AirfoilAD: A benchmark aerodynamic dataset for machine learning-based modelling of lift and drag coefficients

AirfoilAD: A Benchmark Aerodynamic Dataset for Machine Learning-Based Modeling of Lift and Drag Coefficients

Noha MM. Abdelnapi[a], Ahmed Gamal Abdellatif[b], Rania Ahmed[c], Mohamed A. Aziz[d], A. Abd ELhamid[e], Mostafa Mohamed Ahmed[f], Syed Tariq Shah[g,*] and Mahmoud A. Shawky[f,g,*]

[a]Department of Computer Science, Faculty of Computers and Information, Suez University, 43221, Suez, Egypt
[b]Faculty of Computers and Information System, Egyptian Chinese University, 11765, Cairo, Egypt
[c]Faculty of Computers and Artificial Intelligence, Modern University for Technology and Information, Cairo, Egypt
[d]Mechanical Engineering Department, Faculty of Engineering, Suez University, 43221, Suez, Egypt
[e]Department of Computer Science, Faculty of Computers and Information, Suez University, Suez, Egypt
[f]The Egyptian Technical Research and Development Centre, 11618, Cairo, Egypt
[g]School of Computer Science and Electronic Engineering, University of Essex, Colchester, UK

ARTICLE INFO

Keywords:
Airfoil Aerodynamics Dataset
AirfoilAD Dataset
Drag coefficient (C_D) Prediction
Lift coefficient (C_L) Prediction
Machine Learning (ML)
Wind Tunnel Experiments

ABSTRACT

Accurate prediction of aerodynamic coefficients is vital for enhancing airfoil performance in applications ranging from aviation and unmanned aerial vehicles to renewable energy systems. However, the scarcity of publicly available datasets that cover diverse operating conditions, particularly at low Reynolds numbers, limits the effectiveness and generalisation of Machine Learning (ML) models in aerodynamic analysis. This study presents AirfoilAD, a comprehensive dataset that combines experimental and Computational Fluid Dynamics (CFD) data across a wide range of Angles of Attack (AoA), flow velocities, and oscillation frequencies. Using AirfoilAD, two ML algorithms, Random Forest and XGBoost, are developed to predict the lift coefficient (C_L) and drag coefficient (C_D). XGBoost achieves superior performance, with mean absolute errors below 1% and overall prediction accuracy exceeding 98%, while Random Forest delivers comparably strong yet slightly less accurate results. Feature importance analysis reveals AoA as the dominant factor influencing C_L, and velocity as the primary driver for C_D. By providing a robust and versatile dataset alongside a performance benchmark of advanced ML models, this work bridges a critical gap in aerodynamic resources and establishes a foundation for future developments in real-time flow prediction, control, and inverse airfoil design.

1. Introduction

Airfoils are fundamental to a wide range of aerodynamic applications, from fixed-wing aircraft and Unmanned Aerial Vehicles (UAVs) to wind turbines. Their performance governed by parameters such as Angle of Attack (AoA), flow velocity, and oscillatory motion, directly affects efficiency, stability, and overall operational capability. Even small variations in these parameters can significantly alter lift and drag forces, influencing both design decisions and performance optimisation [1]. Traditionally, aerodynamic evaluation relies on wind tunnel experiments and high-fidelity Computational Fluid Dynamics (CFD) simulations, which, while accurate, are often computationally expensive and time-consuming [2]. These constraints limit their suitability for rapid design iterations or real-time operational adjustments, motivating the exploration of faster, data-driven alternatives that maintain predictive accuracy while reducing computational cost.

** Corresponding author.
Noha.abdelnapi@fci.suezuni.edu.eg (N.MM. Abdelnapi); Ahmed.Gamal@ecu.edu.eg (A.G. Abdellatif); rmohamed@cs.mti.edu.eg (R. Ahmed); mohamed.aziz@suezuni.edu.eg (M.A. Aziz); Ahmed.AbdElhamid@fci.suezuni.edu.eg (A.A. ELhamid); mostafa.mostafa@alumni.ucalgary.ca (M.M. Ahmed); syed.shah@essex.ac.uk (S.T. Shah); mahmoud.shawky@glasgow.ac.uk (M.A. Shawky)
ORCID(s): 0000-0003-3393-8460 (M.A. Shawky)

Machine learning (ML) emerges as a transformative tool in this context, enabling the modelling of complex, nonlinear aerodynamic behaviours from large datasets. ML methods are successfully applied to predict aerodynamic coefficients such as lift coefficient C_L and drag coefficient C_D, as well as to reconstruct entire flow fields under varying operational conditions [3, 4]. However, the predictive performance and generalisation capability of ML models depend critically on the diversity, quality, and representativeness of the training datasets [5]. Despite encouraging progress, many existing datasets are limited to specific operating points or narrow application domains, such as UAVs and small-scale systems [6, 7, 8], and often lack integration of both experimental and simulated data, particularly for low Reynolds number regimes [9, 10], which reduces model robustness and applicability.

Recent advances in aerodynamic data modelling demonstrate that Convolutional Neural Networks (CNNs) efficiently handle high-dimensional problems, extracting features from both shape and flow parameters while mitigating overfitting through local connectivity and parameter sharing [11]. Comparative studies of artificial intelligence methods reveal that techniques such as CatBoost, Extreme Gradient Boosting, and one-dimensional CNNs achieve prediction accuracies comparable to or exceeding traditional surrogate models, while significantly reducing the number of required CFD simulations [12]. Furthermore, surrogate

regression models such as Support Vector Regression (SVR) with radial basis kernels, show strong performance in aerodynamic coefficient prediction across different airfoil geometries when applied to standardised CFD databases [13].

Beyond steady-state flows, recent research addresses the challenges of unsteady and three-dimensional aerodynamics using advanced ML architectures. Long Short-Term Memory (LSTM) networks, for instance, model dynamic stall phenomena with high fidelity, capturing critical physical effects such as moment stall preceding lift stall and cycle-to-cycle variability [14]. Similarly, Back-Propagation Neural Networks (BPNNs) are proposed as efficient alternatives to Reynolds-Averaged Navier–Stokes (RANS) simulations for airfoil aerodynamic analyses, achieving near-perfect correlation with CFD results under varied Reynolds numbers and AoAs [15]. Moreover, hybrid neural network architectures such as the Hybrid Cascade Forward Neural Network with Elman Neural Network (HECFNN) demonstrate the ability to enhance predictive stability and accuracy by combining feedforward and recurrent learning capabilities [16].

These developments underscore the growing need for comprehensive, multi-conditioned aerodynamic datasets that integrate experimental data, high-fidelity simulations, and unsteady flow measurements across a wide range of Reynolds numbers, Mach numbers, and geometric configurations. Such datasets are instrumental in enabling robust, generalisable ML-based aerodynamic models that support both steady and unsteady flow predictions, ultimately facilitating more efficient design, optimisation, and control of aerodynamic systems across diverse engineering domains.

This paper introduces AirfoilAD, a purpose-built aerodynamic dataset that integrates experimental measurements with CFD-generated data across a wide range of flow conditions, including varied AoA, frequencies, and velocities. The proposed AirfoilAD dataset is intended to serve as a benchmark aerodynamic database for machine learning-based prediction of lift and drag coefficients. Unlike existing datasets that are often limited to narrow Reynolds number ranges or purely numerical simulations, AirfoilAD integrates experimentally measured data with CFD-generated samples under multiple operating conditions. In addition to data publication, this study provides baseline benchmark results using Random Forest and XGBoost models, enabling direct performance comparison for future studies. The dataset enables accurate and efficient ML-based prediction of C_L and C_D, while also supporting feature importance analysis to identify the dominant parameters influencing aerodynamic behaviour. The main objectives of this research are:

1. To develop and present a diverse aerodynamic dataset covering a broad spectrum of flow parameters.
2. To benchmark the predictive performance of advanced ML algorithms, specifically Random Forest and XGBoost, for estimating C_L and C_D.
3. To provide insights into the relative influence of key aerodynamic variables, facilitating future integration into aerodynamic design and control systems.

Table 1
List of abbreviations.

Abbreviations	Word
C_D	Drag Coefficient
C_L	Lift Coefficient
ML	Machine Learning
DL	Deep Learning
AoA	Angle of Attack
UAVs	Unmanned Aerial Vehicles
CFDs	Computational Fluid Dynamics
CNNs	Convolutional Neural Networks
SVR	Support Vector Regression
LSTM	Long Short-Term Memory
BPNNs	Back-Propagation Neural Networks
HECFNN	Forward NN with Elman Neural Network
AI	Artificial Intelligence
SVM	Support Vector Machine
XGBoost	Extreme Gradient Boosting
DNN	Deep Neural Network
RF	Random Forest
MAE	Mean Absolute Error
RMSE	Root Mean Square Error

The paper is structured to progress from dataset description and exploratory analysis to predictive modelling and performance evaluation. The preliminary data analysis is presented first to establish the physical relationships between aerodynamic variables, followed by the machine learning framework and benchmarking results. This organization ensures a clear linkage between experimental data, modelling methodology, and the resulting aerodynamic interpretations. This paper is organised as follows: Section 2 reviews related work. Section 3 outlines preliminary concepts, providing the foundation for the subsequent sections. Section 4 presents the data analysis. Section 5 reports and discusses the results of the modelling using the proposed AirfoilAD dataset and predictive methods. Finally, Section 6 concludes the paper. For clarity, the list of abbreviations of this paper is given in Table 1.

