Machine Learning for performance prediction in smart buildings: photovoltaic self-consumption and life cycle cost optimization

Abstract

The application of Photovoltaic (PV) system in buildings is growing rapidly in response to the need for clean energy sources and building decarbonization targets. Nonetheless, enhancing PV self-consumption through technical solutions such as Energy Storage Systems (ESS) is getting higher importance to increase the profitability of PV plants, by minimizing the building-grid interaction. In this context, analyzing PV self-consumption of different energy storage configurations becomes more relevant and crucial in building energy modeling although it is heavily time-consuming and complicated, particularly within a multi-objective optimization related to the ESS design. As a solution to resolve this issue, this paper evaluates the accuracy, training, and prediction speed of 24 Machine Learning (ML) models to be used as surrogate models for analyzing PV self-consumption in smart buildings. Furthermore, the performance of short-term Thermal Energy Storage (TES) to increase PV self-consumption is assessed and presented using ML models. The results showed the Gaussian Process Regression (GPR), Neural Networks (NN) including bilayered and trilayered NN models, Support Vector Machines (SVM) including the fine gaussian and cubic SVM models, and Ensembles of Trees (EoT) as superior ML models. The results also revealed that TES systems can efficiently increase PV self-consumption in the building equipped with electric heat pumps to provide heating, cooling, and domestic hot water. Moreover, the TES size optimization regarding the Life Cycle Cost (LCC) showed that the LCC-based optimum TES size can yield 7.1% savings within 30 years of the building service life. The novelties of this research are first to provide a reference to select the most suitable ML models in predicting PV self-consumption, second to implement Machine Learning for analyzing the performance of short-term thermal energy storage to enhance PV self-consumption in buildings, and third to carry out an LCC-based optimization on TES size using ML-based prediction models.

Keywords: PV self-consumption, Prediction, Machine Learning, Accuracy, Thermal energy storage, LCC

Highlights:

1. Machine Learning is known as a reliable technique to predict the energy performance of buildings.
2. The accuracy, training, and prediction speed of 24 ML models are evaluated and compared in this study.
3. Top-ten most accurate ML models in predicting PV self-consumption are introduced by the obtained results.
4. The TES system is an effective solution for improving PV self-consumption in buildings.
5. The LCC-based optimum size of TES leads to 7.1% saving in the total life cycle cost of the building case study.

Nomenclature

AI	Artificial Intelligence	MSE	Mean Squared Error
COP	Coefficient of Performance	NN	Neural Networks
DHW	Domestic Hot Water	R^2	Coefficient of Determination
EER	Energy Efficiency Ratio	RE	Relative Error
EoT	Ensembles of Trees	RES	Renewable Energy Sources
ESS	Energy Storage System	RMSE	Root Mean Square Error
GPR	Gaussian Process Regression	RT	Regression Trees
HP	Heat Pump	SVM	Support Vector Machines
LR	Linear Regression	TES	Thermal Energy Storage
LCC	Life Cycle Cost	$\eta_{distribution}$	Efficiency of distribution systems
MAE	Mean Absolute Error	$\eta_{regulation}$	Efficiency of regulation systems
ML	Machine Learning	$\eta_{emission}$	Efficiency of emission systems

1. Introduction

Within the last few years, the share of Photovoltaic (PV) systems to supply electricity has been rapidly growing provoked by building and industrial decarbonization goals (Shukhobodskiy & Colantuono, 2020) (Gallego-Castillo et al., 2021). Moreover, the significant decline in the market price of PV systems (López Prol & Steininger, 2020) leads to its large-scale installation worldwide which has made it the second-leading absolute growth of all renewable sources so far (Del Pero et al., 2021). Nevertheless, in the countries where the grid parity for PV power plants has already been reached, improving self-consumption is known as the main driver for the profitability of photovoltaic systems (Hirschburger & Weidlich, 2020). Enhancing PV self-consumption yields substantial economic profits and becomes of paramount importance due to its potential in improving the environmental performance and grid stability by reducing peak-power injection to the electricity grid as well as lowering the grid-supplied electricity consumption (Ahmadiahangar et al., 2022).

In this context, several research works are carried out to introduce the concepts, policies, indicators, and technical solutions to measure and enhance the PV self-consumption in buildings. For instance, Luthander et al. (Luthander et al., 2015) defined self-consumption as the share of the total PV generation directly consumed in the building. They further classified the relevant metrics into four categories based on the type of metric and the type of required data and highlighted that metrics such as self-consumption and self-sufficiency belong to the category named load-matching which deals with the overlap between on-site load and generations in the buildings. Several metrics related to the overlap of on-site load and generation of electricity in buildings might differ by name and refer to the same concept (Luthander et al., 2015). A recent review discussed the key performance indicators in smart buildings including self-consumption highlighting that evaluating such indicators helps to underline the effectiveness of energy storage systems in improving buildings' energy performance (Al Dakheel et al., 2020).

PV self-consumption can be expressed as the ratio of self-consumed PV generation in the building to the total PV generation in the same time frame. While considering the performance of Energy Storage Systems (ESSs) in increasing PVself-consumption in buildings, it is important not to count the energy losses attributed to the ESSs (Luthander et al., 2015).

Regarding the growing capacity of installed PV systems around the world, the importance to measure and enhance PV self-consumption attracts more attention due to the economic benefits of PV systems in the absence of energy subsidies and lower price of feed-in tariffs of PV-supplied electricity (Yu, 2021). Moreover, improving PV self-consumption is highly encouraged to reduce the feed-in power to the grid to avoid overload and consecutive damage to the electricity grid (Amini Toosi et al., 2022). Energy Storage Systems (ESS) are recommended as technical solutions to improve PV self-consumption, including electrical and thermal storage in buildings (Baniasadi et al., 2020) (Amini Toosi et al., 2022). However, the results reported in the published articles vary noticeably due to different PV and ESSs sizes, the variety of on-site load, and generation profile in buildings around the world alongside the ESSs technology modeled in each study. For instance, Braun et al. (Braun et al., 2009) showed that installing 4.6 kWh lithium-ion battery in a residential building with annual electricity consumption equal to 5.5 MWh and 5 MWh as the annual PV generation can enhance the PV self-consumption from 30% to approximately 50% in Germany (Schreiber & Hochloff, 2013). Another study evaluated a building with a lower electricity consumption and PV size that is equipped with a larger battery system and showed that the PV self-consumption can be enhanced up to 72%, representing 41% improvement compared to the base scenario without ESSs (Schreiber & Hochloff, 2013). An extensive review elaborated that the potential of improving PV self-consumption in buildings equipped with battery systems ranges between approximately 13-24% if the battery capacity ranges between 0.5-1 kWh per kW PV power (Luthander et al., 2015). The self-consumption rate of an average Central European household equipped with PV can increase to 65-75% utilizing demand-side management and decentralized storage systems. However, residential consumers are likely to achieve a self-consumption rate of around 30% in the absence of such measures (European Commission, 2015).

Although numerous studies have evaluated the effectiveness of battery systems to improve PV self-consumption, few studies have addressed the potential of Thermal Energy Storage (TES) for this purpose. While thermal energy storage has been proved to be also effective for enhancing PV self-consumption (Thygesen & Karlsson, 2014), a recent study by Amini Toosi et al. (Amini Toosi et al., 2022) showed that the TES application in buildings equipped with electric heat pumps can be yield higher economic and environmental benefits when the PV system is designed to supply heat pumps' electricity consumption for heating, cooling, and domestic hot water demand.

Apart from the lack of assessing the potential of TES in evaluating PV self-consumption, the complexity of energy modeling in this field where several electrical and thermal systems are modeled leads to a high computational time, especially in the case of multivariable/ objective optimization of building energy systems. Artificial Intelligence (AI) based models such as Machine Learning (ML) are of those techniques to resolve this complexity which replaces a simulation-based model with a prediction model to forecast the performance of the whole scenarios in a design process based on few known results as the input of prediction model (Amini Toosi et al., 2022).

Although ML models are well-known and widely applied in the field of energy modeling, further investigations are required to verify the accuracy of different algorithms to be employed in an ML-based optimization model. Therefore, this paper aims to evaluate the accuracy and performance of several available ML models and propose the most reliable ones for energy modeling applications. The main novelties of this research are first, assessing the accuracy of multiple ML models in predicting PV self-consumption, second, evaluating the potential of thermal energy storage systems to improve PV self-consumption in buildings and third implementing an ML-based optimization model of TES size regarding the Life Cycle Cost (LCC) of the building case study.

In the following sections, the state of the art in ML application in energy modeling is briefly reviewed and a case study is developed to measure the accuracy and performance of 24 ML models in predicting PV self-consumption alongside demonstrating the performance of TES in enhancing PV self-consumption in buildings and LCC-based optimization of TES size.

2. Machine Learning for predicting building energy performance

Predicting the energy performance of buildings is of utmost importance to assist designers and engineers in defining their designs' energy, economic and environmental performance under different circumstances and uncertainties (Pham et al., 2020). Predicting energy performance can be carried out by physics-based, statistical, and Machine Learning models; although the first is accurate but requires considerable time and computation power. On the other hand, statistical models are easier to implement but may lack accuracy (Yu et al., 2020). In this agenda, the accuracy and efficiency of Machine Learning models draw researchers' attention to advance the buildings' energy demand predictions (Walker et al., 2020).

Machine Learning as an application of Artificial Intelligence (AI) can be described as a process by which the machine learns from past experiences and previous input data and realizes the correlation among input variables and outputs without a massive effort for preparing simulation-based results or programming (Hong et al., 2020) (Bourhnane et al., 2020).

Machine Learning algorithms can be categorized into supervised and unsupervised learning algorithms (Bourhnane et al., 2020). The supervised ML algorithms aim to create a model for predicting outputs based on previous experiences using a set of known labeled input and output data. They can be further categorized into classification and regression learning algorithms, which can be used to predict discrete and continuous responses, respectively. Regression techniques predict continuous responses, for example, changes in temperature or fluctuations in power demand. They can be further categorized into six ML models categories including Regression Trees (RT), Ensembles of Trees (EoT), Linear Regression Models (LR), Gaussian Process Regression models (GPR), Support Vector Machines (SVM), and Neural Networks (NN) (Mathworks, 2022).

