Namal University Mianwali

Department of Computer Science

Soil analysis using machine learning

Student:
Mahmood Yousaf
(2018-uet-nml-cs-11)

Supervisor:
Dr. Malik Jahan Khan

January 15, 2022

Contents

1 Terminologies

Following terminologies will be used in the document.

- Convolutional neural network (CNN)
- Artificial neural network (ANN)
- Phosphorus (P)
- Organic matter (OM)
- Electrical conductivity (EC)

2 Abstract

Agriculture is the backbone of Pakistan economy and millions of peoples are directly or indirectly linked with agriculture sector.The main source for agriculture is soil for crops production.If soil is healthy than crops productions automatically boosts up.But in Pakistan average farmer cannot afford to test their soils in the laboratory which is quite time consuming and costs a lot.So in this work, we trying to facilitate the farmers with application which can check the quality of soil by using their mobile phone images and recommend different fertilizers according to the results of soil sample. About 1064 images are used for machine learning models training with laboratory tested labels.Different machine learning models like ANN, CNN with regression and classification models are trained on images.

3 Literature review

In the year 2014 [3], the author Vinay Kumar determined soil PH values by using digital image processing technique.In this work fifty soil samples were collected and determined their PH values using PH meter.Digital camera was used for sampling the soil images. By using below formula, PH index is calculated.

PH Index = (Red / Green) / Blue [3].

After calculating PH Index, correlation between PH values using PH meter and PH index by using RGB images are determined. [3].

In the year 2016 [4], the author CS ManikandaBabu determined physical and chemical properties of soil by using digital image analysis.For soil sampling, digital camera was used.They used fractal dimension calculation using box counting method to obtain physical recognition.They determined soil PH by using RGB colors.By using plane extraction method, each plane value is calculated and PH index is calculated.By using PH index and color of soil, PH of soil sample was calculated [4]

In the year 2019, the author Maneesha G Nair [2] proposed android application which was capable to use digital image processing. By capturing image of soil, soil PH value was determined and crop suggestion as output.After prepossessing,RGB values are extracted from image to calculate PH index. PH index value was compared with stored PH value in database. [2]

4 Project source code

Project source code which is available on the github. Source Code

5 Introduction

Pakistan economy is heavily based on agriculture.Total yield of crops is based on the quality of soil.If soil nutrients are sufficiently available than crops production automatically boosts up. Pakistani agriculture industry holds about 18.9 percents of GDP and creates about 43.3 percent jobs for labours [5].The main aim of this work to check quality of soil using machine learning.

The main motivation for using machine learning as alternative of laboratory testing is that laboratory testing of soil is time and cost consuming. Moreover farmer cannot afford such costly laboratory testing for soil because they are limited in their resources.We are developing mobile application for farmers which will take image in the input and will provide the report in the local language which will contain different soil nutrients values and recommendation of fertilizer according to the generated report.

We have 701 soil samples. Soil samples values are measured in the soil laboratory by expert using different soil laboratory tools.There are four features measured in the laboratory like pH value, phosphorous (P), electrical conductivity (EC) and organic matter (OM).There are also 701 soil samples images are available which were captured during the collection of soil samples in the field.

Using images of soil samples, different machine learning models are trained to predict soil pH values,phosphorous (P), organic matter (OM) and electrical conductivity (EC).On the basis of predicted values of soil by machine learning models, mobile application will generate report and will suggest different fertilizers for farmer which will improve the quality of soil to get maximum yield.

5.1 Objectives

- Using machine learning to check the quality of soil
- Checking different machine learning models performance
- Recommendation system for farmers which will recommend fertilizer

6 Methodology

6.1 Dataset

We have 1064 soil images which are captured by using smartphone camera.Two images per sample are captured like outer surface image and inner surface image.Location for soil sampling are Khusab,Talagang and Mianwali.Below is the sample image captured using mobile camera.

Figure 1: Soil image captured using mobile camera

6.2 Pre-Processing

All images are captured at uncontrolled field conditions.So there is higher chances of noise in the dataset in form of light,some unwanted corners etc.In the image pre-processing, all images are resized using opencv.

6.2.1 Region of interest

After resizing, specific part of image is extracted to get region of interest (RIO).

Figure 2: Region of interest extraction

6.2.2 Applying the sharpening kernel

After extraction of region of interest (RIO), sharpening kernel is applied because there are some images which have blurr effect, contrast and shade issues. In order to tackle such issues sharpening kernel is applied on all images [1].

