BULLETIN OF THE POLISH ACADEMY OF SCIENCES
TECHNICAL SCIENCES, Vol. 74(3), 2026, Article number: e158305
DOI: 10.24425/bpasts.2026.158305

The use of Internet of Things technology and artificial intelligence tools in analyzing the impact of environmental pollution on the performance of photovoltaic panels

Damian MAZUR[1], **Tomasz KOSSOWSKI**[1]*, **Grzegorz DRAŁUS**[1], **and Aneta ŁOBODZIŃSKA**[2]

[1] Faculty of Electrical and Computer Engineering, Rzeszow University of Technology, Poland
[2] Faculty of Mechanical Engineering and Aviation, Rzeszow University of Technology, Poland

Abstract. The impact of pollution on the performance of photovoltaic (PV) panels and the risk of hot spots is an important issue in the context of optimizing renewable energy systems. Dust, leaves, and bird droppings cause uneven illumination of the panel surface, leading to a decrease in efficiency and local overheating of the cells, which can result in permanent damage. The use of Internet of Things (IoT) technology in PV panel monitoring enables continuous tracking of pollution levels and their impact on system performance. Smart sensors located in various places and on the drones provide real-time data on temperature and air pollution levels of various chemical compounds. The collected information is sent to artificial intelligence-based systems, which analyze patterns and identify potential threats, such as the formation of hot spots or a drop in module performance. Authors method, based on a proprietary structure and selection of deep LSTM network parameters, outperforms other specified machine learning methods in terms of relative prediction accuracy for dust. Proposed algorithm also predicts more accurately than other machine learning methods. Thermal cameras combined with AI algorithms can accurately detect temperature anomalies on the surface of the panels and predict future problems. This allows to optimize cleaning schedule and make maintenance decisions based on actual data rather than periodic inspections.

Keywords: photovoltaic panels; thermal imaging; hotspots; natural dirt; energy efficiency.

1. INTRODUCTION

Drones, or unmanned aerial vehicles (UAVs), have become a key tool in data collection in a variety of fields in recent years. Their growing popularity comes from unique advantages such as the ability to move quickly over difficult terrain, reach places inaccessible by traditional methods, and relatively low operating costs compared to manned aircraft or ground-based measurement systems [1, 2]. Air pollution and the efficiency of photovoltaic (PV) installations are key issues in today's energy transition. In both cases, drones equipped with environmental sensors, thermal imaging cameras, and advanced AI modules are becoming an indispensable diagnostic tool [3, 4]. Their role includes not only ongoing monitoring of the environment and infrastructure, but also the identification of anomalies such as microcracks, hot spots, and pollutant accumulations that affect the degradation of the performance of the photovoltaic system. Air pollutants such as particulate matter (PM2.5, PM10), nitrogen oxides (NO_x) and volatile organic compounds (VOCs) have a direct impact on human health, but also on the technological components of PV installations [5]. Particle deposition on the surface of the panels leads to shading, reducing their efficiency, while acidic pollutants promote microcracks and corrosion of contacts. The efficiency of PV cells can be reduced by up to 30% in the case of long-term contamination. Mathematical models allow losses to be quantified. The atmosphere acts as a natural "transport system" for pollutants, and weather conditions have a major impact on their spread. Temperature influences chemical reactions in the atmosphere, e.g., increased temperature promotes the formation of tropospheric ozone. In high humidity, aerosol particles can absorb more water, which affects their interaction with light and their chemical composition. Strong winds can disperse pollutants and reduce their local concentration, while the absence of wind causes them to accumulate. Rain "washes" pollutant particles out of the air, reducing their concentration. Predictive models must take into account emission sources and their variability over time. Traffic generates nitrogen oxides (NO_x), particulate matter (PM2.5 and PM10) and other pollutants, especially during rush hour. Power plants, manufacturing plants and factories can emit large amounts of gases such as SO_2, CO_2, NO_x and volatile organic compounds (VOCs). During the heating season, emissions of particulate matter and CO_2 from coal-fired boilers, gas boilers, and fireplaces increase. Nitrogen fertilizers and animal husbandry contribute to emissions of ammonia (NH_3), which reacts with other pollutants to form secondary aerosols. Air pollution varies depending on the season, so predictive models must take this into account. In winter, temperature inver-

*e-mail: t.kossowski@prz.edu.pl

Manuscript submitted 2025-07-02, revised 2026-02-23, initially accepted for publication 2026-03-16, published in May 2026.

sions are more frequent, causing a "trap" for pollutants and high concentrations of smog in cities. Strong sunlight in summer promotes the formation of tropospheric ozone, which can be harmful to health. In autumn and winter, the amount of fuel burned increases, increasing emissions of particulate matter [6, 7].

The issue of PV panel efficiency and reliability is particularly important due to their rapidly growing share in the energy markets of many countries around the world. Both AI tools supporting the assessment of the condition and soiling of PV panels [8, 9] and various algorithms for their analysis are being tested [10]. Research related to the analysis and prediction of air pollution [11–13] can also be found. However, there is no comprehensive approach to the correlation between soiling and the energy efficiency of PV panels. The authors of the publication proposed collecting data related to pollutants such as PM10, PM2.5, PM4, PM1, and volatile chemical compounds HCN, HCl, VOC, HCHO, CO, CO_2, NO, and NO_2 using ground-based measuring stations and sensors placed on a drone [14]. This ensured accurate measurements and analyses at the location of the photovoltaic farm. The results are therefore very accurate and correlated with the location of the devices. The level of pollution can thus be directly related to the amount of energy produced (using information from the PV farm's energy system). The authors then proposed an algorithm that allows for more precise determination of trends and forecasting of changes in these pollutants than in other publications [15–17]. This can be directly correlated with the impact and effectively produced energy. To this end, research was conducted to analyze the impact of dirt on the operation of PV panels, which is presented in the following chapters. The most significant contribution of the authors is the development of a proprietary algorithm that improves the accuracy of air pollution predictions and, at the same time, determines their impact on future energy production based on field research. The developed algorithm is based on artificial intelligence tools, and the details of the solution and performance results are presented at the end of the publication [18].

