

PAPER • OPEN ACCESS

Enhancement of multirotor UAV conceptual design through Machine Learning algorithms

To cite this article: C. Pliakos *et al* 2024 *J. Phys.: Conf. Ser.* **2716** 012066

View the [article online](#) for updates and enhancements.

You may also like

- [A Data-driven Method for Realistic Covariance Prediction of Space Object with Sparse Tracking Data](#)
Hong-Kang Liu, Bin Li, Yan Zhang et al.
- [Emerging machine learning strategies for diminishing measurement uncertainty in SPM nanometrology](#)
Linh Thi Phuong Nguyen and Bernard Haochih Liu
- [Projecting armed conflict risk in Africa towards 2050 along the SSP-RCP scenarios: a machine learning approach](#)
Jannis M Hoch, Sophie P de Bruin, Halvard Buhaug et al.



HONOLULU, HI
Oct 6–11, 2024

Abstract submission deadline:
April 12, 2024

Learn more and submit!



Joint Meeting of

The Electrochemical Society
•
The Electrochemical Society of Japan
•
Korea Electrochemical Society

Enhancement of multirotor UAV conceptual design through Machine Learning algorithms

C. Pliakos^{1,2}, D. Terzis^{1,2}, S. Vlachos^{1,2}, C. Bliamis^{1,2}, K. Yakinthos^{1,2}

¹ Laboratory of Fluid Mechanics and Turbomachinery, Department of Mechanical Engineering, Aristotle University of Thessaloniki, 54124 Thessaloniki, Greece

² UAV Integrated Research Center (UAV-iRC), Center for Interdisciplinary Research and Innovation (CIRI), Aristotle University of Thessaloniki, 57001 Thessaloniki, Greece

kyak@auth.gr

Abstract. Designing an efficient and optimized multirotor UAV requires laborious trade-off analyses, involving numerous design variables and mission requirement parameters, especially during the early conceptual design phase. The large number of unknown parameters, as well as the associated design effort often leads to non-optimal designs, for the sake of time efficiency. This work presents the implementation of a machine learning (ML) framework to assist and expedite the conceptual design phase of multirotor UAVs. The framework utilizes information from a comprehensive database of commercial lightweight multirotor UAVs. The database contains an extensive collection of crucial sizing parameters, performance metrics, and features associated with foldability and indoor guidance (e.g., obstacle avoidance sensors). These attributes specifically pertain to multirotor UAVs weighing less than 2kg, which exhibit diverse design and performance characteristics. The proposed ML framework employs multiple regression models (e.g. k-nearest neighbors regression, multi-layer perceptron regression) to predict the sizing parameters during a multirotor UAV's conceptual design phase. This enables designers to make quick informed decisions, while also significantly reducing computational time and effort. Finally, the ML framework's predictive capability is validated by comparing the predicted values with real-world data from an "unseen" test dataset.

1. Introduction

In recent years, the utilization of multirotor Unmanned Aerial Vehicles (UAVs) has witnessed a remarkable surge across various industries and applications [1]. These versatile aerial platforms have transcended their initial roles in military and surveillance operations to become indispensable tools in sectors as diverse as agriculture, cinematography, infrastructure inspection, and emergency response [2]. Their ability to hover, maneuver in tight spaces, and capture high-resolution imagery has revolutionized data collection and remote sensing. As a result, the demand for tailored multirotor UAVs has escalated significantly, prompting the need for more efficient conceptual design approaches.

In the realm of Unmanned Aerial Vehicles (UAVs), the design process for fixed-wing aircraft has reached a mature stage, marked by a systematic approach beginning with tailored user requirements aligned with mission objectives. Established methodologies, often rooted in historical trends from manned aviation, leverage semi-empirical correlations between requirements and design parameters [3]. Extensive efforts have been invested in extracting similar tools for UAVs, bridging the gap between conventional aircraft design and unmanned systems [4]. In stark contrast, the design landscape for



multirotor UAVs remains in its infancy, relying heavily on individual designer expertise and utilizing low-fidelity tools coupled with manufacturer data for components. While tools like ecalc and Drive Calculator [4] expedite the comparison of various components, they often overlook installed performance factors, leading to complexities arising from the vast permutations.

