energies

Article

Providing Convenient Indoor Thermal Comfort in Real-Time Based on Energy-Efficiency IoT Network †

Bouziane Brik [1,*], Moez Esseghir [2], Leila Merghem-Boulahia [2] and Ahmed Hentati [2]

1 DRIVE EA1859, University of Bourgogne Franche-Comté, 58000 Nevers, France
2 Laboratory of Computer Science and Digital Society (LIST3N), University of Technology of Troyes (UTT), 10000 Troyes, France; moez.esseghir@utt.fr (M.E.); leila.merghem_boulahia@utt.fr (L.M.-B.); ahmad.hentati@gmail.com (A.H.)
* Correspondence: bouziane.brik@gmail.com

† This paper is an extended version of our paper published in: Brik, B.; Esseghir, M.; Merghem-Boulahia, L.; Snoussi, H. ThermCont: A machine Learning enabled Thermal Comfort Control Tool in a real time. 2019 15th International Wireless Communications Mobile Computing Conference (IWCMC), 2019; pp. 294–300. doi:10.1109/IWCMC.2019.8766697.

Abstract: Monitoring the thermal comfort of building occupants is crucial for ensuring sustainable and efficient energy consumption in residential buildings. It enables not only remote real-time detection of situations, but also a timely reaction to reduce the damage made by harmful situations in targeted buildings. In this paper, we first design a new Internet of Things (IoT) architecture in order to provide remote availability of both indoor and outdoor conditions, with respect to the limited energy of IoT devices. We then build a multi-output prediction model of indoor parameters using a random forest learning algorithm, and based on a longitudinal real dataset of one year. Our prediction model considers outdoor conditions to predict the indoor ones. Hence, it helps to detect discomfort situations in real-time when comparing predicted variables to real ones. Furthermore, when detecting an indoor thermal discomfort, we provide a new genetic-based algorithm to find the most suitable values of indoor parameters, enabling the improvement of the indoor occupants' thermal comfort. Numerical results show the efficiency of our prediction scheme, reaching an accuracy of 96%, as well as our genetic-based scheme in optimizing the indoor thermal parameters by 85%.

Keywords: IoT network; energy efficiency; indoor thermal comfort monitoring; machine learning; genetic algorithm

Citation: Brik, B.; Esseghir, M.; Merghem-Boulahia, L.; Hentati, A. Providing Convenient Indoor Thermal Comfort in Real-Time Based on Energy-Efficiency IoT Network. *Energies* **2022**, *15*, 808. https://doi.org/10.3390/en15030808

Academic Editor: Andrea Frazzica

Received: 17 December 2021
Accepted: 20 January 2022
Published: 23 January 2022

1. Introduction

Nowdays, people spend 87% of their time indoors, in addition to 6% in an enclosed vehicle, according to a U.S. Environmental Protection Agency (EPA) survey [1]. Moreover, thousands of people have died due to adverse climate conditions in the world [1,2]. For instance, the Office of National Statistics (ONS) of the U.K said that more than 45,000 people died in 2015 due to the cold, and 1000 more people died in summer 2018 than average, due to a heatwave [2]. Hence, occupants' thermal comfort depends on both indoor and outdoor conditions, and has strong implications for people's health, performance, and well-being. Therefore, providing the convenient parameters of indoor thermal comfort is becoming primordial.

Indoor thermal comfort monitoring has attracted a great deal of research attention in recent years, since it improves not only the occupants' comfort but also the efficiency of energy consumption in residential buildings [3–5]. This immense attraction is mainly due to the emergence of Internet of Things (IoT) technology that connects objects to the Internet through ubiquitous sensors [6]. IoT technology enables the collection of thermal comfort parameters in real-time from targeted buildings. Hence, the IoT helps to deal with the indoor thermal comfort monitoring in real time [7].

Thus, thermal comfort is defined as a "condition of mind that expresses satisfaction with the thermal environment and is assessed by subjective evaluation" [8]. In fact, a range of work has studied the indoor thermal comfort measurement [8]. In [8], Fanger proposed two thermal comfort indexes named Predicted Mean Vote (PMV) and Predicted Percentage Dissatisfied (PPD), which became the basis of many standards including ISO 7730 and ASHRAE.

The PMV predicts the thermal sensation of a larger group of people on a seven-point scale from cold to hot (see Table 1) while PPD predicts the percentage of thermally dissatisfied people who feel too cool or too warm (cf. Figure 1). However, PMV is a very complex mathematical formula since it depends on two human parameters: human activity and clothing insulation, and four environmental parameters: air relative humidity, air temperature, mean radiant temperature, and air velocity, which makes it difficult to measure in order to assess the indoor thermal comfort, especially in real time. Therefore, work, in such a context, focused on the PMV value prediction using machine learning algorithms such as Multiple Linear regression, Decision Tree, K-Neighbors, etc. [9–11]. The predicted PMV value is then compared to the real indoor thermal comfort in addition to Fanger'saim PMV to determine the prediction accuracy. However, these works did not deal with the indoor thermal comfort improvement when observing a thermal discomfort. This may be possible especially with the emergence of IoT networks. Moreover, to the best of our knowledge, no work has studied how outdoor environmental variables may affect Fanger's indoor parameters in order to provide indoor thermal comfort.

Table 1. Thermal Sensation Scale.

Value	Sensation
+3	Hot
+2	Warm
+1	Slightly warm
0	Neutral
−1	Slightly cool
−2	Cool
−3	Cold

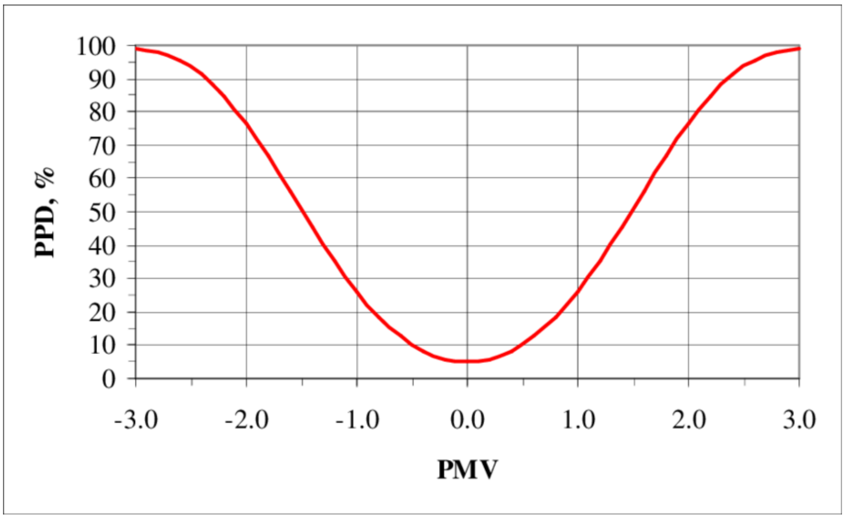

Figure 1. Predicted percentage dissatisfied (PPD) as a function of predicted mean vote (PMV).

