



Article A COP Prediction Model of Hybrid Geothermal Heat Pump Systems based on ANN and SVM with Hyper-Parameters Optimization

Jihyun Shin¹, Jinhyun Lee² and Younghum Cho^{3,*}

- ¹ Enertecunited, Busan 48059, Republic of Korea; sjh1010@ynu.ac.kr
- ² Institute of Industrial Technology, Yeungnam University, Gyeongsan 38541, Republic of Korea; jhl211@ynu.ac.kr
- ³ School of Architecture, Yeungnam University, Gyeongsan 38541, Republic of Korea
- * Correspondence: yhcho@ynu.ac.kr; Tel.: +82-53-810-3081

Abstract: When the geothermal heat pump system is operated due to an imbalance in the heating and cooling load, the system performance is lowered due to the occurrence of a thermal environment problem in the ground. To solve the performance degradation, a hybrid geothermal heat pump system with an added auxiliary heat source is used. For the efficient operation of the system, it is necessary to check the performance coefficient of the hybrid geothermal system. The coefficient of performance can be monitored based on a mathematical model using a measuring instrument. However, in the case of mathematical models, there are a lot of input data required, and many measurement sensors are required for this. If there is an input factor that is omitted among the necessary input factors, the accuracy of the predicted performance coefficient is lowered or a problem occurs that it is impossible to predict. In this study, we intend to create a model that predicts the coefficient of performance (COP) by using ANNs and SVMs that can accurately predict at low cost using small input factors. Hyper-parameter optimization is performed to increase prediction accuracy in machine learning models. We compared the accuracy of ANN and SVM-based prediction models. In this study, the ANN model showed higher CvRMSE by 5.4% and SVM by 8%. It is expected that the predictive model will be able to be used in the operation of the hybrid geothermal system in the future.

Keywords: hybrid geothermal heat pump system; artificial neural network; support vector machine; hyper-parameter; coefficient of performance

1. Introduction

The operation of a hybrid geothermal heat pump system aims to prevent the performance degradation of a system due to geothermal environmental problems [1–3]. This study aimed to reach the maximum coefficient of performance of the entire system by considering the energy of the heat pump and the energy saving of the circulation pump when controlling the flow rate of circulating water on the load side and the heat source side. To this end, it is necessary to monitor the current coefficient of performance according to the operation of a hybrid geothermal heat pump system and predict the coefficient of performance of the next stage. When installing a general geothermal system, a shortterm coefficient of performance measurement is performed to check the steady state of the system and long-term coefficient of performance monitoring is not performed. The coefficient of performance monitoring and prediction of the coefficient of performance in the next operation can be confirmed based on a mathematical model using the measured operational data. On the other hand, in the case of a mathematical model, many necessary input factors require many measurement sensors. When an input factor is omitted from the



Citation: Shin, J.; Lee, J.; Cho, Y. A COP Prediction Model of Hybrid Geothermal Heat Pump Systems based on ANN and SVM with Hyper-Parameters Optimization. *Appl. Sci.* 2023, *13*, 7771. https:// doi.org/10.3390/app13137771

Academic Editor: Francesca Scargiali

Received: 18 May 2023 Revised: 23 June 2023 Accepted: 24 June 2023 Published: 30 June 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https:// creativecommons.org/licenses/by/ 4.0/). necessary input factors, the accuracy of the predicted performance coefficient is lowered, or an unforeseen problem can occur [4,5].

A machine learning predictive model uses several theoretical techniques to find and predict the relationship between input and output data. Compared to a mathematical model, prediction is possible through relatively few input factors and requires a large amount of data. Various data collections are made using a BAS (Building Automation System) for system operation and monitoring in building facility systems, and it can be used to develop predictive models through machine learning. A machine learning-based prediction model cannot be estimated from data, unlike parameters that are estimated or learned from data and stored as a part of the trained model. The performance of the model is different. Selecting the optimal hyper-parameter is essential for improving the performance of the prediction model. In this study, a machine learning-based coefficient of the performance prediction model was developed through hyper-parameter optimization to predict the coefficient of performance in real-time and the next stage of performance, which is a necessary element for the efficient operation of a hybrid geothermal system [6].

2. Theoretical Analysis

2.1. Hyper-Parameter

A machine learning-based prediction model cannot be estimated from data, unlike parameters that are estimated or learned from data and stored as a part of the trained model. The performance of the model is different. Hyper-parameters are the input values of machine learning and deep learning models and control the model to train generalized inference performance from the target data characteristics. These hyper-parameters are composed of various variables that directly affect the model training performance, such as learning rate, learning rate scheduling method, loss function, number of training iterations, weight initialization method, regularization method, and the number of layers to be stacked.

Selecting the optimal hyper-parameter is essential for improving the performance of a prediction model. The prediction performance can be improved compared to the initial prediction model. In machine learning, each model has a range of hyper-parameters; Table 1 lists the types and descriptions of hyper-parameters.

Hyper-Parameter	Contents
Learning rate	A variable that determines how fast it will move in the gradient direction
Cost function	A function that estimates the difference between the expected value and the actual value according to the input
Regularization parameter	Using regularization method to avoid and solve overfitting problems
Mini-batch size	Splitting the entire training data to perform a batch set
Training loop	Variables that determine early termination of learning
Hidden unit	Learning Optimization Determinants on Training Data
Weight initialization	A determinant of performance

Table 1. Type of hyper-parameter.

