BRNO UNIVERSITY OF TECHNOLOGY

Real-time RSET Estimation Based on Simulation Dataset Using Machine Learning: A Complex Geometry Case Study

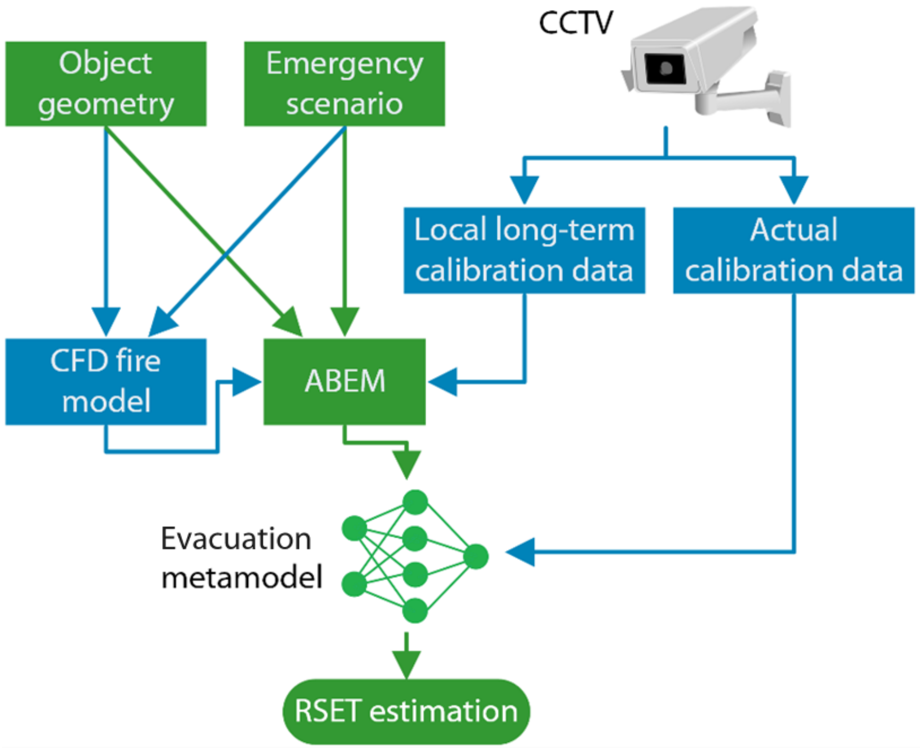
Ondřej Uhlík

E-mail: ondrej.uhlik@vut.cz

Institute of Computer Science, Faculty of Civil Engineering

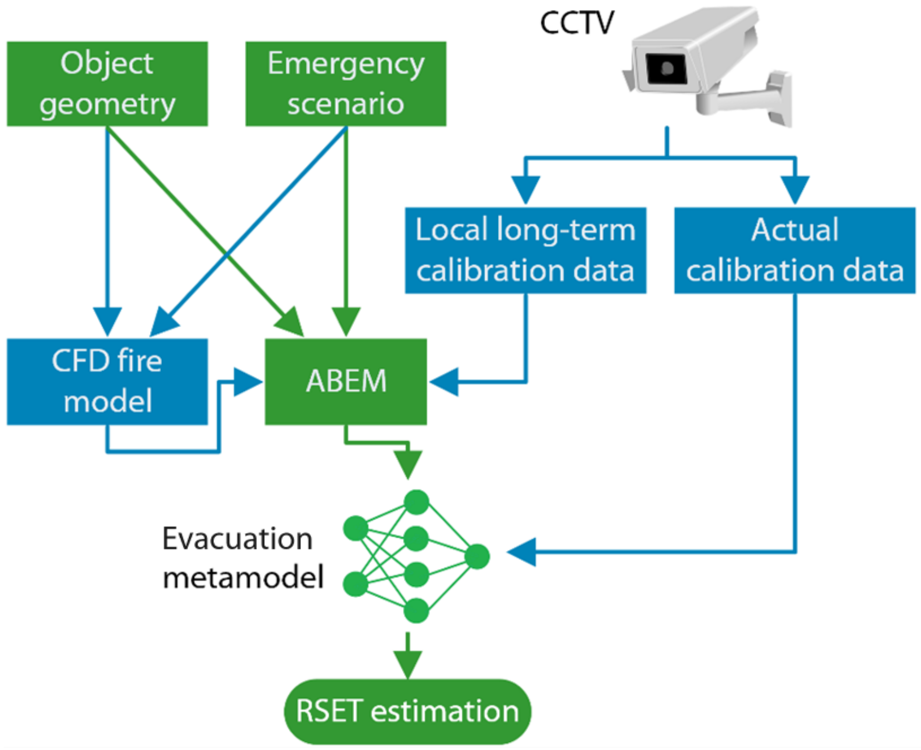
DO WE NEED REAL-TIME ESTIMATES?

