УДК 697.9:004.8

Киргизбаев Данияр Мусаевич

Руководитель группы КИП, отдел Техобслуживания, Тенгизшевройл,

г. Алматы, Казахстан

МЕТОДИКА ОЦЕНКИ КЛИМАТИЧЕСКИ ОБУСЛОВЛЕННЫХ РИСКОВ НАДЕЖНОСТИ ЭКСПЛУАТАЦИИ СИСТЕМ HVAC КРИТИЧЕСКОЙ ИНФРАСТРУКТУРЫ С ПРИМЕНЕНИЕМ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА

Аннотация: Разработана методика оценки климатической нагрузки на HVAC критической инфраструктуры по индексу CARI. На данных ERA5 и CAMS EAC4 за 2020–2024 гг. для четырех городов Казахстана и США рассчитаны шесть климатических факторов. Веса CRITIC определены по 192 наблюдениям 2020–2023 гг.; 48 наблюдений 2024 г. использованы для проверки. Средние CARI за 2024 г. составили 0,8417 в Астане, 0,8243 в Караганде, 0,7969 в Финиксе и 0,6196 в Хьюстоне. Исследована чувствительность к температуре, осадкам и параметрам нормирования. Random Forest снизил ошибку прогноза PM2.5 на 11,9%. По базе LBNL изучены интервалы неисправной работы двух HVAC-установок; на фактических рядах CARI выполнена имитация условных отказов.

Ключевые слова: критическая инфраструктура, HVAC, климатические риски, надежность, индекс CARI, реанализ ERA5, CAMS EAC4, искусственный интеллект, Random Forest.

Daniyar Kirgizbayev,

Head of Instrumentation and Control (I&C) Group,

Maintenance Division, Tengizchevroil,

Almaty, Kazakhstan

METHODOLOGY FOR ASSESSING CLIMATE-RELATED RISKS TO THE OPERATIONAL RELIABILITY OF CRITICAL

INFRASTRUCTURE HVAC SYSTEMS USING ARTIFICIAL INTELLIGENCE

Abstract: A methodology for assessing climate-related exposure of critical infrastructure HVAC systems using the CARI index was developed. Six climatic factors were evaluated using ERA5 and CAMS EAC4 data for four cities in Kazakhstan and the United States over 2020–2024. CRITIC weights were calculated from 192 observations for 2020–2023, while 48 observations from 2024 were used for validation. Mean CARI values for 2024 were 0.8417 in Astana, 0.8243 in Karaganda, 0.7969 in Phoenix, and 0.6196 in Houston. Sensitivity to temperature, precipitation, and normalization parameters was assessed. Random Forest reduced the PM2.5 forecasting error by 11.9%. Faulty operating periods of two HVAC units from the LBNL database were examined, and hypothetical failures were simulated using actual CARI time series.
Keywords: critical infrastructure, HVAC, climate-related risks, reliability, CARI index, ERA5 reanalysis, CAMS EAC4, artificial intelligence, Random Forest.

For data centers, telecommunications facilities, and healthcare institutions, reliable HVAC operation is essential for maintaining indoor environmental conditions, as these conditions affect equipment loads. General issues of infrastructure vulnerability and building cooling strategies have been examined in studies [1; 2]. Continuity of technological processes is particularly important for critical infrastructure. ASHRAE recommends an inlet air temperature of 18–27 °C for certain classes of IT equipment [3]. Meanwhile, the scientific literature presents machine learning methods for predicting building energy consumption and thermal loads [4], as well as intelligent HVAC control [5]. However, forecasting climatic parameters characterizes expected external conditions, whereas assessing actual equipment reliability requires information on its technical condition and recorded failures. Accordingly, developing a reproducible composite indicator that enables comparisons of climate-related loads and

assessment of the influence of individual assumptions on the results is of particular interest.

The aim of this study is to refine the methodology for calculating the CARI index through the statistical determination of weighting coefficients for climatic factors, conduct a comparative evaluation for Kazakhstan and the United States, assess the sensitivity of the index to model parameters, and investigate the potential application of Random Forest for forecasting the underlying climatic variables.

Astana and Karaganda (Kazakhstan) and Phoenix and Houston (United States) were examined over the period 2020–2024. Daily NASA POWER data [6] were used for source verification, while exposure calculations were based on hourly ERA5 data accessed through Open-Meteo [7] and monthly PM2.5 concentrations from CAMS EAC4 [8]. The dataset comprises 240 city-month observations and 175,392 hourly ERA5 records, with no missing temperature values. PM2.5 concentrations were obtained from the nearest EAC4 grid points and converted to μg/m³.