Recent strides in the conceptual design of multirotor UAVs have seen a focused exploration of specific components. For instance, Biczyski et al. [5] introduced a methodology that optimizes motor-propeller combinations, factoring in estimated flight duration. L. de Angelis et al. [6] delved into the energy system during hover, emphasizing pivotal rotor features affecting flight endurance. Meanwhile, holistic approaches to design have emerged, integrating correlations and trends. Delbecq et al. [7] proposed techniques utilizing scaling laws and similarity models to expedite multirotor UAV conceptual design, a sentiment echoed by M. Budinger et al. [8] in the preliminary design phase. A notable innovation comes from Yuyao et al. [9] who employed regression tools for installed performance. Their study, involving a database of 16 drones, explored the relationships between components and overall maximum takeoff weight (MTOW).

The core focus of this study lies in harnessing the power of machine learning (ML) methodologies applied to existing multirotor UAV platforms. The target is to utilize the predictive capabilities of ML, turning machine learning algorithms into invaluable tools that enable swift and accurate predictions of crucial design parameters during the conceptual design phase. By seamlessly incorporating these predictive models into the design loop, designers can make well-informed decisions with minimal input, ushering in a new era of efficiency and precision in the multirotor UAV conceptual design process.

2. Data Acquisition

The initial step in any ML endeavor involves data acquisition for model training. In this study, a custom database was created to meet specific drone analysis criteria.

2.1. Database population

In curating the database, several stringent criteria were applied. Firstly, drones within the defined weight limit of less than 2 kg were exclusively chosen. Secondly, the focus was on drones designed for professional applications, excluding toy-grade variants, ensuring relevance to practical uses like aerial photography, surveillance, and mapping. The database was diversified, encompassing various brands, models, and configurations with all available equipment, sourced directly from manufacturers. Importantly, the selected databases align precisely with the research objectives of the machine learning process. The database is publicly available via a GitHub repository¹.

2.2. Database parameters

The design parameters for each drone were systematically categorized into geometrical/layout, performance, mission-related, and equipment aspects, constituting a total of 26 distinct metrics. These encompass crucial details such as maximum take-off weight, dimensions, and specialized applications. Comprehensive specifics of these metrics are outlined in Table 1.

Table 1. All the parameters that were documented during the data acquisition phase

Category	Parameters collected
Design	Propellers (Count, Diameter), Dimensions (Length, Height, Width, Diagonal), MTOW
Performance	Max Forward, Ascend/Descend and Max Wind Velocities, Flight Time
Energy	Battery (Type, Cells, Capacity, Weight), Charging Time
Control	RC Range, Ceiling
Equipment	Max Sensors, Camera (Video Resolution), GPS Presence
Features	Obstacle Avoidance, Indoor Flight, Payload Type

¹ https://github.com/christospliakos/quadcopter_database_EASN_2023

2.3. Descriptive statistics

In the dataset, which comprises information from 87 diverse drones, there is a small fraction of missing data, approximately 5%. These gaps primarily stem from incomplete manufacturer data. Due to the variety of documented categories, sampled parameters exhibit values of drastically different scales. For example, the propellers' diameter ranges from 2.25" to 18.5" while the battery capacity (measured in mAh) ranges from 450mAh to 7500mAh. The latter is two orders of magnitude larger than the diameter of propellers. This poses the threat of misleading the machine learning algorithms into thinking that the battery capacity is a more biased/weighted parameter, hence a more important one. In chapter 3.6 there is a discussion on the treatment of this issue.

3. Methodology

In the forthcoming sections, the research methodology is outlined, providing an objective overview of the procedures employed, which is outlined in Figure 1. The workflow encompasses three pivotal components: data processing (exploratory data analyses and pre-processing), model construction, and model validation. Notably, this research utilizes the Python programming language, leveraging open-source libraries such as Pandas, Scikit-Learn, Matplotlib, and NumPy for implementation.