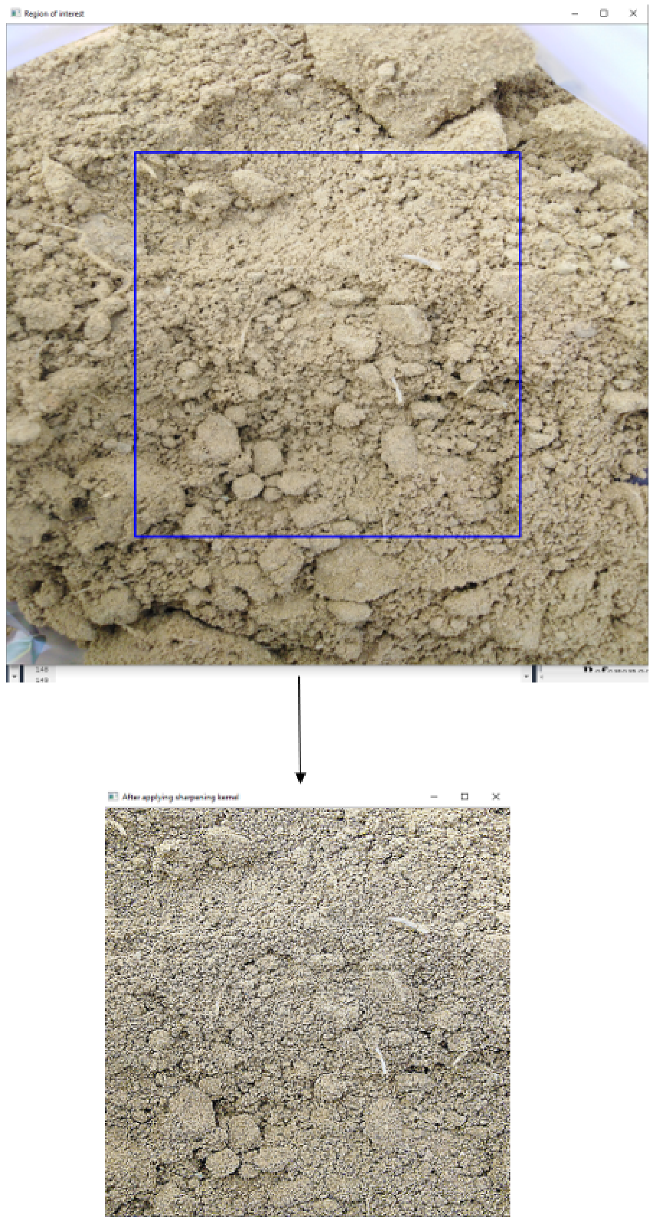

Figure 3: Image Sharpening

6.2.3 pH indexes extraction

By using the following formula, pH index is extracted for each image.

PH Index = (Red / Green) / Blue [3].

6.2.4 Features extraction

By using the following formula, feature is extracted for each image.

x = median(Green) + median(Blue)) + median(Red)
feature = mean(x)

Feature for each image which will be used for machine learning models training and testing.Above formula is derived after experimentation .

6.3 ANN regression model on pH indexes

After pH index extraction for each image,each pH index is labeled with actual pH value and passed with the following ANN architecture for training and testing.

```
Model: "sequential_4"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
dense_15 (Dense)             (None, 128)               256
_________________________________________________________________
dense_16 (Dense)             (None, 256)               33024
_________________________________________________________________
dense_17 (Dense)             (None, 256)               65792
_________________________________________________________________
dense_18 (Dense)             (None, 256)               65792
_________________________________________________________________
dense_19 (Dense)             (None, 1)                 257
=================================================================
Total params: 165,121
Trainable params: 165,121
Non-trainable params: 0
_________________________________________________________________
```

Figure 4: ANN architecture

Similarly pH index is labeled with other soil parameters like OM, EC and P separately to perform estimation on same ANN regression model.

6.4 ANN Regression Model on features

After feature extraction for each image by using experimentation based formula,each feature is labeled with pH, OM, P and EC separately.After labeling, they are passed separately through ANN regression model 7 for training and testing.

