PREDICTION OF GREENHOUSE PLANT USING DEEP LEARNING

Mrs. K Sheetal[1] .,E.Pavani Sreeja[2].,K.Shivani[3].,K.Krishna Lavanya[4] .,Y.Sai Chandana[5]

1 Assistant Professor, Department of CSE., Malla Reddy College of Engineering for Women., Mismanaged., Medical., TS, India (✉sheetalkulkarni.925@gmail.com)

2, 3, 4,5 B. Tech CSE, (19RG1A05J9, 19RG1A05L0, 19RG1A05L5, 19RG1A05P9),

Malla Reddy College of Engineering for Women., Maisammaguda., Medchal., TS, India

.

Abstract

For greenhouse growers and farmers in general, accurate forecasts of plant development and productivity are critical. Growers can enhance environmental control, match supply with demand, and minimize costs by developing models that accurately forecast growth and output. Powerful new analytical tools may be gained from recent advances in Machine Learning and in particular, Deep Learning (DL). Tomato yield forecasting and Ficus Benjamin stem development will be predicted using ML and DL approaches in a controlled greenhouse setting in the proposed project. In the prediction formulations, we use a novel deep recurrent neural network (RNN) based on the LSTM neuron model. The RNN architecture models the intended growth parameters using the previous yield, growth, and stem diameter measurements, as well as microclimate circumstances. Support vector regression and random forest regression are compared in a researchutilising the mean square error criteria to assess the effectiveness of the various approaches. Results from the EU Interreg SMARTGREEN project (2017-2021) in two greenhouses in Belgium and the UK have shown great promise, according to the statistics given.

Keywords:

Deep neural networks for deep learning (RLSNNNs) and recurrent long short-term memory (RLSNNNs)

INTRODUCTION

Plant development, like many other bio-systems, is a highly complex and dynamic system that is intertwined with the environment. Growing and yielding models are thus an important scientific problem. There are a lot of ways to model a problem (including, scale of interest, level of description, integration of environmental stress, etc.). "Knowledge-driven" or "data-driven" modelling are two primary techniques, according to (Todorovski and Demoski, 2006; Atanasova et al., 2008). The method based on existing domain knowledge is referred to as a "knowledge driven" one. Data-driven modelling, on the other hand, does not need the use of domain expertise to create a model. There are a variety of data-driven models (DDMs) that incorporate machine learning approaches such as neural networks, support vector machines, and generalized linear models (Poteau et al., 2012). Methods like this offer a wide range of desired.

features, such as the capacity to approximate nonlinear functions, the ability to forecast accurately and the ability to handle a wide range of inputs (Buhmann, 2003). Regression analysis, linear polarizations, wavelet-based filtering and vegetation indices (NDVI) are the most often used approaches for assessing agricultural data, according to Singh and Liakos et al. (2016) and Liakos et al. (2018) respectively. A new approach, deep learning (DL), has lately gained traction in addition to the aforementioned methods (Goodfellow et al., 2016). According to Wikipedia, DL is part of the machine learning area and is quite similar to an ANN. Deep learning is about "deeper" neural networks that give a hierarchical representation of the data through different operations. This provides for more learning capabilities, resulting in greater performance and accuracy. As a result of feature learning, raw data may be automatically analyzed for characteristics that can be used to build higher-level hierarchical features (Goodfellow et al., 2016). Because of the complexity of the linked models, DL is especially good at solving increasingly complicated issues (Pan and Yang, 2010). It is possible to improve classification and regression accuracy with the use of DL's sophisticated models, but only if big data sets are available to describe the issue. Comparative studies of ANN, SVR, M5-prime, KNN and Multiple Linear Regression for agricultural yield prediction were reported by Gonzalez-Sanchez and colleagues (2019). There were four accuracy measures that were employed in their study: root mean square error (RMS), root relative square error (RRSE), normalized mean absolute error (MAE), and correlation factor. Among the crop yield models tested, M5-Prime exhibited the fewest mistakes. M5-Prime, SVR, ANN, and MLR were found to be the top and worst algorithms rated according to RMSE, RRSE, R, and MAE in that research. Nair and Yang-Won (2016) used four machine learning algorithms to predict corn yield in Iowa State,

