Enhancing efficiency in closed agricultural greenhouses: A data-driven predictive model for energy consumption

Ikhlas Ghiat and Tareq Al-Ansari*

College of Science and Engineering, Hamad Bin Khalifa University, Qatar Foundation, Doha P.O. Box 5825, Qatar

*E-mail: talansari@hbku.edu.qa

Abstract. Predicting energy consumption in agricultural greenhouses is essential to effectively allocate resources, enhance plant growth, and minimize energy inefficiencies. Various factors affect the energy consumption inside the greenhouse including external climate conditions and internal microclimate. Proper understanding of these factors is crucial for maintaining an ideal growing environment and optimizing energy efficiency. This drives the need to investigate the interaction between these factors and greenhouse energy consumption, encompassing the energy needed for cooling and the supply of water and nutrients. This work aims at developing a dynamic model that predicts the total energy consumption of a closed agricultural greenhouse to improve microclimate control and energy efficiency. The study is conducted within a closed-loop agricultural greenhouse with no natural ventilation. Inside, the air is cooled and continuously circulated without being exchanged with ambient air through a heating, ventilation, and air conditioning (HVAC) system. The data-driven model encompasses external climate parameters such solar radiation, ambient temperature, and relative humidity; along with microclimate parameters such as internal temperature, humidity, and CO_2 concentration to predict overall energy consumption. The study examines two machine learning models, deep neural networks (DNN) and extreme gradient boosting (XGBoost), for forecasting energy consumption, and assesses their performance using the coefficient of determination (R^2), the root mean square error (RMSE), and the mean absolute error (MAE). Results reveal that the DNN model surpasses the XGBoost model, exhibiting a superior predictive performance with an R^2 of 80.9%, RMSE of 171.1 kWh and MAE of 130.3 kWh. This study demonstrates its practicality in assisting with energy consumption analyses and identifying inefficient energy usage patterns within closed agricultural greenhouses.

Keywords: Machine learning, Energy consumption, Agriculture, Greenhouse, Deep neural networks, Extreme gradient boosting.

1. Introduction

Water resources face significant pressures and environmental challenges, particularly within the agricultural sector, underscoring the urgent need for smart and sustainable solutions [1,2]. Greenhouses emerge as a sustainable solution, offering a fivefold increase in water use efficiency compared to open fields. The concept of closed or semi-closed greenhouse systems, strategically designed for microclimate management, represents a revolutionary shift in agricultural practices. The deployment of various cooling systems, including evaporative cooling and heating, ventilation, and air conditioning (HVAC), proves indispensable for hot regions, requiring precise control over microclimate conditions.

Content from this work may be used under the terms of the Creative Commons Attribution 4.0 licence. Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.
Published under licence by IOP Publishing Ltd

However, there is a need to reduce natural ventilation usage to maintain high CO_2 levels achieved through enrichment practices. CO_2 enrichment in greenhouses not only boosts yield but also reduces water consumption through decreased evapotranspiration. Various methods, such as carbon capture and on-site CO_2 production using burners, are employed for CO_2 enrichment [1,3]. Operating a semi-closed to closed greenhouse provides additional advantages, including superior microclimate control, reduced impact from external factors like insects and fungi, leading to decreased pesticide use and water usage—a crucial consideration for arid and hyper-arid regions [4]. Recognizing that energy consumption can contribute up to 50% of production costs in greenhouses [5], accurate prediction of energy consumption becomes imperative for enhancing resource management.

The application of machine learning (ML) to forecast greenhouse parameters linked to energy consumption has gained considerable attention. Notably, Patil et al. [6] utilized external climatic data to estimate internal climatic conditions in a tropical greenhouse, employing an auto-regressive neural network. Similarly, studies have employed diverse ML techniques for predicting microclimate in greenhouses, such as in hyper-arid regions, using methods such as support vector regression (SVR), and a combination of computational fluid dynamics (CFD) and a reduced-order modelling approach [7,8]. Other studies delved into the application of various ML models, including extreme gradient boosting (XGBoost), Deep Neural Networks (DNN), and SVR, to predict parameters crucially impacting energy consumption in greenhouses, such as transpiration [9]. Recognizing the crucial role of accurate prediction in control strategies, methodologies like model predictive control (MPC) have emerged as viable solutions for greenhouse operations. Different ML models, including Artificial Neural Networks (ANN), SVR, and Long Short-Term Memory (LSTM), have been integrated into MPC frameworks to forecast greenhouse microclimate, enabling control over diverse variables such as fan and HVAC parameters [10,11].

