Predicting Plant Growth and Yield in Greenhouse Environments Using Deep Learning

KOPILA RAVI CHAND1, GUMMADI TIRUMALA2

#1Assistant Professor, Department of CSE, PBR Visvodaya Institute of Technology and Science, Kavali

#2Assistant Professor, Department of CSE, PBR Visvodaya Institute of Technology and Science, Kavali

Abstract_ Effective plant boom and yield prediction is an indispensable mission for greenhouse growers and for agriculture in general. Developing fashions which can correctly mannequin increase and yield can assist growers enhance the environmental manipulate for higher production, healthy grant and market demand and decrease costs. Recent trends in Machine Learning (ML) and, in particular, Deep Learning (DL) can supply effective new analytical tools. The proposed find out about utilises ML and DL strategies to predict yield and plant increase version throughout two exclusive scenarios, tomato yield forecasting and Ficus benjamina stem growth, in managed greenhouse environments. We installation a new deep recurrent neural community (RNN), the use of the Long Short-Term Memory (LSTM) neuron model, in the prediction formulations. Both the former yield, increase and stem diameter values, as nicely as the microclimate conditions, are used by using the RNN structure to mannequin the focused increase parameters. A comparative find out about is presented, the use of ML methods, such as aid vector regression and random wooded area regression, utilising the suggest rectangular error criterion, in order to consider the overall performance completed by way of the distinct strategies

1.INTRODUCTION

As with many bio-systems, plant increase is a relatively complicated and dynamic environmentally linked system. Therefore, boom and yield modeling is a big scientific assignment . Modeling processes fluctuate in a variety of factors (including, scale of interest, stage of description, integration of environmental stress, etc.). According to (Todorovski and Dzeroski, 2006; Atanasova et al., 2008) two simple modeling strategies are possible, namely, "knowledge-driven" or "data-driven" modeling. The understanding pushed method depends mostly on current area knowledge. In contrast, a data-driven modeling method is succesful of formulating a mannequin entirely from gathered information barring always the usage of area knowledge. Data pushed fashions (DDM) encompass classical Machine Learning techniques, manmade neural networks (Daniel et al., 2008), assist vector machines (Pouteau et al., 2012), and generalized linear models. Those strategies have many suitable characteristics, such as imposing fewer restrictions, or assumptions, the capability to approximate nonlinear functions, sturdy predictive abilities, and the flexibility to adapt to inputs of a multivariate machine

(Buhmann, 2003). According to Singh et al., 2016 and reviewed by using Liakos et al., 2018 Machine Learning (ML), linear polarizations, wavelet-based filtering, vegetation indices (NDVI) and regression evaluation are the most famous methods used for inspecting agricultural data. However and without the aforementioned techniques, a new methodology which is currently gaining momentum is deep studying (DL)(Goodfellow et al., 2016). DL belongs to the computing device mastering computational discipline and is comparable to ANN. However, DL is about “deeper” neural networks that furnish a hierarchical illustration of the information by way of capability of a variety of operations. This lets in large mastering capabilities, and consequently greater overall performance and precision. A sturdy benefit of DL is characteristic learning, i.e., automated characteristic extraction from uncooked data, with points from greater stages of the hierarchy being fashioned by using composition of decrease degree facets (Goodfellow et al., 2016). DL can clear up greater complicated issues especially well, due to the fact of the greater complicated associated fashions (Pan and Yang, 2010). These complicated fashions employed in DL can expand classification accuracy and minimize error in regression problems, furnished there are safely giant data-sets handy describing the problem. Gonzalez-Sanchez et al.(2019) introduced a comparative find out about of ANN, SVR, M5-prime, KNN ML methods and Multiple Linear Regression for crop yield prediction in ten crop datasets. In their study, Root Mean Square Error (RMS), Root Relative Square Error (RRSE), Normalized Mean Absolute Error (MAE) and Correlation Factor (R) have been used as accuracy metrics to validate the models. Results confirmed that M5-Prime executed the lowest blunders across the produced crop yield models. The consequences of that learn about ranked the strategies from the satisfactory to the worst, in accordance to RMSE, RRSE, R, and MAE resulting, in the following order: M5-Prime, kNN, SVR, ANN and MLR. Another find out about by using (Nair and Yang-Won, 2016) utilized 4 ML techniques, SVM, Random Forest (RF), Extremely Randomized Trees (ERT) and Deep Learning (DL) to estimate corn yield in Iowa State. Comparisons of the validation information confirmed that DL supplied greater secure results, overcoming the overfitting hassle.

2.LITERATURE SURVEY

2.1 Chlingaryan, A., Sukkarieh, S. & Whelan, B. 2018, "Machine learning approaches for crop yield prediction and nitrogen status estimation in precision agriculture: A review", Computers and Electronics in Agriculture, vol. 151, pp. 61-69.

