Industrial Engineering Journal
ISSN: 0970-2555
Volume : 53, Issue 4, April : 2024

CROP YIELD PREDICTION AND DISEASE DETECTION USING MACHINE LEARNING

Radhika Rani L, HOD, CSIOT Department, Loyola Academy, Secunderabad
radhikarani79@gmail.com

Abstract

In this project we are combining plant leaf diseases image dataset with Crop Yield dataset to forecast or estimate future crop growth. To implement this project, we have used various crop attributes such as Potassium, Phosphorous and Nitrogen properties of soil and environment data such as Humidity, temperature, and PH values and then from crop leaf images we have extracted all and sum and average all infected leaf's and included in CROP dataset.

Above process Kernel Ridge dataset will be trained with various machine learning algorithms such as Lasso, Kernel Ridge, ENET and Stacking of all 3 algorithms and then evaluate each algorithm performance in terms of MEAN SQUARE ERROR (MSE) and in all algorithms ENET and LASSO giving less MSE values.

We are using OPENCV inbuilt API'S to cluster leaf data based on colour values and then these values will be weighted to get spots whose part is not in green colour. Below code is used to extract or identify infected parts from the leaf.

Keywords: Lasso, Kernel Ridge, ENET, OPENCV, MEAN SQUARE ERROR (MSE

1. Introduction

Agriculture, since its invention and inception, is the prime and pre-eminent activity of every culture and civilization throughout the history of mankind. Not only is it a huge part of the expanding economy, but it is also necessary for our survival. It is also a vital sector for the future of humanity and the Indian economy. Additionally, it makes up a disproportionate amount of jobs. As time goes on, the amount of output that is needed has grown dramatically. People are misusing technology to make goods in large quantities. Every day, new kinds of hybrid types are created. However, these varieties do not provide the essential contents as naturally produced crop. These unnatural techniques spoil the soil. It all ends up in further environmental harm. Most of these unnatural techniques are wont to avoid losses. But when the producers of the crops know the accurate information on the crop yield it minimizes the loss.

Crop prediction is made by the system using historical data collection. The information is provided using historical data on weather, temperature, and several other variables. Our built application runs the algorithm and displays a list of crops with anticipated yield values that are appropriate for the entered data.

Using an Indian government dataset, Aruvansh Nigam, Saksham Garg, and Archit Agrawal [1] conducted trials and found that the Random Forest machine learning algorithm yields the best yield forecast accuracy. Simple Recurrent Neural Network, a sequential model, predicts rainfall more accurately than Long Short-Term Memory (LSTM) models.

The study combines variables such as area, temperature, rainfall, and season to forecast yield. The best classifier, according to the results, is Random Forest when all the parameters are combined. Leo Brieman [2,3] is an expert in the random forest algorithm's strength, accuracy, and correlation.

