Article

Not peer-reviewed version

An Ensemble Multi-Task Learning Model for Predictive Performance Evaluation of Air Handling Units in HVAC Systems

Tobi Micheal Alabi , Adedayo Johnson Ogungbile * , Favour David Agbajor *

Posted Date: 15 June 2026

doi: 10.20944/preprints202606.1155.v1

Keywords: HVAC; deep learning; multi-task learning; ensemble model; machine

Article

An Ensemble Multi-Task Learning Model for Predictive Performance Evaluation of Air Handling Units in HVAC Systems

Tobi Micheal Alabi [1], Adedayo Johnson Ogungbile [2*] and Favour David Agbajor [3*]

[1] Midas Analytics, Tat Chee Avenue, Kowloon, Hong Kong SAR, China
[2] School of Science and Technology, Hong Kong Metropolitan University, Hong Kong SAR, China
[3] School of Architecture and Built Environment, Deakin University, Geelong, VIC 3220, Australia
* Correspondence: jogungbi@hkmu.edu.hk (A.J.O.); f.agbajor@deakin.edu.au (F.D.A.)

Abstract

With urbanization resulting in increased demand for indoor comfort, HVAC (heating, ventilation, and air-conditioning) systems, particularly air handling units (AHUs), are essentials for indoor climate control. The advent of big data and artificial intelligence (AI) have opened new avenues for enhanced safety and reliability in HVAC operations. Hence, this study focused on the predictive performance evaluation of AHUs, which is receiving less attention compared to its fault detection and optimal control issues. Utilizing real-time operational data from Oak National Laboratory, the proposed model employs multi-task learning (MTL) to refine prediction accuracy for AHU return air properties, including temperature, moisture content, and power consumption. This is achieved without allowing any single task to dominate others during the training phase. Moreover, the model introduces an ensemble approach that synergizes the capabilities of the different MTL algorithms using a boosting technique via gradient boosting regression tree (GBRT). This novel strategy has demonstrated superiority over conventional data-driven approaches in terms of performance. The paper culminates by showcasing the significant role of the proposed model as a metric for AHU performance evaluation and its contribution to smart decision-making in a real-world context. In essence, the developed model is poised to facilitate optimal decision-making regarding HVAC components and foster proactive strategies to ensure consistent operation and extend the lifespan of HVAC systems.

Keywords: HVAC; deep learning; multi-task learning; ensemble model; machine

---

1. Introduction

In recent decades, urbanization has led to a substantial majority of the global populace residing in urban areas, with individuals dedicating over 90% of their lifetime within built environments [1]. Consequently, there is an intensified demand for infrastructural amenities capable of delivering optimal comfort and functionality to support the daily activities of inhabitants. On a global scale, space heating and cooling accounts for 40% of building total energy consumption [2,3], and the heating, ventilation, and air conditioning (HVAC) systems serve as the primary equipment for achieving this purpose. Their primary function is to regulate the indoor climate conditions to ensure occupant thermal comfort by supplying controlled air with specific properties (temperature, humidity content, clean air, and flow rate) into the controlled space.

Most commercial HVAC systems comprise of chilled water or hot water production side (chiller plant and associated components or boiler) and the air handling units (AHUs) components. The AHU is responsible for regulating the supply air properties using the interrelationship properties between the outdoor air, supplied chilled water or boiler temperature, and the expected thermal requirements to achieve thermal comfort. This is achieved by controlling the damper positioning to control the

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 15 June 2026 doi:10.20944/preprints202606.1155.v1

amount of outdoor and return air that constitutes the fresh air supply, the supply air humidity control via humidifier or dehumidifier, and airflow rate via supply and return air fans. The system also receives the return air from the conditioned space via the extraction duct and the return fan [4], if the return air is not stale or contaminated, a specific percentage of it is mixed with the outdoor air in the AHU duct for fresh air supply. AHU return air properties; air temperature and relative humidity provide valuable insight in evaluating the system's thermal performance and efficiency. While chiller plants are identified as the largest HVAC component energy consumption, AHU constitutes about 25% of the total HVAC energy [5,6]. Meanwhile, an efficient and optimal AHU system has a ripple effect on chiller plant energy efficiency. To this end, maintaining stable operation of AHUs is crucial for optimal efficiency and prolonging HVAC systems' lifespan.

In this current era of big data availability via advanced information communication technologies (ICT), improved sensors and actuators, and advanced data analytics, the adoption of artificial intelligence (AI) to facilitate HVAC safety and reliability is inevitable. Numerous researches have been conducted on HVAC systems proposing data-driven approaches using machine learning/deep learning algorithms. The research span across HVAC fault detection and diagnosis [7,8], prediction [9], and optimal control [10].

With focus on AHU, Jee et al [5] applied a co-simulation approach for the real-time prediction and control of condenser water temperature and discharged air temperature of water-cooled AHU system. Several machine learning algorithms were compared for the prediction of hospital AHU short-term conditions in [11] with Prophet forecasting algorithm exhibiting superior performance. Yaddarabullah et al [12] developed an innovative method for predicting and adjusting AHUs induction motor frequency using modified Chen's fuzzy timeseries model. The potential energy savings associated with AHU economizer by using free-cooling effect was enhanced in [13] by integrating artificial neural network (ANN) with EnergyPlus simulation.

Bezyan et al. [14] proposed a hybrid machine learning model for the detection and diagnosis of multiple AHUs faults. The proposed method outperformed ANN, decision tree, and random forest. An explainable machine learning technique to determine the most contributing features to AHU fault detection was introduced in ref. [15]. Similarly, a semi-supervised transfer learning method was introduced in ref. [16] for AHU fault detection and diagnosis. Substantial contributions have also been made in using a data-driven approach to predict AHU parameters. Tagliabue et al. [17] applied ANN to predict the indoor air quality in educational facilities. The predicted metric is indoor CO2 concentration, and the forecasted value influences the tuned AHU control measure. An improved gated recurrent network (RNN) for HVAC system indoor air temperature based on multivariate transfer entropy in ref. [18]. Furthermore, Goopyo et al.[19] developed a data-driven model for AHU supply air temperature prediction using ANN.

Most research efforts in the aforementioned literature were channeled towards predictive and control of AHU, less attention is paid to the predictive performance of the system. Return air properties (such as return air temperature and relative humidity) provide insights in evaluating HVAC performance in meeting thermal comfort while maintaining optimal system efficiency. For instance, a return air temperature or relative humidity below or above an expected range can indicate overcooling, undercooling, or faulty components. While extensive research has been conducted on HVAC fault diagnosis, this approach provided no insight into system efficiency or performance. Further, optimal control using data-driven approaches like MPC (model predictive control), fuzzy logic and deep reinforcement learning algorithms do not provide information about HVAC health or impact of the control variables in real-time. Therefore, robust research efforts aimed at the prompt, adaptive and accurate predictive performance analysis of AHUs is very imperative to address these limitations.

AHU return air property is a multi-output task, adopting the conventional machine learning methods that focus only on single-task learning may result in insufficient training and under-fitting. Although single task learning has been widely adopted in the literature for multi-output tasks. However, for more complex problems, there exists an abundant correlation information between each task that single task learning capability may not fully captured. Caruana proposed a multi-task learning (MTL) mechanism to address this challenge [20]. The approach involves a mechanism improving the learning capability of individual tasks via information sharing among multiple related tasks. The research results indicate that the multi-task approach improves the accuracy of each task.

In addition, the combination of several prediction models (also known as ensemble approach) to achieve higher prediction accuracy has been explored by scholars in the energy management field [21]. The combination of extreme learning machine and artificial neural network was proposed by Zhang et al [22] for power load prediction. Ref. [23] generated multiple results using the same prediction model, applied clustering and group, and obtained the final power load prediction by weighting the clustered results. Similarly, Guoyin applied empirical mode decomposition and DBN for cooling load prediction [24]. While MTL approach and ensemble technique have proven to be a valuable method, only a few studies have considered the integration of the two methods. Xuan et al [25] proposed a deep multi-task learning and ensemble approach for multi-energy load predictions with minimal prediction error. Moreover, the two methods have been rarely adopted for AHU prediction model nor the adoption of their integration to tackle AHU prediction complexity.

Research Gaps and Novel Contribution

Without downplaying the painstaking research contributions of various scholars on predictive analytics on HVAC systems, there are still some gaps that demand intuitive exploration. Some of the noticeable research lapses are 1) most of the studies focus on predicting AHU decision or controllable actions such as supply air temperature, this approach is not feasible in a practical scenario, supply air properties into a conditioned space are a controllable action which is influence by various equipment parts; a breakdown or faulty operation of one of the components will greatly influence the supply air properties. 2) Most of the proposed predictive approaches consider building thermal properties as part of the input features, this become a challenge in extending the model to a global scale, as different buildings exhibit dynamic thermal properties, and 3) there is a rare focus on AHU predictive performance, an accurate prediction does not equivalent to optimal equipment performance.

In line with the gaps, our focus is shifted to the rare scenario in literature but the uttermost importance: An AHU predictive model that can be used for performance evaluation. Further, this study will be one of the first studies to apply an ensemble multi-task learning model for AHU prediction. Multi-task learning (MTL) model helps improve the training accuracy of each individual output (temperature, humidity, and power consumption) without one task dominating during training [26]. On the other hand, the ensemble approach is used to train a meta-model that integrates various MTL algorithms and generates a superior model with optimal performance [27]. The main novel contribution of the study is as follows:

1. Development of a predictive model for AHU that can be used to evaluate the system's performance based on the control actions by the building energy management system (BMS). The model can predict the return air temperature, return air relative humidity, and RTU (Roof Top Unit) power consumption. These predicted values serve as metrics to evaluate the system's performance, facilitate optimal decision making, identify components' faults, and encourage proactive measures.
2. Multi-task learning is adopted as the training algorithm for the multi-output prediction of the model. The approach addresses overfitting or underfitting of one of the predicted outputs compared to single-task learning.
3. A robust meta-model that synergizes the capability of various multi-task learning algorithms is proposed to improve model accuracy. We achieve this by considering the popular multi-task learning algorithms as the base models, followed by training a meta-model using an ensemble boosting technique.

The remaining section of the paper is structured as follows. Section 2 presents the working process of AHU as part of HVAC systems. The proposed methodology process, case study, ensemble and multi-task learning architecture are presented in Section 3. Section 4 analyzes the contributions of the proposed model and discusses the influence on lag period on the model performance. Section

5 presents practical scenarios to show the impact of the proposed model on AHU performance evaluation and smart decision-making. Section 6 draws the conclusion.