Agriculture plays big role in populated countries like Bharat. AsCrop yields are basically subject to climate. A developing exact writing models this relationship so as to extend environmental change impacts on the part. We describe an approach to predict millet crop yield prediction, which can be done by taking high dimensional datasets. By using Random Forest Classifier, we obtained 99.74% of accuracy in calculating the millet crop yield prediction by taking various input fields like soil, min temp, max temp, humidity, rainfall, etc.Thus, Millet Crop Yield Prediction is an important agricultural problem this work will help the farmers to identify the crop losses and prevent it in future. We would like to extend this work to predict and find out the accuracy of millet crop yield for both Support Vector Machine(SVM) and Linear Regression (LR).

2.2 Daniel, J., Andrés, P., Héctor, S., Miguel, B. & Marco, T. 2008, "A survey of artificial neural network-based modeling in agroecology" in Soft Computing applications in industry Springer, , pp. 247-269

Agroecological structures are hard to mannequin due to the fact of their excessive complexity and their nonlinear dynamic behavior. The evolution of such structures relies upon on a giant range of ill-defined strategies that range in time, and whose relationships are regularly particularly non-linear and very regularly unknown. According to Schultz et al. (2000), there are two principal issues when dealing with modeling agroecological processes. On the one hand, there is an absence of tools capable to seize statistics in an correct way, and on the different hand there is a lack of understanding about such systems. Researchers are consequently required to build-up fashions in wealthy and poor-data situations, by using integrating special sources of data, even if this facts is noisy, incomplete, and imprecise. In order to mannequin an agroecological system, we can proceed by way of thinking about the modeling hassle as a regression or a classification problem. For instance, we deal with a regression trouble when modeling herbal approaches such as crop yield, local weather and physiological variables, vegetation dynamics, greenhouse conditions, severity of a given pest and/or disease, etc., given that the structured variables are continuous. On the different hand, when we deal with a classification trouble when we favor to mannequin phenomena such as environmental variability, yield satisfactory and quantity, genetic variation, soil properties, land cover, etc., given that the dependant variables of the device are categories, and that the important thinking consists on assigning the identical type to folks with comparable aspects (i.e., by means of forming groups).

3.PROPOSED SYSTEM

In this paper author is predicting ficus plant growth/crop yield by evaluating performance of various machine learning algorithms such as SVR (Support Vector Regression), Random Forest Regression (RF) and LSTM (Long Short Term Memory) deep neural network algorithm. SVR and RF are the traditional old algorithms whose performance of prediction will be low due to unavailable of deep learning technique. To overcome from this problem author is using LSTM deep neural network algorithm to predict plant growth.

Deep Learning extends classical ML through adding greater "depth" (complexity) into the model, as properly as remodeling the records the usage of quite a number features that create records representations in a hierarchical way, via numerous stages of abstraction. A sturdy benefit of DL is characteristic learning, i.e., automated function extraction from uncooked data, with facets in greater degrees of the hierarchy being fashioned via composition of decrease stage features.

DL can clear up complicated troubles specifically properly and fast, due to the extra complicated fashions used, which additionally enable huge parallelization. These complicated fashions employed in DL can extend classification accuracy, or decrease error in regression problems, supplied there are correctly massive datasets on hand describing the problem. DL consists of unique components, such as convolutions, pooling layers, absolutely related layers, gates, reminiscence cells, activation functions, encoding/decoding schemes, relying on the community structure used, e.g., Convolutional Neural Networks, Recurrent Neural Networks and Unsupervised Networks.

The LSTM mannequin is introduce with the goal of modelling lengthy time period dependencies and finding out the top of the line time lag for time collection problems. A LSTM community is composed of one enter layer, one recurrent hidden layer, and one output layer. The fundamental unit in the hidden layer is the memory block, containing reminiscence cells with self-connections memorizing the temporal country and a pair of adaptive, multiplicative gating devices controlling facts waft in the block. The reminiscence telephone is exceptionally a often self-connected linear unit, referred to as Constant Error Carousel (CEC), and the telephone nation is represented with the aid of the activation of the CEC. The multiplicative gates research when to open and close. By preserving the community error constant, the vanishing gradient trouble can be solved in LSTM. Moreover, a overlook gate is delivered to the reminiscence mobile stopping the gradient from exploding when mastering lengthy time series.

3.1IMPLEMENTATION

This project consists of following modules

1) upload dataset: using this module we will upload FICUS plant dataset
2) Dataset cleaning: using this module we will find out empty values in the dataset and replace with mean or 0 values.
3) Train & Test Split: Using this module we will split dataset into two parts called and training and testing. All machine learning algorithms take 80% dataset to train classifier and 20% dataset is used to test classifier prediction accuracy. If classifier prediction accuracy high then Mean Square Error, Root Mean Square Error and Mean Absolute Error will be dropped.
4) Run SVR Classifier: Using this module we will train SVR classifier with splitted 80% data and used 20% data to calculate it performance
5) Run Random Forest Classifier: Using this module we will train Random Forest classifier with splitted 80% data and used 20% data to calculate it performance
6) Run LSTM Classifier: Using this module we will train LSTM classifier with splitted 80% data and used 20% data to calculate it performance
7) Predict Plant & Yield Growth: Using this module we will upload test data and then apply LSTM classifier to predict it growth value

