UNIVERSITI PUTRA MALAYSIA

DEVELOPMENT OF A MACHINE VISION SYSTEM FOR WEEDY RICE SEED IDENTIFICATION

RASHIDAH BINTI RUSLAN

FK 2022 44

DEVELOPMENT OF A MACHINE VISION SYSTEM FOR WEEDY RICE SEED IDENTIFICATION

By

RASHIDAH BINTI RUSLAN

Thesis Submitted to the School of Graduate Studies, Universiti Putra Malaysia, in Fulfilment of the Requirements for the Degree of Doctor of Philosophy

December 2021

Abstract of thesis presented to the Senate of Universiti Putra Malaysia in fulfilment of the requirement for the degree of Doctor of Philosophy

DEVELOPMENT OF A MACHINE VISION SYSTEM FOR WEEDY RICE SEED IDENTIFICATION

By

RASHIDAH BINTI RUSLAN

December 2021

Chair : **Assoc. Prof. Siti Khairunniza binti Bejo, PhD**
Faculty : **Engineering**

Weedy rice contamination in certified rice seed has a dramatic impact on the rice seed industry in Malaysia. To ensure the purity of the certified seed, the authorized agency (Department of Agriculture) made a manual inspection of the rice seed samples. The task is laborious and time-consuming as well as very subjective and error-prone as it is influenced by the skills and experience of the operators in identifying the weedy rice seeds within the cultivated rice seed samples. High similarities between weedy rice morphological features and cultivated rice seed make it more challenging to separate the weedy rice effectively. Therefore, this study was formulated to explore the possibility of automating the manual process of distinguishing the weedy rice using a machine vision and machine learning technique. A machine vision prototype (Patent ID: PI2018500018) works as a platform to replace the human vision in identifying the weedy rice seed was developed. The hardware structure configuration includes selecting a suitable imaging system with uniform lighting and designing the seed plate and body case prototype. The finalized prototype was installed with a moving camera attached to the front light and equipped with imaging and features extraction software. Five cultivated rice seeds varieties and weedy rice variants were collected from the Seed Testing Laboratory. The monochrome and RGB images of the seed kernel were acquired using the prototype for classification model development. Each images is comprised of 15 rice seeds acquired on a seed plate. In total, 895 weedy rice and 7350 cultivated rice seed kernels were used. Ninety-four features were extracted from the morphological, colour and textural parameters. Features optimisation was done based on Stepwise Discriminant Analysis (SDA) and Principal Component Analysis (PCA) approaches. The PCA uses features selected from the correlation loading's observation and PCs with the explained variances greater than 10%. The optimised features from the two types of input image were fed to seven machine learning classifiers and trained using a cross-validation technique using single-parameter (RGB Morph, RGB Colour, RGB Texture, Mono Morph, Mono Grey, Mono Texture), and three-parameter-sets (RGB MCT, Mono MGT, RGB Mono MCGT). The models were trained using ML classifiers such as Decision Trees (DT), Discriminant Analysis (DA), Naïve Bayes (NB), K-Nearest Neighbour (KNN), Support Vector Machines (SVM), Ensemble Classifier (EC), and Logistic Regression (LR). The results revealed SDA has a high percentage of features reduction than the CL plot for the single-parameter-set and a low percentage of features reduction for the three-parameters-set. Furthermore, the SDA had higher classification performance among other optimisation methods. For classification performance, RGB MCT dataset (combination of morphology, colour and textural features from RGB images) modeled by the SVM classifier had the best classification accuracy and average correct classification of 98.1% and 93%, respectively. The RGB MCT model used nine morphology, 22 colour, and 12 texture features. The model was proven to achieve high sensitivity (97.4% to 99.8%) and specificity (97.5% to 100%) when tested using different seeds samples. In conclusion, this study contributed to the development of a complete laboratory-scaled machine vision equipped with the classification model using optimised morphology, colour and texture features.

Abstrak tesis yang dikemukakan kepada Senat Universiti Putra Malaysia sebagai memenuhi keperluan untuk ijazah Doktor Falsafah

PEMBANGUNAN SISTEM PENGLIHATAN MESIN BAGI MENGENALPASTI BENIH PADI ANGIN

Oleh

RASHIDAH BINTI RUSLAN

Disember 2021

Pengerusi : Prof. Madya, Siti Khairunniza bt Bejo, PhD
Fakulti : Kejuruteraan

Benih sah yang dicemari oleh biji benih padi angin memberi kesan yang amat besar kepada industri penghasilan benih padi sah di Malaysia. Bagi mengekalkan ketulenan benih padi sah, agensi berautoriti (Jabatan Pertanian Malaysia) telah melakukan pemeriksaan secara manual kepada sampel benih padi. Tugasan ini adalah sukar dan memakan masa yang lama, mudah terdedah kepada kesilapan serta memerlukan penilaian yang subjektif oleh pengendali makmal berdasarkan kemahiran dan pengalaman mereka dalam mengenalpasti benih padi angin di dalam benih padi sah. Terdapat persamaan yang tinggi di antara benih padi angin dan benih padi sah dari segi bentuk morfologi dan ini menyukarkan tugasan untuk memisahkan benih padi angin dengan efektif. Oleh itu, kajian ini telah dirumuskan untuk meneroka kemungkinan untuk mengubah proses manual kepada automatik melalui teknik penglihatan mesin dan pembelajaran mesin. Sebuah prototaip penglihatan mesin (Patent ID: PI2018500018) telah dibangunkan sebagai platform bagi menggantikan proses pengenalan benih padi angin oleh manusia. Konfigurasi struktur perkakasan merangkumi pemilihan sistem pengimejan yang sesuai, sistem penataan cahaya yang seragam, dan mereka bentuk plat benih dan selongsong badan prototaip. Prototaip akhir telah dipasang dengan kamera yang bergerak bersama lampu hadapan dan dilengkapi dengan perisian bagi mengambil gambar biji benih serta mengekstrak ciri-ciri yang terdapat pada imej. Lima jenis benih padi sah dan varian benih padi angin telah dikumpulkan oleh Makmal Pengujian Biji Benih Padi dan digunakan di dalam kajian ini. Imej benih padi telah diambil oleh penglihatan mesin secara monokrom dan RGB untuk digunakan bagi pembangunan model pengelasan. Sembilan puluh empat ciri telah diekstrak merangkumi ciri bagi parameter morfologi, warna dan tekstur. Pengoptimuman ciri-ciri telah dilakukan melalui *Stepwise Discriminant Analysis (SDA)* dan *Principal Component Analysis (PCA).* Kaedah PCA adalah melalui pemilihan ciri-ciri yang di perhatikan di dalam plot kolerasi loading dan juga melalui pemilihan PC yang mempunyai variasi lebih dari 10%. Ciri-ciri optimum yang diambil dari kedua-dua jenis imej telah dimasukkan ke dalam tujuh pengelas pembelajaran mesin dan dilatih dengan kaedah validasi-silang menggunakan set-satu-parameter (RGB Morph, RGB Colour, RGB Texture, Mono