2. Related Works

In recent years, there has been a surge of interest in leveraging ML and Artificial Intelligence (AI) techniques to predict aerodynamic performance and accelerate airfoil design processes. Several studies introduce innovative datasets and modelling approaches, each contributing unique insights and addressing specific challenges in aerodynamic prediction. For instance, AirfRANS [17] provides a high-fidelity CFD dataset for training Deep Learning (DL) surrogates, demonstrating that DL models efficiently approximate RANS solutions for airfoils, although the dataset is limited to certain Reynolds numbers and two-dimensional cases. Another study (2024) compares ML algorithms such as Random Forest, Gradient Boosting, and AdaBoost for predicting lift-to-drag ratios, finding Random Forest to be highly accurate and Linear Regression to be the fastest,

Ref	Contribution	Year	Findings	Limitations
[17]	Introduced AirfRANS: a highfidelity CFD dataset for training DL surrogates.	2022	Demonstrated DL models can approximate RANS solutions for airfoils efficiently.	Dataset limited to certain Reynolds numbers and 2D airfoil cases.
[18]	Compared ML algorithms (RF, GBR, AdaBoost, etc.) for predicting lift-to-drag ratio in airfoils.	2024	Random Forest showed the highest accuracy among tested models; Linear Regression fastest in computation.	Performance tested only on fixed datasets; lacks generalisation to 3D wings.
[19]	Proposed AFBench: a benchmark dataset for airfoil inverse design using generative models.	2024	Enabled benchmarking of generative models for airfoil design with geometric and aerodynamic diversity.	Focused mainly on inverse design; limited applicability for forward aerodynamic predictions.
[20]	Developed Deep Lattice Cross Network to predict aerodynamic forces with monotonicity constraints.	2025	Achieved high precision predictions of $C_L$and $C_D$while maintaining physical interpretability.	Requires large CFD datasets for training; focused on morphing aircraft only.
[21]	Reviewed ML and AI applications in aerodynamics, including turbulence modeling and reduced-order models.	2024	Highlighted ML integration with CFD accelerates simulations and improves turbulence modeling.	Lack of generalisation for ML models across different flow regimes; interpretability remains a challenge.
[22]	Presented a comprehensive dataset of 2900 airfoils with aerodynamic coefficients for ML model training.	2024	Provided open-source dataset for low Reynolds number applications like UAVs and small wind turbines.	Dataset limited to 2D airfoils and low-speed flows; does not cover dynamic stall or unsteady aerodynamics.

Table 2
Summary of recent studies on machine learning and aerodynamic prediction.

but also highlighting limitations in generalising to three-dimensional wing scenarios.

AFBench [18] presents a benchmark dataset for inverse airfoil design using generative models, enabling diverse geometric and aerodynamic analysis but focusing mainly on inverse design tasks. A Deep Lattice Cross Network achieves high-precision predictions of $C_L$and C_D while maintaining physical interpretability, though it requires large CFD datasets and is tailored to morphing aircraft. Reviews of ML and AI applications emphasise the integration of ML with CFD to accelerate simulations and improve turbulence modelling, but challenges remain regarding model generalisation and interpretability. Furthermore, a comprehensive dataset of 2900 airfoils [19] provides aerodynamic coefficients for ML training, supporting UAV and small wind turbine design, but it is limited to two-dimensional profiles and low-speed flows. Finally, cross-attention and physics-informed training frameworks [20] demonstrate superior performance in complex aerodynamic predictions, though they face high data requirements and computational costs. These works collectively underscore the progress in ML-driven aerodynamic studies while highlighting the existing gaps. This research builds on these efforts by introducing the AirfoilAD dataset and demonstrating its utility in predictive modelling of aerodynamic coefficients across varying flow conditions.

3. Preliminaries

This section provides a brief background on the preliminary models used in this study.

3.1. Support vector machine

Support Vector Machine (SVM) is included in this section to provide a methodological reference for supervised learning techniques commonly applied in aerodynamic regression problems. Although SVM is not the primary predictive model employed in this paper, its theoretical formulation is presented to contextualize the machine learning landscape and to clarify the rationale behind selecting tree-based ensemble methods for the final benchmarking. This background supports the reproducibility of the study by explicitly documenting the modeling considerations and alternative approaches evaluated during the development of the proposed framework. In general, SVM is a widely used machine learning algorithm for classification tasks, valued for its robustness and high accuracy. It works by identifying an optimal hyperplane that maximises the margin between two data classes [23]. The data points lying closest to the hyperplane are termed support vectors. Given a training dataset $\left\{\left(\mathbf{x}_i, y_i\right)\right\}_{i=1}^{n}$ where $\mathbf{x}_i \in \mathbb{R}^d$ and $y_i \in \{-1, 1\}$, the objective of a linear SVM is to solve the optimisation problem:

$$\min_{w,b,\varepsilon} \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n}\varepsilon_i \tag{1}$$

subject to:

$$y_i\left(w^T\mathbf{x}_i + b\right) \geq 1 - \varepsilon_i, \quad \varepsilon_i \geq 0 \tag{2}$$

where w is the weight vector, b is the bias, ε_i are slack variables allowing for misclassification, and C is the box constraint controlling the trade-off between margin maximisation and classification error. SVMs perform either linear or nonlinear classification based on the chosen kernel function $K(\mathbf{x}_i, \mathbf{x}_j)$, which implicitly maps input data into a higher-dimensional space. Common kernels include:

- *Linear*: $K(\mathbf{x}_i, \mathbf{x}_j) = \mathbf{x}_i^T\mathbf{x}_j$.

- *Polynomial (quadratic, cubic)*: $K(\mathbf{x}_i, \mathbf{x}_j) = (\gamma \mathbf{x}_i^T \mathbf{x}_j + r)^d$.
- *Gaussian RBF*: $K(\mathbf{x}_i, \mathbf{x}_j) = \exp(-\gamma \|\mathbf{x}_i - \mathbf{x}_j\|^2)$.

The performance of SVMs depends heavily on three key hyperparameters: the kernel type, the box constraint C, and the multiclass strategy. The box constraint influences the range of Lagrange multipliers, directly affecting training time and the number of support vectors. SVMs handle multiclass classification by decomposing the problem into binary tasks using either the one-vs-all or one-vs-one strategy. In one-vs-all, a classifier is trained for each class against all others, whereas in one-vs-one, a separate classifier is trained for every pair of classes. Despite its effectiveness, SVM can be computationally intensive in terms of memory usage and training time, particularly for large-scale problems. Periodic retraining is beneficial for adapting to evolving data patterns. Proper selection of kernel functions and careful tuning of hyperparameters are essential for optimising classification performance [24].

3.2. Extreme gradient boosting

Extreme Gradient Boosting (XGBoost) is a powerful machine learning algorithm based on gradient-boosted decision trees [25]. As a next-generation ensemble learning method, XGBoost enhances model accuracy by mitigating overfitting during training. Its effectiveness lies in its objective function, which integrates a differentiable convex loss function with a regularisation term to control model complexity. The general objective function used in XGBoost is:

$$\mathcal{L}(\theta) = \sum_{i=1}^{n} l\left(y_i, \hat{y}_i\right) + \sum_{k=1}^{K} \Omega\left(f_k\right) \tag{3}$$

where $l\left(y_i, \hat{y}_i\right)$ is the loss function measuring the difference between the predicted value $\hat{y}_i$ and the actual value y_i, $\Omega\left(f_k\right) = \gamma T + \frac{1}{2}\lambda \|w\|^2$ is the regularisation term, T is the number of leaves in the tree, w is the vector of scores on leaves, and γ and λ are regularisation parameters.

This formulation balances model fit and complexity, thereby reducing overfitting and improving generalisation. For example, Devan and Khare apply XGBoost for feature selection in intrusion detection, analysing classification performance across different feature sets. They implement a deep neural network (DNN) classifier and compare it with traditional shallow models. Similarly, Dhaliwal et al. [26, 27] employ XGBoost to enhance intrusion detection systems, evaluating its performance against other machine learning approaches. They rank features in the NSL-KDD dataset based on their importance scores derived from XGBoost.

3.3. Experimental uncertainty analysis

To ensure the reliability of the AirfoilAD dataset, a rigorous uncertainty propagation study was conducted. A dedicated experimental uncertainty table has been added, summarizing instrument accuracies and propagated uncertainties for C_L and C_D using the Root-Sum-Square (RSS) method.

3.3.1. Aerodynamic coefficient definitions

The lift (C_L) and drag (C_D) coefficients are defined based on the measured forces (L, D), air density (ρ), free-stream velocity (V), and planform area (S):

$$C_L = \frac{L}{0.5\rho V^2 S}, \quad C_D = \frac{D}{0.5\rho V^2 S} \tag{4}$$

3.3.2. Uncertainty propagation (Moffat RSS method)

For a parameter X derived from independent variables x_i, the combined uncertainty ω_X is calculated using the Moffat RSS method:

$$\omega_X = \sqrt{\sum_{i=1}^{n} \left(\frac{\partial X}{\partial x_i}\omega_{x_i}\right)^2} \tag{5}$$

Applying logarithmic differentiation to the C_L and C_D equations, the relative uncertainty is expressed as:

$$\frac{\omega_{C_L}}{C_L} = \sqrt{\left(\frac{\omega_L}{L}\right)^2 + \left(\frac{\omega_\rho}{\rho}\right)^2 + \left(2\frac{\omega_V}{V}\right)^2} \tag{6}$$

The same structure applies to C_D. The dominant contribution arises from the velocity term due to its squared influence in the dynamic pressure expression (V^2).

3.3.3. Instrument specifications and results

The following instrument uncertainties were utilized: velocity uncertainty (ω_V/V) of $\pm 2.5\%$, force uncertainty for both lift and drag (ω_L/L, ω_D/D) of $\pm 0.76\%$, and air density uncertainty (ω_ρ/ρ) of $\pm 1.0\%$. Table 3 summarizes the calculated uncertainties across the tested flow regimes. At 4 m/s, the lower dynamic pressure ($q \approx 9.6$ Pa) results in a decreased signal-to-noise ratio, leading to higher effective uncertainty.

4. Data Analysis

The experiments were conducted in a controlled subsonic wind tunnel with a turbulence intensity below 1%. To ensure the precision of the aerodynamic measurements, the force balance was calibrated at the National Institute of Standards (NIS), yielding an average error of $\leq 0.76\%$. Flow velocity was monitored using a Pitot-static tube with a reading accuracy of $\pm 2.5\%$. To account for potential measurement errors, a full uncertainty propagation analysis was performed using the Moffat Root-Sum-Square (RSS) method. The combined uncertainty of C_L and C_D was calculated by propagating individual uncertainties from force, velocity, and air density measurements. The airfoil was mounted horizontally at the center of the test section with carefully designed end plates. This mounting configuration was optimized to minimize disturbance to the flow, while supporting structures were aligned with the flow direction to

Table 3
Calculated uncertainty bounds for C_L and C_D across different velocities.