Table 1 Summary of the papers that implement Machine Learning models to predict energy performance of building

Author	Building Type	ML algorithm	ML prediction target	Highlights/ Notes
(Zhou & Zheng, 2020)	Office	Multiple linear regression, Support Vector Machine (SVM), backpropagation neural network	Building energy load	Predicting electricity consumption of heating, cooling, lighting, and devices
(Yu et al., 2020)	Educational	Nonlinear autoregressive with exogenous inputs and Artificial Neural Network NARX-ANN	Building thermal load	
(Yang et al., 2020)	Office-educational	Nonlinear autoregressive exogenous-ANN	Control HVAC	A model predictive control system with ML for building automation and control
(Wang & Hong, 2020)	Non-residential	Generative Adversarial Network (GAN)	Electrical load profile	The model tested on the data of numerous buildings
(Walker et al., 2020)	47 Commercial buildings	Multiple ML algorithms: boosted three, random forest, SVM-linear, quadratic, cubic, fine gaussian, ANN	Electricity demand	Compared the performance of different algorithms
(Jonas et al., 2020)	Office	Differential evolution online sequential extreme learning machine (DE-OSELM)	Occupants' presence	
(Vela et al., 2020)	Residential and gym	k-Nearest Neighbor (kNN), Support Vector Machine (SVM), Decision Trees (DT)	Occupancy level	Predicting occupancy level by 3 ML model based on temperature, humidity, and pressure
(Sajjad et al., 2020)	Residential	Gated Recurrent Unit (GRU), Support Vector Machine (SVR)	Heating/cooling load	Multiple output prediction
(Sadeghi et al., 2020)	Residential	Deep Neural Network (DNN), ANN	Heating/cooling load	Accuracy of ANN and DNN are compared
(Ruiz et al., 2020)	Educational	Trees, Support Vector Machine (SVR), Neural Network	Energy consumption	ML algorithms are compared
(Pham et al., 2020)	Multiple buildings	Random Forest (RF)	Short term hourly energy consumption	
(Parzinger et al., 2020)	Residential	Autoregressive with Exogenous variables (ARX)	Fault detection in HVAC	
(Ngarambe et al., 2020)	Simulated Test room	Generalized linear model, Deep Neural Network (DNN), random forest, gradient boosting model	Indoor daylight illuminance	
(Mawson & Hughes, 2020)	Manufacturing buildings sector	Deep Neural Network: feedforward and recurrent	Energy consumption, air temperature, and humidity	The accuracy of two DNN algorithms is compared.
(Martínez-Comesaña et al., 2020)	Public library	eXtreme Gradient Boosting (XGBoost), Support Vector Regression (SVR) and Multi-Layer Perceptron (MLP) neural network	Heat Loss Coefficient (HLC)	
(Maljkovic & Basic, 2020)	District level	Regression trees, random forest, Support Vector Machine (SVM)	Ranking the importance of parameters affecting heat consumption in district heating systems	
(Liu et al., 2020)	Residential	Holt winters-Exterme learning machine network (HW-ELM)	Electricity consumption	
(Ivanko et al., 2020)	Residential	ANN, Prophet, and XGBoost	Domestic Hot Water (DHW) heat use	
(Hwang et al., 2020)	Educational	Deep learning, Gradient boosting, Random forest	Electricity consumption, coefficient of performance of heat pumps	
(Bui et al., 2020)	Residential	Electromagnetism-based Firefly Algorithm - Artificial Neural Network (EFA-ANN)	Heating/ cooling load	
(Zekić-Sušac et al., 2021)	Public	Deep ANN, Decision trees, and random forest	Energy consumption	
(Sha et al., 2021)	Institutional	Gradient tree boosting	Cooling load	
(Amasyali & El-Gohary, 2021)	Multiple buildings	Classification and regression trees (CART), ensemble bagging trees (EBT), artificial neural networks (ANN), and deep neural networks (DNN).	Energy consumption	
(Alduailij et al., 2021)	Multiple buildings	long short term memory (LSTM) network, Artificial Neural Network	Peak energy demand	

Table 2. Summary of Machine Learning application in the field of photovoltaic systems

Author	ML algorithm	ML prediction target	Highlights/ Notes
(Shivam et al., 2021)	Naïve, SVM, ANN, Res-DCNN	PV generation, energy load	Residual dilated causal convolutional network (Res-DCCN) has a higher accuracy followed by SVM and ANN
(Zhou et al., 2020)	SVM, BPNN, ELM, SDA-SVM, SDA-BPNN, SDA-ELM, SDA-GA-ELM	PV power output	The new hybrid SDA-GA-ELM prediction model has higher accuracy than others
(Feng et al., 2020)	ELM, SVM, GRNN, M5T, AE	Daily global solar radiation	The new proposed hybrid PSO-ELM prediction model has a higher accuracy
(Barthwal & Rakshit, 2021)	ANN	Annual thermal and exergy output	
(Natarajan et al., 2021)	DBN, BPNN, ANN, RBNN	PV generation	The new proposed RBNN model has higher accuracy than others
(Ramadhan et al., 2021)	SVR, RF, RNN, LSTM, GRU	Solar irradiance, PV power	LSTM and GRU, which are specialized for modeling sequential data and powerful for handling a large volume of data, were recommendable
(X. Wang et al., 2022)	Bayesian Ridge Regressor, Linear Regressor, Gaussian Process Regressor, MLP Regressor, SVR, Gradient Boosting Regressor, Random Forest Regressor, Lasso, Regressor, Ridge Regressor,	PV power output based on weather type classification	Results showed that in general, Lasso regressor, Random forest regressor, Gradient boosting regressor and SVR has a better accuracy. However all ML models have high accuracy when the local meterological conditions are relatively stable. If not, the SVR model has higher accuracy.
(Khan et al., 2022)	ANN, LSTM, Bagging and DSE-XGB (an improved generally applicable stacked ensemble algorithm proposed based on utilizing two deep learning algorithms namely artificial neural network (ANN) and long short-term memory (LSTM) as base models	PV generation	The proposed DSE-XGB model has a higher accuracy compared to the other models and results in improving R^2 by 10-12% regardless of weather variations.
(Visser et al., 2022)	Regression, SVR, Ensemble learning, Deep learning and physical based techniques	PV power	Ensembles and Deep learning performs better form a technical viewpoint
(Jia et al., 2022)	SVM, Generalized linear modeling, Random Forest	Daily global and diffuse solar radiation	SVM has a higher accuracy than other models and demonstrate a high reliability under slight pollution and stable weather conditions
(Abubakar Mas'ud, 2022)	kNN, Multiple Regression and Decision Trees Regression	PV power	kNN outperformed other models while all tested ML models demonstrated reasonable accuracy
(Ramos et al., 2022)	Random Forest, Gradient Tree Boosting, SVM	PV generation and export to the grid	They showed that Gradient Tree Boosting has a higher accuracy for predicting PV generation and export than other models.
(Zazoum, 2022)	Gaussian Process Regression, SVM	PV power	The results showed that GPR_Matern 5/2 has a higher accuracy to predict PV power based on weather data as the predictors.

Numerous studies have already implemented Machine Learning to predict building performances from multiple perspectives and at different levels. A recent review (Fathi et al., 2020) showed that most published research works concentrated on forecasting electricity consumption and heating cooling load by 44% and 39% of all published works, respectively. They also elaborated that Artificial Neural Network (ANN), Support Vector Regression (SVR), and Random Forest (RF) are the most popular ML algorithms implanted in forecasting the energy performance of buildings.

Table 1 represents the implemented ML algorithms in the selected scientific articles published in 2020 and 2021. As shown in table 1, most research papers implemented the ML models to predict the electricity consumption, heating, and cooling load of residential and other building types. Moreover, the literature covers other prediction targets such as occupant behavior and presence, daylight illuminance, and HVAC fault detection. As an essential requirement of the efficient operation of smart buildings, the peak energy demand, HVAC control parameters, electricity load profile prediction, etc., are also included in the literature.

More in detail, Table 2, summarizes the selected scientific articles published between 2020 to 2022 that implemented Machine Learning models in predicting photovoltaic systems performance. As shown in table 2 most papers employed ML models to predict PV power/ generation based on the variation of weather data in different locations.

Several researchers employed different ML models and compared their accuracy in predicting PV power/ generation. For instance, Zazoum (Zazoum, 2022) showed that GPR_Matern 5/2 is superior to predict PV power compared to the SVM models. Another study performed by Mas'ud (Abubakar Mas'ud, 2022) concluded that kNN has a higher accuracy among the three tested ML models including multiple regression and decision trees regression models. While Visser et al. (Visser et al., 2022) highlighted the higher accuracy of Support Vector Regression models in operational day-ahead solar power forecasting. Some authors also put effort to propose and develop novel hybrid ML models for predicting solar radiation, PV power, and generation in different climatic and geographical conditions and prove the advantage of their proposal by comparing the prediction precision of existing models (Zhou et al., 2020).

ML models are vastly employed to forecast solar radiation, PV power, and generation under different weather conditions (table 2). Nonetheless, predicting PV self-consumption is less addressed in the literature and the application field of ML-based prediction models. Moreover, although several researchers have compared the accuracy of different ML models and provides insightful information, few papers reported and compared a comprehensive set of existing ML models. In this context, support vector machines are the most tested models while several ML models are less evaluated and compared. Therefore, the state of the art barely demonstrates a clear and comprehensive comparison among existing ML models' accuracy for predicting photovoltaic performance.

Furthermore, regarding the supporting solutions to improve PV self-consumption such as ESSs, there are almost no scientific articles demonstrating the performance of ML models in designing thermal/ electrical energy storage systems.

Thus, the three research gaps including the lack of ML application in predicting PV self-consumption, the lack of employing Machine Learning as predictive models in designing ESSs for improving PV self-consumption, and the absence of a comprehensive comparative assessment of existing ML models' accuracy are chosen as the main objectives and novelties of this research. The following sections present a case study to compare the accuracy of 24 ML models in predicting PV self-consumption of different TES capacities.

3. Materials and methods

Aiming at evaluating the accuracy of multiple ML models in predicting PV self-consumption, first a parametric energy model of the building case study is developed allowing a parametric analysis to measure the energy consumption of the building to provide heating, cooling, and Domestic Hot Water (DHW) demand. The parametric model is also enabled to measure the amount of required electricity to be imported and exported from and to the electricity grid in different configurations of ESSs. The building is equipped with air-water heat pump units and photovoltaic panels to supply the building's energy heating/ cooling and DHW energy demand. The size of Thermal Energy Storage (TES) tanks is chosen as the design variable in this study.

A massive number of combinations of TES sizes for heating/cooling and DHW is usually allowable, which requires considerable time to conduct simulation-based analysis. Thus, instead of simulating all design scenarios, Machine Learning is used to enable the designer to analyze the whole design scenario through a prediction process based on simulation-based results of a few design scenarios of TES size. Therefore, six groups of ML models containing 24 different ML algorithms are trained and used to predict the PV self-consumption in each design scenario. Then, these algorithms are evaluated by comparing the statistical indicators and computational time efficiency including each algorithm's training time and prediction speed. Finally, the most accurate ML model is utilized to predict PV self-consumption of all possible TES sizes. Given the accurate surrogate model to predict PV self-consumption, the selected ML model will be used to find the cost optimum size of TES.

3.1. Building energy system configuration

The proposed case study is a multi-family residential building located in Bagnolo in Piano, Reggio Emilia, Italy (figure 1). The building was constructed in 1985 and retrofitted in 2022 as part of the Horizon 2020 HEART project (HEART, 2022). It consists of four floors including three residential floors with 12 apartments and a total net surface area of 636 m^2. The HVAC system is equipped with high-efficiency air-water heat pumps coupled with a water loop emission system (i.e. decentralized water to air heat pumps) able to increase the heating/cooling temperature of the water loop connected to the centralized heat pump, reducing at the same time the distribution losses.

Building's envelope and energy systems features are presented in table 3. Figure 2, illustrates the general operational schemes of the building, including PV and TES (filled with PCM materials) components. To reduce the uncertainty and complexity of the analysis related to the technical plants, in this study, the building energy performance is simulated parametrically in Grasshopper and EnergyPlus considering a traditional emission system (i.e. standard fancoil units) rather than the water loop one and neglecting the PCM storage capacity. The thermal energy demand of the building for heating, cooling, and DHW is 28,884 kWh_{th}, 10,785 kWh_{th}, and 13,226 kWh_{th} respectively and the total electricity consumption by heat pump units is estimated at 22.51 kWh/m^2.year.

Figure 1. Case study residential building, Bagnolo, Italy

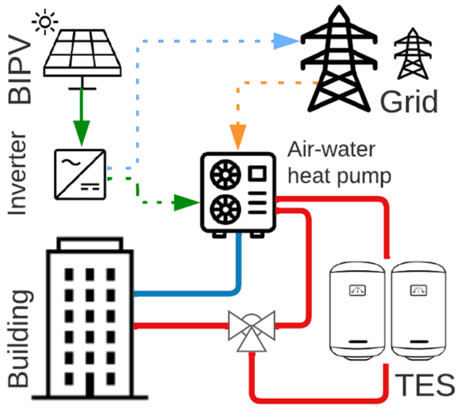

Figure 2. General operational scheme of the building

The Coefficient of Performance (COP) of heat pump units is calculated by the equation in table 3 based on the performance map of the components (Miglioli et al., 2019) (Amini Toosi et al., 2022). The TES system is made of two thermal storage tanks to store the thermal energy required for heating/cooling and DHW separately. The size of TES is modeled parametrically as the design variable between 0 to 5000 liters according to the space constraints in the building.

