ARCHIVES OF TRANSPORT | *ISSN (print):* 0866-9546
Volume 75, Issue 3, 2025 | *e-ISSN (online):* 2300-8830
DOI: *10.61089/aot2025.w98msn71*

ANALYSIS OF FACTORS AFFECTING THE PERFORMANCE OF AN AVIATION SYSTEM USING SELECTED MODELS

Piotr GŁADYSZ[1], Jerzy MERKISZ[2], Anna BORUCKA[3]

[1] *Doctoral School of Poznan University of Technology, Poznan University of Technology, Poland*

[2] *Faculty of Civil and Transport Engineering, Poznan University of Technology, Poland*

[3] *Faculty of Logistics, Military University of Technology, Warsaw Poland*

Abstract:

The aim of this study is to quantitatively analyze the factors influencing energy consumption in unmanned aerial vehicles (UAVs) based on operational data from 177 flights of the DJI Matrice 30T drone. Battery consumption modelling was proposed using variables available at the UAV user level. A comparison of analytical methods (linear regression, LASSO) and machine learning algorithms (Random Forest, XGBoost) was performed. The models were then evaluated using the coefficient of determination R^2 and the root mean square error RMSE. Analytical methods show moderate effectiveness (R^2 = 0.425, RMSE = 14.87%), while machine learning models show significantly higher predictive accuracy: Random Forest achieved R^2 = 0.983 and RMSE = 0.328%, and XGBoost R^2 – 0.973 and RMSE = 3.26%. The analysis of variable significance shows that the greatest impact on energy consumption is exerted by: flight time, distance traveled, and discharge current. Seasonal factors also proved to be significant, indicating the impact of weather conditions on battery discharge dynamics. The results confirm the superiority of adaptive machine learning methods over classical analytical models in forecasting UAV energy consumption based on operational data and indicate the direction for further research taking into account detailed meteorological data. Unlike previous studies, this study is based on operational data from actual UAV missions and uses only variables available from the user's perspective. It also provides a methodical comparison of analytical approaches and machine learning algorithms on a single real-world flight log dataset and additionally considers the impact of seasonality and operating conditions on battery consumption – an aspect largely overlooked in the literature.

Keywords: *UAV, battery, energy consumption, energy management*

To cite this article:

Gładysz, P., Merkisz, J., Borucka, A. (2025). Analysis of factors affecting the performance of an aviation system using selected models. Archives of Transport, 75(3), 25-39. https://doi.org/10.61089/aot2025.w98msn71

Contact:
1) phddron@gmail.com [https://orcid.org/0000-0002-4676-2006] – corresponding author; 2) jerzy.merkisz@put.poznan.pl [https://orcid.org/0000-0002-1389-0503; 3) anna.borucka@wat.edu.pl [https://orcid.org/0000-0002-7892-9640]

1. Introduction

A key parameter in all processes involving unmanned aerial vehicles is flight time. It is expected to be as long as possible to enable the efficient execution of missions such as monitoring, inspections, deliveries, or rescue operations. Long flight times allow for greater coverage, fewer landings, and increased operational efficiency. This is particularly important in sectors such as the military, agricultural, surveying, as well as rescue and search missions, where every minute can be critical. Achieving maximum flight time depends on many factors, and engineers are constantly working on new technologies, particularly advanced batteries, to extend drones operational time and make them more autonomous. Therefore, battery life modeling and consequently, the flight duration of unmanned aerial vehicles is a key aspect of numerous scientific studies, which are gradually refined and updated as technology advances and UAV capabilities improve. This makes it an important research area in aviation.

The authors, considering the importance of flight time as the main parameter of flight operations, reviewed the literature and formulated the following research objectives, which constitute a significant contribution to the development of knowledge in the field of UAVs.

1. Developing a battery consumption model using data accessible to any user.
2. Comparing the effectiveness of analytical methods and machine learning, providing insights into the application of various modeling techniques by analyzing their accuracy in predicting battery consumption.
3. Conducting a detailed analysis for a single UAV. Establishing a solid foundation for future meta-analyses, enabling broader comparisons and generalizations of findings.
4. Identifying factors affecting battery lifespan under different operational conditions.

The structure of the article is as follows. After the introduction, which shows the main research assumptions, a literature review is presented, followed by a separate section describing the methods used. Next the UAV under study and the collected data, which served as the basis for empirical research, are discussed. The final sections provide an analysis of the results, a summary of the research, and suggestions for future research directions.

The novelty of this study manifests itself in three key aspects. First, the analysis is based on actual UAV operational data (over 2.2 million records from 177 flights), disregarding the laboratory conditions typical for simulation work. Second, unlike previous publications focusing on single modeling approaches, this article directly compares analytical models (LR, LASSO) and machine learning algorithms (RF, XGBoost) to assess the trade-off between interpretability and predictive accuracy. Third, the study takes into account operational seasonal conditions and variables available only from the perspective of a UAV user, which increases the potential of practical application of the results in mission planning and energy management.

2. Literature review

The essence of the problem with modeling battery life and consequently the flight time of unmanned aerial vehicles is highlighted in scientific research that is gradually being refined and updated as new technologies and improvements in drone capabilities emerge.

Among the pioneering studies on UAV battery life modeling, Traub's (2011) article is noteworthy, as it introduced the first analytical models of energy consumption for small, unmanned platforms. A few years later, Shibl et al. (2023) expanded on these models, emphasizing the necessity of considering various mission profiles and changing environmental conditions, which aligns with the scope of this study. Research on drone flights in different atmospheric conditions has also been proposed in (Kim et al., 2018; Szczupak et. al., 2025), where a comprehensive mathematical model was developed, considering battery type and flight characteristics. In (Di Franco & Buttazzo, 2015), route planning algorithms were presented, incorporating energy consumption, terrain topography, and weather conditions, while in (Karunarathne et al., 2012), a real-time energy management system was introduced, dynamically adjusting flight parameters in response to current energy consumption and battery status. Empirical validation of theoretical models based on a series of tests with various battery configurations and loads is the result of scientific considerations presented in (Gong et al., 2023), where the authors focused on assessing deviations between theoretical models and actual flight measurements. Many studies address batteries themselves and their optimal operating conditions, e.g., (Luo et al., 2025; Poorani et al., 2025; Josephin Shermila et al., 2025; Li et al., 2024).

In the literature, it is also evident that the dynamic development of unmanned systems is accompanied by the increasingly frequent use of machine learning techniques as an alternative to traditional models (Ziółkowski et al., 2024; Chen et al., 2024). Of course, each of these techniques has its advantages and limitations. Traditional mathematical models typically offer better interpretability and compliance with established physical principles, which significantly enhances their practical usefulness. On the other hand, machine learning techniques excel in detecting complex patterns and dependencies in data, which is particularly valuable in case of numerous and dynamic changes. Therefore, it is crucial to find the right balance between these approaches, and this concept guided the authors of this study.

