Automation in Construction 119 (2020) 103350

Contents lists available at ScienceDirect

Automation in Construction

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

Augmenting building performance predictions during design using generative adversarial networks and immersive virtual environments

Chanachok Chokwitthaya[a,*], Yimin Zhu[a], Supratik Mukhopadhyay[b], Edward Collier[b]

[a] Department of Construction Management, Louisiana State University, Baton Rouge 70803, USA
[b] Department of Electrical Engineering and Computer Science, Louisiana State University, Baton Rouge 70803, USA

ARTICLE INFO

Keywords:
Building design
Building performance model
Contextual factor
Human-building interaction
Immersive virtual environment
Artificial neural network
Generative adversarial network

ABSTRACT

Existing building performance models (*existing BPMs*) often lack the capability for addressing human-building interactions in future buildings or buildings under design because they are mainly derived using data in existing buildings. The limitation may contribute to discrepancies between simulated and actual building performance. In a previous study, the authors discussed a framework using an artificial neural network (ANN)-based greedy algorithm which combines context-aware design-specific data obtained from immersive virtual environments (IVEs) with an *existing BPM* to enhance the simulations of human-building interactions in new designs. Although the framework has revealed the potential to improve simulations, it cannot determine the appropriate combination between context-aware design-specific data and the *existing BPM*.

In this paper, the authors present a new computational framework (the GAN-based framework) to determine an appropriate combination based on a given *performance target* to achieve. Generative adversarial networks (GANs) are used to combine data of an *existing BPM* and context-aware design-specific data using a *performance target* as a guide to produce an *augmented BPM*. The effectiveness and the reliability of the GAN-based framework were validated using an IVE of a single occupancy office. Thirty people participated in an experiment on the simulation of artificial lighting switch uses using the IVE. Their light switch uses data under different work area illuminance were collected and analyzed. The building performance models (BPMs) proposed by Hunt and Da Silva were selected as the *existing BPM* and the *performance target* respectively. The data of each participant was used to generate an *augmented BPM* using the GAN-based framework and an *updated BPM* using the previous framework (i.e., ANN-based greedy algorithm framework). The thirty pairs of the *augmented* and *updated BPMs* were compared. Specifically, the errors measured between the *updated BPMs* and the *performance target* (E_1) and the errors measured between the *augmented BPMs* and the *performance target* (E_2) were analyzed using *t-tests* ($\alpha = 0.05$). In 22 out of 30 cases, the performance of the *augmented BPMs* was significantly better than the *updated* BPMs, and in four cases, the performance of the two was similar. Only in four other cases, the performance of the *updated BPMs* was better. The results confirmed the efficacy of the framework. However, future research is needed to study the *performance target* and uncertainties associated with IVE experiments to better understand and control the reliability of the framework.

1. Introduction

The design stage of a building project is a critical step to make decisions and establish directions for engineering building components, affecting the characteristics, functions, and performance of a building. To optimally translate design goals and objectives into the performance of a building, designers and engineers usually apply building performance models (BPMs) during the design stage such as simulations of building energy consumptions and human-building interactions to understand, investigate, and predict building performance, as well as support decision-making. Nevertheless, the application of BPMs cannot eliminate the significant performance discrepancies between the simulated and the actual performances that have been widely reported [1–3]. For example, studies have reported as much as 150% of differences between predicted and the actual performance of a building [4].

Many factors influence the simulations of building performance, especially human-building interactions such as occupant responses to building contexts and occupant habitual behaviors [5]. Human-building interactions are highly context-depended and sensitive to several contexts [6,7] in which contexts are described by situational

* Corresponding author.
E-mail addresses: cchokw1@lsu.edu (C. Chokwitthaya), yiminzhu@lsu.edu (Y. Zhu), supratik@csc.lsu.edu (S. Mukhopadhyay), ecoll28@lsu.edu (E. Collier).

https://doi.org/10.1016/j.autcon.2020.103350
Received 19 December 2019; Received in revised form 3 June 2020; Accepted 11 July 2020

factors that are not directly included in a model or simulation [8]. These situational factors are often assumed to remain constant across different applications of the model or simulation. For instance, "context" may be physical or natural factors (e.g., building characteristics, building surrounding and environmental factors, and climate conditions), and socio-technical factors (e.g., participant's cultural background, racial/ethnicity, and tasks to perform), which may not be included as variables in a BPM. However, such factors can have an impact on analysis using the BPM during the design of a specific space, whose situational factors may be different from what the BPM has assumed and cannot be treated as constant across different applications. In such cases, these situational factors in relation to any BPM need to be identified, analyzed, and integrated in building performance analyses. Often, BPMs are developed using data obtained from existing buildings, where the contexts of which differs from the contexts of a building under design. Applying such BPMs to understand, investigate, and predict human-building interactions in a building under design may contribute to the discrepancy between predicted and actual performance. Therefore, being able to address human-building interactions responding to specific contexts in new designs (e.g., the context embodied) can potentially enhance the accuracy of BPMs leading to reductions of the discrepancy between predicted and actual performance of a building.

Immersive virtual environments (IVEs) have demonstrated their potential in simulations and data collections in many disciplinary areas, especially engineering fields such as emergency evacuations [9,10], building designs [11], and human-building interactions [12–14]. IVEs provide several advantages over other data collection methods such as sensing, field studies, and surveys. For instance, IVEs can replicate certain context for buildings under design, especially when the contexts cannot be possibly, cost-effectively, or safely replicated in reality. Additionally, IVEs allow users to fully handle experimental conditions, and customize experimental models as desired. Human-building interactions in buildings under design may not be directly observed and analyzed. As a result, the application of IVEs can be an alternative for generating and examining the context-aware design-specific data of a new design. Following Sowa's definition of context [8], "context-aware" refers to the capability of a method, simulation, or model to address the impact of identified contextual factors in analysis. Therefore, by using the method, simulation or model, users are able to consider human-building interactions responding to contexts of a specific design. For example, in the application discussed in the paper, the context-aware design specific data of the proposed computation framework included contextual factors such as types of office task and locations of light switch.

To improve the accuracy of existing building performance models (*existing BPMs*), the authors have offered a framework for customizing *existing BPMs* to address contextual factors of a building under design. The framework using an artificial neural network (ANN)-based greedy algorithm has been developed to combine an *existing BPM* with context-aware design-specific data obtained from IVE experiments [15]. The framework has shown the potential to enhance the prediction accuracy of an *existing BPM*. However, its major limitation is that it lacks the capability to determine the appropriate combination of an *existing BPM* and context-aware design-specific data in a principled way rather than through trial and error, that can entail excess resource and time consumption. Hence, the principal goal of this study is to improve the capability of the framework to be able to determine the appropriate combination without trial and error. The new computational framework applies generative adversarial networks (GANs) to combine an *existing BPM* with context-aware design-specific data obtained from IVE experiments, and uses a *performance target* as a guide during computation to determine the appropriate mix without trial and error. The GAN-based framework produces an augmented building performance model (an *augmented BPM*) representing the appropriate combination that satisfies the *performance target*.

In the following, the authors first discuss comparison of the GAN-based framework and the ANN-based greedy algorithm framework, and then provide the research objective followed by an expression of the GAN-based framework and the explanation of applying the framework on a single-occupancy office to validate the framework. The design and administration of the IVE experiment are explained in detail. Finally, results, discussions, and limitations of the study, as well as conclusions and directions of future work are provided.

2. Comparison of the GAN-based framework and the ANN-based greedy algorithm framework

This section discusses major differences and relationships between the GAN-based framework and the ANN-based greedy algorithm framework. In parametric approaches (e.g., Gaussian mixture model), mixture models mix datasets derived from assumed probability distribution functions such as normal, binomial, and exponential [16]. Often, datasets do not fully comply with assigned distributions leading to the generation of inaccurate mixture models. Consequently, the ANN-based greedy algorithm framework was proposed in the previous work of the authors [15]. The framework was non-parametric so that users did not have to assume distributions for the underlying datasets to be mixed. The framework augmented an *existing BPM* by combining its dataset with context-aware design-specific data obtained from IVE experiments. However, this framework had a major limitation. It only allowed users to apply a linear combination using an assigned mixture ratio (number between 0 and 1) to mix data from the two datasets. Thus, if the probability distributions corresponding to the two datasets were f_1 and f_2 respectively, the greedy algorithm would produce a mixture distribution, $(1-\alpha)f_1 + \alpha f_2$, where $0 \leq \alpha \leq 1$. This was called a linear mixture using α as the mixture ratio. The limitation obstructs the framework to create a mixture distribution that was close to a *performance target* by nonlinearly mixing the probability distributions of the two datasets. In addition, the mixture ratio was not directly related to any *performance target*. Even in the case of linear mixtures, the framework did not provide any algorithm to determine the appropriate mixture ratio that generates the "best mix". Due to this limitation, an *augmented BPM* could only be constructed through trial and error. Users of the framework had to manually define the mixture ratios to combine datasets. They needed to perform several trials to obtain an appropriate combination (for deriving an *augmented BPM*) and, sometimes, appropriate combinations might be infeasible to obtain.

To overcome the disadvantages of the ANN-based greedy algorithm framework, the GAN-based framework is proposed and compared with the ANN-based greedy algorithm framework. The GAN-based framework uses a generative adversarial network (GAN) [17] to combine an *existing BPM* and context-aware design-specific data. Like the ANN-based greedy algorithm, the GAN allows a nonparametric approach to generate mixture models, where users do not have to assume any distribution (e.g., normal) for the underlying datasets getting mixed or for the mixture. In contrast with the ANN-based greedy algorithm, the GAN-based framework allows automatic determination of the appropriate mixture guided by a building performance target. This enables avoiding the trial and error techniques required in the ANN-based greedy algorithm framework and allows users to obtain an appropriate combination of datasets in one shot.

Filtering based approaches such as Kalman filter [18], requires manual determination of the filter type (e.g., linear, extended, and unscented) that will result in an appropriate mixture. In contrast, the GAN-based framework allows automatic determination of the appropriate mixture guided by a building performance target.

3. Research objective

The aim of this study is to create a new GAN-based framework that enables users to better perform building performance simulations

Table 1
The definition of the errors to prove the hypothesis.

Error	Measurement	
E_1	| predicted outcomes of an *updated BPM* – a *performance target* |	(1)
E_2	| predicted outcomes of an *augmented BPM* – a *performance target* |	(2)

during design. To achieve the goal, the objective of this study is twofold: 1) to investigate efficacy of the GAN-based framework in enhancing the prediction accuracy of BPMs, and 2) to examine the reliability of the GAN-based framework using experiments.

To determine the reliability of the framework, the authors conducted experiments on thirty college students to acquire data and statistically tested thirty comparisons between an *augmented BPM* and an *updated BPM*. The GAN-based framework generates an *augmented BPM* and the ANN-based greedy algorithm framework generates an *updated BPM* [15]. The comparison is based on the hypothesis that an *augmented BPM* is more accurate than an *updated BPM*. The absolute error measured discrepancies between predicted outcomes of an *updated BPM* and a *performance target* (E_1) and the absolute error measured discrepancies between predicted outcomes of an *augmented BPM* and a *performance target* (E_2) are shown in Table 1 and calculated using Eqs. (1) and (2) respectively. They are used to develop the hypothesis.

To test the performance of an *augmented BPM*, we hypothesize:

H_0: mean of E_1 - mean of E_2 = 0

H_1: the null hypothesis is not true.