This work involved directly predicting energy consumption in greenhouses situated in hyper-arid regions using external weather and microclimate data, employing two machine learning models, XGBoost and DNN. The unique focus on forecasting variations in total energy consumption, including energy for cooling and irrigation, under diverse environmental conditions aims to enhance resource management efficiency and minimize operational costs of the greenhouse.

2. Materials and methods

This study entailed the prediction of the total energy consumption of an agricultural greenhouse located in a hyper-arid region. A data-driven approach is adopted; whereby external weather conditions along with microclimate conditions are used as predictors for the total energy consumption in the greenhouse using two machine learning models. The total energy consumption involves the energy required for cooling, along with the energy required for irrigation supply and distribution within the greenhouse. External weather conditions are directly linked to the energy required for cooling to maintain the optimal microclimate conditions. Additionally, solar radiation and microclimate conditions directly impact the water consumption of the plants through transpiration which translates into energy consumption for irrigation. This later includes the energy required for pumping and desalinating brackish water and that of irrigation supply and distribution (figure 1).

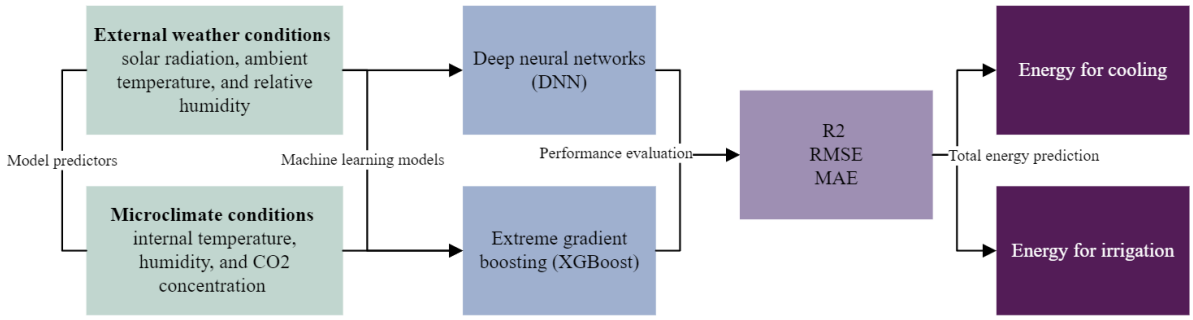

Figure 1. Methodological framework for predicting total energy consumption in greenhouses.

2.1. System description

This work was conducted in a cooling-based agricultural greenhouse located in the state of Qatar. The greenhouse features a closed system by incorporating a heating ventilation and air conditioning (HVAC) system for microclimate control. Additionally, the greenhouse employs a hydroponic setting with drip irrigation, with water supplied between sunrise and sunset to ensure drainage at night and efficient water usage. The drainage water accounts for 30% of irrigation water to help maintain optimal soil moisture levels. A solution containing Nitrogen, Phosphorous, and Potassium was utilized as part of the fertigation system to support plant growth. Irrigation water was supplied from brackish groundwater with an initial salinity of 5 – 20 mS/cm which is desalinated to 0.2 mS/cm using reverse osmosis. The greenhouse cooling and irrigation systems require significant amounts of energy, which highlights the importance of precise energy consumption predictions for operational planning and resource management in this application.

External weather and microclimate data were collected over a one-year period at 5-minute intervals, then averaged daily. A weather station captured external weather conditions (solar radiation, outside temperature and relative humidity), while internal sensors within the greenhouse provided microclimate data (internal temperature, relative humidity and CO_2 concentration). The daily energy consumption of cooling and irrigation systems was recorded throughout the year.

