest method during training (20.16 s). This is followed by the CNN (49.94 s), BOW (143.19 s), the proposed CW-GLCM (188.27 s), SPM (223.04 s), the proposed F-SPM (360.31 s) and CLAHE+GLCM+ELM (681.80 s). The CLAHE+GLCM+ELM method required the longest training time due to the computationally extensive of the CLAHE approach. The CLAHE approach is usually used for image enhancement in off-line application (Reza, 2004). Although the Texture analysis is the fastest method during training, it obtained the lowest average performances for precision, recall and accuracy compared to others. The testing time of the proposed CW-GLCM is the fastest (0.15 s) and followed by the CNN (0.27 s), CLAHE+GLCM+ELM (0.32 s), the proposed F-SPM (0.61 s), Texture analysis (0.87 s), SPM (1.27 s) and BOW (1.33 s). The BOW requires the longest testing time among all due to high computational cost of vector quantization step.

5.4 Discussion

A new image recognition method based on SPM method for apple classification namely, F-SPM has been presented. In comparison with other image recognition methods, the proposed F-SPM method presents the most promising results in the binary-class classification of defective and non-defective apples and multi-class classification between types of defects. The proposed F-SPM method achieved the highest accuracy on the NDDAW and DA datasets. For the NDDA dataset, the proposed F-SPM, CW-GLCM and SPM methods outperformed others with similar precision (96.30%), recall (100%) and accuracy (98.15%). However, lower classification accuracy is observed in the SPM method when tested on NDDAW dataset with 71.43% precision, 68.97% recall and 69.64% accuracy. In contrast, the proposed F-SPM achieved more than 91.43% precision, 84.85% recall and 86.33% accuracy on all datasets. The results indicate that the proposed F-SPM is the most accurate among all the methods including the proposed CW-GLCM. The accuracy difference between the F-SPM and CW-GLCM is 1.96% on NDDAW dataset and 1.13% on DA dataset. Meanwhile, the proposed CW-GLCM is the most suitable method to recognize the Bruise defect as it obtained the perfect measurement (100%) of precision, recall and accuracy on DA dataset.

5.5 Summary

This chapter described the proposed F-SPM method, one of the contributions in this thesis. The F-SPM method is proposed to improve the detection of the defective apples and reduce the unnecessary SPM features. The proposed F-SPM method is evaluated for binary-class and multi-class classifications. The experimental results demonstrated that the proposed F-SPM method improves the classification accuracy of the SPM method for both binary-class and multi-class classifications especially in classifying the defective

apple. The proposed methods also able to classify the binary-class of defective and non-defective apple images including the low-quality region images. In light of the outcomes, the proposed methods have the potential to be implemented in real-time systems for vision classification in smart manufacturing.

CHAPTER 6: CONCLUSION

This chapter concludes the research work presented in this thesis and relates the findings to the objectives presented in the first chapter. The contributions of the research work are also discussed. Finally, the limitations of this research work are outlined and future directions are suggested.

6.1 Conclusion

This thesis presents two new image recognition methods for apple classification. These methods are proposed to increase the classification accuracy for the binary-class classification of defective and non-defective apples even with low-quality region images and multi-class classification between types of defects. The contribution of each proposed method is individual, but associated with the feature extraction stage on the image recognition methods.

The first proposed method is CW-GLCM. The CW-GLCM is a feature extraction of fusion-features with Decision Tree classifier. The CW-GLCM is proposed to improve the ability of the GLCM Texture analysis method in detecting features on the low-quality region of the apple image. To achieve this, the proposed method introduces the fusion-features using the Curvelet and Wavelet transform based on the GLCM method. The Curvelet and Wavelet transform are introduced to enhance the low-quality region in the apple image by improving their texture information. The main reason for fusing the Curvelet features is to detect curves, corners and profiles at the low-quality region of the apple image. In addition, the original GLCM method is also modified using the Wavelet coefficient to enhance the detection of the texture information in the low-quality region by capturing the directional edges in different resolution levels preserving the low and

high frequency information. This leads the proposed method to extract better texture information from the low-quality region of the apple image.

The second proposed method of F-SPM concentrates on improving the drawbacks of the proposed CW-GLCM method in classifying defective apples. From the binary-class classification of defective and non-defective, the apple classification is extended to multi-class classification between types of defects. The proposed F-SPM method is a visual-dictionary feature extraction with SVM classifier. The proposed method includes the spatial layout information of the SPM for encoding the spatial distribution features. This spatial layout information is important to discriminate between defects and natural parts of stem end or calyx. The key contribution in the second proposed method is the detection and selection only significant features for further classification to reduce the unnecessary SPM features through Fuzzy logic detection. The Fuzzy logic detection highlights the high-frequency components of the defect region on the apple skin images.

The proposed methods are tested on three new apple datasets. These datasets are created due to shortage of defective and non-defective datasets that comprise low-quality region for agriculture product. Each dataset focuses on different aspect of classification. The NDDA and NDDAW are both binary-class datasets containing defective and non-defective apple images, with NDDAW contains more low-quality region images compared to the NDDA. Conversely, the DA dataset contains five types of defective apple images, which are Scab, Rot, Cork Spot, Blotch and Bruise. The evaluations of the proposed methods on these datasets used a similar experimental setup and compared with five existing image recognition methods.

In the binary-class classification of defective and non-defective apple including the low-quality region images, the proposed CW-GLCM method outperformed the existing methods with at least 86.79% precision, 91.01% recall and 89.11% accuracy as presented

in section 4.3.3. For the multi-class classification between types of defects, the proposed CW-GLCM method outperformed others in classifying the Bruise defect with a perfect score of 100% in precision, recall and accuracy as discussed in section 5.3.4.2. The proposed F-SPM method also scored 100% in precision, recall and accuracy for classifying the Blotch defect and outperformed the other methods. For classifying the Scab defect, the proposed F-SPM method outperformed the others with 100% precision and 87.50% accuracy. The method also attained the highest precision, recall and accuracy compared to other methods with a minimum value of 85.00% on recall for overall classification performance from five defect classes in the DA dataset.

Finally, the presented results confirm the achievement of the outlined research objectives in detecting and classifying apple images. The apple classification includes for binary-class classification of defective and non-defective apples even with the low-quality region images and multi-class classification between types of defects. Also, the results indicate that both proposed methods have the potential to be implemented for vision classification in smart manufacturing.