Despite extensive research on unmanned aerial vehicles, further studies remain necessary. This is driven not only by the rapid technological advances but also by the need for standardization in this field. This need has been highlighted, for instance, in the work of the European Union Aviation Safety Agency (Bassi, 2019; Wanner et al., 2024) and the U.S. Federal Aviation Administration (FAA) (Ravich, 2019; Barnhart et al., 2021). Moreover, each subsequent study can serve as a foundation for comprehensive meta-analyses that yield valuable insights. For example, the analysis conducted in (Eskandari et al., 2020) which compared the effectiveness of various modeling approaches from classical analytical and empirical models to advanced AI techniques demonstrates the potential of such research. A study encompassing 78 publications from 2015 to 2023 revealed that hybrid approaches combining physical principles with adaptive machine learning algorithms achieve the highest accuracy (on average, 92% alignment with real-world data) while maintaining result interpretability. Therefore, future research should emphasize the creation of standardized benchmarks and openly accessible datasets that enable objective comparison across modeling approaches. Close collaboration among academia, industry, and regulatory bodies will be essential to translate methodological advances into safe and efficient UAV operations. Ultimately, the development of unified framework combining interpretable physics-based models with data-driven adaptability will accelerate the deployment of unmanned aerial vehicles in a variety of real-world applications.

Despite numerous publications on modeling the energy consumption in unmanned aerial vehicles, an analysis of the state of research reveals significant gaps that limit the practical usefulness of existing models. First, most studies focus on highly idealized analytical approaches based on simplified aerodynamic assumptions and constant operating conditions (Traub, 2011; Di Franco & Buttazzo, 2015). These models have limited generalizability in the context of variable mission profiles and random environmental disturbances. Second, there is a growing use of machine learning techniques in the literature (Shibl et al., 2023; Gong et al., 2023), but most of these studies use laboratory data or extended sensor sets that are not available in standard UAV platforms, which significantly limits the possibility of replicating the results in real operating conditions. Third, there is a lack of systematic comparative analyses of analytical methods and AI-based methods conducted on a uniform dataset, which prevents an objective assessment of the trade-off between model interpretability and predictive accuracy (Eskandari et al., 2020). Another research problem is the marginalization of the influence of operational and environmental factors such as seasonality, ambient temperature flight profile, and energy load, despite the significant impact of these parameters on the discharge dynamics of lithium-polymer batteries, as confirmed in the literature (Kim et al., 2018; Li et al., 2021). Furthermore, there is a noticeable lack of studies using massive UAV operational data (flight logs), which enable the identification of statistically significant determinants of energy consumption in real flight missions. This gap justifies the need for research based on operational data, while applying comparative methods with high diagnostic value.

This study addresses the identified gaps by proposing a methodology that integrated analytical models and machine learning algorithms to evaluate their predictive effectiveness under real UAV operating conditions, considering only variables available from the perspective of the system user. The analysis of factors affecting the performance of the aviation system using selected models is significant extension of the existing state of knowledge, as it was carried out on the basis of data obtained from batteries used in UAV missions during rescue operations conducted in diverse environmental and atmospheric conditions. This distinguishes this work from other available scientific studies, as the database is based on authentic, non-laboratory empirical material, which increases its uniqueness and cognitive value.

3. Material and methods:

For the preliminary assessment of dependencies in the model, a linear regression model was used. The regression model is based on examining the correlation between the dependent variable and the predictors (Sang et al., 2025; Grzelak et al., 2022) Parameter estimation involves estimating the parameters of the model and thus analyzing the impact of the predictors on the dependent variable. The parameters of the regression model are estimated using the least squares method (LSM), which allows for identification of trends and seasonal fluctuations in time series. For a linear model, the relationship between the dependent variable and the predictors is represented by the equation (Kozłowski et al., 2024):

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots \beta_k X_k + \varepsilon, \tag{1}$$

where :

Y - dependent variable,

$X_1, X_{2,\dots,} X_k$ - independent variables,

ε - random variable with the distribution $N(0, \sigma^2)$.

The coefficient β_k describe how, on average, the value of the dependent variable y will change if the value of the independent variable x_k, to which they refer, will change by a unit, assuming a fixed level of the other independent variables. The random component in the model reflects an incomplete fit to the empirical data. To test the significance of the multiple correlation coefficient, we use the F-test. A working null hypothesis is created at the significance level α :

H_0: $R = 0$ - no significant effect of at least one of the predictors on the dependent variable,

against the alternative hypothesis:

H_1: $R \neq 0$ - at least one of the predictors has a significant effect on the dependent variable,

If H_0 is true, then the statistic

$$F = \frac{R^2}{1-R^2}\frac{n-k-1}{k} = \frac{\sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2}{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}\frac{n-k-1}{k} \tag{2}$$

has an F distribution with k and $n-k-1$ degrees of freedom. From the tables, we read the quantile of order $1-\alpha$ for the F distribution with $(k,\ n-k-1)$ degrees of freedom. If $F \le F_{1-\alpha}\ (k, n-k-1)$ then at the level of significance α there are no grounds for rejecting H_0, otherwise the hypothesis H_0 is rejected in favor of H_1. To assess the statistical significance of the parameters, the t-test is used. For each $i = 0,1,\dots,k$, at the α level of significance, a null hypothesis is created:

H_0: $\beta_i = 0$ – the regression coefficient (structural parameter) is not significantly different from zero, which means that the predictor has no effect on the dependent variable, contrary to the alternative hypothesis H_1: $\beta_i \neq 0$ - the regression coefficient (structural parameter) is different from zero, which means a significant effect of the variable x_i on explaining the variation of the dependent variable.

Test statistic:

$$t_i = \frac{\beta_i}{S(\beta_i)} \tag{3}$$

has a t-Student distribution with $n-k-1$ degrees of freedom, where the mean error of the β_i parameter estimation is defined as:

$$S(\beta_i) = \sqrt{D_{ii}} \text{ for } i = 0,1,\dots,k \tag{4}$$

while the covariance matrix of the model parameters is of the form:

$$D = S_\varepsilon^2 (\mathrm{X}^T X)^{-1} \tag{5}$$

$$S_\varepsilon^2 = \frac{1}{n-k-1}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \tag{6}$$

For a t-Student distribution with $n-k-1$ degrees of freedom, we determine quantiles of the order $\frac{\alpha}{2}$ and $1-\frac{\alpha}{2}$. If:

$$-t_{1-\frac{\alpha}{2}}(n-k-1) < t_i < t_{1-\frac{\alpha}{2}}(n-k-1), \tag{7}$$

then at the level of significance α there are no grounds for rejecting the hypothesis H_0 according to the above, the structural parameter β_i is not significantly different from zero, and therefore the variable

x_i has a non-significant effect on explaining the variation of the dependent variable Y..

