



UNIVERSITÄT ZU LÜBECK  
INSTITUTE FOR ROBOTICS  
AND COGNITIVE SYSTEMS

# Machine Learning for plant classification based on chlorophyll detection

*Maschinelles Lernen zur Klassifizierung von Pflanzen basierend  
auf Chlorophylldetektion*

**Bachelor thesis**

Within the scope of the study programme  
**Informatics**  
of the Universität zu Lübeck

Submitted by  
**Alexander Walter**

Issued and maintained by  
**Prof. Dr. Elmar Rückert**

With support of  
**M.Sc. Nils Rottmann**

June 17, 2019



## Abstract

Based on the intention to build an autonomous lawn mower robot, this work examines the viability of a sensor and microprocessor for onboard plant classification using machine learning. Usually, some sort of fencing is required to keep the robot in its intended processing area, so such a sensor would allow the robot to differentiate between grass and e.g. flowers. Also, the drive and blade speed can be adjusted for certain species or plant densities, etc.

For this, a data set was collected utilizing a specific method called chlorophyll fluorescence induction. A series of narrowband LEDs are used to drive this process, while a spectrometer measures the spectral intensity. The plant will fluoresce in specific wavelengths when sufficiently illuminated. With this data, machine learning algorithms are trained to explore if they are capable to classify these plants without further information.

With accuracies up to 98% for three plants commonly present on a lawn and up to 86% for eight plants, the results show that chlorophyll fluorescence is a viable method for classification, even under sunlight using the random forest machine learning algorithm.

## Abstrakt

Basierend auf der Absicht einen autonomen Rasenmäroboter zu konstruieren, wird in dieser Arbeit die Umsetzbarkeit eines Sensors und Mikroprozessors zur Pflanzenklassifikation untersucht. Für gewöhnlich wird eine Form von Abzäunung benötigt, damit der Roboter im beabsichtigtem Bearbeitungsareal bleibt. So ein Sensor würde es dem Roboter erlauben zwischen Gras und z.B. Blumen zu unterscheiden. Außerdem kann die Fahr- und Schnittgeschwindigkeit der Messerbalken an die Pflanzenspezies, -dichte usw. angepasst werden.

Deswegen wurde ein Datensatz unter der Verwendung von Chlorophyllfluoreszenzinduktion aufgenommen. Eine Reihe von Schmalband-LEDs führen diesen Prozess herbei, während ein Spektrometer die spektrale Intensität misst. Ausreichend beleuchtet fluoresziert die Pflanze in speziellen Wellenlängen. Mit diesen Daten werden Algorithmen des maschinellen Lernens trainiert um zu untersuchen ob sie dazu in der Lage sind diese Pflanzen ohne weitere Informationen zu klassifizieren.

Mit Genauigkeitsgraden bis zu 98% für drei gewöhnliche Pflanzen präsent auf einem Rasen und bis zu 86% für acht Pflanzen, zeigen die Ergebnisse das Chlorophyllfluoreszenz zur Klassifizierung realisierbar ist, selbst unter Sonnenlicht.



# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	Related Work . . . . .	1
1.2	Contribution and Organization . . . . .	4
<b>2</b>	<b>Basics</b>	<b>5</b>
2.1	Photosynthesis . . . . .	5
2.2	Chlorophyll fluorescence . . . . .	7
2.3	Spectrometer . . . . .	9
2.4	Phototransistor . . . . .	11
<b>3</b>	<b>Machine Learning</b>	<b>13</b>
3.1	Machine Learning Techniques . . . . .	13
3.1.1	Bias and Variance . . . . .	14
3.1.2	Cross-Validation . . . . .	14
3.2	Descision Tree . . . . .	15
3.3	Random Forrest . . . . .	17
3.4	Neural Network . . . . .	18
3.4.1	Activation Function . . . . .	19
3.4.2	Cross-Entropy Loss . . . . .	20
3.4.3	Backpropagation . . . . .	20
<b>4</b>	<b>Experiment Setup and Measurement</b>	<b>23</b>
<b>5</b>	<b>Classification Results</b>	<b>27</b>
5.1	Selecting LED . . . . .	27
5.2	Finding best Phototransistors . . . . .	27
5.3	Final Parameter Setup . . . . .	31
<b>6</b>	<b>Conclusion</b>	<b>33</b>
	Bibliography . . . . .	38

# Chapter 1

## Introduction

Automatically classifying plants or evaluating their condition is of great interest in many fields such as environmental protection, medicine and agriculture (e.g. detection of disease and weed for removal, which is crucial for effective harvesting). Therefore, it has to be determined which methods of data collection, feature extraction, and evaluation are of practical use.

The goal in this work is to explore the possibility of a plant classification sensor for a lawn mower robot and the eventual development of this sensor. Under these circumstances, specifically the ability to differentiate between grass, dandelion and moss, which are frequently found together. Such a sensor should consist of standard low-cost electronics, such as narrowband phototransistors and LEDs, as it is to become a consumer product, eventually. Also, only limited computational power is available, hence the technique for classification should be simple and fast. With such a sensor, the robot can adjust its driving speed corresponding to the plant, stop or turn around to stay in its designated area and differentiate between species to be cut and those (e.g. flowers) that are not to be damaged. By creating a map [Fig. 1.1], it can also calculate optimal driving paths.

### 1.1 Related Work

Most related work approaches use visual properties such as texture-based and morphological features (e.g. leaf shape, size, aspect ratio) as criteria for classification with machine vision algorithms.

These features are usually pre-defined and chosen arbitrarily or determined by botanists and cannot always be automatically extracted. Using principal component analysis (PCA) for automatic feature separation, accuracies greater than 90% can be achieved [Wu et al., 2007, Elhariri et al., 2014],



Figure 1.1: Diagram showing a possible map created by the lawn mower robot using a plant classification sensor.

but demand certain prerequisites such as leaves not overlapping each other or not changing their color due to environmental influences like the weather, otherwise making them either difficult or impossible to classify.

This can possibly be augmented with efficient image segmentation techniques to remove cluttered background, but in turn requires certain conditions such as uniform illumination.

For these reasons, either standardized images from existing, publicly available data sets are used or they were taken inconsistent and by the groups themselves, like single leaves placed on a white background sheet. That way, precisions above 96% can be accomplished [Saleem et al., 2019].

Another procedure relies on the recording of reflectance spectra. Overlapping leaves or clutter barely interfere with spectral features, but ambient light, plant age, growth condition, and health may affect the results and complicate classification. Moreover, a large number of species proved to be difficult to distinguish [Moshou et al., 2001], so the preferable application is to compare pairs or single species statuses.

Methods to receive data based on electromagnetic wavelengths have been implemented using active and passive broad- or narrowband spectrometer, fluorometer and combinations of both. They have been used effectively to detect diseases, stress or metabolic deficiencies [Kusnierek and Korsæth, 2015, Peteinatos et al., 2016]. For analysis and feature extraction, multivariate statistical techniques and PCA are employed, but classification accuracy varies extensively depending on chosen plant species, data sets, evaluation methods, and their combinations.

## 1.1. RELATED WORK

An alternative method is to use chlorophyll *a* fluorescence (CF), which is specific to plants and emitted by basically all photosynthetically active parts of the plant, including the stem and the abaxial side of leaves. Thereby, treating the leaf with a series of different colors and intensities induces the dissipation of a very small amount of excess light energy (about 0.5-10% of absorbed light) emitted by chlorophyll molecules (more specifically Photosystem II) [Atherton et al., 2014].

Originally, it was recommended to keep the plant in the dark for two hours before placing it under actinic light. Then the changes of CF intensity are monitored as a function of time, giving a fluorescence induction curve that is similar in all plants, reaching its maximum after 200-300 ms following the exposure to light [Johnson and Maxwell, 2000, Rosenqvist and Baker, 2004]. This fluorescence fingerprinting has shown to be species-specific and stable enough to classify plants. Still, the results were of diverse accuracy depending on plant species, genetic uniformity, and signal length, but it was found that even reducing the dark period to one second or having the plant under naturally variable ambient illumination had no significant impact. Neither had varying weather conditions, showing that this method is based on fairly stable characteristics.

CF has also been used to measure stress levels in plants or to detect a destructive disease in citrus plants [Cen et al., 2017] through machine learning.

Classification using CF is invariable to previously mentioned problems like cluttered backgrounds, leaves overlapping each other or changing ambient light. It provides a high number of structural photosynthetic characteristics, making it difficult to determine which are connected to their calculable foundations and suitable for interpretation.

The underlying biophysical and physiological processes and their connection to the various collectible parameters are inherently complicated and not fully understood [Banks, 2017]. Still, computational pattern recognition was already able to yield promising results, showing potential to further study this concept.

Self-evidently, using multiple procedures at once for data collection can be beneficial and successful [Simko et al., 2015], but will increase complexity and requires more equipment and space.

## 1.2 Contribution and Organization

In this work, the focus is to find an effective but minimal setup of one chromatic LED and different narrow bandwidth phototransistors to achieve real-time plant detection with the help of machine learning algorithms runnable on a microprocessor. Thereby, different learning techniques are evaluated in regard to efficiency and complexity, especially those that allow for quick evaluation on a trained model (e.g. decision trees).

In chapter 2, basic principles used throughout this thesis are described, most notably the functionality behind chlorophyll *a* fluorescence, which is the main component probed for viability in plant classification. Chapter 3 introduces the general concepts, benefits and drawbacks of the machine learning algorithms chosen for evaluation. The following chapter explains the experimental setup and reasoning behind the method used for data collection. Also, first examples are interpreted. The 5th chapter further discusses the results, followed by the conclusion.

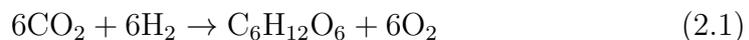
# Chapter 2

## Basics

In this chapter, the underlying principles this thesis is based on are introduced. First, the general biochemical functionalities of photosynthesis are addressed, followed by the key mechanism used in this thesis, the fluorescence of chlorophyll *a* and *b*. The third and fourth points are the techniques applied in modern optical spectrometers and the principal concept of phototransistors.

### 2.1 Photosynthesis

Most green plants and certain bacteria possess chlorophyll, a molecule essential for the process called photosynthesis. These organisms (called photoautotrophs) are able to absorb energy from light with chlorophyll pigments. The energy is then stored in carbohydrates (e.g. sugars) synthesized from carbon dioxide and water, while creating oxygen as a waste product. The general equation can be written as follows:



As chlorophyll primarily absorbs blue and red, it reflects the remaining wavelengths and thus is perceived as green [Fig. 2.1].

