



الجمهورية الجزائرية الديمقراطية الشعبية
People's Democratic Republic of Algeria
وزارة التعليم العالي والبحث العلمي
Ministry of Higher Education and Scientific Research
المدرسة الوطنية العليا للتكنولوجيات المتقدمة
National Higher School of Advanced Technologies



Department of Electrical Engineering and Industrial Computing

End of Study Project to Obtain the Diploma of
Master

Field
Telecommunication

Speciality
Telecommunications Systems and Networks

Subject
**Greenhouse Climate Forecasting:
an RNN-LSTM approach**

Realized by
ZEGHDAR Seif El Islam, SOUIKI Mohamed Chakib

Members of The Jury:

President	Ms. M. BOUTERFAS
Promoter	Ms. L. ABBAD
Promoter	Ms. K. LAKHDARI
Examiner	Mr. E. KHOUMERI
Examiner	Mr. A. HENTOUT

Algiers, the 20 / 06 /2024

Academic year 2023 – 2024

Greenhouse Climate Forecasting: an RNN-LSTM approach

Seif el islam ZEGHDAR, Mohamed chakib SOUIKI
Communications Systems and Networks
National Higher School of Advanced Technologies
Algiers, Algeria
s_zeghdar@ensta.edu.dz, m_souiki@ensta.edu.dz

Abstract—Accurate climate forecasting in greenhouses is crucial for efficient resource management and optimizing crop yields. Traditional forecasting methods often struggle due to the complexity of these ecosystems influenced by variables like temperature, humidity, light, and CO₂ levels. Advanced machine learning techniques, particularly RNNs and LSTMs, excel in managing this complexity by handling sequential data and long-term dependencies. However, each technique has its limitations. To address these limitations, a hybrid RNN and LSTM model was introduced, improving forecasting accuracy and reliability. Trained on the Jena weather time series dataset and evaluated on the Algeria Chapter Green dataset, the model, featuring two layers of RNN and LSTM with 64 cells each, achieved high accuracy with low MAE and MSE values, notably for air pressure (0.73 mbar, 0.12 mbar), temperature (0.76°C, 0.12°C), and air density (3.59 g/m³, 0.58 g/m³). The coefficients of determination (R²) for these parameters were 0.9723, 0.9886, and 0.9899, respectively. These findings demonstrate the potential of advanced neural networks in improving greenhouse management and agricultural productivity.

Keywords— Greenhouse, Climate prediction, RNN, LSTM

I. INTRODUCTION

Climate prediction within controlled environments like greenhouses has become an essential aspect of modern agriculture, enabling more efficient resource management and maximizing crop yields [1]. The complexity of greenhouse ecosystems, influenced by numerous interacting variables such as temperature, humidity, light, and CO₂ levels, poses significant forecasting challenges. Traditional prediction models often struggle to capture the nonlinear and dynamic nature of these systems, prompting the need for more sophisticated approaches [2].

In recent years, advancements in machine learning, particularly deep learning, have shown substantial potential in addressing these challenges. Among these, Recurrent Neural Networks (RNNs) and their enhanced variants, Long Short-Term Memory (LSTM) networks, stand out as powerful tools for time series prediction due to their proficiency in handling sequential data and maintaining long-term dependencies [3].

This article provides a comprehensive overview of the current state-of-the-art techniques in greenhouse climate prediction, highlighting the strengths and limitations of various approaches. It then introduces our proposed solution that leverages a combination of RNN and LSTM models to improve the accuracy and reliability of climate forecasts within greenhouse environments. Our methodology details the entire process from data collection and preprocessing to model training and evaluation, showcasing the effectiveness of these deep learning models in predicting greenhouse climate conditions.

By integrating advanced neural network architectures, this approach seeks to contribute to the growing body of

knowledge in precision agriculture, offering a robust framework for better greenhouse management and increased agricultural productivity.

II. RELATED WORK

The field of greenhouse climate prediction has witnessed significant research efforts over the past few decades, with various methodologies being proposed and implemented. This section reviews key contributions and advancements, covering both traditional and modern approaches.

