ASSESSMENT OF LOW-ENERGY POTENTIAL OF A SCHOOL BUILDING USING OPERATION OPTIMIZATION AND SURROGATE MODELS

Mirko M. STOJILJKOVIĆ[*1], Vladan S. JOVANOVIĆ[1], Dušan J. RANĐELOVIĆ[2], Goran D. VUČKOVIĆ[1], Marko G. IGNJATOVIĆ[1]*

[1]University of Niš, Faculty of Mechanical Engineering in Niš,
Aleksandra Medvedeva 14, Niš — Crveni Krst, Serbia
[2]University of Niš, Faculty of Civil Engineering and Architecture in Niš,
Aleksandra Medvedeva 14, Niš — Crveni Krst, Serbia

[*]Corresponding author; E-mail: mirko.stojiljkovic@masfak.ni.ac.rs

The retrofit of existing buildings has a notable potential to save energy and reduce the environmental impact. This paper presents an approach to assess the primary energy saving potential related to the retrofit of a school building. The retrofit includes wall and roof insulation, fenestration replacement, installation of heat pumps with thermal storage, and usage of photovoltaic panels. The approach relies on building simulations and operation optimization to evaluate a limited number of retrofit combinations. The results are used to formulate two kinds of surrogate models based on gradient boosting: (1) classifier that finds feasible options and (2) regressor that estimates the primary energy consumption. The results show high precision of the predictive models. The F1 score of the classifiers exceeds 0.99 even for very small training samples. The most important feature for estimating feasibility is the heat pump capacity. The coefficient of determination of the regressors is close to 1 and the root mean square error is lower than 1 kWh/m^2, even for moderate sizes of the training set. The most important feature for predicting energy consumption is related to the area and orientation of photovoltaic panels.

Key words: *building simulations, gradient boosting, heat pumps, school retrofit*

1. Introduction

Buildings are responsible for a significant share of global energy consumption and environmental footprint. The retrofit of existing buildings has a notable potential for the reduction of the impact on the environment [1]. It includes the improvement of a building envelope, upgrading heating and cooling systems, installing heat pump (HP) systems, using solar energy with photovoltaic panels (PV), *etc*. During the next 15 years, Serbia should considerably improve the energy efficiency of buildings and increase the share of HPs [2].

Życzyńska *et al.* [3] assessed the impact of the retrofits of 14 public buildings in Poland and reported

37.1% lower energy consumption, due to the improvements to the heating system and insulation. A similar analysis related to schools and kindergartens showed the reduction in heat consumption of 46.8% [4]. Ranđelović *et al.* [5] analyzed various combinations of passive measures for Serbian schools using building energy simulations and estimated that the requirements can be reduced by up to 77% for heating and 79% for cooling. Papadakis and Katsaprakakis [6] compared active and passive retrofit measures for public buildings and found insulation to be of crucial importance, due to the dominance of heating requirements in energy consumption. Sánchez *et al.* [7] considered negative effects of passive measures on the cooling demand in the Mediterranean climate. Ciacci *et al.* [8] optimized retrofit measures for two Italian schools from the cost perspective and concluded that the energy-saving potential (ESP) is 35% and that it is more convenient to improve the system than the building envelope. The impact of climate change on the school retrofit is taken into account in Refs. [9, 10], while sustainability indicators are considered in Ref. [11].

For many years, the bottom-up assessment and optimization of building retrofit measures was carried out using building simulations, based on white-box models. However, these simulations are often computationally intensive and time-consuming. Recently, following the advances in machine learning (ML), simulation models have started being replaced or supplemented with their data-driven black-box surrogates.

Surrogate models or metamodels approximate the behavior of original, usually more complex and computationally intensive calculation procedures. These models rely on ML techniques and data obtained from measurements or simulations. Integrating them into optimization procedures can improve building design by reducing the computational time and increasing the exploration space [12, 13]. For example, combining a model based on an artificial neural network (ANN) — to predict the values of objective functions — with a genetic algorithm (GA) can improve the efficiency of optimizing retrofit options of buildings [14–16]. The power of the surrogate models derived from EnergyPlus [17] simulation results was demonstrated in [18–20], for example. In addition, such models can be applied to calibrate building simulations [21]. However, they still suffer from reusability and interpretability issues [22]. Thrampoulidis *et al.* [1] presented an approach that improves the scalability of surrogates to larger sets of buildings.

Although feed-forward ANNs are the most widely used for surrogate modeling, other ML approaches such as support vector regression (SVR) and tree-based methods are also very effective [22]. For example, Ref. [23] uses a random forest (RF) to classify retrofit options according to their cost-effectiveness, while Ref. [24] applies regressors based on RF, gradient boosting (GB), and extremely randomized trees to predict primary energy (PE) consumption (PEC). Shen and Pan [25] used light GB machines combined with Bayesian optimization of hyperparameters to formulate surrogate models. The models were then applied in a multi-objective optimization (MOO) procedure for the design of green buildings. In addition, they measured the contribution of individual parameters with Shapley additive explanations. Zhu *et al.* [26] presented another MOO approach to the design of green buildings using RF models as surrogates and the non-dominated sorting GA (NSGA) III. Shi and Chen [27] applied several ML regression methods to predict the energy consumption, life-cycle cost, and thermal comfort time for different retrofit options related to the building envelope of a hospital. The results were applied for retrofit MOO with NSGA III. Alaa *et al.* [28] optimized roof parameters with GA based on RF regression models, which were trained using the data obtained with building simulations.

Nguyen *et al.* [29] applied a stacking regression approach to predict energy consumption in retrofitted buildings and reported a significant improvement in predictive performance. They used linear regression

(LR) as a meta-learner. Baseline models were formulated with various boosting ensemble methods and deep ANN. The approach was applied to multiple buildings at once. Li *et al.* [30] considered PV and used a voting binary classification approach to select retrofit options. The voting ensemble performed better than the individual models. Imalka *et al.* [31] presented a life-cycle MOO of PV design with ANN and NSGA II.

Sretenović *et al.* used measured data and SVR to predict daily building heat consumption in [32], while in [33] they combined LR with ANN. Jurišević *et al.* [34] compared LR, decision tree, and ANN for predicting monthly heat consumption in kindergartens and concluded that the last performed the best. Milićević and Marinović [35] used ANN to predict PV energy output.

HPs are very important for progressing towards sustainable heating systems [36]. They enable an increase in the usage of renewable energy [37] for heating, cooling, and hot water preparation. The potential impact and feasibility of a wider adoption of HPs in Serbia were studied in Refs. [38, 39]. The exergy aspect of air-source and ground-source HPs was analyzed in Refs. [40, 41], respectively. Kossi and Rama [42] concluded that HS can significantly improve the flexibility and profitability of HP systems. Li *et al.* [43] emphasized the value of HS in a system with HP and PV.