- The current global security situation requires a **systematic increase in the resilience of buildings** with low security and high concentrations of occupants (soft targets).
- Evacuation models are primarily used to prepare for selected scenarios or to test a building during the design phase.
- Agent-based evacuation models (ABEM) are currently the most commonly used tools for microscopic simulation of building evacuations.
- ABEM simulations are **not suitable for real-time estimates** due to their complexity and computational demands.



AIM OF THIS WORK

- Machine learning (ML) models can be trained on a set of simulations and subsequently used to estimate simulation outputs in real-time.
- The aim of this article is to introduce and test a method for constructing the metamodel using a **case study to estimate RSET** as a key evacuation process indicator.
- 9 machine learning models were compared.

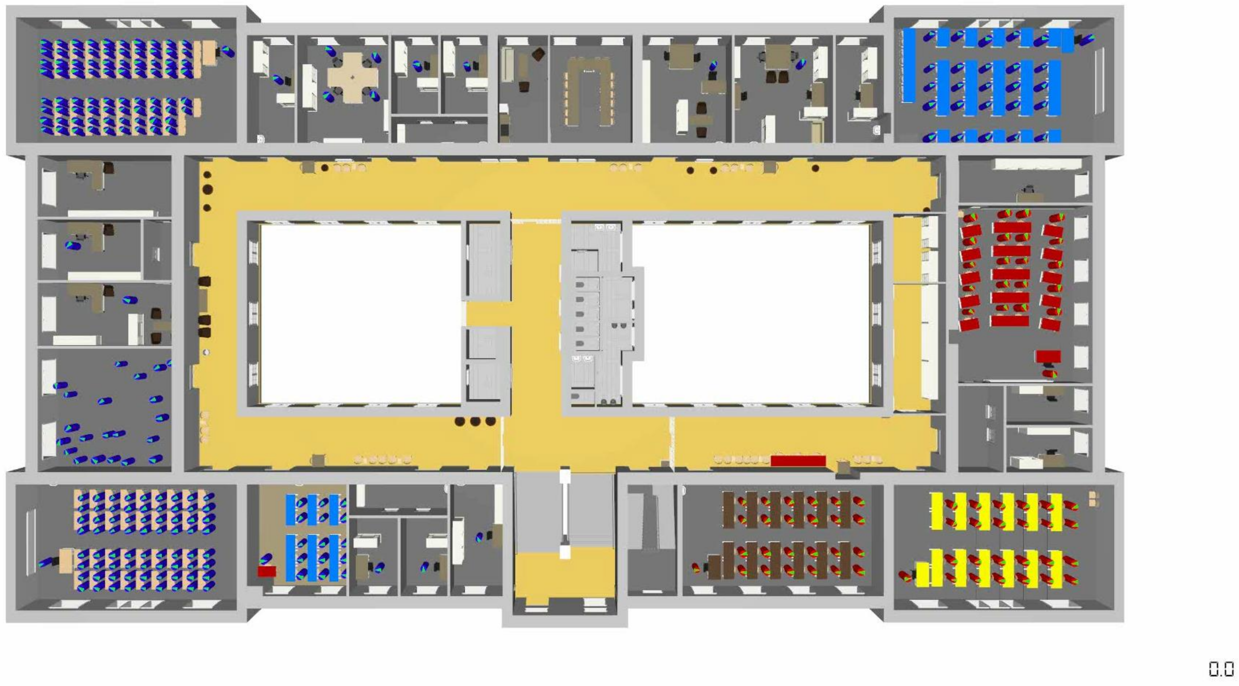


METHODS

Simulator	Pathfinder (version 2023.3.1206)
Simulation dataset generation	LPτ sequence for generation of 100 combination of number of occupants in sectors and mean walking speed (simulation dataset variables)
Simulation performing	Monte Carlo method with 20 runs for each simulation
ML models	Ordinary Least Square, Polynomial Model, k-NN, Random Forest, Gradient Boost, Extreme Gradient Boost, Support Vector Regression, Gaussian Process, Artificial Neural Network
Evaluation metrics	Coefficient of determination (R^2), Maximum Error (ME), Mean Absolute Error (MAE)
ML techniques	Nested Cross validation (outer 5-fold, inner 5-fold) with 20 runs, hyperparameter tuning (randomized search with 300 runs), Min-Max scaling, SHAP analysis for feature contribution examination