Traditional Methods

Early efforts in greenhouse climate prediction primarily relied on statistical and physical models. These models, such as autoregressive integrated moving average (ARIMA) and multiple linear regression (MLR), used historical data to identify patterns and make future predictions. Although these methods laid the groundwork for climate forecasting, they often struggled with the nonlinear and dynamic nature of greenhouse environments [2]. Moreover, physical models, which are based on the principles of thermodynamics and plant physiology, required extensive parameterization and were computationally demanding.

Machine Learning Approaches

With the advent of machine learning, more sophisticated models began to emerge. Algorithms like Support Vector Machines (SVMs), Decision Trees, and Random Forests have been applied to greenhouse climate prediction with varying degrees of success. These methods improved prediction accuracy by capturing more complex relationships within the data compared to traditional statistical models. For instance, LightGBM, a framework for Gradient Boosting Decision Trees (GBDT), has shown effectiveness in greenhouse temperature prediction. Cao et al. [4] demonstrated the good performance of LightGBM in fitting accuracy, highlighting the strong time sequence in greenhouse temperature data and the high correlation of current readings with previous values.

Deep Learning Advances

The introduction of deep learning has revolutionized greenhouse climate prediction. Initially, Convolutional Neural Networks (CNNs) were employed, particularly for spatial data analysis. However, for time-series data, RNN and LSTM networks have proven superior. They can effectively model the sequential nature of climate data, capturing temporal dependencies that other models miss. For example, Moon et al. [5] used an LSTM model to predict CO₂ concentration in greenhouses, achieving an R² of nearly 0.8. Their study highlighted the importance of selecting appropriate time steps and learning rates to optimize model performance. Similarly, Liu et al. [6] demonstrated that LSTM models outperformed RNN and GRU models across multiple climate factors in greenhouses.

Hybrid Models

To further enhance prediction accuracy, hybrid models that combine multiple machine learning and deep learning techniques have been explored. For instance, integrating CNNs with LSTMs allows for the extraction of both spatial and temporal features from the data. Similarly, combining traditional statistical models with modern machine learning algorithms can leverage the strengths of both approaches. An example is the work by Petrakis et al. [7], where a multilayer perceptron (MLP) neural network was used for greenhouse microclimate predictions. Their model, incorporating optimal node selection and data preprocessing, produced reliable results during the testing phase, underscoring the importance of effective model architecture and preprocessing in achieving satisfactory performance. Additionally, Eraliev et al. [8] evaluated multiple time series deep learning models, including LSTM, Gated Recurrent Unit (GRU), and hybrid models, at different time intervals to predict climate variables such as temperature, humidity, and CO₂ levels in indoor hydroponic greenhouses. Their findings indicate that hybrid models often outperform standalone models and that specific models are more suitable for certain prediction intervals, aiding in precise climate control in hydroponic greenhouses.

Comparison and Limitations

Despite significant progress, each approach has its limitations. Traditional models are often inadequate for complex, dynamic environments. Machine learning methods, although more powerful, can suffer from overfitting and require substantial amounts of data. Deep learning models, particularly RNNs and LSTMs, mitigate some of these issues but can be computationally expensive and require careful hyperparameter tuning. For example, Moon et al. [5] highlighted that longer time steps led to decreased test accuracies, stressing the need for careful parameter optimization.

Emerging Trends

Recent research has started exploring the use of advanced architectures like Transformer models and attention mechanisms, which have shown promising results in other time-series prediction domains [9]. Additionally, integrating Internet of Things (IoT) devices for real-time data collection and applying reinforcement learning for adaptive climate control are emerging trends that could transform greenhouse climate prediction [10].

In summary, the evolution of greenhouse climate prediction methods reflects a gradual shift from simple statistical models to sophisticated machine learning and deep learning techniques. The ongoing advancements hold great promise for improving the accuracy and reliability of climate forecasts, ultimately contributing to more efficient and productive agricultural practices.