Design and operation optimization of energy systems is important to thoughtfully estimate and fully exploit their cost and ESP. Optimization of the systems with HPs — which might include HS and PV — is often carried out with mixed integer linear programming (MILP) [44–46], metaheuristic methods such as NSGA II [43], or reinforcement learning in more complex cases [47–49]. Chaoran *et al.* [50] presented a MOO approach to a ground-source HP operation with a surrogate model based on extreme GB that substitutes numerical simulations of HP.

It seems that there is a gap in the literature related to the consideration of optimal operating regimes of energy supply systems in the context of formulating surrogate models to achieve a sustainable building design. In addition, the presented methods mostly use regression to predict energy consumption, costs, thermal comfort, *etc*. The feasibility of retrofit options, *i.e.* predicting whether an energy system will be able to cover the demand, is rarely — if at all — explored with metamodels. This paper aims to fill the gap by providing an approach for defining surrogate models that enable a joint consideration of energy system design and operation together with building envelope improvement. By combining classification and regression tasks, it enables both the identification of feasible solutions and the estimation of optimal ESP, therefore yielding a more comprehensive retrofit assessment and better insights into the improvement opportunity.

The presented approach for the assessment of ESP for school buildings through retrofit. It considers two kinds of interventions: (1) the improvement of the building envelope and (2) the installation of a new energy supply system with HPs. For a limited number of retrofit measure combinations, PE saving potential is estimated with building simulations and operation optimization. The results are used to formulate two kinds of GB surrogate models: (1) classifiers that try to find feasible combinations of retrofit measures and (2) regressors that predict PEC for feasible solutions. The approach can be integrated into a retrofit optimization procedure that uses a metaheuristic method such as GA or NSGA II/III.

The paper also illustrates the application of the presented approach to the case of a school building, using the Serbian PE conversion factors and the typical meteorological year for the city of Niš, Serbia. Having an efficient and accurate methodology for ESP assessment is particularly important for schools due to their regular occupancy, energy efficiency and PV convenience, social importance, and public funding opportunities.

2. Methodology

The methodology for assessing the low-energy potential of school building retrofit measures presented in this paper considers the envelope improvements, in particular the insulation of the exterior walls and roof, as well as the replacement of fenestration. In addition, it assumes the application of HPs for heating and cooling, with optional HS and cool storage (CS), and examines the optimal operation of such energy systems. Finally, installing PV panels is explored as well.

Since both envelope and supply-side improvements are considered, two problems are addressed:

1. Checking feasibility of retrofit options, *i.e.* if HPs with HS and CS can satisfy the building demand.
2. Calculating the specific PEC for feasible options.

The predictor variables that characterize each retrofit option, *i.e.* each combination of measures, are:

- x_1 — the thickness of external wall insulation, in $[\mathrm{cm}]$,
- x_2 — the thickness of roof insulation, in $[\mathrm{cm}]$,
- x_3 — the overall heat transfer coefficient of glazing, in $[\mathrm{W m^{-2}\,K^{-1}}]$,
- x_4 — the total solar energy transmittance of glazing,
- x_5 — the option related to the installation of PV panels or their area, in $[\mathrm{m^2}]$,
- x_6 — the number or capacity of HPs, in $[\mathrm{kW}]$,
- x_7 — the size of HS,
- x_8 — the size of CS.

This paper applies building simulations and the values of variables x_1–x_5 to calculate the electricity, heating, and cooling demand, as well as the PV output for various retrofit options. They act as the inputs of the optimization problem, which yields the PEC for given design values x_6–x_8, if the problem is feasible.

The optimization results are used to formulate separate surrogate models for the feasibility check and PEC estimation. Both kinds of models are based on GB, which is one of the most effective non-linear supervised ML methods and can be trained in a relatively short amount of time. GB is convenient for handling nonlinearities, avoiding overfitting, and providing good performance with small and large datasets, as well as across a number of hyperparameter combinations. Various kinds of GB machines exhibited excellent prediction performance in building-related applications [51, 52]. Compared to ANNs, they often offer a better balance of accuracy and efficiency [53], have lower sensitivity to the choice of hyperparameters, and tend to require smaller amounts of data.

2.1. Operation optimization

This methodology aims to assess the energy saving potential of retrofit options and assumes that the energy supply system will operate in the regime that yields the minimal annual PEC. Operation parameters are determined with MILP. Optimization is performed for the entire typical year at once, although the approaches with a moving horizon [54] and typical days [55] may be applied as well.

The continuous non-negative decision variables of the MILP model, for each time step τ, are:

- $\dot{W}^{\tau}_{\text{e,grid,in}}$ and $\dot{W}^{\tau}_{\text{e,grid,out}}$ — the electricity import from and export to the grid, respectively, in $[\mathrm{kW}]$,

- $\dot{Q}^{\tau}_{\text{h,HP}}$ and $\dot{Q}^{\tau}_{\text{c,HP}}$ — the heating and cooling output for each HP, respectively, in $[\text{kW}]$,
- $\dot{Q}^{\tau}_{\text{h,HS,in}}$ and $\dot{Q}^{\tau}_{\text{c,CS,in}}$ — the heat added to HS and CS, respectively, in $[\text{kW}]$,
- $\dot{Q}^{\tau}_{\text{h,HS,out}}$ and $\dot{Q}^{\tau}_{\text{c,CS,out}}$ — the heat retrieved from HS and CS, respectively, in $[\text{kW}]$,
- t^{τ}_{HS} and t^{τ}_{CS} — the average temperatures of the storage mediums in HS and CS, respectively, in $^{\circ}\text{C}$.

There is one binary variable $\delta^{\tau}_{\text{HP}}$ for each HP and each time step τ that decides whether an HP is operating in the heating ($\delta^{\tau}_{\text{HP}} = 1$) or cooling mode ($\delta^{\tau}_{\text{HP}} = 0$).

Equation (1) defines the objective function — minimization of annual PEC, E_{P}, in $[\text{kWh}]$:

$$\min \quad E_{\text{P}} = \varphi_{\text{e}} \sum_{\tau} \left(\dot{W}^{\tau}_{\text{e,grid,in}} - \dot{W}^{\tau}_{\text{e,grid,out}} \right) \Delta\tau \tag{1}$$

where $\varphi_{\text{e}} = 2.5$ is the PE conversion factor for electricity [56], $\Delta\tau$ is the time step, and τ is the time index.