6.5 CNN regression model on images

CNN provides feature extraction automatically.So pre-processed images are labeled with actual pH, OM, P and EC values separately because every parameter will be predicted by separate CNN model.AlexNet [6] CNN architecture is used for training and testing.AlexNet is a very famous architecture who was competed in the Image-Net Large Scale Visual Recognition Challenge on Sept 30,2012 [6].Following is the model used for predictions.

```
Model: "sequential_1"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
conv2d_5 (Conv2D)            (None, 73, 73, 96)        34944
_________________________________________________________________
batch_normalization_5 (Batch (None, 73, 73, 96)        384
_________________________________________________________________
max_pooling2d_3 (MaxPooling2 (None, 36, 36, 96)        0
_________________________________________________________________
conv2d_6 (Conv2D)            (None, 36, 36, 256)       614656
_________________________________________________________________
batch_normalization_6 (Batch (None, 36, 36, 256)       1024
_________________________________________________________________
max_pooling2d_4 (MaxPooling2 (None, 17, 17, 256)       0
_________________________________________________________________
conv2d_7 (Conv2D)            (None, 17, 17, 384)       885120
_________________________________________________________________
batch_normalization_7 (Batch (None, 17, 17, 384)       1536
_________________________________________________________________
conv2d_8 (Conv2D)            (None, 17, 17, 384)       1327488
_________________________________________________________________
batch_normalization_8 (Batch (None, 17, 17, 384)       1536
_________________________________________________________________
conv2d_9 (Conv2D)            (None, 17, 17, 256)       884992
_________________________________________________________________
batch_normalization_9 (Batch (None, 17, 17, 256)       1024
_________________________________________________________________
max_pooling2d_5 (MaxPooling2 (None, 8, 8, 256)         0
_________________________________________________________________
flatten_1 (Flatten)          (None, 16384)             0
_________________________________________________________________
dense_3 (Dense)              (None, 4096)              67112960
_________________________________________________________________
dropout_2 (Dropout)          (None, 4096)              0
_________________________________________________________________
dense_4 (Dense)              (None, 4096)              16781312
_________________________________________________________________
dropout_3 (Dropout)          (None, 4096)              0
_________________________________________________________________
dense_5 (Dense)              (None, 1)                 4097
=================================================================
Total params: 87,651,073
Trainable params: 87,648,321
Non-trainable params: 2,752
_________________________________________________________________
```

Figure 5: AlexNet CNN architecture

6.6 Discretization for classification models

For the classification models, soil parameters like pH, OM, P and EC are classified by using soil expert opinion.For example, if soil pH is less than 7.0,so it belongs to class1 and if soil pH is greater than 7.0 and less than 7.5 than it belongs to class2. Similarly all soil parameters are classified by following table .

.

OM Classification	
OM	Class
<0.865	OM1
0.87 -- 1.29	OM2
> 1.29	OM3

pH Classification	
pH	Class
<7.0	pH1
7.0 -- 7.5	pH2
7.6 -- 8.0	pH3
8.1 -- 8.5	pH4
> 8.5	pH5

EC Classification	
EC	Class
<4.0	EC1
4.1 -- 8.0	EC2
8.1 -- 16.0	EC3
> 16.0	EC4

P Classification	
P	Class
<3.5	P1
3.6 -- 7.0	P2
7.0 -- 14.0	P3
> 14.0	P4

Figure 6: Discretization Process

6.7 CNN Classification Model

After discretization process, AlexNet [6] CNN architecture is used for training and testing.Following is the model network architecture.

```
Model: "sequential"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
conv2d (Conv2D)              (None, 73, 73, 96)        34944
_________________________________________________________________
batch_normalization (BatchNo (None, 73, 73, 96)        384
_________________________________________________________________
max_pooling2d (MaxPooling2D) (None, 36, 36, 96)        0
_________________________________________________________________
conv2d_1 (Conv2D)            (None, 36, 36, 256)       614656
_________________________________________________________________
batch_normalization_1 (Batch (None, 36, 36, 256)       1024
_________________________________________________________________
max_pooling2d_1 (MaxPooling2 (None, 17, 17, 256)       0
_________________________________________________________________
conv2d_2 (Conv2D)            (None, 17, 17, 384)       885120
_________________________________________________________________
batch_normalization_2 (Batch (None, 17, 17, 384)       1536
_________________________________________________________________
conv2d_3 (Conv2D)            (None, 17, 17, 384)       1327488
_________________________________________________________________
batch_normalization_3 (Batch (None, 17, 17, 384)       1536
_________________________________________________________________
conv2d_4 (Conv2D)            (None, 17, 17, 256)       884992
_________________________________________________________________
batch_normalization_4 (Batch (None, 17, 17, 256)       1024
_________________________________________________________________
max_pooling2d_2 (MaxPooling2 (None, 8, 8, 256)         0
_________________________________________________________________
flatten (Flatten)            (None, 16384)             0
_________________________________________________________________
dense (Dense)                (None, 4096)              67112960
_________________________________________________________________
dropout (Dropout)            (None, 4096)              0
_________________________________________________________________
dense_1 (Dense)              (None, 4096)              16781312
_________________________________________________________________
dropout_1 (Dropout)          (None, 4096)              0
_________________________________________________________________
dense_2 (Dense)              (None, 3)                 12291
=================================================================
Total params: 87,659,267
Trainable params: 87,656,515
Non-trainable params: 2,752
_________________________________________________________________
```

Figure 7: AlexNet CNN architecture for classification

7 Results

Following results which are obtained on unseen datasets after training the models.

7.1 pH, P, EC, and OM prediction on pH indexes

Following graphs showing that actual values of pH, P, EC and OM values along with predicted value of pH, P, EC and OM values by ANN regression model.