Table 3. Building and energy system characteristics, η: efficiency, U value: thermal transmittance

Parameters	Values
Internal Floors	U value = 1.18 W/m^2K
Ground Floor	U value = 3.49 W/m^2K
Roof	U value = 0.55 W/m^2K
External Walls	U value = 0.266 W/m^2K
Windows	U value = 1.2 W/m^2K, Solar heat gain coefficient = 0.6
COP of HP units	= 0.001 (water_temp - Ext_temp+10)² - 0.17 (water_temp - Ext_temp+10)+10
EER of HP units	= COP - 1
$\eta_{distribution}$, $\eta_{regulation}$, $\eta_{emission}$	0.95
Inlet water temperature to HP units	15°C
Outlet water temperature from HP units	35°C in space heating mode 10°C in space cooling mode 45°C in DHW mode
Thermal transmittance of TES envelope	U value = 0.35 W/m^2K
TES size for heating/ cooling	Variable (0 to 5000 liters)
TES size for DHW	Variable (0 to 5000 liters)
Type of the fluid in TES	Water
Thermal energy storage period	1 day ahead
PV peak power	8.5 kW_p
PV systems losses	14%

The thermal energy storage is foreseen for 1 day ahead and the control logic estimates the amount of excess PV generation and activates charging of the TES concerning the predicted electricity need during the next day and the TES capacity as illustrated in figure 3.

The parametric energy model for TES size optimization is developed using Grasshopper and EnergyPlus and the surrogate ML-based energy model to predict all possible scenarios (different combinations of TES sizes) is developed in Matlab. The ML-based surrogate models are trained by 121 simulation-based results (representing around 5% of the whole design space) to predict all combinations of design variables which helps reduce the computational time and preserve accuracy.

3.2.Indices definition

In general, three types of indices are used for this study. First and foremost, PV self-consumption is chosen to measure the impact of different TES sizes on the building energy performance. PV self-consumption can be calculated using the following equation which indicates the share of PV-generated electricity directly used in buildings. PV self-consumption is known as an accepted index to evaluate the optimum sizing of energy systems in buildings for maximizing the use of Renewable Energy Sources (RES) and minimizing the building-grid interaction which will also result in lower dependency on the grid-supplied electricity to operate building energy systems. Equation 1 is used to quantify PV self-consumption in the present assessment.

Equation 1. PV self-consumption

$$PV\ Self\ Consumption\ Rate = \frac{\int_{t1}^{t2} M(t)dt}{\int_{t1}^{t2} P(t)dt}$$

Where:
$M(t) = min\{L(t), P(t)\}$
$P(t) = min\{L(t), P(t) + S(t)\}$
L(t) is instantaneous building power consumption
P(t) is on-site PV power generation
S(t) is the power to and from the storage unit (negative during the charging phase)

Moreover, five statistical indices are used to measure the accuracy of predictions by ML models including the coefficient of determination (R^2), Root Mean Square Error (RMSE), Mean Squared Error (MSE), Mean Absolute Error (MAE), and Relative Error (RE) which can be calculated by the following equations:

Equation 2. R Squared

$$R\ Squared = 1 - \frac{\sum_1^n(Y_i - \widehat{Y_\iota})}{\sum_1^n(Y_i - \bar{Y})}$$

Equation 3. Root Mean Square Error

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}\left(Y_i - \widehat{Y_\iota}\right)^2}{n}}$$

Equation 4. Mean Squared Error

$$MSE = \frac{1}{n}\sum\nolimits_{i=1}^{n}\left(Y_i - \widehat{Y_\iota}\right)^2$$

Equation 5. Mean Absolute Error

$$MAE = \frac{\sum_{i=1}^{n}\left|\widehat{Y_\iota} - Y\right|}{n}$$

Equation 6. Relative Error

$$RE_i = \frac{\widehat{Y_\iota} - Y_i}{Y_i} \times 100$$

Where:
Y_i is the actual simulation-based value of scenario i
$\widehat{Y_\iota}$ is the predicted value of scenario i
$\overline{Y}$ is the mean value of observations
n is the total number of observation/ prediction values

Finally, the Net Present Value (NPV) index as the most used economic indicator in LCC studies (Amini Toosi et al., 2021) is used to find the life cycle cost optimum size of TES. NPV as a recommended economic index by EN 16627 (EN_16627, 2015) for life cycle cost analysis can be calculated using equation 7.

Equation 7. Net Present Value

$$\boldsymbol{NPV} = Investment_{first\ year} + \sum_{i=1}^{n} Operational\_Cost_{first\ yaer}\left(\frac{1 + El\_PIR}{1 + DR}\right)^i$$

Where:
Investment: the initial investment of the TES systems
Operational Cost: the annual cost of electricity considering purchasing electricity from the grid and reimbursement for the feed-in electricity to the grid in the first year.
El_PIR: the electricity price inflation rate
DR: the discount rate
n: the study period (service life) considered in the case study

Table 4. Economic parameters to calculate NPV

Economic parameter	Value	Unit
Electricity price (import from the grid)	0.18	Euro/kWh
Electricity price (export to the grid)	0.10	Euro/kWh
TES investment	300 + (TES volume in litres/ 2)	Euro
Electricity price inflation rate	2.1%	-
Discount rate	3.68%	-
Study period	30	Years
The useful technical life span of TES	30	Years

As shown in table 4, the electricity tariffs for purchasing electricity from grid and the feed-in tariff for electricity are selected from previous studies equal to 0.18 and 0.1 Euro per kWh electricity (Miglioli et al., 2019). It should be noted that the current electricity price has a higher value than that considered in the present study, however, it should be considered a temporary situation. Furthermore, the 10 years average discount rate and electricity price inflation rate in Italy are taken from previous studies equal to 2.1% and 3.68% (Amini Toosi et al., 2020). The life cycle cost analysis is carried out for 30 years of service life of the building equivalent to the average useful life span of the TES components.

3.3.ML modeling and evaluation process

The process starts with creating a parametric building model including the building fabric, HVAC systems, RES, and TES, presented in table 3. The parametric model analyzes the hourly energy consumption, PV generation, direct use of PV-generated electricity, charging and discharging of TES, imported and exported electricity from and to the grid as well as PV-self consumption ratio for the various sizes of TES components.

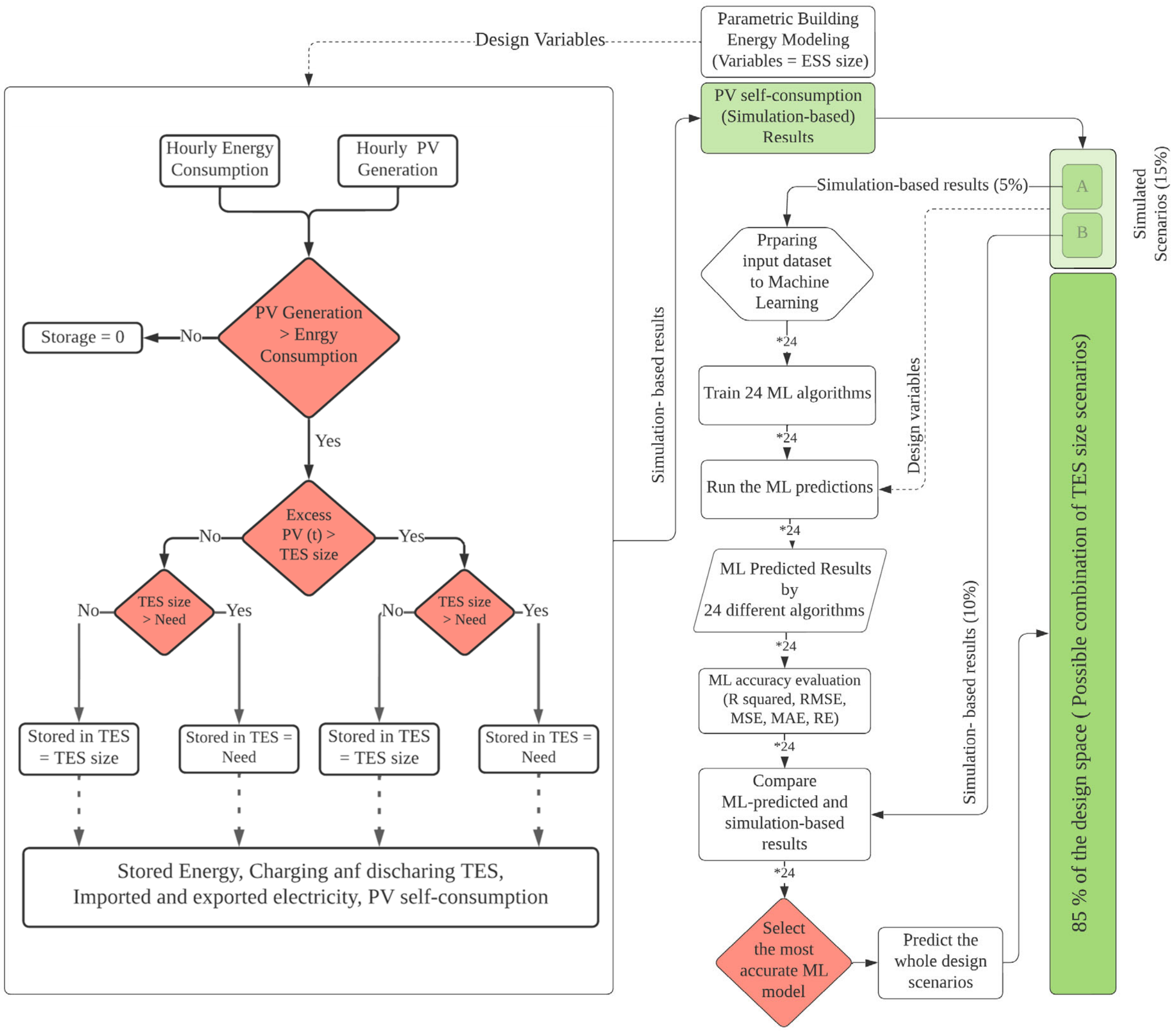

Figure 3. Parametric modeling and machine learning evaluation process

To train the ML models and test their accuracy, 15% of the whole design space (possible combination of TES sizes) is initially simulated and divided into training (A) and test (B) sets. The training set (A) representing 5% of the whole design space is used to train 24 ML models. These trained models are then employed to predict PV self-consumption of the test set (B) to measure the accuracy of ML predictions by comparison to the simulation-based results.

Then the most accurate ML model can be applied to predict the whole design scenarios which assists designers in finding the optimum solutions with high accuracy and minimum computation time and effort. Figure 3 illustrates the evaluation process, including the parametric energy modeling of energy storage systems, training, and testing of multiple ML models.

4. Results and discussion

The initial results of assessing the effectiveness of TES application for improving PV self-consumption showed a significant potential, especially during the midseason and summer period when hourly PV generation is higher than the electricity demand of building energy systems for heating, cooling, and DHW. Moreover, the results showed that the application of TES systems is highly beneficial even during winter except for days with very low solar radiation.

For instance, the comparison between the application of TES (3000 liters for heating/cooling and 1500 liters for DHW) with the base scenario without TES shows that the TES system can increase PV self-consumption from 41.03% to 62.07% leading to 23.19% reduction in imported electricity from the grid. Figure 4 illustrates the energy profile of the building on 21st June as an example day in which the application of TES enhanced the PV self-consumption up to 99.39%. As shown in figure 4, all PV generation during this day is either used directly to supply instantaneous electricity demand or stored in TES to cover the thermal energy demand within the day ahead however the PV self-consumption cannot practically reach 100% due to the energy losses in the TES system.

