**GUIDELINES FOR INDUSTRIAL TRAINING REPORT**

**Institute of Engineering & Technology**



**Guidelines for Industrial Training report**

1. The Industrial Training Report should consist of 30-35 pages.

2. The order of contents of report

a. Cover Page (Format given) b. Declaration (Format given) c. Acknowledgement

d. Abstract

e. Table of Contents (Format given)

f. Chapters

3. Format

Paper: A4

Font:

Header

Footer

Times New Roman,

12pt normal text,

16pt Bold for Chapter headings

14pt Bold for paragraph headings

1.5 line spacing

Upper Left – Name of the Project

Upper Right – Name of the Chapter

Lower Left – Dept. of CEA, GLAU, Mathura

Lower Right – Page No.

4. Numbering of pages:

a. Numbering of pages before the start of the chapter 1 is done in Roman.

Pages will be counted from title page, however number is not mentioned on the title page.

b. For chapters, see the attached report template.

5. The report should be spiral bound.

**Arrangement of Chapters**

The following is suggested format for arranging the project report into variousi

chapters:

**1. Introduction**

This chapter must describe introduction about your project

**1.1 Overview and Motivation**

This section includes, why you are motivated to do this project.

**1.2 Objective:**

This section includes, what is your main aim in the project.

**1.3 Summary of Similar Applications**

This section contains the details about similar applications.

**2. Company Profile**

This chapter must describe introduction about company

**3. Project Design**

The design part must include the following items

Data Flow Diagram

UML diagrams. This UML diagrams must include the following

Class Diagrams

Interaction diagrams-Sequence and Collaboration diagrams

Object Diagrams

Use-case diagrams

Database Design

For database projects, the report must include the following items.

o E-R Diagrams

o Tables – explaining all fields and their data types

o Stored procedures (PL/SQL)

**4. Implementation and User Interface**

May include all user interfaces, output screens and descriptions.

**5. References/Bibliography**

**6. Appendices**

Coding /Code Templates

Consist of coding or code outline for various files

Explain each class with functionality and methods with input and output parameters.

**Table of Contents**

Abstract ii Certificate iii Acknowledgments iv

1. **Introduction** (This chapter must describe introduction about your project) **1**

1.1 Overview and Overview ………………………………........ 1

1.2 Objective .………………………………………………….... 2

2. **Company Profile**

2.1 Define ………………………………………………………..

2.2 ……………… ……………..…………………………………

3. **Project / Training Design**

3.1 ……………………….…………………………

3.2 ……………………..………………….

4. **Project Implementation / Training Details**

4.1 ………….………………………………………

4.2 ……………………..………………….

**5. Implementation and User Interface**

5.1 ……………………………........

5.2 ….……………………………………………...

**References/Bibliography**

**Appendices**

**AN INDUSTRIAL TRAINING REPORT**

**on**

**<Prediction Of Flower Species Using ML>**

**Submitted by**

**PARITOSH PANDEY**

**161500374**

Department of Computer Engineering & Applications

**Institute of Engineering & Technology**



**GLA University Mathura- 281406, INDIA August, 2018**



**ent of Computer Engineering and Applications**

**University, Mathura**

**GLAkm. Stone NH#2, Mathura-Delhi Road, P.O. – Chaumuha, Mathura – 281406**

**Departm**

**17**

**Declaration**

I hereby declare that the work which is being presented in the Training Project Report **“ Prediction of flower Species using ML ”,** in partial fulfillment of the requirements for Industrial Project is an authentic record of my own work carried under the supervision of **, , GLA University, Mathura**.

Sign

Name of Candidate : Paritosh Pandey University Roll No.: 161500374

**Certificate**

This is to certify that the Project ,entitled “**Prediction Of flower Spicies Using ML**” designed on Python with machine learning is an original work carried out by B.Tech student PARITOSH PANDEY from G.L.A University under my guidance from 18-june-2018 to 18-july-2018 has successfully completed the requirements to be recognized as digipodium .

**Project Supervisor**

XaidKamil

**Acknowledgement**

I sincerely wish to express my appreciation for the valuable help,which I received during the cor of the thesis by the following:

**MR. Xaid Kamil** for supervising my project,guiding me thoughout my project.

For his interesting lectures on Python with Machine Learnining and for the many discussion we had on Python.

Without his help and support at crucial times in my project could have gone so far.

i**ABSTRACT**

In Machine Learning, we are using semi-automated extraction of knowledge of data for identifying IRIS flower species. Classification is a supervised learning in which the response is categorical that is its values are in finite unordered set. To simply the problem of classification, scikit learn tools has been used. This paper focuses on IRIS flower classification using Machine Learning with scikit tools. Here the problem concerns the identification of IRIS flower species on the basis of flowers attribute measurements. Classification of IRIS data set would be discovering patterns from examining petal and sepal size of the IRIS flower and how the prediction was made from analyzing the pattern to from the class of IRIS flower. In this paper we train the machine learning model with data and when unseen data is discovered the predictive model predicts the species using what it has been learnt from the trained data. Keywords: Classification, Logistic Regression, K Nearest Neighbour, Machine Learning

**1. INTRODUCTION**

The Machine Learning is the subfield of computer science, according to Arthur Samuel in 1959 told “computers are having the ability to learn without being explicitly programmed”. Evolved from the study of pattern recognition and computational learning theory in artificial intelligence machine learning explores the study and construction of algorithms that can learn from and make predictions on data such algorithms overcome following strictly static program instructions by making data-driven predictions or decisions, through building a model from sample inputs. Machine learning is employed in a range of computing tasks where designing and programming explicitly algorithms with good performance is difficult or unfeasible; example applications include email filtering, detection of network intruders, learning to rank and computer vision. Machine learning focuses on the development of computer programs that can teach themselves to grow and change when exposed to new data. It is a research field at the intersection of statistics, artificial intelligence and computer science and is also known as predictive analytics or statistical learning. There are two main categories of Machine learning. They are Supervised and Unsupervised learning and here in this, the paper focuses on supervised learning. Supervised learning is a task of inferring a function from labeled training data. The training data consists of set of training examples. In supervised learning, each example is a pair of an input object and desired output value. A supervised learning algorithm analyze the training data and produces an inferred function, which can be used for mapping new examples. Supervised learning problems can be further grouped into regression and classification problems. Classification problem is when the output variable is a category, such as “red” or “blue” or “disease” and “no disease”. Regression problem is when the output variable is a real value, such as “dollars” or “weight”. In this paper a novel method for Identification of Iris flower species is presented. It works in two phases, namely training and testing. During training the training dataset are loaded into Machine Learning Model and Labels are assigned. Further the predictive model, predicts to which species the Iris flower belongs to. Hence, the expected Iris species is labeled. This paper focuses on IRIS flower classification using Machine Learning with scikit tools. The problem statement concerns the identification of IRIS flower species on the basic of flower attribute measurements. Classification of IRIS data set would be discovering patterns from examining petal and sepal size of the IRIS flower and how the prediction was made from analyzing the pattern to form the class of IRIS flower. In this paper we train the Machine Learning Model with data and when unseen data is discovered the predictive model predicts the species using what it has learn from trained data.