A *t-test* ($\alpha = 0.05$) is applied to determine whether the performance of an *augmented BPM* is significantly different from that of an *updated BPM*.

4. Overview of the computational framework

The five main components of the GAN-based framework (Fig. 1) are: 1) an existing building performance model (an *existing BPM*), 2) context-aware design-specific data acquired from an IVE experiment, 3) a *performance target*, 4) a computation using generative adversarial network (GAN), and 5) an *augmented BPM*.

In general, "building performance models (BPMs)" is used to describe models of building performance at different building scales. BPMs may include performance models at a small scale such as specific building systems (e.g., lighting, blind, and window usages) to a large scale such as whole buildings (e.g., whole building energy consumptions). For example, at the building system level, Tahmasebi and Mahdavi [19] proposed a BPM for predicting window operations; and Keller et al. [20] developed a BPM to estimate performance of building systems (e.g., gas, electricity, and water). At the whole building level, Cho et al. [21] developed a BPM for estimating the whole building energy performance. Indeed, the framework is parametric, i.e., it takes three generic inputs, an *existing BPM*, context-specific data from IVE, and a *performance target*. These inputs are not related to a particular type of performance simulation. Therefore, the framework can be applied to different types of BPM at different scales and it is not dependent on the nature of BPMs. In the following, details of the framework are discussed.

4.1. Existing building performance model

Existing building performance models (*existing BPMs*) describe historical events and observations, which may not fully consider important contextual factors corresponding to a new building. Thus, contextual factors influencing human-building interactions in a new building are ignored in the *existing BPM* [15]. In addition, the *existing BPM* for predicting human-building interactions may be in different forms such as statistical models or synthetic datasets (generated by the models). When synthetic samples of an *existing BPM* are required, the GAN-based framework offers an approach to produce samples using a statistical approach (e.g., Monte Carlo simulation). The dataset associated with the samples is called an *existing BPM dataset* [15].

4.2. Context-aware design-specific data

Context-aware design-specific data describe key contextual

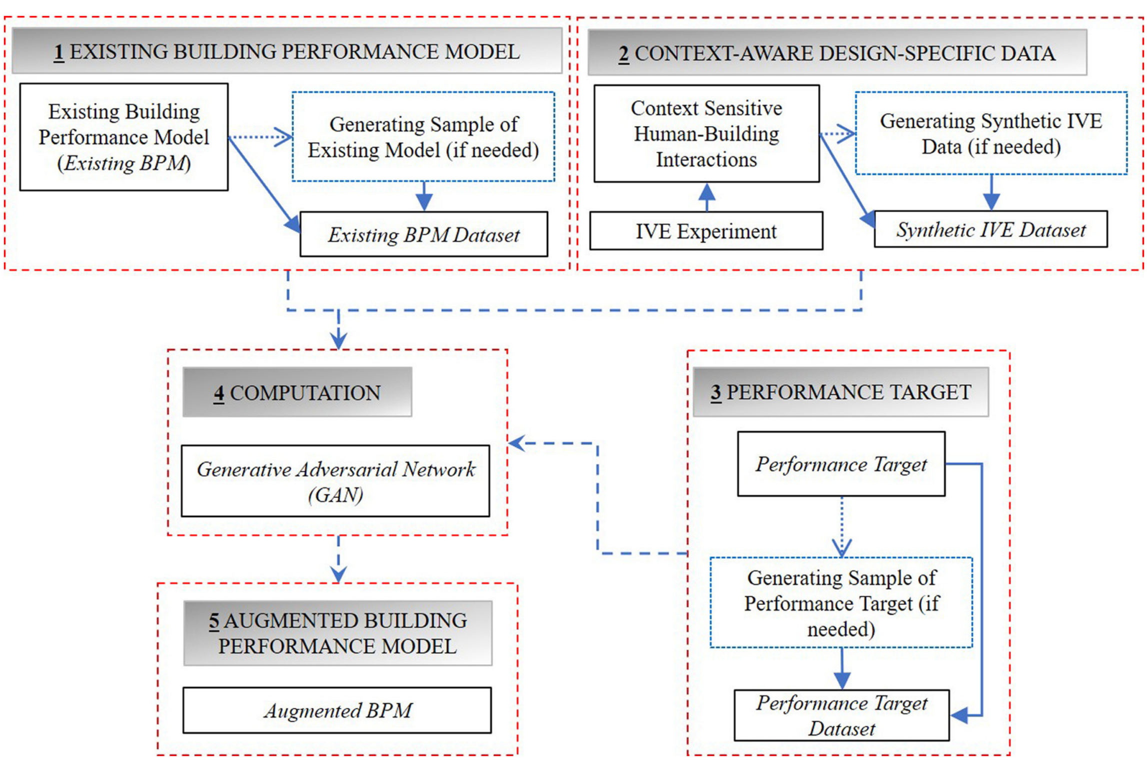


Fig. 1. The GAN-based framework.

conditions of a new design, where human-building interactions take place. For instance, the Hunt model [22] only models the relationship between the use of artificial lighting and the work area illuminance. On the other hand, the types of task (e.g., reading, meeting, and drafting) and the locations of light switch (e.g., a switch is by a door or on a desk) influence preferences of occupants interacting with the light switch. In this case, the types of task and the locations of light switch are contextual factors. When a design needs to explicitly consider such contextual factors to augment an *existing BPM*, immersive virtual environments (IVEs) can be used as tools to obtain context-aware design-specific data for a new design. Nevertheless, conducting IVE experiments sometimes is time consuming and each experiment session is often limited to 30–40 min. Therefore, IVE experiments usually result in small data samples with specific experiment conditions [23–25]. To overcome such a limitation, the GAN-based framework uses a data synthesis technique such as the Gaussian mixture model (GMM) to generate a large independent and identically distributed (IID) dataset, called a *synthetic IVE dataset* [15].

4.3. Performance target

During design, designers often consider and balance different criteria such as comfort, cost, energy efficiency and sustainability to satisfy design objectives [26]. Based on such objectives, various performance metrics do exist such as the energy intensity of a building. However, for the GAN-based framework to work, building performance metrics need to be converted into operational *performance targets* to support computations. This process is still an open question, which requires further research attention.

In this study, the authors assumed an operational *performance target*. The *performance target* may be created using empirical performance data of similar buildings and represented in the form of a statistical model or a set of data. On the other hand, since a *performance target* is used to evaluate with data generated by an *existing BPM*, components in both *performance target* and *existing BPM* need to be comparable.

4.4. Computation

4.4.1. Generative adversarial networks (GANs)

Since Goodfellow et al. [17] proposed the generative adversarial networks (GANs) in 2014, the method has been successfully applied in various domains, especially in deep learning-based studies [27–30], and image syntheses and analyses [31–33].

GANs have two parts: a generator, and a discriminator. The generator is an ANN that attempts to learn a probability distribution and tries to generate an output that follow a target distribution. The discriminator is an ANN discriminating the output of the generator and the target distribution. Conceptually, the generator and the discriminator play a two-player minimax game, where they undermine each other. The undermining continues until an equilibrium point is reached, where the generator and the discriminator do not change their performance regardless of what the opposition may do. Theoretically, in each epoch, the generator tries to produce the output that follows the target distribution. The discriminator observes the output of the generator as well as the target distribution. It tries to accurately discriminate the output of the generator whether they are from the target distribution. The feedback from the discriminator is used to train the generator through backpropagation. In every epoch, the generator keeps trying to produce an output that is follows the target distribution, while the weights of the discriminator are adjusted through backpropagation to accurately discriminate the generator outputs. The process continues until it reaches an equilibrium point, where the generator produces an output whose distribution is close to the target distribution and the discriminator accurately discriminate the generator's outputs and the target distribution [34].

In the GAN-based framework, the generator is trained using combinations of data associated with the *existing BPM* and context-aware design-specific data. The discriminator is trained using the generator's predictions and the *performance target dataset*. The generator has the responsibility to produce an output that is close to the *performance target dataset*. The discriminator has responsibilities to discriminate the output of the generator and the *performance target dataset*. A prediction of the generator that is closest to the *performance target* is considered as an *augmented BPM*.

In GANs, the performance of the discriminator relatively relies on the complexity and dimension of the input. If the input is complex and has only one dimension, the discriminator may be inferior to discriminate the generator's predictions and the *performance target dataset* [35]. To avoid such circumstances and enhance the performance of the framework, the concept of conditional GANs [36] is applied to condition on the generator as well as the discriminator, i.e., the inputs of the generator are fed into the input layer of the discriminator.

5. Application of the GAN-based framework

The application aims at understanding the efficacy and the reliability of the GAN-based framework by testing the hypothesis. The application used the lighting predictions in a single occupancy office as the studied case. An IVE configuration was created based on general recommendations of office designs [37], simulating situations related to variables of a BPM, and contextual factors (i.e., work area illuminance (lx), office tasks including reading, having a break, having a meeting, and drafting, and the location of a light switch such as by a door and on a desk). Thirty people participated in the IVE experiment. An *existing BPM* was the Hunt model for predicting lighting uses [22]. The probability of switching on, a Probit model, provided in Da Silva et al. [38] was used as a *performance target*. After computation, *augmented BPMs* were compared with *updated BPMs* obtained using the ANN-based greedy algorithm framework [15] to evaluate the efficacy and the reliability of the GAN-based framework.

5.1. Existing building performance model

The selections of the light uses prediction model proposed by Hunt [22] was relied on reasons including: 1) It has been cited and used as a baseline lighting BPM in lighting use studies [38–41]; 2) It has been applied in several building performance software packages (e.g., ESP-r, DAYSIM, and RADIANCE); and 3) It has only one independent variable, namely work area illuminance allowing the authors to study and demonstrate the inclusion of other factors as contextual factors. One of the successful applications of Hunt model is that it was used as the underlying theory to develop an algorithm (Lightswitch-2002) for simulating dynamic daylight used in DAYSIM and RADIANCE software. Currently, both software packages have been widely used in not only academic but industrial fields. The selection of the Hunt model was only for demonstrating and testing the framework. In fact, the GAN-based framework is generic, which can take any BPM as an *existing BPM*.

Hunt collected data associated with human-building interactions on light switches using the time-lapse photography method in six rooms (e.g., multi-person offices, school classrooms, and open-plan teaching area). The obtained data were fitted using a Probit model as Eq. (3) to predict the probability of switching on using work area illuminance (lx) as an independent variable. Fig. 2 shows the relationships of the probability of switching on and the work area illuminance of the Hunt model. The authors applied Monte Carlo technique to generate samples of the Hunt model as an *existing BPM dataset*. Monte Carlo simulation is a random process to mimic a behavior of real-life systems [42]. It has been widely accepted as a standard technique for generating independent and identically distributed (IID) samples for models in several works [43–45]. The proper number of samples was determined by using a learning curve technique [46,47]. In general, the learning curve

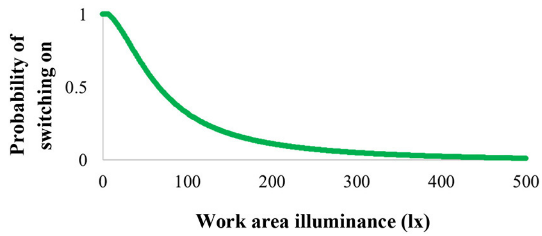


Fig. 2. The Hunt model [15].

technique investigates impacts of number of samples used to train the ANN on the accuracy of predictions of the trained ANN. The technique continuously increases the number of samples until additional samples do not significantly increase in the performance of predictions of the ANN (i.e., reached the knee point). The number of samples at the knee point is taken as the number of samples. The application excluded the analysis of the learning curve since it was demonstrated in Chokwitthaya et al. [47] and the number of samples in the *existing BPM dataset* was carried over from the work.

$$\text{Probability of switching on} = -0.0175 + \frac{1.0361}{(1 + e^{4.0835(\log_{10}(\text{work area illuminance}) - 1.8223)})} \quad (3)$$

5.2. Context-aware design-specific data

5.2.1. Immersive virtual environment (IVE) and experiment

5.2.1.1. The IVE model. A single office was modeled in virtual reality in accordance with the general recommendation of office designs [37]. The dimension of the office was 5.5 × 4.2 × 3.2 m with a net area of 22 square meters as shown in Fig. 3. AutoCAD software was used to create a 3D model of the office. Autodesk 3ds Max was used to assign and render materials of the model. It was also used to estimate the work area illuminance in the application, since Autodesk 3ds Max has shown potentials to simulate indoor illuminance for daylighting [48]. Unreal Engine 4 was applied to simulate the virtual 3D environment (Fig. 4), allowing participants to interact with building components such as the light switch.