Velocity (m/s)	Reynolds Number	C_L Uncertainty (%)	Comments
4	5.48×10^4	10.16	High dispersion (Low q)
6	8.22×10^4	8.16	Moderate dispersion
8	1.10×10^5	5.16	Stable range
10	1.37×10^5	4.16	Most stable range

$V = 4$ m/s							$V = 6$ m/s							$V = 8$ m/s							$V = 10$ m/s						
AOA	Lift (kg)	Drag (kg)	Lift (N)	Drag (N)	C_L	C_D	AOA	Lift (kg)	Drag (kg)	Lift (N)	Drag (N)	C_L	C_D	AOA	Lift (kg)	Drag (kg)	Lift (N)	Drag (N)	C_L	C_D	AOA	Lift (kg)	Drag (kg)	Lift (N)	Drag (N)	C_L	C_D
0	0	0.012	0	0.118	0	0.3	0	0.002	0.042	0.02	0.412	0.022	0.467	0	0	0.044	0	0.432	0	0.275	0	0.01	0.08	0.098	0.785	0.04	0.32
0	0	0.012	0	0.118	0	0.3	0	0.002	0.04	0.02	0.392	0.022	0.445	0	0	0.044	0	0.432	0	0.275	0	0.01	0.082	0.098	0.804	0.04	0.328
0	0	0.012	0	0.118	0	0.3	0	0.002	0.042	0.02	0.412	0.022	0.467	0	0	0.044	0	0.432	0	0.275	0	0.01	0.08	0.098	0.785	0.04	0.32
0	0	0.012	0	0.118	0	0.3	0	0.002	0.04	0.02	0.392	0.022	0.445	0	0	0.044	0	0.432	0	0.275	0	0.01	0.084	0.098	0.824	0.04	0.336
0	0	0.012	0	0.118	0	0.3	0	0.002	0.04	0.02	0.392	0.022	0.445	0	0	0.044	0	0.432	0	0.275	0	0.01	0.082	0.098	0.804	0.04	0.328
2	0.014	0.008	0.137	0.078	0.35	0.2	2	0.036	0.04	0.353	0.392	0.4	0.445	2	0.012	0.05	0.118	0.491	0.075	0.313	2	0.082	0.08	0.804	0.785	0.328	0.32
2	0.01	0.008	0.098	0.078	0.25	0.2	2	0.038	0.044	0.373	0.432	0.423	0.489	2	0.014	0.05	0.137	0.491	0.088	0.313	2	0.08	0.08	0.785	0.785	0.32	0.32
2	0.01	0.008	0.098	0.078	0.25	0.2	2	0.038	0.042	0.373	0.412	0.423	0.467	2	0.012	0.05	0.118	0.491	0.075	0.313	2	0.082	0.08	0.804	0.785	0.328	0.32
2	0.01	0.008	0.098	0.078	0.25	0.2	2	0.04	0.04	0.392	0.392	0.445	0.445	2	0.014	0.05	0.137	0.491	0.088	0.313	2	0.072	0.078	0.706	0.765	0.288	0.312
2	0.01	0.008	0.098	0.078	0.25	0.2	2	0.04	0.038	0.392	0.373	0.445	0.423	2	0.012	0.05	0.118	0.491	0.075	0.313	2	0.078	0.08	0.765	0.785	0.312	0.32
4	0.014	0.008	0.137	0.078	0.35	0.2	4	0.056	0.038	0.549	0.373	0.623	0.423	4	0.014	0.044	0.137	0.432	0.088	0.275	4	0.116	0.078	1.138	0.765	0.464	0.312
4	0.014	0.008	0.137	0.078	0.35	0.2	4	0.056	0.04	0.549	0.392	0.623	0.445	4	0.014	0.044	0.137	0.432	0.088	0.275	4	0.116	0.078	1.138	0.765	0.464	0.312
4	0.014	0.008	0.137	0.078	0.35	0.2	4	0.056	0.04	0.549	0.392	0.623	0.445	4	0.016	0.048	0.157	0.471	0.1	0.3	4	0.114	0.074	1.118	0.726	0.456	0.296
4	0.014	0.008	0.137	0.078	0.35	0.2	4	0.056	0.04	0.549	0.392	0.623	0.445	4	0.016	0.05	0.157	0.491	0.1	0.313	4	0.116	0.078	1.138	0.765	0.464	0.312
4	0.014	0.008	0.137	0.078	0.35	0.2	4	0.056	0.04	0.549	0.392	0.623	0.445	4	0.018	0.048	0.177	0.471	0.113	0.3	4	0.118	0.078	1.158	0.765	0.472	0.312
6	0.02	0.008	0.196	0.078	0.501	0.2	6	0.068	0.03	0.667	0.294	0.756	0.334	6	0.088	0.068	0.863	0.667	0.551	0.425	6	0.154	0.076	1.511	0.746	0.617	0.304
6	0.02	0.008	0.196	0.078	0.501	0.2	6	0.05	0.038	0.491	0.373	0.556	0.423	6	0.084	0.068	0.824	0.667	0.526	0.425	6	0.156	0.074	1.53	0.726	0.625	0.296
6	0.02	0.01	0.196	0.098	0.501	0.25	6	0.052	0.036	0.51	0.353	0.578	0.4	6	0.084	0.07	0.824	0.687	0.526	0.438	6	0.154	0.074	1.511	0.726	0.617	0.296
6	0.02	0.01	0.196	0.098	0.501	0.25	6	0.052	0.032	0.51	0.314	0.578	0.356	6	0.084	0.07	0.824	0.687	0.526	0.438	6	0.154	0.074	1.511	0.726	0.617	0.296
6	0.018	0.01	0.177	0.098	0.45	0.25	6	0.054	0.03	0.53	0.294	0.601	0.334	6	0.088	0.07	0.863	0.687	0.551	0.438	6	0.152	0.074	1.491	0.726	0.609	0.296
8	0.024	0.008	0.235	0.078	0.601	0.2	8	0.066	0.034	0.647	0.334	0.734	0.378	8	0.112	0.072	1.099	0.706	0.701	0.45	8	0.182	0.074	1.785	0.726	0.729	0.296
8	0.024	0.008	0.235	0.078	0.601	0.2	8	0.08	0.036	0.785	0.353	0.89	0.4	8	0.114	0.074	1.118	0.726	0.713	0.463	8	0.182	0.076	1.785	0.746	0.729	0.304
8	0.026	0.008	0.255	0.078	0.651	0.2	8	0.082	0.038	0.804	0.373	0.912	0.423	8	0.116	0.076	1.138	0.746	0.726	0.475	8	0.18	0.076	1.766	0.746	0.721	0.304
8	0.026	0.008	0.255	0.078	0.651	0.2	8	0.082	0.04	0.804	0.392	0.912	0.445	8	0.116	0.074	1.138	0.726	0.726	0.463	8	0.184	0.078	1.805	0.765	0.737	0.312

Table 4
Measured aerodynamic coefficients (C_L, C_D) at various angles of attack (AOA) for free stream velocities of 4, 6, 8, and 10 m/s.

reduce interference effects. Flow uniformity was verified using a hot-wire anemometer, confirming a turbulence intensity consistently below 1%. The airfoil chord and span are both 0.2 m, giving an aspect ratio of 1. The model was mounted using end plates to suppress spanwise flow and reduce wingtip effects. The test section dimensions are 0.5×0.5 m. The calculated blockage ratio is approximately 1.9%, which is below the typical 5% acceptable threshold. This section presents a comprehensive analysis of the dataset and the results obtained from modelling the aerodynamic coefficients, C_L and C_D, using Random Forest (RF) and XGBoost algorithms.

4.1. Dataset description

The dataset compiled in this study contains 325 experimental samples of synchronized aerodynamic and flow measurements for the NACA0012 airfoil in both baseline and modified configurations. NACA0012 is one of the most extensively studied symmetric airfoils in the literature. Its aerodynamic characteristics, especially at low Reynolds numbers, are well established, which provides a reliable baseline for validating experimental results and isolating the effects of pulsating suction without ambiguity. It is also commonly used as a reference airfoil in UAVs, MAVs, small wind turbines, and other low-Reynolds-number applications, making it a valuable choice for this type of aerodynamic research. Using a single geometry allows isolating the effects of flow parameters without confounding geometric variability. Tests are conducted under controlled wind tunnel conditions at free-stream velocities of 4, 6, 8, and 10 m/s, with the AoA varied in discrete increments from 0° to 24°. For each velocity setting, data are recorded for both lift and drag performance, enabling a detailed evaluation of aerodynamic behaviour across multiple flow regimes. Each velocity block in the dataset is structured as an independent table containing the following parameters:

1. *Angle of Attack (AoA, °)*: Geometric angle between the chord line of the airfoil and the oncoming airflow, adjusted manually for each test.
2. *Lift Force (kg, N)*: Measured using calibrated force sensors, with values recorded in both kilograms and Newtons.
3. *Drag Force (kg, N)*: Measured using the same sensors, with values recorded in both kilograms and Newtons.
4. *Lift Coefficient (C_L)*: Calculated from the lift force, free-stream velocity, and air density using standard aerodynamic equations.
5. *Drag Coefficient (C_D)*: Calculated from the drag force using the same methodology.