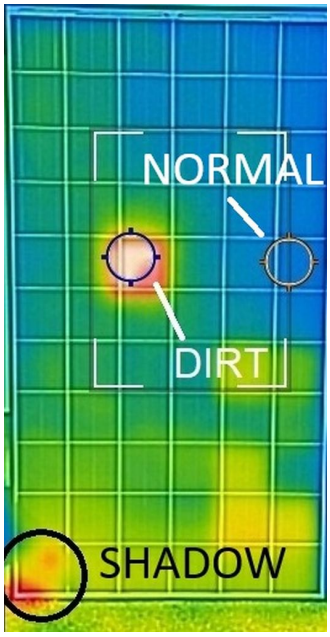

Fig. 1. Hot-spots from dirt and shadows, taken by the authors

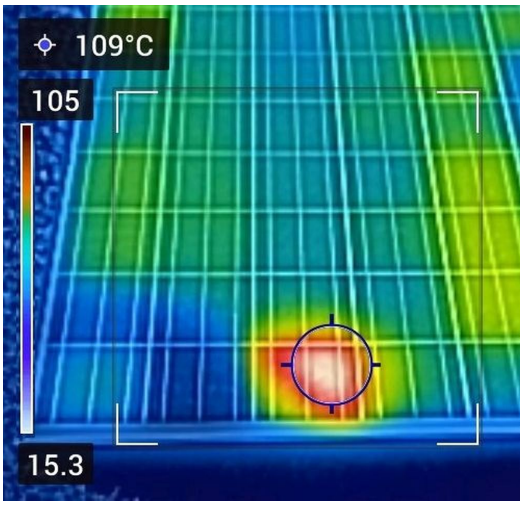

Fig. 2. Hot-spot from shadow, taken by the authors

2. METHODOLOGY

2.1. Hot-spots and dust

The hot-spot phenomenon is one of the most serious threats to the durability of PV panels. It occurs when part of the cell is shaded, dirty, or damaged, causing local overheating. The causes of hot spots can include partial shading of the panel (e.g., by leaves, bird droppings, branches), uneven dirt on the cell surface, or mechanical damage (e.g., microcracks). This can result in local overheating of the module (even above 150°C), which can cause permanent damage to the cell structure, leading to its degradation [18, 19]. The long-term effect is a reduction in the efficiency of the entire panel or a risk of fire in the PV installation. An example photo of a panel showing both the effect of dirt and shading, taken by the authors, is shown in Fig. 1. This photo was taken with a FLIR C5 thermal imaging camera. Photo 2, taken by the authors, shows the thermal effect of only 10 minutes of exposure (during irradiation at 1120 W/m^2. In this example, dirt from excrement was simulated by sticking a 3×3 cm piece of black adhesive tape. The rest of the module reached a temperature of 109°C, which can pose a real threat.

The above examples clearly show that the impact of dirt, especially spot dirt, can be significant for the proper operation and safety of PV panels. It should also be noted that dirt has a significant impact on the entire active surface of photovoltaic devices. The greater the accumulation, the lower the energy production. To explain the scale and magnitude of these relationships, graphs are presented in Figs. 3 and 4. Figure 3 shows the output current of a single photovoltaic panel consisting of 66 modules connected in series. The nominal voltage is 38–40 V, the peak current is 8 A, and the nominal power is 250 W. The panel dimensions are 985×1985 mm. The measurements were taken under variable sunlight conditions (a sunny day with partial cloud

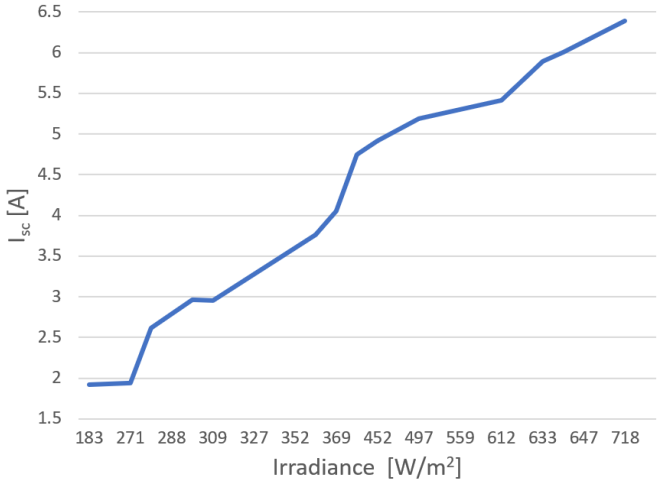

Fig. 3. Current vs irradiance

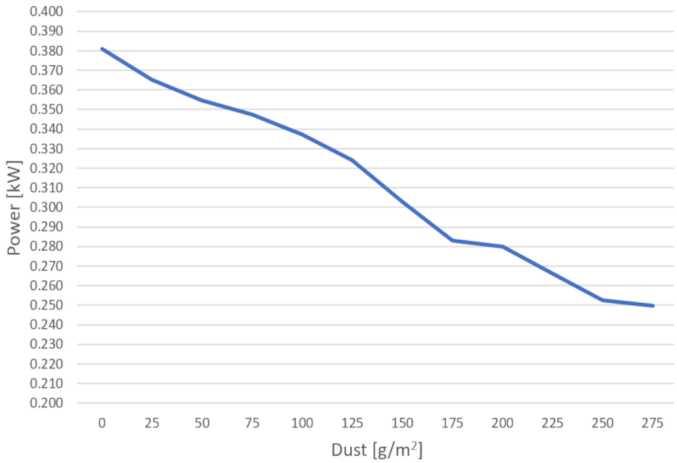

Fig. 4. Power via dust

cover and strong winds in the upper atmosphere). Importantly, the irradiance, current and voltage were measured simultaneously using a PV-dedicated meter, ensuring full-time correlation between all parameters. Thanks to the dynamic changes in sunlight, it was possible to record the electrical characteristics of the panel across almost the entire range of irradiance levels typical for the tested location (Rzeszów, Poland).

The measuring device was a SONEL PVM-1020 meter dedicated to PV panel measurements, together with an IRM-1 irradiance sensor. The device ensured synchronous measurement of irradiance, temperature, short-circuit current and voltage, which guarantees the comparability of the results despite natural fluctuations in sunlight. Figure 4 shows the effect of the decrease in energy production due to the increasing amount of contamination. In this experiment all electrical parameters (irradiance, short-circuit current, voltage and module temperature) were recorded using this measurement setup, ensuring full time correlation between all measured quantities. The test was designed as a controlled case study illustrating the trend of performance degradation with increasing contamination. The tests were carried out on a sunny and cloudless day, ensuring the stability of the measurements. Each portion of sand was weighed on a scale with an accuracy of 1 g and then evenly spread over the entire surface of the PV panel using a sieve. This test simulated the increasing amount of pollution on PV panels from natural sources over a long period of time. Their impact on the amount of energy produced is very clear, which is why it is so important to properly analyze and correlate pollution levels with energy production, maintenance, cleaning, and forecasting. It can also be supported by AI tools [20]. Table 1 lists all recorded values from the measuring device. The simulations were based on the standard single-diode photovoltaic model with series and