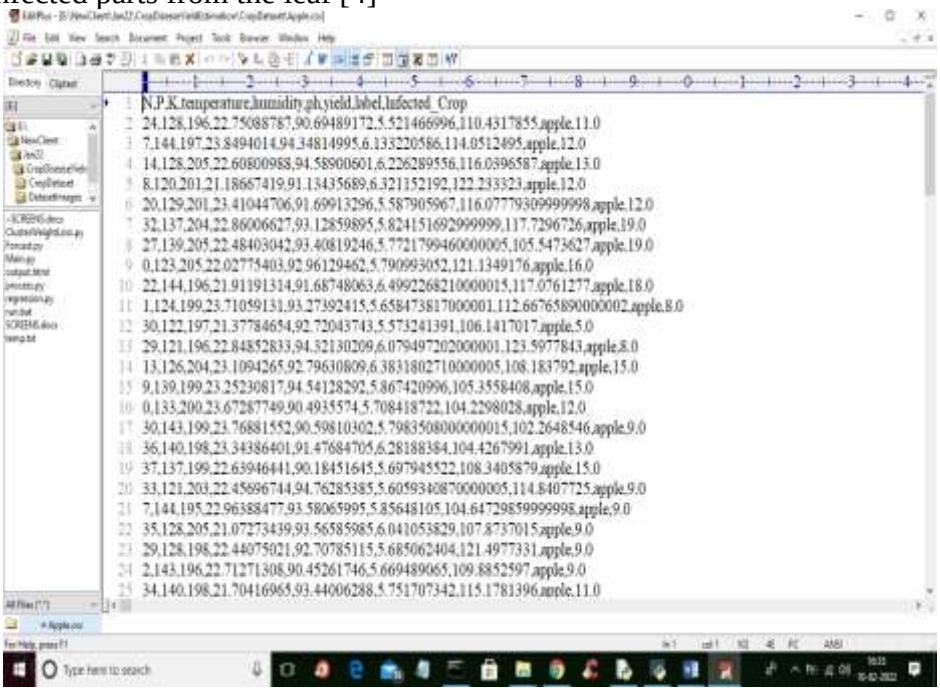
2. Materials & Methods

2.1 PROPOSED SYSTEM

Above process dataset will be trained with various machine learning algorithms such as Lasso, Kernel Ridge, ENET and Stacking of all 3 algorithms and then evaluate each algorithm performance in terms of MEAN SQUARE ERROR (MSE) and in all algorithms ENET and LASSO giving less MSE values.

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 53, Issue 4, April : 2024

By using OPENCV inbuilt API'S to cluster leaf data based on colour values and then this value will be weighted to get spots whose part is not in green colour. Below code is used to extract or identify infected parts from the leaf [4]



In above screen first row represents dataset column names and remaining are the dataset values and in above screen N refers to Nitrogen, K refers to potassium and P refers to phosphorus soil properties and remaining values are the normal environment data with infected and yield crop[5].