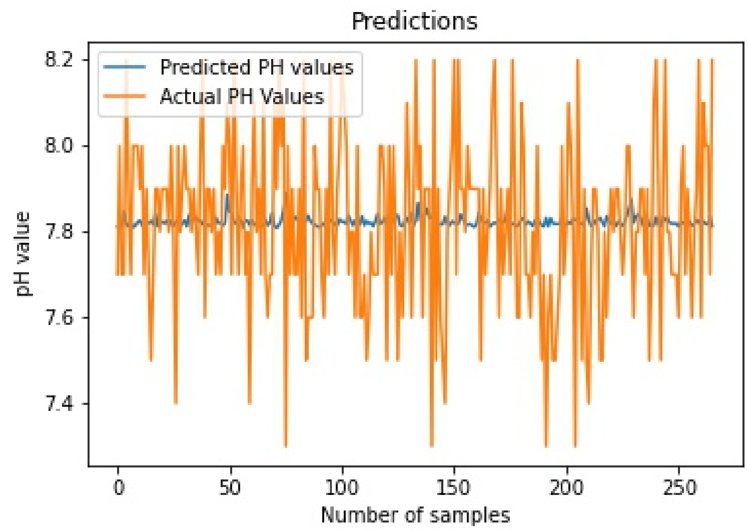

Figure 8: pH estimation using pH indexes

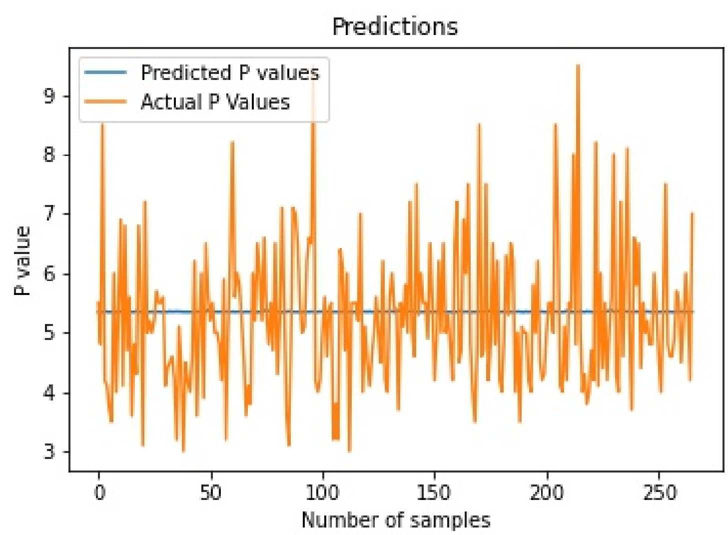

Figure 9: P estimation using pH indexes

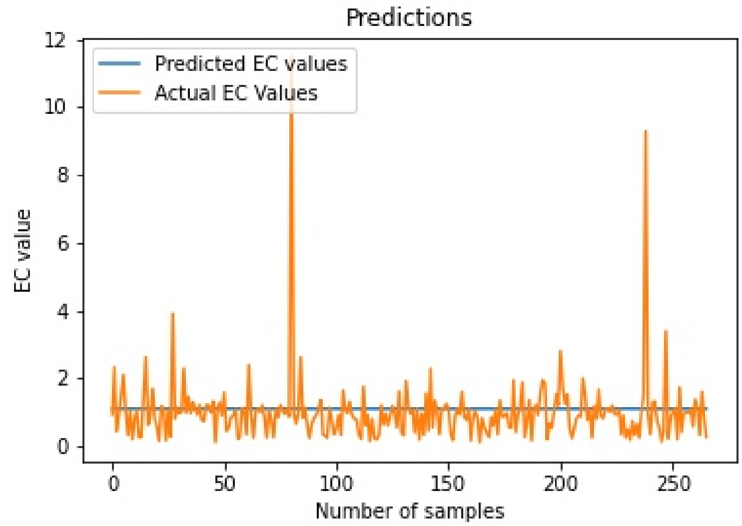

Figure 10: EC estimation using pH indexes

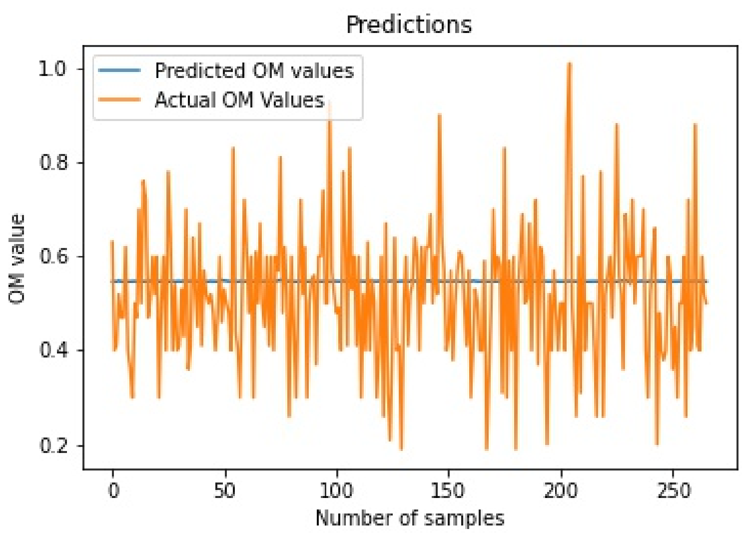

Figure 11: OM estimation using pH indexes

7.2 pH, P, EC, and OM prediction on experimentation based formula

Following graphs showing that actual values of pH, P, EC and OM values along with predicted value of pH, P, EC and OM values by ANN regression model.For this model features are extracted by experimentation based formula instead of pH indexes.