Hyper-parameters can be optimized using various techniques. The manual search method is where the user calculates the optimal hyper-parameter through repeated execution based on intuition or experience. The random search method is a method to find the optimal combination by repeatedly extracting random values within the range, setting the minimum and maximum values of each hyper-parameter, and narrowing the extraction range based on the previously extracted and evaluated results. Hence, Bayesian optimization can optimize the hyper-parameters. Figure 1 shows the conceptual diagram of grid search and random search methods among hyper-parameter optimization methods.



Figure 1. Conceptual diagram of hyper-parameter optimization method.

2.2. COP Prediction Model Based on Machine Learning

For optimal control and energy saving of HVAC systems, it is necessary to accurately predict the coefficient of performance of a geothermal heat pump system [7]. For the coefficient of performance of the heat pump, there are theoretical calculations through mathematical models, short-term monitoring through simulations, long-term monitoring through measuring instruments [8,9], or predictive model development through machine learning [10,11]. Theoretical calculations using a mathematical model require a large number of input values, as shown in Figure 2. Hence, the model is expensive due to the need for various measuring equipment to measure it. Long-term on-site monitoring is the most accurate and reliable way of providing operational management guidance to maintain proper operation. On the other hand, the long-term monitoring method is expensive and time-consuming.



Figure 2. Input data for mathematical model.

A simulation is performed using commercial software, such as TRNSYS and theoretical models. The simulation results often appear more accurate than experimental field results because the simulation conditions are theoretical and simplified, ignoring real uncertain factors. In the case of a simulation, it requires professional knowledge and information collection on various conditions for the simulation, and it takes considerable time. A performance coefficient prediction through machine learning is a method of expressing the dynamic behavior of a system through measurement data. It does not require a high degree of expertise compared to other methods and does not require the selection of various variables [7,12].

Regarding the prediction of the coefficient of performance of geothermal systems, prediction techniques using mathematical models, such as multiple linear regression (MLR) [13,14], calculation of the coefficient of performance through simulation, and various machine learning models [15,16], such as artificial neural network (ANN)s and support vector machine (SVM)s, have been conducted in previous studies. Table 2 lists the contents of the existing literature research on the coefficient of a performance prediction model using various machine-learning models.

Category	Methods	Input	Author
Mathematical method	Polynomial	7 point (Heating load, etc.)	Nam et al.
	Random Forest	9 point (Power consumption, etc.)	Cho
	Recurrent Neural Networks	12 point (Environmental variable, etc.)	Sun et al.
	ANN	5 point (Power consumption, etc.)	Benli
	SVM	8 points (Heating capacity, etc.)	Esen et al.
Adaptive neu s	Adaptive neuro fuzzy inference system	8 points (Heating capacity, etc.)	Kecman
Machine learning	ANN	7 point (Power consumption, etc.)	Yilmaz
inaciante rearrang	K-NN	8 points (Water flow rate, etc.)	Akhlaghi
	Random Forest	12 point (Power consumption, etc.)	Swider
	ANN	29 point (Power consumption, etc.)	Park et al.
	ANN	5 point (Ground temperature, etc)	Esen et al.
ANN		3 point (Ground temperature, etc)	Esen et al.
	Back-propagation Neural Network	23 point (Water flow rate, etc)	Yan et al.
Random Forest and Back propagation Neural Network		16 point (Power consumption, etc.)	Lu et al.

Table 2. Previous studies of performance prediction.

Simon applied MLR to model the COP of a geothermal heat pump but used simulated data. Akhlaghi used a multiple polynomial regression approach to evaluate the effects of the intake air temperature, relative humidity, flow rate, and working air ratio on the COP of a dew point air cooler using simulation data. Park et al. developed MLR and ML neural network models to evaluate the essential parameters for predicting the COP of geothermal heat pump systems in hospital buildings [14]. Nam predicted the coefficient of performance of the geothermal system through a mathematical model, and Benli developed a prediction model using an ANN model with heat pump energy consumption as input [17]. Esen et al. developed a coefficient of the performance prediction model [18].

In the existing coefficient of performance prediction study, the coefficient of performance was predicted using various methods, but the coefficient of the performance prediction was mainly performed by machine learning using the directly related input values used for mathematical calculation of the coefficient of performance, such as power consumption and flow. In addition, things that are difficult to measure during operation or measured through multiple monitoring sensors, such as heat capacity, were used as input values. In addition, the coefficient of the performance prediction model was developed mainly using the data in the limited temperature range provided by the manufacturer or the data limited to the simulation, and studies related to the optimization of hyper-parameters to improve the accuracy of the prediction model are lacking. This study developed a coefficient of a performance prediction model that shows high accuracy at a low cost using the input values that are generally monitored and easy to measure during the operation of a heat pump system. The coefficient of a performance prediction model was developed using the operational data of geothermal systems and measurement data in various temperature ranges through experiments. The accuracy was improved through hyper-parameter optimization.

3. Methodology

Figure 3 shows a flow chart of the development of the performance prediction model. The development of the performance coefficient prediction model proceeds in the order of data collection and preprocessing for predictive model development, input variable selection, initial prediction model development, final prediction model development through hyper-parameter optimization, and prediction performance analysis.



Figure 3. Process of the development of the performance prediction model.

In this study, R studio (Ver. 1.2.1335) has been used for research, and an industrial application was used as a language for data mining and machine learning for developing the predictive model.