6.2 Future Work Direction

There are several future directions and gaps can be explored to further improve and validate the proposed methods. In this research, the proposed methods are evaluated on small sample datasets of apple images (550 images of NDDA, 560 images of NDDAW and 200 images of DA). This is due to limited apple image datasets available publicly despite apple having the highest production rate and steadily increase over the year as reported by USDA (USDA, 2017). Moreover, the available datasets insufficient to describe the targeted problem highlighted in this research (Kamilaris & Prenafeta-Boldú, 2018; X. Song et al., 2016). Therefore, small datasets of apple images are mainly sampled

to classify defective and non-defective apple including low-quality region and also multi-class between types of defects.

In this research, it is challenging to meet the large-scale dataset requirement of deep learning method such as ImageNet dataset. ImageNet dataset consists of 3.2 million cleanly labelled images and aim to contain 50 million images in the dataset (Deng et al., 2009). The large-scale datasets or big data are normally harvested automatically from large users or crowds population using crawling techniques, crowd source or application programming interface that provided by social media providers (B. Jiang, 2015). The optimization issues might arise when using deep learning approach of the pre-trained models on a small dataset that is significantly different due to the models' complexity and hardware restrictions (Chaturvedi et al., 2018; X. Cheng et al., 2017; He et al., 2015; Kamilaris & Prenafeta-Boldú, 2018; Krizhevsky et al., 2012; Y. Zhang et al., 2019).

Therefore, future research can consider increasing the diversity of apple image dataset using data augmentation technique for more advanced deep learning method. However, the key challenge for data augmentation is its computationally expensive to generate enough samples for training on a large neural network. Furthermore, due to the particularities of the low-quality region, various types of defects, severity and cultivar of apple images in this research, it will become a major challenge for the data augmentation. The overly augmented and redundant augmentation may also introduce biases into the dataset and can slow down the training (Graham, 2014; Ho, Liang, Stoica, Abbeel, & Chen, 2019). The potential of the data augmentation and more advance deep learning method to optimize the performance deserve further study.

Second, the proposed F-SPM method is highly dependent on Fuzzy logic detection to select significant features for apple classification. The selection is performed by detecting edges or contours in apple images to highlight only the high frequency components. The

edges are detected by comparing the intensity between two neighboring pixels. However, the small intensity differences between two neighboring pixels do not always represent the edge due to the image complexity. Other factors related to the image properties such as noise, lighting, blurred images and dynamic background may also influence the edge detection, which consequently can reduce the classification accuracy (Anas, 2016; Lakshmi & Sankaranarayanan, 2010; J. Song et al., 2019). Addressing this issue may improve the performance of the proposed method. The aforementioned issues can be addressed by using feature selection technique. Feature selection technique allows the elimination of noisy features, irrelevant and redundant features (Adegoke, Ola, Omotayo, & No, 2014). In future research, feature selection technique such as Relief, Fisher score and Information Gain based methods can be employed in the proposed method to improve the current results. This is possible since feature selection technique can improve the quality of the data for classification (G.-Z. Li, Yang, Liu, & Xue, 2004; Tang, Alelyani, & Liu, 2014).

Finally, considering the achievement of the proposed methods for apple classification discussed in this thesis, it is also believed that the output from the classifier can be further explored. The output from the classifier which is defective and non-defective including types of defects in apple production has the potential to be used as the input for data analytics and visualization. The analytics process can identify the patterns based on the current production data and learns for future planning and prediction; which potentially helps improving the apple growth and processing efficiency.

REFERENCES

- Abdullah, M., Hazem, M., & Reham, R. (2017). Image contrast enhancement using fast discrete curvelet transform via wrapping (FDCT-Wrap). *Int. J. Adv. Res. Comput. Sci. Technol*, 5(2), 10-16.
- Abdulrahman, M., Gwadabe, T. R., Abdu, F. J., & Eleyan, A. (2014). *Gabor wavelet transform based facial expression recognition using PCA and LBP*. Paper presented at the 2014 22nd Signal Processing and Communications Applications Conference (Siu).
- Aborisade, D. (2010). Fuzzy logic based digital image edge detection. *Global Journal of Computer Science and Technology*.
- Acharya, U. R., Raghavendra, U., Fujita, H., Hagiwara, Y., Koh, J. E., Hong, T. J., . . . Gudigar, A. (2016). Automated characterization of fatty liver disease and cirrhosis using curvelet transform and entropy features extracted from ultrasound images. *Computers in biology and medicine*, 79, 250-258.
- Adegoke, B., Ola, B., Omotayo, M., & No, P. (2014). Review of feature selection methods in medical image processing. *IOSR J Eng (IOSRJEN)*, 4(01), 01-05.
- Affonso, C., Sassi, R. J., & Barreiros, R. M. (2015). Biological image classification using rough-fuzzy artificial neural network. *Expert systems with Applications*, 42(24), 9482-9488.
- Agaian, S. S., Mulawka, M. M. A., Rajendran, R., Rao, S. P., KM, S. K., & Rajeev, S. (2016). A comparative study of image feature detection and matching algorithms for touchless fingerprint systems. *Electronic Imaging*, 2016(15), 1-9.
- Agarwal, J., & Bedi, S. S. (2015). Implementation of hybrid image fusion technique for feature enhancement in medical diagnosis. *Human-centric Computing and Information Sciences*, 5(1), 3.
- Al-Hammadi, M. H., Muhammad, G., Hussain, M., & Bebis, G. (2013). *Curvelet transform and local texture based image forgery detection*. Paper presented at the International Symposium on Visual Computing.
- Al-Kadi, O. S. (2015). A multiresolution clinical decision support system based on fractal model design for classification of histological brain tumours. *Computerized Medical Imaging and Graphics*, 41, 67-79.
- Aldavert, D., Rusinol, M., Toledo, R., & Lladós, J. (2015). A study of Bag-of-Visual-Words representations for handwritten keyword spotting. *International Journal on Document Analysis and Recognition*, 18(3), 223-234. doi:10.1007/s10032-015-0245-z
- Anas, E. (2016). *Edge detection techniques using fuzzy logic*. Paper presented at the 2016 3rd International Conference on Signal Processing and Integrated Networks (SPIN).