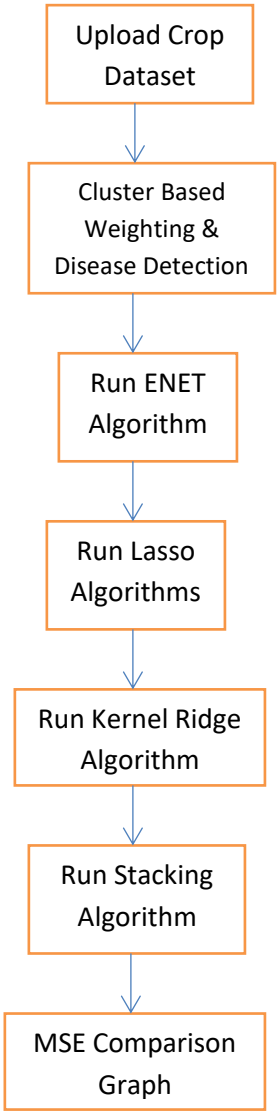


Fig.3.1 Block diagram of Proposed Method

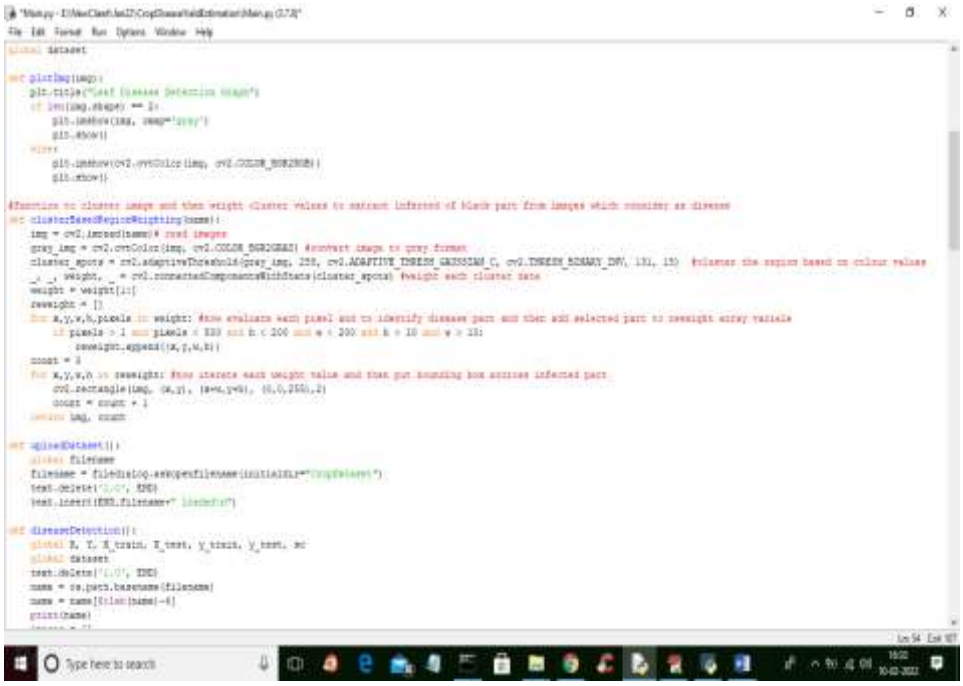
3. Results and Discussion

In this present work we are combining plant leaf diseases image dataset with Crop Yield dataset to forecast or estimate future crop growth. To implement this project, we have used various crop attributes such as Potassium, Phosphorous and Nitrogen properties of soil and environment data such as Humidity, temperature and PH values and then from crop leaf images we have extracted all and sum and average all infected leaf's and included in CROP dataset[6].

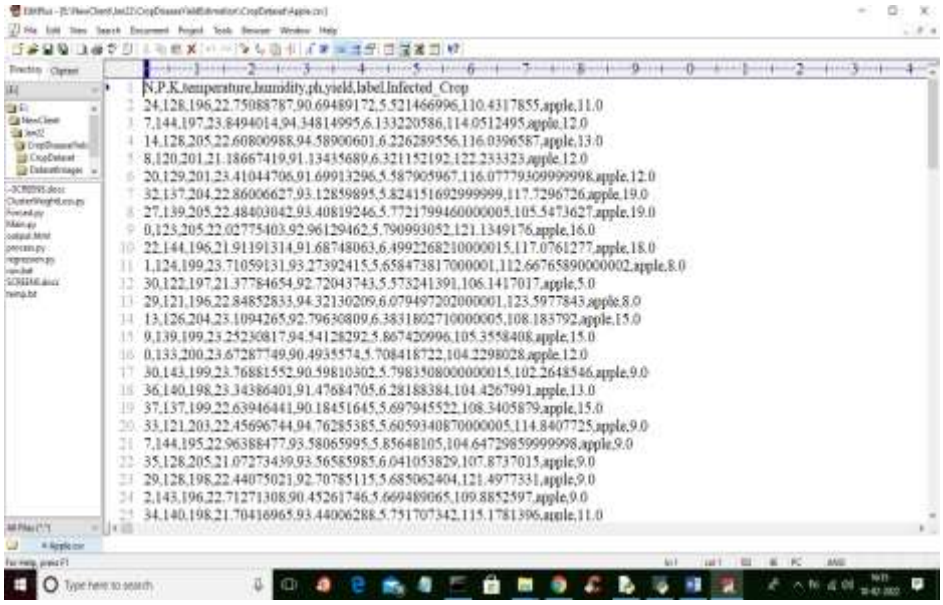
Above process dataset will be trained with various machine learning algorithms such as Lasso, Kernel Ridge, ENET and Stacking of all 3 algorithms and then evaluate each algorithm performance in terms of MEAN SQUARE ERROR (MSE) and in all algorithms ENET and LASSO giving less MSE values.

By using OPENCV inbuilt API'S to cluster leaf data based on colour values and then these values will be weighted to get spots whose part is not in green colour. Below code is used to extract or identify infected parts from the leaf[7].