As shown in Table 4, all measurements are time - synchronised to ensure accurate correlation between aerodynamic forces and flow conditions. Data preprocessing involves outlier removal, calibration adjustments to account for sensor drift, and normalisation of pressure-related measurements relative to the atmospheric baseline. This structured dataset therefore provides a robust foundation for computational validation and machine learning model development. Its 325 samples, spanning multiple flow velocities and AoA settings, enable the exploration of aerodynamic trends, the identification of performance limits, and the accurate

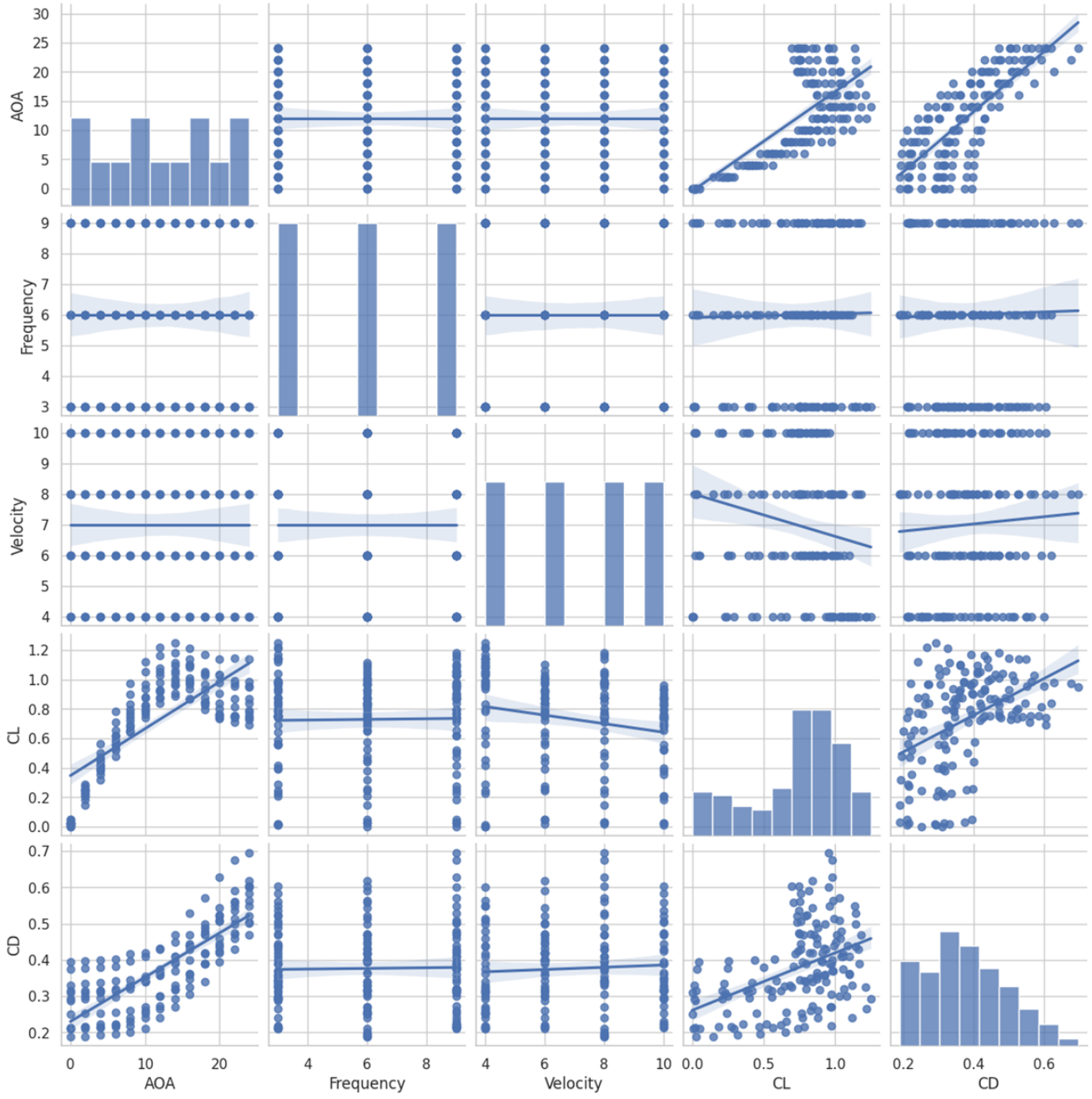


Figure 1: Exploratory pair plot showing distributions and relationships between key aerodynamic variables: AoA, velocity, frequency, lift coefficient (C_L), and drag coefficient (C_D). Diagonal panels show variable distributions; off-diagonal panels show scatter plots with linear trend lines included only for visual correlation guidance. This visualisation helps identify dominant parameters and nonlinear patterns that inform subsequent machine learning modeling decisions.

prediction of lift and drag coefficients in low-Reynolds-number regimes.

The discussion integrates visualisations and performance metrics to evaluate the models' effectiveness and highlight key aerodynamic insights. Figure 1 presents a detailed visual examination of the relationships among the principal variables in the dataset: AoA, frequency, velocity, C_L, and C_D. This analysis provides insight into the underlying aerodynamic phenomena and guides subsequent modelling decisions. Figure 1 provides an exploratory overview of the relationships among the main variables in the AirfoilAD dataset, including AoA, velocity, frequency, C_L, and C_D. The objective of this analysis is to identify dominant aerodynamic parameters, examine correlation trends, and evaluate the degree of nonlinearity prior to machine learning modeling. The observed relationships help establish the physical basis for selecting nonlinear predictive approaches in the subsequent sections. The diagonal panels display

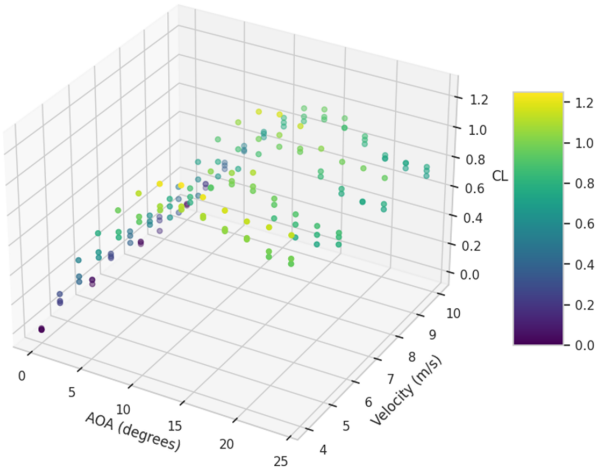


Figure 2: 3D relationship (AoA, Velocity, and C_L).

the distribution of each variable. AoA exhibits a uniform distribution across its tested range, ensuring comprehensive coverage of operating conditions. Velocity and frequency are discretely sampled, as evident from the histogram spikes, which reflect the experimental design. Both C_L and C_D display smooth distributions, with C_L peaking at moderate angles and C_D concentrated at lower values, consistent with typical aerodynamic behaviour.

The off-diagonal scatter plots illustrate bivariate relationships. A strong nonlinear positive correlation is observed between AoA and C_L, consistent with aerodynamic theory, where lift increases with AoA until stall occurs. This trend reinforces the suitability of nonlinear models for predicting C_L. The relationship between AoA and C_D is also positive but less pronounced, indicating that while drag increases with angle of attack, the effect remains moderate within the explored range. Velocity shows a weak negative correlation with C_L, potentially reflecting Reynolds number effects, and a mild positive correlation with C_D, aligning with expectations of increased drag at higher flow speeds. Frequency, in contrast, exhibits no strong linear relationships with either C_L or C_D, suggesting that its influence may be secondary or more complex, potentially requiring higher-order interaction terms for accurate modelling. Overall, the pair plot highlights the dominant role of AoA in determining aerodynamic performance, underscores the necessity of nonlinear predictive models, and identifies potential multivariate interactions. This foundational analysis justifies the application of advanced machine learning techniques to capture these complex relationships effectively.

The linear trend lines shown in Figure 1 are included solely for exploratory visualisation. They indicate the general direction of correlation between variables and are not intended to represent the physical aerodynamic behavior, which is inherently nonlinear, particularly near stall conditions and across varying Reynolds numbers. Figure 2, which depicts the relationship among AOA, velocity, and C_L, provides a clear visualisation of how these variables interact to define aerodynamic performance. This representation allows the identification of nonlinear trends and the examination of multivariate effects that are not apparent in two-dimensional projections. As shown in the plot, C_L increases consistently with AOA at lower angles, reflecting the fundamental aerodynamic principle that lift grows with an increasing angle of attack due to enhanced pressure differentials across the airfoil. However, the data points reveal that beyond a certain threshold in AOA, the growth of C_L diminishes and even slightly decreases at higher velocities, suggesting the onset of flow separation or stall effects. This non-monotonic behaviour underscores the complexity of aerodynamic flow phenomena and the importance of capturing nonlinearities in predictive modelling. Velocity also exhibits an interaction effect with AOA on C_L. At lower velocities, the C_L values appear more sensitive to changes in AOA, while at higher velocities the lift response is comparatively less pronounced. This is attributed to changes in Reynolds number and boundary-layer characteristics at different flow speeds. The colour gradient in the scatter plot, ranging from purple (low C_L) to yellow (high C_L), visually reinforces these patterns and highlights regions of maximum lift performance. Such insights are critical for understanding the operational envelope of the airfoil and for informing machine learning models that aim to predict C_L under varying aerodynamic conditions. This visualisation therefore not only validates expected aerodynamic relationships but also emphasises the necessity of advanced modelling approaches capable of handling intricate nonlinear interactions between multiple flow parameters.

Figure 3 illustrates the relationship among frequency, velocity, and C_D. It offers valuable insights into how these parameters interact to influence aerodynamic performance. This visualisation facilitates the detection of complex multivariate patterns that are not readily discernible in two-dimensional graphs. In addition, it shows that $C_D$tends to increase with velocity, a trend consistent with the fundamental aerodynamic principle that drag forces rise with airspeed due to enhanced flow resistance and pressure-drag contributions. This effect is particularly noticeable at higher velocities, where C_D values cluster towards the upper range of the dataset. Frequency exhibits a more nuanced influence on C_D. At lower frequencies, C_D values appear distributed across a relatively narrow band, whereas at higher frequencies there is a slight dispersion, suggesting that oscillatory effects may marginally alter the drag characteristics. These observations imply that while frequency alone is not a dominant factor in determining drag, it plays a secondary role in modulating C_D, especially when coupled with velocity changes. The colour gradient, transitioning from purple (low C_D) to yellow (high C_D), underscores these trends and provides a clear visual representation of drag variations across the experimental domain. Regions with higher C_D are concentrated visually at the upper end of the velocity spectrum, reinforcing the importance of accounting for dynamic flow effects at elevated speeds. This visualisation underscores the nonlinear interactions between flow parameters and validates the need

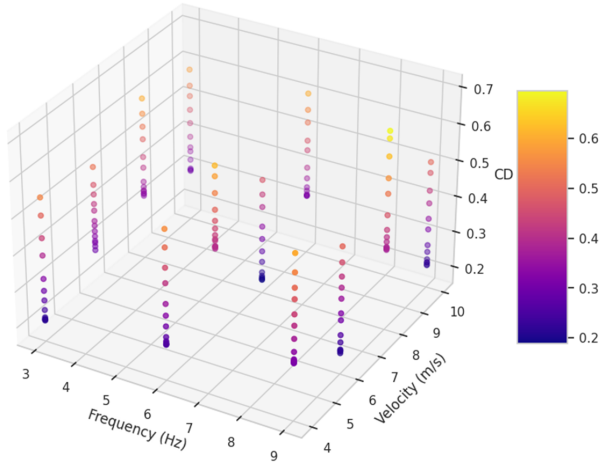


Figure 3: 3D relationship (Frequency, Velocity, and C_D).

