ISSN:2229-6107

INTERNATIONAL JOURNAL OF PURE AND APPLIED SCIENCE & TECHNOLOGY

E-mail :
editor.ijpast@gmail.com
editor.ijpast@.in

www.ijpast.in

INTERNATIONAL JOURNAL OF PURE AND APPLIED SCIENCE & TECHNOLOGY
IJPAST

ISSN 2229-6107 www.ijpast.in

Vol 7,Issuse 2.May 2017

Utilizing Deep Learning to Forecast Crop Development and Yield in Greenhouse Environments

Ch Venkatesh, M Mnajula, P Kranthikiran, SAPAVATHU SUDHAKAR

Abstract

Greenhouse growers and farmers in general rely heavily on accurate predictions of plant growth and harvest. Improvements in environmental management, increased output, improved supply-and-demand matching, and reduced costs may all result from the development of models that can properly simulate growth and yield. Recent advances in ML, and especially Deep Learning (DL), may give robust new analytical tools. Researchers seek to employ ML and DL methods to estimate production and plant growth variance in two greenhouse settings: tomato yield forecasting and Ficus benjamina stem growth. In the prediction formulations, we use a novel deep recurrent neural network (RNN) based on the Long Short-Term Memory (LSTM) neuron model. In order to simulate the desired growth parameters, the RNN design takes into account both the historical values for yield, growth, and stem diameter, as well as the microclimate circumstances. We propose a research that uses the mean square error criteria to compare the results of several ML approaches, such as support vector regression and random forest regression. Positive findings are given based on information collected from two greenhouses in Belgium and the United Kingdom as part of the EU

Keywords: Prediction, deep learning, recurrent LSTM neural networks, growth, yield, tomato, ficus, stem diameter.

INTRODUCTION

Plant development, like many other biosystems, is a highly complex and dynamically coupled environmental system. Consequently, predicting growth and yield is a formidable scientific obstacle. There are several ways in which modeling strategies diverge (including, scale of interest, level of description, integration of environmental stress, etc.). It has been suggested that there are two primary ways to approach modeling: either via a "knowledge-driven" lens or a "data-driven" lens (Todorovski and Dzeroski, 2006; Atanasova et al., 2008). With a knowledge-driven strategy, you primarily make use of your pre-existing expertise in the field. Data-driven modeling, on the other hand, may create a model from raw data without any prior domain expertise.Examples of data-driven models (DDM) include traditional approaches to machine learning including artificial neural networks, support vector machines, and generalized linear models (Pouteau et al., 2012). Those techniques have several positive traits, such as being able to approximate nonlinear functions, being highly predictive, and being flexible enough to respond to inputs from a multivariate system with relative ease

ASSISTANT PROFESSOR[1,2,3], STUDENT[4]
Department of CSE
Arjun College Of Technology & Sciences
Approved by AICTE& Affiliated to JNTUH

SPONSORED BY BRILLIANT BELLS EDUCATIONAL SCOITEY

(Buhmann, 2003). Machine learning (ML), linear polarizations, wavelet-based filtering, vegetation indices (NDVI), and regression analysis are among the most used methods for assessing agricultural data (Singh et al., 2016; Liakos et al., 2018). Yet, aside these methods, deep learning (DL) is a relatively new approach that has been gaining popularity (Goodfellow et al., 2016). When it comes to computing, DL is quite similar to ANN and both are part of the machine learning discipline. On the other hand, DL is all about "deeper" neural networks that provide a hierarchical data representation through a number of operations. Because of this, we may expand our capacity for learning, leading to enhanced efficiency and accuracy. Feature learning, or the automated extraction of features from raw data, is a major benefit of DL. Features at higher levels of the hierarchy are produced by the composition of features at lower levels of the hierarchy (Goodfellow et al., 2016). Due to the progressively complicated nature of the associated models, DL excels at solving issues of increasing complexity (Pan and Yang, 2010). If sufficiently big data-sets are available, the sophisticated models used in DL may improve classification accuracy and decrease error in regression issues.