Table 1
Influence of pollution on PV performance

No.	Irradiance [W/m^2]	Temperature [°C]	Voltage [V]	Short circuit current [A]	Ic STC	Pollution [g/m^2]	Power considering irradiance [kW]
1	128	16.6	37.9	1.13	8.9	0	0.335
2	125	15.4	38.7	1.23	9.91	0	0.381
3	137	15.4	38.8	1.29	9.48	25	0.365
4	140	15.1	38.8	1.28	9.09	50	0.355
5	143	15.4	38.8	1.28	9.00	75	0.347
6	145	15.0	38.8	1.26	8.73	100	0.337
7	146	14.7	38.8	1.22	8.40	125	0.324
8	146	14.6	38.8	1.14	7.83	150	0.303
9	145	14.3	38.7	1.06	7.37	175	0.283
10	142	14.7	38.6	1.03	7.35	200	0.280
11	137	14.9	38.4	0.95	6.93	225	0.266
12	132	15.1	38.3	0.87	6.66	250	0.252
13	127	15.3	38.2	0.83	6.55	275	0.250

shunt resistances. The current-voltage relationship of the PV module was obtained by numerically solving the implicit diode equation (1):

$$I = I_L - I_0\left(e^{\frac{V+IR_S}{nN_SV_{th}}} - 1\right) - \frac{V+IR_S}{R_{sh}}, \tag{1}$$

which was implemented using an iterative Newton-Raphson scheme by the authors. The model parameters (photocurrent, saturation current, diode ideality factor, series and shunt resistances) were calibrated to match the nominal characteristics of the tested 250 W module. Contamination was represented as a proportional reduction of the photocurrent, reflecting the decrease in effective irradiance on the cell surface. The hot-spot scenario was simulated by increasing the series resistance to emulate the effect of a locally overheated or damaged cell. This modelling approach is widely used in photovoltaic analysis and enables the full I-V and P-V characteristics to be evaluated under different degradation conditions.

Table 2 describes tests illustrating the impact of various types of dirt, such as sand, soil, excrement, and natural dirt from dusty trees. This chart is for illustrative purposes only, to show how important it is to forecast the impact of these factors on planned maintenance work for photovoltaic farms.

Table 2
Data on irradiance and power

No.	Irradiance	Power [kW]	Description
1	983	0.242	at an angle of 30 degrees
2	981	0.318	natural dirt
3	997	0.323	clean with droplets, after washing
4	998	0.317	clean, dry
5	1003	0.313	mist of water
6	1002	0.274	ash
7	1006	0.320	partial ash
8	1015	0.189	large amount of sand
9	1015	0.264	small amount of sand
10	1020	0.316	remnants of sand after rain
11	1005	0.324	dusting with clay
12	1005	0.323	simulation of small droppings
13	1016	0.298	simulation of large droppings
14	1039	0.257	small amount of black soil
15	1045	0.317	single soil contaminations
16	1069	0.311	sand 50 g/m^2
17	1020	0.313	sand 75 g/m^2
18	1080	0.313	clean, shaded pointwise

Simulation tests were also performed in the MATLAB environment (2024b) to illustrate (Fig. 5) the effect of various contamination on the current-voltage characteristics over the entire range [21]. The measurement data described above was used to perform the simulations [22]. Similar results can be observed in the work of other scientists [8].

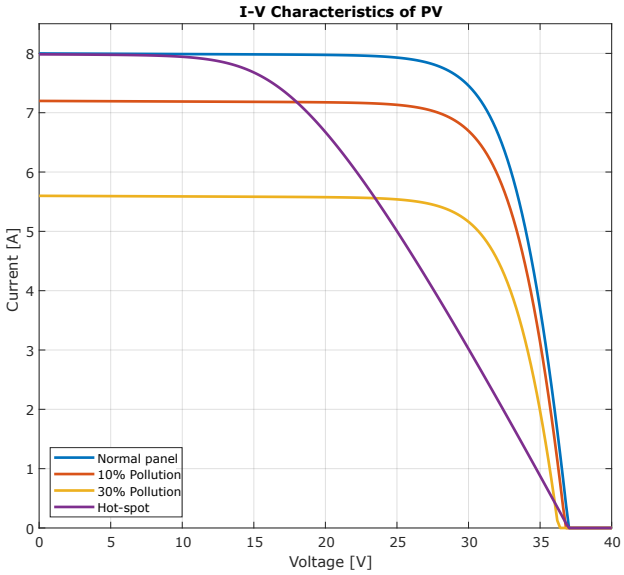

Fig. 5. Simulated I–V characteristics of PV panel

2.2. System-level modeling strategy

Unlike studies that focus solely on optimizing a standalone predictive model, the proposed methodology adopts a system-oriented modeling strategy. The LSTM predictor is embedded within a unified processing pipeline comprising synchronized multi-source data ingestion, deterministic feature engineering, lag-structure generation, cyclic temporal encoding, standardized normalization, and automated evaluation procedures. This design ensures methodological consistency across all compared algorithms (LSTM, RF, SVR, CatBoost, k-NN), thereby eliminating variability introduced by heterogeneous preprocessing strategies. By enforcing identical input representations and validation splits, the proposed framework isolates the true modeling capacity of sequential and non-sequential learners. Consequently, the contribution lies not only in predictive performance, but in establishing a reproducible and controlled benchmarking environment suitable for operational deployment.

2.3. Structured feature-configuration framework

To systematically evaluate the influence of information structure on forecasting performance, a variant-based feature configuration framework was introduced. Instead of relying on a single predefined input set, multiple pollutant–meteorological–energy variable combinations were constructed and evaluated under identical temporal conditions. This controlled feature-variant strategy enables:

- sensitivity analysis of the predictor to exogenous variables,
- identification of dominant information contributors,
- robustness assessment across heterogeneous environmental scenarios.

Such structured exploration of the feature space extends beyond conventional model comparison studies, where the input configuration is typically fixed. The proposed approach therefore contributes at the modeling-strategy level by explicitly quantifying how input structure interacts with temporal learning mechanisms.

2.4. Model selection rationale (LSTM vs. SOTA models)

Although Transformer-based architectures, Temporal Convolutional Networks, and N-BEATS models have demonstrated strong performance in large-scale time-series benchmarks, their advantages are most evident in high-volume datasets and computationally intensive training regimes. In the considered medium-sized environmental dataset (4 296 hourly observations), LSTM provides a favorable balance between predictive accuracy, computational efficiency, and deployment feasibility. Given that pollution forecasting represents one component of a broader IoT–AI operational system, model selection was guided by robustness and scalability rather than exhaustive architectural optimization.

3. DATA ACCURACY

Raw data was recorded as character strings containing sensor identifiers and corresponding measurement values, e.g., AP, 0003, 2 = 2, 4 = 581, 6 = 149, 8 = 50... Before the measurement data could be used to train predictive models, it had to be properly prepared [23]. The data from the sensors and measuring stations contained measurement errors, incorrect or missing values, which had to be removed or replaced. This was especially true for incomplete readings or transmission disruptions during flight [24]. In order for the data to be used effectively by machine learning models, it was standardized or normalized. This included unifying measurement units and aligning value ranges, which facilitates neural network processing and reduces the impact of scale variability. The dataset was then divided into three subsets—training, validation, and testing—to enable effective model training, hyperparameter tuning, and objective performance evaluation, for example, in tasks such as air pollution classification or environmental hazard prediction [25].