3.1. Data Collection and Pre-Processing

A model was developed for predicting the coefficient of performance of a hybrid geothermal heat pump system by collecting data from the target building where the existing geothermal system was installed and operated and data from a test bed equipped with equipment that simulates the geothermal system. The coefficient of performance of the hybrid geothermal heat pump system is equivalent to predicting the coefficient of performance of a geothermal heat pump system. When operating a hybrid geothermal system, the circulating water heat exchanged in the ground undergoes additional heat exchange from an auxiliary heat source, which changes only the temperature entering the heat pump. Hence, a hybrid geothermal system coefficient of the performance prediction model can be developed using the data from the geothermal system. The actual geothermal system operation data were collected through the BAS system for the buildings in which the geothermal system was installed. Table 3 shows an overview of the building for geothermal system operation data collection.

Additional data were collected from the test bed to develop a coefficient of a performance prediction model using the data of various temperature ranges that may occur during geothermal and hybrid geothermal systems operation. Figure 4 shows the configuration of the entire system used in the experiment, and the overview of the heat pump system is listed in Table 4.

Category		Contents
	Use	Office
Building	Total floor area	21,492 m ²
	Size	B1/F9
	Туре	Vertical closed-loop
	ĒĀ	7
Cooth arrest al arrestance	Capacity (Cooling)	105.6 kW
Geothermal system	Capacity (Heating)	101.4 kW
	Number of borehole	70
	Interval of borehole	5 m
Period of d	ata collection	5 February 2014~31 August 2019
Flow meter	Circulation pump compressor	

Table 3.	Overview	of buildings	for geothermal	system of	peration d	ata collection

Source side Load side Conde tank tank Circulation pump Ø Flow meter PID Expansion Draining Draining Auxiliary heat valve valve valve source system

Figure 4. Configuration of systems for operational data collection.

Table 4. Overview of heat pump systems.

Classification	Cooling	Heating
Capacity (kW)	190.61	178.93
Power (W)	38.03	44.52
Flow rate (LPM)	600	600
COP	5.01	4.02

The test bed used a constant temperature tank connected to a heat pump to simulate the heat source and load side and the heat source side. It was used as the cooling and heating source and for maintaining the temperature through an air-cooled auxiliary heat source. The load and heat source sides were simulated by connecting a constant temperature tank to a heat pump to supply cooling and heating sources and maintain the temperature through an air-cooled auxiliary heat source. Table 5 outlines the experimental system, and Figure 5 shows the systems installed in the target building.

In both systems, the factors affecting the performance coefficient of the geothermal heat pump, such as inlet and outlet temperature difference, flow rate, and power consumption, were collected for the performance evaluation in both systems, as well as the outside air temperature. Table 6 gives an overview of the operational data collection in both systems.

Table 5. Overview of auxiliary systems.

Classification	Spec.	Contents
Heat source tank	18.7 ton	Circulation flow rate on heat source side: 5~10 Ton
Load tank	18.7 ton	Circulation flow rate on load side: 5~100 Ton
Auxiliary heat source	10 HP	Air-cooled heat pump



Figure 5. View of systems installed in the target building.

Table 6.	Overview	of operation	data	collectior
----------	----------	--------------	------	------------

Category	Building	Test-Bed
Data collection pariod	1 January 2018~31	25 August 2019~27
Data collection period	December 2018	August 2019
Data collection cycle	3 h	1 min
Data collection method	BAS	Measurement
Temperature range (cooling)	24~43 °C	20~49 °C
Temperature range (heating)	11~24 °C	8.1~26.8 °C

In the case of a building with a geothermal system installed, data monitored through BAS were collected at a three-hour cycle, and the data from January 1 to 31 December 2018, where data in various temperature ranges among the total collected data were analyzed, were used to predict the coefficient of performance. In the test bed, the inlet/outlet temperature of the heat source and load side at one-minute intervals utilized an RTD (Resistance Temperature Detector) temperature sensor and was stored as a data logger (MV2000). The flow rate was measured through the load and heat source side thermostat system, and the flow meter (FMAG5000) was installed on the water pipeline. The power consumption was measured by installing a watt-hour meter (CW240) on the heat pump. The outside air temperature and humidity were measured, and data were collected through the HOBO U12. Figure 6 shows the measurement items and devices measured in the experiment and Table 7 shows the accuracy of the measurement devices.



Figure 6. Equipment for measuring coefficient of performance.

Measurement Device	Specific Information	Ranges	Accuracy
Tomporature concor (°C)	RTD (OMEGA)	$-50 \sim 200$	$\pm 0.15\%$
Temperature sensor (°C)	HOBO U12 (ONSET)	$-20 \sim 70$	$\pm 0.35\%$
Flow meter (m ³ /h)	FMAG 5000 (SIEMENS)	0~400	$\pm 0.4\%$
Power meter	CW240 (YOKOGAWA)	0~3000	$\pm 0.6\%$

Table 7. Accuracy of the measurement devices.

For performance evaluation in a broader range of heat source inlet-temperature conditions considering actual operating conditions, each item was measured while changing the heat source inlet temperature in the range of 8.1–26.8 °C during the heating operation and 20-49 °C during the cooling operation. The experiment was carried out as the water in the water tank on the heat source, and the load side was circulated and heated or cooled to the initial set temperature. After opening various valves connecting the heat pump and the water tank on the heat source side to circulate water, the temperature and flow rate of water at the inlet side of the heat pump heat-source-side were set to suit the experimental conditions. Various valves connecting the heat pump and the water tank on the load side were opened to circulate water. The temperature and flow rate of water at the inlet of the load side of the heat pump was set according to the experimental conditions. When a steady state was reached, each measurement item required to derive the heat pump performance coefficient was measured and recorded according to the measurement cycle. Table 8 lists the operational data items collected for the two systems. As for the measurement uncertainty, as shown in Table 9, the uncertainty of the temperature measuring device is 0.2 and the uncertainty of the flow meter is 0.26~0.32.