Morph, Mono Grey, Mono Texture), dan set-tiga-parameter (RGB MCT, Mono MGT, RGB MCGT). Latihan model telah menggunakan pengelas pembelajaran mesin seperti *Decision Trees (DT), Discriminant Analysis (DA), Naïve Bayes (NB), K-Nearest Neighbour (KNN), Support Vector Machines (SVM), Ensemble Classifier (EC),* dan *Logistic Regression (LR).* Keputusan kajian menunjukkan SDA mempunyai peratusan tinggi dalam menurunkan pemilihan ciri-ciri berbanding CL plot bagi set-satu-parameter dan peratusan yang rendah bagi set-tiga-parameter. Sementelah itu, SDA juga mempunyai peratusan yang tinggi bagi keputusan pengkelasan berbanding kaedah pengoptimuman ciri-ciri yang lain. Bagi keputusan pengelasan, data RGB MCT yang telah melalui pengelas SVM telah berjaya meningkatkan kejituan dan purata pengelasan betul dari set-satu-parameter kepada 98.1% dan 93%. Model itu telah menggunakan sembilan ciri morfologi, 22 ciri warna dan 12 ciri tekstur. Ia juga telah dipilih sebagai model terbaik bagi pengkelasan benih padi angin dari benih padi sah dan terbukti mencapai tahap kepekaan (97.4% to 99.8%) dan kekhukusan (97.5% to 100%) yang tinggi apabila diuji dengan sampel benih yang baru. Secara rumusan, kajian ini telah memberi sumbangan di dalam membangunkan sebuah sistem penglihatan mesin pada skala makmal dan dilengkapi dengan model pengkelasan yang menggunakan ciri-ciri optimum morfologi, warna dan tekstur.

ACKNOWLEDGEMENT

With the name of Allah the Most Compassionate and Most Merciful

Alhamdulillah, all praise and thanks to Almighty Allah, with His blessing giving me the strength and passion, could manage to finish the research until this manuscript completed be compiled.I would like to express my deep gratitude and appreciation to my supervisor, Associate Profesor Dr Siti Khairunniza bt Bejo, for her passion and patience in my research. I would also like to thank her for her guidance and dedication. Special thanks to my supervisory committee, Dr Mahirah Jahari and Dr Ir Ibni Hajar Rukunuddin, for their supports and encouragement throughout this study.

I want to acknowledge the Ministry of Higher Education, Malaysia, for sponsoring this research under research grant PRGS for the prototype development and the opportunity for SLAB scholarships. Thank you to Universiti Putra Malaysia for once again I can be part of this prestigious university. To Universiti Malaysia Perlis, thank you for the study leave opportunity and sustenance for my family.

Finally, thank you very much for the endless prayer, support, and patience to my family and friends. Without them, this journey may not have been completed.
.

This thesis was submitted to the Senate of Universiti Putra Malaysia and has been accepted as fulfilment of the requirement for the degree of Doctor of Philosophy. The members of the Supervisory Committee were as follows:

Siti Khairunniza binti Bejo, PhD
Associate Professor
Faculty of Engineering
Universiti Putra Malaysia
(Chairman)

Mahirah binti Jahari, PhD
Senior Lecturer
Faculty of Engineering
Universiti Putra Malaysia
(Member)

____
ZALILAH MOHD SHARIFF, PhD
Professor and Dean
School of Graduate Studies
Universiti Putra Malaysia

Date: 9 June 2022

Declaration by Members of the Supervisory Committee

This is to confirm that:

- the research and the writing of this thesis were done under our supervision;
- supervisory responsibilities as stated in the Universiti Putra Malaysia (Graduate Studies) Rules 2003 (Revision 2015-2016) are adhered to.

Signature: ______________________

Name of Chairman of Supervisory Committee: Associate Professor Dr. Siti Khairunniza binti Bejo

Signature: ______________________

Name of Member of Supervisory Committee: Dr Mahirah binti Jahari

TABLE OF CONTENTS

			Page
ABSTRACT			i
ABSTRAK			iii
ACKNOWLEDGEMENTS			v
APPROVAL			vi
DECLARATION			viii
LIST OF TABLES			xiii
LIST OF FIGURES			xvii
LIST OF ABBREVIATIONS			xix

CHAPTER

1	**INTRODUCTION**		1
	1.1	Research background	1
	1.2	Problem statement	3
	1.3	Objective	4
	1.4	Project scope and limitations	4
	1.5	Structure of thesis	5
2	**LITERATURE REVIEW**		
	2.1	Introduction	6
	2.2	Weedy rice issues and challenges	6
		2.2.1 The origin and evolution of weedy rice	7
		2.2.2 Weedy rice traits	7
		2.2.3 Weedy rice management and control	10
		2.2.4 Paddy seed certification scheme	11
	2.3	Application of machine vision in the rice industry	13
	2.4	Machine vision system	18
		2.4.1 Image acquisition	18
		2.4.2 Image illumination	18
		2.4.3 Image processing and analysis	19
	2.5	Machine learning	21
		2.5.1 Parameters optimisation	22
		2.5.2 Machine learning models	23
	2.6	Summary of the literature review	32
3	**MATERIALS AND METHODS / METHODOLOGY**		
	3.1	Overview of the methodology	33
	3.2	Samples collection	33
	3.3	Hardware component development	35
	3.4	Image acquisition setup	36
	3.5	Camera movement and control	39
	3.6	Image processing	40
	3.7	Parameters extraction	43
		3.7.1 Morphological features extraction	43
		3.7.2 Colour and greyscale features extraction	44
		3.7.3 Texture features extraction	45
	3.8	Parameters optimisation	49