The dataset covers the period from October 2025 to March 2026 and consists of 4 296 hourly observations collected from two environmental monitoring stations. Each record contains 18 input features, including three temporal variables, nine air pollution indicators, and six atmospheric parameters. No missing or erroneous measurements were detected in the dataset; therefore, no imputation procedures were required. Basic descriptive statistics (minimum, maximum, mean, and standard deviation) were calculated for the five target pollutants (PM10, PM2.5, NO_2, CO, and O_3), which are predicted based on all input features, and are reported in Table 3. These statistical values were calculated based on the entire data set. Temporal variables exhibiting cyclical behavior (e.g., hour of day, day of month) were encoded using sine and cosine transformations to preserve periodic continuity. The dataset was divided into three mutually exclusive subsets: 70% training, 10% validation, and 20% testing data (859 records). A forward-chaining (time-series split) strategy was applied to preserve chronological order and prevent information leakage from future observations. To ensure strict control over data leakage, all preprocessing operations, including normalization and sequence construction for machine learning and LSTM models, were fitted exclusively on the training dataset and subsequently applied to the validation and test sets. Continuous features were normalized using linear Min–Max scaling to the interval [0, 1]. Cyclical sine/cosine features were not further scaled, as their values are naturally bounded.

Table 3
Basic descriptive statistics for the five target pollutants (real data)

Pollut.	Min [μg/m^3]	Max [μg/m^3]	Mean [μg/m^3]	Std [μg/m^3]
PM2.5	0.600	180.27	26.809	18.032
PM10	3.600	194.28	29.693	23.547
O_3	1.060	127.79	36.012	21.096
NO_2	3.175	119.72	21.982	13.224
CO	0.140	41.330	0.497	0.485

List of input (explanatory) factors for the model:

- day of the year,
- day of the week,
- time (hour, minute),
- nitrogen oxides NO_x (μg/m^3),
- carbon monoxide CO (μg/m^3),
- nitrogen oxide NO (μg/m^3),
- nitrogen dioxide NO_2 (μg/m^3),
- particulate matter PM10 (μg/m^3),
- particulate matter PM2.5 (μg/m^3),
- organic chemical compound C_6H_6 (benzene),
- sulfur dioxide SO_2 (μg/m^3),
- ozone O_3 (μg/m^3),
- air temperature (°C),
- wind speed (km/h),
- cloud cover (%),
- pressure (hPa),
- rainfall (mm).

4. AI MODELS EVALUATION

Artificial intelligence tools are used to solve a wide variety of problems. They must therefore be optimized for a specific purpose and data set in order to achieve the best possible results [26–28]. Selection of the optimal machine learning algorithm for a prediction task, e.g., forecasting pollution levels (PM10) based on input data (PM2.5, PM10, temperature, time of day) or detecting environmental anomalies. The selection process is based on multiple simulations, data analysis, and evaluation of model performance in the context of the specific task [29]. The choice of the appropriate algorithm depends on [8]:

- Data type: Continuous (e.g., PM10 measurements) or categorical (e.g., classification of environmental alerts).
- Nature of the task: Regression (prediction of continuous values), classification (assignment to categories) or time series analysis (forecasting over time).

- Data complexity: Linear or non-linear relationships, noise in the data, missing values.
- Operational requirements: Prediction speed, model interpretability, computational resources.
- Evaluation metrics: For regression – MSE (Mean Squared Error), RMSE, MAE; for classification – accuracy, precision, sensitivity, F1-score.

Below are the most commonly used algorithms, their characteristics, and potential applications in the context of pollution prediction from data collected, for example, by drones [17]:

- Linear regression – The model assumes linear relationships between input variables and the target. Simple and interpretable. Prediction of continuous values, e.g., PM10 levels based on PM2.5 and temperature, when the relationships are linear [30].
- Decision trees – A model based on hierarchical decisions, dividing the data space into regions based on feature values. Classification (e.g., pollution alerts: low/medium/high) or regression [31].
- Random Forest – A set of multiple decision trees that averages their predictions, reducing the risk of overfitting. Regression or classification in tasks with non-linear relationships and noise in the data. High accuracy, resistance to overfitting, support for large data sets [32].
- Support Vector Machines (SVM) – A model that maximizes the separation margin between classes (classification) or fits a hyperplane (regression). Classification (e.g., anomaly detection) or regression in small and medium-sized data sets. Anomaly detection in drone data (e.g., sudden increase in PM2.5) [33].
- Convolutional Neural Networks (CNN) – Specialize in analyzing spatial data, such as images or grid data. Processing visual data from drones (e.g., images of pollution, damage) [34].
- Lasso regression – An extension of linear regression with L1 regularization, eliminating less important features. Regression in data with many variables, some of which may be irrelevant.
- Holt method (Holt-Winters) – A time series forecasting method that takes into account trends and seasonality. Forecasting pollution in data with clear time patterns (e.g., daily or weekly cycles). Predicting daily PM10 fluctuations in cities.

4.1. Model selection process

- Data analysis: Determining the type of data (continuous, categorical, time series) and its characteristics (linearity, noise, missing values).
- Model testing: Conducting experiments with different algorithms on a validation set, comparing metrics (e.g., RMSE for regression, F1-score for classification).
- Hyperparameter tuning: Use techniques such as Grid Search or Random Search for the best models.
- Evaluation on a test set: Select the model with the best performance and generalization on unseen data.
- Consideration of operational context: For drones, prediction speed (e.g., real-time) and resilience to changing conditions may be critical [16].

Our method is based on an LSTM model structure specifically designed for time series forecasting, taking into account complex temporal dependencies. The key element of our approach is a multi-layered LSTM architecture that combines carefully selected components, ensuring both the ability to capture long-term dependencies in the data and resistance to overfitting [16, 17].

The model structure consists of the following layers:

- The first LSTM layer with a 'tanh' activation function, returning sequences, which allows the transmission of complete temporal information to subsequent layers.
- A Dropout layer following the first LSTM layer, which randomly deactivates a portion of neurons, reducing the risk of overfitting and improving model generalization.
- A second LSTM layer, also with a 'tanh' activation function, but with half the number of units, enabling efficient compression of information and further processing of temporal features.
- Another Dropout layer, reinforcing model regularization.
- Two Dense layers with 'tanh' activation, which gradually reduce the dimensionality of the data, preparing it for the final prediction.
- A final Dense layer, generating a single output value corresponding to the predicted target variable.

This structure was optimized using a systematic grid search, testing various configurations of the number of neurons, dropout rates, optimizers, and other parameters. The model incorporates input data such as weather parameters and pollutants, enhancing its ability to model complex environmental dependencies.

