JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

Using Deep Learning to Predict Plant Growth and Yield in Greenhouse Environments

Mudugula Babu, IV B.Tech Student, Dept of IT, Sreenidhi Institute of Science and Technology (A), Hyderabad. babumudugula@gmail.com

Talla Umesh, IV B.Tech Student, Dept of IT, Sreenidhi Institute of Science and Technology (A), Hyderabad. umeshtalla2000@gmail.com

Dr. Sreenivas Mekala, Associate Professor, Dept of IT, Sreenidhi Institute of Science and Technology (A), Hyderabad.

ABSTRACT:

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.

1. INTRODUCTION:

Deep Learning extends classical ML by adding more "depth" (complexity) into the model, as well as transforming the data using various functions that create data representations in a hierarchical way, through several levels of abstraction. A strong advantage of DL is feature learning, i.e., automatic feature extraction from raw data, with features in higher levels of the hierarchy being formed through composition of lower level features. DL can solve complex problems particularly well and fast, due to the more complex models used, which also allow massive parallelization. These complex models employed in DL can increase classification accuracy, or reduce error in regression problems, provided there are adequately large datasets available describing the problem. DL includes different components, such as convolutions, pooling layers, fully connected layers, gates, memory cells, activation functions, encoding/decoding schemes,

depending on the network architecture used, e.g., Convolutional Neural Networks, Recurrent Neural Networks and Unsupervised Networks.

The LSTM model is introduce with the objective of modelling long term dependencies and determining the optimal time lag for time series problems. A LSTM network is composed of one input layer, one recurrent hidden layer, and one output layer. The basic unit in the hidden layer is the memory block, containing memory cells with self-connections memorizing the temporal state and a pair of adaptive, multiplicative gating units controlling information flow in the block. The memory cell is primarily a recurrently self-connected linear unit, called Constant Error Carousel (CEC), and the cell state is represented by the activation of the CEC. The multiplicative gates learn when to open and close. By keeping the network error constant, the vanishing gradient problem can be solved in LSTM. Moreover, a forget gate is added to the memory cell preventing the gradient from exploding when learning long time series.

2. LITERATURE REVIEW:

As with many bio-systems, plant growth is a highly complex and dynamic environmentally linked system. Therefore, growth and yield modeling is a significant scientific challenge. Modeling approaches vary in a number of aspects (including, scale of interest, level of description, integration of environmental stress, etc.). According to (Todorovski and Dzeroski, 2006; Atanasova et al., 2008) two basic modeling approaches are possible, namely, "knowledge-driven" or "data-driven" modeling. The knowledge driven approach relies mainly on existing domain knowledge. In contrast, a data-driven modeling approach is capable of formulating a model solely from gathered data without necessarily using domain knowledge. Data driven models (DDM) include classical Machine Learning techniques, artifical neural networks (Daniel et al., 2008), support vector machines (Pouteau et al., 2012), and generalized linear models. Those methods have many desirable characteristics, such as imposing fewer restrictions, or assumptions, the ability to approximate nonlinear functions, strong predictive abilities, and the flexibility to adapt to inputs of a multivariate system (Buhmann, 2003).

According to Singh et al., 2016 and reviewed by Liakos et al., 2018 Machine Learning (ML), linear polarizations, wavelet-based filtering, vegetation indices (NDVI) and regression analysis are the most popular techniques used for analyzing agricultural data. However and besides the aforementioned techniques, a new methodology which is recently gaining momentum is deep learning (DL)(Goodfellow et al., 2016).

DL belongs to the machinelearning computational field and is similar to ANN. However, DL is about “deeper” neural networks that provide a hierarchical representation of the data by means of various operations. This allows larger learning capabilities, and thus higher performance and precision. A strong advantage of DL is feature learning, i.e., automatic feature extraction from raw data, with features from higher levels of the hierarchy being formed by composition of lower level features (Goodfellow et al., 2016).

DL can solve more complex problems particularly well, because of the more complex related models

(Pan and Yang, 2010). These complex models employed in DL can increase classification accuracy and reduce error in regression problems, provided there are adequately large data-sets available describing the problem. Gonzalez-Sanchez et al.(2019) presented a comparative study of ANN, SVR, M5-prime, KNN ML techniques 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) were used as accuracy metrics to validate the models. Results showed that M5-Prime achieved the lowest errors across the produced crop yield models.

