JOURNAL OF NANO- AND ELECTRONIC PHYSICS — *ЖУРНАЛ НАНО- ТА ЕЛЕКТРОННОЇ ФІЗИКИ*
Vol. **17** No 3, 03038(6pp) (2025) — Том **17** № 3, 03038(6cc) (2025)

OPEN ACCESS

REGULAR ARTICLE

Novel Hybrid Approaches for Occupancy Prediction Using Temperature, Light and CO_2 Level Supporting Electrical Energy Management

M. Ennejjar[1,*], M. Ezzini[2], M.A. Jallal[1,3], S. Chabaa[1,4], A. Zeroual[1]

[1] *I2SP Research Team, Physics Department, Faculty of Sciences Semlalia, Cadi Ayyad University, Morocco*
[2] *Fluid Mechanics and Energetic Laboratory, Physics Department, Faculty of Sciences Semlalia, Cadi Ayyad University, Marrakesh, Morocco*
[3] *Univ. Grenoble Alpes, CEA, Liten, Campus Ines, 73375, Le Bourget-du-Lac, France*
[4] *Industrial Engineering Department, National School of Applied Sciences, Ibn Zohr University, Morocco*

(Received 23 March 2025; revised manuscript received 24 June 2025; published online 27 June 2025)

This paper introduces novel hybrid approaches for predicting office space occupancy by combining conventional models with artificial neural networks. Specifically, we propose two hybrid models: the Naive Bayes Classifier integrated with a Multi-Layer Perceptron (NBC-MLP) and a Logistic Mixed-Output Perceptron (LMOP). These models use environmental factors such as temperature, light, and CO_2 levels to predict occupancy. The hybrid models are designed to leverage the strengths of both conventional models and neural networks, enhancing predictive accuracy while maintaining simplicity. The Naive Bayes Classifier, known for its simplicity with categorical data, complements the Multi-Layer Perceptron's ability to capture complex relationships in data. The results show that the proposed hybrid models significantly outperform conventional models, with the LMOP model achieving an accuracy of 99.28 %. This indicates the hybrid models' effectiveness in modeling complex occupancy patterns. Moreover, the models are robust against noisy data and fluctuations in environmental conditions, making them suitable for real-world applications. These models also have practical applications for optimizing space utilization and improving energy efficiency. By predicting occupancy more accurately, they enable better control of HVAC systems and lighting, reducing energy consumption.

Keywords: Occupancy, Energy management, Multi-Layer Perceptron, Hybrid approach, Prediction.

DOI: 10.21272/jnep.17(3).03038 PACS numbers: 07.05.Kf, 88.10.gc

1. INTRODUCTION

Over the past decade, building occupancy modeling has garnered increasing attention due to the observation that more than 50 % of energy in commercial buildings is wasted during unoccupied hours [1]. Occupancy status plays a fundamental role in determining internal loads and organizing the operational schedules of equipment and devices, which is critical for managing and operating energy systems. For instance, in residential buildings, occupancy profiles can be used in building performance simulations to improve the energy efficiency of heating, ventilation, and air conditioning (HVAC) systems. These improvements can be achieved through energy recovery devices, selecting more efficient HVAC equipment, and applying intelligent control systems based on occupancy, significantly reducing energy waste [2, 3].

Previous studies have demonstrated the effectiveness of machine learning models in the field of building occupancy prediction. Techniques like statistical and stochastic algorithms, including multiple linear regression [4], support vector machine (SVM) [5], logistic regression [6] (for binary output) and random forest (RF) [7], have revolutionized this domain. In 2016, Candanedo et al. [4] proposed a model for occupancy detection using measurements of light, humidity, CO_2, and temperature, applying classification and regression trees (CART), RF, and LDA. Similarly, Zhou et al. (2021) applied RF to predict office window opening behaviors with an accuracy of up to 80 %, outperforming SVM and extreme gradient boosting (XGBoost) algorithms [8]. Table 1, presents a summary of some previous study.