**1.1 Overview**

Many methods have been presented for Identification of Iris Flower Species. Every method employs different strategy. Review of some prominent solutions is presented. The methodology for Iris Flower Species System is described [1]. In this work, IRIS flower classification using Neural Network. The problem concerns the identification of IRIS flower species on the basis of flower attribute measurements. Classification of IRIS data set would be discovering patterns from examining petal and sepal size of the IRIS flower and how the prediction was made from analyzing the pattern to form the class of IRIS flower. By using this pattern and classification, in future upcoming years the unknown data can be predicted more precisely. Artificial neural networks have been successfully applied to problems in pattern classification, function approximations, optimization, and associative memories. In this work, Multilayer feed-forward networks are trained using back propagation learning algorithm. The model for Iris Flower Species System is described [2]. Existing iris flower dataset is preloaded in MATLAB and is used for clustering into three different species. The dataset is clustered using the k-means algorithm and neural network clustering tool in MATLAB. Neural network clustering tool is mainly used for clustering large data set without any supervision. It is also used for pattern recognition, feature extraction, vector quantization, image segmentation, function approximation, and data mining. Results/Findings: The results include the clustered iris dataset into three species without any supervision. The model for Iris Flower Species System is described [3]. The proposed method is applied on Iris data sets and classifies the dataset into four classes. In this case, the network could select the good features and extract a small but adequate set of rules for the classification task. For Class one data set we obtained zero misclassification on test sets and for all other data sets the results obtained are comparable to the results reported in the literature

**1.2 Objective of the Project**

The main objective of the this project is to predict the flower species using machine learning in an optimal in terms of run time onto the embedded system. Various algorithms and methodologies are studied and hard are resources planning will be done to achieve the goal.

**1.3 Main Purpose**

Digipodium in Lucknow. Computer Training Institutes with Address, Contact Number, Photos, Maps. View Digipodium, Lucknow on Justdial.

Digipodium in Hazratganj, Lucknow has been offering professional training to students. It specialises and is well-known for training students as well as working professionals in accounting, web designing, programming languages, hardware and networking. It is run and managed by a seasoned professionals who leads a team of educators and trainers having relevant domain expertise. At this institution, one can get trained in the subject of their choice by opting from a wide range of courses. These easy-to-follow courses are primarily aimed at students, working professionals as well as IT professionals who want to enhance their knowledge and further their career prospects. Located Lower Ground Floor, Rajaram Kumar PLaza, you can find this institution with relative ease at Rajaram Kumar Plaza in Hazratganj. Undoubtedly it is one of the best computer training institutes in Hazratganj, Lucknow.

**Services Offered at Digipodium**

Digipodium in Hazratganj offers short-term courses and certificate courses. Inclusive of comprehensive learning, the long-term programmes feature subjects such as web development, financial accountancy, computer application and programming, information technology, multimedia and web-designing. Some of the short-term courses cover topics like Windows XP, 7, 8, 10, Vista, MS Office, DTP (Desk Top Publishing), Web Designing, Multimedia, Tally ERP 9, C, C++ , and Visual Basic. Walk into this centre all through the week between 10:00 - 20:30.

**1.4 Aim of the Project**

The aim of this project is to identify Flower species using Machine Learning.

**1.5 Project Scope**

Both segmentation and recognition are two major fields in computer vision, there are over hundreds of new publications every year. In the following overview of the related works, we try to cover the works that are mostly related to ours to our best knowledge. Co-segmentation is a relatively new topic. But the idea of unsupervised foreground segmentation has made it increasingly important.

**Company profle**

Digipodium in Lucknow. Computer Training Institutes with Address, Contact Number, Photos, Maps. View Digipodium, Lucknow on Justdial.

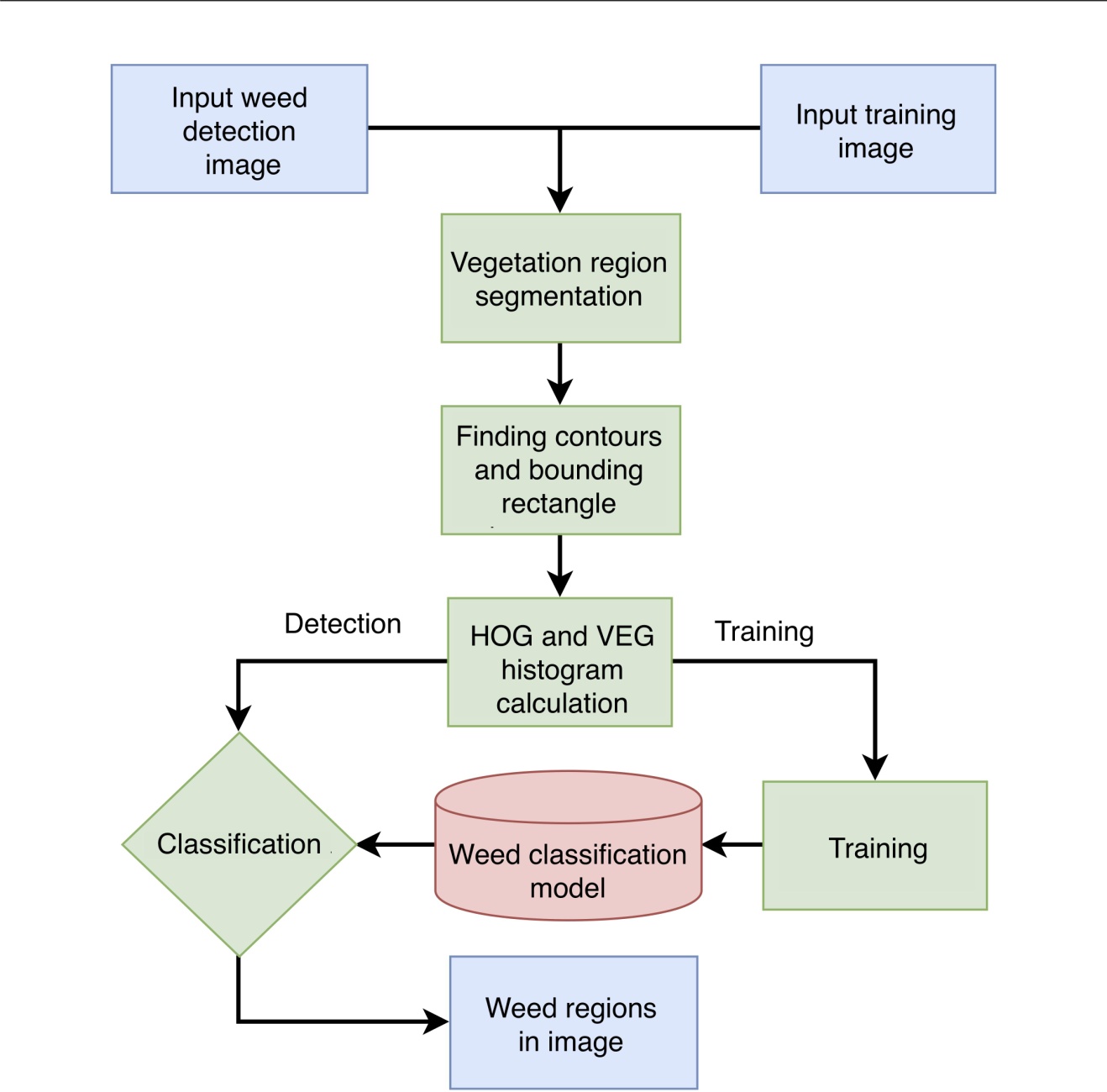
Digipodium in Hazratganj, Lucknow has been offering professional training to students. It specialises and is well-known for training students as well as working professionals in accounting, web designing, programming languages, hardware and networking. It is run and managed by a seasoned professionals who leads a team of educators and trainers having relevant domain expertise. At this institution, one can get trained in the subject of their choice by opting from a wide range of courses. These easy-to-follow courses are primarily aimed at students, working professionals as well as IT professionals who want to enhance their knowledge and further their career prospects. Located Lower Ground Floor, Rajaram Kumar PLaza, you can find this institution with relative ease at Rajaram Kumar Plaza in Hazratganj. Undoubtedly it is one of the best computer training institutes in Hazratganj, Lucknow.