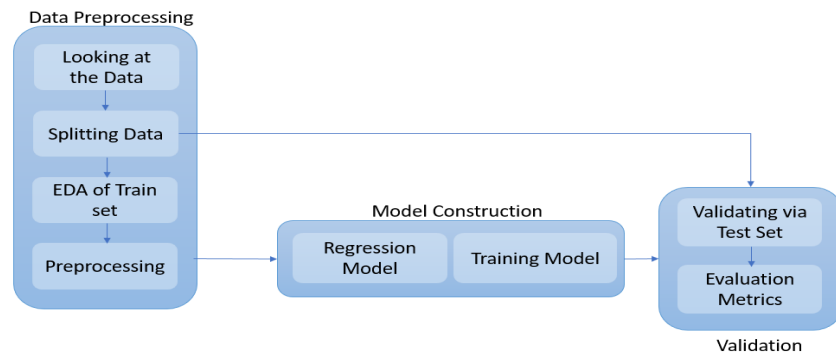


Figure 1. Workflow followed from data preprocessing to the validation of ML algorithms

3.1. Machine Learning

In the realm of machine learning, the power of prediction unfolds as systems delve into data patterns without explicit programming, a technique pivotal in this study. Supervised learning, links features to labels [10]. Through rigorous training on the feature matrix (X_{train} – independent variables) and corresponding labels (y_{train} – dependent variables), algorithms minimize the divergence between predicted outcomes (\hat{y}) and actual labels (y_{train}) using specialized loss functions. The iterative optimization of internal parameters (θ) through techniques like gradient descent refines the models, enabling them to generalize to unseen data (X_{test}) and predict vital design parameters.

The research harnesses regression methods to estimate multirotor UAV design parameters during the conceptual phase swiftly. The precision of this estimation relies on the performance of four distinct models: k-nearest Neighbors, Random Forest, Support Vector Regression, and Multi-layer Perceptron (MLP) regression [10,11]. The evaluation of these models includes key metrics such as Mean Absolute Error (MAE), Mean Squared Error (MSE), and the coefficient of determination R^2 [12], providing comprehensive insights into their predictive capabilities. This in-depth analysis forms the bedrock of the methodology, ensuring the accuracy and reliability of the results.

3.2. Data Leakage Precautions

Guarding against the pervasive threat of Data Leakage, a phenomenon prevalent in numerous machine learning projects [13], was crucial in this study. Specifically, diligent attention was paid to ensuring a clear and early separation between the training and test datasets, addressing the fundamental challenges posed by unseparated data. Illegitimate feature usage during training was also carefully avoided. In the

subsequent section, the division of the dataset into distinct training and test subsets will be explained, followed by a detailed exploration of the careful selection of model features in section 3.4.

3.3. Creating the Training and Test Sets

The dataset was divided into a 20% test subset and an 80% training subset due to the limited data available, a strategy to ensure a balance between model evaluation and training robustness. The training phase, crucial for model development, utilized the 80% training subset, allowing the model to learn patterns within the data. The remaining 20% held out for validation served as a critical checkpoint, enabling us to assess the model's generalization and performance on unseen data. This division and validation process ensured a comprehensive evaluation of the model's predictive capabilities while guarding against overfitting, thereby enhancing the reliability of the results.

3.4. Feature Engineering

In addition to the features initially recorded in the database, new parameters were derived from existing ones, a common strategy to enrich datasets with valuable information [14]. These newly generated features, chosen for their relevance in the UAV design domain or their significance as performance and design indicators, enhance the depth of this analysis. The study introduces several new parameters:

- Battery Energy [Wh]: Calculated by leveraging the nominal voltage of LiPo/Li-Ion batteries (assumed to be 3.7V) and the reported battery capacity.
- Disk Loading [gr/ft^2]: Derived from the propeller area (another constructed parameter) and the maximum takeoff weight, this metric, is widely used in rotary aircraft analysis.
- Volume [cm^3]: Total volume of the quadrotor UAV calculated by length, width, and height.

3.5. Identifying correlations

The next step in the process was to identify relationships between the variables to be regressed. To achieve that, two methods, and therefore two indicators, were utilized: Spearman's correlation coefficient and Mutual Information between the variables.

