iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 14 **Issue:** X **Month of publication:** October 2026

DOI: https://doi.org/10.22214/ijraset.2026.85491

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue X Oct 2026- Available at www.ijraset.com

Machine Learning-Based Forecasting of Outlet Air Temperature in an Earth-to-Air Heat Exchanger

Gubbala Bindu Madhavi[1], Ravi Kumar Boddu[2], Lakshmi Tulasi Chavali[3], Suresh Kumar Mandru[4]
[1]*M.Tech. Scholar, Department of Mechanical Engineering, St. Ann's College of Engineering & Technology, Chirala, Andhra Pradesh, India*
[2]*Assistant Professor, Department of Mechanical Engineering, St. Ann's College of Engineering & Technology, Chirala, Andhra Pradesh, India*
[3]*Professor, Department of Mechanical Engineering, St. Ann's College of Engineering & Technology, Chirala, Andhra Pradesh, India*
[4]*Associate Professor, Department of Mechanical Engineering, St. Ann's College of Engineering & Technology, Chirala, Andhra Pradesh, India*

Abstract: ***Earth-to-air heat exchangers (EAHEs) exchange sensible heat between flowing air and the surrounding ground. Forecasting outlet air temperature over short horizons may support monitoring and future supervisory control studies. This study benchmarks four established regression models—Linear Regression, Decision Tree, Random Forest, and HistGradientBoosting—for one-, three-, and six-hour-ahead prediction using an hourly EAHE analysis dataset. A persistence forecast, which carries the latest measured outlet temperature forward, provides a reference. The 3,120 observations span 9 November 2023 to 17 March 2024. Data were ordered chronologically and divided into training and testing segments using an 80:20 split; median imputation was fitted within the training pipeline. Model performance was assessed using mean absolute error (MAE), root mean squared error (RMSE), and the coefficient of determination (R^2). Random Forest produced the best one-hour result (MAE 0.0578 °C, RMSE 0.0815 °C, R^2 0.9855) and the best three-hour result (MAE 0.1974 °C, RMSE 0.2755 °C, R^2 0.8341). At six hours, Linear Regression was the strongest evaluated regressor (MAE 0.3750 °C, RMSE 0.5124 °C, R^2 0.4257). Errors increased with forecast horizon, and the best model depended on lead time. The contribution is a transparent, dataset-specific comparison rather than a new learning algorithm. External validation and direct measurement of energy or comfort outcomes are required before practical benefits can be established.***

Keywords: Earth-to-air heat exchanger; outlet-air temperature; machine learning; forecasting; Random Forest; Linear Regression; multi-horizon prediction; persistence baseline.

I. INTRODUCTION

Earth-to-air heat exchangers (EAHEs) use the ground as a thermal reservoir to moderate the temperature of air passing through buried ducts. Heat transfer involves convection between the air and pipe wall and conduction through the wall into or from the surrounding soil. Performance depends on inlet-air conditions, ground temperature and properties, pipe geometry, airflow, burial depth, and operating history. Reviews of EAHE modelling describe analytical, reduced-order, and numerical approaches, each balancing physical detail, data requirements, and computational cost [1,2]. Experimental and three-dimensional modelling studies have also investigated the sensitivity of EAHE behaviour to design and operating conditions [3–5,11,12].

Data-driven prediction can complement physical modelling when the objective is a fast forecast from recorded measurements. However, outlet-temperature prediction must be distinguished from heat-transfer rate, cooling capacity, coefficient of performance, and building-level energy savings; these quantities require additional variables or measurements. Machine learning has already been applied to EAHE outlet-temperature forecasting and broader performance prediction. Kifouche et al. [7] studied neural-network-based outlet-temperature prediction, while Yue et al. [8,9] examined multi-performance prediction and interpretable analysis of meteorological influences. Hybrid EAHE systems have also been evaluated using theoretical and experimental methods [10]. Thus, the novelty of the present work is not the use of machine learning itself.