Overview of Key Research

In this section, an overview of the current state of the art and related work in the realm of greenhouses climate forecasting is presented. “Table I” summarizes key research details: methodologies, results, and key findings from recent studies. This comprehensive summary aims to highlight advancements and ongoing challenges in the existing literature of greenhouse climate prediction, offering valuable insights and a solid foundation for further research and development in the field.

TABLE I. SUMMARY OF RECENT STUDIES

Methodology	Results and key findings	Related paper
LightGBM, a Gradient Boosting Decision Trees framework.	- The model demonstrated good fitting accuracy. - High correlation between current and previous greenhouse temperature readings led to superior early-stage fitting compared to later stages.	[4]
LSTM model.	- An 80% accuracy is achieved. - Longer time steps decreased accuracy, stressing optimal time step selection. - Various learning rates had minimal impact on accuracy. - Data beyond 720 minutes was insignificant for prediction due to rapid changes. - Unused plant growth data could improve model robustness. - External CO₂ concentrations offered limited assistance, being relatively constant.	[5]
GCP_LSTM used for prediction, compared to RNN and GRU models.	- LSTM outperformed RNN and GRU across all climate factors. -Time sliding window is crucial for accurate trend analysis and future anticipation. - RNNs excel over ANNs in greenhouse climate forecasting. - Narrow focus on isolated factors poses reliability challenges in climate forecasting. - Optimal number of units balances detail capture and overfitting risk in the model.	[6]
MLP neural network with 10 input nodes (one for each variable), a single hidden layer, and 2 output nodes (representing inside temperature and relative humidity).	- The model, guided by statistical indices like MAE, RMSE, and R², produced reliable results during testing. - Effective preprocessing and proper model architecture selection enhance the model's reliability and applicability in real-world greenhouse management scenarios.	[7]
Evaluated LSTM, GRU, and hybrid models at different time intervals.	- Detailed performance metrics showcased accuracy and efficiency across various prediction intervals. - Hybrid models, notably combining LSTM with other architectures, outperformed standalone models.	[8]

Datasets overview

This section provides a comprehensive overview of the datasets used in both our study and previously referenced studies. Each dataset is briefly summarized, detailing its name, source, location and other details. This information offers valuable insight into the scope and nature of the data used for analysis and modeling purposes. “Table II” presents the datasets alongside their essential attributes.

TABLE II. DATASET OVERIVIEWS

Dataset	Time period	Crop type	Variables measured	Frequency of measurement	Number of samples	Related paper
iFLYTEK 2020 online competition dataset	March 14, to April 13, 2019	Not specified	- In / Out temperature. - In / Out humidity. - In / Out air pressure.	1 minute	25,904	[4]
Greenhouse environmental data	February 2, 2017 to May 31, 2018	Mango trees	- In / Out temperature. - In / Out relative humidity. - In / Out atmospheric pressure. - Inside CO2 concentration. - Inside light intensity. - Soil temperature. - Soil moisture content. - Soil electrical conductivity. - Wind direction. - Wind velocity.	10 minutes	69,684	[5]
Tomato, cucumber and pepper greenhouses datasets	- Tomato: May 14, 2017 to August 6, 2017 - Cucumber: May 14, 2017 to July 31, 2017 - Pepper: May 14, 2017 to July 10, 2017	Tomato, cucumber and pepper	- Temperature. - Humidity. - Illumination. - Soil temperature. - Soil humidity. - CO2 concentration.	1 minute	104,836	[6]
The University of Patras greenhouse dataset	February 14, 2022 to April 18, 2022.	Not specified	- In / Out temperature. - In / Out relative humidity. - Wind speed. - Solar irradiance.	10 minutes	8,594	[7]
Indoor hydroponic greenhouses climate data	March 16, to March 23, 2023	Not specified	- Temperature. - Humidity. - CO2 levels.	1 minute	Not specified	[8]
Jena weather time series dataset	between 2009 and 2016	Not specified	- Air pressure. - Temperature. (Absolute, Potential and Dew). - Humidity (Relative and Specific). - Vapor pressure (Maximum, Actual and deficit). - Water vapor concentration. - Air density. - Wind velocity (Absolute and Maximum). - Wind direction.	10 minutes	420,551	[11]
Algeria Chapter Green dataset	Not specified	Not specified	- Indoor temperature. - Indoor humidity. - Indoor CO2 concentration. - Inside light intensity. - Soil temperature. - Soil moisture content.	1 minute	1,440	[12]