2.2. Model development

In this work, two machine learning models were examined to forecast the total energy consumption in the greenhouse: Extreme Gradient Boosting (XGBoost) and Deep Neural Networks (DNN). The XGBoost model is a scalable implementation of the gradient boosting algorithm that builds numerous decision trees in sequence, with each subsequent tree aimed at correcting the errors of the previous one to progressively enhance the accuracy of the prediction [12]. The DNN model consists of a network of multiple interconnected layers that learn intricate patterns in the data. The network consists of an input layer that receives the data, one or more hidden layers, and an output layer that produces the predictions. The DNN model introduces non-linearities through activation functions that are embedded in these layers and that can learn complex relationships within the data. Additionally, the use of optimizers helps fine-tune the learning process of the DNN model by adjusting weights to minimize the loss function [9]. In this work, the Rectified Linear Unit (Relu) and the hyperbolic tangent (tanh) activation functions are applied to the DNN's input layer and the hidden layers respectively. The Adam optimiser is employed to fine-tune the model's training process.

Three feature combinations were examined in this work to explore their effect on predicting the total energy consumption. The first set entailed feeding the models with only external weather data encompassing solar radiation, outside temperature, and outside relative humidity. The second set comprised microclimate data, including the temperature, relative humidity and CO_2 concentration inside the greenhouse. The third set combined both external weather data and microclimate data to investigate their collective impact on the performance of the models. The dataset was randomly divided into 80% for training and 20% for testing. The XGBoost and DNN models were implemented using the xgboost,

and Keras libraries respectively in Python v3.7. A hyperparameter tuning with a five-fold cross validation was implemented through the GridSearchCV tool for both models. The hyperparameters investigated for both models are given in table 1.

Table 1. Hyperparameter tuning.

Model	**Hyperparameter ranges**				
XGBoost	Learning rate	Number of rounds	Subsample	Max depth	Min child weight
	0.01, 0.1, 0.5, 1	200, 400, 600	0.5, 0.7	6, 8, 10, 12	1, 2, 3
DNN	Batch size	Epochs	Layers	Neurons	-
	5, 10, 20, 30	10, 50, 100, 200	1, 2 ,3, 4	5, 10, 20, 50	-

The performance of the two models was examined using three statistical indicators encompassing the coefficient of determination (R^2), the root mean square error (RMSE), and the mean absolute error (MAE) (equations 1-3).

$$R^2 = \frac{\sum_{i=1}^{n}[(\hat{y}_i-\bar{\hat{y}}_i)(y_i-\bar{y}_i)]^2}{\sum_{i=1}^{n}(\hat{y}_i-\bar{\hat{y}}_i)^2 \sum_{i=1}^{n}(y_i-\bar{y}_i)^2} \tag{1}$$

$$RMSE = \sqrt{\sum_{i=1}^{n}\frac{(y_i-\hat{y}_i)^2}{n}} \tag{2}$$

$$MAE = \frac{\sum_{i=1}^{n}|y_i-\hat{y}_i|}{n} \tag{3}$$

3. Results and discussion

The hyperparameter tuning resulted in a 2-layer DNN model with 20 neurons in each layer, a batch size of 5, and an epochs number of 100. The optimal hyperparameters for the XGBoost were found as a learning rate of 0.1, 600 rounds, a subsample of 0.5, a maximum depth of 12, and a minimum child weight of 3. Additionally, the optimal regularization parameters were identified as 0 for gamma, 1 for both the L1 and L2 regularization.

The comparison between the XGBoost and DNN models incorporating different feature combinations of external weather and microclimate data revealed different performance trends. In general, the DNN consistently outperformed the XGBoost on the testing datasets, across the varying input subsets, demonstrating higher predictive accuracy. When integrating both external weather and microclimate data, both models exhibited significant performance improvements, with the DNN showcasing superior predictive capabilities in terms of R^2 score (80.9%), RMSE (171.1 kWh), and MAE (130.3 kWh) on the testing data (table 2). While the XGBoost displayed tendencies towards overfitting by performing better on the training data, the DNN proved more adept at generalizing and predicting accurately across combined feature sets. figure 2, displaying actual versus predicted values, supports the DNN model's ability to accurately predict energy consumption as it shows the model's ability to align closely with actual values across a diverse range of data points.

Table 2. Model performance results.