Instead, this study provides a consistent benchmark of four accessible regression models against a persistence baseline on the supplied hourly analysis table. The forecasts are evaluated at 1-, 3-, and 6-hour horizons using a chronological hold-out test. The research questions are: (i) how does accuracy vary with forecast horizon; (ii) do the tested models improve on persistence; and (iii) which model provides the lowest errors at each horizon? Conclusions are limited to the dataset and evaluation protocol used.

©IJRASET: All Rights are Reserved

II. DATASET AND FORECASTING FORMULATION

A. Dataset and quality screening

The analysis table contains 3,120 hourly records from 9 November 2023 00:00 to 17 March 2024 23:00. The timestamp audit reported no duplicate or invalid timestamps, no missing target values, and a continuous sequence of 3,119 hourly intervals. Outlet temperature (OUT_y) has a mean of 19.25 °C, standard deviation of 1.43 °C, and a range of 16.86–23.80 °C. The table is described as derived from the public EAHE-D dataset introduced by Alidrissi et al. [6]; the exact raw-file and preprocessing lineage should be verified and documented before submission.

Predictors available at forecast origin t were OUT_y, IN_y, GroundLeft_y, GroundRight_y, Tair_Avg, RH_Avg, BP_CS100_Avg, WS_Avg, WD_Avg, WD_Std, and WSgust_Max. The three solar-radiation columns (GHI_corr_Avg, DHI_corr_Avg, and DNI_corr_Avg) each had approximately 83.04% missing values and were excluded from the primary models. Pipe geometry, airflow, fan power, and direct soil property measurements were not included in the analysis table; consequently, this study does not estimate causal effects or system efficiency.

Table 1. Dataset properties.

Dataset property	Value
Number of observations	3,120 hourly records
Period	2023-11-09 00:00 to 2024-03-17 23:00
Target	Outlet temperature, OUT_y
Target statistics	19.25 ± 1.43 °C; range 16.86–23.80 °C
Duplicate timestamps / missing target	0 / 0
Solar-radiation feature missingness	≈83.04% in each; excluded

B. Mathematical formulation

Let $x_t \in R^p$ denote the vector of p predictor measurements available at hour t, and let y_t denote the measured outlet temperature OUT_y at that time. For forecast horizon $h \in \{1, 3, 6\}$, the supervised learning dataset is defined as:

$$D_h = \{(x_t, y_{t+h}) : t = 1, \ldots, N - h\}.$$

A forecasting model f_h is fitted on the training observations and produces $\hat{y}_{t+h|t} = f_h(x_t)$. The notation emphasizes that only information available at forecast origin t is used to predict the future target. The persistence baseline is defined as $\hat{y}_{t+h|t} = y_t$, i.e., the latest measured outlet temperature is carried forward for each horizon.

For n test predictions, the evaluation metrics are:

$$MAE = (1/n) \Sigma_{i=1}^{n} |y_i - \hat{y}_i|$$

$$RMSE = \sqrt{[(1/n) \Sigma_{i=1}^{n} (y_i - \hat{y}_i)^2]}$$

$$R^2 = 1 - [\Sigma_i (y_i - \hat{y}_i)^2 / \Sigma_i (y_i - \bar{y})^2]$$

Here, y_i is the observed outlet temperature, $\hat{y}_i$ is the prediction, and $\bar{y}$ is the mean of the observed test targets. MAE and RMSE are expressed in °C; RMSE penalizes larger errors more strongly. R^2 is dimensionless and can be negative when predictions perform worse than the test-set mean reference.

III. MACHINE-LEARNING METHODS AND EXPERIMENTAL PROTOCOL

Four established regression methods were evaluated: Linear Regression, Decision Tree, Random Forest, and HistGradientBoosting. Linear Regression provides a linear reference; a Decision Tree captures threshold-based relationships; Random Forest averages tree predictions to reduce variance; and HistGradientBoosting builds a sequence of trees to improve the current ensemble. The models were compared using the same predictors and horizon-specific target alignment. No new learning algorithm is proposed.

