FORECASTING AGRICULTURAL CROP YIELD VARIATIONS USING BIG DATA AND SUPERVISED MACHINE LEARNING

Kariyawasan Jalath Thanthrige Dimuthu Lakmal

(188004E)

Thesis/Dissertation submitted in partial fulfillment of the requirements for the degree Master of Science

Department of Computer Science and Engineering

University of Moratuwa
Sri Lanka

March 2020

DECLRATION

I declare that this is my own work and this thesis does not incorporate without acknowledgement any material previously submitted for a Degree or Diploma in any other University or institute of higher learning and to the best of my knowledge and belief it does not contain any material previously published or written by another person except where the acknowledgement is made in the text.

Also, I hereby grant to University of Moratuwa the non-exclusive right to reproduce and distribute my thesis, in whole or in part in print, electronic or other medium. I retain the right to use this content in whole or part in future works (such as articles or books).

Signature: Date:

The above candidate has carried out research for the Masters dissertation under my supervision.

Signature of the supervisor: Date:

ABSTRACT

The government of Sri Lanka is struggling to make appropriate policy decisions regarding paddy cultivation due to absence of accurate and timely data to estimate the paddy yield, land usage for paddy cultivation and area affected by various paddy diseases. Remote sensing data based machine learning implementations can be identified as a potential solution for the above issue, as remote sensing data can be used for accurate and timely estimations. However, the traditional remote sensing data resources have failed to generate accurate estimates regarding cultivated paddy extent estimations. In this study, novel optical remote sensing data resources and a hybrid approach are employed to mitigate previously reported issues. Furthermore, a multi-temporal approach is used instead of traditional mono-temporal approach by leveraging deep neural networks. This study also consists of a comprehensive comparison on novel optical remote sensing data resources and the evaluations of the capability of using deep neural networks for temporal remote sensing analysis. Outcomes of the study shows quite impressive results over 97% of accuracy in terms of cultivated paddy area detection using optical remote sensing imagery. Moreover, the research was extended to identify cultivated paddy areas using synthetic aperture radar (SAR) imagery. It also outputs a promising result over 96% of accuracy in terms of detecting cultivated paddy regions. The study then extends to detect Brown Planthopper attacks in cultivated paddy fields.

Brown Planthopper is considered as the most destructive insect in paddy cultivation. There are no previous studies for identifying Brown Planthopper attacks using satellite remote sensing data under field conditions. In this study, ratio and standard difference indices derived from optical imagery are fed into a Support Vector Machine model to identify the regions affected by Brown Planthopper attacks. Using the results of cultivated paddy fields detection model as a filter, SVM model results are improved. The combined approach shows accuracy over 96% for detecting Brown Planthopper attacks.

Keywords: Remote sensing, Synthetic Aperture Radar, Agriculture, Rice, Deep Neural Networks, SVM, Brown Planthopper, Paddy Yield, Paddy Extent

ACKNOWLEDGEMENT

I would first like to express my gratitude to all my supervisors; Mrs. Vishaka Nanayakkara, Dr. Amal Shehan Perera, Mr. Suranga Jayasena and Mr. Lasantha Fernando who always shared their wisdom and perception to make this research study a success. They constantly steered me in right direction whenever they found I needed it. Further, I would like to thank Dr.Harsha De Silva, honorable state minister of National Policies and Economic Affairs who emphasized practical challenges in policy making during a discussion which inspired me to conduct this study. I appreciate the support I received from the Ministry of National Policies and Economic Affairs and the Ministry of Agriculture during the implementation of my research.

I would also like to acknowledge all the field officers and other officials attached to Ministry of Agriculture for committing their precious time towards this project by supporting me to collect ground data and sharing their valuable insights regarding paddy cultivation.

I am sincerely grateful for Prof. Sanath Jayasena and Dr. Supunmali Ahangama for reviewing this research study at progress reviews and providing their valuable suggestions to improve the project outcomes.

I would also like to thank Mr. Kumaran Kuganathan for reviewing the drafts of research papers and helping me to collect ground data without any hesitation. Further, I thank both academic and nonacademic staff of Department of Computer Science and Engineering for giving their fullest support towards this project.

Last but not least, I would like to express my gratitude to all the members of DataSEARCH center of Department of Computer Science and Engineering for their constant monitoring and invaluable suggestions to make this study a success.