Multiple Linear Regression, Artificial Neural Networks, Support Vector Regression, K-Nearest Neighbor ML Techniques, and other methods were compared for their ability to forecast agricultural yields in 10 different datasets by Gonzalez-Sanchez et al. (2019). Accuracy measures used to verify the models included the Root Mean Square Error (RMS), Root Relative Square Error (RRSE), Normalized Mean Absolute Error (MAE), and Correlation Factor (R). The results indicated that among all of the generated crop yield models, M5-Prime had the lowest error rate. In that research, the methods were rated from best to worst based on RMSE, RRSE, R, and MAE, with M5-Prime, kNN, SVR, ANN, and MLR coming out on top. In a separate investigation, (Nair and Yang-Won, 2016) estimated corn yield in the state of Iowa using SVM, Random Forest (RF), Extremely Randomized Trees (ERT), and Deep Learning (DL) as four ML methods. To avoid the overfitting issue, DL offered more consistent outcomes, as shown by validation statistical comparisons.

Plants' stem diameter is one of the most significant metrics used to characterize their development throughout the vegetative phase. In addition, stem diameter variation has been employed in a variety of species to develop proxies for plant water status, which is then used in optimization algorithms for plant-based irrigation scheduling. Plant stem diameter variation (SDV) is the day-to-day fluctuation in stem size that occurs in all plants. This diurnal variation is directly correlated with the amount of water available to plants and may be used as a gauge of how much water the plant is receiving. Cultivated plants depend on photosynthesis and the transport of carbohydrates from the site of synthesis to the sink organs for energy throughout the vegetative growth and development stages (Yu et al., 2015). Quite a few sources cover the basics of stem diameter variations (Vandegehuchet et al., 2014). The known sensitivity of SDV to water and nutrient conditions and its tight relationship to the responses of agricultural plants to changes in environmental circumstances suggest that SDV may be an important factor in the success or failure of crop production (Kanai et al., 2008). Diameter of the stem during the vegetative growth stage is a key characteristic in characterizing the development of agricultural plants when subjected to abiotic stress. As a result, models of plant growth and development and environmental factors that affect stem diameter (SDV) are necessary. Researchers have found that SDV models are useful for determining how the environment affects agricultural yields, but they say they need to be reviewed and improved (Hinckley and Bruckerhoff, 2011). Predictions of the yearly growth rate of balsam fir (Abies balsamea L.) may now be made with high precision thanks to SDV daily models (Duchesene and Houle, 2011). Predictions of the potential growth response to climate may be enhanced by include daily data in growth-climate models, since this allows for the identification of specific climatic occurrences that may be missed using a conventional dendroclimatic method (Duchesene and Houle, 2011). So far, however, environmental variable-based models for forecasting SDV and plant development have been inadequate.

Few models have been investigated for the complex and dynamic system that is tomato cultivation in a greenhouse. TOMGRO and TOMSIM (Jones et al., 1999) and (Heuvelink, 1996) are two of the most widely used dynamic growth models in the literature. Those models depict biomass partitioning, crop growth, and yield as a function of a number of climatic and physiological factors, all of which are reliant on physiological processes. Growers have been slow to adopt these technologies because of their complexity, lack of generalizability, and the challenges of predicting values for starting parameters and validating in different circumstances.

Tomato production in terms of fruit weight was established as the Tompousse model by (Abreu et al., 2000). The model was created by investigating the correlation between climate control variables in southern French greenhouses. This model relied on the assumption that there was a linear connection between the pace at which flowers bloomed and the size of the resulting fruit. Testing the concept in Portugal's unheated plastic greenhouses yielded disappointing results. Adams (Adams, 2002) suggested a different approach for estimating tomato yields using a graphical simulation tool. The primary motivation behind the model was to capture variations in greenhouse tomato output on a weekly basis, specifically with regards to fruit size and harvest frequency. The development of the leaf truss and the number of flowers produced were estimated using hourly climatic data. Solar radiation and air temperature went through cycles, which in turn influenced yield swings throughout the year. Many resources exist to aid farmers in decision-making, as stated by (Qaddoum et al., 2013). These have the potential to estimate yield rates, give guidance on climate management measures, and coordinate harvesting with consumer demand.