The amount of exported electricity is limited with the PV production during each time step with eq. (2), while the outputs of all HPs are constrained by their capacities and operating modes with eqs. (3) and (4):

$$\dot{W}^{\tau}_{\text{e,grid,out}} \leq \dot{W}^{\tau}_{\text{e,PV}} \qquad \forall\tau \tag{2}$$

$$\dot{Q}^{\tau}_{\text{h,HP}} \leq \delta^{\tau}_{\text{HP}} \dot{Q}^{\tau}_{\text{h,HP,cap}} \qquad \forall\tau \ \ \forall\,\text{HP} \tag{3}$$

$$\dot{Q}^{\tau}_{\text{c,HP}} \leq \left(1 - \delta^{\tau}_{\text{HP}}\right) \dot{Q}^{\tau}_{\text{c,HP,cap}} \qquad \forall\tau \ \ \forall\,\text{HP} \tag{4}$$

where $\dot{W}^{\tau}_{\text{e,PV}}$ is the electricity generated by the PV system for each time step τ, in $[\text{kW}]$, while $\dot{Q}^{\tau}_{\text{h,HP,cap}}$ and $\dot{Q}^{\tau}_{\text{c,HP,cap}}$ are the heating and cooling capacities for all HPs, for each time step τ, also in $[\text{kW}]$. PV electricity is obtained with building simulations, while HP capacities are obtained from manufacturers and vary according to the temperatures of the exterior air and heating or cooling water.

Equations (5)–(7) ensure the satisfaction of the electrical, heating and cooling demand, respectively:

$$\dot{W}^{\tau}_{\text{e,grid,in}} - \dot{W}^{\tau}_{\text{e,grid,out}} - \sum_{\text{HP}} \left(\frac{\dot{Q}^{\tau}_{\text{h,HP}}}{\text{COP}^{\tau}_{\text{HP}}} + \frac{\dot{Q}^{\tau}_{\text{c,HP}}}{\text{EER}^{\tau}_{\text{HP}}} \right) + \dot{W}^{\tau}_{\text{e,PV}} = \dot{W}^{\tau}_{\text{e,demand}} \qquad \forall\tau \tag{5}$$

$$\sum_{\text{HP}} \dot{Q}^{\tau}_{\text{h,HP}} + \dot{Q}^{\tau}_{\text{h,HS,out}} - \dot{Q}^{\tau}_{\text{h,HS,in}} = \dot{Q}^{\tau}_{\text{h,demand}} \qquad \forall\tau \tag{6}$$

$$\sum_{\text{HP}} \dot{Q}^{\tau}_{\text{c,HP}} + \dot{Q}^{\tau}_{\text{c,CS,in}} - \dot{Q}^{\tau}_{\text{c,CS,out}} = \dot{Q}^{\tau}_{\text{c,demand}} \qquad \forall\tau \tag{7}$$

where $\dot{W}^{\tau}_{\text{e,demand}}$, $\dot{Q}^{\tau}_{\text{h,demand}}$, and $\dot{Q}^{\tau}_{\text{c,demand}}$ are the electricity, heating, and cooling demand, for each time step τ, respectively, in $[\text{kW}]$, while $\text{COP}^{\tau}_{\text{HP}}$ and $\text{EER}^{\tau}_{\text{HP}}$ are the coefficient of performance (COP) and energy efficiency ratio (EER) for all HPs, for each time step τ. The demands are calculated with building simulations, while COP and EER are obtained from manufacturers and also vary with the air and water temperatures.

For HS and CS, the temperatures of the storage mediums are constrained with predefined lower and upper bounds, t^{τ}_{lb} and t^{τ}_{ub}, in eq. (8), while eq. (9) represents the energy balance during the time step τ [57]:

$$t^{\tau}_{\text{lb}} \leq t^{\tau} \leq t^{\tau}_{\text{ub}} \qquad \forall\tau \tag{8}$$

$$t^{\tau} = t^{\tau-1} \exp\left(-\frac{UA}{mc}\Delta\tau\right) + \left(t^{\tau}_{0} + \frac{\eta^{\tau}_{\text{in}}\dot{Q}^{\tau}_{\text{in}} - \eta^{\tau}_{\text{out}}\dot{Q}^{\tau}_{\text{out}}}{UA}\right)\left(1 - \exp\left(-\frac{UA}{mc}\Delta\tau\right)\right) \qquad \forall\tau \tag{9}$$

where t_0 is the ambient temperature, in $[^\circ\mathrm{C}]$, U is the overall heat transfer coefficient of the storage envelope, in $[\mathrm{kWm^{-2}K^{-1}}]$, A is the storage envelope area, in $[\mathrm{m^2}]$, m is the storage medium mass, in [kg], c is the medium specific heat capacity, in $[\mathrm{kWhkg^{-1}K^{-1}}]$, while η_{in} and η_{out} account for the charging and discharging losses.

2.2. Feasibility prediction

A retrofit option is considered feasible when an energy supply system can satisfy a building's demand for electricity, heating, and cooling. The feasibility criterion can depend on the design approach. One way is to use the demand values obtained with a building simulation as inputs and execute the optimization of energy system operating parameters. If the optimization problem is feasible, so is the retrofit option, and vice versa. However, it is enough to use only the demand for the design days and obtain just the first feasible solution from the optimization procedure, which verifies the feasibility [54, 55].

After determining the feasibility for a limited number of retrofit options using simulations and optimization, it is possible to formulate a surrogate model that predicts whether an option is feasible. This is a binary classification problem. The predictors are x_1–x_8. The target is 1 if the option is feasible and 0 otherwise.

When training the model, it is possible to have an imbalanced data set, *i.e.* that the counts of feasible and infeasible solutions differ significantly. The imbalance is handled by assigning higher weights to the observations of the minority class and lower weights to the majority class. The weights are inversely proportional to the square root of the class frequencies.

The performance of the classification is evaluated using the metrics suitable for imbalanced data sets, *i.e.* the precision, p, recall, r, and their harmonic mean, F_1 score, as defined in eq. (10):

$$p = \frac{n_{\mathrm{tp}}}{n_{\mathrm{tp}} + n_{\mathrm{fp}}}, \quad r = \frac{n_{\mathrm{tp}}}{n_{\mathrm{tp}} + n_{\mathrm{fn}}}, \quad F_1 = \frac{2pr}{p + r} \tag{10}$$

where n_{tp}, n_{fp}, and n_{fn} are the numbers of true positive, false positive, and false negative predictions.

2.3. Primary energy prediction

Predicting PEC is a regression problem. The predictors are again x_1–x_8. The predicted variable is the specific PEC, in $[\mathrm{kWhm^{-2}}]$, obtained by dividing E_{P} with the building floor area.

The applied evaluation metrics are: the dimensionless coefficient of determination, R^2, as defined in eq. (11), as well as the root mean square error (RMSE), mean absolute error (MAE), and median absolute error (MedAE), expressed in $[\mathrm{kWhm^{-2}}]$, as given in eq. (12):

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \tag{11}$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}, \quad \mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i|, \quad \mathrm{MedAE} = \operatorname*{median}_{i=1}^{n}(|y_i - \hat{y}_i|) \tag{12}$$

where n is the number of observations, y_i is the true value of the observation i, $i = 1, \ldots, n$, $\hat{y}_i$ is the predicted value of the observation i, and $\bar{y}$ is the arithmetic mean of all values y_i, *i.e.* $\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$.