EVACUATION MODEL DESCRIPTION

- 4th floor of an university building with 300 occupants and 1 exit in the 3rd floor.
- Hallways are equiped with safety grilles.
- Floor is divided into 2 sectors.
- Log-normal distribution of pre-evacuation time: μ = 2,972 s, σ = 1,015 s, min: 5 s, max: 120 s.
- Variability of RSET 100–280 s.



DATASETS

Simulation dataset

Variable	Interval
Number of occupants in sector A	0–229
Number of occupants in sector B	0–71
Mean walking speed	0,9–1,4 $m.s^{-1}$

ML dataset

Input variable	Mathematic form
Number of Occupants (N_{occ})	$N_{occ} = \sum_{i=1}^{n} i$
Minimum distance of occupants to exits (ExD_{min})	$ExD_{max} = \frac{1}{S} \sum_{i=1}^{S} \min(\lVert \vec{x}_k - \vec{y}_j \rVert)$
Mean distance of occupants to exits (ExD_{mean})	$ExD_{mean} = \frac{1}{S \cdot N_{occ} \cdot E} \sum_{i=1}^{S} \sum_{j=1}^{n} \sum_{k=1}^{E} \lVert \vec{x}_k - \vec{y}_j \rVert$

DATASETS

ML dataset

Input variable	Mathematic form
Maximum distance of occupants to exits (ExD_{max})	$ExD_{max} = \frac{1}{S}\sum_{i=1}^{S} \max(\lVert \vec{x}_k - \vec{y}_j \rVert)$
Minimum distance between occupants ($OccD_{min}$)	$OccD_{min} = \frac{1}{S}\sum_{i=1}^{S} Q_{0,05}(\lVert \vec{y}_j - \vec{y} \rVert)$
Mean distance between occupants ($OccD_{mean}$)	$OccD_{mean} = \frac{1}{S \cdot N_{occ}}\sum_{i=1}^{S}\sum_{j=1}^{n} \lVert \vec{y}_j - \vec{y} \rVert$
Mean walking speed (V_{mean})	$V_{mean} = \frac{1}{N_{occ}}\sum_{j=1}^{n} v_j$
Output variable	**Mathematic form**
Mean required safe egress time ($RSET_{mean}$)	$RSET_{mean} = \frac{1}{S}\sum_{s=1}^{S} RSET_s$