In this paper, we first design a new Internet of Things (IoT) architecture in order to provide remote availability of both indoor and outdoor conditions, with respect to the limited energy of IoT devices. We then propose to predict the indoor Fanger's parameters by considering the outdoor parameters, rather than predicting the complicated PMV equation value. To do so, we generated a multi-output prediction model using the random forest algorithm [12] based on a longitudinal real dataset of one year [13]. Thus, our prediction model can then be exploited to predict the indoor parameters in real time. It is worth noting that there are some works which aim to predict indoor thermal comfort [14–16]; however, almost all of them are based on either occupants' perception or indoor thermal conditions to determine the suitable indoor thermal comfort. Moreover, when detecting an indoor thermal discomfort, we propose a new genetic-based algorithm to find the most suitable values of indoor parameters, enabling the improvement of the indoor occupants' thermal comfort. We also note that genetic algorithms rely on biologically-based operators such as crossover, selection, and mutation, to find good solutions for search and problems [17,18]. Experimental results show the efficiency, on one hand, of our random forest-based scheme in improving the prediction accuracy of the indoor thermal parameters by 96%, and, one the other hand, of our genetic-based scheme to determine the suitable parameter values of the indoor thermal comfort, with an accuracy of 85%. We note that our developed model can be applied to any air-conditioned office buildings to predict the indoor thermal comfort based on the output thermal conditions.

The remainder of this paper is organized as follows. Section 2 is about related work in the field of thermal comfort measurement. Section 3 describes our general approach to collect, predict, and control parameter values of the indoor thermal comfort. Evaluation of our schemes by simulation and obtained experimental results are discussed in Section 4. Section 5 concludes the paper. In addition, for ease of reading, Table 2 illustrates the list of abbreviations used in this paper.

Table 2. List of Acronyms.

Acronym	Meaning
AGT	Average Ground Truth
ARE	Average Relative Error
BP	Back Propagation
EPA	Environmental Protection Agency
GA	Genetic Algorithm
IoT	Internet of Things
LRAB	Locally Weighted Regression with Adaptive Bandwidth
ONS	Office of National Statistics
PPD	Predicted Percentage Dissatisfied
PMV	Predicted Mean Vote
RPL	Routing ProtocoL
SVC	Support Vector Classification
SVR	Support Vector Regression
QoS	Quality of Service

2. Related Work

A range of work has been proposed in order to study indoor thermal comfort through Fanger's PMV index and using machine learning algorithms. These works deal with indoor thermal comfort by either extending or reducing the considered Fanger's parameters.

2.1. Works Extending the Considered Parameters

The authors used the Support Vector Regression (SVR) algorithm in order to predict indoor thermal comfort through the PMV index in [19]. They considered Fanger's parameters in addition to the external weather conditions to estimate the relation between them, and, hence, to reduce energy consumption in the studied building. The results show a difference of "0.25" between both the desired and predicted PMV in terms of measurement accuracy.

In [20], the authors have optimized the PMV index by adding two new parameters, number of people in the room and indoor CO_2 concentration, to Fanger's parameters.

The authors then used the Back Propagation (BP) neural network in order to predict the PMV value. However, the authors did not give any prediction accuracy comparing to the traditional Fanger's PMV index.

Another study was performed in a clinical healthcare facility in Belgium, in [21], to analyze patients' thermal comfort. The authors compared between both objective (personal and environmental) parameters and subjective measurements (questionnaires). This study targets 99 patients of maternity, neurology, oncology, abdominal surgery, gastro-enterology, and thoraco-vascular surgery wards. In this work, both PPD and PMV were determined from a designed method by Lin and Deng [22], and ISO 7730 [23].

In [9], the authors used a genetic algorithm in order to reduce energy consumption and achieve a suitable thermal comfort. The authors used new parameters in addition to Fanger's parameters: the power consumption, the outdoor temperature, the outdoor humidity, and the interior volume size. Results show that energy consumption can be saved up to 52.02%.

2.2. Works Reducing the Considered Parameters

In [24], the authors proposed an optimized PMV index by simplifying the parameters of the PMV equation. They considered only two parameters, air temperature and mean radiant temperature. The results show that there is only a difference of "0.01" between Fanger's PMV and the simplified PMV by regression equation, as well as energy reduction by 5.6%.

In [25], a literature review on the thermal comfort of hospital's occupants has been done. The authors concluded that it is highly required to study the relationship between indoor thermal comfort conditions and the productivity of the hospital staff.

In [26], the authors compare three classifiers to predict the indoor thermal comfort and select the Support Vector Classification (SVC). To identify the best algorithm, the authors used Matlab's "Classification Learner" application to classify the data and to compare the accuracy by using SVC (95.80%), KNN (90.40%), and Decision Tree (85.80%). Thus, the learned classifiers consider both temperature and relative humidity as input data and feedback from the users about their thermal sensation as output data.

A Robust Locally Weighted Regression with Adaptive Bandwidth (LRAB) was used in order to predict the thermal comfort in [27]. The learned model considers only three parameters from Fanger's parameters, including air temperature, mean radiant temperature, and humidity. The results show that the used method is better than Fanger's PMV and the Nadaraya–Watson kernel methods in predicting thermal comfort. In addition, they also confirmed that the performance achieves a better prediction with more samples. The accuracy using LRAB was 69.27% and, using Nadaraya–Watson, it was 50.19%, while, using the PMV method, it was only 38.47%.

2.3. Comparison and Discussion

Table 3 compares between the above work according to several criteria such as the used machine learning algorithm, used parameters, achieved accuracy, new considered parameters, etc.

Table 3. Comparative table of the Related Work.

Work	Year	Algorithm	Used Parameters	Accuracy	New Parameters	Energy Saving	Difference with the Real PMV
Han et al. [24]	2014	Regression	-Air temperature -Radiant temperature			5.6%	0.01
Javed et al. [26]	2017	SVC	-Air temperature -Relative humidity	91.62%			
Viani et al. [19]	2017	SVR	-Air temperature -Outdoor temperature -Relative humidity		-Outdoor temperature		0.25
Nian et al. [20]	2017	BP neural Network	-Fanger's parameters -Indoor CO_2 -Number of people in the room		-Indoor CO_2 -Number of people in the room		
Ying et al. [9]	2016	Genetic Algorithm	-Fanger'sparameters -Power consumption -Outdoor temperature -Outdoor humidity -Interior volume size		-Power consumption -Outdoor temperature -Outdoor humidity -Interior volume size	52.02%	
Manna et al. [27]	2013	LRAB	-Air temperature -Humidity -Radiant temperature	69.27%			38.47% (Fanger's PMV)
		Nadaraya-Watson Kernel		50.19%			

Existing works focus on studying the indoor thermal comfort through the PMV index by either extending [19,20] or reducing the considered Fanger's parameters [24,26–28]. Also, these work aimed to either improve the accuracy of indoor thermal comfort measurement as compared to its measurement by Fanger's PMV index [26,27,29], or saving energy consumption in the target building [9,24,30]. In other words, most of the work focuses on how to obtain an indoor thermal comfort value that is as close as possible to the real indoor thermal comfort compared to Fanger's PMV index. However, they did not study how to improve and assess the indoor thermal comfort in real-time, especially when observing an indoor thermal discomfort. Moreover, and to the best of our knowledge, no work deals with the impact of outdoor conditions on the indoor thermal comfort and studies how to provide the convenient values of the indoor thermal comfort parameters, rather than to predict the complicated PMV equation value.