All parameters were searched using the grid method. List of parameters of the Author's model, search ranges in the form of a list of values:

- Number of delayed factor values (`n_steps`): [1, 2, 3, 6, 12, 24]
- Number of neurons in the LSTM layer (`units`): [32, 64, 100, 128, 200, 256, 512]
- Number of training epochs: [60, 100]
- Batch size: [8, 16, 32, 64, 128]
- Dropout rate (`dropout_rate`): [0.0, 0.1, 0.2]
- Optimizer: [`adam`, `rmsprop`, `sgd`]

4.2. Metrics

When evaluating the effectiveness of predictive models, it is crucial to use appropriate quality metrics. Mean Square Error (MSE) is commonly used to evaluate model accuracy, Mean Absolute Error (MAE) provides a more intuitive interpretation of results, and Mean Absolute Percentage Error (MAPE) is particularly useful for comparing models. Mean Squared Error computes the average squared difference between actual and predicted values. It penalizes larger errors more heavily – equation (2).

$$\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 . \tag{2}$$

Mean Absolute Error calculates the average absolute difference between actual and predicted values. It provides better unit in-

terpretability (e.g., µg/m^3 for pollution measurements) – equation (3).

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i|\,. \tag{3}$$

Mean Absolute Percentage Error expresses error as a percentage of the actual value, facilitating comparisons across models and scales. It is useful for evaluating accuracy across different models but can be sensitive to very small actual values, potentially leading to high MAPE – equation (4).

$$\mathrm{MAPE} = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \times 100\%. \tag{4}$$

The Relative Percentage Error (RPE) measures the total absolute difference between predicted values y_i and actual values y_i as a percentage of the total actual values. It is commonly used to evaluate the accuracy of forecasting or regression models—lower values indicate better performance – equation (5).

$$\mathrm{RPE} = \frac{\sum|\hat{y}_i - y_i|}{\sum|y_i|} \times 100\%\,. \tag{5}$$

4.3. Time-series cross-validation

To assess the robustness of the forecasting model under temporal constraints, a five-fold time-series cross-validation procedure based on a forward-chaining strategy was applied. The dataset was iteratively split into sequential training and validation folds, ensuring that each validation subset strictly followed the corresponding training data in time. This approach provides a more realistic evaluation for time-dependent environmental data and prevents information leakage from future observations. The cross-validation results for the LSTM model across all target pollutants are reported in Table 4.

The results presented in Table 4 demonstrate consistent predictive performance across all analyzed pollutants under five-fold time-series cross-validation. For PM10, the model achieves an MAE of 4.890 ± 1.773 µg/m^3 and an R^2 of 0.804 ± 0.106, indicating a stable and satisfactory level of explanatory power under temporal validation constraints. The relative error measures (MAPE = 17.817% ± 3.791; RPE = 15.915% ± 3.684) remain within an acceptable range for environmental forecasting tasks characterized by inherent variability. Similarly, PM2.5 exhibits strong performance (0.817 ± 0.163) with relatively low dispersion of error metrics, indicating good temporal generalization. NO_2 shows slightly higher variability (R^2 = 0.766 ± 0.162), yet maintains stable error levels across folds. For O_3, the model achieves an MAE of 4.74 ± 0.506 µg/m^3 and an R^2 of 0.812 ± 0.143, confirming comparable explanatory power to particulate pollutants. Although the relative percentage error is higher (MAPE = 25.805% ± 4.841), this is consistent with the inherently greater temporal variability and nonlinear atmospheric dynamics of ozone. Importantly, the low standard deviation of MAE confirms stable performance across validation folds, while the second relative error metric (RPE = 14.016% ± 2.379) remains within an acceptable range. Overall, the relatively small standard deviations across metrics confirm stable generalization performance under forward-chaining validation. The results indicate that the proposed forecasting framework maintains robustness across multiple pollutant types despite differing physical and chemical behavior patterns.

5. RESULTS

The authors' method achieves an optimal balance between model complexity and generalization capability, as confirmed by minimizing MAE, MAPE, RPE, MSE errors and maximizing the R^2 coefficient compared to other publications [15–17, 29, 32, 35]. The effectiveness of the method is demonstrated through air pollution forecasting, as shown in Tables 6 and 7. They are also compared in graphs on Figs. 6 and 7. The author's model best parameters for five input type was shown in Table 7.

The author's model best parameters for five input types was shown in Table 7. The comparative results of the proprietary algorithm with several other selected algorithms (CatBoost, KNN, RF, and SVM) for the selected example of PM10 particulate matter are presented in Table 5.

To verify the statistical significance of the observed performance differences, bootstrap-based hypothesis testing was conducted. Five forecasting accuracy metrics were evaluated for each of the five target pollutants (PM10, PM2.5, NO_2, CO, and O_3). For each metric, confidence intervals were estimated using repeated bootstrap resampling, allowing assessment of performance stability and statistical relevance. The detailed results and corresponding interpretations are summarized in Table 8.

Table 4

Time-series cross-validation results for author-developed models (mean ± std for each metric and pollutant)

Pollutant	MAE [µg/m^3]	MAPE [%]	MSE	R^2	RPE [%]
PM2.5	4.092 ± 0.792	19.042 ± 3.875	24.319 ± 5.725	0.817 ± 0.163	15.101 ± 2.449
PM10	4.890 ± 1.773	17.817 ± 3.791	48.757 ± 13.831	0.804 ± 0.106	15.915 ± 3.684
O_3	4.740 ± 0.506	25.805 ± 4.841	50.129 ± 9.370	0.812 ± 0.143	14.016 ± 2.379
NO_2	3.421 ± 0.787	19.105 ± 3.884	34.017 ± 12.169	0.766 ± 0.162	16.423 ± 2.386
CO	0.163 ± 0.077	24.379 ± 3.417	0.030 ± 0.018	0.557 ± 0.147	21.572 ± 3.464