2. AHU Working Process

AHUs are integral components of HVAC systems responsible for regulating and circulating specific airflows to create and maintain optimal indoor air quality and thermal comfort. These units comprise various parts, including fans, heating and cooling , filters, dampers, and actuator valves [28]. Based on the functioning principle as illustrated in Figure 1, the units facilitate air circulation, filtration, heating, cooling, humidification, dehumidification, and energy recovery. The fan within the AHU draws in air through filters and propels it through the cooling/heating coil to the supply duct, facilitating temperature adjustment to the desired level. In other words, the air passing through the filter undergoes particle capture, enhancing air quality by trapping particles within the filter medium. In the heating mode, the AHU may incorporate a heating coil or heat exchanger, which utilizes hot water, steam, or electric resistance to warm the air.

The dampers installed within the AHU enable airflow control by regulating the opening and closing of specific sections, ensuring precise distribution and temperature control throughout different building areas. Typically, dampers regulate the quantity of recirculated air that is reused as fresh air supply and allow free flow of outdoor air into the conditioned space when economizer mode is activated [29]. During cooling mode, the cooling coil or refrigeration system cools the air by extracting heat and subsequently discharges it back into the atmosphere as exhaust air. Meanwhile, to ensure comfort conditions are met, it is imperative to analyze the energy content of the supply air prior to its dissemination into the conditioned space using AHU energy analysis. In real-world cases during summer months, the energy content of fresh air often surpasses that of the supply air, necessitating the AHU to mitigate the surplus energy via the cooling coil. Since the AHU's discharge air temperature (DAT) significantly influences HVAC energy consumption, some of this return air may be recirculated to conserve energy [5]. It is noteworthy that the return air temperature and moisture content are key AHU performance indicators, influencing energy efficiency, system balance, air quality, equipment performance, and occupant comfort. During operation, monitoring these parameters allows for proactive maintenance, optimization, and energy savings within the HVAC system.

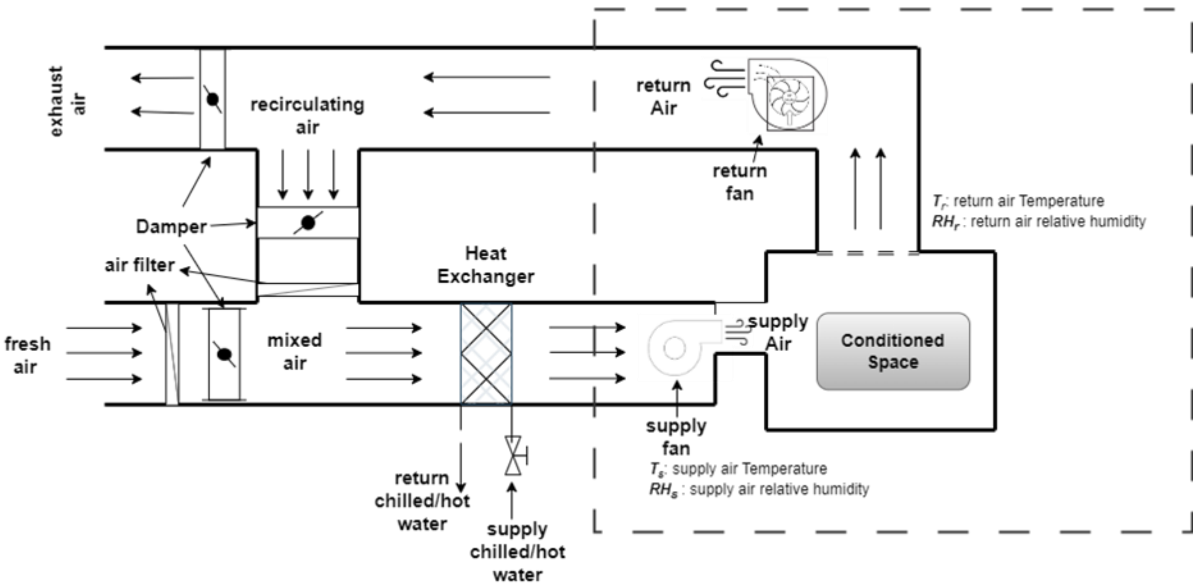


Figure 1. Air handling Unit (AHU).

3. Proposed Model

In this section, the proposed methodology flow adopted in achieving the conceived contributions of this study is presented. Figure 2. illustrates the methodology interactions. The first step is data retrieval followed by cleaning and data preprocessing for the model training. The AHU operational data is publicly available for research purposes is adopted in this study, thanks to Oak Ridge Laboratory research team efforts [30]. Other components of the methodology will be discussed in the proceeding sections.

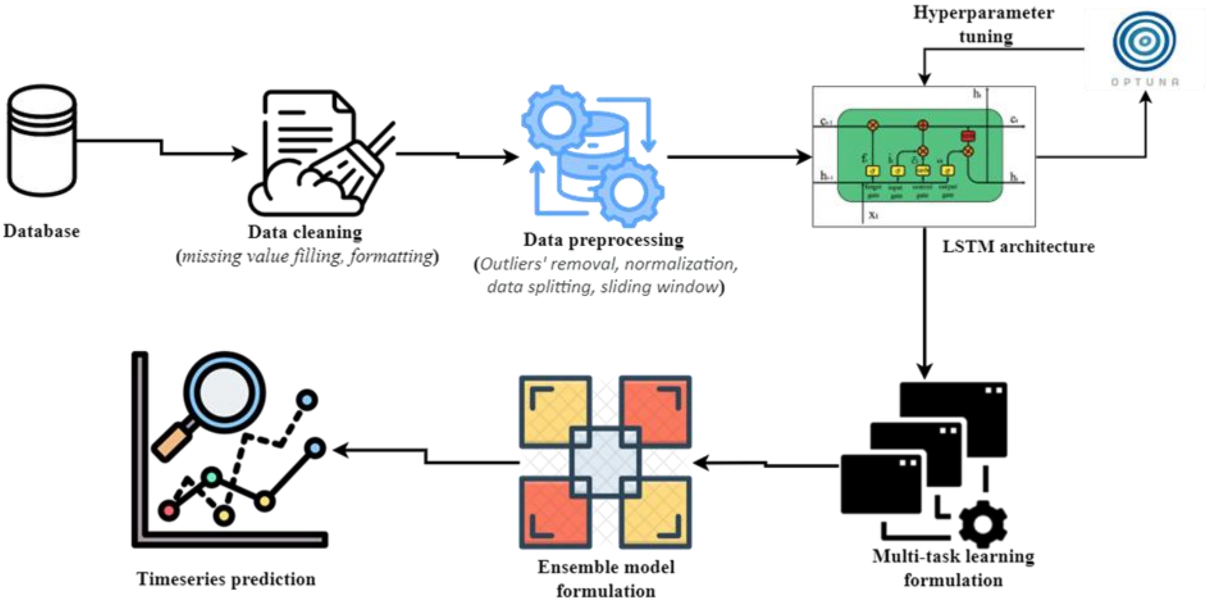


Figure 2. Proposed methodology flow.

3.1. Case Study

In this study, the real-time HVAC RTU (roof top units) operational data provided by Oak Ridge National Laboratory is adopted as the experimental data for the proposed model [31]. The operational data was collected from the building management system (BMS) that is responsible for the RTU sensor measurement storage, actuators command information, and the control platform. The facility for the experiment is a two-story Flexible research platform building located within the Laboratory territory. The RTU is the main HVAC system providing required thermal comfort needs for the facility with a capacity of 44kW and 9.6 energy efficiency rating. It is installed with a direct expansion (DX) cooling coil and heating coil. There are ten (10) thermal zones in the building, and each zone is connected to a dedicated variable air volume (VAV) box to meet their specific needs. Figure 3 shows the sensor types and location within the building and the HVAC systems. More than 500 sensors were installed according to the report, and the operation data was collected by data logger (CR 3000) at 1 min resolution. Table 1 presents the collected data information. To meet the purpose of this study, only data during the Baseline test period in cooling session (28/07/2021-04/08/2021) were extracted.

Figure 4 illustrates the supply air temperature and relative humidity by the RTU to meet the total cooling demand of the thermal zones. The supplied air properties by the RTU are further conditioned by each zone VAV using re-heat coil. The return air properties are captured by temperature and humidity sensor, and the measured value gives insight into the performance of the system or the influence of the supplied air in meeting the thermal comfort. Figure 5 illustrates RTU power usage which is a summation of the condenser, compressor, re-heat coil, and RTU fans (supply and return) energy consumption.