For each horizon, the eligible records were retained in time order and divided into the first 80% for training and the final 20% for testing. Median imputation was fitted using training data only within the model pipeline, preventing test-set values from influencing imputation statistics. The persistence baseline was evaluated on the same horizon-specific test targets. The reported results are from a single chronological hold-out; no independent-site validation or rolling-origin evaluation was performed.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

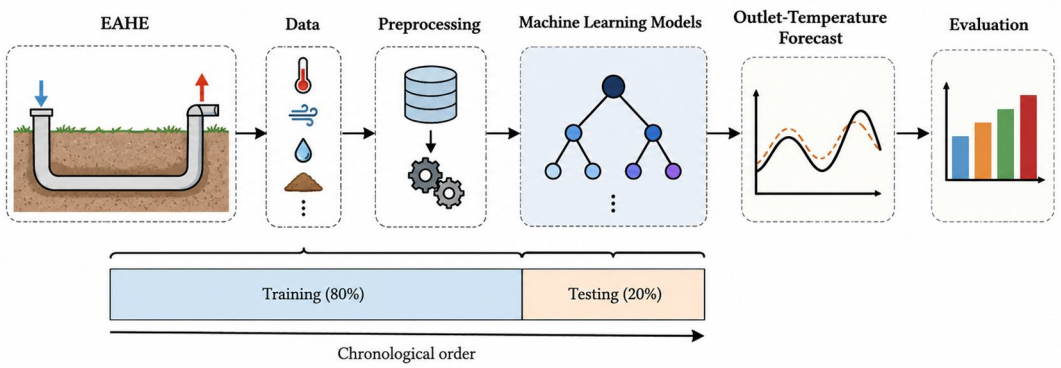


Figure 1. Workflow for EAHE outlet-temperature forecasting and chronological evaluation.

IV. RESULTS AND DISCUSSION

A. *Dataset Description and Quality Assessment*

The supplied analysis table contains 3,120 hourly observations from 09 November 2023 00:00 to 17 March 2024 23:00. Timestamps were parsed and sorted chronologically. The audit identified 0 duplicate timestamps and 0 missing target values. Outlet temperature had a mean of 19.25 °C, standard deviation 1.43 °C, and range 16.86–23.80 °C.

Each of the three solar-radiation columns (GHI_corr_Avg, DHI_corr_Avg, DNI_corr_Avg) has approximately 83.04% missing values. They were excluded from the primary modelling to avoid extensive imputation. This choice should be reported as a data limitation rather than treating imputed solar radiation as measured evidence.

Table 2. Dataset characteristics and quality checks.

Dataset property	Value
Observations	3,120
Sampling interval	1 hour
Start	2023-11-09 00:00:00
End	2024-03-17 23:00:00
Target	Outlet air temperature (OUT_y)
Mean ± SD	19.25 ± 1.43 °C
Minimum / maximum	16.86 / 23.80 °C
Duplicate timestamps	0
Missing target values	0
Missing in each radiation variable	83.04%

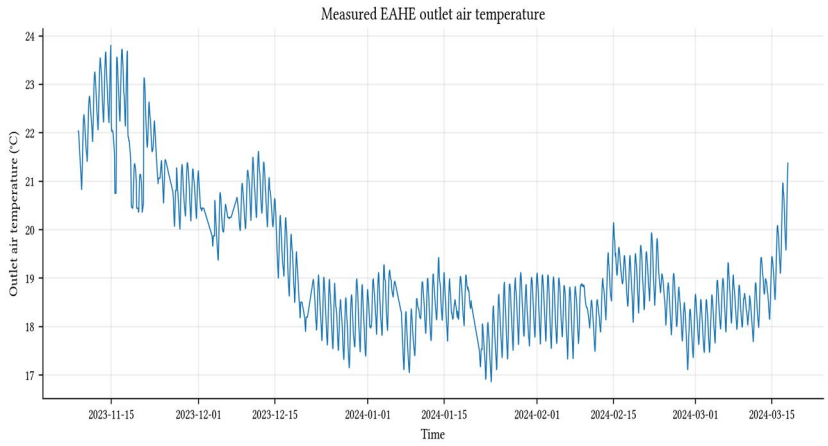


Figure 2. Hourly variation in measured EAHE outlet air temperature.

