International Journal of Electrical and Computer Engineering (IJECE)
Vol. 15, No. 2, April 2025, pp. 2392~2403
ISSN: 2088-8708, DOI: 10.11591/ijece.v15i2.pp2392-2403

Enhancing accuracy in greenhouse microclimate forecasting through a hybrid long short-term memory light gradient boosting machine ensemble approach

Mokeddem Kamal Abdelmadjid, Seddiki Noureddine, Bourouis Amina, Benahmed Khelifa
Information Processing and Telecommunications Laboratory (L.T.I.T), Department of Math and Computer Sciences, Faculty of Exact Sciences, Tahri Mohamed University, Bechar, Algeria

Article Info

Article history:

Received Jun 10, 2024
Revised Oct 15, 2024
Accepted Dec 14, 2024

Keywords:

Deep learning
Greenhouse
Machine learning
Microclimate
Time series

ABSTRACT

Greenhouse cultivation is one of the main methods for improving agricultural yield and quality. With the world needing more and more production, improving greenhouses using innovative technology becomes a must. These high-tech, aka, smart greenhouses depend much on the accuracy and availability of sensor data to perform at their best. In challenging situations such as sensor malfunctions or data gaps, utilizing historical data to predict microclimate parameters within the greenhouse is essential for maintaining optimal growing conditions and effective sustainable resource management control. In this work, and by employing a synthesis technique across various time series models, we forecast internal temperature and humidity, the two main parameters for a greenhouse, by incorporating diverse characteristics as input into a customized forecasting model. The selected architecture integrates deep learning and nonlinear learning models, specifically long short-term memory (LSTM) and light gradient boosting machine (LightGBM) as an ensemble approach, providing a comprehensive framework for time-series prediction, evaluated through mean absolute error (MAE), root mean square error (RMSE), and coefficient of determination (R^2) metrics. With a focus on improving accuracy in anticipating environmental changes, we have achieved high precision in predicting temperature (98.45%) and humidity (99.61%).

Corresponding Author:

Mokeddem Kamal Abdelmadjid
Department of Math and Computer Sciences, Faculty of Exact Sciences, Tahri Mohamed University
49 Taleb Abdelallah Road, Bechar 08000, Algeria
Email: mokeddem.kamal@univ-bechar.dz

1. INTRODUCTION

Greenhouses are vital for sustainable agriculture and food security, providing a controlled environment for year-round crop production, optimizing resource use, and increasing agricultural output. As climate change and rising food demand create new challenges, adopting smart greenhouses becomes crucial for ensuring a reliable food supply. An in-depth study of greenhouse dynamics shows that the delicate balance of microclimate factors significantly influences plant growth and resource utilization [1]. Consequently, enhancing the accuracy in predicting internal climate conditions, particularly temperature and humidity, as they are pivotal factors [2], is essential due to their complex and nonlinear fluctuations that can be impacted by external environmental factors [3]. To address these issues, especially when failures in request or sensor malfunctions may compromise data input, implementing sophisticated smart systems within greenhouses is necessary to analyze and adjust these parameters [4] dynamically.

Diverse methodologies, including mechanistic models, time series analysis, and machine learning (ML) techniques, have been employed to improve efficient and accurate smart agricultural systems [2]. Among these, artificial neural network models have gained prominence. Specifically, multi-layer perceptrons (MLPs) in [5]–[9] display accuracies of 96%, 97.7%, 99.9%, 99.9%, and a root mean square error (RMSE) of 3.7 °C, respectively. Additionally, the LightGBMRegressor in [10] and [11] emerged as a standout model, achieving a remarkable precision of mean absolute error (MAE) by 1.485 °C, and an accuracy of 98% in temperature predictions, respectively.

The recent trend toward employing deep learning (DL) techniques for time series predictions [7], [12] has demonstrated significant potential. Models like gated recurrent units (GRU) in [3] outperformed traditional ML methods by achieving an average accuracy of 91% in predicting temperature, even with limited meteorological data. Furthermore, the study [13] evaluates various DL models in forecasting temperature where the long short-term memory (LSTM) model yielded precision rates between 95% and 99%. This work underscores the importance of model selection and time intervals for accurate climate predictions. Tawfeek *et al.* [14] proposes an adaptive one-dimensional convolutional neural network (CNN) with an exactness of 97.56%.

Models predictive has been successfully implemented to emulate greenhouse functionality with methods like the Mamdani fuzzy inference system in [15], enhancing the tracking of internal parameters. Additionally, the constrained model predictive strategy discrete or combined in [16] and [17] has effectively managed indoor temperatures through simulations. While hybrid approaches combining fuzzy logic and GRU in [18] have achieved an RMSE of 0.292.

Recent studies invest in heterogeneous ensemble learning methods that can significantly improve time-series data predictions while minimizing noise-related errors. The study [19] utilizing ensemble techniques among seven regression models achieved, where the best-performing models are selected and combined, an impressive accuracy of 0.96 in predicting indoor humidity. The findings in [20] indicated that all hybrid resampling models outperformed the linear regression (LR) model. Additionally, the paper [21] focuses on using advanced supervised learning techniques, specifically extreme gradient boosting (XGBoost) and recurrent neural networks (RNN) combined with LSTM. The LSTM-RNN model yielded the best results during the summer season, achieving a precision of 99.94%. El Alaoui *et al.* [22] conducted a comparison of four MLs including bagging and boosting trees methods against computational fluid dynamics (CFD) models for predicting indoor temperature and humidity where the ML models excel well with 98% accuracy.

This work addresses the critical need for accurate forecasting of influential microclimate parameters, which are vital for informed decision-making and proactive measures against future scenarios. Our proposed methodology employs the ensemble technique that integrates the strengths of models within a carefully structured framework to boost predictive accuracy. By thoroughly examining the interrelationships of microclimate parameters and incorporating diverse time-series features, we enhance the effectiveness of our predictive models. The remainder of this article is structured as follows: Section 2 details our methodology, including data preprocessing, model development, and evaluation metrics. Section 3 presents the results, analysis, and discussion. Finally, section 4 summarizes our findings and suggests potential applications and future research directions.