- Anguita, D., Ghelardoni, L., Ghio, A., Oneto, L., & Ridella, S. (2012). *The K-fold Cross Validation*. Paper presented at the ESANN.
- Antonucci, A., & Corani, G. (2017). The multilabel naive credal classifier. *International Journal of Approximate Reasoning*, 83, 320-336.
- Arya, D., Singh, R. S., Kumar, A., & Mandoria, H. (2018). Texture, shape and color based classification of satellite images using GLCM & Gabor Filter, Fuzzy C Means and SVM. *International Research Journal of Engineering and Technology (IRJET)*, 05(04).
- Attamimi, M., Araki, T., Nakamura, T., & Nagai, T. (2013). Visual Recognition System for Cleaning Tasks by Humanoid Robots. *International Journal of Advanced Robotic Systems*, 10.
- Auria, L., & Moro, R. A. (2008). Support vector machines (SVM) as a technique for solvency analysis. 811.
- Avila, S. (2013). *Extended bag-of-words formalism for image classification*. Université Pierre et Marie Curie-Paris VI,
- Bay, H., Ess, A., Tuytelaars, T., & Van Gool, L. (2008). Speeded-up robust features (SURF). *Computer Vision and Image Understanding*, 110(3), 346-359.
- Bay, H., Tuytelaars, T., & Van Gool, L. (2006). *Surf: Speeded up robust features*. Paper presented at the European conference on computer vision-Vol3951, Berlin, Heidelberg.
- Candes, E., Demanet, L., Donoho, D., & Ying, L. (2006). Fast discrete curvelet transforms. *Multiscale Modeling & Simulation*, 5(3), 861-899.
- Capizzi, G., LO SCIUTO, G., Napoli, C., Tramontana, E., & WOŹNIAK, M. (2016). A Novel Neural Networks-Based Texture Image Processing Algorithm for Orange Defects Classification. *International Journal of Computer Science & Applications*, 13(2).
- Chan, T. H., Jia, K., Gao, S. H., Lu, J. W., Zeng, Z. N., & Ma, Y. (2015). PCANet: A Simple Deep Learning Baseline for Image Classification? *Ieee Transactions on Image Processing*, 24(12), 5017-5032. doi:10.1109/Tip.2015.2475625
- Chanti, D. A., & Caplier, A. (2018). Improving bag-of-visual-words towards effective facial expressive image classification. *arXiv preprint arXiv:1810.00360*.
- Chaturvedi, I., Ragusa, E., Gastaldo, P., Zunino, R., & Cambria, E. (2018). Bayesian network based extreme learning machine for subjectivity detection. *Journal of The Franklin Institute*, 355(4), 1780-1797.
- Chen, R. (2018). *Machine Learning Methods for Autonomous Object Recognition and Restoration in Images*. University of Sheffield,

- Chen, X., Kopsaftopoulos, F., Wu, Q., Ren, H., & Chang, F.-K. (2019). A Self-Adaptive 1D Convolutional Neural Network for Flight-State Identification. *Sensors*, *19*(2), 275.
- Cheng, H.-D., Shi, X., Min, R., Hu, L., Cai, X., & Du, H. (2006). Approaches for automated detection and classification of masses in mammograms. *Pattern Recognition*, *39*(4), 646-668. Retrieved from <https://www.sciencedirect.com/science/article/pii/S0031320305002955?via%3Dihub>
- Cheng, X., Zhang, Y., Chen, Y., Wu, Y., & Yue, Y. (2017). Pest identification via deep residual learning in complex background. *Computers and Electronics in Agriculture*, *141*, 351-356.
- Chowdhury, S., Verma, B., & Stockwell, D. (2015). A novel texture feature based multiple classifier technique for roadside vegetation classification. *Expert systems with Applications*, *42*(12), 5047-5055.
- Chui, K. T., & Lytras, M. D. (2019). A Novel MOGA-SVM Multinomial Classification for Organ Inflammation Detection. *Applied Sciences*, *9*(11), 2284.
- Chung, C.-L., Huang, K.-J., Chen, S.-Y., Lai, M.-H., Chen, Y.-C., & Kuo, Y.-F. (2016). Detecting Bakanae disease in rice seedlings by machine vision. *Computers and Electronics in Agriculture*, *121*, 404-411.
- Cong, I., & Duan, L. (2015). Quantum Discriminant Analysis for Dimensionality Reduction and Classification. *arXiv preprint arXiv:1510.00113*.
- Csurka, G., Dance, C., Fan, L., Willamowski, J., & Bray, C. (2004, 11–14 May 2004). *Visual categorization with bags of keypoints*. Paper presented at the Workshop on statistical learning in computer vision, ECCV, Prague, Czech Republic.
- Cui, D. (2019). *Enhanced extreme learning machines for image classification*. Doctoral thesis, Nanyang Technological University, Singapore.
- curvelet.org. CurveLab. Available online:<http://www.curvelet.org/software>. html (accessed on 7 February 2018).
- Da Xu, L., He, W., & Li, S. (2014). Internet of things in industries: A survey. *IEEE Transactions on industrial informatics*, *10*(4), 2233-2243.
- Danades, A., Pratama, D., Anggraini, D., & Anggriani, D. (2016). *Comparison of accuracy level K-nearest neighbor algorithm and support vector machine algorithm in classification water quality status*. Paper presented at the Frontiers of Information Technology (FIT), 2016 International Conference on.
- de Haan, R. R., Visser, J. B., Pons, E., Feelders, R. A., Kaymak, U., Hunink, M. M., & Visser, J. J. (2017). Patient-specific workup of adrenal incidentalomas. *European journal of radiology open*, *4*, 108-114.

- Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., & Fei-Fei, L. (2009). *Imagenet: A large-scale hierarchical image database*. Paper presented at the 2009 IEEE conference on computer vision and pattern recognition.
- Do, T.-N., Lenca, P., Lallich, S., & Pham, N.-K. (2010). Classifying very-high-dimensional data with random forests of oblique decision trees. In *Advances in knowledge discovery and management* (pp. 39-55): Springer.
- Domingos, P. & Pazzani, M. (1996). *Beyond independence: conditions for the optimality of the simple Bayesian classifier*. Paper presented at the Machine Learning: Proceedings of the Thirteenth International Conference', Morgan Kaufmann.
- Donahue, J., Jia, Y., Vinyals, O., Hoffman, J., Zhang, N., Tzeng, E., & Darrell, T. (2014). *Decaf: A deep convolutional activation feature for generic visual recognition*. Paper presented at the International conference on machine learning.
- dos Santos Ferreira, A., Freitas, D. M., da Silva, G. G., Pistori, H., & Folhes, M. T. (2017). Weed detection in soybean crops using ConvNets. *Computers and Electronics in Agriculture*, 143, 314-324.
- Dougherty, E. R., Hua, J., & Sima, C. (2009). Performance of feature selection methods. *Current genomics*, 10(6), 365-374.
- Druzdzel, M. J., & Onisko, A. (2008). *The impact of overconfidence bias on practical accuracy of Bayesian network models: an empirical study*. Paper presented at the BMA.
- Duda, R. O., Hart, P. E., & Stork, D. G. (2001). Pattern classification 2nd ed. *John Wiley & Sons Inc.*
- Duda, R. O., Hart, P. E., & Stork, D. G. (2012). *Pattern classification*: John Wiley & Sons.
- Duneja, A., & Puyalnithi, T. (2017). Enhancing classification accuracy of k-nearest neighbours algorithm using gain ratio. *International Research Journal of Engineering and Technology (IRJET)*, 1385-1388.
- Ebrahimi, M., Khoshtaghaza, M., Minaei, S., & Jamshidi, B. (2017). Vision-based pest detection based on SVM classification method. *Computers and Electronics in Agriculture*, 137, 52-58.
- Fahrurozi, A., Madenda, S., & Kerami, D. (2016, 2016). *Wood Texture Features Extraction by Using GLCM Combined With Various Edge Detection Methods*. Paper presented at the Journal of Physics: Conference Series-Volume 725.
- Fan, C., Jin, H., Wang, F., Zhang, G., & Li, Y. (2019). Combining and matching keypoints and lines on multispectral images. *Infrared Physics & Technology*, 96, 316-324.
- Fei-Fei, L., & Perona, P. (2005). *A bayesian hierarchical model for learning natural scene categories*. Paper presented at the Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on.