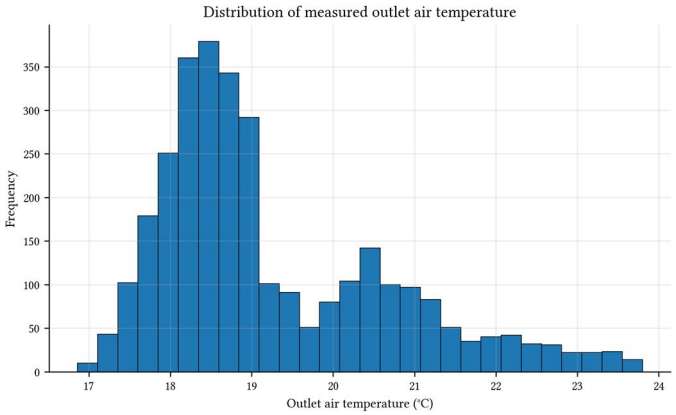


Figure 3. Distribution of outlet air temperature in the analysis dataset.

B. Forecasting Problem and Experimental Protocol

At forecast origin t, the input vector included measurements available at t: OUT_y, IN_y, GroundLeft_y, GroundRight_y, Tair_Avg, RH_Avg, BP_CS100_Avg, WS_Avg, WD_Avg, WD_Std, and WSgust_Max. The target was OUT(t+h), with h = 1, 3, or 6 hours. Thus, the task is a forecast from the current state, not a same-timestamp reconstruction.

For each horizon, eligible observations were kept in chronological order; the first 80% were used for training and the final 20% for testing. Models were Linear Regression, Decision Tree, Random Forest, and HistGradientBoosting. The persistence baseline predicts OUT(t+h) using OUT(t). Any median imputation was performed inside the model pipeline and fitted on the training portion. Model quality was assessed using MAE, RMSE, and R^2 on the held-out later period.

MAE is the mean absolute difference between predictions and measurements, in °C. RMSE is the square root of mean squared error and penalizes large errors more strongly. R^2 compares model predictions with a test-set mean predictor. These metrics should be interpreted together rather than treating R^2 alone as proof of engineering usefulness.

C. One-Hour-Ahead Prediction Results

Random Forest gave the lowest one-hour-ahead error (MAE = 0.0578 °C, RMSE = 0.0815 °C, R^2 = 0.9855). The persistence baseline obtained MAE = 0.0901 °C and RMSE = 0.1030 °C. Relative to this baseline, Random Forest reduced MAE by 35.8% and RMSE by 20.8% on the test period. HistGradientBoosting and Linear Regression also performed competitively.

Table 3. One-hour-ahead performance on the chronological test set.

Model	MAE (°C)	RMSE (°C)	R^2
Random Forest	0.0578	0.0815	0.9855
HistGradientBoosting	0.0633	0.0853	0.9841
Linear Regression	0.0649	0.0870	0.9835
Decision Tree	0.0676	0.0952	0.9802
Persistence baseline	0.0901	0.1030	0.9768

The high short-horizon scores should be interpreted in light of thermal and temporal continuity: the current outlet temperature is an input and is likely strongly associated with the value one hour later. The results support short-horizon forecasting for this dataset; they do not by themselves establish a new heat-transfer mechanism or guarantee performance at another site.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

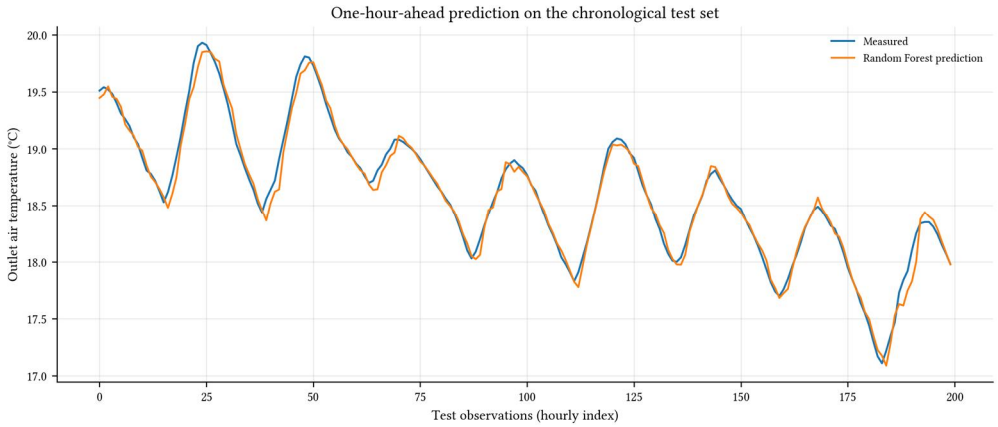


Figure 4. Measured and Random Forest-predicted outlet temperature over the first 200 test observations.