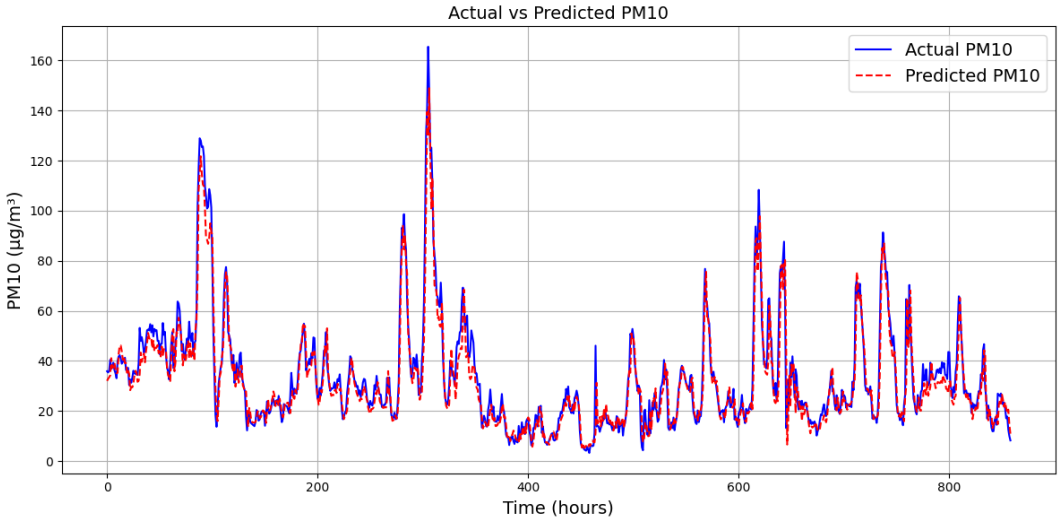

Fig. 6. Prediction results for PM10 pollution using the best model

Table 5
Prediction quality measures for pm10 factor on training data

Model	Factor	MAE [$\mu g/m^3$]	MSE	MAPE [%]	RPE [%]
Custom	PM10	3.411	7.935	13.819	11.482
CatBoost	PM10	4.386	9.957	19.545	16.762
KNN	PM10	4.012	10.035	19.121	16.664
RF	PM10	4.229	11.653	20.277	17.236
SVM	PM10	4.675	12.837	21.571	17.736

Table 6
Error metrics for the best-performing author-developed models for individual pollutants

Pollutant	MAE	MAPE	RPE	MSE	R^2
PM2.5	**3.088**	17.881	**12.537**	**21.398**	**0.936**
PM10	4.482	**14.910**	13.294	43.588	0.913
O_3	5.212	27.909	12.540	56.866	0.918
NO_2	3.982	15.883	15.428	40.795	0.854
CO	**0.09**	19.095	19.445	**0.019**	0.723

Table 7
Parameterization of the best models for individual pollutants

Pollutant	n_steps	units	epochs	batch_size	optimizer
PM2.5	2	64	44	16	`adam`
PM10	2	100	55	8	`rmsprop`
O_3	2	**256**	95	16	`rmsprop`
NO_2	**3**	96	98	**24**	`rmsprop`
CO	2	160	**100**	8	`rmsprop`

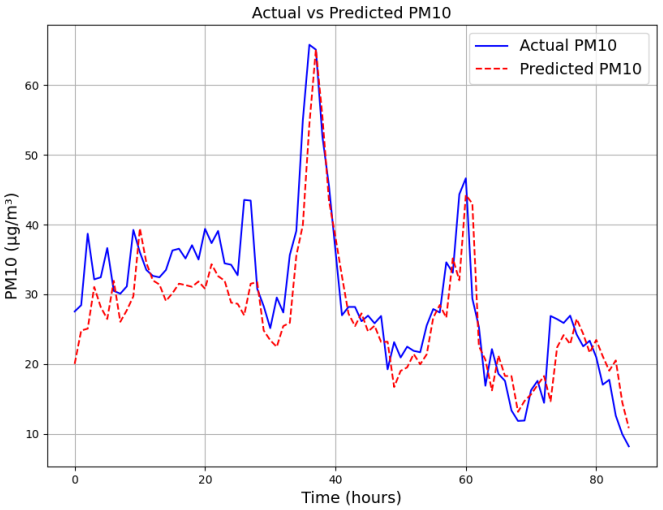

Fig. 7. Zoom of the 10% of PM10 pollution prediction results for the best model

The bootstrap analysis confirms the stability and robustness of the forecasting results across all five target factors. For PM10, error metrics (MAE = 4.482; MAPE = 14.91%. RPE = 13.4294) exhibit relatively narrow confidence intervals, while the coefficient of determination (R^2 = 0.913; CI: 0.835–0.991) indicates strong and stable explanatory power. For PM2.5, lower absolute errors (MAE = 3.982; RPE = 12.537) and a high R^2 value (0.936; CI: 0.883–0.989) further confirm model robustness, with limited variability across bootstrap samples. Although MAPE shows moderate dispersion, it does not materially affect overall model quality. Similar stability patterns were observed for NO_2, CO and O_3. Overall, the narrow confidence intervals across all evaluated metrics demonstrate low estimation variability and confirm the reliability of the proposed multi-factor forecasting framework.

Table 8
Bootstrap confidence intervals and significance test outcomes

Metric	Estimate	CI95%	Std (bootstrap)
PM2.5			
MAE	3.088	[2.987, 3.189]	0.101
MAPE	17.881	[16.838, 18.924]	1.043
RPE	12.537	[12.190, 12.884]	0.347
MSE	21.398	[19.534, 23.262]	1.864
R^2	0.936	[0.883, 0.989]	0.053
PM10			
MAE	4.482	[4.309, 4.655]	0.173
MAPE	14.910	[14.526, 15.294]	0.384
RPE	13.294	[12.934, 13.654]	0.360
MSE	43.588	[39.566, 47.610]	4.022
R^2	0.913	[0.835, 0.991]	0.078
O_3			
MAE	5.212	[5.045, 5.379]	0.167
MAPE	27.909	[26.416, 29.402]	1.493
RPE	12.540	[12.147, 12.933]	0.393
MSE	56.866	[52.724, 61.008]	4.142
R^2	0.918	[0.8604, 0.9756]	0.0576
NO_2			
MAE	3.982	[3.840, 4.124]	0.142
MAPE	15.883	[15.370, 16.396]	0.513
RPE	15.428	[14.934, 15.922]	0.494
MSE	40.795	[36.782, 44.808]	4.013
R^2	0.854	[0.8408, 0.8672]	0.0132
CO			
MAE	0.090	[0.0869, 0.0931]	0.0031
MAPE	19.095	[18.478, 19.712]	0.617
RPE	19.445	[18.478, 19.712]	0.572
MSE	0.019	[0.0172, 0.0208]	0.0018
R^2	0.723	[0.7020, 0.7440]	0.0210

6. DISCUSSION

6.1. Scope and limitations

Ablation studies and exhaustive comparisons with state-of-the-art forecasting architectures were not included in the present work. While such analyses are valuable in model-centric studies, the primary objective of this paper is the integration of environmental forecasting within an IoT-enabled operational framework for photovoltaic systems. The forecasting module is designed to support system-level decision-making rather than to serve as an isolated benchmark model. Future work will explore architectural extensions and comparative evaluations as part of targeted model optimization studies.