The following assumptions are made for linear regression models:

- the variance of the random component is equal for all observations - $D^2(\varepsilon_i) = \sigma^2$ for $i = 1, 2, \ldots, n$),
- there is no significant impact of factors not included in the model on the mean value of the dependent variable - $E(\varepsilon_i) = 0$ for $i = 1, 2, \ldots, n$),
- there is no autocorrelation of random components,
- the random component ε_i has a distribution of $N(0, \sigma)$ for $i = 1, 2, \ldots, n$).

The Variance Inflation Factor (VIF), or more precisely its generalization the Generalized Variance Inflation Factor (GVIF), which allows the analysis of collinearity for categorical variables with more than two levels (Lai et al., 2020; Zhang et al. 2022), was used to test the collinearity of variables in the regression model, and such a variable was the month variable. This factor is expressed by the formula:

$$GVIF = 1/(1 - R_i^2)^{df}, \tag{8}$$

where: R_i^2 - the coefficient of determination from the regression model in which the i-th independent variable is predicted by the other independent variables in the model,

df - the degrees of freedom of the i-th independent variable.

In a situation where variables with different numbers of degrees of freedom are compared, the $GVIF^{1/(2*Df)}$ coefficient is checked. This is a transformation that allows a fairer comparison of collinearity between categorical variables with multiple levels and continuous variables. This value is interpreted in the same way as the standard VIF coefficient. Values above 2, and especially above 3, may indicate problematic collinearity.

After standard linear regression, the LASSO (Least Absolute Shrinkage and Selection Operator) model was applied to improve model interpretability and reduce overfitting. This approach was used to verify whether the most significant variables were selected, enhancing the model's interpretability and improving its generalization capability (Sfyridis & Agnolucci, 2023; Xi et al., 2023). By applying LASSO, only the variables that have a real impact on the dependent variable remain in the model.

In the context of linear regression, the LASSO cost function is minimized to determine the optimal regression coefficients. The formula for the LASSO cost function is:

$$J(\beta) = \Sigma(y_i - \beta_0 - \Sigma\beta_j x_{ij})^2 + \lambda\Sigma|\beta_j| \tag{9}$$

Where:

$J(\beta)$ is the cost function,

y_i is the value of the dependent variable for the i-th observation,

x_{ij} is the value of the j-th independent variable for the i-th observation.

After calculations, values that do not contribute any predictive value are eliminated from the model.

As an alternative to analytical methods (linear regression and LASSO), machine learning models were used: Random Forest (RF) and XGBoost (Fatima et al., 2023; Aslan, 2025). Random Forest is a machine learning algorithm that belongs to ensemble learning methods. It employs multiple decision trees to improve prediction accuracy and stability. From the original dataset, multiple random bootstrap samples are generated. For each sample, an independent decision tree is built, and the final prediction is obtained by averaging the predictions of individual trees. At each node split, a random subset of features is selected. Each tree is constructed using random samples and a random subset of features. The node split is determined based on an optimization criterion. The most commonly used optimization criterion is the minimization of the mean square error:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y})^2, \tag{10}$$

Where:

n is the number of samples in a node.

During the split, the variable and threshold value that maximize the reduction of variance in the newly created nodes are selected. Out-of-Bag (OOB) error is used as a method of model validation, which in the case of regression is calculated as:

$$OOB\_Error = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_{OOB})^2 \tag{11}$$

Where:
$\hat{y}_{OOB}$ is the average prediction for i observations, derived from only those trees that did not see it during training.
To enhance model accuracy and efficiency, another machine learning algorithm was proposed XGBoost (eXtreme Gradient Boosting). While it also utilizes decision trees, it differs from Random Forest in model construction, optimization, and regularization methods. XGBoost is generally more accurate than Random Forest.
XGBoost is based on the concept of gradient boosting, an ensemble learning technique that combines multiple weak models (typically decision trees) to create a strong predictive model. Unlike Random Forest, which builds trees in parallel, gradient boosting constructs trees sequentially. Each subsequent tree attempts to correct the errors made by the previous ones by minimizing the loss function. This is achieved through an iterative process in which new trees optimize the gradient of the loss function.
To evaluate and compare the performance of linear regression, LASSO regression, Random Forest, and XGBoost models, two key metrics were used (Chicco et al., 2021):
- coefficient of determination R^2 - which measures how well the model explains variation in the data. This coefficient is expressed by the formula

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}. \quad (12)$$

A higher value of R^2 means a better model fit.
- RMSE (Root Mean Squared Error), which measures the average difference between predicted and actual values, according to the equation:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \quad (12)$$

A lower RMSE value means a better model fit.

4. Study on battery consumption in an unmanned aerial vehicle

4.1. Research subject and data

The subject of this study is the advanced unmanned aerial vehicle Matrice 30T, manufactured by DJI, the global leader in the drone market (year of manufacture: 2022), as shown in Fig. 1. This platform belongs to the Enterprise segment and is an upgraded version of the Matrice 30, equipped with a thermal camera in addition to the standard optical camera. The DJI Enterprise segment consists of devices designed for advanced operations in public safety, surveying, energy, construction, forestry, and agriculture. The DJI Matrice 30T is a comprehensive quadcopter weighing approximately 3,770 g (including two batteries), engineered for flights in challenging weather conditions. With an IP55-rated housing, the drone can operate in rain, snow, and dusty environments. Additionally, it is designed to function within a temperature range of -20°C to 50°C. The drone reaches a top speed of 83 km/h and withstands wind gusts of up to 54 km/h. The platform is equipped with optical hardware (cameras: FPV, thermal, visible-light, and zoom), as well as a laser rangefinder, ToF (Time of Flight) obstacle detection sensors, and an ADS-B receiver, which warns the operator about approaching manned aircraft. The drone is powered by an electric propulsion system using TB30 batteries (Fig. 2). The batteries operate in pairs, which enhances the safety level of flight operations. The TB30 battery is a lithium-polymer (Li-Po) battery with a capacity of 5880 mAh, a voltage of 26.1 V, and an energy rating of 131.6 Wh. The battery weighs approximately 685 g, and its design allows it to function within a temperature range of -20°C to 50°C. The battery lifespan ensures at least 400 charge cycles. The TB30 features a self-heating function that automatically activates when the battery temperature drops below 10°C. Additionally, it supports a hot-swap function, enabling battery replacement without shutting down the drone. This saves time and ensures smooth flight operation in challenging situations.

4.2. Preliminary data processing

The study was conducted based on a dataset of TB30 battery log observations, collected from a total of 177 flight operations performed between December 5, 2023, and February 27, 2025, using the DJI Matrice 30T UAV. The historical data included the following variables:

1. Total time duration of flight "time [ms]": the time measured from the moment the UAV's motors are started until they are turned off (https://app.airdata.com/wiki/Help/Downloadable+Flight+Data+CSV 2025.03).