Table 8. Operation data list.

Category	Unit	Symbol
Outdoor air temperature	°C	T _{OA}
Source side inlet temperature	°C	$T_{S,i}$
Source side outlet temperature	°C	$T_{S,o}$
Load side inlet temperature	°C	$T_{L,i}$
Load side outlet temperature	°C	$T_{L,o}$
Heat pump power	kW	W _{HP}
COP	-	СОР
Pump power	kW	W_P

Table 9. Measurement uncertainty.

Category	Uncertainty (%)
Temperature	0.2
Flow rate	0.26~0.32

A large amount of data collected through the heat pump operation in the target building were analyzed to select the input variable by assessing the correlation between the coefficients of performance. An input variable was selected by quantitatively evaluating the relationship with a performance coefficient to be predicted using a Pearson correlation coefficient and a coefficient of determination. The Pearson correlation coefficient was calculated using Equation (1), and the coefficient of determination was calculated as the square of the Pearson correlation coefficient.

$$\mathbf{r} = \frac{\sum_{i=1}^{n} (X_i - \overline{X}) (Y_i - \overline{Y})}{\sqrt{\sum_{i=1}^{n} (X_i - \overline{X})^2} \sqrt{\sum_{i=1}^{n} (Y_i - \overline{Y})^2}}$$
(1)

r: Pearson correlation coefficient

n: Number of data

- X_i : Input data
- X: Average of input data Y_i: Output data
- \overline{X} : Average of output data

The Pearson correlation coefficient is a number from 1 to -1. As the value of the correlation coefficient approaches 0, the relationship between the input and output data is weaker. The coefficient of determination indicates the correlation between two variables to be measured; a coefficient closer to 1 indicates better performance. The inlet temperature, outlet temperature, load-side inlet temperature, outlet temperature, and outdoor temperature were used as the input. The monitored items with a high correlation coefficient between input data and output data of less than 0.2 and a coefficient of determination of less than 0.04 among the performance coefficient theoretical equation were selected as variables. Table 10 lists the analysis results of the determination coefficient and Pearson correlation coefficient for the input data for the selected input variable.

Table 10. Correlation analysis between input data and output data.

Category	r	r^2
Outdoor air temperature	-0.72	0.52
Source side inlet temperature	-0.76	0.58
Source side outlet temperature	0.59	0.35
Load side inlet temperature	0.60	0.39
Load side outlet temperature	0.32	0.10

The heat-source-side outlet temperature showed the highest correlation among the input variables, with a Pearson correlation coefficient of -0.76 and a coefficient of determination of 0.58. The outdoor temperature had a Pearson correlation coefficient of 0.32 and a coefficient of determination of 0.10, indicating a relatively small relationship among the input variables.

The data of the input variables, such as the inlet temperature on the heat source side, outlet temperature on the heat source side, load side inlet temperature, load side outlet temperature, and outdoor temperature, are the learning data by deleting the data in the non-operational state of the geothermal system and excluding abnormal data (null) values. (Training data), validation data (Testing data), and evaluation data (Validation data) to generate training data for predictive model development.

3.2. Accuracy Metrics

In this study, to develop an initial predictive model and improve the performance of the initial model, hyper-parameters for each machine learning are selected and set. The predictive model optimization and optimized hyper-parameters were calculated by performance analysis of the predictive model. The predictive model development and performance were analyzed using the parameters applied.

The accuracy was evaluated using the coefficient of variation, the Mean Bias Error (MBE), and the Coefficient of Variation (CV) to evaluate the performance of the predictive model. The MBE means the total error of the predicted value, and the coefficient of variation is a method of analyzing the error through the degree of variance. Using the criteria provided by ASHRAE Guideline 14 (MBE of 10% and Coefficient of variation of the Root Mean Squared Error (CvRMSE) of less than 30%), the accuracy of the prediction model was evaluated, as shown in Equations (2)–(4) below [19,20].

MBE =
$$\frac{\sum_{i=1}^{n} (P_i - M_i)}{n} \times 100$$
 (2)

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{n} (P_i - M_i)^2}{n}}$$
(3)

$$C_{\rm v} \rm RMSE = \frac{\rm RMSE}{\overline{P}} \times 100 \tag{4}$$

P: Predicted value

M: Measured value (or simulation value)

n: Number of measured data

 \overline{P} : Average of predicted value

4. COP Prediction Model Development with Hyper-Parameter Optimization

4.1. Initial COP Prediction Model Development

In this study, an initial coefficient of performance prediction model was developed using an ANN and an SVM model to analyze the suitability for predicting the heat pump performance coefficient by machine learning. At this time, each prediction model was developed using the built-in function of R studio. The initial model performance and optimization performance are compared and analyzed. Table 11 shows the statistical analysis of parameters through the random classification of input variables and training data selected for developing the initial coefficient of the performance prediction model [21].

Table 11. Statistical analysis of ANN model.