3. Indoor Thermal Comfort Parameters Assessment: An IoT-Based Architecture

In this section, we present our IoT-based architecture supporting the indoor thermal comfort parameters monitoring remotely. Section 3.1 gives an overview of our layered IoT-based architecture and Section 3.2 presents our problem modeling. Our machine learning algorithm to predict the indoor thermal comfort parameters is introduced in Section 3.3, and, finally, we present our indoor parameters assessment and optimization tool in Section 3.4.

3.1. Remote Availability: How to Collect Targeted Data Remotely and in Real-Time?

Figure 2 illustrates our indoor parameters prediction process that comprises three main steps: (i) sensor deployment in targeted buildings and data gathering; (ii) data storage and analysis using the machine learning algorithm; (iii) feedbacks to end users to make the most suitable decision and, hence, to guarantee an adequate indoor thermal comfort.

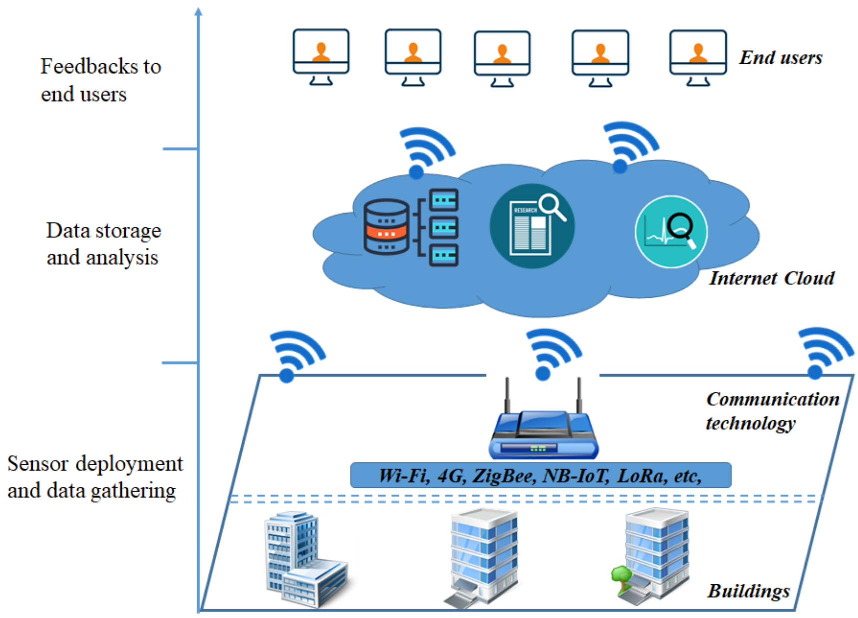

Figure 2. Illustration of the proposed architecture.

Therefore, our IoT-based architecture comprises three main layers: perception, network, and application layer.

The perception layer is responsible for sensing targeted data from the physical world using various sensing equipment. In our case, we need to collect the state of windows and fans from the indoor building through sensor nodes while the outdoor parameters may be collected from the Internet such as outdoor temperature, humidity, air velocity, etc.

The collected data are then sent to centralized entities for further storage and analysis, e.g., data centers in the Internet Cloud, through the network layer and using communication protocols including cellular networks (2G/3G/4G), narrowband IoT (NB-IoT), short-range communication networks (Bluetooth, Wi-fi, Zigbee, etc.), and long-range wireless networks (LoRa, openRF, etc.).

The application layer interacts with end users for further analysis and decision-making. In other words, when detecting an indoor thermal discomfort, the application layer gives building users feedback to help them in making the more adequate decision.

3.2. Machine Learning: How to Model the Indoor Thermal Comfort Parameters Prediction?

We consider long-term data including both outdoor parameters such as outdoor ambient temperature, outdoor humidity, outdoor air velocity, etc., and their corresponding indoor influencing parameters on occupant's thermal comfort such as air temperature, air relative humidity, air velocity, etc. The dataset comprises m representative samples (rows) and h features (columns), and is divided into two sets: the training (or learning) set, to train and create the model, and the testing set, to evaluate the thermal comfort prediction. We used a supervised machine learning algorithm which takes both input and output (desired) data to learn a mapping function, $F(x)$, from an h-dimensional input vector, $x = \left(x^{(1)}, \ldots, x^{(h)}\right)^T \in \mathbb{R}^h$, to a t-dimensional output vector $y = \left(y^{(1)}, \ldots, y^{(t)}\right)$, or $y = F(x)$.

Supervised learning includes two classes of algorithms: regression, for continuous-response values, and classification, for categorical response values. As our y_i are continuous in nature, our problem is a regression problem. We used the random forest regressor algorithm to create a prediction model of indoor occupants' thermal comfort variables across all seasons. Random forest regressor is the most common form of linear regression analysis, which attempts to learn a multi-target regression model consisting of finding the function F that assigns to each instance, given by a vector x, a vector y of t target values:

$$F : \Omega_{x_1} \times \Omega_{x_2} \times \cdots \times \Omega_{x_h} \to \Omega_{y_1} \times \Omega_{y_2} \times \cdots \times \Omega_{y_t}$$
$$x = (x_1, \ldots, x_h) \to y = (y_1, ..., y_t)$$

where Ω_{x_i} and Ω_{y_j} are the sample spaces of each predictive variable x_i, for all $i \in 1, \ldots, h$, and each target variable y_j, for all $j \in 1, \ldots, t$, respectively.

3.3. Parameters Prediction: How to Predict the Indoor Thermal Comfort Parameters?

Assuming that we receive, periodically, the outdoor parameter values which can be collected throughout the Internet or the sensors, we aim to predict the indoor thermal comfort parameters in real-time in order to provide the convenient thermal comfort in the target building. In other words, if we predict a thermal discomfort in the building, we give feedback to end users, hence, allowing them to make adequate decisions, improving the indoor thermal comfort such as turning on/off the AC, opening windows or doors for comfort ventilation, and changing occupants' clothing. Our work focuses on detecting discomfort situations to inform users in real-time. However, deciding which decisions that users should take and how to organize their feedback are out of the scope of this work. In addition, in our study, we leveraged the cloud environment only when training our prediction model. However, once building our learning model, the latter will be deployed locally in the building to predict the indoor thermal comfort.

Figure 3A illustrates our prediction module which is responsible for training, offline, the prediction model by means of both the random forest regressor algorithm and the longitudinal data of one year. As a result, we create a model that is able to predict the indoor thermal comfort parameter values at any time throughout the year.

