Prediction of Early Vigor from Overhead Images of Carinata Plants

A Thesis Submitted to the
College of Graduate and Postdoctoral Studies
in Partial Fulfillment of the Requirements
for the degree of Master of Science
in the Department of Computer Science
University of Saskatchewan
Saskatoon

By
Anupama Das

Permission to Use

In presenting this thesis in partial fulfilment of the requirements for a Postgraduate degree from the University of Saskatchewan, I agree that the Libraries of this University may make it freely available for inspection. I further agree that permission for copying of this thesis in any manner, in whole or in part, for scholarly purposes may be granted by the professor or professors who supervised my thesis work or, in their absence, by the Head of the Department or the Dean of the College in which my thesis work was done. It is understood that any copying or publication or use of this thesis or parts thereof for financial gain shall not be allowed without my written permission. It is also understood that due recognition shall be given to me and to the University of Saskatchewan in any scholarly use which may be made of any material in my thesis.

Requests for permission to copy or to make other use of material in this thesis in whole or part should be addressed to:

Head of the Department of Computer Science
176 Thorvaldson Building
110 Science Place
University of Saskatchewan
Saskatoon, Saskatchewan
Canada
S7N 5C9

Or

Dean
College of Graduate and Postdoctoral Studies
University of Saskatchewan
116 Thorvaldson Building, 110 Science Place
Saskatoon, Saskatchewan S7N 5C9
Canada

Abstract

Breeding more resilient, higher yielding crops is an essential component of ensuring ongoing food security. Early season vigor is significantly correlated with yields and is often used as an early indicator of fitness in breeding programs. Early vigor can be a useful indicator of the health and strength of plants with benefits such as improved light interception, reduced surface evaporation, and increased biological yield. However, vigor is challenging to measure analytically and is often rated using subjective visual scoring. This traditional method of breeder scoring becomes cumbersome as the size of breeding programs increase. In this study, we used hand-held cameras fitted on gimbals to capture images which were then used as the source for automated vigor scoring. We have employed a novel image metric, the extent of plant growth from the row centerline, as an indicator of vigor. Along with this feature, additional features were used for training a random forest model and a support vector machine, both of which were able to predict expert vigor ratings with an 88.9% and 88% accuracies respectively, providing the potential for more reliable, higher throughput vigor estimates.

Acknowledgements

I would first like to deeply thank my supervisors Dr. Ian Stavness and Dr. Kevin Stanley of the department of computer science at University of Saskatchewan for their unparalleled support and guidance through the process of my master study. I would also like to thank Dr. Christina Eynck of the Agriculture and Agri-Food Canada (AAFC) for providing me with all the necessary data during my research work.

I am very grateful to Dr. Mark Eramian and Dr. Kevin Stanley for everything I learned in image processing and machine learning areas due to their teaching styles, although I know that there is so much more to learn.

Finally, I must express my huge gratitude to my parents, my siblings and my friends for providing me with unconditional support and encouragement throughout my years of study and researching this thesis. This accomplishment would not have been possible without them. Thank you.

I dedicate this to my mother and grandmother.

Contents

List of Tables

List of Figures

List of Abbreviations

HTP High Throughput Phenotyping
RF Random Forests
SVM Support Vector Machine
HSV Hue Saturation Value
CART Classification And Regression Tree
RBF Radial Basis Function
ID3 Iterative Dichotomiser 3
ANN Artificial Neural Network
NDVI Normalised Difference Vegetation Index

1 Introduction

As of 2018, the world population stands at 7.6 billion and is expected to soar to 11.2 billion in 2100 [24]. When this population surge meets rising incomes in developing countries, it will create a critical shortage in the global food supply [28]. Among many initiatives that will be required to increase the food supply to meet the demand, the introduction of novel crop species that can be grown in more marginal environments, producing more nutrition for fewer inputs, will be necessary. Breeding programs largely depend on the ability of expert breeders to quickly assess the physical characteristics (phenotypes) of different plants, allowing a higher number of potential crosses to be tested. Phenotype data collection for breeding programs has remained mostly unchanged from the 20^{th} Century, creating a bottleneck in the breeding pipeline. In particular, the need for every breeding plot to be visually observed by a human expert places a limit on either the number of plots considered, limiting the number of potential crosses, or the frequency and completeness of observation, compromising the quality of the observations. Efforts in digital phenotyping have been undertaken in an attempt to address this bottleneck. Farmers are embracing advanced technologies which involve the use of sensors, robots, GPS and data-analytics software to collect agronomic information across the farms, thus making the farming operations more efficient and insight driven. Stationary sensors, mobile robots, and camera drones gather data and send images of crops and information (like plant height, lodging, plant structure, soil moisture) to a computer, which looks for indicators of health, stress and yield [22, 77, 88]. Farmers and plant breeders receive the feedback in real time and then provide water and fertilizer in requisite amounts to only the regions that require it. The availability of such computing tools can assist farmers in monitoring plants and decision making. As a result, water and chemicals are judiciously used, and healthier crops and higher yields are produced, all of which benefit farmers and create better breeds of crops, faster.

The development of mechanized plant phenotyping platforms has begun in the recent past. Various methods of high throughput phenotyping (HTP) are advancing with time, with applications making contributions to the domain of precision farming. Imaging technologies do play a role in automated phenotyping. They are non-destructive and can be employed for collecting phenotypic data from controlled environments or open environments throughout their lifecycle, reducing the time and human labor required [84].

Estimating crop canopy cover and leaf senescence from aerial imaging data of a Maize field trial [60], detecting and counting of soybean flowers digitally [112], or identifying field-level crop type from Landsat data [53] are some of the many recent high throughput phenotyping (HTP) applications. Such research studies exploit the potential of image processing and machine learning techniques in assessing crop growth and productivity.

Not only for the plant structure above the ground, different imaging approaches have been used for plant root phenotyping. An experimental setup was developed for RGB and hyperspectral imaging of root systems that could quantify the root development over time [11]. In another study [18], a new image acquisition system and software platform were developed for the high-throughput phenotyping of three-dimensional root traits.

Compared with the above-mentioned plant characteristics (like canopy cover, root systems), the automated phenotyping of early vigor has received less attention. Early vigor is an essential indicator of the overall health of a crop and is employed by plant breeders as an early estimate of the viability of a line [48]. Early vigor is defined as the rapid development of plant foliage during the early growth stages [8, 92]. It implies the advantage of strength and vitality through the early growth of plants. Plants of high vigor result in various physiological and agronomic benefits. Early vigor cultivars have a high potential for increased production in future warm climates [58], a high specific leaf area (SLA) [82, 93], and improved root growth, which increases the nitrogen absorption [4]. Early flowering is associated with early vigor crops, which is a benefit in dry climates [5]. Further benefits of greater early vigor are improved water use efficiency through reduced soil evaporation [81, 90] and quicker canopy growth [90]. Thick ground cover reduces the light availability in the soil, suppressing the growth of weeds, improving the light interception, potentially enhancing biomass and growth rate [108].

Phenotyping early vigor digitally is more challenging than other traits as it does not directly map to either a morphological characteristic like leaf area [25, 41] or to a spectral measure like NDVI [46]. Vigor has traditionally been a measure made by trained experts, correlated to, but not dependent on, measurable characteristics of the plant. As a trait, based at least in part on intuition, it is reasonable to hypothesize that the correlation between traits will be best captured by a non-linear model across a number of features related to vigor, and that this model should be built using machine learning techniques.

1.1 Research Objectives

The primary objectives of this research work are

1. Identify features for digitally phenotyping early plant vigor.
2. Develop a machine-learning-based prediction model to estimate the plant vigor level.
3. Provide a detailed approach to evaluate the machine learning models that will be employed.
4. Establish a baseline performance that can be used in future research as a benchmark.

To accomplish the task of identifying early vigor from the plant images, a set of features related to the size, symmetry and abundance of crop in a single row breeding plot will be extracted. To accomplish the development of the automated model to predict the vigor ratings, the performance of the different machine learning algorithms like random forest and support vector machine will be tested.

1.2 Contributions

In this work, along with a set of features on area and perimeter covered by the plants, we present a novel feature which is the extent of plant growth from the center line that can be used for estimating early vigor from the images of single-row breeding plot. We show that a random forest model based on the crop perimeter, area, and mean and variance of distance from the centerline, can be used to estimate reported vigor (on a scale of 1 to 3) of Carinata to an accuracy of approximately 89%. Further analysis of the structure of the decision tree demonstrates that 'mean of distances at observation 2', 'variance of distances at observation 2' and 'mean of distances at observation 1' are the most distinct features, providing the greatest diagnostic power. We are able to classify the data samples in the low vigor (1) and high vigor (3) categories with exceptional accuracies (90% and 98%, respectively), but have a more difficult time with category 2 (81%). We expect that this is due to the difficulty of discerning crops at the boundaries of vigor classes for both the automated algorithm and the human expert. The findings here both inform the utility of automated vigor selection and highlight the shortcomings in the practice of manual collection.

1.3 Organization of Thesis

This thesis is comprised of seven chapters. Chapter 2 provides the background information on the fundamental concepts entailed in this work. Existing research work associated with predicting or identifying early vigor using various forms of technology and their limitations are discussed in Chapter 3. Chapter 4 discusses the image acquisition techniques, the feature extraction and the data used for the study. Chapter 5 describes the machine learning models employed. Chapter 6 presents the results and finally, Chapter 7 discusses the contributions and implications of this research work.

2 Background

2.1 Image Processing

Image processing deals with the processing of digital images using computer algorithms to extract useful information or to eliminate unwanted information or noise. Digital image processing encompasses the operations performed on an image, in order to get maximum information yield from the images. Various image preprocessing techniques, described in chapter 4, prepared the image dataset suitable for object identification. Detecting the row centerline, which is one of the primary objectives of this research work, was accomplished by implementing the Hough Transform.

2.1.1 Hough Transform

The Hough transform is a technique to detect regular shapes in an image by identifying their features [75]. It can detect regular curves like line, circle, ellipse because the desired features of the shape can be represented in a parametric form. A regular curve is defined as a smooth curve which has a non-zero velocity at every point in the curve. The main advantage of the Hough transform is that it is robust to partial occlusions in feature descriptions and image noise [75].

It was first invented by Paul Hough in 1959 for machine analysis in pictures of a bubble chamber. It was patented as U.S. Patent 3,069,654 in 1962 with the name 'Method and Means for Recognizing Complex Patterns'. This patent uses a slope-intercept parametrization for straight lines, which results in an unlimited transform space as the slope can extend to infinity. Later, in 1972, Duda first described the usage of the rho-theta parametric form of the Hough transform [26].

The transform can be described as a mapping function which involves conversion of an image point into its corresponding shape (for example, a straight line) in the Hough parametric space (ρ, θ) using the equation 2.1, where ρ is the distance of the line from the origin, θ is the angle between the line and the horizontal axis, x and y are the coordinates of the line.

$$\rho = xcos\theta + ysin\theta \tag{2.1}$$

It is possible to find an association between each line of the image and a (ρ, θ) pair. The Hough transform algorithm begins by creating a two-dimensional array, called the accumulator array, to detect the existence of a line described by $\rho = xcos\theta + ysin\theta$. The number of the rows and columns of the array is determined

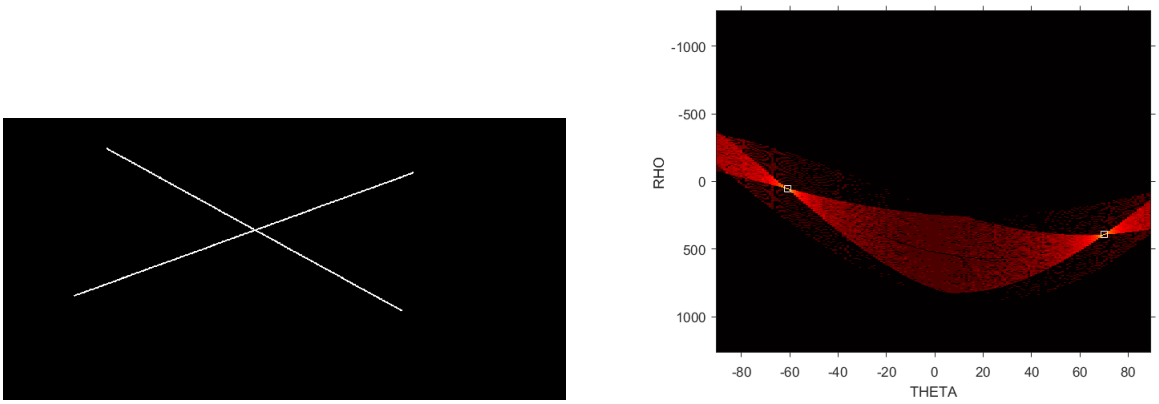


Figure 2.1: Illustration of Hough space for a sample set of lines.

by number of quantization levels of ρ and θ respectively. The idea is that an image point is mapped to a set of possible lines, that passes through that point.

When an edge point is mapped into the Hough space, bins in the accumulator array is incremented for the family of lines that could pass through that point [37]. This results in some bins with the highest values corresponding to the lines in the image that can be extracted [98]. One can detect lines by setting a threshold for the bin values and interpreting values greater than the threshold as a line. The visualization of the Hough parametric space is shown in Figure 2.1. The two points of intersection of curves are the Hough parameters of the two lines.

The Hough transform was chosen for line detection because of its many desirable features. Occlusion can hamper the success of a shape detection technique but Hough transform can recognize partially deformed shapes as it treats each image point independently. Additionally, it is robust to random data generated by poor image segmentation. It is less likely that random points gather more votes to a single bin of the accumulator. Also, the Hough transform can collect evidence for different instances of a particular shape in the same image. It produces a distinct peak in the accumulator array for each instance of the shape.

2.2 Supervised Machine Learning

This section begins with a short introduction to supervised learning followed by the detailed description of the classifiers used in this work.

Supervised learning is a type of machine learning algorithm which uses a training dataset comprised of known pairs of input variables with each of their corresponding output values. The machine learns the mapping rule from the labeled training examples to build the model. During the process of learning, the model identifies the optimal set of features which helps in reaching a decision or predicting a label for a data sample with the least error. A test dataset is usually used to evaluate the performance of the model. Supervised learning algorithms can be categorized into two groups namely:

- Regression algorithms: Regression algorithms are used when the goal of a model is to predict a continuous value.

- Classification algorithms: Classification algorithms are used when the goal of a model is to predict a categorical value.

2.2.1 Classifiers Used

This research work aims to build an intelligent model which is able to give a plant vigor rating on a scale of 3 with 1 being the lowest vigor and 3 being the highest vigor. Since the target value to be predicted by the model is a categorical value, in other words, the objective of the model is to assign the plants to one of the three vigor groups, this problem is modelled as a classification problem. Random forest was chosen for this research work because of its various advantages described in the following paragraph. For comparing the performance of the random forest model, support vector machines was chosen as the baseline model.

Random Forests

Random forest is a supervised algorithm which can be used for both regression and classification tasks. As the name suggests, this algorithm constructs a forest or an ensemble of trees. For classification, the predicted label is the mode of the classes of the individual decision trees, and for regression, the predicted value is the mean calculated from the predictions made by the individual trees. Random forest splits the dataset into random subsets on which, a decision tree is built. The operation of a random forest is captured in Figure 2.2. Since the result generated by the random forest is influenced by each decision tree, the following paragraphs give an insight on the decision tree.

The decision tree is one of the most widely used learning algorithms and they have been applied in diverse areas of the scientific sphere ranging from computational biology, business management, healthcare operations, precision farming [54, 74, 100]. The key ingredient of the success of this method is its interpretability, that makes the model transparent and understandable to any user, its flexibility, that makes itself fit in any realm of the problem, and its ease of use, that makes it accessible even to non-specialists. When combined with ensemble methods, it provides fairly decent results concerning predictive accuracy. Notably in the case of high-throughput datasets, tree-based methods are also highly scalable from a computational point of view.

The fundamental premise of a decision tree is to split the data set into smaller subsets based on the feature values until a decision is made, once the tree reaches a requisite subset that contains data samples belonging to a class. The decision is made by asking a succession of questions about the features associated with the instances [45]. Each question is represented in a node, and for every possible answer to the question, a child node is created [45]. An instance is classified into a class when the hierarchy of questions leads to a leaf node. The decision on which attribute to split on is made based on different metrics like information gain, Gini impurity, and variance reduction. Some of the tree algorithms like the CART (Classification and Regression

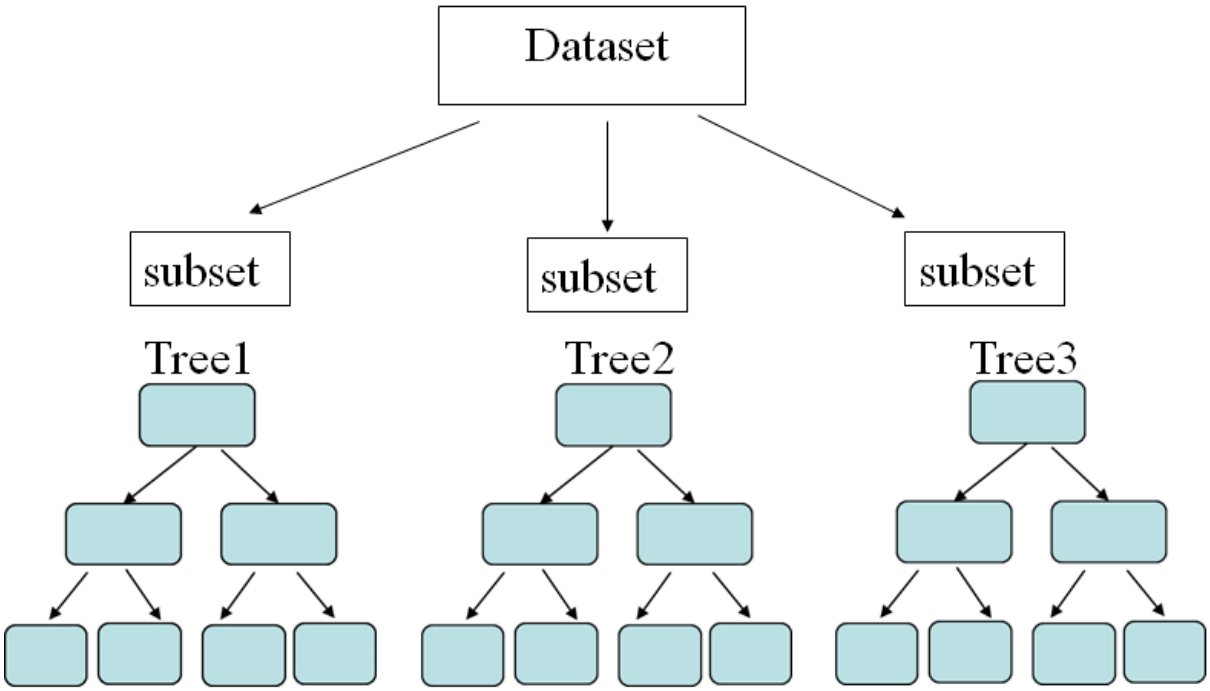


Figure 2.2: Illustration of the functioning of random forest lines

Trees) uses the Gini Index as a metric for classification, Iterative Dichotomiser 3 (ID3) uses entropy function and information gain as metrics.

Advantages:

- Ease in analyzing the tree: It is easy to look through the tree to evaluate the decision paths and understand why specific samples are classified or predicted in a specific manner rather than calculating classification scores possibly based on some non-linear model questions.
- Tree algorithms are nonparametric and non-linear: This algorithm does not make strong assumptions about the mapping function. As such, it can learn from any functional form and better generalizes to previously unseen data. Nonparametric methods are excellent in cases of large training datasets.

One disadvantage of decision trees is that they tend to overfit the data. Over-fitting is the phenomenon in which the model perfectly fits in the training data to the degree that it ends up with branches with strict constraints of sparse data. This affects the model performance when learning from a previously unseen dataset. However, this issue can be avoided in Random Forests because the data set is divided into smaller random subsets and then a decision tree is built for each subset. The average of the outcomes resulted from the ensemble of decision trees is considered as the predicted value [79].

Support Vector Machine

Support vector machine (SVM) proposed by Vapnik and his research group at AT&T Bell Laboratories in the 1990s, is based on the idea of finding a boundary or hyperplane that separates a set of data samples [107]. It provides high generalizability in practical scenarios [20]. In the SVM model, decision hyperplanes are decided based on identified support vectors to generate a plane that separates two class samples with the maximal margin [107] as shown in Figure 2.3

Support vector machine (SVM) is a supervised machine learning algorithm that can be used for both classification or regression tasks. This algorithm works by plotting each data sample as a point in the n-dimensional space where n is the number of features in the dataset. The algorithm classifies by identifying an optimal hyperplane which groups the data samples. The best hyperplane is the one that maximizes margins between the classes.

In two dimensional space, the frontier is a line dividing the plane into two parts. In such a case, the dataset is known to be linearly separable. In cases of non-linear separation problems, an intrinsic technique of SVM, called the kernel trick, transforms low dimensional input space to a higher dimensional space making it a linearly separable problem as shown in Figure 2.4.

Various kernels exist:

- Gaussian RBF kernel: It is a general-purpose kernel which is used when there is no prior knowledge about the data. The equation is given by

$$k(x_i, x_j) = exp\big(-\frac{\|x_i - x_j\|^2}{2\sigma^2}\big), \tag{2.2}$$

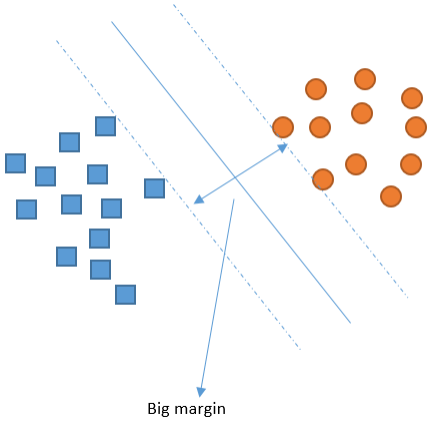


Figure 2.3: Hyperplane classifying the samples

where x_i and x_j are two data samples. The adjustable parameter sigma influences the performance of the kernel, and must be tuned to the problem [111]. If it is overestimated, the exponential is quite likely to behave linearly [32] and the higher-dimensional space loses its non-linear power [111]. In this case, the decision boundary becomes flexible and smooth and tends to make wrong predictions in classification [111]. On the other hand, if the value of sigma is underestimated, the decision boundary becomes strict and sharp and tends to overfit [111]. The decision boundary becomes highly sensitive to noise in training data and therefore the classifier will lack its generalizability [111].

