Greenhouse management using artificial intelligence

Mustafa Sultan
Department of Artificial Intelligence
Nile University
El Sheihk Zayed City, Egypt

Saif Eldin Mohamed
Department of Artificial Intelligence
Nile University
El Sheihk Zayed City, Egypt

Omar Shaalan
Department of Artificial Intelligence
Nile University
El Sheihk Zayed City, Egypt

Ahmed Kamal
Department of Artificial Intelligence
Nile University
El Sheihk Zayed City, Egypt

In this study, we developed a comprehensive machine learning approach to predict five key environmental parameters in greenhouse agriculture. The project leverages multiple artificial intelligence techniques, each tailored to specific environmental aspects: Convolutional Neural Networks (CNN) for external temperature and wind speed prediction, Support Vector Regression (SVR) for internal temperature modeling, Random Forest for humidity prediction, and an ensemble learning approach combining linear regression, decision trees, and random forest for carbon dioxide levels. Data preprocessing included feature selection through correlation analysis, outlier removal, and normalization. The CNN models processed monthly data transformed into image-based subplots, while other models utilized direct sensor measurements. The Random Forest humidity model, utilizing 10 trees with a maximum depth of 3, demonstrated effective performance while avoiding overfitting. The ensemble learning approach for CO2 prediction improved upon individual model performance, showing better consistency and accuracy compared to single-algorithm implementations. Notable findings include the superiority of integrated modeling approaches over single-algorithm solutions, particularly in handling the complex interactions between greenhouse environmental parameters. This work provides a foundation for advanced greenhouse monitoring and control systems.

I. Introduction

This project aims to improve greenhouse agriculture by utilizing artificial intelligence techniques. Some of the important aspects of the greenhouse microclimate like Temperature, Humidity, and Carbon dioxide emissions are analyzed to create an accommodating machine-learning model for each aspect. This project mainly focuses on 5 aspects: Outer temperature of the greenhouse, inner temperature of the greenhouse, wind speed, Humidity in the greenhouse and the amount of carbon dioxide present in the air inside the greenhouse. Each aspect is predicted with a different module. The outer temperature and wind speed were predicted using a CNN model, the inner temperature was predicted using an SVR model, the humidity was predicted using a random forest model, and the carbon dioxide model was built using an ensemble learning model utilizing different machine learning algorithms. The dataset that was used to build these models contained various contains most of the important metrics of a greenhouse collected by various sensors and computers.

II. Litreture review

Recent advancements in greenhouse technology have increasingly focused on integrating predictive modeling and machine learning techniques to optimize environmental conditions and enhance crop cultivation. Kumar et al. (2024) demonstrated the effectiveness of predictive modeling in managing greenhouse climates by analyzing key environmental parameters such as temperature, humidity, and CO_2 levels. Their study highlighted how data-driven models can significantly improve plant growth and yield, particularly in controlled environments where small fluctuations in climate can greatly impact productivity. Furthermore, Escamilla-García et al. (2020) underscored the critical role of artificial neural networks (ANNs) in smart agriculture, particularly for their ability to process complex and nonlinear data. ANNs have been successfully applied to predict and regulate crucial climate factors, providing real-time insights and ensuring optimal growing conditions for plants. Their research also highlighted the broader potential of smart agriculture technologies in fostering sustainable agricultural practices.

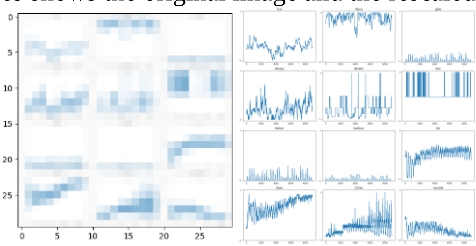
In addition to these efforts, Rambier (2023) explored the application of machine learning frameworks for forecasting greenhouse climates, using advanced competition datasets to refine predictive accuracy. This study emphasized the importance of leveraging historical and real-time climate data to anticipate changes and implement timely corrective actions, thus reducing resource wastage and improving energy efficiency. Collectively, these studies provide a comprehensive overview of the transformative potential of predictive modeling and artificial intelligence in greenhouse environments, offering innovative solutions to challenges such as resource management, energy conservation, and crop quality enhancement. As the demand for sustainable agricultural practices continues to grow, these advancements pave the way for smarter, more efficient greenhouse systems capable of supporting global food security.

III. Methodology

A. Data Preprocessing

1) CNN models: For the models that utilized CNN, the data was first turned into an images that contains the subplots of the data with each picture containing the readings of a month. Before being turned into subplots, the data is first normalized into the value of a range from –1 to 1, the images are generated in (1600x1600) dimensions, and before fitting to the model they were rescaled to (30x30).

Figues shows the original image and the rescaled one



2) General Preprocessing: Before creating the model, a feature selection is performed in order to determine which feature will be used to help predict the targeted feature. This is done by seeing the relationship of the features using correlation coefficient and then put into a confusion matrix. The data is then removed from nulls and outliers in order to ensure that there are no rows that will cause trouble to the model. The features are then normalised in order to ensure an unbiased prediction.