2.4. Model formulation

Formulating a surrogate model for both the classification and the regression problem starts with a random selection of a limited number of retrofit options, *i.e.* combinations of retrofit measures. Each retrofit option is one observation. Then, the following procedure, which is illustrated in fig. 1, is applied:

1. Determine the values of the predicted variables for each selected observation using building simulations and system optimization. If only feasibility is predicted, design-day simulations and first-solution optimization are sufficient.
2. Split all observations into the training and test subsets.
3. Train and validate regression GB models for PEC prediction with the k-fold cross-validation (CV) using the training set, considering only feasible retrofit options.
4. If the regressor predictive performance obtained with CV is satisfactory, formulate and train the final model. Use the test set to evaluate it. Proceed to the assessment of the feasibility models. Otherwise, randomly select additional retrofit options and repeat the procedure.
5. Train and validate classification (feasibility) GB models with the CV, using observations.
6. If the classifier performance is satisfactory, train and test the final model. Otherwise, randomly select additional retrofit options and repeat the procedure only for feasibility.

The methodology is implemented in the Python programming language by the first author of the paper, using the Gurobi Optimizer [58] and Scikit-learn library [59].

3. Results and discussion

The presented approach is applied to the case of a two-story primary school building located in the urban part of the city of Niš, Serbia. The building is 63 years old and has the floor area of $4613\ \mathrm{m}^2$. The area of the pitched roof is $3381\ \mathrm{m}^2$. The exterior uninsulated masonry brick walls occupy $3793\ \mathrm{m}^2$ and the fenestration $1267\ \mathrm{m}^2$. The school is connected to the district heating system, while split systems are used for cooling. The annual specific heat consumption is around $150\ \mathrm{kWh/m}^2$. The school operates in two shifts, from 07:00 to 20:00 hours, from Monday to Friday, with around 1000 persons per day. The peak occupancy is from 08:00

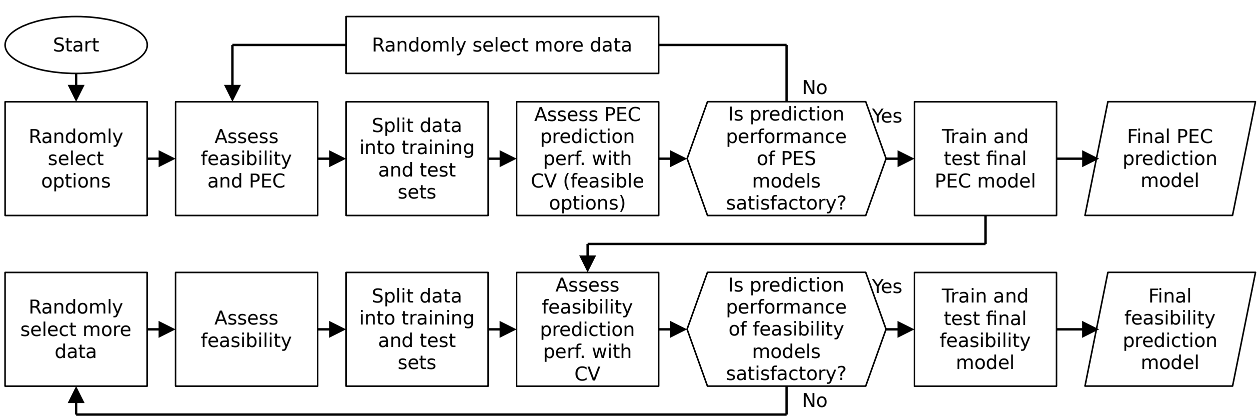

Figure 1. Workflow diagram for model formulation

to 11:30. A classroom is typically occupied by $20 - 25$ students and one teacher. The school year lasts from the beginning of September to mid June, with the three-week break in January. It also operates during the last two weeks of June and the last two weeks of August with reduced capacity.

The analyzed retrofit measures are:

- 11 options for the thickness of external wall insulation made of mineral wool, ranging from 0 to $30\ \mathrm{cm}$,
- 11 options for the thickness of roof insulation made of mineral wool, ranging from 0 to $30\ \mathrm{cm}$,
- 14 options for new fenestration with 10 distinct values of the overall heat transfer coefficient and five values of the total solar energy transmittance,
- five options for PV panels, ranging from 0 to $1300\ \mathrm{m}^2$, where option 0 refers to the case without PV, option 1 includes PV on the south side, option 2 puts PV on the south and east side, option 3 includes the south, east, and west side, while option 4 uses the total available space,
- five options for HPs with the heating capacities ranging from 336 to $672\ \mathrm{kW}$ at the external air temperature of $-15°\mathrm{C}$,
- six options of water HS, ranging from 0 (no storage) to $48\ \mathrm{m}^3$,
- six options of water CS, ranging from 0 (no storage) to $48\ \mathrm{m}^3$.

Multiplying the option counts for all variables yields 1524600 retrofit combinations in total.

The electrical, heating, and cooling demand, as well as the electrical output of PV panels are obtained using building simulation in the EnergyPlus software (version 23) and the typical meteorological year. The simulation results have hourly resolution. The absolute difference between the measured annual heat consumption and the results of simulations formulated according to the current state is less than 10%. The operation optimization is conducted with the Gurobi Optimizer (version 11). The size of the training sample is chosen from the learning curves to satisfy the conditions $F_1 \geq 0.995$ and $\mathrm{RMSE} \leq 1$. The test sample has the equal size as the training sample.

The classification of the retrofit combinations, *i.e.* observations according to feasibility is performed with the GB classifier from the Scikit-learn library (version 1.6). Around 90% of the observations have the target value of 1, which means that the data is highly imbalanced.

Figure 2 illustrates the improvement of the classification model performance with the number of sampled observations. The light areas represent the mean value ± 1 standard deviation. For very small sample sizes, the models tend to overfit. However, even for the case of 1000 observations, *i.e.* less than 0.07% of the total number of combinations, the results are already satisfactory. The means of the precision, recall, and F_1 score for the validation data exceed 0.99. The values of all evaluation metrics approach 1 as the sample grows.

An example classification model trained with 10000 observations and tested with another 10000 observations yields the following result for the test set:

- 977 correctly classified infeasible combinations,
- 9009 correctly classified feasible combinations,
- 11 incorrectly classified infeasible combinations,
- 3 incorrectly classified feasible combinations.

The infeasible solutions wrongly classified as feasible could be identified and dropped in a later stage of the investigation, when the promising options are examined in greater depth. However, the incorrectly classified feasible solutions would probably be removed from any further analysis and lost.