5.2.1.2. The design of the IVE for experiment. The concept of spatial-temporal event-driven modeling (STED) [23] was partially applied to the design of the IVE experiment. The STED models critical events during a day in a chronological order representing longitudinal observations in reality to minimize the negative impact of IVE technologies on participants. It uses categorical values to model such critical events in IVEs. It involves four major components (i.e., states, contexts, events, and human-building interactions). In this application, the design of the IVE experiment ignored the chronological order, so that events were discretely modeled and do not influence next adjacent

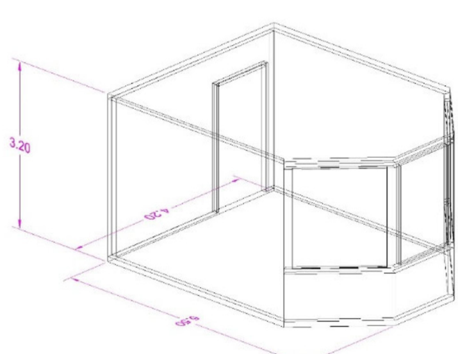


Fig. 3. The model of the office.

Fig. 4. The IVE configuration.

events, i.e., a finished state of an event was not transferred to be an initial state of a next adjacent event as described in the STED.

According to the STED, the four main components (i.e., states, contexts, events, and human-building interactions) were defined as follows for the IVE experiment:

- States were the on/off status of the light switch, which was initially set to off in all scenarios.
- Contexts were conditional factors describing the states. There could be many contextual factors. In this study, two commonly identified factors, office tasks and light switch locations, were selected based on previous literature discussing the impact of tasks [49–51] and in particular the location of light switches [12]. Corresponding to the Hunt model, the factor considered as the independent variable was work area illuminance. The level of work area illuminance was defined using the following recommendations: 1) The recommended lighting levels provided by the U.S. general services administration (GSA) [52], where 500 lx is the suggested light intensity at work areas for conducting office tasks, and; 2) Previous studies have shown a significantly low probability of switching on when the work area illuminance is higher than 200 lx [5,39,53]. Therefore, the IVE experiment was designed to use the 500 lx as the maximum work area illuminance. The work area illuminance between 200 and 500 lx was assigned with a 150-lx interval. The smaller interval of illuminance (i.e., 50 lx) was assigned between 0 and 200 lx to capture more possible fluctuations of human-building interactions on light switching in this range. Table 2 describes details of the office tasks, light switch locations and work area illuminance considered in the IVE experiment. However, it has to be noted that even if the levels of the work area illuminance may be possibly defined with smaller intervals to simulate continuous illuminance, it will costly increase the duration of an IVE experiment. The choice of the interval in this case was based on the consideration to finish an experiment for each participant within 60 min including a training session.
- Events were triggers that cause occupants to change or maintain the states. In this application, combinations of the contextual factors (4

Table 2
Variables and their values considered in the IVE experiment.

Contextual factor		Independent variable	Dependent variable
Office task	Light switch location	Work area illuminance (lx)	Likelihood of switching on
Intensive reading	By the door	50	Very unlikely
Having a break	On the desk	100	Not likely
Having a meeting		150	Neutral
Drafting		200	Likely
		350	Very likely
		500	
Total = 4	*Total = 2*	*Total = 6*	

office tasks and 2 light switch locations) and the dependent variable (6 work area illuminance) led to 4 × 2 × 6 = 48 events. Accordingly, each participant generated 48 data points, one data point for each event.
- Human-building interactions (dependent variables) were the likelihood of switching on or off (the dependent variable) described in Table 2.

5.2.1.3. IVE experiment and data collection. The experiment had two main sessions, namely the training and the experimental session. In the following, details of the IVE experiment are explained.

The training session was designed to get participants familiarized with the IVE experiment. This session allowed participants to explore themselves in the virtual environment along with practices of responding to questions in the experiment. The training session took around 10 min for each participant.

In the experimental session, participants initially sat on a chair, which was at the desk in the virtual environment, as if they were about to perform some tasks in the office. The participants were told that they were the sole occupant of the office, who could interact with a light switch freely. In each event, the participants were exposed with one of the work area illuminance levels presented in Table 2 and audio cues presented in Table 3 informing the participants about the current conditions of the office. After listening to the audio cues, the participants were asked to determine the likelihood of switching on under the condition of the office at that moment. There were five available choices of the likelihood constructed based on Likert scale [54] and the choices were mapped to the probability of switching on as shown in Table 4. The following example illustrates how the audio cues and questions were presented to the participants. If the task and the switch location were "*intensive reading*" and "*by the door*" in a particular event respectively under any work area illuminance, the audio cue was "*You are going to intensively read research papers for at least an hour. If the light switch is by the door and you have to walk to turn it on, please select your need of turning the light on under the provided situation.*" The 48 events mentioned previously were assigned to the participants randomly and each participant took around 40 min to finish the experiment.

A total of 30 students (18 males, 12 females) participated in the research study. Before conducting the experiment, the participants had to sign the consent form and completed the motion sickness screening questionnaire. The experiment received approval from the local institutional review board (IRB). The participants used a head-mounted display to conduct the experiment in IVE during the experiment. Fig. 5 shows a participant exploring the office and making a decision on the use of the lighting during the experiment.

5.2.2. Generating synthetic IVE data

Since the sample from the IVE experiment was relatively small, the authors augmented the dataset by creating a *synthetic IVE dataset*. The Gaussian mixture model (GMM) [55] has been proven to have higher performance in clustering data than many other data clustering methods such as k-means [56], k-nearest neighbor [57], and multivariate kernel density (MVKD) [58]. Moreover, it has been used to generate IID samples in research studies [59,60]. In this application, the GMM was used to augment the IVE dataset by increasing the number of IID samples based on the data obtained from the IVE experiment, i.e., "*synthetic IVE dataset*". The GMM learned the IVE experimental data by using mixtures of Gaussian, where the data were categorized into z groups based on the Gaussian (i.e., normal) distribution. Each group had its own mean (μ) and variance (σ^2). It was assumed that each data point (x) probabilistically belonged to z and the distribution for each group was separately inferred.

The GMM was constructed according to what was done by Chokwitthaya et al. [61]. The context-aware design-specific data obtained from the IVE experiments were sourced by each individual participant, resulting in thirty datasets. To train the GMM, the K-mean algorithm was used to initialize the GMM parameters [62]. The full covariance type was applied. The convergence criterion for training the GMM was 10^{-2}. The trained GMM was implemented to generate IID samples of the context-aware design-specific data, called the *synthetic IVE dataset*. Since thirty datasets were used to train the GMM individually, there was a total of thirty sets of the *synthetic IVE dataset*. Similar to the *existing BPM dataset*, the learning curve technique determined the number of samples in the *synthetic IVE datasets* [47].

5.3. Performance target

In this application, the Probit regression model of light switch uses proposed by Da Silva et al. [38] described in Eq. (4) was selected as a *performance target*. The model was selected for the following reasons: 1) Data used to construct the model were obtained from eight single-occupancy offices, which were similar to the design of the single occupancy office in this application; 2) Probit regression models were accepted and applied to represent data associated with human-building interactions in many research studies [41,63,64]; 3) The model used work area illuminance as an independent variable similar to the *existing BPM*; and 4) The large discrepancy between the Hunt and Da Silva models set a high target, which increased the challenge for the framework to generate an *augmented BPM* that could meet the performance target. The challenge might prevent an *augmented BPM* to meet the *performance target* in some cases, which opened the opportunity for the authors to explore and discuss such cases. Indeed, the framework can take any performance model as a *performance target*. A target is used to guide the combination of an *existing BPM* and context-aware design-specific data in the framework. Therefore, if a different *performance target* is used, the framework will generate a different *augmented BPM* that is correlated with the characteristics of the new *performance target*.

$$\text{Probability of switching on} = \frac{1}{(1 + e^{-(2.035+(-0.003)*\text{work area illuminance})})} \tag{4}$$

The authors applied Monte Carlo simulation to generate samples of the Da Silva model using Eq. (4) and the characteristic of the model is presented in Fig. 6. The obtained dataset of samples was called a *performance target dataset*.

5.4. Computation

5.4.1. Data preprocessing

The *existing BPM dataset*, the *synthetic IVE datasets*, and the

Table 3
Audio cues of office tasks and switch locations.

Office task		Switch location	
Task	Audio cue	Location	Audio cue
Intensive reading	You are going to intensively read research papers for at least an hour.	By the door	The light switch is by the door and you have walk to turn it on.
Having a break	You are going to have a break in your office for at least an hour.	On the desk	The light switch is on the desk which is now reachable.
Having a meeting	You are going to have a meeting in your office for at least an hour.		
Drafting	You are going to work on a drafting task for at least an hour.		

Table 4
Choices of the likelihood of switching on and their interpretation.

Likelihood of switching on	Very unlikely	Not likely	Neutral	Likely	Very likely
Probability of switching on (%)	1	25	50	75	99

Fig. 5. A participant exploring the virtual office and selecting a need of turning the light on.

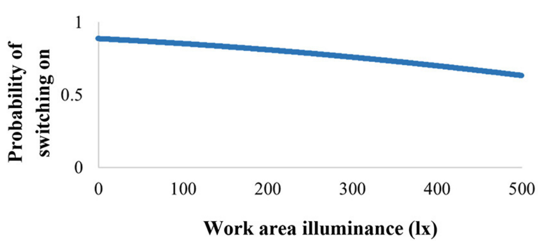


Fig. 6. The Da Silva model.

Table 5
Groups of the *sub-synthetic IVE dataset*.

Group	Combination of the contextual factors		
1	Intensive reading	+	Light switch by the door
2	Having a break	+	Light switch by the door
3	Having a meeting	+	Light switch by the door
4	Drafting	+	Light switch by the door
5	Intensive reading	+	Light switch on the desk
6	Having a break	+	Light switch on the desk
7	Having a meeting	+	Light switch on the desk
8	Drafting	+	Light switch on the desk

performance target dataset were normalized to maintain the compatibility and consistency of datasets during computation. The normalization was done by using Eq. (5) [65]. To reduce the complexity of the *synthetic IVE datasets* during computations, each of the *synthetic IVE dataset* was split into eight groups of *sub-synthetic IVE dataset* based on combinations of the contextual factors (Table 5). The computation used each of the *sub-synthetic IVE datasets* to augment the *existing BPM*.

$$\text{Normalized Data} = \frac{(\text{data–means of the synthetic IVE dataset})}{\text{standard deviation of the synthetic IVE dataset}} \qquad (5)$$

5.4.2. Generative adversarial networks (GANs)

A GAN was applied as the computational component for the framework. The GAN comprised of two major components, which were a generator and a discriminator (Fig. 7). Prior to train the GAN, the *existing BPM dataset* and the *sub-synthetic IVE datasets* were split into training datasets (i.e., the *existing BPM training dataset* and the *sub-synthetic IVE training datasets*) and testing datasets (i.e., the *existing BPM testing dataset* and the *sub-synthetic IVE testing datasets*) with a 70–30 split.