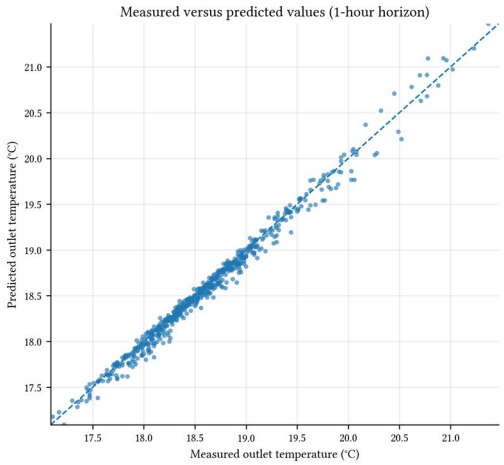


Figure 5. Measured versus predicted outlet temperature for the one-hour Random Forest forecast.

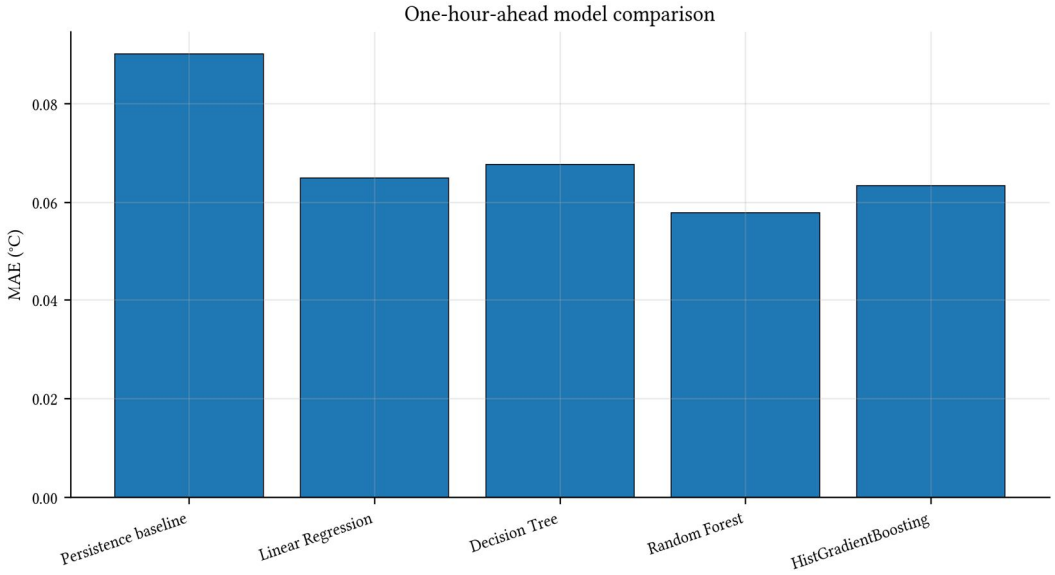


Figure 6. MAE comparison for the one-hour-ahead models and persistence baseline.

D. Effect of Forecast Horizon

At three hours, Random Forest was best by MAE (0.1969 °C; RMSE = 0.2739 °C; R^2 = 0.8355). At six hours, Linear Regression had the lowest MAE (0.3745 °C) and highest R^2 (0.4228) among the tested models. Prediction accuracy declined as the forecast horizon increased, indicating that current-state information becomes less sufficient further into the future. The six-hour results should be described as moderate/limited predictive skill rather than high-accuracy long-range forecasting.