4.DATASET

To implement this project we are using FICUS plant dataset and this dataset saved inside 'dataset' folder. Below are some examples of dataset

CO2,Radiation,diameter,humidity,
outside_temperature,inside_temper
ature,measurement,Yield
35.7, 20.85, 29.53, 0.91, 35.7, 27.48, 2.46, 35.7
35.1, 26.92, 29.77, 0.93, 35.1, 26.92, 2.83, 35.7
55.15, 25.42, 31.27, 0.67, 55.15, 31.8, 9.98, 45.6
54.87, 28.86, 32.39, 0.67, 54.87, 35.73, 9.97, 45.6
66.45, 34.7, 43.11, 0.75, 66.45, 39.12, 9.75, 13.1

In above dataset we have columns as CO2, RADIATION, DIAMETER etc and last value is the YIELD of the crop under above environment values. By using above values we will train classifier and then upload test data to predict future growth or yield. Below are some test environment values but YIELD column is missing and classifier will predict

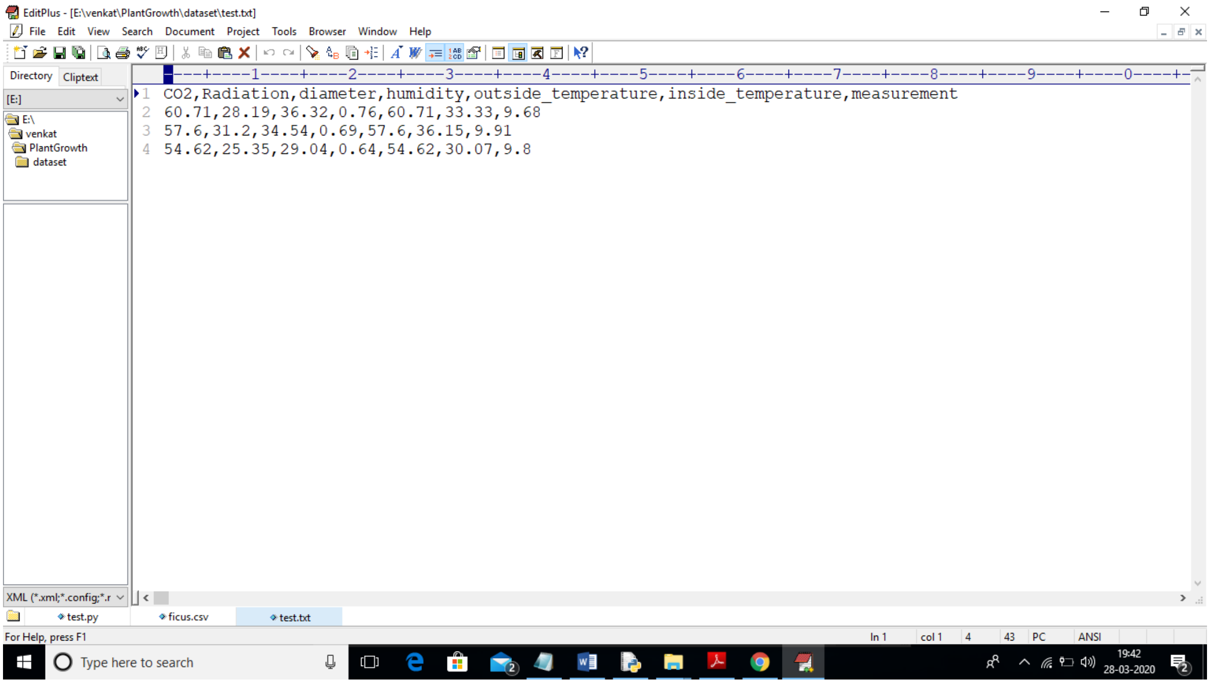


Fig 1:In above test data set we can see we have environment values but yield/growth value is missing and when we apply LSTM classifier on above test data then it will predict future growth for above test data.