Figure 2. Learning curve for the classification task — finding feasible options

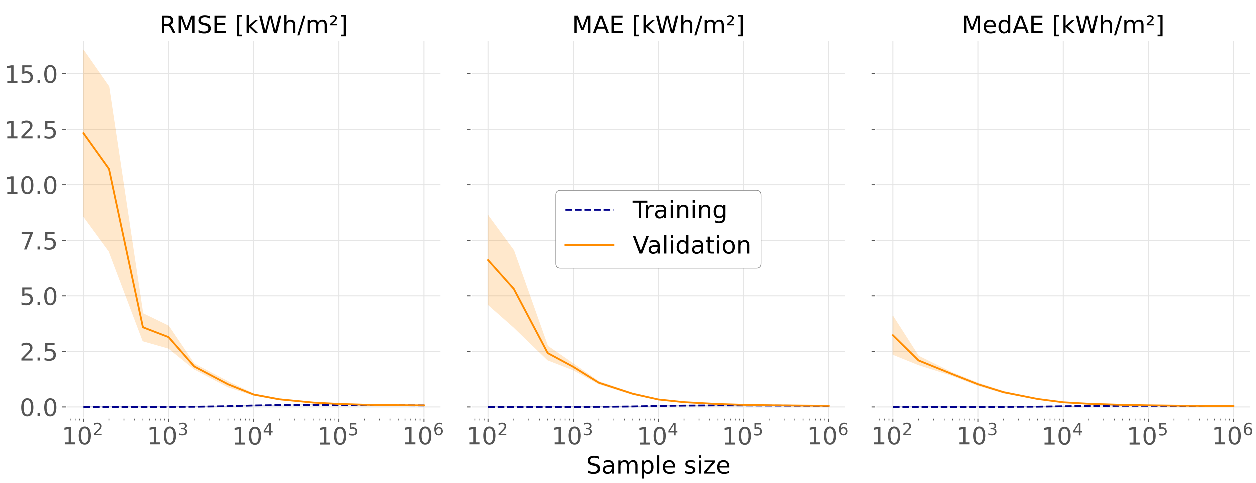

Figure 3. Learning curve for the regression task — estimating specific PEC

GB — as well as other decision-tree-based methods — can provide the information on the feature importance. It indicates how useful a feature is for defining a particular model. The above-mentioned classification model indicates that the most important variable for predicting feasibility is HP capacity (38%), followed by the thickness of wall and roof insulation.

The area under the receiver operating characteristic curve is very close to 1, so this curve is omitted.

The prediction of the specific PEC for feasible retrofit combinations is conducted with the GB regressor from the Scikit-learn library.

Figure 3 shows the decrease in RMSE, MAE, and MedAE with the increase in the sample size. Again, there seems to be some overfit for very small sizes. With 10000 observations, the means of all errors are below $1\ \mathrm{kWh/m^2}$ and further approach zero with the growing sample. The values of the coefficient of determination are close to 1, so they are not visualized.

Figure 4 analyzes the performance of an example regression model trained with 10000 observations and tested with another 10000 observations. The leftmost plot almost corresponds to the straight line of the $45°$ slope, indicating that the predicted and actual values are very close. The scatter diagram in the middle

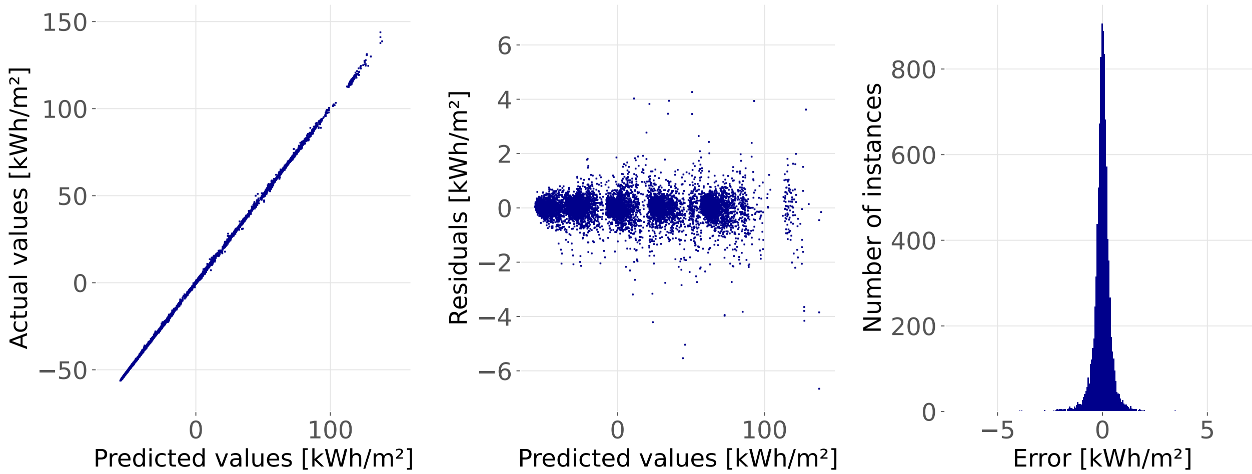

Figure 4. Predicted PEC, actual PEC, and errors for the model trained with 10000 data points

illustrates that most of the residuals have the absolute value below $2\ \mathrm{kWh/m^2}$, with a very few exceptions. Actually, 99.86% of all absolute residuals are lower than 2, and 98.49% are lower than $1\ \mathrm{kWh/m^2}$. The rightmost histogram shows symmetric error distribution.

The most important feature for PES prediction is the PV area and orientation, followed by the insulation thickness of the roofs and walls.

The experiments indicate significant time saving potential due to the use of surrogate models. They are conducted on a laptop computer with an Intel(R) Core(TM) i5-10210U processor, which has four physical cores and supports eight threads. The operating system is Ubuntu 24.04 and the Python version is 3.12.

A single full building simulation, executed on a single core, in a multiprocessing setting, lasted on average approximately 323 s, while a design day simulation took around 28 s. It is worth mentioning that the simulation time heavily depends on the level of details considered. The time required for running one MILP procedure varies significantly and is around 43 s on average, when executed on a single core. However the duration depends on the problem formulation. For example, if heating and cooling seasons are separated, the binary variables can be predefined and the problem is reduced to a much simpler linear programming formulation, which needs only 3 s. On the contrary, more complex problems, such as the ones described in [46], might take several hours to complete. The assessment of all possible combinations requires $11 \cdot 11 \cdot 14 \cdot 5 = 8470$ simulations and 1524600 optimization runs, which would last approximately 10.6 days when executed on eight cores in parallel, with the simplest optimization problem. Proportionally, 20000 observations need 3.33 h.

Obtaining learning curves for 10000 observations, for both classification and regression problems, lasts around 151 s with GB machines. Training and testing the final models with 20000 observations in total, together with predicting all remaining combinations might last up to 20 s. Some GB variants such as extreme GB and histogram GB might be significantly faster. Therefore, formulating surrogate models and full predictions occurs in under 3 min, which is less than the time needed for a single building simulation. The full process, including simulations and optimization, takes less than 3.5 h.