3.5.1. Mutual Information

The first and most important criterion for the variable relationships, Mutual Information (M.I.) was utilized. M.I. of two random variables measures the mutual dependence between the two variables. More specifically, it quantifies the “amount of information” obtained about one random variable by observing the other random variable. Higher values of M.I. indicate a stronger relationship between variables. In figures 2a – 2b, M.I. for two critical design parameters is shown,

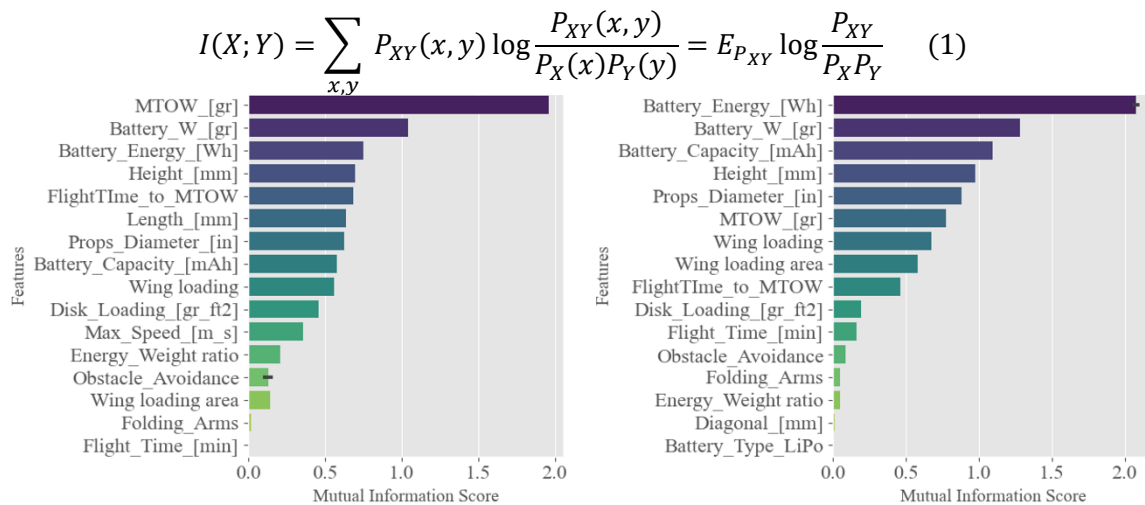


Figure 2a – 2b. Mutual information between parameters for MTOW (a) and Battery Energy (b)

3.5.2. Spearman's Correlation coefficient

Figure 3 visually represents Spearman rank correlation coefficients (ρ) indicating the strength and direction of monotonic relationships (linear or not) between variables (symmetrical figure), ranging from -1 to +1. A coefficient of -1 signifies a perfect negative relationship, 0 indicates no correlation, and +1 indicates a perfect positive relationship.

$$r_s = \rho_{R(X),R(Y)} = \frac{\text{cov}(R(X), R(Y))}{\sigma_{R(X)}\sigma_{R(Y)}} \quad (2)$$