One chlorophyll molecule absorbs one photon and in turn transfers a single electron down an electron transport chain. This ultimately leads to the reduction of nicotinamide adenine dinucleotide phosphate (NADP) to NADPH, a cofactor vital for all cellular life [Spaans et al., 2015]. The lost electron is regained when water is split in a process called photolysis, releasing dioxygen.

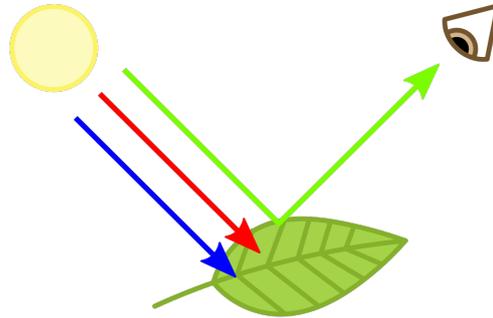
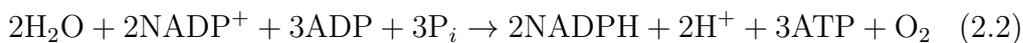


Figure 2.1: Due to chlorophyll strongly absorbing the red and blue wavelength, plants are perceived as green.<sup>1</sup>

The primary photochemical process takes place in units of structural protein complexes, the photosystem. Two of these can be differentiated: Photosystem I (PSI) and Photosystem II (PSII). Both are required for photosynthesis, but the types of chlorophyll used in each photosystem have distinct functions and spectral properties.

The primary pigment and reaction center in PSII is chlorophyll *a*, also called P680. The number represents the wavelength (680 nm) of its corresponding absorption maximum in the light spectrum [Fig. 2.2].

After giving away an electron, the oxidized chlorophyll *a* requires an external source of electrons. In green plants, two water molecules are oxidized by several successive reactions in PSII, yielding four hydrogen ions and a diatomic oxygen molecule. The gained electrons are then used to reduce the oxidized P680, resetting its ability to absorb another photon. The overall equation for this non-cyclic reaction, where the electron does not return to PSII, is [Raven et al., 2005]:



Photosystem I (PSI) works in series with PSII and usually reduces its charged reaction center P700 by electrons coming from PSII. It can also be cyclic, as the displaced electron from the photosystem ultimately returns to PSI, which is not possible in PSII. This will only generate ATP and no NADPH.

<sup>1</sup>(a) Licence: CC BY-SA 3.0, available at [https://upload.wikimedia.org/wikipedia/commons/thumb/a/a3/Why\\_are\\_plants\\_green.svg/1280px-Why\\_are\\_plants\\_green.png](https://upload.wikimedia.org/wikipedia/commons/thumb/a/a3/Why_are_plants_green.svg/1280px-Why_are_plants_green.png)

## 2.2. CHLOROPHYLL FLUORESCENCE

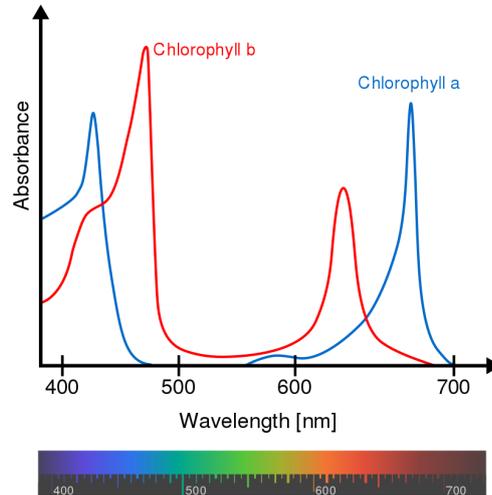


Figure 2.2: The absorbance spectrum of chlorophyll *a* and chlorophyll *b*.<sup>2</sup>

## 2.2 Chlorophyll fluorescence

The energy of an excited molecule of chlorophyll can be used in three different ways. It can be passed down to the photochemical reaction centers (PSI and PSII) and drive the photosynthesis, as mentioned above or it can be used in two processes necessary for the plant's protection. As light intensity increases, its absorbance increases as well, but the capacity for photosynthesis is limited and will eventually get saturated [Fig. 2.3].

An excited chlorophyll is in a singlet state, where all electrons are quantummechanically paired. Their overall angular momentum is zero, meaning  $s = 0$ , which is the net spin quantum number. This state has a chance to form long-lived chlorophyll triplet states (two electrons are unpaired,  $s = 1$ ), which in turn can produce singlet oxygen that may damage pigments, proteins, and other molecules through oxidation.

Therefore, the excited state can return to the ground state by emitting energy as heat (called non-photochemical quenching) or by emitting a photon (called fluorescence) in the far-red spectrum. These three processes are competing with one another, which means a change in the efficiency of one will

---

<sup>2</sup>(a) Licence: CC BY-SA 3.0, available at [https://upload.wikimedia.org/wikipedia/commons/thumb/2/23/Chlorophyll1\\_ab\\_spectra-en.svg/1024px-Chlorophyll1\\_ab\\_spectra-en.svg.png](https://upload.wikimedia.org/wikipedia/commons/thumb/2/23/Chlorophyll1_ab_spectra-en.svg/1024px-Chlorophyll1_ab_spectra-en.svg.png)

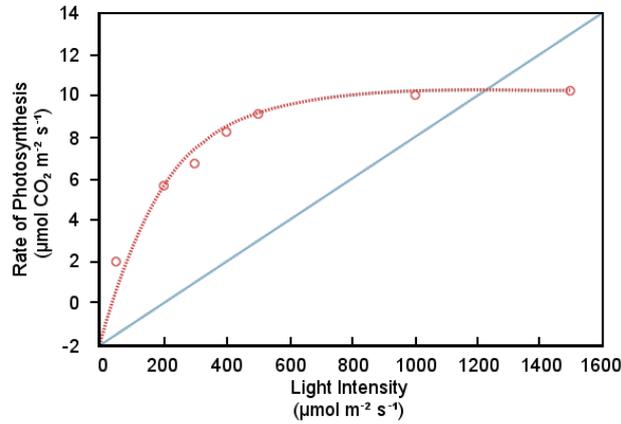


Figure 2.3: The plants ability to assimilate carbon saturates (red), while its absorption of light (blue) increases linearly.<sup>3</sup>

affect the efficiency of the other two.

This changing yield in fluorescence was first observed by [KAUTSKY et al., 1960] and is called the Kautsky Effect. A major contributor to the fluorescence is PSII through its high ratio of chlorophyll *a* (P680), while the impact of PSI is low and comparably constant. This also means that we expect a peak around 680 nanometers, as these were the most absorbed and then emitted by chlorophyll *a* [Fig. 2.4].

The most common application is to use flashes of light and measure fluorescence intensity over time. The values of the initial fluorescence minima  $F_0$  and maxima  $F_M$  following an increase in the first 1-2 seconds are used to calculate PSII efficiency, the impact of photochemical quenching and several other parameters. Plotting the intensity over time gives a fluorescence induction curve that can hold additional information [Kalaji et al., 2017].

Having to measure several seconds using flashes of light on the same spot is barely a decent option for a moving robot. Thus, we induce chlorophyll *a* fluorescence with a short exposure to a LED light source and measure the emitted light. As can be seen in Fig. 2.5, this is enough to receive a proper answer from the plant.

<sup>3</sup>(a) Licence: CC BY-SA 3.0, available at [https://upload.wikimedia.org/wikipedia/commons/7/74/Photosynthetic\\_parameters\\_of\\_plants.png](https://upload.wikimedia.org/wikipedia/commons/7/74/Photosynthetic_parameters_of_plants.png)

### 2.3. SPECTROMETER

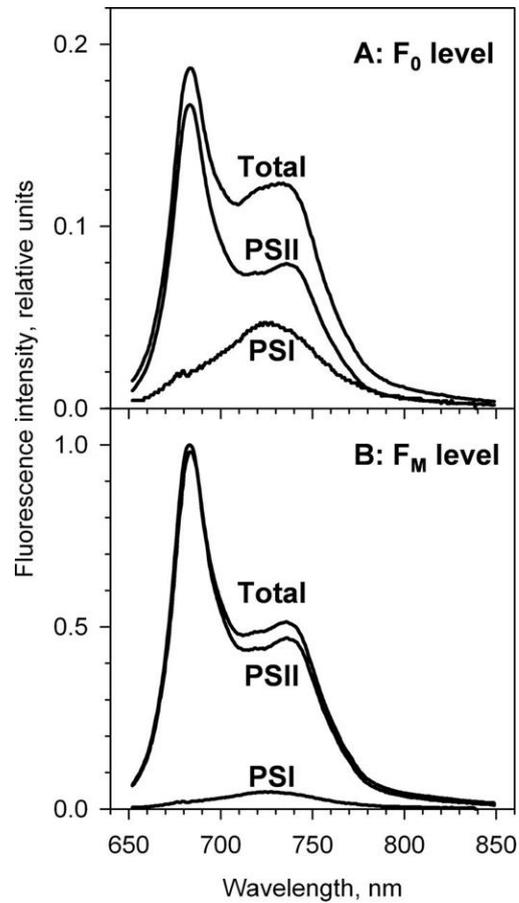


Figure 2.4: The relative spectral contributions of PSI and PSII fluorescence to the total emission spectrum in barley leaves at the  $F_0$  (A) and  $F_M$  (B) levels.<sup>4</sup>

## 2.3 Spectrometer

Optical spectrometers are instruments to separate the wavelengths or frequencies of light and measure its intensity. This is either done by refraction using a prism or diffraction, which is the method modern spectrometers apply.

These instruments usually measure the properties of certain portions of the electromagnetic spectrum, which includes non-visible wavelengths such

<sup>4</sup>Source: [Atherton et al., 2014]. Redrawn after Franck F, Juneau P, Popovic R. 2002. Resolution of the photosystem I and photosystem II contributions to chlorophyll fluorescence of intact leaves at room temperature. *Biochimica et Biophysica Acta* 1556, 239–246. Copyright Elsevier 2002.