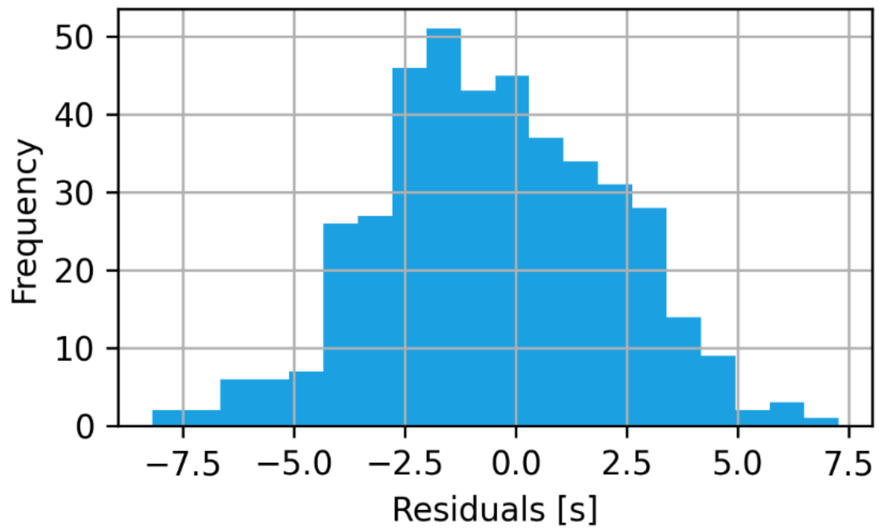
RESULTS OF NESTED CROSS VALIDATION

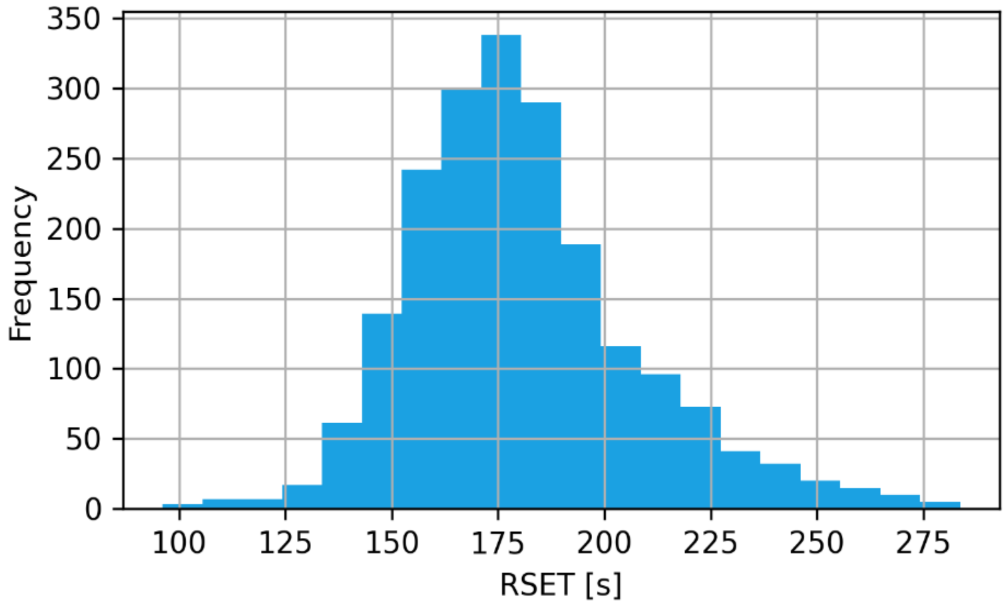
- Nested cross-validation was repeated 20 times for statistical analysis.
- Mean values and standard deviations are presented.

ML	R^2		ME		MAE	
	Train	Test	Train	Test	Train	Test
OLS	0,85 (±0,01)	0,80 (±0,08)	37,67 (±4,68)	26,67 (±9,61)	6,81 (±0,39)	7,63 (±1,30)
POLY	0,91 (±0,07)	-0,34 (±2,31)	22,16 (±9,89)	≫50 (±≫50)	5,32 (±1,69)	12,42 (±5,45)
KNN	0,93 (±0,02)	0,87 (±0,06)	23,91 (±5,77)	20,87 (±7,98)	4,75 (±0,46)	6,15 (±1,28)
RF	0,98 (±0,01)	0,87 (±0,05)	13,41 (±4,36)	22,47 (±6,44)	2,39 (±0,23)	6,05 (±1,51)
GB	1,00 (±0,00)	0,92 (±0,04)	1,44 (±1,54)	15,88 (±4,32)	0,45 (±0,40)	4,80 (±1,05)
XGB	1,00 (±0,01)	0,91 (±0,04)	4,00 (±4,44)	17,29 (±5,68)	0,82 (±0,66)	5,04 (±1,13)
SVR	0,98 (±0,01)	0,86 (±0,23)	10,51 (±2,39)	25,32 (±20,74)	2,13 (±0,46)	4,60 (±1,63)
GP	1,00 (±0,00)	0,92 (±0,04)	1,21 (±2,91)	20,17 (±10,54)	0,16 (±0,37)	4,46 (±1,24)
ANN	**0,99 (±0,01)**	**0,94 (±0,05)**	**8,80 (±1,85)**	**12,77 (±7,14)**	**2,17 (±0,46)**	**3,81 (±0,75)**