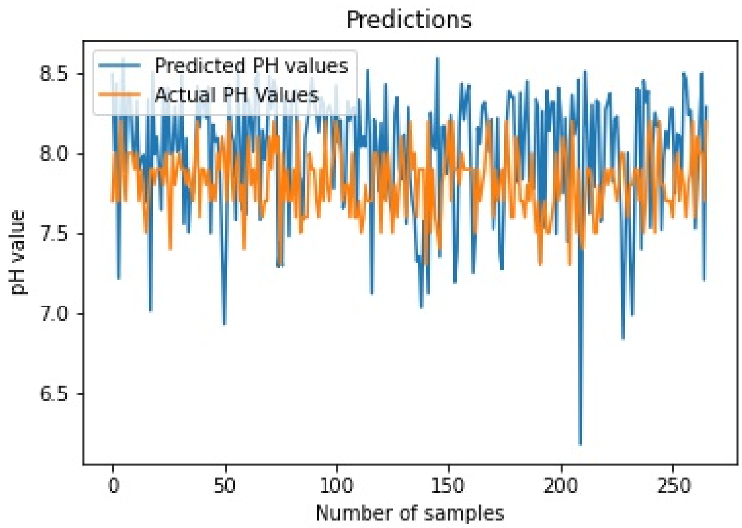

Figure 12: pH prediction using formula

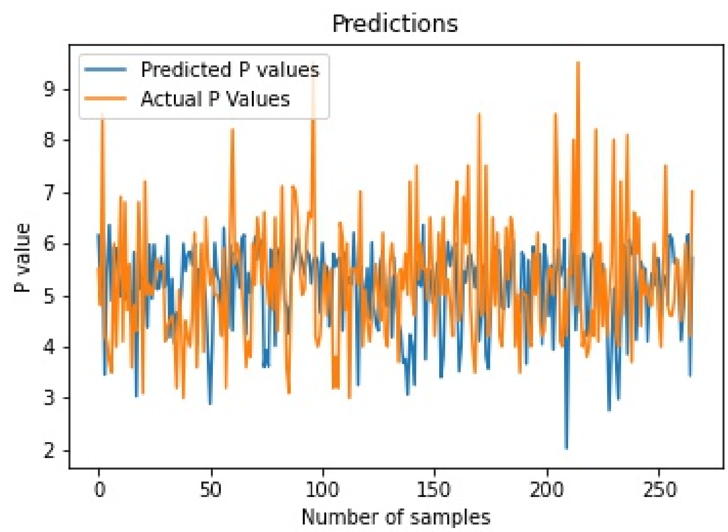

Figure 13: P prediction using formula

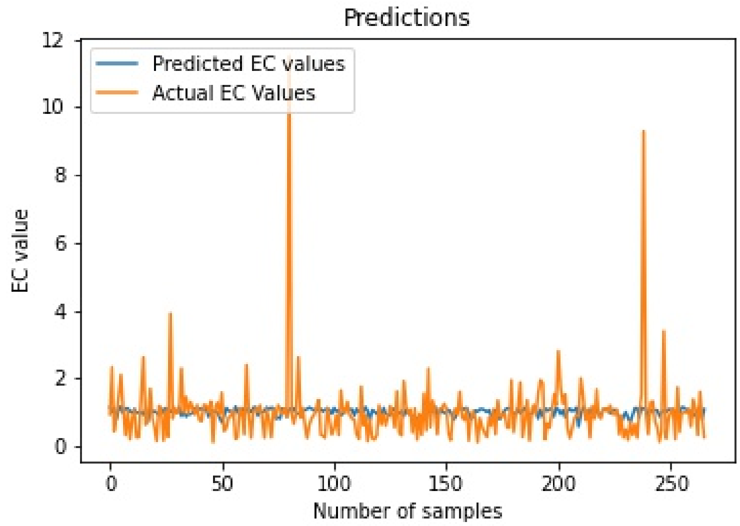

Figure 14: EC prediction using formula

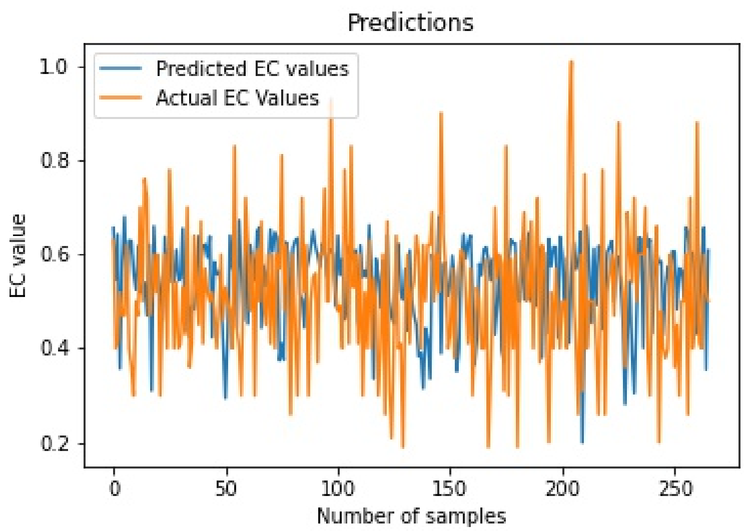

Figure 15: OM prediction using formula

7.3 pH, P, EC and OM prediction using CNN regression models

Following graphs showing that actual values of pH, P, EC and OM values along with predicted value of pH, P, EC and OM values by CNN regression models.All CNN regression models are trained using labeled images to estimate different parameter's of soil.