- polynomial kernel: The equation is given by

$$k(x_i, x_j) = (x_i \cdot x_j + 1)^d \tag{2.3}$$

where x_i and x_j are two data samples and d is the degree of the polynomial. The polynomial kernel adds polynomial features to transform the data.

- sigmoid kernel: For certain range of α and c parameters, the sigmoid kernel behaves like the RBF kernel. Experiments also show that their performance are similar [57].The equation is given by

$$k(x_i, x_j) = tanh(\alpha {x_i}^T x_j + c) \tag{2.4}$$

where x_i and x_j are two data samples.

2.3 Model Evaluation

2.3.1 Cross-Validation

Cross-validation is a method to evaluate the performance of machine learning models. In cross-validation, the data set is split into K folds out of which one subset is treated as a test set while the remaining subsets

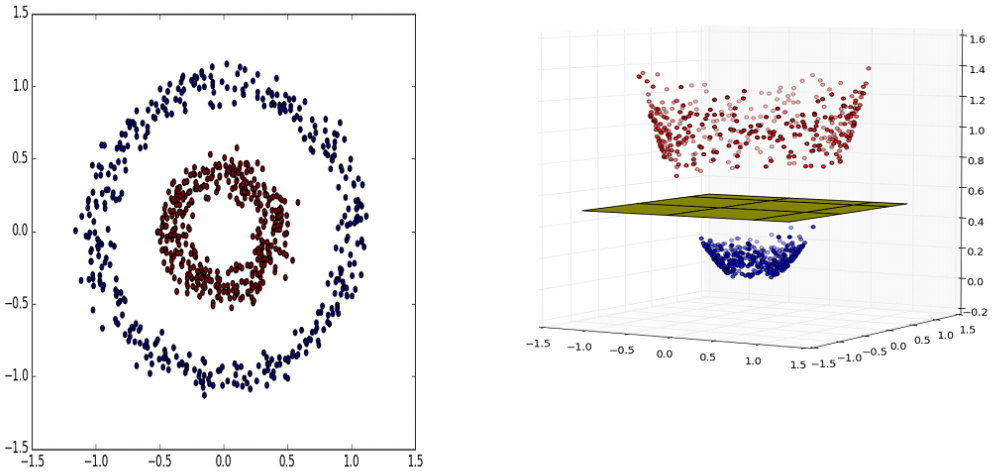


Figure 2.4: An example of the kernel trick.

are used for training as shown in Figure 2.5. In this way, it gives an insight into the generalizability of the model by measuring the accuracy of a machine learning model on unseen data.

2.3.2 Confusion Matrix

Sometimes, classification model accuracy can be misleading, as model generalizability cannot be entirely characterized using simple metrics. A confusion matrix, shown in Figure 2.6, offers insight on a model's generalizability. It is a table describing the performance of the model on test data for which the actual labels are known. The row represents the number of actual class labels in a particular class and column represents the number of samples predicted for the corresponding class. The diagonal line represents the number of samples correctly classified, whereas the non-diagonal cells indicate how samples were misclassified.

The confusion matrix shows the ways in which a classification model is confused while making the predictions. It sheds light not only on the errors being made by a classifier but also on the types of errors. For example, for a binary classification one can calculate the error rate, analyze how precisely the classifier made the predictions, understand the recall ability of the classifier and also find the accuracy of the classifer.

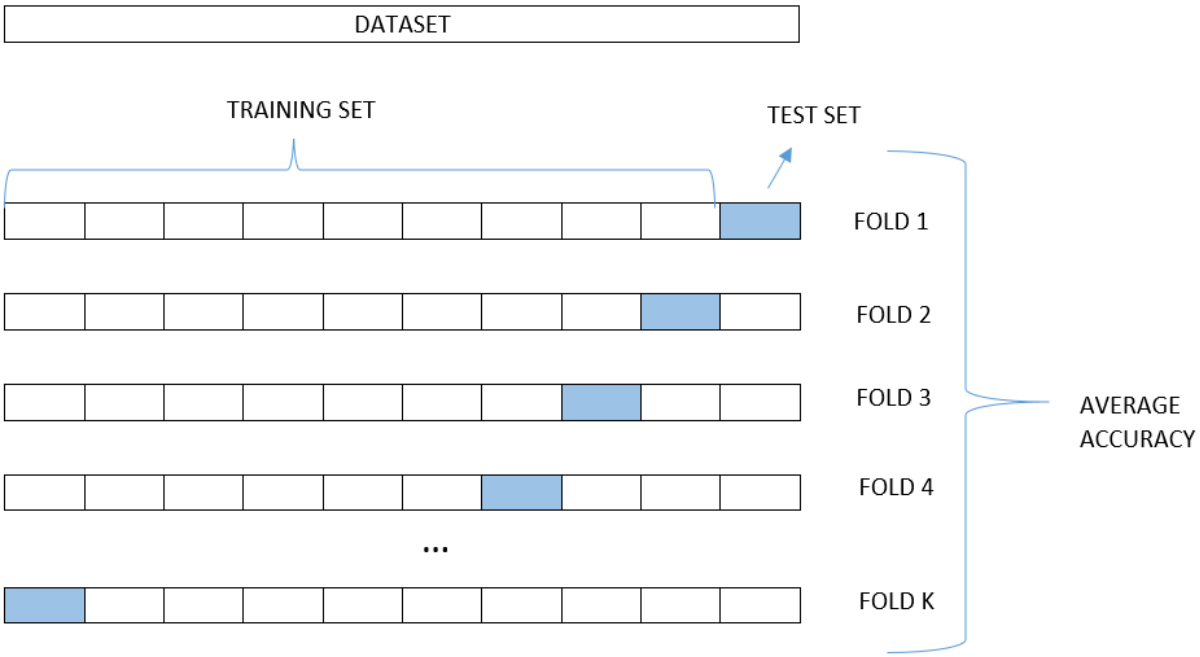


Figure 2.5: Visual representation of K-Folds

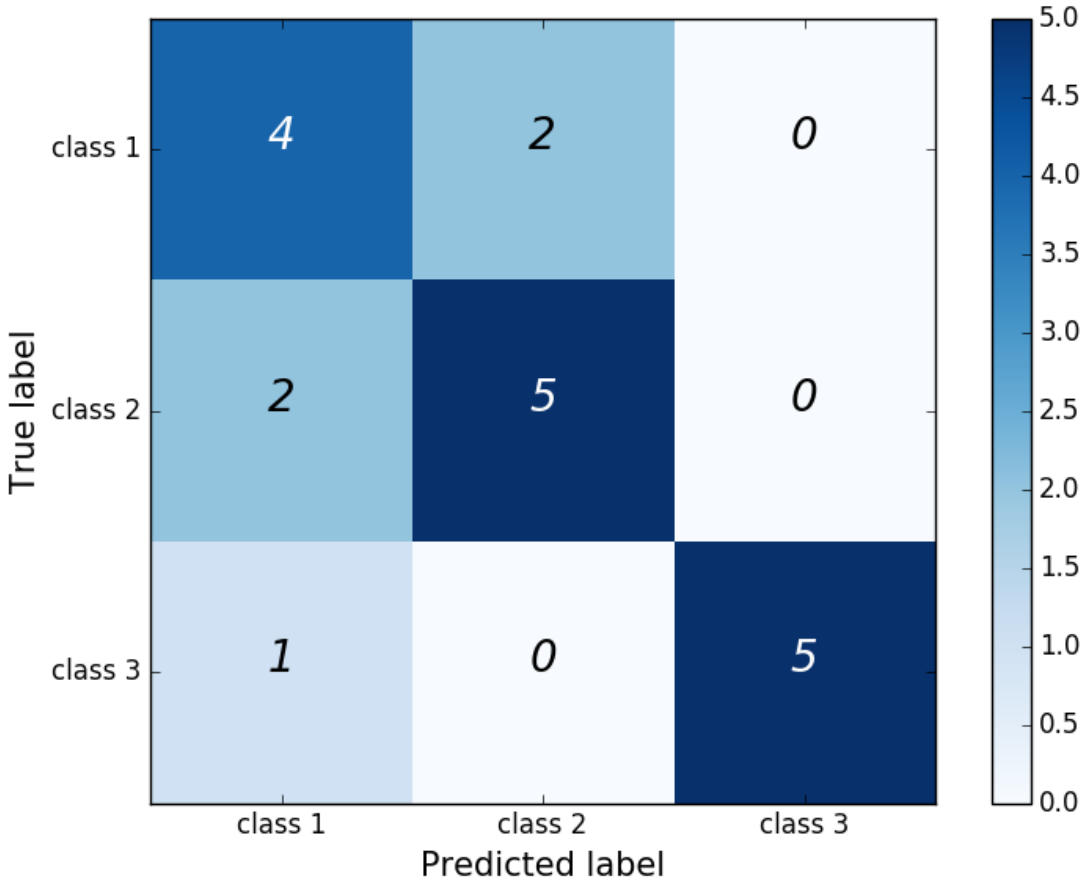


Figure 2.6: Visual representation of confusion matrix

3 Literature Review

3.1 Digital Plant Phenotyping

Image analysis has many applications in designing agricultural processes and calibrating farming practices. This section focuses on the role of image analysis in plant phenotyping. There are many times when decision-making is crucial and done based on the appearance of the subject. For example, taking decisions on the condition and quality of crops such as when to harvest, or whether a disease is present. In the case of plant breeding programs, plants with better growth trajectory are more likely to be of superior varieties and in demand. Labor-intensive and time-consuming tasks such as harvesting, plant counting and plant rating rely extensively on human visual interpretation and subjectivity.

A common indicator of plant performance is its leaf color [105]. Leaf greenness is determined by specific biological properties such as leaf chlorophyll content and leaf morphological characteristics such as leaf shape [105]. Leaf greenness during plant development is affected by a lack of nutrition and water, and environmental conditions such as temperature [105]. The most frequent method of phenotyping leaf greenness is by using the spectral measure such as the normalized difference vegetation index (NDVI) [73,103]. NDVI is calculated using the difference between the reflectance amounts in the visible light spectrum and the near-infrared part of the spectrum (NIR) [105].

For the visible range, red or blue channels are used for the detection of NDVI [55,73,103]. This enables in capturing the digital information of plants in different lighting conditions. NDVI has been used in a number of research works to study phenological variations in crops [42, 47], to analyze vegetation [64, 96], to identify stress [9, 80], and assess nitrogen status [3, 36]

Other leaf characteristics like their shape and size can be readily captured to a first approximation by a two-dimensional outline. A tool called 'Leaf Processor' was developed that performs an integrated analysis of bending energy and other shape parameters [6]. In this way, it gives an accurate quantification of leaf perimeter changes that indicates many differences in the leaf shape. ImageJ, a very popular image processing software, has been used for calculating the leaf area based on the thresholded pixel count measurement [27, 43,67,86]. Another software, Easy Leaf Area, uses the combination of color ratios and connected component analysis to quickly measure leaf area in various images in a few seconds [27].

While the methods in [15, 29, 104] compute leaf area, leaf angle, indent width and height, the authors of [17] define a set of new phenotypes, e.g., junction-tip distance, leaf curvature, integral leaf-skeleton area, leaf-junction angle and stem angle, with a brief explanation on their importance.

The other most direct indicator of plant performance is the plant biomass [105]. This is often monitored in a way by computing the number of pixels covered by an individual plant or the entire canopy of all the plants in a field within an image [51]. Such methods achieved great success in assessing the performance of arabidopsis [38], tobacco [106] or cereal grain crops [78].

Digital image analysis, which is non-destructive in nature, can be used to measure above-ground biomass of plants [30]. Some studies have estimated above-ground forest and canopy biomass from remote sensing, satellite, and airborne images [31, 56, 89]. Many linear and non-linear functions are used to model biomass of the plants [14, 16, 31, 49, 99]. Some studies mentioned in the literature, that focuses on estimation of the biomass of plant, used destructively obtained measurements and digitally acquired parameters [14, 78, 99]. Plants of high early vigor are known to have robust growth and perform very well. As such, different indicators of plant performance mentioned above are indirectly used to assess the level of vigor in plants.

3.2 Analysis of Early Vigor

Scoring early plant vigor through visual judgment and with multiple raters in the field is labor-intensive, biased and expensive [46]. Even experienced raters determine early plant vigor subjectively. However, visual determination of early vigor is still under use because of a lack of high throughput phenotyping methods. Manually counting the tiller number of each plant per plot would probably be an accurate method to detect early plant vigour but implementing it in the field trials in breeding programs is cumbersome [87]. Advanced phenotyping techniques predicting early plant vigor would improve the breeding activities [46]. Phenotypic differences in early plant vigor were reflected in cases where there was an increase in nitrogen levels or sowing densities [40, 52, 97].

Another widely used method of analyzing early vigor has been estimating the canopy cover. Field applications like in [33, 35, 47, 55] analyzed canopy cover by referring to the fraction of the ground that is covered by the canopy. The canopy cover is related to early vigor and can be exploited to see temporal differences, indicating that canopy cover is one of the significant traits for plant phenotyping in the future [19]. It is also possible to estimate the leaf area index (LAI), the number of leaves, and the canopy architecture while constructing the 3D-shape of the canopy from multiple images. There are studies where 3D-reconstruction has been successfully implemented in the lab [1, 70] but it is challenging to be executed in the field [10]. With the advancement of modern imaging capabilities of unmanned aerial vehicles [63, 114] and other remote devices that can efficiently generate plant images in the field from different directions, it should be possible to analyze a 3D model of the plant canopy and further progress in this research topic [105].

Brandão *et al.* used satellite images to monitor cotton vigor [13], and found that normalized difference vegetation index (NDVI) correlated with cotton plant vigor. This study identified the NDVI index across the spatial arrangement and the cotton yield distributions during different phenological stages in crop growth. Geostatistical methods were used to perform analysis using geostatistical package Vesper [109]. Yield monitor

data and multispectral satellite images at 56m spatial resolution were collected in a rainfed cotton field on two dates to monitor the plant vigor.

Anzola *et al.* designed a Wireless Sensor Network to collect and characterize the leaf angle and the amount of chlorophyll to judge the plant vigor [2]. The proposed method was able to judge the vigor based on data fusion of fuzzy comprehensive evaluation, thus, avoiding the inaccuracy of a single factor. However, this study was able to fulfill the objective in a controlled environment monitoring ten cotton plants.

De Bei *et al.* developed a free computer app, VitiCanopy, that runs image analysis methods on the captured images and returns the leaf area index (LAI) and plant area index (PAI) of the vine canopy [23]. VitiCanopy uses the front camera and global positioning system (GPS) of smartphones and tablet PCs, to run image analysis algorithms on digital images of plant canopies and extract useful information related to canopy architecture [23]. Results generated by VitiCanopy on the images of grapevines, correlated fairly well with traditional methods of measuring LAI and PAI [23].

3.3 Crop Row Determination

Because linear breeding plots are considered in this work, an important component of the image processing pipeline is determining the plot row. Søgaard *et al.* estimated the row position through a combination of the point estimation and weighted linear regression methods [95]. In contrast, the authors of [85] detected crop rows by dividing the image into horizontal patches and marking the centers of gravity of each of those patches. The marked points helped in determining the crop rows. Other work [83] selected the crop row from all the candidate lines that met at a vanishing point. Bossu *et al.* explained the extraction of rows based on the fact that crop rows are localized in a frequency domain and a mother wavelet function with this frequency could be chosen [12]. The most popular method for detecting crop rows has been the Hough transform. Gée *et al.* applied a double Hough transform assuming that the crop rows are represented by the lines of the image that converge to the vanishing point [34]. Guerrero *et al.* applied the Hough transform to the points which represent the edges of the crops [39] and Montalvo *et al.* implemented the randomized Hough transform for identifying crop rows [61].

Apart from agricultural applications, the Hough transform was popular among other real-time applications. Bar deceleration strips and row well covers were detected using the Hough transform with 89% and 93% accuracy respectively [59]. Multiple object instances in images were detected by redeveloping the Hough transform within a probabilistic, energy-based framework [7]. We chose to use the Hough transform method to identify the row lines in the images because it is a widely used computer vision method to detect the geometric primitives and it works well in case of gaps in the crop rows due to missing plants.

3.4 Machine Learning Applications in Agriculture

The evolution of high-throughput plant phenotyping technologies led to an explosion of data like high-quality images and sensor readings [91]. However, identifying relevant patterns and features for phenotyping plants from this vast collection of data calls for the application of the advanced machine learning tools. In plant phenotyping and breeding activities, machine learning methods are usually employed during four crucial phases of decision making which are identification, classification, quantification and prediction [91].

Machine learning approaches provide a modular strategy for performing data analysis for its varied applications in plant stress phenotyping. Recent studies based on machine learning for phenotyping using images obtained from UAV-based platforms to detect weeds in wheat (*Triticum aestivum L.*) [102], maize [72], and sunflower (*Helianthus annuus L.*) [101] have paved a new path for better farming practices on a spatial and temporal basis.

Examples include numerous success stories in various domains of phenotyping like phenotyping water stress from thermal images of spinach (*Spinacia oleracea L.*) using support vector machine (SVM) and Gaussian processes classifier [94], prediction of disease in rice using SVM and generalized regression neural network (GRNN) [44], identifying plants with infestation using random forests (RF) and SVM spatial matching kernel, prediction of yield from satellite imagery data using artificial neural network (ANN) [69]. The building of the prediction model for vigor rating is the primary objective of this study and adopting machine learning techniques to achieve it seemed promising.

Compared to these literature outlined above, this research work proposed a novel metric to phenotype plant vigor which is the extent of the plant growth from the center line. Compared to other remote sensing approaches, the images in this work were acquired using a novel hand-held remote sensing platform called the GrowPro which is low-cost and can be built easily using commonly available pieces of equipment. Additionally, the system is easy to assemble, easy to operate in different environmental conditions and can be used without specialized training. An automated prediction model based on machine learning algorithms was designed to estimate the plant vigor ratings which is first of its kind in the literature.

4 Experimental Setup

4.1 Data Acquisition

4.1.1 Image Data

Images of carinata plants were used for digitally phenotyping early vigor. High-resolution overhead (top-down) images of the plants were collected from a nursery located in Saskatoon, Saskatchewan, operated by Agriculture and Agrifood Canada in the summer of 2017 [68]. The nursery was divided into three rectangular blocks, each containing 50 ranges. Each range (as shown in Figure 4.1) contained 46 rows of Carinata plants, where two rows were grouped to make one plot of the same genotype of Carinata. In summary, there were 23 plots from each of the 50 ranges from each of the three blocks which amounted to a total of 3450 plots. To analyze crop vigor in our study, we considered only those ranges for which there were images captured during two stages of plant development: the true leaf stage and canopy closure stage. Images were taken on six different dates: June 7, June 19, June 23, June 26, June 28, and July 5. Images of only 5 ranges were captured on consecutive dates. We used the manual vigor ratings which were performed on June 21, 2017 by the Carinata breeder as our ground-truth data. Thus, we considered only those images that were taken before June 21, 2017 so our dataset consisted images of the five ranges taken on June 7 and June 19, leaving a total of 1380 raw images of rows of Carinata plants from the three blocks from the two dates. Images of the plots two weeks prior to June 21 were included to allow for differential comparisons within the machine learning algorithm, potentially encoding growth trajectory as well as point measurements.

The images were acquired with the GrowPro system, which consisted of an action camera (GoPro Hero 5) attached to a monopod with a three-axis gimbal to stabilize the camera to point downward as shown in Figure 4.2. The GrowPro was hand-held and moved across the plots, row by row. The filming procedure is showed in Figure 4.3. While filming, the operator moved evenly and with regular constant pace. A step-by-step guide has been chalked out in [68] for the operator to avoid errors and difficulties. The resulting images covered halves of the crop rows from both the sides. These images were stitched together using Agisoft Photoscan to form an orthomosaic picture of each complete range (as shown in Figure 4.1).

The full process can be found in [68], but the Agisoft process can be summarized as:

1. Align the input pictures with medium accuracy to create a sparse point cloud.

2. Build a 3D model based on the sparse cloud.

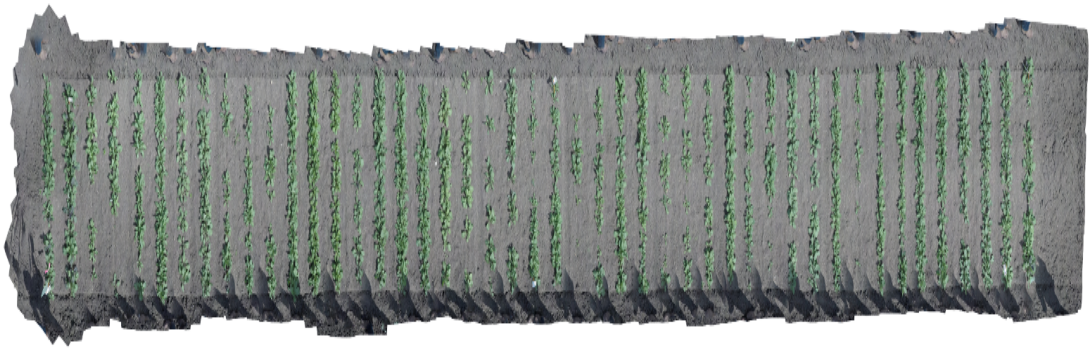


Figure 4.1: Stitched image of a range captured on June 19,2017.

Figure 4.2: The image acquisition tool used called GrowPro system, reproduced from: [68]

Figure 4.3: The filming procedure, reproduced from: [68]

3. Build a mesh

4. Generate an orthomosaic

5. Export the orthomosaic

From the final stitched image, the rows were segmented, using a semi-automatic segmentation tool developed in our lab. Samplest of segmented images is shown in Figure 4.4.