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 53, Issue 4, April : 2024



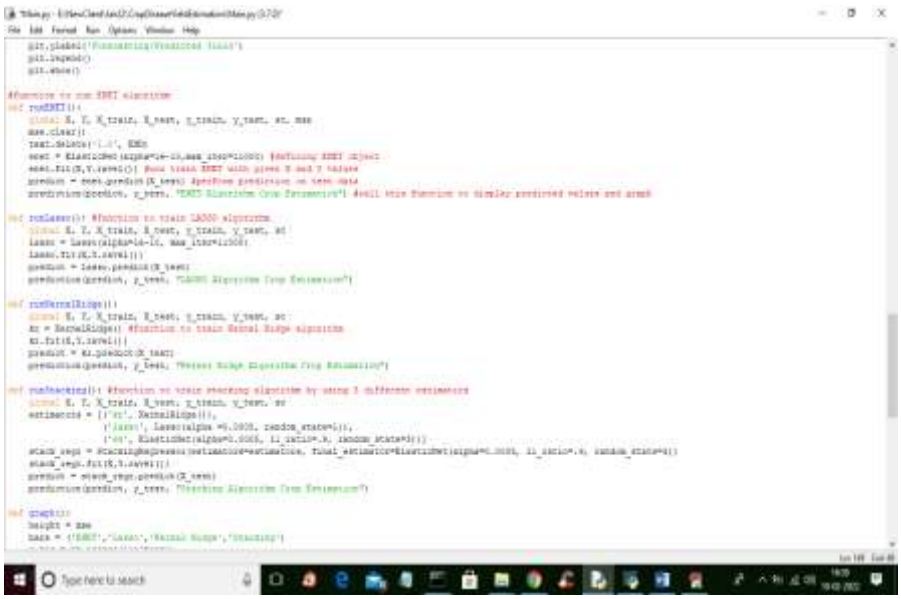
In above screen read red colour comments to detect and extract disease part from images and after applying above function on images will get final dataset with crop infected values as below screen In above screen we have 4 crop details and now open each file to view its content like below screen[8].



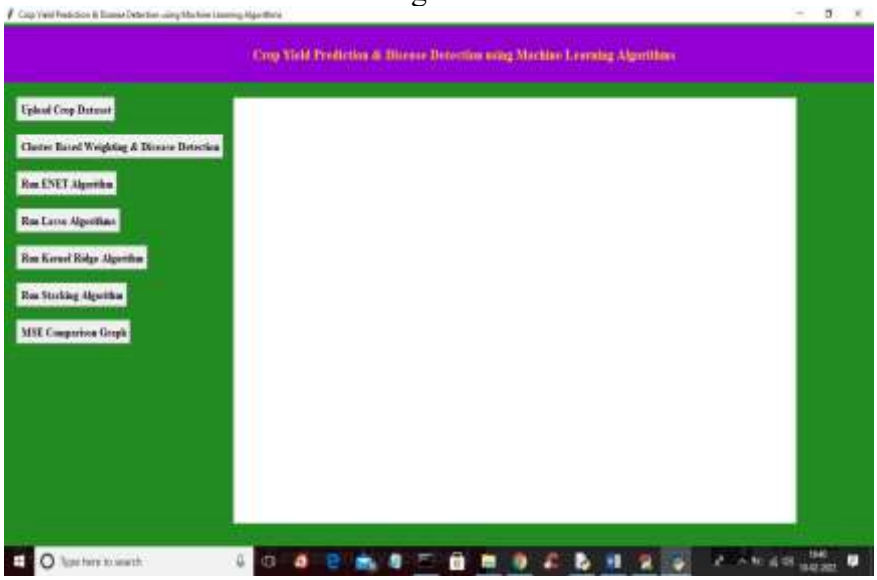
```
N,P,K,temperature,humidity,ph,yield,label,Infected_Crop
24,128,196,22.75088787,90.69489172,5.521466996,110.4317855,apple,11.0
7,144,197,23.8494014,94.34814995,6.133220586,114.0512495,apple,12.0
14,128,205,22.60800988,94.58900601,6.226289556,116.0396587,apple,13.0
8,120,201,21.18667419,91.13435689,6.321152192,122.233323,apple,12.0
20,129,201,23.41044706,91.69913296,5.587905967,116.0777936999998,apple,12.0
32,137,204,22.86006627,93.12859895,5.824151692999999,117.7296726,apple,19.0
27,139,205,22.48403042,93.40819246,5.772179946000005,105.5473627,apple,19.0
0,123,205,22.02775403,92.96129462,5.790993052,121.1349176,apple,16.0
22,144,196,21.91191314,91.68748063,6.499226820000015,117.0761277,apple,18.0
1,124,199,23.71059131,93.27392415,5.658473817000001,112.6676589000002,apple,8.0
30,122,197,21.37784654,92.72043743,5.573241391,106.1417017,apple,5.0
29,121,196,22.84852833,94.32130209,6.079497202000001,123.5977843,apple,8.0
13,126,204,23.1094265,92.79630809,6.383180271000005,108.185792,apple,15.0
9,139,199,23.25230817,94.54128292,5.867420996,105.3558408,apple,15.0
0,133,200,23.67287749,90.4935574,5.708418722,104.2298028,apple,12.0
30,143,199,23.76881552,90.59810302,5.798350800000015,102.2648546,apple,9.0
36,140,198,23.34386401,91.47684705,6.28188384,104.4267991,apple,13.0
37,137,199,22.63946441,90.18451645,5.697945522,108.3405879,apple,15.0
33,121,203,22.45696744,94.76285385,5.605934087000005,114.8407725,apple,9.0
7,144,195,22.96388477,93.58065995,5.85648105,104.6472985999998,apple,9.0
35,128,205,21.07273439,93.56585985,6.041053829,107.8737015,apple,9.0
29,128,198,22.44075021,92.70785115,5.685062404,121.4977331,apple,9.0
2,143,196,22.71271308,90.45261746,5.669489065,109.8852597,apple,9.0
34,140,198,21.70416965,93.44006288,5.751707342,115.1781396,apple,11.0
```