2. METHOD

Introducing artificial intelligence (AI) as a solution to analyze meteorological features offers a promising approach to advancing agricultural practices. Through AI integration, we aim to develop a dynamic control system that utilizes real-time data from various sensors within the greenhouse. As illustrated in Figure 1, our main strategy involves incorporating AI models to refine learning methodologies. The approach integrates ensemble models that combine the strengths of both bagging and stacking techniques. In this framework, the base models, which form the bagging ensemble, are trained on various subsets of the data. This strategy helps to mitigate overfitting and enhances the model's ability to generalize to new data. The predictions resulting from these base models are then used to train a second model, known as the meta-model. The meta-model learns to synthesize the predictions from the base models to produce a final and more accurate prediction [19], [22].

Applying this ensemble approach can greatly enhance the precision of greenhouse microclimate predictions. To achieve this aim, we will test a range of ML and DL models, to analyze and predict the complex data present in greenhouse environments [7], [12]. A thorough comparative study of these models will be essential to identify the most effective option for microclimate forecasting. These models exhibit diverse strengths and capabilities in handling such data, ranging from traditional ML techniques like support vector machines (SVM) to more advanced DL architectures such as CNNs and LSTMs. Furthermore, we emphasize the importance of time series prediction, which enables us to incorporate temporal dynamics into our analysis. This means that the mode considers not only current environmental conditions but also how those conditions change over time.

By incorporating bagging and stacking into the learning architecture for greenhouse microclimate predictions, we can significantly enhance the robustness and accuracy of the forecasting model by leveraging the diversity of multiple models trained on different subsets of the data. The initial data structuring involves a lag approach to facilitate transformation. Subsequently, rolling mean and standard deviation calculations are applied with diverse window sizes, a common practice in time-series data analysis [11], [23]. Both measures offer valuable insights into the distribution and variability of the data. The next step includes partitioning the dataset into training and testing sets, dedicating 80% of the available data for model training, and reserving the remaining 20% for testing model performance.

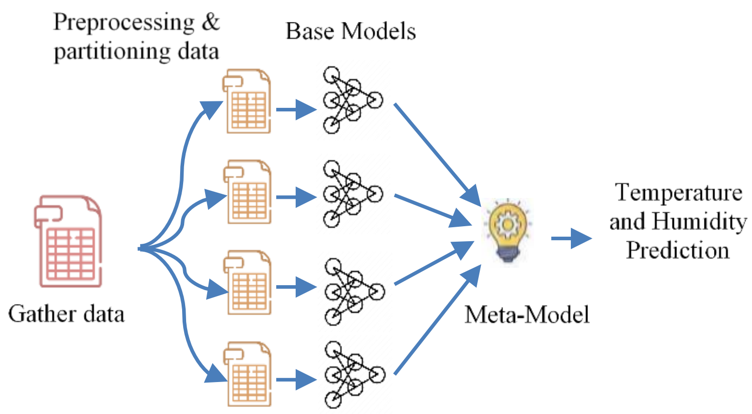


Figure 1. The proposed architecture

2.1. Data collection

To achieve our goal of enhancing accuracy by building the greatest predictive model, we must first deal with the data aspect. We initiate this process by gathering a dataset containing measurements of various microclimate-related characteristics obtained from Kaggle. This dataset includes measurements at 10-minute intervals spanning from March 3rd, 2021, to July 3rd, 2021. Table 1 provides an enumeration and description of the columns comprising this dataset.

The dataset contains 17,542 records, primarily consisting of float64 data types, except for the timestamp field “created”, which is in *DateTime* format. It encompasses eight variables related to meteorological measurements both inside and outside the greenhouse, with varying degrees of missing data. Detailed statistical information about the dataset, excluding the timestamp, is provided in Table 2.

Table 1. Dataset columns enumeration and description

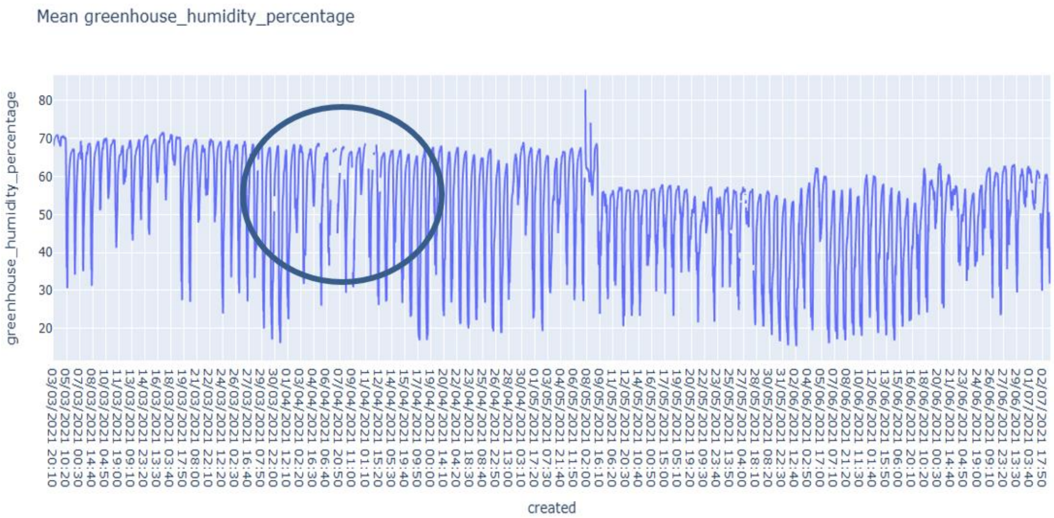
Number	Column	Description	Unit
1	*greenhous_temperature_celsius*	The inside temperature	Celsius
2	*greenhouse_humidity_percentage*	The inside humidity	%
3	*greenhouse_illuminance_lux*	The inside illuminance	lux
4	*online_temperature_celsius*	The outside online temperature	Celsius
5	*online_humidity_percentage*	The outside online humidity	%
6	*greenhouse_total_volatile_organic_compounds_ppb*	The inside total volatile organic compounds	ppb
7	*greenhouse_equivalent_co2_ppm*	The inside equivalent CO_2	ppm
8	Created	The timestamp of when the record was created	date/time

