A CONVOLUTIONAL NEURAL NETWORK (CNN) BASED APPROACH TO RECOGNIZE MEDICINAL PLANTS BY ANALYZING PLANT LEAF

W.D.L.S.A Costa

(179312U)

M.Sc. in Computer Science

Department of Computer Science and Engineering

University of Moratuwa

Sri Lanka

May 2019

A CONVOLUTIONAL NEURAL NETWORK (CNN) BASED APPROACH TO RECOGNIZE MEDICINAL PLANTS BY ANALYZING PLANT LEAF

W.D.L.S.A Costa

(179312U)

Dissertation submitted in partial fulfilment of the requirements for the Degree of M.Sc. in Computer Science specializing in Software Architecture

Department of Computer Science and Engineering

University of Moratuwa

Sri Lanka

May 2019

DECLARATION

Student Declaration:

I declare that this is my own work and this dissertation 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 dissertation, 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:……

Student Name: WDLSA Costa

Student Registration Number: 179312U

Supervisor Declaration:

The above candidate has carried out research for the Master's Dissertation under my supervision.

Signature: …….

Date:……

Supervisor Name:

ABSTRACT

Sri Lanka is naturally gifted with a vast spread of flora and fauna throughout the country. Most of these plants can be used for therapeutic purposes like when in ancient days, our ancestors used them vastly and considered them to be one of the most highly efficient medical systems in the world at that time. However, with the dawn of modern medicine, indigenous medicine has been decreasing in usage due to factors such as the lack of knowledge about medical plants, the desire for fast recovery, and the reducing interest in traditional treatments due to their smells and appearances. Out of the above, the lack of knowledge about medical plants has been identified as the most contributing factor that demotivates the general public from using traditional medicine. Hence it is evident that a reliable and easy-to-use application to identify and analyse medical plants would be a timely solution to increase the use of traditional medicine in society today. The main objective of this research was to review the features used for leaf recognition, evaluate existing leaf-based medicinal recognition systems, and design a system that would address the loopholes in the available solutions. As such, the researcher carried out a comprehensive literature survey and reviewed existing classification methodologies like Support Vector Machine, Principle Component Analysis, Probabilistic Neural Network and Conventional Neural Network to assess what the best methodology for the above task would be. Due to its feature-extraction capability and high levels of accuracy, Conventional Neural Network was selected as the best approach for this study. Based on the selected method, the researcher designed and developed a feasible application that is capable of finding medical plants by the features of its leaf and medical values. Once the system was built, the researcher distributed surveys and conducted interviews in order to critically test and evaluate the application among industrial experts as well as general users. An overall recognition rate of 85% was recorded by the system, which was well-appreciated by the experts. However, recommendations like probable scope expansions and the need for higher response times were also suggested in the feedback received.

Keywords: medicinal plant identification, Conventional Neural Network, medical plant values, identification of leaf properties

ACKNOWLEDGMENT

First and foremost, I thank and praise Almighty God without whose love and blessings this research would never be completed.

I thank my loving wife and my parents who were always there for me at every step of the way with their prayers and words of encouragement. I also thank my sister and Dilshan who were always there to lend me their support.

I thank my supervisor Dr. Indika Perera, for agreeing to supervise my project, and for his patience, support and guidance.

Finally, I would like to thank Trisha, Tiran, Vilochane, Ryan, Niroshani, Bimsara, Shashi and Dulanjaya for their never-ending support, suggestions and help.

Table of Contents

DECLARATION 2
ABSTRACT 3
ACKNOWLEDGMENT 4
TABLE OF FIGURES 8
LIST OF TABLES 9
1 INTRODUCTION 10
1.1 Introduction 10
1.2 Introduction to Medicinal Plants 10
1.3 Project Background 10
1.4 Objectives 11
1.5 Research Question 12
1.6 Motivation 12
1.7 Organization of the report 12
1.7.1 Chapter 02 13
1.7.2 Chapter 03 13
1.7.3 Chapter 04 13
1.7.4 Chapter 05 13
1.7.5 Chapter 06 13
1.8 Summary 14
2 LITERATURE REVIEW 15
2.1 Introduction 15
2.2 Introduction to Plant Anatomy 15
2.2.1 Flowering Plants 15
2.2.2 Non-Flowering Plants 16
2.2.3 Leaf Arrangement 16
2.2.4 Leaf Type and Leaf Shape 17
2.3 Pre-processing and Feature Extraction 18
2.3.1 Introduction to Pre-processing and Feature Extraction 18
2.3.2 Shape based calculations 19
2.3.3 Colour-based Calculations 20
2.3.4 Venation-based Calculations 20