Our study addresses the gap in the literature regarding the use of basic models that may not provide the best accuracy for occupancy prediction and the exploration of more complex models to improve prediction performance. The key contributions of this study are:

- Development of two novel hybrid approaches: NBC-MLP and LMOP, which combine basic machine learning models with MLP to improve prediction accuracy.
- Combination of simple models and complex architectures: The hybridization of Naive Bayes Classifier (NBC) and Logistic Regression (LR) with MLP neural networks allows us to leverage the strengths of simpler models (speed and effectiveness with small datasets) while benefiting from the optimization power of MLP.
- Enhanced prediction precision: The integration of these hybrid models offers a promising solution to im-

* m.ennejjar.ced@uca.ac.ma

prove occupancy prediction accuracy, which can ultimately lead to more effective energy management and reduced energy waste in buildings.

In summary, we focus on improving the accuracy of MLP-based models for occupancy prediction by introducing novel hybrid approaches that combine both simple and complex machine learning models. This approach holds the potential to significantly improve the efficiency of building energy system

The structure of the presented study is organized as follows: the second section introduces the adopted methodology, which includes a detailed explanation of the proposed hybrid models. The third section then presents and discusses the obtained results. Subsequently, the fourth section provides a comparative analysis, and the final section concludes with a summary of the study.

Table 1 – Summary of some research on occupancy prediction

Ref	Methods	Architecture	Inputs	Accuracy
[4]	RF	Office room	Light, CO_2	97.41
	LDA		CO_2, Temperature	87.62
[9]	SVM	household	Electric power consumption (W)	90
	KNN			88
	HMM			87
[10]	HMM	high-density library building	Occupancy data for 12 months	–
[11]	LDA, SVM, RNN	Residence	Indoor and outdoor environmental measurement	85 % and 83 %
[12]	TS-ANN	Office	Occupancy and time	97.4 %
[13]	CatBoost	Laboratory	Air quality index, CO_2 concentration, Temperature and humidity	92.9 %

2. METHODOLOGICAL APPROACH

This study aims to enhance prediction accuracy and minimize input variables by hybridizing linear, probabilistic models with neural networks. This is key for building occupancy prediction, balancing accuracy, computational efficiency, and simplicity. Two hybrid models are proposed: the first combines Logistic Regression (LR) with a Multilayer Perceptron (MLP), and the second merges the Naive Bayes Classifier (NBC) with an MLP. LR provides interpretability, while NBC is effective with noisy or limited data. Fig. 1 illustrates the methodology.