2. Date of the flight: the date and time of the flight. Recording according to the UTC time standard (https://app.airdata.com/wiki/Help/Downloadable+Flight+Data+CSV 2025.03).
3. Height above the ground elevation at the point of takeoff "height [feet]": the barometric altitude measurement relative to the ground elevation at the takeoff point (https://app.airdata.com/wiki/Help/Downloadable+Flight+Data+CSV 2025.03).
4. Speed of aircraft |"speed [mph]": the UAV's horizontal speed during the flight operation, expressed in miles per hour (https://app.airdata.com/wiki/Help/Downloadable+Flight+Data+CSV 2025.03).
5. Total distance traveled at the currently indicated time "mileage [feet]": the total distance traveled by the UAV from the moment of takeoff, expressed in feet (https://app.airdata.com/wiki/Help/Downloadable+Flight+Data+CSV 2025.03).
6. Aircraft pitch (rotation around the y-axis) in degrees "pitch [degrees]": the aircraft's displacement along the positive Y-axis in the body coordinate system, corresponding to a rightward shift, expressed in degrees (https://app.airdata.com/wiki/Help/Downloadable+Flight+Data+CSV 2025.03; https://developer.dji.com/doc/mobile-sdk-tutorial/en/basic-introduction/basic-concepts/flight-control.html#body-coordinate-system 2025.03).
7. Aircraft roll (rotation around the x-axis) in degrees "roll [degrees]": the aircraft's displacement along the positive X-axis in the body coordinate system, corresponding to a forward shift, expressed in degrees (https://app.airdata.com/wiki/Help/Downloadable+Flight+Data+CSV 2025.03; https://developer.dji.com/doc/mobile-sdk-tutorial/en/basic-introduction/basic-concepts/flight-control.html#body-coordinate-system 2025.03).
8. Level of current being used by the aircraft "current [A]": the battery discharge current at a given moment during the flight, expressed in amperes (A) (https://app.airdata.com/wiki/Help/Downloadable+Flight+Data+CSV 2025.03).

Fig. 1. Matrice 30T (multirotor) with a set of TB30 batteries

The research was conducted in the R environment, version 4.2.2. As part of the preliminary data processing, a quality control check was performed, including missing values analysis using the is.na() function and a report of missing values at the variable level; no missing values were found in the dataset. The time variable was converted to UTC format, and the "month" component was separated as a categorical variable. The data had a uniform sampling frequency resulting from native telemetric recording, so no additional resampling or time interpolation was necessary. The models were trained on the original variable scales. For LASSO regression, internal predictor normalization was used (default cv.glmnet procedure), while Random Forest and XGBoost models were trained on unscaled data due to their insensitivity to variable scale differences. The data was divided into training and test sets in an 80/20 ratio using a random method (createDataPartition). The random generator seed was set (set.seed(123)) to ensure repeatability of the division. Cross-validation (cv.glmnet) was used in the LASSO model. The study used the default hyperparameter settings for RF and XGBoost models, as the goal was to compare the methods rather than optimize them. The analyses were performed in R environment using the following libraries: dplyr, caret, glmnet, randomForest, and xgboost. The management of dataset with a large number of obseervations was performed using frame methods (dplyr), which ensured the efficiency of the calculations.

Each time, the significance of the impact of the identified variables and the strength of this impact were examined using the models discussed in Chapter 3.

Fig. 2. TB30 battery set

4.3. Analytical models

The simplest linear regression model was proposed first. The estimated parameter values of this model are presented in Tab. 1.

All estimated parameters proved to be statistically significant. The coefficient of determination was $R^2 = 0.4257$, indicating that only about 42.57% of the variability in energy consumption was explained by the factors included in the model. This result suggests that other significant variables were not considered in the analysis, or that the assumed model structure does not fully reflect the actual dependencies. The forecast errors on the test set were: MSE = 221.1914 and RMSE = 14.8725. To examine the residual distribution, the Kolmogorov-Smirnov test was used. The test statistic was D = 0.035727, with a p-value < 2×10^{-16}. The null hypothesis assumes that the residuals follow a normal distribution. Therefore, there is no basis to accept the null hypothesis the residual distribution does not meet the normality assumption. However, it is important to emphasize that the analyzed sample consists of 2,239,732 observations. In the case of very large datasets, normality tests often reject the null hypothesis even for minor deviations from normality. This occurs because as the number of observations increases, these tests become highly sensitive to even minimal differences between the empirical and normal distributions. In such cases, it is better to assess normality using a histogram, which is presented in Fig. 3. It can be concluded that the distribution is approximately normal.

Table 1. Estimated parameter values of linear regression models

Parameter	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	87.96	6.562×10^{-2}	1340.45	$< 2\times 10^{-16}$
time [ms]	-2.323×10^{-6}	9.358×10^{-9}	-248.28	$< 2\times 10^{-16}$
height [feet]	8.68×10^{-3}	1.104×10^{-4}	78.65	$< 2\times 10^{-16}$
Speed [mph]	0.357	1.723×10^{-3}	207.10	$< 2\times 10^{-16}$
Distance [feet]	7.728×10^{-4}	4.702×10^{-6}	164.36	$< 2\times 10^{-16}$
Mileage [feet]	-2.078×10^{-3}	2.296×10^{-6}	-905.18	$< 2\times 10^{-16}$
Current [A]	-2.051	6.524×10^{-3}	314.44	$< 2\times 10^{-16}$
Pitch [degrees]	5.849×10^{-2}	2.065×10^{-3}	-28.32	$< 2\times 10^{-16}$
Roll [degrees]	1.087×10^{-2}	2.493×10^{-3}	4.36	1.3×10^{-5}
January -	1.288	5.601×10^{-2}	-23.00	$< 2\times 10^{-16}$
February	-2.738	5.579×10^{-2}	-49.07	$< 2\times 10^{-16}$
March	-1.372	4.361×10^{-2}	-31.47	$< 2\times 10^{-16}$
April	-9.533	6.298×10^{-2}	-151.36	$< 2\times 10^{-16}$
May	4.458	6.615×10^{-2}	67.39	$< 2\times 10^{-16}$
June	3.381	6.698×10^{-2}	50.47	$< 2\times 10^{-16}$
July	5.512	6.630×10^{-2}	83.14	$< 2\times 10^{-16}$
August	2.546	6.226×10^{-2}	40.90	$< 2\times 10^{-16}$
September	12.75	5.644×10^{-2}	-225.82	$< 2\times 10^{-16}$
October	-6.971	5.542×10^{-2}	-125.78	$< 2\times 10^{-16}$
November	9.998	7.255×10^{-2}	137.81	$< 2\times 10^{-16}$