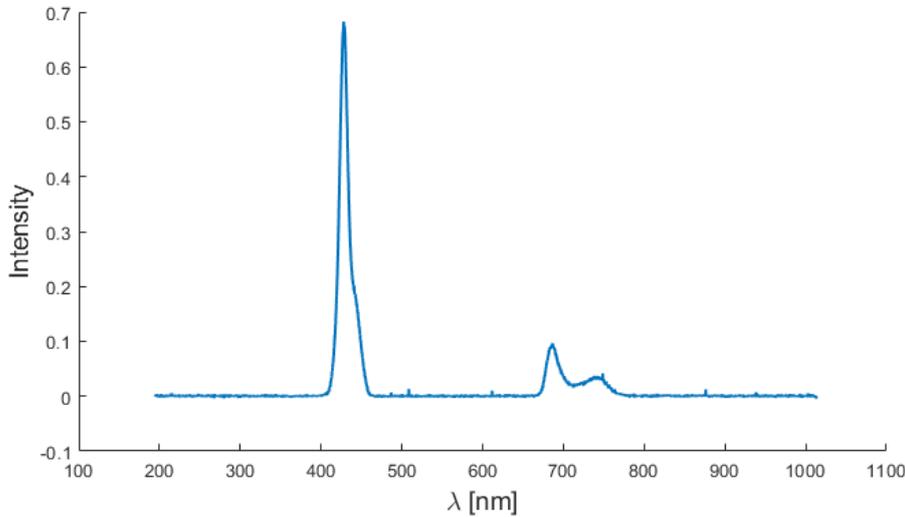


Figure 2.5: An example scan from dandelion scan 256, using the 5th LED without ambient light. The first spike at roughly 428 nm is the reflected light from the LED itself. The following hill matches with the expected fluorescence, having its maximum at around 685 nm.

as gamma rays, X-rays and infrared. Thus, spectrometers find use in a wide variety of fields.

For example, it allows astronomers to determine the chemical composition of stars or the atmospheres of planets. Atoms or molecules absorb specific wavelengths, leaving a unique fingerprint in the reflected light. Most notably, this method is completely passive, as long as the object emits enough light. Otherwise, a light source can be used, such as LEDs or lasers of specific wavelengths and electromagnetic range [Fig. 2.6].

The standard technique used is diffraction grating, applying the Huygens-Fresnel principle. It states that every point on a wavefront acts as a new point-source. The grating of the spectrometer is a repeating pattern of slits or a mirror with grooves. Each groove then acts as a new point-source from which the light propagates. The path length to each groove varies, changing the phases of the emanating waves [Fig. 2.7]. Through additive and destructive interference, the resulting diffracted light is composed of peaks and valleys.

With the light coming in at an angle  $\Theta_i$  and the diffracted light going out at angle  $\Theta_m$ , the optical path length can be calculated.

$$d(\sin \Theta_i + \sin \Theta_m) = m\lambda; m \in \mathbb{Z}$$

## 2.4. PHOTOTRANSISTOR

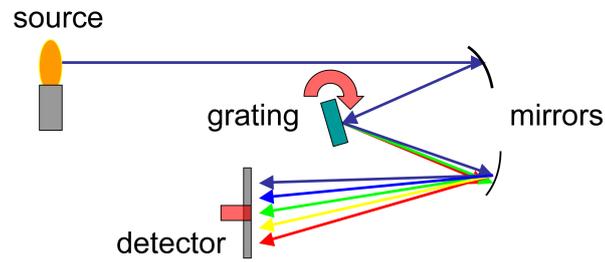


Figure 2.6: The general schematics of a spectrometer applying grating.<sup>5</sup>

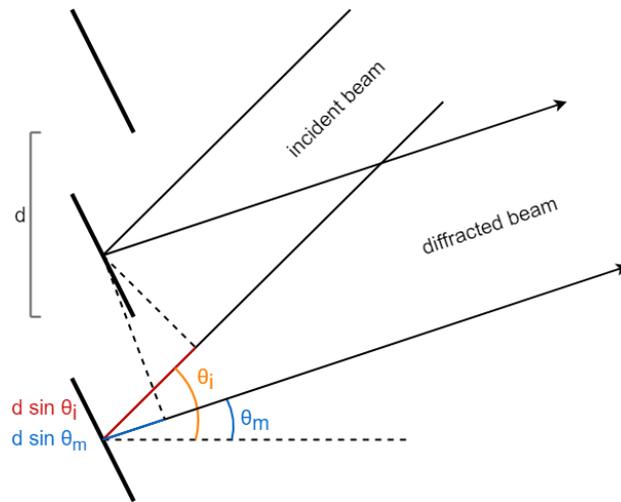


Figure 2.7: Diffraction grating diagram with spacing  $d$ , angle  $\theta_i$  for the incident beam and angle  $\theta_m$  for the diffracted beam.

If the path from adjacent grooves is  $\frac{\lambda}{2}$ , the waves cancel each other out. Comparably, at a path difference of  $\lambda$ , the waves are in phase and create maxima. Arrays of photodetectors are then able to measure the intensities of the deriving spectra and their wavelengths.

## 2.4 Phototransistor

A phototransistor (or photodiode) is a semiconductor device that generates an electric current when photons are absorbed. Its primary functionality is achieved through a p-n junction, a crystal with a "p" (positive) side con-

<sup>5</sup>(a) Licence: CC BY-SA 3.0, available at [https://upload.wikimedia.org/wikipedia/commons/f/f6/Spectrometer\\\_schematic.gif](https://upload.wikimedia.org/wikipedia/commons/f/f6/Spectrometer\_schematic.gif)

## CHAPTER 2. BASICS

taining an excess of electron holes and an "n" (negative) side containing an excess of electrons in the outer shells of its electrically neutral atoms.

This structure is elemental to many powerful electronic devices, such as LEDs and transistors. In the case of photodiodes, it creates a current when struck by a photon through the photoelectric effect. Holes move toward the anode and electrons toward the cathode, producing a photocurrent measured in electrons per photon or amps per watt. The energy of a photon is inversely proportional to its wavelength. The detected range of wavelengths for a phototransistor can be manipulated by using filters and different materials, e.g. silicon or germanium as its semiconductor.

# Chapter 3

## Machine Learning

Reducing the effort and complexity of a task has always been of major interest. In recent years, machine learning has become a new powerful tool, as programmable computers and the necessary computational power have become readily available. Section 3.1 gives a general overview of machine learning, while sections 3.2, 3.3, and 3.4 go into more detail for three specific machine learning algorithms.

### 3.1 Machine Learning Techniques

Nowadays, the field of Machine Learning can be divided into three subareas:

#### **Supervised Learning**

teaches machines the relationship between certain variables to a target variable and then extrapolate for new input data. They are most commonly applied in:

- Classification
- Regression

#### **Unsupervised Learning**

applies algorithms that learn by themselves. No target variables are given, so it is able to find hidden relations and patterns in the given data. The main principles are:

- Dimensionality reduction
- Clustering

**Reinforced Learning**

gives one or more agents feedback from the environment. They adapt their decisions on their own and are given either a positive or negative reward. Based on this reward, the agent reevaluates its decisions, steadily improving its performance over time.

There exist many options to implement and use machine learning algorithms. In this work, the more than sufficient functionalities and tools provided by MatLab's libraries (version R2019a) are utilized.

**3.1.1 Bias and Variance**

There is always a tradeoff between bias and variance in machine learning. These properties are sources of error and can prevent machine learning algorithms from generalizing for new data. Thus, their impact should be minimized.

- Selection bias causes the algorithm to miss the important relations between features and target variables due to erroneous assumptions (underfitting).
- Variance causes the algorithm to rather learn the random noise during training. This leads to strongly differentiating results due to high sensibility to fluctuations in the input (overfitting).

A high-bias algorithm is susceptible of overrepresenting features from the training set, but will therefore often give similar results (low variance). In turn, learning algorithms with high-variance may overfit and learn the noise in the training data itself, leading to drastically different results if the input is slightly changed (low bias).

**3.1.2 Cross-Validation**

To assess a model's ability to predict new data from the testing set, cross-validation is a common analysis technique. Its goal is to detect selection bias or overfitting by partitioning the data set into complementary training and testing sets. Repeating this process while changing different partitions, variance can be reduced by averaging over the results of every iteration.

One such method is called  $k$ -fold cross-validation, which randomly partitions the data into  $k$  equal sized subsamples [Fig. 3.1]. One is retained for validation, while the remaining  $k-1$  subsets are used for training. This is

### 3.2. DECISION TREE

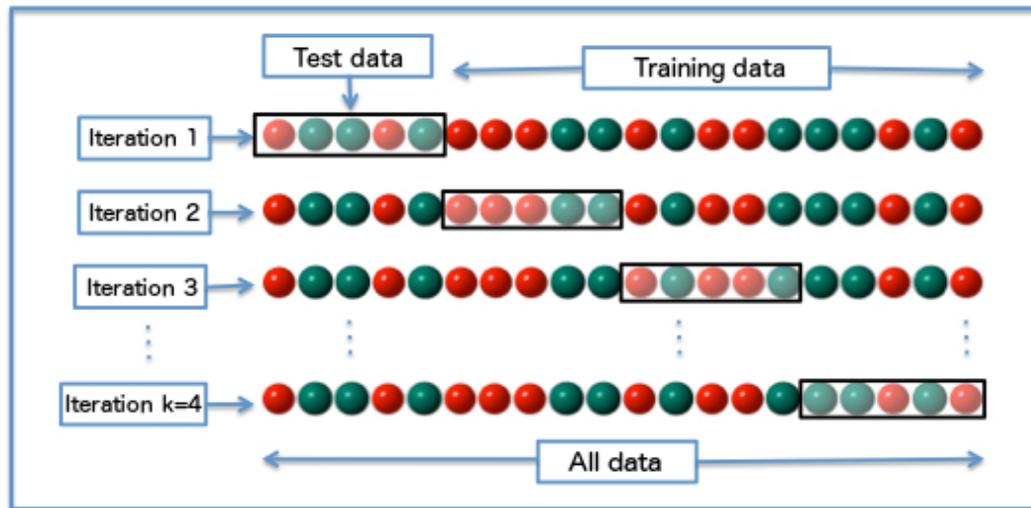


Figure 3.1: Diagram of  $k$ -fold cross-validation with  $k = 4$ . Generally,  $k$  can be chosen freely.<sup>6</sup>

repeated  $k$  times with every subset becoming the testing set once. By combining the results, a single outcome can be gained for which all observations were used for validation exactly once.

## 3.2 Decision Tree

A binary decision tree is a structure built on a sequential decision process. It consists of three components.

### Nodes

where the value of a certain feature is tested.

### Edges

that connect to the next node or leaf depending of the tests outcome.

### Leaves

which represent the terminal nodes and prediction.