Six factors were selected to compare climatic exposures: temperature (T), humidity (H), precipitation (P), wind (W), PM2.5 (A), and radiation (S). Temperature exposure was defined as the mean hourly exceedance above 28 °C and deviation below −10 °C, while humidity exposure was defined as the proportion of hours with relative humidity (RH) above 80%, dew point above 15 °C, and temperature above 18 °C. For precipitation, the monthly maximum hourly intensity was used; for wind, the exceedance of the maximum wind gust above 12 m/s was considered. For PM2.5, the exceedance of a monthly mean concentration of 5 μg/m³

was considered, while radiation exposure was defined as the mean hourly exceedance above 650 W/m².

Since the selected factors are measured in different physical units, normalization was applied to enable their joint analysis by expressing each exposure on a dimensionless scale:

$$\square_{\square} = \frac{\square_{\square}}{\square_{\square} + \square_{\square}}, \qquad 0 \leq \square_{\square} < 1, \tag{1}$$

where x_i is the monthly exposure of the i-th climatic factor, and the corresponding scale parameter is positive. For temperature, humidity, precipitation, wind, PM2.5, and solar radiation, the scale parameters were set to 5 °C, 0.15, 8 mm/h, 12 m/s, 15 μg/m³, and 60 W/m², respectively. The index was calculated using the following formula:

$$CARI = \prod_{i=1}^{6} (1 - r_i)^{w_i}, \qquad w_i \geq 0, \quad \sum_{i=1}^{6} w_i = 1 \tag{2}$$

The CARI index ranges from 0 (exclusive) to 1 (inclusive), with lower values indicating greater climate-related exposure. The annual index is calculated as the mean of the 12 monthly indices.

The weights were calculated using the CRITIC method [9]. The six variables r_i were normalized to the [0, 1] scale using the minimum and maximum values for 2020–2023, after which the standard deviations σ_i and Pearson correlation coefficients ρ_{ik} were calculated. The information content and weight of each factor were calculated as follows:

$$C_i = \sigma_i \sum_{k=1}^{6} (1 - \rho_{ik}), \qquad w_i = \frac{C_i}{\sum_{j=1}^{6} C_j} \tag{3}$$

The weights were estimated using 192 observations from 2020–2023 (Table 1), while 48 observations from 2024 were reserved for validation.

Sensitivity was assessed using author-defined, equal, and entropy-based weights [10].

Table 1. CARI weighting coefficients

Factor	Author-defined	CRITIC, 2020–2023	Entropy-based, 2020–2024
Temperature	0.300	0.2033	0.2330
Humidity	0.150	0.1892	0.3801
Precipitation	0.150	0.1348	0.0745
Wind	0.100	0.2168	0.0622
PM2.5	0.200	0.0918	0.0572
Radiation	0.100	0.1641	0.1930

The CRITIC method increased the relative weight of the wind component and reduced the weight of PM2.5 compared with the expert-assigned weights. The entropy-based method identified humidity as a significant factor. A comparison of the weighting coefficients is presented in Figure 1.