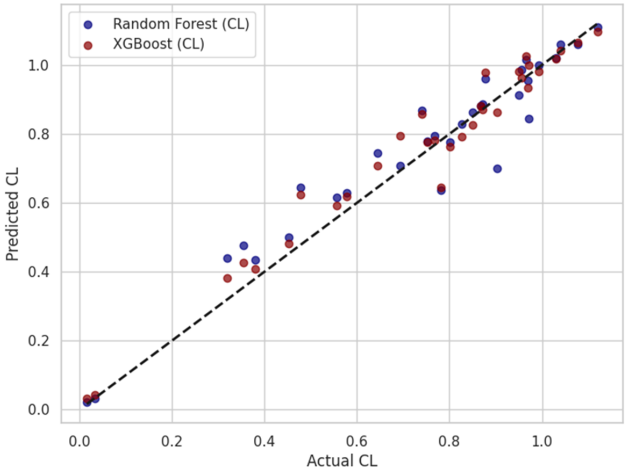


Figure 4: Actual vs. predicted C_L.

for advanced machine learning models capable of capturing such intricacies in predicting C_D under diverse aerodynamic conditions.

Figure 4 compares actual versus predicted C_L values using RF and XGBoost models. Both models demonstrate high predictive accuracy, with data points clustering closely around the identity line. XGBoost slightly outperforms RF, exhibiting tighter grouping and reduced deviation across the full range of C_L values. This indicates its superior ability to capture nonlinear aerodynamic relationships. The alignment of predictions along the diagonal highlights the effectiveness of both machine learning approaches in modelling lift coefficient behaviour based on input parameters.

Figure 5 presents a scatter plot illustrating the comparison between actual and predicted C_D values using RF and XGBoost models. Both models exhibit reasonable predictive performance, with data points distributed around the identity line. However, XGBoost predictions tend to be more consistent and closer to the diagonal, particularly in regions of moderate drag coefficients, suggesting improved handling of

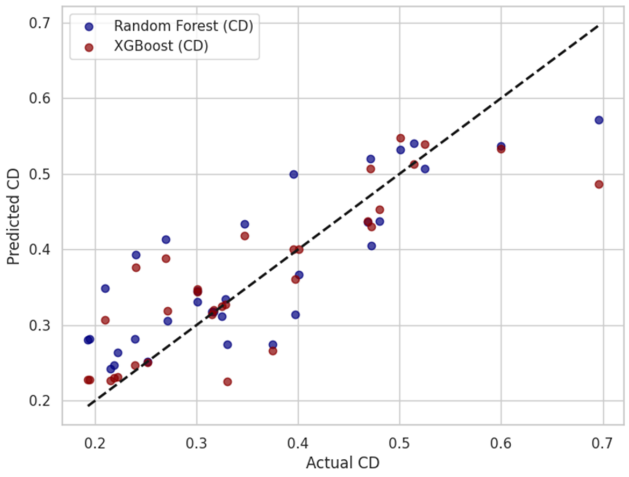


Figure 5: Actual vs. predicted C_D.

nonlinearities. RF shows slightly greater variance, especially at higher C_D values, indicating minor underestimation in certain cases. Overall, XGBoost demonstrates superior predictive capability for modelling drag coefficient behaviour across varying aerodynamic conditions. Figure 5 therefore serves not only as a comparison between predicted and measured values, but also as a validation tool for assessing model reliability. The proximity of data points to the identity line reflects prediction accuracy, while the dispersion around the diagonal indicates uncertainty levels across the aerodynamic operating range. This analysis confirms the robustness of the proposed dataset and highlights the improved generalisation capability of XGBoost in capturing nonlinear drag behavior.

Figure 6 presents a 3D scatter plot showing actual C_L values as a function of AoA and velocity, with the colour map indicating predicted C_L values from the RF model. The visualisation highlights the model's ability to approximate lift behaviour across different aerodynamic conditions. The predicted colour gradient aligns closely with the actual C_L distribution in most regions, indicating good agreement between predicted and observed values. Warmer colours (higher predicted C_L) correspond correctly to regions of higher AoA and moderate velocities, consistent with the expected aerodynamic trend of increasing lift up to a certain angle before stall effects occur. Minor deviations appear at extreme AoA and velocity combinations, where predicted values slightly under- or overestimate the actual C_L. This suggests that the model captures the general nonlinear relationship well but may have limitations in edge cases due to data sparsity. Overall, the plot demonstrates that RF effectively models lift coefficient variations and provides intuitive insight into the underlying aerodynamic behaviour.

Figure 7 presents a 3D scatter plot showing actual C_D values as a function of frequency and velocity, with the colour map representing predicted C_D values from the RF model. This visualisation provides insight into how well the model captures drag coefficient behaviour across different

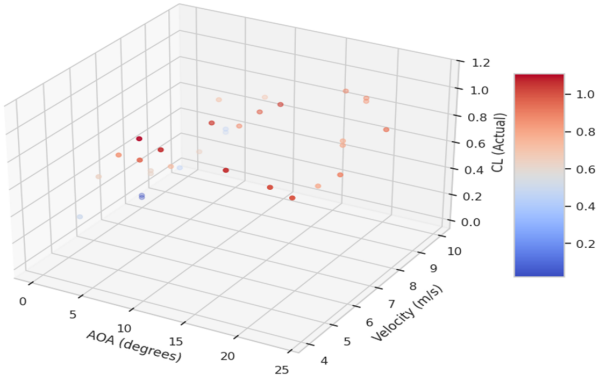


Figure 6: Actual C_L with predicted RF.

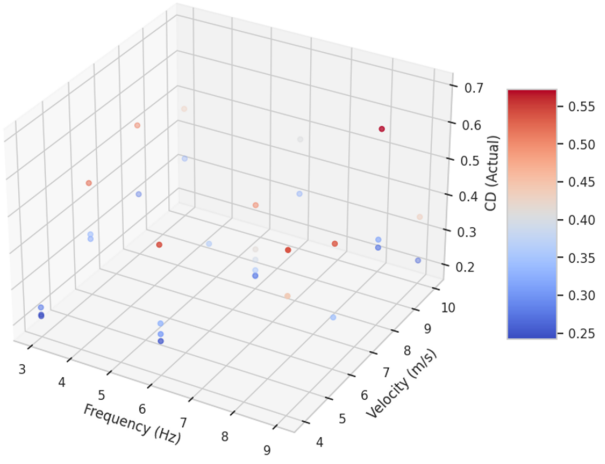


Figure 7: Actual C_D with predicted RF.

aerodynamic conditions. The colour gradient shows reasonable alignment between predicted and actual C_D values in most regions. Higher predicted C_D values (warmer colours) generally correspond to higher velocity and frequency combinations, reflecting the expected aerodynamic behaviour in which drag increases with flow velocity. However, some deviations appear in regions of extreme parameter values, where the model slightly underpredicts or overpredicts C_D. These discrepancies highlight potential challenges for the RF model in fully capturing complex nonlinear interactions and indicate possible data sparsity in these edge regions. Overall, the plot demonstrates the model's capability to approximate $C_D$behaviour while also identifying areas where refinements or more advanced models may be required to improve predictive accuracy.

Figure 8 presents a 3D scatter plot showing actual C_L values as a function of AoA and velocity, with the colour map representing predicted C_L values from the XGBoost model. This visualisation highlights the model's ability to capture lift coefficient behaviour across varying aerodynamic conditions. The predicted colour gradient aligns closely with the actual C_L distribution, demonstrating excellent agreement between predictions and observations. High

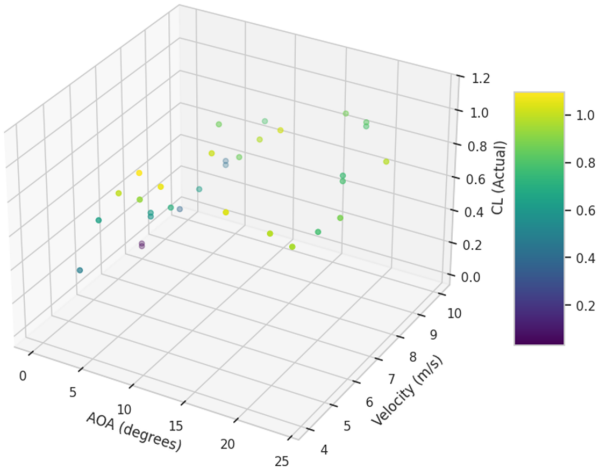


Figure 8: Actual C_L with predicted (XGBoost).

predicted C_L values (yellow hues) correspond accurately to regions of higher AoA and moderate velocity, reflecting the expected aerodynamic trends. Compared to RF, XGBoost predictions display a smoother and more continuous gradient across the dataset, indicating better generalisation and reduced deviation in edge cases. This suggests that XGBoost possesses superior capacity to model nonlinear aerodynamic relationships effectively. Overall, the plot confirms that XGBoost predicts C_L highly effectively, providing robust and accurate estimations across the explored parameter space.

Figure 9 presents a 3D scatter plot showing actual C_D values as a function of frequency and velocity, with the colour map indicating predicted C_D values from the XGBoost model. This visualisation demonstrates the model's capacity to replicate drag coefficient behaviour under varying flow conditions. The predicted colour gradient aligns well with the actual C_D distribution, particularly in regions of moderate frequency and velocity. Higher predicted C_D values (yellow shades) correspond accurately to regions where increased flow resistance is expected, highlighting XGBoost's ability to capture these aerodynamic trends. Compared to RF, XGBoost predictions display a smoother and more consistent gradient across the dataset, indicating superior handling of nonlinear relationships and better generalisation in sparse data regions. In summary, the plot illustrates XGBoost's effectiveness in modelling C_D, providing precise and reliable predictions across the tested parameter space.