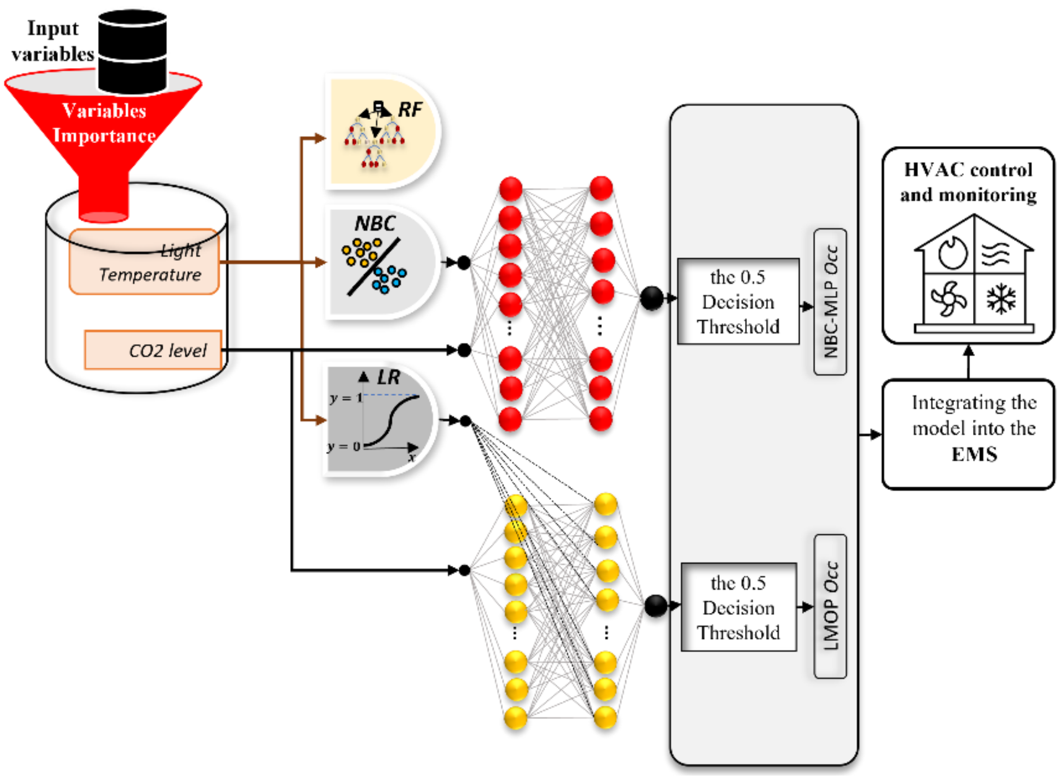


Fig. 1 – The proposed methodology

Both hybrid models were trained to predict building occupancy using inputs like temperature, lighting, and CO_2 concentration, key factors influencing occupancy. By combining neural networks' predictive power with the simplicity of linear and probabilistic models, the study aims to enhance accuracy while keeping the model efficient.

The approach ensures adaptability and scalability, with Logistic Regression and Naive Bayes providing quick, robust components, and the MLP network capturing complex patterns.

The study used three datasets [4] of an office room for training and testing the classification models, summarized in Table 2. These datasets include temperature, humidity, light, CO_2 levels, occupancy status, and timestamps. Variable selection was based on correlation analysis (Fig. 2).

Table 2 – Data set description

Data Set	Number of Observations	Class Distribution (%)
Training	8143	0 (79 %) – 1 (21 %)
Testing 1	9752	0 (79 %) – 1 (21 %)
Testing 2	2665	0 (64 %) – 1 (36 %)

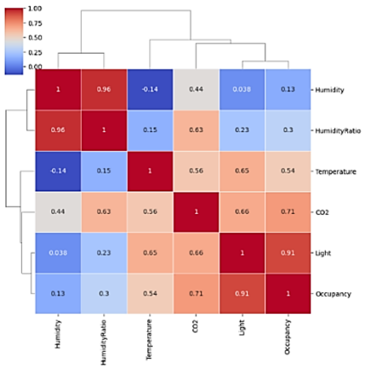


Fig. 2 – Correlation matrix between variables

2.1 Classic Classification Models

Naive Bayes Classifier (NBC): NBC is a Boolean classification method based on Bayes' law. In our study the model tries to predict whether the space is occupied (1) or is not occupied (0). So, the principle is to choose the most probable value of the output (Y) based on the vector of inputs $X = [v_1, v_2, \ldots, v_n]$ as expressed in equation (1) [14, 15].

$$Y \leftarrow \arg\max_{y_k} \frac{P(Y=y_k)\prod_i P(X_i|Y=y_k)}{\sum_j P(Y=y_j)\prod_i P(X_i|Y=y_j)} \qquad (1)$$

which simplifies to the following equation (because the denominator does not depend on y_k).

$$Y \leftarrow \arg\max_{y_k} P(Y=y_k)\prod_i P(X_i|Y=y_k) \qquad (2)$$

Logistic Regression (LR): Logistic regression is a statistical method used in binary classification problems, the output of which can be 1 or 0. The output is transformed into probability using sigmoid function which maps any real-valued number into the (0, 1) interval, and it is given by [16]:

$$P = \frac{1}{1+e^{-(\alpha_0+\alpha_1 x_1+\alpha_2 x_2+\cdots+\alpha_n x_n)}} \qquad (3)$$

Where P is the probability of the output, α_0 is the intercept, and $\alpha_1, \alpha_2, \ldots, \alpha_n$ are the coefficients of the inputs $x_1, x_2, \ldots, x_n$.

Random Forests (RF): Random Forest (RF) is one of machine learning algorithm which combines a plenty of decision tree models to make regression or classification [17].

2.2 Proposed NBC-MLP Hybrid Occupancy Predictor

This study introduces a hybrid approach for predicting office space occupancy, combining the Naive Bayes Classifier (NBC) with a Multi-Layer Perceptron (MLP) neural network. The model uses indoor temperature (T) and light intensity as inputs for the NBC, which generates a probabilistic occupancy prediction. This output is then fed into the MLP along with CO_2 levels, refining the prediction into a continuous value. A thresholding technique is applied to convert this output into a binary occupancy status ($y = 1$ for occupied, $y = 0$ for unoccupied). Fig. 3 illustrates the model structure.