III. IMPLEMENTATION

A. Materials and methods

1) Data Description

In this study, two datasets were utilized. The first dataset was employed during the model-building phase, while the second dataset, serving as an exemplary dataset representing the type of data the model will process in real-world applications. This second dataset was used to evaluate the model's performance. Duplicating the testing step on a fully new and unseen data ensures that the model is effective and reliable for its intended use.

Dataset 1

The first dataset, used during the model-building phase, is the Jena weather time series dataset recorded by the Max Planck Institute for Biogeochemistry in Germany [11].

This dataset includes 14 different features such as air temperature, atmospheric pressure, and humidity. These were collected every 10 minutes, beginning in 2003. For efficiency, only the data collected between 2009 and 2016 were utilized, as prepared by François Chollet for his book Deep Learning with Python.

The dataset, comprising 420,551 records, was first preprocessed and underwent future engineering before being fed to the model. To optimize efficiency, we used a time step of 30 minutes, reducing the dataset size to 140,182 records, which was sufficient for our analysis. The data was then split into training, validation and test datasets with portions of 70%, 20%, and 10%, respectively. To ensure the data could be divided into windows of consecutive samples and to provide more realistic validation and test results, it was not randomly shuffled before splitting.

Further Statistical information about Dataset 1 is provided in “Table III”.

TABLE III. DATASET 1 OVERVIEW

Feature	Description	Range	Mean
P (mbar)	Air pressure in millibars	913.60 - 1015.35	989.21
T (°C)	Temperature in degrees Celsius	-22.76 - 37.28	9.45
T _{pot} (K)	Potential temperature in Kelvin	250.85 - 311.21	283.49
T _{dew} (°C)	Dew point temperature in degrees Celsius	-24.80 - 23.06	4.95
RH (%)	Relative humidity as a percentage	13.88 - 100.00	76.00
VP _{max} (mbar)	Maximum vapor pressure in millibars	0.97 - 63.77	13.57
VP _{act} (mbar)	Actual vapor pressure in millibars	0.81 - 28.25	9.53
VP _{def} (mbar)	Vapor pressure deficit in millibars	0.00 - 46.01	4.04
SH (g/kg)	Specific humidity in grams per kilogram	0.51 - 18.07	6.02
H2Oc (mmol/mol)	Water vapor concentration in millimoles per mole	0.81 - 28.74	9.64
rho (g/m ³)	Air density in grams per cubic meter.	1059.45 - 1393.54	1216.05

WV (m/s)	Wind velocity in meters per second	-9999.00 - 14.01	1.70
max. WV (m/s)	Maximum wind velocity in meters per second	-9999.00 - 23.50	3.03
WD (deg)	Wind direction in degrees	0.00 - 360.00	174.72

Also, the evolution of a few selected features over time in this dataset is presented in “Fig. 1”

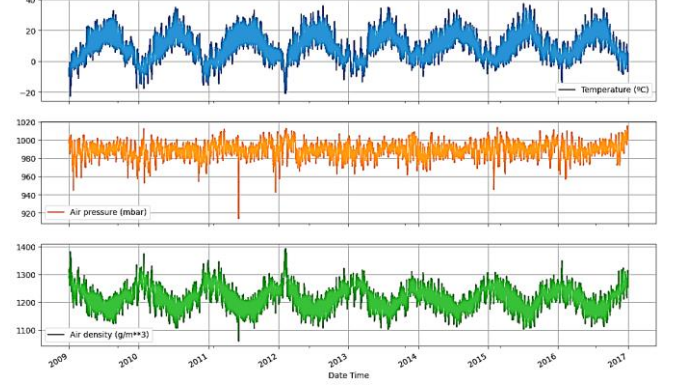


Fig. 1. Dataset 1 features evolution over time.