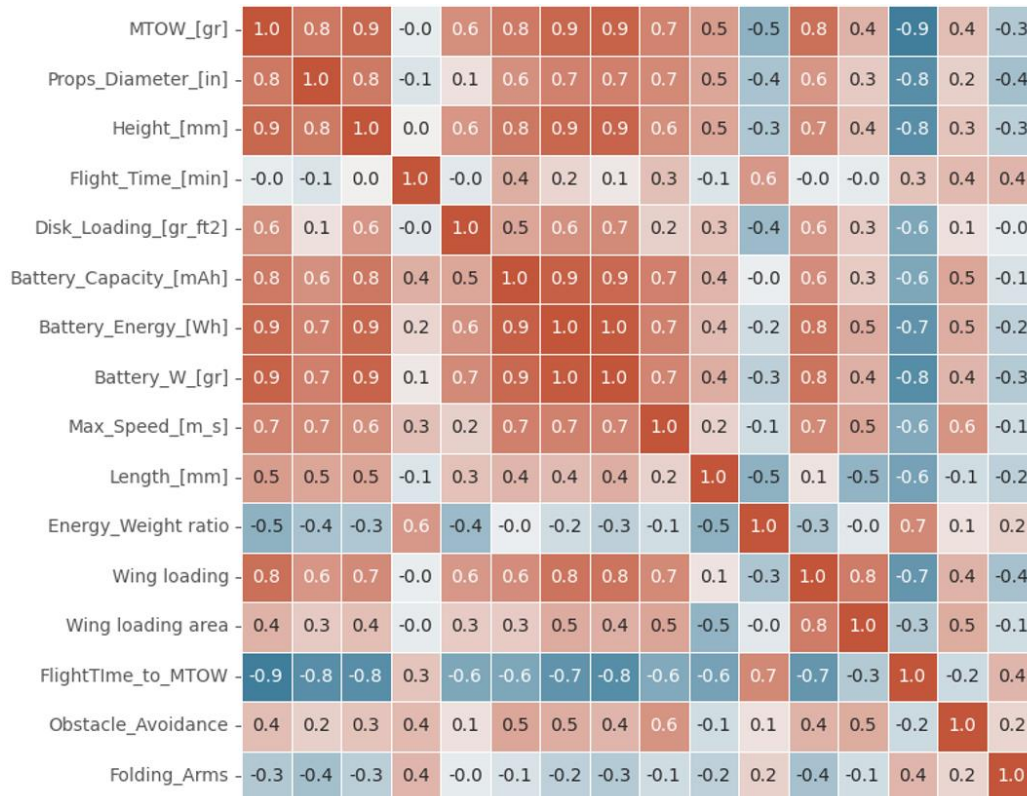


Figure 3. Heatmap of spearman correlation coefficient for numerous parameters

It was observed that certain "areas of interest" were formed, primarily around the performance and geometrical variables. Consequently, these relationships were investigated further. On the other hand, categorical variables generally exhibited poor performance in correlating with other variables.

3.6. Models Selection

Given the wealth of features, the research demanded a strategic focus on specific models and parameter combinations. Among the numerous variables, two crucial parameters profoundly impacting multirotor UAV design were identified: the **Maximum Take-off Weight (MTOW)** and the total available **Battery Energy**. These parameters, extracted through analysis of correlations and mutual information, emerged as pivotal variables shaping the UAV's conceptual design. The selected features for training the ML models are summarized in Table 2, with each row representing an independent feature.

Table 2. The two models that were selected to train the ML algorithms.

	Model 1	Model 2
Y variable (Dependent)	Maximum Take-off Weight (gr)	Battery Energy (Wh)
X variables (Independent)	Battery Weight (gr)	Propellers Diameter (in)
	Battery Energy (Wh)	UAV Length (mm)
	UAV Length (mm)	Maximum Take-off Weight (gr)
	Maximum Forward Velocity (m/s)	Total flight Time (min)
	Obstacle Avoidance feature (0/1)	

3.7. Preparing Data for Regression Models

In the final preprocessing stage, the challenge of varied feature ranges was tackled as it is showcased in section 2.3. Standardization, ensuring zero mean and unit standard deviation, was crucial for unbiased algorithm performance, excluding Random Forest. Additionally, the 5% missing data was carefully handled using the Scikit-learn library's Iterative Imputing technique, ensuring a complete and reliable dataset for subsequent analysis.

4. Results

For the ML algorithms, basic tuning was performed using a grid search technique which led to the hyper-parameters shown in Table 3.

Table 3. Hyper-parameters for the 4 ML algorithms for the 2 district models

Model	RFR	kNN	MLP	SVR
MTOW	n_estimators=125, min_samples_split=2, criterion=absolute_error	n_neighbors=8, metric=euclidean, p=1.5	layer_sizes= (200, 100, 50), activation=relu, learning_rate_init=0.05, alpha=0.001	C=100, degree=3, epsilon=0.1, kernel=linear
Battery Energy	n_estimators=35, min_samples_split=4, criterion=friedman_mse	n_neighbors=3, metric=minkowski, p=3	layer_sizes= (100,100,100), activation=relu, learning_rate_init=0.05, alpha=0.001	C=500, degree=3, epsilon=0.1, kernel=rbf

In Table 4 the resulting metrics for each of the 2 models are displayed. The validation and the calculation of the metrics is always done on the test subset (unseen data). This is an indication that the ML algorithm can generalize well and capture the complex relationships in unseen data.

Table 4. Metrics of predictions on test subset for the ML algorithms

ML Algorithm	MTOW Model			Battery Energy Model		
	MAE ^a	MSE ^a	R ^{2b}	MAE ^a	MSE ^a	R ^{2b}
RFR	135.19	30472.27	0.89	6.11	71.87	0.89
kNN	35.99	6532.86	0.98	6.67	81.33	0.88
MLP	198.92	66287.48	0.77	8.85	113.98	0.83
SVR	139.95	49069.37	0.83	7.61	91.58	0.87

^a smaller value is better.