- Galloway, M. M. (1974). Texture analysis using grey level run lengths. *NASA STI/Recon Technical Report N, 75*.
- Ganesan, P., & Sajiv, G. (2017). *A comprehensive study of edge detection for image processing applications*. Paper presented at the 2017 International Conference on Innovations in Information, Embedded and Communication Systems (ICIIECS).
- Ganganagowder, N. V., & Kamath, P. (2017). Intelligent classification models for food products basis on morphological, colour and texture features. *Acta Agronómica, 66*(4), 486-494.
- George, M., & Zwiggelaar, R. (2019). Comparative study on local binary patterns for mammographic density and risk scoring. *Journal of Imaging, 5*(2), 24.
- Girisha, A., Chandrashekhar, M., & Kurian, M. (2013). Texture feature extraction of video frames using GLCM. *International Journal of Engineering Trends and Technology, 4*(6), 2718-2721.
- Girshick, R., Donahue, J., Darrell, T., & Malik, J. (2014). *Rich feature hierarchies for accurate object detection and semantic segmentation*. Paper presented at the Proceedings of the IEEE conference on computer vision and pattern recognition.
- Graham, B. (2014). Fractional max-pooling. *arXiv preprint arXiv:1412.6071, V4, 1412.6071*.
- Guyon, I., & Elisseeff, A. (2003). An introduction to variable and feature selection. *Journal of machine learning research, 3*(Mar), 1157-1182.
- Hagargi, P. A., & Shubhangi, D. (2018). Brain tumor MR image fusion using most dominant features extraction from wavelet and curvelet transforms. *Brain, 5*(05), 33-38.
- Han, T., & Jiang, D. (2016). Rolling bearing fault diagnostic method based on VMD-AR model and random forest classifier. *Shock and Vibration, 2016*.
- Hanbay, K., Talu, M. F., & Özgüven, Ö. F. (2016). Fabric defect detection systems and methods—A systematic literature review. *Optik, 127*(24), 11960-11973.
- Haq, I., Anwar, S., Shah, K., Khan, M. T., & Shah, S. A. (2015). Fuzzy logic based edge detection in smooth and noisy clinical images. *Plos One, 10*(9), e0138712.
- Harada, T., Ushiku, Y., Yamashita, Y., & Kuniyoshi, Y. (2011). *Discriminative spatial pyramid*. Paper presented at the Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on.
- Harris, C., & Stephens, M. (1988, 31st August - 2nd September). *A combined corner and edge detector*. Paper presented at the Alvey vision conference, University of Manchester.
- Hashemi, M. (2019). Enlarging smaller images before inputting into convolutional neural network: zero-padding vs. interpolation. *Journal of Big Data, 6*(1), 98.

- Hassan, R., Kasim, S., Jafery, W. A. Z. W. C., & Shah, Z. A. (2017). Image Enhancement Technique at Different Distance for Iris Recognition. *International Journal on Advanced Science, Engineering and Information Technology*, 7(4-2), 1510-1515.
- Hastie, T., Tibshirani, R., Friedman, J., & Franklin, J. (2005). The elements of statistical learning: data mining, inference and prediction. *The Mathematical Intelligencer*, 27(2), 83-85.
- He, K., Zhang, X., Ren, S., & Sun, J. (2015). Spatial pyramid pooling in deep convolutional networks for visual recognition. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 37(9), 1904-1916. Retrieved from <https://ieeexplore.ieee.org/document/7005506>
- He, K., Zhang, X., Ren, S., & Sun, J. (2016). *Deep residual learning for image recognition*. Paper presented at the Proceedings of the IEEE conference on computer vision and pattern recognition.
- Hinneburg, A., Aggarwal, C. C., & Keim, D. A. (2000). *What is the nearest neighbor in high dimensional spaces?* Paper presented at the 26th Internat. Conference on Very Large Databases.
- Ho, D., Liang, E., Stoica, I., Abbeel, P., & Chen, X. (2019). Population Based Augmentation: Efficient Learning of Augmentation Policy Schedules. *arXiv preprint arXiv:1905.05393*, V1, 1905.05393.
- Hsu, C.-W., Chang, C.-C., & Lin, C.-J. (2003). A practical guide to support vector classification.
- Idris, M., Arof, H., Tamil, E., Noor, N., & Razak, Z. (2009). Review of feature detection techniques for simultaneous localization and mapping and system on chip approach. *Information Technology Journal*, 8(3), 250-262.
- Idris, M. Y. I., Warif, N. B. A., Arof, H., Noor, N. M., Wahab, A. W. A., & Razak, Z. (2019). ACCELERATING FPGA-SURF FEATURE DETECTION MODULE BY MEMORY ACCESS REDUCTION. *Malaysian Journal of Computer Science*, 32(1), 47-61.
- Ismail, A., Idris, M. Y. I., Ayub, M. N., & Por, L. Y. (2018). Vision-Based Apple Classification for Smart Manufacturing. *Sensors*, 18, 4353.
- Ismail, A., Idris, M. Y. I., Ayub, M. N., & Yee, L. (2019). Investigation of Fusion Features for Apple Classification in Smart Manufacturing. *Symmetry*, 11(10), 1194.
- Jadhav, S. D., & Channe, H. (2016). Comparative study of K-NN, naive Bayes and decision tree classification techniques. *International Journal of Science and Research (IJSR)*, 5(1), 1842-1845.
- Jain, G., & Kaur, S. (2017). A Review on Various Edge Detection Techniques in Distorted Images. *International Journal*, 7(5).
- Jiang, B. (2015). Big data is not just a new type, but a new paradigm. *arXiv preprint arXiv:1511.01242*.