including SVM, Random Forest (RF), Extremely Randomized Trees (ERT), and Deep Learning (DL). DL was shown to be more stable, addressing the issue of overfitting, based on comparisons of validation data. During the vegetative development stage, stem diameter is one of the most significant criteria to describe plant growth. Stem diameter variation has been utilized extensively to develop proxies for plant water status, which has been employed in a broad variety of species to optimize plant-based irrigation scheduling. An indication of changes in plant water content may be found in plant stem diameter variation (SDV), which is a measure of plant stem shrinkage and recovery over the course of the day and night. Crop plants depend on photosynthesis and the transport of photo-assimilates from the site of synthesis to sink organs for glucose during active vegetative growth and development (Yu et al., 2015). In a large body of literature, the foundations of stem diameter variation have been thoroughly examined and recorded (Vandeguchte et al., 2014). Soil dissolved volume (SDV), which is intimately linked with agricultural plant responses to environmental changes, has been established (Kanai et al., 2008). Crop plants under abiotic stress during the vegetative development stage have a significant impact on stem diameter. In order to forecast the reaction of SDV to environmental changes and plant development under diverse situations, stem diameter growth models must be developed. SDV models for assessing the influence of the environment on crop development need to be critically reviewed and improved, according to several research (Hinckley and Bruckerhoff, 2011). Accurate inter-annual variance in yearly growth of the balsam fir (Abies Bernama L) has been reliably predicted using SDV daily models (Duchesne and Houle, 2011). By including daily data into growth-climate models, researchers may better forecast the possible growth response to climate by detecting specific climatic events that would otherwise go unnoticed by a dendroclimatic method (Duchesne and Houle, 2011). However, models that use environmental factors to predict SDV and plant development are still in their infancy. Few models have been investigated for tomato crop growth in greenhouse environment as a dynamic and complicated system. As far as dynamic growth models are concerned, Jones et al. (Jones, 1999) and Heuvel ink (1996) use TOMGRO and TOMSIM (Heuvel ink, 1996). Models of crop development and production are based on physiological processes, and they are influenced by a variety of climatic and physiological variables. As a result of their complexity, limited practical use, difficulties in determining starting parameter values and the need of calibration and validation in each new environment, growers have been restricted in their acceptance. Based on the weight of gathered fruits, the Tomosesmodel was created by Abreu et al. (2000). Environmental characteristics in a heated greenhouse in southern France were examined to construct the model. This model assumed a linear connection between the pace of blooming and the rate of fruit development. When tested in Portugal in unheated plastic greenhouses, the model's performance was low. Using a graphical modelling program, Adams (Adams, 2002) came up with another tomato yield model (Adams, 2002) Modeling weekly variations in greenhouse tomato productivity in terms of fruit size and harvest rate was the primary goal. Leaf truss growth and flower output were estimated using hourly climatic data. In general, seasonal differences in solar radiation and air temperature have a significant impact on crop yields. Many tools exist to assist farmers in making choices (Qadim et al., 2013), according to (Qadim, 2013). It's possible to anticipate yield rates, provide climate control suggestions, and time agricultural production to match market needs using these. An environmental (CO2, humidity, radiation, outside temperature, inside temperature) as well as actual yield and stem diameter variation measurement-trained deep learning model is proposed in this paper and has the ability to accurately predict stem diameter problems in Ficus trees and tomato yield problems in tomatoes. Listed below are the sections of this document. Described in Section 2 are the suggested technique and datasets that will be used. Section 3 displays the outcomes, and Section 4 concludes and plans for the future..

MATERIALS AND METHODS

Conventional Machine Learning

Machine learning (ML) algorithms can solve complex, non-linear problems on their own, employing data from a variety of different sources. Artificial intelligence can make better decisions and take more informed actions in real-world situations with little or no human involvement. Data-driven decision making may be applied to a broad range of industries, including agriculture, thanks to this strong and versatile platform. Accurate predictions of plant growth, yield, and output have been made using various machine learning algorithms in recent years. For example, artificial neural networks, SVR, M5-prime regression trees, Random Forests, and K-Nearest Neighbors are the most effective algorithms (Culinarian et al., 2018). SVR and RF models are utilized as baseline models in this work to forecast plant production and growth.

Support vector regression (SVR)

Vapnik's Generalized Portrait technique was generalized nonlinearly to produce Support vector regression (SVR) (Cortes and Vapnik, 1995). It uses a kernel function to transform the input data into a higher-dimensional space and a hyperplane to segregate various types of data. The regularization parameter c regulates the trade-off between margin and mistakes. Radial basis kernel functions (Serb) use K(xi, Xu) = exp (xi – Xu–2) as the kernel function. As you can see, the radial basis function requires that y be a constant.