In above screen first row represents dataset column names and remaining are the dataset values and in above screen N refers to Nitrogen, K refers to potassium and P refers to phosphorus soil properties and remaining values are the normal environment data with infected and yield crop.
By using above dataset to train all 4 machine learning algorithms and in below screen we can see all algorithm details

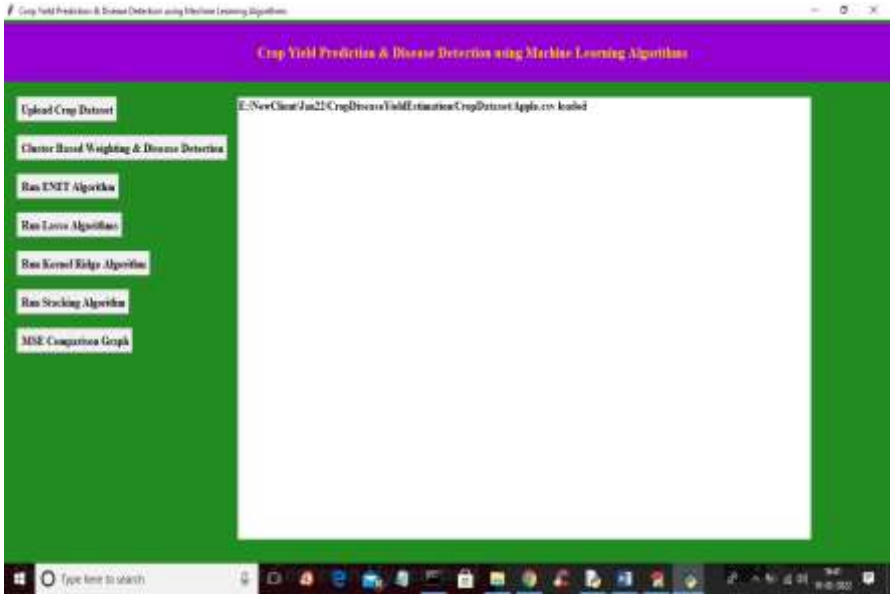
Industrial Engineering Journal
ISSN: 0970-2555
Volume : 53, Issue 4, April : 2024