- Jiang, L., Zhu, B., Cheng, X., Luo, Y., & Tao, Y. (2009). 3D surface reconstruction and analysis in automated apple stem-end/calyx identification. *Transactions of the ASABE*, 52(5), 1775-1784.
- Kahraman, I., Karas, I., & Akay, A. (2018). Road Extraction Techniques from Remote Sensing Images: A Review. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.*, 42, 339-342.
- Kamilaris, A., & Prenafeta-Boldú, F. X. (2018). Deep learning in agriculture: A survey. *Computers and Electronics in Agriculture*, 147, 70-90.
- Kanungo, T., Mount, D. M., Netanyahu, N. S., Piatko, C. D., Silverman, R., & Wu, A. Y. (2002). An efficient k-means clustering algorithm: Analysis and implementation. *Ieee Transactions on Pattern Analysis and Machine Intelligence*, 24(7), 881-892.
- Karandikar, J., McLeay, T., Turner, S., & Schmitz, T. (2015). Tool wear monitoring using naive Bayes classifiers. *The International Journal of Advanced Manufacturing Technology*, 77(9-12), 1613-1626.
- Karbauskaitė, R., & Dzemyda, G. (2016). Fractal-based methods as a technique for estimating the intrinsic dimensionality of high-dimensional data: a survey. *Informatica*, 27(2), 257-281.
- Kaur, R., Chawla, M., Khiva, N. K., & Ansari, M. D. (2018). Comparative Analysis of Contrast Enhancement Techniques for Medical Images. *Pertanika Journal of Science & Technology*, 26(3), 965-978.
- Kazmi, W., Garcia-Ruiz, F. J., Nielsen, J., Rasmussen, J., & Andersen, H. J. (2015). Detecting creeping thistle in sugar beet fields using vegetation indices. *Computers and Electronics in Agriculture*, 112, 10-19.
- Kejriwal, N., Kumar, S., & Shibata, T. (2016). High performance loop closure detection using bag of word pairs. *Robotics and Autonomous Systems*, 77, 55-65. doi:10.1016/j.robot.2015.12.003
- Koch, C., Georgieva, K., Kasireddy, V., Akinci, B., & Fieguth, P. (2015). A review on computer vision based defect detection and condition assessment of concrete and asphalt civil infrastructure. *Advanced Engineering Informatics*, 29(2), 196-210.
- Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012, Dec 3, 2012 - Dec 6, 2012). *Imagenet classification with deep convolutional neural networks*. Paper presented at the Advances in neural information processing systems, Lake Tahoe, Nevada, USA.
- Kumar, A. D., Selvam, R. P., & Kumar, K. S. (2018). Review on prediction algorithms in educational data mining. *International Journal of Pure and Applied Mathematics*, 118(8), 531-537.
- Kurtulmuş, F., & Ünal, H. (2015). Discriminating rapeseed varieties using computer vision and machine learning. *Expert systems with Applications*, 42(4), 1880-1891.

- Lakshmi, S., & Sankaranarayanan, D. V. (2010). A study of edge detection techniques for segmentation computing approaches. *IJCA Special Issue on "Computer Aided Soft Computing Techniques for Imaging and Biomedical Applications" CASCT*, 35-40.
- Langarizadeh, M., & Moghbeli, F. (2016). Applying naive bayesian networks to disease prediction: a systematic review. *Acta Informatica Medica*, 24(5), 364.
- Lazebnik, S., Schmid, C., & Ponce, J. (2006). *Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories*. Paper presented at the Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on.
- Le, Q., & Mikolov, T. (2014). *Distributed representations of sentences and documents*. Paper presented at the International Conference on Machine Learning.
- Lee, M. H., Cho, M., & Park, I. K. (2015). Feature description using local neighborhoods. *Pattern Recognition Letters*, 68, 76-82.
- Li, D., Cong, A., & Guo, S. (2019). Sewer damage detection from imbalanced CCTV inspection data using deep convolutional neural networks with hierarchical classification. *Automation in Construction*, 101, 199-208.
- Li, G.-Z., Yang, J., Liu, G.-P., & Xue, L. (2004). *Feature selection for multi-class problems using support vector machines*. Paper presented at the Pacific Rim International Conference on Artificial Intelligence.
- Li, L., Ota, K., & Dong, M. (2018). Deep Learning for Smart Industry: Efficient Manufacture Inspection System with Fog Computing. *IEEE Transactions on industrial informatics*.
- Li, Q., Peng, J., Li, Z., & Ren, Y. (2017). *An image classification algorithm integrating principal component analysis and spatial pyramid matching features*. Paper presented at the Image Information Processing (ICIIP), 2017 Fourth International Conference on.
- Li, W., Chen, Y., Sun, W., Brown, M., Zhang, X., Wang, S., & Miao, L. (2019). A gingivitis identification method based on contrast-limited adaptive histogram equalization, gray-level co-occurrence matrix, and extreme learning machine. *International Journal of Imaging Systems and Technology*, 29(1), 77-82.
- Li, W., Dong, P., Xiao, B., & Zhou, L. (2016). Object recognition based on the Region of Interest and optimal Bag of Words model. *Neurocomputing*, 172, 271-280.
- Li, Y., Wang, S., Tian, Q., & Ding, X. (2015). Feature representation for statistical-learning-based object detection: A review. *Pattern Recognition*, 48(11), 3542-3559.
- Lin, W. C., Tsai, C. F., Chen, Z. Y., & Ke, S. W. (2016). Keypoint selection for efficient bag-of-words feature generation and effective image classification. *Information Sciences*, 329, 33-51. doi:10.1016/j.ins.2015.08.021