Table 2. Statistical analysis of the dataset

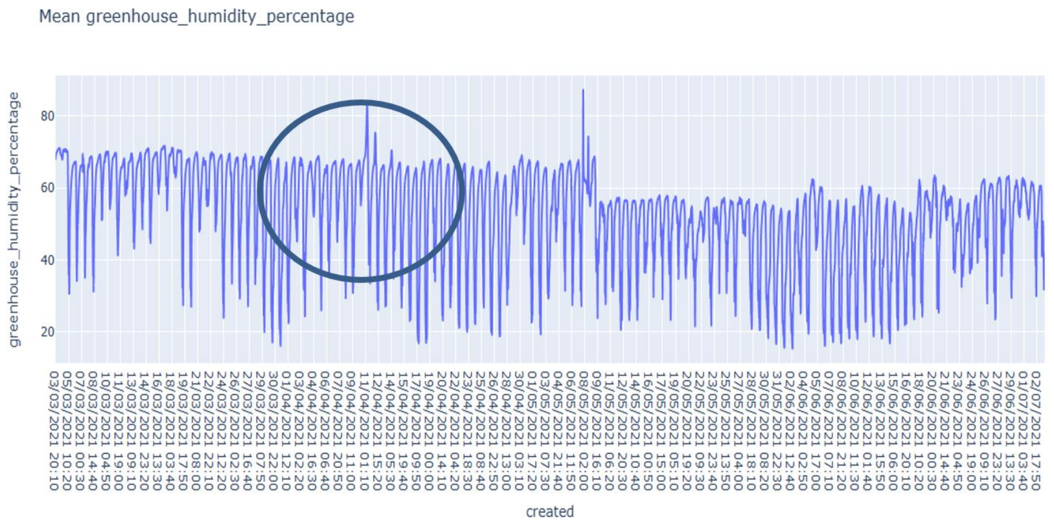
Columns	Mean	Std	min	max
Humidity	53.126	13.125	15.314	8.720e+01
Temperature	25.619	13.082	-4.120	7.757e+01
illuminance_lux	3654.316	5833.280	0.000	3.649e+04
online_temperature	11.039	6.702	-4.900	3.230e+01
online_humidity	73.818	17.852	18.000	1.000e+02
total_vol_org_comp	972.689	1052.417	0.000	1.311e+04
equivalent_co2_ppm	1214.514	1230.192	0.000	1.811e+04

2.2. Data preprocessing and engineering

A fundamental step in data analysis is understanding the integrity of data. To do this, we cleaned the dataset by handling missing values. Identifying these missing values is crucial as they can significantly impact our analyses and results. In our case for handling missing data in our air quality dataset, we have delved into "Quadratic interpolation". Quadratic interpolation is a technique that estimates missing values by fitting a quadratic function to the neighboring data points. This approach is valuable when the data exhibits nonlinear patterns. Figure 2 exemplifies the treatment of the humidity percentage data. Figure 2(a) shows the humidity percentage before any process and Figure 2(b) illustrates the handling of missing values after. The time series within our dataset is characterized by a frequency of 1.0, a key attribute that offers vital insights into the temporal intervals between each recorded data point.



(a)

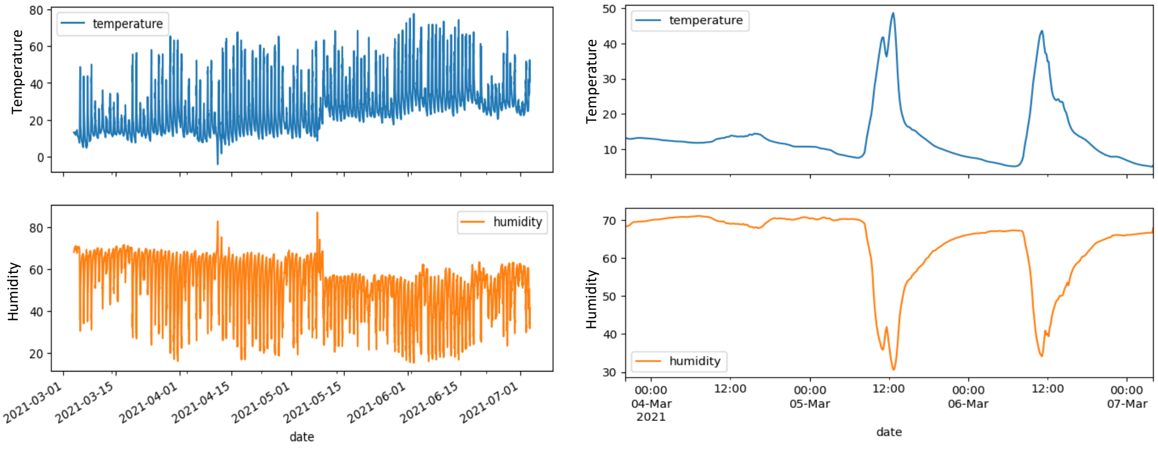


(b)

Figure 2. Evolution of humidity percentage over time before and after correction (a) original humidity percentage representation and (b) projected data after quadratic interpolation application.

Figure 3 depicts the evolution over time of temperature and humidity. The temporal progression is showcased in Figure 3(a), while Figure 3(b) highlights a segment of three days. To untangle the complex interactions among various parameters and identify the key input combinations that significantly influence modeling results, rigorous importance on correlation analysis is essential [2], [20]. We have chosen to utilize the simple yet powerful Pearson's correlation coefficient r, defined as (1), between x and y [19]:

$$r = \Sigma_i((x_i - mean(x))(y_i - mean(y)))\ (\sqrt{\Sigma_i(x_i - mean(x))^2}\ \sqrt{\Sigma_i(y_i - mean(y))^2})^{-1} \quad (1)$$



(a) (b)

Figure 3. Example of data showing the evolution of temperature and humidity over time (a) temporal progression of temperature and humidity and (b) three-day segment high-lighting temperature and humidity

This strategic approach ensures that the model is built on a foundation of well-understood relationships between variables. In our modeling process, we adopted a selection criterion centered around negative correlation coefficients upper than 0.50, as shown in Figure 4. The method involved singling out variables that displayed the most pronounced negative correlation coefficients among all input variables. This led to the identification of four key input combinations resumed in Table 3.

