Journal of Wind Engineering & Industrial Aerodynamics 226 (2022) 105027

Contents lists available at ScienceDirect

Journal of Wind Engineering & Industrial Aerodynamics

journal homepage: www.elsevier.com/locate/jweia

Explainable Machine Learning (XML) to predict external wind pressure of a low-rise building in urban-like settings

D.P.P. Meddage [a,b], I.U. Ekanayake [c], A.U. Weerasuriya [d,*], C.S. Lewangamage [e], K.T. Tse [d], T.P. Miyanawala [f], C.D.E. Ramanayaka [g]

[a] *Department of Civil and Environmental Engineering, Faculty of Engineering, University of Ruhuna, Hapugala, Sri Lanka*
[b] *Department of Civil Engineering, Faculty of Engineering, Sri Lanka Institute of Information Technology, Malabe, Sri Lanka*
[c] *Department of Computer Engineering, Faculty of Engineering, University of Peradeniya, Peradeniya, Sri Lanka*
[d] *Department of Civil and Environmental Engineering, The Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong*
[e] *Department of Civil Engineering, Faculty of Engineering, University of Moratuwa, Kadubedda, Sri Lanka*
[f] *Department of Mechanical Engineering, Faculty of Engineering, University of Moratuwa, Katubedda, Sri Lanka*
[g] *School of Design and the Built Environment, Faculty of Humanities, Curtin University, Perth, Western Australia, Australia*

ARTICLE INFO

Keywords:
Machine learning
Explainable machine learning
SHAP
Wind pressure
Low-rise building

ABSTRACT

This study used explainable machine learning (XML), a new branch of Machine Learning (ML), to elucidate how ML models make predictions. Three tree-based regression models, Decision Tree (DT), Random Forest (RF), and Extreme Gradient Boost (XGB), were used to predict the normalized mean ($C_{p,mean}$), fluctuating ($C_{p,rms}$), minimum ($C_{p,min}$), and maximum ($C_{p,max}$) external wind pressure coefficients of a low-rise building with fixed dimensions in urban-like settings for several wind incidence angles. Two types of XML were used — first, an intrinsic explainable method, which relies on the DT structure to explain the inner workings of the model, and second, SHAP (SHapley Additive exPlanations), a post-hoc explanation technique used particularly for the structurally complex XGB. The intrinsic explainable method proved incapable of explaining the deep tree structure of the DT, but SHAP provided valuable insights by revealing various degrees of positive and negative contributions of certain geometric parameters, the wind incidence angle, and the density of buildings that surround a low-rise building. SHAP also illustrated the relationships between the above factors and wind pressure, and its explanations were in line with what is generally accepted in wind engineering, thus confirming the causality of the ML model's predictions.

1. Introduction

Pressure distributions must be known in detail in order for a structure's wind loads to be calculated (Tamura et al., 1999; Tanaka et al., 2012; Zu and Lam, 2018; Hu et al., 2019), a building's natural ventilation rate to be estimated (Van Moeseke et al., 2005; Gough et al., 2018; Pan et al., 2019; Charisi et al., 2021), or for the wind field near a building to be modeled (Hui et al., 2013; Xing et al., 2018; Kim et al., 2019; Wang et al., 2020; Meddage et al., 2022). Wind tunnel experiments and Computational Fluid Dynamics (CFD) simulation are by far the most popular methods for obtaining such wind pressure data, but this should in no way downplay the importance of making full-scale field measurements of wind pressure. As much as such measurements can be sporadic and be obtained under uncontrolled/uncontrollable conditions, they can still prove very useful, as they reveal the true nature of wind pressure fluctuations (Fu et al., 2012). In contrast, even though wind tunnel experiments and CFD simulations are fully controlled, they have such limitations as similarity issues (for wind tunnel experiments) (Cermak, 1984) and inherent numerical errors (for CFD simulation) (Thordal et al., 2019). Besides, the latter techniques are expensive, time-consuming, and require not only expertise but also specific facilities. All of the above therefore necessitate a fast, economical, and less resource-intensive way to estimate wind pressure.

Since the advent of artificial intelligence (AI) as driven by the proliferation of computational power, interactive computer programming languages and sophisticated databases, Machine learning (ML) has made tremendous progress. ML now offers a fast, economical solution for almost all applications in the manufacturing and service sectors and may

---

* Corresponding author.
E-mail address: auw@connect.ust.hk (A.U. Weerasuriya).

https://doi.org/10.1016/j.jweia.2022.105027
Received 12 December 2021; Received in revised form 12 April 2022; Accepted 3 May 2022
Available online 10 May 2022
0167-6105/© 2022 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

even be able to solve wind engineering's three "non-" issues: non-stationarity, non-Gaussianity, and non-linearity (Kareem, 2020), as well as provide an innovative approach for solving unsteady, multiscale, complex fluid flows, including wind (Brunton et al., 2020a,b). In fact, fluid dynamics ML applications have already been projected to grow rapidly in the near future, particularly driven by synergy between big data, high-performance computing, advanced algorithms, and considerable investment in research (Brunton et al., 2020a,b).

Given its ability to describe and interpret complex, non-linear relationships between inputs and outputs, ML can be a fitting substitute for wind tunnel experiments and CFD simulations when predicting wind pressure on structures (Kareem, 2020). One can consider this as an optimistic insight with respect to the reality of concurrent use of ML applications with wind tunnel testing and CFD simulation. Amalgamations of ML with physical and numerical modeling result in large cost reductions and high efficiency in evaluating alternative building designs in the early design stage (see Weerasuriya et al., 2020), quick modeling of wind environment in the surrounding of buildings (see Weerasuriya et al., 2021), and urban-like settings (see Moonen and Allegrini, 2015). Stathopoulos and Wu (1994) made one of the earliest attempts to develop such tools for wind engineering: inspired by prior attempts, Stathopoulos and Wu (1994) created a knowledge-based computer system based on provisions given in National Building Code of Canada (NBCC) and previous experimental data to estimate wind loads that pertain to low-rise buildings with various roof shapes and opening configurations. However, arguably, ML applications in wind engineering really only began with the advent of artificial neural networks (ANN). ANN has allowed wind engineers to establish complex non-linear relationships between a building's dimensions, flow properties of the approaching wind, and wind pressure on the building. For example, Chen et al. (2002) used backpropagation ANN to predict the time series of external wind pressure on a roof panel of a flat-roofed low-rise building and confirmed ANN's accuracy and robustness when predicting pressure at locations affected by corner vortices, as well as for cases with wind directions and upstream terrains slightly different from the training dataset. Kwatra et al. (2002) developed an ANN model using a limited dataset to predict wind pressure on the gable roof of a low-rise building with or without other buildings in the surroundings. The model required half of the effort of wind tunnel tests to predict wind pressure on the roof, even for roof slopes that were not part of the training dataset. More recently, several studies employed novel ANN architectures such as feed-forward multilayer ANN (Bre et al., 2018), and fuzzy neural networks (Fu et al., 2007) to improve accuracy for predicting the wind pressure on buildings.

There has also been a noticeable surge in ML applications in wind engineering recently. Various studies have for example employed ML techniques to predict the wind pressure on circular cylinders (Hu and Kwok, 2020) and buildings (Hu et al., 2020; Kim et al., 2021; Lamberti, G., & Gorlé, 2021), to estimate wind speed (Khosravi et al., 2018; Mohandes and Rehman, 2018; Yang and Chen, 2019; Vassallo et al., 2021; Hur, 2021), to reconstruct flow fields near bluff bodies (Ti et al., 2020; Fukami et al., 2020), and to design wind-sensitive structures (Ding and Kareem, 2020; Li et al., 2021). Most of these studies followed supervised ML such as various forms of Decision Trees (Hu and Kwok, 2020; Hu et al., 2020; Lin et al., 2021a,b, 2022), ANN (Ti et al., 2020; Lamberti, G., & Gorlé, 2021), reinforced learning (Ding and Kareem, 2020; Li et al., 2021); others were sporadically unsupervised ML, such as K-mean, K-medoid, and DBSCAN (Li et al., 2020; Kim et al., 2021; Lin et al., 2021). Despite their superior performance and high prediction accuracy, however, the inner workings of these ML models are unknown to end-users; end-user knowledge of a model is limited how a few hyper-parameters are calibrated during training, which is to say, end-users do not know what important role the inputs play in the model's predicting its outputs, or what forms of relationship exist between the inputs and outputs. The absence of such knowledge diminishes end-user confidence in using ML models, reducing the reliability of the predictions, thereby preventing ML from having broader applications in wind engineering.

Emerging as a new branch of ML, explainable Machine Learning (XML) intends to give end-users a better understanding of how ML models make predictions on an overall and per instance basis. XML has gained favor among various stakeholders including data scientists, business owners, model risk analysts, regulators, and consumers, owing to its ability to detail the correctness, robustness, bias, improvement, transferability, and human comprehensibility of ML models (Belle and Papantonis, 2021). With these features, XML can turn opaque, *black-box* ML models (e.g., ensemble and deep learning models) into clear, *glass-box* models, unveiling their inner workings. XML also illuminates how ML models make their predictions, revealing the dependency between input and output as well as the importance of the inputs (Xu et al., 2019; Roscher et al., 2020).

To the best of our knowledge, no previous studies have used XML to explain the performance of ML models in wind engineering applications. Therefore, the primary objective of the current study is to demonstrate how XML can improve end users' confidence in applying ML in wind engineering applications. The study further argues that it is important to understand a model's inner workings, as that affects the accuracy of any predictions that the model makes. The study also uses XML explanations to improve the reliability and trustworthiness of the predictions given by black-box ML models, and further delineates how XML can help distinguish between influencing building parameters and dominant pressure patterns, in order to cross-validate ML models' predictions against existing knowledge of external wind pressure of low-rise buildings. Finally, the study demonstrates that using XML does not mean necessarily sacrificing accuracy or complexity, but rather underpins a model's predictions by providing human-comprehensible explanations.

Since the wind engineering community is new to XML, this study begins by introducing its basics, then details a particular XML technique: SHAP (SHapley Additive exPlanations) in Section 2. Section 3 outlines the wind tunnel test on the external wind pressure of a low-rise building in an urban-like setting, which this study uses as a case study. Section 4 introduces three tree-based regression models that predict external wind pressure on the low-rise building. Section 5 compares the prediction accuracy of each ML model, and Section 6 provides global and local explanations for the predictions of the best-performing ML model. Finally, Section 7 presents several concluding remarks, and Section 8 discusses some limitations of this study and recommendations for future studies.

2. Explainable machine learning (XML)

2.1. Basics of XML

Despite its popularity, XML does not have a universally accepted definition. Miller (2019) argues that XML is not just "more AI" but a technique to address the human-agent (i.e., user-computer) interaction problem in machine learning. Based on this idea, Miller (2019) loosely defines XML as an explanatory agent that reveals the causes that underlie a model's predictions or another agent's decision-making. Explanation has sometimes been described as post-hoc interpretability (Lipton, 2018), where interpretability refers to the degree to which an observer can comprehend the cause of a decision (Biran and Cotton, 2017; Kim et al., 2016). Thus, explanation can be considered as a mode of interpretability. In ML literature, the term "explanation" has often been equated to or interchangeably used for "interpretability" (Miller, 2019; Molnar, 2020), "understandability" (Andrzejak et al., 2013), "comprehensibility" (Luštrek et al., 2016), and "intelligibility" (Weld and Bansal, 2019).

Despite its ambiguous definitions, XML methods can be clearly categorized according to one of several ways. Firstly, XML can be broadly categorized as intrinsic or post-hoc, depending on whether the end-user can understand the inner workings of the model by looking at its structure (i.e., intrinsic) or will require special explainability techniques after model training (i.e., post-hoc). ML models whose inner structures are simple (e.g., linear regression models or Decision Trees with limited tree depth) are easily comprehensible; those with complex inner workings require post-hoc techniques. Secondly, XML can be classified as whether it is limited to a specific type of ML model (i.e., model-specific) or is generically applicable (i.e., model agnostic). Some model-specific XML techniques are intrinsically interpretable (e.g., weights of a linear regression model); others are post-hoc methods (e.g., SHAP, LIME). While model-specific techniques must access and analyze a model's inner structure, model-agnostic methods examine input and output features as a pair. Thirdly, XML can be grouped according to whether explanations are given at the global or local level. Global-level explanations elucidate an entire model's behavior by revealing how its input impacts overall prediction accuracy and laying out any unrealistic relationships between input and output. However, such global-level explanations are difficult for ML models with highly correlated inputs (Ribeiro et al., 2016, 2018). Global explanations also comment on the performance of an ML model but do not explain a specific model's prediction or clarify how data impacts the particular instance (Veiber et al., 2020). Currently, several XML techniques such as partial dependency test, feature importance, individual conditional expectations have been used to obtain global-level explanations. Local explanations, on the other hand, reveal the reasons for making a specific prediction for an instance using relevant inputs and outputs of a black-box model. Local explanations also provide details of relevant inputs for the outputs, rank features by importance in either ascending or descending order and unveil relationships between input and output prediction that are human-comprehensible. Some popular local explanation techniques are LIME (Local Interpretable model-agnostic explanation) (Ribeiro et al., 2016), Anchors (or Scoped Rules) (Ribeiro et al., 2018), and SHAP (SHapley Additive exPlanations) (Lundberg and Lee, 2017). Among them, SHAP is the most advantageous, as it has a solid theoretical foundation, is able to provide global explanations via combining individual local explanations, compares an instance's prediction against the average prediction (i.e., contrastive explanations), uses fairly distributed prediction among the inputs, and follows an approach to unify different XML techniques such as LIME and Shapely value (Molnar, 2020). The current study therefore chooses SHAP to explain both globally and locally the predictions of an ML model in a wind engineering application.

2.2. SHAP (SHapley additive exPlanations)

Lundberg and Lee (2017) proposed SHAP to explain individual predictions based optimal Shapley values from the game theory. SHAP explains the prediction of an instance by computing the contribution of each input in terms of its Shapley value according to the coalitional game theory (Molnar, 2020). Here, the inputs are players in a coalitional game, and the prediction is the payout, which is fairly distributed among players. SHAP satisfies three desirable properties, namely, local accuracy, missingness, and consistency, while finding a single unique solution for the coalitional game theory (Lundberg and Lee, 2017). Local accuracy requires the explanation model, *g*, to at least equal the output of the original model, *f*, for a simplified input, *x'*, which corresponds to the original input, *x*. Missingness constrains no attributed impact of any absent features on the prediction. Consistency requires that even if a model changes, the marginal contribution of an input increases or stays the same regardless of other inputs but should not decrease. Lundberg and Lee (2017) present a model-agonistic KernelSHAP and model-specific LinearSHAP for linear models, DeepSHAP for deep neural networks, and later TreeSHAP for tree-based models (Lundberg et al., 2018).

The current study employs TreeSHAP, which uses a linear explanation model and Shapley values (Eq. (1)) to locally approximate the original prediction model *f*.

$$g(z') = \varphi_o + \sum_{i=1}^{M} \varphi_i z'_i \tag{1}$$

where g is the explanation model, $z' \in \{0,1\}^M$ represents simplified features or the coalition vector, M is the maximum coalition size, and $\varphi \in \mathbb{R}$ is the feature attribution for feature j (i.e., its Shapley value). Lundberg and Lee (2017) propose Eq. (2) to calculate the attribution of each feature:

$$\varphi_i = \sum_{S \subseteq N \setminus \{i\}} \frac{|S|!(M-|S|-1)!}{M!}\left[f_x(S \cup \{i\}) - f_x(S)\right] \tag{2}$$

where

$$f_x(S) = f(h_x(z')) = E[f(x)|x_S] \tag{3}$$

In Eq. (2), S is a subset of the input features, and M is the set of all input features. In Eq. (3), h_x is a mapping function between the coalition vector and the original input space, and $E[f(x)|x_S]$ is the expected value of the function conditioned on the subset S.