In this research, we present a deep learning model that can accurately predict either ficus stem diameter or tomato yield given environmental (CO_2, humidity, radiation, outside temperature, interior temperature), as well as actual yield and stem diameter fluctuation observations. The remaining sections of this work are laid out as follows. Brief descriptions of the proposed methodology and the used datasets are provided in Section 2. Results are presented in Section 3, and discussion of the findings and recommendations for further research are provided in Section 4.

MATERIALS AND METHODS

Conventional Machine Learning

Machine learning (ML) methods have the potential to solve complex non-linear issues using data from a variety of sources, which is a major benefit. When applied to real-world circumstances, ML allows for more informed decision making and action with little to no human participation. It offers a robust and adaptable framework for making decisions based on data, which has many potential uses outside of agriculture. Recent years have seen the use of several ML algorithms to accurately estimate plant growth, yield, and output across a variety of crops. Artificial Neural Networks, Support Vector Regression, M5-prime Regression Trees, Random Forests, and K-Nearest Neighbors are some of the most effective methods (Chlingaryan et al., 2018). In this article, we employ SVR and RF models as our baseline for estimating future crop yields and plant sizes.

Support vector regression (SVR)

After Vapnik's Generalized Portrait method, support vector regression (SVR) was created as a nonlinear version of the approach (Cortes and Vapnik, 1995). Through the use of a kernel function and a hyperplane, it projects the input data into a higher dimensional space, where it may be used to divide the data into several categories. The regularization parameter c regulates the balance between margin and errors. Using radial-basis-function (rbf) kernels, SVR (system for visually recognizing faces) $K(xi,xj)=\exp(-y\|xi-xj\|2)$. Here y is a constant used in the radial basis function.

Random forest (RF)

Originally developed by Ho in, RF is a kind of ensemble learning method (Ho, 1998). RF employs decision trees as the ensemble's root learner. The basic premise of ensemble learning is that it takes more than one predictor to reliably estimate a target value from a set of test data. The reason for this is because a single predictor cannot separate noise from patterns in sample data. RF builds many separate regression trees, selecting a bootstrap sample of the training data at each tree node. As a result, data is continuously added to the regression tree until it has expanded to its maximum size. Instead, the sum of all the predicted values of the regression trees is used to form the final prediction value (Breiman 2001).

Deep learning (DL)

By incorporating more "depth" (complexity) into the model and processing the input using a number of algorithms that provide hierarchical data representations through many levels of abstraction, Deep Learning goes beyond traditional ML. Feature learning, or the automated extraction of features from raw data, is one of DL's main benefits. Features at higher levels of the hierarchy are produced through composition of features at lower levels. Due to the more complicated models utilized, which also allow for huge parallelization, DL is able to tackle difficult problems exceptionally effectively and quickly. With sufficiently big datasets available, the complicated models used in DL may improve classification accuracy or decrease error in regression tasks. Convolutions, pooling layers, fully connected layers, gates, memory cells, activation functions, encoding/decoding schemes, and so on are all part of DL. These components vary depending on the network architecture used (Convolutional Neural Networks, Recurrent Neural Networks, Unsupervised Networks, etc). (Kamilaris et al., 2018).

Long short-term memories (LSTM)