		3.8.1 Stepwise Discriminant Analysis (SDA)	50
		3.8.2 Principal Component Analysis (PCA)	50
	3.9	Classification model development	51
	3.10	Model testing	54
	3.11	Summary	55
4	**RESULTS AND DISCUSSION**		
	4.1	Overview of the results and discussion	57
	4.2	Development of the machine vision system	57
		4.2.1 The image acquisition setup	57
		4.2.2 Final prototype of the machine vision	64
	4.3	SDA analyses result	66
		4.3.1 Tests of equality of group means	66
		4.3.2 Summary of the SDA analyses	70
	4.4	PCA analyses result	76
		4.4.1 Parameters optimisation using the Explained Variance	76
		4.4.2 Parameters optimisation using the Correlation Loading (CL) plot	78
		4.4.3 Summary of the PCA analyses	80
	4.5	Comparison of the features reduction between SDA and CL plot	85
	4.6	The classification results	89
		4.6.1 RGB Morph classification result	91
		4.6.2 RGB Colour classification result	103
		4.6.3 RGB Texture classification result	107
		4.6.4 Mono Morph classification result	112
		4.6.5 Mono Grey classification result	115
		4.6.6 Mono Texture classification result	118
		4.6.7 RGB MCT classification result	123
		4.6.8 Mono MGT classification result	127
		4.6.9 RGB Mono MCGT classification result	132
	4.7	Comparison of the classification performance of the single parameter between RGB and Monochrome image	136
		4.7.1 The Morphology parameter	136
		4.7.2 The Colour/Greyscale parameter	137
		4.7.3 The Texture parameter	139
	4.8	Comparison of the classification performance of three parameters combination using the RGB and Monochrome image features	141
	4.9	Comparison of the classifiers' performance among the best-optimised model.	143
	4.10	The classification results of the testing dataset	148
	4.11	The proposed model implementation in industry	152
5	**CONCLUSION AND RECOMMENDATIONS**		
	5.1	Conclusions	
	5.2	Recommendations	

REFERENCES	156
APPENDICES	162
BIODATA OF STUDENT	215
LIST OF PUBLICATIONS	216
LIST OF AWARDS AND ACHIEVEMENT	217

LIST OF TABLES

Table		Page
2.1	The weedy rice plant characteristics	7
2.2	The weedy rice seed morphotype characteristics	8
2.3	Cultural weed management practices in Malaysia	10
2.4	Standard requirement for registered and certified seed under the Paddy Seed Certification Scheme	12
2.5	Application of machine vision in rice seed quality evaluation	14
2.6	Summary of device or software program developed using machine vision in the grain industry.	16
2.7	Machine learning classifiers and their description	24
2.8	Summary of advantages and limitations of machine learning classifier	25
2.9	Overview of machine learning application in the rice industry	28
3.1	Cultivated rice seed and weedy rice samples used in this study	35
3.2	Morphological features and their descriptions	43
3.3	Colour features extracted from RGB image	44
3.4	Greyscale features extracted from the monochrome image	45
3.5	Textural features extracted using GLCM array	46
3.6	Textural features from RGB and monochrome image	48
3.7	Description of the dataset created for the analyses	49
3.8	ML classifiers option by MATLAB	52
3.9	The sign convention of the confusion matrix	53
3.10	Confusion matrix for WR and CR classification model	53
3.11	The formula and descriptions of the performance measure	54
4.1	Area scan camera specification	58
4.2	Various lens specification	60

4.3	Industrial light specification	62
4.4	Seed characteristics based on length and width	63
4.5	Summary of the equality test of the group means for all features	67
4.6	Stepwise discriminant analysis result for the RGB Morph dataset.	70
4.7	Summary of the variables entry and removal at step 9 of the stepwise analysis for the RGB Morph dataset	71
4.8	Summary of the features removed and selected from the SDA analysis	72
4.9	Cumulative explained variance in percentage and eigenvalues of each PC	77
4.10	Summary of the parameter reduction using PCs from PCA	78
4.11	Loading values of the RGB Morph variables on each PC	79
4.12	Summary of the features removed and retained from the CL plot	81
4.13	Percentage of features reduction and the features selected by the SDA and CL for single-parameter-set	87
4.14	Percentage of features reduction and similarities on the features selected by the SDA and CL for three-parameter-set	88
4.15	The classification model accuracy for RGB Morph dataset	91
4.16	The best model accuracy for RGB Morph	93
4.17	Confusion matrix of the machine learning model for RGB Morph using all features	95
4.18	Confusion matrix of the machine learning model for RGB Morph using SDA optimised features	97
4.19	Confusion matrix of the machine learning model for RGB Morph using CL plot optimised feature	99
4.20	Confusion matrix of the machine learning model for RGB Morph using explained variance optimised features	101
4.21	Confusion matrix of the best-optimised parameter for RGB Morph	103
4.22	The best model accuracy for RGB Colour	104

4.23	Confusion matrix of the best-optimised parameter for RGB Colour.	107
4.24	The best model accuracy for RGB Texture.	108
4.25	The confusion matrix of the best-optimised parameter for RGB Texture.	111
4.26	The best model accuracy for Mono Morph.	112
4.27	Confusion matrix of the best-optimised parameter for Mono Morph.	115
4.28	The best model accuracy for Mono Grey	116
4.29	The confusion matrix of the best-optimised parameter for Mono Grey	118
4.30	The best model accuracy for Mono Texture	119
4.31	Confusion matrix of the best-optimised parameter for Mono Texture	122
4.32	The best model accuracy for RGB MCT	123
4.33	The confusion matrix of the best-optimised parameter for RGB MCT	126
4.34	The best model accuracy for Mono MGT	127
4.35	The confusion matrix of the best-optimised parameter for Mono MGT.	131
4.36	The best model accuracy for RGB Mono MCGT	132
4.37	The confusion matrix of the best-optimised parameter for RGB Mono MCGT	135
4.38	Classifiers performance of the best-optimised parameter for RGB Morph and Mono Morph	125
4.39	Classifier's performance of the best-optimised parameter for RGB Colour and Mono Grey	138
4.40	Classifiers performance of the best-optimised parameter for RGB Texture and Mono Texture	140
4.41	Classifiers performance of the best-optimised parameter for RGB MCT, Mono MGT, and RGB Mono MCGT	142
4.42	The summary of the optimisation method and its ML classifiers' performance	144