Table 4. Best model and persistence baseline at each forecast horizon; errors are in °C

Horizon	Best ML model	MAE	RMSE	R^2	Persistence MAE	Persistence RMSE	Persistence R^2
1	Random Forest	0.0578	0.0815	0.9855	0.0901	0.1030	0.9768
3	Random Forest	0.1974	0.2755	0.8341	0.2567	0.2923	0.8134
6	Linear Regression	0.3750	0.5124	0.4257	0.4530	0.5232	0.4011

.

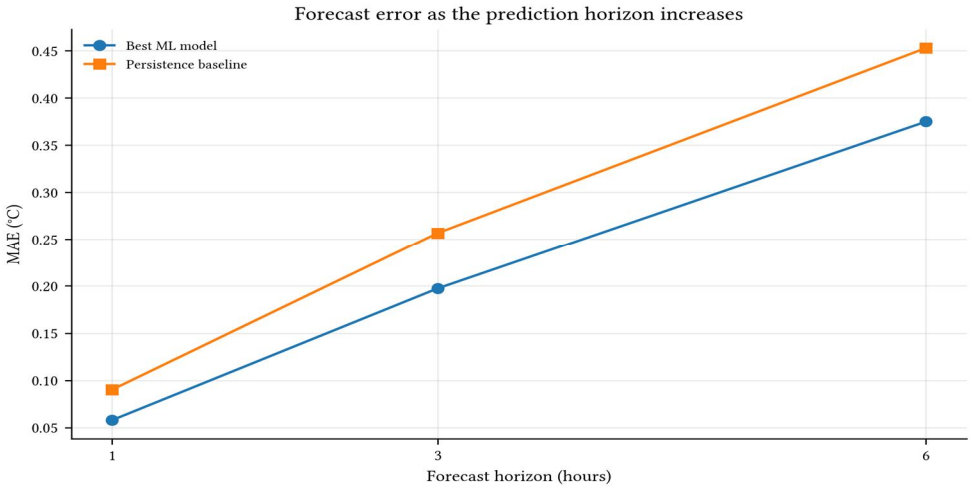


Figure 7. MAE of the best model and persistence baseline as forecast horizon increases.

E. Feature Importance and Engineering Interpretation

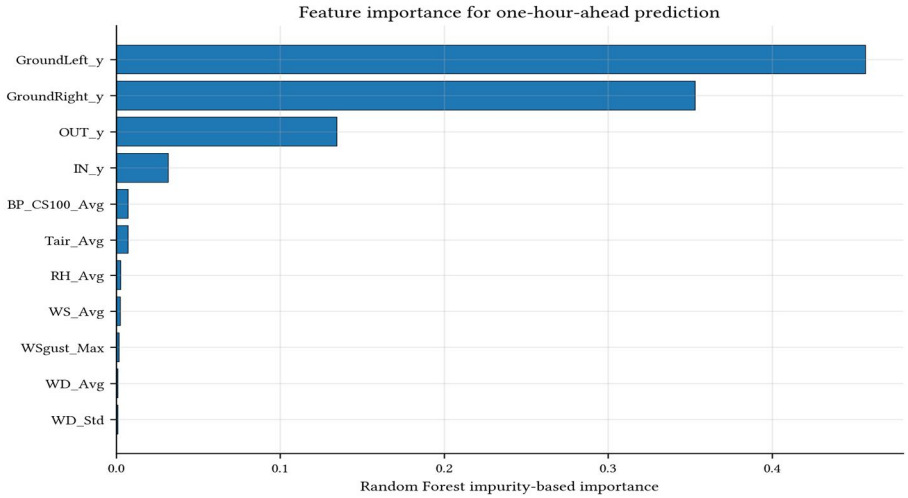


Figure 8. Random Forest impurity-based feature importance for one-hour-ahead prediction.

The feature-importance analysis is dominated by current outlet temperature. This is plausible for a one-hour forecast because temperatures typically change gradually. Other inlet, ground, and ambient measurements may provide additional context about thermal conditions. However, impurity-based importance is not causal evidence and can favor correlated predictors. A feature-group ablation experiment would be needed to quantify the incremental value of each group.