- Longitudinal Data Collection
  We consider a publicly available dataset [13]. The collected dataset is from a medium size office (5388.3763 square meters) of 24 occupants at Philadelphia for a whole year starting July 2012. The data were collected each 15 minutes via both online daily surveys and data logger measurements of the indoor thermal environment, occupants' behavior, and outdoor environmental parameters. The dataset comprises eight features as input data including absolute time, outdoor ambient temperature, outdoor humidity, outdoor air velocity, occupant age, floor number, the state of the fan and the state of window, and six parameters as outputs: indoor air temperature, indoor humidity, indoor air velocity, mean radiant temperature, metabolism rate, and clothing insulation. The size of input space is $[m; h]$. The number of rows is $m = 24 \times 24 \times 4 \times 365 = 840{,}960$ rows and the number of columns corresponds to the number of input features $h = 8$. The size of the output space is $[m; 6]$ which corresponds to the six parameter values associated to the input features. We note that once we predict the six thermal parameters, we extract their corresponding PMV value and, hence, it is easy for us to then determine the discomfort situation;
- Data Preparation Process
  Due to the fact that the machine-learning algorithm learns from data, it is critical to feed it with the accurate and meaningful data and to make sure that the data are in a useful format and scale. The learning algorithm is then applied to produce a regressor.
  - Pre-processing. The input parameter values must be on a similar range and scale. Therefore, we use the common method Min-Max Normalization that normalizes the data values from 0 to 1:
    $$v' = \frac{v - min_f}{max_f - min_f} \tag{1}$$
    where v is the current value of feature f, and max_f and min_f are, respectively, the maximum and minimum values of feature f;
  - Split the Data into training and test set. In this step, we split the data into training and test (evaluation) subsets, usually with a ratio of 70–80 percent for training and 20–30 percent for test. The machine-learning algorithm then uses the training data to train the model and uses the evaluation data to evaluate the predictive accuracy of the trained model;
- Machine Learning Algorithm
  After having processed the data, we use random forest regressor to create the prediction model of occupants' thermal comfort parameters. We note that, to deal with the overfitting problem, we used the standard k-fold cross-validation, which divides the data into k subsets, named folds. Then, it iteratively trains the algorithm on $k - 1$ folds, while using the remaining fold as the test set. Then, we evaluate the accuracy of the model using several metrics (Section 4.1).
  Algorithm 1 illustrates our machine-learning algorithm.

Algorithm 1 Random Forest Regressor Algorithm

Require: Input ($X[m, h]$) and output ($y[m, 6]$) dataset.
Ensure: Prediction model of the indoor thermal comfort parameters values

```
//Pre-processing the dataset (similar scale and range)
X_proc = Pre_processing(X)
y_proc = Pre_processing(y)
//Split the data into training (75%) and test (25%)set
X_train, X_test, y_train, y_test = train_test_split
(X_proc, y_proc, test_size = 0.25)
//Call Random Forest algorithm
Regressor = RandomForestRegressor()
Regressor.fit (X_train, y_train)
//Predict the indoor parameters values
Para_predic = Regressor.predict(X_test)
//Evaluate model accuracy
Evaluate(Regressor)
return Regressor
```

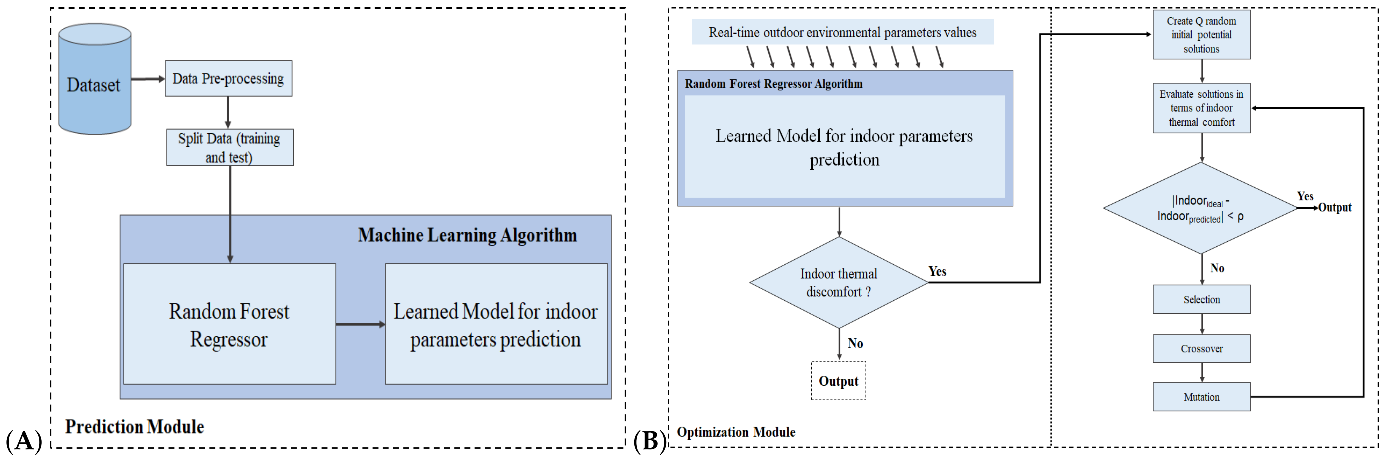

Figure 3. Our modules of indoor parameters assessment. (**A**) Indoor parameters prediction module. (**B**) Indoor parameters optimization module.

3.4. Indoor Thermal Comfort Assessment: How to Optimize the Indoor Thermal Comfort?

When predicting an indoor thermal discomfort, we used the meta-heuristic Genetic Algorithm (GA) to search the most adequate combination of indoor parameters' values (solutions), improving the indoor thermal comfort. This results in a GA comprising different steps, as illustrated in Figure 3B and detailed below.

1. Create Q potential solutions: we create a set of $Q = \{\sigma^1, \sigma^2, \ldots, \sigma^n\}$ random potential solutions (also called chromosomes or individuals), where n is the initial population size. We denote by $\sigma^i = \{\sigma^i_1, \sigma^i_2, \ldots, \sigma^i_k\}$ the configuration parameters vector of the individual i, with σ^i_j denoting a target parameter value, e.g., the temperature or humidity of individual i;
2. Evaluation of solutions performance: the objective is to define the evaluation function (fitness) of each solution (chromosome). For each feasible solution, this function takes as its input the configuration parameters vector to estimate the PMV value and, hence, the indoor thermal comfort. Then, according the required indoor thermal comfort, the absolute difference between predicted and ideal value of the indoor thermal comfort is used as the solution's fitness.
   As we mentioned before, we aim to improve the indoor thermal comfort. Therefore, if the absolute difference is below a threshold (ρ), the chromosome j is selected as a

solution and its parameter values are sent to the building. Otherwise, the optimization passes to the next steps;

3. Selection: to generate a new population of chromosomes from the current population, this step of GA consists of selecting chromosomes (individuals) from the population to be parents to crossover (next step). To do so, we used a rank selection method that ranks the chromosomes based on their fitness;
4. Crossover: to produce children solutions by combining part of the genetic information from their parent solutions. Hence, crossover enables the generation of new chromosomes which are better than their parents if they inherit the best parents' features. We used the arithmetic recombination method, which takes the weighted average of the two parents. If X and Y are the parents, the crossover function returns:

$$Child = \alpha.X + (1-\alpha).Y \tag{2}$$

where α is a random value between $[0,1]$;
5. Mutation: aims to maintain genetic diversity from one population to the next. The mutation operator alters one or more gene values in a chromosome from its initial state in order to reach a better solution. We used the *random resetting* method, in which a random gene is selected and replaced by a random value from the set of permissible values.

We note that the found indoor parameters' values, by our genetic algorithm, are sent to end users, which can help them to make adequate decisions in order to improve their indoor thermal comfort.

Algorithm 2 illustrates the main steps of our GA.