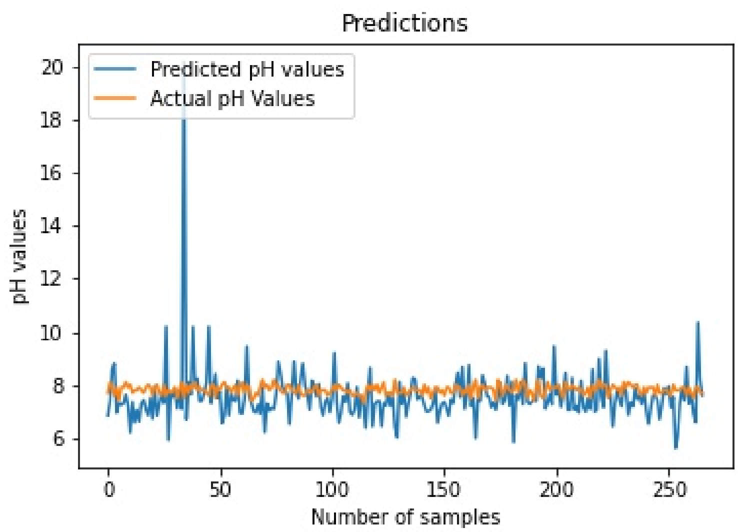

Figure 16: pH prediction using AlexNet CNN Regression model

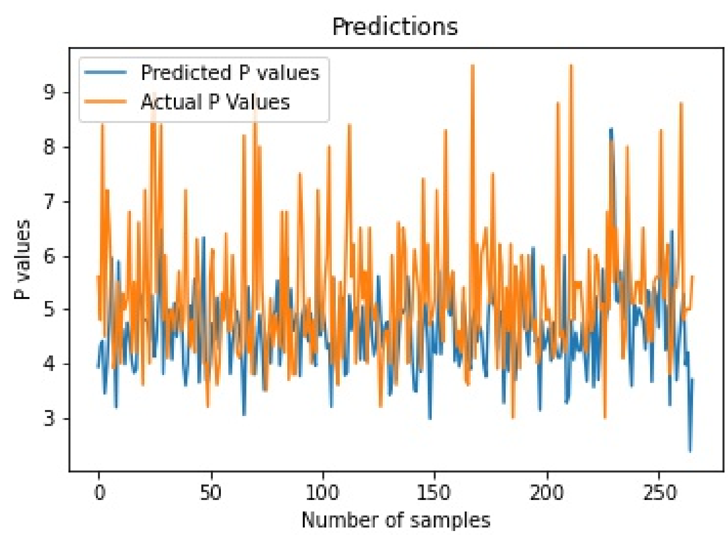

Figure 17: P prediction using AlexNet CNN Regression model

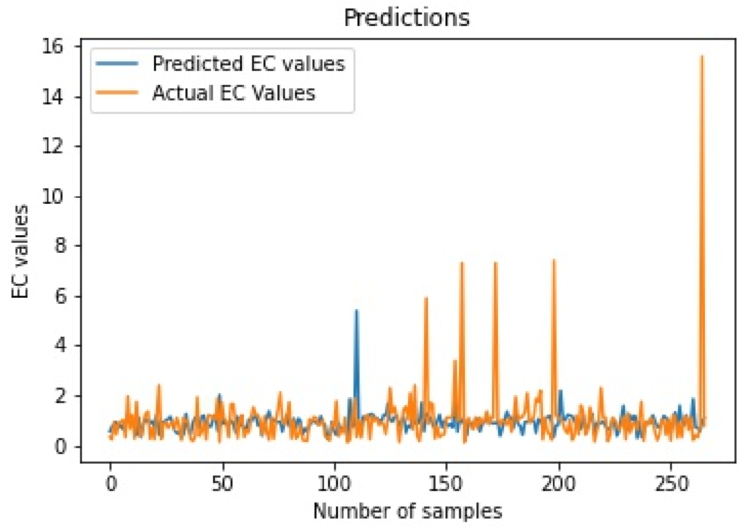

Figure 18: EC prediction using AlexNet CNN Regression model

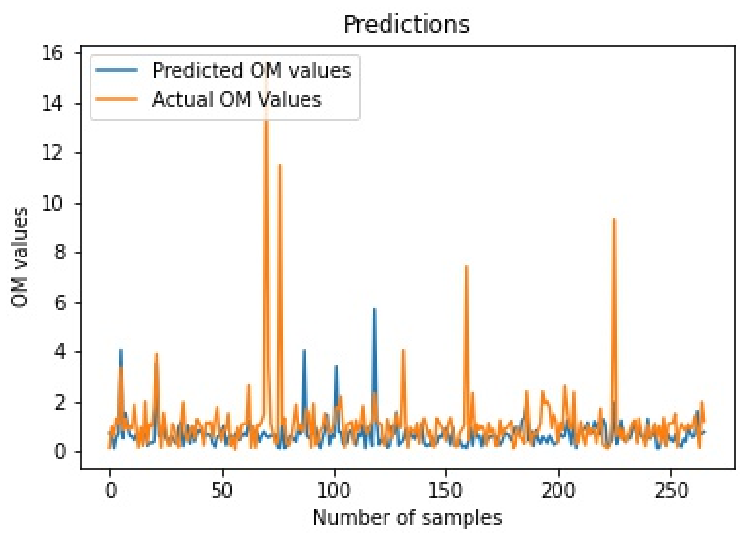

Figure 19: OM prediction using AlexNet CNN Regression model

7.4 pH, P, EC and OM prediction using CNN classification models