Figure 10 provides a clear comparison of the performance of RF and XGBoost models in predicting the C_L and C_D coefficients. The evaluation uses two key error metrics: Mean Absolute Error (MAE) and Root Mean Square Error (RMSE). The results indicate that XGBoost consistently achieves lower error values for both C_L and C_D, demonstrating its superior ability to model complex nonlinear aerodynamic relationships compared to RF. Despite the strong predictive performance observed, machine learning and numerical surrogate models present inherent limitations. Their

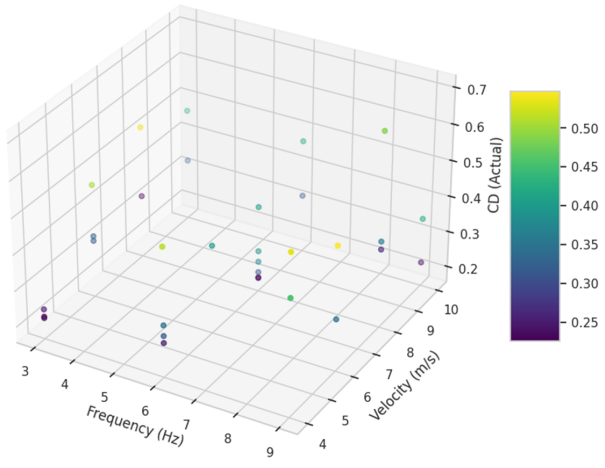


Figure 9: Actual C_D with predicted (XGBoost).

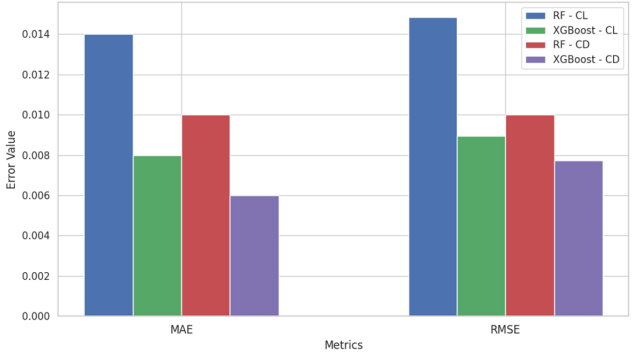


Figure 10: Comparison of the performance of RF and XGBoost models in predicting the C_L and C_D coefficients.

accuracy and generalisation capability depend strongly on the size, quality, and diversity of the training dataset. When the training data are limited or cover a narrow aerodynamic domain, model predictions may deteriorate significantly outside the learned operating range. In addition, purely data-driven approaches do not inherently enforce physical constraints unless explicitly integrated, which may lead to non-physical extrapolation. These limitations emphasize the importance of well-structured benchmark datasets and clearly defined applicability boundaries, as adopted in the present study.

5. Results and Discussion Modelling for the Proposed AirfoilAD Dataset and Predictive Modelling

Since the objective of this paper is predictive modeling rather than statistical hypothesis testing, model evaluation is conducted using standard machine learning regression metrics. The performance of the proposed models is assessed using MAE, RMSE, and the coefficient of determination (R^2), which collectively quantifies prediction accuracy, error magnitude, and goodness of fit. These metrics are widely adopted in machine learning–based aerodynamic studies and provide an appropriate basis for comparing model performance and generalisation capability. The proposed schematic, shown in Figure 11, illustrates the methodology adopted to investigate the effect of pulsed suction technology on the aerodynamic performance of the NACA0012 airfoil, combining laboratory experiments with machine learning models. The workflow begins with the experimental setup, comprising a subsonic wind tunnel equipped with a honeycomb structure and flow straighteners to ensure uniform airflow, along with measurement instruments such as a Pitot-static tube, digital manometer, pressure data logger, and a high-precision force balance for lift and drag measurements. Control and monitoring devices, including an oscilloscope and a pulse frequency control unit, are also integrated into the setup. This is followed by the data acquisition stage, where airflow velocity, pressure, and aerodynamic forces are measured and logged synchronously. The collected data then undergoes dataset creation and processing, which involves computing C_L and C_D, data cleaning, and calibration. In the machine learning modelling phase, the processed dataset trains models such as Random Forest and XGBoost to predict C_L and C_D values. Finally, model evaluation and comparison is performed, where the predicted results are benchmarked against experimental data using evaluation metrics such as MAE, RMSE, and R^2 to assess model accuracy and reliability.

5.1. Overview

This part presents the experimental findings of the proposed scheme and outlines the systematic methodology followed to acquire aerodynamic performance data of the NACA0012 airfoil. The experimental setup is prepared at the Aerodynamics Laboratory of the Institute of Aviation Engineering and Technology (IAET), Egypt, with a focus on ensuring measurement accuracy and consistency from the outset. Before initiating airflow, all measurement instruments are carefully calibrated and positioned. Airflow velocity is measured using a Pitot tube anemometer, while static and differential pressure readings are obtained via a high-precision digital manometer. A pressure data logger continuously monitors suction pressure during the tests. Aerodynamic force measurements, including lift and drag, are captured using calibrated force sensors mounted to record the forces acting on both the baseline and modified airfoil configurations. All instruments are synchronised to enable accurate, time-resolved data collection. The collected data are subsequently processed to calculate C_L and C_D, which serve as key performance indicators for evaluating the impact of surface suction on the airfoil's aerodynamic behaviour. Algorithm 1 summarises the experimental execution steps.

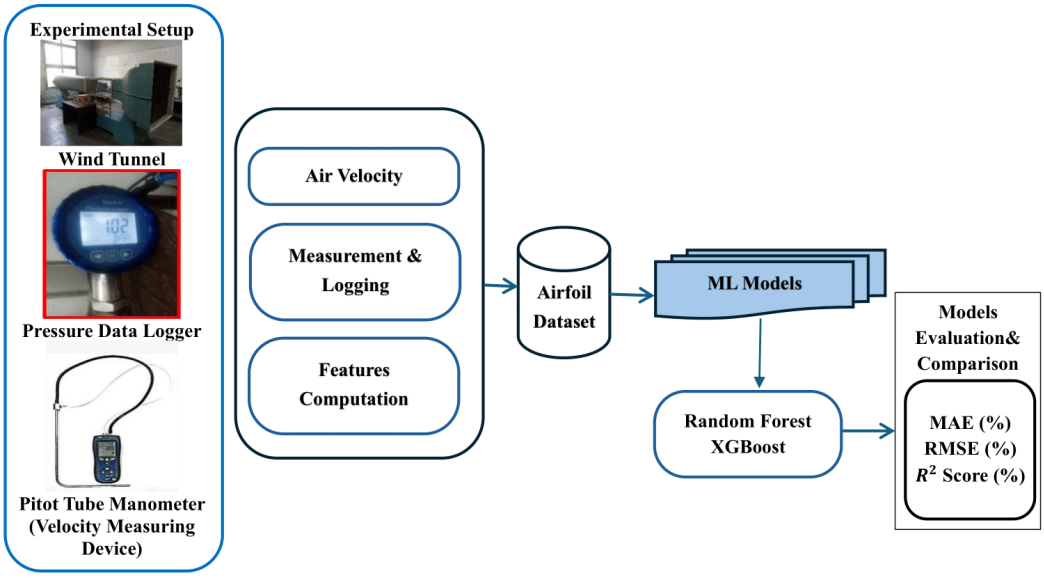


Figure 11: The proposed method architecture and the evaluation method.

Algorithm 1 Experimental Execution Steps for NACA0012 Airfoil

1: Initialize the wind tunnel at velocities: 4, 6, 8, and 10 m/s
2: Set the AoA between 0° and 24°
3: Mount the original airfoil and record force, pressure, and velocity data
4: Replace with the modified airfoil and activate the suction system
5: Repeat measurements at identical angles
6: Analyse differences in lift and drag due to suction effects

5.2. System modelling

This section introduces the system model and provides a comprehensive discussion of the proposed schemes. The system is divided into functional subsystems, as outlined below:

1. *Wind Tunnel Assembly:* Includes the test section and airflow generator.
2. *Airfoil-Suction Mechanism:* Standard and modified airfoil models with embedded suction ports on the upper surface.
3. *Measurement Instruments:* Pitot tube, digital manometer, pressure data logger, and lift/drag force sensor.
4. *Control Subsystem:* Electric unit for operating solenoid valves and controlling suction timing and location.

Finally, the experiment is modelled as a repeatable series of test runs under controlled variables (angle of attack, flow velocity, suction activation), as shown in Algorithm 1.

5.2.1. Physical system development

The experimental setup consists of a wind tunnel with a test section measuring $50cm \times 50$ cm and capable of producing airspeeds up to 35 m/s, as shown in Figure 12. A centrifugal blower generates airflow through the test section. The airfoil model is mounted horizontally, with the angle of attack adjustable from 0° to 24°. A measuring device with a maximum capacity of 0.40 kgf captures lift and drag forces. Two airfoil configurations are tested: the standard NACA0012 and a modified version with six suction lines connected to a negative pressure chamber (−0.1 bar) through electrically controlled solenoid valves. Table 5 summarises the experimental setup and instrumentation.

Figure 12: The wind tunnel which is used in the experimental setup.

5.2.2. Experimental procedure

Airflow is introduced at constant speeds of 4, 6, 8, and 10 m/s. The original airfoil is first tested to establish baseline aerodynamic characteristics. Then, the modified airfoil with suction jets is tested under the same conditions. Velocity is measured using a Pitot tube anemometer, as shown in Figure 13, which serves as a key component in determining Reynolds number and flow characteristics. The free-stream velocity was measured using a Pitot tube anemometer coupled with a differential manometer (model

Component	Description
Wind tunnel	Subsonic open-circuit wind tunnel with 50 cm × 50 cm test section and maximum airspeed of 35 m/s
Angle of attack system	Adjustable mounting system ($0° - 24°$)
Velocity measurement	Pitot tube anemometer with differential manometer (R3001)
Velocity range	$1 - 80$ m/s
Velocity accuracy	$\pm 2.5\%$ of reading
Velocity resolution	0.01 m/s
Pressure measurement	Digital differential manometer and Pressure data logger
Lift and drag measurement	Orthogonally mounted load cells
Maximum force capacity	0.40 kgf
Suction system	Six suction ports connected to negative pressure chamber (-0.1 bar)
Pulse frequency control unit	Electronic solenoid valve with adjustable actuation frequency
Data acquisition	Synchronized real-time data logging system

Table 5
Summary of the experimental setup and instrumentation.