In above screen read red colour comments to know about all algorithms training.
To run project double click on 'run.bat' file to get below screen

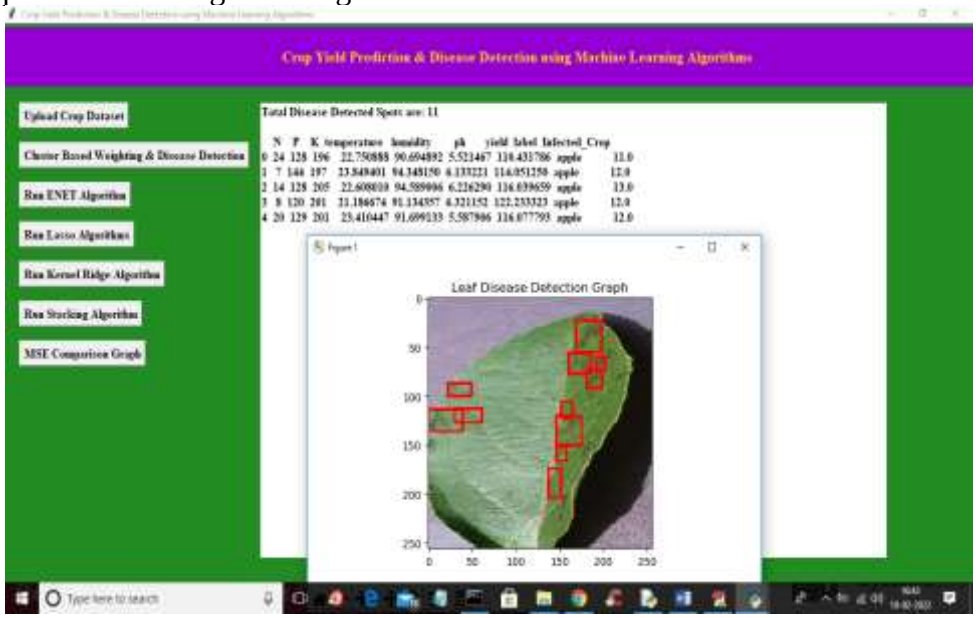


In above screen click on 'Upload Crop Dataset' button to upload dataset and to get below screen
In above screen selecting and uploading 'Apple.csv' dataset file and then click on 'Open' button to get below screen



Industrial Engineering Journal
ISSN: 0970-2555
Volume : 53, Issue 4, April : 2024

In above screen dataset loaded and now click on 'Cluster Based Weighting & Disease Detection' button to read images and then detect infected parts from the leaf and then add to dataset and for sample output I am showing one image of infected leaf like below screen



In above screen all infected leaf details added to dataset and then displaying total dataset details and then displaying one sample leaf showing disease part surrounded with red colour bounding boxes and now dataset is ready with crop and disease details and now close above image and then click on 'Run ENET Algorithm' button to get below output

Test Data Yield	Predicted Yield
111.7263395	110.0079
105.40047509999998	107.13186
102.26485459999998	108.17088
117.61029149999999	111.04815
119.1025189	120.03182
117.66028270000001	120.29226
121.6622761	121.835396
105.6941544	106.599976
109.8852597	108.62111
109.33835520000002	115.10777
104.4267991	102.915794
116.7366261	116.66921
105.02413290000003	108.798256
120.4359949	116.91329
108.183792	110.31094
105.85543510000001	106.3612
116.0396587	114.30057
104.2298028	109.169044

In above screen in text area we can see ENET got MSE value as 10 and in graph red line refers to TEST Crop DATA and green line refers to predicted crop yield and we can see in both lines very close difference is there so ENET is good in prediction and in browser we can see actual test data values and predicted values

In above screen first column showing TEST data values and second column showing predicted crop yields in QUINTAL from ENET algorithm and now close above screen and graph and then click on 'Run Lasso Algorithm' button to get below output