For TreeSHAP, Lundberg et al. (2018) propose a modern Shapely interaction index, Φ (Eq. (4)) by considering pairwise interactions of inputs that lead to the estimation of the impact of all feature pairs on a model prediction.

$$\Phi_{i,j} = \sum_{S \subseteq N \setminus \{i,j\}} \frac{|S|!(M-|S|-2)!}{2(M-1)!}\nabla_{ij}(S) \tag{4}$$

when $i \neq j$, and

$$\nabla_{ij}(S) = f_x(S \cup \{i,j\}) - f_x(S \cup \{i\}) - f_x(S \cup \{j\}) + f_x(S) \tag{5}$$

$$= f_x(S \cup \{i,j\}) - f_x(S \cup \{j\}) - \left[f_x(S \cup \{i\}) - f_x(S)\right] \tag{6}$$

In Eq. (4), Φ between features i and j is divided equally between each feature such that $\Phi_{i,j} = \Phi_{j,i}$, and the total interaction effect is $\Phi_{i,j} + \Phi_{j,i}$. The main effect of a feature for a prediction is defined as the difference between its SHAP value and the SHAP interaction values:

$$\Phi_{i,i} = \varphi_i - \sum_{j \neq i} \Phi_{i,j} \tag{7}$$

The current study implements TreeSHAP using Python's Scikit-learn, matplot, NumPy, pandas, and Shap libraries.

3. Wind tunnel experiment

3.1. Wind pressure of a low-rise building in urban-like settings

The ML models of the current study were developed based on wind tunnel test data of mean and fluctuating wind pressures on a low-rise building in an urban-like setting. Wind pressure distributions of this target building was more complex than isolated buildings due to the numerous interactions it had with other buildings in the surroundings and the incoming wind flow (Xie and Gu, 2007; Kim et al., 2011; Hui et al., 2017), and the fact that it was affected by geometric parameters, architectural features, incident wind directions (or orientation of the building), approaching wind conditions (Akon and Gregory, 2018; Fern'andez-Cabán & Masters, 2018), and nearby buildings (Stathopoulos, 1984). Effects from nearby buildings would depend mainly on the buildings' dimensions and arrangements, their orientations with respect to the incident wind direction, and distances to the target building (English, 1990; Khanduri et al., 1998). All these factors make it difficult to estimate the wind pressure of buildings in an urban-like setting with accuracy, which is nonetheless imperative for many wind-related processes pertaining to groups of buildings, such as wind force/pressure acting on the buildings' façades and claddings (Elshaer et al., 2017), the buildings' natural ventilation (King et al., 2017), pollutant dispersion in built-up areas (Aristodemou et al., 2018), and extreme wind patterns in built-up areas (Biao et al., 2019). Current understandings of wind pressure on buildings in urban-like settings are largely based on data from wind tunnel tests and CFD simulations.

3.2. Building models and inflow wind conditions

The current study uses data of external wind pressure on a low-rise building in an urban-like setting from the aerodynamic database of non-isolated low-rise buildings created by Tokyo Polytechnic University in Japan (http://www.wind.arch.t-kougei.ac.jp/info_center/windpressure/grouplowrise/mainpage.html). The database includes wind tunnel test data for low-rise buildings with three types of roofs (flat, gable, and hipped) surrounded by similar low-rise buildings. The heights of the buildings are 60 mm, 120 mm, or 180 mm and are configured in either regular, staggered, or random form.

This study chose a portion of the database that contains the external wind pressure of a flat-roofed, low-rise building surrounded by buildings arranged in each of the "regular" arrangements in Fig. 1. The arrangements had various densities (C_A) of 0.1, 0.15, 0.2, 0.25, 0.30, and 0.40, where C_A was defined as the ratio between the area occupied by buildings to the total area of the site. All buildings in the building group were 120 mm in height (H), 160 mm in breadth (B), and 240 mm in depth (D) (Fig. 2). These building models were tested for five wind incidence angles: $\theta = 0°$, 22.5°. 45°, 67.5°, and 90°, making up 30 cases in the dataset (Fig. 3). Of Note that, although the number of wind tunnel test models are 30, they have 7680 data instances, of which each instance contains a unique pressure record corresponds to a particular combination of pressure tap location, wind incidence angle, and density of surrounding buildings. Among them, five cases (filled red squares in Fig. 3) were selected as validation cases, and the rest were used for training the ML models. The size of training (25 cases) and validation (5 cases) datasets was chosen by following a trial-and-error method, in which dataset size varied in a range of 3–27 cases (or 10%–90% of total number of cases). The analysis shown in Figure A1 in Annex A indicates an insignificant impact on dataset size for model training but a considerable effect on prediction accuracy of ML models. The accuracy of the latter fluctuated considerably as can be seen in variation in correlation coefficient (R^2) in a range of 0.32–0.94 as the validation dataset size varied from 27 cases to 3 cases. However, R^2 was larger than 0.8 for most of the trials with 3–6 validation cases and had the maximum R^2 for 5 cases. Therefore, this study employed 25 and 5 cases for training and validation of ML models, respectively. Training and validation datasets contained four pressure coefficients (Eqs. 8–11) of each pressure taps together with their coordinates, and parameters to characterize the surrounding and inflow wind stored as a matrix. The sizes of matrices were 6400 × 13 (training) and 1280 × 13 (validation) with the rows represented pressure taps of each test case, and columns contained the corresponding data. Of note that the size of the training dataset duly satisfies a '1:10' rule of thumb between the model parameters and number of datasets (Haldar, 2015) as the maximum number of model parameters in this study is seven (Figure A1 in Annex). The ML models were trained following the K-fold cross-validation method with K set to 5 to evaluate the model.

Each building configuration was tested in an atmospheric boundary layer (ABL) wind flow over a suburban terrain specified as terrain category III in AIJ (2004). The simulated ABL wind flow followed a power-law type wind profile with an exponent of 0.20 and had a 7.8 m/s wind speed at 100 mm height. The turbulence intensity profile showed a proper decay with height and had 25% turbulence intensity at 100 mm height. Wind pressure on the external walls of the target building was measured using 256 pressure taps (Fig. 4). All pressure taps synchronously recorded wind pressure for 18 s at a sampling frequency of 781.25 Hz. The pressure data collected by each pressure tap subsequently passed through a low-pass filter at 300 Hz to estimate the four types of normalized pressure coefficients, as per Eqs. 8–11.

$$C_{p,mean} = \frac{\overline{p}}{\overline{p}_H} \tag{8}$$

$$C_{p,rms} = \frac{\sqrt{\overline{(p(t) - \overline{p})^2}}}{\overline{p}_H} \tag{9}$$

$$C_{p,\max} = \frac{p_{\max}(t)}{\overline{p}_H} \tag{10}$$

$$C_{p,\min} = \frac{p_{\min}(t)}{\overline{p}_H} \tag{11}$$

where $C_{p,mean}$, $C_{p,rms}$ $C_{p,\max}$, and $C_{p,\min}$ are the mean, RMS, maximum, and minimum pressure coefficients, $\overline{p}$ is the time-average pressure, $p(t)$ is the instantaneous pressure at time t, $p_{\max}(t)$, and $p_{\min}(t)$ are the maximum and minimum instantaneous pressure, $\overline{p}_H$ is the reference dynamic wind pressure calculated at the roof height as $0.5\rho V_H^2$, where V_H is the mean longitudinal wind speed at roof height H, and ρ is the air density. It should be noted that peak pressure coefficients: $C_{p,\max}$, and $C_{p,\min}$ are calculated using the single worst peak method (Stathopoulos, 1979) because of pressure time histories are available for one trial for each test case in the database.

Table 1 summarizes the descriptive statistics of the geometric

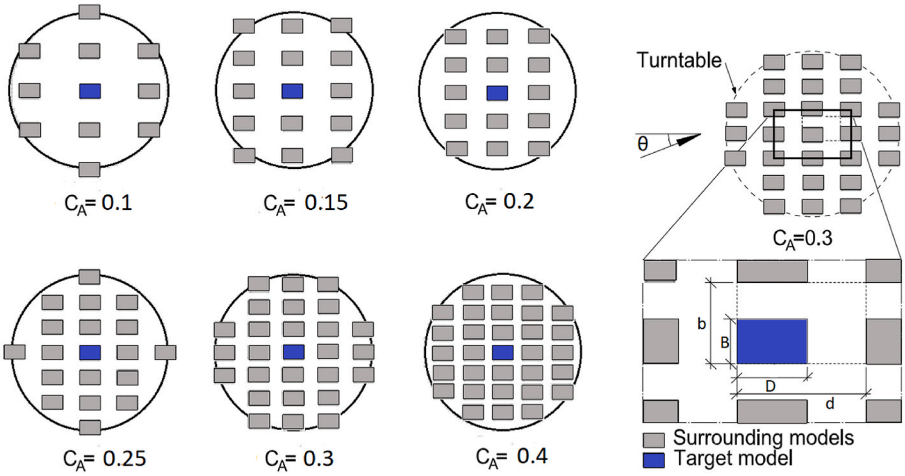


Fig. 1. Regular arrangements of buildings with various area densities (C_A) (grey rectangles – surrounding buildings, blue rectangle – target building). (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

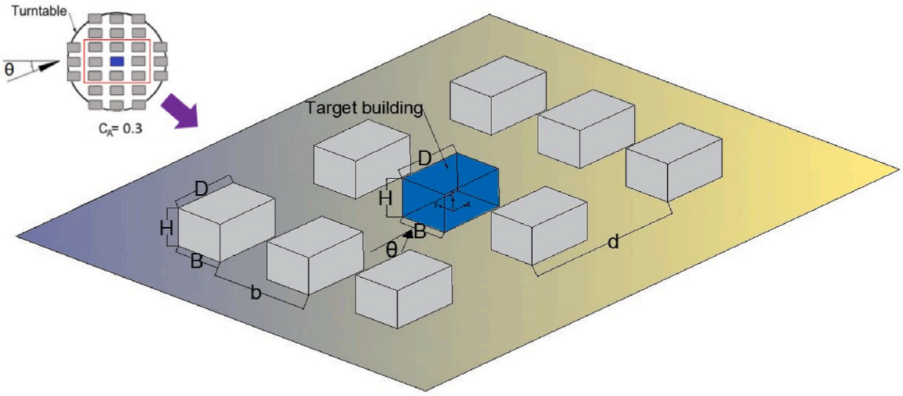


Fig. 2. Dimensions of the target and surrounding buildings.

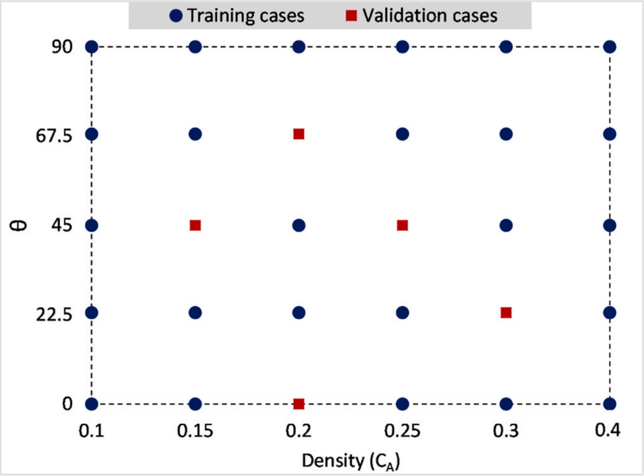


Fig. 3. The training and validation datasets for developing ML models (θ – wind incidence angle; C_A – area density).

parameters and pressure measurements from the wind tunnel tests. Of note is that this study did not consider building dimensions as inputs because they were constant for all test cases. The geometric parameters, x/H, y/H, and z/H are the coordinate of the pressure taps on the building's walls. The area density C_A characterizes how closely the buildings in the surroundings of the target building are arranged. Its minimum and maximum values (0.10 and 0.40, respectively) represent the sparsest and densest arrangements of the surrounding buildings. The wind incidence angle θ relates variations in external wind pressure with the incident wind direction. Fluctuating and positive extreme wind pressure, $C_{p,max}$, and $C_{p,rms}$, contain only positive values, whereas all $C_{p,min}$ values are negative. $C_{p,mean}$ has both positive and negative values and varies in a wide range from 0.78 to −1.96 with a mean value of −0.22.

Fig. 5 shows the pairwise correlations between the geometric parameters and pressure coefficients. The pairwise correlation plots show that the geometric parameters and pressure coefficients have no definite trends, thus it can be concluded that the two types of parameters are mutually independent. However, some pressure coefficients exhibit weak or moderate correlations with each other. An example is the discernible positive correlations between $C_{p,max}$ and $C_{p,mean}$ ($R = 0.73$), and $C_{p,min}$ and $C_{p,mean}$ ($R = 0.68$). A perceptible weak negative correlation exists between $C_{p,rms}$ and $C_{p,min}$ ($R = -0.62$), but the rest of the pressure coefficients show no distinct variations with each other. Without distinct correlations between the geometric parameters and pressure coefficients, it is not possible to develop an accurate linear regression to predict external wind pressure of the target building. On the other hand, variations in the pressure coefficients can be related to the geometric parameters via non-linear relationships, which can be explored using advanced predictive models such as ML models. Therefore, this study employs ML models to predict external wind pressure using the geometric parameters and wind flow conditions as described in the following subsection.

4. Machine learning (ML) models

Three tree-based regression algorithms, namely, the Classification And Regression Tree (CART), Random Forest, and Extreme Gradient Boosting (XGBoost) were used to develop the ML models in this study. Tree-based regression algorithms are popular as predictive ML models due to their simple but powerful conceptual structures that can deal with nonlinear interactive relationships between inputs and outputs (Hastie et al., 2009). In addition, tree diagrams created by tree-based regression algorithms show the relationships between inputs and outputs (transparent ML model), and these relationships are easily explainable. Therefore tree-based regression models can be classified as intrinsic XML. In particular, Decision Trees with controlled tree depths and complexity can reveal important inputs and individual contributions to the prediction (Gottard et al., 2020).

The three tree-based regression models developed for this study, namely, Decision Tree (DT) based on CART, Random Forest (RF) based on Random Forest, and XGB based on XGBoost, have various degrees of model complexity and explainability. DT has the simplest structure of a single tree with natural visualization presenting it as an intrinsic XML

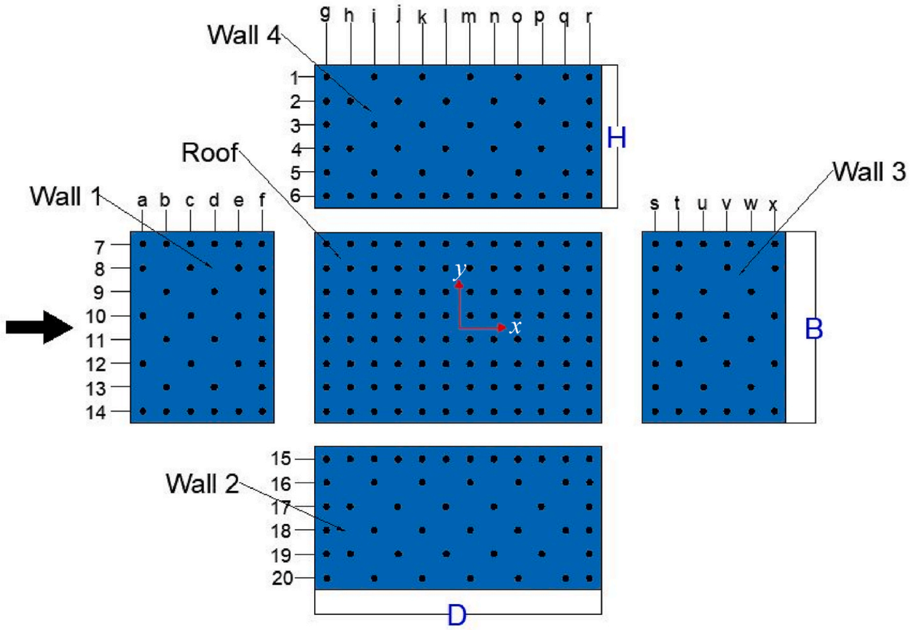


Fig. 4. Arrangement of pressure taps on the external walls of the target building (H – height; B – breadth; D – depth of the target building).