4.43 Summary of the SDA selected features for the best-optimised RGB MCT, Mono MGT, and RGB Mono MCGT model 146

4.44 Classification results on the testing dataset between WR and MR297, WR and MR219, WR and MR220CL2, WR and RC2, WR and RC8 and WR and CR 150

LIST OF FIGURES

Figure		Page
2.1	Morphotypes of weedy rice found in Peninsular Malaysia, modern elite cultivar (CV-MR219) and (OR) wild rice *Oryza rufipogon*	9
2.2	Manual weedy rice identification from the cultivated rice seed sample	12
3.1	The sample collection of the cultivated rice seed and weedy rice seed used in this study	34
3.2	Schematic design setup of the machine vision system	36
3.3	Sample of (a) one seed holder (b) a seed plate that combined 24 seed holders on perspex glass.	37
3.4	Seed images acquired from the machine vision	37
3.5	The camera movement control in LabVIEW (a) front panel (b) block diagram	39
3.6	The final GUI of the front panel for image processing and features extraction	40
3.7	Block Diagram for RGB image processing and features extraction	41
3.8	Flowchart of the image processing	42
3.9	Flowchart of the ML process to classify the cultivated rice seed and weedy rice seed	56
4.1	The effect of focal length and platform distance on the magnification of the image plane	59
4.2	Image acquired using simple LED configuration taken from a different working distance a) 14 cm b) 23 cm	60
4.3	The design of the seed holder	63
4.4	The finalized design and colour of the seed holder placed on black background	64
4.5	The overall design of the machine vision (a) internal look of the system assembly (b) overall look of the machine vision	65

4.6	Finalized seed plate with four by six seed holders on a perspex plate used for image acquisition	65
4.7	Cumulative explained variance plot of the RGB Morph dataset	77
4.8	The CL plot of the RGB Morph dataset at PC1 and PC2	80
4.9	Process flow diagram of the analysis presented for subchapter 4.6	90
4.10	The CL plot for RGB MCT on PC1 and PC2	145
4.11	Seed sample distribution (score) on the PC1 and PC2 for RGB MCT dataset	148

LIST OF ABBREVIATIONS

ALS	Acetolactate synthase
ANN	Artificial Neural Network
CCD	Charge Couple Device
CIE	Commission International de L'Eclairge
CL	Correlation Loading
CMOS	Complementary metal oxide semiconductor sensor
CNN	Convolutional Neural Network
CR	Cultivated Rice
DA	Discriminant Analysis
DT	Decision Tree
EC	Ensemble Classifier
FDR	False Discovery Rate
FN	False Negative
FNR	False Negative Rate
FOR	False Omission Rate
FP	False Positive
FPR	False Positive Rate
GLCM	Grey Level Co-Occurrence Matrix
GLRM	Grey Level Run Length Matrix
HIS	Hue Intensity Saturation
IADA	Integrated Agricultural Development Area
IMI	Imidazolinone
KADA	Kemubu Agricultural Development Authority
KNN	K Nearest Neighbour

LED	Light Emitting Diode
LR	Logistic Regression
MADA	Muda Agricultural Development Authority
MCGT	Morphology-Colour-Grey-Texture
MCT	Morphology-Colour-Texture
MGT	Morphology-Grey-Texture
ML	Machine Learning
MONO	Monochrome
NAP	National Agriculture Policy
NIPALS	Nonlinear Iterative Partial Least Square
NPV	Negative Predictive Value
PC	Principal Component
PCA	Principal Component Analysis
RGB	Red Green Blue
SDA	Stepwise Discriminant Analysis
SFFS	Sequential Forward Floating Selections
SJPM	Standard Jabatan Pertanian Malaysia
SVM	Support Vector Machine
TN	True Negative
TNR	True Negative Rate
TP	True Positive
TPR	True Positive Rate

CHAPTER 1

INTRODUCTION

1.1 Research Background

Rice consumption is a defining feature of Malaysian culture. Domestic paddy production in Malaysia is concentrated to ten granary areas in Peninsular with an enormous contribution from Muda Agricultural Development Authority (MADA) area, about 38.8%, followed by Kemubu Agricultural Development Authority (KADA) at 9.1% and Integrated Agricultural Development Area (IADA) Barat Laut Selangor at 8.1% (Agrofood Stat, 2016). On average, Malaysia produced 4.47 MT/ha (Firdaus et al., 2020) over a total land cropping area of 0.7 million hectares. Malaysia's rice production remained relatively constant compared to other ASEAN countries such as Vietnam, Thailand, Philippines and Indonesia, which have shown an increasing trend since 1990 due to relatively constant harvested area in the ten granary areas (Che Omar et al., 2019).

Malaysians consume at least three meals a day, consisting of rice in many forms. The rice consumption is expected to increase from 2.7 mil MT to 3.2 mill MT in 2027 as the national population grows. Malaysia produced 1.8 mill MT/ha in 2017 (Firdaus et al., 2020), leaving a wide gap between rice production and consumption. The current gap between production and consumption trend was filled with net imports from Vietnam, Thailand, and other Asia countries.

The government introduced the Seed Incentive program in 2007 under the Third National Agricultural Policy (NAP3) to boost paddy yield production. It outlines the involvement of the private sector for an adequate supply of quality seeds in the country. The private sector's participation increases the supply significantly to 80,000 MT (as of 2016) of high-quality seed from the initial capacity of 24,000 MT in 2007. Together with the Seed Incentive program, the Paddy Seed Certification Scheme (Skim Pengesahan Benih Padi) was introduced to ensure that the farmers supplied high-quality certified paddy seeds. Certified seeds are ensured of quality by laboratory testing on purity, germination, seed moisture content, and the search for unwanted weeds or other seeds.