TABLE OF CONTENT

Declration i

Abstact ii

Acknowledgement iii

List of Figures vi

List of Tables vii

List of Abbreviations viii

1 Introduction 1

1.1 Remote Sensing 2

1.2 Motivation and Objectives 3

1.2.1 Motivation 3

1.2.2 Aims and Objectives 4

2 Background and Related Work 6

2.1 Vegetation Indices 6

2.1.1 Normalized Vegetation Index 6

2.1.2 Land Surface Water Index 7

2.2 Neural Networks 7

2.2.1 Feed Forward Neural Networks 8

2.2.2 Recurrent Neural Networks 8

2.2.3 Long Short Term Memory Network 9

2.2.4 Convolution Neural Networks 11

2.2.5 Support Vector Machines 12

2.3 Passive Remote Sensing Data with Low Spatial and Temporal Resolution for Crop Land Classification 153

2.4 Passive Remote Sensing Data with High/Medium Spatial and Temporal Resolution for Crop Land Classification 153

2.5 Active Remote Sensing Data for Crop Land Classification 14

2.6 Paddy Disease Detection 15

3 Cultivate Paddy Extent Detection 17

3.1 Cultivated Paddy Extent Detection Using Passive Remote Sensing Data 17

3.1.1 Study Area 17

3.1.2 Study Data 19

3.1.3 Preprocessing Steps 21
3.1.4 Data Partitioning 25
3.1.5 Experiment Setup 25
3.1.5.1 Paddy Land Classification using NDVI 25
3.1.5.2 Paddy Land Classification using PlanetScope NDVI and Sentinel-2 LSWI 26
3.1.5.3 Paddy Land Classification using Sentinel-2 NDVI and Sentinel-2 LSWI 26
3.1.5.4 Paddy Land Extent Estimation 2930
3.1.6 Results 30
3.2 Cultivated Paddy Extent Detection Using Active Remote Sensing Data 35
3.2.1 Study Area 35
3.2.2 Study Data 35
3.2.3 Preprocessing Steps 37
3.2.4 Experiment Setup 40
3.2.5 Results 40
4 Brown Planthopper Attacks Detection 42
4.1 Study Area 43
4.2 Study Data 44
4.3 Preprocessing Steps 44
4.4 Experiment Setup 50
4.5 Results 52
5 Agri AI Platform 54
6 Conclusion 57
7 References 60

LIST OF FIGURES

		Page
Figure 1.1	Active vs Passive Remote Sensing	3
Figure 2.1	Architecture of Feed Forward Neural Network	8
Figure 2.2	Architecture of a simple Recurrent Neural Network	9
Figure 2.3	Architecture of a LSTM unit cell	11
Figure 2.4	A simple architecture of a 1D CNN	12
Figure 3.1	Agro-ecological regions and locations of sample paddy fields	19
Figure 3.2	PlanetScope NDVI time series graphs	22
Figure 3.3	Sentinel-2 NDVI time series graphs	23
Figure 3.4	Sentinel-2 LSWI time series graphs	23
Figure 3.5	Preprocessing Procedure	24
Figure 3.6	Paddy extent estimation for a paddy field in dry zone	33
Figure 3.7	Paddy extent estimation for a paddy field in intermediate zone	34
Figure 3.8	Paddy extent estimation for a paddy field in wet zone	34
Figure 3.9	Agro-ecological regions and locations of sample paddy fields	36
Figure 3.10	Preprocessing steps applied for SAR images	38
Figure 3.11	GLCM Variance Time Series	39
Figure 3.12	Performance of CNN in terms of Cultivated Land Detection	41
Figure 4.1	Brown Planthopper	42
Figure 4.2	Brown Planthoppers feed by sucking the juice in Paddy Plants	43
Figure 4.3	An area with Hopperburn	43
Figure 4.4	Preprocessing Steps for BPH attack detection	47
Figure 4.5	Time series characteristics of spectral bands	47
Figure 4.6	Time series characteristics of ratio indices	48
Figure 4.7	Time series characteristics of standard difference indices	49
Figure 4.8	Identified damaged regions by BPH attacks	51
Figure 4.9	Results of the SVM model after filtering out non-cultivated areas	51
Figure 5.1	Agro AI Architecture	55
Figure 5.2	Agri AI Web Interface	56
Figure 5.3	User can draw a region in the map to get cultivated extent.	56

LIST OF TABLES

		Page
Table 3.1	Performance Evaluation of Machine Learning Models built based on PlanetScope NDVI	28
Table 3.2	Performance Evaluation of Machine Learning Models built based on Sentinel-2 NDVI	28
Table 3.3	Performance Evaluation of Machine Learning Models built based on PlanetScope NDVI and Sentinel-2 LSWI	28
Table 3.4	Performance Evaluation of Machine Learning Models built based on Sentinel-2 NDVI and Sentinel-2 LSWI	29
Table 3.5	Paddy Extent Estimation Performance Evaluation	33
Table 4.1	Index attributes with correlations greater than 0.7	50

LIST OF ABBREVIATIONS

Abbreviation	Description
SAR	Synthetic Aperture Radar
BPH	Brown Planthopper
NDVI	Normalized Difference Vegetation Index
LSWI	Land Surface Water Index
NDWI	Normalized Difference Water Index
RI	Ratio Index
SDI	Standard Difference Index
CNN	Convolution Neural Network
LSTM	Long Short Term Memory
SVM	Support Vector Machine
STARFM	Spatial and Temporal Adaptive Reflectance Fusion Model
MODIS	Moderate Resolution Imaging Spectroradiometer