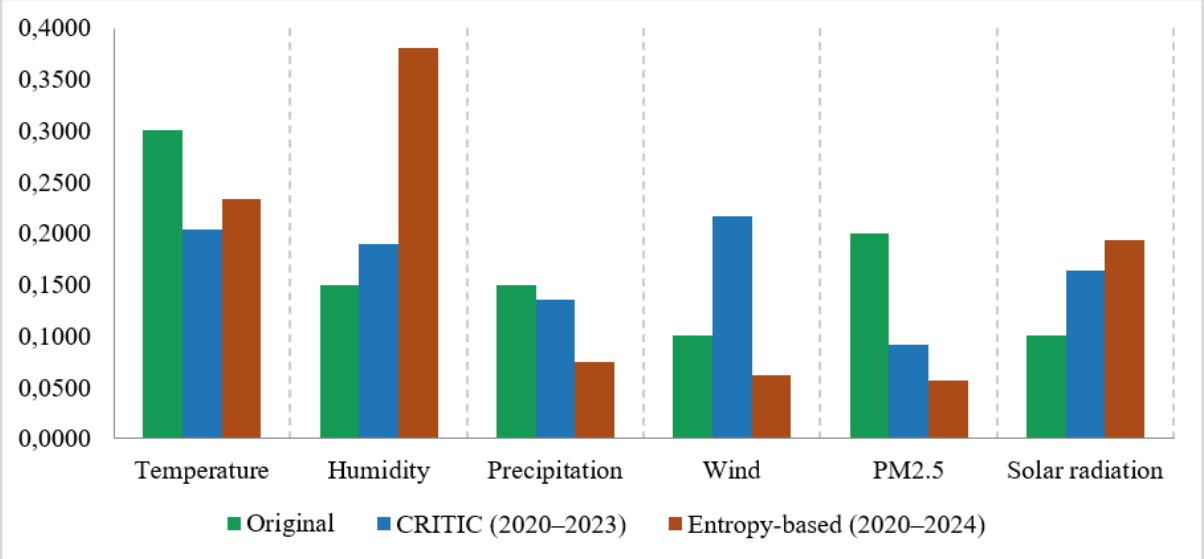


Fig. 1. Comparison of climatic factor weights: author-defined values, CRITIC (2020–2023), and the entropy-based method. Source: author's calculations.

Using CRITIC weights, the mean CARI for 2020–2024 was 0.8307 in Astana, 0.8170 in Karaganda, 0.7981 in Phoenix, and 0.6367 in Houston. The highest relative climate-related exposure was observed in Houston. A comparison of the 2024 results obtained using different weighting methods is presented in Fig. 2.

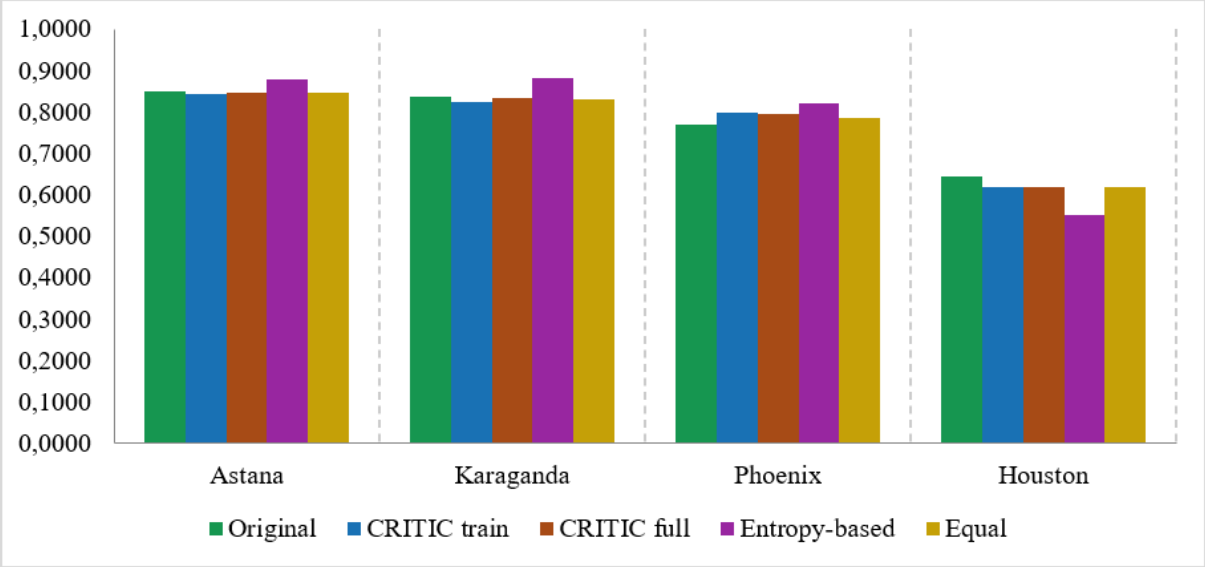


Fig. 2. Comparison of mean CARI values for 2024 using different weighting methods. Source: author's calculations.

When all normalization parameters were varied by ±20%, the mean CARI values for 2024 ranged from 0.8136 to 0.8622 in Astana, 0.7937 to 0.8468 in Karaganda, 0.7649 to 0.8210 in Phoenix, and 0.5752 to 0.6546 in Houston. The index is sensitive to normalization scales; therefore, these parameters must be calibrated for the specific HVAC system.