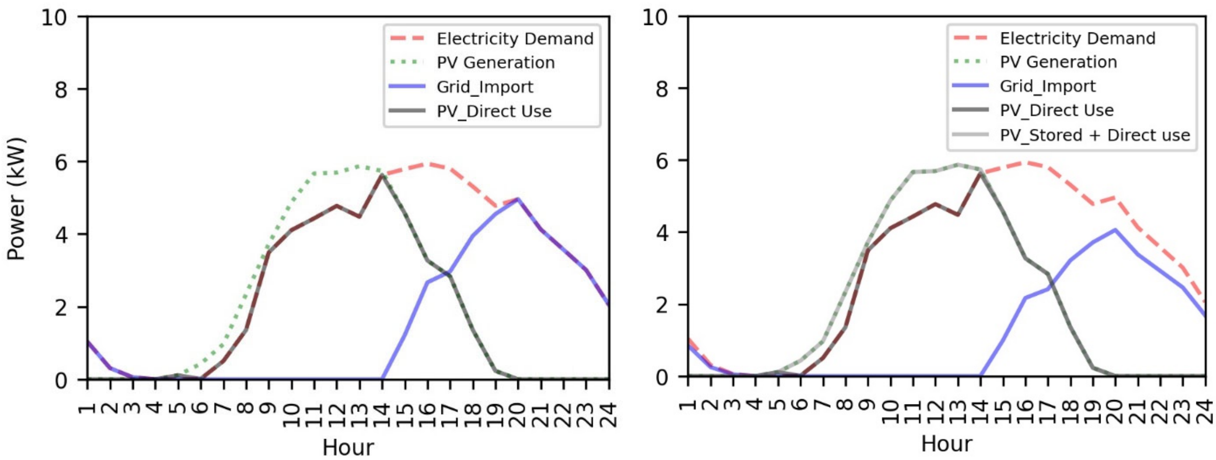

Figure 4. Hourly energy profile on 21st June,
Left) without TES application
Right) with 3000 liters and 1500 liters of TES for heating/cooling and DHW

The next sections will present the results of testing ML models' accuracy in predicting PV self-consumption for different TES sizes. Such prediction models help to create and estimate the energy profile of buildings for any TES size configuration to realize the impact of different TES sizes on PV self-consumption with lower computational time and cost during the design phase.

4.1.ML models' prediction accuracy

As described in the methodology section, 392 simulation-based results were obtained through hourly energy simulation, representing around 15% of all possible combinations of TES size as the design variables. After training Machine Learning models on 5% of simulated scenarios (A: train set), 24 ML models were evaluated regarding the prediction accuracy in terms of R^2, RMSE, MAE, MSE, and RE by comparing the predicted and simulated results on around 10% of the remaining simulated scenarios (B: test set).

With respect to the coefficient of determination (R^2), it is revealed that in general ML algorithms belonging to Gaussian Process Regression (GPR), Neural Network (NN) bilayered and trilayered models, Ensembles of Trees (EoT), and fine gaussian SVM represent higher accuracy among all other ML algorithms. Table 5 summarizes the evaluation results of all employed algorithms.

Table 5. Statistical indices for accuracy assessment of ML-based prediction models

Model Type	ML algorithms	R^2	RMSE	MSE	MAE
Regression Trees	Fine_Tree	0.95	1.27	1.620	0.843
	Medium_Tree	0.918	1.833	3.361	1.237
	Coarse_Tree	0.660	3.684	13.572	3.027
Ensembles of Tress	Boosted_Trees	0.988	2.305	5.314	2.158
	Bagged_Trees	0.981	1.373	1.885	1.013
Linear Regression Models	Linear	0.747	2.987	8.923	2.366
	Linear_Interactions	0.759	2.945	8.671	2.341
	Linear_Robust	0.744	3.112	9.682	2.347
	Linear_Stepwise	0.747	2.987	8.923	2.366
Gaussian Process Regression Models	GPR_Rational Quadratic	1.000	0.108	0.012	0.082
	GPR_Squared Exponential	0.998	0.254	0.065	0.202
	GPR_Matern 5/2	1.000	0.072	0.005	0.053
	GPR_Exponential	0.999	0.158	0.025	0.080
Support Vector Machines	SVM_Linear	0.728	3.413	4.304	2.417
	SVM_Quadratic	0.901	2.075	4.304	1.505
	SVM_Cubic	0.967	1.050	1.103	0.895
	SVM_Fine Gaussian	0.990	0.804	0.647	0.607
	SVM_Medium Gaussian	0.961	1.309	1.713	0.867
	SVM_Coarse Gaussian	0.821	3.002	9.015	1.897
Neural Networks	NN_Narrow	0.747	2.987	8.923	2.366
	NN_Medium	0.961	1.211	1.467	1.011
	NN_Wide	0.961	1.211	1.467	1.011
	NN_Bilayered	0.998	0.236	0.056	0.194
	NN_Trilayered	0.997	0.328	0.107	0.250

Regarding the coefficient of determination (R^2), Linear Regression (LR) models, along with Coarse_Tree, SVM_Linear and NN_Narrow are found among the models with the lowest accuracy to predict the energy performance of design scenarios. Amongst the models belonging to the Regression Trees, the Fine Tree showed a higher accuracy with R^2 equal to 0.95. On the other side, Gaussian Process Regression Models represent a significant accuracy. All GPR models along with NN_Bilayered, and NN_Trilayered belonging to the Neural Network models, SVM_Fine_Gaussian, SVM_Cubic, Boosted Trees, and Bagged_Trees are found among the top-ten most accurate ML models in this case study.

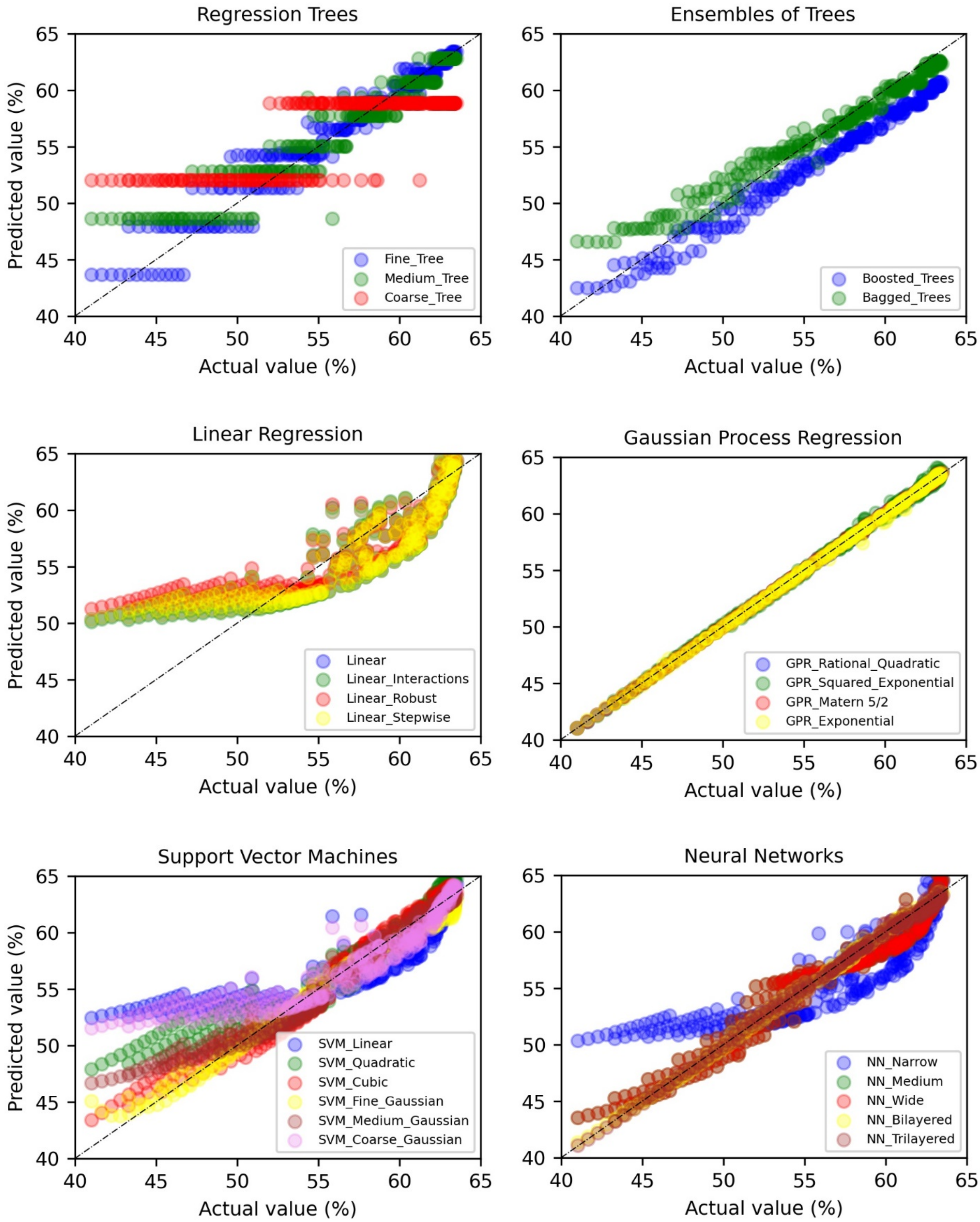

Figure 5. Actual simulation-based values versus the predicted values by all employed ML models

Figure 5, compares the actual simulation-based values and the predicted values by all employed ML models. Gaussian Process Regression models result in a negligible deviation between predicted and simulated results, which highlights a significant precision to predict PV self-consumption. In contrast, the ML algorithms in Linear Regression and Regression Trees do not perform accurately.

Figure 6, represents the Relative Error (RE) of predictions by different ML models. RE indicated the extent to which each predicted result is overestimated or underestimated compared to the actual results produced by simulation-based analysis.

Boosted_Trees and Bagged_Trees performed differently in terms of accuracy. Boosted_Trees algorithm mainly underestimated the results in comparison to the actual results. In contrast, Bagged_Trees always estimates the values higher than the values predicted by Boosted_Trees and in some cases higher than the actual simulated results.

The Linear Regression (LR) model's algorithms presented a lower precision compared to Regression Trees (RT), except for the coarse_tree which is found as the least precise model in this case study. SVM algorithms perform more precisely in general compared to LR algorithms except for SVM_Linear and SVM_Coarse_Gaussian models that represent an accuracy as low as LR models. The most accurate SVM model is SVM_Fine_Gaussian representing R^2 equal to 0.99.