Table 1
Descriptive statistics of various input and output parameters from wind tunnel data.

Parameter Type	Parameter	Mean	Maximum	Minimum	Standard deviation
Geometric	x/H	0	1	−1	0.72
	y/H	0	0.67	−0.67	0.51
	z/H	0.71	1	0.08	0.32
	C_A	0.23	0.40	0.10	0.10
	Θ	45°	90°	0°	32°
Pressure	$C_{p,\ mean}$	−0.22	0.78	−1.96	0.25
	$C_{p,\ rms}$	0.19	0.88	0.10	0.09
	$C_{p,\ min}$	−1.16	−0.07	−6.62	0.70
	$C_{p,\ max}$	0.56	3.22	0	0.59

method if the tree depth is small. The other two models have nexus structures created by complex algorithms, thus requiring post-hoc explanations (Meddage et al., 2021). The follow subsections further detail and discuss each ML model.

4.1. Decision tree (DT)

A DT has a tree-like structure with several branches and leaf nodes created by a recursive binary partitioning of the instance space (Singh and Gupta, 2014). Each branch represents a choice between several alternatives, and each leaf node offers classification or decisions until the DT arrives at the outcome. Data are split at nodes based on a learning algorithm such as Iterative Dichotomies 3 (ID3) (Quinlan, 1986), Classification And Regression Tree (CART) (Breiman et al., 1984), CHi-squared Automatic Interaction Detector (CHAID) (Kass, 1980). The current study employs CART with the minimization of the sum of square error between prediction and observation as criterion for splitting data at the nodes. After the final DT has been built, any branches that are not useful, i.e., they reflect noise or outliers in the data, are pruned from the tree based on the Variable Importance (VI) score. Pruning also helps to avoid data over-fitting and minimize model complexity (Zimmerman et al., 2016).

Although DT is generally explainable and computationally efficient, it cannot deal with linear relationships between inputs and outputs and is extremely sensitive to any changes in the input. Such changes will make the DT less robust, resulting in unstable model structures (Priyanka & Kumar, 2020). Moreover, DT quickly loses explainability as the tree grows in depth (Molnar, 2020). These drawbacks can be minimized by choosing proper values for such hyper-parameters as maximum depth, maximum number of leaf nodes, and minimum number of data samples at a leaf node (Hu and Kwok, 2020). Following the grid search method, this study has found the most suitable values for these hyper-parameters to be: maximum depth = 14; minimum number of leaf nodes = 2, minimum number of data samples at a leaf node = 2; maximum number of features = 5 (Figure A2 inAppendix A). Figure B1 in Appendix B shows how accuracy varies with these hyper-parameters.

4.2. Random Forest (RF)

Breiman (2001) proposed the RF — an ensemble tree model — to minimize DT's lack of robustness, low bias and high variance, which result from the data overfitting. Ensemble tree models use not one but numerous DTs, in parallel or in series, as *weak learners*, and average their predictions to produce an overall prediction. The trees in an RF are created in parallel, and predictions are made using the Bagging (Bootstrap AGGregatING) algorithm (Breiman, 1996). Bagging has two key processes: bootstrap sampling, and aggregation of individual predictions. Bootstrap sampling produces data subsets for training individual weak learners using sampling with replacement (Efron and Tibshirani, 1994). Aggregation averages the weak learners' predictions to estimate the output of the RF. In addition to Bagging, RF incorporates a randomized feature selection at its nodes to minimize the correlation (similarity) between the *weak learners*. At the nodes, RF chooses m out of a total of p features available for data splitting, such that $m = \log_2(p+1)$ (Breiman, 1996), or $m = p/3$, or $m = \sqrt{p}$ (Han and Kim, 2019). The rest of the model training is similar to that of DT with the calibration of a set of hyper-parameters including the number of weak learners, number of selected features, and maximum depth of trees. In this study, the RF hyper-parameters were chosen to be: maximum tree depth = 14; number of weak learners = 100; minimum samples at a leaf = 2; minimum sample split = 2. RF is advantageous for reducing the overfitting problem of DT and can handle both categorical and continuous inputs. However, RF requires more computational power and training time as it uses many trees to make predictions. Its ensemble of trees hinders explainability, impeding the identification of important inputs for

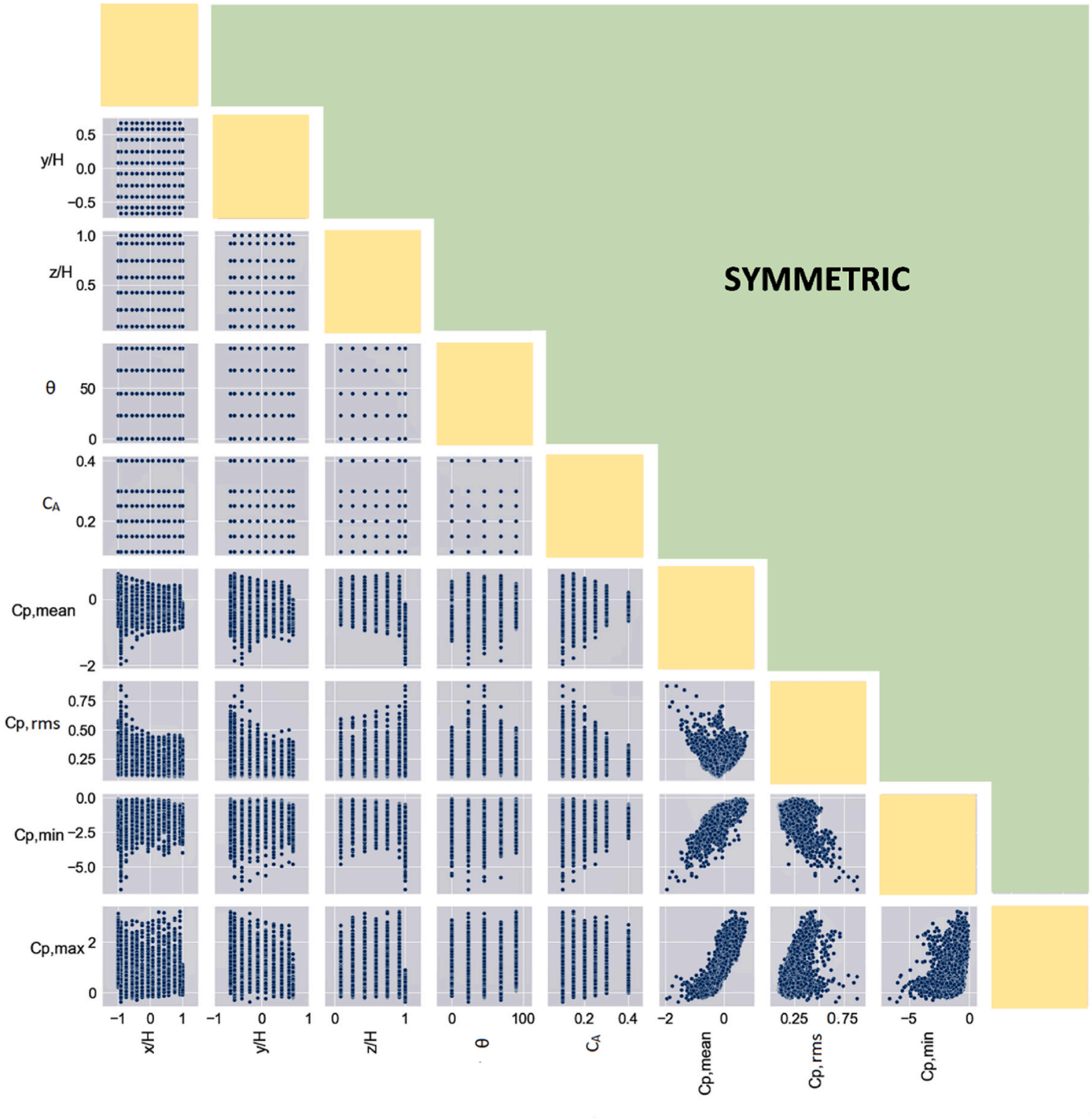


Fig. 5. Pairwise correlations between geometric parameters and normalized pressure coefficients.

making predictions (Choudhury and Gupta, 2019).

4.3. XGBoost (XGB)

XGB also uses an ensemble of trees, assembled based on a gradient boosting framework. Here boosting refers to converting *weak learners* whose predictions are slightly more accurate than random guesses to *strong learners* with near-perfect performance (Zhou, 2012). A boosting algorithm creates *weak learners* sequentially to minimize the difference between the observation and aggregated predictions of all *weak learners* produced in previous iterations. The gradient boosting technique — a variant of boosting — uses gradient descent algorithms to minimize the loss function based on the squared difference between the observation and aggregated predictions. Chen and Guestrin (2016) have shown how XGB is an advanced version of the gradient boosting technique, as it incorporates system optimization and several other algorithmic improvements. For example, XGB uses parallelized implementation for sequential tree building and defines the maximum tree depth to improve the computational performance of the algorithm. XGB also optimizes hardware usage by allocating internal buffers in each thread to store gradient statistics and *out-of-core* computing to handle large data sets that do not fit into the memory. XGB penalizes complex tree structures being overfitted to training data by using regularization schemes such as Lasso, and Ridge, and its embedded sparsity awareness handles sparsity patterns in the training data. XGB employs weighted quantile sketches to find the optimal split points among weighted datasets, and its built-in cross-validation method specifies the exact number of boosting iterations required in a single run without adding another algorithm for this search. However, overfitting to training data and prolonged model training as XGB's hyper-parameters are too numerous have been identified as its shortcomings (Egas-López and Gosztolya, 2020). These shortcomings can be minimized by optimizing its hyper-parameters: the minimum loss reduction required to further partition a leaf node, the learning rate, the sub-sample ratio of the training instance, maximum depth of a tree, and the number of trees. After the grid search (Figure B3 in Annex), the current study has identified these optimum values for the hyper-parameters: number of trees = 400 with, maximum tree depth = 5; $\gamma = 0.0001$; learning rate = 0.3923; reg_α = 0.0001 (Figure A1 in Annex1). Similar to RF, the complex model structure of XGB hinders self-explanation of the model, thus necessitating post-hoc explainable methods to elucidate model performance.

5. Performance evaluation of ML models

Before applying XML, the prediction accuracy of the three tree-based regressors was qualitatively estimated by comparing the predicted $C_{p,mean}$, $C_{p,rms}$, $C_{p,max}$, and $C_{p,min}$ with the wind tunnel test data for the five validation cases (1280 data points) (Fig. 6). The comparison shows satisfactory accuracy of the ML models in predicting positive $C_{p,mean}$, which deviated less than 20% from the corresponding wind tunnel test data. However, a considerable number of predictions of negative $C_{p,mean}$ by DT and RF were more than 20% different from the wind tunnel test data. In contrast, XGB showed high accuracy in predicting negative $C_{p,mean}$ with most of its deviations within ±10% of the wind tunnel test data. XGB showed similar competence in predicting $C_{p,rms}$, which deviated less than 10% from the wind tunnel test data. Compared to XGB, both DT and RF were less accurate in predicting $C_{p,rms}$, as their predictions.

Scattered across the lines that marked ±20% deviations from the wind tunnel test data. More dispersed predictions of $C_{p,min}$ were observed for DT and RF, where the deviations were higher than 20%. Indeed, the worst prediction accuracy of all three regressors was found to be for $C_{p,min}$, which scattered over a wide range, and many data points lay outside the ±20% error margin. Compared with predictions for $C_{p,mean}$, and $C_{p,rms}$, XGB was less accurate in predicting $C_{p,max}$, yet most of its predictions had less than ±20% error compared to the large prediction error of DT and RF. High error margin of $C_{p,max}$ and $C_{p,min}$ could be related to fact that their calculation is based on the single worst peak method rather than an ensemble average method. A single peak can be caused by rarely occurring events such as sudden increases of turbulence intensity due to infrequent eddies generated by surrounding buildings. Peak pressure caused by such events weakly depends on the inputs considered for the ML models in this study. As a result, the prediction accuracy of $C_{p,max}$ and $C_{p,min}$ was smaller than that of $C_{p,mean}$, and $C_{p,rms}$. Moreover, Fig. 6 shows the relatively invariant predictions of DT, in particular, for $C_{p,mean}$, $C_{p,rms}$, and to some extend for $C_{p,min}$ with respect to the wind tunnel test data. This invariance resulted from the "piece-wise approximation" method used for predictions by DT. RF also displayed (to a lesser extent) invariant predictions for these three pressure coefficients. In contrast, XGB's predictions closely followed the trend of the wind tunnel test data, confirming its superiority as a predictive model.

The prediction accuracy of the three regressors was quantitatively evaluated by calculating four validation metrics, namely, correlation coefficient (R), coefficient of determination (R^2), root-mean-squared-error (RMSE), and fractional bias (FB) (Table 2). The R-value, which tracks how closely the predictions follow the wind tunnel test data, is close to 1 for all cases, indicating that all three regressors accurately captured the variations in wind pressure as in the wind tunnel test. However, R^2 of the regressors indicates different degrees of deviation of predictions from the wind tunnel test data. For example, DT and XGB have the lowest and highest R^2 values for $C_{p,min}$ (0.88 and 0.94), indicating that DT's predictions deviated more from the wind tunnel test data than XGB. R^2 also provides basis for comparing the prediction accuracies of different types of pressure coefficients. For instance, all three regressors had lower accuracy predicting $C_{p,min}$ than predicting $C_{p,mean}$. The high and low prediction accuracies of the mean and extreme wind pressures are further confirmed by RMSE. For instance, smaller RMSE values (close to 0) of the three regressors for $C_{p,mean}$, and $C_{p,rms}$ substantiate their high accuracy in predicting mean and fluctuating wind