Variables		Statistical Parameters			
		Means	Standard Deviation	Minimur	n Maximum
	Training	2.90	0.71	1.17	5.46
COP	Testing	2.76	0.55	1.22	3.77
I and side inlat temperature	Training	17.29	1.71	11.60	25.00
Load side miet temperature	Testing	17.30	1.75	12.10	24.20
Load side outlet temperature	Training	16.02	2.12	9.00	24.40
Load side outlet temperature	Testing	15.99	2.16	9.10	23.60
Source side inlet temperature	Training	33.52	3.81	24.20	43.50
Source side inter temperature	Testing	33.56	3.74	24.60	43.00
Source side outlet	Training	37.52	3.81	28.20	47.50
temperature	Testing	37.56	3.74	28.60	47.00
Outsido tomporaturo	Training	25.43	2.21	22.90	33.70
	Testing	2530	2.22	23.00	33.10

The training data and the validation data were evenly distributed for all parameters. Statistical analysis of the parameters was performed by random classification of the input variables. The training data selected for developing the initial coefficient of the performance prediction model are shown in Table 12. The training and validation data were distributed evenly for all parameters [22].

The initial ANN utilization prediction model was constructed as an ANN model with a structure of one hidden layer and two hidden neurons with input variables. Table 13 lists the structure of the ANN, and Table 14 shows the structure of the initial prediction model of the SVM [23,24].

Variables		Statistical Parameters				
		Means	Standard Deviation	Minimum	Maximum	
COD	Training	2.90	0.71	1.17	5.46	
COP	Testing	2.76	0.55	1.22	3.77	
Load side inlet temperature	Training	17.29	1.71	11.60	25.00	
	Testing	17.30	1.75	12.10	24.20	
Load side outlet temmenature	Training	16.02	2.12	9.00	24.40	
Load side outlet temperature	Testing	15.99	2.16	9.10	23.60	
Course side inlet term ensture	Training	33.52	3.81	24.20	43.50	
Source side miet temperature	Testing	33.56	3.74	24.60	43.00	
Course aide outlet temperature	Training	37.52	3.81	28.20	47.50	
Source side outlet temperature	Testing	37.56	3.74	28.60	47.00	
Outside temperature	Training	25.43	2.21	22.90	33.70	
	Testing	2530	2.22	23.00	33.10	

Table 12. Statistical analysis of SVM model.

Table 13. Structure of initial ANN model.

	Contents		
	Acti	ivation	Sigmoid
E	Ι	LOSS	Mean squared error
Function	Optimizati	ion algorithm	Adam
	E	Epoch	
	Input lavor	Number of layer	1
	niput layer	Number of neuron	5
Classification	Liddon lavor	Number of layer	1
Structure	Hidden layer	Number of neuron	2
	Outrout lawor	Number of layer	1
	Output layer	Number of neuron	1

Table 14. Structure of initial SVM model.

Category	Contents
Туре	Eps-regression
Kernel	Polynomial
Epsilon	0.1
Cost	1
Gamma	1

The performance of the initial prediction model was evaluated based on the coefficient of determination, CvRMSE, and calculation time by comparing the predicted heat pump performance coefficient and the measured heat pump performance coefficient by inputting the test data into the developed prediction model. Table 15 presents the performance evaluation result, and Table 16 lists the calculation time analysis result. The ANN model satisfied the decision coefficient criterion exceeding 0.8 and the CvRMSE criterion. For the SVM, the CvRMSE was 15.9, which satisfied the criterion, but the coefficient of determination did not. The prediction accuracy of the initial model was 17.8% for the ANN model and 15.9% for the SVM, showing that the SVM was higher.

Table 15. Performance analysis of initial ANN model.

Calagory		Accuracy	
Category	R ²	CvRMSE	Error
ANN	0.89	17.8	-12~11
SVM	0.65	15.9	$-14 \sim 16$

Category	Normalization	Train Time (s)
ANN	Min-Max	1.43
	Standardization	0.83
CX7N4	Min-Max	2.63
SVM	Standardization	6.64

Table 16. Computation time analysis of initial ANN model.

The computation time of the prediction model for the environment to be checked in real-time should be considered, along with the accuracy of the prediction. The ANN terminates the operation when the indicator specifying the error between the measured value and the predicted value reaches a certain level. Table 17 lists the computation time analysis results of the initial coefficient of the performance prediction model developed using ANNs and SVMs.

Table 17. Hyper-parameter optimization method.

Machine Learning Model	Hyper Parameter	Optimization Method
	Training iterations	Grid search
ANN	Number of hidden layer	Bayesian optimization
	Number of hidden node	Bayesian optimization
	Kernel	Grid search
SVM	Number of gamma	Random search
	Number of cost	Random search

The ANN model has a shorter computation time than the SVM. A real-time prediction is possible because both models were calculated within one minute. The performance of the initial predictive model was improved by optimizing the structure and parameters of the predictive model.

4.2. Hyper-Parameter Optimization

When developing a machine learning-based prediction model, it cannot be estimated from data, unlike the parameters that are estimated or learned from data and stored as a part of the trained model. The performance of the model was different. It is essential to select the optimal hyper-parameter to improve the performance of the prediction model. The prediction performance can be improved compared to the initial prediction model. Table 17 lists the hyper-parameters analyzed for each machine learning model in this study and the optimization method.

Hyper-parameters can be optimized using a range of techniques. In the manual search method, the user calculates the optimal hyper-parameter through repeated execution based on intuition or experience. The random search method is used to find the optimal combination by repeatedly extracting random values within the range, setting the minimum and maximum values of each hyper-parameter, and narrowing the extraction range based on the previously extracted and evaluated results. Hence, Bayesian optimization optimizes hyper-parameters.

