Applied Mathematics & Information Sciences
An International Journal

http://dx.doi.org/10.18576/amis/190209

A Machine Learning Approach to Microclimate Monitoring and Fault Detection

Nurkamilya Daurenbayeva[1], *Lyazzat Atymtayeva*[2,*], *Almas Nurlanuly*[3], *Artem Bykov*[1], *Bakhytzhan Akhmetov* [4], *Gabit Shuitenov* [5], *Umut Turusbekova* [6]

[1]Department of Computer Engineering, International Information Technology University, A15H7X9 Almaty, Kazakhstan
[2]Department of Information Systems, SDU University, 043801 Kaskelen, Kazakhstan
[3]Department of Aviation Equipment and Technology, Academy of Civil Aviation, A35X2Y6 Almaty, Kazakstan
[4]Department of Informatics and Informatization of Education, Kazakh National Pedagogical University named after Abay, 050010 Almaty, Kazakhstan
[5]Department of Information Systems and Technologies, Esil University, 010005 Astana, Kazakhstan
[6]Department of Artificial Intelligence Technologies, L.N. Gumilyov Eurasian National University, 010008 Astana, Kazakhstan

Received: 1 Oct. 2024, Revised: 22 Nov. 2024, Accepted: 21 Dec. 2024
Published online: 1 Mar. 2025

Abstract: The integration of Machine Learning (ML) into heating, ventilation, and air conditioning (HVAC) systems significantly enhances fault detection and diagnosis (FDD), crucial for improving energy efficiency, as buildings account for around 40% of global energy consumption. However, the presence of inaccurate and noisy data can complicate FDD efforts, as traditional methods often struggle in real-world conditions. This study explores FDD through an experiment conducted in two distinct environments: residential and non-residential buildings. Using a Node MCU microcontroller and over 16 sensors, data on microclimate parameters such as temperature, humidity, and CO2 levels were collected and analyzed in real time. The findings highlighted the variability of microclimate conditions and identified challenges associated with existing FDD methods, including the limitations of Principal Component Analysis (PCA) in noisy environments. Recent literature categorizes ML-based FDD methods into three groups: traditional machine learning, deep learning, and hybrid models, demonstrating their superiority over conventional approaches. However, challenges such as data variability and the need for real-time processing still exist. To develop intelligent fault diagnosis systems, the CRISP-DM methodology is proposed, encompassing phases from business understanding to deployment, while addressing potential noise and inaccuracies. The system architecture includes sensors for monitoring climate and air quality, a microcontroller for data processing, a user interface for real-time notifications, and analysis algorithms for anomaly detection. Overall, this research underscores the potential of ML in optimizing Heating, Ventilation, and Air Conditioning performance and emphasizes the need for adaptive models and IoT integration to enhance data collection efficiency, marking an important step toward sustainable energy practices in building microclimate control.

Keywords: Heating, Ventilation, and Air Conditioning (HVAC), Cross-Industry Standard Process for Data Mining(CRISP-DM), Machine Learning (ML) in HVAC Systems, Fault Detection and Diagnosis (FDD), Energy Efficiency in Buildings, Microclimate Parameters, Data Variability and Noise in FDD, Principal Component Analysis (PCA).

1 Introduction

Microclimate control in buildings has become a major global concern due to rising energy consumption and carbon dioxide emissions. Buildings contribute to over a third of global energy use, with commercial buildings accounting for about 41% of this total. In Kazakhstan, 90% of the energy in the housing and utilities sector is used for building operations, with residential buildings consuming the largest portion — 50-55% [1]. The microclimate, including temperature, humidity, and air movement, plays a crucial role in energy efficiency and comfort within buildings. Modern automated systems primarily regulate temperature, but optimal health conditions require consideration of other factors as well. For example, in winter, it is important to maintain indoor air temperature within $23-24\,^{\circ}$C. to prevent fatigue, and the floor temperature should be slightly lower than the air temperature to avoid colds and improve

* Corresponding author e-mail: lyazzat.atymtayeva@sdu.edu.kz

thermoregulation[4]. Microclimate systems consume significant energy resources, which necessitates the development of technologies to improve their energy efficiency. Enhancing these systems not only boosts comfort but also reduces overall energy consumption, which is especially important in the context of climate change and limited resources[5], [6]. The integration of ML into HVAC systems represents a significant advancement in improving FDD. With buildings accounting for approximately 40% of global energy consumption and HVAC systems responsible for over half of that figure, optimizing their performance is crucial for energy efficiency and sustainability. However, the development of effective FDD methods is complicated by several challenges, including the complexity of HVAC operations, the dynamic nature of buildings, and the variability of faults.