One of the threats impacting, lowering the yield of paddy production comes from the weedy rice infestation. Weedy rice is a type of weed that looks similar to the cultivated rice plant. The first occurrence of weedy rice was reported to occur sporadically in the Muda area in Kedah state in 1990 (Azmi and Baki., 2007). The problem of weedy rice has now become widespread. In major rice granaries in Malaysia, weedy rice plants have been competitive with cultivated rice (Oryza sativa L.) and dominant throughout the rice planting field (Sudianto et al., 2016). Factors such as the transition from transplanting the seedling to direct seeding (since 1990s) and sharing machinery such as combine harvesters from one granary to another have increased the proliferation of weedy rice across Malaysia's paddy field and made it difficult to control the spread (Ruzmi et al., 2017). The wide adoption of direct seeding increased the weedy rice infestation and spread where manual weeding is difficult.

Various efforts have been made either by private and government agencies to tackle the issues. The introduction of the Clearfied® rice program in FELCRA Seberang Perak in 2010 can control the weedy rice infestation. However, the potential leaching of imidazolinone (IMI) herbicides in the soils and the emerging of acetolactate synthase (ALS) resistant weedy rice is found to nullify the efficacy of Clearfield® rice in long terms usage (Sudianto et al., 2013). Although the IMI herbicides have been discontinued, cultivated rice seed variety MR 220 CL2 developed for Clearfield rice is a popular choice among farmers due to its resistance to pests and diseases and lack of varieties selection.

Reducing and avoiding weedy rice contamination in the cultivated rice seed is crucial. One of the approaches taken by the seed processing plant is to install the indented cylinder machine right before the seed bagging operation. The indented cylinder works by separating the seed according to its seed length. The seed kernel passes from the inlet housing into the machine's interior with pockets or indents. As the machine rotates, seeds fall in the indents are carried and fall out of the pockets into the trough and discharged by screw conveyor. Seed kernels longer or shorter than the indents are remained in the shell and emptied by the outlet casing (Cimbria, n.d). Despite the indented cylinder's installation, there are weedy rice seeds that fall into the trough and escape into the bagging process. Therefore, these measures are not 100% effective due to weedy rice having a similar cultivated rice seed size.

To date, the standard protocol for seed purity and determination of unwanted weed is through a manual sampling process by the laboratory operators. Seed testing requires laboratory workers to count, identify and distinguish between the cultivated rice seed and weedy rice seed/unwanted seed or off-type seed. The maximum allowable unwanted or dangerous weed seeds are ten seeds/kg for certified rice. If the seed producers did not comply with the standard, the seed lot would be rejected, giving the producers losses. The losses also will reduce the required quota for farmers' supply. The effort and standard operating protocols have proved to minimise weedy rice contamination in seed bags produced by the processing plant; however, the task is laborious and exposed to the paddy seeds' misidentification.

Machine vision technology has been utilised in several agricultural sectors, such as application in precision agriculture practice, fresh produce quality assessment, and sorting and classification. The imaging technology coupled with machine learning offers many advantages in sensing as it is relatively accurate, non-destructive, and yields consistent results (Rehman et al., 2019). The ability of digital camera machine vision to recognise two-dimensional data attributes through the pixel size in the image that represents the shape, size, colour and texture in the visible colour region (Chen et al., 2002) gives the advantage to understand the external quality parameter of agricultural samples.

Some of the machine vision applications on grain quality inspection were demonstrated by Kaur and Singh (2013), Pazoki et al. (2014), Golpour et al. (2014), Chaugule and Mali (2014), Anami et al. (2015), Huang and Chien (2017), Ansari et al. (2021). These researchers had proved that the purity assessment using imaging technology and seed separation using classification analysis on different rice seed varieties is possible with high accuracy.

1.2 Problem Statement

It is crucial to avoid contamination of weedy rice seed in the certified rice seed produced by the appointed seed producers due to difficulty controlling the spread in the rice field. The preventive measures in installing the indented cylinder during seed processing help remove the weedy rice from the cultivated rice seed based on its length. However, weedy rice seeds often have a similar size length to the cultivated rice seed escape the indented cylinder and appear in the seed bag. The high similarities of the seed features exhibited in the weedy rice seed are understood as the hybridization of wild Oryza population and Indica rice subspecies. In the meantime, manual quality inspection by the Seed Testing Laboratory is based on the distinctive physical appearance, such as size, hull and kernel colour. However, this task depends on the laboratory personnel's experience in identifying the weedy rice among the cultivated rice seed sample. Furthermore, the task becomes laborious during peak season and exposed to the risk of seed misidentification.

The current techniques that rely on the physical separation based on seed length and physical appearances seen by human eyes do not guarantee effective separation of the weedy rice. Earlier work on weedy rice classification by Aznan et al. (2016) had proved the weedy rice seed is distinguishable from the cultivated rice seed using discriminant function analysis. However, this study only used morphological features of the seed and tested on one rice seed variety. Therefore, there is room for improvements in the weedy rice classification problem beyond the physical separation and the utilization of the morphological features from the image.

Besides the physical appearances, other potential parameters are available such as colour and textural features of the seed, which can be captured in an image and extracted using image processing. The combination of morphology, colour, and texture parameters extracted from seed images is expected to increase weedy rice's classification rate using various machine learning techniques. Therefore, it is vital to automate the weedy rice seed identification process using the image captured through machine vision technique. This research explores the possibility of replacing the manual process of weedy rice seed identification with a machine vision system to reduce the laborious work for inspection and human error. Specifically, this project emphasises developing a laboratory-scale image recognition system to identify weedy rice among the cultivated rice seed.

1.3 Objective

This study aims to develop a machine vision prototype for weedy rice identification within local Malaysian cultivated rice seed varieties using image processing and machine learning technique.

The specific objectives are:

1. To construct the machine vision's hardware configuration systems for image acquisition consists of camera setting, lighting setup and seed plate design, and software programme development for image processing and features extraction.

2. To optimize parameters extracted from the Monochrome and RGB seed kernel images using Stepwise Discriminant Analysis (SDA) and Principal Component Analysis (PCA).

3. To identify the most suitable weedy rice and cultivated rice seed classification model developed using machine learning technique.