6.2. Operational impact of improved pollution forecasts

The proposed IoT–AI architecture focuses on forecasting airborne particulate matter (PM10), which indirectly affects photovoltaic (PV) performance through soiling-induced efficiency degradation. While the quantitative example focuses on PM10, the proposed framework simultaneously models PM2.5, NO_2, CO, and O_3, which may further enhance operational decision support under complex pollution conditions. Although the system does not directly predict PV power output, its practical value lies in enabling optimized operational and maintenance (O&M) strategies. The LSTM model reduced the mean absolute error of PM10 prediction from 4.2 to 3.4 $\mu g/m^3$ compared to Random Forest ($\Delta MAE = 0.8\ \mu g/m^3$). Improved forecasting precision increases the reliability of detecting high-risk pollution events and allows proactive cleaning scheduling. Soiling losses in PV systems are commonly reported in the range of 2–10%, depending on site-specific environmental conditions. To quantify the potential operational benefit, consider a 100 MW PV plant with an average daily energy yield of 800 MWh. A conservative 1% loss due to soiling corresponds to 8 MWh per day. If enhanced PM10 forecasting reduces these losses by 10% (i.e., recovery of 0.1% of total daily production), the energy gain is approximately: 0.8 MWh/day ≈ 292 MWh/year. At an electricity price of 0.4 PLN/kWh, this corresponds to approximately 120 000 PLN annually. This analysis demonstrates that even moderate improvements in environmental prediction accuracy can translate into significant economic benefits at utility scale. Therefore, the proposed IoT–AI framework provides tangible operational value beyond standard machine learning performance metrics. It should be noted that the presented economic assessment represents a conservative lower-bound estimate, as it considers only the direct conversion of recovered energy into market value based on average electricity prices. In practice, forecasting improvements may generate additional benefits through reduced imbalance penalties, improved day-ahead scheduling reliability, and enhanced coordination with energy storage systems. Therefore, the actual operational impact of improved PM10 prediction accuracy may exceed the values reported in this study.

7. CONCLUSIONS

Regular monitoring of PV panels using IoT and artificial intelligence not only increases the efficiency of installations, but also extends their service life by minimizing the risk of damage caused by contamination. The implementation of advanced diagnostic technologies, such as intelligent thermal imaging analysis, is a key element of modern photovoltaic system management and increases the energy efficiency of renewable energy sources.

Over a 5% improvement in anomaly detection accuracy was achieved for current methods based on the average of scientific publications. A proprietary algorithm was developed and compared with known algorithms (e.g., KNN, SVM, CatBoost, Random Forest) in terms of prediction accuracy. The analysis and development of the algorithms was carried out with a view to processing data from air pollution sensors, temperature sensors, etc.

Authors' method, based on a proprietary structure and selection of deep LSTM network parameters, outperforms other specified machine learning methods in terms of relative pre-

diction accuracy (MAPE, RPE metrics) for PM10 dust. Our method (algorithm) also predicts PM10 absolute values (MAE and MSE metrics) more accurately than other machine learning methods. Authors' method also predicts other pollutants with high accuracy.

The presented results and developed algorithms can be used, after appropriate adaptation, in tools supporting the management of photovoltaic farms in terms of electricity production efficiency in relation to air quality and the amount of pollutants. This can bring measurable benefits in terms of a significant improvement in efficiency and reduction of energy production losses. This is very important in view of the largest development of the PV sector in Europe in Poland, although the solution is universal and can be applied on a global scale. Accurate prediction of pollution and related production losses can enable the planning of appropriate service or maintenance actions, ensuring high efficiency and reliability of RES sources. The presented results are therefore not only of a research and scientific nature, but also have practical implications for the renewable energy sector.

ACKNOWLEDGEMENTS

This scientific publication resulted from the project carried out under the grant programme for R&D works of scientific entities, Call IV, Competition No. 1/2024 Project title "Ewaluacja rozwiązań technologii Internet of Things i AI w kontekście monitorowania zanieczyszczeń środowiska", Grant agreement number 2/PRZ/1/DP/PCI/2024 – co-financed under the project "Podkarpackie Center for Innovation 2029" implemented under the regional programme European Funds for Podkarpacie 2021–2027, priority: FEPK.01 Competitive and digital economy, measure: FEPK.01.01, Research and development, Type of project: Building the potential of innovation centers Podkarpackie Center for Innovation. Publication co-funded by the Ministry of Science and Higher Education of the Republic of Poland to maintain the research potential of the disciplines of Technical Informatics and Telecommunications.

REFERENCES

[1] S. Sai, A. Garg, K. Jhawar, V. Chamola, and B. Sikdar, "A comprehensive survey on artificial intelligence for unmanned aerial vehicles," *IEEE Open J. Veh. Technol.*, vol. 4, pp. 713–738, 2023, doi: 10.1109/OJVT.2023.3316181.

[2] A. Puente-Castro, D. Rivero, A. Pazos, and E. Fernandez-Blanco, "A review of artificial intelligence applied to path planning in uav swarms," *Neural Comput. Appl.*, vol. 34, no. 1, pp. 153–170, 2022, doi: 10.1007/s00521-021-06569-4.

[3] S. Osowski and F. Gil, "Melanoma recognition using an ensemble of deep cnn structures," *Bull. Pol. Acad. Sci. Tech. Sci.*, vol. 72, no. 6, p. e151675, 2024, doi: 10.24425/bpasts.2024.151675.

[4] F. Haghshenas, A. Krzyżak, and S. Osowski, "Comparative study of deep learning models in melanoma detection," in *IAPR Workshop on Artificial Neural Networks in Pattern Recognition*. Springer, 2024, pp. 121–131, doi: 10.1007/978-3-031-71602-711.

[5] V. Soni, D.P. Bhatt, and N.S. Yadav, "Hais-ids: A hybrid artificial immune system model for intrusion detection in iot," *Bull. Pol. Acad. Sci. Tech. Sci.*, vol. 73, p. e152211, 2025, doi: 10.24425/bpasts.2024.152211.

[6] T. Kaczorek, J. Klamka, and A. Dzieliński, "Descriptor continuousand discrete-time linear systems with zero transfer matrices," *Bull. Pol. Acad. Sci. Tech. Sci.*, vol. 73, p. e152710, 2025, doi: 10.24425/bpasts.2024.152710.

[7] T. Ding, W. Zhong, and Y. Cai, "Multi-unmanned aerial vehicle odor source location based on improved artificial fish swarm algorithm," *Bull. Pol. Acad. Sci. Tech. Sci.*, vol. 72, no. 4, p. e150329, 2024, doi: 10.24425/bpasts.2024.150329.

[8] G.B. Balachandran, M. Devisridhivyadharshini, M.E. Ramachandran, and R. Santhiya, "Comparative investigation of imaging techniques, pre-processing and visual fault diagnosis using artificial intelligence models for solar photovoltaic system – a comprehensive review," *Measurement*, vol. 232, p. 114683, 2024, doi: 10.1016/j.measurement.2023.114683.

[9] C. Prasshanth, S.B. Narayanan, N.V. Sridharan, and S. Vaithiyanathan, "Fault detection in photovoltaic systems using unmanned aerial vehicle-captured images and rough set theory," *Sol. Energy*, vol. 290, p. 113348, 2025, doi: 10.1016/j.solener.2024.113348.

[10] S. Bhatia, S. Sachdeva, and P. Goswami, "Air pollution prediction and hotspot detection using machine learning," *J. Stat. Manage. Syst.*, vol. 25, no. 7, pp. 1553–1564, 2022, doi: 10.48175/IJARSCT-27774.