Algorithm 2 Our Genetic Algorithm (GA)

Require: Initial population (Q) and their fitness values, size of population (n), number of generations (g), ideal indoor thermal comfort ($Indoor_{ideal}$).
Ensure: Solution σ^*.
Initialization
for $j = 1$ to g **do**
  //Calculate the fitness value of each feasible solution
  for all $\sigma^i \in Q$ **do**
    Fitness(σ^i) $= |Indoor(\sigma^i) - Indoor_{ideal}|$
  end for
  Rank selection based method
  Select the best e solution
  //Crossover and mutation
  Number of **Crossover** $n_c = (n-e)/2$
  for $t = 1$ to n_c **do**
    Randomly **select** two solutions σ^a and σ^b
    Generate σ^c by arithmetic recombination **Crossover** of σ^a and σ^b
    Mutate a randomly gene of σ^c using random resetting method
  end for
end for
return The best solution σ^*

4. Performance Evaluation

In this section, we present the experimental study that we performed to evaluate both our prediction and optimization algorithms.

4.1. Experiments Setup

We used scikit-learn library on Python to implement our random forest regressor algorithm. Once we created our indoor parameters prediction model, we then implemented our genetic algorithm to optimize the indoor thermal comfort parameters when observing an indoor thermal discomfort. In addition, to assess the performance of the random forest algorithm in predicting the indoor thermal comfort parameters, we compared it to three other regression-learning algorithms: (i) Multiple Linear regression, which models the relationship between two or more input variables and output variables by fitting a linear equation to input data; (ii) Decision Tree Regression, which develops a regression model in the form of a decision tree structure; (iii) K-Neighbors Regression, which predicts output data based on similarity measures (e.g., distance functions) between both input and output data. To be fair in our comparison, we note that, for each algorithm, we tried different configurations and selected only the best configuration, i.e., that gave us better accuracy. Moreover, to choose the suitable heuristic, we compared the genetic algorithm with the Tabu-search algorithm. The obtained results showed that the genetic algorithm highly outperforms the Tabu-search algorithm and, hence, Tabu-search is not suitable for our study. Therefore, we chose to present only the results related to the genetic algorithm.

Table 4 shows the main parameters of both the used dataset and genetic algorithm implementation.

Table 4. Implementation parameters.

Parameters	Values
Dataset and Machine Learning	
Time period	From July 2012 to August 2013
Number of occupants	24 occupants
Number of samples	840,984 samples
Number of input variables	8 variables
Number of output variables	6 variable
Percentage of training set	75% of the dataset
Percentage of test set	25% of the dataset
Machine learning algorithm	Random Forest Algorithm
Genetic Algorithm	
Q size (s)	$[5, 50]$ individuals
Number execution iteration	$[5, 50]$ iterations
Threshold ρ	0.01
Crossover α	0.7
Wireless Sensors	
Simulation Time	3600 s
Communication Standard	IEEE 802.15.4
Application payload	30 bytes
Transport Layer	UDP
Network Layer	IPv6 + RPL
Buffer size	8 packets
Mac reliability (ACK)	Enabled
MAC max. retransmission	3
Channel check rate	8 HZ
Max. frame size	127 bytes

We considered the following metrics to evaluate the random forest regressor's performance:

- *Modeling and predicting time*: reflecting the time complexity of the machine learning algorithm to create the prediction model and to make new predictions by using training and test sets, respectively;
- *R square of the test set*: *R square* is a measure of how close the data are to the fitted regression line of our model (model accuracy). It is also known as the coefficient of determination;
- *Adjusted R* square of the test set: due to the fact that the *R* square always increases as we add more predictors (input variables) to the model, *Adjusted R* square attempts to

correct this overestimation and might decrease if the new predictor does not improve the model. *Adjusted R* square is determined by the following formula:

$$R^2_{Adjusted} = 1 - \left[\frac{(1 - R^2)(m - 1)}{m - h - 1} \right] \quad (3)$$

where m and h are, respectively, the number of samples and input variables of the dataset;

- Average Relative Error (ARE): represents the sample standard deviation of the differences between predicted and real (observed) values. For each output datum, we predict its value and evaluate the performance against the real value in terms of the ARE as follows, $ARE = \frac{\sum_{i=1}^{m}(y_i - \widehat{y}_i)^2}{m}$;
- Average Ground Truth (AGT): for each output datum, AGT explains the average output data value away from the ground truth value when making predictions on the test set. The AGT is deduced from ARE as follows: $AGT = \sqrt{ARE}$.

Furthermore, we implement an indoor wireless sensor IoT-based network using Contiki 3 OS and Cooja emulator [31] in order to collect indoor data, such as the states of the fans and windows that we need as input data to then predict the indoor thermal comfort parameters. We note that outdoor data such as ambient temperature, humidity, air velocity, etc. may be collected through the Internet network.

Cooja is an IoT-based network emulator that runs on the ContikiOS. It supports the simulation of several types of sensor nodes and each node can be programmed individually with its own code. We note that we emulate the hardware of the real sensor nodes Tmote Z1 which is available in the Cooja emulator. In fact, Tmote Z1 sensors comprise ultra-low power MCU and a 2.4 GHz Transceiver. They include mainly two sensors: ADXL345 accelerometer and TMP102 temperature sensors. In our emulation, these sensors can communicate with each other using IEEE 802.15.4 standard, which is a short range wireless communication standard that offers low complexity, low data rate transmission, and low energy consumption. Moreover, we use the Powertrace tool to measure the energy consumption of each sensor node. Table 5 illustrates setting parameters in our emulation and the protocol stack.

Table 5. Simulation Parameters of wireless sensors.

Parameters	Values
Simulation Time	3600 s
Communication Standard	IEEE 802.15.4
Application payload	30 bytes
Transport Layer	UDP
Network Layer	IPv6 + RPL
Buffer size	8 packets
Mac reliability (ACK)	Enabled
MAC max. retransmission	3
Channel check rate	8 HZ
Max. frame size	127 bytes

Our network comprises three types of nodes: (i) terminal sensor nodes which are in charge of sensing targeted data periodically before sending them toward intermediate nodes; (ii) intermediate nodes to send collected data from terminal nodes to the sink node; (iii) a sink node, which is in charge of transferring sensed data to the cloud environment in which we predict the indoor thermal comfort parameters based on both collected indoor and outdoor data.

4.2. Performance Evaluation of Machine Learning Algorithms

Table 6 shows a performance comparison between all the machine-learning algorithms. We observe that the random forest algorithm improves the prediction accuracy (*R square* and *adjusted R*) as compared to the other algorithms. We also note that the random forest algorithm generates a long modeling time, 52.6 s, compared to the other algorithms. However, the modeling time is not that important as it is executed only once. Even the time complexity of the Decision Tree Regressor is better than that of our random forest algorithm; we chose the random forest because it clearly improves prediction accuracy (96%), which makes it more suitable to evaluate and control the indoor thermal comfort parameters.

Figure 4A–C show, respectively, the relative prediction error, the ground truth, and the prediction accuracy of the predicted indoor temperature value when varying the number of predicted samples. We remark that all algorithms generate a stable performance as the number of predicted samples is increased. Thus, both random forest and decision tree algorithms outperform the other algorithms.

Table 6. Results for different models applied to suggested dataset.