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 53, Issue 4, April : 2024

Test Data Yield	Predicted Yield
111.7763395	110.00778
105.40047509999998	107.13201
102.26485459999998	108.171135
117.61029149999999	111.048355
119.1025189	120.03269
117.66028270000001	120.29319
121.6622761	121.836975
105.6941544	106.60014
109.8852597	108.62133
109.33835520000002	115.10788
104.4267991	102.91533
116.7366261	116.66918
105.02413290000003	108.797935
120.4359949	116.91325
108.183792	110.31101
105.85543510000001	106.360695
116.0396587	114.300095
104.2298028	109.16947

In above two screens you can see LASSO output screens and now click on 'Run Kernel Ridge Algorithm' button to get below output

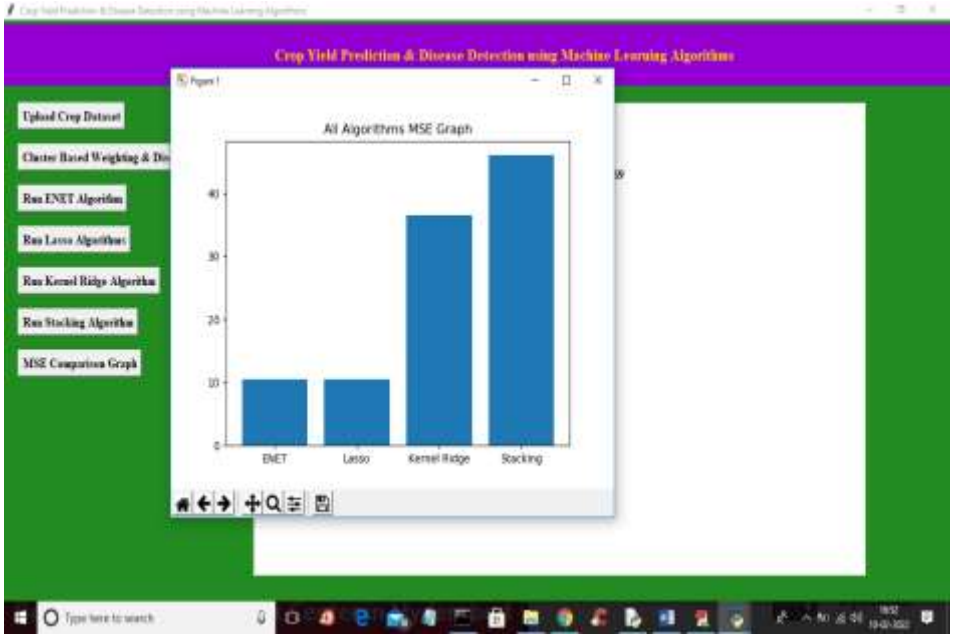
Test Data Yield	Predicted Yield
111.7763395	105.34914692791865
105.40047509999998	110.63563616218964
102.26485459999998	104.22418438513304
117.61029149999999	108.24981765287009
119.1025189	113.02498285686514
117.66028270000001	113.09664334820368
121.6622761	112.57409774830111
105.6941544	107.98026917746591
109.8852597	112.39697997860131
109.33835520000002	111.7230945284565
104.4267991	106.01535707641602
116.7366261	114.29071649555374
105.02413290000003	107.15390909678027
120.4359949	107.60697581036204
108.183792	103.60610893551873
105.85543510000001	108.52788288714198
116.0396587	107.16540270993104
104.2298028	111.71637505124683

In above two screens we can see Kernel Ridge output and now click on 'Run Stacking Algorithm' button to get below output