It was with the intention of modeling long-term dependencies and pinpointing the best time lag for time series issues that the LSTM model was first presented in (Hochreiter and Schmidhuber, 1997). A long short-term memory (LSTM) network has three layers: input, hidden, and output. The memory block is the fundamental unit of the hidden layer, consisting of memory cells with self-connections that store the current state in time and a pair of adaptive, multiplicative gating units that regulate the flow of information inside the block. Cell state is indicated by the degree of activation of a linear unit inside the memory cell, which is referred to as the Constant Error Carousel (CEC). When to open and when to shut, the multiplicative gates are taught. The vanishing gradient issue may be addressed in LSTM by fixing the network error to a fixed value. When learning lengthy time series, a forget gate is included into the memory cell to stop the gradient from exploding. You may summarize LSTM's structure and functioning as follows:

$$i_t=\sigma(w_i x_t+\mathrm{U}_i m_{t-1}+\mathrm{b}_i)$$

$$s_t = \tanh(w_s x_t+\mathrm{U}_s m_{t-1}+\mathrm{b}_s)$$

$$f_t = (w_s x_t+\mathrm{U}_s m_{t-1}+\mathrm{b}_s)$$

$$f_t=\sigma\left(w_f x_t+\mathrm{U}_f m_{t-1}+\mathrm{b}_f\right)$$

$$c_t=\mathrm{c}_{t-1} \circ f_t+\mathrm{s}_t \circ i_t$$

$$m_t=\mathrm{s}_t \circ o_t$$

where it, $it$and$yt$are denoted as input gate, forget gate and output gate at time t respectively, and$mt$and ctrepresent the hidden state and cell state of the memory cell at time t.

Microclimatic measurements

As a first experiment, we used the DL and ML models on information about Ficus plants (Ficus benjammina) grown on four tables in a 90m2greenhouse section at the Ornamental Plant Research Centre (PCS) in Destelbergen, Belgium. It was estimated that there were fifteen pots per square meter of plant space, and that each pot had three cuttings. The greenhouse's microclimate was adjusted by adjusting the openings of the greenhouse's windows, the temperature of the air heating system, the intensity of the assimilation light, and the amount of carbon dioxide (CO2) added to the air. Automatic flood irrigation was used to water the plants, with the timing and total radiation doses being the determining factors. The microclimate and irrigation control settings mimicked those of professional greenhouses. The greenhouse's microclimate was tracked in real time. The levels of carbon dioxide (CO2) and photosynthetically active radiation (PAR) were measured using a Vaisala CARBOCAP GMP343 carbon dioxide probe and a LI-190 Quantum Sensor (both from LI-COR, Lincoln, Nebraska, USA). A temperature and relative humidity probe (Campbell Scientific CS215, Logan, UT, USA) was set up in a vented radiation shield to take readings. A linear variable displacement transducer (LVDT, Solartron, Bognor Regis, UK) sensor was used to track the diameter of the stems of these plants in real time. The rate of change in stem diameter (mm h1) was estimated by taking the difference between the present stem diameter and the stem diameter measured an hour earlier for a specific time point.

In the second trial, environmental (CO_2, humidity, radiation, outside temperature, internal temperature), as well as yield, real measurements were used to train DL and ML models. In contrast to the weekly frequency with which the yield was recorded, the environmental data were taken every hour. We did data augmentation, interpolating weekly data to provide daily data measurements, to address these data features. Similarly, we averaged the environmental data collected every hour so that we could have consistent daily representations.

Separate datasets were used for training, testing, and validating in each trials. Sixty percent of the data collected was used to create the training set, fifteen percent the validation set, and twenty-five percent the test set.

Prediction evaluation

These prediction models have been evaluated using the Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and Mean Squared Error (MSE). The equations below illustrate the formulas of these several metrics used for evaluation:

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

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

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

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

RESULTS AND DISCUSSION

In order to forecast plant growth and yield in greenhouse settings, we have built and evaluated DL (LSTM), SVR, and RFR prediction models for: a) ficus growth prediction using the SDV indicator, and b) tomato yield prediction. The parameters of each model were determined using the popular grid search technique. For the SVR model development, the gamma and C parameters were critical. When developing the RF model, it was crucial to take into account both the total number of trees and the maximum depth of the tree. For the DL LSTM model, it was crucial to get the details right about the number and size of hidden layers.

There were three stages to the method that was ultimately used:

Data cleansing and preparation is a must.

• Separating data into a training set, a validation set, and a test set.

Models using DL/LSTM, SVR, and RF may be created and used to make a prediction one step into the future.