The generator was an ANN with a three-layer perceptron, namely an input, a hidden, and an output layer. In each training epoch, the *existing BPM training dataset* and the *sub-synthetic IVE training dataset* were used to train the generator. The generator took data of the work area illuminance as the input in the input layer. The hidden layer included 20 hidden neurons with the rectified linear unit activation function (ReLU). Elastic net regularization, a combination of L1 (Laplacian) and L2 (Gaussian) penalties, was applied to prevent overfitting [66]. The input in the output layer of the generator was the probability of switching on. The output layer applied the sigmoid activation function since the output was probability. The loss function, learning rate, and regularization were binary cross entropy (logistic regression), 10^{-4}, and 10^{-4} respectively. Before training the GAN, the generator was pre-trained using the *existing BPM training dataset* and the *sub-synthetic IVE training dataset* to initialize its weights and biases as well as increase the robustness of its learning. The generator learned the relationship of work area illuminance and probability of switching on associated with the *existing BPM training dataset* and the *sub-synthetic IVE training dataset*. In every training epoch, the generator predicted the probability of switching on based on the work area illuminance of the *existing BPM testing dataset* and the *sub-synthetic IVE testing dataset*. The probability of switching on predicted by the generator that was closest to that in the *performance target* was used to construct an *augmented BPM*.

The discriminator was an ANN discriminating the outputs from the generator and the *performance target dataset*. The discriminator comprised of a three-layered ANN with similar structure as the generator but different input datasets. Before training the discriminator, the work area illuminance as the input in the generator was paired with the probability of switching on predicted by the generator. The paired data were labeled as 0. The *performance target dataset* was labels as 1. The labels were assigned to distinguish the dataset from the generator and the *performance target dataset*. Indeed, labels could be any number, besides 0 and 1. Then, the paired data were concatenated with the *performance target dataset* (the purple box in Fig. 7) and used as the input in the input layer of the discriminator. This step generated conditioning situation on to the generator and the discriminator. The labels were taken as the input in the output layer.

The *synthetic IVE dataset* of each participant was applied to train the GAN resulting in the *augmented BPM* of a participant. Therefore, a total of 30 *augmented BPMs* are obtained.

5.5. Overview of the ANN-based greedy algorithm [15,47]

The ANN-based greedy algorithm framework [15] comprised of four major elements as shown in Fig. 8: 1) *an existing BPM*, 2) context-aware design-specific data obtained from an IVE experiment, 3) computation, and 4) an *updated BPM*.

The *existing BPM*, the context-aware design-specific data, the *synthetic IVE datasets* and the data preprocessing were identical to those in the application of the new framework.

The computation uses the ANN to combine the *existing BPM dataset* and the *synthetic IVE dataset* by using a mixture ratio (α) to guide the combination. The structure of the ANN is illustrated in Fig. 9. In this application, the ANN was a three-layered perceptron with an input, a hidden, and an output layer. The ANN took the tasks, light switch locations, and work area illuminance as inputs in the input layer. The input in the output layer was the probability of switching on. The hidden layer had 20 neurons with a rectified linear unit activation function (ReLU) and the elastic net regularization. The sigmoid activation function was applied to the output layer. The loss function, learning rate, and regularization were the binary cross entropy (logistic

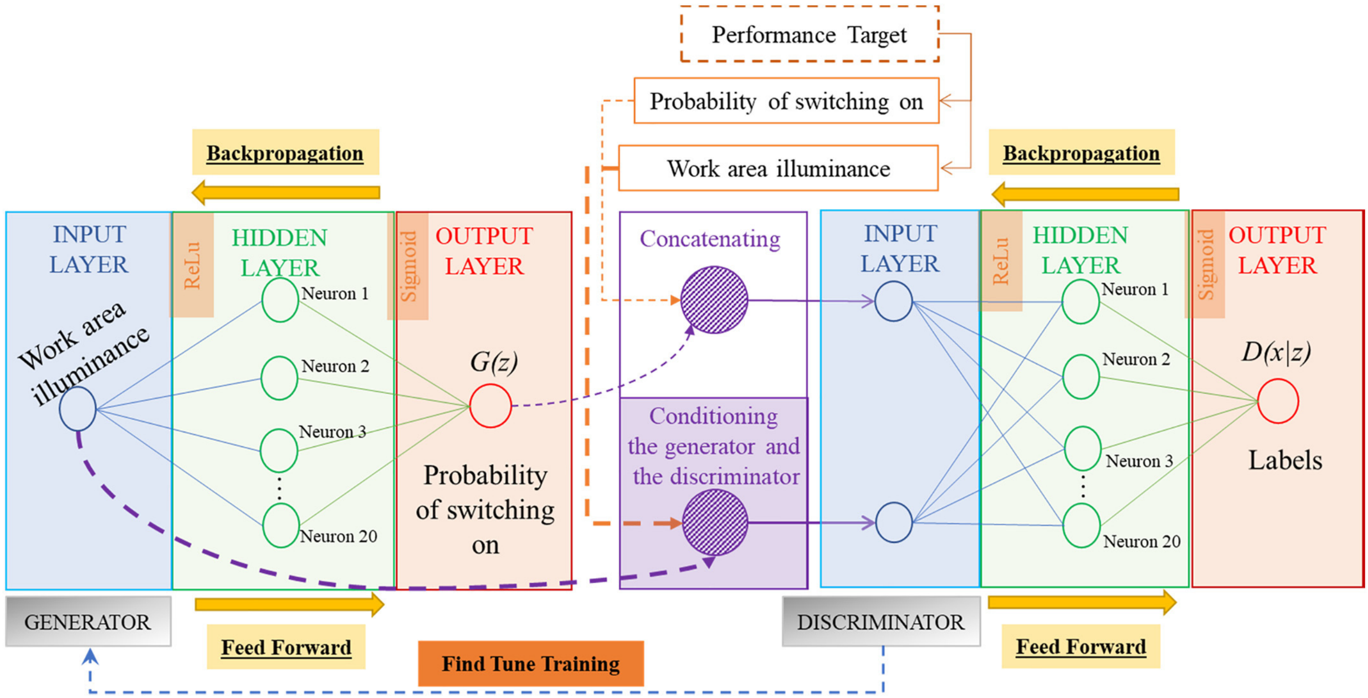


Fig. 7. The scheme of the generative adversarial network (GAN) in the application.

regression), 10^{-4}, and 10^{-4} respectively. The *existing BPM dataset* and the *synthetic IVE datasets* were divided into training datasets (i.e., the *existing BPM training dataset* and the *synthetic IVE training datasets*) and testing datasets (i.e., the *existing BPM testing dataset* and the *synthetic IVE testing datasets*) with a 70–30 split.

The training algorithm used an efficient greedy algorithm (Fig. 10) proposed in Chokwitthaya et al. [15] to determine which training dataset to be used to train the ANN (either the *existing BPM training dataset* or the *synthetic IVE training dataset*). In each training epoch, two values of the mean absolute error (MAE) were calculated: 1) the MAE measuring the difference between the predictions of the ANN and the *synthetic IVE testing dataset* (MAE^{SI}), and 2) the MAE measuring the difference between the predictions of the ANN and the *existing BPM testing dataset* (MAE^{EX}). During training the ANN, a mixture ratio (α) was used to keep the proportion of MAE^{SI} and MAE^{EX} using Eq. (7). If Eq. (7) was true, the ANN was trained on the *synthetic IVE training dataset* in the next epoch. Otherwise, the ANN was trained on the *existing BPM training dataset* in the next epoch.

$$\frac{MAE^{SI}}{MAE^{EX}} > \frac{1-\alpha}{\alpha} \tag{6}$$

The computation was performed by using one of the *synthetic IVE datasets* at a time resulting in thirty computational cases. A pseudorandom number, in the interval [0,1], was generated and used as the mixture ratio (α) for each case as shown in Table 6. Accordingly, a total of thirty *updated BPMs* were generated. The *updated BPMs* were further used to test the hypothesis.

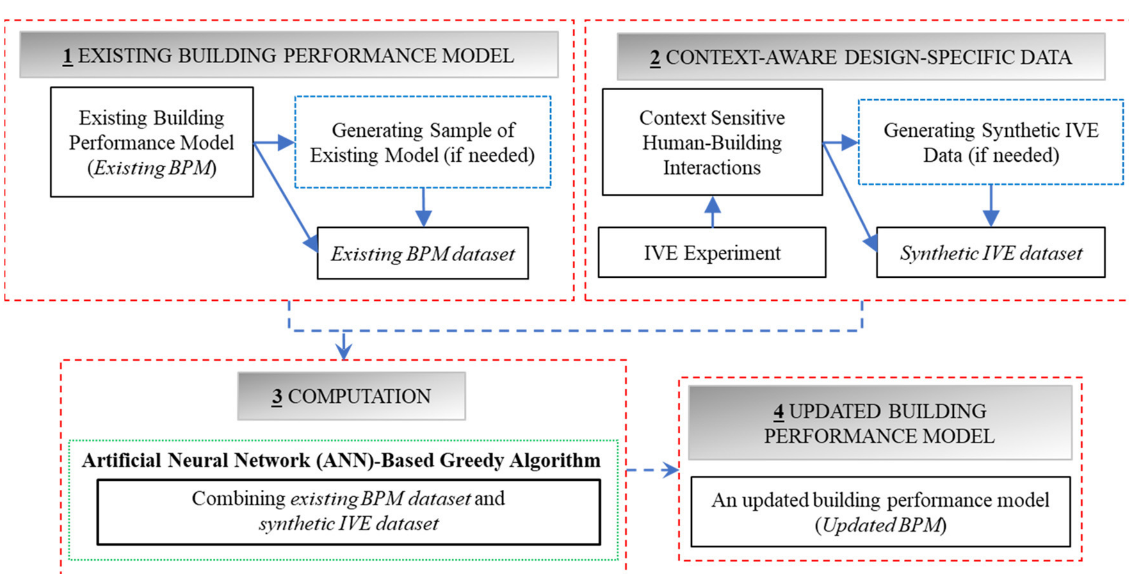


Fig. 8. Scheme of the ANN-based greedy algorithm [15].

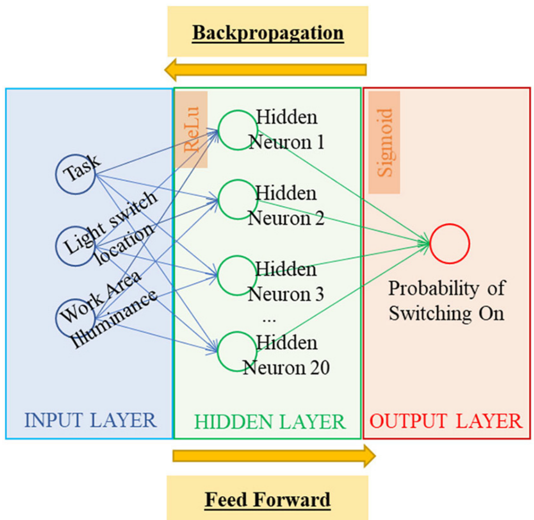


Fig. 9. The scheme of the ANN of this application [15].