4.2.1. ANN Hyper-Parameter Optimization

In the ANN-based coefficient of the performance prediction model, the learning degree changes according to the number of hidden layers (Hidden Layer, HL) and the number of neurons in the hidden layer (Hidden Neuron, HN). As the number of hidden layers increases, the ANN is called "deep". In the case of a simple I/O relationship, high learning performance is possible even if the number of hidden layers and neurons is small. It is important to select the number of hidden layers and the number of neurons in the predictive model. In the hyper-parameter range of Table 18, the optimal number of hidden layers and neurons are derived through Bayesian optimization.

Category	Parameters
ANN	1, 2, 3, 4
SVM	1~17

Table 18. Hyper-parameter for optimizing ANN model.

Figure 7 shows the prediction model performance analysis according to 100 repetitions of learning. The performance of the performance coefficient prediction model was stabilized at 60 repetitions of learning. After a steady decrease in RMSE up to 40 times, unstable performance was analyzed during 40 to 50 repetitions. This required iterative learning at least 60 times to develop a performance coefficient prediction model using an ANN model.

The optimization of the number of hidden layers and the number of neurons proceeds through two steps. First, the coefficient of determination, CvRMSE, was analyzed when the number of hidden layers was increased from 1 to 4 and the number of neurons was fixed. With the number of hidden layers with the highest performance calculated using this method, the combination of the optimal number of hidden layers and the number of neurons was derived through the change in the number of neurons. The analysis result is shown in Figure 8. The best performance was shown when the number of hidden layers was two and the number of neurons in the hidden layer was 11.



Figure 7. Prediction model performance analysis for optimizing number of iterations.

4.2.2. SVM Hyper-Parameter Optimization

The hyper-parameters of SVM model were configured as shown in Table 19, and model development and accuracy analysis were performed.



Figure 8. Prediction model performance analysis for hidden layer structure.

Table 19.	Hyper-paramet	er for optin	nizing SVM	l model.
			()	

Category Parameters	
Kernel	Radial, Linear, Polynomial, Sigmoid
Cost	1, 2, 4, 8, 16
Gamma	0.00001, 0.0001, 0.001, 0.01, 0.1, 0

In this study, four models were developed according to the type of kernel, and each model was evaluated by applying the cost and gamma configuration using a random research method that sets the maximum and minimum values for cost and gamma and repeats and optimizes random values within the range. The parameters changed in the development of the SVM model are shown in the table. The cost was changed in the range of 0, 1, 2, 4, 8, and 16, and gamma was changed in the range of 0.00001, 0.001, 0.001, 0.01, 0.1, and 0. The performance was analyzed in a total of 120 configurations. According to the kernel, it was divided into SVM Model 1 (Radial), SVM Model 2 (Linear), SVM Model 3 (Polynomial), and SVM Model 1 (Sigmoid). Tables 20–23 show the performance of each model, and the shaded cell in each table is the value representing the highest performance in the model. Table 20 shows the performance results for each configuration of Model 1. In Model 1, when the cost was 16 and the gamma was 0, the performance was highest at 0.8. When analyzing the performance of Model 2, in which the kernel was applied linearly, it shows similar performance in most configurations. When the cost was 1,2, and 4, it showed the same performance in all gammas, and there was no change in performance when the cost was 8 and 16. When analyzing the prediction performance through simulation, when the cost was 16, and the gamma was 0.00001, 0.0231 showed the best performance (Table 21). Table 22 presents the performance results for each configuration of Model 3 to which the kernel was applied as a polynomial. In Model 3, when gamma was 0, high performance was shown regardless of cost, and when gamma was 0, a lower cost indicated higher performance. Performance analysis of each combination showed that the highest performance was 11.52 when the cost was 1 and the gamma was 0. As shown in the analysis results in Table 23, Model 4 shows the best performance at 20.84 when the cost is 2, and the gamma is 0.001. In the case of Model 4, the performance was much lower when gamma was 0.1 or 1, so it was not shown in the analysis results.

Category			Cost		
Gamma	1	2	4	8	16
0.00001	49.6	47.0	43.9	41.4	39.6
0.0001	39.3	38.5	37.9	37.7	36.5
0.001	36.5	34.1	29.7	26.5	24.7
0.01	20.4	12.3	5	2.6	1.7
0.1	3.2	4.4	1.9	2.4	2.5
0	5.2	3.1	2.5	1.2	0.8

Table 20. Prediction model performance analysis according to hyper-parameter (SVM model 1).

Table 21. Prediction model performance analysis according to hyper-parameter (SVM model 2).

Category			Cost		
Gamma	1	2	4	8	16
0.00001	2.334	2.334	2.313	2.310	2.310
0.0001	2.334	2.334	2.313	2.310	2.310
0.001	2.334	2.334	2.313	2.310	2.310
0.01	2.334	2.334	2.313	2.310	2.310
0.1	2.334	2.334	2.313	2.310	2.310
0	2.334	2.334	2.313	2.310	2.310

Table 22. Prediction model performance analysis according to hyper-parameter (SVM model 3).