As can be seen from the outcomes of both tests, the DL/LSTM model performs better than the SVR and RF models. Table 1 displays the accuracy (in MSE, RMSE, and MAE) achieved in both trials by applying each of the three (trained) models to the test datasets.

Table 1. Performance of the DL/LSTM model compared to those of SVR and RF models for plant yield and growth prediction.

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**

In Figure 1 we see how well various prediction models do (RF, SVR and LSTM). It's evident that the LSTM model beat the RF and SVR implementations in predicting Ficus growth (SVD). Figure 2 demonstrates how the LSTM model generalized better than the RF and SVR models by following the trend of the actual yield value and storing a better representation of the temporal nature of the provided data.

CONCLUSIONS

In this research, we describe a deep learning (DL) strategy that makes use of LSTM to accurately forecast the development of a Ficus tree (represented by the standard deviation vector; SDV) and the production of tomatoes. In terms of mean squared error (MSE), root mean squared error (RMSE), and mean absolute error (MAE), experimental findings showed that the DL approach (using an LSTM model) outperformed other conventional ML techniques like SVR and RF. As a result, our project's primary objective is to create DL methods for forecasting plant growth and production in greenhouse settings. Research into the long-term effects of a) increasing the quantity of data used to train the proposed DL methods and b) expanding the DL method to perform multi-step (at a weekly, or a multiple of weeks basis) prediction of growth and yield in a wide range of greenhouses in the UK and Europe is warranted.

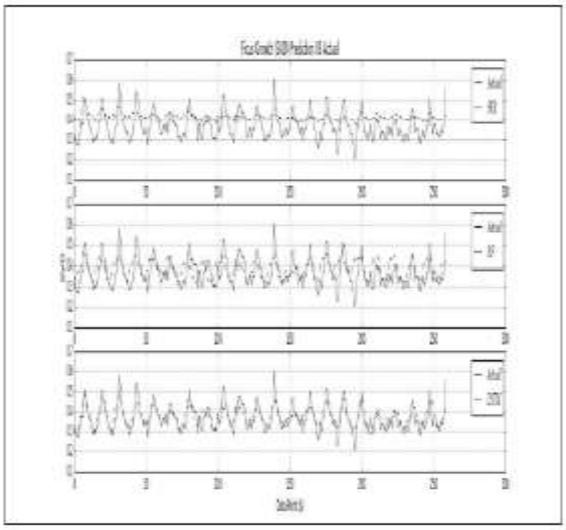

Figure 1. Testing results and performance comparison of Ficus growth (SVD) predictions.

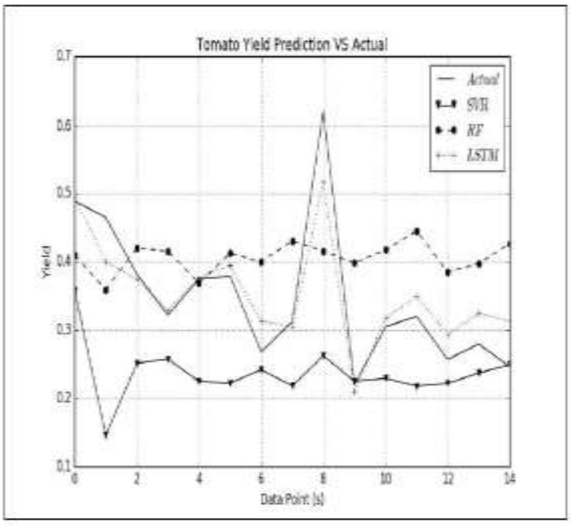

Figure 2. Testing results and performance comparison of Tomato Yield predictions.

Literature cited

1. In 1998, Abreu, Meneses, and Gary published "Tompousse, a model of yield prediction for tomato crops: calibration study for unheated plastic greenhouses" in XXV International Horticultural Congress, Part 9: Computers and Automation, Electronic Information in Horticulture 519, pp. 141.