- Liu, B.-D., Meng, J., Xie, W.-Y., Shao, S., Li, Y., & Wang, Y. (2019). Weighted Spatial Pyramid Matching Collaborative Representation for Remote-Sensing-Image Scene Classification. *Remote Sensing*, *11*(5), 518.
- Liu, C., Shirowzhan, S., Sepasgozar, S. M., & Kaboli, A. (2019). Evaluation of Classical Operators and Fuzzy Logic Algorithms for Edge Detection of Panels at Exterior Cladding of Buildings. *Buildings*, *9*(2), 40.
- Liu, Y., Zhang, Y.-M., Zhang, X.-Y., & Liu, C.-L. (2016). Adaptive spatial pooling for image classification. *Pattern Recognition*, *55*, 58-67.
- Loncomilla, P., Ruiz-del-Solar, J., & Martínez, L. (2016). Object recognition using local invariant features for robotic applications: A survey. *Pattern Recognition*, *60*, 499-514.
- Lowe, D. G. (2004). Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, *60*(2), 91-110.
- Lu, Y., Yi, S., Zeng, N., Liu, Y., & Zhang, Y. (2017). Identification of rice diseases using deep convolutional neural networks. *Neurocomputing*, *267*, 378-384.
- Luo, H.-L., Wei, H., & Lai, L. L. (2011). Creating efficient visual codebook ensembles for object categorization. *Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on*, *41*(2), 238-253.
- Luo, J., Song, D., Xiu, C., Geng, S., & Dong, T. (2014). Fingerprint classification combining curvelet transform and gray-level cooccurrence matrix. *Mathematical Problems in Engineering*, *2014*.
- Ma, J., Xu, F., Huang, K., & Huang, R. (2017). GNAR-GARCH model and its application in feature extraction for rolling bearing fault diagnosis. *Mechanical Systems and Signal Processing*, *93*, 175-203.
- Ma, Z., Zhou, C., Hepburn, D. M., & Cowan, K. (2016). *Fractal theory based pattern recognition of motor partial discharge*. Paper presented at the 2016 International Conference on Condition Monitoring and Diagnosis (CMD).
- Mandelbrot, B. B. (1983). *The fractal geometry of nature/Revised and enlarged edition*. New York, *WH Freeman and Co.*, 1983, 495 p.
- Maule, P., Shete, A., Wani, K., & Dawange, A. (2015). A Review on GLCM Feature Extraction in Retinal Image. *International Journal of Advanced Research in Computer Engineering & Technology (IJARCET)*, *4*(9).
- Mehdizadeh, S. A., Minaei, S., Hancock, N. H., & Torshizi, M. A. K. (2014). An intelligent system for egg quality classification based on visible-infrared transmittance spectroscopy. *Information Processing in Agriculture*, *1*(2), 105-114.
- Mikolajczyk, K., & Schmid, C. (2005). A performance evaluation of local descriptors.

- Moallem, P., Serajoddin, A., & Pourghassem, H. (2017). Computer vision-based apple grading for golden delicious apples based on surface features. *Information Processing in Agriculture*, 4(1), 33-40.
- Mohammed, I. S., & Alhamdani, I. M. (2019). A fuzzy system for detection and classification of textile defects to ensure the quality of fabric production. *International Journal of Electrical & Computer Engineering (2088-8708)*, 9.
- Mokhtar, U., El Bendary, N., Hassenian, A. E., Emary, E., Mahmoud, M. A., Hefny, H., & Tolba, M. F. (2015). SVM-based detection of tomato leaves diseases. In *Intelligent Systems' 2014* (pp. 641-652): Springer.
- Mondal, D., Kole, D. K., & Roy, K. (2017). Gradation of yellow mosaic virus disease of okra and bitter gourd based on entropy based binning and Naive Bayes classifier after identification of leaves. *Computers and Electronics in Agriculture*, 142, 485-493.
- Moyne, J., & Iskandar, J. (2017). Big data analytics for smart manufacturing: Case studies in semiconductor manufacturing. *Processes*, 5(3), 39.
- Nagorny, K., Lima-Monteiro, P., Barata, J., & Colombo, A. W. (2017). Big data analysis in smart manufacturing: a review. *International Journal of Communications, Network and System Sciences*, 10(3), 31-58.
- Nasirahmadi, A., & Ashtiani, S.-H. M. (2017). Bag-of-Feature model for sweet and bitter almond classification. *Biosystems Engineering*, 156, 51-60.
- Neves, D. P., Mehdizadeh, S. A., Tschärke, M., de Alencar Nääs, I., & Banhazi, T. M. (2015). Detection of flock movement and behaviour of broiler chickens at different feeders using image analysis. *Information Processing in Agriculture*, 2(3-4), 177-182.
- Ojala, T., Pietikäinen, M., & Harwood, D. (1996). A comparative study of texture measures with classification based on featured distributions. *Pattern Recognition*, 29(1), 51-59.
- Ojala, T., Pietikäinen, M., & Mäenpää, T. (2002). Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Transactions on Pattern Analysis & Machine Intelligence*(7), 971-987.
- Olaniyi, E. O., Adekunle, A. A., Odekuoye, T., & Khashman, A. (2017). Automatic system for grading banana using GLCM texture feature extraction and neural network arbitrations. *Journal of Food Process Engineering*, 40(6).
- Oni, D., Ojo, J., Alabi, B., Adebayo, A., & Amoran, A. (2018). Patterned Fabric Defect Detection and Classification (FDDC) Techniques: A Review [J]. *International Journal of Scientific & Engineering Research*, 9(2), 1156-1165.
- Onisko, A., & Druzdzel, M. J. (2003). *Effect of imprecision in probabilities on Bayesian network models: An empirical study*. Paper presented at the Working notes of the European Conference on Artificial Intelligence in Medicine (AIME-03):

Qualitative and Model-based Reasoning in Biomedicine. Protaras, Cyprus (October 18–22 2003).

- Oniśko, A., & Druzdzel, M. J. (2013). Impact of precision of Bayesian network parameters on accuracy of medical diagnostic systems. *Artificial intelligence in medicine*, 57(3), 197-206.
- Ozkan, K., Ergin, S., Isik, S., & Isikli, I. (2015). A new classification scheme of plastic wastes based upon recycling labels. *Waste management*, 35, 29-35. doi:10.1016/j.wasman.2014.09.030
- Panchal, P., Panchal, S., & Shah, S. (2013). A comparison of SIFT and SURF. *International Journal of Innovative Research in Computer and Communication Engineering*, 1(2), 323-327.
- Papakostas, G. A., Koulouriotis, D. E., Karakasis, E. G., & Tourassis, V. D. (2013). Moment-based local binary patterns: a novel descriptor for invariant pattern recognition applications. *Neurocomputing*, 99, 358-371.
- Patel, K. K., Kar, A., Jha, S., & Khan, M. (2012). Machine vision system: a tool for quality inspection of food and agricultural products. *Journal of food science and technology*, 49(2), 123-141.
- Patel, P., & Bhandari, A. (2019). A Review on Image Contrast Enhancement Techniques. *INTERNATIONAL JOURNAL ONLINE OF SCIENCE*, 5(7), 5-5.
- Pellegrini, T. (2015). *Comparing SVM, Softmax, and shallow neural networks for eating condition classification*. Paper presented at the Sixteenth annual conference of the international speech communication association.
- Penatti, O. A., Silva, F. B., Valle, E., Gouet-Brunet, V., & Torres, R. d. S. (2014). Visual word spatial arrangement for image retrieval and classification. *Pattern Recognition*, 47(2), 705-720.
- Petitjean, F., Buntine, W., Webb, G. I., & Zaidi, N. (2018). Accurate parameter estimation for Bayesian network classifiers using hierarchical Dirichlet processes. *Machine Learning*, 107(8-10), 1303-1331.
- ping Tian, D. (2013). A review on image feature extraction and representation techniques. *International Journal of Multimedia and Ubiquitous Engineering*, 8(4), 385-396.
- Pinto, C., Furukawa, J., Fukai, H., & Tamura, S. (2017). *Classification of Green coffee bean images basec on defect types using convolutional neural network (CNN)*. Paper presented at the 2017 International Conference on Advanced Informatics, Concepts, Theory, and Applications (ICAICTA).
- Prabuwono, A. S., Usino, W., Yazdi, L., Basori, A. H., Bramantoro, A., Syamsuddin, I., . . . Allehaibi, K. H. S. (2019). Automated Visual Inspection for Bottle Caps Using Fuzzy Logic.
- Qayyum, A., Malik, A. S., Saad, N. M., Iqbal, M., Faris Abdullah, M., Rasheed, W., . . . Bin Jafaar, M. Y. (2017). Scene classification for aerial images based on CNN