Random forest (RF)

Ensemble learning methods like RF date back to Ho's work from (Ho, 1998). RF utilizes decision trees as the ensemble's primary base learner. It is argued in ensemble learning that it is not possible to accurately forecast the intended value of test data by using just one predictor. The reason for this is because a single predictor cannot tell the difference between random noise and actual patterns based on sample data. A bootstrap sample of the training data is selected for each separate regression tree. Because of this, the regression tree grows until it reaches its maximum size. A weighted average of the predictions from each regression tree is used to arrive at final prediction values (Bierman 2001).

Deep learning (DL)

Traditional machine learning (ML) models are augmented with "depth" (complexity) and other functions that generate data representations hierarchically, via numerous abstraction layers. To get the most out of DL, it's important to use feature learning, which involves automatically extracting features from raw data and then composing higher-level features from the constituent parts. Due to the use of increasingly complicated models and huge parallelization, DL is able to tackle difficult problems quickly and efficiently. If the datasets characterizing the issue are big enough, the sophisticated models used in DL may improve classification accuracy or minimize regression error. Convolutions, pooling layers, fully connected layers, gates, memory cells, activation functions, encoding/decoding techniques, depending on the network architecture employed, such as Convolutional Neural Networks, Recurrent Neural Networks, Unsupervised Networks (Kamila is et al., 2018).

Long short-term memories (LSTM)

Initial development of the LSTM model was aimed at modelling long-term dependencies and establishing the best time delay for time series issues (Hochreiter and Schmid Huber, 1997). One input layer, one recurrent hidden layer, and one output layer make up an LSTM network. There are two types of adaptive, multiplicative gating units in the hidden layer: one controls information flow and contains memory cells with self-connections to save the temporal state. Constant Error Carousel (CEC) is the memory cell's primary recurrently self-connected linear unit, and the cell state is represented by the CEC activation. Learning to open and shut the multiplicative gates takes place over time. The vanishing gradient issue in LSTM can be overcome by keeping the network error constant. When learning lengthy time series, a forget gate is added to the memory cell, which prevents the gradient from expanding. An explanation of the workings of LSTM may be summarized in this way:

$$i_t=\sigma(w_i x_t+U_i m_{t-1}+b_i)$$

$$s_t = \tanh(w_s x_t+U_s m_{t-1}+b_s)$$

$$f_t = (w_s x_t+U_s m_{t-1}+b_s)$$

$$f_t=\sigma(w_f x_t+U_f m_{t-1}+b_f)$$

$$c_t=c_{t-1} \circ f_t+s_t \circ i_t$$

$$m_t=s_t \circ o_t$$

Input, forget, and output gates are marked by it, it, and yet, while the hidden state and cell state of the memory cell are represented by mt and CT respectively.

Microclimatic measurements

DL and ML models were applied to data from four cultivation tables in a 90 m2 greenhouse compartment at the Ornamental Plant Research Centre (PCS) in Nesselberger, Belgium, where we studied Ficus plants. Approximately 15 pots per m2 were filled with 3 cuttings each. Window apertures, a thermal screen, an air heating system, assimilation light, and a CO2 addition system were all used to regulate the greenhouse's environment. A time and radiation sum-controlled automated flood irrigation system was used to water the plants. The microclimate and irrigation controls were set at the same level as those used in commercial greenhouses. The greenhouse's microclimate was monitored constantly. A LI-190 Quantum Sensor (LI-COR, Lincoln, Nebraska, USA) and a carbon dioxide probe (Vaisala CARBOCAP GMP343, Vantaa, Finland) were used to detect PAR and CO2 concentration, respectively. A temperature and relative humidity probe (Campbell Scientific CS215, Logan, UT, USA) was put in a vented radiation shield to detect temperature and relative humidity. A linear variable displacement transducer (LVDT, Solectron, Bogner Regis, UK) sensor was used to continually measure the stem diameter of these plants. When the present stem diameter was compared to the diameter measured an hour earlier, the hourly fluctuation rate (mm d 1) was determined as the difference. Data from a UK Greenhouse farm was used to train the DL and ML models in the second experiment. This included both environmental (CO2, humidity, radiation, outside temperature, interior temperature) as well as yield real measurements. The yield was measured monthly, while environmental data were recorded hourly. Because of the irregularity of the weekly readings, we used data augmentation to provide daily data measurements. We also averaged the hourly environmental data to get daily representations that were comparable. These datasets were then divided into training, testing and validation sets in each studies. In all, 60 percent of the collected data was allocated to the training set, 15 percent to the validation set, and 25 percent to the test set.