**Services Offered at Digipodium**

Digipodium in Hazratganj offers short-term courses and certificate courses. Inclusive of comprehensive learning, the long-term programmes feature subjects such as web development, financial accountancy, computer application and programming, information technology, multimedia and web-designing. Some of the short-term courses cover topics like Windows XP, 7, 8, 10, Vista, MS Office, DTP (Desk Top Publishing), Web Designing, Multimedia, Tally ERP 9, C, C++ , and Visual Basic. Walk into this centre all through the week between 10:00 - 20:30.

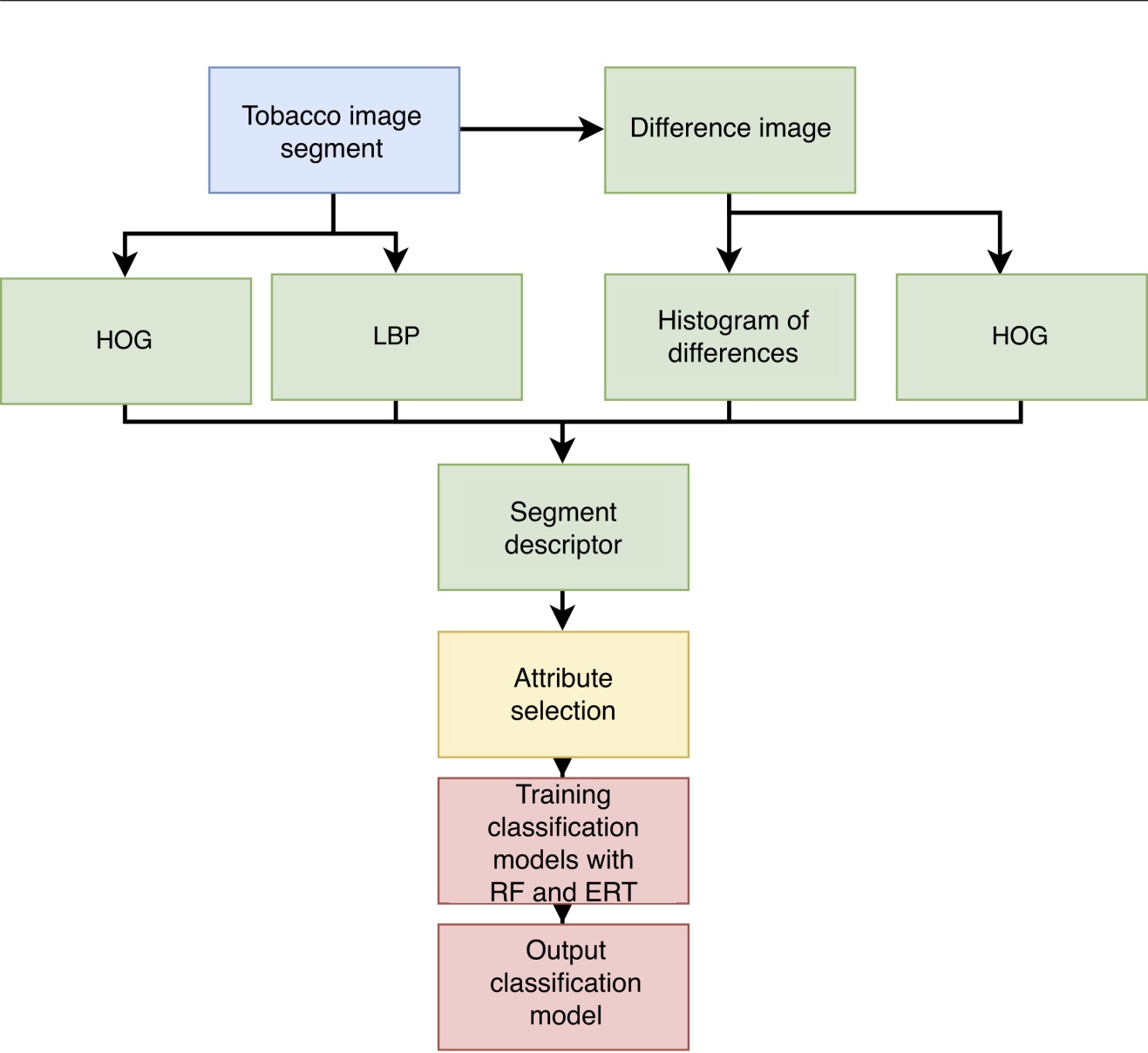
**Design poject**

**3.1 Data Flow Diagram**

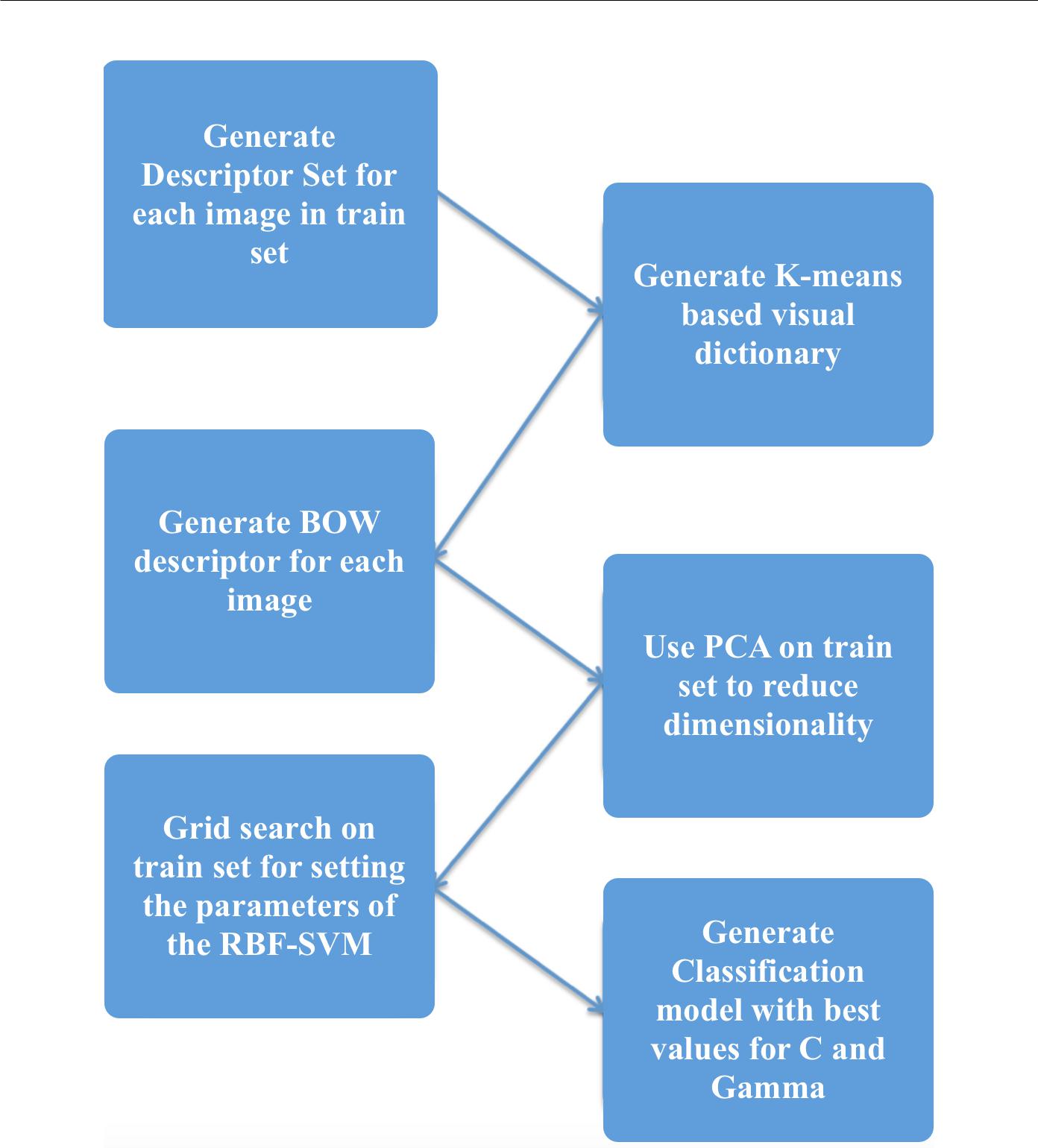


**3.2 Class Diagram**

class hierarchy for recognizition of flower species -we can edit the template and create our own diagram. Creately diagrams can be exported and added to Word ,PPT(powerpoint),Excel,Visio or any other document.



**3.3a sequence diagram**

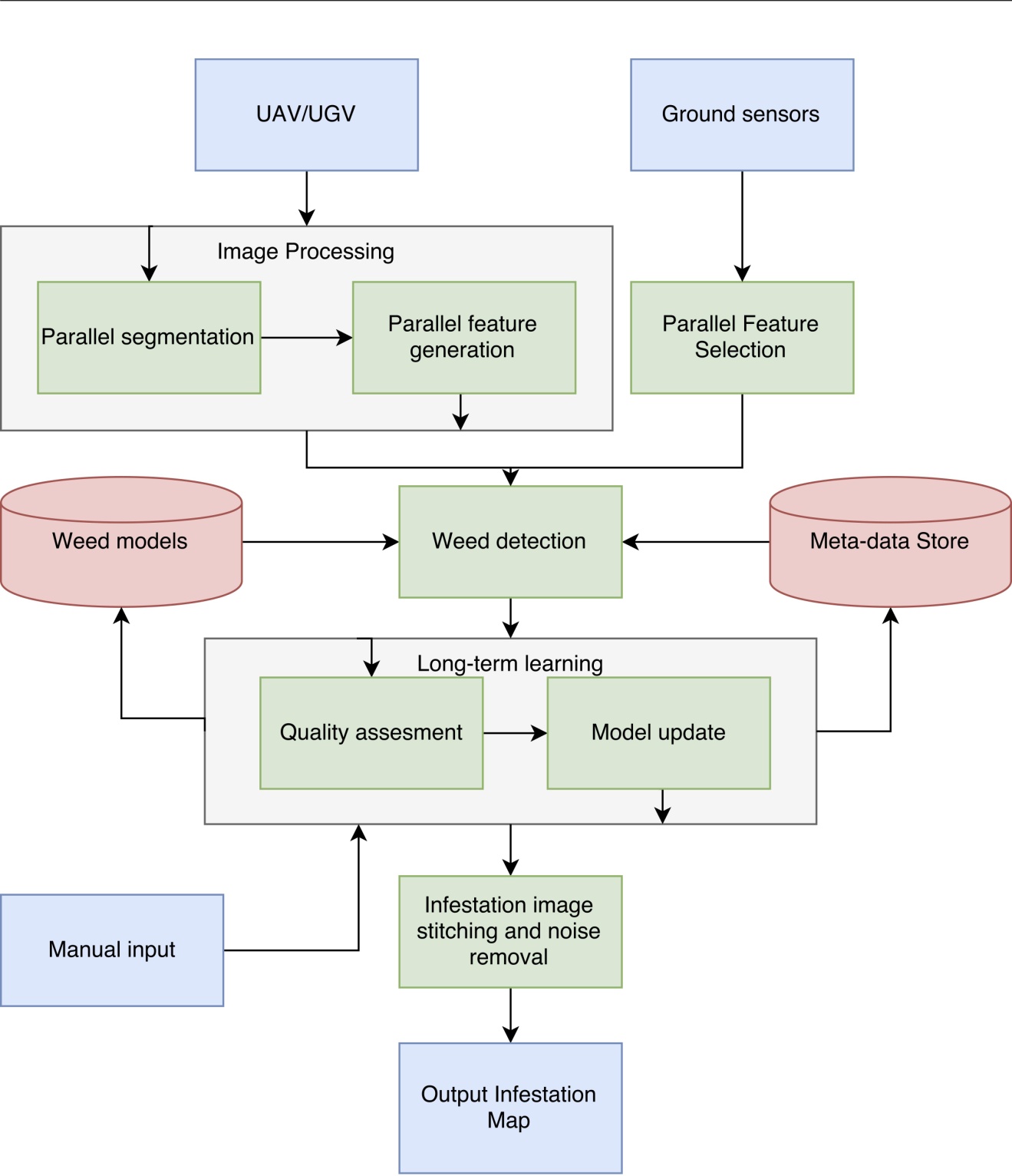


**Collaboration Diagram for Treatment at Hospital:-**

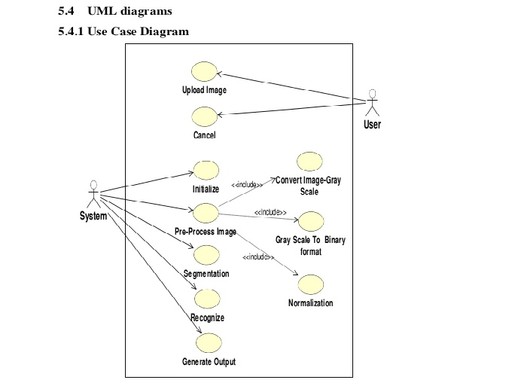
**Collaboration Diagram for Discharge from Hospital :**