2 Related works

FDD in HVAC systems is essential for improving energy efficiency and system reliability. Traditional methods often struggle with noisy or incomplete data, high-lighting the need for more adaptive solutions. Recent advancements in machine learning (ML) offer promising alternatives, with techniques like Support Vector Machines (SVM), Random Forest, Deep Learning, and Hybrid Models Enhancing fault detection accuracy. Additionally, statistical methods such as PCA, Independent Component Analysis (ICA) and Partial Least Squares (PLS) are commonly used to process complex, high-dimensional data, improving fault detection even in noisy environments. This section explores various fault detection methods, emphasizing the role of ML and statistical analysis in overcoming challenges posed by noisy data and improving system performance. According to research conducted in the field of predictive maintenance[14], the use of time series models and machine learning methods for predicting equipment failures has proven to be highly effective in industry. The integration of Machine Learning into heating, ventilation, and air conditioning systems significantly enhances fault detection and diagnosis, which is crucial for improving energy efficiency in buildings, as buildings account for around 40% of global energy consumption. Effective energy management systems rely on accurate data, which can be influenced by various factors, such as the positioning and sensitivity of temperature regulators. In this regard, the insights from recent studies on smart temperature regulators and Building Energy Management Systems (BEMS) highlight the critical role of sensor placement and sensitivity levels for precise decision-making and control over building systems. These factors are similarly important for Machine Learning-based fault detection methods, where accurate and real-time data collection is essential for successful anomaly detection and system optimization. The study also acknowledges challenges like noise and variability in sensor data, which can complicate real-time decision-making processes, as seen in both traditional and advanced FDD methods[7]. In recent years, various methodologies have been used to diagnose faults in heat pumps (HPs), especially with the development of machine learning methods. A study in the field of Microclimate Systems conducted by Barandier et al. (2024)[18] showed that machine learning methods such as algorithms with a teacher can significantly improve fault diagnosis in HVAC systems by analyzing real-time data obtained from multiple sensors. In particular, the study highlights the importance of accurate and reliable data collection to improve troubleshooting algorithms and overcome problems arising from data noise. The present study continues this concept by applying a link weighting method to assess the importance of each component in heat pumps, with the aim of reducing the number of signs by 50% while maintaining the effectiveness of the system for fault diagnosis. The results show that the K-nearest neighbor (K-NN) algorithm demonstrated the best results, with an accuracy of more than 99%, which corresponds to the high performance indicators found in the above-mentioned study of HVAC systems[14]. Recent advancements in FDD for HVAC systems have shown that Artificial Intelligence (AI), particularly ML and Deep Learning (DL) methods, offer significant advantages in improving system performance and energy efficiency. A comprehensive review of AI-based FDD methods conducted by Jian Bi et al. (2024)[19] highlights the evolution of these techniques, categorizing them into three main types: traditional machine learning, deep learning, and hybrid models. The study emphasizes that, compared to traditional physics-based methods, AI approaches have garnered substantial research interest due to their higher accuracy and reduced dependency on expert knowledge in dynamic environments. However, the review also points out that challenges persist, particularly in addressing issues related to noisy and dynamic data, as well as achieving resolution in complex systems. Integrating Machine Learning into systems like HVAC has shown great promise in improving fault detection and diagnosis, especially in real-world scenarios with noisy and inaccurate data. Since buildings contribute significantly to global energy consumption, using ML to optimize HVAC systems is crucial for boosting energy efficiency. Similarly, ML-driven Facial Expression Recognition (FER) systems, utilizing deep learning models like Convolutional Neural Networks (CNNs), excel at interpreting non-verbal cues in Computer Vision. Both applications involve complex data analysis in imperfect conditions, whether for microclimate data in HVAC systems or images for facial expression recognition.While both domains deal with different types of data (environmental parameters vs. visual data), they share common challenges such as data noise, real-time processing needs, and the necessity of adaptive, efficient models for high accuracy. These studies highlight the growing importance of ML across diverse fields, and how advanced ML models can significantly improve performance in both operational efficiency (HVAC) and Deep Learning for FER[21], [22] In the study by Sam et al. (2011), a fault modeling approach is presented that utilizes statistical machine learning and information theory, along with sensor information flow analysis for fault diagnosis in HVAC systems. In contrast, our research focuses on real-time data collection from over 16 sensors in residential and commercial buildings, with the data being transmitted to Google Sheets for analysis, without the use of fault modeling or sensor information flow analysis. *Challenges with Noisy Data* The data generated by HVAC systems can often be inaccurate or noisy due to sensor malfunctions, extreme environmental conditions, or external interferences. Traditional FDD approaches, which often rely on expert analysis and predefined rules, may be inadequate when faced with these variations, highlighting the need for more adaptable and robust solutions.