Dataset 2

Serving as an exemplary dataset, the Algeria Chapter Green dataset represents the type of data the model will process in real-world applications [12]. This dataset was used to evaluate the model's performance ensuring its effectiveness and reliability in the real world context. It includes all the features, such as air temperature and humidity, that the model will encounter in actual use.

The data were collected during the Green Algeria Project every minute for one day, resulting in 1,440 records. This amount was insufficient for a thorough performance evaluation, so data augmentation and time shifting were applied, expanding the dataset by a factor of 6 to a total of 8,640 records, which is acceptable for model evaluation.

This second dataset underwent the same preprocessing, feature engineering, and train-validation-test split steps as the first dataset.

Further statistical information about Dataset 2 is provided in “Table IV”.

TABLE IV. DATASET 2 OVERVIEW

Feature	Description	Range	Mean
indoor_temp	Temperature inside the greenhouse	4.34 - 39.13	16.10
indoor_humidity	Level of humidity inside the greenhouse	29.91 - 85.31	70.95
indoor_lighting	Intensity of light inside the greenhouse	-0.32 - 346.00	65.73
indoor_CO2	Concentration of carbon dioxide inside the greenhouse	498.75 - 921.13	762.14
soil_temp	Temperature of the greenhouse soil	9.20 - 19.44	13.91
soil_moisture	Moisture level of the greenhouse soil	20.64 - 23.30	22.01

Also, the evolution of a few selected features over time in this dataset is presented in “Fig. 2”

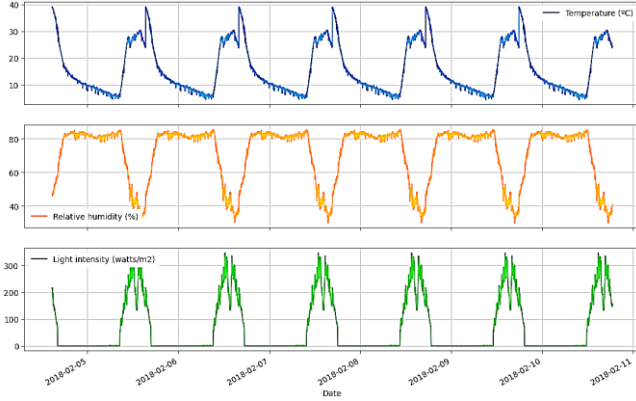


Fig. 2. Dataset 2 features evolution over time.

2) Data Preprocessing

Potential equipment abnormalities during data collection might result in gaps or anomalies in the datasets [13]. Moreover, the data augmentation process applied to Dataset 2 may introduce anomalies and erroneous values. Upon inspecting the original data, erroneous minimum values were noted for Wind velocity “wv (m/s)” and Maximum wind velocity “max. wv (m/s)” in Dataset 1, as well as for Indoor lighting in Dataset 2.

To address these issues, corrective measures were implemented to ensure that the data fell within reasonable ranges.

3) Feature Engineering

In this phase, it's important to ensure that the data is appropriately formatted for the model. Due to their unsuitable nature, some data features were modified based on prior experience indicating their limited usefulness. Specifically, Wind velocity “wv (m/s)” and Maximum wind velocity “max. wv (m/s)” from Dataset 1 were converted into x and y vectors along with the Wind direction “wd (deg)” feature. These vectors facilitate the model's interpretation of wind-related data distribution.

Additionally, useful insights can be derived from the DateTime column in the two datasets by transforming it into sine and cosine signals to represent ‘Time of day’ and ‘Time of year’. This allows the model to access important frequency features.

4) Data Normalization

It is important to scale features before training a neural network. A common method is normalization: subtract the mean and divide by the standard deviation of each feature for each sample [13]. Importantly, the mean and standard deviation should be computed using only the training data so that the model does not have prior access to the values in the validation and test sets. The formula for calculating this normalization is clarified in (1) below:

$$X_N = \frac{X - \mu_{\text{train}}}{\sigma_{\text{train}}} \quad (1)$$

Where X_N is the normalized value, X is the original value, μ_{train} is the train set mean and σ_{train} is the train set standard deviation.