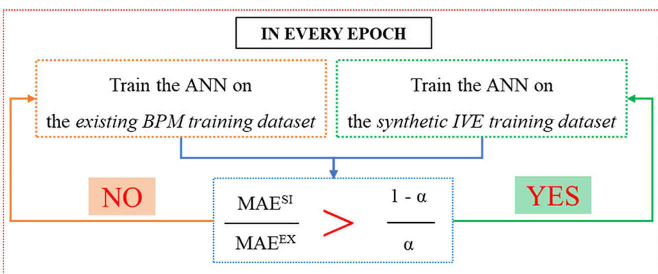


Fig. 10. Efficient greedy algorithm [15].

6. Results

6.1. The context-aware design-specific data acquired from the IVE experiment

Fig. 11 presents the means and the standard deviations of the context-aware design-specific data obtained from the IVE experiment of all participants classified by the office tasks and the light switch locations. Two observations show the qualitative effectiveness of the data:

1) The probability of switching on under low work area illuminance was higher than high work area illuminance regardless the assigned tasks and the light switch locations. This general pattern matched the previous studies [39,67,68].
2) The probability of switching on with respect to the assigned task of "*having a break*" was slightly lower than the probability of switching on with respect to the other assigned tasks. Previous studies showed that the lighting needs were different based on task types [50,51], thus the results showed a certain level of validity qualitatively.

On the other hand, the probability of switching on with respect to the location of light switches was visually similar under each plot in Fig. 11. This observation was different from what literature suggests in general [12], that the location of switches influenced participants' use of switches.

6.2. Descriptive comparisons between the Augmented BPMs and the Updated BPMs

Fig. 12 presents plots of the probability of switching on versus the work area illuminance using the mean of the *augmented BPMs*, and the mean of the *updated BPMs* classified by the office tasks and the switch locations. The mean was calculated based on the 30 *augmented* and the 30 *updated BPMs* respectively. The *performance target dataset* is also included in Fig. 12. Several observations can be drawn based on Fig. 12:

- The overall probability of switching on of the *augmented BPMs* can be visually noticed that they are greater than the overall probability of switching on of the *updated BPMs,* when the *augmented BPMs* and the *updated BPMs* are evaluated with the *performance target.*
- The probability of switching on of the *updated BPMs* between 0 and 50 lx is nearer to the *performance target* compared to the probability of switching on of the *augmented BPMs*. One possible cause may be that the IVE experiment did not include illuminance between 0 and 50 lx, which may impact the computation procedure.
- The *augmented BPMs* have a lower probability of switching on than the *performance target*. The circumstance may imply that the *augmented BPMs* do not meet the *performance target*. Some contributing factors are discussed in the Discussions section.

6.3. Hypothesis testing

To test the hypothesis, two absolute errors were calculated including: 1) the absolute errors measured between the probability of switching on associated with the *updated BPM* and that associated with the *performance target dataset*, i.e., E_1 calculated using Eq. (7) in Table 7, and 2) the absolute errors measured between the probability of switching on associated with the *augmented BPM* and that associated with the *performance target dataset*, i.e., E_2 calculated using Eq. (8) in Table 7.

6.3.1. Tests on the mean of E_1 and E_2 in all cases

A two tailed *t-test* ($\alpha = 0.05$) was first used to analyze whether E_1 and E_2 were significantly different. The hypothesis is described as follow:

H_0: mean of E_1 - mean of $E_2 = 0$

H_1: mean of E_1 - mean of $E_2 \neq 0$

Table 8 shows the statistical test of significant differences between E_1 and E_2 of the thirty individuals. In 26 of 30 cases, the differences between means of E_1 and E_2 were significant, in which the *p*-values were lower than 0.05.

6.3.2. Tests on the mean of E_1 and E_2 in cases 2, 10, 15 and 18

Specifically, further tests were conducted to determine if the *updated BPM* was statistically more accurate than the *augmented BPM* in the four cases (i.e., cases 2, 10, 15, and 18) as suggested by their means.

Table 6
The mixture ratio (α) of computational cases.

Case	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
α	0.73	0.75	0.18	0.45	0.64	0.59	0.13	0.85	0.34	0.85	0.36	0.84	0.96	0.33	0.61

Case	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
α	0.83	0.40	0.97	0.57	0.54	0.21	0.60	0.20	0.41	0.29	0.05	0.08	0.35	0.14	0.38

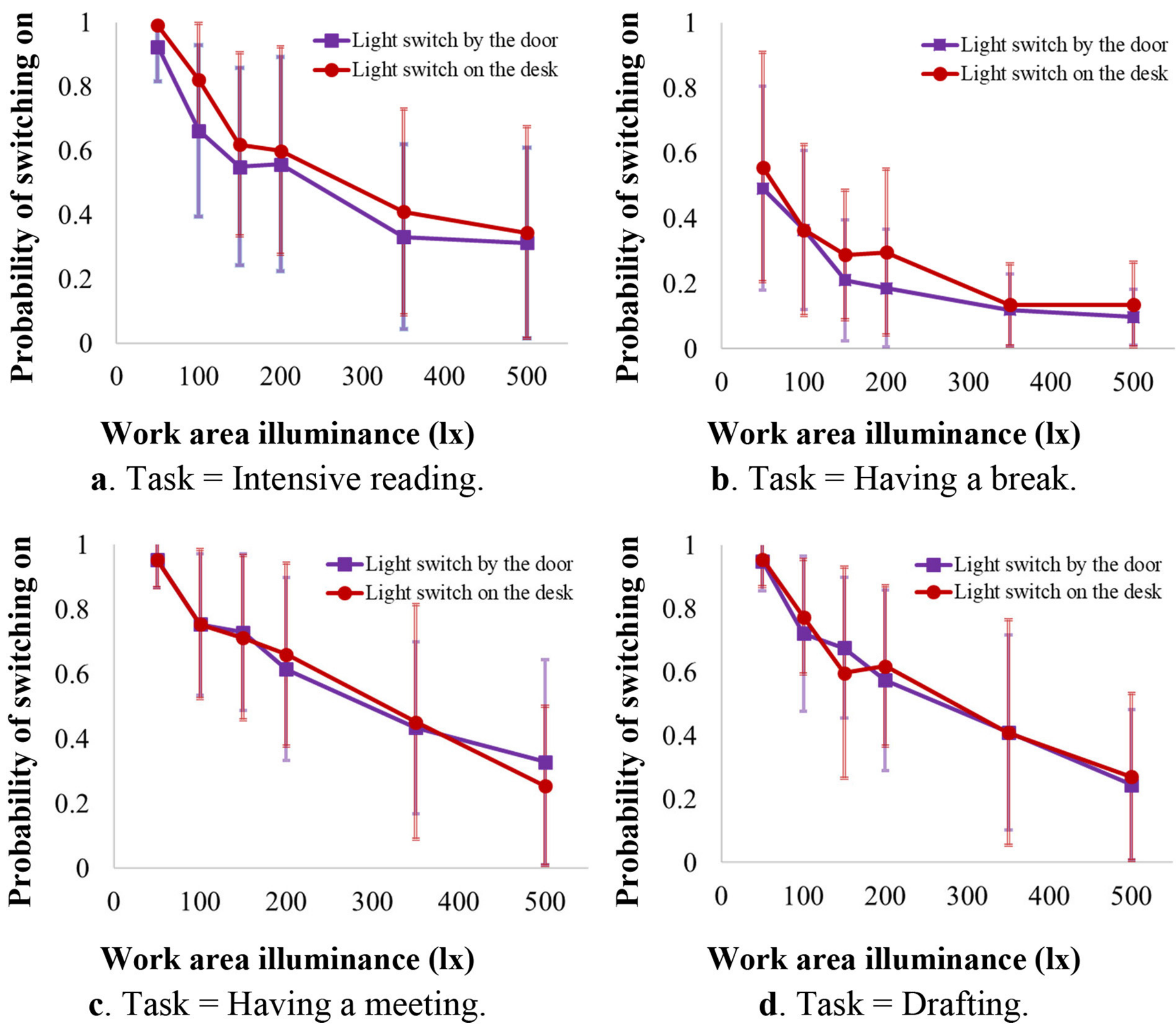


Fig. 11. The context-aware design-specific data.

Tests were performed by a one tailed *t-test* ($\alpha = 0.05$) with the following hypothesis:

H_0: mean of E_1 - mean of $E_2 = 0$

H_1: mean of E_1 - mean of $E_2 < 0$

From Table 9, the null hypotheses were rejected in all cases, which demonstrate that the *updated BPM* has significantly better accuracy compared to the *augmented BPM* in all four cases.

6.3.3. Tests on the rest of cases

For the rest 22 cases, their means showed that the *augmented BPMs* were more accurate than the *updated BPMs*. Since the *augmented BPMs* were expected to have significant smaller errors than the *updated BPMs*, the hypothesis was defined as follow:

H_0: mean of E_1 - mean of $E_2 = 0$

H_1: mean of E_1 - mean of $E_2 > 0$

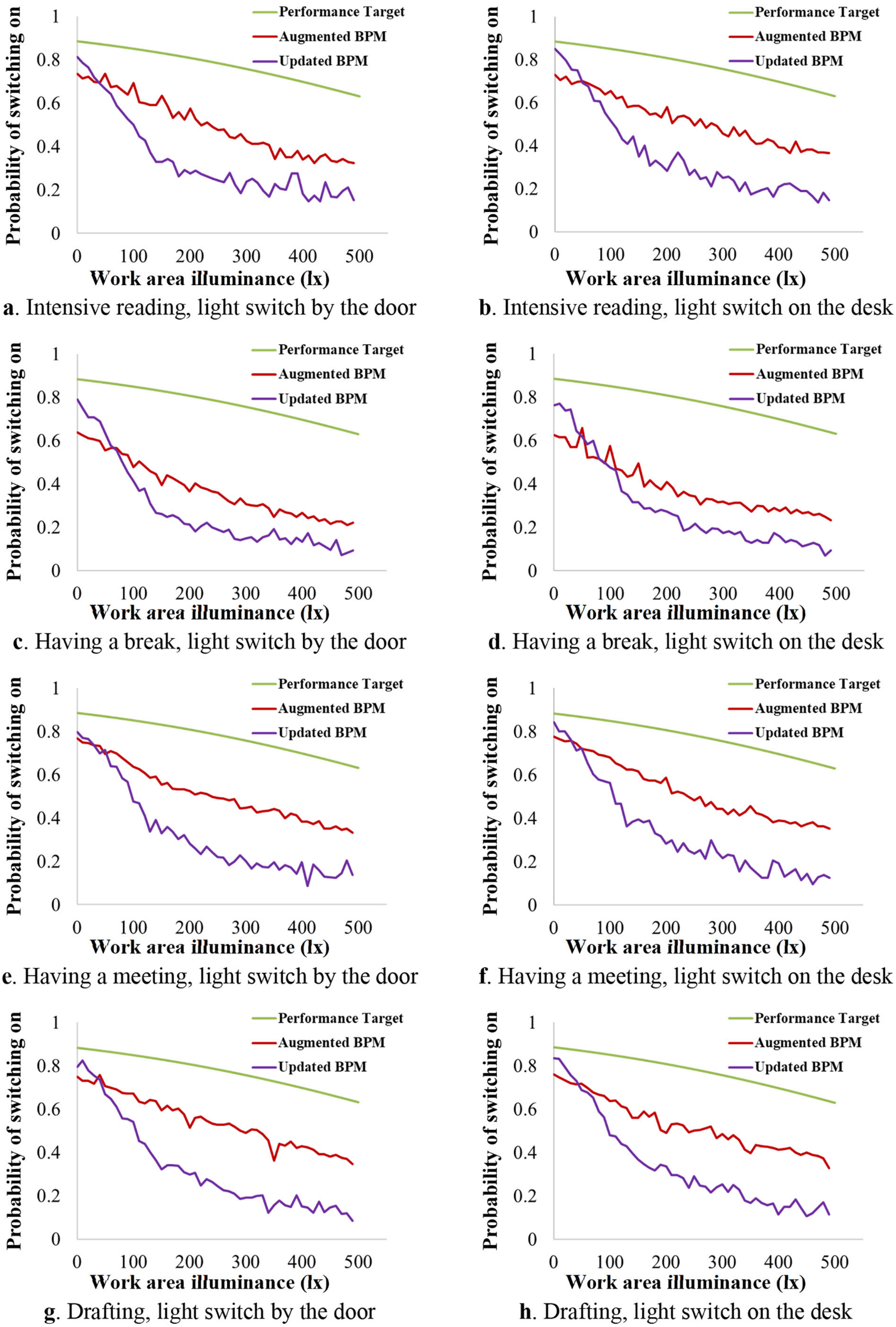
A one tailed *t-test* ($\alpha = 0.05$) was applied to test the hypothesis. The results of the hypothesis testing are shown in Table 10.