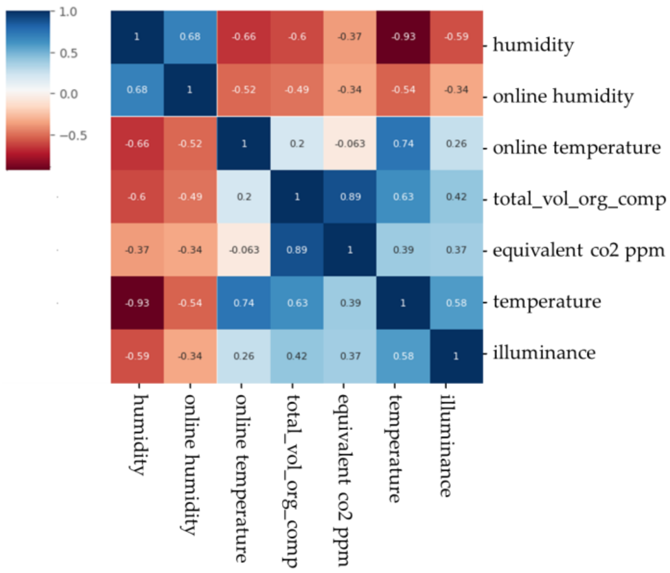


Figure 4. Correlation analysis

Table 3. The four combinations of inputs were selected

N°	Combination	Correlation	To predict
1	Temperature, Humidity	-0.93	Temperature
2	Online temperature, Humidity	-0.66	Humidity
3	Online humidity, Temperature	-0.54	Temperature
4	Illuminance, Humidity	-0.59	Humidity

After identifying all combinations, the next step is reshaping these time series datasets into a supervised learning problem that relies on employing a sliding window methodology [24], [25]. In the chosen transformation, we utilize the value from the preceding time step ($t-1$) and the current time value (t) to predict the value at the time step ($t+1$), with a window width (also known as lag or sliding method) set to one. The four input combinations involve information from both previous times (*i.e.*, data at $t-1$) and current times (*i.e.*, data at t) to forecast the greenhouse temperature or humidity at time $t+1$. The primary goal is to predict temperature or humidity, resulting in three input features and one output value for prediction in each training pattern. In Table 4, an example of the pair tuple (temperature and humidity) likely demonstrates how are utilized together to forecast temperature values.

The proposal advocates for a comprehensive approach to time series tasks, advice on the integration of external attributes, and the application of internal feature engineering methodologies, underscoring their inherent benefits. This strategic framework involves the incorporation of key features, notably lag values of numerical attributes [23]. These statistical attributes, encompassing metrics such as mean, rolling, standard deviation, month, week, day, hour, minute, and day of the week, offer valuable insights into the broader trends of environmental factors across the temporal spectrum. Moreover, they play a pivotal role in counteracting the impact of outliers during the model training phase [11]. We utilize three sets of lagged values, representing data from the previous day, a 7-day retrospective, and a 30-day retrospective, to serve as proxies for capturing metrics from the last week and last month. This incorporation of lagged values as statistical features aims to capture temporal dependencies and historical patterns, empowering the model with the ability to leverage past information for improved accuracy in time series analysis. As a result, we have, for each combination, a total of twenty-eight inputs for each pair tuple.

Table 4. Data lagged with a window width of one

Humidity (t-1)	Temperature (t-1)	Humidity (t)	Temperature (t+1)
68.055576	13.223257	68.271154	13.1496152
68.271154	13.149615	68.402229	13.0943393
68.402229	13.094339	68.483559	12.9893954
68.483559	12.989395	68.556920	12.9019275
68.556920	12.901927	68.746014	12.8791106

3. RESULTS AND DISCUSSION

3.1. Comparative analysis of forecasting models

To find the optimal approaches for temperature and humidity prediction, and during the selection of the performant model, we employed a diverse range of models such as CNN-LSTM, LSTM, SVM-RBF, Prophet, LightGBM, and XGBoost on training the dataset based on specific input combinations. Our primary objective is to evaluate each model's performance by a synthesis analysis. The evaluation metrics is crucial and is informed by the searcher in the literature. Given the regression nature of our problem, we have opted for mean absolute error (MAE), root mean squared error (RMSE), and R-squared (R^2) metrics. These metrics are commonly used in similar studies and provide a comprehensive assessment of model performance.

Table 5 provides a summary of the results obtained during the test phases for all models. The distinction between the LSTM and LSTM Series lies in the sequence of features taken as inputs (name). Figure 5 illustrates the performance of ML and DL models across various prediction combinations. Figure 5(a) illustrates ML models in temperature prediction, while Figure 5(b) shows DL models in humidity prediction for the first combination. For the second combination, Figure 5(c) displays the performance of ML models for humidity prediction, and Figure 5(d) depicts DL models for the same task. The third combination features Figures 5(e) and 5(f), which represent ML and DL models, respectively, for temperature prediction. Finally, Figure 5(g) illustrates the accuracy of ML models in humidity prediction using the fourth combination, whereas Figure 5(h) demonstrates the accuracy of DL models. These representations demonstrate and provide a comprehensive view of the strengths of both ML and DL approaches across different forecasting combinations.

The four ML models demonstrated exceptional proficiency in capturing intricate data relationships, yielding impressive results in prediction. Prophet, a time series forecasting model developed by Facebook, showcased consistent performance in all metrics. However, the results were subpar with the data combination 3, yielding an MAE of 6.860, an RMSE of 11.349, and an R^2 of 19.6%. This outcome negatively impacts our selection of this model compared to the earlier results from other models. In contrast, the comparison of LightGBM with XGBoost and SVM-RBF favors the LightGBM model due to its consistent reduction in MAE and RMSE values across temperature and humidity predictions. This consistent improvement, also in the work of [11] and [21], underscores LightGBM's reliability and effectiveness in capturing the underlying data patterns, making it the preferred choice for temperature and humidity forecasting tasks.

Table 5. Models' performance summary for temperature and humidity prediction

Indicators / Model	Combination 1 to predict temperature			Combination 2 to predict humidity			Combination 3 to predict temperature			Combination 4 to predict humidity		
	MAE	RMSE	R^2	MAE	RMSE	R^2	MAE	RMSE	R^2	MAE	RMSE	R^2
LightGBM	0.875	1.888	0.977	0.614	0.972	0.993	1.054	2.314	0.966	0.527	0.853	0.994
XGBoost	1.012	2.158	0.971	0.741	1.157	0.991	1.216	2.626	0.957	0.596	0.937	0.994
SVM-RBF	0.918	1.859	0.978	0.819	1.409	0.986	0.846	1.771	0.980	0.702	1.242	0.989
Prophet	0.479	0.871	0.995	0.469	0.752	0.996	6.860	11.349	0.196	0.463	0.760	0.996
LSTM	1.366	1.850	0.979	0.973	1.210	0.989	-	-	-	0.925	1.254	0.989
LSTM Series	0.021	0.032	0.966	0.036	0.049	0.917	0.023	0.029	0.972	0.020	0.027	0.974
CNN-LSTM	0.033	0.041	0.944	0.031	0.042	0.939	0.019	0.036	0.956	0.031	0.037	0.951