4.1.2 Ground-Truth

Vigor was rated by an expert who observed the plants in the field as part of standard practice. Vigor was subjectively scored as 1 (low vigor), 2 (moderate vigor) or 3 (high vigor). In rare cases, 0 (poor vigor) and 4 (exceptional vigor) ratings were assigned, but accounted for less than 2% of plots, and were removed from consideration.

4.2 Feature Extraction

Because the data used in this study is in the form of images, the feature extraction workflow involves a set of image processing steps each of which is described in the following subsections. A summary of the main steps involved in extracting features is illustrated in the Figure 4.5. The workflow starts with the image

Figure 4.4: Sample set of input images from June 7 and June 19.

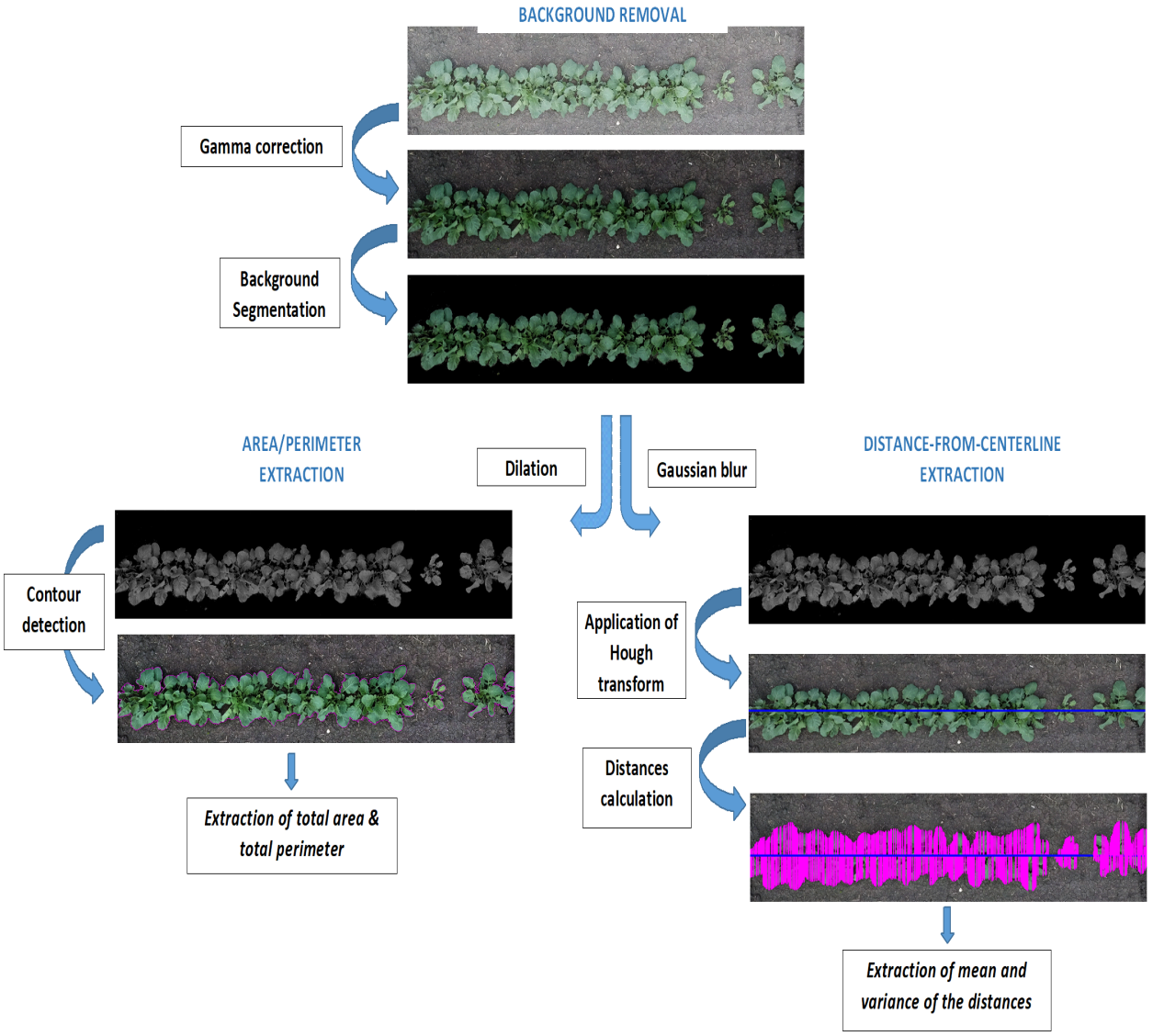


Figure 4.5: Feature extraction workflow.

enhancement, followed by background removal. After which, two different set of techniques were employed to extract the desired features.

4.2.1 Image Resizing

All image processing tasks were carried out in Python using `OpenCV` 3.1 module [71]. The segmented images were of varying sizes. To obtain a uniform number of pixels for height and width, the images were first subjected to resize operation (while keeping their aspect ratio) using the `INTER_AREA` interpolation algorithm available in `OpenCV`.

Figure 4.6: Before and after gamma correction for gamma=0.65 (top) and gamma=0.75 (bottom).

4.2.2 Image Enhancement

The dataset consisted of images of varying brightness, some looked appropriate, whereas some looked bleached out. This situation had risen because the filming of the crops took place on different dates and times of the day. Hence, gamma correction was used to enhance the images. Gamma correction, also called as power law transform, is used to correct the brightness of an image which follows a non-linear relationship between the input values and the mapped output values.

So, the image pixel intensities were scaled from the range [0, 255] to [0, 1.0]. From there, the output gamma-corrected image was obtained by applying the following equation:

$$O = I^{(1/G)}$$

where I is the input image and G is the gamma value. The resulting image O is then converted back to the range [0, 255]. Different gamma values ranging from 0.65 to 0.95 were used by manually tuning the parameter for different image sets.

Figure 4.7: Background removal after the gamma correction.

4.2.3 Background Removal

The next task after gamma correction was to threshold the region of interest from the images. This was done by a sequence of two operations, color detection and bitwise AND. The gamma corrected images were first transformed to the HSV (Hue, Saturation, Value) color space and a manually-determined threshold was applied for a range of green color. The green portions of the plants were prominent in this color space and made the color segmentation effective. The value of green color in HSV space is 120 which is in range $[0, 360]$. But `OpenCV` uses a different scale for H value which is $[0, 180]$. However, S and V are still in range $[0, 255]$. As a result, the value of H for green is 60 because the H value is halved to fit the range $[0, 255]$. Therefore, a mask of the green plants was created by setting the lower bound to $[30, 30, 50]$ and the upper bound to $[100, 255, 255]$. All the green objects in this colorspace were easy to extract as tints, tones and shades of green have the same hue. The masking operation resulted in a binary image. The bitwise AND operation was then performed. It computes per-element bit-wise logical conjunction. The `cv2.bitwise_and` of `OpenCV` takes three arguments: two input arrays and an 8-bit operation mask, that specifies elements of the output image to be changed. The gamma-corrected image and the original image were the two input arrays in this case and the advantage of the third argument was exploited by setting the binary image generated before as the mask. The output was a clean segmented image displaying only the pixels that have a corresponding white value in the mask. The results of the above-mentioned steps are summarized in Figure 4.7

After separating out the plants from the background, two different sets of methods were applied to the color segmented images.The following two paragraphs describe the two sets of methods in detail.

4.2.4 Area and Perimeter Extraction

The first set of methods aimed to extract the area and perimeter covered by the plants. The segmented images were converted to grayscale using the color-space conversion method `COLOR_BGR2GRAY` available in

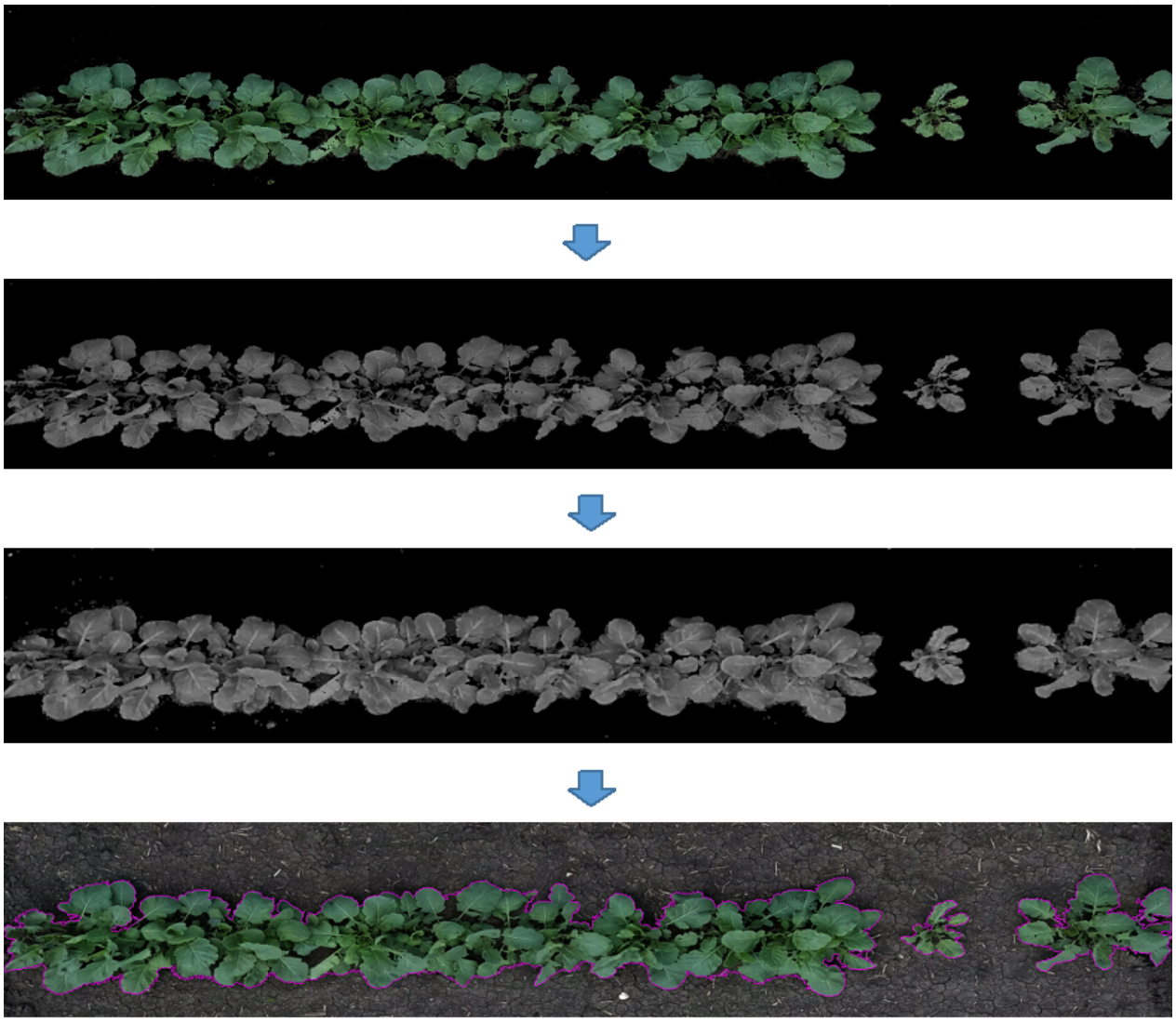


Figure 4.8: Steps involved in extracting area: conversion to grayscale, dilation, contour detection.

`OpenCV`, to enable application of morphological operations. Sometimes, there might be some unwanted small spurious holes in the images. To resolve this issue, dilation was employed to close small holes in the grayscale image prior to further analysis using a square structuring element of size 3×3. As dilation adds pixels to the boundaries of the object in an image, the gap between the disconnected regions gets smaller and sometimes filled. Now, the contour detection algorithm was run using the `OpenCV` library `findContours`. The results of these four steps are shown in Figure 4.8. Contours are the boundaries of a shape with same intensity [62]. Contours can be ordinarily defined as a curve passing through all the continuous boundary points having same color or intensity, therefore, they are useful for shape extraction and object detection [62]. The area and perimeter of the plants in units of pixels were computed using the contour feature methods called `contourArea` and `arcLength`.

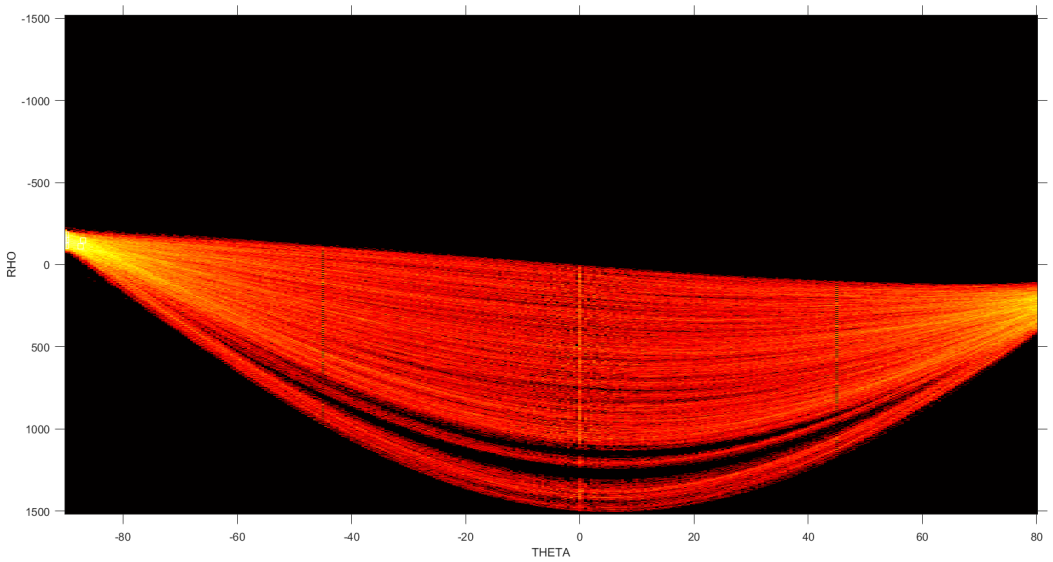


Figure 4.9: Peak in the hough space.

4.2.5 Distance-from-centerline Extraction

The second set of methods was intended to extract the extent of the growth from the centerline. The grayscale image was then smoothed using Gaussian blur. This step was necessary to remove the noise from the image. A Gaussian kernel of size 3×3 was used for the blurring process to remove noise from the image. After cleaning the image, it is then used as the input for the detection of the crop row lines. When discussing with breeders how they scored vigor, the concept of leaf area and leaf coverage shape featured prominently in their descriptions. To operationalize the idea of leaf coverage shape we used the distance of each leaf pixel to the centerline of the plot as a base metric, and the mean and variance of this measure as the features. To determine measures of the size of symmetry of the plants in the image, we used a measure of the distance from the centerline, which required computing the centerline using the well known Hough transform. Prior to applying the Hough transform, the input image needs to be in the form of an edge image. To serve this purpose, the blurred image was skeletonized [115]. The `skeletonize` function removes the outermost region of the object and leaves a skeleton remnant preserving the shape and topology of the object. Rather an edge image, we used a skeleton image as an input to the Hough transform. The threshold parameter of the algorithm was set to 300, which means the minimum number of votes it should get for it to be confirmed as a line. The algorithm outputs a vector of lines from which the first most significant line can be chosen. The skeleton of the object consisting of multiple line segments help the Hough transform in detection of the line. This second set of processing steps are summarized in the Figure 4.11. The votes accumulated in the accumulator array corresponding to the line can be visualized in Figures 4.9 and 4.10.

Once the centerline was detected, we calculated the extent of growth with respect to it. For this, the

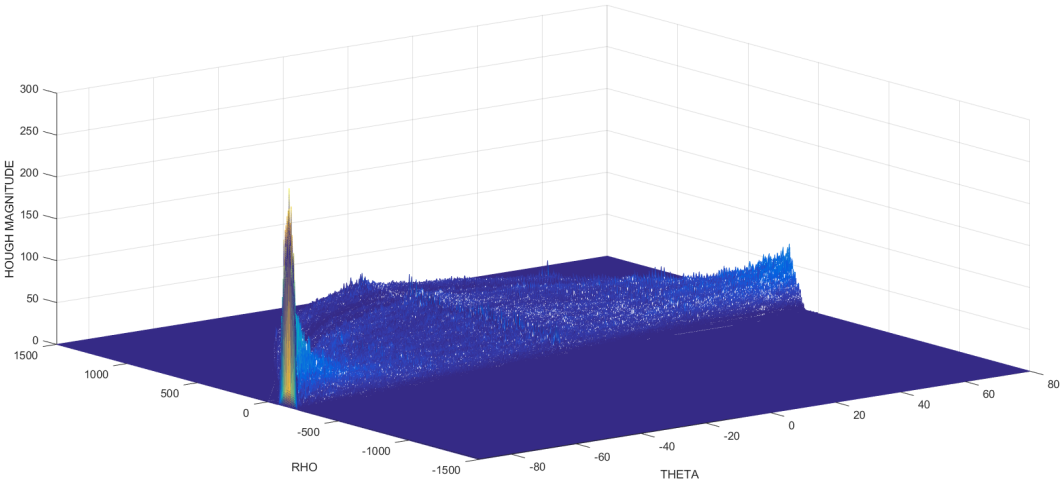


Figure 4.10: A 3D plot of the hough space.

points lying on the boundary of the objects (referred to as boundary points for the remainder of the paper) were considered. The perpendicular distance between the boundary points and the row centerline was the basis for feature computation. The row centerline was chosen because it provided the best reference for determining the spread of plant foliage from the hinge point at the intersection of stem and root systems, assuming that the hinge points of the plants lie approximately on the row centerline. The distances from the centerline was calculated by first obtaining the line equations of the centerline and the perpendicular line (the line representing the distance from the centerline to the boundary point). After which, the length of the perpendicular line was calculated using the equation 4.4. The line equation of the row centerline can be represented as:

$$y = ax + c \tag{4.1}$$

where a is the slope of the line and c is the intercept. The line equation of the perpendicular line passing through a boundary point can be represented as

$$y = bx + d \tag{4.2}$$

where b is the slope of the line and d is the intercept. Next, the points of intersection of these two lines were found using the equation

$$\left(\frac{d-c}{a-b}, \frac{ad-bc}{a-b}\right) \tag{4.3}$$

The derivation of the equation 4.3 can be found in [110]. The perpendicular distance from the centerline was calculated as follows:

$$\sqrt{(x_2-x_1)^2+(y_2-y_1)^2} \tag{4.4}$$

where (x_1,y_1) and (x_2,y_2) are the coordinates of the two end-points of the line. The resultant perpendicular distances between the boundary points and the centerline for an image is shown in Figure 4.11.

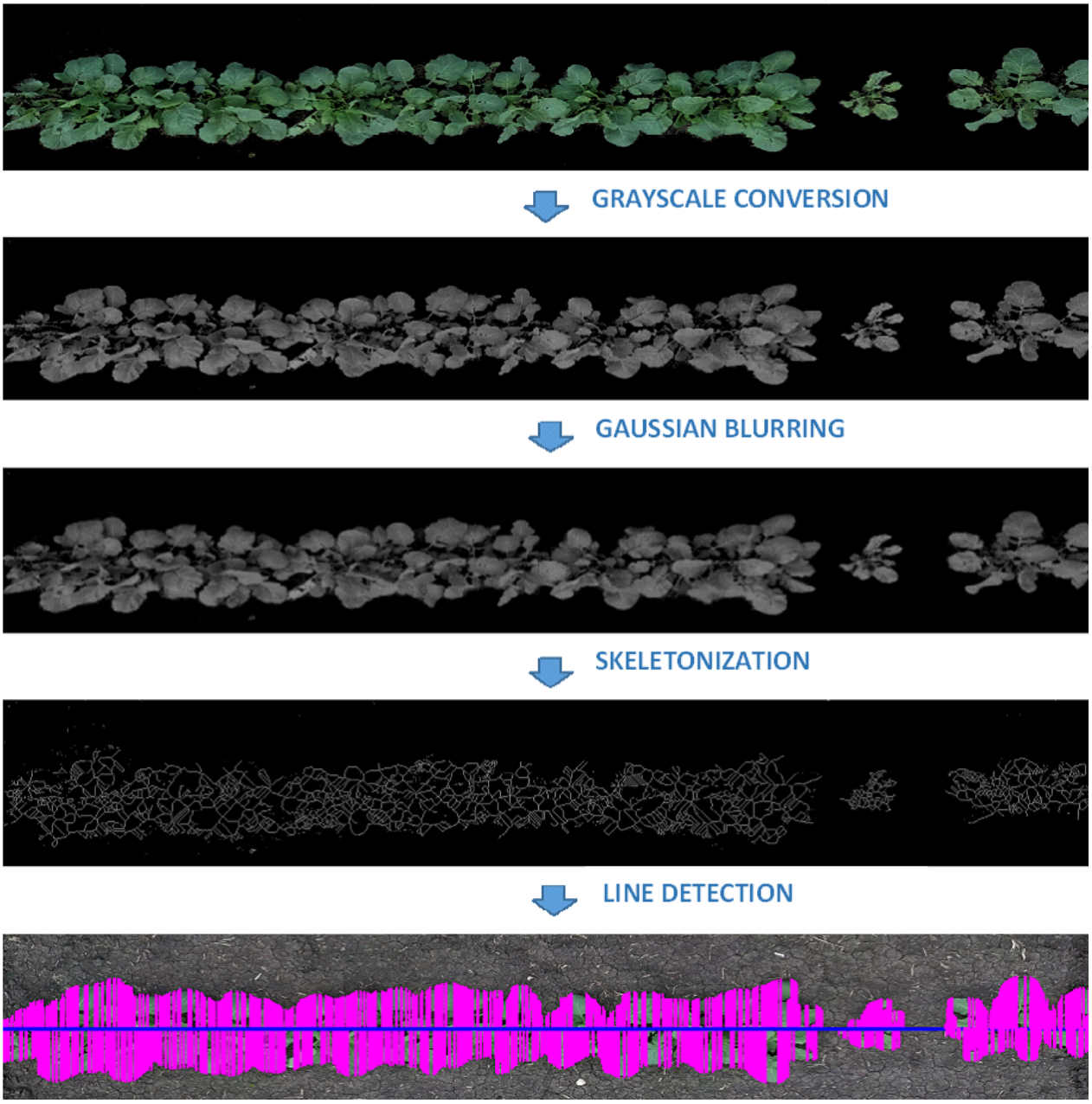


Figure 4.11: Steps involved in detection of centerline.