Category	Cost				
Gamma	1	2	4	8	16
0.00001	51.21	47.00	50.05	51.05	53.05
0.0001	40.70	39.20	38.30	49.60	47.00
0.001	50.47	49.69	48.22	46.50	44.63
0.01	52.99	52.94	52.85	53.03	53.02
0.1	15.31	20.84	18.38	16.38	14.95
0	11.52	13.52	13.41	20.28	15.31

Table 23. Prediction model performance analysis according to hyper-parameter (SVM model 1).

Category			Cost		
Gamma	1	2	4	8	16
0.00001	51.21	49.55	47.00	43.92	41.41
0.0001	43.07	40.78	39.25	38.46	37.89
0.001	24.70	20.84	32.85	33.78	26.47
0.01	38.27	37.65	36.12	46.72	53.05

5. COP Prediction Model Accuracy Analysis with Hyper-Parameter Optimization

The final coefficient of the performance prediction model was produced by developing an initial coefficient of the performance prediction model for ANNs and SVMs, and the accuracy of the coefficient of the performance prediction model was improved by hyperparameter optimization. Table 24 lists the composition of the performance coefficient prediction model using the ANN developed by applying the derived optimal hyperparameters. By applying the derived optimal cost and gamma, four SVM models were developed, as shown in Table 25.

Category			Contents	
Function	Activation		Sigmoid	
	Ι	Loss		
	Optimizat	Optimization algorithm		
	Ē	5000		
Structure	Input layer	Number of layer	1	
		Number of neuron	4	
	TT: J.J. an Jacon	Number of layer	2	
	Hidden layer	Number of neuron	11	
	Output layer	Number of layer	1	
		Number of neuron	1	

Table 24. Construction of an ANN model with hyper-parameter optimization.

Table 25. Construction of a SVM model with hyper-parameter optimization.

Category	Contents		
Туре	Eps-regression		
Kernel	Radial		
Epsilon	16		
Cost	0		
Gamma	0.1		
Number of support vector	1116		

Table 26 and Figure 9 show the accuracy analysis results of each prediction model using the two machine learning models. The performance coefficient prediction model was analyzed using an ANN model. The MBE was -3.6%, and the CvRMSE was 5.4%, which is the verification standard for the hourly measurement data according to ASHRAE Guideline 14 (MBE is 10%, CV (RMSE) is less than 30%). The performance of the predictive model was excellent. The coefficient of the performance prediction model using the SVM also satisfied the criteria with MBE = -9.8% and CvRMSE = 8.1%. For the coefficient of determination, the ANN model was 0.95, and the SVM was 0.96, indicating a high relationship. Both models showed high performance, but the ANN model was better. Hence, the utility of the hybrid geothermal system was higher in this study.

Table 26. Performance analysis of prediction model with hyper-parameter optimization.

Category		Acc	uracy	
	MBE	RMSE	CvRMSE	R ²
ANN SVM	-3.6% -9.8%	15.96% 43.96%	5.4% 8.10%	0.953 0.962

Figure 10 shows the continuous data of the coefficient of performance predicted through the coefficient of performance prediction model using the ANN and the measured performance coefficient.



Figure 9. Comparison of prediction performance of prediction model with hyper-parameter optimization.



Figure 10. Comparison of measured performance and prediction performance through an ANN model with optimal hyper-parameter.

6. Conclusions

A machine learning-based coefficient of a performance prediction model was developed to predict the coefficient of performance in real-time and the coefficient of performance of the next stage, which is essential for the efficient operation of hybrid geothermal heat pump systems. The performance coefficient prediction model development dealt with data collection and preprocessing for predictive model development, input variable selection, initial prediction model development, and final prediction model development and verification through prediction model optimization. The details are as follows.

- (1) The data of the target building where the actual geothermal system was installed and operated were collected to develop a predictive model for the coefficient of performance of the hybrid geothermal heat pump system. The data were measured at the test bed installed with equipment that simulated a geothermal system. The learning data comprised one set of outdoor air temperature, a heat source side inlet/outlet temperature, load side inlet/outlet temperature, wattage, and COP.
- (2) The input variables were selected by quantitatively evaluating the relationship between the Pearson correlation coefficient and the performance coefficient to be predicted using the Pearson correlation coefficient and the determination coefficient for

the collected learning data. The outlet temperature of the heat source had the highest correlation among the input variables with a Pearson correlation coefficient of -0.76 and a coefficient of determination of 0.58, and the outdoor temperature showed a relatively small relationship among the input variables with a Pearson correlation coefficient of 0.32 and a coefficient of determination of 0.10.

The prediction model optimization that calculates the hyper-parameters of the initial performance coefficient prediction model of the ANN and SVM and the prediction model development and performance applying the optimized hyper-parameter were verified. The MBE was -3.6%, CvRMSE was 5.4%, and the coefficient of a performance prediction model using an SVM also showed predictive performance with -9.8% for MBE and 8% for CvRMSE.

Author Contributions: Conceptualization, methodology and writing-original draft, J.S.; writing editing and visualization, J.L.; supervision, project administration and funding acquisition, Y.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the National Research Foundation of Korea (NRF) grant funded by the Korean government (MSIT). (NRF-2021R1A4A1031705).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data sharing not applicable.