	Adjusted R	Modelling Time (s)	Predicting Time (s)
Multiple Linear Regressor	0.42685	0.51362	0.02094
KNeighbors Regressor	0.70651	806.0191	56.06211
Decision Tree Regressor	0.8467	7.13492	0.1835
Random Forest Regressor	0.9654	52.6785	1.6652

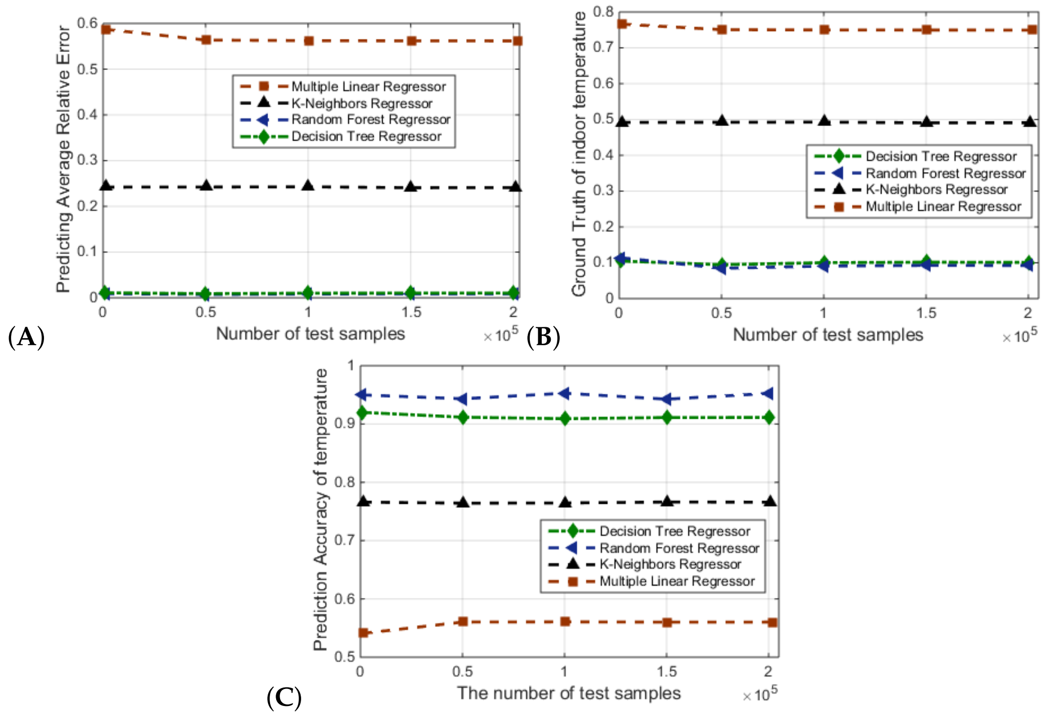

Figure 4. Comparison between machine learning algorithms in predicting the indoor temperature. (**A**) Prediction error. (**B**) Ground Truth. (**C**) Prediction accuracy.

In the following, we evaluate the prediction models of only random forest and decision tree algorithms, as they clearly outperform the other algorithms in terms of prediction accuracy. Figure 5A,B compare between the predicted and real value of indoor temperature and humidity, respectively, during a day (24 h). As we can see, both algorithms generate the same performance except at 9 a.m., 10 a.m., and 11 a.m. for the predicted temperature, and at 3 a.m. and 5 a.m. for the predicted humidity, where the random Forest regressor outperforms the "Decision Tree Regressor." These results confirm the aforementioned results (see Table 6 and Figure 4), showing the efficiency and the accuracy of the random forest algorithm in predicting the indoor thermal comfort parameters, noting that, in our accuracy study, we focus more on temperature and humidity, since we also obtained the same performance for the other parameters of indoor thermal comfort.

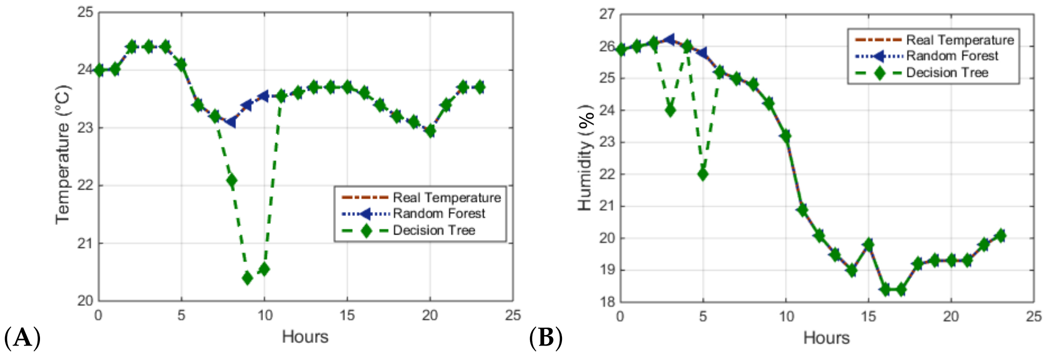

Figure 5. Comparison between predicted and real value of indoor temperature and humidity, during a day (24 h).

4.3. Performance Evaluation of Our Genetic Algorithm (GA)

To evaluate our GA, we assumed that we received, periodically, the input variables of our learned model including outdoor ambient temperature, outdoor humidity, outdoor air velocity, building occupant age, building floor number, and the state of both the fans and windows in the building. We then predict the indoor parameters of thermal comfort in the building in order to detect a discomfort situation as explained in section III-D. In the following, we focus on the control of indoor ambient temperature that we have to maintain between 15 °C and 25 °C.

Figure 6A compares the periodically predicted ambient temperature value in the building to the ideal indoor temperature value (desired value) and the optimized indoor temperature value, which is given by our GA in the case of detecting a thermal discomfort. As we can see, when an indoor thermal discomfort is observed, our GA improves the indoor temperature by adjusting its value and making it close or equal to the desired value, with an accuracy of 85%. Thus, we note that both predicted and optimized temperature values for May, June, and October are equal, because we did not observe a thermal discomfort. Therefore, we did not apply our GA.

Figure 6B,C show the predicted temperature value of the best chromosome (solution) when varying, respectively, the initial population size and the number of iterations of the GA. We observe that the GA cannot improve the indoor ambient temperature when the initial population size is less than 20 solutions. However, above 20 individuals, our GA starts to find the most suitable values of indoor temperature that improve the indoor thermal comfort when compared to the best temperature value. Likewise, the GA tends to improve the indoor temperature when the number of iterations increases. From these results, we can deduce that our GA tends to obtain good configuration of parameters when the population size and the number of iterations are equal to 20 and 25, respectively.

4.4. Performance Evaluation of Our Remote IoT Network

Figure 7 depicts the performance evaluation of the IoT-based network in terms of messages overhead (number of packets), delay overhead, and power consumption of sensor nodes. After one hour of simulation, we obtained different results for every case (four cases).

Figure 7A shows that the number of sent packets increases over the simulation time. As every sensor has its own data to send, it is normal that the more sensors in the simulation, the more packets we will have.

Figure 7B shows the average total power consumed by the sensors. We can observe that the power consumed by different sensors is high at the beginning of the simulation due to the initialization process that happened in the beginning and, after a while, it starts to be stabilize as sensor nodes perform the same operations periodically (channel listening, data sensing, and sending). Thus, it is clear that the consumed power with the "5 sensors" case is less than the "20 sensors" case. This is due to the fact that more sensor nodes implicate more radio traffic that shall be intercepted by the other sensors, which means more power consumption.