^b bigger value is better.

In Figures 4 and 5, depicting MTOW and Battery Energy models respectively, the graphical representation illustrates the comparison between predicted values from the test dataset (X-axis) and the corresponding real unseen values (Y-axis) (left side). In contrast, on the right, the residuals of the prediction (actual – predicted value) are presented over the real values. In an ideal scenario, on the predicted versus real values, data points would precisely align along the $Y = X$ line indicating a flawless match. On the residuals versus the real values the residuals should lay on the $Y = 0$ line which indicates a perfect prediction. The residual plot is a way to unveil any biased pattern of the ML algorithms. Specifically, it is important to ensure that there are no repeating patterns such as “U” shapes, and that the dispersion of data points remains random around the $Y = 0$ line.

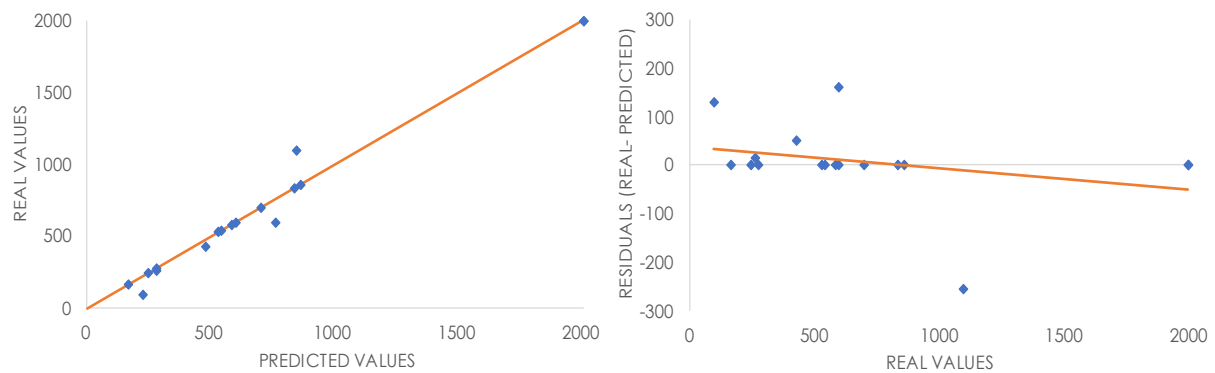


Figure 4a – 4b. Real values versus the predicted values (4a) and residual values (real - predicted) versus the real values (4b) for the kNN algorithm and MTOW model

Notably, the kNN algorithm stands out in the MTOW model, displaying superior performance with an impressive R^2 of 0.98. In the context of the Battery Energy model, the RFR algorithm exhibits noteworthy results, achieving a R^2 of 0.89. These remarkably high R^2 scores underscore the models' exceptional accuracy in capturing the intricate relationships within the data, showcasing their robust predictive capabilities.

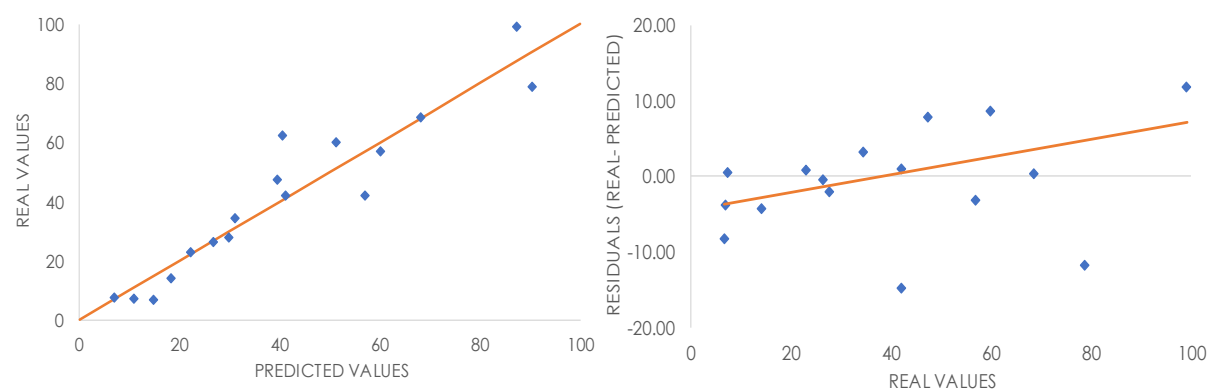


Figure 5a – 5b. Real values versus the predicted values (4a) and residual values (real - predicted) versus the real values (4b) for RFR algorithm and Battery Energy model