The results of the hypothesis testing in Table 10 show that the *p*-values were smaller than 0.05 for all 22 cases, in which the null hypothesis was rejected. The tests suggested that the probability of switching on predicted by the *augmented BPMs* was significantly better than that predicted by the *updated BPMs*.

7. Discussions

The hypothesis testing at individual level shows mixed results. In the following, discussions regarding the context-aware design-specific data, the *augmented BPMs*, and the results of hypothesis tests are presented.

- The context-aware design-specific data involved variances associated with the probability of switching on (Fig. 11). As mentioned in literature (e.g., [15,69]), participants are clearly a source of the variances since different people may respond to the IVE experiment differently. In addition, factors such as the uses of virtual cues and experimental settings (e.g., the display quality such as brightness and color, and participants' perception about the IVE) may contribute to the variances of the context-aware design-specific data. Such variances can affect the accuracy of the *augmented BPMs*.
- Although in most individual cases (i.e., 26 out of 30 cases), the *augmented BPMs* did not underperform compared to the *updated BPMs*, the means of the individual *augmented BPMs* were not close to the *performance target* (Fig. 12). This result does not necessarily discount the effectiveness of the *augmented BPMs*. The discussion of the issue may begin with the *existing BPM*. If the *existing BPM* significantly lacks the ability to address the characteristics of a design such as building configurations and occupant profiles, by using the *existing BPM*, the *augmented BPM* may be significantly biased by data



a. Intensive reading, light switch by the door

b. Intensive reading, light switch on the desk

c. Having a break, light switch by the door

d. Having a break, light switch on the desk

e. Having a meeting, light switch by the door

f. Having a meeting, light switch on the desk

g. Drafting, light switch by the door

h. Drafting, light switch on the desk

Fig. 12. Plots of the mean of the *augmented BPMs*, the mean of the *updated BPMs*, and the *performance target*.

Table 7
The absolute errors to prove the hypothesis.

Error	Measurement	
E_1	|The probability of switching on from the *updated BPM* - The probability of switching on from the *performance target dataset*|	(7)
E_2	|The probability of switching on from the *augmented BPM* - The probability of switching on from the *performance target dataset*|	(8)

Table 8
The summary of tests of significant difference between E_1 and E_2.

Case	Mean E_1	Std. E_1	Mean E_2	Std. E_2	*p*-Value	H_o
1	0.259	0.186	0.181	0.128	0.597	*Accept*
2	0.159	0.147	0.213	0.156	< 0.05	Reject
3	0.388	0.206	0.219	0.176	< 0.05	Reject
4	0.589	0.186	0.238	0.189	< 0.05	Reject
5	0.279	0.146	0.208	0.183	< 0.05	Reject
6	0.557	0.221	0.275	0.187	< 0.05	Reject
7	0.572	0.196	0.247	0.153	< 0.05	Reject
8	0.328	0.152	0.291	0.171	< 0.05	Reject
9	0.572	0.206	0.350	0.184	< 0.05	Reject
10	0.078	0.055	0.153	0.108	< 0.05	Reject
11	0.586	0.193	0.288	0.206	< 0.05	Reject
12	0.327	0.175	0.268	0.149	< 0.05	Reject
13	0.215	0.211	0.260	0.201	0.178	*Accept*
14	0.585	0.191	0.363	0.158	< 0.05	Reject
15	0.312	0.144	0.335	0.154	< 0.05	Reject
16	0.347	0.226	0.295	0.132	0.068	*Accept*
17	0.584	0.192	0.273	0.188	< 0.05	Reject
18	0.295	0.154	0.325	0.176	< 0.05	Reject
19	0.346	0.212	0.338	0.191	< 0.05	Reject
20	0.497	0.131	0.470	0.193	0.067	*Accept*
21	0.587	0.186	0.215	0.162	< 0.05	Reject
22	0.551	0.236	0.191	0.177	< 0.05	Reject
23	0.573	0.209	0.251	0.189	< 0.05	Reject
24	0.586	0.168	0.178	0.204	< 0.05	Reject
25	0.578	0.193	0.232	0.202	< 0.05	Reject
26	0.582	0.199	0.328	0.210	< 0.05	Reject
27	0.585	0.186	0.300	0.180	< 0.05	Reject
28	0.573	0.178	0.332	0.203	< 0.05	Reject
29	0.581	0.195	0.121	0.164	< 0.05	Reject
30	0.584	0.186	0.126	0.111	< 0.05	Reject

Table 9
The summary of tests of significant difference between E_1 and E_2.

Case	Mean E_1	Std. E_1	Mean E_2	Std. E_2	*p*-Value	H_o
2	0.159	0.147	0.213	0.156	< 0.05	Reject
10	0.078	0.055	0.153	0.108	< 0.05	Reject
15	0.312	0.144	0.335	0.154	< 0.05	Reject
18	0.295	0.154	0.325	0.176	< 0.05	Reject

Table 10
The summary of the hypothesis testing.

Case	Mean E_1	Std. E_1	Mean E_2	Std. E_2	*p*-Value	H_o
3	0.388	0.206	0.219	0.176	< 0.05	Reject
4	0.589	0.186	0.238	0.189	< 0.05	Reject
5	0.279	0.146	0.208	0.183	< 0.05	Reject
6	0.557	0.221	0.275	0.187	< 0.05	Reject
7	0.572	0.196	0.247	0.153	< 0.05	Reject
8	0.328	0.152	0.291	0.171	< 0.05	Reject
9	0.572	0.206	0.350	0.184	< 0.05	Reject
11	0.586	0.193	0.288	0.206	< 0.05	Reject
12	0.327	0.175	0.268	0.149	< 0.05	Reject
14	0.585	0.191	0.363	0.158	< 0.05	Reject
17	0.584	0.192	0.273	0.188	< 0.05	Reject
19	0.346	0.212	0.338	0.191	< 0.05	Reject
21	0.587	0.186	0.215	0.162	< 0.05	Reject
22	0.551	0.236	0.191	0.177	< 0.05	Reject
23	0.573	0.209	0.251	0.189	< 0.05	Reject
24	0.586	0.168	0.178	0.204	< 0.05	Reject
25	0.578	0.193	0.232	0.202	< 0.05	Reject
26	0.582	0.199	0.328	0.210	< 0.05	Reject
27	0.585	0.186	0.300	0.180	< 0.05	Reject
28	0.573	0.178	0.332	0.203	< 0.05	Reject
29	0.581	0.195	0.121	0.164	< 0.05	Reject
30	0.584	0.186	0.126	0.111	< 0.05	Reject

from the *existing BPM* and not reflect the *performance target*. Furthermore, the IVE experiment may fail to address contextual factors that the *performance target* addresses, which in turn affects the *augmented BPM*. Finally, the *performance target* may be unrealistic and impossible to achieve. As a result, a thorough investigation and evaluation of this issue is needed.

- In the four cases, where the *updated BPMs* had better performance than the *augmented BPMs*, the mixture ratios (α) are high (i.e., 0.75, 0.85, 0.61, and 0.97 corresponding to the cases 2, 10, 15, and 18 respectively). In those cases, the *updated BPMs* were constructed using more knowledge of the *synthetic IVE datasets* than the knowledge of the *existing BPM datasets*. In addition, the probability of switching on of the *synthetic IVE datasets* of those cases was close to or higher than the *performance target*. Consequently, high αs and the probability of switching on of the *synthetic IVE datasets* caused the *updated BPMs* closer to the *performance target* than the *augmented BPMs*. Nevertheless, high αs contribute to heavily biases of the *updated BPMs* toward the *synthetic IVE datasets*, which may not always generate better results, if the *existing BPM datasets* are closer to the *performance target*. On the other hand, the GAN-based framework tries to appropriately mix (i.e., with minimum bias) the *existing BPM dataset* and the *synthetic IVE dataset* toward the *performance target*, therefore, the *augmented BPM* incorporate balanced knowledge of the *existing BPM dataset* and the *synthetic IVE dataset*.
- Even though the previous studies stated that office tasks [49–51] and switch locations [12] influence human-building interactions on lighting uses, the context-aware design-specific data showed a consistent pattern compared to previous studies with respect to tasks but not switch locations. The situation with respect to light switch locations may be explained by several reasons: 1) The effectiveness of stimuli (e.g., visual and audio cues): The experiment assigned the office tasks to the participants by using audio cues as stated previously. The participants did not perform actual tasks. By using only audio cues, the participants might not be stimulated enough to realize how much lighting intensity they really need; 2) The data collection procedure: For the switch locations, the experiment informed the participants about where the switch was by using audio and visual cues. The participants did not actually interact with the switch (e.g., walking to the switch at the door). Therefore, the participants might not attempt to differentiate the difference in terms of access to the switch at different locations; and 3) The choice of locations: It is also possible that indeed the two locations should not have any difference.
- The selection of the contextual factors (i.e., types of office task and locations of light switch) depends on the application and the knowledge of new contextual factors in literature. In addition, the selection depends on the virtual reality technologies because the contextual factors need to be model and experimented on in IVE. Therefore, to demonstrate and test the framework, in the meantime not to be limited by the time and resource needs for developing IVEs, we chose the two contextual factors that have been commonly identified in literature [12,49–51].

The framework does not limit the number or types of contextual factor that can be included. If other contextual factors are identified to have significant influence on human-building interactions related to lighting switch uses, they can be considered in this study. Their inclusion may completely change the characteristics of context-aware design-specific data. Consequently, the *augmented BPM* may change as well. Other types of contextual factor may increase the complexity of input parameters. However, the complexity of input parameters does not affect the application of the framework. With the limitation of IVE at this moment, some types of contextual factor (e.g., senses and climates) may not be easy to effectively simulate in IVE. Furthermore, the more contextual factors included in IVE experiments, the higher demand for experimental cost, resources, and times. To demonstrate the efficacy of the framework and due to limitations of IVE technologies, we only considered two contextual factors (i.e., the office tasks and the light switch locations) in the experiment. In summary, the framework does not preclude any additional contextual factors. Nevertheless, users of the framework need to consider the tradeoff between the desired number of contextual factors and the increase in time and resource needs, and complexity.