The results of that study ranked the techniques from the best to the worst, according to RMSE, RRSE, R, and MAE resulting, in the following order: M5-Prime, kNN, SVR, ANN and MLR. Another study by (Nair and Yang-Won, 2016) applied four 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 statistics showed that DL provided more stable results, overcoming the overfitting problem. Stem diameter is considered as one of the important parameters describing the growth of plants during vegetative growth stage.

Also, the variation of stem diameter has widely been used to derive proxies for plant water status and, is therefore applied in optimisation strategies for plant-based irrigation scheduling in a wide range of species. Plant stem diameter variation (SDV) refers to plant stem periodic shrinkage and recovery movement during the day and night, and this periodic variation is related to plant water content and can be used as an indicator of the plant water content change. During active vegetative growth and development, crop plants rely on the carbohydrate gained from photosynthesis and the translocation of photo-assimilates from the site of synthesis to sink organs (Yu et al., 2015). The fundamentals of stem diameter variations have been well documented in a substantial amount of literature (Vandegehuchet et al., 2014).

It has been documented that SDV is sensitive to water and nutrient conditions and is closely related to the responses of crop plants to the changes of environmental conditions (Kanai et al., 2008). The stem diameter is an important parameter describing the growth of crop plants under abiotic stress during vegetative growth stage. Therefore, it is important to generate stem diameter growth models able to predict the response of SDV to environmental changes and plant growth under different conditions. Many studies emphasize the need to critically review and improve SDV models for assessment of environmental impact on crop growth (Hinckley and Bruckerhoff, 2011). SDV daily models have been developed to accurately predict inter-annual variation in annual growth in balsam fir (Abies balsamea L) (Duchesene and Houle, 2011). Inclusion of daily data in growth-climate models can improve predictions of the potential growth response to climate by identifying particular climatic events that escape to a classical dendroclimatic approach (Duchesene and Houle, 2011).

However, models for predicting SDV and plant growth using environmental variables have so far remained limited. Tomato crop growing in

greenhouse environment is considered as a dynamic and complex system, with few models having been studied for it up to now. In the literature TOMGRO and TOMSIM (Jones et al., 1999), (Heuvelink, 1996) are considered as the main applicable dynamic growth models. Those models are dependent on physiological processes, and they represent biomass partitioning, crop growth, and yield as a function of several climate and physiological parameters. However, due to their limited application to practicalsettings, their complexity, the difficulty in estimating initial parameter values and the need for calibration and validation in every new environment, growers uptake has been limited. The Tompousse model was developed by (Abreu et al., 2000) to predict tomato yield in terms of the weight of harvested fruits.

The model was developed by examining the relationship between environmental parameters in a heated greenhouses in the Southern part of France. A linear relationship between flowering rate and fruit growth was the basic assumption used in this model. However, the model performance was poor when tested in unheated plastic greenhouses in Portugal. Another tomato yield model was proposed by Adams (Adams, 2002), based on a form of graphical simulation tool. The main objective of the model was to represent weekly fluctuations of greenhouse tomato yield in terms of fruit size and harvest rate. Hourly climate data were used to estimate the rate of growth of leaf truss and the flower production. Yield seasonal fluctuations were generally infuenced by periodic variations of solar radiation and air temperature. According to (Qaddoum et al., 2013), there is a large number of tools that can help farmers in making decisions.

3. METHODOLOGY:

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

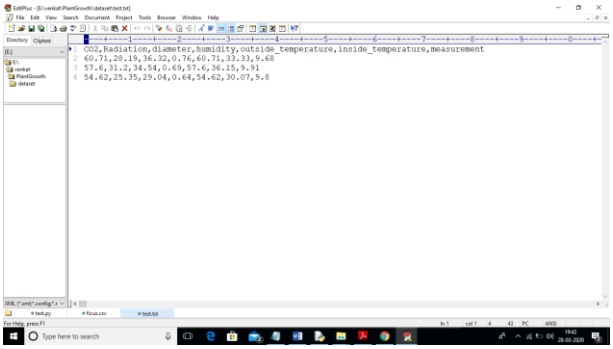
4. RESULTS AND DISCUSSIONS:

Dataset information

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_temperature, 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



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.