1.4 Project Scope and Limitations

This project's scope covers the process of weedy rice seed identification as being practiced in the Seed Testing Laboratory, Jabatan Pertanian Malaysia. Seed testing is one of the quality measures of certified seeds. The proposed machine vision prototype is expected to automate the manual weedy rice seed identification. This project's outcome can also be used in any laboratory whose required to check on the weedy rice seed contamination as prevention before sending samples to the Seed Testing Laboratory for certification.

This project is limited to the following scope.

1. The sample used in this study is based on the five popular rice seed varieties produced in season 2/2018 collected from the Seed Testing Laboratory, Jabatan Pertanian Malaysia in Teluk Cengai, Kedah. The varieties used were MR220 CL2, MR219, RC2 RC8 and MR297. The cultivated rice seed samples were fully matured and obtained from the bag of the certified seeds produced by the processing plants. The seed sample represents the shape, size, and yield of the cultivated rice planted in the Muda area.

2. The samples of weedy rice used in this study are limited to the available weedy rice collection saved by the Seed Testing Laboratory, Jabatan Pertanian Malaysia in Teluk Cengai, Kedah from season 1/2018 to season 2/2018. These seeds have undergone a manual selection and were only found during weedy rice search from the cultivated rice seed samples.

3. The development of the machine vision system considers the standard procedure conducted by the Seed Testing Laboratory, Jabatan Pertanian Malaysia. The weedy rice seed needs to be separated, identified, and collected among the cultivated rice seed. Therefore, the sampling method was designed to use a seed plate instead of scattered seed placement for easier identification of the weedy rice. The seed plate also allows for consistent data collection of the seed images specifically for model development.

4. The machine learning (ML) classifier used in this study is limited to supervised ML.

1.5 Structure of Thesis

Chapter 2 presents a review of the literature. The weedy rice issues in the Malaysian context were summarised. Further, the weedy rice characteristics and the potential of machine vision and machine learning techniques to address the issues are discussed. The framework of developing the machine vision, the advantages of machine learning techniques, the previous studies involving machine learning and grain classification were also reviewed and summarised in Chapter 2.

Chapter 3 presents a proposed methodology to conduct this study. The development of the machine vision was outlined starting from the hardware component development. Then the image acquisition setup was presented, and the method for image acquisition was highlighted. Next, the acquired images were processed using LabVIEW explicitly written for this project. The parameters used in this study was presented, and the method for parameter extractions was discussed. The features reduction and optimisation method were also highlighted. Finally, the classification model parameter measures were outlined and defined for the evaluation of the models. Validation of the best model performance using the testing dataset was explained at the end of Chapter 3.

Chapter 4 present the results and discussions on the findings of this research. The first result discussed the selection and limitations of the prototype development, which involved the image acquisition setup. Then, the results of parameters optimisation and the selected features using the stepwise discriminant analyses and the correlation loading plot from the principal component analysis were highlighted. Further, the results of the classification model employed using the machine learning classifiers were presented, and comparative analyses on the best-optimised model were made. Finally, the result of the testing dataset was discussed.

Finally, Chapter 5 presented the conclusions and summarised the findings of this research. The main contributions of this thesis are clearly outlined. Some future recommendations are also presented.

REFERENCES

Abdullah, M. Z., Guan, L. C., and Mohd Azemi, B. M. N. (2001). Stepwise Discriminant Analysis for Colour Grading of Oil Palm Using Machine Vision System. *Food and Bioproducts Processing*, *79*, 223–231.

Anami, B. S., Malvade, N. N., and Palaiah, S. (2019). Automated recognition and classification of adulteration levels from bulk paddy grain samples. *Information Processing in Agriculture*, *6*, 47–60.

Anami, B. S., Naveen, N. M., and Hanamaratti, N. G. (2015). Behavior of HSI Color Co-Occurrence Features in Variety Recognition from Bulk Paddy Grain Image Samples. *International Journal of Signal Processing, Image Processing and Pattern Recognition*, *8*, 19–30.

Ansari, N., Ratri, S. S., Jahan, A., Ashik-e-rabbani, M., and Rahman, A. (2021). Inspection of paddy seed varietal purity using machine vision and multivariate analysis. *Journal of Agriculture and Food Research*, *3*, 100109.

Aznan, A. A., Rukunudin, I. H., Shakaff, A. Y. M., Ruslan, R., Zakaria, A., and Saad, F. S. A. (2016). The use of machine vision technique to classify cultivated rice seed variety and weedy rice seed variants for the seed industry. *International Food Research Journal*, *23*, S31–S35.

Aznan, A. A., Ruslan, R., Rukunudin, I. H., Azizan, F. A., and Hashim, A. Y. (2017). Rice seed varieties identification based on extracted colour features using image processing and artificial neural network (ANN). *International Journal on Advanced Science, Engineering and Information Technology*, *7*, 2220–2225.

Ben-David, A. (2008). About the relationship between ROC curves and Cohen's kappa. *Engineering Applications of Artificial Intelligence*, *21*, 874–882.

Bro, R., and Smilde, A. K. (2014). Principal component analysis. *Analytical Methods*.

Chaugule, A. A., and Mali, S. N. (2016). Identification of paddy varieties based on novel seed angle features. *Computers and Electronics in Agriculture*, *123*, 415–422.

Chaugule, A., and Mali, S. N. (2014). Evaluation of Texture and Shape Features for Classification of Four Paddy Varieties. *Journal of Engineering (United States)*, *2014*.

Chauhan, B. S. (2013). Strategies to manage weedy rice in Asia. *Crop Protection*, *48*, 51–56.

Chen, X., Xun, Y., Li, W., and Zhang, J. (2010). Combining discriminant analysis and neural networks for corn variety identification. *Computers and Electronics in Agriculture*, *71*, 48–53.

Cheng, F., Ying Y.B., and Li, Y.B. (2006). Detection of Defects in Rice Seeds Using Machine Vision. *Transactions of the ASABE, 49*, 1929–1934.

Choudhary, R., Paliwal, J., and Jayas, D. S. (2008). Classification of cereal grains using wavelet, morphological, colour, and textural features of non-touching kernel images. *Biosystems Engineering*, *99*, 330–337.

Chung, C. L., Huang, K. J., Chen, S. Y., Lai, M. H., Chen, Y. C., and Kuo, Y. F. (2016). Detecting Bakanae disease in rice seedlings by machine vision. *Computers and Electronics in Agriculture*, *121*, 404–411.