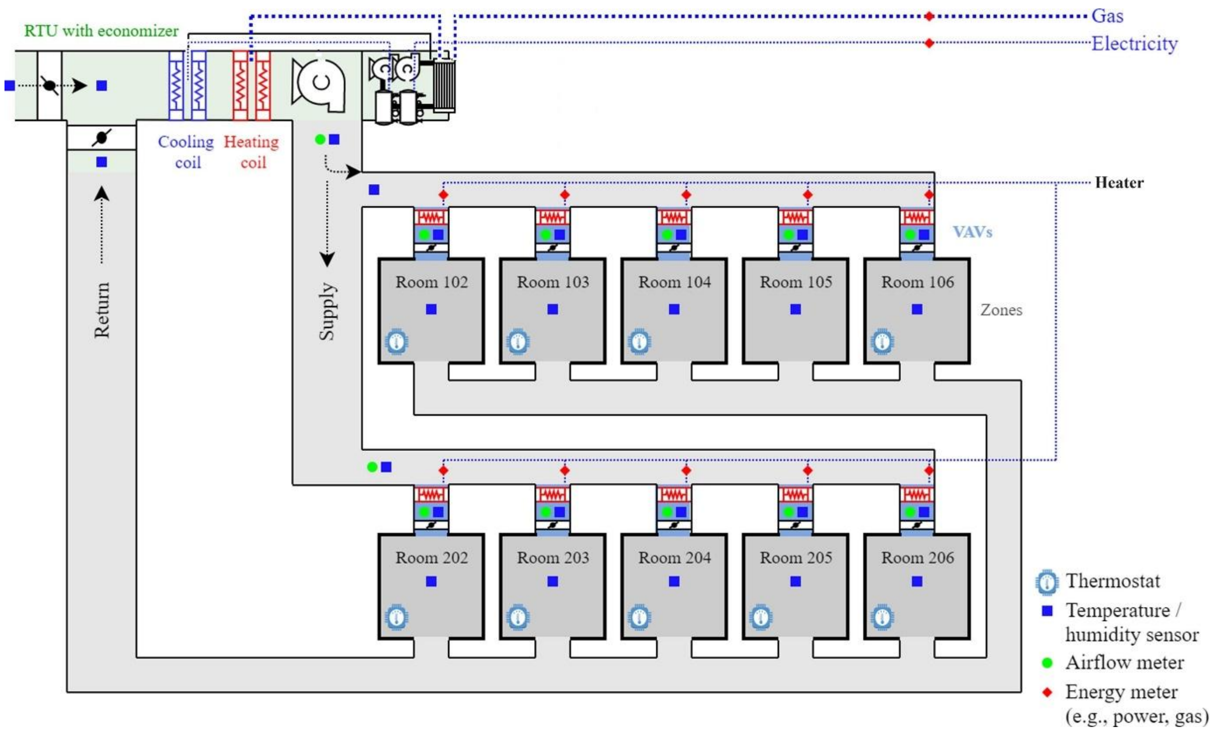
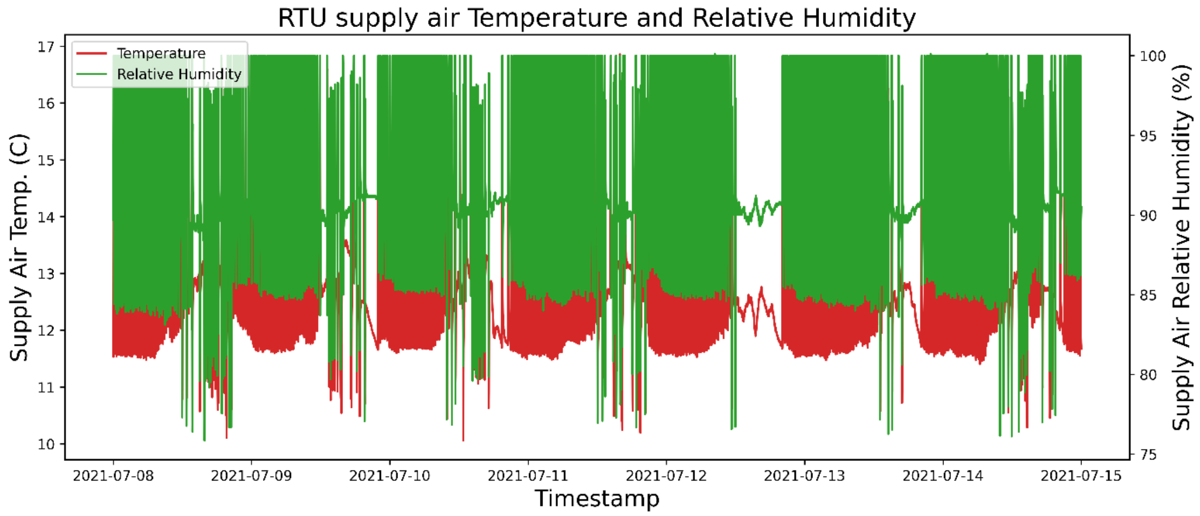


Figure 3. HVAC system operation diagram [31].

Table 1. Sensor data description.

Category	Data description	Unit
Indoor condition	Indoor air temperature	°C
	Indoor air relative humidity	%
Supply air	Supply air temperature (RTU)	°C
	Supply air temperature (VAV boxes)	°C
	Supply air relative humidity (RTU)	%
	Supply air relative humidity (VAV boxes)	%
Return air	Return air temperature	°C
	Return air relative humidity	%
Energy consumption	Compressor electric consumption	Wh
	Condenser electric energy consumption	Wh
	Re-heat coil electric energy consumption	Wh
	Fan electric energy consumption	Wh
Airflow rate	Airflow rate (RTU)	m^3/h
	Airflow rate (VAV boxes)	m^3/h



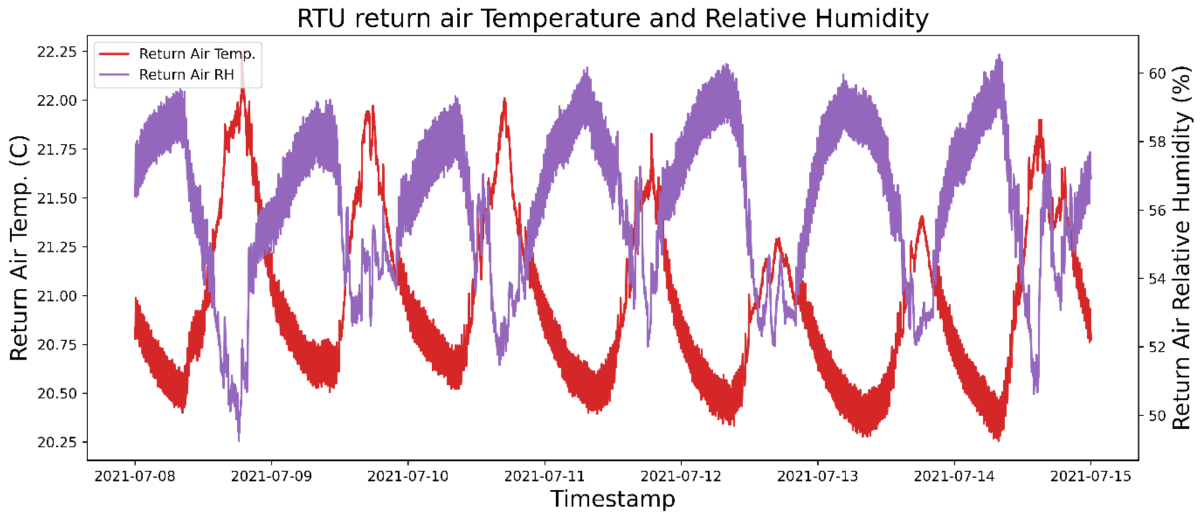


Figure 4. RTU supply and return air properties.

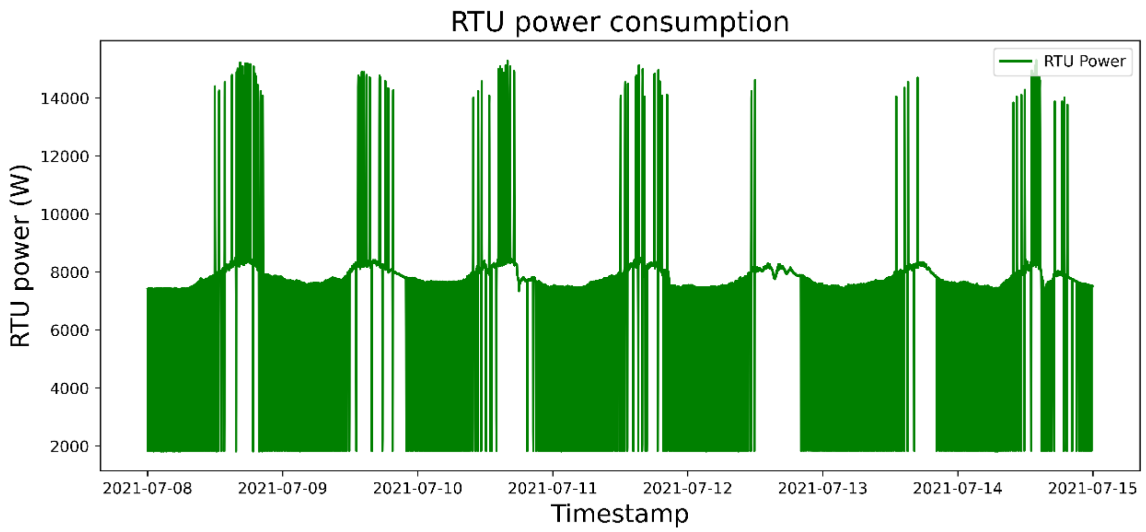


Figure 5. RTU total power consumption.

3.2. Data Preparation

The data preparation step encompasses data cleaning, preprocessing and feature extraction. In this study, the date column in the AHU operational data was formatted, followed by missing values filling using moving average method. Since our focus is on RTU operation, the VAV (variable air volume) measurements for each zone were excluded. However, we obtained the average indoor temperature and relative humidity of all the zones that are served by the RTU while each zones' data was excluded. After cleaning the data, the final features consideration is presented in Table 2. To avoid data leakage during model training, the obtained cleaned data was firstly split into training, validation, and testing using a proportion of 8:1:1. Outlier check was carried out on the training data, as shown in Figure 6, the RTU power consumption feature has some outliers while it is very minimal on other features. As a result, we adopted the interquartile range (IQR) to handle the outliers. Furthermore, the training data is normalized using a standardization approach to unify different magnitudes of the feature measurements and contribute to ease convergency of the model during training. Notably, the obtained normalized properties of the training data (i.e., the standard deviation and the mean) were used to normalize both validation and the test data.

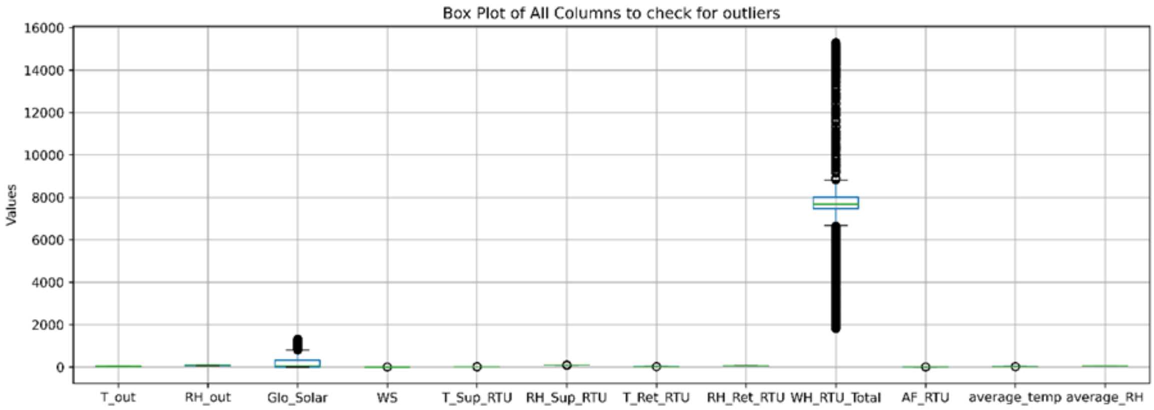


Figure 6. Outlier identification.

Table 2. Selected features.

Feature descriptions	Unit	Data resolution
Outdoor air Temperature	°C	1 min
Outdoor air relative humidity	%	1 min
Average zone indoor temperature	°C	1 min
Average zones indoor relative humidity	%	1 min
RTU supply air temperature	°C	1 min
RTU supply air relative humidity	%	1 min
RTU supply airflow rate	m^3/s	1 min
RTU Return air temperature	°C	1 min
RTU Return air relative humidity	%	1 min
RTU power consumption	Wh	1 min

Timeseries day-ahead predictions are influenced by the cumulative effects of previous timesteps magnitude. To handle this effect, the application of lag period or lookback period is a primary approach, a window size of 15 is with a stride of 1 is adopted in this study. Meanwhile, a sensitivity analysis on the effect of different window sizes will be conducted during experimentation.