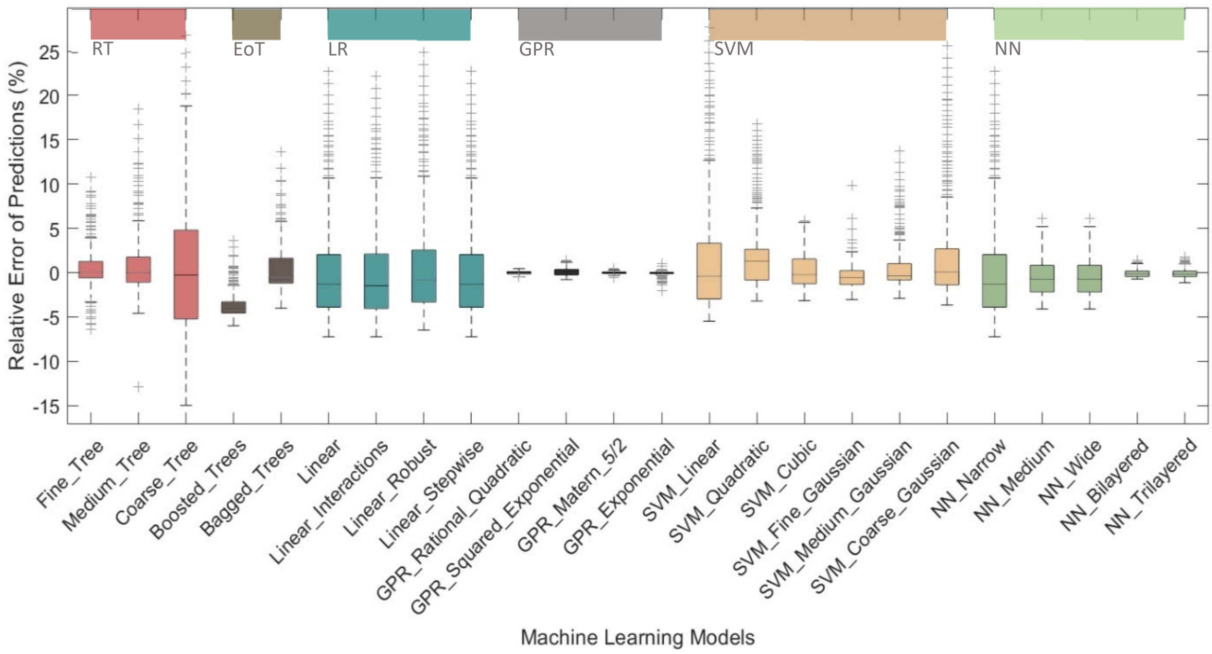

Figure 6. Relative error of predictions, RT: Regression Trees, EoT: Ensembles of Trees, LR: Linear Regression, GPR: Gaussian Process Regression, SVM: Support Vector Machines, NN: Neural Networks

The results shown in Figure 6 along with those presented in figure A1 in appendix A illustrate comparable and detailed information about the accuracy of each ML model tested in this case study. As shown in figure A1, the relative error of an overwhelming majority of results predicted by GPR models, bilayered and trilayered neural networks fall between -1% to 1% confirming the precision and the reliability of these ML models implemented in the present study.

Figure 7 shows a complicated relationship between the ML models' accuracy and training time. While some low-accurate algorithms such as Coarse_Tree or SVM_Coarse Gaussian offered a significant training speed, there are other ML models such as GPR models that provide high accuracy and training speed simultaneously. The results show accuracy and training time are not correlated in ML models. This fact highlights the importance of considering both accuracy and training speed for a smart selection of ML algorithms. The suitable ML algorithm must be chosen amongst the most accurate ones while considering the one with the lowest training time. Likewise, the prediction speed is an important performance indicator of ML algorithms, particularly for case studies with a giant number of predictions.

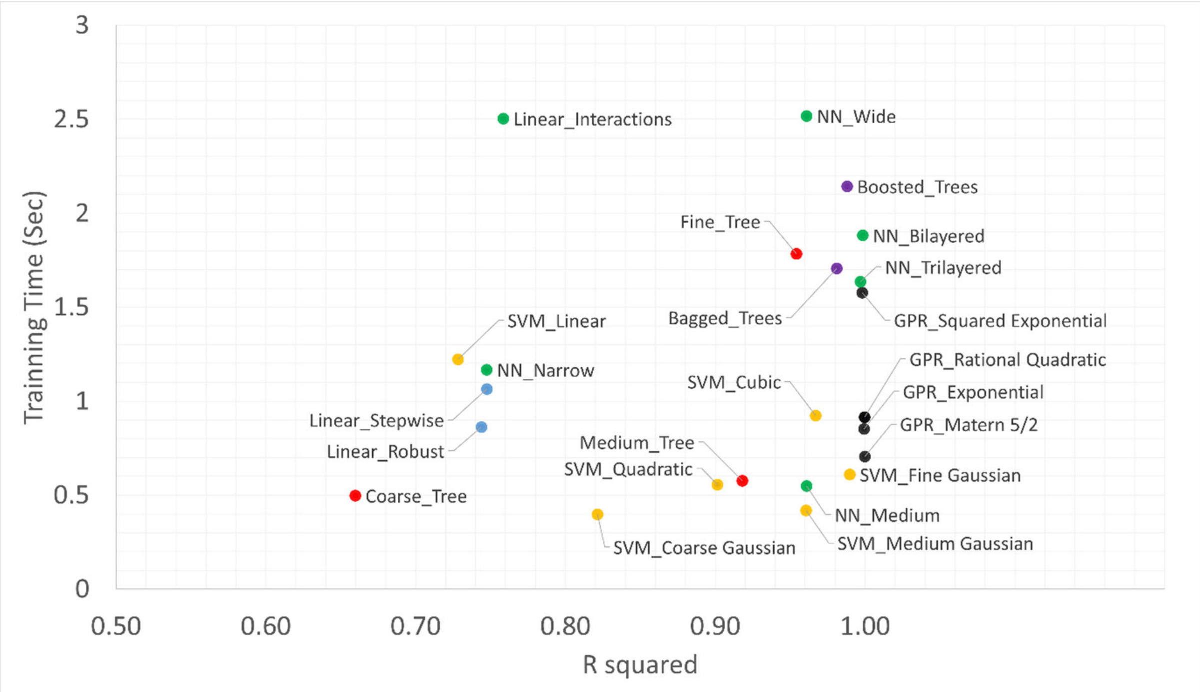

Figure 7. Training time and R squared of tested ML models
Note: Linear model is excluded due to its high training time

Figure 8 illustrates the prediction speed of all tested ML models and their R^2 as an accuracy indicator. Like training speed, some models such as Medium_Tree or SVM_Coarse Gaussian offered a high prediction speed however due to the lack of accuracy they cannot be recommended as appropriate ML models in this case study. On the other hand, most GPR and Neural Network models provided relatively high prediction speed and significant accuracy simultaneously nominating them as suitable ML algorithms for the performance prediction in this case study.

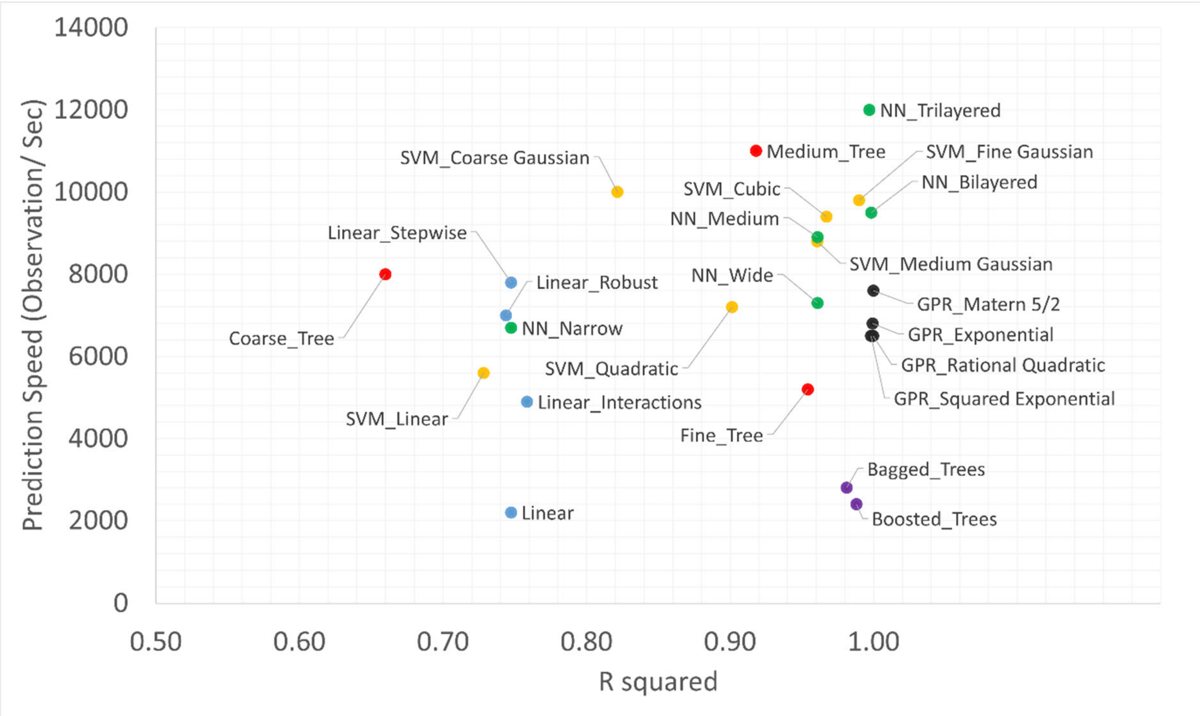

Figure 8. Prediction speed and R squared of tested ML models

Figures 7 and 8 deliver useful information on ML models concerning their accuracy, training time, and prediction speed and therefore can be utilized as a reference to select the most suitable ML algorithms for building energy performance prediction. For the case studies with a high required number of predictions, rapid forecasting becomes significantly important as a criterion to select the ML algorithms. Although the accuracy of prediction should not be compromised, the training and prediction speed must also be considered when a large number of predictions are expected.

Finally, it can be realized that amongst those ML models providing a coefficient of determination (R^2) higher than 0.99, all GPR models alongside bilayered and trilayered neural networks outperform in terms of training and prediction speed. In addition to the aforementioned ML models, SVM_Fine Gaussian, SVM_Cubic, Boosted_Trees, and Bagged_Trees are found among the top-ten most accurate ML models.

4.2.PV-self consumption of different TES scenarios

Given the results provided in the previous sections to evaluate the accuracy, training time, and prediction speed of different Machine Learning models in forecasting the PV self-consumption for multiple design scenarios of thermal energy storage, GPR_Matern 5/2 is selected in this section. The trained GPR_Matern 5/2 model in the previous section is used to predict the results of all possible combinations of TES sizes including 2601 scenarios. Figure 9 illustrates the results both for the simulated results and the predicted ones.

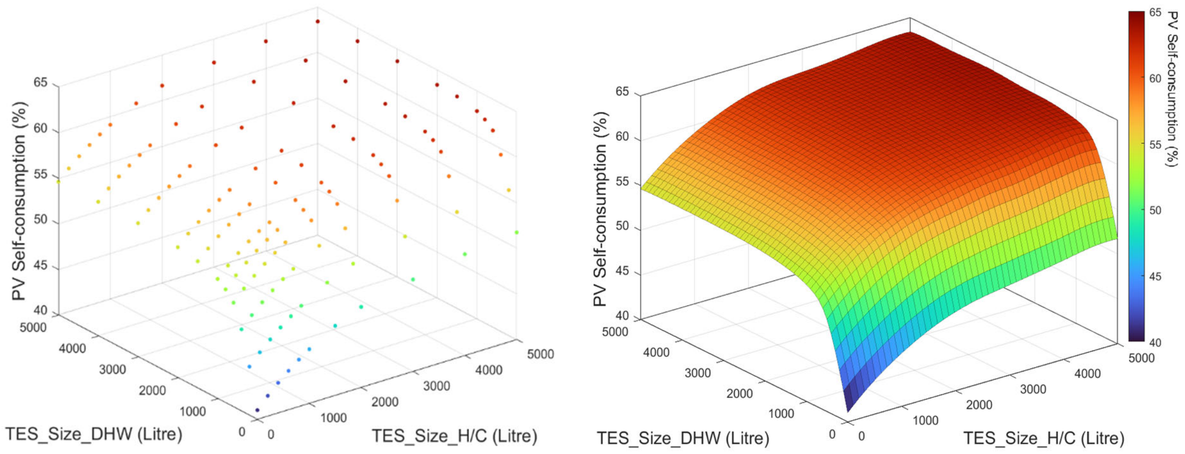

Figure 9. PV self-consumption for different TES sizes,
Left: The actual simulation-based results,
Right: Predicted results for whole design scenarios by GPR_Matern 5/2 model.