RESIDUALS PLOT

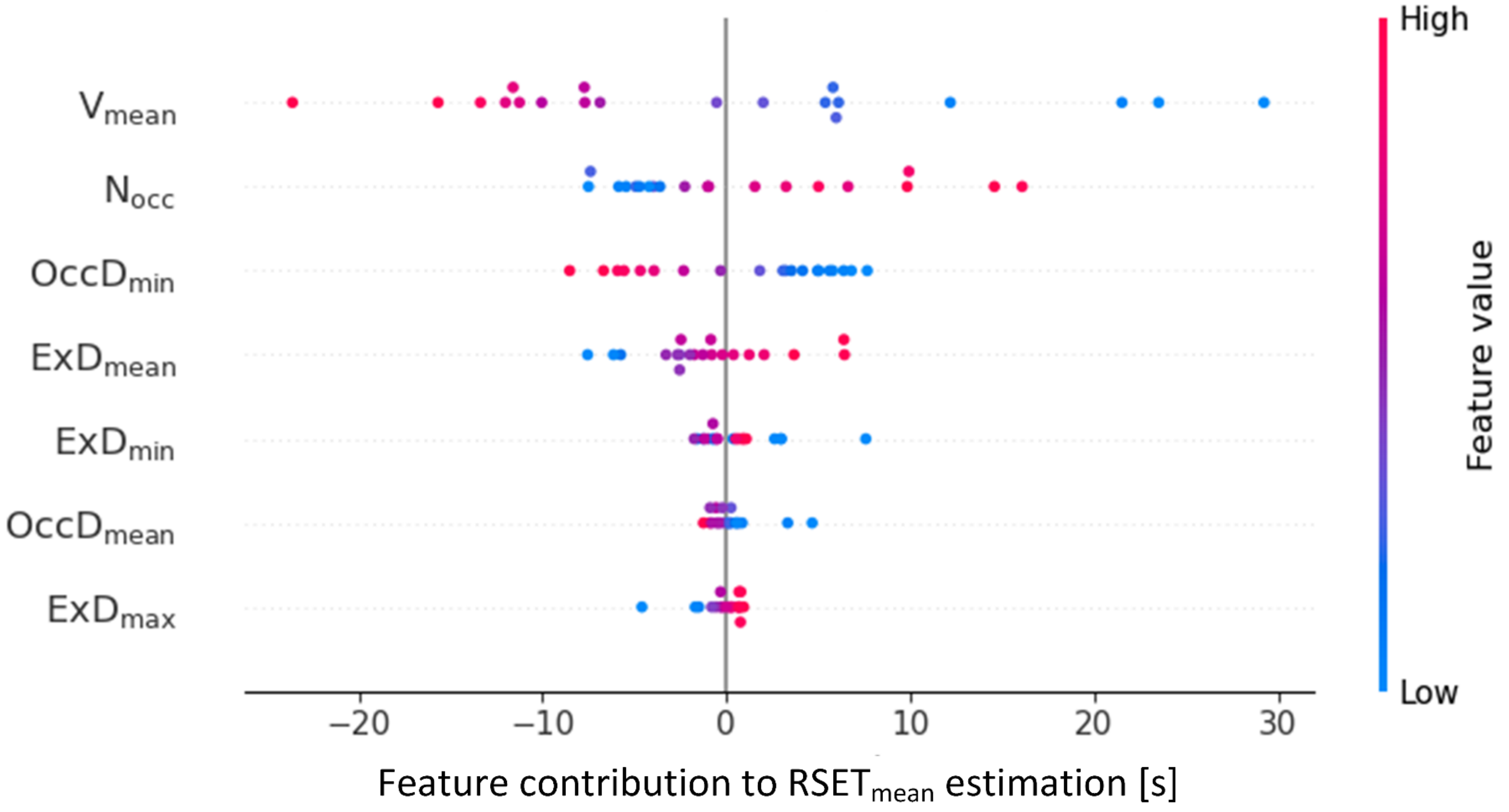
- Residuals plot of ANN metamodel in context of histogram of RSET for all Pathfinder simulations.
- Majority of residuals are distributed in the range of -5 to 5 seconds.





FEATURES CONTRIBUTION

- SHAPLEY method for feature impact analysis.



DISCUSSION

- The results showed that the task of estimating RSET using machine learning metamodels **can be solved** with relatively sufficient accuracy.
- Of all 9 tested metamodels, the **ANN metamodel achieved the best results** in terms of all the metrics used.
- The goal of subsequent research should be the optimization of the input features of the simulation dataset in terms of their number and characteristics, with the aim of achieving even more reliable results.