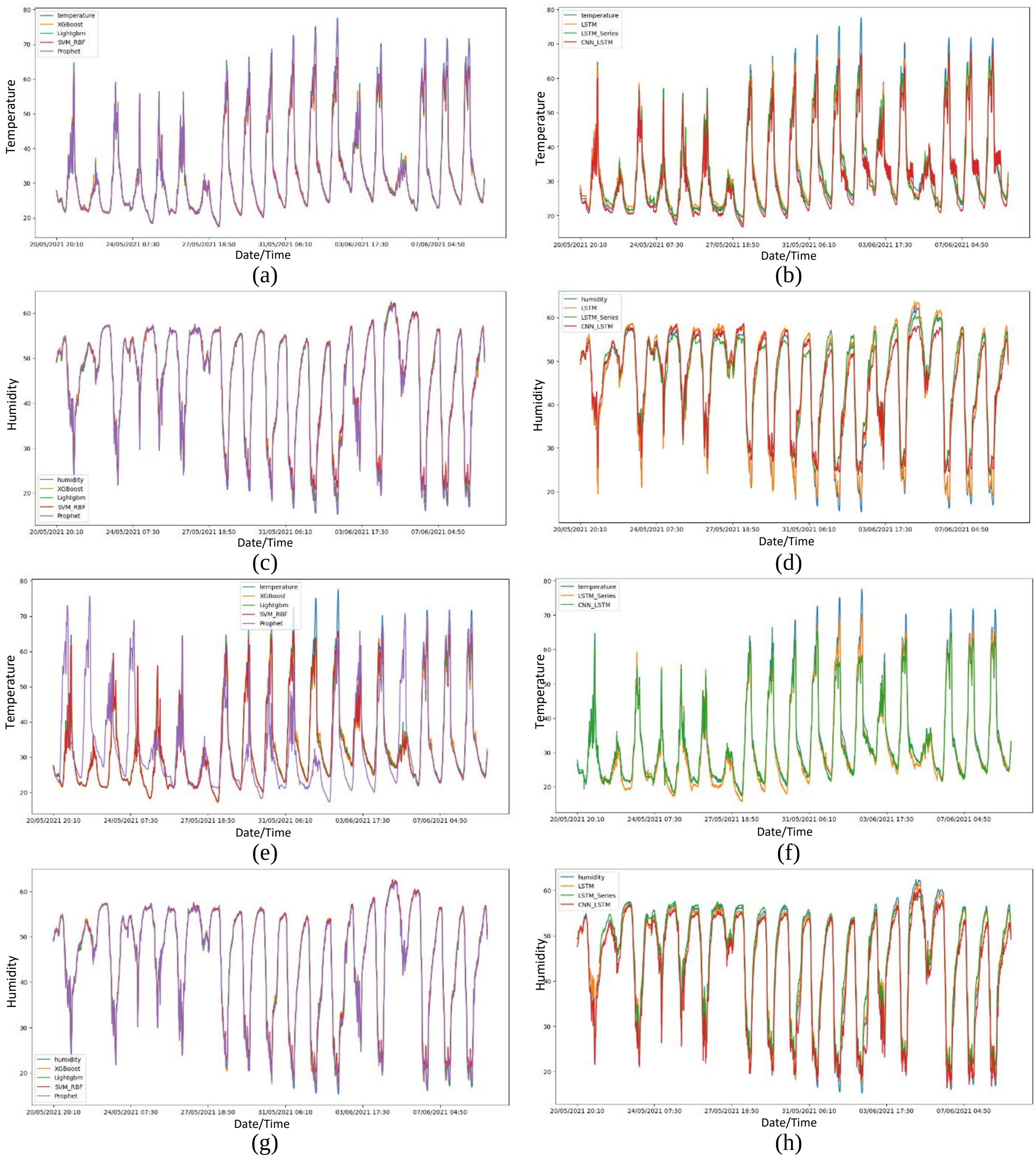


Figure 5. Visual representation of (a) models' performance in temperature, (b) models' performance in humidity prediction, (c) the performance of ML models for humidity prediction, (d) the performance of DL models for humidity prediction, (e) the performance of ML models for temperature prediction, (f) the performance of DL models for temperature prediction, (g) the accuracy of ML models in humidity prediction using the fourth combination, and (h) the accuracy of DL models

The LSTM model demonstrated weak predictive abilities, especially if we refer to the results of the previously mentioned models, as indicated by its MAE, RMSE, and R^2 scores across different combinations. Similar to the Prophet model, for the third combination, the LSTM also undershoots with negative values for R^2. To address this, we explored an alternative approach utilizing a specialized implementation of LSTM tailored for time series analysis, dubbed LSTM Series. We reorganized the datasets while preserving the sequential order of observations. Here, the data from the previous one hundred forty-four (144) records, representing one day, will be utilized to predict temperature or humidity readings for the following ten-minute interval [26]. The details of the LSTM Series model proposed are given in Table 6. The model yielded compelling results with notable reductions in MAE and RMSE if compared against the prior results of all models and the CNN-LSTM which falls short of being the optimal solution.

Table 6. The layers and parameters of the LSTM series

Layer number	Layer type	Output shape	Parameters
1	LSTM	(None, 144, 128)	80384
2	LeakyReLU	(None, 144, 128)	0
3	LSTM	(None, 144, 128)	131584
4	LeakyReLU	(None, 144, 128)	0
5	Dropout	(None, 144, 128)	0
6	LSTM	(None, 64)	49408
7	Dropout	(None, 64)	0
8	Dense	(None, 1)	65
Total parameters: 261,441			

The analysis underlines the diverse strengths of each model in forecasting temperature and humidity, highlighting the critical importance of selecting models that align with our ensemble models. A thorough evaluation by metrics, alongside a comparison with other published works presented in Table 7, demonstrates the effectiveness of our preliminary results. This stage of comparative analysis is essential for identifying the most suitable forecasting models based on their predictive performance.