As shown in figure 9, the GPR_Matern 5/2 model produced the results for whole possible combinations of TES sizes with significant accuracy as examined in the previous section. The application of the ML model in this case study accelerated the whole design/ assessment process around 20 times faster than the simulation-based process. The results also show that the application of TES systems is effectively improving the PV self-consumption. The annual PV self-consumption can be improved from 41.03% for the base scenario (no TES installation) up to 63.4% for the maximum size of the TES tanks. The application of TES can also increase the PV self-consumption by more than 90% on certain days as elaborated in the previous section. The minimum TES size to reach at least 60% as the PV self-consumption rate is found as 1700 liters and 1200 liters for heating/cooling and DHW respectively in this case study. Increasing the TES size beyond this volume to the maximum allowed size can improve PV self-consumption by only around 3% since the excess size of TES will not be fully utilized, considering the hourly profiles of building electricity demand, PV generation, and the control logic illustrated in figure 3.

PV self-consumption also depends on the installed power of PV systems and the period of thermal energy storage which in this case study is one-day storage. Increasing the storage period to more than one day can also increase the PV self-consumption by application of any type of ESSs. This section provided the results of PV self-consumption for different TES capacities and the next section uses the same methodology for the life cycle cost optimization of TES size.

4.3. The LCC-based optimum size of TES

Improving PV self-consumption in buildings by application of TES yields economic profits by reducing the required electricity purchase from the grid while the Life Cycle Cost (LCC) of TES scenarios depends on both the initial investment of TES installation and the subsequent savings achieved by higher PV self-consumption rate. Thus, the GPR_Matern 5/2 model is trained on the economic performance of simulation-based results (test set) and implemented to forecast the LCC performance of whole TES scenarios in terms of Net Present Value (NPV) to find the LCC-based optimum size of TES in this case study. Figure 10, presents the LCC results of the different combinations of TES sizes. The GPR_matern 5/2 offered a high accuracy in predicting the NPV of TES scenarios representing an R^2 equal to 1. Nonetheless, even the application of an accurate ML-based surrogate model into a decision-making process may lead to a quasi-optimum solution instead of the actual optimal solutions if the difference among the scenarios' performance is very small. Therefore, the quasi-optimum solutions determined by ML prediction are re-evaluated using the simulation-based process to obtain the actual results and compare them to the ML-based predicted results. As shown in figure 10.b and 10.c, the optimal TES size obtained from ML predicted results is 2700 liters while the optimum size according to the actual simulation-based is 2600 liters. Nevertheless, both results are enough close to the best economic performance based on the NPV indicator and therefore could be considered the optimal solution.

More in-depth, the optimum size of TES to minimize the NPV (i.e. the discounted costs of each TES scenario including initial investment and operational energy cost over the 30 years of the building's life span) is found as 1600 liters and 1000 liters for heating/cooling and DHW respectively. The optimum TES size can yield 7.1% saving in the net present value of the building's life cycle costs with respect to the electricity tariffs, price of the components, and the macroeconomic parameters (discount and inflation rates) used in this study.

The LCC-based optimum size of TES improves the PV self-consumption rate up to 59.4% and reduces the electricity consumption by approximately 20.27% compared to the base scenario. The optimal scenario also decreases the annual electricity export and the peak power injection

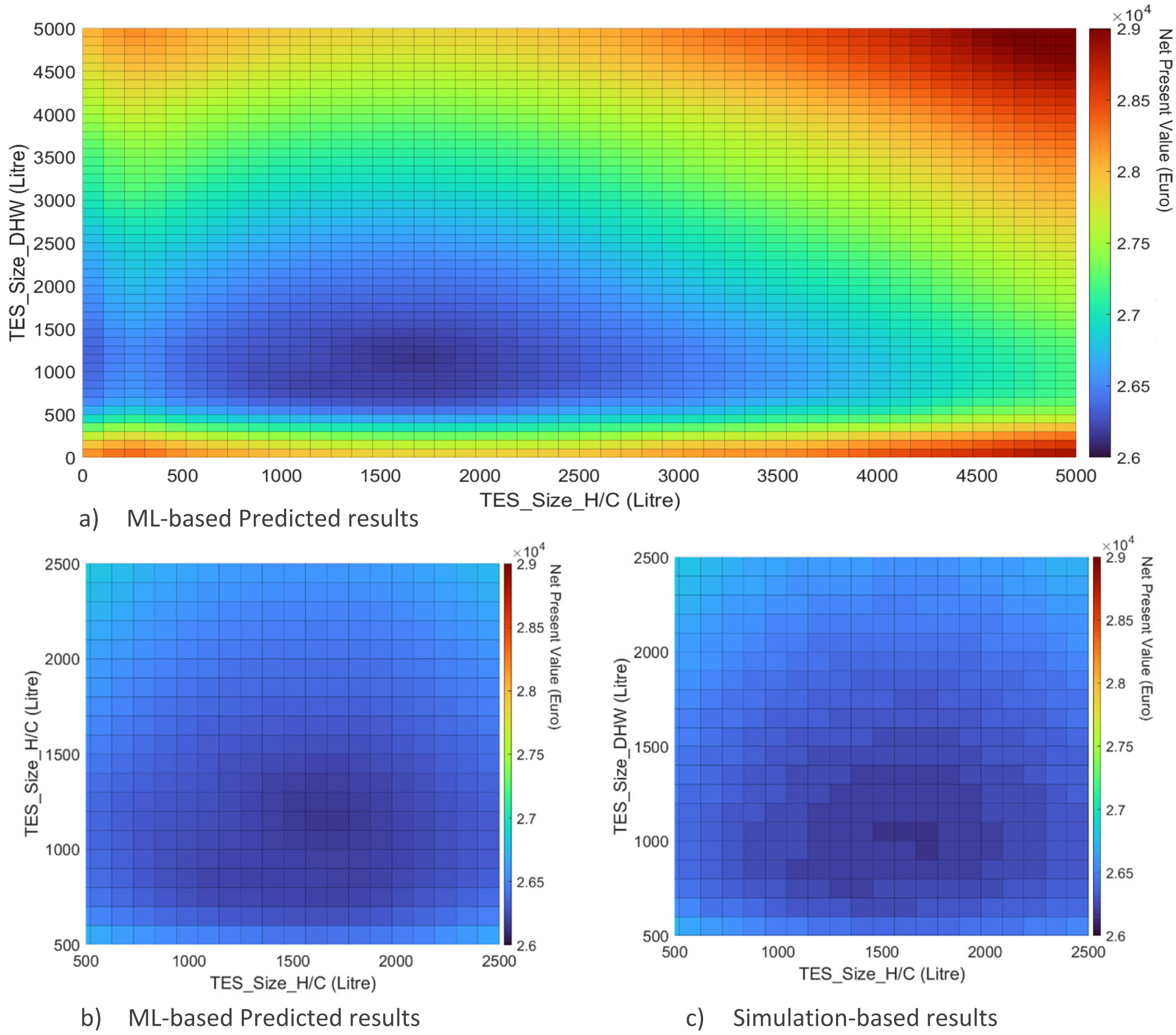

Figure 10. Life cycle cost of different TES sizes
a) LCC results of the whole design space,
b) LCC results of quasi-optimum scenarios (based on the prediction by GPR_matern 5/2 model)
c) LCC results of quasi-optimum scenarios (based on actual simulation)

to the electricity grid by 31.73% and 21.23% respectively which benefits the grid stability. The results show that although increasing TES size beyond the cost optimum scenarios will improve the self-consumption rate marginally, it will cause higher life cycle costs due to a larger initial investment. The LCC calculation is carried out based on the economic parameters provided in the previous section while considering lower installation cost for TES components, lower discount rate, and higher electricity price inflation rate over the building life span in LCC calculation can justify larger TES size as the LCC-based optimum scenario. Moreover, extending the storage period (e.g. more than one day) or the service life (e.g. 50 years) might result in economic justification for larger TES size as well, although the cost of replacing TES components should also be taken into account in case the service life of the building is assumed longer than TES technical life span.

5. Conclusion

In this paper, the precision and performance of 24 different Machine Learning models areevaluated. The accuracy, training time, and prediction speed of the ML algorithms as the main performance criteria were chosen in this evaluation. The evaluation was conducted on predicting PV self-consumption for various TES sizes of a residential building. Therefore, first, a parametric energy model of the building case study was created to analyze five percent of all possible combinations of energy storage sizes to feed and train the ML models, and then the predicted results by each model were assessed and compared to the actual simulation-based results for realizing the prediction accuracy.

The evaluation confirmed that some ML models are noticeably superior. Almost all linear Machine Learning algorithms showed lower accuracy in this case study compared to other tested models. Furthermore, in general, Support Vector Machine algorithms outperformed all Linear Regression models and offered a high precision still providing lower accuracy compared to Gaussian Regression Process and most Neural Network models. In summary and according to the R^2 of each prediction model, the GPR models including GPR_Matern 5/2, GPR_Rational_quadratic, GPR_Exponential, and GPR_Squared_exponential alongside NN_Bilayered, and NN_Trilayered belonging to the Neural Network models, SMV models including SVM_Fine_Gaussian, and SVM_Cubic, and EoT models containing Boosted_Trees and Bagged_Trees and are found as the top-ten most accurate ML models in this case study.

Although all tested ML algorithms showed significant training and prediction speed, an intelligent selection of suitable ML algorithms must also consider not only the models' accuracy but also the training and prediction speed. In this study, no linear relation is found between the accuracy and training/prediction speed which indicates that higher training/prediction speed does not necessarily lead to higher or lower accuracy and could be recommended as independent selection criteria, particularly in the cases with huge numbers of required predictions where the training and prediction speed matter.

The results provided in this paper can be used as a reference for researchers and practitioners to understand the performance variation among several existing Machine Learning models and therefore help choose the most suitable ML models for predicting the energy performance of buildings during the design process.

The application of TES as a solution to improve PV self-consumption was found to be noticeably beneficial. For instance, as the results of ML-based analysis presented, PV self-consumption in this case study can be improved from 41.03% to 62.07% by installing TES with 3000 liters and 1500 liters for heating/cooling and DHW respectively. Moreover, the optimum size of TES regarding the life cycle cost analysis in this case study is found as 1600 liters and 1000 liters for heating/cooling and DHW respectively. The LCC-based optimum size of TES yields both a high PV self-consumption rate (59.4%) and 20.27% reduction in the grid-supplied electricity consumption leading to 7.1% decrease in the total life cycle cost.

Future studies can extend this research by testing hybrid ML models, application of ML models to other energy systems' configurations, and replicating the proposed methodology in various geographic areas to assess PV self-consumption and the life cycle costs during designing the buildings and the energy systems using Machine Learning-based surrogate models.

Acknowledgment

This study and project are financially supported by EU Research and Innovation programme Horizon 2020 through number 768921 – HEART. The authors would like to thank the European Commission for enabling the funding of this project.