It is a supervised learning algorithm that uses training examples to pick appropriate tests. The feature that maximizes information gain is placed at the top, from which the decision tree is built hierarchically [Fig. 3.2]. Then new examples are submitted to a series of tests to determine their class label. It

<sup>6</sup>(a) Licence: CC BY-SA 3.0, available at [https://upload.wikimedia.org/wikipedia/commons/1/1c/K-fold\\_cross\\_validation\\_EN.jpg](https://upload.wikimedia.org/wikipedia/commons/1/1c/K-fold_cross_validation_EN.jpg)

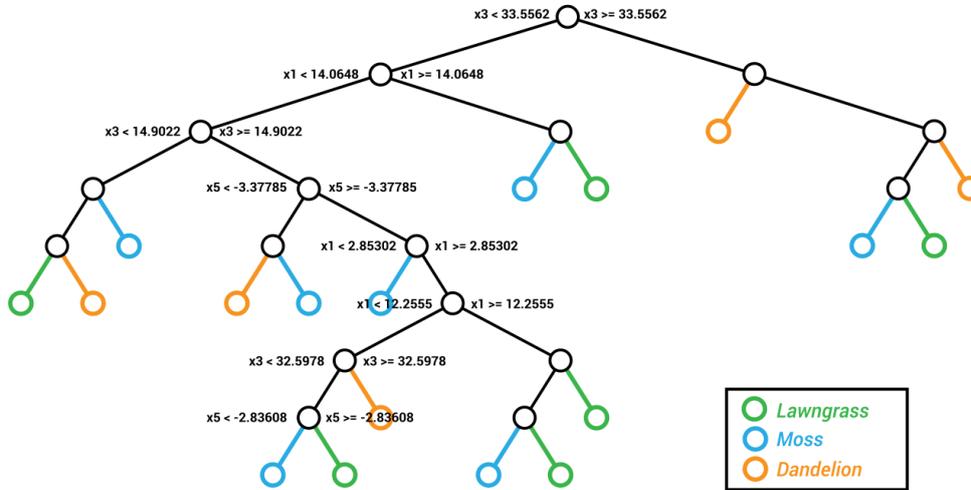


Figure 3.2: Example of a decision tree trained in MatLab. The tree was trained with the spectral answers of one LED and ambient light for lawn grass, dandelions and moss. The resulting structure reached an accuracy of over 93% consistently.

works for categorical and continuous input and output and is mostly used in classification problems.

### Finding optimal splits using entropy

The key to decision tree is to construct it from the given data alone. The best feature to split the data needs to be determined so that the resulting data subsets are as homogeneous as possible. A quantitative way to measure such quality is the Shannon entropy.

For a discrete probability distribution  $D$  with  $(p_1, p_2, \dots, p_n)$  the Shannon entropy is:

$$E(p_1, \dots, p_n) = - \sum_{i=1}^n p_i \log(p_i) \quad (3.1)$$

There are  $n$  possible outcomes and the probability that an instance drawn from  $D$  results in the outcome  $k$  is  $p_k$ . The Shannon's entropy function calculates a numerical value that describes how well the data is split into its classes. In the worst case each subset is uniformly distributed across all classes, resulting in an entropy of 1. The best scenario is a perfect split where every subset only has data from one class.

To construct a proper decision tree, the change in entropy needs to be quantified, so it can be minimized. With a data set  $S$ , a feature  $A$  with

### 3.3. RANDOM FORREST

values  $v \in V$  and Shannon's entropy function  $E$ , the gain of a split along the feature  $A$ , denoted  $G(S, A)$  can be calculated.

$$G(S, A) = E(S) - \sum_{v \in V} \frac{|S_v|}{|S|} E(S_v) \quad (3.2)$$

where  $S_v$  denotes the subset of  $S$  for which the feature  $A$  has value  $v$ .

The entropies of each part with an appropriate weight depending on each piece's size is subtracted off the entropy before the split. If the entropy increases the value will be small. Should the split separate the classes appropriately, each subset  $S_v$  will possess little entropy and  $G(S, A)$  will be large. Once the decision tree is constructed, exactly one terminal leaf with a unique path of rules exists for every single input. Also, classifying new data by checking the rules at every node until a leaf is reached is simple and very fast.

## 3.3 Random Forrest

Random Forest (RF) is a frequently used technique in data science to solve various problems. It uses an ensemble of decision trees [Fig. 3.3], which tend to have high variance and low bias, making them unstable. Averaging over multiple decision trees minimizes the variance, making it superior to other techniques such as logistic regression, which has low variance, but very high bias. The trees are trained on different parts of the same data set and since RF averages over multiple decision trees that might overfit, it lowers the variance of the final result. Generally, it improves the resulting model's performance, but the bias increases slightly and some of the interpretability is lost.

The basic functionality of RF is a technique called bootstrap aggregation (bagging). From a training set  $X = x_1, \dots, x_n$  with response  $Y = y_1, \dots, y_n$  a random sample with replacement (elements may appear multiple times) is selected  $B$  times from the training set. For every sample  $X_b, Y_b$  a classification (or regression) tree is trained as described in the previous section. After training, a majority vote on all the individual predictions of the individual classification trees is taken to decide the final prediction.

While a single tree's prediction might be sensible to noise in its training set, the average over many uncorrelated trees is not. The number of trees  $B$  can be chosen freely. Depending on the nature and size of the data set, hundreds or thousands of trees may be required. Using cross-validation, an optimal number can be found.

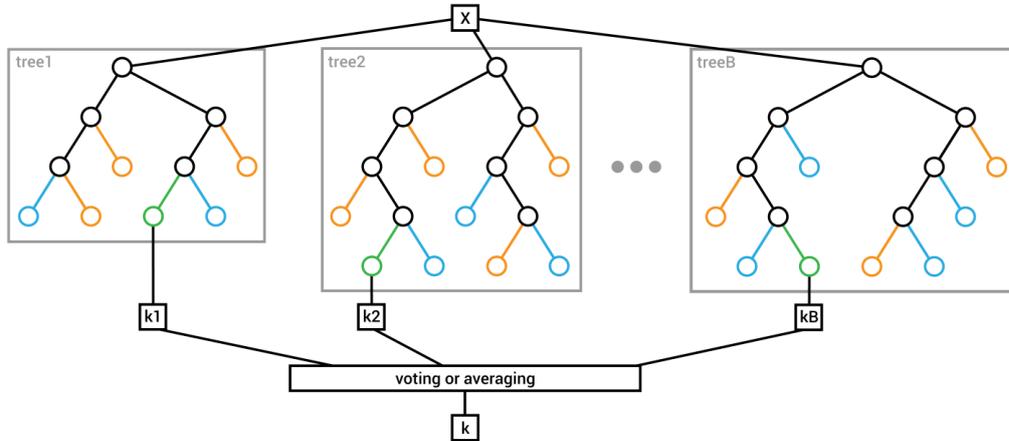


Figure 3.3: General architecture of the random forest model. It trains a number of  $B$  decision trees and votes or averages over their collective results.

### 3.4 Neural Network

Neural networks (NN) are inspired by biological brains and their networks of neurons. They can be used as a supervised, unsupervised or reinforced learning system for many different tasks. A NN consists of layers, typically an input and an output layer of nodes (called artificial neurons) and a number of hidden layers in between [Fig. 3.4]. These layers and nodes are connected by edges (representing the synapses of a brain) that transmit information if a node is activated. These edges have weights, that represent the impact of the signal passing through them. Thus, activations in one layer determine the activations of the next layer. As an information processing mechanism, the important detail is how those activations are propagated from one layer to the next so that it reaches an intelligent conclusion.

The intuitively understandable concept is that the input layer of nodes consists of many single or small datapoints of fixed size and position. Specific values activate the neuron, which leads to a signal being sent to the next layer of nodes. In turn, these add up all incoming signals and if their threshold is reached (called bias), they will be activated. This adds complexity with every layer until the output nodes are reached. These then activate according to the confidence in the prediction, usually measured as a probability. For classification, the output nodes are labeled with the possible classes.

In case of a fully connected NN in which information travels forward-only, it is called a feedforward neural network.

### 3.4. NEURAL NETWORK

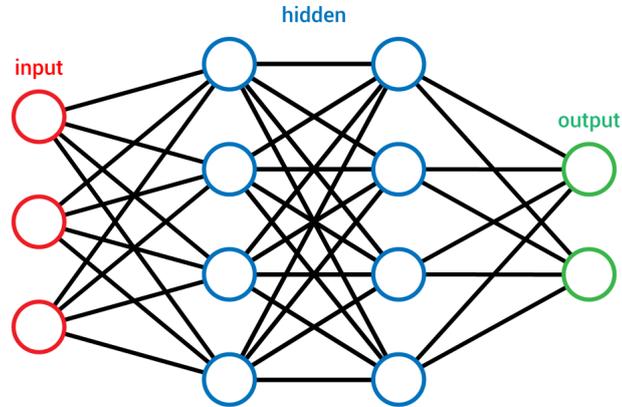


Figure 3.4: A neural network has a single input and output layer with an arbitrary number of hidden layers.

#### 3.4.1 Activation Function

If a neuron is to activate above a threshold and do nothing below it, a function with the same properties is required. The most basic would be the step function, which can only put out two values, 0 or 1. In case of continuous input, a continuous function is required.

A first approximation of the step function is the sigmoid function, but the softmax or rectifier linear unit (ReLU) activation functions are the most commonly used. In the following example

$$a_k^L = \text{softmax}(z_k^L) = \frac{e^{z_k^L}}{\sum_c e^{z_k^c}}$$

is used for the last layer  $L$  and

$$a_k^l = \text{relu}(z_k^l) = \max(0, z_k^l)$$

for a hidden layer annotated with  $l$ . The indices  $k, j$  and  $i$  denote neurons in the layers  $l, l-1$  and  $l-2$  respectively.

Basically, these functions calculate a value for any given node by taking the weighted sum  $z_k^l$  from all the nodes in the previous layer and a bias  $b$  specific to it.

$$z_k^l = b_k^l + \sum_j W_{kj}^l a_k^{l-1} \quad (3.3)$$

This is done for every node until the last layer is reached and the networks output determined.

### 3.4.2 Cross-Entropy Loss

With the network having reached an arbitrary output, it requires a method to identify its accuracy. With a labeled data set, it is possible to compare the networks conclusion with the correct answer. Starting with the output layer, the degree of every node's inaccuracy is calculated through an error function  $E$ . For regression problems, the mean squared error (MSE) is usually taken, which simply squares the difference between the networks result and the actual label for a given input.

For classification, cross-entropy loss is common practice and also the default algorithm in MatLab.

$$E_{\text{Cross-Entropy}} = - \sum_d t_d \log(a_k^L) = - \sum_d t_d (a_d^L - \log \sum_c e^{z_c^L})$$

with  $t_k$  as a binary indicator (1 or 0) for the correct classification of the  $k^{\text{th}}$  neuron in the output layer.

(3.4)

### Gradient Descent

With a method to measure the loss, the weight metric  $W_{kj}^l$  can be updated. This is done by negating the derivative of  $E$  with respect to the weights using a gradient descent algorithm. With a learning rate  $\alpha$  to control the step size the derivatives can be calculated as

$$\Delta W_{kj}^l = -\alpha \frac{\partial E}{\partial a_k^l} = -\alpha \frac{\partial E}{\partial a_k^l} \frac{\partial a_k^l}{\partial z_k^l} \frac{\partial z_k^l}{\partial W_{kj}^l}$$

### 3.4.3 Backpropagation

The update  $\frac{\partial E}{\partial a_k^l}$  is calculated for the last layer first. It determines the importance of every weight and bias of every node in the previous layer in relation to an output node.

This is done for every node, which allows for the weights and biases to be adjusted relative to their impact on this neuron. That is then incrementally propagated back for every previous layer with a recursive definition of  $\delta_k^l$  and  $\delta_k^{l-1}$ , which is why it is called backpropagation.

#### Updating the last layer

We simplify gradient descent to

$$\frac{\partial E}{\partial W_{kj}^L} = \frac{\partial E}{\partial z_k^L} \frac{\partial z_k^L}{\partial W_{kj}^L}$$

### 3.4. NEURAL NETWORK

to calculate the derivation. For this, we use the definition of  $E$  in formula 3.4. For the first half:

$$\begin{aligned}
 \frac{\partial E}{\partial z_k^L} &= - \sum_d t_d \left( 1_{d=k} - \frac{1}{\sum_c e^{z_c^L}} e^{z_k^L} \right) \\
 &= - \sum_d t_d (1_{d=k} - a_k^L) \\
 &= \sum_d t_d a_k^L - \sum_d t_d 1_{d=k} \\
 &= a_k^L - t_k
 \end{aligned} \tag{3.5}$$

with  $1_{d=k}$  as the identity function

$$1_{d=k} = \begin{cases} 1 & \text{if } d = k \\ 0 & \text{otherwise} \end{cases}$$

For  $a_k^L$  softmax gives a normalized probability for every output node, representing the networks prediction.  $t_k$  gives the value 1 for the correct label and 0 otherwise. With this, we receive a loss value for the node holding the correct label depending on the networks calculated probability.

We define this as  $\delta_k^L$

$$\delta_k^L = \frac{\partial E}{\partial z_k^L} = a_k^L - t_k$$

The second half of the equation remains to be solved. With the definition of  $z_k^L$  in formula 3.3, this is trivial.

$$\frac{\partial z_k^L}{\partial W_{kj}^L} = a_j^{L-1}$$

This propagates the loss in the final layer to the activations in the previous one. Thus, the resulting updates for the last layers weights are:

$$\frac{\partial E}{\partial W_{kj}^L} = \frac{\partial E}{\partial z_k^L} \frac{\partial z_k^L}{\partial W_{kj}^L} = \delta_k^L a_j^{L-1}$$

The derivation for the bias is similar.

$$\frac{\partial E}{\partial b_{kj}^L} = \frac{\partial E}{\partial z_k^L} \frac{\partial z_k^L}{\partial b_{kj}^L} = \delta_k^L (1) = \delta_k^L$$

### Updating the previous layer

We can get the equations to calculate the adjustments for the second to last layer in similar fashion.

$$\frac{\partial E}{\partial W_{ji}^{l-1}} = \frac{\partial E}{\partial a_j^{l-1}} \frac{\partial a_j^{l-1}}{\partial z_j^{l-1}} \frac{\partial z_j^{l-1}}{\partial W_{ji}^{l-1}} \quad (3.6)$$

The derivations of the three components

$$\begin{aligned} \frac{\partial E}{\partial a_j^{l-1}} &= \sum_k \frac{\partial E}{\partial z_k^l} \frac{\partial z_k^l}{\partial a_j^{l-1}} \\ &= \sum_k \delta_k^l W_{kj}^l \end{aligned} \quad (3.7)$$

$$\frac{\partial a_j^{l-1}}{\partial z_j^{l-1}} = f'(z_j^{l-1}) \quad (3.8)$$

$$\frac{\partial z_j^{l-1}}{\partial W_{ji}^{l-1}} = a_i^{l-2} \quad (3.9)$$

give us the equation

$$\frac{\partial E}{\partial W_{ji}^{l-1}} = a_i^{l-2} f'(z_j^{l-1}) \sum_k \delta_k^l W_{kj}^l. \quad (3.10)$$

By defining  $\delta_j^{l-1} = \frac{\partial E}{\partial a_j^{l-1}} \frac{\partial a_j^{l-1}}{\partial z_j^{l-1}} = f'(z_j^{l-1}) \sum_k \delta_k^l W_{kj}^l$

we can write 3.10 as

$$\frac{\partial E}{\partial W_{ji}^{l-1}} = \delta_j^{l-1} a_i^{l-2}.$$

And for the bias

$$\frac{\partial E}{\partial b_j^{l-1}} = \frac{\partial E}{\partial a_j^{l-1}} \frac{\partial a_j^{l-1}}{\partial z_j^{l-1}} \frac{\partial z_j^{l-1}}{\partial b_j^{l-1}} = \delta_j^{l-1} (1) = \delta_j^{l-1}$$

Using these derivatives, every layer can be updated accordingly.

## Chapter 4

# Experiment Setup and Measurement

The general setup consists of eight LEDs with different wavelengths that illuminate the plant and a spectrometer to measure the fluorescence induced by the light. To figure out if chlorophyll induction is a viable approach for plant detection for a robot working outside, data collection for training was intentionally not done under laboratory conditions. A robot would not be able to guarantee prerequisites such as constraints on lighting and angle or distance to plant matter in a real world scenario.

Hence, the plants fluorescence was mostly measured in sunlight with occasional changes in the angle to the plant. The distance was roughly kept approximately at 10 cm, but since leaves and stems grow in 3 dimensions, primary focus was to ensure that enough surface of the plant could be properly illuminated by the LEDs. The ambient light was artificially reduced in various degrees by shading. Also, data was collected on several days. Detailed information about changes to lighting condition and plants can be found in the respective protocol on the CD attached to the thesis.

To measure the fluorescence, a highly sensitive, broad spectrum spectrometer was used. The CCS200/M from Thorlabs has a wavelength range of 200 - 1000 nm and outputs two arrays with 3648 values each. One contains the measured light intensity and the other the corresponding wavelength for every value. It is connected to an adjustable lens via a fiber patch cable and a computer. Around the lens are the eight LEDs, which are also connected to the same computer [Fig. 4.1].

Through a custom written C# assembly, both can communicate with each other and be accessed in MatLab. A single scan consists of nine measurements, of which eight are a scan for each LED and one without to measure

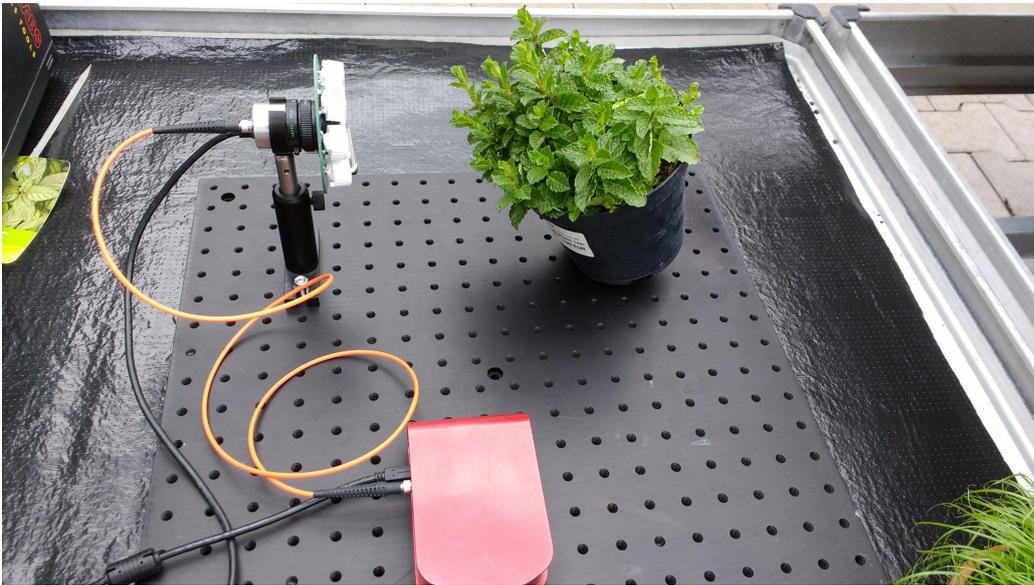


Figure 4.1: The general setup used for data collection. The plant's position was changed regularly, but the distance was kept around 10 cm and the green should receive sufficient light from every LED.

the ambient light. This concludes one measurement, with  $9 \times 3648$  data points for intensity, doubled for the corresponding wavelength array. Over 200 such scans were taken for several plants each, especially for those of great interest, such as lawngrass, dandelions, and moss. An example scan can be seen in Fig. 4.2.

### Measuring process

As soon as a scan is started through MatLab, it uses the custom assembly to initiate the spectrometer. Then, a signal is sent to the microprocessor, activating the first LED until the spectrometer confirms that it has completed its measurement for the given integration time, which is the time frame for measurement. All scans used a consistent integration value of 0.5 (500 ms). Following this, it receives the two arrays containing the spectral intensities with their corresponding wavelengths. This is repeated for every LED, followed by a final scan for ambient light intensity and is then saved to an innumerated file for the corresponding plant by MatLab. One such complete scan takes about 25 seconds.

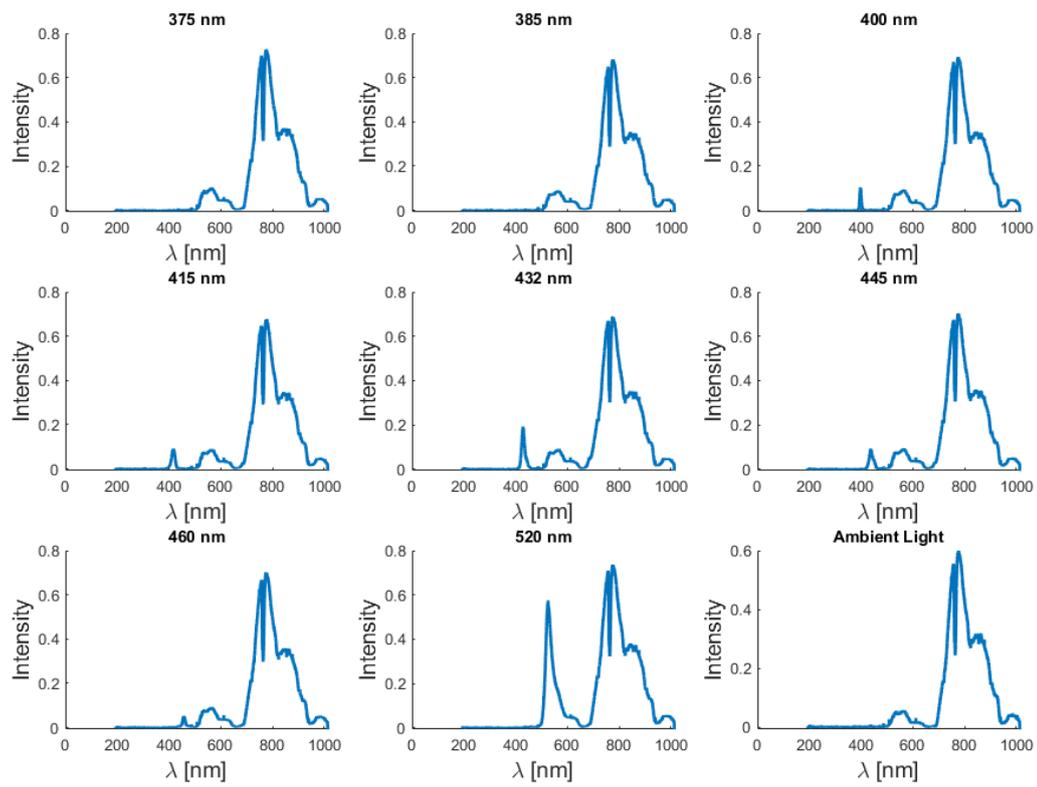


Figure 4.2: An example of a complete scan circle with 8 LEDs and one without (lawnglass scan 4).

*CHAPTER 4. EXPERIMENT SETUP AND MEASUREMENT*

# Chapter 5

## Classification Results

To evaluate the recorded data, the three machine learning methods described in chapter 3 are used and evaluated: Random Forest (RF), Decision Tree (DT), Neural Network (NN), Neural Network using Cross-Validation (NN CV) and Random Forest using Cross-Validation (RF CV). The goal is to find and use available phototransistors instead of a broadband spectrometer. Therefore, the relative spectral sensitivities of 24 reasonable phototransistors were acquired.

Before the machine learning algorithms are trained, the scan data is multiplied with the spectral sensitivities of the chosen phototransistors. This is to simulate the real world capabilities of these components, rather than the sensitivity of a broadband spectrometer. Then it is integrated over the resulting, interpolated intensity curve. Also, an option to subtract the ambient light is provided.

### 5.1 Selecting LED

The first goal is to reduce the search space by reducing the number of considered parameters. Since the sensor should need a minimal amount of LEDs the most relevant ones are determined in this step. Training a random forest with all LEDs, a single phototransistor and all plants, we can extract the feature importance from the tree [Fig. 5.1].

### 5.2 Finding best Phototransistors

LED 5 is picked and the three main plants to search for the most promising phototransistors. The complete tables with all results can be found in the

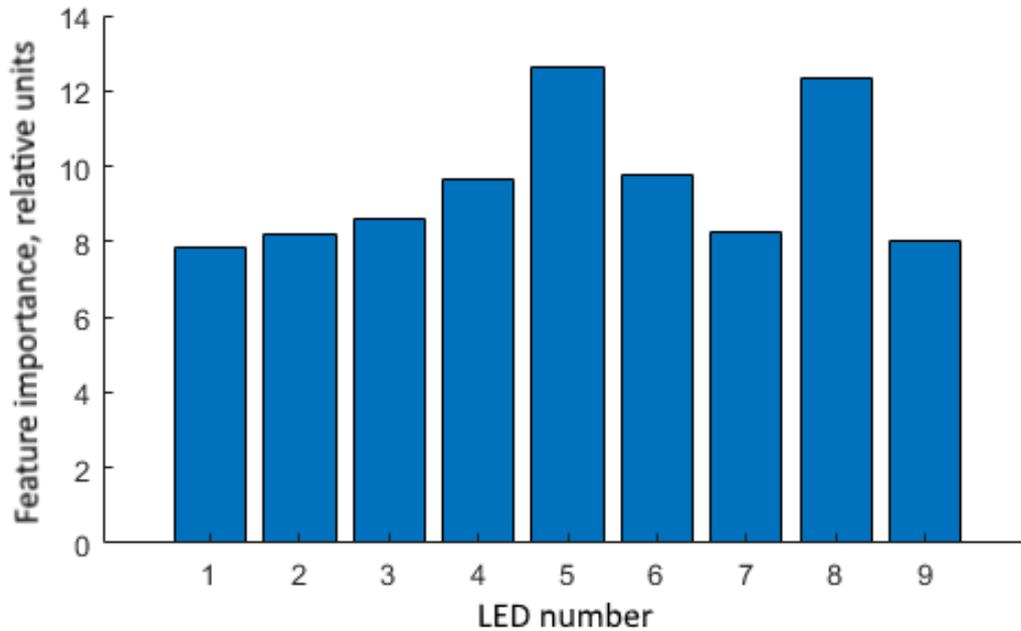


Figure 5.1: LED relevance in a trained random forest model according to its out-of-bag permuted variable delta error. LED number 5 and 8 were the most significant.

document `Classification_Evaluation.gsheet` on the CD attached to this thesis. The five best performing phototransistors, averaged over four training sessions once with ambient light and once without can be seen in table 5.1 containing the missclassification probabilities. The phototransistors 12, 5, 1, 8 and 2 reach best average accuracies below 66%. To improve the results, combinations of three out of all 24 are evaluated.

The three best performing combinations were then used for training, further confirming that the phototransistor with designation 12 is the most valuable candidate [Table 5.2]. Transistor 12 was part in 80 of the 100 best combinations. These considerably increased performance, with RF and RF CV consistently achieving about 94% accuracy. Thus, the classification for all plants (plant numbers 2,3,5,7,8,9,10,13) was tested [Table 5.3]. As expected, performance decreased with the higher number of plants to 80%. We can also see that the neural network was barely effected by any of the previous changes. The main observation is that for the data that does include the ambient light, RF and RF CV remain the best methods. For the data that removes it, NN CV is significantly better.

## 5.2. FINDING BEST PHOTOTRANSISTORS

Single Phototransistor with LED 5 and plants 3, 7, 8.					
LED	5	5	5	5	5
Plants	3,7,8	3,7,8	3,7,8	3,7,8	3,7,8
Phototransistors	12	5	1	8	2
RF no AL	0.32967	0.39560	0.39011	0.41978	0.43681
DT no AL	0.30165	0.38736	0.37747	0.39835	0.38901
NN no AL	0.28834	0.27684	0.29299	0.29857	0.31831
NN CV no AL	0.23780	0.28256	0.28122	0.29236	0.30355
RF CV no AL	0.33239	0.39281	0.38347	0.41970	0.43348
RF & AL	0.45963	0.39533	0.42243	0.41869	0.40657
DT & AL	0.42128	0.37860	0.38639	0.39937	0.38466
NN & AL	0.33167	0.31610	0.29071	0.31334	0.31951
NN CV & AL	0.29823	0.27023	0.27433	0.27924	0.27877
RF CV & AL	0.45241	0.39708	0.42448	0.42332	0.40803
Avg	<b>0.34530</b>	<b>0.34925</b>	<b>0.35236</b>	<b>0.36627</b>	<b>0.36787</b>

Table 5.1: Table showing the training evaluations for Random Forest (RF), Decision Tree (DT), Neural Network (NN), Neural Network using Cross-Validation (NN CV) and Random Forest using Cross-Validation (RF CV). Top half are the results when Ambient Light (AL) was removed, while it was kept for the bottom half. Best candidate is number 12 and coloured red.

3 best performing combinations of 3 phototransistors.			
LEDs	5	5	5
Plants	3,7,8	3,7,8	3,7,8
Phototransistors	1,12,24	2,12,24	9,12,16
RF no AL	<b>0.05330</b>	<b>0.04670</b>	<b>0.04890</b>
DT no AL	0.08956	0.08022	0.07253
NN no AL	0.25467	0.24805	0.22007
NN CV no AL	0.15616	<b>0.06348</b>	<b>0.05712</b>
RF CV no AL	<b>0.05166</b>	<b>0.04740</b>	<b>0.04343</b>
RF & AL	<b>0.05421</b>	<b>0.05450</b>	<b>0.05940</b>
DT & AL	0.08881	0.08939	0.08679
NN & AL	0.28801	0.32136	0.28458
NN CV & AL	0.19906	0.20660	0.19613
RF CV & AL	<b>0.05965</b>	<b>0.05997</b>	<b>0.06142</b>
Avg	<b>0.12951</b>	<b>0.12177</b>	<b>0.11304</b>

Table 5.2: On average over all ML methods, best performances were reached with phototransistors 9,12 and 16.

CHAPTER 5. CLASSIFICATION RESULTS

3 best performing combinations of 3 phototransistors.			
LED	5	5	5
Plants	all	all	all
Phototransistors	1,12,24	2,12,24	9,12,16
RF no AL	0.31426	0.29669	0.28636
DT no AL	0.38740	0.37851	0.36839
NN no AL	0.2832	0.28458	0.27633
NN CV no AL	0.18334	0.22240	0.18735
RF CV no AL	0.31529	0.30372	0.28781
RF & AL	0.20190	0.20079	<b>0.17730</b>
DT & AL	0.29810	0.29092	0.28511
NN & AL	0.28533	0.28415	0.28742
NN CV & AL	0.22403	0.22602	0.23163
RF CV & AL	0.21067	0.20524	<b>0.18608</b>
Avg	<b>0.27035</b>	<b>0.26930</b>	<b>0.25738</b>

Table 5.3: Phototransistors 9, 12 and 16 remained best combination for all plants

### Finding best LEDs

With the best values merely reaching 80% accuracy, we will try to further improve this by adding additional LED data, as some plants showed detectable fluorescence for other LEDs. That is done in similar fashion as with the phototransistors. First, every LED is evaluated on its own, followed by combinations of the best performing. Looking at the averages in table 5.4, we see that LEDs 5, 8 and 4 are the best performing. Furthermore, the data shows a big differences between ambient light (AL) and no AL, with all LEDs giving considerably better performance with AL. Probing combinations of these

Single LEDs							
LED	2	3	4	5	6	7	8
Plants	all	all	all	all	all	all	all
Photos	2,12,24	2,12,24	2,12,24	2,12,24	2,12,24	2,12,24	2,12,24
Avg	0.4818	0.3507	<b>0.3254</b>	<b>0.2675</b>	0.3616	0.4180	<b>0.3190</b>

Table 5.4: LED number 5 achieved best values. LED 1 and 2 proved to be equally insignificant.

LEDs, it is found that LEDs 5 and 8 give the best result [Table 5.5]. Adding LED 4 results in a minor improvement of around 15%. Since NN does not show promising results in any setup, it will not further be evaluated. DC will

### 5.3. FINAL PARAMETER SETUP

also be discarded, as it gives decent result for small number of plants only, but never better than RF or RF CV. Occasionally, NN CV achieves similar values as RF and RF CV, but most of the time its error is way higher.

Combinations of best performing LEDs					
LED	4,5	5,8	4,5,8	5,8	4,5,8
Plants	all	all	all	3,7,8	3,7,8
Photos	2,12,24	2,12,24	2,12,24	2,12,24	2,12,24
RF no AL	0.27913	0.21136	0.19835	<b>0.028022</b>	<b>0.023077</b>
NN CV no AL	0.17051	0.20589	0.18948	0.11165	<b>0.047753</b>
RF CV no AL	0.27851	0.21446	0.20248	<b>0.030254</b>	<b>0.021969</b>
RF & AL	0.16939	0.14021	0.13699	<b>0.02278</b>	<b>0.030565</b>
NN CV & AL	0.21607	0.24494	0.20973	0.24435	0.2413
RF CV & AL	0.17928	0.14627	0.14145	<b>0.030289</b>	<b>0.032886</b>
Avg	0.2155	<b>0.1939</b>	<b>0.1797</b>	<b>0.0779</b>	<b>0.0663</b>

Table 5.5: LEDs 5 and 8 are the best pair and LEDs 4, 5 and 8 the best 3-tuple.

## 5.3 Final Parameter Setup

Thus, LEDs 5 (432 nm) and 8 (520 nm) are the most effective pair, with LED 4 (415 nm) as an optional addition for slight improvement. After testing the three previously determined combinations of phototransistors as well as several others, the already best achieving tuple of numbers 9, 12 and 16 [Fig. 5.2] consistently gave the best results. Other phototransistors that often appeared in well performing groups were numbers 1, 2, 11, 17 and 24. Adding one or more of them might improve the results, but testing revealed only minor enhancement or even decrease in accuracy. With all parameters optimized, it is trained again [Table 5.6].

Through this, classification for the plants 3 (dandelion), 7 (lawngress) and 8 (moss) achieved 96-98% accuracy with slightly better results if the ambient light was removed from the data. If the algorithms are trained for all eighth plants, 86% accuracy is reached for data including AL, a significant difference to the 81% without.

Using the minimalistic setup with one LED, another 4-5% of accuracy are lost for the Random Forest methods and NN CV achieved the best results across all methods with 85% without AL. Additionally, it can be trained for the ninth LED, which would be all LEDs deactivated, meaning sunlight only.

CHAPTER 5. CLASSIFICATION RESULTS

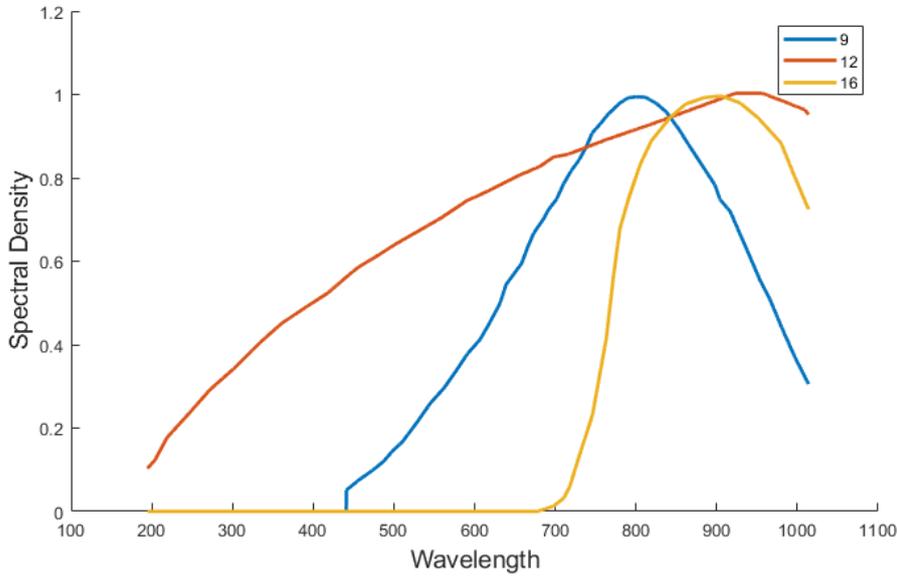


Figure 5.2: The spectral densities of phototransistors 9 (blue), 12 (red) and 16 (yellow).

Evaluation applying best parameters			
LED	4,5,8	4,5,8	5
Plants	3,7,8	2,3,5,7,8,9,10,13	2,3,5,7,8,9,10,13
Photos	9,12,16	9,12,16	9,12,16
RF no AL	<b>0.020192</b>	0.18719	0.28905
NN CV no AL	<b>0.067992</b>	0.18187	<b>0.15041</b>
RF CV no AL	<b>0.024725</b>	0.19163	0.28926
RF & AL	<b>0.030277</b>	<b>0.13684</b>	<b>0.17458</b>
NN CV & AL	0.248760	0.22721	0.24704
RF CV & AL	<b>0.032437</b>	<b>0.13972</b>	0.18571

Table 5.6: With three LEDs, RF and RF CV are the most robust methods. With only one LED, NN CV outperformed all other methods.

This still achieved up to 80% accuracy for the three main plants and 60% for all eight.

# Chapter 6

## Conclusion

The results show that using chlorophyll *a* fluorescence induction and ML algorithms is a viable approach for plant detection. Achieving 98% accuracy for three common plants present on a lawn and 86% for six more and rather arbitrary species is promising, even though these values dropped when the sensor complexity was reduced.

### Sources of Error

The target setup of one LED and three phototransistors (PT) only achieved about 84% accuracy for eight plants. One possible reason for this can be found in the small data sample. Plants had to be changed regularly, as a noticeable decline in fluorescence was noticed after using a plant and angle for a while. This can be explained by non-photochemical quenching, which starts later than the photon emitting process, as mentioned in section 2.2. Thus, either an area not used was illuminated or a new member of the same species was taken.

Generally, this is advantageous to the learning process and improves robustness for changes in age, size and health of individual plants, which can not be controlled in nature. But this necessitates a huge training set to avoid miss-classifications and confusions with other plants. Further proof can be found when we consider that the data for lawngrass and dandelions was collected separately from the other plants. Their fluorescence was measured on the same days, switching between the two species after 50 scans. Through this, the data would contain less fluctuations. Therefore, classification for these two plants is expected to be better than for the rest, but less applicable to the real world as the data lacks natural anomalies, while the data for the other plants lacks the necessary data quantity for the machine learning algorithms to learn all the details properly.

	2	153	1	9	10	4	7	13	22
Dandelion	1	293	13	8	5	5	15	9	
Daisy	14	14	137	4	13	19	10	18	
Lawnglass	9	8	2	232	2	7	28	11	
Moss	4	2	9	2	171	12	15	4	
9	7	1	22	6	18	158	5	12	
10	7	10	10	14	14	6	148	10	
13	22	4	18	10	10	14	10	171	
		2	Dandelion	Daisy	Lawnglass	Moss	9	10	13
		Predicted Class							

Figure 6.1: Confusion chart for the minimal setup of LED 5, PTs 9,12,16 for all 8 plants. Lawnglass and dandelion had the fewest miss-classifications.

The confusion chart [Fig. 6.1] further supports this argument. Indicated through the color and respective numbers, we see that lawnglass and dandelion had the best results and fewest miss-classifications across all eight species. Other sources of error are reflections from background objects and the plants itself. Reflecting background was avoided as much as possible, especially white or generally reflective surfaces. But some plants have considerable stronger reflecting, shiny leaves than others. This can be considered as just another feature that will positively reflect the data, but the major concern here is that the spectrometer can get "blinded" by this, meaning overly saturated at the LED's wavelength, resulting in a decreasing sensibility for the wavelengths of interest. An example of such a plant with "waxy" leaves is plant number 5, the european daisy (*Bellis perennis*) [Fig. 6.2]. The spectrometer maxed out over a span of wavelengths. If compared to a scan with an adjusted, more advantageous position of the same plant, the difference is clearly noticable [Fig. 6.3]. Adding to this is the fact that we scan for the

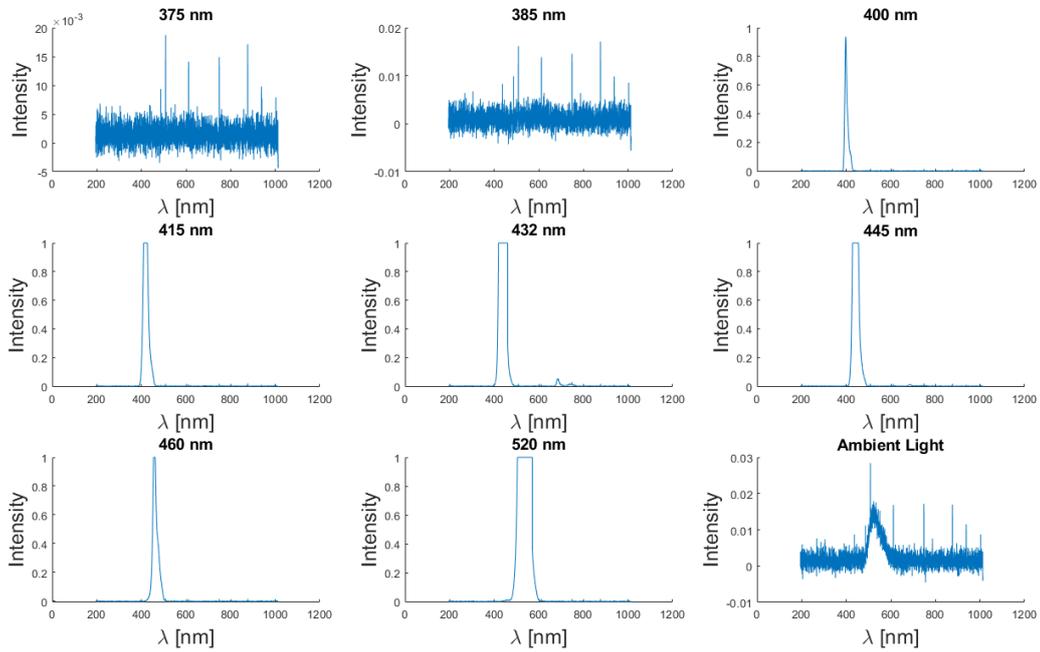


Figure 6.2: Example of an erroneous scan for a daisy (scan 4). Fluorescence is measurable and barely visible for LED 5 (432 nm) only.

ambient light only once. The entire scan beginning with the first LED and ending with the final AL measurement takes roughly 25 seconds. During this time, light conditions can easily change drastically (e.g. due to clouds). This can be improved by measuring AL before or after every LED, which would double the scan time. Alternatively, a second, identical spectrometer could be used in parallel with a slightly changed setup to guarantee consistency.

## The Machine Learning Algorithms

Since the collected data is labeled with the corresponding plant, supervised learning algorithms were chosen for evaluation. The methods come with different benefits and drawbacks that need to be considered.

A decision tree is an intuitive and fast predictor even capable of handling unnormalized datasets, as its internal structure is not influenced by the values assumed by each feature. They perform well for few variables and are easy to interpret and explain. Also, they are easy to implement on a microprocessor. But without proper constraints, the decision tree might grow until only a small sample number is properly represented in its nodes. This

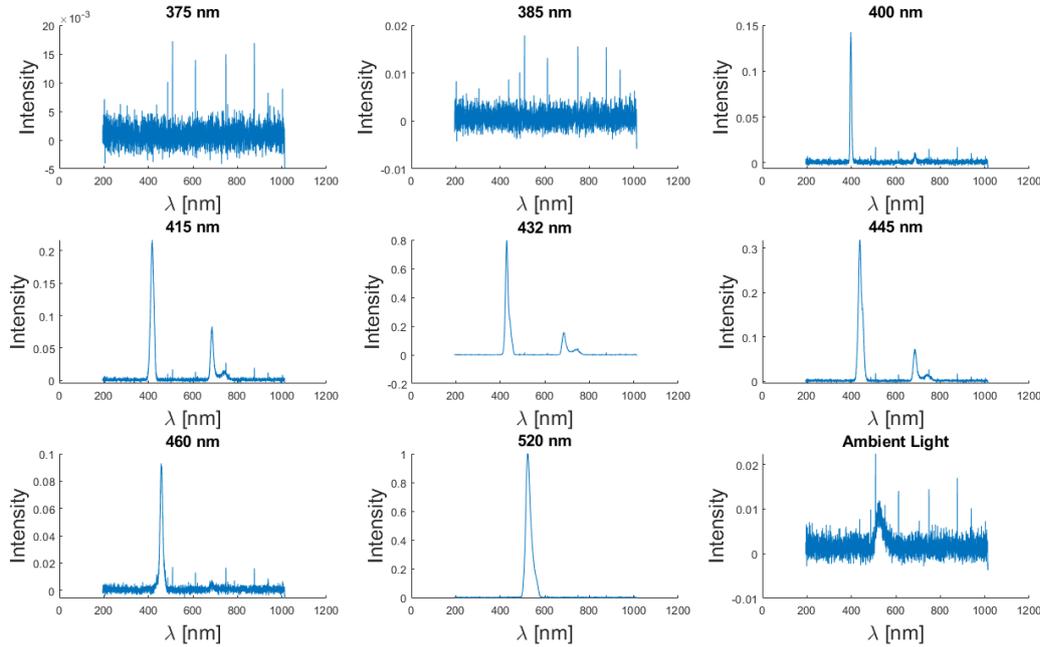


Figure 6.3: Example of an improved scan for the same daisy (scan 10). Fluorescence is clearly visible for LEDs 4, 5 and 6 and still measurable for LEDs 3 and 7.

is called overfitting and the structure will no longer be able to generalize correctly.

For the purpose of this thesis, DT classification accuracy rapidly declined with parameter numbers and proved inadequate. Unless only a few plants need to be detectable, it will not be able to consistently classify accurately.

Therefore, random forest was implemented. As it averages over many trees, variance is reduced. To lower bias we also used a RF variant that applied  $k$ -fold cross-validation. Each tree is trained independently, using a random sample of the data. This makes the algorithm robust and also makes parallel computing very effective. They have a lower classification error than decision trees, can easily handle high dimensional spaces and a large number of training examples. In exchange, some of the interpretability is lost.

Both RF and RF CV showed the most promising and consistent results. Across all evaluations, they performed equally well, with a minor edge to RF without CV. An explanation might be that RF CV was more prone to fluctuation in the performance of the system it was trained on, as it would

requiere more ressources.

To add an unrelated machine learning method, we used neural networks for classification as well. They excell in finding non-linear data with many input features and are widely used in the industry. Generally, NNs are able to learn almost any function, even if the type of function or data isn't known beforehand. But this comes at the complete loss of interpretability. It can be regained, but requieres additional effort. Also, it is difficult to tell if they will generalize well to new data, trained models depend crucially on initial parameters and they are difficult to troubleshoot if they do not work as intended.

In context of this work, NN can be discarded. NN CV on the other hand showed rather anomalous behavior. It underperformed most of the time, but just small changes to the parameters drastically improved classification. As mentioned in this paragraph, it is inherently difficult to determine the origin of this behavior. There are many possible explanations, two of which we will discuss. First, the network could generally be able to classify the plants, potentially even better than RF (as can be seen in the final data table). But the sample size is simply too small and all the other, above mentioned sources of error and the sensible nature of NNs lead to many miss-classifications. Another possibility could be that under certain conditions the NN with CV managed to memorize the entire data set including noise, as discussed in [Arpit et al., 2017].

### **Recommendation**

The recommended and best LED is number 5 with a wavelength of 432 nm. To this, a tuple of phototransistors consisting of number 9, 12 and 16 is suggested. Using these components, a low-cost sensor can be designed and used to collect new spectral data. The recommended classifier is RF with CV, since it proved to be the most reliable and robust method.

*CHAPTER 6. CONCLUSION*

# Bibliography

- [Arpit et al., 2017] Arpit, D., Jastrzebski, S., Ballas, N., Krueger, D., Bengio, E., Kanwal, M. S., Maharaj, T., Fischer, A., Courville, A., Bengio, Y., and Lacoste-Julien, S. (2017). A closer look at memorization in deep networks. In *Proceedings of the 34th International Conference on Machine Learning - Volume 70*, ICML'17, pages 233–242. JMLR.org.
- [Atherton et al., 2014] Atherton, J., Porcar-Castell, A., Tyystjärvi, E., van der Tol, C., Flexas, J., Pfündel, E. E., Moreno, J., Frankenberg, C., and Berry, J. A. (2014). Linking chlorophyll a fluorescence to photosynthesis for remote sensing applications: mechanisms and challenges. *Journal of Experimental Botany*, 65(15):4065–4095.
- [Banks, 2017] Banks, J. M. (2017). Continuous excitation chlorophyll fluorescence parameters: a review for practitioners. *Tree Physiology*, 37(8):1128–1136.
- [Cen et al., 2017] Cen, H., Weng, H., Yao, J., He, M., Lv, J., Hua, S., Li, H., and He, Y. (2017). Chlorophyll fluorescence imaging uncovers photosynthetic fingerprint of citrus huanglongbing. *Frontiers in Plant Science*, 8:1509.
- [Elhariri et al., 2014] Elhariri, E., El-Bendary, N., and Hassanien, A. E. (2014). Plant classification system based on leaf features. In *2014 9th International Conference on Computer Engineering Systems (ICCES)*, pages 271–276.
- [Johnson and Maxwell, 2000] Johnson, G. N. and Maxwell, K. (2000). Chlorophyll fluorescence—a practical guide. *Journal of Experimental Botany*, 51(345):659–668.
- [Kalaji et al., 2017] Kalaji, H. M., Schansker, G., Brestic, M., Bussotti, F., Calatayud, A., Ferroni, L., Goltsev, V., Guidi, L., Jajoo, A., Li, P., Losciale, P., Mishra, V. K., Misra, A. N., Nebauer, S. G., Pancaldi, S., Penella,

## BIBLIOGRAPHY

- C., Pollastrini, M., Suresh, K., Tambussi, E., Yannicari, M., Zivcak, M., Cetner, M. D., Samborska, I. A., Stirbet, A., Olsovska, K., Kunderlikova, K., Shelonzek, H., Rusinowski, S., and Baba, W. (2017). Frequently asked questions about chlorophyll fluorescence, the sequel. *Photosynthesis Research*, 132(1):13–66.
- [KAUTSKY et al., 1960] KAUTSKY, H., APPEL, W., and AMANN, H. (1960). [chlorophyll fluorescence and carbon assimilation. part xiii. the fluorescence and the photochemistry of plants.]. *Biochemische Zeitschrift*, 332:277—292.
- [Kusnierek and Korsaeht, 2015] Kusnierek, K. and Korsaeht, A. (2015). Simultaneous identification of spring wheat nitrogen and water status using visible and near infrared spectra and powered partial least squares regression. *Computers and Electronics in Agriculture*, 117:200 – 213.
- [Moshou et al., 2001] Moshou, D., Vrindts, E., Ketelaere, B. D., Baerde-maeker, J. D., and Ramon, H. (2001). A neural network based plant classifier. *Computers and Electronics in Agriculture*, 31(1):5 – 16.
- [Peteinatos et al., 2016] Peteinatos, G. G., Korsaeht, A., Berge, T. W., and Gerhards, R. (2016). Using optical sensors to identify water deprivation, nitrogen shortage, weed presence and fungal infection in wheat. *Agriculture*.
- [Raven et al., 2005] Raven, P. H., Evert, R. F., Eichhorn, S. E., et al. (2005). *Biology of plants*. Macmillan.
- [Rosenqvist and Baker, 2004] Rosenqvist, E. and Baker, N. R. (2004). Applications of chlorophyll fluorescence can improve crop production strategies: an examination of future possibilities. *Journal of Experimental Botany*, 55(403):1607–1621.
- [Saleem et al., 2019] Saleem, G., Akhtar, M., Ahmed, N., and Qureshi, W. (2019). Automated analysis of visual leaf shape features for plant classification. *Computers and Electronics in Agriculture*, 157:270 – 280.
- [Simko et al., 2015] Simko, I., Jimenez-Berni, J. A., and Furbank, R. T. (2015). Detection of decay in fresh-cut lettuce using hyperspectral imaging and chlorophyll fluorescence imaging. *Postharvest Biology and Technology*, 106:44 – 52.

## BIBLIOGRAPHY

- [Spaans et al., 2015] Spaans, S., Weusthuis, R., Van Der Oost, J., and Kengen, S. (2015). Nadph-generating systems in bacteria and archaea. *Frontiers in Microbiology*, 6:742.
- [Wu et al., 2007] Wu, S. G., Bao, F. S., Xu, E. Y., Wang, Y., Chang, Y., and Xiang, Q. (2007). A leaf recognition algorithm for plant classification using probabilistic neural network. In *2007 IEEE International Symposium on Signal Processing and Information Technology*, pages 11–16.