5) Data structuring

In the time series forecasting domain, data can be passed to the model one time step at a time or in multiple time steps simultaneously, known as a window of data. This approach allows the model to generate one or multiple time step predictions. Therefore, the data should be structured into windows of successive time steps, with inputs (X) and labels (Y). The size of these windows can be adjusted depending on the prediction requirements.

6) Forecasting model

Most of the studies referenced used a single model for forecasting and examined its performance. However, a less explored solution is to combine two models to leverage the advantages of each for optimal performance.

In our study, two models were combined: RNN and LSTM. This combined approach enables the model to effectively process and learn from sequential data, leveraging the RNN's capability to handle sequences and the LSTM's ability to manage long-term dependencies. Addressing challenges such as the vanishing gradient, the LSTM enhances prediction accuracy and reliability over longer sequences compared to a standard RNN. Consequently, the model captures both short-term patterns and long-term trends, providing a comprehensive approach for time series forecasting.

Delving into the model architecture, the network initiates with an RNN layer comprising 64 cells and ReLU activation, followed by a Dropout layer to prevent overfitting by randomly setting 20% of the input units to zero at each update. Next, an LSTM layer with 64 cells using the tanh activation function is incorporated. Following this, a Dense layer with 8 units and ReLU activation is introduced to capture a more complex representation of the data, accompanied by another Dropout layer to further mitigate overfitting. Lastly, a Dense layer featuring a linear activation function is employed to perform linear transformations on the output of previous layer, complemented by a Reshape layer to tailor the output shape to match the desired format.

“Fig. 3”, provides a summary of the model architecture.

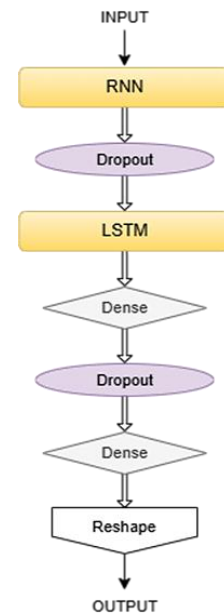


Fig. 3. Proposed model architecture

B. Results and Discussion

1) Evaluation Indicator

For evaluation, the mean square error (MSE), mean absolute error (MAE), and R-square coefficient of determination (R²) were used to assess the model performance. The formulas for calculating the evaluation indicator are:

$$\text{MSE} = \frac{1}{N} \sum_{i=1}^N \|y_i - \hat{y}_i\|^2 \quad (2)$$

$$\text{MAE} = \frac{1}{N} \sum_{i=1}^N \|y_i - \hat{y}_i\| \quad (3)$$

$$R^2 = 1 - \frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{\sum_{i=1}^N (y_i - \bar{y})^2} \quad (4)$$

where y_i is the actual value and $\hat{y}_i$ is the predicted value of the i -th group of performance evaluation data, $\bar{y}$ is the average value of data, and N is the total number of samples. For MSE and MAE, a smaller index value indicates a better prediction. The R² coefficient of determination indicates the degree of fitting between the predicted value and the actual value.

2) Analysis of Results

In the present study, the model was used to predict temperature, air density, and air pressure inside the greenhouse. To determine the optimal architecture, multiple configurations were tested with varying numbers of cells and different structures. Evaluation metrics MAE, MSE, and R² indicated that the best architecture was found to be two layers of RNN and LSTM, each with 64 cells.

The model training process concluded after 13 epochs, with the best epoch determined by the minimum validation MSE of approximately 1.35×10^{-2} , occurring at the 2nd epoch.

The model's performance in predicting air pressure was satisfactory, with MAE and MSE values of 0.73 mbar and 0.12 mbar², respectively. Similarly, temperature predictions yielded low MAE and MSE values of 0.76°C and 0.12°C², respectively. Air density predictions also showed small MAE and MSE values of 3.59 g/m³ and 0.58 g²/m⁶, respectively. The coefficients of determination (R²) were high for all parameters: 0.9723 for air pressure, 0.9886 for temperature, and 0.9899 for air density, demonstrating the model's strong performance. The following "Table V" summarize the statistics about the model's performance.

