



Cite this: *Phys. Chem. Chem. Phys.*,
2026, **28**, 5138

Uncertainty-aware machine learning-based prediction of plasma parameters in a microwave atmospheric pressure plasma jet

Suryasunil Rath,^a Priyabrata Das,^b Pulak Mohan Pandey^b and Satyananda Kar^{ib*}

Microwave atmospheric pressure plasma jets (MW-APPJs) exhibit significant potential for diverse applications, *i.e.*, hydrogen production, CO₂ dissociation, water treatment, material processing and waste treatment due to their stable operation at atmospheric pressure and generation of highly tunable reactive species. For effective utilization of MW-APPJs, a detailed understanding of the operational conditions that influence plasma parameters is essential. The present work proposes an uncertainty-aware, multi-output, interpretable supervised machine learning (ML) framework to predict eight plasma parameters, *viz.* electron excitation temperature (T_{exc}), electron number density (n_e), four reactive species (OH, N₂, H_α and O), gas temperature (T_g), and plume length. A dataset comprising 441 experimental runs was generated by varying the input powers (700–1000 W), sliding short positions (0.95–1.05λ_g/2) and argon flow rate (5–15 lpm). Six regression models, namely *k*-nearest neighbours (KNN), extra trees (ET), random forest (RF), artificial neural networks (ANN), gradient boosting (GB), and extreme gradient boosting (XGB), were optimized using Bayesian hyperparameter tuning and evaluated using both accuracy and reliability metrics. While XGB achieved competitive pointwise accuracy, the optimized GB model emerged as the most balanced performer when predictive accuracy, calibration behaviour, and uncertainty reliability were jointly considered. On a held-out test set, the GB model achieved mean absolute percentage errors below 3% and R^2 values exceeding 0.97 across all plasma parameters. Bootstrap-based uncertainty quantification demonstrated near-nominal 90% prediction interval coverage with comparatively narrow uncertainty bounds, and calibration analysis confirmed statistically consistent uncertainty estimates. Experimental validation using 30 independent plasma operating conditions, separated into interpolated and extrapolated regimes, further confirmed robust generalization, with increased epistemic uncertainty appropriately accompanying extrapolative predictions. SHapley Additive exPlanations (SHAP) based interpretability analysis identified microwave power as the dominant controlling feature for most plasma parameters, while gas flow rate governed the intensity of OH emission. Overall, this uncertainty-aware ML framework provides a reliable foundation for data-driven plasma diagnostics and future optimization of MW-APPJ-based processes.

Received 12th November 2025,
Accepted 27th January 2026

DOI: 10.1039/d5cp04364f

rsc.li/pccp

1. Introduction

Plasma, often referred to as the fourth state of matter, is an ionized gas consisting of electrons, ions, neutral particles, and excited species. This state arises when external energy input exceeds the ionization potential of a neutral gas, which leads to the ionization of atoms or molecules. Unlike solids, liquids, or gases, plasma exhibits collective behaviour due to the presence of charged particles that respond to electromagnetic fields. These unique properties, including high electrical conductivity,

tunable reactivity, and strong coupling to electromagnetic fields, render plasmas indispensable in fields ranging from applications across diverse scientific and industrial domains. Atmospheric pressure plasma jets (APPJs) provide significant advantages, such as the elimination of vacuum systems, thereby reducing operational expenses and providing better control over plasma-chemical reactions.^{1–4} Furthermore, their open-air configuration permits precise spatial and temporal control over reactive species generation, enabling tailored plasma-chemical processes for applications in hydrogen production,^{5,6} CO₂ dissociation,⁷ surface engineering,⁸ food processing,⁹ biomedical treatments,^{10,11} and environmental remediation.^{12,13} Microwave atmospheric pressure plasma jets (MW-APPJs) represent a distinct subclass of APPJ systems, characterized by their utilization of high-frequency (commonly 915 MHz and 2.45 GHz) EM

^a Department of Energy Science and Engineering, Indian Institute of Technology Delhi, Hauz Khas, New Delhi 110016, India. E-mail: satyananda@dese.iitd.ac.in

^b Department of Mechanical Engineering, Indian Institute of Technology Delhi, Hauz Khas, New Delhi 110016, India

waves for plasma generation in both continuous and pulsed modes.^{14,15} A key feature of MW-APPJs is the excitation of surface waves at the plasma dielectric interface, which significantly enhances discharge stability and efficient power coupling.^{16–21} Various microwave coupling configurations have been developed for plasma generation, with waveguide-based geometries being particularly effective due to their high power (kW) handling capabilities with lower losses. These systems offer significant advantages over electrode-based designs, including the elimination of electrode erosion and plasma contamination through electrodeless operation. Among various waveguide configurations, the surfaguide geometry is preferable due to its tapered applicator section, which facilitates localized electric field intensification, thereby promoting efficient plasma generation. This combination of features enables robust plasma operation across wide parameter ranges.^{22–24}