Having the validated surrogate models, it is possible to quickly make predictions for all combinations of retrofit measures and conduct analyses or run a metaheuristic optimization algorithm that uses the predictions

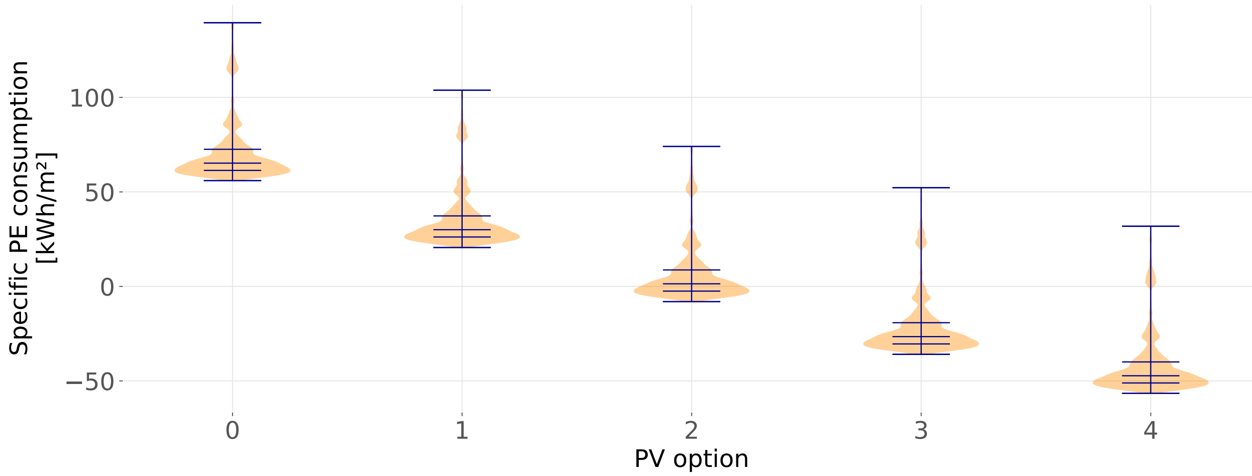

Figure 5. Impact of PV choice on the specific PEC

to evaluate objective functions.

The analysis of the considered improvements indicates that the most significant measure in terms of PEC is the installation of PV. It can reduce the specific PEC at best to $27\ \mathrm{kWh/m^2}$. The combination of PV and other active measures, *i.e.* HPs and storage, can achieve $25\ \mathrm{kWh/m^2}$. The second most effective measure is the insulation of exterior walls, which can yield the specific PEC of $87\ \mathrm{kWh/m^2}$ when HPs without storage are used for heating and cooling. All considered passive measures, *i.e.* insulation and new fenestration, can decrease PEC to $57\ \mathrm{kWh/m^2}$. Neither active nor passive measures can achieve zero PEC solely, but when combined, they can yield the specific PEC below $-50\ \mathrm{kWh/m^2}$.

Figure 5 shows the impact of the chosen PV option on the values of the specific PEC. Without PV, the specific PEC is always above $55\ \mathrm{kWh/m^2}$. For option 1, minimal PEC drops to approximately $20\ \mathrm{kWh/m^2}$. Option 2 is included in some combinations that generate more energy than it is consumed. Actually around 42% of all combinations with this option have a negative value of the specific PEC. Options 3 and 4 are related dominantly to the retrofit combinations with negative PEC values.

The middle PV option 2 is further explored from the wall insulation point of view, as shown in fig. 6. First, it can be noted that the specific PEC drops quickly with the rise of the insulation thickness for lower thickness values and much more slowly later, which is expected and is a well known consequence of the mathematical expression for the overall heat transfer coefficient for a multi-layer wall. The other important observation is that PEC can have a negative value for the thickness above $6\ \mathrm{cm}$. However, for all thickness values, there is still a considerable number of retrofit options with positive PEC.

From the practical point of view, the surrogate models presented in this paper might be applied as a part of an analysis or optimization procedure. They can provide a tool that assesses retrofit measure combinations rapidly yet accurately, but also enables the tightening of the search space and allows researchers to focus on deeper exploration of promising solutions. For example, in an economic analysis or cost optimization that focuses only on the solutions with negative PEC (*e.g.* exploring zero-energy potential), one could eliminate PV options 0 and 1 because they never satisfy this requirement. The case without wall insulation can be discarded for the same reason. One could also notice that the values of wall insulation thickness above $20\ \mathrm{cm}$ have a very low additional contribution to PE savings, so they might be neglected as well. These

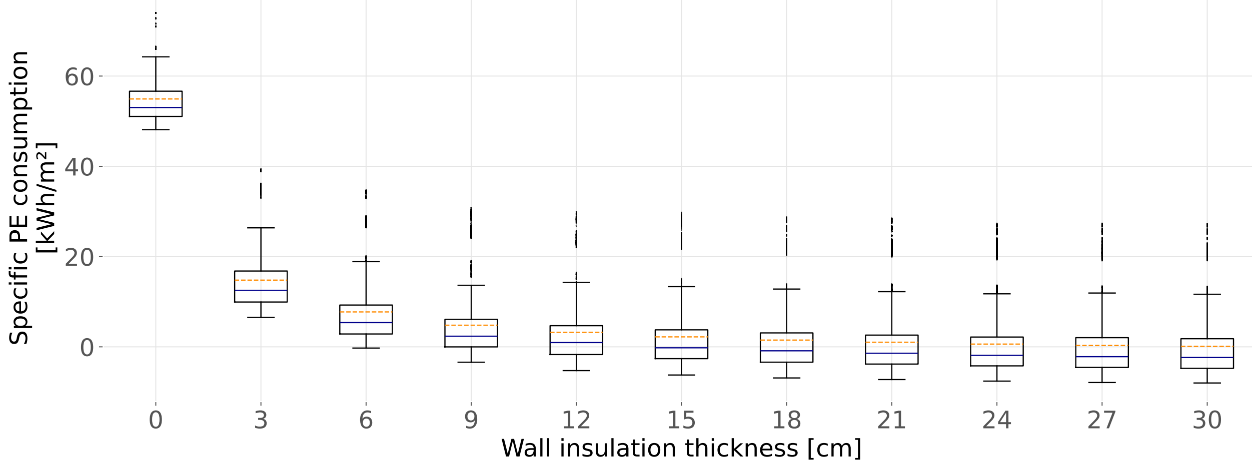

Figure 6. Impact of wall insulation thickness on the specific PEC for the PV option 2

considerations can reduce the number of explored solutions by 67.3%.

Black-box models with low computational requirements and high predictive performance of energy consumption are especially important for energy planning at the neighborhood, district, or city level. They can be used to aggregate energy consumption data for multiple buildings and help decision makers to estimate potential savings, plan the development of infrastructure or decarbonization measures, ensure the compliance of building performance with the existing or future codes, identify the buildings and measures of the highest priority for retrofit, *etc*. In addition, such models can facilitate simulating multiple retrofit scenarios or joint cost optimization of multiple buildings of the same kind. This is particularly convenient in the case of schools, which have regular occupancy patterns, large potential for energy savings, PV installation, and net-zero-energy performance, as well as high priority in public funding. Since geographical information systems (GIS) can be combined with building simulation and optimization to optimize retrofitting [60], there is a potential to couple GIS and ML building surrogate models in such applications.