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

References

- Kelly, J.A.; Fu, M.; Clinch, J.P. Residential home heating: The potential for air source heat pump technologies as an alternative to solid and liquid fuels. *Energy Policy* 2016, *98*, 431–442. [CrossRef]
- Kang, E.C.; Riederer, P.; Yoo, S.Y.; Lee, E.J. New approach to evaluate the seasonal performance of building integrated geothermal heat pump system. *Renew. Energy* 2013, 54, 51–54. [CrossRef]
- 3. Fischer, D.; Madani, H. On heat pumps in smart grids: A review. Renew. Sustain. Energy Rev. 2017, 70, 342–357. [CrossRef]
- 4. Zhou, C.H.; Ni, L.; Wang, J.; Yao, Y. Investigation on the performance of ASHP heating system using frequency-conversion technique based on a temperature and hydraulic-balance control strategy. *Renew. Energy* **2020**, *141*, 141–154. [CrossRef]
- 5. Kazjonovs, J.; Sipkevics, A.; Jakovics, A.; Dancigs, A.; Bajare, D.; Dancigs, L. Performance analysis of air-to-water heat pump in Latvian climate conditions. *Environ. Clim. Technol.* **2014**, *14*, 18–22. [CrossRef]
- Szreder, M. Economical and technical aspects of using air source heat pumps for hot water. In Proceedings of the E3S Web of Conferences, Sanya, China, 19–21 November 2018; 46, p. 00014.
- 7. Zhou, C.; Ni, L.; Li, J.; Lin, Z.; Wang, J.; Fu, X.; Yao, Y. Air-source heat pump heating system with a new temperature and hydraulic-balance control strategy: A field experiment in a teaching building. *Renew. Energy* **2019**, *141*, 148–161. [CrossRef]
- 8. Lazzarin, R.; Noro, M. Lessons learned from long term monitoring of a multisource heat pump system. *Energy Build*. **2018**, 174, 335–346. [CrossRef]
- 9. Cho, S.H.; Kim, W.T.; Tae, C.S.; Zaheeruddin, M. Effect of length of measurement period on accuracy of predicted annual heating energy consumption of buildings. *Energy Convers. Manag.* **2004**, *45*, 2867–2879. [CrossRef]
- 10. Bourdeau, M.; Zhai, X.Q.; Nefzaoui, E.; Guo, X.; Chatellier, P. Modeling and forecasting building energy consumption: A review of data-driven techniques. *Sustain. Cities Soc.* **2019**, *48*, 101533. [CrossRef]
- 11. Alobaidi, M.H.; Chebana, F.; Meguid, M.A. Robust ensemble learning framework for day-ahead forecasting of household based energy consumption. *Appl. Energy* **2018**, *212*, 997–1012. [CrossRef]
- 12. Asaee, S.R.; Ugursal, V.I.; Beausoleil-Morrison, I. Techno-economic feasibility evaluation of air to water heat pump retrofit in the Canadian housing stock. *Appl. Therm. Eng.* **2017**, *111*, 936–949. [CrossRef]
- 13. Akhlaghi, Y.G.; Ma, X.; Zhao, X.; Shittu, S.; Li, J. A statistical model for dew point air cooler based on the multiple polynomial regression approach. *Energy* **2019**, *181*, 868–881. [CrossRef]
- Park, S.K.; Moon, H.J.; Min, K.C.; Hwang, C.; Kim, S. Application of a multiple linear regression and an artificial neural network model for the heating performance analysis and hourly prediction of a large-scale ground source heat pump system. *Energy Build*. 2018, 165, 206–215. [CrossRef]
- Dai, B.; Qi, H.; Liu, S.; Ma, M.; Zhong, Z.; Li, H.; Song, M.; Sun, Z. Evaluation of transcritical CO₂ heat pump system integrated with mechanical subcooling by utilizing energy, exergy and economic methodologies for residential heating. *Energy Convers. Manag.* 2019, 192, 202–220. [CrossRef]

- Yan, L.; Hu, P.; Li, C.; Yao, Y.; Xing, L.; Lei, F.; Zhu, N. The performance prediction of ground source heat pump system based on monitoring data and data mining technology. *Energy Build.* 2016, 127, 1085–1095. [CrossRef]
- 17. Kim, J.; Nam, Y.J. A Numerical Study on System Performance of Groundwater Heat Pumps. Energies 2016, 9, 4. [CrossRef]
- 18. Esen, H.; Inalli, M.; Sengur, A.; Esen, M. Forecasting of a ground-coupled heat pump performance using neural networks with statistical data weighting pre-processing. *Int. J. Therm. Sci.* **2008**, *47*, 431–441. [CrossRef]
- 19. Sayegh, M.A.; Danielewicz, J.; Nannou, T.; Miniewicz, M.; Jadwiszczak, P.; Piekarska, K. Trends of European research and development in district heating technologies. *Renew. Sustain. Energy Rev.* **2017**, *68*, 1183–1192. [CrossRef]
- 20. Rivière, P.; Adnot, J.; Marchio, D.; Pérez-Lombard, L.; Ortiz, J.A. A method to reduce European chiller hourly load curves to a few points. In Proceedings of the Climamed 2005–2nd Mediterranean Congress of Climatization, Madrid, Spain, 24–25 February 2005.
- 21. Eberhart, R.C.; Dobbins, R.W. Neural Network PC Tools: A Practical Guide; Academic Press: San Diego, CA, USA, 1990.
- 22. Zurada, J.M. Introduction to Artificial Neural Systems; West Publishing Company: St. Paul, MN, USA, 1992.
- 23. Fausett, L.V. Fundamentals Neural Networks: Architecture, Algorithms, and Applications; Prentice-Hall: Hoboken, NJ, USA, 1994.
- 24. Hassoun, M.H. Fundamentals of Artificial Neural Networks; MIT Press: Cambridge, MA, USA, 1995.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.