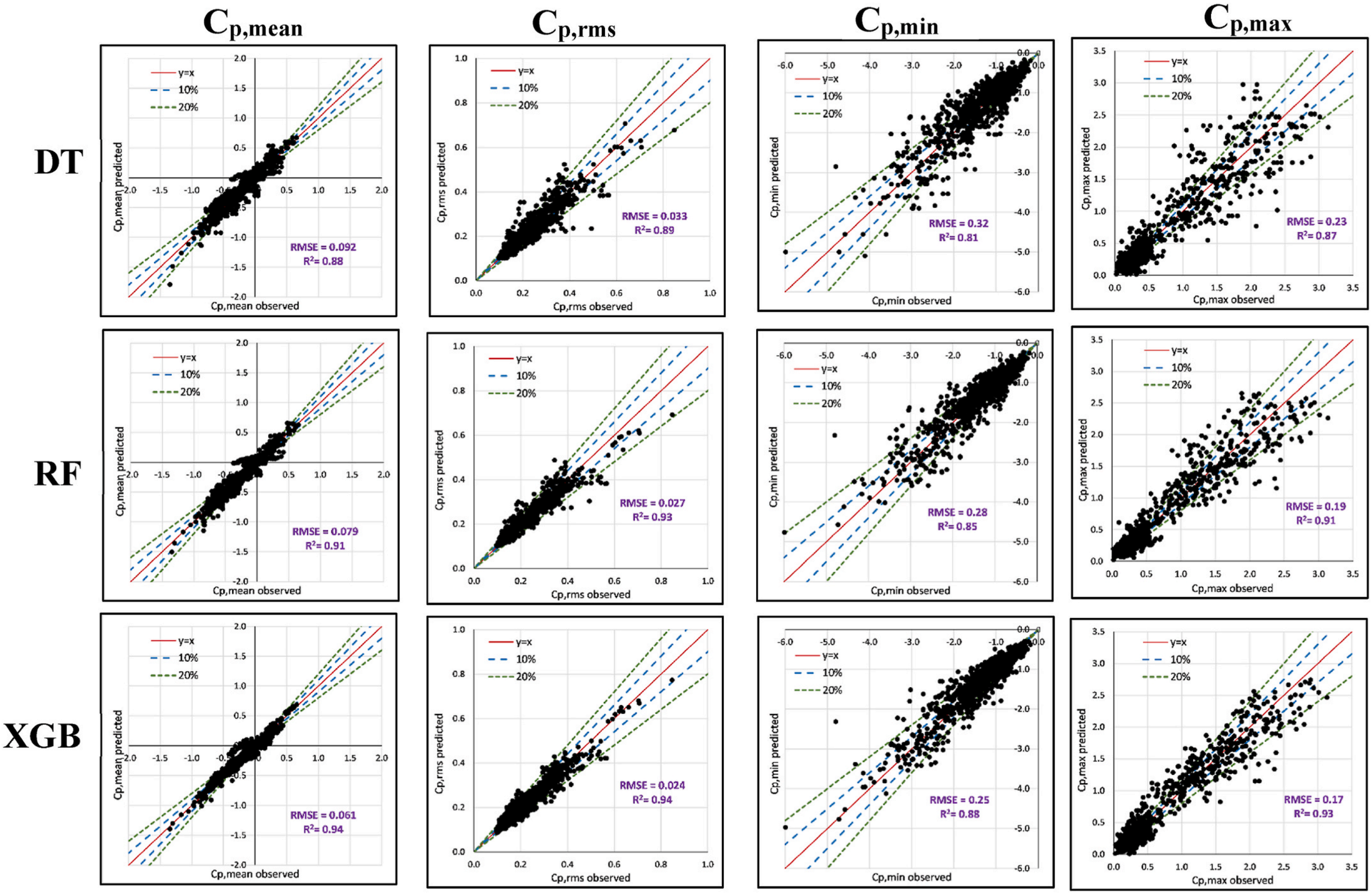


Fig. 6. Comparison of $C_{p,mean}$, $C_{p,rms}$, $C_{p,min}$, and $C_{p,max}$ predicted by ML models and measured in the wind tunnel test.

pressure compared to their predictions for extreme pressure such as C_{pmin} and $C_{p,max}$, which has RMSE >0.17. These errors have mainly resulted from the three regressors underestimating their predictions, as indicated by the negative FB values across the validation datasets. Despite these minor imperfections, the three regressors duly fell within the acceptable ranges for the selected validation metrics: R > 0.8; $R^2 >$ 0.8; −0.3 < FB < 0.3; 0.2 < RMSE <0.5, generally used in wind-related studies (Chang and Hanna, 2004; Goricsán et al., 2011; Moonen and Allegrini, 2015; Weerasuriya et al., 2021).

Tree regressors tend to overfit the training dataset, resulting in abnormally high and low prediction accuracies for the training and validation datasets, respectively (i.e., low bias and high variance of ML models). Fig. 7 investigates this common drawback of tree regressors by comparing R^2 values for the training and validation datasets. Across all pressure coefficients, R^2 values of the training dataset were 5%–16% higher than those of the validation dataset. This confirms that the regressors did not over-fit the training dataset and were sufficiently flexible to make predictions for *out-of-bag* data.

Fig. 8 compares $C_{p,mean}$, $C_{p,rms}$, $C_{p,min}$, and $C_{p,max}$ as predicted by DT, RF, and XGB against wind tunnel test data at the pressure taps along the centerline of the building to estimate the models' accuracy in predicting local pressures. Overall, the three regressors were more accurate predicting positive pressure coefficients than negative pressure coefficients such as $C_{p,mean}$ and $C_{p,min}$. Interestingly, RF and XGB had similar average errors for $C_{p,mean}$; XGB in fact had slightly smaller average errors: 6.9%, 5.9%, 29.8% for Case 1 ($C_A = 0.15$ and $\theta = 45°$), Case 2 ($C_A = 0.2$ and $\theta = 0°$), and Case 5 ($C_A = 0.3$ and $\theta = 22.5°$), while RF performed better in Cases 3 ($C_A = 0.2$ and $\theta = 67.5°$) and 4 ($C_A = 0.25$ and $\theta = 45°$). Similar average errors were observed for $C_{p,rms}$ predicted by RF and XGB, but the standard deviation and maximum prediction error of RF were higher than XGB. It should be noted that RF needed a deeper (14 levels) and more complex tree structure than XGB (5 levels) to achieve similar accuracy. Moreover, XGB outperformed the other two regressors when predicting extreme pressures such as $C_{p,min}$, and $C_{p,max}$. For example, XGB's average prediction error for $C_{p,min}$ was 3% and 5.4% smaller than RF and DT's. XGB also had a smaller average prediction error for $C_{p,max}$ than RF, except for Case 5. However, the average prediction error of DT, which also has 14 levels, was 1%–14% higher than the ensemble tree models of XGB and RF across the validation cases. This confirms the superior models' performance of ensemble trees over a single DT. However, of note in Fig. 8 is that deviations of $C_{p,max}$ from wind tunnel test data are larger than that of $C_{p,min}$ in contrary to its high overall prediction accuracy observed in Fig. 6. This observation indicates considerable variations in local prediction accuracy, which cannot be discerned from the error graphs of overall prediction accuracy.

Fig. 9 illustrates the distribution of $C_{p,mean}$ on the building's walls and the roof at $\theta = 67.5°$ as measured in wind tunnel tests and predicted by XGB. Overall, the variations in wind pressure and special features of $C_{p,mean}$ captured by XGB were similar to the wind tunnel test results. For example, XGB's prediction and wind tunnel test results both showed intense suction pressure along the windward edge of the roof. XGB's prediction of $C_{p,mean}$ on Wall 1 was also similar to the wind tunnel tests' observations — an increasing trend of mean wind pressure from the ground toward the roof. The percentage error of XGB's $C_{p,mean}$ prediction for the roof and Wall 1 ($(C_{p,mean}\ of\ WT - C_{p,mean}\ of\ XGB)/\overline{C}_{p,mean}\ of\ WT$; where $C_{p,mean}$ of WT and $C_{p,mean}$ of XGB are mean pressure coefficients at each pressure taps from wind tunnel tests and XGB, and $\overline{C}_{p,mean}\ of\ WT$ is the average of absolute pressure coefficients of the wind tunnel test) was within −30%–50%. For Wall 2, XGB's percentage error was a maximum of ±30% with an average of −25%.

Table 2
Validation metrics calculated for the predictions of DT, RF, and XGB.

Pressure coefficient	Metric	DT	RF	XGB
$C_{p,mean}$	R	0.936	0.952	0.970
	R^2	0.880	0.910	0.940
	RMSE	0.092	0.079	0.061
	FB	−0.033	−0.030	−0.013
$C_{p,rms}$	R	0.944	0.962	0.970
	R^2	0.890	0.930	0.940
	RMSE	0.033	0.027	0.024
	FB	−0.004	−0.008	−0.011
$C_{p,min}$	R	0.901	0.923	0.937
	R^2	0.810	0.850	0.880
	RMSE	0.323	0.280	0.250
	FB	−0.010	−0.006	−0.006
$C_{p,max}$	R	0.934	0.953	0.962
	R^2	0.870	0.910	0.930
	RMSE	0.229	0.193	0.170
	FB	−0.007	−0.001	−0.015

6. XML application for model predictions

6.1. Intrinsic model explanation

Since intrinsic model explanations can be used for Decision Trees with simple structures, this study applied it onto DT. However, DT developed for the current study had a deep tree structure with 14 levels, which made it difficult to comprehensively elucidate the model's inner workings using intrinsic model explanations. Therefore, this study has chosen the first three levels of DT (Fig. 10) to demonstrate how intrinsic model explanations are applied to DT's $C_{p,mean}$ predictions.

Up to depth = 3, DT uses four features: $X[0] = \theta$, $X[1] = C_A$, $X[2] = x/H$, $X[3] = y/H$, and $X[4] = z/H$ for recursive binary splitting of the data. Since the splitting criterion was based on the minimization of the mean square error (MSE), the CART algorithm sequentially chose features to lower the MSE as the tree grew in depth. DT chose $x/H \leq 0.958$ for splitting data at the first (root) node, which in effect indicates that x/H was the most dominant feature for predicting $C_{p,mean}$. At the second level, θ and y/H were the features used for data splitting, thus signifying their high impact on predicting $C_{p,mean}$. The four nodes at the third level utilized three features, C_A, x/H, and θ, for data splitting. The repetitive uses of the features x/H and θ with different magnitudes highlight how different features impact the values of the predictions. In addition, a set of 'IF-THEN' rules can be inferred from tracing the DT (i.e., following the arrows) from top to bottom. For instance, the "IF-THEN" rule associated with the left-most arrows, marked red, can be stated as "IF $x/H \leq 0.958$ AND $\theta \leq 56.5°$ AND $C_A \leq 0.275$ AND … THEN *Prediction*". Note that the deep tree structure of DT which contained many AND conditions preceding *Prediction* at level 14 made it difficult to deduce simple and comprehensible relationships between the features and predictions. This shows that Decision Trees with deep tree structures do require post-hoc explanation methods such as SHAP to elucidate their model behavior.

6.2. SHAP global explanation

Fig. 11 shows how SHAP global explanations can interpret the way five features, x/H, C_A, y/H, θ, and z/H, contributed to XGB's predictions of $C_{p,mean}$, $C_{p,rms}$, $C_{p,min}$, and $C_{p,max}$. The features are ranked in descending order of contribution (i.e., feature importance) to each type of prediction. Feature importance is estimated by calculating mean absolute SHAP values for each feature multiplying the sign (i.e., positive, or negative) of its correlation to the relevant prediction. For $C_{p,mean}$, all features but C_A contributed positively, with the highest and lowest contributions from x/H and z/H, respectively (Fig. 11(a)). On average, x/H and y/H caused a 0.1 and 0.08 rise in $C_{p,mean}$, while z/H on average increased it by 0.02. This indicates a greater influence of horizontal position (x/H, and y/H) of pressure taps than their vertical position (z/H). θ also contributed less to the prediction of $C_{p,mean}$ (SHAP value = +0.04), while C_A (SHAP = −0.02) only decreased $C_{p,mean}$ slightly. x/H was also the largest positive contributor (SHAP = +0.05) to $C_{p,rms}$,

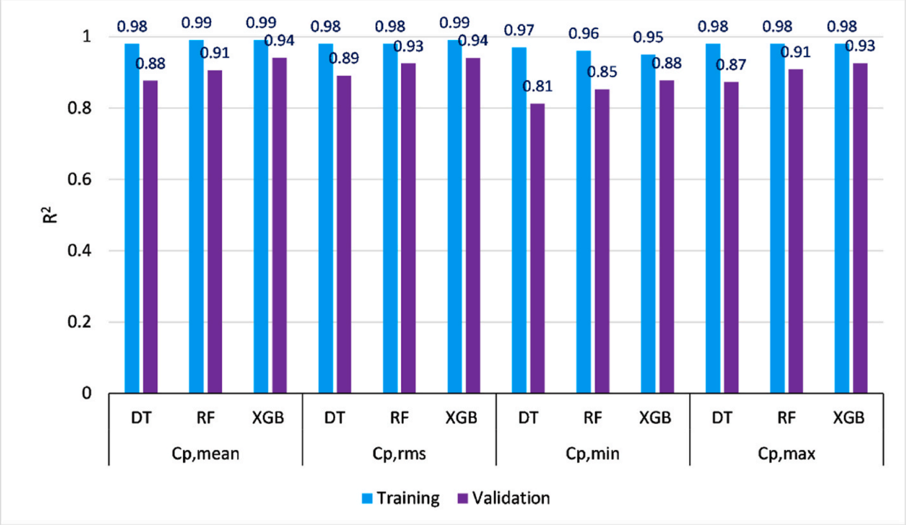


Fig. 7. Comparison of R^2 of DT, RF, and XGB for predicting $C_{p,mean}$, $C_{p,rms}$, $C_{p,min}$, and $C_{p,max}$ for the training and validation datasets.

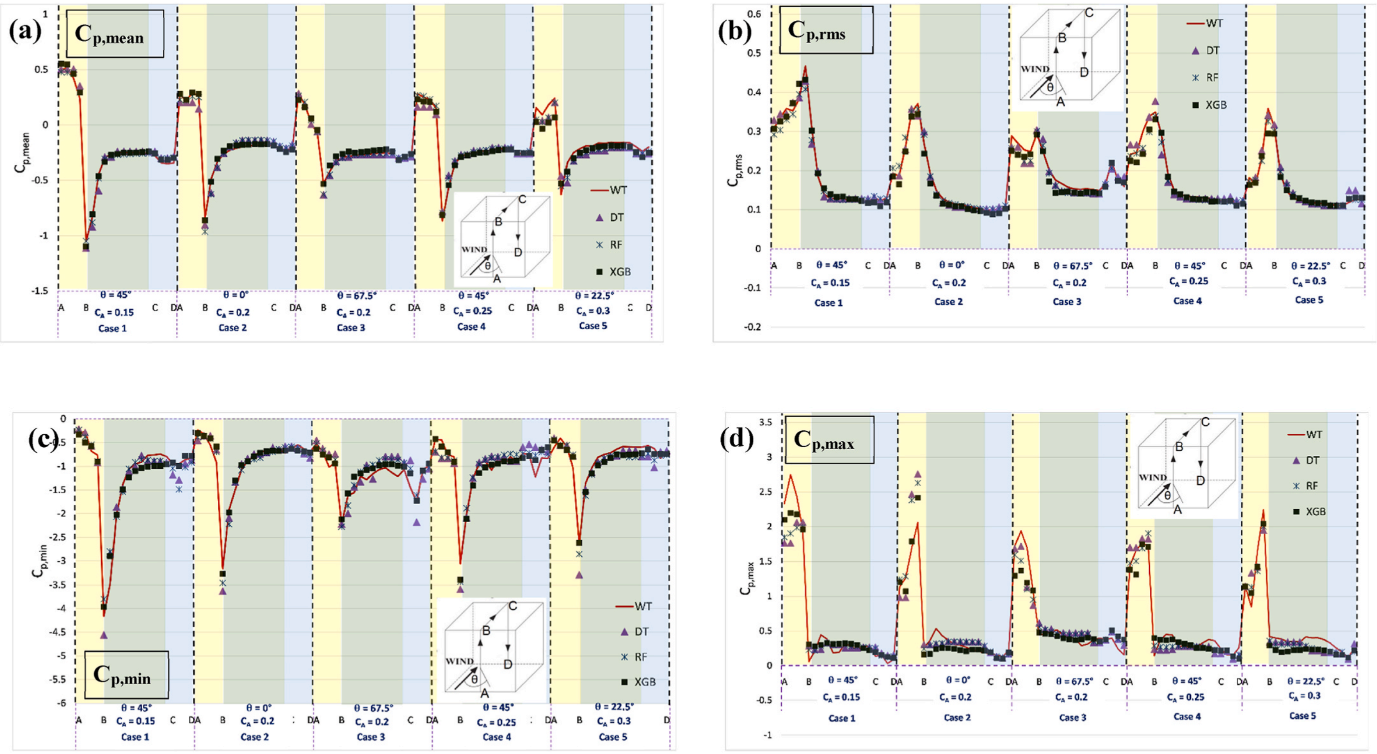


Fig. 8. Comparison of ML models' predictions with the wind tunnel test data of (a) $C_{p,mean}$, (b) $C_{p,rms}$, (c) $C_{p,min}$, and (d) $C_{p,max}$ along the vertical center plane of the target building in the surrounding with area density (C_A) = 0.1, 0.15, 0.20, 0.25, 0.30, 0.4.

followed by y/H (SHAP = +0.03) (Fig. 11(b)). In contrast to $C_{p,mean}$, C_A increased $C_{p,rms}$ by 0.02 while θ and z/H tended to decrease $C_{p,rms}$ by 0.02 and 0.01, respectively. For $C_{p,min}$, θ and z/H were the positive contributors, which decreased negative maximum pressure, while the rest of the inputs increased $C_{p,min}$, with the highest contribution coming from x/H(SHAP = +0.31)(Fig. 11(c)). Interestingly, the type of feature contributions to $C_{p,max}$ was polar opposite to that of $C_{p,min}$ despite magnitudes were different. For example, the contributions of x/H and y/H to $C_{p,max}$ were positive and approximately similar in magnitudes (+0.26, and +0.25) (Fig. 11(d)). Both θ and z/H had negative impacts on $C_{p,max}$ while the latter magnitude was a half of the former in contrast to similar positive magnitudes found for $C_{p,min}$.