Appendix A: The frequency of relative errors in the predicted results of the test set

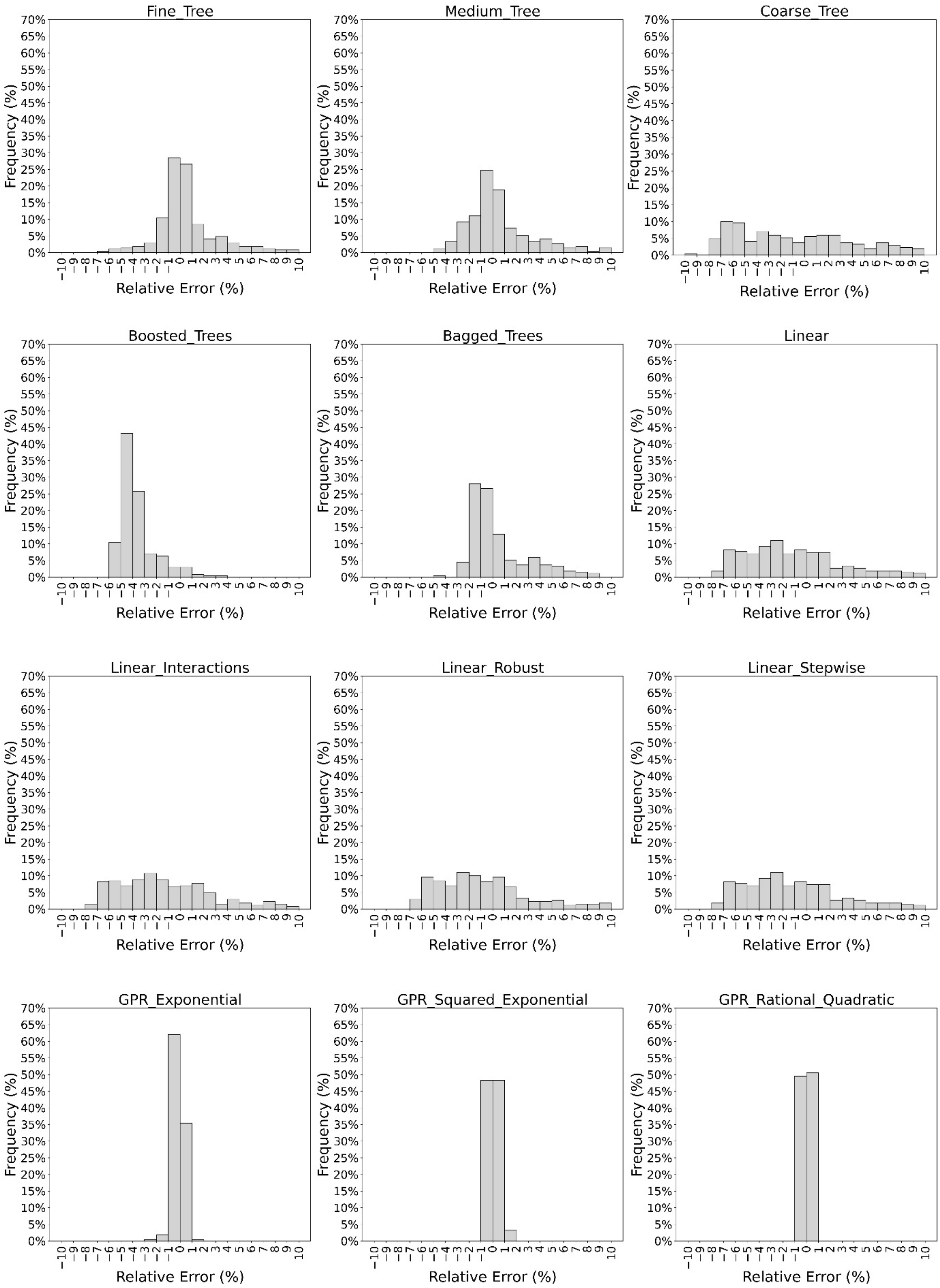

Figure A1. The frequency of relative error in each ML model's predicted results for the test set (only the relative errors between -10% to +10% are shown)

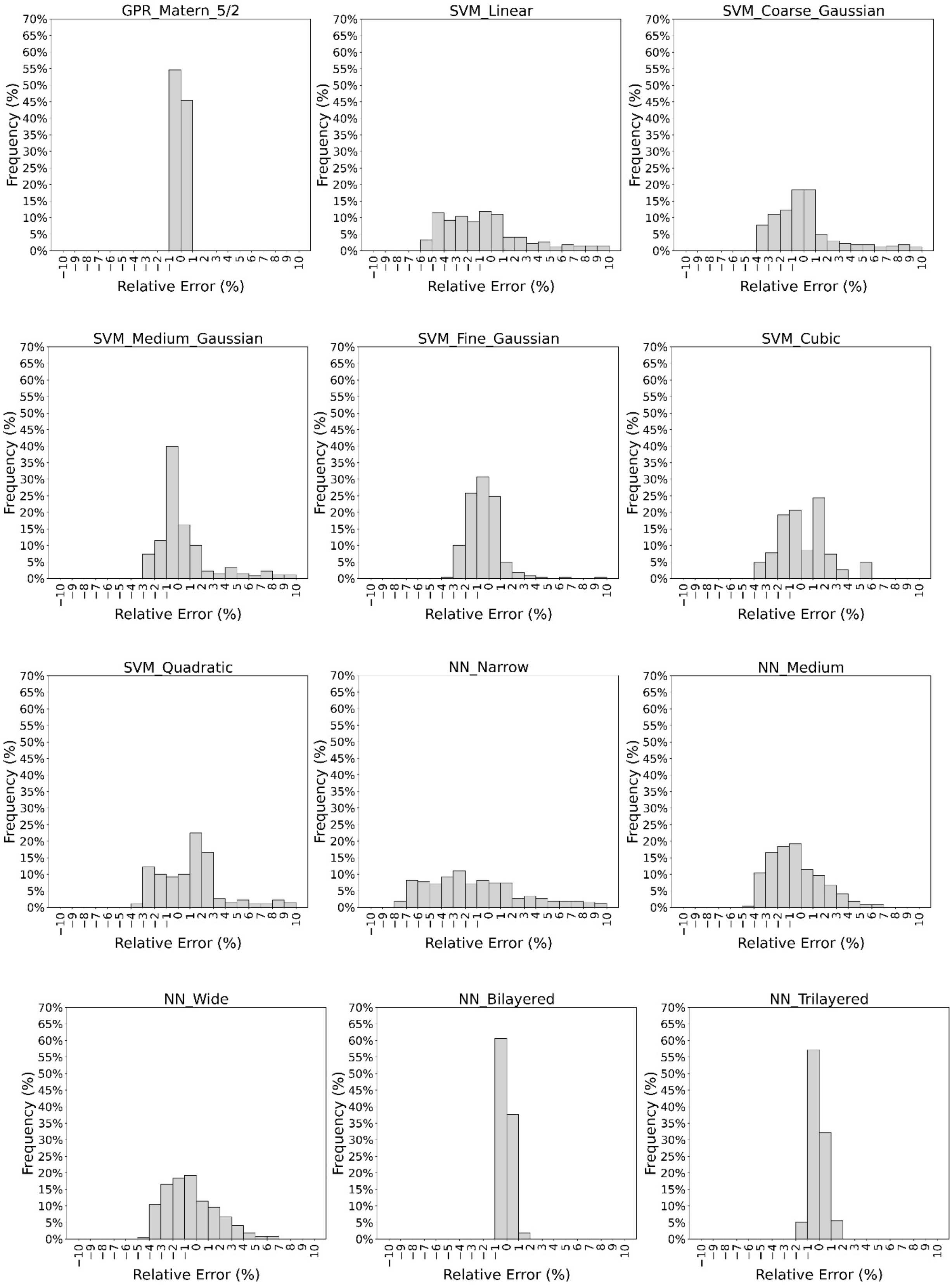

Figure A1 [continue]. The frequency of relative error in each ML model's predicted results for the test set (only the relative errors between -10% to +10% are shown)

References

A. Miglioli, C. Del Pero, F. Leonforte, & N. Aste. (2019). Load matching in residential buildings through the use of thermal energy storages. *International Conference on Clean Electrical Power (ICCEP)*, 272–279.

Abubakar Mas'ud, A. (2022). Comparison of three machine learning models for the prediction of hourly PV output power in Saudi Arabia. *Ain Shams Engineering Journal*, *13*(4). https://doi.org/10.1016/j.asej.2021.11.017

Ahmadiahangar, R., Karami, H., Husev, O., Blinov, A., Rosin, A., Jonaitis, A., & Sanjari, M. J. (2022). Analytical approach for maximizing self-consumption of nearly zero energy buildings- case study: Baltic region. *Energy*, *238*. https://doi.org/10.1016/j.energy.2021.121744

Al Dakheel, J., Del Pero, C., Aste, N., & Leonforte, F. (2020). Smart buildings features and key performance indicators: A review. In *Sustainable Cities and Society* (Vol. 61). Elsevier Ltd. https://doi.org/10.1016/j.scs.2020.102328

Alduailij, M. A., Petri, I., Rana, O., Alduailij, M. A., & Aldawood, A. S. (2021). Forecasting peak energy demand for smart buildings. *Journal of Supercomputing*, *77*(6), 6356–6380. https://doi.org/10.1007/s11227-020-03540-3

Amasyali, K., & El-Gohary, N. (2021). Machine learning for occupant-behavior-sensitive cooling energy consumption prediction in office buildings. *Renewable and Sustainable Energy Reviews*, *142*(January), 110714. https://doi.org/10.1016/j.rser.2021.110714

Amini Toosi, H., Lavagna, M., Leonforte, F., del Pero, C., & Aste, N. (2020). Life Cycle Sustainability Assessment in Building Energy Retrofitting; A Review. In *Sustainable Cities and Society* (Vol. 60, p. 102248). Elsevier Ltd. https://doi.org/10.1016/j.scs.2020.102248

Amini Toosi, H., Lavagna, M., Leonforte, F., del Pero, C., & Aste, N. (2021). *Implementing Life Cycle Sustainability Assessment in Building and Energy Retrofit Design—An Investigation into Challenges and Opportunities*. 103–136. https://doi.org/10.1007/978-981-16-4562-4_6

Amini Toosi, H., Lavagna, M., Leonforte, F., del Pero, C., & Aste, N. (2022). A novel LCSA-Machine learning based optimization model for sustainable building design-A case study of energy storage systems. *Building and Environment*, *209*. https://doi.org/10.1016/j.buildenv.2021.108656

Baniasadi, A., Habibi, D., Al-Saedi, W., Masoum, M. A. S., Das, C. K., & Mousavi, N. (2020). Optimal sizing design and operation of electrical and thermal energy storage systems in smart buildings. *Journal of Energy Storage*, *28*. https://doi.org/10.1016/j.est.2019.101186

Barthwal, M., & Rakshit, D. (2021). Artificial neural network coupled building-integrated photovoltaic thermal system for indian montane climate. *Energy Conversion and Management*, *244*. https://doi.org/10.1016/j.enconman.2021.114488

Bourhnane, S., Abid, M. R., Lghoul, R., Zine-Dine, K., Elkamoun, N., & Benhaddou, D. (2020). Machine learning for energy consumption prediction and scheduling in smart buildings. *SN Applied Sciences*, *2*(2), 1–10. https://doi.org/10.1007/s42452-020-2024-9

Braun M, Büdenbender K, Magnor D, & Jossen A. (2009). Photovoltaic self-consumption in Germany: using lithium-ion storage to increase self-consumed photovoltaic energy. *24th European Photovoltaic Solar Energy Conference (PVSEC).*

Bui, D. K., Nguyen, T. N., Ngo, T. D., & Nguyen-Xuan, H. (2020). An artificial neural network (ANN) expert system enhanced with the electromagnetism-based firefly algorithm (EFA) for predicting the energy consumption in buildings. *Energy*, *190*, 116370. https://doi.org/10.1016/j.energy.2019.116370

del Pero, C., Aste, N., & Leonforte, F. (2021). The effect of rain on photovoltaic systems. *Renewable Energy*, *179*, 1803–1814. https://doi.org/10.1016/j.renene.2021.07.130

EN_16627. (2015). *EN 16627:2015 Sustainability of construction works. Assessment of economic performance of buildings. Calculation methods*.