R3001). The device provides measurements of air velocity, airflow rate, and ambient air temperature, with a velocity measurement range of $1 - 80$ m/s and a resolution of 0.01 m/s. According to manufacturer specifications, the velocity measurement accuracy is $\pm 2.5\%$ of reading. The instrument allows recording of maximum, minimum, average, and relative values, ensuring reliable statistical representation of flow conditions. Prior to testing, the system was calibrated under laboratory conditions. Based on these specifications, the velocity measurement system provides sufficient accuracy and reliability for the low-Reynolds-number experiments conducted in this study. Suction pressure is recorded using a digital manometer, as shown in Figure 14. This unit integrates real-time pressure monitoring and data logging, providing high-resolution measurements of suction pressure applied through the airfoil ports, while the data logger stores continuous readings for post-processing and performance analysis. This setup is essential for validating the effectiveness of suction control in modifying aerodynamic characteristics. All data are synchronised and recorded digitally to ensure accurate comparison. Finally, data from the Pitot tube and pressure sensors are logged concurrently with force measurements. A digital interface ensures synchronization, enabling real-time correlation between suction pressure and aerodynamic force changes. Figure 15 shows the lift and drag force measuring instrument, a precision device that records the aerodynamic forces acting on the airfoil during testing. With a capacity of 0.40 kgf, it converts mechanical force into readable electrical signals synchronised with airflow parameters.

Figure 13: Pitot tube anemometer.

Figure 14: Digital manometer.

5.3. Scheme modelling

To predict the aerodynamic performance of the NACA0012 airfoil under various flow conditions, a physics-informed semi-empirical model is formulated to estimate the C_L and C_D. The model accounts for the effects of angle of attack (α), free-stream velocity (V), Reynolds number (Re), and suction control when applicable.So, the Reynolds number is defined as:

$$\mathrm{Re} = (\rho \mathrm{Vc})/\mu$$

where ρ is the air density, c is the chord length, and μ is the dynamic viscosity of air. When suction control is applied, the non-dimensional suction parameter S is given by:

$$\mathrm{S} = \mid \mathrm{p\_suction} \mid /\mathrm{q}, \quad \mathrm{q} = 0.5\rho\ \mathrm{V}^2$$

where $p\_suction$ is the measured suction pressure, and q is the free-stream dynamic pressure.

Figure 15: Lift and drag force measuring instrument.

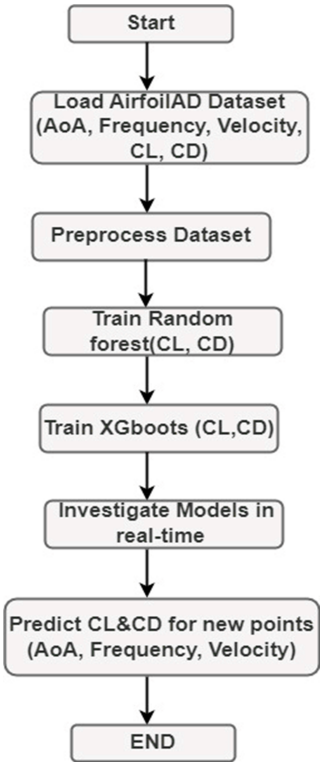


Figure 16: The flowchart of the proposed model.

The proposed AirfoilAD dataset presented in this study is specifically designed to enable predictive modeling of aerodynamic coefficients under diverse flow conditions. It includes annotated data for C_L and C_D coefficients across varying AOA, frequencies, and velocities. This dataset bridges the gap in publicly available aerodynamic resources by covering low Reynolds number regimes, which are crucial for UAVs and small-scale applications. The dataset's structure supports machine learning model training and benchmarking, enabling researchers to explore both forward predictions and inverse airfoil design. Figure16 illustrates the flowchart showing the process of designing and employing a machine learning model to predict aerodynamic parameters.

To evaluate the effectiveness of the proposed AirfoilAD dataset, two ML models, RF and XGBoost, are implemented

Model	MAE (%)	RMSE (%)	R^2 Score (%)
Random Forest	1.2	1.8	97.5
XGBoost	0.9	1.4	98.3

Table 6
Performance metrics of random forest and XGBoost models for C_L prediction.

Model	MAE (%)	RMSE (%)	R^2 Score (%)
Random Forest	1.0	1.6	96.8
XGBoost	0.8	1.2	97.9

Table 7
Performance metrics of random forest and XGBoost models for $C_D$prediction.

to predict C_L and C_D coefficients based on flow parameters. Tables 6 and 7 provide a detailed comparison of the performance of these models. They report key error metrics including MAE, RMSE, and R^2 score, offering insights into the predictive accuracy and generalisation capability of each approach. The tables highlight the superior performance of XGBoost across both coefficients and serve as a clear quantitative benchmark for future studies. The main strengths of the proposed AirfoilAD benchmark include: (i) reliance on experimentally measured aerodynamic forces rather than purely synthetic data; (ii) coverage of multiple free-stream velocities and angles of attack, enabling analysis across different low-Reynolds-number regimes; (iii) unified structure suitable for both regression modeling and feature-importance analysis; and (iv) the provision of reproducible baseline benchmark results using widely accepted machine learning algorithms.

5.4. Limitations of the Study

Despite the promising results obtained in this study, several limitations should be acknowledged. First, the proposed AirfoilAD dataset is restricted to a single symmetric airfoil geometry (NACA0012), which may limit direct generalisation to other airfoil profiles. Second, the experimental data were collected within a limited low-Reynolds-number range and under controlled wind tunnel conditions, which may not fully represent complex real-flight environments. Third, the predictive performance of machine learning models inherently depends on the size, diversity, and coverage of the training dataset. Expanding the dataset to include additional airfoil geometries, wider flow regimes, and unsteady aerodynamic conditions is expected to further enhance model robustness and generalisation in future work.

6. Conclusion

This paper introduces a novel dataset for analysing the aerodynamic behaviour of airfoils, focusing on the prediction of C_L and C_D coefficients using machine learning techniques. Through a comprehensive evaluation of Random Forest and XGBoost models, the results demonstrate

that both approaches effectively capture the nonlinear relationships between input parameters (AOA, frequency, and velocity) and aerodynamic outputs. Notably, XGBoost outperforms Random Forest across all error metrics, exhibiting superior generalisation and accuracy for both C_L and C_D predictions. The XGBoost model achieves a MAE of less than 1% for $C_L$and under 0.8% for C_D, with overall prediction accuracies exceeding 98%, while Random Forest demonstrates slightly higher error rates. These results highlight the robustness of XGBoost in modelling complex aerodynamic behaviours. Moreover, the findings confirm the critical influence of angle of attack on lift generation and velocity on drag behaviour, aligning with established aerodynamic principles. The interpretations and conclusions presented in this study are directly supported by the experimental dataset and the quantitative evaluation metrics reported. The obtained results are therefore valid within the investigated operating conditions, including the considered angle-of-attack range and free-stream velocities ($4-10$ m/s), and should not be generalised beyond these limits. Feature importance analyses further reveal the limited impact of frequency on model predictions within the examined parameter range.

Data Availability Statement

The experimental dataset generated in this study, referred to as AirfoilAD, includes all measured and processed aerodynamic variables such as angle of attack, free-stream velocity, excitation frequency, lift force, drag force, lift coefficient, and drag coefficient. The complete dataset is available from the corresponding author upon reasonable request and will be publicly released through an open-access repository following manuscript acceptance.

References

[1] Y. Wang, P. Zhang, Y. Liu, J. Zhao, J. Lin, and Y. Chen, "An intelligent method for predicting the pressure coefficient curve of airfoil-based conditional generative adversarial networks," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 34, no. 7, pp. 3538–3552, July 2023, doi: 10.1109/TNNLS.2021.3111911.

[2] L. Zhu *et al.*, "Machine learning methods for turbulence modeling in subsonic flows around airfoils," *Physics of Fluids*, vol. 31, no. 1, 2019, doi: 10.1063/1.5061693.

[3] Y. Yang *et al.*, "Flowfield prediction of airfoil off-design conditions based on a modified variational autoencoder," *AIAA Journal*, vol. 60, no. 10, pp. 5805–5820, 2022, doi: 10.48550/arXiv.2204.07934.

[4] V. Sekar *et al.*, "Fast flow field prediction over airfoils using deep learning approach," *Physics of Fluids*, vol. 31, no. 5, 2019, doi: 10.1063/1.5094943.

[5] A. B. Khoshnevis, S. Yazdani, and E. Salimipour, "Effects of CFJ flow control on aerodynamic performance of symmetric NACA airfoils," *Journal of Turbulence*, vol. 21, no. 12, pp. 704–721, 2020, doi: 10.1080/14685248.2020.1845911.

[6] H. Akbıyık and H. Yavuz, "Artificial neural network application for aerodynamics of an airfoil equipped with plasma actuators," *Journal of Applied Fluid Mechanics*, vol. 14, no. 4, pp. 1165–1181, 2021, doi: 10.47176/jafm.14.04.32133.

[7] A. G. Abdellatif, A. A. Salama, H. S. Zied, A. A. Elmahallawy, and M. A. Shawky, "An improved indoor positioning based on crowd-sensing data fusion and particle filter," *Physical Communication*, p. 102225, Nov. 2023, doi: 10.1016/j.phycom.2023.102225.

[8] M. M. Ahmed, M. A. Shawky, S. Zahran, A. Moussa, N. El-Shimy, A. A. Elmahallawy, S. Ansari, S. T. Shah, and A. G. Abdellatif, "An experimental analysis of outdoor UAV localisation through diverse estimators and crowd-sensed data fusion," *Physical Communication*, vol. 66, p. 102475, 2024.