Double click on 'run.bat' file to get below screen

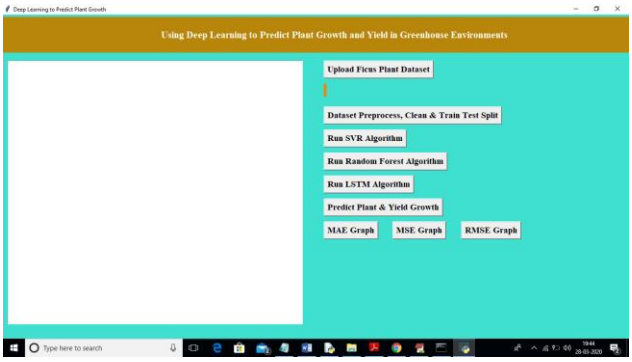


Fig 1: In above screen click on 'Upload Ficus Plant Dataset' button and upload dataset

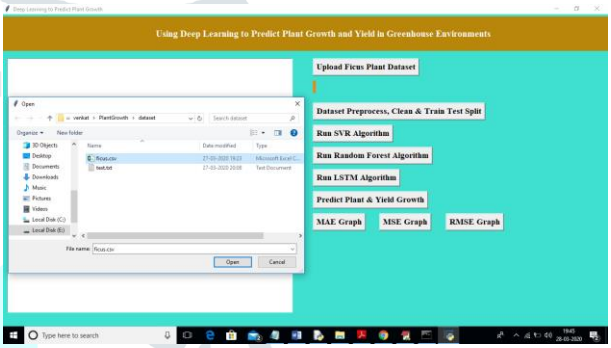


Fig 2: In above screen I am uploading 'ficus.csv' dataset file and after uploading dataset will get below screen

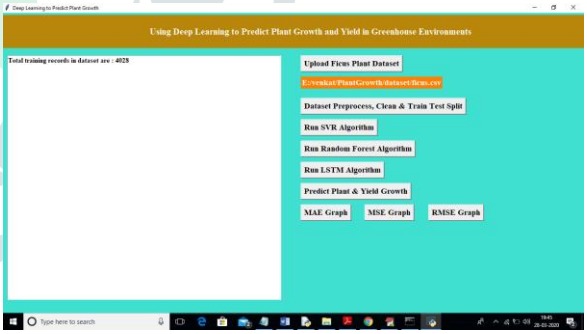


Fig 3: In above screen we can see dataset loaded and dataset contains total 4028 records. Now click on 'Dataset Preprocess, Clean & Train Test Split' button to clean dataset and to split dataset into train and test part

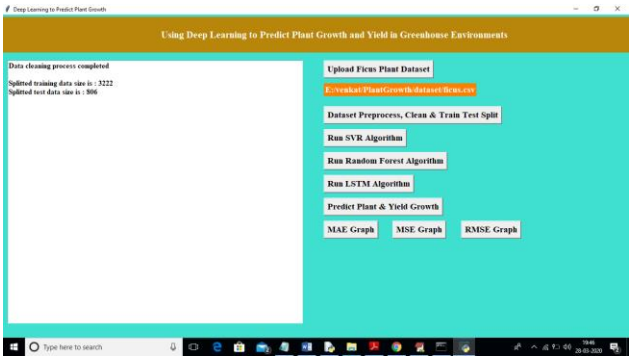


Fig 4: In above screen we can see application split dataset into 80 and 20% and application using 3222 records for training and 806 for testing. Now dataset loaded and splitted and now click on 'Run SVR Algorithm' button to train SVR algorithm

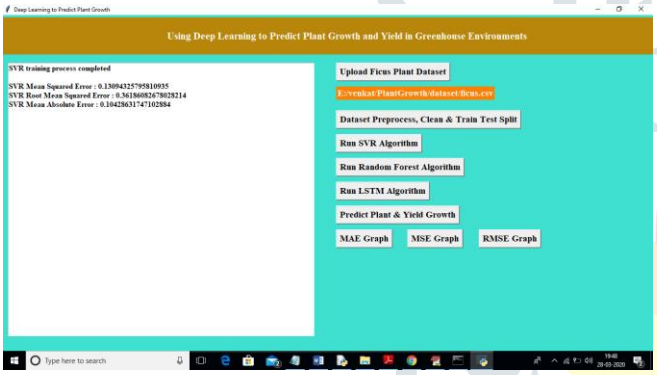


Fig 5: In above screen we got RMSE, MAE and MSE error for SVR algorithm and now click on 'Run Random Forest Algorithm' button to train random forest algorithm