Fig. 11 provides an overall picture of how each feature contributed to the predictions of XGB but does not show how each feature's contribution varied with its magnitude. SHAP summary plots (Fig. 12) offer such information by combining feature contribution (SHAP values on the x-axis), feature importance (y-axis), and feature value (color of marker). For example, low feature values of x/H had high impact on the predictions of both positive and negative $C_{p,mean}$ (Fig. 12(a)), while its high values contributed less to the predictions. In contrast, large and small magnitudes of z/H had distinct positive and negative contributions to $C_{p,mean}$, respectively. However, such a distinct division in contribution for $C_{p,mean}$ cannot be identified for some features such as θ. Other than the type of contribution, Fig. 12 indicates possible relationships between the

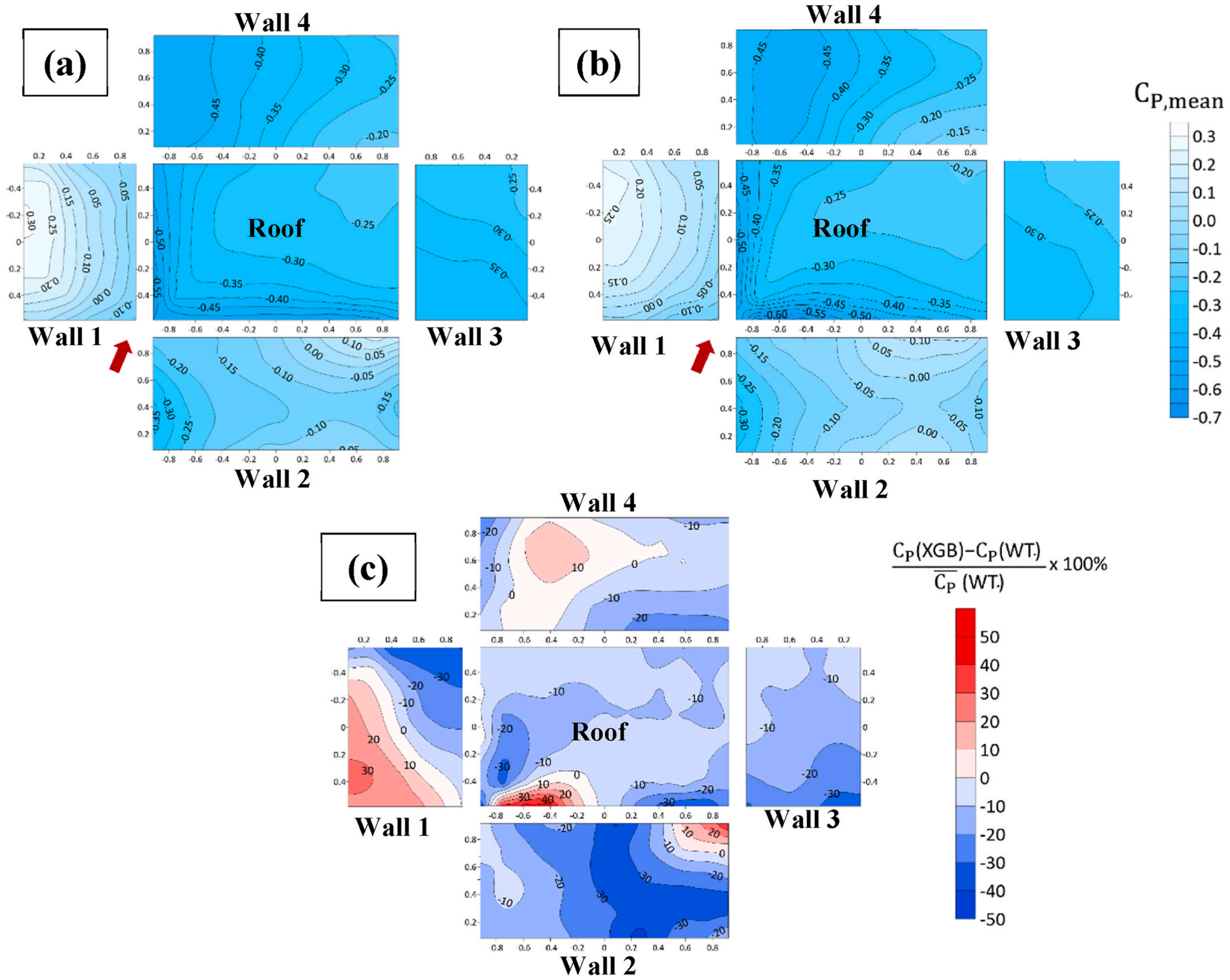


Fig. 9. Distribution of $C_{p,mean}$ on the target building in the surrounding with $C_A = 0.2$ at $\theta = 67.5°$ from (a) wind tunnel test, (b) XGB, (c) percentage of error of XGB.

inputs and outputs. For instance, Fig. 12(b) depicts inversely proportional relationships between fluctuating pressure ($C_{p,rms}$) and x/H, y/H, and C_A and a direct proportionality with z/H. These relationships were inverted for $C_{p,min}$, as x/H, y/H, C_A had a directly proportional while z/H had an inversely proportional relationship with $C_{p,min}$ (Fig. 12(c)). Interestingly, such distinct relationships cannot be established for θ with any pressure coefficient, as its high and low values contributed both positively and negatively to model output. The contributions of features for $C_{p,max}$ and $C_{p,min}$ were polar opposites of each other (Fig. 12(d)). For example, all x/H, y/H, and C_A had inversely proportional relationships with $C_{p,max}$, while maximum pressure was directly proportional to z/H and θ.

6.3. SHAP local explanation

SHAP summary plots, as in Fig. 12, can also be used as explanations for local pressure predictions. Local explanations can be provided for specific zones by clustering pressure taps in those areas or point-like locations using individual pressure taps. Fig. 13 shows such local explanations for $C_{p,mean}$ on the roof, and Walls 1, 2, and 3 using SHAP summary plots. For the interpretation, note that the constant features of the selected areas such as of z/H of the roof or x/H of Wall 1 appear as low feature values (marked blue) in the SHAP summary plots. Fig. 13(a) presents x/H as the dominant factor for predicting $C_{p,mean}$ on the roof, followed by y/H, θ, and C_A. Furthermore, Fig. 13(a) shows direct proportionalities between $C_{p,mean}$ of the roof and x/H, y/H, and C_A. Extreme low and high values of C_A, in particular, induced distinct negative and positive effects on $C_{p,mean}$ on the roof. On the other hand, as of the global explanations, θ had mixed contributions to $C_{p,mean}$ of the roof across values tested in the wind tunnel. At the roof level, the constant z/H's effect was minimum, as it ranked as the least important feature on the list. However, its SHAP value varied between -0.12 and 0.2 as many pressure taps at that height were influenced to various degrees by other features. In contrast to the global explanations in Fig. 12, θ was the second most dominant factor (after $x/H = -1$) for $C_{p,mean}$ on Wall 1 (Fig. 13(b)). There, the impact of θ varied with $C_{p,mean}$ in two distinct ways: first, low to moderate values of θ increased $C_{p,mean}$ on Wall 1; second, its extreme high values such as $\theta = 90°$ always reduced $C_{p,mean}$. Similarly, high C_A values, which related to closely arranged buildings, reduced $C_{p,mean}$ on Wall 1, while its extreme low values increased mean wind pressure. Interestingly, $C_{p,mean}$ on Wall 1 had inversely proportional relationships with features with low values, except for x/H. In

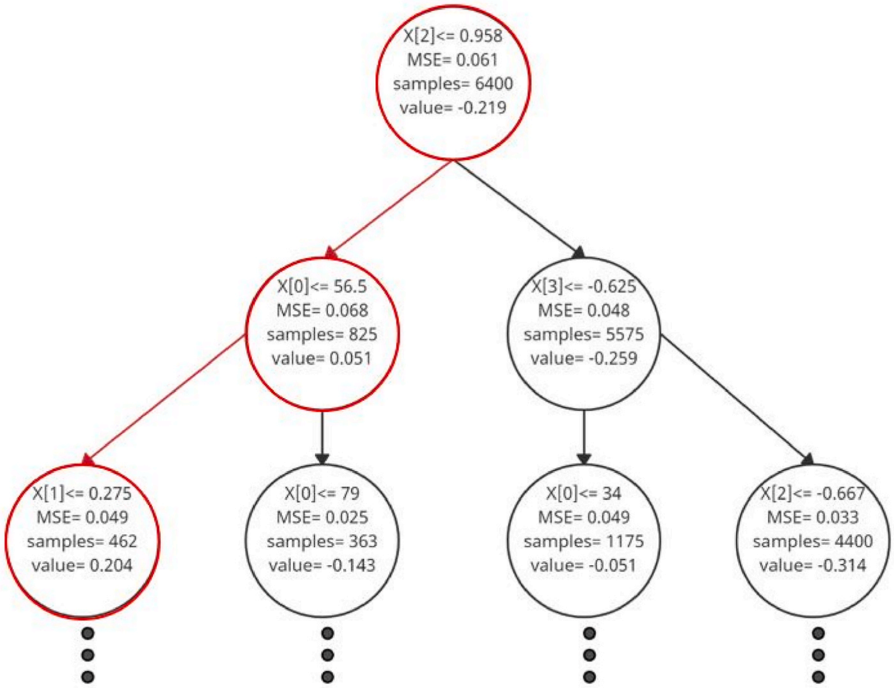


Fig. 10. The structure of the DT for predicting $C_{p,mean}$ (up to depth = 3).

contrast, $C_{p,mean}$ of Wall 2 displayed direct proportionalities with θ and x/H — the two most dominant features of the case (Fig. 13(c)). The other two features, z/H and C_A, had limited but indistinguishable types of contribution to $C_{p,mean}$ on Wall 2 across the wind tunnel test values. On Wall 3, $C_{p,mean}$ was mostly governed by a negative contribution from y/H for its entire range of values (Fig. 13(d)). The second most dominant feature was θ for $C_{p,mean}$ on Wall 3 with positive or negative contributions of θ strongly dependent on its low and high feature values. Interestingly, though C_A and z/H ranked third and fourth for feature importance for $C_{p,mean}$ on Wall 3, their contributions could still be distinctly divided as positive and negative, according to their high and low feature values.

Figs. 14 and 15 demonstrate SHAP local explanations for $C_{p,mean}$, and $C_{p,min}$ predicted by XGB at two pressure taps. One of the pressure taps (e12 in Fig. 4) was at 3/4 height of Wall 1 ($x/H = -1$; $y/H = -0.25$; $z/H = 0.75$), and the other (g7 in Fig. 4) was at a corner of the roof ($x/H = -0.92$; $y/H = 0.58$; $z/H = 1$). In each SHAP explanation plot, the base value indicates average $C_{p,mean}$, or $C_{p,min}$ estimated for all test cases. The prediction $f(x)$ for a given case and its magnitude are boldfaced. The red and blue arrows indicate an increase and decrease in $f(x)$ from the base value, as a result of positive and negative contributions from the features. The magnitude of contribution is proportional to the arrow's

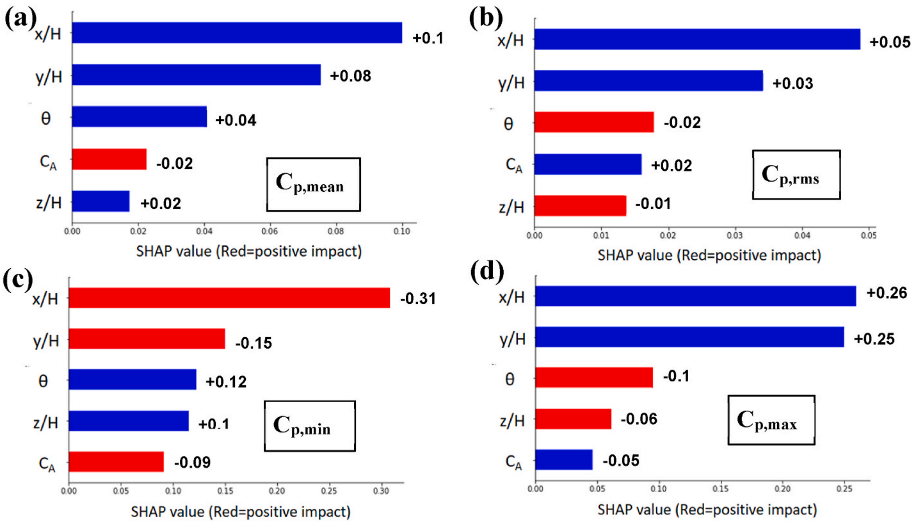


Fig. 11. SHAP feature importance measured as the mean absolute SHAP values for the predictions of (a) $C_{p,mean}$, (b) $C_{p,rms}$, (c) $C_{p,min}$, and (d) $C_{p,max}$ by the XGB.

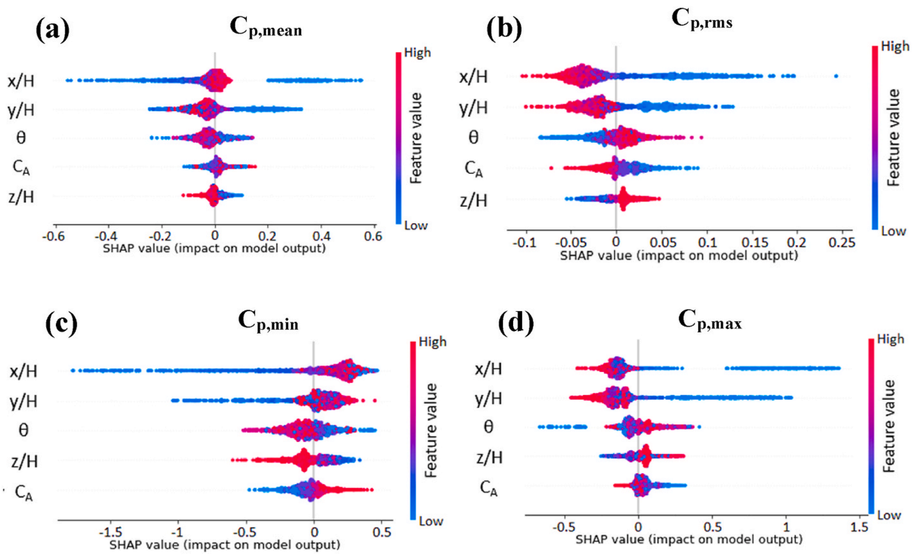


Fig. 12. SHAP summary plot for (a) $C_{p,mean}$, (b) $C_{p,rms}$, (c) $C_{p,min}$, and (d) $C_{p,max}$ predicted by XGB.

length. For instance, the base value of $C_{p,mean}$ at pressure tap e12 was −0.2188 but increased to 0.53 for the case with $\theta = 45°$ and $C_A = 0.15$ (Fig. 14(a)). The rise of $C_{p,mean}$ is attributed to the positive contributions of all features. In contrast, θ had a negative impact on $C_{p,mean}$ for the case with $\theta = 67.5°$ and $C_A = 0.2$, but its influence was overcome by the collective positive contributions from *x/H*, *y/H*, and C_A (Fig. 14(b)). However. The resultant $C_{p,mean} = 0.08$ was considerably smaller than in the previous case. Case 3 with $\theta = 0°$ and $C_A = 0.2$ also had positive contributions from all features similar to Case 1(Fig. 14(c)), but with smaller magnitudes compared to Case 1. These smaller positive contributions increased C_{pmean} from its base value but to a lower value of 0.37.