- using sparse coding technique. *International Journal of Remote Sensing*, 38(8-10), 2662-2685.
- Qi, X., Wang, T., & Liu, J. (2017). *Comparison of Support Vector Machine and Softmax Classifiers in Computer Vision*. Paper presented at the 2017 Second International Conference on Mechanical, Control and Computer Engineering (ICMCCE).
- Raghupathi, W., & Raghupathi, V. (2014). Big data analytics in healthcare: promise and potential. *Health information science and systems*, 2(1), 3.
- Rahim, A., Chin, C., Abdullah, S., Singh, S., Nuawi, M. Z., & Hassan, F. (2018). Characterization of wavelet decomposition strain signal using the k-mean clustering method. *International Journal of Engineering and Technology (UAE)*, 7(3), 158-162.
- Reza, A. M. (2004). Realization of the contrast limited adaptive histogram equalization (CLAHE) for real-time image enhancement. *Journal of VLSI signal processing systems for signal, image and video technology*, 38(1), 35-44.
- Riyadi, S., Azizah, L. M. r., Damarjati, C., & Hariadi, T. K. (2018). *Evaluation of Mangosteen Surface Quality using Discrete Curvelet Transform*. Paper presented at the 2018 International Conference on Information and Communication Technology Convergence (ICTC).
- Roberts, D. R., Bahn, V., Ciuti, S., Boyce, M. S., Elith, J., Guillera-Aroita, G., . . . Thuiller, W. (2017). Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. *Ecography*, 40(8), 913-929.
- Rosten, E., Porter, R., & Drummond, T. (2010). Faster and better: A machine learning approach to corner detection. *Ieee Transactions on Pattern Analysis and Machine Intelligence*, 32(1), 105-119.
- Rüßmann, M., Lorenz, M., Gerbert, P., Waldner, M., Justus, J., Engel, P., & Harnisch, M. (2015). Industry 4.0: The future of productivity and growth in manufacturing industries. *Boston Consulting Group*, 9.
- Sachdeva, V. D., Fida, E., Baber, J., Bakhtyar, M., Dad, I., & Atif, M. (2017, 27-28 Dec. 2017). *Better object recognition using bag of visual word model with compact vocabulary*. Paper presented at the Emerging Technologies (ICET), 2017 13th International Conference on, Islamabad, Pakistan.
- Saeys, Y., Inza, I., & Larrañaga, P. (2007). A review of feature selection techniques in bioinformatics. *bioinformatics*, 23(19), 2507-2517.
- Sarala, J., & Sivanantham, E. (2014). Design of Multilevel Two Dimensional-Discrete Wavelet Transform For Image Processing Applications. *International Journal of Computing Communication and Information System*, 6(1), 1-6.
- Semeniuta, O., Dransfeld, S., Martinsen, K., & Falkman, P. (2018). Towards increased intelligence and automatic improvement in industrial vision systems. *Procedia Cirp*, 67, 256-261.

- Shin, S.-J., Woo, J., & Rachuri, S. (2014). Predictive analytics model for power consumption in manufacturing. *Procedia Cirp*, 15, 153-158.
- Shukla, M., & Changlani, S. (2013). A comparative study of wavelet and curvelet transform for image denoising. *IOSR Journal of electronics and communication engineering*, 7(4), 63-68.
- Silva, C., Bouwmans, T., & Frélicot, C. (2015, 11–14 March 2015). *An extended center-symmetric local binary pattern for background modeling and subtraction in videos*. Paper presented at the Proceedings of the 10th International Conference on Computer Vision Theory and Applications-Volume 1: VISAPP, Berlin, Germany.
- Singh, A., Thakur, N., & Sharma, A. (2016). *A review of supervised machine learning algorithms*. Paper presented at the 2016 3rd International Conference on Computing for Sustainable Global Development (INDIACom).
- Singha, M., & Hemachandran, K. (2012). Content based image retrieval using color and texture. *Signal & Image Processing*, 3(1), 39.
- Sladojevic, S., Arsenovic, M., Anderla, A., Culibrk, D., & Stefanovic, D. (2016). Deep neural networks based recognition of plant diseases by leaf image classification. *Computational intelligence and neuroscience*, 2016.
- Sokolova, M., & Lapalme, G. (2009). A systematic analysis of performance measures for classification tasks. *Information Processing & Management*, 45(4), 427-437.
- Solomon, C., & Breckon, T. (2011). *Fundamentals of Digital Image Processing: A practical approach with examples in Matlab*: John Wiley & Sons.
- Song, J., Wang, J., & Li, S. (2019). *Dynamic Fuzzy Inference System for Edge Detection of Stone Inscriptions*. Paper presented at the Proceedings of the 2019 3rd International Conference on Digital Signal Processing.
- Song, X., Zhang, G., Liu, F., Li, D., Zhao, Y., & Yang, J. (2016). Modeling spatio-temporal distribution of soil moisture by deep learning-based cellular automata model. *Journal of Arid Land*, 8(5), 734-748.
- Sthevanie, F., & Ramadhani, K. (2018, 2018). *Spoofing detection on facial images recognition using LBP and GLCM combination*. Paper presented at the Journal of Physics: Conference Series.
- Syaliman, K., Nababan, E., & Sitompul, O. (2018). *Improving the accuracy of k-nearest neighbor using local mean based and distance weight*. Paper presented at the Journal of Physics: Conference Series.
- Tang, J., Alelyani, S., & Liu, H. (2014). Feature selection for classification: A review. *Data classification: Algorithms and applications*, 37.
- Tunio, N., Memon, A. L., Khuhawar, F. Y., & Abro, G. M. (2019). Detection of Infected Leaves and Botanical Diseases using Curvelet Transform. *International Journal of Advanced Computer Science and Applications*, 10(1), 516-520.