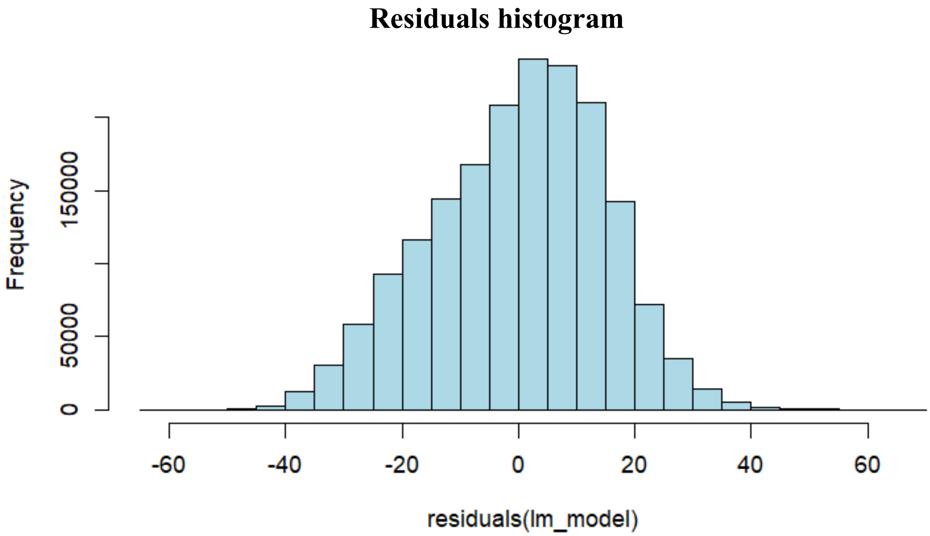

Fig. 3. Distribution of the residuals of the regression model

The Durbin-Watson test was used to examine the autocorrelation of residuals. The test statistic was D-W Statistic = 0.0039, with a p-value = 0, meaning that there is no basis to accept the null hypothesis of no autocorrelation in the residuals. Thus, despite the significant impact of all predictors, the model does not meet the assumptions of classical linear regression, as the residuals do not follow a normal distribution and exhibit strong autocorrelation. As a consequence, the estimators may be biased and inefficient. Additionally, it was checked whether the independent variables used in the model exhibit multicollinearity. For this purpose, Variance Inflation Factors (VIFs) were calculated for each independent variable, and the results are presented in Tab. 2.

Table 2. Calculated values of Variance Inflation Factor for dependent variables

	GVIF	Df	$GVIF^{1/(2*Df)}$
time [ms]	1.388	1	1.178
height [feet]	1.163	1	1.078
Speed [mph]	1.543	1	1.242
Distance [feet]	1.542	1	1.242
Mileage [feet]	1.293	1	1.137
Current [A]	1.103	1	1.050
Pitch [degrees]	1.387	1	1.178
Roll [degrees]	1.014	1	1.007
month	2.022	11	1.033

The results show that all VIF values are relatively low, with most of them being below 2, which suggests that multicollinearity between the independent variables in the model is not an issue.

In the next step, a model incorporating variable selection using the LASSO operator (Least Absolute Shrinkage and Selection Operator) was proposed. The operator indicates which coefficients, significantly different from zero, should remain in the model as significant predictors.

The set of variables selected by the LASSO model, which best predicts the battery level, is presented in Fig. 4.

The coefficients that the model considered less important were removed from the model (no values are shown on the chart). The optimal lambda value in the loss function was found to be λ =0.01548771, which strikes a good balance between fitting the data and simplifying the model (eliminating less significant variables). The LASSO model evaluation was performed by calculating the following error values: MSE = 221.3078, RMSE = 14.87642, and R^2 = 0.4249529. Computed values show that no significant improvement was achieved compared to the linear regression model.

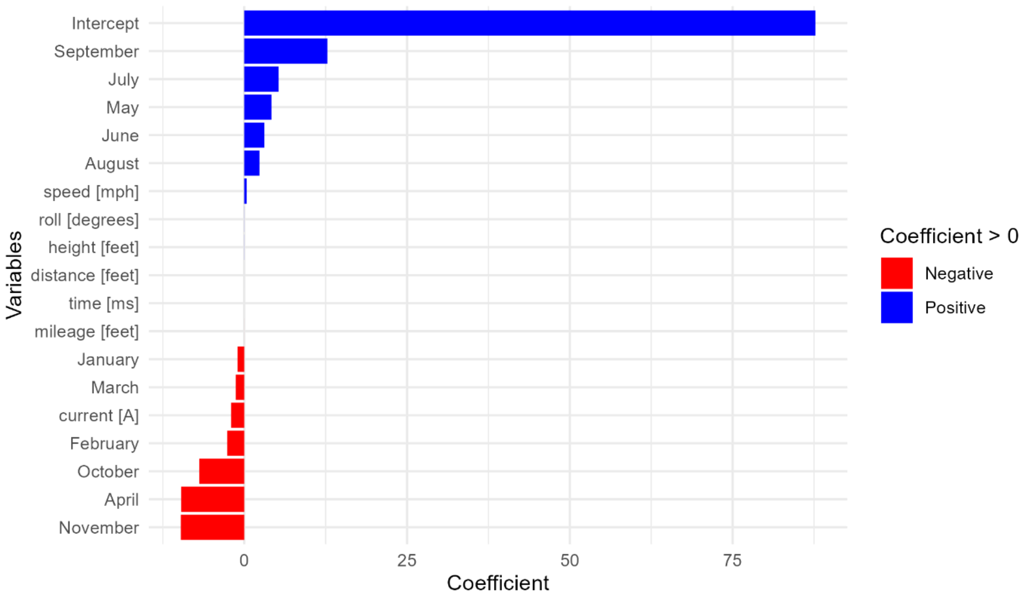

Fig. 4. Variable importance in the LASSO Model

4.4. Machine Learning Models

To improve model fitness quality, more advanced modeling methods were utilized, specifically Random Forest and XGBoost machine learning models. The Random Forest model, applied to predict dependent variable battery [%], was trained on training data using 100 trees. It achieved an extremely high ability to explain the variability in the data, explaining 99.84% of the variability of the dependent variable, which is an exceptionally high result. The mean of squared residuals is 0.600009, which indicates the average difference between the actual and predicted values of battery [%], meaning the model fit the data well, although there remains some difference between the actual and predicted values. The number of variables considered at each split was 3. The variable importance analysis according to the RF model is presented in Fig. 1. The importance was assessed based on the percentage increase in the MSE error (%IncMSE) and the increase in node purity (IncNodePurity). %IncMSE indicates how much the mean squared error increased in the model when a specific variable was removed. A higher %IncMSE value means that the variable has a greater impact on the model quality the greater the increase in MSE after removing the variable, the more significant that variable is for the model. The IncNodePurity measure indicates how much a given variable improves the purity of the nodes in the trees of the model. Higher values of IncNodePurity indicate that the variable helps to better separate the observations, leading to purer nodes and improving the model's ability to distinguish between different cases.