ML-Based FDD Methods Recent studies classify ML-based FDD methods into three categories: traditional machine learning methods, deep learning, and hybrid models. Traditional methods, such as Support Vector Machines (SVM) and Random Forest, have shown promising results in fault detection. For example, Hamayat et al. (2023)[11] demonstrated that Extra Tree, Random Forest, and CatBoost classifiers outperformed others in terms of accuracy and fault detection metrics. Deep learning methods utilize neural networks to analyze complex patterns, enhancing HVAC optimization, although widespread adoption is hindered by issues such as data quality and model interpretability. Hybrid models combine both approaches and effectively address changes in operating conditions. Statistical analysis methods, including Principal Component Analysis (PCA), In-dependent Component Analysis (ICA), and Partial Least Squares (PLS), are commonly used for fault detection. PCA identifies the most significant features of multivariate data by focusing on uncorrelated directions with the largest variance. When applied to Multivariate Time Series (MTS), it enhances fault detection efficiency. For example, Li et al.[16] combined PCA, ICA, PLS, and the Kalman filter to improve fault detection accuracy by projecting data into a subspace along the fault area. In ICA, data is treated as linear combinations of independent components, while PLS, which combines PCA and canonical correlation analysis, has limitations in handling nonlinear MTS and imbalanced data. Various fault detection techniques exist, such as those used by Xiangjun et al. [17], who applied methods like information fusion, AI, neural networks, fuzzy logic, and genetic algorithms. They highlight that different types of fault data require different processing methods. Combining multiple data sources helps reduce noise, overcome the limitations of individual protection systems, and improve fault detection accuracy and reliability. In addition to prefiltering and principal component analysis, other methods combine different data types to enhance anomaly detection. For instance, [2] proposes algorithms that jointly process geoelectric and seismoacoustic signals, enabling real-time monitoring of geotechnical changes despite external noise.

3 Materials and Methods

The CRISP-DM methodology offers a systematic framework for developing intelligent fault diagnosis systems in building climate control. The process begins with the stage of defining business objectives, which sets the goals and limitations of the project, as well as the key problems of microclimatic control, taking into account the interests of stakeholders. Further, in the data understanding stage, information from various sources such as sensor data and maintenance logs is collected and analyzed, with particular attention to data quality, especially in noisy environments. Data preparation involves cleaning, filtering, and normalization to eliminate errors before building models. At the modeling stage, machine learning algorithms are selected and tested to create predictive models of fault diagnosis taking into account noise to increase reliability. The effectiveness of the model is evaluated by metrics such as accuracy and completeness, which allows you to check its performance in real-world use, taking into account the variability of data and the presence of noise. Finally, during the deployment phase, the system is integrated into the work environment with the provided monitoring and maintenance to ensure stable operation. The use of CRISP-DM allows organizations to effectively implement systems that can cope with data variability and noise, ensuring reliable diagnostics[14], [16].

Description of the experiment The experiment aimed to identify and analyze faults or anomalies in microclimate parameters both inside and outside of two distinct locations: residential and non-residential buildings. Using a hardware setup with over 16 sensors, the study continuously collected data on various microclimate characteristics such as temperature, humidity, and carbon dioxide levels. This real-time data was then transmitted to Google Sheets for analysis[13]. System architecture

1. Sensors Temperature and humidity: Measure climatic conditions. CO2 and TVOC (Total Volatile Organic Compounds): Evaluate the air quality. Pressure: Measures atmospheric pressure. Electrical parameters: Monitor the status of power supply and consumption. 2. Microcontroller Processes data from sensors and sends it for analysis. 3. Data processing Server or cloud storage: Stores the collected data. Analysis algorithms: Use machine learning to identify anomalies and faults. 4. User Interface Display: Displays the current system parameters and status. Application or web interface: Allows the user to receive notifications and manage the system. 5.