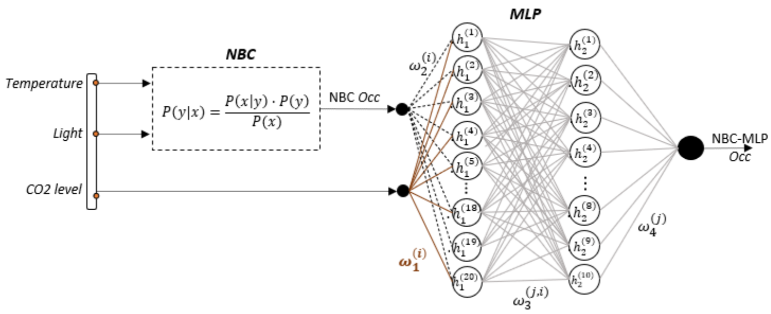


Fig. 3 – Structure of the proposed NBC-MLP hybrid model, which consists of two hidden layers {20, 10} with ReLU activation and a sigmoid activation function in the output layer. The CO_2 level and the output of Naive Bayes Classifier (NBC) model serve as the inputs of MLP

Equation (6) formalizes the relationship governing the overall model's output, integrating the probabilistic prediction from the NBC with the continuous refinement from the MLP. This hybrid approach aims to enhance prediction accuracy by leveraging the complementary strengths of probabilistic reasoning and neural network-based learning.

$$NBC - MLP \text{ Occ} = \frac{1}{1+e^{\left[-\left[\sum_{j=1}^{10} \omega_4^{(j)} . h_2^{(j)} + b_3\right]\right]}} \tag{4}$$

$$NBC - MLP \text{ Occ} = \frac{1}{1+e^{\left[-\left[\sum_{j=1}^{10} \omega_4^{(j)} . \max\left(0, \sum_{i=1}^{20} \omega_3^{(j,i)} . h_1^{(i)} + b_2^{(j)}\right) + b_3\right]\right]}} \tag{5}$$

$$NBC - MLP \text{ Occ} = \frac{1}{1+e^{\left[-\left[\sum_{j=1}^{10} \omega_4^{(j)} . \max\left(0, \sum_{i=1}^{20} \omega_3^{(j,i)} . \max(0, \omega_1^{(i)} . CO_2 + \omega_2^{(i)} . \widehat{NBC\, Occ} + b_1^{(i)}) + b_2^{(j)}\right) + b_3\right]\right]}}, \tag{6}$$

where:

- $\omega_1^{(i)}$, $\omega_2^{(i)}$, $\omega_3^{(j,i)}$, *and* $\omega_4^{(j)}$ are the weight associated respectively with CO_2, the output of NBC for each neuron *i* of the first hidden layer of the MLP, the weight between the neuron *i* of the first layer and the neuron *j* of the second hidden layer, and the weight between the neuron *j* of the second layer and the final output.
- $b_1^{(i)}, b_2^{(j)}$ *and* b_3 Are respectively: the bias associated with each neuron *i* of the first layer, the bias associated with each neuron *j* of the second layer, and the bias of the final output of the MLP.

2.3 Proposed LMOP Hybrid Occupancy Predictor

This section presents a hybrid approach for predicting office occupancy, combining Logistic Regression (LR) with a Multi-Layer Perceptron (MLP) neural network. The process begins with using indoor temperature (*T*) and light intensity as inputs for the LR model, which generates an initial occupancy prediction. This output is then fed into the second hidden layer of the MLP, alongside CO_2 levels as the input for the first hidden layer. The MLP refines the prediction by integrating the CO_2 level. The final continuous output from the MLP is converted into a binary classification: occupied ($y = 1$) if the output exceeds 0.5, and unoccupied ($y = 0$) otherwise. Fig. 4 illustrates the model structure, and Equation (7) defines the output relationship between the LR and MLP models..