**-**

**3.5 object diagram**



**3.6 use case diagrams**

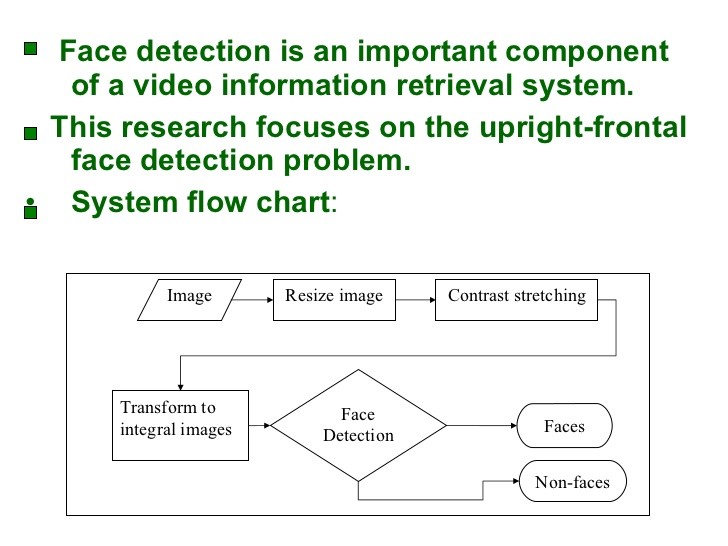


**database design**

**3.7 Entty Relatonship Diagram**

A basic ER model consists of objects called entities and specifies relationship among those entities. Purpose of this diagram is not to define any functionality rather show association and dependency among entities. ER diagram is drawn with "rectangular boxes" as entities and the "straight lines" showing the relationship between these boxes. An entity is an object or a thing that has an independent existence and can be easily differentiated from others. Each entity has some attributes like name, age, address, department etc.

Entities consisting of similar atributes make the entita sets. These entities have some association among each other hich make a relationship. These relationships can be "one to one" or "one to mana" or "mana to mana".



**3.8 SYSTEM IMPLEMENTATION**

Deep learning algorithms have given very good results in image classification and object recognition from images. In this thesis we apply SegNet semantic segmentation algorithm for the task of weed segmentation from RGB images taken under various light conditions. Examples from the dataset are presented in Figures 3-8, 3-9 and 3-10. The dataset is described in Chapter 2.3.

By using the deep learning architecture we are able to directly classify each pixel as weed, plant or ground. In this way the pixels that represent weed will be detected by directly training the model using SegNet. There is no additional need for feature extraction and processing because the deep learning architectures combine these steps into one.

For the purpose of segmenting the green areas from the land areas in the images

24

Plant Species Recognition based on Image Processing and Machine Learning



Figure 3-9: Example image from the carrot dataset (2)



Figure 3-10: Example image from the carrot dataset (3)

**MACHINE LEANRING With OpenCV**

For this project python with machine learning and open cv module is used.

OpenCV offers a good face detection and recognition [module](http://opencv.willowgarage.com/wiki/FaceRecognition) by [Philipp Wagner](http://www.bytefish.de/)). It contains algorithms which can be used to perform some cool stuff. In this guide I will roughly explain how face detection and recognition work; and build a demo application using OpenCV which will detect and recognize faces. (Also, there is a nice video of the result at the end).

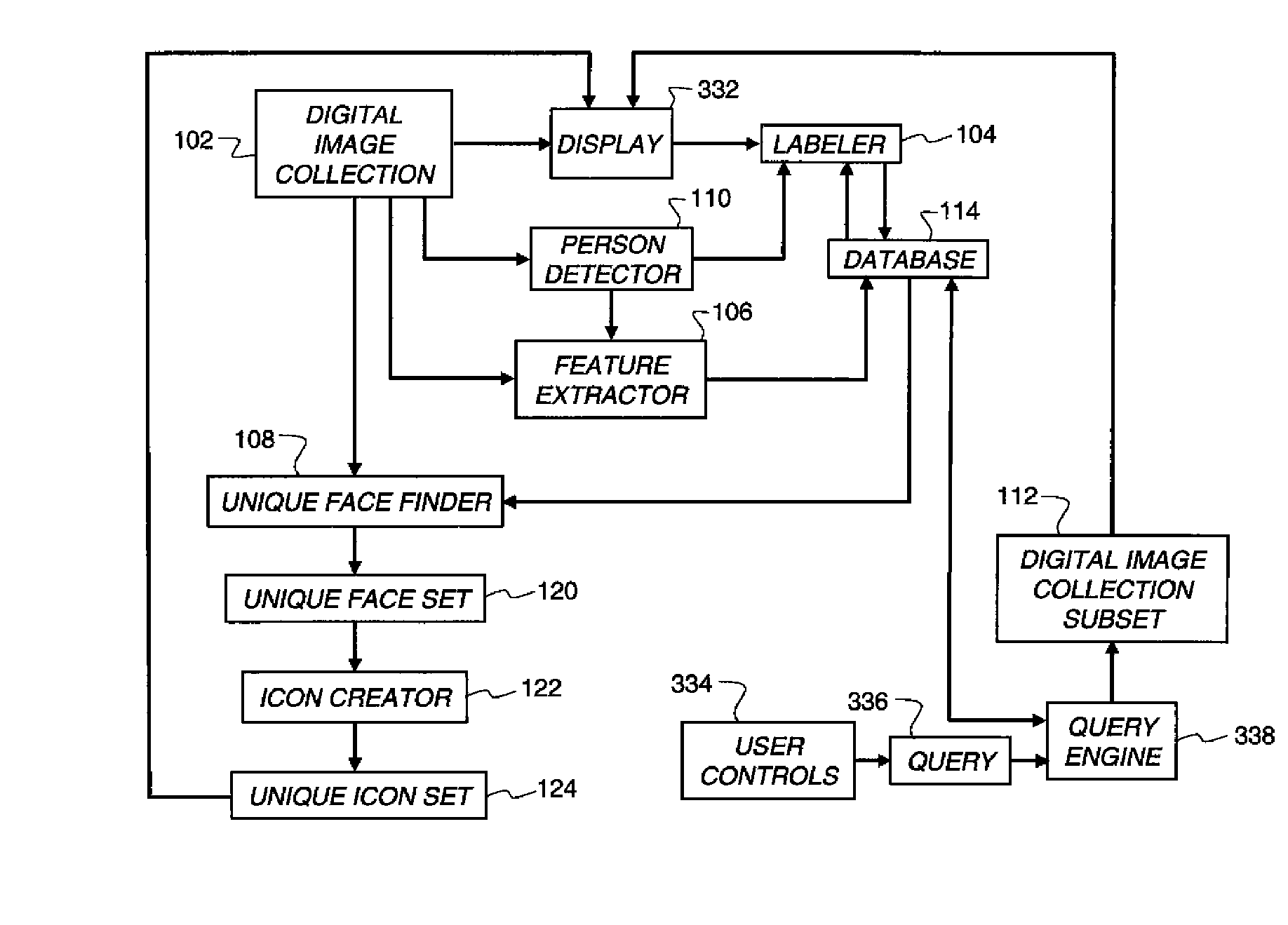
All images for this example were chosen to have a frontal face perspective. They have been cropped, scaled and rotated to be aligned at the eyes, just like this set of George Clooney image

**4. System Testing**

**4.1 User Interface Layout**

A method of organizing an image collection includes detecting faces in the image collection, extracting features from the detected faces, determining a set of unique faces by analyzing the extracted features, wherein each face in the set of unique faces is believed to be from a different person than the other faces in the set; and displaying the unique faces to a user.