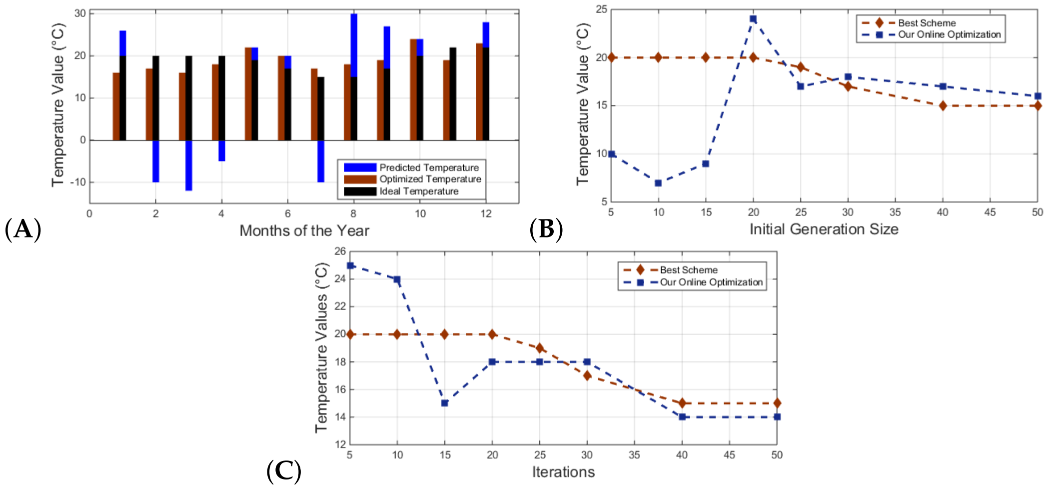

Figure 6. Performance evaluation of our GA. (**A**) Comparison between ideal, predicted, and optimized temperature value. (**B**) Optimized temperature value when varying initial population size. (**C**) Optimized temperature value when varying the number of iterations.

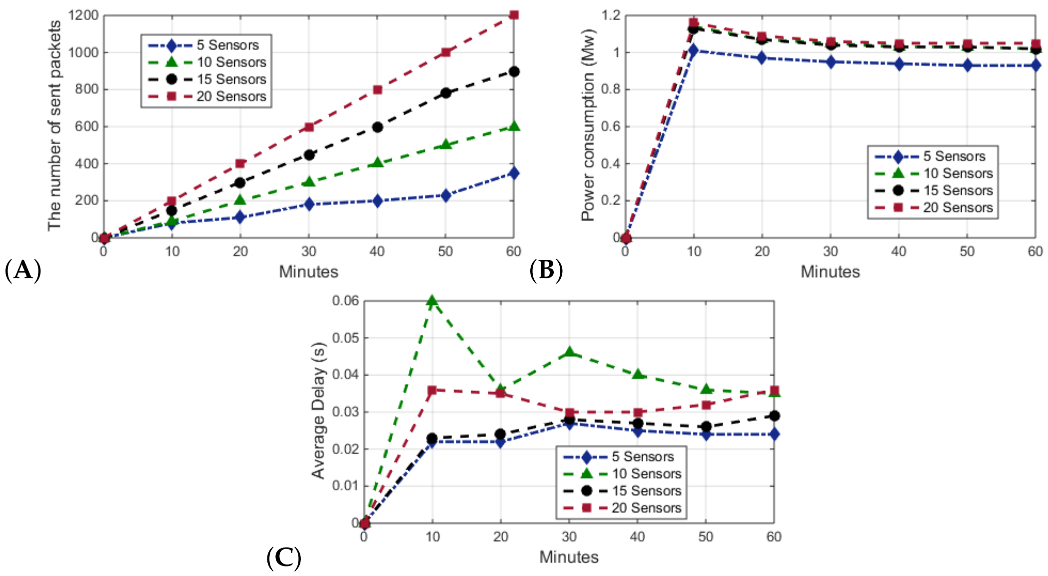

Figure 7. Performance evaluation of our IoT network. (**A**) Number of sent packets (messages overhead). (**B**) Power consumption of sensor nodes. (**C**) Sending delay (delay overhead).

Figure 7C shows the average delay to send a data packet from a sensor node to a receiver node, propagation delay, in our network. It is clear that the more sensors we have, the more the delay will increase due to the increased number of exchanged packets. As we mentioned before, this is mainly due to the Euclidean distance separating sensor nodes from their receiver nodes, which also confirms the previous statement in which we claimed that the power consumption may also be affected due to the distance difference.

5. Conclusions

In this paper, we propose an IoT-based architecture supporting the prediction and control of indoor thermal comfort parameters, remotely. Our architecture enables sensed data to be sent to remote Cloud servers for further storage and analysis in order to measure and predict the indoor thermal comfort parameters. Thus, our architecture is built based on both an efficient, machine learning algorithm to predict and asses the indoor parameters values, and a genetic algorithm to optimize the indoor thermal comfort parameter values in real-time when observing an indoor thermal discomfort. Hence, our architecture enables not only the prediction of the indoor thermal comfort but also the giving of feedback to end-users in real time, in order to improve the indoor thermal comfort.

Experiment results showed that our schemes generate a good and acceptable performance in terms of achieved prediction accuracy and network QoS (message and delay overhead, data sending rate, etc.), and may optimize and control indoor thermal comfort parameters in real-time. On one hand, our random forest-based scheme improves the prediction accuracy of the indoor thermal parameters by 96%. One the other hand, our genetic-based scheme succeeds in determining the suitable parameter values of the indoor thermal comfort with an accuracy of 85%.

As a future work, we are planning to study the energy consumption of the proposed system, as compared to the traditional thermal stat control systems.