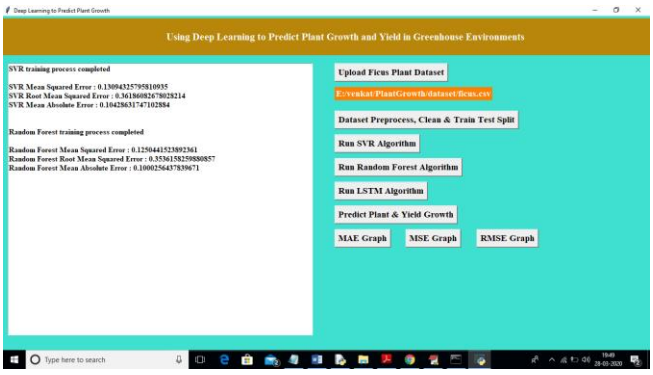


Fig 6: In above screen we got random forest MSE, RMSE, MAE error and now click on 'Run LSTM Algorithm' button to train dataset with LSTM algorithm

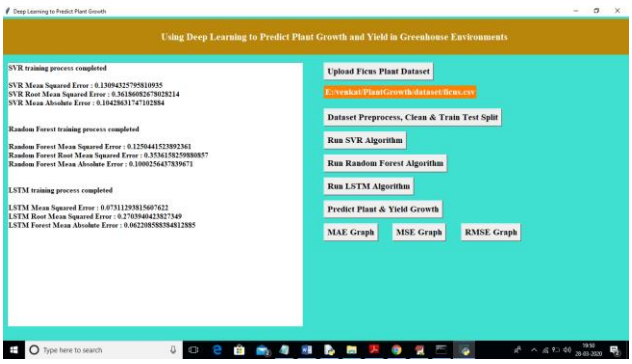


Fig 7: In above screen we can see LSTM got less MSE, RMSE and MAE error compare to traditional algorithm. Now all algorithms training process completed and now we can upload test file and predict its growth

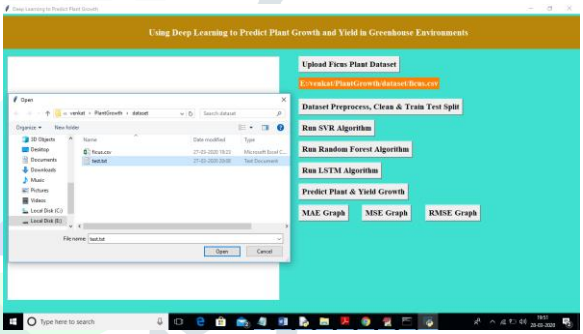


Fig 8: In above screen I am uploading 'test.txt' file and now click on 'Open' button to predict growth for test data

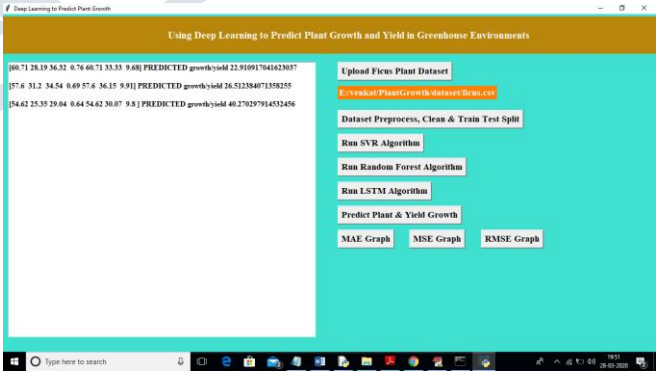


Fig 9: In above screen for first record growth prediction is 22% and second record 26% and third record having 40% growth prediction. Similarly u can add new records to test data and can predict its growth. Now click on 'MAE Graph' button to see MAE comparison graph between all algorithms