Models and Inputs	Testing			Training		
	R^2 (%)	RMSE (kWh)	MAE (kWh)	R^2 (%)	RMSE (kWh)	MAE (kWh)
External weather data: solar radiation, outside temperature, outside relative humidity						
XGBoost	45.5	289.3	234.4	74.4	207.4	146.7
DNN	46.9	285.4	224.3	50.3	286.1	215.2
Microclimate data: greenhouse temperature, greenhouse relative humidity, greenhouse CO_2 concentration						
XGBoost	54.7	263.9	204.9	73.9	209.3	156.4
DNN	56.9	257.3	202.1	57.5	264.6	201.5
External weather and microclimate data: Solar radiation, outside temperature, outside relative humidity, greenhouse temperature, greenhouse relative humidity, greenhouse CO_2 concentration						
XGBoost	73.8	200.3	157.9	87.8	142.9	96.5
DNN	80.9	171.1	130.3	78.6	187.7	136.2

The investigation into the performance of the XGBoost model, despite employing hyperparameter tuning, early stopping with 10 rounds, 5-fold cross-validation, and L1/L2 regularization, reveals a trend of higher performance on the training set than the testing set. Despite measures to counter overfitting tendencies, the model's capacity to learn intricate patterns within the training data might lead to overfitting, impacting its ability to generalize well to unseen test data. The discrepancy between training and testing performances (e.g., R^2 of 73.8% vs 87.8%), although not severe, suggests that while attempted measures such as early stopping and regularization were in place, there might still be some level of overfitting. The DNN model showcasing lower accuracy on the training set as compared to the testing set suggests a more balanced learning approach (table 2). The lower training performance indicates that the model refrains from overfitting the data. Simultaneously, the better testing performance suggests its ability to generalize and predict accurately on unseen data, reducing the risk of underfitting. This balanced performance indicates the DNN model's suitability for predicting the total energy consumption in the greenhouse, demonstrating its capacity to capture underlying patterns in the data without compromising its ability to generalize to unseen datasets. Precise energy consumption predictions are vital for greenhouse applications for efficient resource management. The DNN model has the capability to understand the complex interplay within external weather conditions, microclimate dynamics, and energy consumption which gives valuable insights for optimising energy usage, thereby reducing operational costs and enhancing the overall productivity in greenhouses.

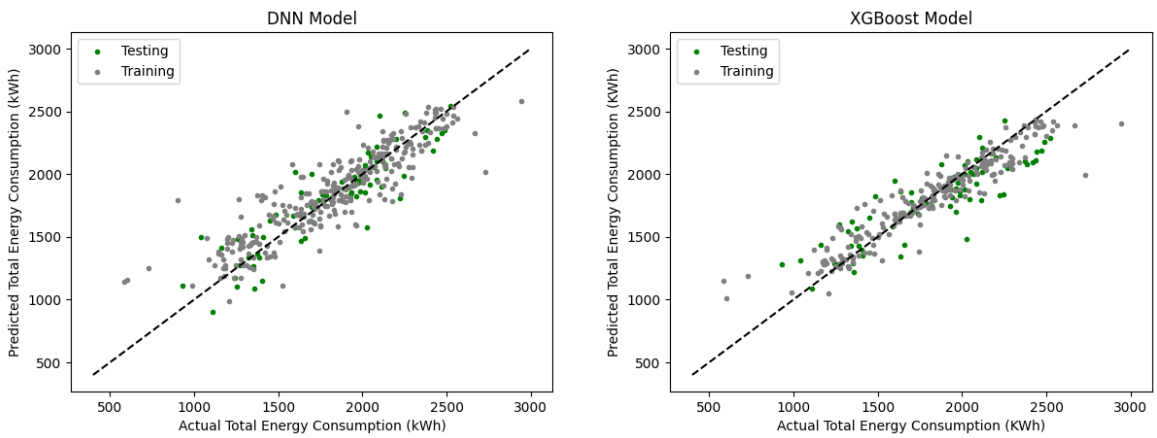

Figure 2. Actual Vs predicted total energy consumption.