The results of the variable importance analysis in the Random Forest model indicate that the variables mileage [feet], time [ms], and distance [feet] have the greatest impact on the model's quality, both in terms of MSE error and node purity. These are the variables that significantly affect the model's accuracy and help improve its predictive power. Other variables, such as height [feet], roll [degrees], current [A], and month, also play an important role, although their impact is smaller.

The XGBoost model was trained for 100 iterations on data with 8 features. The RMSE metric systematically decreases over the iterations, indicating that the model is fitting well to the training data, as shown in Fig. 6.

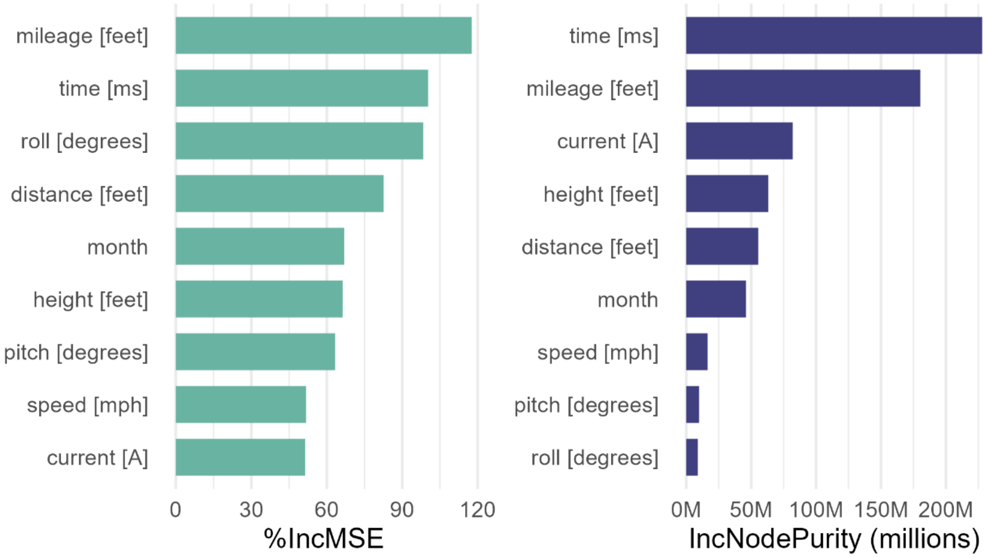

Fig. 5. Variable importance in the RF Model

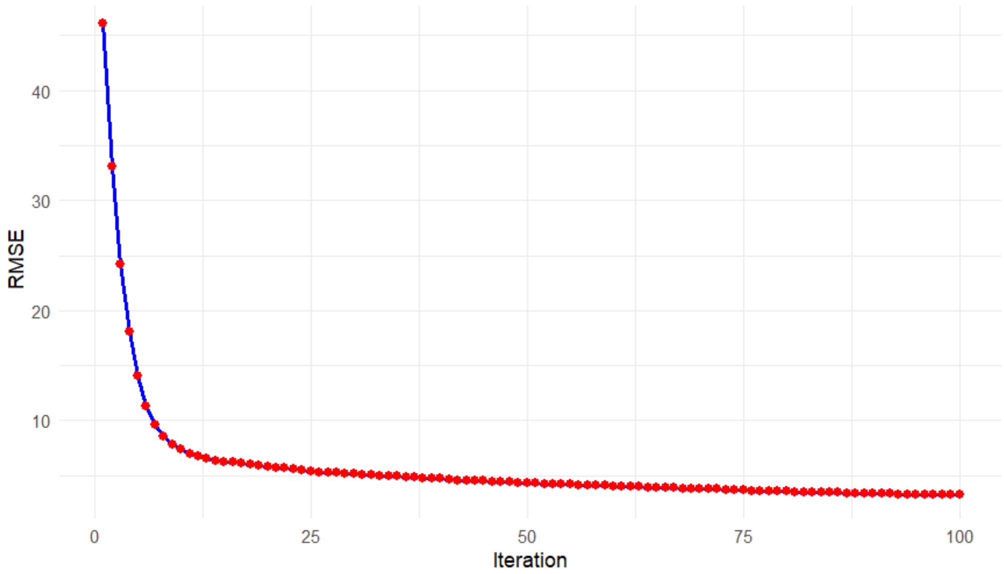

Fig. 6. RMSE vs Iterations in XGBoost

The importance of features in the XGBoost model was determined based on three key metrics for each variable, i.e. gain (how much information a feature contributes to the model - the higher the value, the more it influences the model's decisions), cover (how often the feature is used in tree splits - the higher the value, the more frequently the feature is used), frequency (how often the feature appears in a decision tree split). The results are presented in table 3.

The XGBoost model identifies time as the most important feature (gain = 0.454). This feature is also used relatively frequently during tree splits (Cover = 0.227) and appears most often in splits (Frequency = 0.249). The model also designates mileage [feet] and current [A] as secondary important features. The seasonal factor (month) also plays a significant role in the model. The complete ranking of feature importance is presented in Fig. 7

Table 3. Importance of features according to the XGBoost model

Feature	Gain	Cover	Frequency
time [ms]	0.454	0.227	0.249
mileage [feet]	0.192	0.182	0.165
current [A]	0.111	0.103	0.089
distance [feet]	0.067	0.113	0.117
height [feet]	0.060	0.223	0.160
September	0.040	0.016	0.019
April	0.012	0.004	0.007
January	0.011	0.009	0.025
July	0.010	0.014	0.017
May	0.008	0.017	0.020
November	0.008	0.020	0.015
March	0.007	0.009	0.015
speed [mph]	0.006	0.028	0.039
August	0.004	0.005	0.008
October	0.004	0.007	0.006
February	0.003	0.003	0.009
roll [degrees]	0.003	0.017	0.034
June	0.002	0.002	0.005