Deng, X., Zhu, Q., and Huang, M. (2014). Semi-supervised classification of rice seed based on hyperspectral imaging technology Written for presentation at the 2014 ASABE and CSBE / SCGAB Annual International Meeting Sponsored by ASABE Montreal , Quebec Canada, *7004*.

Du, C. J., and Sun, D.W. (2004). Recent developments in the applications of image processing techniques for food quality evaluation. *Trends in Food Science & Technology*, *15*, 230–249.

Duan, L., Yang W., Bi, K., Chen, S., Luo, Q., Liu, Q Chen, S., Y., Luo, Q., and Liu, Q. (2016). Fast discrimination and counting of filled/unfilled rice spikelets based on bi-modal imaging. *Computers and Electronics in Agriculture*, *7*, 55–63.

Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recognition Letters*, *27*, 861–874.

Ganganwar, V. (2012). An overview of classification algorithms for imbalanced datasets. *International Journal of Emerging Technology and Advanced Engineering*, *2*, 42–47.Retrieved from http://www.ijetae.com/files/Volume2Issue4/IJETAE_0412_07.pdf

Golpour, I., Parian, J. A., and Chayjan, R. A. (2014). Identification and classification of bulk paddy, brown, and white rice cultivars with colour features extraction using image analysis and neural network. *Czech Journal of Food Sciences*, *32*, 280–287.

Guo, Q., Wu, W., Massart, D. L., Boucon, C., and De Jong, S. (2002). Feature selection in principal component analysis of analytical data. *Chemometrics and Intelligent Laboratory Systems*, *61*, 123–132.

Guzman, J. D., and Peralta, E. K. (2008). Classification of philippine rice grains using machine vision and artificial neural networks,. *World conference on agricultural information and IT, IAALD AFITA WCCA*, 41–48. Retrieved from https://pdfs.semanticscholar.org/7c6d/c5fbf223d2fee465ddb26e790f4ada0c8b3d.pdf

Hong, P. T. T., Hai, T. T. T., Lan, L. T., Hoang, V. T., Hai, V., and Nguyen, T. T. (2016). Comparative Study on Vision Based Rice Seed Varieties Identification. *Proceedings - 2015 IEEE International Conference on Knowledge and Systems Engineering, KSE 2015*, 377–382.

Huang, K. Y., and Chien, M. C. (2017). A novel method of identifying paddy seed varieties. *Sensors (Switzerland)*, *17*.

Husin, N. A., Khairunniza-bejo, S., and Abdullah, A. F. (2020). Classification of Basal Stem Rot Disease in Oil Palm Plantations Using Terrestrial Laser Scanning Data. *Agronomy*, *10*.

Jahari, M., Yamamoto, K., Miyamoto, M., Kondo, N., Ogawa, Y., Suzuki, T., Habaragamuwa, H., and Ahmad, U. (2015). Double lighting machine vision system to monitor harvested paddy grain quality during head-feeding combine harvester operation. *Machines*, *3*, 352–363.

Kaur, H., and Singh, B. (2013). Classification and Grading Rice Using Multi-Class SVM. *International Journal of Scientific and Research Publications*, *3*, 1–5. Retrieved from www.ijsrp.org

Koklu, M., and Ozkan, I. A. (2020). Multiclass classification of dry beans using computer vision and machine learning techniques. *Computers and Electronics in Agriculture*, *174*, 105507.

Kong, W., Zhang, C., Liu, F., Nie, P., and He, Y. (2013). Rice seed cultivar identification using near-infrared hyperspectral imaging and multivariate data analysis. *Sensors (Basel, Switzerland)*, *13*, 8916–8927.

Kuo, T.Y., Chung, C. L., Chen, S. Y., Lin, H.A., and Kuo, Y.F. (2016). Identifying rice grains using image analysis and sparse-representation-based classification. *Computers and Electronics in Agriculture*, *127*, 716–725.

Kuo, T., Chen, S., Lin, H., Chung, C., and Kuo, Y. (2015). Identifying Rice Grains of Various Cultivars Using Machine Vision Written for presentation at the 2015 ASABE Annual International Meeting Sponsored by ASABE New Orleans , Louisiana, *7004*.

Liakos, K. G., Busato, P., Moshou, D., Pearson, S., and Bochtis, D. (2018). Machine learning in agriculture: A review. *Sensors (Switzerland)*, *18*, 1–29.

Lin, P., Chen, Y., He, J., and Fu, X. (2018). Determination of the Varieties of Rice Kernels Based on Machine Vision and Deep Learning Technology. *Proceedings - 2017 10th International Symposium on Computational Intelligence and Design, ISCID 2017*, *2018-Janua*, 169–172.

Lurstwut, B., and Pornpanomchai, C. (2016). Application of image processing and computer vision on rice seed germination analysis. *International Journal of Applied Engineering Research*, *11*, 6800–6807.

Lurstwut, Benjamaporn, and Pornpanomchai, C. (2017). Image analysis based on color, shape and texture for rice seed (Oryza sativa L.) germination evaluation. *Agriculture and Natural Resources*, *51*, 383–389.

Majumdar, S., and Jayas, D. S. (2000a). Classification of cereal grains usingmachine vision: I. Morphology models. *Transactions of the ASAE*, *43*, 1669–1675.

Majumdar, S., and Jayas, D. S. (2000b). Classification of cereal grains usingmachine vision: II. Color models. *Transactions of the ASAE*, *43*, 1677–1680.

Majumdar, S., and Jayas, D. S. (2000c). Classification of cereal grains using machine vision: III. Texture models. *Transactions of the American Society of Agricultural Engineers*, *43*, 1681–1687.

Majumdar, S., and Jayas, D. S. (2000d). Classification of cereal grains using machine vision: IV Combined morphology, color and texture models. *Transactions of the ASAE*, *43*, 1689–1694.

Mandal, S., Roy, S., and Tanna, H. (2012). A low-cost image analysis technique for seed size determination. *Current Science*, *103*, 1401–1403.

Mebatsion, H. K., Paliwal, J., and Jayas, D. S. (2013). Automatic classification of non-touching cereal grains in digital images using limited morphological and color features. *Computers and Electronics in Agriculture*, *90*, 99–105.