Table 7. Comparison of model's performance: literature results vs. our findings

Model	Reference	Metric	Prediction	
			Temperature	Humidity
LightGBM	[10]	R^2	0.980	-
	[19]		-	0.962
	[11]	MAE	1.485	-
	Our	R^2/MAE	0.977 / 0.875	0.994
XGBoost	[11]	MAE	1.626	-
	Our		1.012	-
SVM	[10]	R^2	0.977	-
	[22]		0.984	0.984
	[3]		0.790	-
	Our		0.980	0.989
LSTM	[3]	R^2	0.810	-
	[13]		0.950	0.940
	Our		0.972	0.974

Notably, the revised LSTM model outperformed other models, closely aligning with the trend line of actual values, particularly evident in its superior MAE and RMSE metrics. Given the exemplary performance of this model, we aim to further enhance R^2 metrics. To achieve this, we propose leveraging our ensemble approach by integrating nonlinear learning and deep-learning methodologies, combining the strengths of LSTM and LightGBM. Their demonstrated accuracy positions them as prime candidates for enhancing predictive performance in time-series forecasting.

3.2. Implementing the proposed architectural design

The architecture designed for our study involves four LSTM models with distinct combinations from the dataset as inputs, while the predictions of each LSTM model are fed into LightGBM as inputs. We have opted to leverage the robust gradient-boosting framework of LightGBM with the parameters presented in Table 8. This choice is driven by LightGBM's effectiveness in regression tasks and time series forecasting, aligning well with the nature of the models we are employing. The adopted process of selecting and preparing inputs for the LightGBM algorithm can be summarized as follows:

Algorithm 1. The meta-model (LightGBM)
Inputs: Prediction values resulting from the base models.
Outputs: Predictions of temperature and humidity values.

1. Store the R^2 accuracy values of the four base models
2. Identify all base models with R^2 values greater than 0.9
3. Retrieve predictions from all selected base models
4. Reshape the predicted values from the time series into a supervised learning format
5. The reshaped predicted values and corresponding real values as inputs for the meta-model
6. Optimize the hyperparameters of the model using the best parameter setting
7. Train the meta-model on the combined dataset
8. Test the model on a separate test set to evaluate its performance
9. Evaluate the results, including metrics such as R^2, MAE, RMSE.

Table 8. Parameters used to train the LightGBM

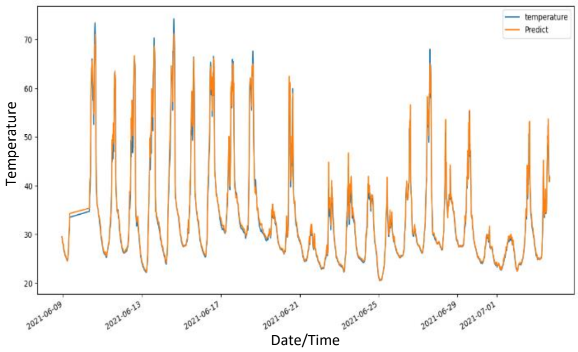
Parameters	Values
Number of boosting iterations to build	454
Learning rate	0.182
Maximum depth of each tree	4.570
Minimum sum of instance weight needed in a child node	6.975
Subsample	0.782
Columns sample by tree	0.995

The constructed architecture demonstrated significant improvements in regression tasks and time series forecasting, as indicated in Table 9. Notably, the model achieved an impressive accuracy of 98.45% in predicting temperature, with a minimal MAE of 0.7556 and RMSE of 1.2841. Similarly, for humidity prediction, the model achieved an accuracy of 99.61%, with a remarkable reduction in MAE and RMSE to 0.4383 and 0.6951, respectively.

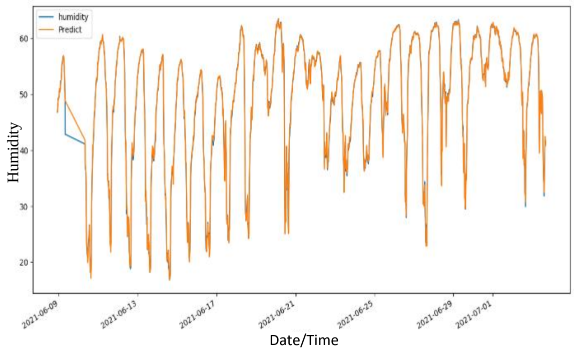
Table 9. Forecasting performance metrics of the LSTM-LightGBM

Indicators / Model	Predict temperature			Predict humidity		
	MAE	RMSE	R^2	MAE	RMSE	R^2
LSTM-LightGBM model	0.7556	1.2841	0.9845	0.4383	0.6951	0.9961

The visual representation in Figure 6 serves similarly as a compelling improvement of the effectiveness of integrating LSTM and LightGBM within the model architecture. Figure 6(a) showcases the accuracy of temperature prediction, whereas Figure 6(b) highlights the performance in humidity prediction. The graph vividly portrays the enhanced predictive capabilities and success of this approach, also the comparison with the previous in Table 10 and the precision R^2 visually depicted in Figure 7 reinforces the notion that the fusion of these algorithms leads to improved model efficiency and accuracy.



(a)



(b)

Figure 6. Performance visualization of LSTM and LightGBM implementation (a) temperature forecasting accuracy comparison and (b) humidity forecasting accuracy comparison

Table 10. Comparison of diverse ensemble models' performance with our LSTM-LightGBM using R^2 metric

Reference	Ensemble model	Prediction	
		Temperature	Humidity
[22]	Bagging trees	0.970	0.984
	Boosting trees	0.983	0.985
[19]	Stacking heterogeneous	-	0.9651
Our	LSTM-LightGBM	0.9845	0.9961
Our	LSTM	0.9720	0.9740
Our	LightGBM	0.9770	0.9940

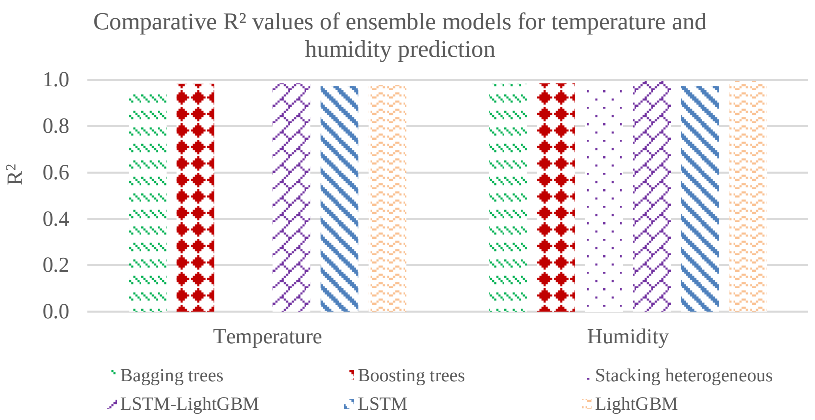


Figure 7. The values of R^2 in predicting temperature and humidity by different ensemble models

It is important to emphasize that comparing our results with previous studies is crucial for refining our strategies and outcomes. However, these studies utilize various datasets, which can lead to differing results if we apply our approach. This variability can present challenges in preprocessing and feature engineering when evaluating the performance of other predictive models against our proposed model. Moreover, the content of these datasets varies, as they categorize specific meteorological parameters relevant to their respective regions. Access to these datasets, whether controlled or open, can also pose difficulties in obtaining the best comparisons. For instance, open-access datasets often require formal requests, which can be time-consuming. Additionally, most of these datasets are not publicly available.