5.RESULTS AND DISCUSSION

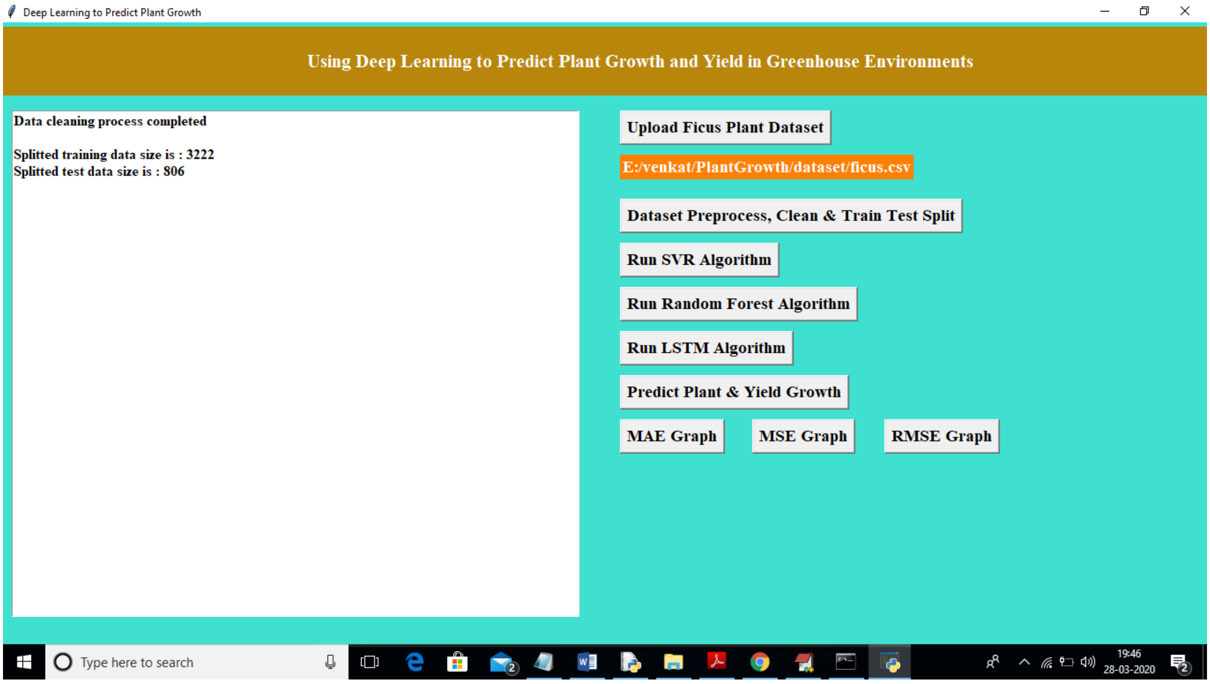


Fig 2:In above screen we can see application split dataset into 80 and 20% and application using 3222 records for training and 806 for testing. Now dataset loaded and splitted and now click on 'Run SVR Algorithm' button to train SVR algorithm

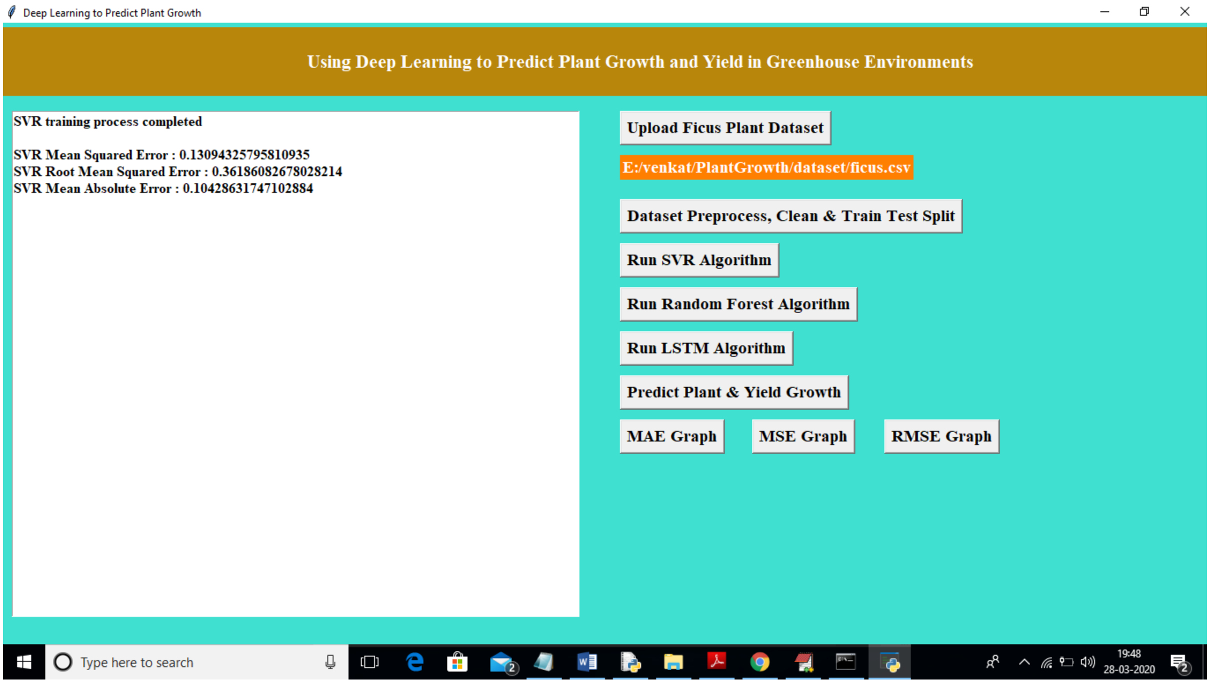


Fig 3:In above screen we got RMSE, MAE and MSE error for SVR algorithm and now click on 'Run Random Forest Algorithm' button to train random forest algorithm