The scenario involving a 2 °C temperature increase and a 20% increase in precipitation, with all other variables held constant, yielded the following CARI values for 2024: 0.8424 (Astana), 0.8241 (Karaganda), 0.7828 (Phoenix), and 0.6034 (Houston). Changes relative to the baseline index were +0.0007, −0.0002, −0.0141, and −0.0161, respectively. Cold exposure decreased in Astana, while heat-related loads increased in the United States.

To investigate the potential for forecasting climate-related loads, an additional experiment was conducted using the Random Forest machine learning algorithm [11] to predict monthly mean temperature, maximum hourly precipitation, and PM2.5 concentrations. Each model comprised 150 trees (maximum depth of 7, with a minimum of 2 observations per leaf). The input

features included city, cyclically encoded month, the preceding month, and the corresponding month of the previous year. The models were trained on 144 observations from 2021–2023 and validated on 48 observations from 2024. The actual value for the corresponding month of the previous year served as the seasonal baseline.

Forecast accuracy was evaluated using the mean absolute error (MAE), defined as follows:

$$MAE = \frac{1}{N}\sum_{j=1}^{N} |y_j - \hat{y}_j| \quad (4)$$

where N is the number of validation observations, y_j is the actual value, and $\hat{y}_j$ is the predicted value. For temperature, the MAE values of Random Forest and the seasonal baseline model were 2.246 and 1.990 °C, respectively; for precipitation, 2.100 and 2.073 mm/h; and for PM2.5, 1.834 and 2.083 µg/m³. An 11.9% reduction in forecasting error was achieved only for PM2.5. Comparisons of the forecasts and errors are presented in Figures 3 and 4.

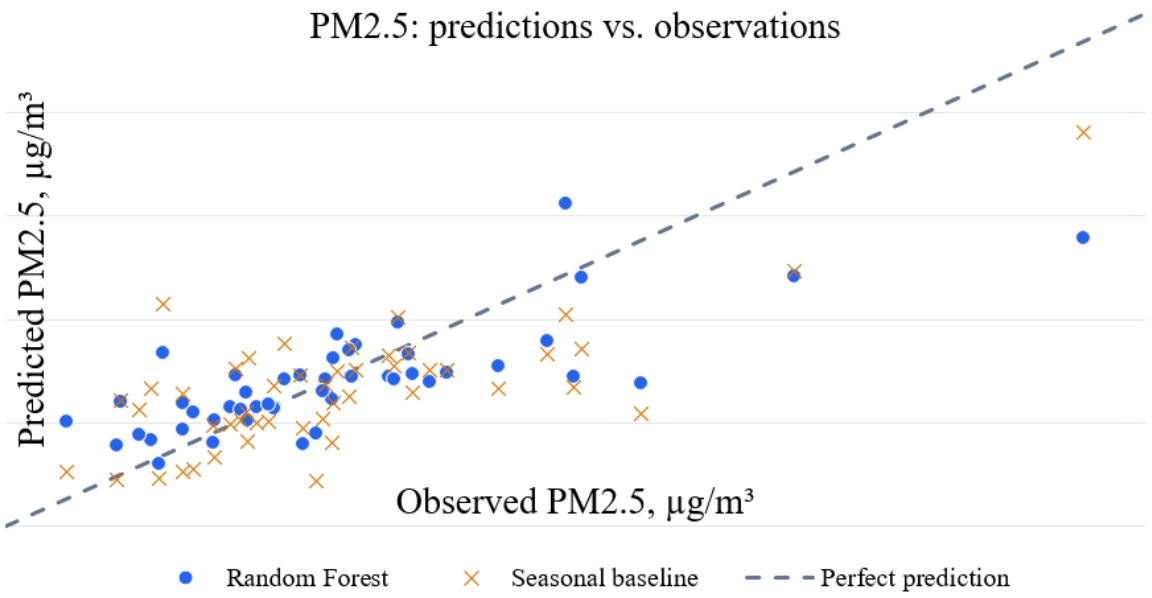


Fig. 3. Actual PM2.5 values and forecasts generated by Random Forest and the seasonal baseline model for the 2024 test dataset. Source: author's calculations.