Plot number	MeanDis7	VarDis7	TotalArea7	TotalPeri7	MeanDis19	VarDis19	TotalArea19	TotalPeri19	Score
2608	7.18	30.95	550	671.41	26.28	280.31	9392	1009.98	1
1188	9.41	47.60	846	301.22	48.05	793.84	92827	15395.79	2

Table 4.1: Data sample.

4.3 Data Overview

From the images, eight features (in pixel units) were extracted from two different dates: June 7 and June 19 as shown in the table. The feature values (shown in Table 4.1) for every plot were computed and saved in a CSV file which acted as the input training set for our machine learning based prediction model. The final dataset consisting of the eight features and that used for training the vigor rating prediction models are shown in Table 4.1.

The final dataset comprised of 294 data points grouped into three different classes where 44 data points belonged to class 1 (plants of low vigor), 217 data points belonged to class 2 (plants of medium vigor) and 33 data points belonged to class 3 (plants of high vigor). As shown in Figure 4.12 the number of class 2 samples is highest compared to classes 1 and 3. This could be because the process of rating is subjective and was done by two people. Additionally, the priors of the vigor classes are not equal. This imbalance in data might cause a bias in the process of prediction, prompting the model to classify samples in class 2. To avoid this behavior, the classes 1 and 3 were augmented. This was done by resampling each class 1 and 3, to 150 data points before training the model using the `resample` module provided by `scikit-learn` (version 0.19.1) [71]. The essence of resampling is to use only the sample data and to resample from that data to generate various realizations of the experimental outcomes. In resampling new data points were generated by randomly drawing the features from all possible permutations of the data samples, ensuring that new samples have similar attributes to that of the original data sample. This was achieved by setting the hyperparameter `replace` to `False`. Resampling produces a unique sampling distribution on the basis of the original data. Additionally, it yields unbiased estimates as it generates data points based on the unbiased samples of all the possible results of the original data. The random forest and support vector machine models were trained on the augmentated dataset. Details on building these prediction models follow.

Figures 4.13 and 4.14 represent the histogram distributions of all the eight features after upsampling the minority classes, i.e. vigor rating 1 and vigor rating 3. Blue bars correspond to class 1 samples, green bars correspond to class 2 samples and class 3 samples are represented by red bars. The x-axis represent the range of feature values and the y-axis represent the number of samples falling in the bins. The number of bins was set to 20 and the bin size can be obtained by

$$\frac{(maximum\,feature\,value - minimum\,feature\,value)}{20} \tag{4.5}$$

. The blue bars dominate the lower range of the feature values as they belong to class 1, i.e. low vigor

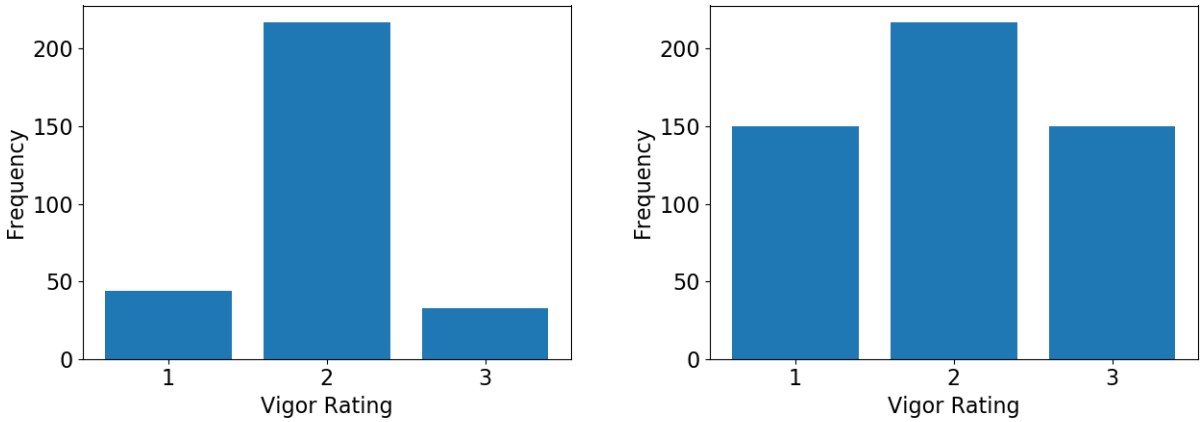
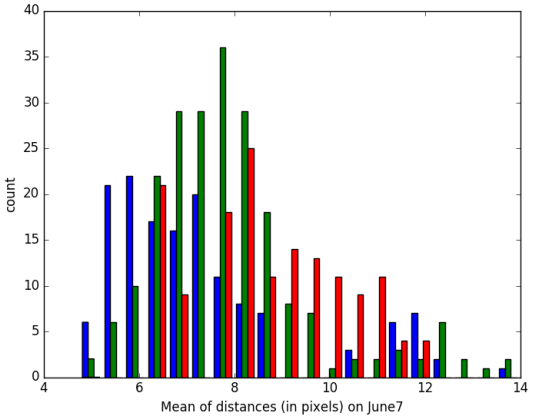
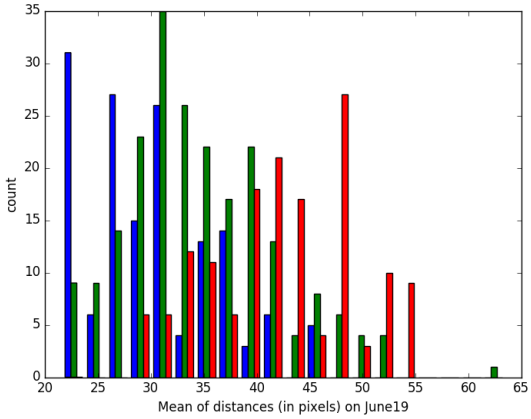


Figure 4.12: Histogram of vigor classes before and after augmentation.

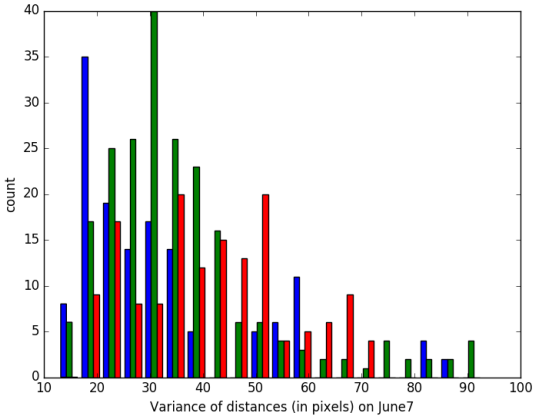
samples. Because they are rated as low vigor, their extent of growth from the centerline and the ground cover are expected to be of low magnitude. Similarly, the high vigor samples represented by the red bars are on the higher range of the feature values because of their larger growth. The green bars rendering the samples of medium vigor, are spread all over the graph, overlapping with other classes' feature values. However, the green peaks are observed in the middle range of the features.



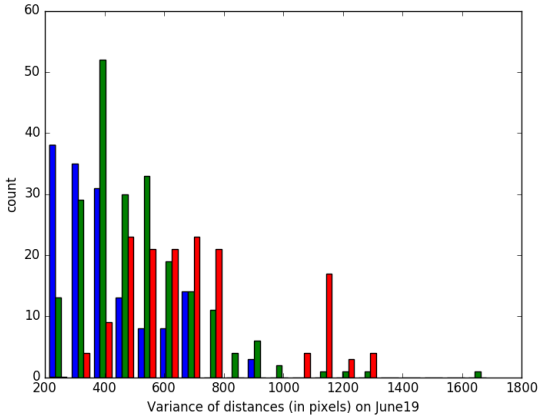
(a) Mean of distances on June 7.



(b) Mean of distances on June 19.



(c) Variance of distances on June 7.

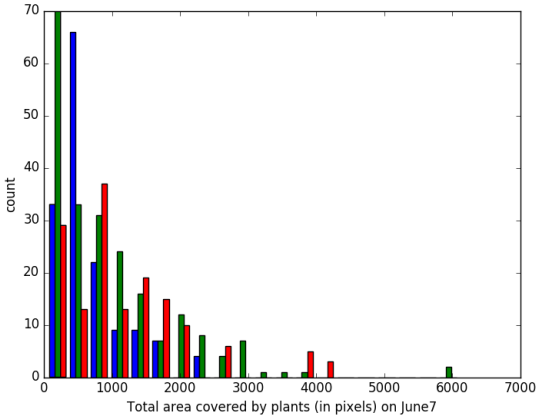


(d) Variance of distances on June 19.

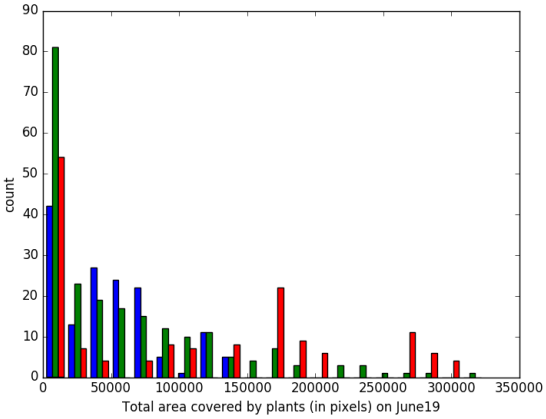
Figure 4.13: Feature distribution for each class.

Correlations and scatter plot matrices were used to determine if relationships exist between the measured features and the vigor ratings. The correlation matrix is shown in Figure 4.16 and scatter plot matrices are shown in Figure 4.15. Correlation value calculates the strength of linear relationship between two variables. It lies between -1 and $+1$, where $+1$ indicates perfect positive linear relationship while -1 indicates perfect negative linear relationship.Correlation value of 0 indicates no linear relationship at all. Correlation values above 0.5 indicate moderate linear relationship. In Figure 4.15 it can be seen that there is high correlation between variance of distances and the mean of distances on the same day. Similarly, high correlation can be found between area and the perimeter of the plants on the same day, while on the other hand there is very less correlation among other feature pairs. In Figure 4.16 the main diagonal of the matrix contains the covariance between each variable and itself. The non-diagonal values in the matrix represent the covariance between the two variables. Figure 4.16 shows low correlation values ranging from 0.0 to 0.25 indicating the presence of a weak linear relationship between the extracted features and the vigor class labels.

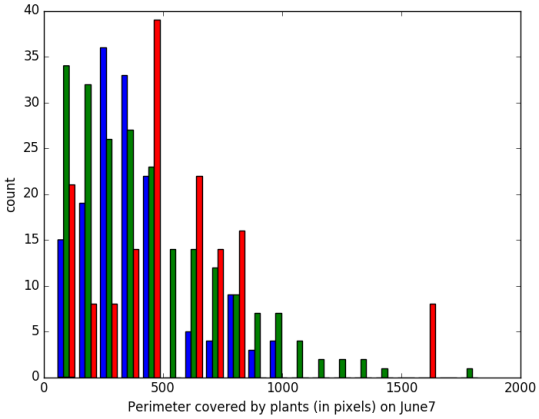
This indicates that linear models are not sufficient to predict the vigor ratings. Machine learning models



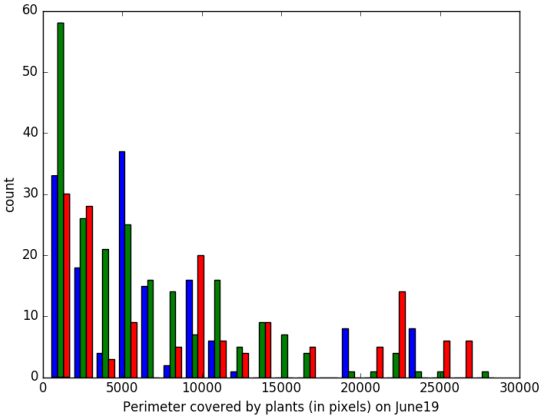
(a) Total area covered by plants on June 7.



(b) Total area covered by plants on June 19.



(c) Total perimeter covered by plants on June 7.



(d) Total perimeter covered by plants on June 19.

Figure 4.14: Feature distribution for each class.

are able to capture the non linearity in data and can perform well in identifying pattern. Random forest and support vector machine were experimented with the dataset and the results of their performance are outlined in chapter 5.

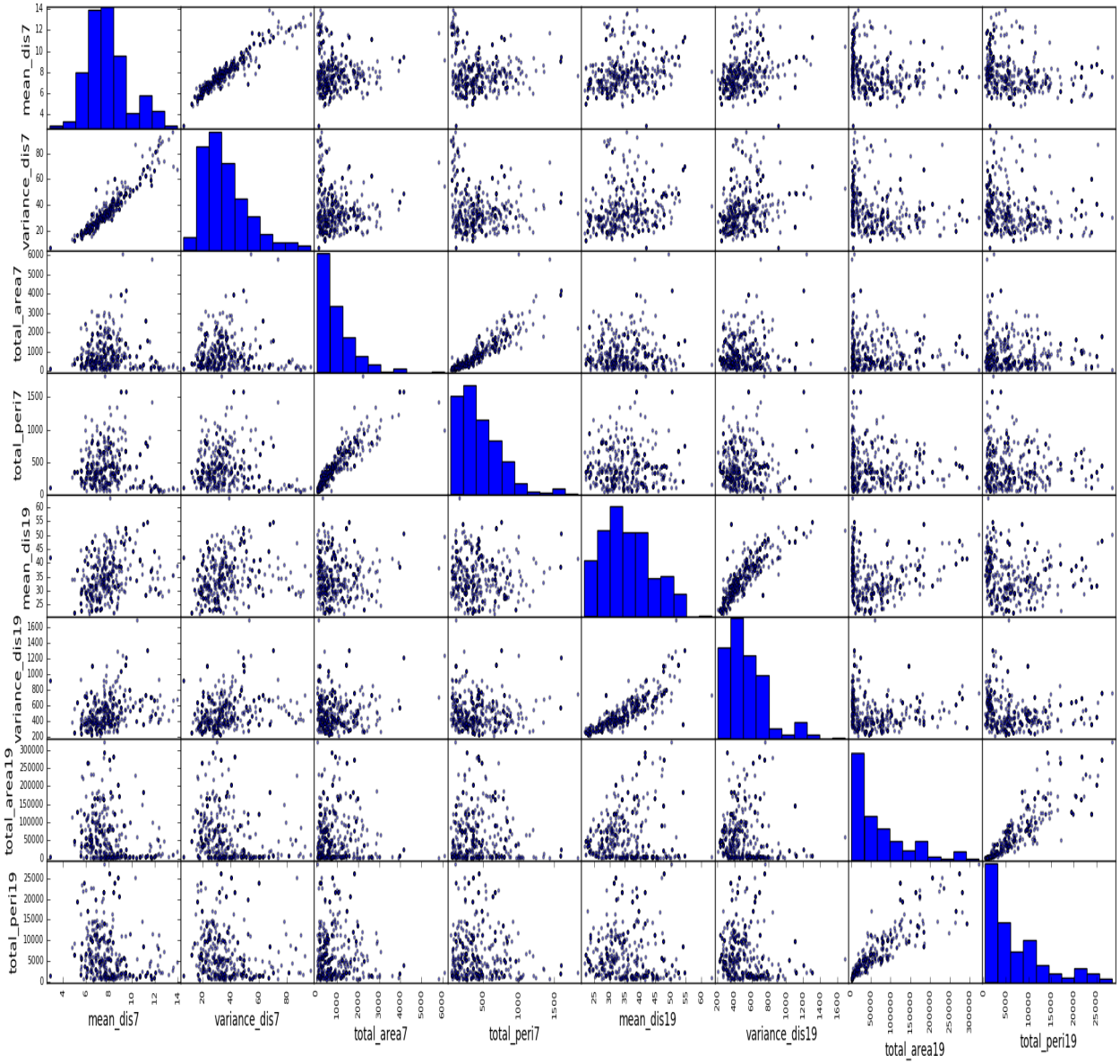


Figure 4.15: Scatter plot matrices of correlation among features.

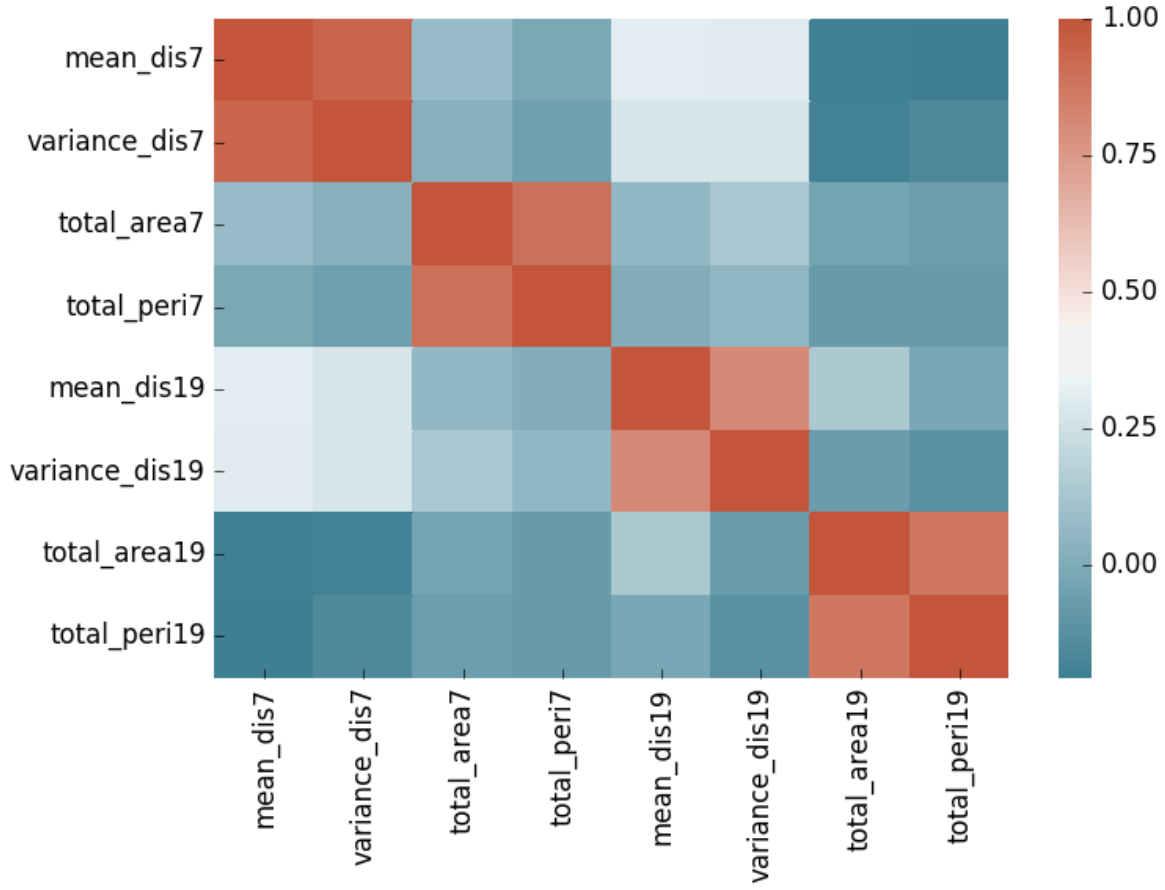


Figure 4.16: Correlation matrix of features.

5 Methods

The major research question this work attempts to answer is whether or not the crop vigor rating of a human can be estimated automatically from images. To learn vigor ratings from the extracted features, two popular classification models, random forests and support vector machine, were selected. Machine learning models were created using `scikit-learn` [71] module.

Performance of a classification model is measured by using a test set consisting of labeled data samples. Initially, the classifier is trained on a training dataset, and then tested on a previously unseen test dataset. Finally, the accuracy of the classifier is calculated by comparing the predicted labels with the true labels. K-fold cross-validation is used to evaluate the quality of the random forests and support vector machine models. During the cross-validation, the data set is divided into K subsets of data from which one set is treated as a test set while the remaining $K-1$ sets of data are used for training the model. During each run, different subsets are treated as test sets exactly once. In this way, the performance of the machine learning model is tested on previously unseen data. For this work, K was set to 10 leaving 10% of the dataset for testing and 90% of the dataset for training. In this way, cross-validation helps in knowing how the machine learning model would generalize to an independent data set.

5.1 Random Forest

Tree algorithms usually adopt a top-down approach, by choosing a feature and its value at each node that best splits the set of samples. The different metrics (e.g. Gini impurity, information gain) that decide the best splitting criterion, generally measure the similarity of the target variable within the subsets. These metrics are applied to each subset of samples and the resulting values are combined to evaluate the quality of the split. In reality, both Gini impurity and Entropy usually generates similar results [76]. The metric used in this study is Gini impurity. The splits were chosen in such a way that the Gini impurity in the resulting subsets is reduced. The Gini impurity is the difference between the sum of the squared probabilities of each class from one. Mathematically, it is defined as

$$G = 1 - \sum_{i=1}^{C} p_i^2 \tag{5.1}$$

where p_i denotes the proportion of instances belonging to the class $i (i = 1 \cdots C)$. Gini impurity is minimum (i.e. zero) when all data samples in the node are classified into a single class.

Several model parameters are present in random forests, which were selected by random grid search

techniques.

5.1.1 Number of Estimators

The number of estimators represents the number of decision trees random forests employs on various sub-samples of the dataset. Usually, a higher number of trees leads to better learning of data. However, increasing the number of trees can increase the computation time of the training process considerably. Several experiments were carried out to find the right value of the hyperparameter and the results are shown in Figure 5.1. Accuracy increases with the number of estimators as it employs several decision trees to learn from the data. Higher accuracy was noticed with 12 estimators.