TABLE V. MODEL PERFORMANCE METRICS

	MAE	MSE
Air pressure	0.73 mbar	0.12 mbar ²
Temperature	0.76 °C	0.12 °C ²
Air density	3.59 g/m ³	0.58 g ² /m ⁶

"Fig. 4" illustrates the time series of predicted and observed parameters: air pressure, temperature, and air density over a two-day period from the test dataset. This visual representation showcases a close match between predicted and observed values, indicating a high degree of alignment and underscoring the accuracy of the model.

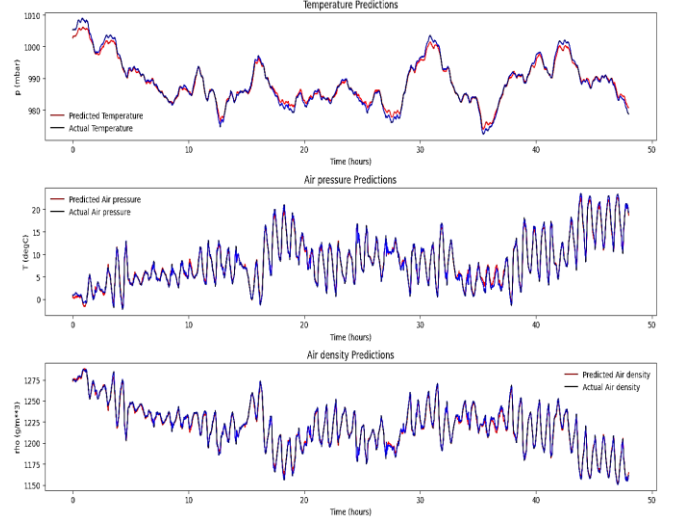


Fig. 4. Parameters predictions vs actual values

IV. DISCUSSION

The study employed three different types of error metrics to evaluate and compare the predicted and observed values for the three parameters under study: indoor air pressure, indoor temperature, and indoor air density. This comprehensive approach facilitated a detailed comparison with a wide range of literature, as different studies employ various formulas to evaluate errors. The low MAE and MSE values calculated highlighted the robustness of the model. Additionally, extensive literature review confirmed that the coefficients of determination (R²) for all three parameters were very high, with values close to 1, underscoring the model's exceptional predictive capability.

A key benefit of the hybrid RNN and LSTM model lies in its capacity to capture both sequential and long-term dependencies within the data, which is essential for precise prediction of complex environmental variables in greenhouses. Despite the computational complexity and the need for careful tuning, the outcomes demonstrate that such sophisticated models can significantly outperform traditional methods. The low MAE and MSE values, along with high R² scores, indicate that the model is not only accurate but also reliable for practical applications in greenhouse management.

The robustness of the model is further demonstrated by its strong performance across various weather datasets. The successful application to both the Jena weather time series and the Algeria Chapter Green dataset underscores the model's flexibility and reliability. This versatility indicates that the model can be effectively utilized for different climatic conditions, enhancing its utility in diverse greenhouse environments worldwide.

V. CONCLUSION

This study effectively utilized two distinct datasets to develop and validate a neural network model for predicting greenhouse conditions, such as temperature, air density, and air pressure. The first dataset, the Jena weather time series, was used during the model-building phase and underwent extensive preprocessing and feature engineering to optimize the data for training. The second dataset, which simulates real-world greenhouse data, was used to evaluate the model's performance, ensuring its reliability for practical applications.

By testing various configurations, the optimal model architecture was identified as two layers of RNN and LSTM, each with 64 cells. The training process concluded after 13 epochs, with the best performance observed at the 2nd epoch, achieving a minimum validation MSE of 1.35×10^{-2} .