Author Contributions: Conceptualization, methodology, and writing B.B.; supervision and funding acquisition, M.E. and L.M.-B.; validation and analysis, A.H. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Environmental Protection Agency. Available online: www.epa.gov (accessed on 21 November 2018).
2. Telegraph. 2015. Available online: http://www.telegraph.co.uk/news/weather/11382808/Winter-death-toll-to-exceed-40000.html (accessed on 20 November 2018).
3. Brik, B.; Esseghir, M.; Merghem-Boulahia, L.; Snoussi, H. Indoor Thermal Comfort Collection of People with Physical Disabilities. In Proceedings of the 2018 International Symposium on Networks, Computers and Communications (ISNCC), Rome, Italy, 19–21 June 2018; pp. 1–6. [CrossRef]
4. Brik, B.; Esseghir, M.; Merghem-Boulahia, L.; Snoussi, H. ThermCont: A machine Learning enabled Thermal Comfort Control Tool in a real time. In Proceedings of the 2019 15th International Wireless Communications Mobile Computing Conference (IWCMC), Tangier, Morocco, 24–28 June 2019; pp. 294–300. [CrossRef]
5. Brik, B.; Esseghir, M.; Merghem-Boulahia, L.; Snoussi, H. An IoT-based deep learning approach to analyse indoor thermal comfort of disabled people. *Build. Environ.* **2021**, *203*, 108056. [CrossRef]
6. Sisinni, E.; Saifullah, A.; Han, S.; Jennehag, U.; Gidlund, M. Industrial Internet of Things: Challenges, Opportunities, and Directions. *IEEE Trans. Ind. Inform.* **2018**, *14*, 4724–4734. [CrossRef]
7. Brik, B.; Esseghir, M.; Merghem-Boulahia, L.; Snoussi, H. ThingsGame: When sending data rate depends on the data usefulness in IoT networks. In Proceedings of the 2018 14th International Wireless Communications Mobile Computing Conference (IWCMC), Limassol, Cyprus, 25–29 June 2018; pp. 886–891. [CrossRef]
8. Fanger, P. *Thermal Comfort: Analysis and Applications in Environmental Engineering*; Danish Kr. 76, 50; Danish Technical Press: Copenhagen, Denmark, 1970.
9. Chang, Y.; Lin, Y. PMV-Based Genetic Algorithms for Indoor Temperature Control System. In Proceedings of the 2016 International Symposium on Computer, Consumer and Control (IS3C), Xi'an, China, 4–6 July 2016; pp. 295–298.
10. Farhan, A.A.; Pattipati, K.; Wang, B.; Luh, P. Predicting individual thermal comfort using machine learning algorithms. In Proceedings of the 2015 IEEE International Conference on Automation Science and Engineering (CASE), Gothenburg, Sweden, 24–28 August 2015; pp. 708–713.
11. Tamani, N.; Brik, B.; Lagraa, N.; Ghamri-Doudane, Y. Vehicular Cloud Service Provider Selection: A Flexible Approach. In Proceedings of the GLOBECOM 2017—2017 IEEE Global Communications Conference, Singapore, 4–8 December 2017; pp. 1–6. [CrossRef]
12. Peterek, T.; Dohnálek, P.; Gajdoš, P.; Šmondrk, M. Performance evaluation of Random Forest regression model in tracking Parkinson's disease progress. In Proceedings of the 13th International Conference on Hybrid Intelligent Systems (HIS 2013), Gammarth, Tunisia, 4–6 December 2013; pp. 83–87. [CrossRef]

13. Langevin, J.; Gurian, P.L.; Wen, J. Tracking the human-building interaction: A longitudinal field study of occupant behavior in air-conditioned offices. *J. Environ. Psychol.* **2015**, *42*, 94–115. [CrossRef]
14. Qavidel Fard, Z.; Sadat Zomorodian, Z.; Sadat Korsavi, S. Application of Machine Learning in Thermal Comfort Studies: A Review of Methods, Performance and Challenges. *Energy Build.* **2021**, *256*, 111771. [CrossRef]
15. Yu, K.H.; Chen, Y.A.; Jaimes, E.; Wu, W.C.; Liao, K.K.; Liao, J.C.; Lu, K.C.; Sheu, W.J.; Wang, C.C. Optimization of thermal comfort, indoor quality, and energy-saving in campus classroom through deep Q learning. *Case Stud. Therm. Eng.* **2021**, *24*, 100842. [CrossRef]
16. Gao, N.; Shao, W.; Rahaman, M.S.; Zhai, J.; David, K.; Salim, F.D. Transfer learning for thermal comfort prediction in multiple cities. *Build. Environ.* **2021**, *195*, 107725. [CrossRef]
17. IEE Colloquium on 'Applications of Genetic Algorithms' (Digest No.1994/067). In Proceedings of the IEE Colloquium on Applications of Genetic Algorithms, London, UK, 15 March 1994.
18. Brik, B.; Khan, J.A.; Ghamri-Doudane, Y.; Lagraa, N.; Lakas, A. GSS-VC: A game-theoretic approach for service selection in vehicular cloud. In Proceedings of the 2018 15th IEEE Annual Consumer Communications Networking Conference (CCNC), Las Vegas, NV, USA, 12–15 January 2018; pp. 1–6. [CrossRef]
19. Viani, F.; Polo, A. A forecasting strategy based on wireless sensing for thermal comfort optimization in smart buildings. *Microw. Opt. Technol. Lett.* **2017**, *59*, 2913–2917. [CrossRef]
20. Nian, F.; Wang, K. Study on indoor environmental comfort based on improved PMV index. In Proceedings of the 2017 3rd International Conference on Computational Intelligence Communication Technology (CICT), Ghaziabad, India, 9–10 February 2017; pp. 1–4.
21. Verheyen, J.; Theys, N.; Allonsius, L.; Descamps, F. Thermal comfort of patients: Objective and subjective measurements in patient rooms of a Belgian healthcare facility. *Build. Environ.* **2011**, *46*, 1195–1204. [CrossRef]
22. Lin, Z.; Deng, S. A study on the thermal comfort in sleeping environments in the subtropics—Developing a thermal comfort model for sleeping environments. *Build. Environ.* **2008**, *43*, 70–81. [CrossRef]
23. International Organization for Standardization. *ISO 7730: Moderate Thermal Environments—Determination of the PMV and PPD Indices and Specification of the Conditions for Thermal Comfort*; ISO: Geneva, Switzerland, 1994.
24. Han, H.; Lee, J.; Kim, J.; Jang, C.; Jeong, H. Thermal Comfort Control Based on a Simplified Predicted Mean Vote index. *Energy Procedia* **2014**, *61*, 970–974. [CrossRef]
25. Khodakarami, J.; Nasrollahi, N. Thermal comfort in hospitals—A literature review. *Renew. Sustain. Energy Rev.* **2012**, *16*, 4071–4077. [CrossRef]
26. Javed, M.; Li, N.; Li, S. Personalized thermal comfort modeling based on Support Vector Classification. In Proceedings of the 2017 36th Chinese Control Conference (CCC), Dalian, China, 26–28 July 2017; pp. 10446–10451.
27. Manna, C.; Wilson, N.; Brown, K.N. Personalized Thermal Comfort Forecasting for Smart Buildings via Locally Weighted Regression with Adaptive Bandwidth. In Proceedings of the 2nd International Conference on Smart Grids and Green IT Systems, Aachen, Germany, 9–10 May 2013; SMARTGREENS: Rehovot, Israel, 2013.
28. Brik, B.; Lagraa, N.; Ghamri-Doudane, Y.; Lakas, A. Finding the most adequate public bus in Vehicular Clouds. In Proceedings of the 2016 International Conference on Wireless Networks and Mobile Communications (WINCOM), Fez, Morocco, 26–29 October 2016; pp. 67–74. [CrossRef]
29. Azzaoui, N.; Korichi, A.; Brik, B.; Fekair, M.e.A. Towards Optimal Dissemination of Emergency Messages in Internet of Vehicles: A Dynamic Clustering-Based Approach. *Electronics* **2021**, *10*, 979. [CrossRef]
30. Brik, B.; Lagraa, N.; Lakas, A.; Cherroun, H.; Cheddad, A. ECDGP: Extended Cluster-Based Data Gathering Protocol for Vehicular Networks. *Wirel. Commun. Mob. Comput.* **2016**, *16*, 1238–1255. [CrossRef]
31. Osterlind, F.; Dunkels, A.; Eriksson, J.; Finne, N.; Voigt, T. Cross-Level Sensor Network Simulation with COOJA. In Proceedings of the 2006 31st IEEE Conference on Local Computer Networks, Tampa, FL, USA, 14–16 November 2006; pp. 641–648.