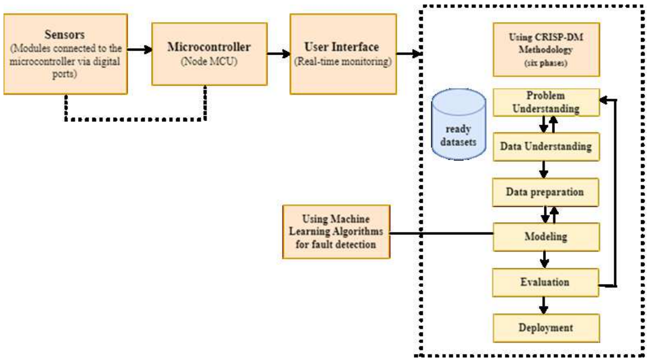

Fig. 1: Architecture of an ML-based microclimate system.

Notification system Notifies users of faults found and recommendations for elimination.

Management of Noise

In the experimental setup, preliminary data filtering techniques were applied to reduce the impact of noise before analysis. Employing methods such as moving aver-ages and Kalman filtering allowed for enhanced data quality, which in turn aids the re-liability of subsequent analyses, including PCA. In the experimental setup, a moving average and a Kalman filter were used to reduce the effect of noise, which improved the quality of the data. This contributes to the reliability of subsequent algorithms, including the principal component Method. Alternative approaches to noise reduction include wavelet transforms and adaptive filters, which also effectively suppress noise while preserving meaningful signals. In addition, alternatives to PCA, such as t-SNE and Uniform Manifold Approximation and Projection (UMAP) methods, are possible for dimensionality reduction tasks, especially when the data is highly noisy and it is necessary to preserve nonlinear dependencies. However, in the conditions of our experiment — with a moderate noise level and the need for real—time data processing - the use of moving averages and the Kalman filter turned out to be the optimal solution. These methods were chosen for their computational efficiency and the ability to preserve key data characteristics, which ensures the reliability of subsequent analysis, including the use of the PCA method.

4 Data analysis and results

The data analysis focused on detecting anomalous values or discrepancies that might indicate issues with the microclimate control system or other abnormal conditions. By employing a scientific approach, the analysis not only identified specific faults but also revealed trends and patterns, helping to develop effective management strategies for microclimate control in buildings.

The analysis showed that the presence of noise in the data significantly reduces the accuracy of the PCA method at high noise intensity, which emphasizes the need for additional filtering measures. As shown in the study on the modification of PCA to work with noisy data and omissions, where a weighted version of the EM-PCA algorithm is used to reduce the influence of heteroscedastic noise, more complex methods can significantly improve the accuracy of the results and reduce the influence of interference. Future research may consider multi-channel filtering or combined approaches with hybrid machine learning models to improve diagnostic accuracy[13]. Variability of Microclimate Parameters The unique characteristics of the two locations highlighted the variability of microclimate parameters.