Compared to dominant positive feature contributions at pressure tap e12, extreme negative pressure ($C_{p,min}$) at the pressure tap (g7) at the corner of the roof was shaped by both positive and negative feature contributions (Fig. 15). Its base value $C_{p,min} = -1.148$ indicates intense suction pressure often detected at the corners of flat roofs (Surry, 1991; Sarkar et al., 1997; Wu et al., 2001). The features *x/H*, θ, and *z/H* further increased $C_{p,min}$ of Case 1 with $\theta = 22.5°$ and $C_A = 0.3$, while C_A and *y/H* reduced its magnitude (Fig. 15(a)). In particular, the highest contribution of *x/H* points to the disadvantageous geometric location, which was highly vulnerable to adverse suction pressure. However, $C_{p,min}$ decreased from −2.12 of Case 1 to −1.66 of Case 2 as θ increased from 22.5° to 67.5°, and C_A decreased from 0.3 to 0.2 (Fig. 15(b)). This reduction is attributed to the positive contribution of $\theta = 67.5°$ of Case 2, contrary to its negative contribution at $\theta = 22.5°$ in Case 1. However, the moderately dense building arrangement of Case 2 ($C_A = 0.2$) tended to increase negative extreme pressure at the roof' corner compared to the reduction observed for the densely arranged buildings of Case 1 ($C_A =$

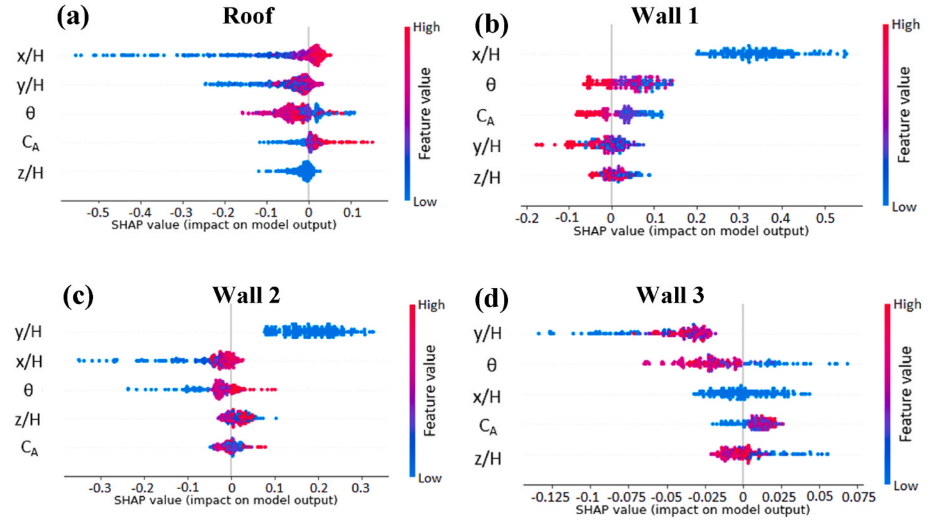


Fig. 13. SHAP summary plot for $C_{p,mean}$ on (a) roof, (b) wall 1, (c) wall 2, and (d) wall 3 predicted by XGB.

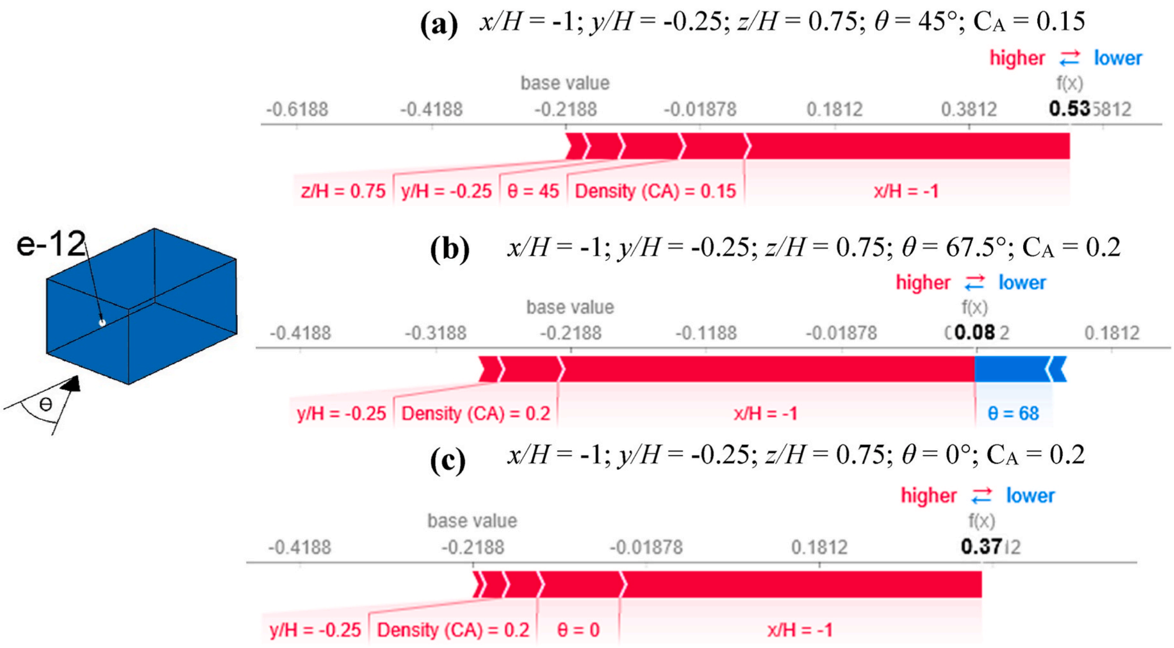


Fig. 14. SHAP local explanation for $C_{p,mean}$ of the pressure tap (e12) for (a) $\theta = 45^\circ$ and $C_A = 0.15$, (b) $\theta = 67.5^\circ$ and $C_A = 0.2$, and (c) $\theta = 0^\circ$ and $C_A = 0.2$ (θ – wind incidence angle; C_A – area density).

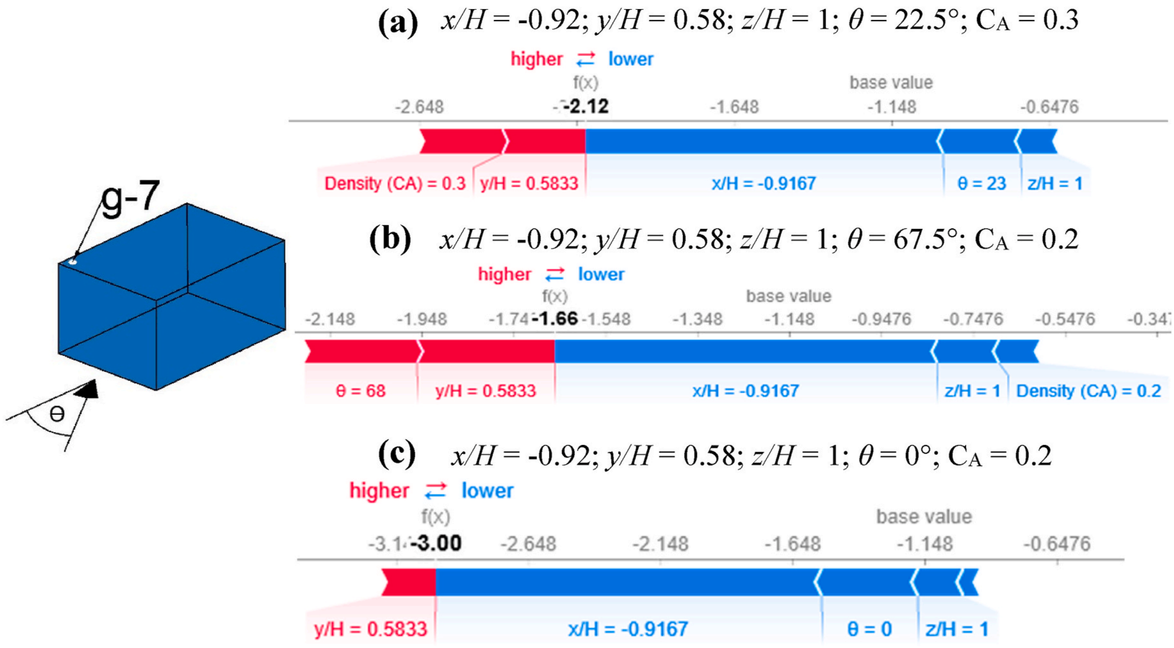


Fig. 15. SHAP local explanation for $C_{p,min}$ of the pressure tap (g7) for (a) $\theta = 22.5^\circ$ and $C_A = 0.3$, (b) $\theta = 67.5^\circ$ and $C_A = 0.2$, and (c) $\theta = 0^\circ$ and $C_A = 0.2$ (θ – wind incidence angle; C_A – area density).

0.3). The negative impact of $C_A = 0.2$ became smaller at $\theta = 0°$ in Case 3 but the wind incidence angle contributed to a further decrease of $C_{p,min}$ to its lowest value of -3.0 (Fig. 15(c)). Therefore, the building configuration of Case 3 exemplified the worst conditions for generating the severest suction pressure at the windward corner of a flat roof.

7. Concluding remarks

The following includes several important findings of applying ML and XML in wind engineering, particularly to predict external wind pressure, as identified by this study:

- Ensemble tree regressors such as RF and XGB are more accurate in predicting external wind pressure than a single decision tree (DT). However, the DT, and also the RF require deeper trees to achieve similar accuracy as XGB. The deep tree structures of DT and RF and/or a complex ensemble of many trees (such as the case of RF and DT) prevent the use of intrinsic explanation methods, thus requiring post-hoc interpretation techniques to be applied to explain model behaviors.
- Advanced ML models such as XGB are advantageous for capturing the effect of location, wind incidence angle, and characteristics of the surroundings on mean, fluctuating, and extreme wind pressures. These models can be used to accurately estimate (with less than ±20% error) any of $C_{p,mean}$, $C_{p,rms}$, $C_{p,min}$, or $C_{p,max}$ on the external walls of buildings, or a specific zone (e.g., roof), or a point location.
- SHAP can reveal the type of relationship (directly or inversely proportional) between the geometric parameters of a building, the wind incidence angle, buildings in the surroundings, and external wind pressure. It can also estimate the impacts of features for a range of feature values. This knowledge can be employed to select critical design factors for optimizing building designs to minimize wind effects in the early design stage.

A set of empirical evidence can be formed by using SHAP local explanations to evaluate the causality of ML models. For example, SHAP local explanations for wind pressure at critical building locations such as the windward corner of the roof, at 2/3 height of the windward wall, a point closer to the windward edge of a sidewall can be compared with well-established scenarios for external wind pressure at these locations. For example, SHAP indication of the shift of the contribution of θ for $C_{p,mean}$ on the windward wall shifts from positive to negative as θ changes from 0° to 90° can be explained as the wall's direct exposure to incoming wind at $\theta = 0°$ and under the separated wind flow at $\theta = 90°$.

8. Limitations and future work

Although this study has provided some valuable insight on employing SHAP as an XML to elucidate wind pressure predictions by XGB for a low-rise building in an urban-like setting, it also has some limitations in scope and methodology, which should be avoided and rectified in future research as discussed below:

- The three tree regressors were trained for the current study using five parameters: x/H, y/H, z/H, θ, and C_A as inputs for predicting the outputs $C_{p,mean}$, $C_{p,rms}$, $C_{p,min}$, $C_{p,max}$. However, the TPU aerodynamic database contains data with numerous other design parameters such as various roof types (e.g., hip, gable), roof angles, building arrangements (e.g., staggered, and random), building dimensions, etc. It is also necessary to consider how pressure distribution of the target building changes with wind speed and turbulence intensity of approaching winds for accurate ML models' predictions. Adding these parameters to the ML models will increase complexity but provide a great opportunity to explore ML and XML's capability of handling sophisticated wind engineering applications.
- The current study chose tree-based regression models because their simplest (e.g., single decision tree) to complex model structures (e.g., RF and XGB) allow intrinsic and post-hoc explanation methods to be applied. However, this study does not rule out the possibility that other advanced ML models such as ANN, deep neural networks (DNN), generative adversarial network (GAN), etc may have superior performance in predicting external wind pressure on buildings. Therefore, it is necessary to examine the accuracy of other advanced ML models than tree regressors for wind engineering applications.
- It is also imperative to apply XML for various ML models to investigate how differently ML models use inputs, establish relationships between inputs and outputs, and identify feature importance for predictions. For instance, Park and Park (2021) have shown differences in feature importance selected by eight ML models to predict the natural ventilation of a building using the same dataset.
- SHAP has proved its competence as a post-hoc XML method by providing very useful and accurate explanations. On the other hand, instead of SHAP, several other XML techniques such as LIME, Anchors, counterfactual explanations can be applied to explain ML model predictions. These XML methods perhaps point out different feature importance, relationships between inputs and outputs, feature contribution, etc. Therefore, it is recommended that these other XML techniques be used for the same set of predictions to compare whether all post-hoc methods produce similar explanations, as well as feature importance and rankings.

CRediT authorship contribution statement

D.P.P. Meddage: Methodology, Investigation, Writing – original draft. **I.U. Ekanayake:** Methodology, Investigation. **A.U. Weerasuriya:** Conceptualization, Supervision, Writing – review & editing. **C.S. Lewangamage:** Supervision, Funding acquisition. **K.T. Tse:** Funding acquisition, Writing – review & editing. **T.P. Miyanawala:** Writing – review & editing. **C.D.E. Ramanayaka:** Writing – review & editing.