- The framework is generic and parametric. It can take any BPM and context-aware design-specific data. For example, if there is a BPM modeling the performance of a multioccupancy space and effective and reliable IVE to observe human-building interactions in such a space, the framework can generate an *augmented BPM* for modeling the performance of the multioccupancy space. Therefore, characteristics of BPMs and the nature of human-building interactions do not affect the performance of the framework. However, the complexity in the development of BPMs and data collection of human-building interactions may vary, which influences the application of the framework. For instance, developing BPMs and collecting data in IVE for analyzing building performance in multioccupancy spaces is more complex. If such models and IVE data are available, the framework can produce an *augmented BPM*. Unfortunately, virtual reality technologies currently lack the capability to simulate human-building interaction scenarios in multioccupancy spaces. This is the reason we chose a space with a single occupancy, since the goal of the application is to show the efficacy of the framework.

8. Limitations of the study

Major limitations of the study can be discussed in the following:

- An approach to establish a *performance target* is not included in the framework. An approach to map design goals and objectives of buildings into a computational target is needed.
- As mentioned in the discussion, the most appropriate mixture may be obtained when context-aware design-specific data are relatively close to a *performance target*. If a target is unrealistic and context-aware design-specific data are relatively close to a *performance target*, an *augmented BPM* may be unrealistic as well. However, the framework does not yet have a method to assess whether a *performance target* is realistic.
- Since IVE experiments cannot be conducted for long period of time, the capability of IVEs for collecting longitudinal data is limited [23,70]. Accordingly, the IVE experiment is constructed using discrete events, which may not thoroughly cover all possible situations.
- Currently, visual simulation is one of the most matured IVE capabilities. To simulate other sensations (e.g., thermal, and scent), there is a need to integrate IVE with other equipment or devices (e.g., an external heating/cooling device to simulate thermal sensation). With limited resources (e.g., times, costs, and tools), we selected lighting performance to demonstrate the efficacy of the framework. Future work is needed to test the performance of the framework using different categories of building performance models.
- Uncertainties of the components in the framework such as the *existing BPM*, the context-aware specific data, the *performance target*, and structures of the computation may affect the development of an *augmented BPM*. The framework lacks uncertainty and sensitivity analyses for the components. Being able to analyze the uncertainty and sensitivity of the framework can significantly contribute to the improvement of the framework.

9. Conclusions and future work

The results of the hypothesis tests have shown that in most cases the *augmented BPM* has higher accuracy than the *updated BPM*, which suggests that the GAN-based framework is in general better in performance than the previous ANN-based greedy algorithm. However, in a few cases, the opposite is observed. Causes of the instability in performance of the framework require further research. In general, the selection of the *performance target* and the IVE experiments are potentially the main causes. Therefore, further research is needed to create a technique that can analyze the uncertainty, sensitivity, and robustness of the framework including data from IVE experiments, and a method to map *performance targets* between the design level and the computational level. Furthermore, methods to effectively and efficiently identify contextual factors (e.g., causality analysis [71], unsupervised approaches [72], and feature ranking [15]) need more research attention.

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.

Acknowledgements

This paper was partially supported by the U.S. National Science Foundation Award #1640818. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author (s) and do not necessarily reflect the views of the National Science Foundation.

References

[1] G. Branco, B. Lachal, P. Gallinelli, W. Weber, Predicted versus observed heat consumption of a low energy multifamily complex in Switzerland based on long-term experimental data, Energy and Buildings 36 (6) (2004) 543–555, https://doi.org/10.1016/j.enbuild.2004.01.028.
[2] C. van Dronkelaar, M. Dowson, C. Spataru, D. Mumovic, A review of the energy performance gap and its underlying causes in non-domestic buildings, Frontiers in Mechanical Engineering 1 (2016) 1–14, https://doi.org/10.3389/fmech.2015.00017.
[3] B. Frei, C. Sagerschnig, D. Gyalistras, Performance gaps in Swiss buildings: an analysis of conflicting objectives and mitigation strategies, Energy Procedia 122 (2017) 421–426, https://doi.org/10.1016/j.egypro.2017.07.425.
[4] C.M. Clevenger, J.R. Haymaker, M. Jalili, Demonstrating the impact of the occupant on building performance, J. Comput. Civ. Eng. 28 (1) (2014) 99–102, https://doi.org/10.1061/(ASCE)CP.1943-5487.0000323.
[5] H.B. Gunay, W. O'Brien, I. Beausoleil-Morrison, S. Gilani, Development and implementation of an adaptive lighting and blinds control algorithm, Build. Environ. 113 (2017) 185–199, https://doi.org/10.1016/j.buildenv.2016.08.027.
[6] T. Hong, S. D'Oca, W.J.N. Turner, S.C. Taylor-lange, An ontology to represent energy-related occupant behavior in buildings. Part I: introduction to the DNAs framework, Build. Environ. 92 (2015) 764–777, https://doi.org/10.1016/j.buildenv.2015.02.019.
[7] V. Fabi, S. Corgnati, Rune Vinther Andersen, B.W. Olesen, R.V. Andersen, S. Corgnati, B.W. Olesen, Occupants' window opening behaviour: A literature review of factors influencing occupant behaviour and models, Build. Environ. 58 (2012) 188–198, https://doi.org/10.1016/j.buildenv.2012.07.009.
[8] J.F. Sowa, Syntax, semantics, and pragmatics of contexts, International Conference on Conceptual Structures, 954 1995, pp. 1–15, , https://doi.org/10.1007/3-540-60161-9_25.
[9] M. Kinateder, M. Müller, M. Jost, A. Mühlberger, P. Pauli, Social influence in a virtual tunnel fire - influence of conflicting information on evacuation behavior, Appl. Ergon. 45 (6) (2014) 1649–1659, https://doi.org/10.1016/j.apergo.2014.05.014.
[10] M. Kobes, I. Helsloot, B. De Vries, J. Post, Exit choice,(pre-) movement time and (pre-) evacuation behaviour in hotel fire evacuation - behavioural analysis and validation of the use of serious gaming in experimental research, Procedia Engineering 3 (2010) 37–51, https://doi.org/10.1016/j.proeng.2010.07.006.

[11] Y. Hong, P. Michalatos, LumiSpace: A VR architectural daylighting design system, SIGGRAPH ASIA 2016 Virtual Reality Meets Physical Reality: Modelling and Simulating Virtual Humans and Environments, 2016, pp. 1–2, , https://doi.org/10.1145/2992138.2992140.

[12] A. Heydarian, J.P. Carneiro, D. Gerber, B. Becerik-Gerber, Immersive virtual environments, understanding the impact of design features and occupant choice upon lighting for building performance, Build. Environ. 89 (2015) 217–228, https://doi.org/10.1016/j.buildenv.2015.02.038.

[13] S. Saeidi, T. Rizzuto, Y. Zhu, R. Kooima, Measuring the effectiveness of an immersive virtual environment for the modeling and prediction of occupant behavior, Sustainable Human–Building Ecosystems, 2015, pp. 159–167, , https://doi.org/10.1061/9780784479681.017.

[14] S. Niu, W. Pan, Y. Zhao, A virtual reality integrated design approach to improving occupancy information integrity for closing the building energy performance gap, Sustain. Cities Soc. 27 (2015) 275–286, https://doi.org/10.1016/j.scs.2016.03.010.

[15] C. Chokwitthaya, Y. Zhu, R. Dibiano, S. Mukhopadhyay, Combining context-aware design-specific data and building performance models to improve building performance predictions during design, Autom. Constr. 107 (2019) 102917, https://doi.org/10.1016/j.autcon.2019.102917.

[16] A. Lijoi, R.H. Mena, I. Prünster, Controlling the reinforcement in Bayesian non-parametric mixture models, Journal of the Royal Statistical Society: Series B (Statistical Methodology) 69 (4) (2007) 715–740, https://doi.org/10.1111/j.1467-9868.2007.00609.x.

[17] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, Y. Bengio, Generative adversarial nets, Advances in Neural Information Processing Systems, 2014, pp. 2672–2680 http://papers.nips.cc/paper/5423-generative-adversarial-nets (Accessed date: July 14, 2019).

[18] R.E. Kalman, A new approach to linear filtering and prediction problems, J. Basic Eng. 82 (1) (1960) 35–45, https://doi.org/10.1115/1.3662552.

[19] F. Tahmasebi, A. Mahdavi, An inquiry into the reliability of window operation models in building performance simulation, Build. Environ. 105 (2016) 343–357, https://doi.org/10.1016/j.buildenv.2016.06.013.

[20] M. Keller, J. O'Donnell, K. Menzel, Integrating the specification, acquisition and processing of building performance information, Tsinghua Science and Technology 13 (1) (2008) 1–6 (doi: 10.1.1.457.374).

[21] C.S. Cho, D. Chen, S. Woo, Building information modeling (BIM) - based design of energy efficient buildings, International Symposium on Automation and Robotics in Construction, 2011, pp. 1079–1084, , https://doi.org/10.22260/isarc2011/0198.

[22] D.R.G. Hunt, Predicting artificial lighting use- a method based upon observed patterns of behavior, Light. Res. Technol. 12 (1) (1980) 7–14, https://doi.org/10.1177/096032718001200102.

[23] S. Saeidi, C. Chokwitthaya, Y. Zhu, M. Sun, Spatial-temporal event-driven modeling for occupant behavior studies using immersive virtual environments, Autom. Constr. 94 (2018) 371–382, https://doi.org/10.1016/j.autcon.2018.07.019.

[24] P.S. Bordnick, K.M. Graap, H.L. Copp, J. Brooks, M. Ferrer, Virtual reality cue reactivity assessment in cigarette smokers, CyberPsychol. Behav. 8 (5) (2005) 487–492, https://doi.org/10.1089/cpb.2005.8.487.

[25] D.A. Das, K.A. Grimmer, A.L. Sparnon, S.E. McRae, B.H. Thomas, The efficacy of playing a virtual reality game in modulating pain for children with acute burn injuries: a randomized controlled trial, BMC Pediatr. 5 (1) (2005) 1–10, https://doi.org/10.1186/1471-2431-5-1.

[26] F.W.H. Wong, P.T.I. Lam, E.H.W. Chan, Optimising design objectives using the Balanced Scorecard approach, Des. Stud. 30 (4) (2009) 369–392, https://doi.org/10.1016/j.destud.2008.10.004.

[27] Y. Lecun, Y. Bengio, G. Hinton, Deep learning, Nature 521 (7553) (2015) 436–444, https://doi.org/10.1038/nature14539.

[28] E. Collier, R. Dibiano, S. Mukhopadhyay, CactusNets: layer applicability as a metric for transfer learning, International Joint Conference on Neural Networks, 2018, pp. 1–8, , https://doi.org/10.1109/IJCNN.2018.8489649.

[29] S. Basu, S. Ganguly, S. Mukhopadhyay, R. DiBiano, M. Karki, R. Nemani, DeepSat - a learning framework for satellite imagery, International Conference on Advances in Geographic Information Systems, 2015, pp. 1–10, , https://doi.org/10.1145/2820783.2820816.