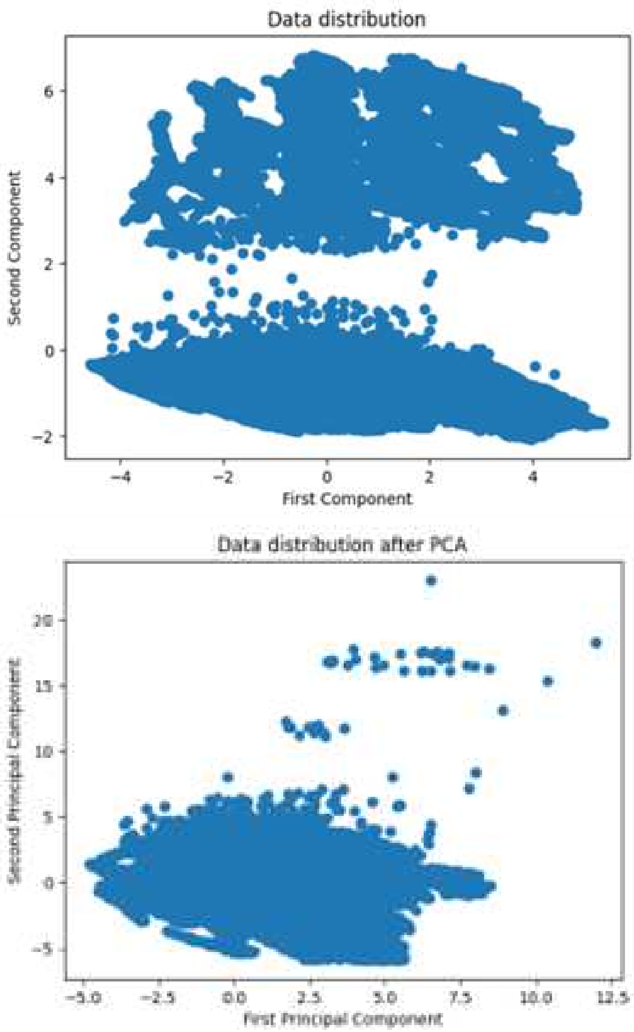

Fig. 2: Figure 2. Data distribution before and after PCA.

In the country house, factors like heating and ventilation systems, geographical location, and room size influenced conditions. In contrast, the kindergarten's microclimate was particularly critical due to children's sensitivity to temperature and humidity, with factors like occupancy and activity levels playing a significant role. Conducting the experiment in these diverse settings provided a more comprehensive understanding of HVAC systems. *FDD Approaches* Knowledge-Based Methods These methods utilize expert knowledge and predefined rules derived from historical performance data. They often employ if-then rules or heuristics to identify faults based on established operating conditions. While effective in complex scenarios, they require substantial input from domain experts and may struggle with variability. *Data-Driven Methods* In recent research on fault detection and diagnosis in microclimate control systems, a study by Daurenbayeva et.al (2023) provides a broader view on the topic, emphasizing the importance of controlling environmental parameters such as temperature, humidity, and air speed for both human comfort and energy efficiency. The study dis-cusses various microclimate control systems, including those used in greenhouses, where maintaining an optimal microclimate is crucial for plant growth. This subcategory analyzes historical data using statistical techniques and machine learning algorithms to identify patterns and anomalies. These methods do not rely on explicit rules but instead learn from past data, making them adaptable to changing conditions. However, they require high-quality datasets and can face challenges in interpretability. *Comparison of Methods* Comparison of traditional statistical methods against AI-driven methods reveals significant advantages of AI in adaptability and accuracy. Traditional methods often rely on fixed thresholds that may fail to account for the dynamic nature of HVAC systems, while AI methods can learn from trends and patterns in data, providing more reliable diagnostics. *3.1 Understanding Variance in Principal Component Analysis (PCA)* In Principal Component Analysis, understanding variance plays a crucial role in grasping the essence of the technique and its outcomes. Variance serves as a fundamental concept in PCA, delineating how much information each principal component retains from the original dataset. PCA, while a powerful tool for dimensionality reduction and data representation, has its limitations, particularly when derived from noisy data. The main limitation of using FDD methods in HVAC systems is their sensitivity to data noise caused by sensor inaccuracy or external influences. For example, methods such as PCA may not adequately represent the variability of data in conditions of high noise, which can lead to false anomalies. This limitation makes it difficult to interpret the results and reduces the overall accuracy of the diagnosis. *Preprocessing Before PCA Application* It may be beneficial to conduct preliminary data processing—such as filtering and noise reduction—before applying PCA. This approach enhances the algorithm's effectiveness by ensuring cleaner data input. Although PCA is advantageous, the presence of noise can significantly skew results, indicating that other techniques like t-SNE or UMAP might be considered for dimensionality reduction in particularly noisy datasets.
Explained Variance Ratios for the residential building:
PC1: 34.06%
PC2: 25.77%
PC3: 20.77%
PC4: 19.40%
Explained Variance Ratios for the non-residential building:
PC1: 40.08%
PC2: 21.21%
PC3: 20.94%
PC4: 17.76% *T*hough PCA is effective for reducing dimensions, the explained variance ratios may not accurately reflect the true variability, especially in the presence of noise. The cumulative explained variance for the first two principal components indicates substantial variability capture, further emphasizing the need for cautious interpretation given PCA's susceptibility to noise in data. *Orthogonal Transformation* To preserve the total variance, the transformation applied must be orthogonal, ensuring the trace remains invariant. The traditional use of PCA is to retain only the first k¡p principal components to maximize the variance explained by these components while minimizing the variance of the remaining components. *PCA Limitations* It is important to note that a single parameter does not solely define the variance captured by PCA. The actual proportion of total variance captured can vary due to noise and other influencing factors [23].