For evaluating the classification models, the most common measures which are used world widely are precision, recall and F1 score measures.Following are the results on different parameters of soil using classification models.

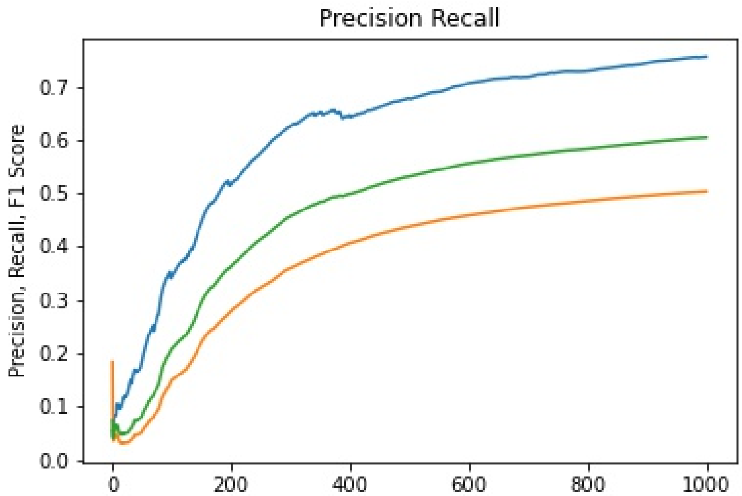

Figure 20: Precision, recall and F1 score measures for CNN classification model for pH value

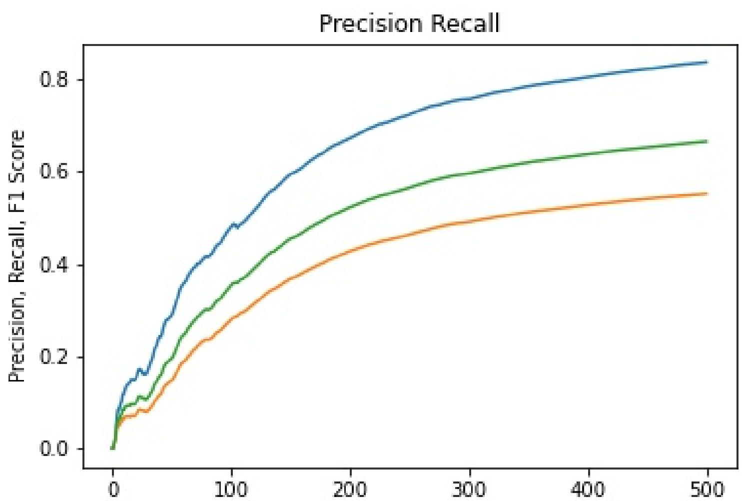

Figure 21: Precision, recall and F1 score measures for CNN classification model for P value