3.3. GRU Model Architecture

LSTM is a special type of recurrent neural network that is famous for its temporal correlation and nonlinearity consideration when applied on sequential data [32]. Specifically, it consists of three gate structures that enable it to resolve vanish gradient challenge: input gate, forget gate, and output gate. Among LSTM variants, GRU is the most popular due to its reset gate feature where input and output gate are merged. This gate controls the presence or absence of previous hidden state information. As a result, GRU has only two gates (update gate and reset gate). In addition, GRU also has some other improvements that simply the network structure by mixing the cell state and the hidden state. This approach contributes to a reduction in training time and while ensuring prediction accuracy. Figure 7 illustrates GRU structure.

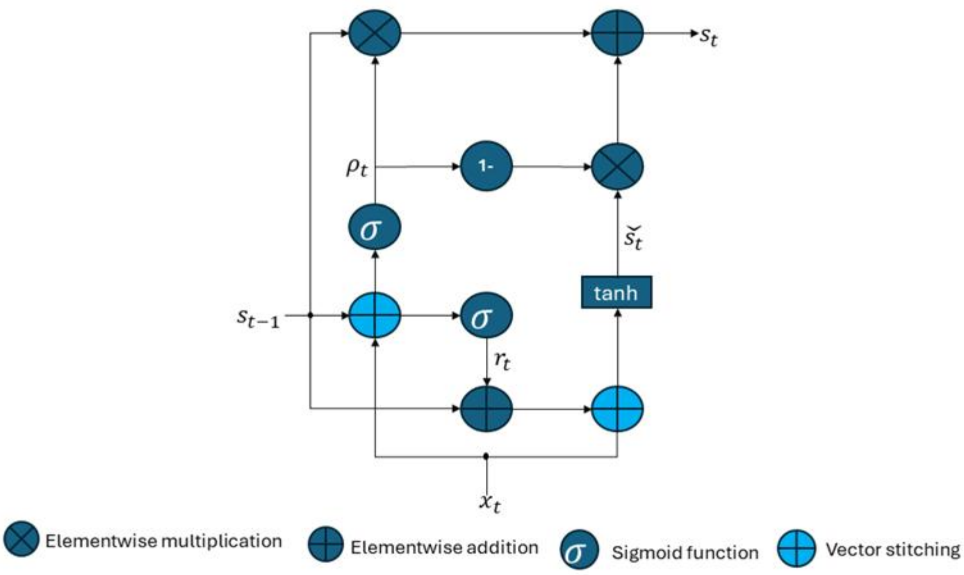


Figure 7. GRU structure.

Mathematically, GRU intuitive approach is described in eq (1) – eq (4) where W and U represent the associated each gate. At current time t, GRU update gate ρ_t receives the current input x_t and the previous time hidden state S_{t-1}, it uses the received information to determine whether the neurons will be activated, this is done by executing a matrix computation and the use of sigmoid function on the computation output. The reset gate r_t also receives the same information (x_t and S_{t-1}) to determine the quantity of historical information that will be forgotten as shown in eq (3). The candidate hidden state $\check{S}_t$ is computed by combining the output of r_t and input x_t and passing them into tanh activation function. In the end, the final output S_t is computed in eq (4) by using the update gate output ρ_t, $\check{S}_t$, and the previous hidden state S_{t-1}.

$$r_t = \sigma_{sig}(U_r S_{t-1} + W_r x_t) \quad (1)$$

$$\rho_t = \sigma_{sig}(U_z S_{t-1} + W_z x_t) \quad (2)$$

$$S_t = \varphi_{tanh}(U_h(r_s . S_{t-1}) + W_h x_t) \quad (3)$$

$$S_t = (1 - \rho_t) . S_{t-1} + \rho_t . \check{S}_t \quad (4)$$

3.4. Multi-Task Learning

Single-task learning has been widely applied on AHU prediction has discussed in Section 2.2. For instance, the prediction of AHU return air temperature and relative humidity can be segmented into two separate tasks and develop a model separately for each task. However, there is a correlation between the two predictions since they are both air properties from the same source. A single task learning approach will not capture this correlation, which results in limited information during individual task learning and weak generalization. Multi-task learning addresses these challenges. It also contributes to reducing computational costs by summing the training time of each task while

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 15 June 2026 doi:10.20944/preprints202606.1155.v1

improving the generalization ability of the prediction accuracy [25]. A typical MTL approach is a linear combination of all specific-task loss as shown in eq (5).

$$\mathcal{L}_{MTL}^{(t)} = \sum_{i=1}^{N_T} \omega_i \mathcal{L}_i(\varphi_s^{(t)}, \varphi_i^{(t)}) \quad (5)$$

where N_T is the number of different tasks, the shared layers parameters of the model are denoted by $\varphi_s^{(t)}$, $\varphi_i^{(t)}$ is the task i-specific layers at training step t, and ω_i represents the corresponding task-specific weights. Finally, the model parameters (shared and task-specific) are updated based on a mini-batch sample of the data, using gradient descent algorithms:

$$\varphi^{t+1} = \varphi^t - \eta \nabla_{\varphi(t)} \mathcal{L}_{MTL}^{(t)} \quad (6)$$

where η is the learning rate.

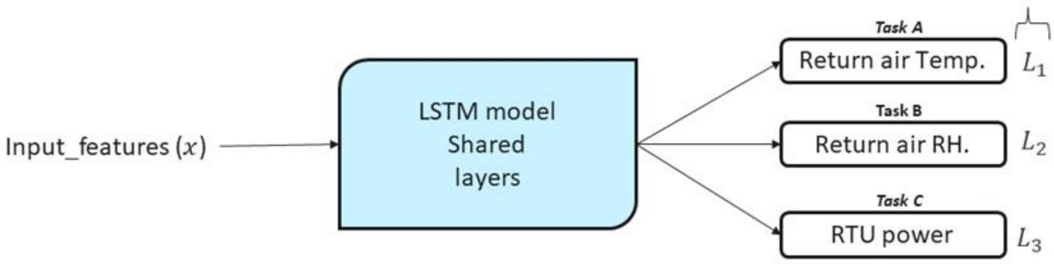


Figure 8. Multi-task prediction from shared layer.

The distinct feature of MTL is that different tasks involve different losses, the simply approach of adding up all losses during the training can lead to domineering effect of one of the tasks while other tasks may underfit or have no or minimal effect on the learning of the shared layers. To mitigate this challenge, some researchers have proposed some innovative approach such as weighted approach [33], loss dynamics weighting during training [34], the use of geometric loss [35] to avoid manual fine tuning of the tasks' loss weights, and uncertainty weighting strategy [36]. The weighted approach involves manual allocation of weight to each task loss during training, it is expressed mathematically as shown in eq (5). The weight is allocated based on the perceived influence of each task on the overall model. On the other hand, Dynamic weighting strategy involved weight allocation and update at each epoch during the model training as illustrated in eq (7) – eq (8), each task allocated loss are updated by a certain ratio based on the increase in loss at each epoch. Geometric loss function is another innovative approach; the geometric mean of all the task losses is taken as the loss function for shared layers weights parameters update during the training. The approach alleviates manual tunning of the task weight as shown in eq (9). Uncertainty weighting, which is also referred to as homoscedastic uncertainty is another approach to weigh losses in multi-task learning. It involves adding noise parameters to each task loss function involved in multi-task learning in conjunction with task weight, the weight is a learnable parameter, and it is updated using gradient information during model training. The approach is mathematically represented in eq (10):

$$\mathcal{L}_{total}^{(t)} = \omega_i^{(t)} \mathcal{L}_k^{(t)} + \omega_k^{(t)} \mathcal{L}_k^{(t)} \quad (7)$$

$$\omega_i^{(t)} = \check{\mathcal{L}}_k^{(t-1)} \Big/ \check{\mathcal{L}}_i^{(t-1)}, \qquad \omega_k^{(t)} = 1.0 \quad (8)$$

where $\check{\mathcal{L}}_T^{(t-1)}$ denotes task i average loss over the previous epoch.

Geometric Loss:

 doi:10.20944/preprints202606.1155.v1

$$\mathcal{L}_{total} = \left(\prod_{\tau=1}^{T} \mathcal{L}_\tau\right)^{\frac{1}{T}} \tag{9}$$

Homoscedastic uncertainty weighting strategy:

$$\mathcal{L}(W, \sigma_i) = \sum_i \frac{1}{2\sigma_i^2} \mathcal{L}_i(W) + \log \sigma_i^2 \tag{10}$$

where $\mathcal{L}_i(W)$ is denotes i-th task loss, either regression or classification loss. The coefficient of the first term in the equation and the second term is obtained by maximizing the log likelihood of the model with respect to parameters W and observation noise parameter σ. More details about the mathematical derivations of the log likelihood are presented in [36].

3.5. Meta-Learning Model via Ensemble Approach

In numerous cases, the idea of ensemble approach which involves integrating the output of several trained single models has been proving to have better performance to a certain degree. Currently, different algorithms have been introduced for the effectiveness of this approach, the famous ones are bagging and boosting [27]. Gradient boosting regressor tree (GBRT) is one of the famous ensemble frameworks that uses boosting technology [37]. It has some advantages such as high operation efficiency, generalization ability, and ability to prevent being trapped into local minimum.