- USDA. (2017). *Global Agricultural Information Network Report (CH17058)*. Retrieved from USDA Foreign Agricultural Service, Washington: <https://www.fas.usda.gov>
- Vijayakumar, J., & Durai, L. J. (2017). A Review and Performance Analysis of Image Edge Detection Algorithms. In: IJFRCSCCE.
- Wan, J., Tang, S., Li, D., Wang, S., Liu, C., Abbas, H., & Vasilakos, A. V. (2017). A manufacturing big data solution for active preventive maintenance. *IEEE Transactions on industrial informatics*, 13(4), 2039-2047.
- Wang, J., Ma, Y., Zhang, L., Gao, R. X., & Wu, D. (2018). Deep learning for smart manufacturing: Methods and applications. *Journal of Manufacturing Systems*, 48, 144-156.
- Wang, Z., Hu, M., & Zhai, G. (2018). Application of deep learning architectures for accurate and rapid detection of internal mechanical damage of blueberry using hyperspectral transmittance data. *Sensors*, 18(4), 1126.
- Warif, N. B. A., Wahab, A. W. A., Idris, M. Y. I., Ramli, R., Salleh, R., Shamshirband, S., & Choo, K.-K. R. (2016). Copy-move forgery detection: survey, challenges and future directions. *Journal of Network and Computer Applications*, 75, 259-278.
- Warif, N. B. A., Wahab, A. W. A., Idris, M. Y. I., Salleh, R., & Othman, F. (2017). SIFT-symmetry: a robust detection method for copy-move forgery with reflection attack. *Journal of Visual Communication and Image Representation*, 46, 219-232.
- Wei, W., Polap, D., Li, X., Woźniak, M., & Liu, J. (2018). *Study on remote sensing image vegetation classification method based on decision tree classifier*. Paper presented at the 2018 IEEE Symposium Series on Computational Intelligence (SSCI).
- Wei, X., Yang, Z., Liu, Y., Wei, D., Jia, L., & Li, Y. (2019). Railway track fastener defect detection based on image processing and deep learning techniques: A comparative study. *Engineering Applications of Artificial Intelligence*, 80, 66-81.
- Wright, S., & Marwala, T. (2008). Artificial intelligence techniques for steam generator modelling. *arXiv preprint arXiv:0811.1711*.
- Wu, C., Liu, Z., & Jiang, H. (2014). Catenary image enhancement using wavelet-based contourlet transform with cycle translation. *Optik-International Journal for Light and Electron Optics*, 125(15), 3922-3925.
- Xiao, M., Jiang, M., Li, G., Xie, L., & Yi, L. (2017). An evolutionary classifier for steel surface defects with small sample set. *Eurasip Journal on Image and Video Processing*, 2017(1), 48.
- Xiao, Z., Zhang, X., Geng, L., Zhang, F., Wu, J., & Liu, Y. (2019). Research on the Method of Color Fundus Image Optic Cup Segmentation Based on Deep Learning. *Symmetry*, 11(7), 933.
- Xie, L., Lee, F., Liu, L., Yin, Z., Yan, Y., Wang, W., . . . Chen, Q. (2018). Improved Spatial Pyramid Matching for Scene Recognition. *Pattern Recognition*.

- Yan, S., Xu, X., Xu, D., Lin, S., & Li, X. (2012). Beyond spatial pyramids: A new feature extraction framework with dense spatial sampling for image classification. In *Computer Vision–ECCV 2012* (pp. 473-487): Springer.
- Yan, S., Xu, X., Xu, D., Lin, S., & Li, X. (2015). Image classification with densely sampled image windows and generalized adaptive multiple kernel learning. *Cybernetics, IEEE Transactions on*, *45*(3), 381-390.
- Zhang, B., Gu, B., Tian, G., Zhou, J., Huang, J., & Xiong, Y. (2018). Challenges and solutions of optical-based nondestructive quality inspection for robotic fruit and vegetable grading systems: A technical review. *Trends in food science & technology*, *81*, 213-231.
- Zhang, B., Huang, W., Gong, L., Li, J., Zhao, C., Liu, C., & Huang, D. (2015). Computer vision detection of defective apples using automatic lightness correction and weighted RVM classifier. *Journal of food Engineering*, *146*, 143-151.
- Zhang, B., Huang, W., Li, J., Zhao, C., Fan, S., Wu, J., & Liu, C. (2014). Principles, developments and applications of computer vision for external quality inspection of fruits and vegetables: A review. *Food Research International*, *62*, 326-343.
- Zhang, H.-B., Zhang, Y.-X., Zhong, B., Lei, Q., Yang, L., Du, J.-X., & Chen, D.-S. (2019). A Comprehensive Survey of Vision-Based Human Action Recognition Methods. *Sensors*, *19*(5), 1005.
- Zhang, J., Wang, N., Yuan, L., Chen, F., & Wu, K. (2017). Discrimination of winter wheat disease and insect stresses using continuous wavelet features extracted from foliar spectral measurements. *Biosystems Engineering*, *162*, 20-29.
- Zhang, J. H., Meng, J., Wang, R., & Hou, G. L. (2014). Fault Classification of Waste Heat Recovery System Based on Support Vector Machine. *Proceedings of the 2014 9th Ieee Conference on Industrial Electronics and Applications (Icica)*, 733-737. Retrieved from <Go to ISI>://WOS:000349638300139
- Zhang, Y., Chen, Y., Huang, C., & Gao, M. (2019). Object Detection Network Based on Feature Fusion and Attention Mechanism. *Future Internet*, *11*(1), 9.
- Zhang, Y., Cui, X., Liu, Y., & Yu, B. (2018). Tire defects classification using convolution architecture for fast feature embedding. *International Journal of Computational Intelligence Systems*, *11*(1), 1056-1066.
- Zhang, Z. X., Tan, T. N., Huang, K. Q., & Wang, Y. H. (2012). Three-Dimensional Deformable-Model-Based Localization and Recognition of Road Vehicles. *Ieee Transactions on Image Processing*, *21*(1), 1-13. doi:10.1109/Tip.2011.2160954
- Zhao, Z.-Q., Zheng, P., Xu, S.-t., & Wu, X. (2019). Object detection with deep learning: A review. *Ieee Transactions on Neural Networks and Learning Systems*.
- Zolfaghari, S., Noor, S. B. M., Rezazadeh Mehrjou, M., Marhaban, M. H., & Mariun, N. (2018). Broken rotor bar fault detection and classification using wavelet packet signature analysis based on fourier transform and multi-layer perceptron neural network. *Applied Sciences*, *8*(1), 25.