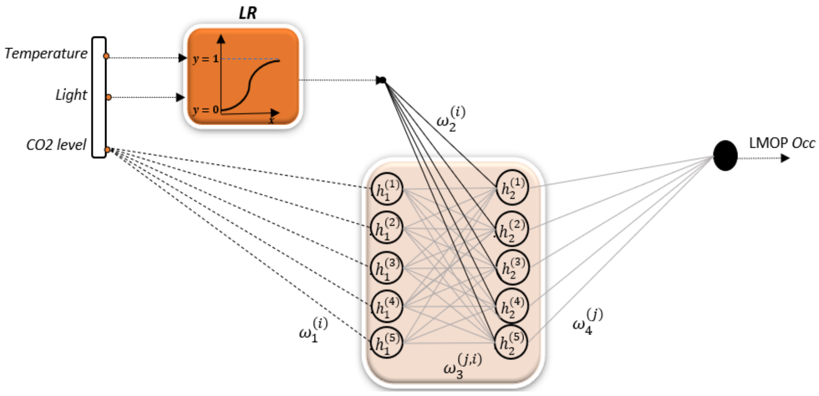


Fig. 4 – Structure of the proposed LMOP hybrid model, which consists of two hidden layers {5, 5} with ReLU activation and a sigmoid activation function in the output layer. The CO_2 level serves as the input to the first layer, while the output of the linear regression (LR) model is linked as an additional input to the second layer

$$\begin{cases} \boldsymbol{Occ = 1 \ ; \ \ if \ \ \ LMOP\_ \text{Occ}} = \dfrac{1}{1+e^{\left[-\left[\sum_{j=1}^{5} \omega_4^{(j)} . \max\left(0, \ \omega_2^{(i)} . \widehat{LR\_Occ} + \sum_{i=1}^{5} \omega_3^{(j,i)} . \max(0, \omega_1^{(i)} . CO_2 + b_1^{(i)}) + b_2^{(j)}\right) + b_3\right]\right]}} > \boldsymbol{0.5} \\ \boldsymbol{Occ = 0 \ ; \ \ if\ else} \end{cases} \tag{7}$$

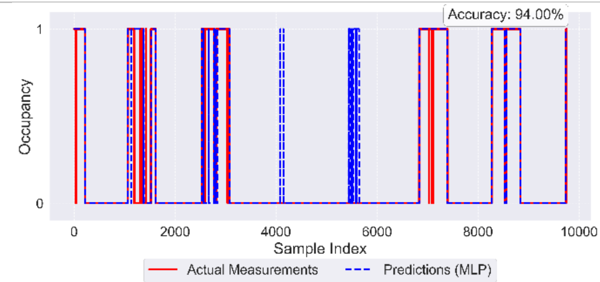
3. RESULTS AND DISCUSSION

This section provides an analysis of the results from the proposed approaches, comparing them with traditional models. Table 3 shows a detailed comparison of accuracy across all models: NBC, RF, LR, MLP, and the hybrid models, NBC-MLP and LMOP. The hybrid models, NBC-MLP and LMOP, significantly outperform the others, with the NBC-MLP model achieving 98.91 % accuracy and the LMOP model reaching 99.28 %. These results highlight the effectiveness of hybrid models in enhancing predictive accuracy, demonstrating their potential to outperform standard MLP models and emphasizing the value of hybrid approaches for optimizing accuracy.

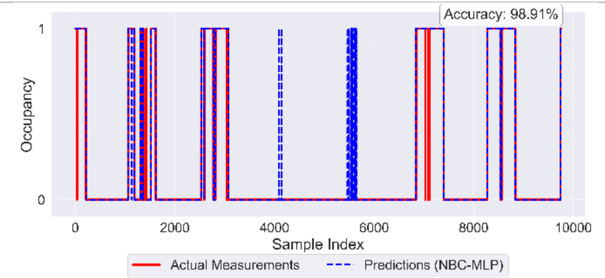
Table 2 – Models' accuracies across different scenarios in the test and training phases

		Accuracy (%)		
Models	**Inputs**	**Training**	**Test 1**	**Test 2**
NBC	*T*, light, CO_2 level	97.78	98.74	97.74
	light, CO_2 level	98.35	99.01	97.71
	CO_2 level	90.22	78.64	87.27
RF	*T*, light, CO_2 level	100	96.24	94.74
	light, CO_2 level	100	95.81	94.33
	CO_2 level	97.53	63.80	78.94