[9] A. Teimourian *et al.*, "Airfoil aerodynamic performance prediction using machine learning and surrogate modeling," *Heliyon*, vol. 10, no. 8, 2024.

[10] H. Chen *et al.*, "Multiple aerodynamic coefficient prediction of airfoils using a convolutional neural network," *Symmetry*, vol. 12, no. 4, p. 544, 2020, doi: 10.3390/sym12040544.

[11] B.-W. Zan *et al.*, "High-dimensional aerodynamic data modeling using a machine learning method based on a convolutional neural network," *Advanced Aerodynamics*, vol. 4, no. 1, p. 39, 2022, doi: 10.1186/s42774-022-00128-8.

[12] S. Yetkin, S. Abuhanieh, and S. Yigit, "Investigation on the abilities of different artificial intelligence methods to predict the aerodynamic coefficients," *Expert Systems with Applications*, vol. 237, p. 121324, 2024.

[13] E. Andrés-Pérez and C. Paulete-Periáñez, "On the application of surrogate regression models for aerodynamic coefficient prediction," *Complex & Intelligent Systems*, vol. 7, no. 4, pp. 1991–2021, 2021, doi: 10.1007/s40747-021-00307-y.

[14] G. Vijayakumar *et al.*, "Enhancement of unsteady and 3D aerodynamics models using machine learning," in *Journal of Physics: Conference Series*, vol. 1452, no. 1, IOP Publishing, 2020, doi: 10.1088/1742-6596/1452/1/012065.

[15] S. Ahmed *et al.*, "Aerodynamic analyses of airfoils using machine learning as an alternative to RANS simulation," *Applied Sciences*, vol. 12, no. 10, p. 5194, 2022, doi: 10.3390/app12105194.

[16] M. S. Alkhasawneh, "Hybrid cascade forward neural network with Elman neural network for disease prediction," *Arabian Journal for Science and Engineering*, vol. 44, no. 11, pp. 9209–9220, 2019, doi: 10.1007/s13369-019-03829-3.

[17] F. Bonnet, J. A. Mazari, P. Cinnella, and P. Gallinari, "AIRFRANS: High fidelity computational fluid dynamics dataset for approximating Reynolds-Averaged Navier–Stokes solutions," in *Advances in Neural Information Processing Systems*, 2022.

[18] A. Teimourian *et al.*, Airfoil aerodynamic performance prediction using machine learning and surrogate modeling, *Heliyon*, vol. 10, no. 8, Apr. 2024, doi: 10.1016/j.heliyon.2024.e29377.

[19] J. Liu *et al.*, AFBench: A large-scale benchmark for airfoil design, Oct. 2024, [Online]. Available: http://arxiv.org/abs/2406.18846.

[20] J. Zhao, L. Zeng, A. Lin, and X. Shao, Deep learning prediction method for aerodynamic forces on morphing aircraft considering physical monotonicity, *Adv. Aerodyn.*, vol. 7, no. 1, Dec. 2025, doi: 10.1186/s42774-024-00197-x.

[21] Y. Wang, P. Zhang, Y. Liu, J. Zhao, J. Lin, and Y. Chen, Aerodynamic coefficients prediction via cross-attention fusion and physical-informed training, 2025. [Online]. Available: https://www.aaai.org.

[22] K. Agarwal, V. Vijaykrishnan, D. Mohanty, and M. Murugaiah, A comprehensive dataset of the aerodynamic and geometric coefficients of airfoils in the public domain, *Data (Basel)*, vol. 9, no. 5, May 2024, doi: 10.3390/data9050064.

[23] K. Shaukat, S. Luo, V. Varadharajan, I. A. Hameed, and M. Xu, A survey on machine learning techniques for cyber security in the last decade, *IEEE Access*, vol. 8, 2020, doi: 10.1109/ACCESS.2020.3041951.

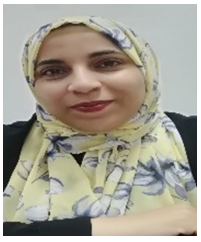
[24] W. S. Noble, What is a support vector machine?, *Nat. Biotechnol.*, vol. 24, no. 12, pp. 1565–1567, 2006, doi: 10.1038/nbt1206-1565.

[25] T. Chen and C. Guestrin, XGBoost: A scalable tree boosting system, in *Proc. ACM SIGKDD Int. Conf. Knowl. Discov. Data Min.*, 2016, doi: 10.1145/2939672.2939785.

[26] P. Devan and N. Khare, An efficient XGBoost–DNN-based classification model for network intrusion detection system, *Neural Comput.*

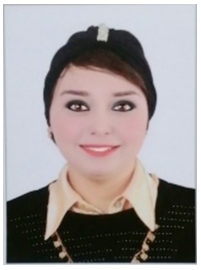
Appl., vol. 32, no. 16, pp. 12417–12431, 2020, doi: 10.1007/s00521-020-04708-x.

[27] S. S. Dhaliwal, A. Al Nahid, and R. Abbas, Effective intrusion detection system using XGBoost, *Inf.*, vol. 9, no. 7, p. 149, 2018, doi: 10.3390/info9070149.



Dr. Noha MM. Abdelnapi is an Assistant Professor at the Computer Science Department, Faculty of Computers and Information, Suez University. She received her MSc in Computer Science in 2016, focusing on Cloud Computing Security, and earned her PhD in 2019 on Applying IoT in Healthcare. With over 10 years of teaching experience at various universities, she has published research papers and book chapters in reputable international journals and conferences. Dr. Noha has also served as the Quality Assurance Unit Manager at the faculty since 2021. Her research interests include High Performance Computing, Cloud Computing, Artificial Intelligence (AI), Internet of Things (IoT), Data Analysis, and Cybersecurity.

Dr. Ahmed Gamal Abdellatif Ibrahim is a dedicated lecturer in the Department of Communications and Electronics Engineering at the Air Defense College in Alexandria, Egypt. He was born in 1985 in Sharkia Governorate, Egypt. In 2007, Dr Gamal graduated with a B.Sc. degree with honours in Electronics and Electrical Engineering from the Air Defense College at Alexandria University. He continued his academic journey and obtained a master's degree in Electronics and Electrical Engineering from Alexandria University in 2017. Dr. Gamal has also demonstrated his commitment to expanding his knowledge and gaining international experience. In 2019, he became a PhD student visitor at the prestigious Research Center of Geomatics (CIRGEO) at the University of Padua in Italy. This valuable experience allowed him to broaden his horizons and enrich his research pursuits. Dr. Gamal successfully completed his Ph.D. degree in Electronics and Communications Engineering in 2022. His research interests reflect his diverse background and multidisciplinary approach, including navigation, indoor positioning, tracking, filtering, information security, and image processing. Finally, He was a reviewer for several journals and major conferences.



Dr. Rania Ahmed is an Assistant Professor at the Faculty of Computers and Artificial Intelligence, Modern University for Technology and Information. She earned her PhD in Computer Science, specializing in Artificial Intelligence, from Menoufia University. She holds both BSc and MSc degrees in Computer Science from Helwan University. Since 2008, she has been affiliated with the Modern University. Her research has been published in international journals and conferences, including venues such as Springer, IEEE Access, and ScienceDirect (Elsevier).

Dr. Mohamed A. Aziz holds a Ph.D. in Aerospace Engineering from Cairo University, Egypt (2014). He served as an Assistant Professor at the Institute of Aviation Engineering and Technology from 2015 to 2018. Currently, he is an Associate Professor in the Department of Mechanical Engineering, College of Engineering, Suez University. Dr. Aziz

is also a member of the Egyptian Engineers Syndicate.

A. Abd Elhamid is currently an Assistant Lecturer in the Computer Science Department at the Faculty of Computers and Information, Suez University, Suez, Egypt. He previously held the position of Demonstrator and then Assistant Lecturer in the Engineering Sciences Department at the Faculty of Petroleum and Mining Engineering, Suez University. He holds a PhD degree in Computer Science from Port Said University. His research interests focus on Machine Learning, Image Processing, Cloud Computing, and Neural Networks. His diverse academic background reflects a strong interest in interdisciplinary research bridging engineering and computer science, supporting his academic and educational activities.

Mostafa Mohamed Ahmed received his B.Sc. degree (with honours) in Communication and Electrical engineering and the M.Sc. degree in Electrical engineering from Alexandria University, Alexandria, Egypt, in 2005 and 2013, respectively, and the Ph.D. degree in electrical, computer, and geomatics engineering from the University of Calgary, Calgary, AB, Canada, in 2017. He is currently a doctor at the Air Defense College, Alexandria, Egypt and a research assistant in the Mobile Multi-Sensor Systems (MMSS) Research Group, Department of Geomatics, University of Calgary. He has experience in Unmanned Aerial Vehicles (UAVs) navigation systems, His research interests include Global Navigation Satellite System (GNSS) denied environment navigation systems for UAVs, low-cost multi-sensors fusion, machine learning, adaptive digital signal processing, radar signal processing, advanced estimation techniques, computer vision, and image processing.

Syed T. Shah is a highly accomplished academic and research professional with a background in Electrical and Electronic Engineering. He received his Master's and Ph.D. degrees from the Department of Electrical and Electronic Engineering at Sungkyunkwan University in Suwon, South Korea, in 2015 and 2018, respectively. Currently, he is working as a Postdoctoral Fellow at the University of Glasgow in the UK and also serving as an Associate Professor with the Department of Electrical Engineering, BUITEMS, Pakistan. His research interests include 5G and beyond networks, Open RAN, AI-enabled wireless networks, RF energy harvesting, and Intelligent reflecting surfaces. Dr Shah is also an Editor of the Electronics Journal and a reviewer of several IEEE Transactions, Letters, and Magazines.

Mahmoud A. Shawky was born in 1990 in Saudi Arabia. He received his B.Sc. degree in Electronics and Electrical Engineering in 2012 from Air Defence College, Alexandria University, M.Sc. (Eng.) degree in Authentication Mechanisms in Computer Network Protocols from Alexandria University, Alexandria, Egypt. He received his PhD degree from the James Watt School of Engineering, University of Glasgow, UK. His research interests are in the areas of cryptography and number theory, digital signatures, authentication in wireless communications and cyber security.