Prediction evaluation

For assessing the accuracy of these forecasting models, researchers employed the MAE, RMSE, and MSE metrics (mean squared errors). In the following equations, the formulas for various assessment measures are shown:

$$MSE = \frac{1}{n}\sum_{t=1}^{n}\left(\frac{A_t - F_t}{A_t}\right)^2$$

$$MAE = \frac{1}{n}\sum_{t=1}^{n}\frac{|A_t - F_t|}{|A_t|}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{t=1}^{n}\left(\frac{A_t - F_t}{A_t}\right)^2}$$

where A_t is the actual value and F_t is the predicted value.

RESULTS AND DISCUSSION

Plant yield and growth in greenhouse conditions may be predicted using DL (LSTM), SVR, and RFR prediction models that we've created and validated. These models have been used to forecast Ficus growth using the SDV indicator, as well as tomato production predictions. The parameters of each model were determined using the grid search approach. The SVR model design relied heavily on the gamma and C parameters. The number of trees and the maximum depth of the tree were critical considerations in the creation of the RF model. For the DL LSTM model, it was critical to consider the number and size of hidden layers. There were three stages in the strategy that was put into action: • Data pre-processing and data cleansing. • Training, validation, and test datasets may be divided into three separate datasets. • Design and implementation of DL/LSTM, SVR, and RF models for one-step forward prediction. In both studies, the DL/LSTM model beats the SVR and RF models, demonstrating its superiority. MSE, RMSE, and MAE are shown in Table 1 for each of the three (trained) models on the test datasets in both studies. Table 1 compares the DL/LSTM model's performance to those of the SVR and RF models in predicting plant production and growth.

Datasets	Tomato Yield			Ficus Growth(SDV)		
Models	SVR	RF	LSTM	SVR	RF	LSTM
MSE	0.015	0.040	**0.002**	0.006	0.006	**0.001**
RMSE	0.125	0.200	**0.047**	0.073	0.062	**0.042**
MAE	0.087	0.192	**0.03**	0.070	0.063	**0.030**

Figure 1 shows how prediction models perform (RF, SVR and LSTM). In terms of predicting Ficus growth (SVD), the LSTM model beat both RF and SVR. RF and SVR models were unable to generalize as well as the LSTM model, as shown in Figure 2, which exhibited a greater capacity to capture the temporal aspect of the presented data.

CONCLUSIONS

Researchers in this study used an artificial neural network (ANN) to predict the SDV of Ficus and tomato yields, with good accuracy in both cases. In terms of MSE, RMSE, and MAE error criterion, experimental findings showed that the DL approach (using an LSTM model) outperformed other classic ML techniques like SVR and RF. That's why it's important to us that we apply deep learning methods to anticipate plant growth and productivity in a greenhouse setting. Next steps include: a) expanding the DL approach to conduct multi-step (weekly, or multiple-week) prediction of growth and yield in various greenhouses in the UK and Europe; and b) increasing the quantity of

gathered data that are utilized for training the suggested DL methods.

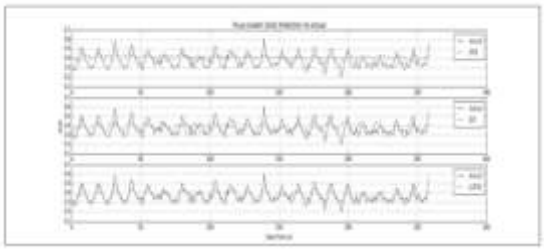

Figure 1. Testing results and performance comparison of Ficus growth (SVD) predictions.

ACKNOWLEDGEMENTS

The EU Interreg SMARTGREEN project is the source of this study (2017-2021). To the farmers (UK and EU) who provided the data, we would like to express our gratitude for their useful input, recommendations, and criticisms.

Literature cited.

Abreu, P., Meneses, J. & Gary, C. 1998, "Tomoses, 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. 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. Atanasova, N., Todorovski, L., Demoski, 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.

Barandiaran, I. 1998, "The random subspace method for constructing decision forests", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 20, no. 8. Bierman, L. 2001, "Random forests", Machine Learning, vol. 45, no. 1, pp. 5-32.