[11] C. Wu, S. Lu, J. Tian, L. Yin, L. Wang, and W. Zheng, "Current situation and prospect of geospatial ai in air pollution prediction," *Atmosphere*, vol. 15, no. 12, p. 1411, 2024, doi: 10.3390/atmos15121411.

[12] R. Ravi, N.S. Kumari, P. Geethika, K.V. Rao, and M.S. Rao, "Air pollution forecasting using deep learning algorithms: A review," in *XVIII International Conference on Data Science and Intelligent Analysis of Information*. Springer, 2023, pp. 511–517, doi: 10.1007/978-3-031-51167-749.

[13] A. Rogalski, "Reconfigurable, non-volatile neuromorphic photovoltaics," *Light-Sci. Appl.*, vol. 12, no. 1, p. 185, 2023, doi: 10.1038/s41565-023-01446-8.

[14] O. Alvear *et al.*, "A discretized approach to air pollution monitoring using uav-based sensing," *Mob. Netw. Appl.*, vol. 23, pp. 1693–1702, 2018, doi: 10.1007/s11036-018-1065-4.

[15] T.M. Lei, S.W. Siu, J. Monjardino, L. Mendes, and F. Ferreira, "Using machine learning methods to forecast air quality: A case study in macao," *Atmosphere*, vol. 13, no. 9, p. 1412, 2022, doi: 10.3390/atmos13091412.

[16] Q. Liu, B. Cui, and Z. Liu, "Air quality class prediction using machine learning methods based on monitoring data and secondary modeling," *Atmosphere*, vol. 15, no. 5, p. 553, 2024, doi: 10.3390/atmos15050553.

[17] W. Mao, W. Wang, L. Jiao, S. Zhao, and A. Liu, "Modeling air quality prediction using a deep learning approach: Method optimization and evaluation," *Sustain. Cities Soc.*, vol. 65, p. 102567, 2021, doi: 10.1016/j.scs.2020.102567.

[18] A. Farahbakhsh, D. Zarifi, and M. Mrozowski, "Design of mmwave broadband rotary joint and 360° beam-steering rotenna based on gap waveguide technology," *IEEE Trans. Antennas Propag.*, vol. 73, pp. 4373–4383, 2025, doi: 10.1109/TAP.2025.3552222.

[19] R. Szmurło and S. Osowski, "Relabeling the imperfect labeled data to improve recognition of face images using cnn." *Prz. Elektrotechniczny*, vol. R100, no. 6, pp. 27–30, 2024, doi: 10.15199/48.2024.06.05.

[20] A. Di Tommaso, A. Betti, G. Fontanelli, and B. Michelozzi, "A multi-stage model based on yolov3 for defect detection in pv panels based on ir and visible imaging by unmanned aerial vehicle," *Renew. Energy*, vol. 193, pp. 941–962, 2022, doi: 10.1016/j.renene.2022.05.093.

[21] J. Fu *et al.*, "Non-volatile and gate-controlled multistate photovoltaic response in wse2/h-bn/graphene semi-floating gate field-effect transistors," *Adv. Opt. Mater.*, vol. 12, no. 21, p. 2400638, 2024, doi: 10.1002/adom.202400638.

[22] MathWorks, "Deep learning toolbox – matlab – mathworks," 2024, accessed: 2025-05-09. [Online]. Available: https://www.mathworks.com/products/deep-learning.html

[23] B. Blachowski and N. Pnevmatikos, "Neural network based vibration control of seismically excited civil structures," *Period. Polytech. Civ. Eng.*, vol. 62, no. 3, pp. 620–628, 2018, doi: 10.3311/PPci.11601.

[24] L. Vračar, D. Marinković, M. Stojanović, and M. Milovančević, "Automated inspection system with gps and deep learning in urban rail safety and efficiency," *Acta Polytech. Hung.*, vol. 22, no. 4, 2025, doi: 10.12700/APH.22.4.2025.4.2.

[25] J. Kabziński, M. Jasiński, and J. Rąbkowski, "New control algorithms, modern devices, and materials in electric energy conversion systems," *Bull. Pol. Acad. Sci. Tech. Sci.*, vol. 72, p. e151676, 2024, doi: 10.24425/bpasts.2024.151676.

[26] F. Mohsen, H.R. Al-Absi, N.A. Yousri, N. El Hajj, and Z. Shah, "A scoping review of artificial intelligence-based methods for diabetes risk prediction," *NPJ Digit. Med.*, vol. 6, no. 1, p. 197, 2023, doi: 10.1038/s41746-023-00933-5.

[27] I. Felde, "Artificial intelligence techniques and biomimetic methods supporting heat treatment processes," *Acta Polytech. Hung.*, vol. 21, no. 10, 2024, doi: 10.12700/APH.21.10.2024.10.20.

[28] S. Sayah, M. Birane, and K. Benmouiza, "Deep learning-based control perspective for single-phase grid-connected inverter using gated recurrent units," *Acta Polytech. Hung.*, vol. 22, no. 3, pp. 269–287, 2025, doi: 10.12700/APH.22.3.2025.3.14.

[29] M. Castelli, F. M. Clemente, A. Popovič, S. Silva, and L. Vanneschi, "A machine learning approach to predict air quality in California," *Complexity*, vol. 2020, no. 1, p. 8049504, 2020, doi: 10.1155/2020/8049504.

[30] J. Łęski and N. Henzel, "Generalized ordered linear regression with regularization," *Bull. Pol. Acad. Sci. Tech. Sci.*, vol. 60, no. 3, pp. 481–489, 2012, doi: 10.2478/v10175-012-0061-2.

[31] R. Iranzad and X. Liu, "A review of random forest-based feature selection methods for data science education and applications," *Int. J. Data Sci. Anal.*, vol. 20, p. 197–211, 2025, doi: 10.1007/s41060-024-00509-w.

[32] A. Setiawan, U.L. Wibowo, A. Mubarok, K. Larasati, and J.A. Hammad, "Random forest algorithm to measure the air pollution standard index," *Knowl. Eng. Data Sci.*, vol. 7, no. 1, pp. 86–100, 2024, doi: 10.17977/um018v7i12024p86-100.

[33] A. Shmilovici, "Support vector machines," in *Data mining and knowledge discovery handbook*. Springer, 2010, pp. 231–247, doi: 10.1007/978-0-387-09823-412.

[34] L. Alzubaidi *et al.*, "Review of deep learning: concepts, cnn architectures, challenges, applications, future directions," *J. Big Data*, vol. 8, p. 53, 2021, doi: 10.1186/s40537-021-00444-8.

[35] A. Mathew *et al.*, "Air quality analysis and $PM_{2.5}$ modelling using machine learning techniques: A study of Hyderabad city in India," *Cogent Eng.*, vol. 10, no. 1, p. 2243743, 2023, doi: 10.1080/23311916.2023.2243743.