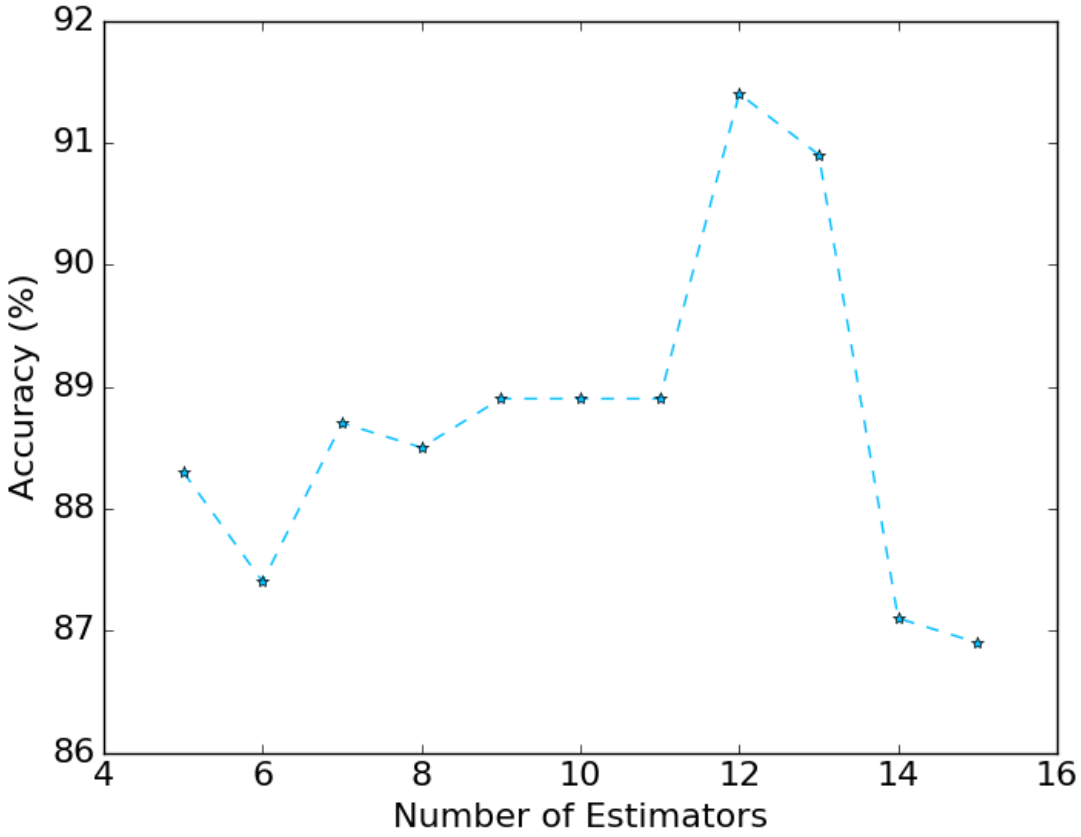


Figure 5.1: Classification accuracy with variations in the parameter.

5.1.2 Maximum Depth

Maximum depth represents the maximum depth of each tree in the random forests. Once the tree reaches the maximum depth it stops growing by not splitting nodes further during the building of the initial decision tree. More information about the dataset is captured by deep trees as deeper trees tend to have more splits.

However, the depth of the tree needs to be controlled to reduce the risk of over-fitting. Accuracies for different tree depths are shown in Figure 5.2

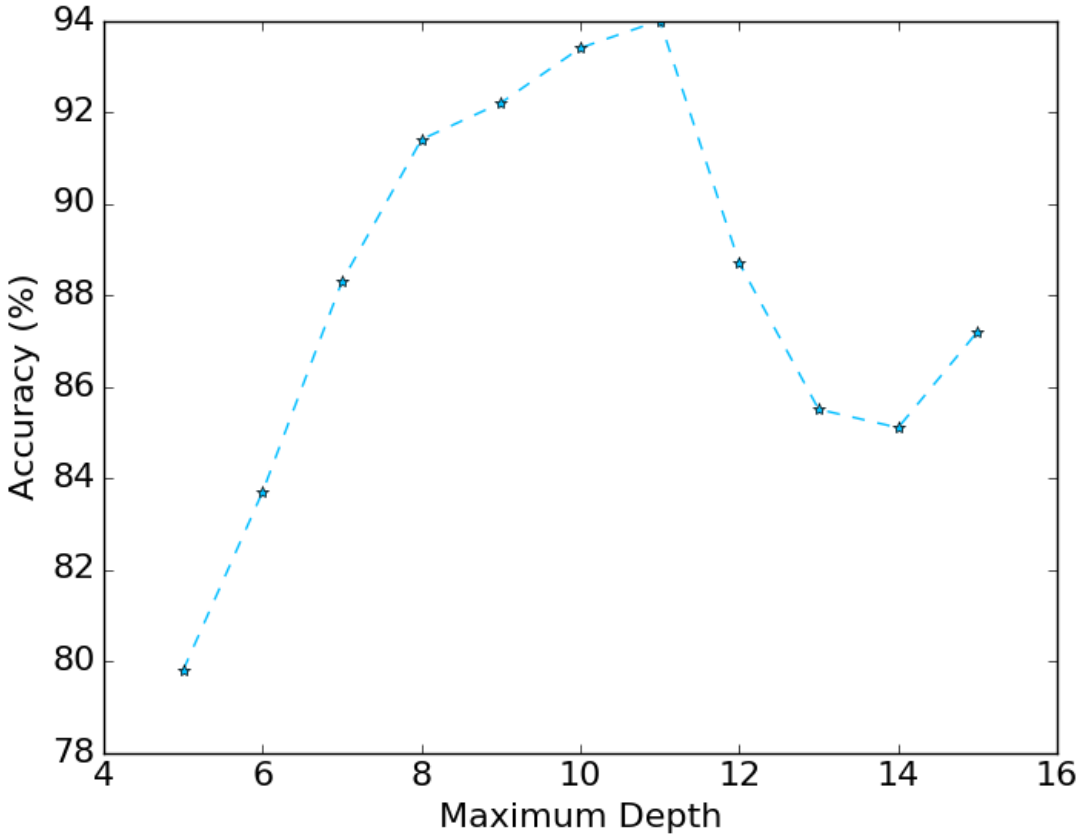


Figure 5.2: Classification accuracy with variations in the parameter.

5.1.3 Minimum Samples to Split

This parameter specifies the minimum number of samples a node must have before it can be split. Figure 5.3 shows the random forests accuracy for a various minimum number of samples. As indicated at a higher number of samples model performance decreases, this is likely because the model has to consider high number of samples while splitting a node. Whereas, at lower values, the model learns merely on two to three samples at a time.

After carrying out detailed analysis, the final values chosen for the hyperparameters of the random forest model were:

1. `Number of Estimators` =10

2. `Criterion` = Gini

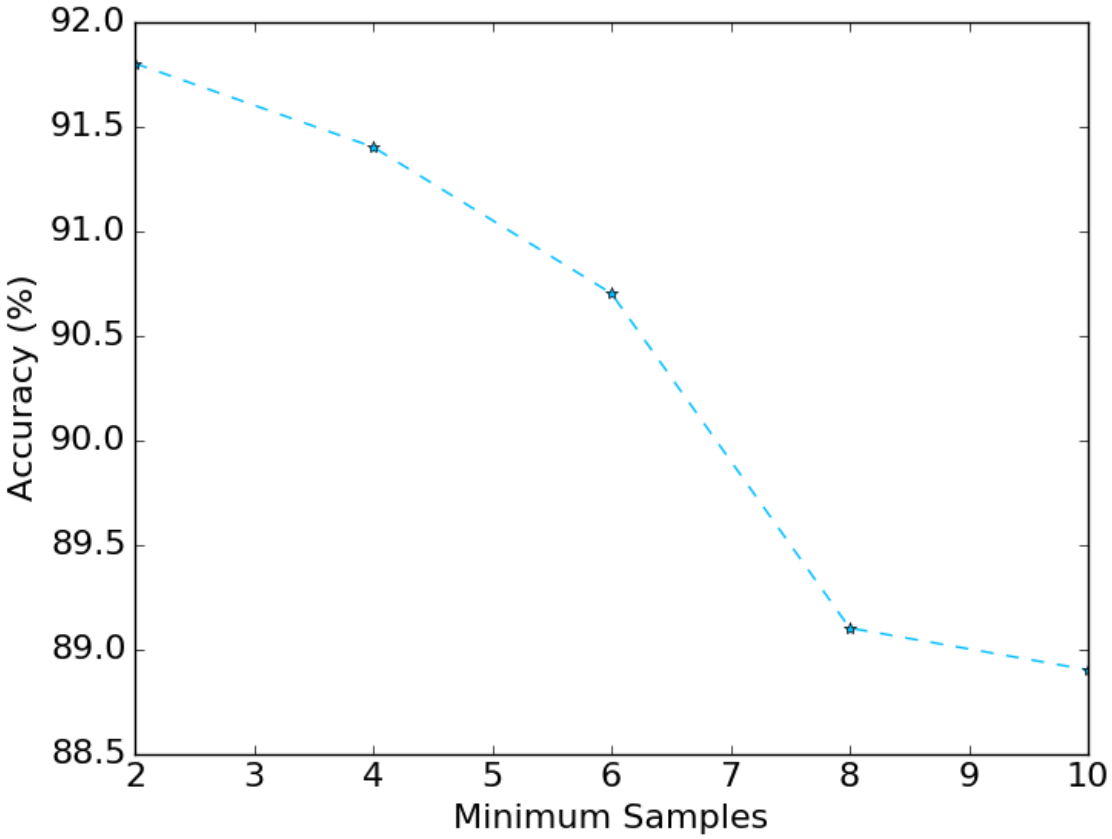


Figure 5.3: Classification accuracy with variations in the parameter.

Kernel	Accuracy
Poly	88%
Rbf	84%
Sigmoid	40%

Table 5.1: Accuracy for different kernels

3. `Max Depth` = 8

4. `Minimum Samples to Split` =5

5.2 Support Vector Machine

The support vector machine is a very popular classification algorithm, that is can solve both binary and non-binary classification problems. This algorithm serves as a comparison for random forests. We used the `module SVC (version 3.2.4.3.1.)` from `Scikit-learn` [71]. SVMs are very sensitive to feature scales. Feature values were scaled using `Scikit-learn's StandardScaler` which transforms the data such that its distribution will have a mean value 0 and standard deviation of 1. The pre-processed and scaled training data was used as input to the SVM classifier for training the model. The hyper-parameters of SVM were tuned to optimize the performance of the classifier. The tuning process involved running many experiments to find the best set of parameter values. The values were chosen through `Scikit-learn's Grid Search` technique. The effect of the parameters to the classifiers' behavior is as follows.

5.2.1 Kernel Parameter

It has been observed in chapter 4 that our dataset has no presence of linearity and thus, machine learning models are used to make the predictions. One approach in handling non-linear datasets is adding more polynomial features, which sometimes results in linearly separable datasets. However, at low polynomial degree, the difficulty arises in dealing with complex and large datasets whereas, at high polynomial degree, the number of features increases making the model slow. The 'kernel trick' of SVM simulates the result one gets by adding many polynomial features, without actually adding them. It acts as a mapping function that transforms the data into higher dimensional space. For this dataset, the polynomial kernel proved to have the best accuracy compared to other kernel as shown in the Table 5.1.

5.2.2 Gamma Parameter

Gamma, also known as kernel coefficient, defines the extent of influence of a single training data sample. High gamma values mean less extent and each data sample has a low impact on the new features. Small gamma

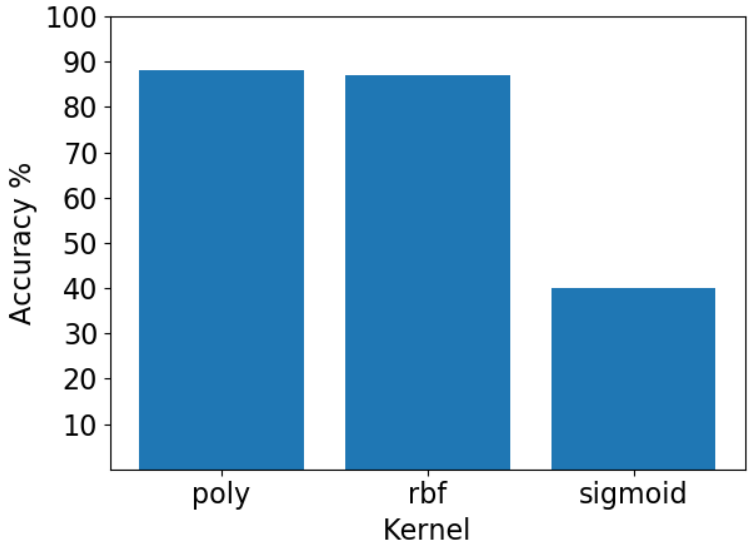


Figure 5.4: SVM accuracy for different kernels.

implies the data instances have large degree of influence on the transformed data and make the decision frontier smoother. Figure 5.5 shows the accuracies of the SVM model for different gamma values that were used while performing the analysis.

5.2.3 C Parameter

The C parameter trades off mis-classification of training samples against the simplicity of the decision boundary. Sometimes it is hard to find linearly separable planes in a dataset, in such cases, C acts as a penalty parameter that allows margin violations. A low C enables margin violations making the decision boundary smooth and large which leads to better classifications. To the contrary, a high C does not allow much margin violations resulting in many mis-classified samples due to small class boundaries. Figure 5.6 shows the accuracies of the SVM model that were obtained while analyzing the performance of the SVM model for different C values.

After carrying out detailed analysis, the final values chosen for the hyperparameters of the support vector machine model were:

1. `Kernel` =Poly
2. `Gamma` = 2
3. `C` = 0.1

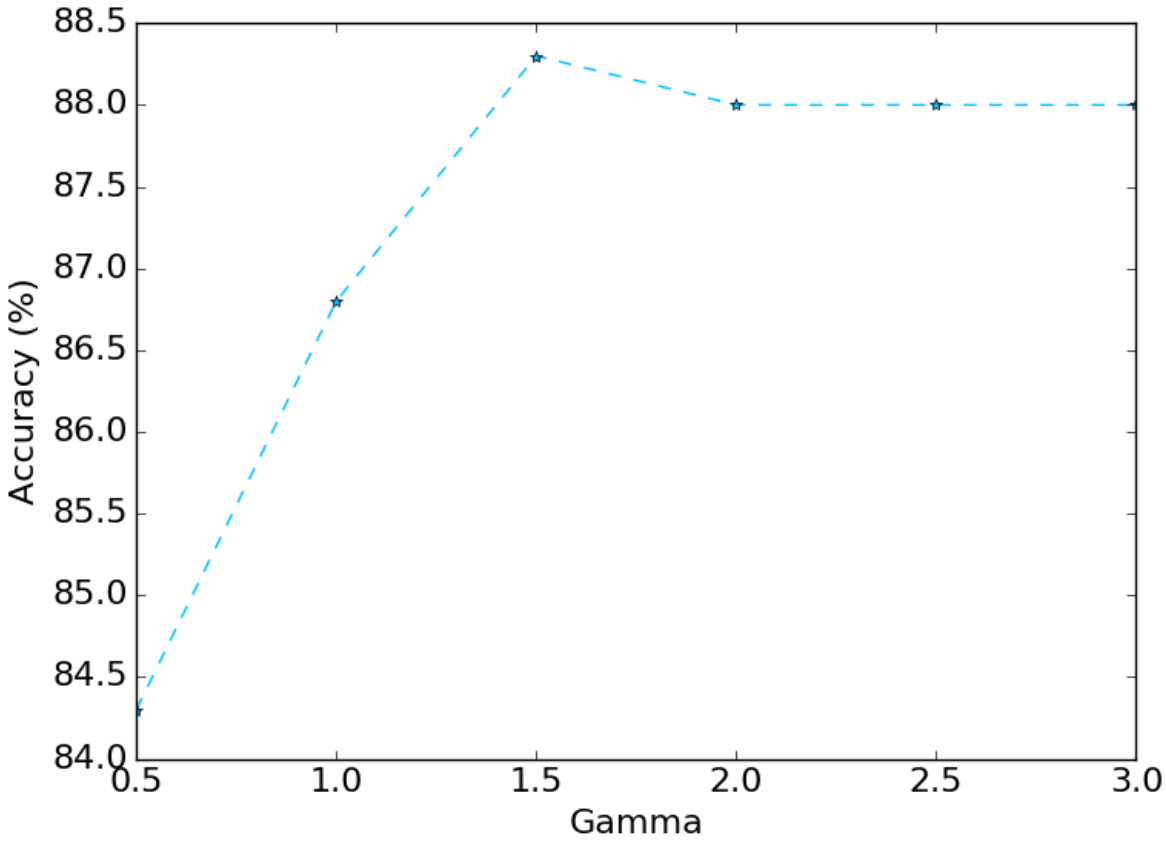


Figure 5.5: Classification accuracy with variations in the parameter.

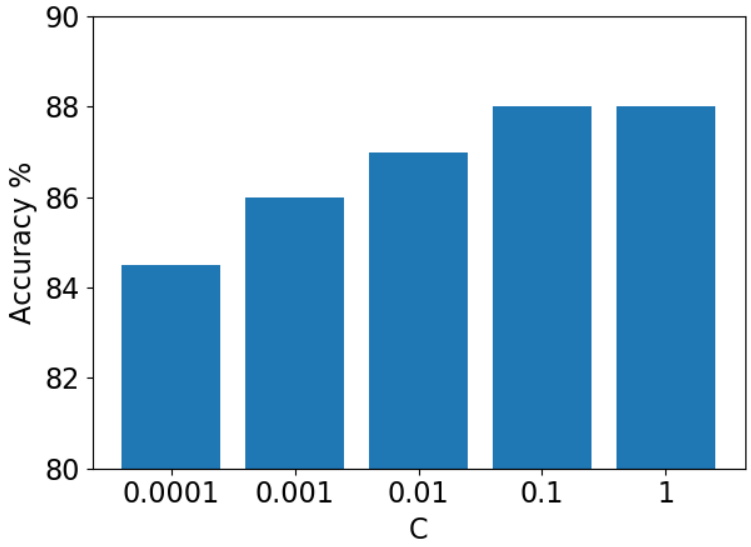


Figure 5.6: Classification accuracy with variations in the parameter.

6 Results

This chapter presents the critical evaluations of the created machine learning models, in other words, how well are they performing in predicting the vigor ratings. Ten-fold cross-validation was used to evaluate the models as K-fold cross-validation is a powerful measure to discover overfitting. It iteratively runs the algorithm on K-1 subsets of the whole dataset while using the remaining subset as the test set. So, using all the different parts of the training set as validation sets, cross validation gives a fairly good indication of how well the model generalizes to unseen data. Accuracy percentage and confusion matrices were used as performance metrics for the models.

The output of cross-validation is a confusion matrix. Whenever a data sample is used as a part of the test set, it has not been used for training that classifier. Hence, each data sample is used as a test sample exactly once. The confusion matrix provided by cross-validation is intuitively a fair indicator of the performance of the learning algorithm on independent test samples. Moreover, classification model accuracy can sometimes be misleading, as model generalizability cannot be adequately characterized using simple metrics. A confusion matrix was used to interpret the model fit.

6.1 Random Forests

6.1.1 Accuracy

Compared to the manual vigor ratings, the random forests model classified the plants with an accuracy of 88.97% which is better than previous studies using statistical methods [2, 13, 21]. The accuracy of each fold in 10-folds validation is presented in Figure 6.1. Here, the x-axis represents the accuracy and the y-axis represents the fold. The model performed reasonably well across all folds with accuracies above 80% and exceedingly well across folds $3, 4, 9, 10$ with accuracies greater than 90%. Compared to other folds, in fold 5, the model performed with the lowest accuracy of around 80%. From the Figure 6.1, we can say that there were no high fluctuations in the accuracy of the model.

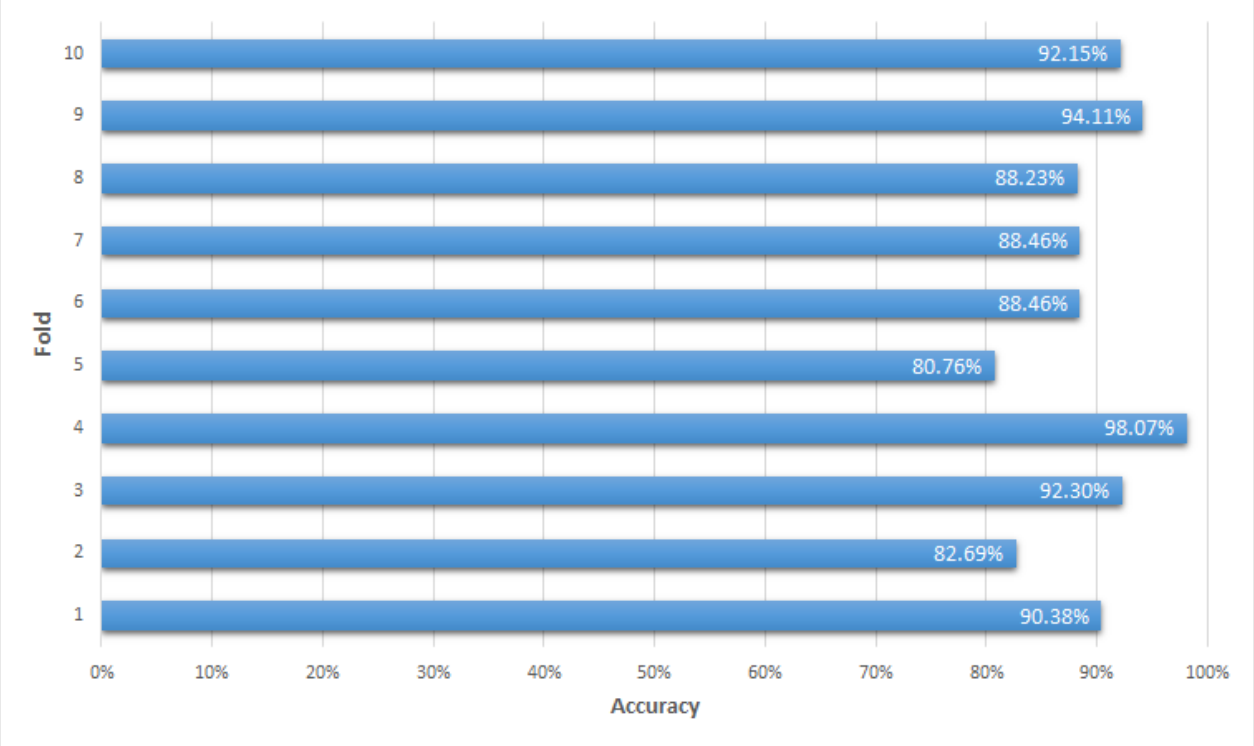


Figure 6.1: Accuracy of Random Forests model for each fold in 10-fold Cross-validation.