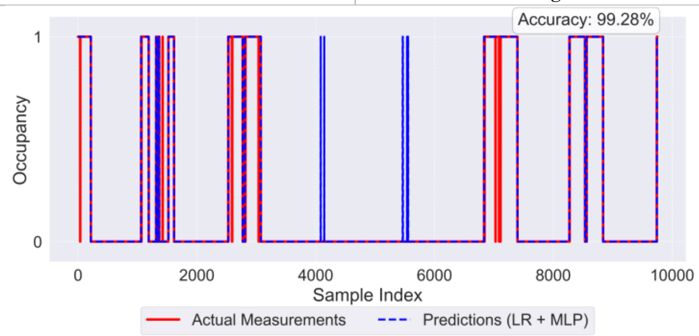
LR	*T*, light, CO_2 level	98.60	99.16	97.86
	light, CO_2 level	98.83	99.31	97.82
	CO_2 level	90.18	78.67	87.24
MLP	*T*, light, CO_2 level	99.23	91.29	93.36
	light, CO_2 level	98.88	94.00	97.26
	CO_2 level	91.87	60.60	86.19
NBC-MLP	*T*, light, CO_2 level	98.82	98.72	97.64
	light, CO_2 level	98.85	98.91	97.75
LMOP	*T*, light, CO_2 level	98.80	99.28	97.86
	light, CO_2 level	98.48	99.22	97.86



(a) Predicted occupancy vs. Measured occupancy using a standard MLP after applying the 0.5 decision threshold



(b) Predicted occupancy vs. Measured occupancy using the proposed NBC-MLP hybrid occupancy predictor after applying the 0.5 decision threshold



(c) Predicted occupancy vs. Measured occupancy using the proposed LMOP hybrid occupancy predictor after applying the 0.5 decision threshold

Fig. 5 – Predicted occupancy vs. Measured occupancy during Test 1 phase for the standard MLP, Proposed NBC-MLP, and LMOP hybrid occupancy predictors after applying the 0.5 decision threshold across optimal scenarios

Fig. 5 further compares reconstructed occupancy profiles with actual profiles before and after applying the 0.5 decision threshold for the MLP, NBC-MLP, and LMOP models. The MLP model struggled to capture occupancy variability, while the hybrid models accurately reconstructed the profiles.

4. COMPARISON ANALYSIS

In this section, we compare the performance of our hybrid approaches with existing methods in the literature, using the same dataset for occupancy prediction in office spaces. The study [4] tested several classification methods, including Random Forest (RF), Gradient Boosting Machine (GBM), and Classification and Regression Trees (CART), achieving accuracies of 98.06 %, 96.10 %, and 96.52 %, respectively, using variables such as temperature, humidity, light and CO_2 level. Another study [14] used the Naive Bayes Classifier (NBC) with similar variables, achieving 97.7 % accuracy. In comparison, our hybrid models, NBC-MLP and LMOP, demonstrated significant improvements. The LMOP model, in particular, achieved 99.28 % accuracy, outperforming all other models tested.

5. CONCLUSION

In this work, we proposed two novel hybrid approaches – NBC-MLP and LMOP – for predicting occupancy in office spaces by combining conventional classification models with neural networks. By integrating environmental variables such as temperature, light, and CO_2 levels, these models significantly outperformed conventional methods like Naive Bayes Classifier, Random Forest, Gradient Boosting Machine, and Classification and Regression Trees. Notably, the LMOP model achieved a remarkable accuracy of 99.28 %, setting a new performance benchmark. Comparisons with existing literature further highlighted the superiority of our approaches, demonstrating enhanced accuracy, robustness, and reliability. Overall, this study makes a significant contribution to indoor occupancy modeling, offering promising solutions for smart building management, energy optimization, and occupant comfort improvement.