In this study, we propose a meta-learning model that inherits the strength of each multi-task learning weighting strategy. The meta-learning model is based on GBRT as illustrated in Figure 9. Specifically, four (4) multi-task learning will be trained using the same architecture but different weighting strategies. Furthermore, the weight importance of each model will be obtained using the validation data, both the predicted output by each model on the validation data and their associated weight importance will serve as input into the meta-model for training.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 15 June 2026 doi:10.20944/preprints202606.1155.v1

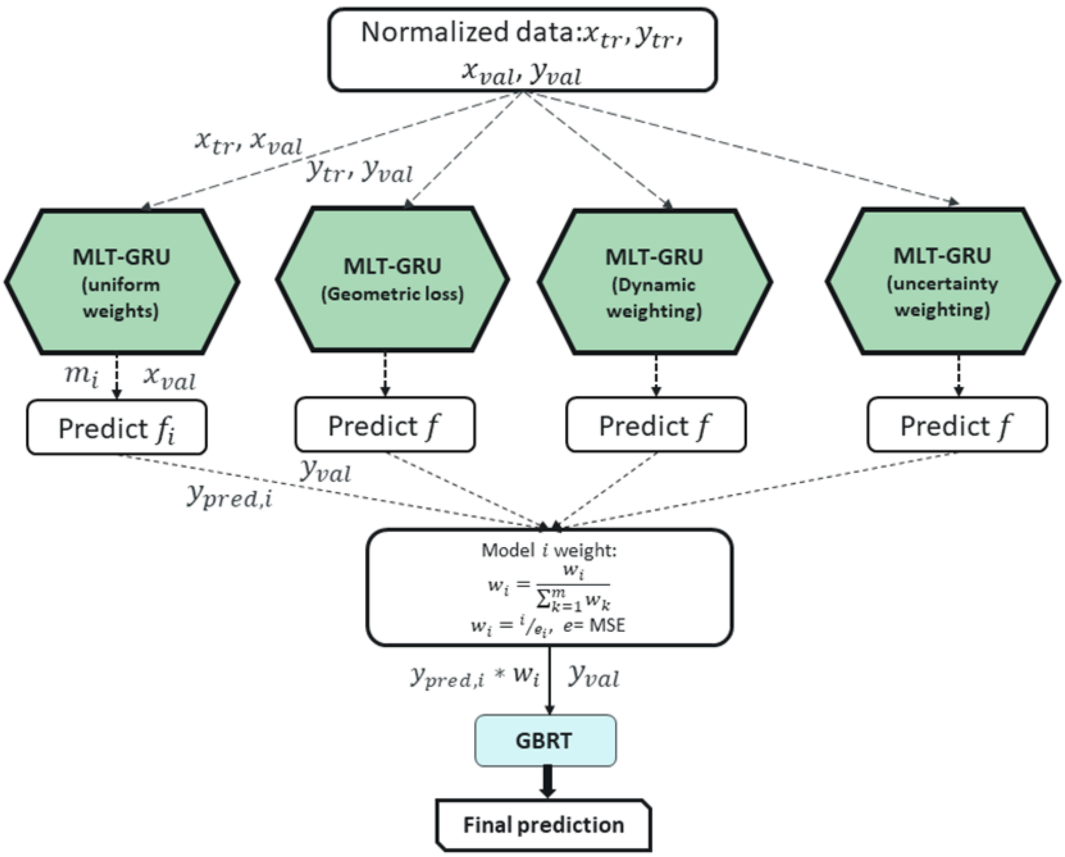


Figure 9. Meta-learning model.

3.6. Overall Training Process

The training process in this study is in two stages. In the first stage, individual MTL models were independently trained, for their shared and task-specific layers to learn optimal parameters value that will achieve minimal loss between actual and predicted values. Since all the tasks are regression problems, we adopted mean squared error (MSE) as the training loss and selected Adam optimizer for the model's weight update via gradient descent algorithm. Hyperparameters selection has major influence on deep learning model performance and convergency, we eliminate trial and error approach by executing hyperparameter tuning on the model architecture and learning rate range using Optuna [38]. Meanwhile, we stick to a batch size of 64 to reduce hyperparameters selection search space and computational demand. Model overfitting is also alleviated during the model training, we achieve this by (1) using a callback function that terminate the training process when the model error is not reducing after 10 consecutive epochs on the validation data, (2) by introducing an intuitive function that check and compare the model validation loss at previous and current epoch and save the best model state with the list error, and (3) introducing learning rate scheduler that adjust the learning rate by 10% after 5 consecutive non-improvement.

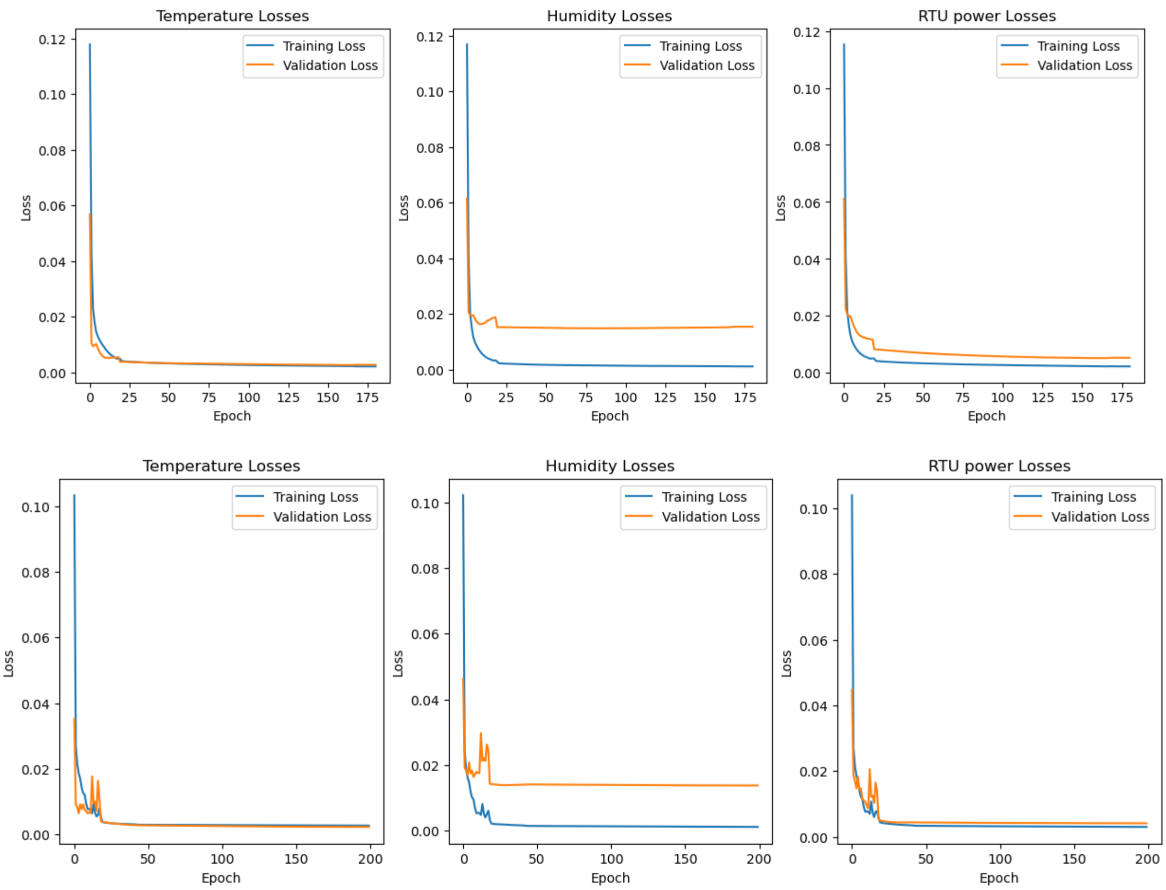
The second phase involves training of our meta-learning model i.e. GBRT via ensemble approach. After successful training of the base models, their corresponding predicted values on the validation data along with their weight importance forms the meta-features. The model's weight importance was computed based on its relative performance on the validation data in comparison to other base models. Finally, the base models' prediction on the validation data and their associated weight importance serves as input features for the meta model. The parameters for the GBRT are set as follows: 500 is chosen for the boosting stages, maximum depth of 3, 0.2 as the learning rate, and using least square regression as the loss function.

4. Experimental Results

In this section, we present the results obtained during various MTL algorithms experimentation. Since hyperparameters value play a major role in model performance, our first task is to select the best hyperparameters value combination. This is achieved by carrying out preliminary training for 10 epochs round on the training data and using the validation MSE as the objective function. After 50 successful trials, the obtained hyperparameters value includes: two (2) layers of GRU units with 64 hidden units and 0.00019 as the learning rate.

4.1. Multi-Task Learning Loss Curve

Figure 10 demonstrates the loss curves of each MTL model, the horizontal axis depicts the training epoch while the vertical axis represents the average loss value in the corresponding epoch. Generally, the average loss value drops significantly at the beginning of training for all the models, indicating that gradient descent is performed, and the chosen learning rate is suitable. After a certain number of epochs, the loss curve tends to be stable. However, it can be observed that loss curves for the return air temperature and RTU power prediction converge early with optimal performance for all the models. Meanwhile, the smoothness of the curves differs, which indicates the variation in learning capability of the MTL algorithms in terms of shared layer and each task layer learning.



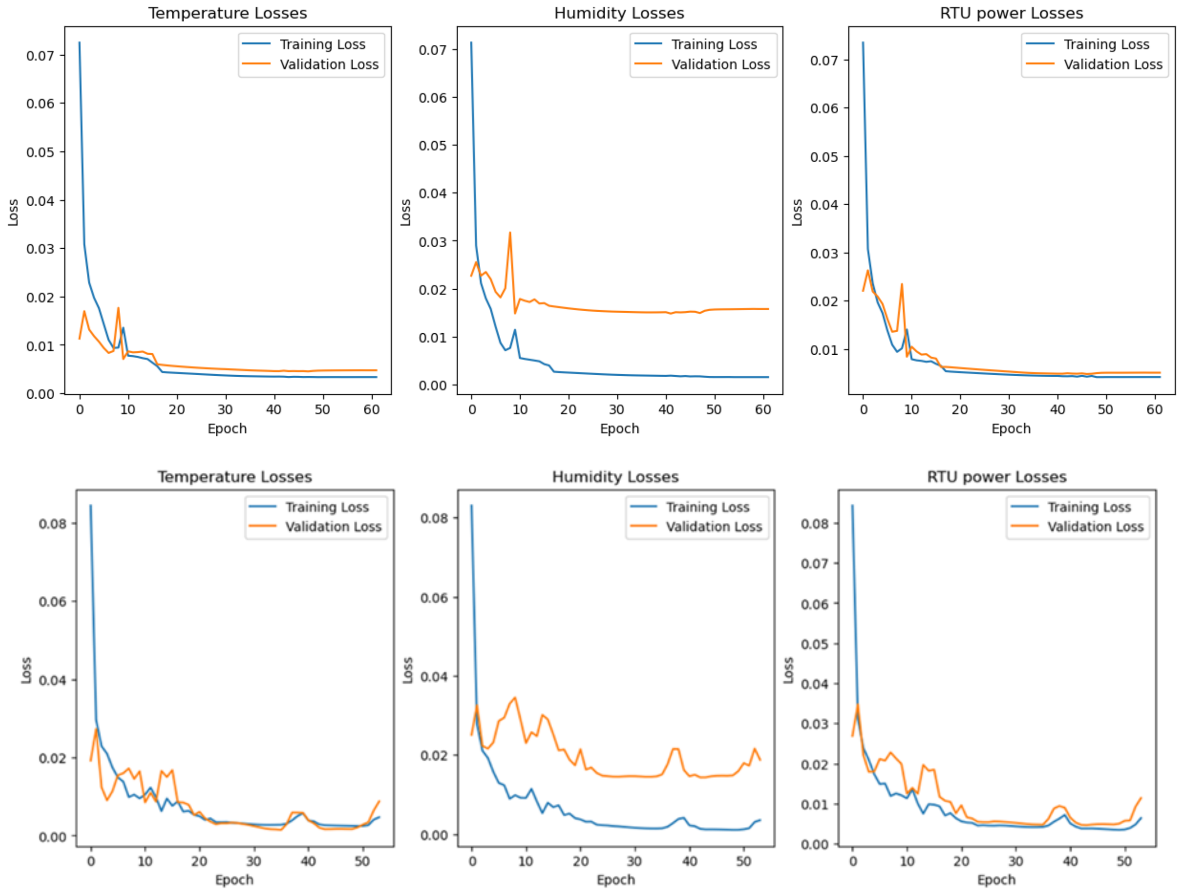


Figure 10. multi-task learning models loss curve (a) average weighing (b) Geometric loss (c) dynamic weighting (d) uncertainty weighting.