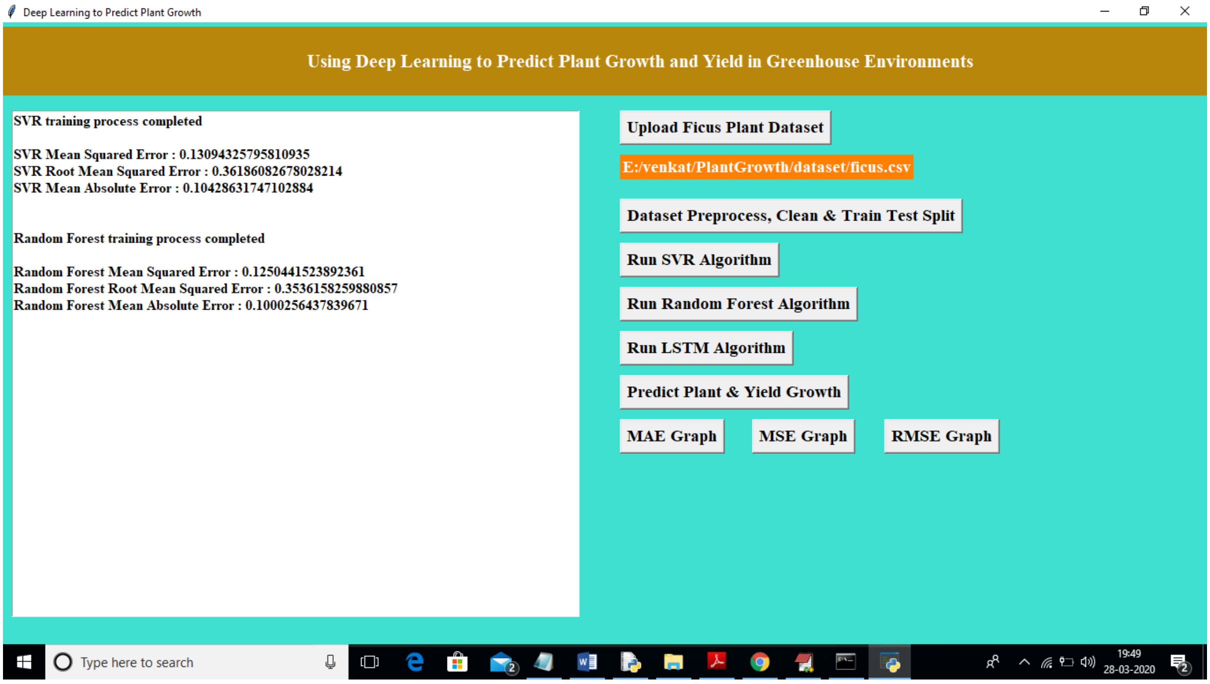


Fig 4:In above screen we got random forest MSE, RMSE, MAE error and now click on 'Run LSTM Algorithm' button to train dataset with LSTM algorithm