REFERENCES

1. O.T. Masoso, L.J. Grobler, *Energy Build.* **42** No 2, 173 (2010).
2. A.I. Dounis, C, Caraiscos, *Renew. Sustain. Energy Rev.* **13**, 1246 (2009).
3. M. Ennejjar, M. Ezzini, N. El Assri, M.A. Jallal, S. Chabaa, A. Zeroual, *Phys. Scr.* **100**, 055227 (2025).
4. L.M. Candanedo, V. Feldheim, *Energy Build.* **112**, 28 (2016).
5. W. Wang, J. Chen, T. Hong, *Autom. ConStruct.* **94**, 233 (2018).
6. N. Li, J. Li, R. Fan, H. Jia, *Renew. Energy* **73**, 84 (2015).
7. Z. Wang, T. Hong, M.A. Piette, M. Pritoni, *Build Environ.* **158**, 281 (2019).
8. X. Zhou, J. Ren, J. An, D. Yan, X. Shi, X. Jin, *J. Build. Eng.* **42**, 102514 (2021).
9. W. Kleiminger, C. Beckel, T. Staake, S. Santini, *Proceedings of the 5th ACM Workshop on Embedded Systems For Energy-Efficient Buildings* (ACM, Rome, Italy: 2013).
10. B. Alfalah, M. Shahrestani, L. Shao, *J. Build. Eng.* **73**, 106795 (2023).
11. E.D.R. De La Roy, T. Recht, A. Zemmari, P. Bourreau, L. Mora, *Build. Environ.* **231**, 110019 (2023).
12. Y. Jin, D. Yan, X. Zhang, J. An, M. Han, *Build. Simulat.* **14**, 219 (2021).
13. J. Dutta, S. Roy, *Appl. Soft Comput.* **119**, 108536 (2022).
14. M. Ennejjar, S. Chabaa, M.A. Jallal, A. Zeroual, *ITM Web of Conferences* **69**, 01006 (2024)
15. T. Bayes, *Naive Bayes Classifier. Article Sources and Contributors, 1-9. Logistic Regression* (1968).
16. M. Ennejjar, M. Ezzini, S. Chabaa, A. Zeroual, *2024 International Conference on Decision Aid Sciences and Applications (DASA)* (2024).

Нові гібридні підходи для прогнозування заповненості з використанням температури, освітлення та рівня CO_2 для підтримки управління електроенергією

M. Ennejjar[1], M. Ezzini[2], M.A. Jallal[1,3], S. Chabaa[1,4], A. Zeroual[1]

[1] *I2SP Research Team, Physics Department, Faculty of Sciences Semlalia, Cadi Ayyad University, Morocco*
[2] *Fluid Mechanics and Energetic Laboratory, Physics Department, Faculty of Sciences Semlalia, Cadi Ayyad University, Marrakesh, Morocco*
[3] *Univ. Grenoble Alpes, CEA, Liten, Campus Ines, 73375, Le Bourget-du-Lac, France*
[4] *Industrial Engineering Department, National School of Applied Sciences, Ibn Zohr University, Morocco*

У статті представлені нові гібридні підходи до прогнозування заповнюваності офісних приміщень шляхом поєднання традиційних моделей зі штучними нейронними мережами. Зокрема, запропоновано дві гібридні моделі: наївний баєсівський класифікатор, інтегрований з багатошаровим перцептроном (NBC-MLP), та логістичний перцептрон зі змішаним виходом (LMOP). Ці моделі використовують фактори навколишнього середовища, такі як температура, освітлення та рівень CO_2, для прогнозування заповнюваності, що відноситься до задач прикладної фізики. Гібридні моделі розроблені для використання сильних сторін як традиційних моделей, так і нейронних мереж, підвищуючи точність прогнозування, зберігаючи при цьому простоту. Баєсівський класифікатор, відомий своєю простотою роботи з категоріальними даними, доповнює здатність багатошарового перцептрона фіксувати складні взаємозв'язки в даних. Аналогічно, логістичний перцептрон зі змішаним виходом інтегрує логістичну регресію з нейронними мережами для покращення можливостей прогнозування. Результати показують, що запропоновані гібридні моделі значно перевершують традиційні моделі, причому модель LMOP досягає точності 99,28%. Це свідчить про ефективність гібридних моделей у моделюванні складних моделей заповнюваності. Більше того, моделі стійкі до шумних даних та коливань умов навколишнього середовища, що робить їх придатними для реальних застосувань. Завдяки точнішому прогнозуванню заповнюваності, вони дозволяють краще контролювати системи опалення, вентиляції та кондиціонування повітря (HVAC) та освітлення, зменшуючи споживання енергії.

Ключові слова: Заповненість, Керування енергією, Багатошаровий персептрон, Гібридний підхід, Прогнозування.