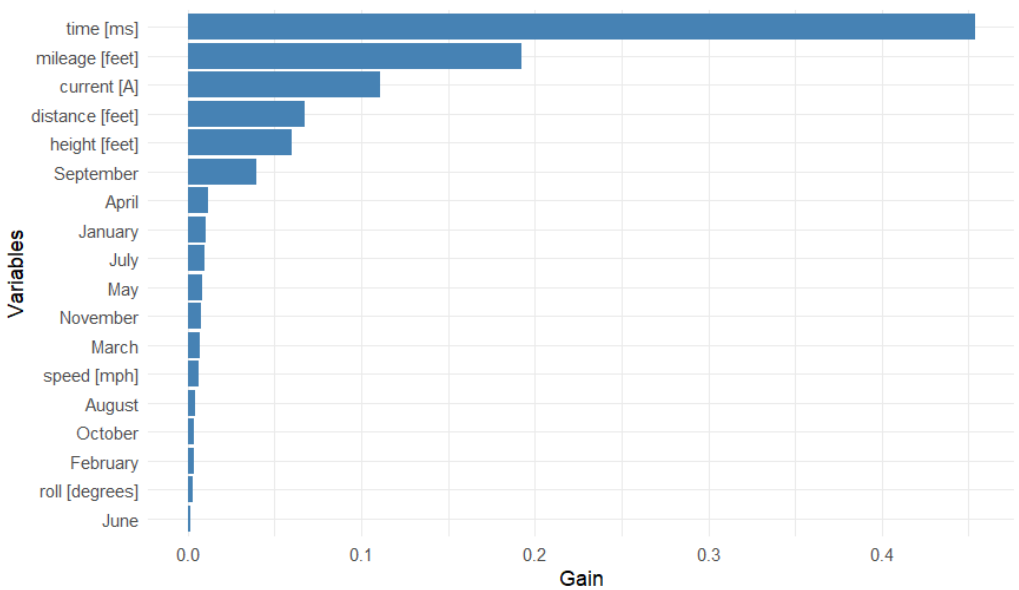

Fig. 7. Importance of features in the XGBoost model

5. Discussion of the results

To evaluate the models, the two key metrics presented in the table 4 , RMSE and R^2, were used.

Table 4. Values of the RMSE and the R^2 metrics for each model

Model	RMSE	R^2
LM	14.873	0.425
LASSO	1212.429	0.425
RF	0.328	0.983
XGBoost	3.26	0.973

The lowest RMSE value of 0.328 and, thus, the highest accuracy was achieved by the Random Forest model. It also has a high fitting score of 98%R^2. The XGBoost model, with an RMSE of 3.26 and a high fit of 97%, produced a significantly better result than the LM and LASSO models, but performed weaker than RF. Linear Regression poorly explains the data variability, yet all variables were statistically significant. LASSO Regression is the least suitable for predictions.
However, a key outcome is identifying the importance of individual variables. Based on the RF and XGBoost models, one may infer that both consistently indicate that mileage [feet] and time [ms] are the most important factors influencing battery consumption, which is a fairly obvious and expected conclusion.
The variable "roll [degrees]" is much more important in the RF model than in XGBoost. Meanwhile, XGBoost ranks "current [A]" and "distance [feet]" in 3rd and 4th positions, respectively. What should be noted is that the "month" variable was included in every model. These results undoubtedly suggest the need for further research, acquisition of more detailed flight information, and the inclusion of more variables in future studies, which will be the focus of the authors' subsequent research.

6. Conclusion

Battery consumption in unmanned aerial vehicles is influenced by many factors; however, not all of them are directly measurable from the user's perspective. Nevertheless, there is a significant need to develop models that describe and predict the relationships between energy consumption and available parameters, which is the basis of this article. To analyze these relationships, mathematical models have been proposed to determine the impact of selected factors on actual energy consumption. This is the first stage of research on UAV flight modeling. In this case, the operation of a single aircraft is being analyzed.
The best results were obtained using machine learning models, but the application of analytical models also confirmed the significant impact of the examined variables on battery consumption. Particularly in the conducted analysis, the notable influence of the "month" variable in all developed models should be emphasized. This variable strongly correlates with weather conditions, which directly affect battery performance. Therefore, in further research, detailed data regarding meteorological conditions will be gathered. The authors assume that wind strength and direction could be key factors. Such an analysis will allow for a better understanding of the impact of individual meteorological factors on battery consumption and enable the development of more accurate predictive models. As a result, this will contribute to the optimization of flight planning and energy management for unmanned aerial vehicles.
It is also important to emphasize the need for modeling in various environmental conditions, including extreme ones, which as demonstrated by the conducted literature review remains a significant challenge. A conclusion of this kind is also drawn in the study by Li et al. (2021), who modeled flight time in an Arctic environment. The importance of such relationships and the necessity of integrating real-time weather data are also emphasized by Mohsan et al. (2022) in their research. Recent discussions in this area clearly point to the need for integrating real-time meteorological data with predictive models (Góra et al., 2022), which will undoubtedly be the subject of further research by the authors.

References

1. Aslan, E. (2025). Temperature Prediction and Performance Comparison of Permanent Magnet Synchronous Motors Using Different Machine Learning Techniques for Early Failure Detection. Eksploatacja i Niezawodność – Maintenance and Reliability, 27(1). https://doi.org/10.17531/ein/192164

2. Barnhart, R. K., Marshall, D. M., Shappee, E. (2021). Introduction to unmanned aircraft systems. Crc Press. https://doi.org/10.1201/9780429347498
3. Bassi, E. (2019, June). European drones regulation: Today's legal challenges. In 2019 international conference on unmanned aircraft systems (ICUAS) (pp. 443-450). IEEE. https://doi.org/10.1109/ICUAS.2019.8798173
4. Chen, X., Cheng, X., Wu, N., Liu, X. (2024). Unmanned aerial vehicle transmission defect detection technology based on edge computing. Diagnostyka, 25. https://doi.org/10.29354/diag/194598
5. Chicco, D., Warrens, M. J., & Jurman, G. (2021). The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. Peerj computer science, 7, e623. https://doi.org/10.7717/peerj-cs.623
6. Di Franco, C., & Buttazzo, G. (2015, April). Energy-aware coverage path planning of UAVs. In 2015 IEEE international conference on autonomous robot systems and competitions (pp. 111-117). IEEE. https://doi.org/10.1109/ICARSC.2015.17
7. Eskandari, R., Mahdianpari, M., Mohammadimanesh, F., Salehi, B., Brisco, B., & Homayouni, S. (2020). Meta-analysis of unmanned aerial vehicle (UAV) imagery for agro-environmental monitoring using machine learning and statistical models. Remote Sensing, 12(21), 3511. https://doi.org/10.3390/rs12213511
8. Fatima, S., Hussain, A., Amir, S. B., Ahmed, S. H., & Aslam, S. M. H. (2023). XGBoost and random forest algorithms: an in depth analysis. Pakistan Journal of Scientific Research, 3(1), 26-31. https://doi.org/10.57041/pjosr.v3i1.946
9. Gong, H., Huang, B., Jia, B., & Dai, H. (2023). Modeling power consumptions for multirotor UAVs. IEEE Transactions on Aerospace and Electronic Systems, 59(6), 7409-7422. https://doi.org/10.1109/TAES.2023.3288846
10. Góra, K., Smyczyński, P., Kujawiński, M., & Granosik, G. (2022). Machine learning in creating energy consumption model for uav. Energies, 15(18), 6810. https://doi.org/10.3390/en15186810
11. Grzelak, M., Borucka, A., & Guzanek, P. (2021, June). Application of linear regression for evaluation of production processes effectiveness. In International Conference Innovation in Engineering (pp. 36-47). Cham: Springer International Publishing.https://doi.org/10.1007/978-3-030-78170-5_4
12. Hang, T., Wen, J., Zheng, B. S., Xiao, J. H., Zhou, F. (2024). Reliability analysis of the vehicle door system EDCU based on Weibull distribution. Eksploatacja i Niezawodność – Maintenance and Reliability, 27(2). https://doi.org/10.17531/ein/195257
13. https://app.airdata.com/wiki/Help/Downloadable+Flight+Data+CSV (Access from 16.03.2025r.)
14. https://developer.dji.com/doc/mobile-sdk-tutorial/en/basic-introduction/basic-concepts/flight-control.html#body-coordinate-system (Access from 16.03.2025r.)
15. Karunarathne, L., Economou, J. T., & Knowles, K. (2012). Power and energy management system for fuel cell unmanned aerial vehicle. Proceedings of the Institution of Mechanical Engineers, Part G: Journal of Aerospace Engineering, 226(4), 437-454. https://doi.org/10.1177/0954410011409995
16. Kim, S. J., Lim, G. J., & Cho, J. (2018). Drone flight scheduling under uncertainty on battery duration and air temperature. Computers & Industrial Engineering, 117, 291-302. https://doi.org/10.1016/j.cie.2018.02.005
17. Kozłowski, E., Borucka, A., Oleszczuk, P., Leszczyński, N. (2024). Evaluation of readiness of the technical system using the semi-Markov model with selected sojourn time distributions. Eksploatacja i Niezawodność – Maintenance and Reliability, 26(4). https://doi.org/10.17531/ein/191545
18. Lai, J., Zou, Y., Zhang, J., & Peres-Neto, P. R. (2022). Generalizing hierarchical and variation partitioning in multiple regression and canonical analyses using the rdacca. hp R package. Methods in Ecology and Evolution, 13(4), 782-788. https://doi.org/10.1111/2041-210X.13800
19. Li, J., Ding, P. (2024). Application of DBN-based KRLS method for RUL prediction of lithium-ion batteries. Eksploatacja i Niezawodność – Maintenance and Reliability, 27(2). https://doi.org/10.17531/ein/194174