5. Conclusions

The potential of machine learning (ML) applications has been explored for rapid and accurate estimation of the design parameters of multirotor UAVs during the conceptual design phase. A concrete database was constructed using data from UAV manufacturers, as well as new features were created to add more

layers of information. After rigorous preprocessing of the data while taking into consideration the risks of data leakage, a framework of 4 ML algorithms was developed. Maximum take-off weight and battery energy were the two variables chosen as the dependent ones (Y variable) since they are deemed as some of the most critical characteristics that drive the design of a multirotor UAV. The performance of the ML framework shows that the kNN and RFR predict the target parameters better than the other algorithms with remarkable precision. This underscores the potential of ML algorithms as indispensable tools for designers, ensuring precise predictions even in scenarios with constrained knowledge.

Acknowledgement

This work has been implemented within the project “MIDRES-Autonomous Robotic Unmanned Aerial Vehicle System for Navigation in Inaccessible Interior Spaces and Human Detection” which has been financially supported by the European Regional Development Fund, Partnership Agreement for the Development Framework (2014-2020), co-funded by Greece and European Union in the framework of OPERATIONAL PROGRAMME: “Competitiveness, Entrepreneurship and Innovation 2014-2020 (EPAnEK)”, Nationwide Action: “Research - Create - Innovate” (project code: T2EDK-00592)

References

- [1] Mohsan S A H, Khan M A, Noor F, Ullah I and Alsharif M H 2022 Towards the Unmanned Aerial Vehicles (UAVs): A Comprehensive Review *Drones* 2022, Vol. 6, Page 147 **6** 147
- [2] Martinez-Alpiste I, Golcarenenrenji G, Wang Q and Alcaraz-Calero J M 2021 Search and rescue operation using UAVs: A case study *Expert Syst Appl* **178**
- [3] Raymer D 2018 *Aircraft Design: A Conceptual Approach, Sixth Edition* (Washington, DC: American Institute of Aeronautics and Astronautics, Inc.)
- [4] Mitridis D, Kapsalis S, Terzis D and Panagiotou P 2023 An Evaluation of Fixed-Wing Unmanned Aerial Vehicle Trends and Correlations with Respect to NATO Classification, Region, EIS Date and Operational Specifications *Aerospace* **10**
- [5] Biczyski M, Sehab R, Whidborne J F, Krebs G and Luk P 2020 Multirotor Sizing Methodology with Flight Time Estimation *J Adv Transp* **2020**
- [6] de Angelis E L, Giulietti F, Rossetti G and Bellani G 2021 Performance analysis and optimal sizing of electric multirotors *Aerosp Sci Technol* **118**
- [7] Delbecq S, Budinger M, Ochotorena A, Reysset A and Defay F 2020 Efficient sizing and optimization of multirotor drones based on scaling laws and similarity models *Aerosp Sci Technol* **102**
- [8] Budinger M, Reysset A, Ochotorena A and Delbecq S 2020 Scaling laws and similarity models for the preliminary design of multirotor drones *Aerosp Sci Technol* **98**
- [9] Yuyao J and Patrick H A user-friendly method for small UAV sizing and performance estimate
- [10] Aurélien Géron 2019 Hands-on machine learning with Scikit-Learn, Keras and TensorFlow: concepts, tools, and techniques to build intelligent systems *O'Reilly Media* 851
- [11] Sarker I H 2021 Machine Learning: Algorithms, Real-World Applications and Research Directions *SN Comput Sci* **2** 160
- [12] Naser M Z and Alavi A H 2021 Error Metrics and Performance Fitness Indicators for Artificial Intelligence and Machine Learning in Engineering and Sciences *Architecture, Structures and Construction*
- [13] Kapoor S and Narayanan A 2023 Leakage and the reproducibility crisis in machine-learning-based science *Patterns* **4**
- [14] Lo D 2018 Feature Engineering for Machine Learning and Data Analytics Feature Engineering for Machine Learning and Data Analytics Xin XIA