Although surrogate models can be formulated for various energy systems, building types, and climates, applying a model to conditions not considered during training is generally not an acceptable approach. In order to extend the application to different cases, one should either: (1) rerun simulations and optimization, with the new weather and other inputs, and then retrain a model with appropriate data, (2) use multiple buildings to formulate a model with a different structure, which includes a wider range of input features, such as climate parameters or building type, or (3) formulate multiple models, one for each considered case.

4. Conclusion

This paper presents and applies an approach for the assessment of energy-saving potential for school buildings through retrofit based on building simulations, operation optimization, and machine learning surrogate models. It examines the improvement of the building envelope and the installation of a new energy system based on heat pumps, with optional thermal storage and photovoltaic panels.

The simulation results act as the inputs of the operation optimization process that calculates minimal primary energy consumption. Primary energy is determined for a limited number of retrofit measure combinations, each of which represents one observation. The predictive models based on gradient boosting are formulated using these observations. These models can separate feasible retrofit combinations from infeasible ones and assess the primary energy consumption.

For the examined case of a school building located in the city of Niš, Serbia, both models were validated and exhibited a very good performance, indicating their applicability, even with a relatively small number of training observations *e.g.* 1000 or 10000. The classification models that predict the feasibility of retrofit options have the average precision, recall, and F_1 score greater than 0.99 for the training sets with 1000 or more observations. The regression models that predict the specific primary energy consumption have the coefficient of determination close to 1 and the average root mean square error below $1\ \mathrm{kWh/m^2}$ for the sample size of 10000 or greater.

Such surrogate models are important because they can assess many retrofit measure combinations accurately, in a short time and with relatively modest computing resources. This allows the tightening of the search space and focusing on exploration of the promising areas, which can be particularly important for energy planning.

In the future, this approach could be applied to multi-objective optimization of retrofit options, possibly in combination with a metaheuristic method.

Further work should also explore the applicability of other machine learning methods, as well as the potential to improve prediction performance in the cases of small training sets, either by optimizing hyperparameters or involving hybrid approaches such as voting or stacking. In addition, using explainable artificial intelligence methods — *e.g.* local interpretable model-agnostic explanations or Shapley additive explanations — might provide additional insights into the obtained results.

Acknowledgment

This research was financially supported by the Ministry of Science, Technological Development and Innovation of the Republic of Serbia (Contract No. 451-03-137/2025-03/200109).

References

[1] Thrampoulidis, E. *et al.*, Approximating optimal building retrofit solutions for large-scale retrofit analysis, *Applied Energy 333* (2023), 120566,

[2] ****, Energy Sector Development Strategy of the Republic of Serbia up to 2040 with Projections up to 2050, *Official Gazette of the Republic of Serbia*, No. 94 (2024) (In Serbian)

[3] Życzyńska, A. *et al.*, Improving the Energy Performance of Public Buildings Equipped with Individual Gas Boilers Due to Thermal Retrofitting, *Energies 14* (2021), No. 6, 1565,

[4] Życzyńska, A. *et al.*, Energy Effects of Retrofitting the Educational Facilities Located in South-Eastern Poland, *Energies 13* (2020), No. 10, 2449,

[5] Ranđelović, D. *et al.*, Investigation of a passive design approach for a building facility: a case study, *Energy Sources, Part A: Recovery, Utilization, and Environmental Effects 47* (2021), No. 1., pp. 8890–8908,

[6] Papadakis, N. and Katsaprakakis, D. A., A Review of Energy Efficiency Interventions in Public Buildings, *Energies 16* (2023), No. 17, 6329,

[7] Sánchez, E. C. *et al.*, Architectural and environmental strategies towards a cost optimal deep energy retrofit for mediterranean public high schools, *Energy Reports 9* (2023), pp. 6434–6448,

[8] Ciacci, C. *et al.*, Evaluation of the Cost-Optimal Method Applied to Existing Schools Considering PV System Optimization, *Energies 15* (2022), No. 2, 611,

[9] Baglivo, C. *et al.*, Dynamic Evaluation of the Effects of Climate Change on the Energy Renovation of a School in a Mediterranean Climate, *Sustainability 13* (2021), No. 11, 6375,

[10] Campagna, L. M. *et al.*, Life cycle cost optimization for schools energy retrofit under climate change: Methodological approach and analyses in five different climates, *Energy and Buildings 335* (2025), 115561,

[11] Jovanović, M. P. *et al.*, Assessing the sustainability of Serbian school buildings by analyse and synthesis parameters under information deficiency method, *Thermal Science 22* (2018), Suppl. 4, pp. S1271–S1283,

[12] Cruz, A. S. *et al.*, Multi-objective optimization based on surrogate models for sustainable building design: A systematic literature review, *Building and Environment 266* (2024), 112147,

[13] Prina, M. G. *et al.*, Machine learning as a surrogate model for EnergyPLAN: Speeding up energy system optimization at the country level, *Energy 307* (2024), 132735,

[14] Asadi, E. *et al.*, Multi-objective optimization for building retrofit: A model using genetic algorithm and artificial neural network and an application, *Energy and Buildings 81* (2014), pp. 444–456,

[15] Sharif, S. A., and Hammad, A., Developing surrogate ANN for selecting near-optimal building energy renovation methods considering energy consumption, LCC and LCA, *Journal of Building Engineering 25* (2019), 100790,

[16] Alexakis, K. *et al.*, Genetic algorithm-based multi-objective optimisation for energy-efficient building retrofitting: A systematic review, *Energy & Buildings 328* (2025), 115216,

[17] ****, EnergyPlus, Energy Simulation Software,

[18] Edwards, R. E. *et al.*, Constructing large scale surrogate models from big data and artificial intelligence, *Applied Energy 202* (2017), pp. 685–699,

[19] Lawrence, C. R. *et al.*, Application of surrogate modelling to improve the thermal performance of single-family homes through archetype development, *Energy & Buildings 237* (2021), 110812,

[20] Bamdad, K. *et al.*, Building energy optimization using surrogate model and active sampling, *Journal of Building Performance Simulation 13* (2020), No. 6, pp. 760–776,

[21] Herbinger, F. *et al.*, Building energy model calibration using a surrogate neural network, *Energy & Buildings 289* (2023), 113057,

[22] Manmatharasan, P. *et al.*, AI-driven design optimization for sustainable buildings: A systematic review, *Energy & Buildings 332* (2025), 115440,

[23] Stojiljković *et al.*, Classification of retrofit measures for residential buildings according to the global cost, *Thermal Science 25* (2021), No. 4A, pp. 2677–2689,