Declaration of competing interest

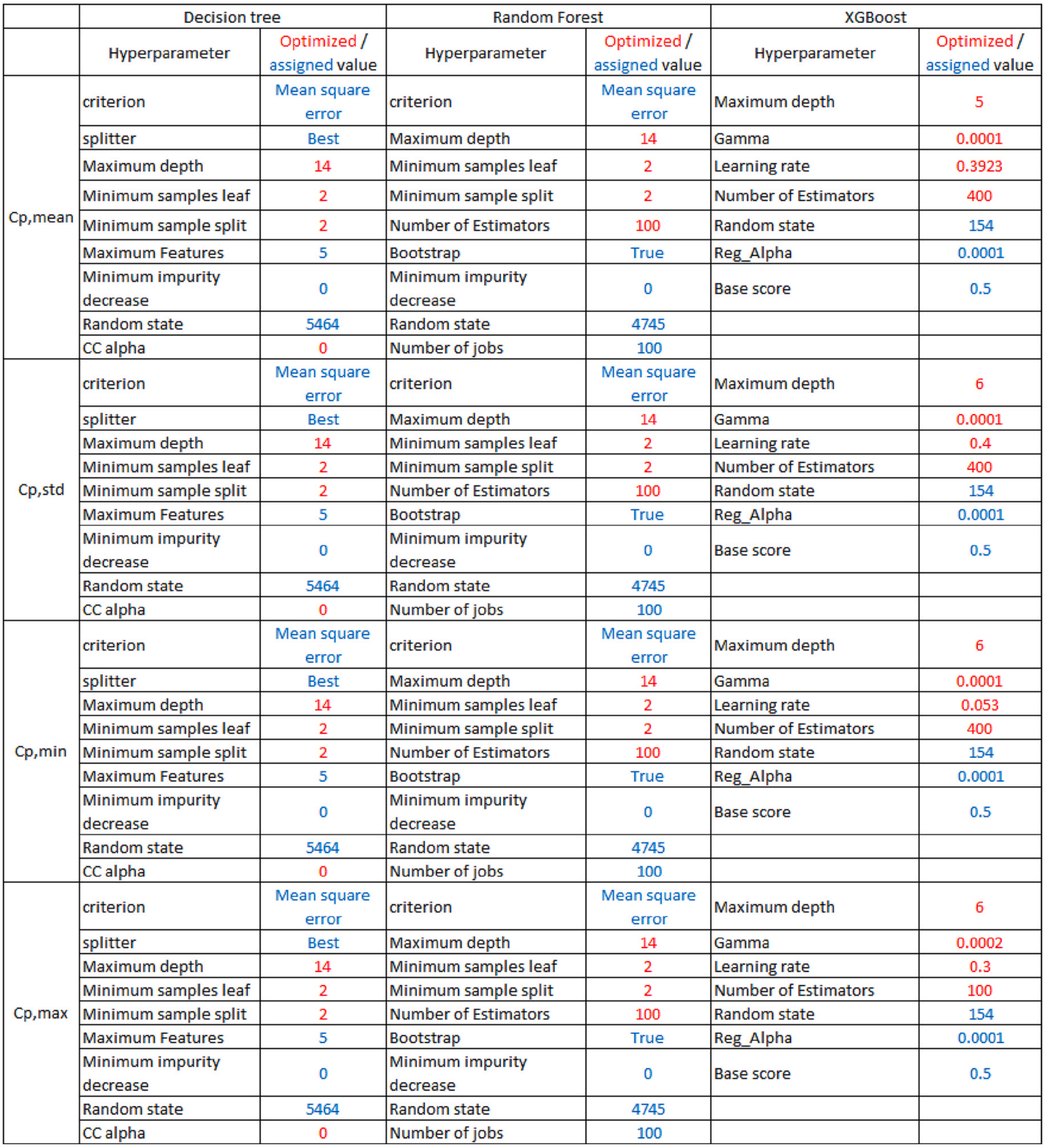
The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

This research study was partly funded by National Science Foundation, Sri Lanka, through grant No. OSTP/2019/17and by the Hong Kong (HK) Research Grants Council (RGC) General Research Fund (GRF) 16207118.

APPENDIX A

Fig. A1. Variation in correlation coefficient (R2) of training and validation of three ML models: DT, RF, and XGB with size of training and validation datasets[]



	Decision tree		Random Forest		XGBoost	
	Hyperparameter	Optimized / assigned value	Hyperparameter	Optimized / assigned value	Hyperparameter	Optimized / assigned value
Cp,mean	criterion	Mean square error	criterion	Mean square error	Maximum depth	5
	splitter	Best	Maximum depth	14	Gamma	0.0001
	Maximum depth	14	Minimum samples leaf	2	Learning rate	0.3923
	Minimum samples leaf	2	Minimum sample split	2	Number of Estimators	400
	Minimum sample split	2	Number of Estimators	100	Random state	154
	Maximum Features	5	Bootstrap	True	Reg_Alpha	0.0001
	Minimum impurity decrease	0	Minimum impurity decrease	0	Base score	0.5
	Random state	5464	Random state	4745		
	CC alpha	0	Number of jobs	100		
Cp,std	criterion	Mean square error	criterion	Mean square error	Maximum depth	6
	splitter	Best	Maximum depth	14	Gamma	0.0001
	Maximum depth	14	Minimum samples leaf	2	Learning rate	0.4
	Minimum samples leaf	2	Minimum sample split	2	Number of Estimators	400
	Minimum sample split	2	Number of Estimators	100	Random state	154
	Maximum Features	5	Bootstrap	True	Reg_Alpha	0.0001
	Minimum impurity decrease	0	Minimum impurity decrease	0	Base score	0.5
	Random state	5464	Random state	4745		
	CC alpha	0	Number of jobs	100		
Cp,min	criterion	Mean square error	criterion	Mean square error	Maximum depth	6
	splitter	Best	Maximum depth	14	Gamma	0.0001
	Maximum depth	14	Minimum samples leaf	2	Learning rate	0.053
	Minimum samples leaf	2	Minimum sample split	2	Number of Estimators	400
	Minimum sample split	2	Number of Estimators	100	Random state	154
	Maximum Features	5	Bootstrap	True	Reg_Alpha	0.0001
	Minimum impurity decrease	0	Minimum impurity decrease	0	Base score	0.5
	Random state	5464	Random state	4745		
	CC alpha	0	Number of jobs	100		
Cp,max	criterion	Mean square error	criterion	Mean square error	Maximum depth	6
	splitter	Best	Maximum depth	14	Gamma	0.0002
	Maximum depth	14	Minimum samples leaf	2	Learning rate	0.3
	Minimum samples leaf	2	Minimum sample split	2	Number of Estimators	100
	Minimum sample split	2	Number of Estimators	100	Random state	154
	Maximum Features	5	Bootstrap	True	Reg_Alpha	0.0001
	Minimum impurity decrease	0	Minimum impurity decrease	0	Base score	0.5
	Random state	5464	Random state	4745		
	CC alpha	0	Number of jobs	100		

Fig. A2. The selected hyperparameters for the DT, RF, and XGB by the grid search method[]

APPENDIX B

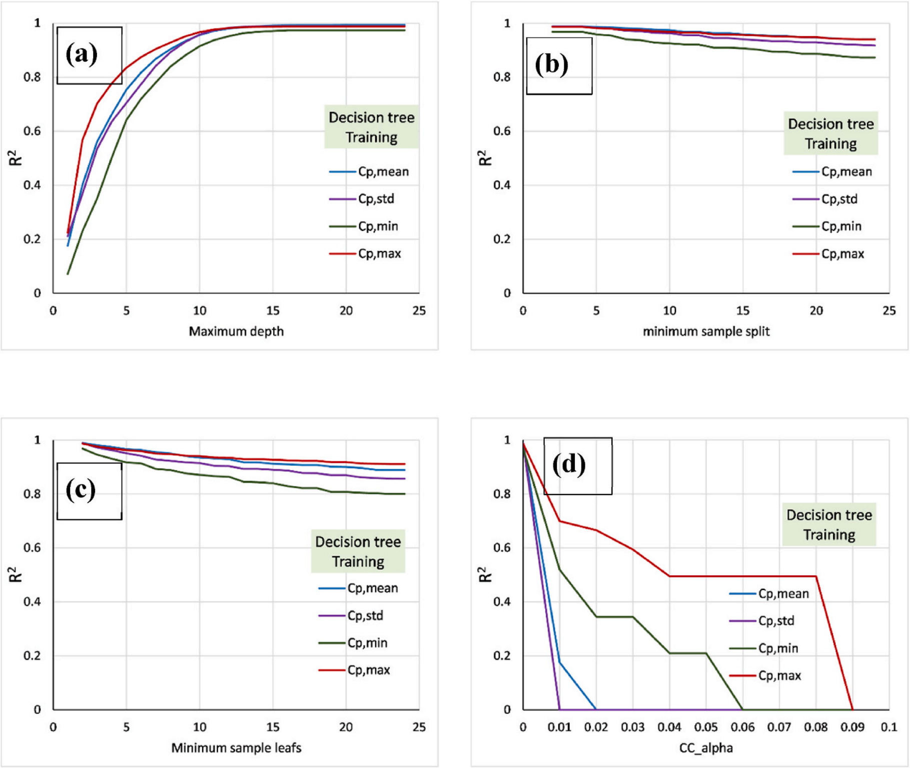


Fig. B1. The variation in coefficient of determination (R2) with (a) maximum depth, (b) minimum sample split, (c) minimum sample leaf, and (d) CC_alpha of the DT[]

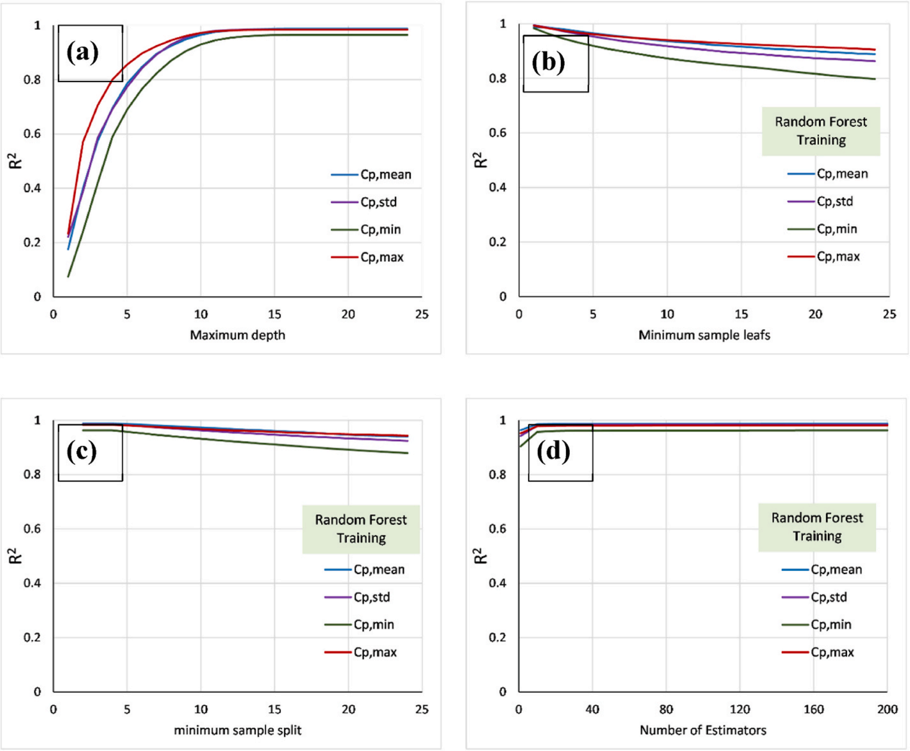


Fig. B2. The variation in coefficient of determination (R^2) with (a) maximum depth, (b) minimum sample leaf, (c) minimum sample split, and (d) Number of estimators of the RF[]

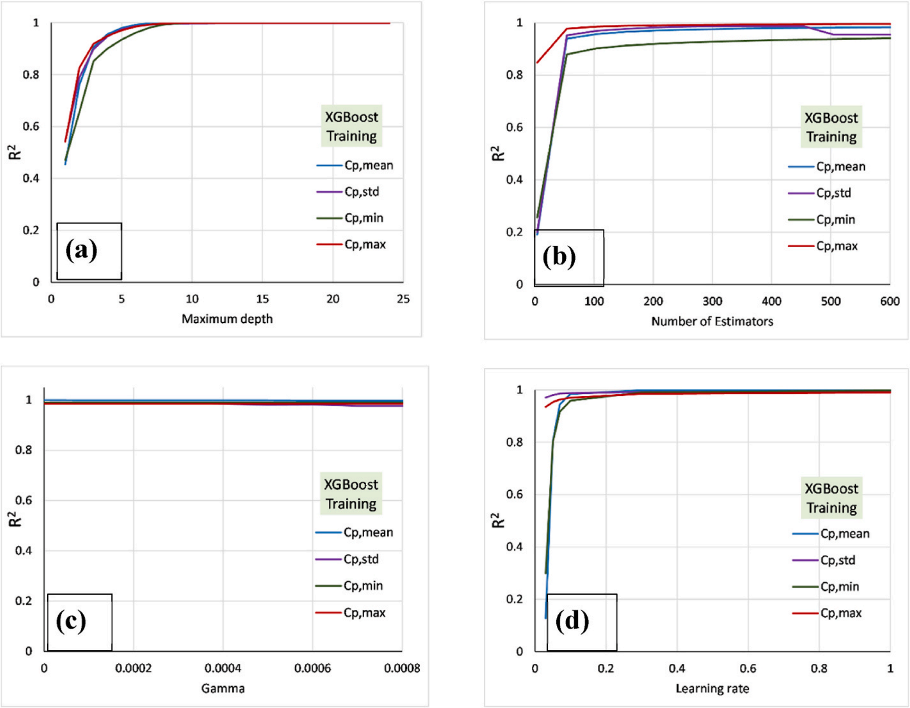


Fig. B3. The variation in coefficient of determination (R^2) with (a) maximum depth, (b) the minimum number of estimators, (c) γ, and (d) learning rate of the XGB[]

References

AIJ, 2004. Recommendations on Loads for Buildings. Architectural Institute of Japan, Japan.

Akon, Abul Fahad, Gregory, A., 2018. Kopp. "Turbulence structure and similarity in the separated flow above a low building in the atmospheric boundary layer. J. Wind Eng. Ind. Aerod. 182, 87–100.

Andrzejak, A., Langner, F., Zabala, S., 2013. April). Interpretable models from distributed data via merging of decision trees. In: IEEE Symposium on Computational Intelligence and Data Mining (CIDM). IEEE, pp. 1–9, 2013.

Aristodemou, E., Boganegra, L.M., Mottet, L., Pavlidis, D., Constantinou, A., Pain, C., et al., 2018. How tall buildings affect turbulent air flows and dispersion of pollution within a neighbourhood. Environ. Pollut. 233, 782–796.

Belle, V., Papantonis, I., 2021. Principles and practice of explainable machine learning. Front. big Data 39.

Biao, L., Cunyan, J., Lu, W., Weihua, C., Jing, L., 2019. A parametric study of the effect of building layout on wind flow over an urban area. Build. Environ. 160, 106160.

Biran, O., Cotton, C., 2017, August. Explanation and justification in machine learning: a survey. IJCAI-17 workshop on explainable AI (XAI) 8 (No. 1), 8–13.

Bre, F., Gimenez, J.M., Fachinotti, V.D., 2018. Prediction of wind pressure coefficients on building surfaces using artificial neural networks. Energy Build. 158, 1429–1441.

Breiman, L., Friedman, J., Olshen, R., Stone, C., 1984. A Classification and Regression Tree. Wadsworth International Group Chapman and Hall, USA.

Breiman, L., 1996. Bagging predictors. Mach. Learn. 24 (2), 123–140.

Breiman, L., 2001. Random forests. Mach. Learn. 45 (1), 5–32.

Brunton, S.L., Hemati, M.S., Taira, K., 2020a. Special issue on machine learning and data-driven methods in fluid dynamics. Theor. Comput. Fluid Dynam. 34, 333–337, 2020.

Brunton, S.L., Noack, B.R., Koumoutsakos, P., 2020b. Machine learning for fluid mechanics. Annu. Rev. Fluid Mech. 52, 477–508.

Cermak, J.E., 1984. Wind-simulation criteria for wind-effect tests. J. Struct. Eng. 110 (2), 328–339.

Chang, J.C., Hanna, S.R., 2004. Air quality model performance evaluation. Meteorol. Atmos. Phys. 87 (1), 167–196.

Charisi, S., Waszczuk, M., Thiis, T.K., 2021. Determining building-specific wind pressure coefficients to account for the microclimate in the calculation of air infiltration in buildings. Adv. Build. Energy Res. 15 (3), 368–389.

Chen, T., Guestrin, C., 2016, August. Xgboost: a scalable tree boosting system. In: Proceedings of the 22nd Acm Sigkdd International Conference on Knowledge Discovery and Data Mining, pp. 785–794.

Chen, Y., Kopp, G.A., Surry, D., 2002. Interpolation of wind-induced pressure time series with an artificial neural network. J. Wind Eng. Ind. Aerod. 90 (6), 589–615.