The model demonstrated strong predictive capabilities, as indicated by the low MAE and MSE values for all target parameters. Specifically, the model achieved MAE and MSE values of 0.73 mbar and 0.12 mbar for air pressure, 0.76°C and 0.12°C for temperature, and 3.59 g/m³ and 0.58 g/m³ for air density. Additionally, the coefficients of determination were notably high: 0.9723 for air pressure, 0.9886 for temperature, and 0.9899 for air density, indicating excellent model performance.

Visual analysis of the time series predictions for air pressure, temperature, and air density showed a close match between predicted and observed values, confirming the model's accuracy. This was further validated by the three different types of error descriptions used to compare predicted and observed values, facilitating a comprehensive comparison with existing literature.

The model's adaptability and reliability are evidenced by its consistent performance across different weather datasets. By successfully applying the model to both the Jena weather time series and the Algeria Chapter Green dataset, the study demonstrates the model's versatility and suitability for real-world applications under various climatic conditions. This flexibility enhances its utility in diverse greenhouse environments worldwide, ensuring it can effectively support agricultural productivity in different settings.

In summary, the study successfully developed a reliable model for predicting key greenhouse parameters by leveraging preprocessing, feature engineering, and a combined RNN-LSTM architecture. This model is well-suited for practical applications, offering high accuracy and robustness, and holds significant potential for improving greenhouse management and control systems.

REFERENCES

- [1] T. Karanisa, Y. Achour, A. Ouammi, and S. Sayadi, "Smart greenhouses as the path towards precision agriculture in the food-energy and water nexus: case study of Qatar," *Environmental Systems and Decisions*, vol. 42, pp. 521–546, 2022.
- [2] O. Alizadeh, "Advances and challenges in climate modeling," *Climatic Change*, vol. 170, p. 18, 2022.
- [3] B. Ghogh and A. Ghodsi, "Recurrent neural networks and long short-term memory networks: Tutorial and survey," *arXiv preprint arXiv:2304.11461*, 2023.
- [4] Q. Cao, Y. Wu, J. Yang, and J. Yin, "Greenhouse temperature prediction based on time-series features and LightGBM," *Applied Sciences*, vol. 13, no. 3, p. 1610, 2023.
- [5] T. Moon, H. Y. Choi, D. H. Jung, S. H. Chang, and J. E. Son, "Prediction of CO₂ concentration via long short-term memory using environmental factors in greenhouses," *Horticultural Science and Technology*, vol. 38, no. 2, pp. 201–209, 2020.
- [6] Y. Liu, D. Li, S. Wan, F. Wang, W. Dou, X. Xu, 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, 2022.
- [7] T. Petrakis, A. Kavga, V. Thomopoulos, and A. A. Argiriou, "Neural network model for greenhouse microclimate predictions," *Agriculture*, vol. 12, no. 6, p. 780, 2022.
- [8] 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, 2023.
- [9] D.-K. Kim and K. Kim, "A Convolutional Transformer Model for Multivariate Time Series Prediction," *IEEE Access*, vol. PP, no. 99, pp. 1–1, Jan. 2022.
- [10] Y. Hu, Y. Liu, A. Kaushik, C. Masouros, and J. Thompson, "Timely Data Collection for UAV-based IoT networks: A Deep Reinforcement Learning Approach," *IEEE Sensors Journal*, vol. 23, no. 11, pp. 12295–12308, 2023.
- [11] Max Planck Institute for Biogeochemistry, "Jena Weather Dataset [Data set]," 2009–2016. Available: <https://www.kaggle.com/datasets/harishedison/jena-weather-dataset/data>. [Accessed: Jun. 10, 2024].
- [12] AGC Dataset Analysis, "Algeria Chapter Green dataset [Data set]," 2022. Available: <https://github.com/adityarags/AGC-Dataset-Analysis/tree/main/Dataset>. [Accessed: Jun. 10, 2024].
- [13] "Data Normalization Explained: Types, Examples, & Methods," *Estuary*, 30-Jul-2023. [Online]. Available: <https://estuary.dev/data-normalization/>. [Accessed: Jun. 17, 2024].