4.2. Model Performance Evaluation

The efficacy of machine learning in real-time cannot be strictly certified by the smoothness of the loss curve during training. In this study, we evaluate the performance of each individual model and the proposed ensemble approach on new data. The adopted evaluation metrics are mean-squared error (MSE), mean absolute error (MAE), and r2 score. Table 3 shows the evaluation performance of the models considered. While all the models exhibit good performance on each output prediction, MTL with average weighting achieves 0.98 r2 score on return air prediction, MTL with dynamic weighting obtains 0.96, while MTL with uncertainty weighting is the best performing on RTU power prediction with 0.88 r2-score. Figure 11 demonstrates the models MSE, the best performing model is expected to achieve the least MSE. The MSE value varies among model output and each MTL algorithm, this justifies the need for ensemble approach to complement individual models' weakness and aggregate their strength.

 doi:10.20944/preprints202606.1155.v1

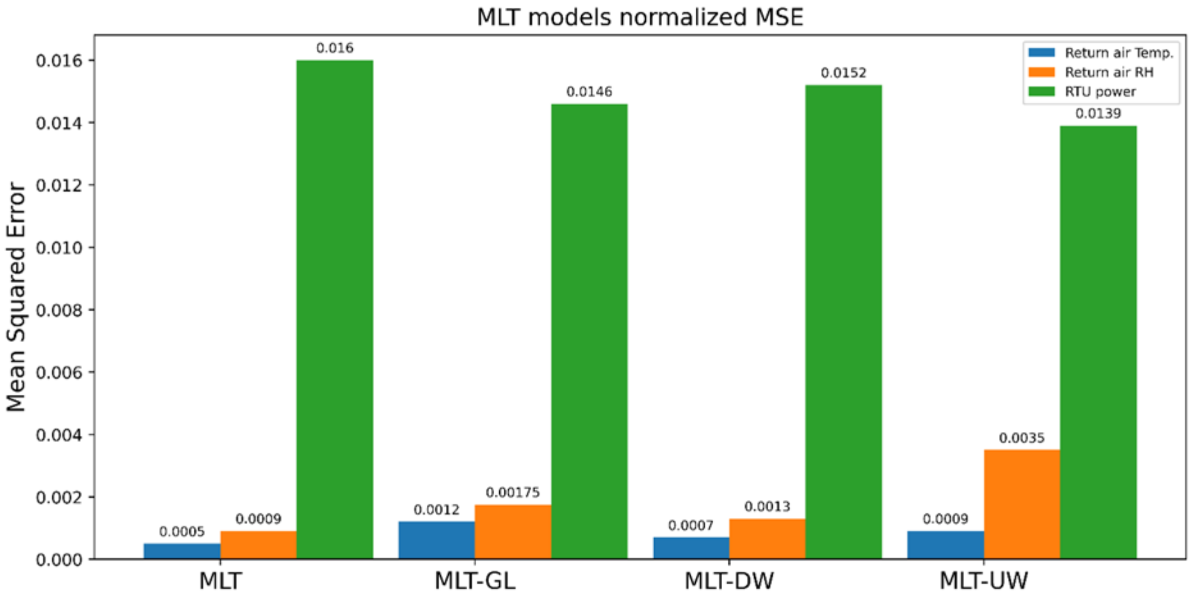


Figure 11. Prediction models mean square error evaluation.

Table 4 presents the evaluation metrics of various ensemble models configurations with different number of base models for training the meta-model. The best performing meta-model over each individual model in Table 11 is E-MTL1 with 2 base models (MTL average weighting and MTL with geometric loss function) with 0.98, 0.96, and 0.88 r2-score on return air temperature, return air relative humidity, and RTU power consumption, respectively. This shows the superiority of ensemble approach over single model. Other meta-models also exhibit superiority over a single model. However, a trade-off must be balance between model's accuracy and memory usage, as this will influence the model's latency and throughput performance at production stage.

Further steps are being taken to prove the supremacy of the proposed approach in real-time prediction. Since the evaluation metric values have shown that the proposed model is superior, there is no need to show other models time-series prediction. Figure 12 illustrates the capability of the proposed model in predicting day ahead RTU expected return air properties, all model performs exceptionally on return air temperature and relative humidity prediction. The most fascinating result is observed in Figure 12b RTU power prediction, the model is able to accurately handle the oscillating nature of RTU power consumption, this indicates the model optimally captured the interrelationship between RTU features in predicting the next timestep value. The performance also indicates that the proposed approach is applicable to both short and long-time predictions.

Table 3. Multi-task learning models' evaluation metrics on test data.

	MSE			**MAE**			**R2-score**		
	Return air Temp.	Return air RH	RTU power	Return air Temp.	Return air RH	RTU power	Return air Temp.	Return air RH	RTU power
MLT*	0.0022	0.1142	7842	0.0391	0.2718	50.0269	0.9873	0.9345	0.8670
MLT-GL	0.0048	0.2124	7040	0.0560	0.3725	40.7777	0.9726	0.9527	0.8804
MTL-DW	0.0027	0.1650	7293	0.0403	0.3427	46.8458	0.9843	0.9632	0.8761
MLT-UW	0.0037	0.4332	6702	0.0495	0.5552	45.6620	0.9791	0.9035	0.8861

MTL*: multi-task learning with average loss; GL: Geometric loss; DW: Dynamic weighting; UW: Uncertainty weighting.

Table 4. Ensemble multi-task learnig models evaluation metrics performance.

MSE	**MAE**	**R2-score**

2 model	Return air Temp.	Return air RH	RTU power	Return air Temp.	Return air RH	RTU power	Return air Temp.	Return air RH	RTU power
E-MTL1	0.0027	0.1650	7293	0.0403	0.3427	46.8458	0.9843	0.9632	0.8761
E-MTL2	0.0048	0.2124	7040	0.0560	0.3725	40.7777	0.9726	0.9527	0.8804
E-MTL3	0.0037	0.4332	6702	0.0496	0.5552	45.6620	0.9791	0.9035	0.8861
E-MTL4	0.0048	0.2124	7040	0.0560	0.3724	40.7777	0.9726	0.9527	0.8804
E-MTL5	0.0037	0.4332	6702	0.4958	0.5552	45.6620	0.9791	0.9035	0.8861
E-MTL6	0.0037	0.4332	6702	0.4958	0.5552	45.6620	0.9791	0.9035	0.8861
3 Models									
E-MTL-31	0.0048	0.2124	7040	0.0560	0.3724	40.7777	0.9726	0.9527	0.8804
E-MTL-32	0.0037	0.4332	6702	0.0496	0.5520	45.6610	0.9791	0.9036	0.8861
E-MTL-33	0.0037	0.4332	6702	0.0496	0.5520	45.6612	0.9791	0.9036	0.8861
4 Models									
E-MTL-4	0.0037	0.4332	6702	0.0495	0.5520	45.6556	0.9791	0.9140	0.8861

E-MTL1: no weight update + DW; E-MTL2: no weight update + GL; E-MTL3: no weight update + UW; E-MTL4: DW+GL; E-MTL5: GL+UW; E-MTL6: DW+UW; E-MTL31: no weight update + DW + GL; E-MTL32: no weight update + DW + UW; E-MTL33: DW+GL+UW; E-MTL4: no weight update + DW + GL + UW.

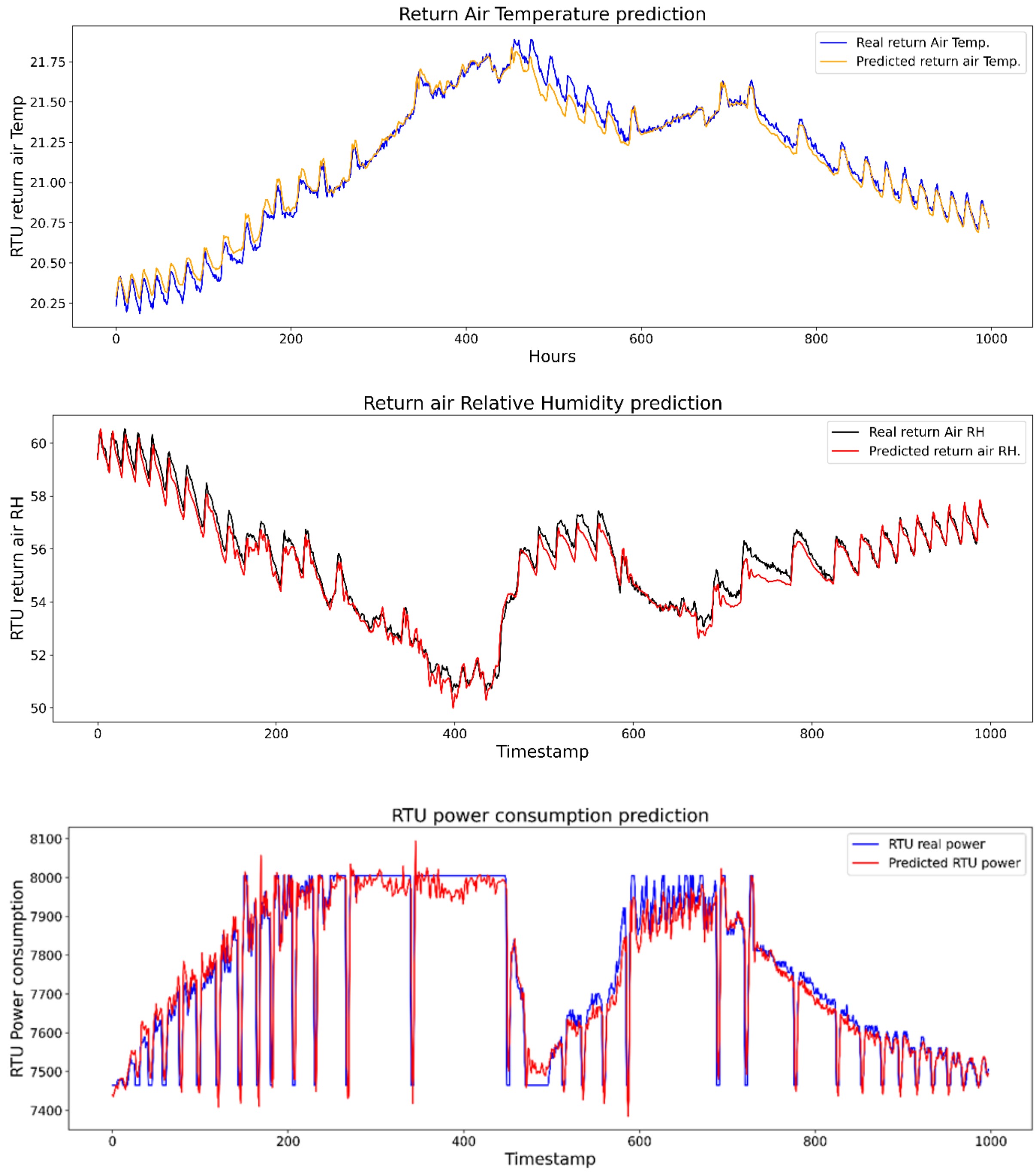


Figure 12. Proposed model predictive performance compared to true values.