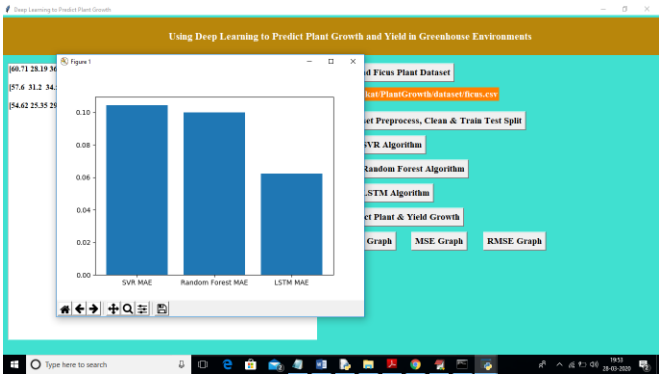


Fig 10: In above graph x-axis represents algorithm name and y-axis represents MAE error. From above graph we can conclude that LSTM got less error and its prediction performance will be best compare to other two.

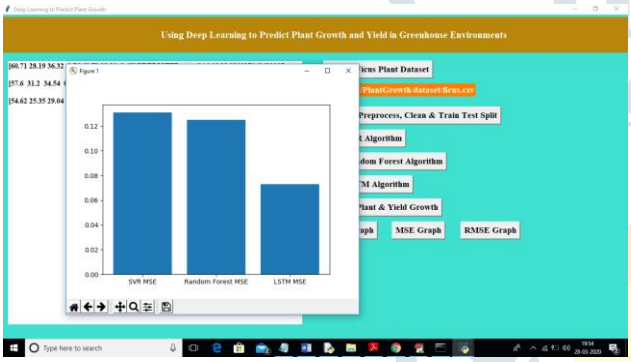


Fig 11: Below MSG error graph

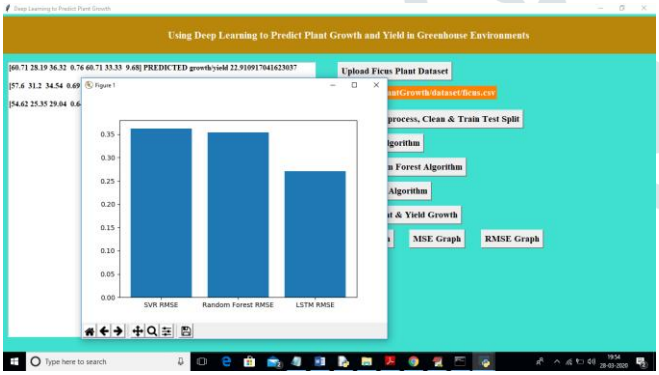


Fig 12: Below RMSE graph

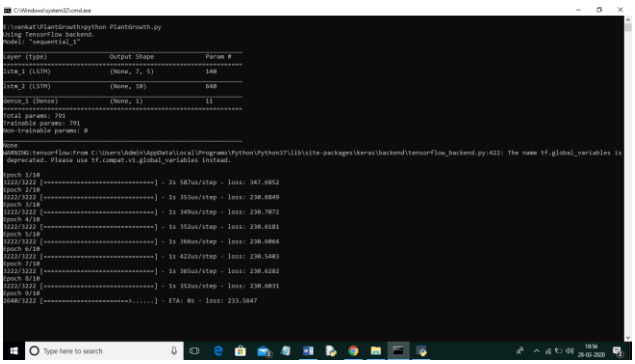


Fig 13: In above black screen we can see training model generation for LSTM and to build this model I am using 10 epoch and in each epoch LSTM will use recent data to train model and forgot old data reference.

5. CONCLUSION:

The paper developed a DL approach using LSTM for Ficus growth (represented by the SDV) and tomato yield prediction, achieving high prediction accuracy in both problems. Experimental results were presented that show that the DL technique (using a LSTM model) outperformed other traditional ML techniques, such as SVR and RF, in terms of MSE, RMSE and MAE error criteria. Hence, the main aim of our project is to develop DL methodologies to predict plants growth and yield in greenhouse environment. Future studies looking at the continuity of : a) greatly increase the number of collected data that are used for training the proposed DL methods; b) extending the DL method so as to perform multi-step (at a weekly, or a multiple of weeks basis) prediction of growth and yield in a large variety of greenhouse.

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 onPattern Analysis and Machine Intelligence, vol. 20, no. 8..