6.1.2 Confusion Matrix

The confusion matrix for Random Forests is shown in Figure 6.2. Here, the row represents the number of true class labels in a particular class whereas the column represents the number of samples predicted by the model for the corresponding class. Hence, the sum of the entries in a row of the matrix gives the total number of samples for the corresponding class. For instance, for class 1 there are 150 samples. Similarly, the total number of samples for class 2 is 217. Finally, class 3 has 150 samples. The diagonal line represents the

number of samples correctly classified, whereas the non-diagonal cells indicate the number of misclassified samples. The matrix represents the overall classifications and misclassifications done by the model from all K folds during the K-fold cross validation.

For class 1, out of 150 samples, 136 samples were correctly predicted as class 1 and 8 samples were wrongly labeled as class 2. For class 2, 176 samples were correctly predicted as class 2 whereas 23 and 18 samples were misclassified as class 1 and class 3 respectively. In the case of class 3, 148 samples were correctly predicted as class 3 and 2 samples were wrongly labeled as class 2. As shown there is a high number of samples across the diagonal, indicating the goodness of the performance of the model. The random forests model is clearly able to distinguish class 1 from 3, but can misclassify few borderline cases in class 2.

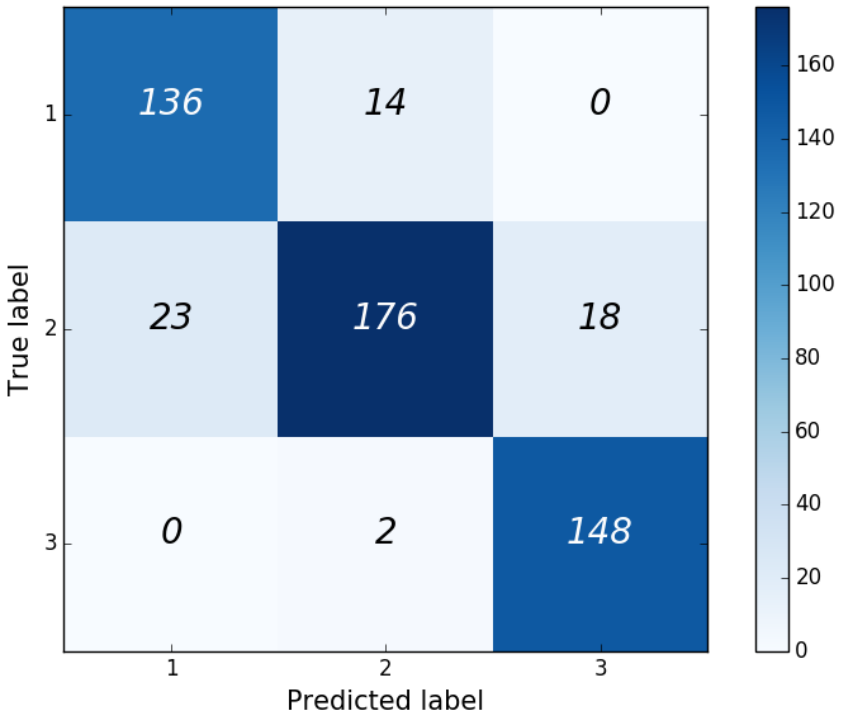


Figure 6.2: Confusion matrix for Random Forests model.

Overall, the model performed very well for class 1, class2 and class 3 with accuracies of 90.66%, 81.10% and 98.66% respectively. The model struggled a bit to generalize the Class 2 samples with a slightly low accuracy of 86% comparatively. The reasons causing this scenario could be that the features of samples belonging to classes 1 and 3, indicating low vigor and high vigor respectively, are quite distinct which helps the model in distinguishing samples and classifying them effectively. Conversely, the features of the samples belonging to class 2 overlap with those of class 1 and 3 leading to the model's confusion in classifying. Manual vigor ratings often come with human bias as it was performed by different people with varying judging abilities and different potential confounds.

6.1.3 Feature Importance

To understand how well the random forest utilized the extracted features in approximating the manual classifications, a schematic diagram representing the feature importances is shown in Figure 6.3. For a random forests model, the feature importance can be measured using the "Gini impurity" or "mean decrease impurity" which is defined as the total reduction in node impurity averaged over all the trees of the forest. Here, the feature importance values are normalized. The higher the value, the more important the feature is to the prediction. From Figure 6.3, it can be noticed that the three most important features were 'mean_dis19', 'variance_dis19', 'mean_dis7', implying that our novel distance from center-line feature was critically important to identifying vigor, and that both dates provided important information. The fact that both dates provide relevant data is important because it highlights the loss of measurement quality due to the requirement for direct measurement by experts in the field. The quality of vigor in the first measurement was not assessed because of the limited time of the expert. The camera operator and algorithm were not limited and therefore were able to incorporate the necessary information.

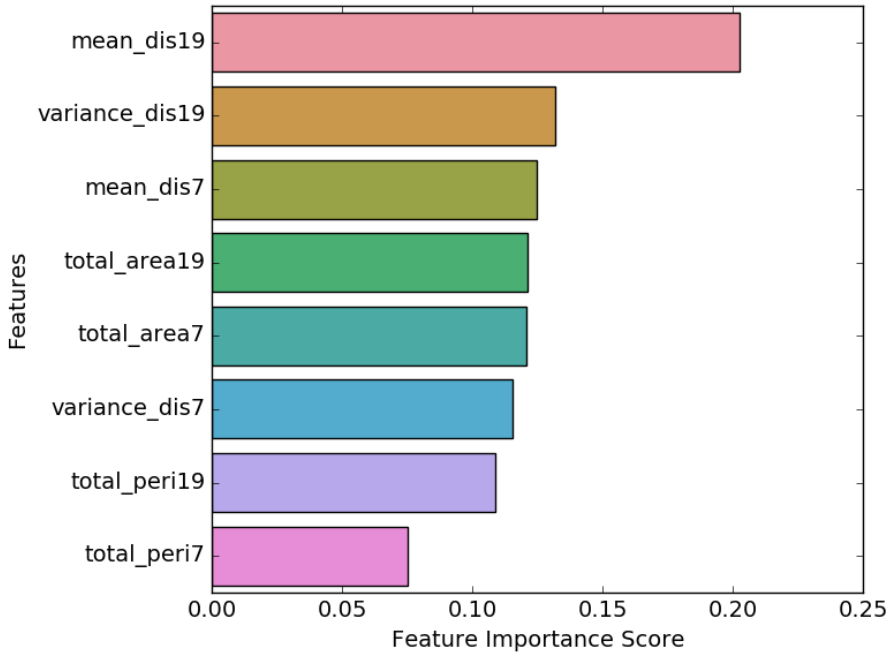


Figure 6.3: Feature importance score.

The value row in each node tells us how many of the observations that were sorted into that node fall into each of our three categories.

6.2 Support Vector Machine

6.2.1 Accuracy

In comparison to the manual vigor ratings, the support vector machine classified the plants with an accuracy of 88% which are better than the results of other studies [2,13,21]. Accuracy of each fold in 10-folds validation is presented in Figure 6.4. Here, the x-axis represents the accuracy and the y-axis represents the fold. The model performed fairly well across all folds with accuracies above 80% and exceedingly well across folds $7, 8, 9$ with accuracies around 90%. Compared to other folds, in folds 2 and 5, the model performed with lowest accuracy around 80%. Generally, the model performed consistently well across all folds with varying training and test sets.

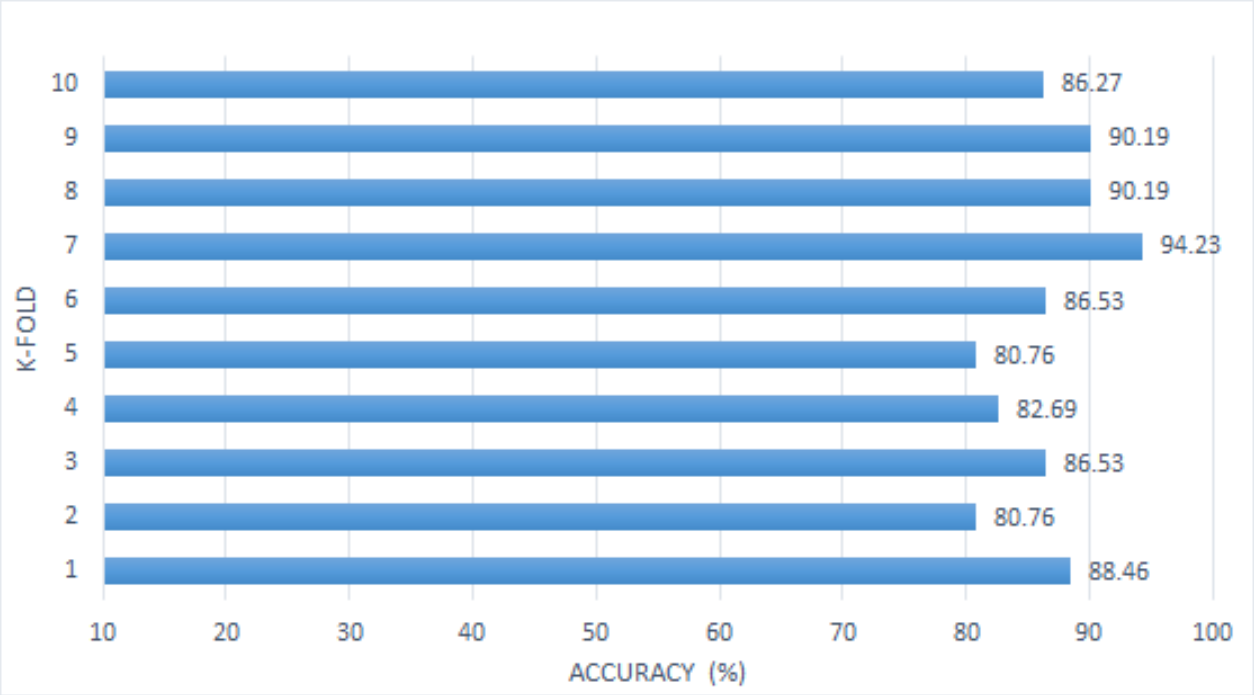


Figure 6.4: Accuracy of support vector machine for each fold in 10-fold Cross-validation.

6.2.2 Confusion Matrix

Confusion matrix for support vector machine model is shown in Figure 6.5. This learning algorithm is on par with the random forest model with a classification accuracy of 88%. From the confusion matrix, it can be said that 146 classifications were made correctly in the case of class 1 whereas in case of class 3, there was just one misclassification and 149 correctly classified samples. Overall, the model performed exceedingly well in making classifications for classes 1 and 3. Compared to the random forest model, the number of wrongly classified samples in case of class 2 is higher. There were 35 samples that were wrongly classified as class 1

and so were 22 samples as class 3. A similar pattern in classification is observed in this case of prediction model too. Samples of classes 1 and 3 were easily distinguishable from each other but not with the class 2 samples. The model classified the class 1 samples with an accuracy of 97.33% and class 3 samples with an accuracy of 99.33%. Class 2 classifications were made with a low accuracy of 73%. However, the number of correctly classified samples are high in number and lie across the diagonal of the confusion matrix indicating an acceptable performance by the model.

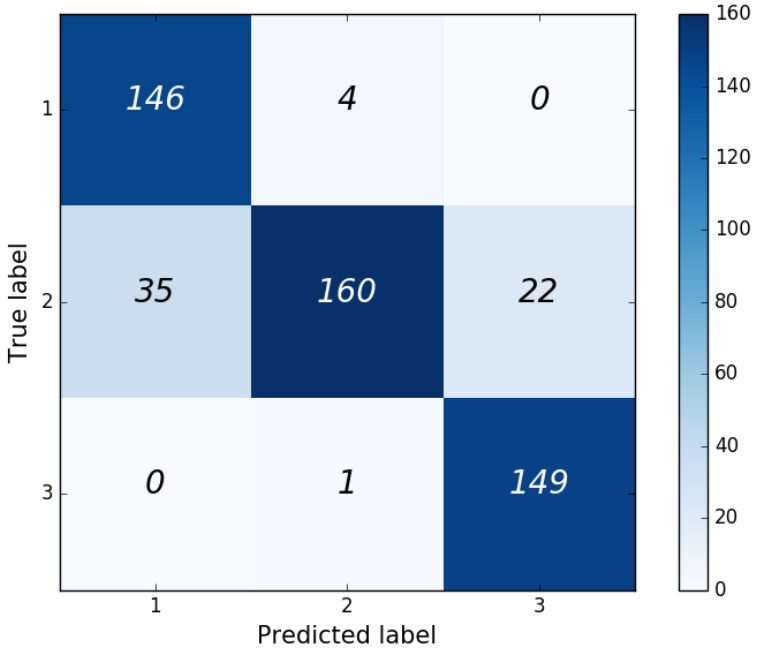
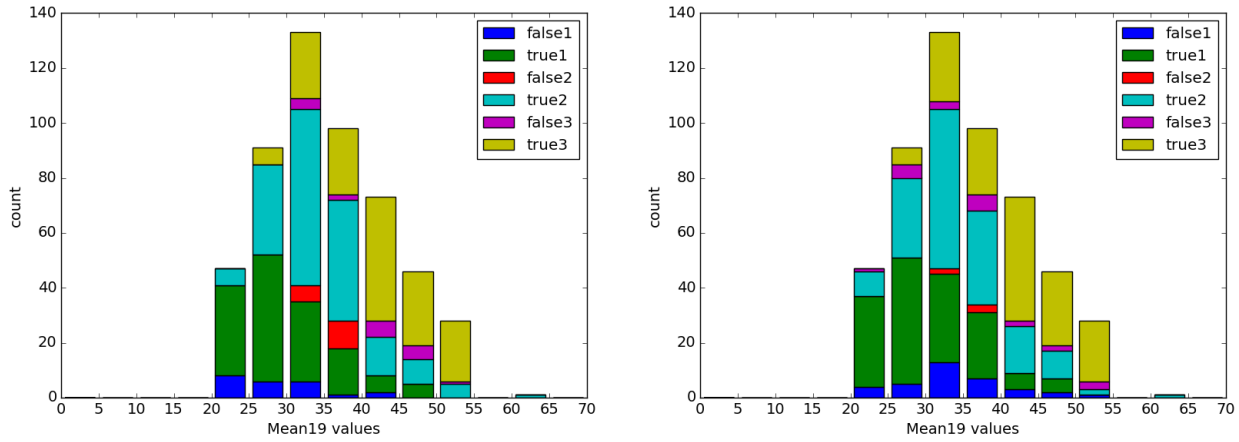


Figure 6.5: Confusion matrix for support vector machine.

6.3 Analysis of the Model Predictions

For further analysis, histograms showing the number of correctly and incorrectly classified samples for each class in each feature bin were plotted. In the following histogram plots as shown in Figures 6.6 to 6.13 , the dark green color represents the samples correctly classified as class 1, the sky blue color represents the samples correctly classified as class 2 and the lime green color represents the samples correctly classified as class 3. Similarly, the blue color stands for the samples wrongly classified as class 1, the red color represents the samples wrongly classified as class 2 and the magenta color represents the samples wrongly classified as class 3.

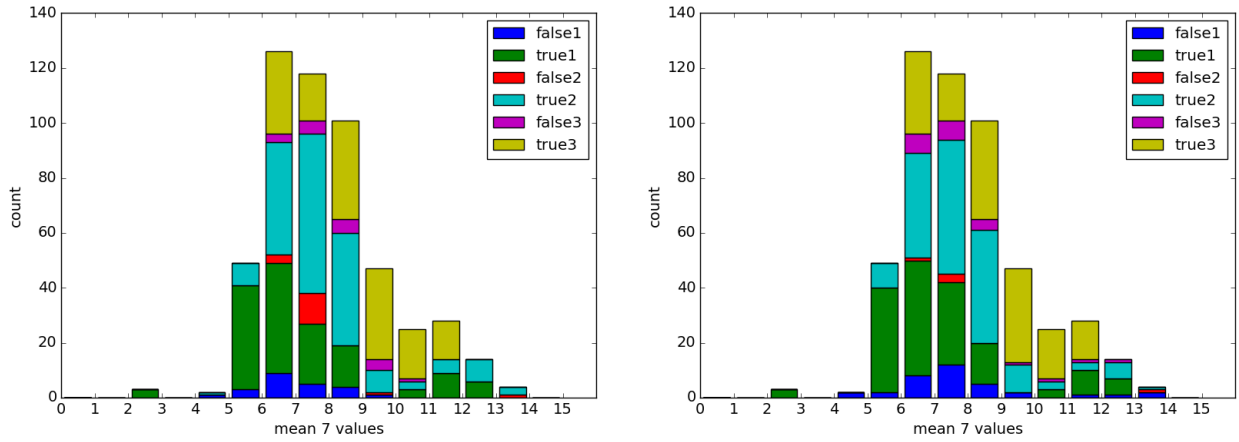
These histogram plots demonstrate that no one feature provided a clear demarcation between classes, as a significant overlap between classes for each feature value can be observed. This clearly indicates that statistical analysis alone is not enough to estimate or predict the plant vigor and joint associations between features need to be considered using non-linear models such as Random Forests.



(a) Histogram of Random Forest.

(b) Histogram of SVM.

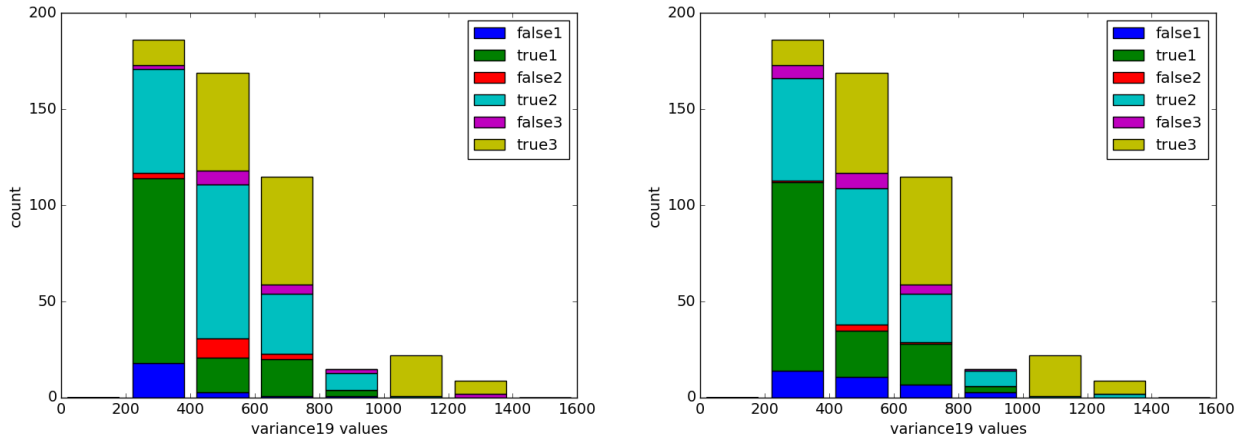
Figure 6.6: Stacked histogram for class predictions based on the mean of distances on June 19.



(a) Histogram of Random Forest.

(b) Histogram of SVM.

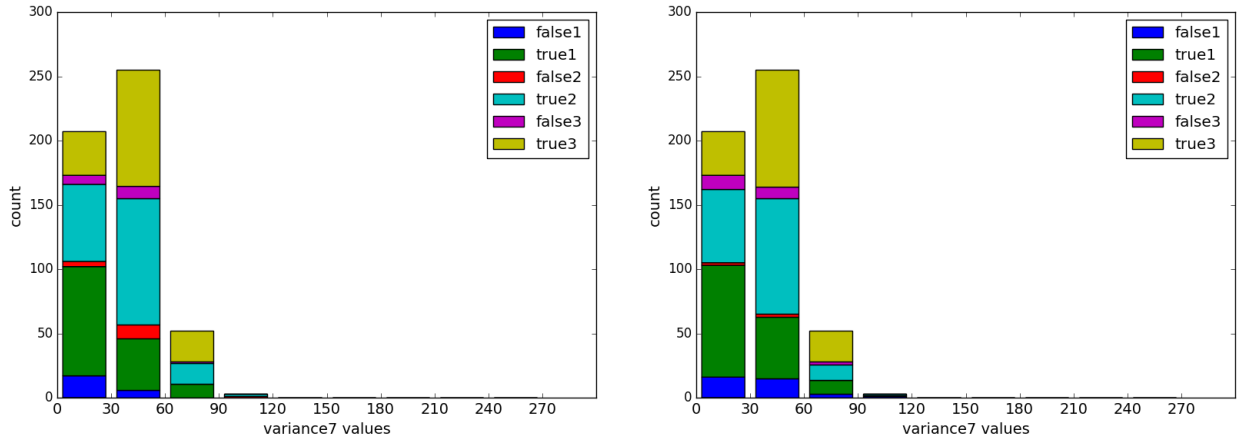
Figure 6.7: Stacked histogram for class predictions based on the mean of distances on June 7.



(a) Histogram of Random Forest.