B. Model Architecture

1) CNN models: The CNN model handles the preprocessod data and use it to predict the estimated value for a week. Various CNN blocks were experimented with but only 3 scored good results for both of the models.

2) SVR model: Firstly, the uneeded features are seperated from the important features. After that the features are handed to an SVR model with a split ratio of 8:2 training to testing data. The reason for picking an SVR model for temperature is its robust mathematical foundation and flexibility.

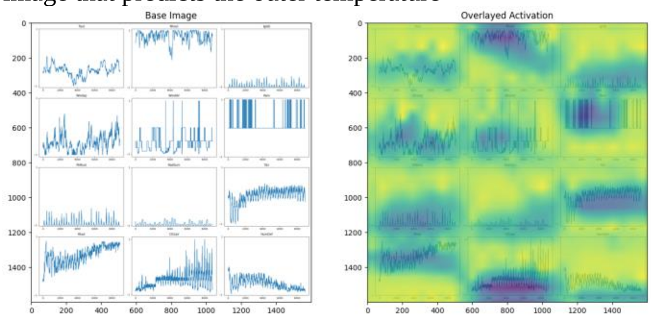
3) Random Forest Model: Random Forest regression model was implemented to predict relative humidity levels The model utlised 10 trees with a maximum of 3 levels for it’s depth in its architecture as adding anymore trees resulted in an overfitting, We focused on three core environmental variables - internal temperature (Tair), humidity deficit (HumDef), and ventilation (t_ventlee_vip) - as these have direct physical relationships with relative humidity. Randomforest was used due to it’s robustness against noisy data while preventing overfitting.

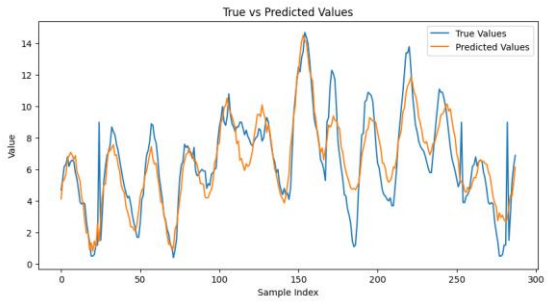
4) Ensemble Learning: This model utilised 3 algorithms in its making, those algorithms linear regression, decision trees and random forest. There was experimentation with using MLP regressor, svr and kernal ridge, but each had its own problem. The MLP regressor scored a low accuracy of an r square of 0.15 and proved to be hard to be improved, the svr scored a very low score of 0.02 so it was not appealing to optimize, and the kernal ridge needed high computational power which at the time was not avaible. The reason for using ensemble learing is that each of the 3 models on their own scored an unsatisfactory r square score, but utilizing the 3 models at once resulted in better average of results and better consistency.

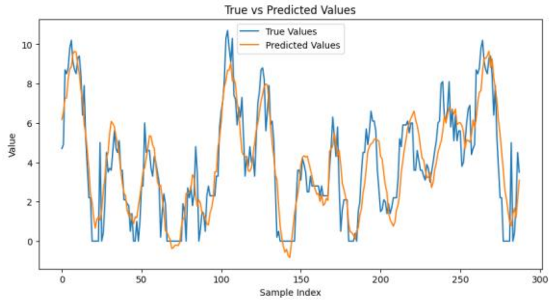
IV. Results

1) *CNN models:* The results for the outside temperature had a r square score of 0.89 and while the score for the wind speed is 0.87

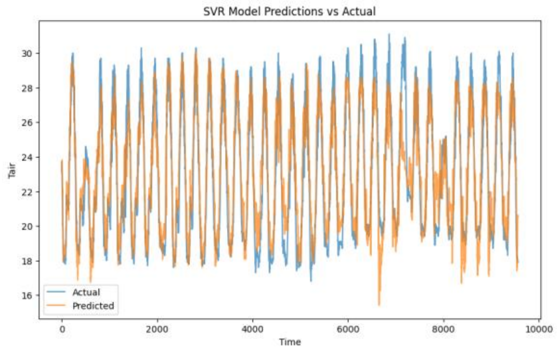
   Figures show the understanding of the model to the image that predicts the outer temperature



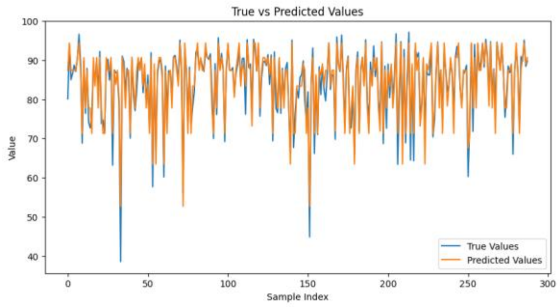




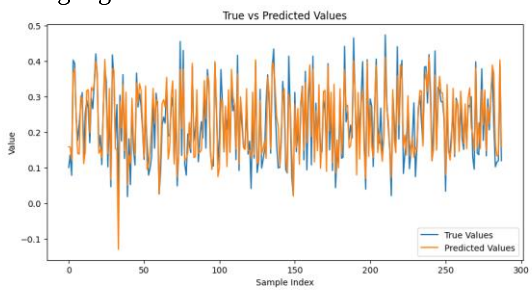
2) *SVR Model:* The SVR gotten an r square score of 0.84