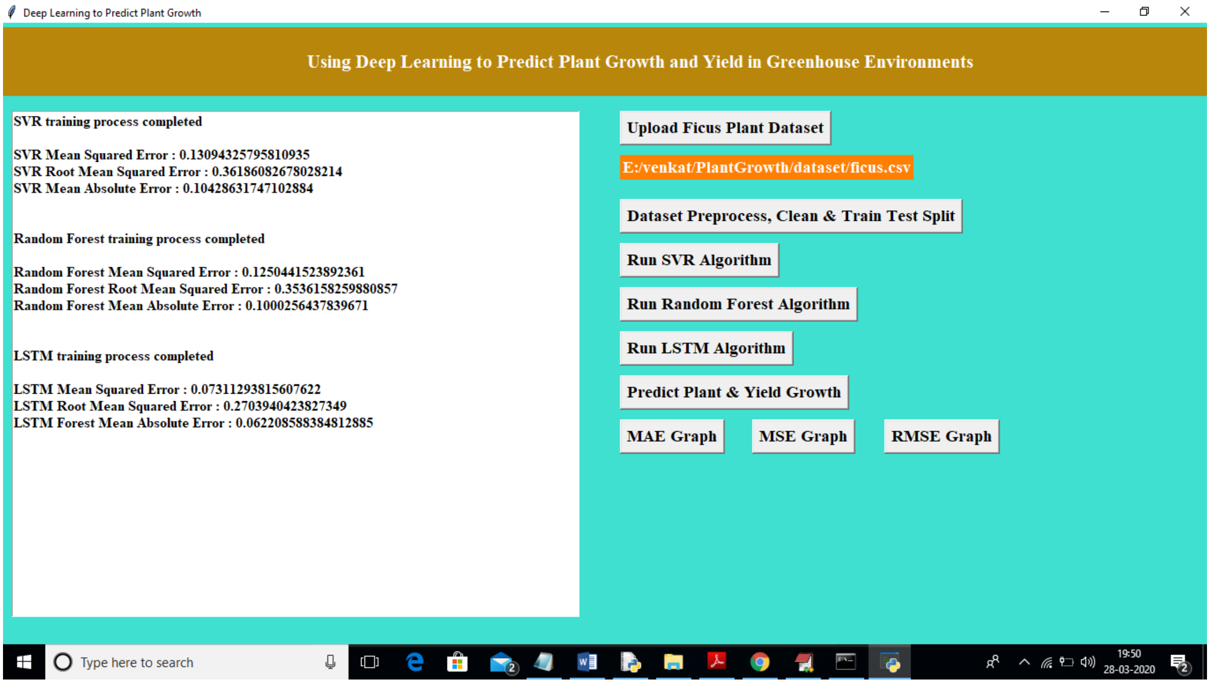


Fig 5:In above screen we can see LSTM got less MSE, RMSE and MAE error compare to traditional algorithm. Now all algorithms training process completed and now we can upload test file and predict its growth

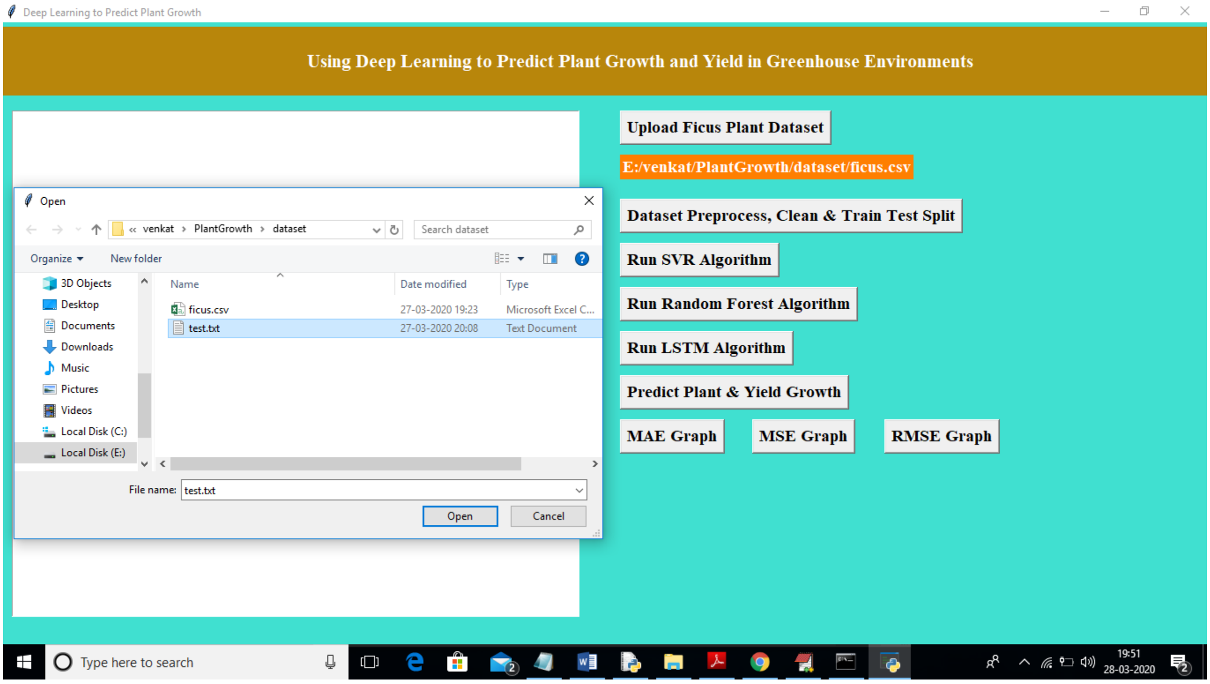


Fig6:In above screen I am uploading ‘test.txt’ file and now click on ‘Open’ button to predict growth for test data