The approach aims to predict vital microclimate parameters, specifically temperature and humidity, using only two key parameters enhanced by additional features grouped in various combinations. In cases of sensor malfunctions or inaccurate actual temperature readings, for example, our model predicts this value based on humidity data-whether from the interior, exterior, or both. This flexibility allows for accurate temperature and humidity forecasting using different parameter combinations. The accuracy of findings derived from an ensemble approach highlights the effectiveness of AI in optimizing greenhouse microclimates to boost production efficiency. With the precision needed and the implementation of an AI-driven model that dynamically controls meteorological parameters, farmers can trust the model's recommendations to enhance yields while conserving energy, especially during adverse conditions that threaten crops.

4. CONCLUSION

Our findings in this work highlight the enhanced performance achieved through the proposed ensemble approach, which significantly boosts overall prediction accuracy, aligning with numerous studies documented in the literature. To further refine our model, we conducted experiments utilizing various algorithms and innovative strategies for data utilization. By experimenting with different combinations of newly introduced and effective variables, we aimed to augment the dataset and improve predictive accuracy. This collaborative strategy not only sets the stage for advanced predictive modeling techniques but also enhances decision-making processes, opening up possibilities for optimizing resource management, improving crop yield, and promoting sustainable practices in greenhouse farming. Furthermore, our future works will also focus on exploring models that can be edge-implemented, also known as TinyML or EdgeAI, to limit the need to centralize decision-making in smart greenhouses that are placed far from the servers. Another point is our hope to take our work from the experimentation phase to the real-life greenhouse implementation.

REFERENCES

[1] M. K. Abdelmadjid, S. Noureddine, B. Amina, B. Khelifa, B. Tariq, and L. Boubakeur, "IoT and WSNs technology for control in the greenhouse agriculture-review," in *2022 3rd International Conference on Embedded and Distributed Systems, EDiS 2022*, Nov. 2022, pp. 136–141. doi: 10.1109/EDiS57230.2022.9996500.

[2] Y. Guo, H. Zhao, S. Zhang, Y. Wang, and D. Chow, "Modeling and optimization of environment in agricultural greenhouses for improving cleaner and sustainable crop production," *Journal of Cleaner Production*, vol. 285, p. 124843, Feb. 2021, doi: 10.1016/j.jclepro.2020.124843.

[3] Z. He *et al.*, "Gated recurrent unit models outperform other machine learning models in prediction of minimum temperature in greenhouse Based on local weather data," *Computers and Electronics in Agriculture*, vol. 202, p. 107416, Nov. 2022, doi: 10.1016/j.compag.2022.107416.

[4] M. Morozova, "Methodology for controlling greenhouse microclimate parameters and yield forecast using neural network technologies," in *Studies in Systems, Decision and Control*, vol. 439, Springer Nature Switzerland, 2024, pp. 245–277, doi: 10.1007/978-3-031-40718-5_7.

[5] I. Ihoume, R. Tadili, N. Arbaoui, M. Benchrifa, A. Idrissi, and M. Daoudi, "Developing a multi-label tinyML machine learning model for an active and optimized greenhouse microclimate control from multivariate sensed data," *Artificial Intelligence in Agriculture*, vol. 6, pp. 129–137, 2022, doi: 10.1016/j.aiia.2022.08.003.

[6] S. Chakir, A. Bekraoui, E. M. Zemmouri, H. Majdoubi, and M. Mouqallid, "Prediction of olive cuttings greenhouse microclimate under Mediterranean climate using artificial neural networks," in *Lecture Notes in Networks and Systems*, vol. 455, Springer International Publishing, 2022, pp. 63–69, doi: 10.1007/978-3-031-02447-4_7.

[7] N. Wang and X. Zhao, "Time-series forecasting based on multi-layer attention architecture," *KSII Transactions on Internet and Information Systems*, vol. 18, no. 1, pp. 1–14, Jan. 2024, doi: 10.3837/tiis.2024.01.001.

[8] T. Petrakis, A. Kavga, V. Thomopoulos, and A. A. Argiriou, "Neural network model for greenhouse microclimate predictions," *Agriculture (Switzerland)*, vol. 12, no. 6, p. 780, May 2022, doi: 10.3390/agriculture12060780.

[9] S. Francik and S. Kurpaska, "The use of artificial neural networks for forecasting of air temperature inside a heated foil tunnel," *Sensors (Switzerland)*, vol. 20, no. 3, p. 652, Jan. 2020, doi: 10.3390/s20030652.

[10] B. Vega-Márquez, J. Pardo-Martínez, M. del M. Villegas-Oliva, and J. C. Riquelme, "Forecasting greenhouse temperature using machine learning models: optimizing crop production in Andalucia," in *Lecture Notes in Networks and Systems*, vol. 749 LNNS, Springer Nature Switzerland, 2023, pp. 239–248, doi: 10.1007/978-3-031-42529-5_23.

[11] Q. Cao, Y. Wu, J. Yang, and J. Yin, "Greenhouse temperature prediction based on time-series features and LightGBM," *Applied Sciences (Switzerland)*, vol. 13, no. 3, p. 1610, Jan. 2023, doi: 10.3390/app13031610.

[12] A. Mahmoud and A. Mohammed, "A survey on deep learning for time-series forecasting," in *Studies in Big Data*, vol. 77, Springer International Publishing, 2021, pp. 365–392, doi: 10.1007/978-3-030-59338-4_19.

[13] O. Eraliev and C. H. Lee, "Performance analysis of time series deep learning models for climate prediction in indoor hydroponic greenhouses at different time intervals," *Plants*, vol. 12, no. 12, p. 2316, Jun. 2023, doi: 10.3390/plants12122316.