3) *Random Forest Model:* The random forest model scored 0.91



4) *Ensemble Learning:* For the ensemble model, the linear regression model scored an r score of 0.75, the decision tree model scored 0.8 and the random random forest model scored 0.91. In the end, the voting regressor scored an r score of 0.87



V. Findings

The findings of this study highlight the potential of machine learning techniques in enhancing greenhouse climate control by accurately predicting key environmental parameters. The results indicate that each algorithm employed served a specific purpose effectively, contributing to a comprehensive and integrated system for greenhouse management. The CNN model demonstrated high accuracy in predicting external temperature and wind speed, showcasing the utility of deep learning in handling image-based data transformations. This is particularly useful in scenarios where time-series data can be visualized as trends over months, allowing for robust pattern recognition. The high R^2 scores (0.89 and 0.87 for external temperature and wind speed, respectively) underscore the efficacy of CNNs in modeling complex, non-linear relationships within climate data.

The SVR model, with an R^2 score of 0.84, proved its suitability for internal temperature prediction. This aligns with its established reputation for handling small to medium-sized datasets and capturing intricate patterns within continuous variables. The choice of SVR for this parameter highlights its robustness and adaptability in greenhouse applications, particularly where precise internal climate regulation is critical for plant growth and development.

The Random Forest model outperformed other models in predicting humidity levels, achieving an R^2 score of 0.91. Its strength lies in its ability to handle noisy data and prevent overfitting, which is crucial when dealing with greenhouse environments where humidity fluctuations can be influenced by multiple interacting variables, such as temperature, ventilation, and water vapor saturation. By utilizing a limited number of trees with shallow depth, the model maintained a balance between complexity and performance, further validating the importance of tuning hyperparameters for optimal results.

The ensemble learning approach for CO_2 level prediction demonstrated the advantages of combining multiple algorithms to enhance prediction accuracy and consistency. The individual models (linear regression, decision trees, and random forest) achieved varying degrees of accuracy, but the ensemble model's voting regressor outperformed them, with an R^2 score of 0.87. This approach leverages the strengths of each algorithm while mitigating their weaknesses, offering a more reliable prediction framework for CO_2 management. This is significant, given the critical role of CO_2 in photosynthesis and plant health.

VI. Discussion

Overall, the study confirms the superiority of integrated machine learning approaches over single-algorithm solutions in managing the complex interactions within greenhouse environments. The combination of diverse models tailored to specific environmental factors resulted in a robust and scalable system for real-time monitoring and control. These findings align with the literature, particularly the works of Kumar et al. (2024) and Escamilla-García et al. (2020), who emphasized the transformative potential of data-driven technologies in greenhouse agriculture. Moreover, the study's ensemble approach for CO_2 prediction supports Rambier's (2023) argument that leveraging historical and real-time data through machine learning frameworks can significantly improve prediction accuracy and resource efficiency.

VII. Limitations

Despite these promising results, some limitations warrant further investigation. The study relied on sensor data, which may be subject to calibration errors and noise, potentially affecting model performance. Additionally, the computational requirements for CNN and ensemble models can pose challenges for real-time deployment in resource-constrained settings. Future research could explore the integration of advanced hardware solutions, such as edge computing, to address these limitations and enhance real-time processing capabilities. Furthermore, expanding the dataset to include additional environmental factors, such as light intensity and soil moisture, could improve the comprehensiveness of the predictive models.

VIII. CONCLUSION

in conclusion, this study provides a foundation for developing advanced greenhouse monitoring and control systems. By employing a diverse range of machine learning techniques, it offers a scalable and efficient solution to optimize greenhouse climates, ultimately contributing to more sustainable and productive agricultural practices.

REFERENCES

[1] Predictive Modeling for Enhanced Plant Cultivation in Greenhouse Environment Y. Jeevan Nagendra Kuma, Ragi Chandan, Sri Harsh Somanini, Suresh Vadtya, Y. Ram Lohit Pranay, Kahtan A. Mohammed, Rakesh Chandrashekar, Lavish Kansal, Ravi Kalr E3S Web of Conf. 507 01066 (2024) DOI: 10.1051/e3sconf/202450701066

[2] Escamilla-García, A., Soto-Zarazúa, G. M., Toledano-Ayala, M., Rivas-Araiza, E., & Gastélum-Barrios, A. (2020). Applications of Artificial Neural Networks in Greenhouse Technology and Overview for Smart Agriculture Development. Applied Sciences, 10(11), 3835. https://doi.org/10.3390/app10113835.

[3] Estelle Rambier. [Private] Forecast climate in greenhouse. https://kaggle.com/competitions/private-forecast-climate-in-greenhouse, 2023. Kaggle.