EUROPEAN COMMISSION. (2015). *Best practices on Renewable Energy Self-consumption*.

Fathi, S., Srinivasan, R., Fenner, A., & Fathi, S. (2020). Machine learning applications in urban building energy performance forecasting: A systematic review. *Renewable and Sustainable Energy Reviews*, *133*(August), 110287. https://doi.org/10.1016/j.rser.2020.110287

Feng, Y., Hao, W., Li, H., Cui, N., Gong, D., & Gao, L. (2020). Machine learning models to quantify and map daily global solar radiation and photovoltaic power. *Renewable and Sustainable Energy Reviews*, *118*. https://doi.org/10.1016/j.rser.2019.109393

Gallego-Castillo, C., Heleno, M., & Victoria, M. (2021). Self-consumption for energy communities in Spain: A regional analysis under the new legal framework. *Energy Policy*, *150*. https://doi.org/10.1016/j.enpol.2021.112144

HEART. (2022). The Holistic Energy and Architectural Retrofit Toolkit (HEART). In *https://heartproject.eu/*. https://heartproject.eu/

Hirschburger, R., & Weidlich, A. (2020). Profitability of photovoltaic and battery systems on municipal buildings. *Renewable Energy*, *153*, 1163–1173. https://doi.org/10.1016/j.renene.2020.02.077

Hong, T., Wang, Z., Luo, X., & Zhang, W. (2020). State-of-the-art on research and applications of machine learning in the building life cycle. *Energy & Buildings*, *212*, 109831. https://doi.org/10.1016/j.enbuild.2020.109831

Hwang, J. K., Yun, G. Y., Lee, S., Seo, H., & Santamouris, M. (2020). Using deep learning approaches with variable selection process to predict the energy performance of a heating and cooling system. *Renewable Energy*, *149*, 1227–1245. https://doi.org/10.1016/j.renene.2019.10.113

Ivanko, D., Sørensen, Å. L., & Nord, N. (2020). Selecting the model and influencing variables for DHW heat use prediction in hotels in Norway. *Energy and Buildings*, *228*, 110441. https://doi.org/10.1016/j.enbuild.2020.110441

Jia, D., Yang, L., Lv, T., Liu, W., Gao, X., & Zhou, J. (2022). Evaluation of machine learning models for predicting daily global and diffuse solar radiation under different weather/pollution conditions. *Renewable Energy*, *187*, 896–906. https://doi.org/10.1016/j.renene.2022.02.002

Jonas, B., Motuzien, V., Vilutien, T., & Indriulionis, A. (2020). Occupancy Prediction Using Differential Evolution Online Sequential Extreme Learning Machine Model. *Energies*, 1–20.

Khan, W., Walker, S., & Zeiler, W. (2022). Improved solar photovoltaic energy generation forecast using deep learning-based ensemble stacking approach. *Energy*, *240*. https://doi.org/10.1016/j.energy.2021.122812

Liu, C., Sun, B., Zhang, C., & Li, F. (2020). A hybrid prediction model for residential electricity consumption using holt-winters and extreme learning machine. *Applied Energy*, *275*(February), 115383. https://doi.org/10.1016/j.apenergy.2020.115383

López Prol, J., & Steininger, K. W. (2020). Photovoltaic self-consumption is now profitable in Spain: Effects of the new regulation on prosumers' internal rate of return. *Energy Policy*, *146*. https://doi.org/10.1016/j.enpol.2020.111793

Luthander, R., Widén, J., Nilsson, D., & Palm, J. (2015). Photovoltaic self-consumption in buildings: A review. In *Applied Energy* (Vol. 142, pp. 80–94). Elsevier Ltd. https://doi.org/10.1016/j.apenergy.2014.12.028

Maljkovic, D., & Basic, B. D. (2020). Determination of influential parameters for heat consumption in district heating systems using machine learning. *Energy*, *201*, 117585. https://doi.org/10.1016/j.energy.2020.117585

Martínez-Comesaña, M., Febrero-Garrido, L., Granada-álvarez, E., Martínez-Torres, J., & Martínez-Mariño, S. (2020). Heat loss coefficient estimation applied to existing buildings through machine learning models. *Applied Sciences (Switzerland)*, *10*(24), 1–18. https://doi.org/10.3390/app10248968

Mathworks. (2022). *Matlab Machine Learning Toolbox*.

Mawson, V. J., & Hughes, B. R. (2020). Deep learning techniques for energy forecasting and condition monitoring in the manufacturing sector. *Energy and Buildings*, *217*, 109966. https://doi.org/10.1016/j.enbuild.2020.109966

Miglioli, A., Pero, C. del, Leonforteand, F., & Aste, N. (2019). Load matching in residential buildings through the use of thermal energy storages. *International Conference on Clean Electrical Power (ICCEP)*, 272–279.

Natarajan, Y., Kannan, S., Selvaraj, C., & Mohanty, S. N. (2021). Forecasting energy generation in large photovoltaic plants using radial belief neural network. *Sustainable Computing: Informatics and Systems*, *31*. https://doi.org/10.1016/j.suscom.2021.100578

Ngarambe, J., Irakoze, A., Yun, G. Y., & Kim, G. (2020). Comparative performance of machine learning algorithms in the prediction of indoor daylight illuminances. *Sustainability (Switzerland)*, *12*(11), 1–22. https://doi.org/10.3390/su12114471

Parzinger, M., Hanfstaengl, L., Sigg, F., Spindler, U., Wellisch, U., & Wirnsberger, M. (2020). Residual analysis of predictive modelling data for automated fault detection in building's heating, ventilation and air conditioning systems. *Sustainability (Switzerland)*, *12*(17). https://doi.org/10.3390/SU12176758

Pham, A., Ngo, N., Thu, T., Truong, H., & Huynh, N. (2020). Predicting energy consumption in multiple buildings using machine learning for improving energy efficiency and sustainability. *Journal of Cleaner Production*, *260*, 121082. https://doi.org/10.1016/j.jclepro.2020.121082

Ramadhan, R. A. A., Heatubun, Y. R. J., Tan, S. F., & Lee, H. J. (2021). Comparison of physical and machine learning models for estimating solar irradiance and photovoltaic power. *Renewable Energy*, *178*, 1006–1019. https://doi.org/10.1016/j.renene.2021.06.079

Ramos, L., Colnago, M., & Casaca, W. (2022). Data-driven analysis and machine learning for energy prediction in distributed photovoltaic generation plants: A case study in Queensland, Australia. *Energy Reports*, *8*, 745–751. https://doi.org/10.1016/j.egyr.2021.11.123

Ruiz, L. G. B., Pegalajar, M. C., Molina-Solana, M., & Guo, Y. K. (2020). A case study on understanding energy consumption through prediction and visualization (VIMOEN). *Journal of Building Engineering*, *30*(February), 101315. https://doi.org/10.1016/j.jobe.2020.101315

Sadeghi, A., Sinaki, R. Y., Young, W. A., & Weckman, G. R. (2020). An intelligent model to predict energy performances of residential buildings based on deep neural networks. *Energies*, *13*(3). https://doi.org/10.3390/en13030571

Sajjad, M., Khan, S. U., Khan, N., Haq, I. U., Ullah, A., Lee, M. Y., & Baik, S. W. (2020). Towards efficient building designing: Heating and cooling load prediction via multi-output model. *Sensors (Switzerland)*, *20*(22), 1–19. https://doi.org/10.3390/s20226419

Schreiber, M., & Hochloff, P. (2013). Capacity-dependent tariffs and residential energy management for photovoltaic storage systems. *IEEE Power and Energy Society General Meeting*.

Sha, H., Moujahed, M., & Qi, D. (2021). Machine learning-based cooling load prediction and optimal control for mechanical ventilative cooling in high-rise buildings. *Energy and Buildings*, *242*, 110980. https://doi.org/10.1016/j.enbuild.2021.110980

Shivam, K., Tzou, J. C., & Wu, S. C. (2021). A multi-objective predictive energy management strategy for residential grid-connected PV-battery hybrid systems based on machine learning technique. *Energy Conversion and Management*, *237*. https://doi.org/10.1016/j.enconman.2021.114103

Shukhobodskiy, A. A., & Colantuono, G. (2020). RED WoLF: Combining a battery and thermal energy reservoirs as a hybrid storage system. *Applied Energy*, *274*. https://doi.org/10.1016/j.apenergy.2020.115209

Thygesen, R., & Karlsson, B. (2014). Simulation and analysis of a solar assisted heat pump system with two different storage types for high levels of PV electricity self-consumption. *Solar Energy*, *103*, 19–27. https://doi.org/10.1016/j.solener.2014.02.013

Vela, A., Alvarado-Uribe, J., Davila, M., Hernandez-Gress, N., & Ceballos, H. G. (2020). Estimating occupancy levels in enclosed spaces using environmental variables: A fitness gym and living room as evaluation scenarios. *Sensors (Switzerland)*, *20*(22), 1–21. https://doi.org/10.3390/s20226579

Visser, L., AlSkaif, T., & van Sark, W. (2022). Operational day-ahead solar power forecasting for aggregated PV systems with a varying spatial distribution. *Renewable Energy*, *183*, 267–282. https://doi.org/10.1016/j.renene.2021.10.102

Walker, S., Khan, W., Katic, K., Maassen, W., & Zeiler, W. (2020). Energy & Buildings Accuracy of different machine learning algorithms and added-value of predicting aggregated-level energy performance of commercial buildings. *Energy & Buildings*, *209*, 109705. https://doi.org/10.1016/j.enbuild.2019.109705

Wang, X., Sun, Y., Luo, D., & Peng, J. (2022). Comparative study of machine learning approaches for predicting short-term photovoltaic power output based on weather type classification. *Energy*, *240*. https://doi.org/10.1016/j.energy.2021.122733

Wang, Z., & Hong, T. (2020). Generating realistic building electrical load profiles through the Generative Adversarial Network (GAN). *Energy and Buildings*, *224*, 110299. https://doi.org/10.1016/j.enbuild.2020.110299

Yang, S., Wan, M. P., Chen, W., Ng, B. F., & Dubey, S. (2020). Model predictive control with adaptive machine-learning-based model for building energy efficiency and comfort optimization. *Applied Energy*, *271*(May), 115147. https://doi.org/10.1016/j.apenergy.2020.115147

Yu, B., Kim, D., Cho, H., & Mago, P. (2020). A nonlinear autoregressive with exogenous inputs artificial neural network model for building thermal load prediction. *Journal of Energy Resources Technology, Transactions of the ASME*, *142*(5), 1–9. https://doi.org/10.1115/1.4045543

Yu, H. J. J. (2021). System contributions of residential battery systems: New perspectives on PV self-consumption. *Energy Economics*, *96*. https://doi.org/10.1016/j.eneco.2021.105151

Zazoum, B. (2022). Solar photovoltaic power prediction using different machine learning methods. *Energy Reports*, *8*, 19–25. https://doi.org/10.1016/j.egyr.2021.11.183

Zekić-Sušac, M., Mitrović, S., & Has, A. (2021). Machine learning based system for managing energy efficiency of public sector as an approach towards smart cities. *International Journal of Information Management*, *58*. https://doi.org/10.1016/j.ijinfomgt.2020.102074

Zhou, Y., & Zheng, S. (2020). Machine-learning based hybrid demand-side controller for high-rise office buildings with high energy flexibilities. *Applied Energy*, *262*(December 2019), 114416. https://doi.org/10.1016/j.apenergy.2019.114416

Zhou, Y., Zhou, N., Gong, L., & Jiang, M. (2020). Prediction of photovoltaic power output based on similar day analysis, genetic algorithm and extreme learning machine. *Energy*, *204*. https://doi.org/10.1016/j.energy.2020.117894