***4.2 output screen***



**AFTER Prediction:-**

Id SepalLengthCm SepalWidthCm PetalLengthCm PetalWidthCm Species

1 5.1 3.5 1.4 0.2 Iris-setosa

2 4.9 3 1.4 0.2 Iris-setosa

3 4.7 3.2 1.3 0.2 Iris-setosa

4 4.6 3.1 1.5 0.2 Iris-setosa

5 5 3.6 1.4 0.2 Iris-setosa

6 5.4 3.9 1.7 0.4 Iris-setosa

7 4.6 3.4 1.4 0.3 Iris-setosa

8 5 3.4 1.5 0.2 Iris-setosa

9 4.4 2.9 1.4 0.2 Iris-setosa

10 4.9 3.1 1.5 0.1 Iris-setosa

11 5.4 3.7 1.5 0.2 Iris-setosa

12 4.8 3.4 1.6 0.2 Iris-setosa

13 4.8 3 1.4 0.1 Iris-setosa

14 4.3 3 1.1 0.1 Iris-setosa

15 5.8 4 1.2 0.2 Iris-setosa

16 5.7 4.4 1.5 0.4 Iris-setosa

17 5.4 3.9 1.3 0.4 Iris-setosa

18 5.1 3.5 1.4 0.3 Iris-setosa

19 5.7 3.8 1.7 0.3 Iris-setosa

20 5.1 3.8 1.5 0.3 Iris-setosa

21 5.4 3.4 1.7 0.2 Iris-setosa

22 5.1 3.7 1.5 0.4 Iris-setosa

23 4.6 3.6 1 0.2 Iris-setosa

24 5.1 3.3 1.7 0.5 Iris-setosa

25 4.8 3.4 1.9 0.2 Iris-setosa

26 5 3 1.6 0.2 Iris-setosa

27 5 3.4 1.6 0.4 Iris-setosa

28 5.2 3.5 1.5 0.2 Iris-setosa

29 5.2 3.4 1.4 0.2 Iris-setosa

30 4.7 3.2 1.6 0.2 Iris-setosa

31 4.8 3.1 1.6 0.2 Iris-setosa

32 5.4 3.4 1.5 0.4 Iris-setosa

33 5.2 4.1 1.5 0.1 Iris-setosa

34 5.5 4.2 1.4 0.2 Iris-setosa

35 4.9 3.1 1.5 0.1 Iris-setosa

36 5 3.2 1.2 0.2 Iris-setosa

37 5.5 3.5 1.3 0.2 Iris-setosa

38 4.9 3.1 1.5 0.1 Iris-setosa

39 4.4 3 1.3 0.2 Iris-setosa

40 5.1 3.4 1.5 0.2 Iris-setosa

41 5 3.5 1.3 0.3 Iris-setosa

42 4.5 2.3 1.3 0.3 Iris-setosa

43 4.4 3.2 1.3 0.2 Iris-setosa

44 5 3.5 1.6 0.6 Iris-setosa

45 5.1 3.8 1.9 0.4 Iris-setosa

46 4.8 3 1.4 0.3 Iris-setosa

47 5.1 3.8 1.6 0.2 Iris-setosa

48 4.6 3.2 1.4 0.2 Iris-setosa

49 5.3 3.7 1.5 0.2 Iris-setosa

50 5 3.3 1.4 0.2 Iris-setosa

51 7 3.2 4.7 1.4 Iris-versicolor

52 6.4 3.2 4.5 1.5 Iris-versicolor

53 6.9 3.1 4.9 1.5 Iris-versicolor

54 5.5 2.3 4 1.3 Iris-versicolor

55 6.5 2.8 4.6 1.5 Iris-versicolor

56 5.7 2.8 4.5 1.3 Iris-versicolor

57 6.3 3.3 4.7 1.6 Iris-versicolor

58 4.9 2.4 3.3 1 Iris-versicolor

59 6.6 2.9 4.6 1.3 Iris-versicolor

60 5.2 2.7 3.9 1.4 Iris-versicolor

61 5 2 3.5 1 Iris-versicolor

62 5.9 3 4.2 1.5 Iris-versicolor

63 6 2.2 4 1 Iris-versicolor

64 6.1 2.9 4.7 1.4 Iris-versicolor

65 5.6 2.9 3.6 1.3 Iris-versicolor

66 6.7 3.1 4.4 1.4 Iris-versicolor

67 5.6 3 4.5 1.5 Iris-versicolor

68 5.8 2.7 4.1 1 Iris-versicolor

69 6.2 2.2 4.5 1.5 Iris-versicolor

70 5.6 2.5 3.9 1.1 Iris-versicolor

71 5.9 3.2 4.8 1.8 Iris-versicolor

72 6.1 2.8 4 1.3 Iris-versicolor

73 6.3 2.5 4.9 1.5 Iris-versicolor

74 6.1 2.8 4.7 1.2 Iris-versicolor

75 6.4 2.9 4.3 1.3 Iris-versicolor

76 6.6 3 4.4 1.4 Iris-versicolor

77 6.8 2.8 4.8 1.4 Iris-versicolor

78 6.7 3 5 1.7 Iris-versicolor

79 6 2.9 4.5 1.5 Iris-versicolor

80 5.7 2.6 3.5 1 Iris-versicolor

81 5.5 2.4 3.8 1.1 Iris-versicolor

82 5.5 2.4 3.7 1 Iris-versicolor

83 5.8 2.7 3.9 1.2 Iris-versicolor

84 6 2.7 5.1 1.6 Iris-versicolor

85 5.4 3 4.5 1.5 Iris-versicolor

86 6 3.4 4.5 1.6 Iris-versicolor

87 6.7 3.1 4.7 1.5 Iris-versicolor

88 6.3 2.3 4.4 1.3 Iris-versicolor

89 5.6 3 4.1 1.3 Iris-versicolor

90 5.5 2.5 4 1.3 Iris-versicolor

91 5.5 2.6 4.4 1.2 Iris-versicolor

92 6.1 3 4.6 1.4 Iris-versicolor

93 5.8 2.6 4 1.2 Iris-versicolor

94 5 2.3 3.3 1 Iris-versicolor

95 5.6 2.7 4.2 1.3 Iris-versicolor

96 5.7 3 4.2 1.2 Iris-versicolor

97 5.7 2.9 4.2 1.3 Iris-versicolor

98 6.2 2.9 4.3 1.3 Iris-versicolor

99 5.1 2.5 3 1.1 Iris-versicolor

100 5.7 2.8 4.1 1.3 Iris-versicolor

101 6.3 3.3 6 2.5 Iris-virginica

102 5.8 2.7 5.1 1.9 Iris-virginica

103 7.1 3 5.9 2.1 Iris-virginica

104 6.3 2.9 5.6 1.8 Iris-virginica

105 6.5 3 5.8 2.2 Iris-virginica

106 7.6 3 6.6 2.1 Iris-virginica

107 4.9 2.5 4.5 1.7 Iris-virginica

108 7.3 2.9 6.3 1.8 Iris-virginica

109 6.7 2.5 5.8 1.8 Iris-virginica

110 7.2 3.6 6.1 2.5 Iris-virginica

111 6.5 3.2 5.1 2 Iris-virginica

112 6.4 2.7 5.3 1.9 Iris-virginica

113 6.8 3 5.5 2.1 Iris-virginica

114 5.7 2.5 5 2 Iris-virginica

115 5.8 2.8 5.1 2.4 Iris-virginica

116 6.4 3.2 5.3 2.3 Iris-virginica

117 6.5 3 5.5 1.8 Iris-virginica

118 7.7 3.8 6.7 2.2 Iris-virginica

119 7.7 2.6 6.9 2.3 Iris-virginica

120 6 2.2 5 1.5 Iris-virginica

121 6.9 3.2 5.7 2.3 Iris-virginica

122 5.6 2.8 4.9 2 Iris-virginica

123 7.7 2.8 6.7 2 Iris-virginica

124 6.3 2.7 4.9 1.8 Iris-virginica

125 6.7 3.3 5.7 2.1 Iris-virginica

126 7.2 3.2 6 1.8 Iris-virginica

127 6.2 2.8 4.8 1.8 Iris-virginica