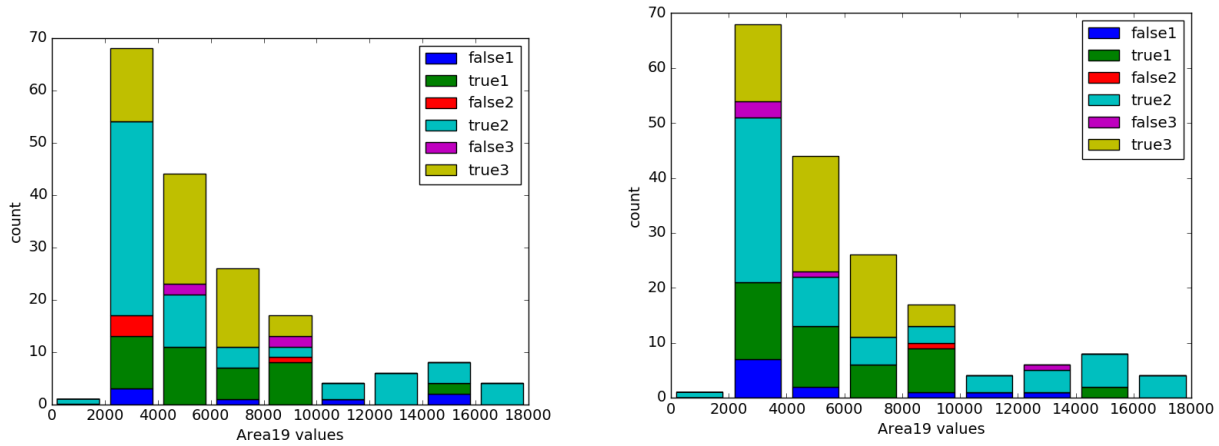
(b) Histogram of SVM.

Figure 6.8: Stacked histogram for class predictions based on the variance of distances on June 19.



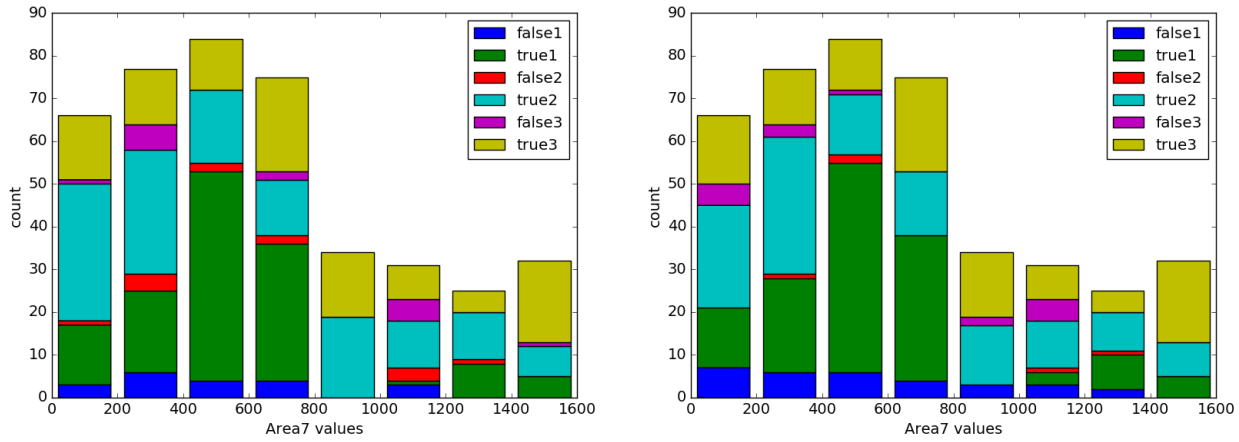
(a) Histogram of Random Forest. **(b)** Histogram of SVM.

Figure 6.9: Stacked histogram for class predictions based on the variance of distances on June 7.



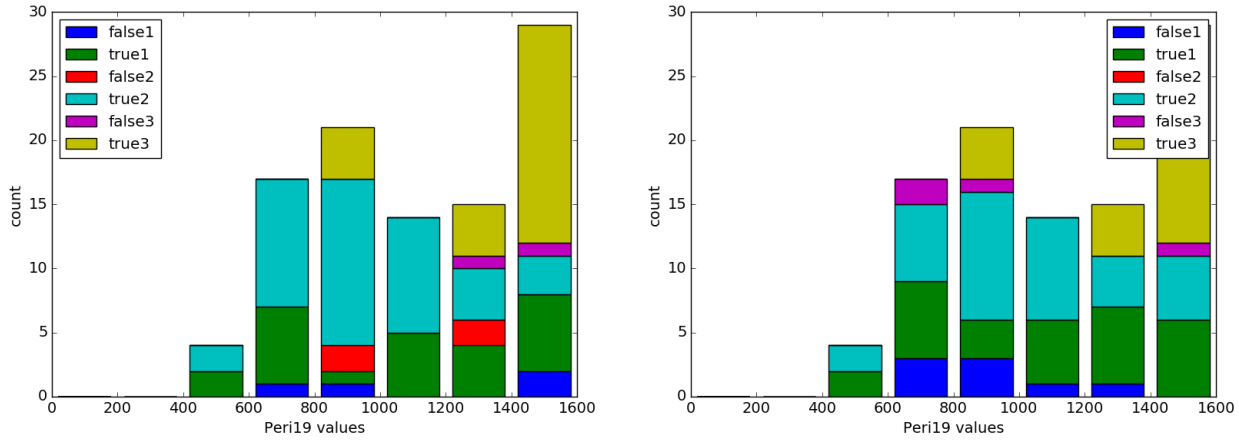
(a) Histogram of Random Forest. **(b)** Histogram of SVM.

Figure 6.10: Stacked histogram for class predictions based on the area covered by plants on June 19.



(a) Histogram of Random Forest. **(b)** Histogram of SVM.

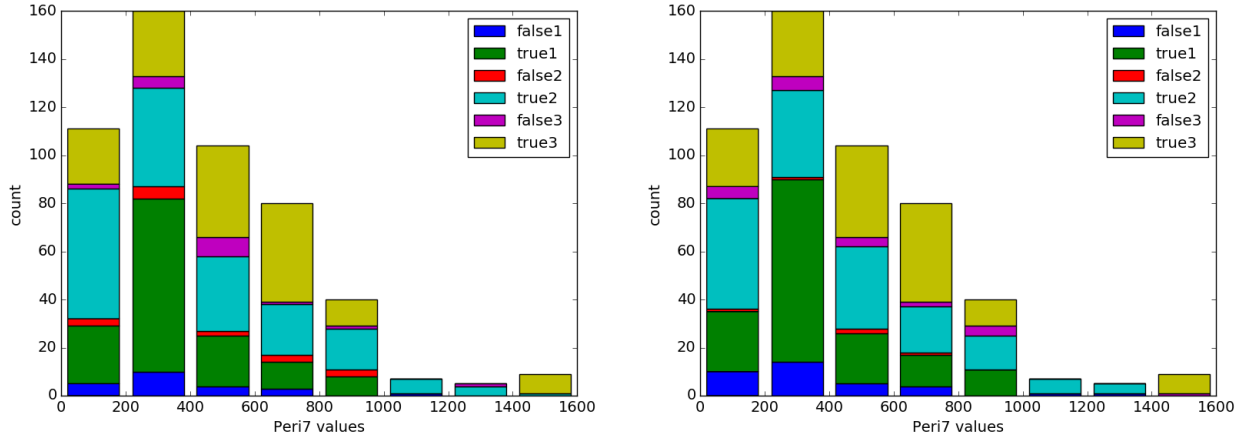
Figure 6.11: Stacked histogram for class predictions based on the area covered by plants on June 7.



(a) Histogram of Random Forest.

(b) Histogram of SVM.

Figure 6.12: Stacked histogram for class predictions based on the perimeter covered by plants on June 19.



(a) Histogram of Random Forest.

(b) Histogram of SVM.

Figure 6.13: Stacked histogram for class predictions based on the perimeter covered by plants on June 7.

7 Discussion and Future Work

One of the most important global issues the world is facing is food insecurity. A core objective of modern agriculture is to have higher food production and create seeds that can withstand future climate scenarios. With high vegetative growth or biomass comes high yield in plants and plants with high vigor have proven to generate greater yield [13] making improving vigor a less breeding activity.

The advancement of effective high-throughput phenotyping platforms (HTPPs) including sensors, drones and high-performance computing are paving the way to increase farming production. However, these technological approaches result in enormous amount of data. In such cases, machine learning tools have proven to be effective in analyzing and extracting relevant information from these large datasets.

This research work addresses the research question we had which was, "Can we build a machine learning based model to predict the reported vigor ratings from digital images of the crops?". We have answered the question by proving that the Random Forest and Support Vector Machines models were 88.97% and 88% accurate respectively in predicting the plant vigor ratings. Overall, as can be seen from their confusion matrices (Figures 6.2 and 2.3) the models performed very well for class 1, class 2 and class 3. Both the models struggled a bit to generalize the class 2 samples. The reasons behind this could be that the features of samples belonging to classes 1 and 3, indicating low vigor and high vigor respectively, are quite distinguishable helping the model to classify them better. On the contrary, the features of the class 2 data samples overlap with those of class 1 and 3 samples leading to the model's uncertainty in classifying.

The traditional method of identifying plant vigor level is done by visual examination. This process is afflicted with inefficiencies and prone to human error. For a trained computer, identifying plant vigor level is essentially pattern recognition. After sorting through thousands of photos of plants, a machine learning algorithm can estimate vigor level and may even prompt the plant breeders to identify the stronger crop varieties.

7.1 Comparison to Previous Studies

In the last years, much work have used technology-based methods to capture vigor information [2,13,21,23,66, 113] due to the benefits that come with this trait, like greater light interception, greater nitrogen absorption and high water use efficiency. However, some of these studies have certain shortfalls. De Bei et al. [23] used a few number of images to assess the vigor based on the canopy size and porosity. The image data were collected using the software application Viticanopy installed on a smartphone or tablet. Anzola et al. [2]

monitored the vigor of just ten plants in the greenhouse for twelve hours. A small sample of data prevents comprehensive analysis of plant features within a single plant and across different varieties. Comparatively, in our study we could acquire more than thousand images capturing the information from different stages of the plant growth.

Anzola et al. [2] looked for the leaf angle for each plant by deploying a sensor node for each plant. This type of data acquisition system becomes expensive and time taking in case of large breeding programs. Brandao et al. used multispectral satellite images to monitor cotton vigor. Satellite images are not easily accessible to farmers and plant breeders whereas, in our study, we have employed an easy to build and cost effective image acquisition system that can be used in large breeding programs.

Coulon Leroy et al. [21] developed a fuzzy inference system (FIS), modeling interactions between environmental factors, agricultural practices and vine growth, that assesses the vigor level of plants based on three different input variables from soil information, rootstock information and inter-row management constraint. The performance of the FIS was optimized using root mean square error (RMSE) and R-squared (R_2) values. The optimized system was evaluated on test data which resulted in the 11 samples out of the total 74 samples, matching with the observed vigor values. To the contrary, our model resulted in 455 samples (over a total of 517 samples), with matching vigor values to that of the original values.

To summarize the differences with the above-mentioned studies:

1. The data acquisition system used in this study, called the GrowPro, is affordable and can be easily built by a user with the its parts commonly available in the market.

2. Early plant vigor was digitally phenotyped on a relatively large dataset comprising of around 1400 images of the Carinata plant.

3. In this research work, the analysis was performed on the temporal image data of the plants, captured from the true leaf stage to the canopy closure stage.

4. In this research work, a detailed analysis of the machine learning based prediction models were carried out. Moreover, the performance of the models were quantified in terms of accuracy.

7.2 Contributions

This study is a substantial improvement on the previous studies in the literature. This is the first research in devising intelligent algorithms for predicting early season vigor. The combination of methods adopted in this study, extracts features from raw images and are then used to train the supervised algorithms to arrive at a phenotyping decision or trait. Also we provided a detailed analysis of the performance of the prediction models along with comparison to a baseline learning algorithm, the SVM. A successful high-throughput image-based phenotyping system is non-invasive along with being an automated data acquisition system that can collect data over different time stamps.

Our analysis is the first attempt to try reproducing the crop vigor rating given by an expert breeder when assessing early vigor in Carinata using a combination of machine learning and appropriate feature selection. A random forest was able to replicate the manual classifications with an accuracy of approximately 88.97% as established through K-fold cross-validation. Similarly, the support vector machine was at par with random forest and predicted vigor ratings with a high accuracy of 88%. The confusion matrices of both the models demonstrated that the classifiers were always able to differentiate a score of 1 (low vigor) from a score of 3 (high vigor). Misclassification always occurred on the boundaries of a score of 2 (moderate vigor). Distinguishing these values was a function of several interdependent features, as evidenced by the lack of a single key identifying feature. In cases with complex discrimination mechanics, a degree of misclassification is expected; indeed, while we have assumed that the metrics provided by the expert are correct, given the complexity of the classification problem, the 11% error rate is likely a combination of the algorithm failing to correctly classify feature vectors near the boundaries and the expert providing somewhat subjective and inaccurate ratings of Carinata on the margins of class boundaries. No research findings in the literature reported an accuracy for their estimated vigor ratings. So, it was difficult to compare the performance of the machine learning models in predicting the vigor ratings.

The work has provided three primary contributions to the field of rapid image-based phenotyping:

Automatic Vigor Rating The combination of the selected features and the machine learning based classifiers (random forest and support vector machine) were able to classify early vigor with an accuracy of around 88% compared to expert manual ratings. This fulfills our second objective of our study as stated in chapter 1. To the best of our knowledge, this is the first attempt to predict expert ratings of early vigor from images. Our accuracy sets the bar for future research in this area.

Novel Features When employing feature-based machine learning methods, it is important to specify high-value features. While our feature set comprised a number of standard features, we also employed the mean and variance of the distance from the centerline as an operationalization of the breeder heuristic "how bushy it has grown out." These features were evaluated as the most important for classification as reported by the random forest implementation we chose. This reveals that these set of features were helpful in identifying the vigor levels of the crops, thus fulfilling the first objective of our study as stated in chapter 1.

Performance Baseline The 88.97% accuracy achieved by random forest reported here is now the current state of the art for automated classification. However, our analysis of the features indicated that the human assignment of cases on the boundary between classes seemed somewhat arbitrary, prompting breeders to reconsider how they rate borderline cases.

7.3 Limitations and Future Work

The work reported here should be applicable to species closely related to Carinata such as Canola and related oilseed and feed crops; however, the features we have chosen may be less discriminatory for root crops such as carrots or potatoes, or grains such as wheat or barley. While the overall approach is still valid, it is likely changes to the feature set will be required for application to different classes of crops. The images were acquired using a novel capture system. The algorithm could be extended to other close-in imaging systems such as cart mounted [50] or manual systems [65], but would be difficult to extend to aerial systems such as drones or helicopters, unless extremely high fidelity images, approximating the resolution of manual or cart-based systems, are available.

While we have made some notable contributions to state of the art, this work is not without limitations, many of which point to future work. First, it is unclear exactly how much spatial resolution is required for accurate classification. A downsampling study would allow us to quantify the resolution of images necessary for classification, and by extension which capture systems could take advantage of the pipeline. With high resolution images, morphological traits of the leaves can be identified and used for distinguishing vigor levels of the crops. Second, we have only reported the results for two types of classification algorithms which are random forest and support vector machine; other algorithms especially deep learning techniques may provide better results. However, deep learning algorithms require large datasets to produce accurate results. Besides for the purpose of prediction, deep learning techniques may also be adapted for feature extraction. A further study of how the feature sets perform under a variety of learning algorithms could be fruitful. Finally, we have only tested on a single species over a single year. While this provides a strong proof of concept, a wider variety of crops grown in different locations under different conditions would provide a better understanding of how robust the pipeline is.

7.4 Conclusion

This research work has demonstrated that using carefully selected features it is possible to rate early season vigor in Carinata from images captured using handheld systems using a random forest and support vector machine. This system is fully automated and could provide value to breeders by providing triage of plots which are clearly of low or high vigor, allowing the breeder to focus on intermediate plots which may require human interpretation. By providing a system which can quantitatively triage work for breeders, we can increase the speed and accuracy of breeding programs, an important step in moving to a more secure food future.

References

[1] Eren Erdal Aksoy, Alexey Abramov, Florentin Wörgötter, Hanno Scharr, Andreas Fischbach, and Babette Dellen. Modeling leaf growth of rosette plants using infrared stereo image sequences. *Computers and electronics in agriculture*, 110:78–90, 2015.

[2] John Petearson Anzola, Vicente García-Díaz, and Andrés Camilo Jiménez. Wsn analysis in grid topology for potato crops for iot. In *Proceedings of the 4th Multidisciplinary International Social Networks Conference on ZZZ*, page 44. ACM, 2017.

[3] Daryl B Arnall, Brenda S Tubaña, Starr L Holtz, Kefyalew Girma, and William R Raun. Relationship between nitrogen use efficiency and response index in winter wheat. *Journal of plant nutrition*, 32(3):502–515, 2009.

[4] S Asseng, NC Turner, and Brian A Keating. Analysis of water-and nitrogen-use efficiency of wheat in a mediterranean climate. *Plant and Soil*, 233(1):127–143, 2001.

[5] Senthold Asseng, IAN Foster, and Neil C Turner. The impact of temperature variability on wheat yields. *Global Change Biology*, 17(2):997–1012, 2011.

[6] Andreas Backhaus, Asuka Kuwabara, Marion Bauch, Nick Monk, Guido Sanguinetti, and Andrew Fleming. Leafprocessor: a new leaf phenotyping tool using contour bending energy and shape cluster analysis. *New phytologist*, 187(1):251–261, 2010.

[7] Olga Barinova, Victor Lempitsky, and Pushmeet Kohli. On detection of multiple object instances using hough transforms. *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2010.

[8] Lammert Bastiaans, Aad van Ast, and DL Zhao. What is the basis of early vigour, being an important trait of weed competitiveness in rice. *PDF File. Cached at http://www. ewrs. org/pwc/pwc_Samsun/Bastiaans. pdf. Date retrieved: August*, 23:2011, 2011.

[9] Jan Behmann, Jörg Steinrücken, and Lutz Plümer. Detection of early plant stress responses in hyperspectral images. *ISPRS Journal of Photogrammetry and Remote Sensing*, 93:98–111, 2014.

[10] Bernhard Biskup, Hanno Scharr, Ulrich Schurr, and UWE Rascher. A stereo imaging system for measuring structural parameters of plant canopies. *Plant, cell & environment*, 30(10):1299–1308, 2007.

[11] Gernot Bodner, Mouhannad Alsalem, Alireza Nakhforoosh, Thomas Arnold, and Daniel Leitner. Rgb and spectral root imaging for plant phenotyping and physiological research: Experimental setup and imaging protocols. *Journal of visualized experiments: JoVE*, (126), 2017.

[12] J Bossu, Ch Gée, G Jones, and F Truchetet. Wavelet transform to discriminate between crop and weed in perspective agronomic images. *computers and electronics in agriculture*, 65(1):133–143, 2009.

[13] Ziany N Brandão, Célia R Grego, Ricardo Y Inamasu, and Lúcio A Jorge. Spectral reflectance of satellite images using geostatistics methods to estimate growth and cotton yield. In *Remote Sensing for Agriculture, Ecosystems, and Hydrology XVI*, volume 9239, page 923920. International Society for Optics and Photonics, 2014.

[14] Lucas Busemeyer, Arno Ruckelshausen, Kim Möller, Albrecht E Melchinger, Katharina V Alheit, Hans Peter Maurer, Volker Hahn, Elmar A Weissmann, Jochen C Reif, and Tobias Würschum. Precision phenotyping of biomass accumulation in triticale reveals temporal genetic patterns of regulation. *Scientific reports*, 3:2442, 2013.

[15] Max Bylesjö, Vincent Segura, Raju Y Soolanayakanahally, Anne M Rae, Johan Trygg, Petter Gustafsson, Stefan Jansson, and Nathaniel R Street. Lamina: a tool for rapid quantification of leaf size and shape parameters. *BMC plant biology*, 8(1):82, 2008.

[16] Dijun Chen, Kerstin Neumann, Swetlana Friedel, Benjamin Kilian, Ming Chen, Thomas Altmann, and Christian Klukas. Dissecting the phenotypic components of crop plant growth and drought responses based on high-throughput image analysis. *The Plant Cell*, pages tpc–114, 2014.

[17] Sruti Das Choudhury, Srinidhi Bashyam, Yumou Qiu, Ashok Samal, and Tala Awada. Holistic and component plant phenotyping using temporal image sequence. *Plant methods*, 14(1):35, 2018.

[18] Randy Clark, Robert MacCurdy, Janelle Jung, Jon Shaff, Susan Rutherford McCouch, Dan Aneshansley, and Leon Kochian. 3-dimensional root phenotyping with a novel imaging and software platform. *Plant physiology*, pages pp–110, 2011.

[19] Joshua N Cobb, Genevieve DeClerck, Anthony Greenberg, Randy Clark, and Susan McCouch. Next-generation phenotyping: requirements and strategies for enhancing our understanding of genotype–phenotype relationships and its relevance to crop improvement. *Theoretical and Applied Genetics*, 126(4):867–887, 2013.

[20] Corinna Cortes and Vladimir Vapnik. Support-vector networks. *Machine Learning*, 20(3):273297, 1995.

[21] Cécile Coulon-Leroy, Brigitte Charnomordic, Marie Thiollet-Scholtus, and Serge Guillaume. Imperfect knowledge and data-based approach to model a complex agronomic feature–application to vine vigor. *Computers and electronics in agriculture*, 99:135–145, 2013.

[22] Andrew M Cunliffe, Richard E Brazier, and Karen Anderson. Ultra-fine grain landscape-scale quantification of dryland vegetation structure with drone-acquired structure-from-motion photogrammetry. *Remote Sensing of Environment*, 183:129–143, 2016.

[23] Roberta De Bei, Sigfredo Fuentes, Matthew Gilliham, Steve Tyerman, Everard Edwards, Nicolò Bianchini, Jason Smith, and Cassandra Collins. Viticanopy: A free computer app to estimate canopy vigor and porosity for grapevine. *Sensors*, 16(4):585, 2016.