Neelamegam, P., Abirami, S., Vishnu Priya, K., and Valantina, S. R. (2013). Analysis of rice granules using image processing and neural network. *2013 IEEE Conference on Information and Communication Technologies, ICT 2013*, *10*, 879–884.

Ngampak, D., and Piamsa-Nga, P. (2015). Image analysis of broken rice grains of Khao Dawk Mali rice. *Proceedings of the 2015-7th International Conference on Knowledge and Smart Technology, KST 2015*, 115–120.

Paliwal, J., Visen, N. S., Jayas, D. S., and White, N. D. G. (2003). Cereal grain and dockage identification using machine vision. *Biosystems Engineering*, *85*, 51–57.

Patrício, D. I., and Rieder, R. (2018). Computer vision and artificial intelligence in precision agriculture for grain crops: A systematic review. *Computers and Electronics in Agriculture*, *153*, 69–81.

Pazoki, A. R., Farokhi, F., and Pazoki, Z. (2014). Classification of rice grain varieties using two artificial neural networks (mlp and neuro-fuzzy). *Journal of Animal and Plant Sciences*, *24*, 336–343.

Pearson, T. (2009). Hardware-based image processing for high-speed inspection of grains. *Computers and Electronics in Agriculture*, *69*, 12–18.

Qiu, Z., Chen, J., Zhao, Y., Zhu, S., He, Y., and Zhang, C. (2018). Variety identification of single rice seed using hyperspectral imaging combined with convolutional neural network. *Applied Sciences (Switzerland)*, *8*, 1–12.

Rehman, T. U., Mahmud, M. S., Chang, Y. K., Jin, J., and Shin, J. (2019). Current and future applications of statistical machine learning algorithms for agricultural machine vision systems. *Computers and Electronics in Agriculture*, *156*, 585–605.

Ruzmi, R., Ahmad-Hamdani, M. S., Abidin, M. Z. Z., and Roma-Burgos, N. (2021). Evolution of imidazolinone-resistant weedy rice in Malaysia: The current status. *Weed Science*, *69*, 598–608.

Ruzmi, R., Ahmad-Hamdani, M. S., and Bakar, B. B. (2017). Prevalence of herbicide-resistant weed species in Malaysian rice fields: A review. *Weed Biology and Management*, *17*, 3–16.

Shahin, M. A., and Symons, S. J. (2001). A machine vision system for grading lentils. *Canadian Biosystems Engineering / Le Genie des biosystems au Canada*, *43*, 77–714.

Shivrain, V. K., Burgos, N. R., Anders, M. M., Rajguru, S. N., Moore, J., and Sales, M. a. (2007). Gene flow between Clearfield??? rice and red rice. *Crop Protection*, *26*, 349–356.

Singh, K. R., and Chaudhury, S. (2016). Efficient technique for rice grain classification using back-propagation neural network and wavelet decomposition. *IET Computer Vision*, *10*.

Song, B. K., Chuah, T. S., Tam, S. M., and Olsen, K. M. (2014). Malaysian weedy rice shows its true stripes: Wild Oryza and elite rice cultivars shape agricultural weed evolution in Southeast Asia. *Molecular Ecology*, *23*, 5003–5017.

Sudianto, E., Neik, T.-X., Tam, S. M., Chuah, T.-S., Idris, A. A., Olsen, K. M., and Song, B.-K. (2016). Morphology of Malaysian Weedy Rice (Oryza sativa): Diversity, Origin and Implications for Weed Management . *Weed Science*, *64*, 501–512.

Sun, Y., Wong, A. K. C., and Kamel, M. S. (2009). Classification of imbalanced data: A review. *International Journal of Pattern Recognition and Artificial Intelligence*, *23*, 687–719.

Tanabata, T., Shibaya, T., Hori, K., Ebana, K., and Yano, M. (2012). SmartGrain: High-throughput phenotyping software for measuring seed shape through image analysis. *Plant Physiology*, *160*, 1871–1880.

Teoh, C. C., and Bakar, B. H. A. (2009). Immature paddy quantity determination using image processing and analysis techniques (Penentuan kuantiti padi tidak matang dengan menggunakan teknik pemprosesan dan penganalisisan imej). *J. Trop. Agric. and Fd. Sc*, *37*, 241–248.

Velesaca, H. O., Suárez, P. L., Mira, R., and Sappa, A. D. (2021). Computer vision based food grain classification: A comprehensive survey. *Computers and Electronics in Agriculture*, *187*.

Visen, N. S., Paliwal, J., Jayas, D. S., and White, N. D. G. (2004). Image analysis of bulk grain samples using neural networks. *Canadian Biosystems Engineering / Le Genie des biosystems au Canada*, *46*.

Vithu, P., and Moses, J. A. (2016a). Machine vision system for food grain quality evaluation: A review. *Trends in Food Science & Technology*, *56*, 13–20.

Vithu, P., and Moses, J. A. (2016b). Machine vision system for food grain quality evaluation: A review. *Trends in Food Science and Technology*, *56*, 13–20.

Wang, K., and Mittleman, D. M. (2007). Guided propagation of terahertz pulses on metal wires. *NATO Science for Peace and Security Series B: Physics and Biophysics*, 55–68.

Yang, X., Tridandapani, S., Beitler, J. J., Yu, D. S., Yoshida, E. J., Curran, W. J., and Liu, T. (2012). Ultrasound GLCM texture analysis of radiation-induced parotid-gland injury in head-and-neck cancer radiotherapy: An in vivo study of late toxicity. *Medical Physics*, *39*, 5732–5739.

Zareiforoush, H., Minaei, S., Alizadeh, M. R., and Banakar, A. (2015). Potential Applications of Computer Vision in Quality Inspection of Rice: A Review. *Food Engineering Reviews*, *7*, 321–345.

Zhang, B., Huang, W., Li, J., Zhao, C., Fan, S., Wu, J., and 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.

Zhao, C., Xu, W., Song, X., Dai, W., Dai, L., Zhang, Z., and Qiang, S. (2018). Early flowering and rapid grain filling determine early maturity and escape from harvesting in weedy rice. *Pest Management Science*, *74*, 465–476.