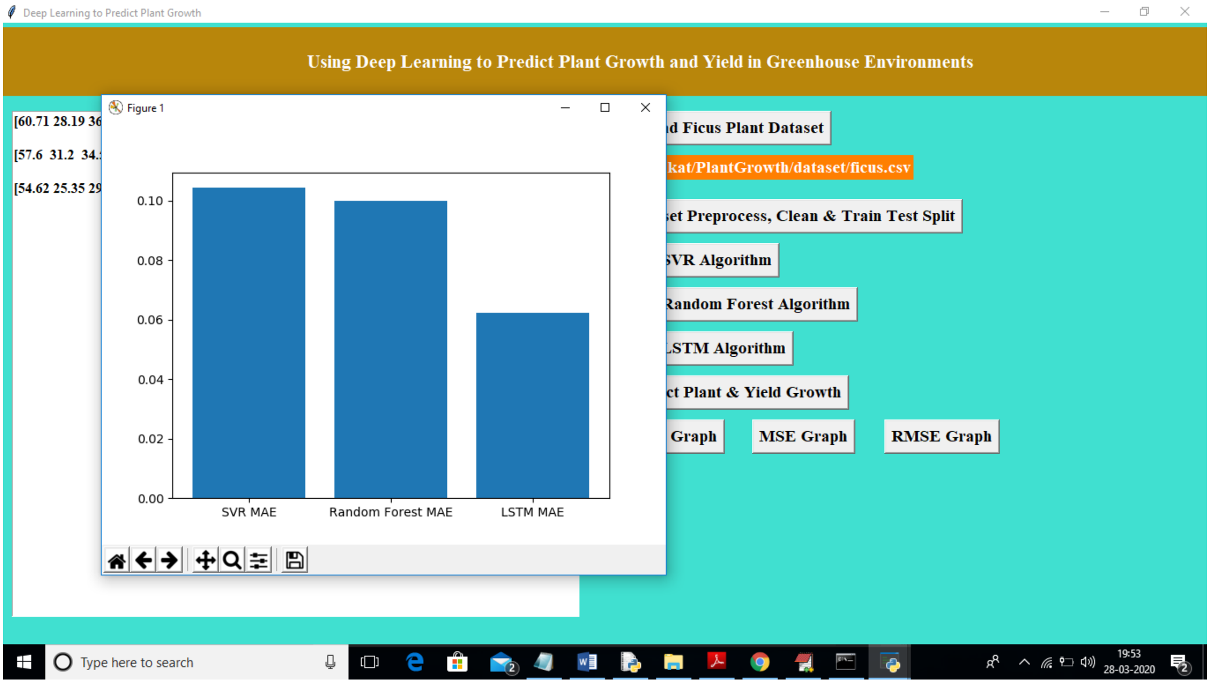
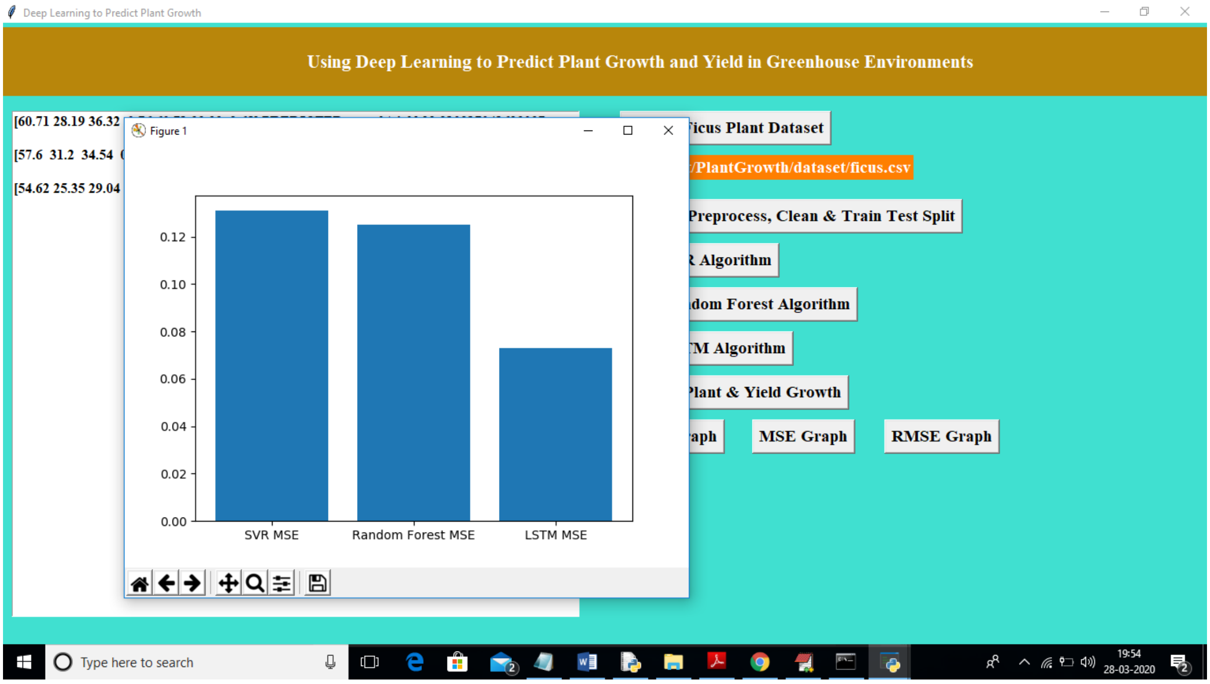


Fig 7:In above graph x-axis represents algorithm name and y-axis represents MAE error. From above graph we can conclude that LSTM got less error and its prediction performance will be best compare to other two.