[24] Desa. *World Population Prospects: The 2010 Revision, Volume II-Demographic Profiles*. UN, 2013.

[25] T Duan, SC Chapman, E Holland, GJ Rebetzke, Y Guo, and B Zheng. Dynamic quantification of canopy structure to characterize early plant vigour in wheat genotypes. *Journal of experimental botany*, 67(15):4523–4534, 2016.

[26] Richard O Duda and Peter E Hart. Use of the hough transformation to detect lines and curves in pictures. *Communications of the ACM*, 15(1):11–15, 1972.

[27] Hsien Ming Easlon and Arnold J Bloom. Easy leaf area: Automated digital image analysis for rapid and accurate measurement of leaf area. *Applications in plant sciences*, 2(7), 2014.

[28] Maarten Elferink and Florian Schierhorn. Global demand for food is rising. can we meet it? 04 2016.

[29] Noah Fahlgren, Malia A Gehan, and Ivan Baxter. Lights, camera, action: high-throughput plant phenotyping is ready for a close-up. *Current opinion in plant biology*, 24:93–99, 2015.

[30] Alfredo Fernández-Landa, José Antonio Navarro, Sonia Condés, Nur Algeet-Abarquero, and Miguel Marchamalo. High resolution biomass mapping in tropical forests with lidar-derived digital models: Poás volcano national park (costa rica). *iForest-Biogeosciences and Forestry*, 10(1):259, 2017.

[31] P Flombaum and OE Sala. A non-destructive and rapid method to estimate biomass and aboveground net primary production in arid environments. *Journal of Arid Environments*, 69(2):352–358, 2007.

[32] G Garg, Vijander Singh, Mudita Grover, J Gupta, et al. Optimal kernel learning for eeg based sleep scoring system. *Int. J. Biol. Med. Res*, 2:1220–1225, 2011.

[33] Claude-Alain Gebhard, Lucie Büchi, Frank Liebisch, Sokrat Sinaj, Hans Ramseier, and Raphaël Charles. Beurteilung von leguminosen als gründüngungspflanzen: Stickstoff und begleitflora. *Agrarforschung Schweiz*, 4(9):384–393, 2013.

[34] Ch Gée, J Bossu, G Jones, and F Truchetet. Crop/weed discrimination in perspective agronomic images. *Computers and Electronics in Agriculture*, 60(1):49–59, 2008.

[35] B Gerard and A Buerkert. Aerial photography to determine fertiliser effects on pearl millet and guiera senegalensis growth. *Plant and Soil*, 210(2):167–178, 1999.

[36] Bruno Gérard, Andreas Buerkert, Pierre Hiernaux, and Horst Marschner. Non-destructive measurement of plant growth and nitrogen status of pearl millet with low-altitude aerial photography. In *Plant Nutrition for Sustainable Food Production and Environment*, pages 373–378. Springer, 1997.

[37] Yihong Gong, Mei Han, Wei Hua, and Wei Xu. Maximum entropy model-based baseball highlight detection and classification. *Computer Vision and Image Understanding*, 96(2):181–199, 2004.

[38] Christine Granier, Luis Aguirrezabal, Karine Chenu, Sarah Jane Cookson, Myriam Dauzat, Philippe Hamard, Jean-Jacques Thioux, Gaëlle Rolland, Sandrine Bouchier-Combaud, Anne Lebaudy, et al. Phenopsis, an automated platform for reproducible phenotyping of plant responses to soil water deficit in arabidopsis thaliana permitted the identification of an accession with low sensitivity to soil water deficit. *New Phytologist*, 169(3):623–635, 2006.

[39] José Miguel Guerrero, María Guijarro, Martin Montalvo, Juan Romeo, Luis Emmi, Angela Ribeiro, and Gonzalo Pajares. Automatic expert system based on images for accuracy crop row detection in maize fields. *Expert Systems with Applications*, 40(2):656–664, 2013.

[40] Peng HOU, Fei LI, et al. Monitoring winter wheat population dynamics using an active crop sensor. *Spectroscopy and Spectral Analysis*, 31(2):535–538, 2011.

[41] Min Huang, Ruichun Zhang, Jiana Chen, Fangbo Cao, Ligeng Jiang, and Yingbin Zou. Morphological and physiological traits of seeds and seedlings in two rice cultivars with contrasting early vigor. *Plant Production Science*, 20(1):95–101, 2017.

[42] Marcus Jansen, Francisco Pinto, Kerstin A Nagel, Dagmar van Dusschoten, Fabio Fiorani, Uwe Rascher, Heike U Schneider, Achim Walter, and Ulrich Schurr. Non-invasive phenotyping methodologies enable the accurate characterization of growth and performance of shoots and roots. In *Genomics of plant genetic resources*, pages 173–206. Springer, 2014.

[43] Kevyn J Juneau and Catherine S Tarasoff. Leaf area and water content changes after permanent and temporary storage. *PLoS One*, 7(8):e42604, 2012.

[44] Rakesh Kaundal, Amar S Kapoor, and Gajendra PS Raghava. Machine learning techniques in disease forecasting: a case study on rice blast prediction. *BMC bioinformatics*, 7(1):485, 2006.

[45] Carl Kingsford and Steven L Salzberg. What are decision trees? *Nature biotechnology*, 26(9):1011, 2008.

[46] Sebastian Kipp, Bodo Mistele, Peter Baresel, and Urs Schmidhalter. High-throughput phenotyping early plant vigour of winter wheat. *European Journal of Agronomy*, 52:271–278, 2014.

[47] Sebastian Kipp, Bodo Mistele, and Urs Schmidhalter. Identification of stay-green and early senescence phenotypes in high-yielding winter wheat, and their relationship to grain yield and grain protein concentration using high-throughput phenotyping techniques. *Functional Plant Biology*, 41(3):227–235, 2014.

[48] Arvind Kumar, Satish Verulkar, Shalabh Dixit, Bhagirath Chauhan, Jerome Bernier, Ramaiah Venuprasad, Dule Zhao, and MN Shrivastava. Yield and yield-attributing traits of rice (oryza sativa l.) under lowland drought and suitability of early vigor as a selection criterion. *Field Crops Research*, 114(1):99–107, 2009.

[49] S Kuyah, J Dietz, C Muthuri, R Jamnadass, P Mwangi, and R Neufeldt Coe. H.(2012). allometric equations for estimating biomass in agricultural landscapes: I. aboveground biomass. *Agriculture, Ecosystems and Environment*, 158:216–224.

[50] Alan M Lefcourt, Ross Kistler, S Andrew Gadsden, and Moon S Kim. Automated cart with vis/nir hyperspectral reflectance and fluorescence imaging capabilities. *Applied Sciences*, 7(1):3, 2016.

[51] Dario Leister, Claudio Varotto, Paolo Pesaresi, Alexandra Niwergall, and Francesco Salamini. Large-scale evaluation of plant growth in arabidopsis thaliana by non-invasive image analysis. *Plant Physiology and Biochemistry*, 37(9):671–678, 1999.

[52] Liuling Li, DC Nielsen, Qiang Yu, Liwang Ma, and LR Ahuja. Evaluating the crop water stress index and its correlation with latent heat and co2 fluxes over winter wheat and maize in the north china plain. *Agricultural Water Management*, 97(8):1146–1155, 2010.

[53] Qiangzi Li, Xin Cao, Kun Jia, Miao Zhang, and Qinghan Dong. Crop type identification by integration of high-spatial resolution multispectral data with features extracted from coarse-resolution time-series vegetation index data. *International journal of remote sensing*, 35(16):6076–6088, 2014.

[54] Konstantinos Liakos, Patrizia Busato, Dimitrios Moshou, Simon Pearson, and Dionysis Bochtis. Machine learning in agriculture: A review. *Sensors*, 18(8):2674, 2018.

[55] Frank Liebisch, Norbert Kirchgessner, David Schneider, Achim Walter, and Andreas Hund. Remote, aerial phenotyping of maize traits with a mobile multi-sensor approach. *Plant methods*, 11(1):9, 2015.

[56] Kevin S Lim and Paul M Treitz. Estimation of above ground forest biomass from airborne discrete return laser scanner data using canopy-based quantile estimators. *Scandinavian Journal of Forest Research*, 19(6):558–570, 2004.

[57] Hsuan-Tien Lin and Chih-Jen Lin. A study on sigmoid kernels for svm and the training of non-psd kernels by smo-type methods. *submitted to Neural Computation*, 3:1–32, 2003.

[58] Fulco Ludwig and Senthold Asseng. Potential benefits of early vigor and changes in phenology in wheat to adapt to warmer and drier climates. *Agricultural Systems*, 103(3):127–136, 2010.

[59] Jing Lv, Yuanyuan Shang, and Hui Ding. Road obstacle detection based on randomized hough transform. *Proceedings of the 2nd International Conference on Teaching and Computational Science*, 2014.

[60] Richard Makanza, Mainassara Zaman-Allah, Jill E Cairns, Cosmos Magorokosho, Amsal Tarekegne, Mike Olsen, and Boddupalli M Prasanna. High-throughput phenotyping of canopy cover and senescence in maize field trials using aerial digital canopy imaging. *Remote Sensing*, 10(2):330, 2018.

[61] M Montalvo, Gonzalo Pajares, José Miguel Guerrero, J Romeo, María Guijarro, Angela Ribeiro, José Jaime Ruz, and JM Cruz. Automatic detection of crop rows in maize fields with high weeds pressure. *Expert Systems with Applications*, 39(15):11889–11897, 2012.

[62] Alexander Mordvintsev and K Abid. Opencv-python tutorials documentation.

[63] David J Mulla. Twenty five years of remote sensing in precision agriculture: Key advances and remaining knowledge gaps. *Biosystems engineering*, 114(4):358–371, 2013.

[64] Wiebe Nijland, Rogier de Jong, Steven M de Jong, Michael A Wulder, Chris W Bater, and Nicholas C Coops. Monitoring plant condition and phenology using infrared sensitive consumer grade digital cameras. *Agricultural and Forest Meteorology*, 184:98–106, 2014.

[65] H Noh, Q Zhang, B Shin, S Han, and L Feng. A neural network model of maize crop nitrogen stress assessment for a multi-spectral imaging sensor. *Biosystems Engineering*, 94(4):477–485, 2006.

[66] Paula Cristina Oliveira, João Paulo Moura, Luís Filipe Fernandes, Elza Maria Amaral, and Ana Alexandra Oliveira. A non-destructive method based on digital image processing for calculate the vigor and the vegetative expression of vines. *Computers and Electronics in Agriculture*, 124:289–294, 2016.

[67] Francesco Orsini, Matilde Paino D'urzo, Gunsu Inan, Sara Serra, Dong-Ha Oh, Michael V Mickelbart, Federica Consiglio, Xia Li, Jae Cheol Jeong, Dae-Jin Yun, et al. A comparative study of salt tolerance parameters in 11 wild relatives of arabidopsis thaliana. *Journal of Experimental Botany*, 61(13):3787–3798, 2010.

[68] Ilya Vladimirovich Ovsyannikov. Growpro: a flexible and high-resolution imaging tool for high-throughput plant phenotyping in the field. Master's thesis, Ilya Vladimirovich Ovsyannikov, 6 2018.

[69] Xanthoula Eirini Pantazi, Dimitrios Moshou, Thomas Alexandridis, RL Whetton, and Abdul Mounem Mouazen. Wheat yield prediction using machine learning and advanced sensing techniques. *Computers and Electronics in Agriculture*, 121:57–65, 2016.

[70] Stefan Paulus, Jan Behmann, Anne-Katrin Mahlein, Lutz Plümer, and Heiner Kuhlmann. Low-cost 3d systems: suitable tools for plant phenotyping. *Sensors*, 14(2):3001–3018, 2014.

[71] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

[72] José M Peña, Jorge Torres-Sánchez, Angélica Serrano-Pérez, Ana I de Castro, and Francisca López-Granados. Quantifying efficacy and limits of unmanned aerial vehicle (uav) technology for weed seedling detection as affected by sensor resolution. *Sensors*, 15(3):5609–5626, 2015.

[73] Nathalie Pettorelli, Sadie Ryan, Thomas Mueller, Nils Bunnefeld, Bogumila Jedrzejewska, Mauricio Lima, and Kyrre Kausrud. The normalized difference vegetation index (ndvi): unforeseen successes in animal ecology. *Climate Research*, 46(1):15–27, 2011.

[74] Vili Podgorelec, Peter Kokol, Bruno Stiglic, and Ivan Rozman. Decision trees: an overview and their use in medicine. *Journal of medical systems*, 26(5):445–463, 2002.

[75] A. Walker R. Fisher, S. Perkins and E. Wolfart. Hough transform. `https://homepages.inf.ed.ac.uk/rbf/HIPR2/hough.htm`, 2003.

[76] Laura Elena Raileanu and Kilian Stoffel. Theoretical comparison between the gini index and information gain criteria. *Annals of Mathematics and Artificial Intelligence*, 41(1):77–93, 2004.

[77] Sajith Rajapaksa, Mark Eramian, Hema Duddu, Menglu Wang, Steve Shirtliffe, Seungbum Ryu, Anique Josuttes, Ti Zhang, Sally Vail, Curtis Pozniak, et al. Classification of crop lodging with gray level co-occurrence matrix. In *2018 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pages 251–258. IEEE, 2018.

[78] Karthika Rajendran, Mark Tester, and Stuart J Roy. Quantifying the three main components of salinity tolerance in cereals. *Plant, cell & environment*, 32(3):237–249, 2009.

[79] Emad Ramadan, Sadiq Alinsaif, and Md Rafiul Hassan. Network topology measures for identifying disease-gene association in breast cancer. *BMC bioinformatics*, 17(7):274, 2016.

[80] Lisandro Rambo, Bao-Luo Ma, Youcai Xiong, and Paulo Regis Ferreira da Silvia. Leaf and canopy optical characteristics as crop-n-status indicators for field nitrogen management in corn. *Journal of Plant Nutrition and Soil Science*, 173(3):434–443, 2010.

[81] RA Richards. Crop improvement for temperate australia: future opportunities. *Field Crops Research*, 26(2):141–169, 1991.

[82] RA Richards. Defining selection criteria to improve yield under drought. *Plant growth regulation*, 20(2):157–166, 1996.

[83] J Romeo, G Pajares, M Montalvo, JM Guerrero, M Guijarro, and A Ribeiro. Crop row detection in maize fields inspired on the human visual perception. *The Scientific World Journal*, 2012, 2012.

[84] Youssef Rouphael, Lukáš Spíchal, Klára Panzarová, Raffaele Casa, and Giuseppe Colla. High-throughput plant phenotyping for developing novel biostimulants: From lab to field or from field to lab? *Frontiers in plant science*, 9, 2018.

[85] Nadir Sainz-Costa, Angela Ribeiro, Xavier P Burgos-Artizzu, María Guijarro, and Gonzalo Pajares. Mapping wide row crops with video sequences acquired from a tractor moving at treatment speed. *Sensors*, 11(7):7095–7109, 2011.

[86] Caroline A Schneider, Wayne S Rasband, and Kevin W Eliceiri. Nih image to imagej: 25 years of image analysis. *Nature methods*, 9(7):671, 2012.

[87] IM Scotford and PCH Miller. Estimating tiller density and leaf area index of winter wheat using spectral reflectance and ultrasonic sensing techniques. *Biosystems engineering*, 89(4):395–408, 2004.

[88] Tri Setiyono, Andrew Nelson, and Francesco Holecz. Remote sensing-based crop yield monitoring and forecasting. *Crop monitoring for improved food security*, 2014.

[89] Sushma Sher-Kaul, Beat Oertli, Emmanuel Castella, and Jean-Bernard Lachavanne. Relationship between biomass and surface area of six submerged aquatic plant species. *Aquatic Botany*, 51(1-2):147–154, 1995.

[90] KHM Siddique, D Tennant, MW Perry, and RK Belford. Water use and water use efficiency of old and modern wheat cultivars in a mediterranean-type environment. *Australian Journal of Agricultural Research*, 41(3):431–447, 1990.

[91] Arti Singh, Baskar Ganapathysubramanian, Asheesh Kumar Singh, and Soumik Sarkar. Machine learning for high-throughput stress phenotyping in plants. *Trends in plant science*, 21(2):110–124, 2016.

[92] David Sloane. *Early vigour: its role in enhancing the productivity of wheat grown in South Australia*. PhD thesis, 1999.

[93] DHG Sloane, GS Gill, and GK McDonald. The impact of agronomic manipulation of early vigour in wheat on growth and yield in south australia. *Australian Journal of Agricultural Research*, 55(6):645–654, 2004.

[94] Hazel K Smith, Graham JJ Clarkson, Gail Taylor, Andrew J Thompson, John Clarkson, Nasir M Rajpoot, et al. Automatic detection of regions in spinach canopies responding to soil moisture deficit using combined visible and thermal imagery. *PloS one*, 9(6):e97612, 2014.

[95] Henning Tangen Søgaard and Hans Jørgen Olsen. Determination of crop rows by image analysis without segmentation. *Computers and electronics in agriculture*, 38(2):141–158, 2003.

[96] Kamel Soudani, G Hmimina, Nicolas Delpierre, J-Y Pontailler, Marc Aubinet, Damien Bonal, B Caquet, A De Grandcourt, Benoît Burban, Christophe Flechard, et al. Ground-based network of ndvi measurements for tracking temporal dynamics of canopy structure and vegetation phenology in different biomes. *Remote sensing of environment*, 123:234–245, 2012.

[97] G Stevens, Allen Wrather, Matthew Rhine, E Vories, and D Dunn. Predicting rice yield response to midseason nitrogen with plant area measurements. *Agronomy journal*, 100(2):387–392, 2008.

[98] Jeremy Storer. Automated detection for baseball batter analysis using image processing and segmentation methods.

[99] Oliver Tackenberg. A new method for non-destructive measurement of biomass, growth rates, vertical biomass distribution and dry matter content based on digital image analysis. *Annals of botany*, 99(4):777–783, 2007.

[100] Zelealem T Temtime, SV Chinyoka, and JPW Shunda. A decision tree approach for integrating small business assistance schemes. *Journal of Management Development*, 23(6):563–578, 2004.

[101] Jorge Torres-Sánchez, Francisca López-Granados, Ana Isabel De Castro, and José Manuel Peña-Barragán. Configuration and specifications of an unmanned aerial vehicle (uav) for early site specific weed management. *PloS one*, 8(3):e58210, 2013.

[102] Jorge Torres-Sánchez, José Manuel Peña, Ana Isabel de Castro, and Fransisca López-Granados. Multi-temporal mapping of the vegetation fraction in early-season wheat fields using images from uav. *Computers and Electronics in Agriculture*, 103:104–113, 2014.

[103] Compton J Tucker. Red and photographic infrared linear combinations for monitoring vegetation. *Remote sensing of Environment*, 8(2):127–150, 1979.

[104] Varun Varma and Anand M Osuri. Black spot: a platform for automated and rapid estimation of leaf area from scanned images. *Plant ecology*, 214(12):1529–1534, 2013.

[105] Achim Walter, Frank Liebisch, and Andreas Hund. Plant phenotyping: from bean weighing to image analysis. *Plant methods*, 11(1):14, 2015.

[106] Achim Walter, Hanno Scharr, Frank Gilmer, Rainer Zierer, Kerstin A Nagel, Michaela Ernst, Anika Wiese, Olivia Virnich, Maja M Christ, Beate Uhlig, et al. Dynamics of seedling growth acclimation towards altered light conditions can be quantified via growscreen: a setup and procedure designed for rapid optical phenotyping of different plant species. *New Phytologist*, 174(2):447–455, 2007.

[107] Haifeng Wang, Bichen Zheng, Sang Won Yoon, and Hoo Sang Ko. A support vector machine-based ensemble algorithm for breast cancer diagnosis. *European Journal of Operational Research*, 267(2):687–699, 2018.

[108] BR Whan, GP Carlton, and WK Anderson. Potential for increasing early vigour and total biomass in spring wheat. i. identification of genetic improvements. *Australian Journal of Agricultural Research*, 42(3):347–361, 1991.

[109] BM Whelan, AB McBratney, and B Minasny. Vesper 1.5–spatial prediction software for precision agriculture. In *Precision Agriculture, Proc. 6th Int. Conf. on Precision Agriculture, ASA/CSSA/SSSA, Madison, WI, USA*, volume 179, 2002.

[110] Wikipedia contributors. Lineline intersection — Wikipedia, the free encyclopedia, 2018. [Online; accessed 23-February-2019].

[111] Lou Xiongwei, Huang Decai, Fang Luming, and Xu Aijun. Study of kernel selection and construction kernel for svm classification. *Sensors & Transducers*, 157(10):374, 2013.

[112] So Yahata, Tetsu Onishi, Kanta Yamaguchi, Seiichi Ozawa, Jun Kitazono, Takenao Ohkawa, Takeshi Yoshida, Noriyuki Murakami, and Hiroyuki Tsuji. A hybrid machine learning approach to automatic plant phenotyping for smart agriculture. In *Neural Networks (IJCNN), 2017 International Joint Conference on*, pages 1787–1793. IEEE, 2017.

[113] Kyosuke Yamamoto, Takashi Togami, and Norio Yamaguchi. Super-resolution of plant disease images for the acceleration of image-based phenotyping and vigor diagnosis in agriculture. *Sensors*, 17(11):2557, 2017.

[114] Chunhua Zhang and John M Kovacs. The application of small unmanned aerial systems for precision agriculture: a review. *Precision agriculture*, 13(6):693–712, 2012.

[115] TY Zhang and Ching Y Suen. A fast parallel algorithm for thinning digital patterns. *Communications of the ACM*, 27(3):236–239, 1984.

Appendix A

Permission

This thesis reuses Figures 4.1, 4.2, and 4.3 from "[71] Ilya Vladimirovich Ovsyannikov. Growpro: a flexible and high-resolution imaging tool for highthroughput plant phenotyping in the field. Master's thesis, Ilya Vladimirovich Ovsyannikov, 6 2018." The author of the thesis confirmed that the materials could be reused. The email is attached: