GIRIJANANDA CHOWDHURY INSTITUTE OF MANAGEMENT AND TECHNOLOGY,

AZARA, HATHKHOWAPARA, GUWAHATI-781017

PROJECT REPORT

ON

"Estimation of chlorophyll value from digital image using Deep Learning"

Submitted in completion of the project for the requirement for the degree of

Bachelor of Technology

in

DEPARTMENT OF

COMPUTER SCIENCE AND ENGINEERING

ASSAM SCIENCE AND TECHNOLOGY UNIVERSITY GUWAHATI

Submitted by:

1. Sahrul Alom Choudhari (170310007044)
2. George Bhokta (170310007019)
3. Abhinab Saikia (170310007002)

Project Guide:

Dr. Utpal Barman

Assistant Professor

Department of CSE,

GIMT, Guwahati

DECLARATION

We hereby declare that this project work entitled **"Estimation of chlorophyll value from digital image using Deep Learning"** was carried out by us under the guidance and supervision of **Dr. Utpal Barman,** Assistant Professor, Department of Computer Science and Engineering, Girijananda Chowdhury Institute of Management and Technology, Guwahati. This project is submitted to Department of Computer Science and Engineering during the academic year 2020-21. The work is never produced before any authority except Assam Science and Technology University for evaluation.

Name of Project Members	Roll No.	Signature
1. Sahrul Alom Choudhari	170310007044	
2. George Bhokta	170310007019	
3. Abhinab Saikia	170310007002	

GIRIJANANDA CHOWDHURY INSTITUTE OF MANAGEMENT AND TECHNOLOGY

SESSION 2017-21

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

CERTIFICATE

This is to certify that George Bhokta, Sahrul Alom Choudhari and Abhinab Saikia, Students of B.Tech 4th year 8th Semester have completed the project titled "**Estimation of chlorophyll value from digital image using Deep Learning**" during the academic session 2020-21 under my guidance and supervision. I approve the project for submission as required for the completion of Bachelor of Technology Degree.

(Signature of the Project Guide)

Dr. Utpal Barman

Assistant Professor
Department of CSE,

GIMT, Guwahati

ACKNOWLEDGEMENT

"Gratitude is not only the greatest of virtues, but the parent of all the others."

We would like to express our heartfelt gratitude to our Principal, **Prof. Thuleswar Nath**, GIMT-Guwahati and to our HOD, **Dr. Th. Shanta Kumar**, Department of Computer Science and Engineering, GIMT-Guwahati for giving us the freedom to work on this project and for also providing occasional guidance when needed. Our sincere gratitude goes to our project guide, **Dr. Utpal Barman**, Assistant Professor, Department of Computer Science and Engineering, GIMT-Guwahati for being with us, encouraging us to complete our project with honesty and sincerity. And lastly, we also convey our sincere thanks to all the lab bearers and office bearers for their active and enthusiastic co-operation.

Name of Project Members	Roll no	Signature
1. Sahrul Alom Choudhari	170310007044	
2. George Bhokta	170310007019	
3. Abhinab Saikia	170310007002	

ABSTRACT

The chlorophyll of leaf can be determined using soil plant analysis development meter or spectrophotometer or **Soil Plant Analysis Development** (SPAD) by agriculture scientists, agriculture experts, and farmers. Usually, these methods are very costly and may not be available to all the farmers and experts. Low greenness of leaf indicates low photosynthesis in the plant, and it creates many problems in the plant. This project put forwards a low-cost smartphone contact image based digital chlorophyll meter to predict the chlorophyll of tea leaf. This project makes an effort to predict chlorophyll of tea leaf using Multiple Linear Regression (MLR), Support Vector Regression(SVR) and Artificial Neural Network(ANN).

CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1: INTRODUCTION

1.1 Overview

The chlorophyll of leaf can be determined using SPAD or Spectrometer used by agricultural experts. Usually, these methods are very costly and may not be available to small framers. And as we know there are 800 tea Estates in Assam and 13000 of them all across India and more than 1000 small tea gardens. Hence, for small farmers the traditional methods are costly and may not be available easily.

Therefore, in our project we are trying to predict the chlorophyll value of leaves using smartphone images which will cut the cost of detecting chlorophyll value which was otherwise not in the case of traditional methods.

1.2 What is chlorophyll ?

$$6\,CO_2 + 6H_2O \xrightarrow[\text{Chlorophyll}]{\text{Sunlight}} C_6H_{12}O_6 + 6O_2$$

Photosynthetic Reaction

Fig. 1 Photosynthetic Reaction

Chlorophyll, in various forms, is bound within the living cells of algae and other phytoplankton found in surface water. Chlorophyll is a key biochemical component in the molecular apparatus that is responsible for photosynthesis, the critical process in which the energy from sunlight is used to produce life-sustaining oxygen. In the photosynthetic reaction below, carbon dioxide is reduced by water, and chlorophyll assists this transfer. Chlorophyll is present in many organisms including algae and some species of bacteria. Chlorophyll a is the most abundant form of chlorophyll within photosynthetic organisms and, for the most part, gives plants their green color. However, there are other forms of chlorophyll, coded b, c, and d, which augment the overall fluorescent signal. These types of chlorophyll, including chlorophyll a, can be present in all photosynthetic organisms but vary in concentrations. Chlorophyll enables plants and other chlorophyll-containing organisms to perform photosynthesis. Chlorophyll is a chelate, or a central metal ion, in this case magnesium, which is bonded to a larger organic molecule called a porphyrin. The porphyrin molecule is composed of carbon, hydrogen, and other elements such as nitrogen and oxygen. The magnesium ion bonded within this ring is thought to be responsible for electron transfer during photosynthesis (below)

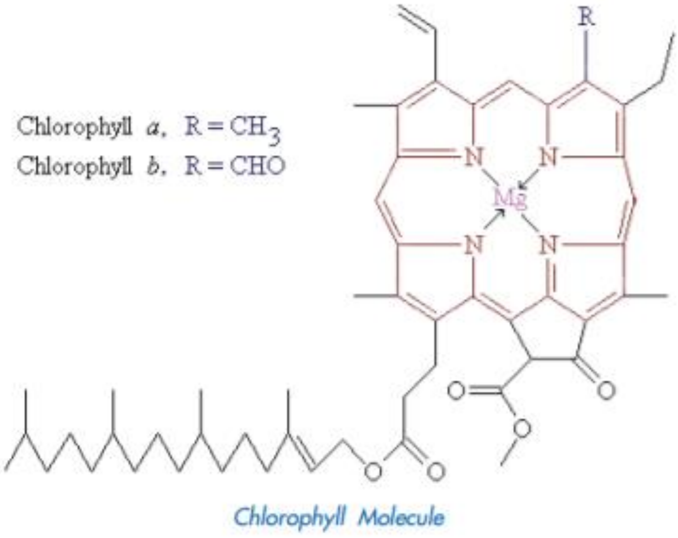


Fig. 2 Chlorophyll Molecule

The Importance of Chlorophyll as a Water Quality Parameter:

The measurement and distribution of microscopic living plant matter, commonly referred to as phytoplankton or algae, have been of interest to scientists, researchers, and aquatic resource managers for decades. An understanding of the phytoplankton population and its distribution enables researchers to draw conclusions about a water body's health, composition, and ecological status. Although researchers often refer to microscopic living plant matter as either algae or phytoplankton, and the two terms are often used interchangeably, each has distinct meaning. Algae refer to simple aquatic organisms, such as seaweed, pond scum, and plankton, that are plant like and contain chlorophyll. Phytoplankton are a subset of algae and are the suspended aquatic microorganisms that contain chlorophyll. For in-situ monitoring, the measured parameter is the chlorophyll contained within the phytoplankton. Chlorophyll is essential to the existence of phytoplankton. Phytoplankton can be used as an indicator organism for the health of a particular body of water. Monitoring chlorophyll levels is a direct way of tracking algal growth. Surface waters that have high chlorophyll conditions are typically high in nutrients, generally phosphorus and nitrogen. These nutrients cause the algae to grow or bloom. When algae populations bloom, then crash and die in response to changing environmental conditions, they deplete dissolved oxygen levels - a primary cause of most fish kills. High levels of nitrogen and phosphorus can be indicators of pollution from man-made sources, such as septic system leakage, poorly functioning wastewater treatment plants, or fertilizer runoff. Thus, chlorophyll measurement can be utilized as an indirect indicator of nutrient levels.

Reasons to Measure Chlorophyll :

Algal content can be tracked in surface water and, with time, databases and quality assurance protocols can be developed to characterize lakes or streams. These characterizations can be used for the indirect monitoring and detection of indicator pollutants, including phosphorus and nitrogen. Monitoring phytoplankton concentration is a less expensive alternative to frequent collection of grab samples for costly and labour-intensive laboratory analysis. (continued) Photosynthetic Reaction 6 CO2 + 6H2OgC6H12O6 + 6O2 Sunlight Chlorophyll Chlorophyll Molecule 0113 T606-01 In general, the amount of chlorophyll in a collected water sample is used as a measure of the concentration of suspended phytoplankton. The use of the measurement of phytoplankton as an indicator of water quality is described in Section 10200 A. of Standard Methods for the Examination of Water and Wastewater. Currently, chlorophyll determinations are made on lakes, rivers, reservoirs, and coastal and ocean waters across the globe. Ocean and coastal studies investigate the distribution of phytoplankton in marine systems. These studies can help track and predict deadly algae blooms. The nutrient enrichment of water bodies is leading to increased production of organic matter and resulting in low levels of dissolved oxygen that are killing marine life. Also, ocean profiling can track and record chlorophyll readings, which may change vertically along a column of water. Rivers and streams are monitored for excessive growth of phytoplankton due to high concentrations of plant nutrients. This excessive growth can lead to eutrophication of the river or stream and cause deadly fish kills. For similar reasons, lake, pond, and reservoir monitoring, including lake profiling studies, also observe excessive algae population distribution and growth. Algae control is a major concern in pond management, especially in smaller bodies of water, where excessive algae growth can quickly become a problem. Measuring chlorophyll concentration is also a step in the process of screening/monitoring for nuisance algal blooms that may influence the taste and odor of drinking water sources. These blooms may actually create conditions that are toxic to fish, wildlife, livestock, and humans. Bodies of water used as drinking water sources are also monitored for phytoplankton concentrations for the early detection of algal blooms to minimize filtration system clogs.

1.3 Traditional Methods of chlorophyll detection

There are various techniques to measure chlorophyll, including spectrophotometry, high performance liquid chromatography (HPLC), and fluorometry. All of these methods are published in Standard Methods for the Examination of Water and Wastewater, 19th Edition. Spectrophotometry is the classical method of determining the quantity of chlorophyll in surface water. It involves the collection of a fairly large water sample, filtration of the sample to concentrate the chlorophyll-containing organisms, mechanical rupturing of the collected cells, and extraction of the chlorophyll from the disrupted cells into the organic solvent acetone. The extract is then analyzed by either a spectrophotometric method (absorbance or fluorescence), using the known optical properties of chlorophyll, or by HPLC. This general method, detailed in Section 10200 H. of Standard Methods, has been shown to be accurate in multiple tests and applications and is the procedure generally accepted for reporting in scientific literature. The fluorometric method also requires the same extraction methods used with spectrophotometry, then uses a fluorometer to measure discrete molecular chlorophyll fluorescence. However, these methods have significant disadvantages. They are time consuming and usually require an experienced, efficient analyst to generate consistently accurate and reproducible results. In addition, they do not lend themselves readily to continuous monitoring of chlorophyll (and thus phytoplankton) because the collection of samples at reasonable time intervals, e.g., every hour, would be extremely time-consuming. YSI has developed optical sensors for chlorophyll determinations both in spot sampling and in continuous monitoring applications. The sensors are based on an alternative method for the measurement of chlorophyll that overcomes these disadvantages, albeit with the potential loss of accuracy. In this procedure, chlorophyll is determined in situ without disrupting the cells as in the extractive analysis. The YSI 6025 chlorophyll sensor and EXO 599102 total algae sensor are designed for these insitu applications, and their use allows the facile collection of large quantities of chlorophyll data in either spot sampling or continuous monitoring applications. It is important to remember, however, that the results of in-situ analysis will not be as accurate as results from the certified extractive analysis procedure. The limitations of the in-situ method should be carefully considered before making chlorophyll determinations with the YSI sonde and sensor. Some sources of inaccuracy can be minimized by combining extractive analysis of a few samples during a sampling or monitoring study with the YSI sensor data. The in-situ studies will never replace the standard procedure. The estimates of chlorophyll concentration from the easy-to-use YSI chlorophyll system are designed to complement the more accurate, but more difficult to obtain, results from more traditional methods of chlorophyll determination.

1.4 Laboratory Method

Step 1. First the leaves are cut into small pieces so that it would grind easily.

Fig. 3 Cutting

Step 2. Then the leaves are grinded to fine particles to break open the cell membrane to release the chlorophyll into the solution.

Fig. 4 Grinding

Step 3. Add small amount of $MgCO_3$ to aid in the process.

Fig. 5 Adding $MgCO_3$

Step 4. Add Acetone to the solution.

Fig. 6 Adding Acetone

Step 5. Put the solution in Spectrophotometer.

Fig. 7 Spectrophotometer

1.5 Using SPAD

It is a lightweight handheld meter for measuring the chlorophyll content of leaves without causing damage to plants. It provides an indication of the amount of chlorophyll present in plant leaves.

Fig. 8 SPAD

1.6 Scope of the project

This project aims to help the small scale Tea Garden Farmers who are not able to monitor there plant health due to lack of necessary equipment. We aim to develop a cheap method to find the chlorophyll value of the tea leaves and hence help the farmers to better organize the tea plantation.

Chapter 2: REQUIREMENT ANALYSIS

2.1 Hardware requirements:

- Minimum RAM 8GB, 12GB Recommended.
- 4GB of available disk space minimum, 8GB Recommended.
- 1280*800 minimum screen resolution.
- 64-Bit Operating System.

2.2 Software Requirements:

- Python 3.6.
- Anaconda Navigator
- Jupyter Notebook

Chapter3: PREREQUISITES

3.1 Keras

Keras is an open-source neural-network library written in Python. It is capable of running on top of TensorFlow, Microsoft Cognitive Toolkit, Theano, or PlaidML. Designed to enable fast experimentation with deep neural networks, it focuses on being user-friendly, modular, and extensible.

3.2 Numpy

Numpy is a general-purpose array-processing package. Its provide a high-performance multidimensional array object, and tools for working with these arrays. It is the fundamental package of scientific computing with Python. Besides its obvious scientific uses, Numpy can also be used as an efficient multidimensional container of generic data.

3.3 OpenCV

OpenCV supporta wide variety of programming language such as C++, Python, Java, etc., and is available on different platforms including Windows, Linux, OS X, Android, and Ios. Interfaces for high-speed GPU operations based on CUDA and OpenCL are also under active development. OpenCV-Python is the Python API for the OpenCV, combining the best qualities of the OpenCV C++ API and the Python language.

3.4 Scikit-Learn

Scikit-learn (formerly scikits.learn) is a free software machine learning library for the Python programming language.It features various classification, regression and clustering algorithms including support vector machines, random forests, gradient boosting, k-means and DBSCAN, and is designed to interoperate with the Python numerical and scentific libraries NumPy and SciPy.

3.5 Matplotlib

Matplotlib is a plotting library for the Python programming language and its numerical mathematics extension NumPy. It provides an object oriented API for embedding plots into applications using general-purpose GUI toolkits like Tkinter, Python, Qt, or GTK+. There is also procedural “pylab” interface based on a state machine(like OpenGL), designed to closely resemble that of MATLAB, through its use is discouraged.

Chapter 4: LITERATURE SURVEY

Sl. No.	Paper	Plant	Color Index	R^2	Model
1.	(Ali et al., 2012)	Tomato	[G-R/3-G/3]/255	0.96	Linear Regression
2.	(Ali et al., 2012)	Broccoli	Logsig = [GR/3-G/3]/255	0.91	Linear Regression
3.	(Dey et al., 2016)	Pan Betel	NRI = R/(R + G + B)	0.95	Linear Regression
4.	(Vesali et al., 2015)	Corn	4 Different Color feature	0.82	ANN

Table 4.1: Literature Survey

Chapter 5: BASIC CONCEPTS USED

5.1 Machine Learning

Machine learning is an application of artificial intelligence (AI) that provides systems the ability to automatically learn and improve from experience without being explicitly programmed. Machine learning focuses on the development of computer programs that can access data and use it learn for themselves. The primary aim is to allow the computers learn automatically without human intervention or assistance and adjust actions accordingly. Machine learning enables analysis of massive quantities of data. While it generally delivers faster, more accurate results in order to identify profitable opportunities or dangerous risks, it may also require additional time and resources to train it properly. Combining machine learning with AI and cognitive technologies can make it even more effective in processing large volumes of information.

Machine learning algorithms are often categorized as supervised or unsupervised.

Supervised machine learning algorithms can apply what has been learned in the past to new data using labelled examples to predict future events. Starting from the analysis of a known training dataset, the learning algorithm produces an inferred function to make predictions about the output values. The system is able to provide targets for any new input after sufficient training. The learning algorithm can also compare its output with the correct, intended output and find errors in order to modify the model accordingly.

Unsupervised machine learning algorithms are used when the information used to train is neither classified nor labelled. Unsupervised learning studies how systems can infer a function to describe a hidden structure from unlabelled data. The system doesn’t figure out the right output, but it explores the data and can draw inferences from datasets to describe hidden structures from unlabelled data.

5.2 Deep Learning

Deep learning is a class of machine learning algorithms that uses multiple layers to progressively extract higher level features from the raw input. For example, in image processing, lower layers may identify edges, while higher layers may identify the concepts relevant to a human such as digits or letters or faces. Deep learning (also known as deep structured learning) is part of a broader family of machine learning methods based on artificial neural networks. Learning can be supervised, semi-supervised or unsupervised.

Deep learning architectures such as deep neural networks, convolutional neural networks and recurrent neural networks have been applied to fields including computer vision, image recognition, speech recognition, natural language processing, audio recognition, social network filtering etc.

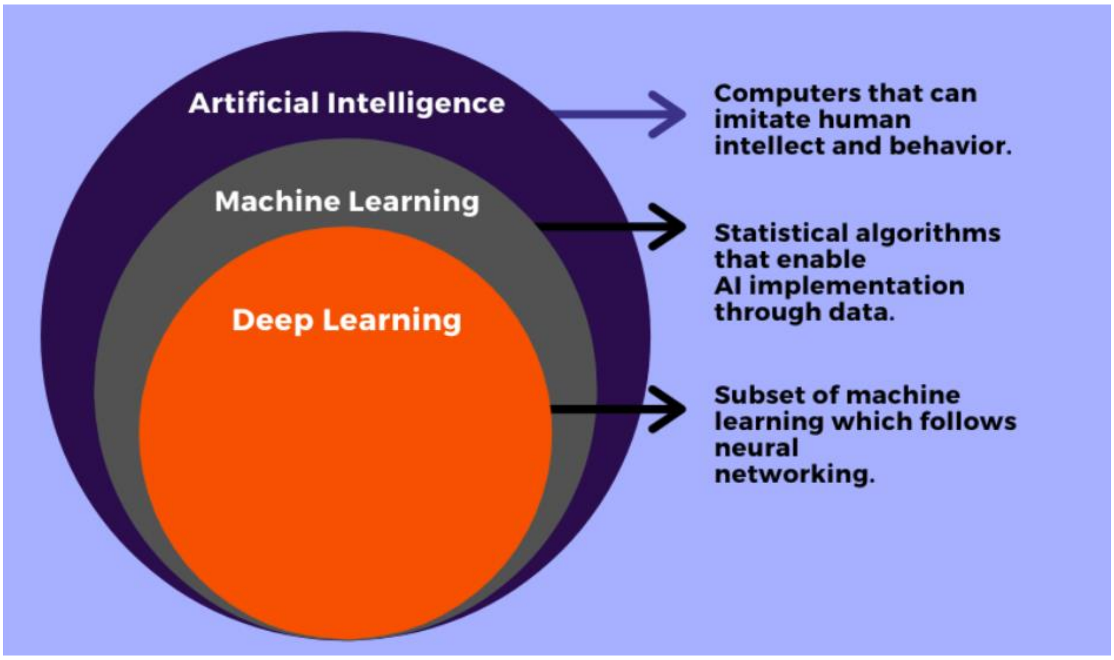


Fig. 9 Relation between AI, ML and DL

5.3 Neural Networks

Neural networks are a series of algorithms that mimic the operations of a human brain to recognize relationships between vast amounts of data. A neural network is a series of algorithms that endeavors to recognize underlying relationships in a set of data through a process that mimics the way the human brain operates. A “neuron” in a neural network is a mathematical function that collects and classifies information according to a specific architecture. A neural network contains layers of interconnected nodes. Each node is a perceptron and is similar to a multiple linear regression. The perceptron feeds the signal produced by a multiple linear regression into an activation function that may be nonlinear. Hidden layers fine-tune the input weightings until the neural network's margin of error is minimal. It is hypothesized that hidden layers extrapolate salient features in the input data that have predictive power regarding the outputs. This describes feature extraction, which accomplishes a utility similar to statistical techniques such as principal component analysis.

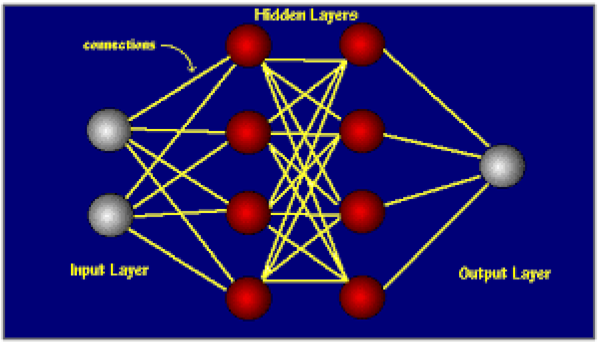


Fig. 10 Simple Neural Network

5.4 Artificial Neural Networks

An Artificial Neural Network in the field of Artificial intelligence where it attempts to mimic the network of neurons makes up a human brain so that computers will have an option to understand things and make decisions in a human-like manner. The artificial neural network is designed by programming computers to behave simply like interconnected brain cells.

There are around 1000 billion neurons in the human brain. Each neuron has an association point somewhere in the range of 1,000 and 100,000. In the human brain, data is stored in such a manner as to be distributed, and we can extract more than one piece of this data, when necessary, from our memory parallelly. We can say that the human brain is made up of incredibly amazing parallel processors.

We can understand the artificial neural network with an example, consider an example of a digital logic gate that takes an input and gives an output. "OR" gate, which takes two inputs. If one or both the inputs are "On," then we get "On" in output. If both the inputs are "Off," then we get "Off" in output. Here the output depends upon input. Our brain does not perform the same task. The outputs to inputs relationship keep changing because of the neurons in our brain, which are "learning."

5.5 Python

Python is an interpreted, high-level, general-purpose programming language. Created by Guido van Rossum and first released in 1991, Python's design philosophy emphasizes code readability with its notable use significant whitespace. Its language constructs and object-oriented approach aim to help programmers write clear, logical code for small and large-scale projects.

Python is dynamically typed and garbage collected. It supports multiple programming paradigms including structured, object-oriented and functional programming. Python is often described as a "batteries language due to its comprehensive standard library.

Python interpreters are available for many operating systems. A global community of programmers develops and maintains C Python an open sources implementation. A non-profit organization, the Python Software Foundation, manages and directs resources for Python and C Python development.

From development to deployment and maintenance, Python helps developers be productive and confident about the software they're building. Benefits that make Python the best fit for machine learning and AI-based projects include simplicity and consistency, access to great libraries and frameworks for AI and machine learning (ML), flexibility, platform independence, and a wide community. These add to the overall popularity of the language.

Chapter 6: ARCHITECTURE OF THE SYSTEM

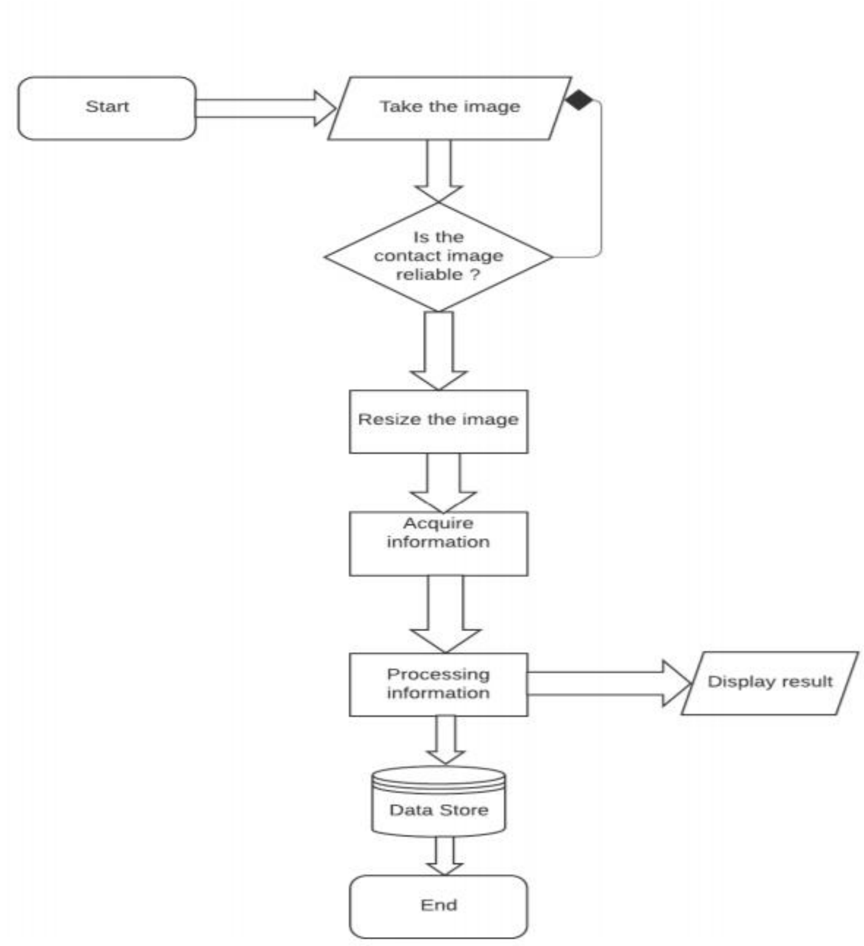


Fig. 11 Model flow chart

Chapter 7: MATERIALS AND METHODS

7.1.Dataset

The dataset is prepared from mature tea leaves obtained from Kokrajhar Tea Estate for the dataset preparation. The dataset contains 1364 rows and 10 columns mean_r , mean_g , mean_b , stddev_r , stddev_g , stddev_b , variance , kurtosis, skewness and Chlorophyll value.

After obtaining the tea leaves following steps are executed:

1. Contact images are taken using a Samsung smartphone camera along with a high intensity beam of torch light and chlorophyll value of the same is taken with SPAD and recorded into an CSV file.

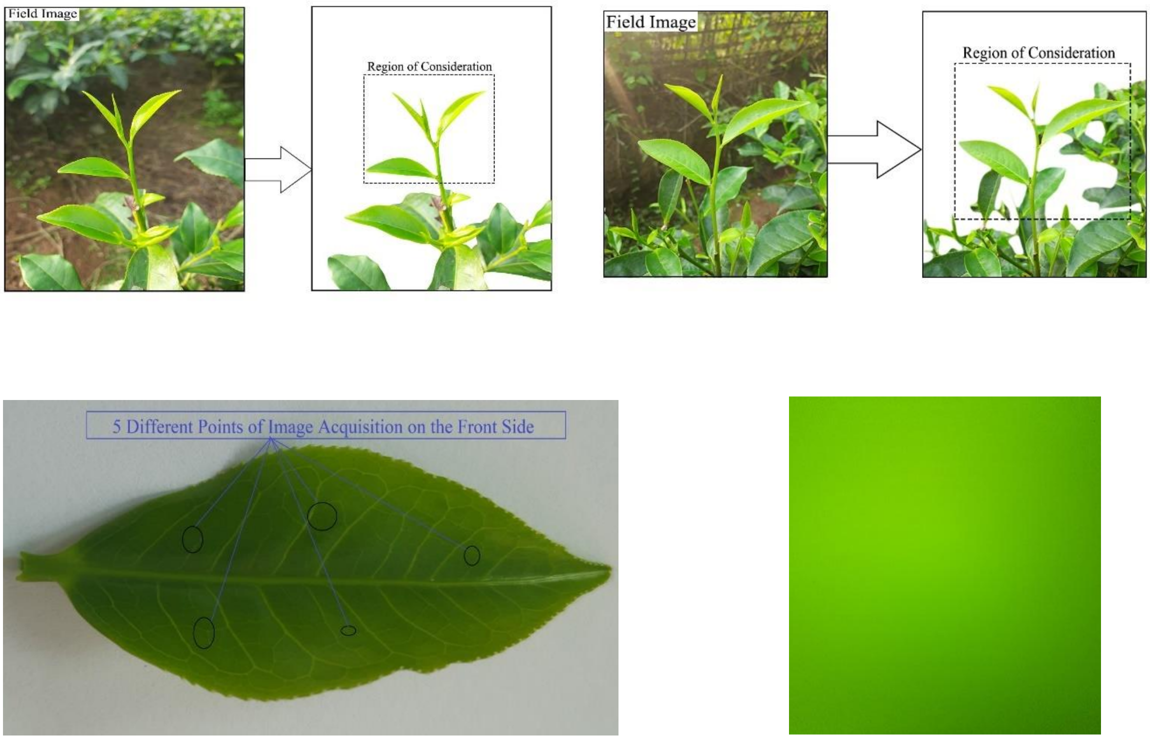


Fig. 12 Acquisition of Contact images

2. Then, different colour features are extracted from the contact images using feature extraction methods available in OpenCV.

```
    red_mean = np.mean(red_channel)
    green_mean = np.mean(green_channel)
    blue_mean = np.mean(blue_channel)

  #  h_mean = np.mean(h_channel)
      s_mean = np.mean(s_channel)
      v_mean = np.mean(v_channel)


    red_std = np.std(red_channel)
    green_std = np.std(green_channel)
    blue_std = np.std(blue_channel)

    h_std = np.std(h_channel)
    s_std = np.std(s_channel)
    v_std = np.std(v_channel)

    kur = kurtosis(img, axis=None)
    sk = skew(img, axis=None)
    var = np.var(img, axis=None)
    mom = moment(img, moment=1)
    histr = cv2.calcHist([img],[0],None,[256],[0,256])

    vector = [red_mean,green_mean,blue_mean,red_std,green_std,blue_std,var,h_std,s_std,v_std,
             kur,sk,mom,histr]

    df_temp = pd.DataFrame([vector],columns=names)
    df = df.append(df_temp)
      df = sorted(df)
    print(file)
return df
```

Fig. 13 Code snippet for Feature extraction.

```
def create_dataset():

    names = ['mean_r','mean_g','mean_b','stddev_r','stddev_g','stddev_b','variance', 'kurtosis' ,'skewness'  ]

    df = pd.DataFrame([], columns = names)

    for file in img_files:
        imgpath = ds_path + "\\" + file
        main_img = cv2.imread(imgpath)

        #Preprocessing
        img = cv2.cvtColor(main_img, cv2.COLOR_BGR2RGB)
        gs = cv2.cvtColor(img,cv2.COLOR_RGB2GRAY)
        hsv_img = cv2.cvtColor(img, cv2.COLOR_BGR2HSV)
        blur = cv2.GaussianBlur(gs, (25,25),0)
        ret_otsu,im_bw_otsu = cv2.threshold(blur,0,255,cv2.THRESH_BINARY_INV+cv2.THRESH_OTSU)
        kernel = np.ones((50,50),np.uint8)
        closing = cv2.morphologyEx(im_bw_otsu, cv2.MORPH_CLOSE, kernel)

        #Color features
        red_channel = img[:,:,0]
        green_channel = img[:,:,1]
        blue_channel = img[:,:,2]
        blue_channel[blue_channel == 255] = 0
        green_channel[green_channel == 255] = 0
        red_channel[red_channel == 255] = 0

        h_channel = hsv_img[:, :, 0]
        s_channel = hsv_img[:, :, 1]
        v_channel = hsv_img[:, :, 2]

        red_mean = np.mean(red_channel)
        green_mean = np.mean(green_channel)
        blue_mean = np.mean(blue_channel)
```

Fig. 14 Code snippet for Feature extraction.

3. Different combinations of features are formed out to better study the correlation between the feature and the value of chlorophyll.
   The combinations are R+G+B, (G-B)/(G+B), (G-R)/(G+R), R/(R+G+B), G/(R+G+B), G/(R+G+B) and (R-B)/(R+B).

7.2. Multiple Linear Regression

Multiple linear regression (MLR), also known simply as multiple regression, is a statistical technique that uses several explanatory variables to predict the outcome of a response variable. The goal of multiple linear regression (MLR) is to model the linear relationship between the explanatory (independent) variables and response (dependent) variable.

First, we fit the features of the tea leaf images with the SPAD calculated chlorophyll values. Then we perform MLR taking the combination of features. And hence predict the value of chlorophyll.

7.3 SVR

In machine learning, support-vector machines (SVMs, also support-vector networks) are supervised learning models with associated learning algorithms that analyse data for classification and regression analysis. When data are unlabelled, supervised learning is not possible, and an unsupervised learning approach is required, which attempts to find natural clustering of the data to groups, and then map new data to these formed groups. The support-vector clustering algorithm, created by Hava Siegelmann and Vladimir Vapnik, applies the statistics of support vectors, developed in the support vector machines algorithm, to categorize unlabelled data, and is one of the most widely used clustering algorithms in industrial applications.

7.4 Deep Learning model

There are many models widely used in deep learning. In this study, we have used an artificial neural network model and 1-D Convolution neural network.

7.4.1 Artificial Neural Network Structure

The neural network used here has 12 inputs, 4 dense layers of output shapes (128, 64, 32, 1). The activation function used in the first three dense layers are ELU and the fourth layer uses Linear activation function. The loss function used in the neural network is the Mean Squared Error (MSE) and the optimizer used here is AdaMax [2]. AdaMax is an extension to the Adam version of gradient descent that generalizes the approach to the infinite norm (max). It is designed to accelerate the optimization process. The above neural network is then fitted with the training set with the batch size of 64 and number of epochs equal to 10.

7.4.1 Training our deep learning model

Artificial Neural Network: Training is basically usage of a set of examples for learning, that is to fit the parameters of the classifier. The model has been trained using 1267 images of tea leaves. The neural network is then fitted with the training set with the batch size of 64 and number of epochs equal to 10.

Chapter 8: DISCUSSIONS AND RESULTS

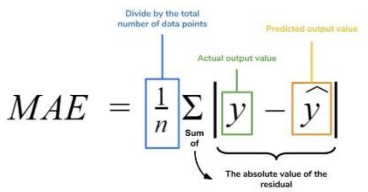
8.1 Matrix evaluation

The idea of building machine learning model works on a constructive feedback principle. We build a model, get feedback from metrics, make improvements and continue until we achieve a desirable accuracy. Evaluation metrics explain the performance of a model. An important aspect of evaluation metrics is their capability to discriminate among model results. Many analysts and aspiring data scientists do not even bother to check how robust their model is. Once they are finished building a model, they hurriedly map predicted values on unseen data. This is an incorrect approach. Simply building a predictive model is not your motive. It’s about creating and selecting a model which gives high accuracy out of sample data. Hence, it is crucial to check the accuracy of your model prior to computing predicted values. The types of evaluation metrics used in the project are: -

1. Mean Absolute Error
2. Mean Squared Error
3. Root Mean Squared Error

8.1.1 Mean Absolute Error

The MAE measures the average magnitude of the errors in a set of forecasts, without considering their direction. It measures accuracy for continuous variables. The equation is given in the library references. Expressed in words, the MAE is the average over the verification sample of the absolute values of the differences between forecast and the corresponding observation. The MAE is a linear score which means that all the individual differences are weighted equally in the average.



8.1.2 Mean Square Error

The mean squared error (MSE) tells you how close a regression line is to a set of points. It does this by taking the distances from the points to the regression line (these distances are the "errors") and squaring them. The squaring is necessary to remove any negative signs. It also gives more weight to larger differences. It's called the mean squared error as you're finding the average of a set of errors. The lower the MSE, the better the forecast.

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \tilde{y}_i)^2$$

8.1.3 Root Mean Square Error

The RMSE is a quadratic scoring rule which measures the average magnitude of the error. The equation for the RMSE is given in both of the references. Expressing the formula in words, the difference between forecast and corresponding observed values are each squared and then averaged over the sample. Finally, the square root of the average is taken. Since the errors are squared before they are averaged, the RMSE gives a relatively high weight to large errors. This means the RMSE is most useful when large errors are particularly undesirable.

$$RMSE = \sqrt{\sum_{i=1}^{n}\frac{(\hat{y}_i - y_i)^2}{n}}$$

8.2 Results

8.2.1 Linear Regression

The results obtained while predicting the chlorophyll value is satisfactory. The following table shows the regression scores of the features.

Features	Regression score
Mean of R	0.024
Mean of G	0.028
Mean of B	0.026
R+G+B	0.026
(G-B)/(G+B)	0.001
(G-R)/(G+R)	0.0006
R/(R+G+B)	0.032
G/(R+G+B)	0.0004
B/(R+G+B)	0.002
(R-B)/(R+B)	0.004

Table 8.1 Regression scores of the features

The following graph shows the actual and the respective predicted values of chlorophyll.

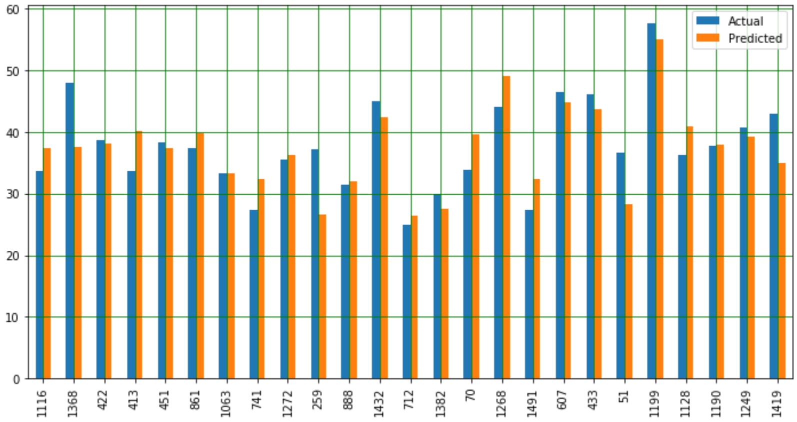


Fig. 16 Actual vs Predicted Chlorophyll values using MLR

The values of MAE, MSE, RMSE and R^2 for LR algorithm are 3.591, 23.533, 4.851 and 0.711 respectively.

8.2.2 Support Vector Regression

The chlorophyll estimation is also done using SVR algorithm. The 12 features of the tea leaf images are first fitted with the actual SPAD calculated chlorophyll values. SVR is built based on the concept of Support Vector Machine or SVM. It is useful in assigning classes when the data is not linearly separable. The results obtained while predicting the chlorophyll value is satisfactory. The following table shows the regression scores of the features.

Actual	Predicted
33.6	32.606
47.9	37.438
38.6	34.767
33.6	37.528
38.3	34.705
37.4	36.706
33.3	32.579

Table 8.2 Actual vs Predicted Chlorophyll values using SVR

The following graph shows the actual and the respective predicted values of chlorophyll.

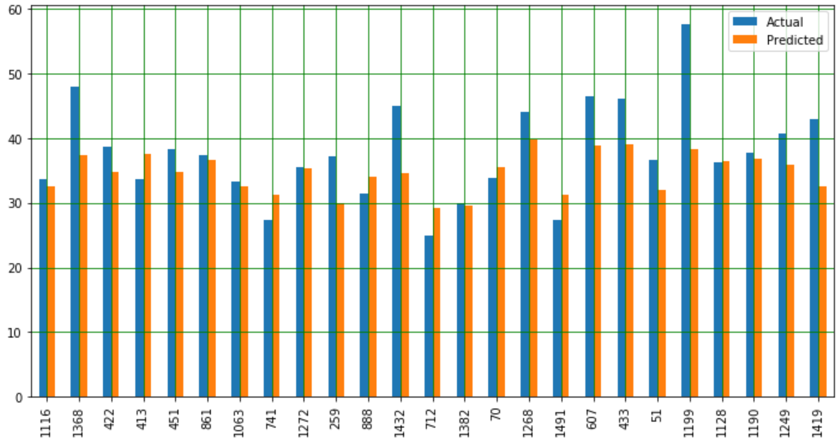


Fig. 17 Actual vs Predicted Chlorophyll values using SVR

The values of MAE, MSE, RMSE and R^2 for SVR are 5.274, 55.904, 7.476 and 0.314 respectively.

8.2.3 Artificial Neural Network

The values of MAE, MSE, RMSE and R^2 obtained for ANN are 4.443, 23.533, 4.851 and 0.711.

The following table shows the comparison between the evaluation parameters of the methods.

Algorithms	**MAE**	**MSE**	**RMSE**	**R^2**
LR	3.591	23.533	4.851	0.711
SVR	5.274	55.904	7.476	0.314
ANN	4.443	23.533	4.851	0.711

Table 8.3 Evaluation metrics for LR, SVR and ANN

Chapter 9: APPLICATIONS

- **Tea plantation:** The chlorophyll detection by deep learning method can be used to lower the dependency of traditional methods of chlorophyll detection in tea plantation. This will help lower the expense of procuring expensive tools like soil plant analysis development meter or spectrophotometer. This ensures that our method can be useful to the low scale farmers.

- **Research field:** This method can help in research of detecting parameters that is most likely to be the deciding factor of an important feature of any other crop.

- **Organise the tea plantation:** This method can help in organising the process of tea plantation. The need for quality tea leaves can be made available more efficiently. Our research will help the small-scale framers to better understand their plant health and hence be more aware and responsible of their crops.

Chapter 10: ADVANTAGES

- **Economical traditional method**: In the traditional methods, the chlorophyll of leaf can be determined using soil plant analysis development meter or spectrophotometer. This is an expensive method and cannot is not easily available to a small scale farmer or an agricultural expert. Our method of detecting chlorophyll using Deep Learning cuts the prices of the above mentioned tools.
- **Readability of large datasets**: In our project Deep Learning has been used which out performs other techniques as the data size is large**.** This approach is effective for new applications, or for applications which will have a relatively big number of output categories. This is a relatively less popular approach because, with the rate of learning and large volumes of data, the networks typically take significantly more time to train but provides faster test results.
- **Removes the need to worry about feature extraction:** When there is lack of domain understanding for feature introspection, Deep Learning techniques outshines others as we have to worry less about feature engineering. In traditional Machine learning techniques, most of the applied features need to be identified by a domain expert in order to reduce the complexity of the data and make patterns more visible to learning algorithms to work. The biggest advantage Deep Learning algorithms is that they try to learn high-level features from data in an incremental manner thus eliminating the need of domain expertise and hardcore feature extraction.

Chapter 11: LIMITATIONS

The limitations of the project are as follows:

- It requires very large amount of data in order to perform better than other techniques.
- While training the model takes lot amount of time.
- The chlorophyll is changed due to the geographic conditions. With the change of chlorophyll, the colour indexes of the leaf are also changed because the colour of the leaf is highly correlated to the chlorophyll of the leaf.

Chapter 12: CONCLUSION

In this project we used three Machine Learning models namely, Linear Regression, Support Vector Regression and Artificial Neural Network. We evaluated the models using evaluation metrics like MAE, MSE and RMSE. The methods that we applied in this project can be proved to be very efficient in detecting chlorophyll using deep learning. The contact imaging used for data collection is a low-cost mechanism and because of the contact imaging technique, the results obtained are reliable and accurate. After studying all the evaluation metrics the Artificial network has proved to be the most accurate to estimate the chlorophyll value. The LR and SVR has too been able to improve the accuracy.

BIBLIOGRAPHY

1. Ali, M.M., Al-Ani, A., Eamus, D., Tan, D.K., 2012. A new image processing based technique to determine chlorophyll in plants. Am.-Eurasian J. Agricult. Environ. Sci. 12, 1323–1328.
2. Barman, U., Choudhury, R.D., 2019. Soil texture classification using multi class support vector machine. Information Processing in Agriculture.
3. Barman, U., Choudhury, R.D., 2019. Prediction of Soil pH using Smartphone based Digital Image Processing and Prediction Algorithm.
4. Barman, U., Choudhury, R.D., 2019. Bacterial and Virus affected Citrus Leaf Disease Classification using Smartphone and SVM.
5. Barman, U., Choudhury, R.D., Saud, A., Dey, S., Kumar, B., 2018. Estimation of Chlorophyll Using Image Processing. Int. J. Recent Sci. Res. 9, 24850–24853. Cortazar, B., Koydemir, H.C., Tseng, D., Feng, S., Ozcan, A., 2015. Quantification of plant chlorophyll content using Google Glass. Lab Chip 15, 1708–1716.
6. Dey, A.K., Sharma, M., Meshram, M.R., 2016. An analysis of leaf chlorophyll measurement method using chlorophyll meter and image processing technique. Procedia Comput. Sci. 85, 286–292.
7. Friedrichs, A., Busch, J., van der Woerd, H., Zielinski, O., 2017. SmartFluo: a method and affordable adapter to measure chlorophyll a fluorescence with smartphones. Sensors 17, 678.
8. Gaviria-Palacio, D., Guáqueta-Restrepo, J.J., Pineda-Tobón, D.M., Pérez, J.C., 2017. Fast estimation of chlorophyll content on plant leaves using the light sensor of a smartphone. Dyna 84, 234–239.