V. CONCLUSION

This study compared Linear Regression, Decision Tree, Random Forest, and HistGradientBoosting for EAHE outlet-temperature forecasting at 1-, 3-, and 6-hour horizons, using a chronological 80:20 split and a persistence baseline. Random Forest provided the best one-hour and three-hour results, whereas Linear Regression was the strongest tested regressor at six hours. The increasing errors with lead time indicate that forecast difficulty depends on the prediction horizon. The contribution is a reproducible, dataset-specific benchmark of established models rather than a novel algorithm. The results should be interpreted within the limitations of a single analysis table, missing solar radiation data, and a single hold-out period. Future work should verify raw-data lineage, use rolling-origin and independent-site validation, improve feature coverage, and assess uncertainty. Any claims about energy use, thermal comfort, or operational benefit require separate measurement or validated system-level analysis.

REFERENCES

[1] Bisoniya, T. S., Kumar, A., & Baredar, P. (2014). Study on calculation models of earth-air heat exchanger systems. Journal of Energy, 2014, Article 859286. https://doi.org/10.1155/2014/859286

[2] Ahmed, S. F., Liu, G., Mofijur, M., Azad, A. K., Hazrat, M. A., & Chu, Y.-M. (2021). Physical and hybrid modelling techniques for earth-air heat exchangers in reducing building energy consumption: Performance, applications, progress, and challenges. Solar Energy, 216, 274–294. https://doi.org/10.1016/j.solener.2021.01.022

[3] Hasan, M. I., Noori, S. W., & Shkarah, A. J. (2019). Parametric study on the performance of the earth-to-air heat exchanger for cooling and heating applications. Heat Transfer—Asian Research, 48(5), 1805–1829. https://doi.org/10.1002/htj.21458

[4] Ahmad, S. N., & Prakash, O. (2023). Three-dimensional simulation of earth air heat exchanger for cooling application and its validation using an experimental test rig. Heat and Mass Transfer, 59(7), 1277–1292. https://doi.org/10.1007/s00231-022-03334-8

[5] Shukla, A., Tiwari, G. N., & Sodha, M. S. (2006). Parametric and experimental study on thermal performance of an earth-air heat exchanger. International Journal of Energy Research, 30(6), 365–379. https://doi.org/10.1002/er.1154

[6] Alidrissi, Y., Wakil, M., Najib, M., Kaitouni, S. I., Mghazli, M. O., & Bakhouya, M. (2024). EAHE-D: A dataset for modeling and performance evaluation of earth to air heat exchangers. Data in Brief, 57, 111110. https://doi.org/10.1016/j.dib.2024.111110

[7] Kifouche, A., Kaddour, A., Lalmi, D., Chenini, N., Alkhafaji, M. A., Chambashi, G., Kaid, N., & Menni, Y. (2024). Enhancing real-time earth–air heat exchanger outlet temperature forecasting in arid climates using artificial neural network: A case study from Bechar, Algeria. International Journal of Low-Carbon Technologies, 19, 2493–2501. https://doi.org/10.1093/ijlct/ctae206

[8] Yue, Y., Yan, Z., Ni, P., Lei, F., & Yao, S. (2024). Machine learning-based multi-performance prediction and analysis of Earth-Air Heat Exchanger. Renewable Energy, 227, 120550. https://doi.org/10.1016/j.renene.2024.120550