Test Data Yield	Predicted Yield
111.7763395	112.37788967614604
105.40047509999998	112.56773459787667
102.26485459999998	113.31247294887679
117.61029149999999	112.44149232823331
119.1025189	111.8293333145682
117.66028270000001	112.28966712777633
121.6622761	112.60713094663001
105.6941544	113.49099787798896
109.8852597	113.16991621178036
109.33835520000002	112.96922957561515
104.4267991	112.73622062097418
116.7366261	112.39794485452386
105.02413290000003	112.9232059964214
120.4359949	112.07606498836593
108.183792	112.78875444499069
105.85543510000001	112.893765412497
116.0396587	111.7524944875991
104.2298028	112.65347381710659

In above 2 screens we can see Stacking algorithm output and now click on 'MSE Comparison Graph' button to get below graph

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 53, Issue 4, April : 2024



In above graph x-axis represents algorithm names and y-axis represents MSE value and in all algorithms ENET and LASSO got less MSE values compare to other 2 algorithms and the lower the MSE the better is the algorithm. Similarly, you can upload other crop dataset and get output for those crops.

4. CONCLUSION:

In the present work by combining plant leaf diseases image dataset with Crop Yield dataset to forecast or estimate future crop growth. To implement this project we have used various crop attributes such as Potassium, Phosphorous and Nitrogen properties of soil and environment data such as Humidity, temperature and PH values and then from crop leaf images we have extracted all and sum and average all infected leafs and included in CROP dataset[9-10].

Above process dataset will be trained with various machine learning algorithms such as Lasso, Kernel Ridge, ENET and Stacking of all 3 algorithms and then evaluate each algorithm performance in terms of MEAN SQUARE ERROR (MSE) and in all algorithms ENET and LASSO giving less MSE values.

We are using OPENCV inbuilt API'S to cluster leaf data based on colour values and then these values will be weighted to get spots whose part is not in green colour. Below code is used to extract or identify infected parts from the leaf [11-13].

References:

[1] Aruvansh Nigam, Saksham Garg, Archit Agrawal "Crop Yield Prediction using ML Algorithms ", 2019

[2] Leo Brieman, "Random Forests", 2001

[3] Priya, P., Muthaiah, U., Balamurugan, M." Predicting Yield of the Crop Using Machine Learning Algorithm", 2015

[4] Mishra, S., Mishra, D., Santra, G. H., "Applications of machine learning techniques in agricultural crop production", 2016

[5] Dr. Jeevan Kumar, "Supervised Learning Approach for Crop Production", 2020

[6] Ramesh Medar,Vijay S, Shweta, "Crop Yield Prediction using Machine Learning Techniques", 2019

[7] Ranjini B Guruprasad, Kumar Saurav, Sukanya Randhawa,"Machine Learning Methodologies for Paddy Yield Estimation in India: A CASE STUDY", 2019

[8] Sangeeta, Shruthi G, "Design and Implementation of Crop Yield Prediction Model In Agriculture", 2020

[9] Automated Vision-Based Diagnosis of Banana Bacterial Wilt Disease and Black Sigatoka Disease, 1st International conference on the use of mobile ICT in Africa At: 9-10th December 2014, Stellenbosch, South Africa 4 June 2019, Godliver Owomugisha, , Busitema University (BSU), John A Quinn, John A Quinn

[10] Ebrahimi, M.A., Khoshtaghaza, M.H., Minaei, S., Jamshidi, B., Vision-based pest detection based on SVM classification method. Comput. Electron. Agric., 137, 52–58, 2017.

[11]. Chung, C.L., Huang, K.J., Chen, S.Y., Lai, M.H., Chen, Y.C., Kuo, Y.F., Detecting Bakanae disease in rice seedlings by machine vision. Comput. Electron. Agric., 121, 404–411, 2016.

[12]. Pantazi, X.E., Moshou, D., Oberti, R., West, J., Mouazen, A.M., Bochtis, D., Detection of biotic and abiotic stresses in crops by using hierarchical self-organizing classifiers. Précis. Agric., 18, 383–393, 2017.

[13]. Moshou, D., Bravo, C., West, J., Wahlen, S., McCartney, A., Automatic detection of "yellow rust" in wheat using reflectance measurements and neural networks. Comput. Electron. Agric., 44, 173–188, 2004.