20. Li, J., Ding, P. (2024). Application of DBN-based KRLS method for RUL prediction of lithium-ion batteries. Eksploatacja i Niezawodność – Maintenance and Reliability, 27(2). https://doi.org/10.17531/ein/194174
21. Li, N., Liu, X., Yu, B., Li, L., Xu, J., & Tan, Q. (2021). Study on the environmental adaptability of lithium-ion battery powered UAV under extreme temperature conditions. Energy, 219, 119481. https://doi.org/10.1016/j.energy.2020.119481
22. Luo, X., Bu, W., Liang, H., Zheng, M. (2025). Convolutional Neural Network - Gated Recurrent Unit combined with Error Correction for Lithium Battery State of Health Estimation. Eksploatacja i Niezawodność – Maintenance and Reliability, 27(4). https://doi.org/10.17531/ein/202184
23. Mohsan, S. A. H., Khan, M. A., Noor, F., Ullah, I., & Alsharif, M. H. (2022). Towards the unmanned aerial vehicles (UAVs): A comprehensive review. Drones, 6(6), 147. https://doi.org/10.3390/drones6060147
24. P, J.S., D, A.D., C, R.L., & R, N. (2025) Efficiency and Reliability: Optimization of Energy Management in Electric Vehicles Apply Monarch Butterfly Algorithm and Fuzzy Logic Control. Eksploatacja i Niezawodność – Maintenance and Reliability, 27(3). http://doi.org/10.17531/ein/200691
25. Poorani, S., Jebarani Evangeline, S., Bagyalakshmi, K., Maris Murugan, T. (2025). Improving Reliability in Electric Vehicle Battery Management Systems through Deep Learning-Based Cell Balancing Mechanisms. Eksploatacja i Niezawodność – Maintenance and Reliability, 27(3). http://doi.org/10.17531/ein/200714
26. Ravich, T. M. (2019). Emerging technologies and enforcement problems: The Federal Aviation Administration and drones as a case study. Loy. U. Chi. J. Reg. Compl., 4, 34.
27. Sang, T., Zhu, K., Shen, J., Yang, L. (2025). An uncertain programming model for fixed charge transportation problem with item sampling rates. Eksploatacja i Niezawodność – Maintenance and Reliability, 27(1). https://doi.org/10.17531/ein/192165
28. Sfyridis, A., & Agnolucci, P. (2023). Factors affecting road traffic: identifying drivers of annual average daily traffic using least absolute shrinkage and selection operator regression. Transportation research record, 2677(5), 1178-1192. https://doi.org/10.1177/03611981221141435
29. Shibl, M. M., Ismail, L. S., & Massoud, A. M. (2023). A machine learning-based battery management system for state-of-charge prediction and state-of-health estimation for unmanned aerial vehicles. Journal of Energy Storage, 66, 107380. https://doi.org/10.1016/j.est.2023.107380
30. Szczupak, P., Kossowski, T., Szostek, K., Szczupak, M. (2025). Tests of pulse interference from lightning discharges occurring in unmanned aerial vehicle housings made of carbon fibers. Eksploatacja i Niezawodność – Maintenance and Reliability, 27(1). https://doi.org/10.17531/ein/193984
31. Traub, L. W. (2011). Range and endurance estimates for battery-powered aircraft. Journal of Aircraft, 48(2), 703-707. https://doi.org/10.2514/1.C031027
32. Wang, Z., Shangguan, W., Peng, C., Cai, B. (2024). Similarity based remaining useful life prediction for lithium-ion battery under small sample situation based on data augmentation. Eksploatacja i Niezawodność – Maintenance and Reliability, 26(1). https://doi.org/10.17531/ein/175585
33. Wanner, D., Hashim, H. A., Srivastava, S., & Steinhauer, A. (2024). UAV avionics safety, certification, accidents, redundancy, integrity, and reliability: a comprehensive review and future trends. Drone Systems and Applications, 12, 1-23. https://doi.org/10.1139/dsa-2023-0091
34. Xi, L. J., Guo, Z. Y., Yang, X. K., & Ping, Z. G. (2023). Application of LASSO and its extended method in variable selection of regression analysis. Zhonghua yu fang yi xue za zhi [Chinese journal of preventive medicine], 57(1), 107-111. https://doi.org/10.3760/cma.j.cn112150-20220117-00063
35. Zhang, Z., Li, L., Li, X., Hu, Y., Huang, K., Xue, B., ... & Yu, Y. (2022). State-of-health estimation for the lithium-ion battery based on gradient boosting decision tree with autonomous selection of excellent features. International Journal of Energy Research, 46(2), 1756-1765. https://doi.org/10.1002/er.7292
36. Ziółkowski, J., Oszczypała, M., Lęgas, A., Konwerski, J., Małachowski, J. (2024). A method for calculating the technical readiness of aviation refuelling vehicles. Eksploatacja i Niezawodność – Maintenance and Reliability, 26(3). https://doi.org/10.17531/ein/187888