5 Discussion

The Principal Component Method is a powerful tool for dimensionality reduction and data structure analysis, but its application in noisy data conditions can be fraught with certain difficulties. In particular, the coefficients of explained variance obtained through PCA do not always accurately reflect the actual variability of the data, since they may include noise-related components. For example, the statement that one component explains 40.08% of the variation may be incorrect if a significant part of the variability is due to random fluctuations or noise in the data. This highlights the need to take into account the limitations of the method and sensitivity to noise. During the analysis of the cumulative explained variance for the first four components, it was found that in the first data set they explain 61.29% of the total variance, and in the second — 59.83%. These values indicate that, despite minor differences between the sets, the first four components in both cases explain most of the variability in the data. However, as shown in the practice of using PCA, for the purposes of further analysis and modeling, it is usually sufficient to use only the first few components, especially if their cumulative explained variance is more than 60%. In our case, despite the fact that the first four components together give high values of the explained variance, it was decided to use the first three components to simplify interpretation and effective modeling. The reasons for this choice are as follows: Cumulative explained variance for the first three components: In the first dataset, the first three components explain 80.6% of the variance (34,06% + 25,77% + 20,77%), which is a fairly high indicator. In the second set, this indicator is 82,23%:(40,08% +21,21%+20,94%), which also indicates a significant degree of information retention. Reducing the complexity of the analysis: Using the first three components allows you to significantly reduce the dimensionality of the data, while preserving most of the important information. It also simplifies data interpretation and visualization, since working with four components can lead to unnecessary complexity and difficulties in analysis. Noise reduction: The inclusion of the fourth component may result in accounting for variations that are not essential to the data structure, especially if these variations may be due to random noise. Using the first three components allows you to focus on the most relevant information and reduce the impact of noise. As a result of the analysis, it was decided to limit ourselves to the first three main components, which made it possible to explain more than 80% of the variance in the data. This approach provides a balance between preserving information and simplifying analysis, minimizing the impact of noise and increasing the interpretability of results. Future directions may include more sophisticated adaptive models for processing data from various sources, as shown in a study using collaborative processing of heterogeneous data to improve monitoring accuracy[8]. A similar approach can be applied in HVAC systems, allowing you to take into account a variety of parameters and effectively identify faults at an early stage.

6 Conclusion

The integration of Machine Learning methods into microclimate management systems for Fault Detection and Diagnosis offers significant opportunities to enhance the efficiency and reliability of these systems. However, implementing ML in this area presents challenges, such as working with noisy data and accounting for varying operational conditions. The conducted research demonstrates that combining data-driven and knowledge-based approaches can significantly improve fault detection processes, contributing to the development of more adaptive and resilient microclimate management systems. The practical significance of the proposed methods lies in improving diagnostic accuracy and enhancing the robustness of algorithms in the presence of sensor malfunctions or failures. These approaches are not only beneficial for managing HVAC systems but also have applications in related fields such as intelligent building systems, industrial facility management, predictive maintenance, and energy optimization in the energy sector. Promising future research directions include developing adaptive models that can effectively handle fluctuating noise levels and automatically adjust diagnostic thresholds to improve accuracy under real-world conditions. The use of advanced architectures, such as deep neural networks and hybrid models, will further enhance diagnostic system stability under high noise conditions. Additionally, research into improving model interpretability remains an important focus. A promising direction for future work is the development of adaptive hybrid models that can account for interactions among various microclimate parameters and variability in operating conditions, thereby increasing the stability and efficiency of FDD systems. For example, hybrid approaches combining physical models with ML techniques, as discussed in[16], can significantly improve FDD processes. Moreover, the use of Principal Component Analysis (PCA) in conjunction with ML techniques for FDD in microclimate systems is a powerful tool for improving their reliability and efficiency. PCA helps filter out noise, highlight significant signals, and improve prediction accuracy. To achieve the best results, it is essential to consider both the characteristics of the data and the specific features of the system. Therefore, PCA-based approaches for data processing in microclimate systems represent a promising and effective tool for system monitoring and early fault detection.