Buhmann, M.D. 2003, Radial basis functions: theory and implementations, Cambridge university press. Culinarian, A., Succories, 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. Cortes, C. &Vapnik, V. 1995, "Support-vector networks", Machine Learning, vol. 20, no. 3, pp. 273-297.

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. Duchesne, L. & Houle, D. 2011, "Modelling day-to-day stem diameter variation and annual growth of balsam fir (Abies balsamic (L.) Mill.) from daily climate",

Forest Ecology and Management, vol. 262, no. 5, pp. 863-872. Fan, X., Kang, M., Heuvel ink, E., De Riffey, P. & Hu, B. 2015, "A knowledge-and-data-driven modeling approach for simulating plant growth: A case study on tomato growth", Ecological Modelling, vol. 312, pp. 363-373. González-Chávez, Ma del Carmen A, Carrillo-González, R., Cuellar-Sánchez, A., Delgado-Alvarado, A., SuárezEspinosa, J., Ríos-Leal, E., Solís-Domínguez, F.A. & Maldonado-Mendoza, I.E. 2019, "Phytoremediation assisted by mycorrhizal fungi of a Mexican defunct lead-acid battery recycling site", Science of the Total Environment, vol. 650, pp. 3134-3144.

Heuvel ink, E. 1996, Tomato growth and yield: quantitative analysis and synthesis, Heuvel ink. Hinckley, T.M. &Bruckerhoff, D.N. 1975, "The effects of drought on water relations and stem shrinkage of Quercus alba", Canadian Journal of Botany, vol. 53, no. 1, pp. 62-72. Hochreiter, S. &Schmid Huber, J. 1997, "Long short-term memory", Neural computation, vol. 9, no. 8, pp. 1735-1780.

Jones, J., Kenig, A. &Vallejos, C. 1999, "Reduced state–variable tomato growth model", Transactions of the ASAE, vol. 42, no. 1, pp. 255. Kamila is, A. &Prenafeta-Boldú, F.X. 2018, "Deep learning in agriculture:

A survey", Computers and Electronics in Agriculture, vol. 147, pp. 70-90. Kanai, S., Adu-Gyamfi, J., Lei, K., Ito, J., Okura, K., Mogahed, R.E., El-Shemmy, H., Mohapatra, R., Mohapatra, P.K. &Shaneka, H. 2008, "N-deficiency damps out circadian rhythmic changes of stem diameter dynamics in tomato plant", Plant science, vol. 174, no.

2, pp. 183-191. Lacuna, Y., Bengio, Y. & Hinton, G. 2015, "Deep learning", Nature, vol. 521, no. 7553, pp. 436. Liakos, K., Busato, P., Mosholu, D., Pearson, S. and Botti's, D., 2018. Machine learning in agriculture: A review. Sensors, 18(8), p.2674. Moon, J.,

Berglund, L.J., Do mire, Z., An, K. & O'Driscoll, S.W. 2009, "Stem diameter and micromotion of press fit radial head prosthesis: a biomechanical study", Journal of shoulder and elbow surgery, vol. 18, no. 5, pp. 785-790. Pan, S.J. & Yang, Q. 2010, "A survey on transfer learning", IEEE Transactions on Knowledge and Data Engineering, vol. 22, no. 10, pp. 1345-1359. Poteau, R.,

Meyer, J., Tupuxuarid, R. & Stoll, B. 2012, "Support vector machines to map rare and endangered native plants in Pacific islands forests", Ecological Informatics, vol. 9, pp. 37-46. Qadim, K., Hines, E. & Iliescu,

D. 2013, "Yield prediction for tomato greenhouse using Fun", ISRN Artificial Intelligence, vol. 2013. Singh, A., Ganapathy Subramanian, B., Singh, A.K. & Sarkar, S. 2016, "Machine learning for high-throughput stress phenotyping in plants",

Trends in plant science, vol. 21, no. 2, pp. 110-124. Todorovski, L. &Demoski, S. 2006, "Integrating knowledge-driven and data-driven approaches to modeling", Ecological Modelling, vol. 194, no. 1-3, pp. 3-13. Vandeguchte, M.W.,

Guyot, A., Hubal, M., De Groote, S.R., De Beadmaker, N.J., Hayes, M., Welti, N., Lovelock, C.E., Lockington, D.A. &