Choudhury, A., Gupta, D., 2019. A survey on medical diagnosis of diabetes using machine learning techniques. In: Recent Developments in Machine Learning and Data Analytics. Springer, Singapore, pp. 67–78.

Ding, F., Kareem, A., 2020. Tall buildings with dynamic facade under winds. Eng. 6 (12), 1443–1453.

Egas-López, J.V., Gosztolya, G., 2020, October. Predicting a cold from speech using Fisher vectors; SVM and XGBoost as classifiers. In: International Conference on Speech and Computer. Springer, Cham, pp. 145–155.

Efron, B., Tibshirani, R.J., 1994. An Introduction to the Bootstrap. CRC press.

Elshaer, A., Gairola, A., Adamek, K., Bitsuamlak, G., 2017. Variations in wind load on tall buildings due to urban development. Sustain. Cities Soc. 34, 264–277.

English, E.C., 1990. Shielding factors from wind-tunnel studies of prismatic structures. J. Wind Eng. Ind. Aerod. 36, 611–619.

Fernández-Cabán, P.L., Masters, F.J., 2018. Effects of freestream turbulence on the pressure acting on a low-rise building roof in the separated flow region. Front. Built. Environ. 4, 17.

Fu, J.Y., Liang, S.G., Li, Q.S., 2007. Prediction of wind-induced pressures on a large gymnasium roof using artificial neural networks. Comput. Struct. 85 (3–4), 179–192.

Fu, J.Y., Wu, J.R., Xu, A., Li, Q.S., Xiao, Y.Q., 2012. Full-scale measurements of wind effects on guangzhou west tower. Eng. Struct. 35, 120–139.

Fukami, K., Fukagata, K., Taira, K., 2020. Assessment of supervised machine learning methods for fluid flows. Theor. Comput. Fluid Dynam. 34 (4), 497–519.

Goricsán, I., Balczó, M., Balogh, M., Czáder, K., Rákai, A., Tonkó, C., 2011. Simulation of flow in an idealised city using various CFD codes. Int. J. Environ. Pollut. 44 (1–4), 359–367.

Gottard, A., Vannucci, G., Marchetti, G.M., 2020. A note on the interpretation of tree-based regression models. Biom. J. 62 (6), 1564–1573.

Gough, H.L., Luo, Z., Halios, C.H., King, M.F., Noakes, C.J., Grimmond, C.S.B., Quinn, A. D., 2018. Field measurement of natural ventilation rate in an idealised full-scale building located in a staggered urban array: comparison between tracer gas and pressure-based methods. Build. Environ. 137, 246–256.

Haldar, M., 2015. How Much Training Data Do You Need? (Last access on. https://medium.com/@malay.haldar/how-much-training-data-do-you-need-da8ec091e956. (Accessed 4 March 2022).

Han, S., Kim, H., 2019. On the optimal size of candidate feature set in random forest. Appl. Sci. 9 (5), 898.

Hastie, T., Tibshirani, R., Friedman, J., 2009. The Elements of Statistical Learning: Data Mining, Inference, and Prediction, second ed. Springer-Verlag.

Hu, G., Kwok, K.C.S., 2020. Predicting wind pressures around circular cylinders using machine learning techniques. J. Wind Eng. Ind. Aerod. 198, 104099.

Hu, G., Liu, L., Tao, D., Song, J., Tse, K.T., Kwok, K.C.S., 2020. Deep learning-based investigation of wind pressures on tall building under interference effects. J. Wind Eng. Ind. Aerod. 201, 104138.

Hu, G., Song, J., Hassanli, S., Ong, R., Kwok, K.C., 2019. The effects of a double-skin façade on the cladding pressure around a tall building. J. Wind Eng. Ind. Aerod. 191, 239–251.

Hui, Y., Tamura, Y., Yang, Q., 2017. Analysis of interference effects on torsional moment between two high-rise buildings based on pressure and flow field measurement. J. Wind Eng. Ind. Aerod. 164, 54–68.

Hui, Y., Tamura, Y., Yoshida, A., Kikuchi, H., 2013. Pressure and flow field investigation of interference effects on external pressures between high-rise buildings. J. Wind Eng. Ind. Aerod. 115, 150–161.

Hur, S.H., 2021. Short-term wind speed prediction using Extended Kalman filter and machine learning. Energy Rep. 7, 1046–1054.

Kareem, A., 2020. Emerging frontiers in wind engineering: computing, stochastics, machine learning and beyond. J. Wind Eng. Ind. Aerod. 206, 104320.

Kass, G.V., 1980. An exploratory technique for investigating large quantities of categorical data. J. Roy. Stat. Soc.: Ser. C (Appl. Stat.) 29 (2), 119–127.

Khanduri, A.C., Stathopoulos, T., Bédard, C., 1998. Wind-induced interference effects on buildings—a review of the state-of-the-art. Eng. Struct. 20 (7), 617–630.

Khosravi, A., Machado, L., Nunes, R.O., 2018. Time-series prediction of wind speed using machine learning algorithms: a case study Osorio wind farm, Brazil. Appl. Energy 224, 550–566.

Kim, B., Khanna, R., Koyejo, O.O., 2016. Examples are not enough, learn to criticize! criticism for interpretability. Adv. Neural Inf. Process. Syst. 29.

Kim, B., Tse, K.T., Yoshida, A., Tamura, Y., Chen, Z., Van Phuc, P., Park, H.S., 2019. Statistical analysis of wind-induced pressure fields and PIV measurements on two buildings. J. Wind Eng. Ind. Aerod. 188, 161–174.

Kim, B., Yuvaraj, N., Tse, K.T., Lee, D.E., Hu, G., 2021. Pressure pattern recognition in buildings using an unsupervised machine-learning algorithm. J. Wind Eng. Ind. Aerod. 214, 104629.

Kim, W., Tamura, Y., Yoshida, A., 2011. Interference effects on local peak pressures between two buildings. J. Wind Eng. Ind. Aerod. 99 (5), 584–600.

King, M.F., Gough, H.L., Halios, C., Barlow, J.F., Robertson, A., Hoxey, R., Noakes, C.J., 2017. Investigating the influence of neighbouring structures on natural ventilation potential of a full-scale cubical building using time-dependent CFD. J. Wind Eng. Ind. Aerod. 169, 265–279.

Kwatra, N., Godbole, P.N., Krishna, P., 2002. Application of artificial neural network for determination of wind induced pressures on gable roof. Wind Struct. 5 (1), 1–14.

Lamberti, G., Gorlé, C., 2021. A multi-fidelity machine learning framework to predict wind loads on buildings. J. Wind Eng. Ind. Aerod. 214, 104647.

Li, D., Liu, B., Cheng, Y., 2020. Wind Pressure Coefficients Zoning Method Based on an Unsupervised Learning Algorithm. Mathematical Problems in Engineering, 2020.

Li, S., Snaiki, R., Wu, T., 2021. A knowledge enhanced deep reinforcement learning based shape optimizer for aerodynamic mitigation of wind sensitive structures. Comput. Aided Civ. Infrastruct. Eng. 36 (6), 733–746.

Lin, P., Ding, F., Hu, G., Li, C., Xiao, Y., Tse, K.T., Kwok, K.C.S., Kareem, A., 2022. Machine learning-enabled estimation of crosswind load effect on tall buildings. J. Wind Eng. Ind. Aerod. 220 https://doi.org/10.1016/j.jweia.2021.104860. In press.

Lin, P., Hu, G., Li, C., Li, L., Xiao, Y., Tse, K.T., Kwok, K.C.S., 2021a. Machine learning-based prediction of crosswind vibrations of rectangular cylinders. J. Wind Eng. Ind. Aerod. 211, 104549.

Lin, P., Ding, F., Hu, G., Li, C., Xiao, Y., Tse, K.T., Kwok, K.C.S., Kareem, A., 2021b. Machine learning-based prediction of crosswind vibrations of rectangular cylinders. J. Wind Eng. Ind. Aerod. 211, 104549.

Lipton, Z.C., 2018. The Mythos of Model Interpretability: in machine learning, the concept of interpretability is both important and slippery. Queue 16 (3), 31–57.

Lundberg, S.M., Lee, S.I., 2017, December. A unified approach to interpreting model predictions. In: Proceedings of the 31st International Conference on Neural Information Processing Systems, pp. 4768–4777.

Lundberg, S.M., Erion, G.G., Lee, S.I., 2018. Consistent Individualized Feature Attribution for Tree Ensembles arXiv preprint arXiv:1802.03888.

Luštrek, M., Gams, M., Martinčić-Ipšić, S., 2016. What makes classification trees comprehensible? Expert Syst. Appl. 62, 333–346.

Miller, T., 2019. Explanation in artificial intelligence: insights from the social sciences. Artif. Intell. 267, 1–38.

Meddage, D.P.P., Lewangamage, C.S., Weerasuriya, A.U., 2022. On the deviation of mean pressure coefficients in wind loading standards for a low-rise, gable-roofed building with boundary walls. Structures 36, 50–64. https://doi.org/10.1016/j.istruc.2021.12.003.

Meddage, D.P.P., Ekanayake, I.U., Weerasuriya, A.U., Lewangamage, C.S., 2021. Tree-based Regression Models for Predicting External Wind Pressure of a Building with an Unconventional Configuration. 2021 Moratuwa Engineering Research Conference (MERCon), 257–262. https://doi.org/10.1109/MERCon52712.2021.9525734.

Mohandes, M.A., Rehman, S., 2018. Wind speed extrapolation using machine learning methods and LiDAR measurements. IEEE Access 6, 77634–77642.

Molnar, C., 2020. Interpretable Machine Learning. Lulu. com.

Moonen, P., Allegrini, J., 2015. Employing statistical model emulation as a surrogate for CFD. Environ. Model. Software 72, 77–91.

Pan, W., Liu, S., Li, S., Cheng, X., Zhang, H., Long, Z., et al., 2019. A model for calculating single-sided natural ventilation rate in an urban residential apartment. Build. Environ. 147, 372–381.

Park, H., Park, D.Y., 2021. Comparative analysis on predictability of natural ventilation rate based on machine learning algorithms. Build. Environ. 195, 107744.

Priyanka, Kumar, D., 2020. Decision tree classifier: a detailed survey. Int. J. Inf. Decis. Sci. 12 (3), 246–269.

Quinlan, J.R., 1986. Induction of decision trees. Mach. Learn. 1 (1), 81–106.

Ribeiro, M.T., Singh, S., Guestrin, C., 2016, August. Why should i trust you?" Explaining the predictions of any classifier. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 1135–1144.

Ribeiro, M.T., Singh, S., Guestrin, C., 2018, April. Anchors: high-precision model-agnostic explanations. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 32. No. 1.

Roscher, R., Bohn, B., Duarte, M.F., Garcke, J., 2020. Explainable machine learning for scientific insights and discoveries. IEEE Access 8, 42200–42216.

Sarkar, P.P., Zhao, Z., Mehta, K.C., 1997. Flow visualization and measurement on the roof of the Texas Tech building. J. Wind Eng. Ind. Aerod. 69, 597–606.

Singh, S., Gupta, P., 2014. Comparative study ID3, cart and C4. 5 decision tree algorithm: a survey. International Journal of Advanced Inform. Sci.Technol. (IJAIST) 27 (27), 97–103.

Stathopoulos, T., 1979. Turbulent Wind Action on Low Rise Buildings. Ph.D.). University of Western Ontario, London, Ontario, Canada.

Stathopoulos, T., 1984. Wind loads on low-rise buildings: a review of the state of the art. Eng. Struct. 6 (2), 119–135.

Stathopoulos, T., Wu, H., 1994. Knowledge-based wind loading for envelope design: beyond building codes. J. Wind Eng. Ind. Aerod. 53 (1–2), 177–188.

Surry, D., 1991. Pressure measurements on the Texas Tech building: wind tunnel measurements and comparisons with full scale. J. Wind Eng. Ind. Aerod. 38 (2–3), 235–247.

Tamura, Y., Suganuma, S., Kikuchi, H., Hibi, K., 1999. Proper orthogonal decomposition of random wind pressure field. J. Fluid Struct. 13 (7–8), 1069–1095.

Tanaka, H., Tamura, Y., Ohtake, K., Nakai, M., Kim, Y.C., 2012. Experimental investigation of aerodynamic forces and wind pressures acting on tall buildings with various unconventional configurations. J. Wind Eng. Ind. Aerod. 107, 179–191.

Thordal, M.S., Bennetsen, J.C., Koss, H.H.H., 2019. Review for practical application of CFD for the determination of wind load on high-rise buildings. J. Wind Eng. Ind. Aerod. 186, 155–168.

Ti, Z., Deng, X.W., Yang, H., 2020. Wake modeling of wind turbines using machine learning. Appl. Energy 257, 114025.

Van Moeseke, G., Gratia, E., Reiter, S., De Herde, A., 2005. Wind pressure distribution influence on natural ventilation for different incidences and environment densities. Energy Build. 37 (8), 878–889.

Vassallo, D., Krishnamurthy, R., Fernando, H.J., 2021. Utilizing physics-based input features within a machine learning model to predict wind speed forecasting error. Wind Energy Sci. 6 (1), 295–309.

Veiber, L., Allix, K., Arslan, Y., Bissyandé, T.F., Klein, J., 2020. Challenges towards production-ready explainable machine learning. In: 2020 {USENIX} Conference on Operational Machine Learning. OpML, 20).

Wang, J., Van Phuc, P., Yang, Q., Tamura, Y., 2020. LES study of wind pressure and flow characteristics of flat-roof-mounted solar arrays. J. Wind Eng. Ind. Aerod. 198, 104096.

Weerasuriya, A.U., Zhang, X., Lu, B., Tse, K.T., Liu, C.H., 2020. Optimizing lift-up design to maximize pedestrian wind and thermal comfort in 'hot-calm'and 'cold-windy'climates. Sustain. Cities Soc. 58, 102146.

Weerasuriya, A.U., Zhang, X., Lu, B., Tse, K.T., Liu, C.H., 2021. A Gaussian Process-Based emulator for modeling pedestrian-level wind field. Build. Environ. 188, 107500.

Weld, D.S., Bansal, G., 2019. The challenge of crafting intelligible intelligence. Commun. ACM 62 (6), 70–79.

Wu, F., Sarkar, P.P., Mehta, K.C., Zhao, Z., 2001. Influence of incident wind turbulence on pressure fluctuations near flat-roof corners. J. Wind Eng. Ind. Aerod. 89 (5), 403–420.

Xie, Z.N., Gu, M., 2007. Simplified formulas for evaluation of wind-induced interference effects among three tall buildings. J. Wind Eng. Ind. Aerod. 95 (1), 31–52.

Xing, F., Mohotti, D., Chauhan, K., 2018. Study on localised wind pressure development in gable roof buildings having different roof pitches with experiments, RANS and LES simulation models. Build. Environ. 143, 240–257.

Xu, F., Uszkoreit, H., Du, Y., Fan, W., Zhao, D., Zhu, J., 2019. October). Explainable AI: a brief survey on history, research areas, approaches and challenges. In: CCF International Conference on Natural Language Processing and Chinese Computing. Springer, Cham, pp. 563–574.

Yang, H.F., Chen, Y.P.P., 2019. Representation learning with extreme learning machines and empirical mode decomposition for wind speed forecasting methods. Artif. Intell. 277, 103176.

Zhou, Z.H., 2012. Ensemble Methods: Foundations and Algorithms. Chapman and Hall/CRC.

Zimmerman, R.K., Balasubramani, G.K., Nowalk, M.P., Eng, H., Urbanski, L., Jackson, M. L., et al., 2016. Classification and Regression Tree (CART) analysis to predict influenza in primary care patients. BMC Infect. Dis. 16 (1), 1–11.

Zu, G.B., Lam, K.M., 2018. Across-wind excitation mechanism for interference of twin tall buildings in staggered arrangement. J. Wind Eng. Ind. Aerod. 177, 167–185.