[24] Stojiljković *et al.*, Predicting primary energy savings of building retrofit measures with decision-tree-based ensemble methods, *Facta Universitatis, Series: Working and Living Environmental Protection 17* (2020), No. 3, pp. 151–162,

[25] Shen, Y. and Pan, Y., BIM-supported automatic energy performance analysis for green building design using explainable machine learning and multi-objective optimization, *Applied Energy 333* (2023), 120575,

[26] Zhu, Y. *et al.*, Application of hybrid machine learning algorithm in multi-objective optimization of green building energy efficiency, *Energy 316* (2025), 133581,

[27] Shi, Y. and Chen, P, Energy retrofitting of hospital buildings considering climate change: An approach integrating automated machine learning with NSGA-III for multi-objective optimization, *Energy & Buildings 319*, (2024) 114571,

[28] Alaa, H. *et al.*, Metaheuristic optimization of roof designs to enhance energy performance and thermal comfort using parametrization and machine learning, *Solar Energy 286* (2025), 113186,

[29] Nguyen, A. T. *et al.*, Meta learning regression framework for energy consumption prediction in retrofitted buildings: A case study of South Korea, *Journal of Building Engineering 96* (2024), 110403,

[30] Li, K. *et al.*, Improving building retrofit Decision-Making by integrating passive and BIPV techniques with ensemble model, *Energy & Buildings 323* (2024), 114727,

[31] Imalka, S. T. *et al.*, Machine learning driven building integrated photovoltaic (BIPV) envelope design optimization, *Energy & Buildings 324* (2024), 114882,

[32] Sretenović, A. A. *et al.*, Support vector machine for the prediction of heating energy use, *Thermal Science 22* (2018), Suppl. 4, pp. S1171–S1181,

[33] Sretenović, A. A. *et al.*, Hybrid artificial intelligence model for prediction of heating energy use, *Thermal Science 26* (2022), No. 1B, pp. 705–716,

[34] Jurišević, N. M. *et al.*, Assessment of predictive models for the estimation of heat consumption in kindergartens, *Thermal Science 26* (2022), No. 1B, pp. 503–516,

[35] Milićević, M. M. and Marinović, B. R., Machine learning methods in forecasting Solar photovoltaic energy production, *Thermal Science 28* (2024), No. 1B, pp. 479–488,

[36] IEA, The Future of Heat Pumps, International Energy Agency, Paris, France, 2022,

[37] Adebayo, P. *et al.*, Development, modeling, and optimization of ground source heat pump systems for cold climates: A comprehensive review, *Energy and Buildings 320* (2024) 114646,

[38] Topić Božič, J. *et al.*, Life cycle assessment of energy green transition goals in Slovenia and Serbia. Heat Pump Example, *Thermal Science 28* (2024), No. 6A, pp. 4709–4721,

[39] Jokić, A. I. *et al.*, Feasibility of closed loop ground source heat pump for residential heating and cooling applications in Serbia, *Thermal Science 24* (2020), No. 2A, pp. 965–976,

[40] Vučković, G. D. *et al.*, Air-source heat pump performance comparison in different real operational conditions based on advanced exergy and exergoeconomic approach, *Thermal Science 25* (2021), No. 3A, pp. 1849–1866,

[41] Aryanfar, Y. *et al.*, A competitive study of a geothermal heat pump equipped with an intermediate economizer for R134a and R513a working fluids, *Thermal Science 27* (2023), No. 6B, pp. 5025–5038,

[42] Kossi, P., Rama, M., Improving the accuracy of heat pump feasibility assessment, *Thermal Science 28* (2024), No. 5B, pp. 4381–4394,

[43] Li, B. *et al.*, Economy and energy flexibility optimization of the photovoltaic heat pump, system with thermal energy storage, *Journal of Energy Storage 100* (2024) 113526,

[44] Krützfeldt H. *et al.*, MILP design optimization of heat pump systems in German residential buildings,

Energy and Buildings 249 (2021) 111204,
[45] Aguilera, J. J. *et al.*, Operation optimization in large-scale heat pump systems: A scheduling framework integrating digital twin modelling, demand forecasting, and MILP, *Applied Energy 376* (2024) 124259,
[46] Stojiljković, M. M. *et al.*, Cost-optimal operation of hybrid heat pump systems with progressive electricity tariffs, *Thermal Science OnLine-First* (2025),
[47] Kang, H. *et al.*, Multi-objective sizing and real-time scheduling of battery energy storage in energy-sharing community based on reinforcement learning, *Renewable and Sustainable Energy Reviews, 185* (2023), 113655,
[48] Kang, H. *et al.*, Reinforcement learning-based optimal scheduling model of battery energy storage system at the building level, *Renewable and Sustainable Energy Reviews, 190* (2024), Part A, 114054,
[49] Dou, Z. *et al.*, Innovative energy solutions: Evaluating reinforcement learning algorithms for battery storage optimization in residential settings, *Process Safety and Environmental Protection, 191* (2024), Part B, pp. 2203–2221,
[50] Chaoran, W. *et al.*, Operational strategy optimization of an existing ground source heat pump (GSHP) system using an XGBoost surrogate model, *Energy & Buildings 318* (2024), 114444,
[51] Ali, U. *et al.*, Urban building energy performance prediction and retrofit analysis using data-driven machine learning approach, *Energy and Buildings 303* (2024), 113768,
[52] Gnekpe, C. *et al.*, Energy Performance of Building Refurbishments: Predictive and Prescriptive AI-based Machine Learning Approaches, *Journal of Business Research 183* (2024), 114821,
[53] Asamoah, P. B. *et al.*, Evaluating the performance of machine learning models for energy load prediction in residential HVAC systems, *Energy and Buildings 334* (2025), 115517,
[54] Stojiljković, M. M. *et al.*, Greenhouse gases emission assessment in residential sector through buildings simulations and operation optimization, *Energy 92* (2015), Part 3, pp. 420–434,
[55] Stojiljković, M. M., Bi-level multi-objective fuzzy design optimization of energy supply systems aided by problem-specific heuristics, *Energy 137* (2017), pp. 1231–1251,
[56] ****, Rulebook on Energy Efficiency of Buildings, *Official Gazette of the Republic of Serbia*, No. 61 (2011) (In Serbian)
[57] ****, EnergyPlus Version 23.2.0 Documentation, Engineering Reference, 2023
[58] ****, Gurobi Optimizer Reference Manual, Gurobi Optimization, LLC, 2024,
[59] Pedregosa, F. *et al.*, Scikit-learn: Machine Learning in Python, *Journal of Machine Learning Research 12* (2011), No. 85, pp. 2825–2830
[60] Ascione, F. *et al.*, Knowledge and energy retrofitting of neighborhoods and districts. A comprehensive approach coupling geographical information systems, building simulations and optimization engines, *Energy Conversion and Management 230* (2021), 113786,

Paper submitted: 27.02.2025
Paper revised: 30.04.2025
Paper accepted: 04.05.2025