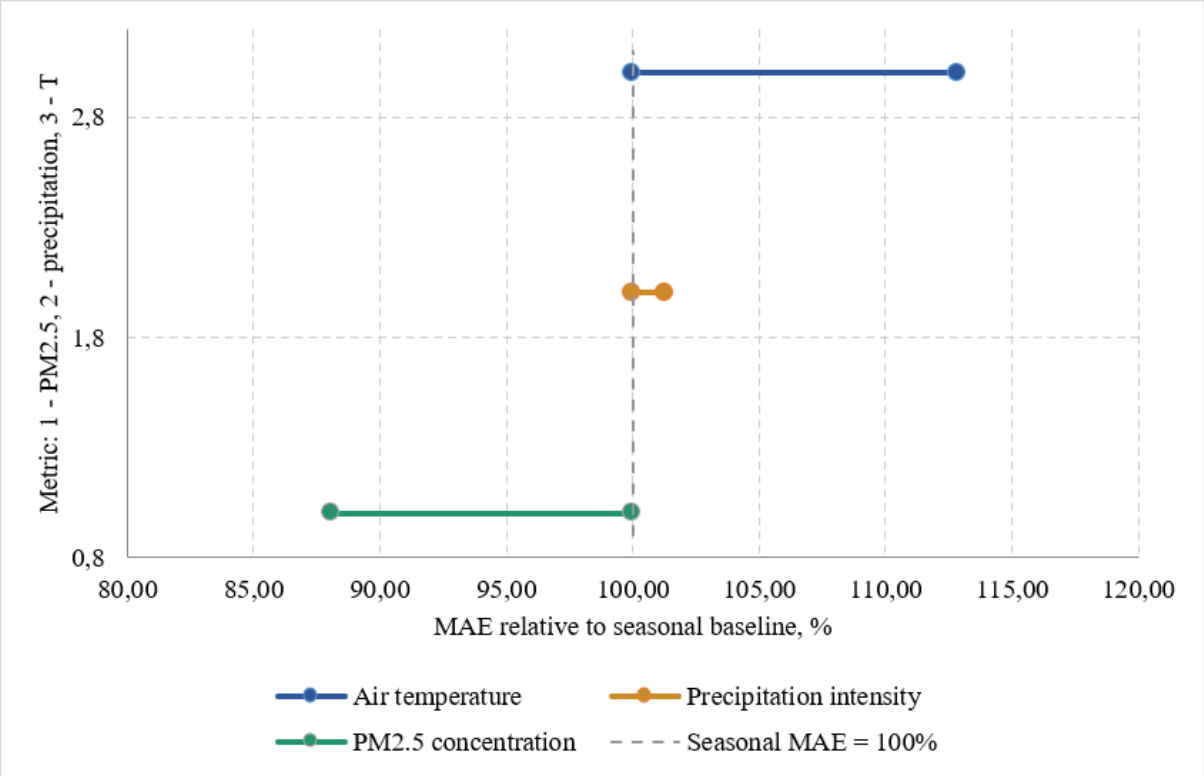


Fig. 4. Relative MAE of Random Forest compared with the seasonal baseline model. Source: author's calculations.

CARI depends on the weighting coefficients and normalization parameters, while the climatic factors are aggregated differently: precipitation and wind are represented by monthly maxima, whereas temperature and radiation are characterized by the duration of threshold exceedances. For PM2.5, the WHO annual guideline value of 5 μg/m³ was used [12]. The LBNL database documents faulty operating conditions of two rooftop units (RTUs): a second-stage compressor failure (Milford) and refrigerant undercharge (Colchester) [13; 14]. The published periods of faulty and normal operation are presented in Table 2.

Table 2. Published intervals of HVAC field observations based on LBNL data [14]

Site	Recorded fault	Days with fault	Days without fault
Milford	Second compressor stage fails to engage	61	51
Colchester	Refrigerant undercharge in Circuit B (40%)	29	184

To demonstrate the relationship between the climate index and quantitative reliability metrics, a simulation experiment was conducted using 240 actual CARI values calculated from ERA5 and CAMS EAC4 data for four cities over 2020–2024 as the climatic basis. One hypothetical HVAC unit operating continuously was assumed for each city. The occurrence rate of simulated failures was defined by the following relationship:

$$\lambda_t = \lambda_0 \cdot exp[\beta \cdot (C_0 - CARI_t)] \quad (5)$$

Here, λ_t is the expected mean number of failures per 1,000 h; $\lambda_0 = 0.35$ failures/1,000 h, $C_0 = 0.80$, and $\beta = 4$ are parameters used exclusively for modeling purposes. As CARI decreases, the failure rate increases according to the assumed function. For a month with H_t calendar hours, the expected number of events is $\mu_t = H_t\lambda_t/1000$. The actual number of simulated events was generated from a Poisson distribution with parameter μ_t using a fixed random seed of 20261029. Each event was assigned hypothetical repair start and end times; repair duration was sampled from a uniform distribution ranging from 3 to 9 h (random seed 20261030). The calculations were performed as follows: MTBF = T_operating / N; MTTR = T_repair / N; $f_{1000} = 1000 \cdot N / T_operating$.