2. IV International Symposium on Models for Plant Growth and Control in Greenhouses: Modeling for the 21st Century-Agronomic and 593, pp. 19. Adams, S. 2001. "Predicting the weekly changes in glasshouse tomato yields."

3. 2008, "Application of automated model discovery from data and expert knowledge to a real-world domain: Lake Glums," Ecological Modelling, vol. 212, no. 1-2, pp. 92-98. Atanasova, N., Todorovski, L., Deroski, S., & Kompare, B.

4. Random subspace approach for creating decision forests. Barandiaran, I. 1998. IEEE Transactions on Pattern Analysis and Machine Intelligence, volume 20, issue 8.

5. Breiman, L. 2001, "Random forests", *Machine Learning*, vol. 45, no. 1, pp. 5-32.
6. Buhmann, M.D. 2003, *Radial basis functions: theory and implementations*, Cambridge university press.

7. 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.
8. Cortes, C. & Vapnik, V. 1995, "Support-vector networks", *Machine Learning,* vol. 20, no. 3, pp. 273-297.
9. 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.
10. Duchesne, L. & Houle, D. 2011, "Modelling day-to-day stem diameter variation and annual growth of balsam fir (Abies balsamea (L.) Mill.) from daily climate", *Forest Ecology and Management,* vol. 262, no. 5, pp. 863-872.
11. Fan, X., Kang, M., Heuvelink, E., De Reffye, 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.
12. González-Chávez, Ma del Carmen A, Carrillo-González, R., Cuellar-Sánchez, A., Delgado-Alvarado, A., Suárez-Espinosa, 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.
13. Heuvelink, E. 1996, *Tomato growth and yield: quantitative analysis and synthesis,* Heuvelink.
14. 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.
15. Hochreiter, S. & Schmidhuber, J. 1997, "Long short-term memory", *Neural computation,* vol. 9, no. 8, pp. 1735-1780.
16. Jones, J., Kenig, A. & Vallejos, C. 1999, "Reduced state–variable tomato growth model", *Transactions of the ASAE,* vol. 42, no. 1, pp. 255.
17. Kamilaris, A. & Prenafeta-Boldú, F.X. 2018, "Deep learning in agriculture: A survey", *Computers and Electronics in Agriculture,* vol. 147, pp. 70-90.
18. Kanai, S., Adu-Gymfi, J., Lei, K., Ito, J., Ohkura, K., Moghaieb, R.E., El-Shemy, H., Mohapatra, R., Mohapatra, P.K. & Saneoka, 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.
19. LeCun, Y., Bengio, Y. & Hinton, G. 2015, "Deep learning", *Nature,* vol. 521, no. 7553, pp. 436.
20. Liakos, K., Busato, P., Moshou, D., Pearson, S. and Bochtis, D., 2018. Machine learning in agriculture: A review. *Sensors*, *18*(8), p.2674.
21. Moon, J., Berglund, L.J., Domire, 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.
22. 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.
23. Pouteau, R., Meyer, J., Taputuarai, 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.
24. Qaddoum, K., Hines, E. & Iliescu, D. 2013, "Yield prediction for tomato greenhouse using EFuNN", *ISRN Artificial Intelligence,* vol. 2013.
25. Singh, A., Ganapathysubramanian, 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.
26. Todorovski, L. & Džeroski, S. 2006, "Integrating knowledge-driven and data-driven approaches to modeling", *Ecological Modelling,* vol. 194, no. 1-3, pp. 3-13.
27. Vandegehuchte, M.W., Guyot, A., Hubau, M., De Groote, S.R., De Baerdemaeker, N.J., Hayes, M., Welti, N., Lovelock, C.E., Lockington, D.A. & Steppe, K. 2014, "Long-term versus daily stem diameter variation in co-occurring mangrove species: Environmental versus ecophysiological drivers", *Agricultural and Forest Meteorology,* vol. 192, pp. 51-58.
28. Yu, S., Lo, S. & Ho, T.D. 2015, "Source–sink communication: regulated by hormone, nutrient, and stress cross-signaling", *Trends in plant science,* vol. 20, no. 12, pp. 844-857.