128 6.1 3 4.9 1.8 Iris-virginica

129 6.4 2.8 5.6 2.1 Iris-virginica

130 7.2 3 5.8 1.6 Iris-virginica

131 7.4 2.8 6.1 1.9 Iris-virginica

132 7.9 3.8 6.4 2 Iris-virginica

133 6.4 2.8 5.6 2.2 Iris-virginica

134 6.3 2.8 5.1 1.5 Iris-virginica

135 6.1 2.6 5.6 1.4 Iris-virginica

136 7.7 3 6.1 2.3 Iris-virginica

137 6.3 3.4 5.6 2.4 Iris-virginica

138 6.4 3.1 5.5 1.8 Iris-virginica

139 6 3 4.8 1.8 Iris-virginica

140 6.9 3.1 5.4 2.1 Iris-virginica

141 6.7 3.1 5.6 2.4 Iris-virginica

142 6.9 3.1 5.1 2.3 Iris-virginica

143 5.8 2.7 5.1 1.9 Iris-virginica

144 6.8 3.2 5.9 2.3 Iris-virginica

145 6.7 3.3 5.7 2.5 Iris-virginica

146 6.7 3 5.2 2.3 Iris-virginica

147 6.3 2.5 5 1.9 Iris-virginica

148 6.5 3 5.2 2 Iris-virginica

149 6.2 3.4 5.4 2.3 Iris-virginica

150 5.9 3 5.1 1.8 Iris-virginica

Laaout part 3

**4.2 Description**

During the work in this doctoral thesis, several experiments were performed using image processing and machine learning algorithms for the purpose of plant species recognition and weed detection.

Three datasets were recorded with regular RGB cameras under different heights and lighting conditions which introduced additional challenges to the task of plant classification for the purpose of weed detection.

In the spinach dataset where the segmentation is simple and each of the seedlings could be separated from the images, we used the contours from the plants to generate descriptors. This approach is rarely possible in practice, because plants and weeds grow simultaneously and often overlap each other on the images. In cases where there is a visible overlapping, like in the tobacco dataset, sliding windows could be used to analyze different regions of the image and perform the classification using texture descriptors. Both of the approaches were promising but not very successful on the obtained datasets, partially due to the low quality of the obtained images and partially due to the similarities in the contours between weeds and plants and the high level of overlap in the tobacco dataset. In the tobacco dataset, even the manual annotation was very difficult because the weed areas were hard to distinguish

31

Plant Species Recognition based on Image Processing and Machine Learning

from the plant areas. The overlapping and different lighting conditions even after normalization and equalizing of the images introduce high noise in the data and the machine learning and even deep learning algorithms weren’t able to successfully build a model for weed detection.

With the carrot dataset, we obtained higher quality images and decided to use the SegNet semantic segmentation architecture. With the SegNet architecture we obtained a precision of around 60% on the data, a result which suggests that this approach could be used for weed segmentation from RGB images under variable lighting conditions, but needs modifications in order to obtain good results. The reported precision in the literature for SegNet is much higher than the obtained one for other datasets. For images that are recorded with multi-spectral cameras, some results are in the range of over 90% precision, however the precision on RGB datasets even under same lighting conditions and fixed height is similar to the results obtained in our work.

From the results we can conclude that the segmentation and weed detection from young seedlings using RGB cameras under different weather and lighting conditions is difficult, however, with the usage of machine learning and image processing algo-rithms the task is not impossible and the obtained results show us several directions that could be taken to improve performance. Multi-spectral cameras could also be employed or constant lighting RGB images could be obtained to improve the results. Larger datasets for training could also improve the models. Furthermore, the over-fitting could be overcome by changes in the deep learning architecture, by further simplification or by using different types of deep learning architectures and layers.

To build a more reliable model, more data would be needed, than the one ob-tained for the purpose of the presented experiments. To achieve this a cloud based architecture is proposed that would allow small farmers to collaborate on the data gathering and to achieve cheaper sensing and processing system.