where T_operating is the total hypothetical operating time, T_repair is the total duration of hypothetical repairs, and N is the number of simulated failures. The calculation results are presented in Table 3:

Table 3. Results of CARI-based HVAC failure simulations (simulated data)

Period	Load class	Simulated failures	Failure rate per 1,000 h	MTBF / MTTR, h
2020–2024	High	33	0.745	1342.3 / 6.04
2020–2024	Normal	43	0.329	3038.0 / 6.14
2024	High	8	0.849	1178.4 / 6.60
2024	Normal	9	0.352	2843.8 / 6.90

A total of 76 simulated failures were obtained. Under high-load conditions, the failure frequency was 0.745 per 1,000 h, compared with 0.329 per 1,000 h

under normal-load conditions, yielding a rate ratio of 2.26 (95% CI: 1.44–3.56). For 2024, the rate ratio was 2.41 (95% CI: 0.93–6.25).

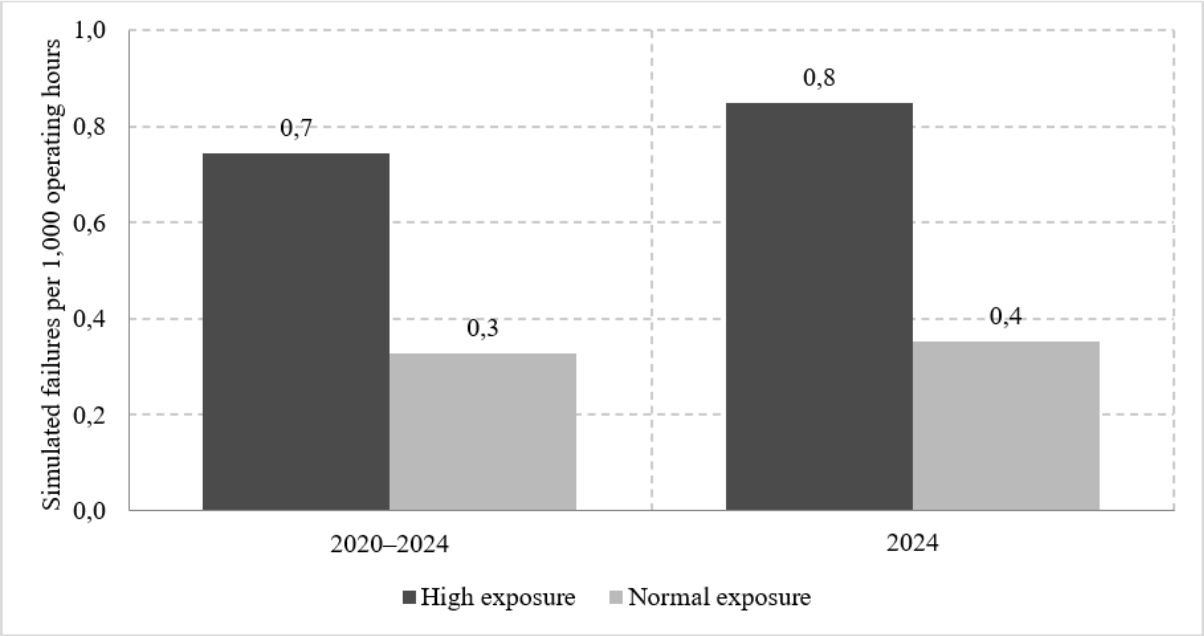


Fig. 5. Comparison of simulated HVAC failure rates at different CARI levels. Source: author's calculations.

The following conclusions were drawn from the results of the study:

1. The methodology for assessing climate-related loads on critical infrastructure HVAC systems was refined through the statistical determination of weighting coefficients for six factors using the CRITIC method.