2.4 Image Processing using Raspberry Pi Processor 21
2.5 Classification Methods 22
2.5.1 Principal Component Analysis (PCA) 22
2.5.2 Kernel PCA 23
2.5.3 Colour PCA 23
2.5.4 Bag of Visual Words Approach 23
2.5.5 Knowledge-based Approach 24
2.5.6 Using Grey Level Co-Occurrence Matrix 24
2.5.7 Support Vector Machine 25
2.5.8 Probabilistic Neural Networks 26
2.5.9 Convolutional Neural Networks 26
2.6 Summary 28
3 METHODOLOGY 29
3.1 Introduction 29
3.2 Overview of the Proposed approach 29
3.3 Front End Layer 30
3.4 Service Layer 31
3.5 Image Processing Layer 32
3.6 Prototype Development Methodology 33
3.7 Summary 34
4 IMPLEMENTATION 35
4.1 Introduction 35
4.2 Technology Stack Overview 35
4.2.1 Language Selection for Front End Development 35
4.2.2 Language Selection for Middle Layer Development 36
4.2.3 Language Selection for Backend Development 36
4.3 Connectivity of Middle Layer and Back End 36
4.4 Implemented Layers 37
4.4.1 Front End Layer 37
4.4.2 Middle Layer 38
4.4.3 Backend Layer 38
4.5 Summary 41
5 TESTING AND EVALUATION 42

5.1 Introduction 42
5.2 Testing 42
5.3 Testing Results 43
5.4 Test Results Summary 54
5.5 Performance Measure of the Classification 54
5.6 Evaluation 55
5.6.1 Evaluation Approach and Selected Evaluators 55
5.7 Summary 60
6 CONCLUSION 62
6.1 Introduction 62
6.2 Achievement of Aim and Objectives 62
6.3 Achievement of Objectives 62
6.4 Problems Encountered During the Research 63
6.5 Limitations 64
6.6 Future Enhancements 64
6.7 Key findings 65
6.8 Conclusion 65
References 66

TABLE OF FIGURES

Figure 2.1:Seedings, leaf venation and floral parts are shown....16
Figure 2.2: Leaf Arrangement....16
Figure 2.3:Leaf Types....17
Figure 2.4:Leaf Shapes....18
Figure 2.5:Types of leaf features [3]....19
Figure 2.6:Leaf length and Leaf Breadth[3]....20
Figure 2.7:Mean and Standard Deviation of the pixels of the leaf....20
Figure 2.8:A block Diagram for Raspberry Processor used Image Processing[4]....21
Figure 2.9:Overview of using Raspberry Processor for Image Processing[4]....22
Figure 2.10:A GLCM based approach[9] 25
Figure 2.11: A sample Convolutional Neural Network [12]....26
Figure 3.1: Overview of the Proposed system....30
Figure 3.2:Front End Overview....31
Figure 3.3:Overview of the service layer....32
Figure 3.4:High level view of the Image Processing part....33
Figure 4.1:Front End of the Application....37
Figure 4.2:A view of the training dataset....39
Figure 4.3:Overview of the Convolutional Neural Network used....39
Figure 4.4:Training Module- Define Layers 40
Figure 4.5:Training Module 40
Figure 4.6:Image Resize and Classification....41
Figure 5.1:Confusion Matrix for Randomly Selected Seven Classes....53
Figure 5.2:Accuracy Measures for the Training Dataset 55
Figure 5.3: Evaluator and User Feedback....59

LIST OF TABLES

Table 5.1:Basic Test Cycle for Plant Leaves 43
Table 5.2:Critical Testing for Aguna 46
Table 5.3:Critical Testing for Kohomba 47
Table 5.4:Critical Testing for Kuppameniya.....48
Table 5.5:Critical Testing for Ranawara 49
Table 5.6: Critical Testing for Katupila 50
Table 5.7:Critical Testing for Thebu 51
Table 5.8:Critical Testing for Hathawariya.....52
Table 5.9:Expert Evaluation1 56
Table 5.10:Expert Evaluation2 57
Table 5.11:Expert Evaluation3 58