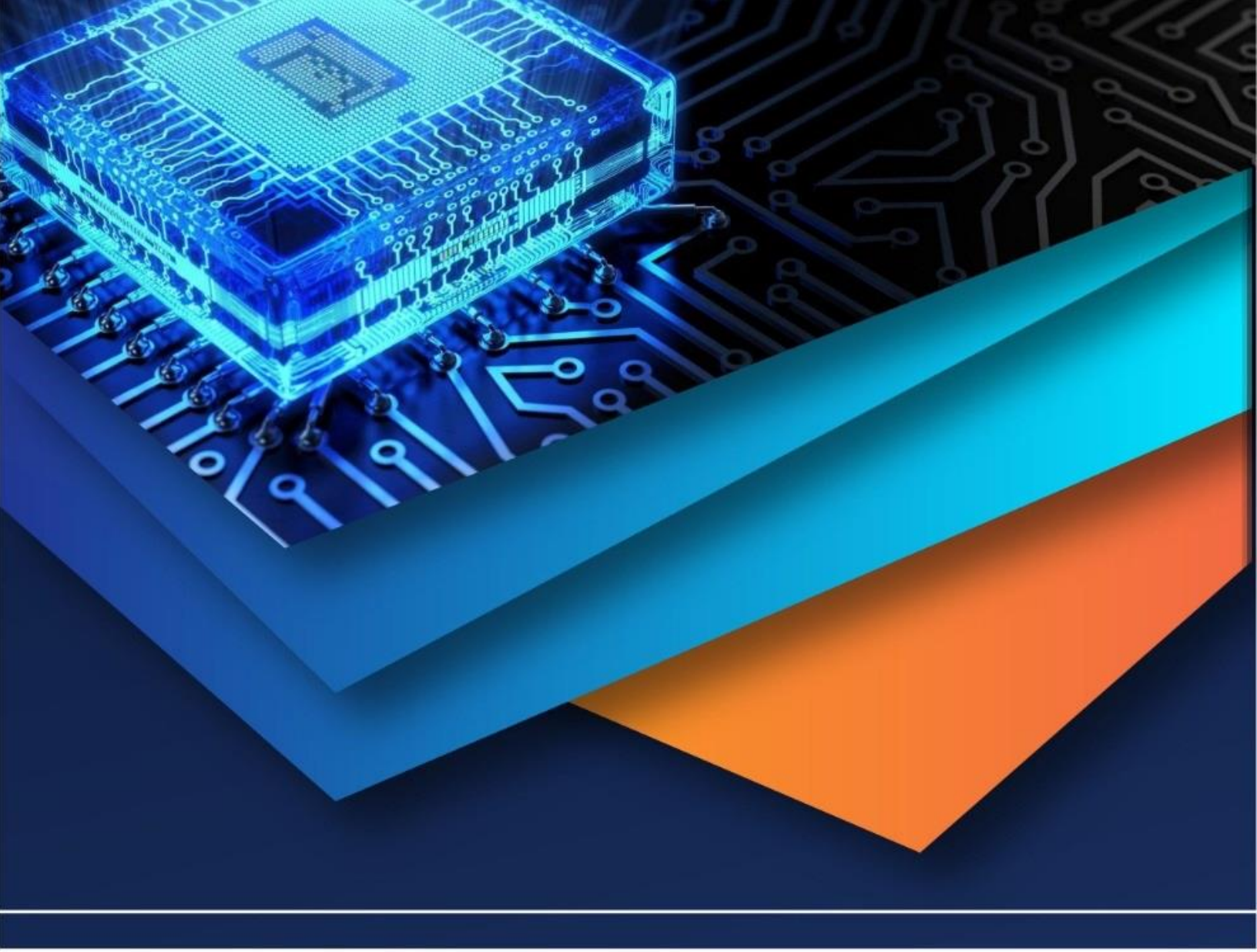
[9] Yue, Y., Yan, Z., Ni, P., Lei, F., & Zhang, D. (2025). Meteorological effects on the thermal performance of earth–air heat exchanger: Insights from interpretable machine learning. Energy and Buildings. https://doi.org/10.1016/j.enbuild.2025.116304

[10] Qi, X., Yang, D., Guo, X., Chen, F., An, F., & Wei, H. (2024). Theoretical modelling and experimental evaluation of thermal performance of a combined earth-to-air heat exchanger and return air hybrid system. Renewable Energy. https://doi.org/10.1016/j.renene.2024.121418

[11] Ahmad, S. N., Prakash, O., & Saini, P. (2025). Performance investigation of earth air heat exchanger: A synergistic approach of 3D simulation and its experimental validation for heating mode. Environmental Progress & Sustainable Energy, 44(4), e14620. https://doi.org/10.1002/ep.14620

[12] EAHE-D dataset. (2024). EAHE-D: A dataset for modeling and performance evaluation of earth to air heat exchangers. Zenodo. https://doi.org/10.5281/zenodo.12820831

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)