Accurately predicting the change in total energy consumption with varying environmental conditions will ensure efficient resource management, help minimize operational costs, and reduce the environmental footprint associated with greenhouse operations. This will build resilient greenhouse operations by supporting early planning for resources and for alternative supplies in case of extreme events or disruptions. This could also help comply with regulations and avail certain incentives related to energy efficiency and environmental sustainability in agricultural practices. Energy is needed to control microclimate and irrigation in the greenhouse, so predicting energy will help ensure that the greenhouse provides ideal, optimal and stable conditions for these factors and minimize events or conditions that can stress or damage the plants. Moreover, overwatering or underwatering of plants can be prevented by using energy predictions to inform irrigation schedules and strategies.

4. Conclusion

This study delves into the complex interplay of external climate conditions and internal microclimate factors to develop a dynamic model for predicting total energy consumption in a closed-loop agricultural greenhouse. Two machine learning models, extreme gradient boosting (XGBoost), and deep neural networks (DNN) were employed for total energy consumption forecasting, with the DNN model exhibiting superior predictive performance compared to the XGBoost, with an R^2 of 80.9%, RMSE of 171.1 kWh and MAE of 130.3 kWh. Results reveal the superior predictive performance of the DNN model over the XGBoost and its practical utility in analysing and optimizing energy usage patterns within closed agricultural greenhouses.

5. References

[1] Ghiat I, Mahmood F, Govindan R and Al-Ansari T 2021 CO2 utilisation in agricultural greenhouses: A novel 'plant to plant' approach driven by bioenergy with carbon capture systems within the energy, water and food Nexus *Energy Convers. Manag.* **228** 113668
[2] Lahlou F-Z, Mackey H R and Al-Ansari T 2023 Towards the development of an improved mass balance and water quality index based grey water footprint model *Environ. Sustain. Indic.* **18** 100236
[3] Ghiat I, AlNouss A, McKay G and Al-Ansari T 2020 Modelling and simulation of a biomass-based integrated gasification combined cycle with carbon capture: Comparison between monoethanolamine and potassium carbonate *IOP Conf. Ser. Earth Environ. Sci.* **463**
[4] Katsoulas N, Sapounas A, De Zwart F, Dieleman J A and Stanghellini C 2015 Reducing ventilation requirements in semi-closed greenhouses increases water use efficiency *Agric. Water Manag.* **156** 90–9
[5] Shen Y, Wei R and Xu L 2018 Energy consumption prediction of a greenhouse and optimization of daily average temperature *Energies* **11**
[6] Patil S L, Tantau H J and Salokhe V M 2008 Modelling of tropical greenhouse temperature by auto regressive and neural network models *Biosyst. Eng.* **99** 423–31
[7] Mahmood F, Ghiat I, Govindan R and Al-Ansari T 2020 Reduced-order Modelling (ROM) Approach for Optimal Microclimate Control in Agricultural Greenhouses *Comput. Aided Chem. Eng.* **48** 1879–84
[8] Mahmood F, Govindan R and Al-Ansari T 2021 Predicting Microclimate of a Closed Greenhouse Using Support Vector Machine Regression *Comput. aided Chem. Eng.* 1229–34
[9] Ghiat I, Govindan R, Bermak A, Yang Y and Al-Ansari T 2023 Hyperspectral-physiological based predictive model for transpiration in greenhouses under CO2 enrichment *Comput. Electron. Agric.* **213** 108255
[10] Jung D-H, Kim H S, Jhin C, Kim H-J and Park S H 2020 Time-serial analysis of deep neural network models for prediction of climatic conditions inside a greenhouse *Comput. Electron. Agric.* **173** 105402
[11] Mahmood F, Govindan R, Bermak A, Yang D, Khadra C and Al-Ansari T 2021 Energy

utilization assessment of a semi-closed greenhouse using data-driven model predictive control *J. Clean. Prod.* **324** 129172

[12] Chen T and Guestrin C 2016 XGBoost: A Scalable Tree Boosting System *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* pp 785–94

Acknowledgement

The authors would like to acknowledge the support from Qatar Fertiliser Company (QAFCO), Hassad Foods Company, and Hamad bin Khalifa University, member of Qatar Foundation, Education City, Doha, Qatar. This research was supported by QNRF-MME Joint Funding Food Security Call grant MME01-0922-190049 from the Qatar National Research Fund (a member of Qatar Foundation).

Reproduced with permission of copyright owner. Further reproduction prohibited without permission.