[30] S. Basu, S. Mukhopadhyay, M. Karki, R. DiBiano, S. Ganguly, R. Nemani, S. Gayaka, Deep neural networks for texture classification—a theoretical analysis, Neural Netw. 97 (2018) 173–182, https://doi.org/10.1016/j.neunet.2017.10.001.

[31] E. Collier, K. Duffy, S. Ganguly, G. Madanguit, S. Kalia, G. Shreekant, R. Nemani, A. Michaelis, S. Li, A. Ganguly, S. Mukhopadhyay, Progressively growing generative adversarial networks for high resolution semantic segmentation of satellite images, International Conference on Data Mining Workshops, 2018, pp. 763–769, , https://doi.org/10.1109/ICDMW.2018.00115.

[32] Q. Liu, S. Mukhopadhyay, M.X.B. Rodriguez, X. Fu, S. Sahu, D. Burk, M. Gartia, A one-shot learning framework for assessment of fibrillar collagen from second harmonic generation images of an infarcted myocardium, International Symposium on Biomedical Imaging, 2020, pp. 839–843, , https://doi.org/10.1109/ISBI45749.2020.9098444.

[33] Q. Liu, E. Collier, S. Mukhopadhyay, PCGAN-CHAR: progressively trained classifier generative adversarial networks for classification of noisy handwritten bangla characters, International Conference on Asian Digital Libraries, 2019, pp. 3–15, , https://doi.org/10.1007/978-3-030-34058-2_1.

[34] J.T. Springenberg, Unsupervised and semi-supervised learning with categorical generative adversarial networks, International Conference on Learning Representations, 2016, pp. 1–20 http://arxiv.org/abs/1511.06390 (Accessed date: August 12, 2019).

[35] C. Chokwitthaya, E. Collier, Y. Zhu, S. Mukhopadhyay, Improving prediction accuracy in building performance models using generative adversarial networks (GANs), International Joint Conference on Neural Networks, 2019, pp. 1–9, , https://doi.org/10.1109/IJCNN.2019.8852411.

[36] M. Mirza, S. Osindero, Conditional generative adversarial nets, ArXiv Preprint ArXiv:1411.1784, 2014, pp. 1–7 http://arxiv.org/abs/1411.1784 (Accessed date: September 23, 2019).

[37] J. Voss, Revisiting Office Space Standards, https://www.brigholme.com/files/resources/Revisiting-office-space-standards-white-paper.pdf (Accessed date: October 03, 2019).

[38] P.C. Da Silva, V. Leal, M. Andersen, Occupants interaction with electric lighting and shading systems in real single-occupied offices: results from a monitoring campaign, Build. Environ. 64 (2013) 152–168, https://doi.org/10.1016/j.buildenv.2013.03.015.

[39] C.F. Reinhart, Lightswitch-2002: a model for manual and automated control of electric lighting and blinds, Sol. Energy 77 (1) (2004) 15–28, https://doi.org/10.1016/j.solener.2004.04.003.

[40] D. Bourgeois, C. Reinhart, I. Macdonald, Adding advanced behavioural models in whole building energy simulation: a study on the total energy impact of manual and automated lighting control, Energy and Buildings 38 (7) (2006) 814–823, https://doi.org/10.1016/j.enbuild.2006.03.002.

[41] G.R. Newsham, Manual control of window blinds and electric lighting: implications for comfort and energy consumption, Indoor and Built Environment 3 (3) (1994) 135–144, https://doi.org/10.1177/1420326X9400300307.

[42] D.P. Kroese, T. Brereton, T. Taimre, Z.I. Botev, Why the Monte Carlo method is so important today, Wiley Interdisciplinary Reviews: Computational Statistics 6 (6) (2014) 386–392, https://doi.org/10.1002/wics.1314.

[43] H. Ishwaran, G. Takahara, Independent and identically distributed Monte Carlo algorithms for semiparametric linear mixed models, J. Am. Stat. Assoc. 97 (460) (2002) 1154–1166, https://doi.org/10.1198/016214502388618951.

[44] M.N. Almarshad, M. Barkat, S.A. Alshebeili, A Monte Carlo simulation for two novel automatic censoring techniques of radar interfering targets in log-normal clutter, Signal Process. 88 (3) (2008) 719–732, https://doi.org/10.1016/j.sigpro.2007.09.013.

[45] T. Homem-De-Mello, On rates of convergence for stochastic optimization problems under non-independent and identically distributed sampling, SIAM J. Optim. 19 (2) (2008) 524–551, https://doi.org/10.1137/060657418.

[46] P. Cunningham, J. Carney, S. Jacob, Stability problems with artificial neural networks and the ensemble solution, Artif. Intell. Med. 20 (3) (2000) 217–225, https://doi.org/10.1016/S0933-3657(00)00065-8.

[47] C. Chokwitthaya, Y. Zhu, R. Dibiano, M. Supratik, A machine learning algorithm to improve building performance modeling during design, MethodsX 7 (2020) 35–49, https://doi.org/10.1016/j.mex.2019.10.037.

[48] C. Reinhart, P.F. Breton, Experimental validation of 3ds Max design 2009 and Daysim 3.0, International Building Performance Simulation Association, 2009, pp. 1514–1521 (doi: 10.1.1.172.3251).

[49] P. Boyce, Lighting research for interiors: the beginning of the end or the end of the beginning, Light. Res. Technol. 36 (4) (2004) 283–293, https://doi.org/10.1191/11477153504li118oa.

[50] P.R. Boyce, N.H. Eklund, S.N. Simpson, Individual lighting control: task performance, mood, and illuminance, J. Illum. Eng. Soc. 29 (1) (2000) 131–142, https://doi.org/10.1080/00994480.2000.10748488.

[51] J.A. Veitch, G.R. Newsham, Lighting quality and energy-efficiency effects on task performance, mood, health, satisfaction and comfort, J. Illum. Eng. Soc. 27 (1) (1998) 107–129, https://doi.org/10.1080/00994480.1998.10748216.

[52] U.S. General Services Administration, https://www.gsa.gov/node/82715 (Accessed date: July 06, 2019).

[53] D.R.G. Hunt, The use of artificial lighting in relation to daylight levels and occupancy, Build. Environ. 14 (1) (1979) 21–33, https://doi.org/10.1016/0360-1323(79)90025-8.

[54] R. Likert, A technique for the measurement of attitudes, Arch. Psychol. 22 (140) (1932) 1–55 https://legacy.voteview.com/pdf/Likert_1932.pdf (Accessed date: August 17, 2019).

[55] C.M. Bishop, Pattern Recognition and Machine Learning, Springer, 2006 (ISBN 13: 9780387310732; ISBN 10: 0387310738).

[56] Z. Wang, C. Da Cunha, M. Ritou, B. Furet, Comparison of K-means and GMM methods for contextual clustering in HSM, International Conference on Changeable, Agile, Reconfigurable and Virtual Production, 2019, pp. 154–159, , https://doi.org/10.1016/j.promfg.2018.12.025.

[57] R.B. Lanjewar, S. Mathurkar, N. Patel, Implementation and comparison of speech emotion recognition system using Gaussian mixture model (GMM) and K-nearest neighbor (*K*-NN) techniques, Procedia Computer Science 49 (1) (2015) 50–57, https://doi.org/10.1016/j.procs.2015.04.226.

[58] G.S. Morrison, A comparison of procedures for the calculation of forensic likelihood ratios from acoustic-phonetic data: multivariate kernel density (MVKD) versus Gaussian mixture model-universal background model (GMM-UBM), Speech Comm. 53 (2) (2011) 242–256, https://doi.org/10.1016/j.specom.2010.09.005.

[59] B. Awwad, S. Hasan, J.Q. Gan, Machine learning and data mining in pattern recognition, International Workshop on Machine Learning and Data Mining in Pattern Recognition, 3587 2005, pp. 96–106, , https://doi.org/10.1007/b138149.

[60] Q. Liu, S. Mukhopadhyay, Y. Zhu, R. Gudishala, Improving route choice models by incorporating contextual factors via knowledge distillation, International Joint Conference on Neural Networks, 2019, pp. 1–8, , https://doi.org/10.1109/IJCNN.2019.8852482.

[61] C. Chokwitthaya, Y. Zhu, A. Jafari, Applying the Gaussian mixture model to generate large synthetic data from a small data set, Construction Research Congress, 2020, pp. 1–10 https://www.researchgate.net/profile/Chanachok_Chokwitthaya/publication/335662593_Applying_the_Gaussian_Mixture_Model_to_Generate_Large_

Synthetic_Data_from_a_Small_Data_Set/links/5d7283a84585151ee4a0f655/Applying-the-Gaussian-Mixture-Model-to-Generate-Lar.
[62] K.Y. Lee, Local fuzzy PCA based GMM with dimension reduction on speaker identification, Pattern Recogn. Lett. 25 (16) (2004) 1811–1817, https://doi.org/10.1016/j.patrec.2004.07.006.
[63] H. Rijal, Investigation of comfort temperature and occupant behavior in Japanese houses during the hot and humid season, Buildings 4 (3) (2014) 437–452, https://doi.org/10.3390/buildings4030437.
[64] G.Y. Yun, K. Steemers, Time-dependent occupant behaviour models of window control in summer, Build. Environ. 43 (9) (2008) 1471–1482, https://doi.org/10.1016/j.buildenv.2007.08.001.
[65] P.V. O'neil, Advanced engineering mathematics, Cengage Learning, 1991 (ISBN 13: 9780534983185; ISBN 10: 0534983189).
[66] Hui Zou, H. Trevor, Regularization and variable selection via the elastic net, Journal of the Royal Statistical Society: Series B (Statistical Methodology) 67 (2) (2005) 301–320, https://doi.org/10.1111/j.1467-9868.2005.00503.x.
[67] C.F. Reinhart, K. Voss, Monitoring manual control of electric lighting and blinds, Light. Res. Technol. 35 (3) (2003) 243–260, https://doi.org/10.1191/1477153503li064oa.
[68] D. Yan, X. Feng, Y. Jin, C. Wang, The evaluation of stochastic occupant behavior models from an application-oriented perspective: using the lighting behavior model as a case study, Energy and Buildings 176 (2018) 151–162, https://doi.org/10.1016/j.enbuild.2018.07.037.
[69] Y. Zhu, S. Saeidi, T. Rizzuto, A. Roetzel, R. Kooima, Potential and challenges of immersive virtual environments for occupant energy behavior modeling and validation: a literature review, Journal of Building Engineering 19 (2018) 302–319, https://doi.org/10.1016/j.jobe.2018.05.017.
[70] C. Chokwitthaya, R. Dibiano, S. Saeidi, S. Mukhopadhyay, Y. Zhu, Enhancing the prediction of artificial lighting control behavior using virtual reality (VR): a pilot study, Construction Research Congress, 2018, pp. 216–233, , https://doi.org/10.1061/9780784481301.022.
[71] A. Nabijiang, S. Mukhopadhyay, Y. Zhu, R. Gudishala, S. Saeidi, Q. Liu, Why do you take that route? Cognitive Science Society, 2019, pp. 2454–2460 https://cogsci.mindmodeling.org/2019/papers/0424/0424.pdf?fbclid=IwAR0sYRRN2_DTRm27DHvM0irXbCJ4D9PcWRJjJALlOMVVWjkx9JQq1eV9ZrY (Accessed date: October 12, 2019).
[72] Q. Liu, S. Mukhopadhyay, Unsupervised learning using pretrained CNN and associative memory bank, International Joint Conference on Neural Networks, 2018, pp. 1–8, , https://doi.org/10.1109/IJCNN.2018.8489408.