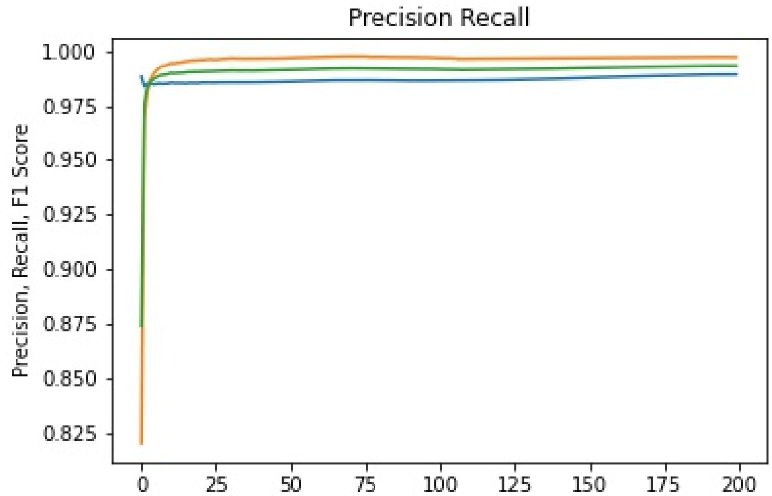

Figure 22: Precision, recall and F1 score measures for CNN classification model for EC value

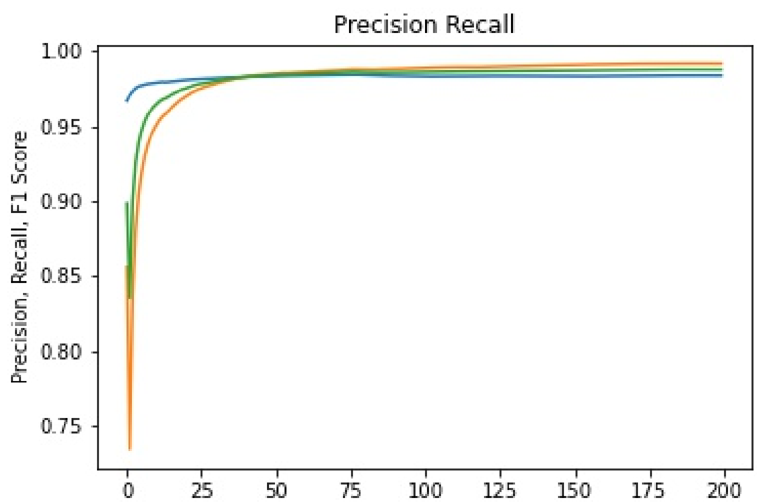

Figure 23: Precision, recall and F1 score measures for CNN classification model for OM value

8 Conclusion

pH index method [3] is failed badly on bigger dataset because they trained their machine learning models on 40 to 50 soil samples in a controlled lightening environment.But in our case, we trained all machine learning models on 1064 soil samples with laboratory tested results as labels.Some models performed very bad on different parameters of soil but some of them performed exceptionally well.Overall P results are quite satisfactory on majority machine learning models.

9 Acknowledgements

Before starting this project I was completely blank. But Dr. Malik Jahan Khan motivated me by sharing his experience and explaining the core of machine learning. I am very thankful to Dr. Malik Jahan Khan to share such amazing knowledge. The most important part is that Dr. Malik Jahan Khan is available at any time to me to discuss any kind of problems I was facing in the project.

References

[1] https://iq.opengenus.org/sharpening-filters.

[2] AV Guilalas, M Devoor, M Mizoguchi International Journal of . . . , and 2019. Determination of soil ph using digital image processing. 2019.

[3] Vinay Kumar, B. K. Vimal, Rakesh Kumar, and Mukesh Kumar. Determination of soil ph by using digital image processing technique. *Journal of Applied and Natural Science*, 6:14–18, 2014.

[4] CS ManikandaBabu and M Arun Pandian. Determination of physical and chemical characteristics of soil using digital image processing. *DIGITAL IMAGE PROCESSING*, 1:2, 2016.

[5] Wikipedia contributors. Agriculture in pakistan — Wikipedia, the free encyclopedia, 2021. [Online; accessed 10-November-2021].

[6] Wikipedia contributors. Alexnet — Wikipedia, the free encyclopedia, 2021. [Online; accessed 23-November-2021].