7 Acknowledgement

This article was made with the support of the Ministry of Science and Higher Education of the Russian Federation in accordance with agreement (N: 075-15-2022-317) dated 20 April 2022 on providing a grant in the form of subsidies from the Federal budget of the Russian Federation. The grant was provided for state support for the creation and development of a World-class Scientific Center Agrotechnologies for the Future.

References

[1] F. Farmani, M. Parvizimosaed, H. Monsef, and A. Rahimi-Kian, Electr. Power Energy Syst. **2018**, 523–536 (2021).
[2] J. Hyvärinen and S. Kärki, IEA Annex 25: Real Time Simulation of HVAC Systems for Building Optimization, Fault Detection and Diagnosis, Technical Report, VTT Building Technology, Espoo, Finland (1996).
[3] A. Nacer, B. Marhic, L. Delahoche, and J.B. Masson, Build. Environ. 142, 484–501 (2018).
[4] A. Nurlanuly and N. Daurenbayeva, World Sci. Eng. Sci. 1, 24–28 (2019).
[5] V.G. Zhitov, Ph.D. Thesis, Irkutsk State Technical University, Irkutsk, Russia (2007).

[6] D. Miljković, Proceedings of the 34th International Convention MIPRO, Opatija, Croatia, pp. 750–755 (2011).
[7] N. Daurenbayeva, A. Nurlanuly, L. Atymtayeva, and M. Mendes, Energies **16**, 3508 (2023). https://doi.org/10.3390/en16083508.
[8] O. Kuzichkin, A. Grecheneva, A. Bykov, N. Dorofeev, and D. Surzhik, International Multidisciplinary Scientific GeoConference: SGEM **1.1**, 877–884 (2018).
[9] S.L. Zhou, A.A. Shah, P.K. Leung, X. Zhu, and Q. Liao, Decarbonization (2023). https://doi.org/10.1016/j.decarb.2023.100023.
[10] Q. Wang, Y. Bai, Y. Zhang, Y. Xu, and Y. Liu, IEEE International Conference on Artificial Intelligence and Technology (AINIT) (2024). https://doi.org/10.1109/ainit61980.2024.10581610.
[11] F. Hamayat, R.F. Ahmad, A. Ud Din, and S. Zubair, 2023 International Conference on Frontiers of Information Technology (FIT), Islamabad, Pakistan, pp. 148–153 (2023). https://doi.org/10.1109/FIT60620.2023.00036.
[12] S. West, Y. Guo, R. Wang, and J. Wall, Journal of Building Performance 2, 15–25 (2011).
[13] S. Bailey, Publ. Astron. Soc. Pac. 124, 1015–1023 (2012). https://doi.org/10.1086/668105.
[14] B. Mateus, M. Mendes, J.T. Farinha, A.B. Martins, and A.M. Cardoso, In Proceedings of IncoME-VI and TEPEN 2021; Springer: Berlin/Heidelberg, Germany, pp. 11–25 (2023).
[15] N. Daurenbayeva, L. Atymtayeva, A. Nurlanuly, Proceedings of the 2023 International Conference on Electrical, Computer and Communication Engineering (ICECCO), pp. 1–4 (2023). https://doi.org/10.1109/ICECCO58239.2023.10147131.
[16] Y. Li, Comput. Intell. Neurosci. **2021**, 6612342 (2021).
[17] Z. Xiangjun, W. Yuanyuan, and X. Yao, Fault Detection, edited by W. Zhang, IntechOpen, Rijeka, Croatia, Chapter 4 (2010).
[18] P. Barandier, M. Mendes, and A.J.M. Cardoso, Energy and Buildings 316, 114342 (2024). https://doi.org/10.1016/j.enbuild.2024.114342.
[19] J. Bi, H. Wang, E. Yan, C. Wang, K. Yan, L. Jiang, and B. Yang, Energy Reviews 3(2), 100071 (2024). https://doi.org/10.1016/j.enrev.2024.100071.
[20] I. Matetić, I. Štajduhar, I. Wolf, and S. Ljubić, Sensors 23, 1 (2023). https://doi.org/10.3390/s23010001.
[21] M. Kanatov, L. Atymtayeva, A.M.A. Musleh, and G. Tulemissova, Appl. Math. Inf. Sci. **13**(5), 859–865 (2019). https://doi.org/10.18576/amis/130520.
[22] L. Atymtayeva, M. Kanatov, and A.M.A. Musleh, Appl. Math. Inf. Sci.17(2), 375–383 (2023). https://doi.org/10.18576/amis/170220.
[23] PCA Explained Variance, https://ro-che.info/articles/2017-12-11-pca-explained-variance (accessed 15 Nov. 2024).