4.3. Influence of Lag Periods on Model Performance

Lag period value is one of the primary factors that influences the performance of time-series prediction. In this study, a sensitivity analysis was conducted as shown in Figure 13 to evaluate the degree of this influence on the model's performance. All the individual models exhibit vary MAPE values in response to lag period value. MTL average weighting and dynamic weighting achieved high MAPE percentage at 5mins lag period, indicating that the models may overfit during training by capturing noise or experience short-term fluctuations. For a long short-term of 20mins, all the models also exhibit unfavorable MAPE values, except for MTL with Geometric loss. These values indicate that the models fail to capture relevant patterns due to a long lag period. A trade-off that balances responsiveness and stability is achieved by using a 10min lag period. Although, MTL-uncertainty weighting has high MAPE for return air RH prediction at 15mins lag period due to variation in model loss computation. However, the disparity in the models' performance is balanced by our proposed ensemble approach.

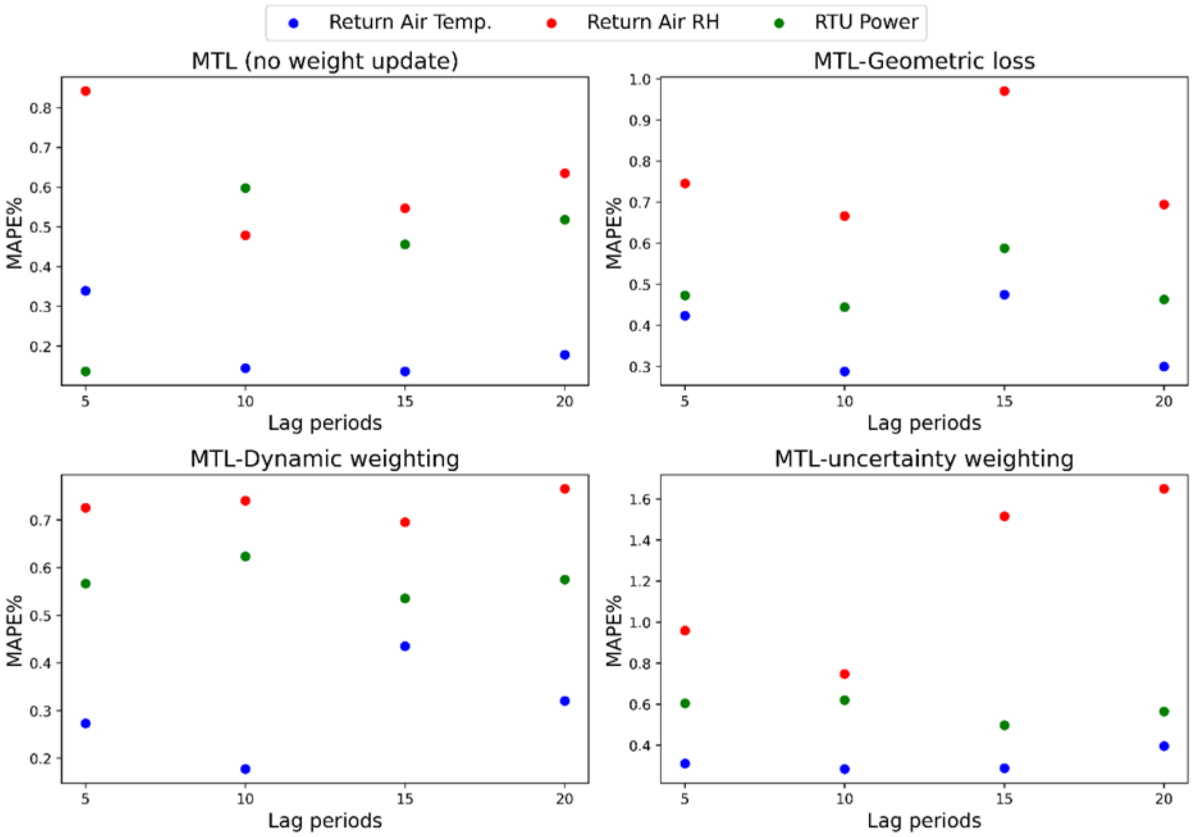


Figure 13. Lag period influence on model performance.

5. Impact of the Proposed Model on AHU Predictive Performance and Smart Decision Making

The integration of deep learning model into the operational framework of AHU signifies a transformative step in advancing the predictive performance of HVAC systems. We have developed an improved predictive model that was meticulously designed to forecast the key parameters that primarily determine the performance of AHU. These include the system's return air temperature, return air relative humidity, and RTU power consumption, leveraging the intricate interdependencies between environmental data, thermal zones sensor readings, and previous RTU states information to enhance the accuracy and reliability of the prediction.

By accurately predicting RTU return air properties, the model ensures the maintenance of optimal environmental conditions, offering a nuanced understanding of the systems' energy dynamics, and facilitates effective energy management strategies. For instance, it will empower the facility managers to make data-driven adjustments to the AHU operations, preemptively addressing inefficiencies and potential system failures. This proactive management not only enhances the longevity of the equipment but also aligns with the broader goals of sustainability by reducing unnecessary energy expenditure. In the following section, we present two (2) scenarios to describe the impact of the proposed model.

Scenario 1

In an office building where maintaining a consistent and comfortable indoor climate is crucial for productivity. The proposed ensemble multi-task learning model is integrated into the Building Management System (BMS). This model is adept at forecasting the next timestep's return air temperature, humidity, and power consumption with remarkable accuracy. For instance, at 8:00 AM, the model may predict a 2 °C increase in return air temperature, a 5% rise in return air relative humidity, and a 10% spike in power consumption by 9:00 AM. These forecasts, derived from current readings, historical data, and external factors like environmental data, enable the BMS to make informed decisions. Consequently, the BMS can proactively adjust supply air parameters, slightly lower the temperature and reducing humidity to maintain stable conditions, it can also optimize fan speeds to prevent energy wastage. This showcases the model's pivotal role in enhancing energy efficiency, indoor environmental quality, and optimal smart decision making.

Scenario 2

In a scenario where a large commercial building's AHU system is under constant operation, the ensemble multi-task learning model is deployed for monitoring the system's performance and energy efficiency. On a day with unexpectedly cool weather, the model detects a significant deviation in energy usage: a surge in AHU power consumption despite a lower than usual thermal load. The return air temperature is predicted to be significantly lower than the setpoint, indicating overcooling. Simultaneously, the model forecasts a high return air humidity level than expected, suggesting inadequate dehumidification.

The model's analysis, which integrates these forecasts with ongoing data and historical patterns, points to a likely stuck cooling valve, which causes excessive cooling regardless of the actual demand. The model also flags the possibility of a faulty sensor, which could mislead the system into understanding the air moisture content. With this insight, the facility management team will be alerted and swift into actions to resolve the issue. This proactive approach not only saves energy but also ensures the continued comfort and health of the building occupants. Moreover, it helps to safeguard the AHU system from further complications and costly downtime. Thus, the ensemble multi-task learning model proves its worth as an essential tool for energy efficiency and system reliability in HVAC management.

6. Conclusion

This research introduces a groundbreaking approach to the predictive analysis of AHU within HVAC systems, leveraging the power of ensemble multi-task learning (MTL) models. The utilization of real-time data from Oak National Laboratory has been instrumental in refining the model's predictive capabilities, ensuring a balanced learning process across various performance metrics. The proposed ensemble MTL model, enhanced by boosting technology via GBRT, has surpassed traditional predictive models, setting a new benchmark in the field. Its application has led to significant improvements in the accuracy of predicting return air properties, which is crucial for the efficient operation of AHUs.

The implications of this model extend beyond mere performance evaluation. It serves as a cornerstone for intelligent decision-making, providing actionable insights that facilitate the optimization of HVAC system components. This proactive approach to maintenance and operation is expected to enhance system reliability, reduce downtime, and extend the service life of AHUs. In essence, the model embodies the synergy between artificial intelligence and operational data analytics, offering a sophisticated tool for engineers and facility managers to achieve unprecedented levels of efficiency in HVAC management. The adoption of such models promises a future where indoor climate control is not only more responsive but also more sustainable.

Author Contributions: Conceptualization, T.M.A. and F.D.A.; Methodology, T.M.A., A.J.O. and F.D.A.; Software, T.M.A..; Validation, T.M.A.; Resources, A.J.O.; Data curation, T.M.A.; Writing—original draft preparation, T.M.A., A.J.O. and F.D.A.; Writing—review and editing, F.D.A.; Visualization, T.M.A. and F.D.A.; Project administration, A.J.O.; Funding acquisition, A.J.O. All authors have read and agreed to the published version of the manuscript.