**5 REFERENCES**

1. D. K. Ray, N. D. Mueller, P. C. West, and J. A. Foley, “Yield trends are in-sufficient to double global crop production by 2050,” PloS one, vol. 8, no. 6, p. e66428, 2013.
2. P. Cosmin, “Adoption of artificial intelligence in agriculture,” Bulletin of Univer-sity of Agricultural Sciences and Veterinary Medicine Cluj-Napoca. Agriculture, vol. 68, no. 1, 2011.
3. D. Schimmelpfennig et al., “Farm profits and adoption of precision agriculture,” United States Department of Agriculture, Economic Research Service, Tech. Rep., 2016.
4. C. Ritter, D. Dicke, M. Weis, H. Oebel, H. P. Piepho, A. B¨uchse,

and R. Gerhards, “An on-farm approach to quantify yield variation and to derive decision rules for site-specific weed management,” Precision Agriculture, vol. 9, no. 3, pp. 133–146, 2008. [Online]. Available: http: //dx.doi.org/10.1007/s11119-008-9061-5

1. K.-H. Dammer and G. Wartenberg, “Sensor-based weed detection and application of variable herbicide rates in real time,” Crop Protection, vol. 26, no. 3, pp. 270 – 277, 2007, weed Science in Time of Transition. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0261219406002456
2. M. Loghavi and B. B. Mackvandi, “Development of a target oriented weed control system,” Computers and Electronics in Agriculture, vol. 63, no. 2, pp. 112 – 118, 2008. [Online]. Available: http://www.sciencedirect.com/science/ article/pii/S0168169908000471
3. A. I. de Castro, M. Jurado-Exp´osito, J. M. Pe˜na-Barrag´an, and F. L´opez-Granados, “Airborne multi-spectral imagery for mapping cruciferous weeds in cereal and legume crops,” Precision Agriculture, vol. 13, no. 3, pp. 302–321, 2012. [Online]. Available: http://dx.doi.org/10.1007/s11119-011-9247-0
4. F. Castaldi, F. Pelosi, S. Pascucci, and R. Casa, “Assessing the potential of images from unmanned aerial vehicles (uav) to support herbicide patch

35

Plant Species Recognition based on Image Processing and Machine Learning

spraying in maize,” Precision Agriculture, pp. 1–19, 2016. [Online]. Available:

http://dx.doi.org/10.1007/s11119-016-9468-3

1. K.-H. Dammer, “Real-time variable-rate herbicide application for weed control in carrots,” Weed Research, vol. 56, no. 3, pp. 237–246, 2016. [Online]. Available: http://dx.doi.org/10.1111/wre.12205
2. T. Berge, S. Goldberg, K. Kaspersen, and J. Netland, “Towards machine vision based site-specific weed management in cereals,” Computers and Electronics in Agriculture, vol. 81, pp. 79 – 86, 2012. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0168169911002602
3. M. Angelovski, P. Lameski, E. Zdravevski, and A. Kulakov, “Application of bci technology for color prediction using brainwaves,” ICT Innovations 2012, Web Proceedings ISSN 1857-7288, p. 253, 2012.
4. P. Lameski, D. Kulakov, E. Zdravevski, and A. Kulakov, “Tumor detection in manually selected regions of mri images,” ICT Innovations 2014, Web Proceed-ings ISSN 1857-7288, pp. 183–190, 2014.
5. P. Lameski, E. Zdravevski, R. Mingov, and A. Kulakov, “Comparison of local image descriptors for plant identification from leaf image,” in Proceedings of the 2013 ICMER Conference, 2013.
6. B. Dikovski, P. Lameski, E. Zdravevski, and A. Kulakov, “Structure from mo-tion obtained from low quality images in indoor environment,” in Proceedings of the 10th Conference for Informatics and Information Technology (CIIT 2013). Faculty of Computer Science and Engineering (FCSE) and Computer Society of Macedonia, 2013.
7. I. Dimitrovski, G. Madjarov, D. Kocev, and P. Lameski, “Maestra at lifeclef 2014 plant task: Plant identification using visual data.” in CLEF (Working Notes), 2014, pp. 705–714.
8. P. Lameski, E. Zdravevski, and A. Kulakov, “Unsupervised weed detection in spinach seedling organic farms,” in Proceedings of the 24th International Elec-trotechnical and Computer Science Conference ERK 2015, September 2015.
9. P. Lameski, E. Zdravevski, A. Kulakov, and D. Gjorgjevik, “Plant images classifi-cation based on the angles between the leaf shape-contour points,” in Proceedings of the 12th International Conference for Informatics and Information Technology (CIIT 2015), At Bitola, Macedonia. Faculty of Computer Science and Engi-neering (FCSE) and Computer Society of Macedonia, 2015.
10. P. Lameski, E. Zdravevski, R. Mingov, and A. Kulakov, “SVM parameter tuning with grid search and its impact on reduction of model over-fitting,” in Rough Sets, Fuzzy Sets, Data Mining, and Granular Computing: 15th International

36

Plant Species Recognition based on Image Processing and Machine Learning

Conference, RSFDGrC 2015, Tianjin, China, November 20-23, 2015, Proceed-ings, Y. Yao, Q. Hu, H. Yu, and J. W. Grzymala-Busse, Eds. Cham, Switzerland: Springer International Publishing, 2015, pp. 464–474.

1. P. Lameski, E. Zdravevski, and A. Kulakov, “Weed segmentation from grayscale tobacco seedling images,” in International Conference on Robotics in Alpe-Adria Danube Region. Springer International Publishing, 2016, pp. 252–258.
2. P. Lameski, E. Zdravevski, A. Kulakov, and V. Trajkovik, “Cloud based architec-ture for automated weed control,” in Proceedings of the 17th IEEE International Conference on Smart Technologies IEEE EUROCON 2017. IEEE, July 2017.
3. P. Lameski, E. Zdravevski, S. Koceski, A. Kulakov, and V. Trajkovik, “Sup-pression of intensive care unit false alarms based on the arterial blood pressure signal,” IEEE Access, vol. PP, no. 99, pp. 1–1, 2017.
4. D. Bogatinov, P. Lameski, V. Trajkovik, and K. M. Trendova, “Firearms train-ing simulator based on low cost motion tracking sensor,” Multimedia Tools and Applications, vol. 76, no. 1, pp. 1403–1418, 2017.
5. P. Lameski, L. Pejov, L. Djinevski, D. Trajanov, and S. Filiposka, “Sampling rare events in hybrid statistical physics quantum mechanical studies of hydrophobic interactions in liquids: A gpu implementation,” in Conference: ICT Innovations 2015, At Ohrid, Macedonia, 2015.
6. P. Lameski, E. Zdravevski, V. Trajkovik, and A. Kulakov, Weed Detection Dataset with RGB Images Taken Under Variable Light Conditions. Cham: Springer International Publishing, 2017, pp. 112–119. [Online]. Available: https://doi.org/10.1007/978-3-319-67597-8\_11
7. E. Zdravevski, P. Lameski, V. Trajkovik, A. Kulakov, I. Chorbev, R. Goleva, N. Pombo, and N. Garcia, “Improving activity recognition accuracy in ambient assisted living systems by automated feature engineering,” IEEE Access, 2017.
8. E. Zdravevski, B. Risteska Stojkoska, M. Standl, and H. Schulz, “Automatic machine-learning based identification of jogging periods from accelerometer measurements of adolescents under field conditions,” PLOS ONE, vol. 12, no. 9, pp. 1–28, 09 2017. [Online]. Available: https://doi.org/10.1371/journal.pone. 0184216
9. E. Zdravevski, P. Lameski, and A. Kulakov, “Towards a general technique for transformation of nominal features into numeric features in supervised learning,” in Proceedings of the 9th Conference for Informatics and Information Technol-ogy (CIIT 2012). Faculty of Computer Science and Engineering (FCSE) and Computer Society of Macedonia, April 2012.
10. A. Dimitrievski, E. Zdravevski, P. Lameski, and V. Trajkovik, “Towards appli-cation of non-invasive environmental sensors for risks and activity detection,” in

37