Nurkamilya Daurenbayeva is a Senior Lecturer in the ”Computer Engineering” Department at the International University of Information Technology. She holds Bachelor’s and Master’s degrees from Kazakh National Technical University, completing her Master’s as part of the ”Shanghai Cooperation Organization University” program in collaboration with ITMO University in Saint Petersburg, Russia. Currently, she is pursuing doctoral studies in ”Computer and Software Engineering” at International Information Technology University. Her research focuses on energy modeling, intelligent microclimate control, and machine learning for fault detection. An active educator since 2016, she has authored several international publications and holds a copyright certificate for a microclimate monitoring hardware complex.

Lyazzat Atymtayeva received the Ph.D and Doctor of Science degree in Mechanics, Mathematics and Computer Science at Al-Farabi National University, Kazakhstan. Now she is working as associate professor in Information Systems at SDU University. Her research interests are in the areas of mechanics, applied mathematics and computer science including the numerical and rigorous mathematical methods and models for mechanical engineering and computer science, intelligent and expert systems in Information Security, Artificial Intelligence and Machine Learning, Project Management and Human-Computer Interaction. She has published research papers in reputed international journals of mathematical and computer sciences. She is a reviewer and an editor of international journals in mathematics and information sciences.

Almas Nurlanuly is a Senior Lecturer with over 15 years of teaching experience. He holds a Bachelor's degree in "Multichannel Telecommunication Systems" from Almaty University of Power Engineering and Telecommunications (AUPET, 2003–2008) and a Master's degree in "Radio Engineering, Electronics, and Telecommunications" from the same university (2012–2014). Additionally, he pursued doctoral studies in "Thermal Power Engineering" at AUPET (2016–2019), though without completing a defense.He has contributed to multiple peer-reviewed journals and databases, holds patents and a copyright certificate for utility models and technical innovations. His scientific interests focus on electronics, robotics, and related technological advancements. In this domain, he has led several research projects aimed at developing innovative solutions.

Artem Bykov received his PhD in Technical Sciences and the academic title of Associate Professor in Systems, Networks, and Telecommunication Devices from Vladimir State University, Vladimir, Russia. He is an Associate Professor at the Department of Computer Engineering, International Information Technology University, Almaty, Kazakhstan. His research topics mainly include data mining, processing of geophysical control data, and designing systems for automated information processing.

Bakhytzhan Akhmetov received his PhD in Automated Design Systems in 1984 and his Doctorate in Social and Economic Systems Management in 2007 from Satbayev University, Kazakhstan. He is a Professor and an Academician of the National Engineering Academy of the Republic of Kazakhstan with a specialization in Informatics, Computer Engineering, and Management. He has 45 years of scientific and pedagogical experience, progressing from Lecturer to Rector. His research focuses on Information and Communication Technologies, Artificial Intelligence, and Cybersecurity. Dr. Akhmetov has authored over 500 scientific publications, including 48 articles indexed in Scopus. He has supervised 2 Doctors of Science and 17 PhD candidates. His citation ratings are Yandex: 8, Web of Science: 4, and Google Scholar: 17.

Gabit Shuitenov is an Associate Professor in the "Information Systems and Technologies" Department at Esil University, Astana, Kazakhstan. He graduated from Aktobe State University with a degree in Physics and Informatics in 1995 and earned his Candidate of Sciences degree at Abay National Pedagogical University in 2007. The author of over 70 publications, including works indexed in Scopus, his research focuses on education informatization, digital technologies, and information system security. He was an expert for the Erasmus+ LMPI project (2017–2020) and is currently involved in two grant-funded scientific and technical projects.

Umut Turusbekova is Associate Professor of the Department of Artificial Intelligence Technologies at L.N. Gumilyov Eurasian National University. She earned her bachelor's degree from Shakarim Semipalatinsk State University. In 2008, she defended her dissertation for the degree of Doctor of Philosophy (PhD) at the L.N. Gumilyov Eurasian National University. Her research interests include natural language processing and machine learning, as well as information security issues. Umut is the author of several publications in international journals and holds a copyright certificate for the electronic textbook "Discrete Mathematics".