Acknowledgments: The work presented in this article is supported by the Centre for Advances in Reliability and Safety (CAiRS), admitted under AIR@InnoHK Research Cluster and the Hong Kong Metropolitan University, Hong Kong SAR, China.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. M. Mannan and S. G. Al-Ghamdi, "Indoor air quality in buildings: A comprehensive review on the factors influencing air pollution in residential and commercial structure," *International Journal of Environmental Research and Public Health*, vol. 18, no. 6, p. 3276, 2021.
2. F. D. Agbajor, M. C. Mewomo, V. B. Umoh, and S. A. Makanjuola, "Building energy management system research in South Africa—A decade overview," *Energy Reports*, vol. 9, pp. 33-39, 2023/10/01/ 2023, doi: https://doi.org/10.1016/j.egyr.2023.05.056.
3. T. M. Alabi *et al.*, "A review on the integrated optimization techniques and machine learning approaches for modeling, prediction, and decision making on integrated energy systems," *Renewable Energy*, vol. 194, pp. 822-849, 2022/07/01/ 2022, doi: https://doi.org/10.1016/j.renene.2022.05.123.
4. M. Kumar Yadav, A. Verma, B. Ketan Panigrahi, and S. Mishra, "User comfort driven time-table linked AHU scheduling for ancillary service maximization of an educational building," *Energy and Buildings*, vol. 225, p. 110317, 2020/10/15/ 2020, doi: https://doi.org/10.1016/j.enbuild.2020.110317.
5. H. J. Jee, S. H. Yeon, J. Park, Y. Yoon, and K. H. Lee, "Machine learning based simultaneous control of air handling unit discharge air and condenser water temperatures set-point for minimized cooling energy in an office building," *Energy and Buildings*, vol. 297, p. 113471, 2023/10/15/ 2023, doi: https://doi.org/10.1016/j.enbuild.2023.113471.
6. T. M. Alabi, F. D. Agbajor, Z. Yang, L. Lu, and A. J. Ogungbile, "Strategic potential of multi-energy system towards carbon neutrality: A forward-looking overview," *Energy and Built Environment*, vol. 4, no. 6, pp. 689-708, 2023/12/01/ 2023, doi: https://doi.org/10.1016/j.enbenv.2022.06.007.
7. H. Zhu, W. Yang, S. Li, and A. Pang, "An Effective Fault Detection Method for HVAC Systems Using the LSTM-SVDD Algorithm," *Buildings*, vol. 12, no. 2, p. 246, 2022. [Online]. Available: https://www.mdpi.com/2075-5309/12/2/246.
8. B. Wu, W. Cai, F. Cheng, and H. Chen, "Simultaneous-fault diagnosis considering time series with a deep learning transformer architecture for air handling units," *Energy and Buildings*, vol. 257, p. 111608, 2022/02/15/ 2022, doi: https://doi.org/10.1016/j.enbuild.2021.111608.
9. H. Sha, P. Xu, C. Hu, Z. Li, Y. Chen, and Z. Chen, "A simplified HVAC energy prediction method based on degree-day," *Sustainable Cities and Society*, vol. 51, p. 101698, 2019/11/01/ 2019, doi: https://doi.org/10.1016/j.scs.2019.101698.
10. J. Hou, X. Li, H. Wan, Q. Sun, K. Dong, and G. Huang, "Real-time optimal control of HVAC systems: Model accuracy and optimization reward," *Journal of Building Engineering*, vol. 50, p. 104159, 2022/06/01/ 2022, doi: https://doi.org/10.1016/j.jobe.2022.104159.
11. R. Al-Aomar, M. AlTal, and J. Abel, "A data-driven predictive maintenance model for hospital HVAC system with machine learning," *Building Research & Information*, vol. 52, no. 1-2, pp. 207-224, 2024/02/17 2024, doi: 10.1080/09613218.2023.2206989.
12. Yaddarabullah;, Y. Alviando;, D. Lestari;, A. B. A. Rahman;, and A. Saad, "Optimizing Air Handling Unit Induction Motor Frequency Predictions: Evaluating and Advancing Forecasting Techniques with a Modified Chen's Fuzzy Time Series Model," *International Journal of Intelligent Engineering and Systems*, vol. 17, no. 2, pp. 595-608, 2024, doi: 10.22266/ijies2024.0430.48.
13. Y. Yoon, B. Seo, and S. Cho, "Potential Cooling Energy Savings of Economizer Control and Artificial-Neural-Network-Based Air-Handling Unit Discharge Air Temperature Control for Commercial Building," *Buildings*, vol. 13, no. 5, 2023, doi: 10.3390/buildings13051174.
14. B. Bezyan and R. Zmeureanu, "Application of a Hybrid Machine Learning Model for the Detection and Diagnosis of Multiple Dependent Faults in an Air Handling Unit," in *Proceedings of the 5th International*

Conference on Building Energy and Environment, Singapore, L. L. Wang *et al.*, Eds., 2023// 2023: Springer Nature Singapore, pp. 2303-2314.

15. J. Belikov *et al.*, "Explainable AI based Fault Detection and Diagnosis System for Air Handling Units," in *Proceedings of the International Conference on Informatics in Control, Automation and Robotics*, 2022, vol. 1, pp. 271-279, doi: 10.5220/0011350000003271. [Online]. Available: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85137929058&doi=10.5220%2f0011350000003271&partnerID=40&md5=aa71082eaf3536a9bdc364caf91731c5
16. V. Martinez-Viol, E. M. Urbano, J. E. Torres Rangel, M. Delgado-Prieto, and L. Romeral, "Semi-Supervised Transfer Learning Methodology for Fault Detection and Diagnosis in Air-Handling Units," *Applied Sciences*, vol. 12, no. 17, p. 8837, 2022. [Online]. Available: https://www.mdpi.com/2076-3417/12/17/8837.
17. L. C. Tagliabue, F. Re Cecconi, S. Rinaldi, and A. L. C. Ciribini, "Data driven indoor air quality prediction in educational facilities based on IoT network," *Energy and Buildings*, vol. 236, p. 110782, 2021/04/01/ 2021, doi: https://doi.org/10.1016/j.enbuild.2021.110782.
18. Z. Li, P. Wang, J. Zhang, and S. Mu, "A strategy of improving indoor air temperature prediction in HVAC system based on multivariate transfer entropy," *Building and Environment*, vol. 219, p. 109164, 2022/07/01/ 2022, doi: https://doi.org/10.1016/j.buildenv.2022.109164.
19. G. Hong and B. S. Kim, "Development of a Data-Driven Predictive Model of Supply Air Temperature in an Air-Handling Unit for Conserving Energy," *Energies*, vol. 11, no. 2, doi: 10.3390/en11020407.
20. R. Caruana, "Multitask Learning," *Machine Learning*, vol. 28, no. 1, pp. 41-75, 1997/07/01 1997, doi: 10.1023/A:1007379606734.
21. S. Buhan and I. Çadırcı, "Multistage Wind-Electric Power Forecast by Using a Combination of Advanced Statistical Methods," *IEEE Transactions on Industrial Informatics*, vol. 11, no. 5, pp. 1231-1242, 2015, doi: 10.1109/TII.2015.2431642.
22. R. Zhang, Z. Y. Dong, Y. Xu, K. Meng, and K. P. Wong, "Short-term load forecasting of Australian National Electricity Market by an ensemble model of extreme learning machine," *IET Generation, Transmission & Distribution*, vol. 7, no. 4, pp. 391-397, 2013, doi: https://doi.org/10.1049/iet-gtd.2012.0541.
23. Y. Wang, Q. Chen, M. Sun, C. Kang, and Q. Xia, "An Ensemble Forecasting Method for the Aggregated Load With Subprofiles," *IEEE Transactions on Smart Grid*, vol. 9, no. 4, pp. 3906-3908, 2018, doi: 10.1109/TSG.2018.2807985.
24. G. Fu, "Deep belief network based ensemble approach for cooling load forecasting of air-conditioning system," *Energy*, vol. 148, pp. 269-282, 2018/04/01/ 2018, doi: https://doi.org/10.1016/j.energy.2018.01.180.
25. W. Xuan, W. Shouxiang, Z. Qianyu, W. Shaomin, and F. Liwei, "A multi-energy load prediction model based on deep multi-task learning and ensemble approach for regional integrated energy systems," *International Journal of Electrical Power & Energy Systems*, vol. 126, p. 106583, 2021/03/01/ 2021, doi: https://doi.org/10.1016/j.ijepes.2020.106583.
26. M. Sugiyama, "Chapter 34 - Multitask Learning," in *Introduction to Statistical Machine Learning*, M. Sugiyama Ed. Boston: Morgan Kaufmann, 2016, pp. 391-404.
27. A. Mohammed and R. Kora, "A comprehensive review on ensemble deep learning: Opportunities and challenges," *Journal of King Saud University - Computer and Information Sciences*, vol. 35, no. 2, pp. 757-774, 2023/02/01/ 2023, doi: https://doi.org/10.1016/j.jksuci.2023.01.014.
28. A. Rosato, M. El Youssef, F. Guarino, A. Ciervo, and S. Sibilio, "Experimental studies of air-handling units' faulty operation for the development of data-driven fault detection and diagnosis tools: A systematic review," *Energy Reports*, vol. 8, pp. 494-503, 2022/11/01/ 2022, doi: https://doi.org/10.1016/j.egyr.2022.10.087.
29. J. E. Seem, J. M. House, G. E. Kelly, and C. J. Klaassen, "A Damper Control System for Preventing Reverse Airflow Through the Exhaust Air Damper of Variable-Air-Volume Air-Handling Units," *HVAC&R Research*, vol. 6, no. 2, pp. 135-148, 2000/04/01 2000, doi: 10.1080/10789669.2000.10391254.
30. O. R. N. Laboratory. https://www.ornl.gov/ (accessed.
31. Y. Yoon, S. Jung, P. Im, and A. Gehl, "Datasets of a Multizone Office Building under Different HVAC System Operation Scenarios," *Sci Data*, vol. 9, no. 1, p. 775, Dec 19 2022, doi: 10.1038/s41597-022-01858-6.

32. W. Wu, W. Liao, J. Miao, and G. Du, "Using Gated Recurrent Unit Network to Forecast Short-Term Load Considering Impact of Electricity Price," *Energy Procedia,* vol. 158, pp. 3369-3374, 2019/02/01/ 2019, doi: https://doi.org/10.1016/j.egypro.2019.01.950.
33. T. Gong *et al.*, "A Comparison of Loss Weighting Strategies for Multi task Learning in Deep Neural Networks," *IEEE Access,* vol. 7, pp. 141627-141632, 2019, doi: 10.1109/ACCESS.2019.2943604.
34. E. Grégoire, M. H. Chaudhary, and S. Verboven, "Sample-level weighting for multi-task learning with auxiliary tasks," *Applied Intelligence,* vol. 54, no. 4, pp. 3482-3501, 2024/02/01 2024, doi: 10.1007/s10489-024-05300-9.
35. S. Chennupati, G. Sistu, S. K. Yogamani, and S. A. Rawashdeh, "MultiNet++: Multi-Stream Feature Aggregation and Geometric Loss Strategy for Multi-Task Learning," *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW),* pp. 1200-1210, 2019.
36. A. Kendall, Y. Gal, and R. Cipolla, "Multi-task Learning Using Uncertainty to Weigh Losses for Scene Geometry and Semantics," *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition,* pp. 7482-7491, 2017.
37. X. Ma, C. Ding, S. Luan, Y. Wang, and Y. Wang, "Prioritizing Influential Factors for Freeway Incident Clearance Time Prediction Using the Gradient Boosting Decision Trees Method," *IEEE Transactions on Intelligent Transportation Systems,* vol. 18, pp. 2303-2310, 2017.
38. T. Akiba, S. Sano, T. Yanase, T. Ohta, and M. Koyama, "Optuna: A Next-generation Hyperparameter Optimization Framework," *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining,* 2019.