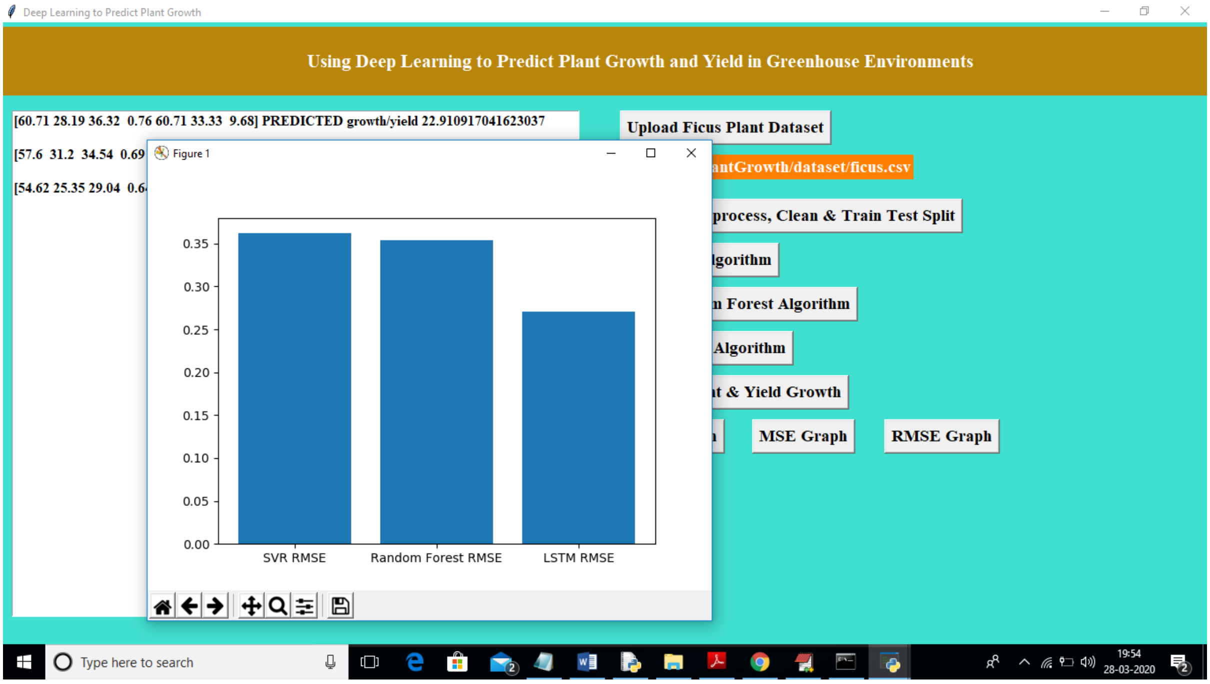


Fig 8: RMSE graph

6.CONCLUSION

The paper developed a DL strategy the use of LSTM for Ficus boom (represented via the SDV) and tomato yield prediction, accomplishing excessive prediction accuracy in each problems. Experimental effects have been introduced that exhibit that the DL method (using a LSTM model) outperformed different usual ML techniques, such as SVR and RF, in phrases of MSE, RMSE and MAE error criteria. Hence, the primary purpose of our assignment is to strengthen DL methodologies to predict plant life boom and yield in greenhouse environment. Future research searching at the continuity of : a) extensively enlarge the wide variety of accrued information that are used for coaching the proposed DL methods; b) extending the DL technique so as to function multi-step (at a weekly, or a a couple of of weeks basis) prediction of increase and yield in a massive range of greenhouse, in the UK and Europe.

REFERENCES

1. Abreu, P., Meneses, J. & Gary, C. 1998, "Tompousse, a model of yield prediction for tomato crops: calibration study for unheated plastic greenhouses", XXV International Horticultural Congress, Part 9: Computers and Automation, Electronic Information in Horticulture 519, pp. 141.
2. Adams, S. 2001, "Predicting the weekly fluctuations in glasshouse tomato yields", IV International Symposium on Models for Plant Growth and Control in Greenhouses: Modeling for the 21st Century-Agronomic and 593, pp. 19.
3. Atanasova, N., Todorovski, L., Džeroski, S. & Kompare, B. 2008, "Application of automated model discovery from data and expert knowledge to a real-world domain: Lake Glumsø", Ecological Modelling, vol. 212, no. 1-2, pp. 92-98.
4. Barandiaran, I. 1998, "The random subspace method for constructing decision forests", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 20, no. 8. Breiman, L. 2001, "Random forests", Machine Learning, vol. 45, no. 1, pp. 5-32.
5. Buhmann, M.D. 2003, Radial basis functions: theory and implementations, Cambridge university press.
6. Chlingaryan, A., Sukkarieh, S. & Whelan, B. 2018, "Machine learning approaches for crop yield prediction and nitrogen status estimation in precision agriculture: A review", Computers and Electronics in Agriculture, vol. 151, pp. 61-69.
7. Cortes, C. & Vapnik, V. 1995, "Support-vector networks", Machine Learning, vol. 20, no. 3, pp. 273-297.
8. Daniel, J., Andrés, P., Héctor, S., Miguel, B. & Marco, T. 2008, "A survey of artificial neural network-based modeling in agroecology" in Soft Computing applications in industry Springer, , pp. 247-269.