2. Differences in climate-related loads were identified among the studied cities in Kazakhstan and the United States. The highest relative load was observed in Houston, while the seasonal characteristics of the other locations were determined by different combinations of temperature, humidity, and other climatic factors. The scenario analysis further confirmed that the results depend on the nature of changes in external conditions.

3. The potential applications of machine learning for forecasting individual climatic variables were identified. The Random Forest experiment demonstrated an 11.9% reduction in mean absolute error compared with the seasonal baseline method when forecasting PM2.5. No comparable advantage was observed for air temperature or maximum hourly precipitation intensity.

Based on LBNL data, field observation intervals corresponding to faulty and normal operating conditions of two HVAC units were systematically analyzed. The durations of the recorded operating states were determined, and the CARI-based simulation demonstrated a 2.26-fold increase in the frequency of hypothetical failures during periods of elevated climate-related loads.

References:

1. IPCC. Climate Change 2022: Impacts, Adaptation and Vulnerability. – Cambridge University Press, 2022. – URL: https://www.ipcc.ch/report/ar6/wg2/.

2. Zhang C., Kazanci O. B., Levinson R. [et al.]. Resilient cooling strategies – A critical review and qualitative assessment // Energy and Buildings. – 2021. – Vol. 251. – Art. 111312. – DOI: 10.1016/j.enbuild.2021.111312.

3. ASHRAE. ASHRAE Handbook – HVAC Applications. Chapter 20. Data Centers and Telecommunication Facilities. – 2023. – URL: https://handbook.ashrae.org/Handbooks/A23/SI/A23_Ch20/a23_ch20_si.aspx.

4. Qin H., Yu Z., Li Z., Li H., Zhang Y. Nearly Zero-Energy Building Load Forecasts through the Competition of Four Machine Learning Techniques // Buildings. – 2024. – Vol. 14, № 1. – Art. 147. – DOI: 10.3390/buildings14010147.

5. Lu S., Zhou S., Ding Y. [et al.]. Exploring the comprehensive integration of artificial intelligence in optimizing HVAC system operations // Results in Engineering. – 2025. – Vol. 25. – Art. 103765. – DOI: 10.1016/j.rineng.2024.103765.

6. NASA POWER. Daily API Documentation. – URL: https://power.larc.nasa.gov/docs/services/api/temporal/daily/ (Accessed: October 19, 2026).

7. Copernicus Climate Change Service. ERA5 hourly data on single levels. – DOI: 10.24381/cds.adbb2d47. – URL: https://open-meteo.com/en/docs/historical-weather-api (Accessed: October 19, 2026).

8. Copernicus Atmosphere Monitoring Service. CAMS global reanalysis (EAC4) monthly averaged fields. – DOI: 10.24381/fd75fff2. – URL: https://ads.atmosphere.copernicus.eu/datasets/cams-global-reanalysis-eac4-monthly (Accessed: October 19, 2026).

9. Diakoulaki D., Mavrotas G., Papayannakis L. Determining objective weights in multiple criteria problems: The CRITIC method // Computers & Operations Research. – 1995. – Vol. 22, № 7. – P. 763–770. – DOI: 10.1016/0305-0548(94)00059-H.

10. Shannon C. E. A Mathematical Theory of Communication // Bell System Technical Journal. – 1948. – Vol. 27, № 3. – P. 379–423. – DOI: 10.1002/j.1538-7305.1948.tb01338.x.

11. Breiman L. Random Forests // Machine Learning. – 2001. – Vol. 45. – P. 5–32. – DOI: 10.1023/A:1010933404324.

12. World Health Organization. WHO global air quality guidelines: particulate matter (PM2.5 and PM10), ozone, nitrogen dioxide, sulfur dioxide and carbon monoxide. – Geneva: WHO, 2021. – URL: https://www.who.int/publications/i/item/9789240034228.

13. Granderson J., Lin G., Chen Y. [et al.]. A labeled dataset for building HVAC systems operating in faulted and fault-free states // Scientific Data. – 2023. – Vol. 10. – Art. 342. – DOI: 10.1038/s41597-023-02197-w.

14. Lawrence Berkeley National Laboratory. Fault Detection and Diagnostics Data Sets: Roof Top Unit (RTU). Dataset inventory, tables F2–F3. – URL: https://faultdetection.lbl.gov/dataset/rtu/ (Accessed: October 25, 2026)