[14] M. A. Tawfeek, N. Yanes, L. Jamel, G. Aldehim, and M. A. Mahmood, "Adaptive deep learning model to enhance smart greenhouse agriculture," *Computers, Materials and Continua*, vol. 77, no. 2, pp. 2545–2564, 2023, doi: 10.32604/cmc.2023.042179.

[15] J. B. Velandia, J. S. C. Quintana, and S. C. V. Ayala, "Environment humidity and temperature prediction in agriculture using Mamdani inference systems," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 11, no. 4, pp. 3502–3509, Aug. 2021, doi: 10.11591/ijece.v11i4.pp3502-3509.

[16] H. Hamidane, S. El Faiz, M. Guerbaoui, A. Ed-Dahhak, A. Lachhab, and B. Bouchikhi, "Constrained discrete model predictive control of a greenhouse system temperature," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 11, no. 2, pp. 1223–1234, Apr. 2021, doi: 10.11591/ijece.v11i2.pp1223-1234.

[17] A. Moufid, N. Boutchich, and N. Bennis, "Linear matrix inequalities tool to design predictive model control for greenhouse climate," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 13, no. 1, pp. 258–269, Feb. 2023, doi: 10.11591/ijece.v13i1.pp258-269.

[18] J. Zhen, R. Xu, J. Li, S. Shen, and J. Wen, "Study on the intelligent control model of a greenhouse flower growing environment," in *Communications in Computer and Information Science*, vol. 1629 CCIS, Springer Nature Singapore, 2022, pp. 127–145, doi: 10.1007/978-981-19-5209-8_9.

[19] S. Rezaei Melal, M. Aminian, and S. M. Shekarian, "A machine learning method based on stacking heterogeneous ensemble learning for prediction of indoor humidity of greenhouse," *Journal of Agriculture and Food Research*, vol. 16, p. 101107, Jun. 2024, doi: 10.1016/j.jafr.2024.101107.

[20] K. Khosravi *et al.*, "Multi-step ahead soil temperature forecasting at different depths based on meteorological data: integrating resampling algorithms and machine learning models," *Pedosphere*, vol. 33, no. 3, pp. 479–495, Jun. 2023, doi: 10.1016/j.pedsph.2022.06.056.

[21] J. M. Esparza-Gómez *et al.*, "Long short-term memory recurrent neural network and extreme gradient boosting algorithms applied in a greenhouse's internal temperature prediction," *Applied Sciences (Switzerland)*, vol. 13, no. 22, p. 12341, Nov. 2023, doi: 10.3390/app132212341.

[22] M. El Alaoui, L. O. Chahidi, M. Rougui, A. Mechaqrane, and S. Allal, "Evaluation of CFD and machine learning methods on predicting greenhouse microclimate parameters with the assessment of seasonality impact on machine learning performance," *Scientific African*, vol. 19, p. e01578, Mar. 2023, doi: 10.1016/j.sciaf.2023.e01578.

[23] M. Kour, "Modelling and forecasting of carbon-dioxide emissions in South Africa by using ARIMA model," *International Journal of Environmental Science and Technology*, vol. 20, no. 10, pp. 11267–11274, Nov. 2023, doi: 10.1007/s13762-022-04609-7.

[24] Y. Liu *et al.*, "A long short-term memory-based model for greenhouse climate prediction," *International Journal of Intelligent Systems*, vol. 37, no. 1, pp. 135–151, Aug. 2022, doi: 10.1002/int.22620.

[25] J. Morales-García, A. Bueno-Crespo, R. Martínez-España, J. L. Posadas, P. Manzoni, and J. M. Cecilia, "Evaluation of low-power devices for smart greenhouse development," *Journal of Supercomputing*, vol. 79, no. 9, pp. 10277–10299, Feb. 2023, doi: 10.1007/s11227-023-05076-8.

[26] M. A. Guillén *et al.*, "Performance evaluation of edge-computing platforms for the prediction of low temperatures in agriculture using deep learning," *Journal of Supercomputing*, vol. 77, no. 1, pp. 818–840, Apr. 2021, doi: 10.1007/s11227-020-03288-w.

BIOGRAPHIES OF AUTHORS

Mokeddem Kamal Abdelmadjid is a dedicated individual currently pursuing a PhD at the University Tahri Mohamed of Bechar. His research interests span across various disciplines including web design, artificial intelligence, wireless sensors, embedded systems, and smart farming within greenhouses. With a keen focus on the integration of artificial intelligence and the internet of things (IoT), He is actively involved in exploring innovative solutions for smart microclimate farming applications within greenhouse environments. His commitment to advancing knowledge in these areas showcases his passion for cutting-edge technologies and their practical applications in agricultural settings. He can be contacted at email: mokeddem.kamal@univ-bechar.dz.

Seddiki Noureddine is a researcher at the University Tahri Mohamed of Bechar. His areas of interest encompass a wide range of fields, including embedded systems, wireless sensor networks, artificial intelligence, the internet of things, and smart agriculture in greenhouses. Noureddine is actively engaged in the investigation of novel approaches for intelligent microclimate applications in agriculture and greenhouses, with a focus on the integration of artificial intelligence and the IoT. His dedication to furthering the areas' understanding is evidence of his passion for cutting-edge technology and its usefulness in farming environments. He can be contacted at email: seddiki.noureddine@univ-bechar.dz.

Bourouis Amina received her CS bachelor's and master's degrees in 2011 and 2013 respectively, from the University Tahri Mohammed of Bechar, Algeria, and her Ph.D. from the University of Tunis El Manar, Tunis, Tunisia in 2019. Currently, she is an associate professor at Tahri Mohammed University, teaching DevOps and algorithms, and grid computing in the Maths and computer sciences department. Her research interests include formal verification, automata theory, web services, and IoT. He can be contacted at email: amina.bourouis@univ-bechar.dz.

Benahmed Khelifa has done his research for a Ph.D. degree in the field of wireless sensor networks at the School of Computing and Mathematical Sciences at Liverpool John Moores University, UK. He has 30 years of experience in education and research at the University Tahri Mohamed of Bechar, Algeria. Khelifa has also several international publications in the fields of wireless sensor networks, security, surveillance, smart grid, internet of things, Smart agriculture, and Smart transport. Khelifa Benahmed holds engineering and magister degrees and a Ph.D. in computer science from the University of Oran Algeria. He can be contacted at email: benahmed_khelifa@yahoo.fr.