The plasma parameters, *viz.*, electron excitation temperature (T_{exc}), electron number density (n_e), emission intensities of reactive species, gas temperature (T_g), and plume length, are crucial for understanding the discharge stability, optimizing the performance for applications, chemical reactions, and the spatial propagation dynamics of the plasma jet.^{4,25} The variation of input parameters significantly influences the output plasma characteristics. The power applied to generate plasma is usually in the form of an electric field or electromagnetic radiation, such as microwaves. The energy transferred to the gas excites the seed electrons, causing them to accelerate and collide with neutral particles, which results in ionization. When microwave power is increased at fixed gas flow rates and sliding short positions (for reducing reflected power), the enhanced energy input per neutral particle leads to greater electron impact ionization and excitation rates. As a result, an increment in the key plasma parameters, including the T_{exc} , n_e , emission intensities of reactive species, T_g , and plume length, is observed. In contrast, gas flow rate variation at fixed power and sliding short position exhibits a non-monotonic effect. Since gas flow determines the neutral particle density available for the generation of plasma, resulting in higher plasma parameters until an optimum flow rate is reached. Beyond this point, the gas flow increases, diluting the energy per particle and reducing ionisation due to insufficient power density. In both cases, lower and higher flow rates signify lower power coupling efficiency. For that, optimal operations are needed for specific operations. Meanwhile, the sliding short position critically governs microwave power coupling efficiency, controlling the reflection coefficient.

Recent studies have demonstrated the efficacy of machine learning (ML) algorithms in predicting the plasma parameters and optimizing plasma-based processes.^{26–29} The investigation conducted by Bakhtiyari *et al.*³⁰ demonstrated the efficacy of machine learning approaches, specifically artificial neural networks (ANN) and adaptive neuro-fuzzy inference systems (ANFIS), in predicting spatiotemporal variations of plasma temperature and electron density in cadmium laser-induced plasmas (LIP). This prediction highlights the capability of data-driven models to handle the temporal and transient behaviour

characteristic of laser-produced plasmas. Furthermore, Suresh *et al.*³¹ integrated random forest regression (RFR) and gradient boosting regression (GBR) algorithms with a collisional radiative (CR) model to estimate electron temperature and density in gallium laser-produced plasmas. Their comprehensive validation demonstrated remarkable consistency of electron temperature with three diagnostic methods: the OES-CR model (0.67 eV), ML (RFR-0.66 eV, GBR-0.67 eV), and line ratio methods (0.67 eV). The use of ML in the domain of plasma catalysis was reported by Li *et al.*,³² who have used ANN and reinforcement learning (RL) algorithms for both prediction and optimization of plasma-based conversion of CO_2 and CH_4 , particularly aiming to increase the methanol production. The same research group later applied supervised learning and RL models to optimize plasma-catalytic dry reforming of methane (DRM) over $\text{Ni}/\text{Al}_2\text{O}_3$ catalysts in a dielectric barrier discharge (DBD) reactor.³³ They identified the optimal conditions (60 W power, 9.5 wt% $\text{Ni}/\text{Al}_2\text{O}_3$, 74 mL min^{-1} flow rate) that achieved 36% conversion at 34 eV per molecule energy cost. Bong *et al.*,³⁴ developed a hybrid deep learning framework combining convolutional autoencoder (CAE) feature extraction with dense neural network (DNN) regression to quantitatively analyze plasma images. Their optimized model achieved 96% prediction accuracy for both gas flow rate (± 3.08 slpm) and process gas concentration (± 300 ppm), demonstrating the viability of computer vision techniques for non-invasive plasma process monitoring. The study conducted by Witman *et al.*³⁵ implemented RL with sim-to-real transfer to control substrate temperature during APPJ treatment. The RL agent, which was trained on simulated plasma-substrate thermal dynamics, exhibited robust temperature regulation across substrates with varying thermal and electrical properties. Lin *et al.*³⁶ applied physics-informed neural networks (PINN) to predict species emission intensities of reactive species, gas temperature, and electron temperature in cold atmospheric plasma (CAP) jets and validated with the literature.

Although recent studies have demonstrated the potential of ML for plasma diagnostics, most existing approaches are limited to a limited set of algorithms and a small number of predicted parameters, often neglecting uncertainty and generalization behaviour. In atmospheric plasma systems, output parameters such as T_{exc} , n_e , reactive species, T_g , and plume length exhibit distinct and often nonlinear dependencies on input conditions (power, gas flow rates, and sliding short position). To address these limitations and enable a comprehensive understanding of plasma behaviour, this study introduces an uncertainty-aware, multi-output machine learning framework capable of capturing complex interdependencies across a broad input–output space. The performance of six distinct ML algorithms *i.e.*, k nearest neighbour (KNN), extra tree (ET), random forest (RF), artificial neural network (ANN), gradient boosting (GB) and extreme gradient boosting (XGB) is evaluated for predicting eight key plasma parameters *viz.* T_{exc} , n_e , four reactive species, T_g , and plume length from three different input features (power, gas flow rates, position of sliding short) before and after hyperparameter tuning. Beyond pointwise accuracy, model reliability is assessed through bootstrap-based uncertainty quantification,

calibration analysis, and experimental validation under both interpolative and extrapolative operating regimes. Furthermore, the interpretability of the best performing model is assessed using SHapley additive exPlanations (SHAP) analysis to elucidate the relative influence of input parameters on individual plasma characteristics.

2. Methodology for predictive modelling

2.1 Feature space design and data collection

The dataset used in this study was prepared by varying three features, namely power, Ar gas flow rates, and position of the sliding short within the ranges of 700–1000 W, 5–15 lpm, and $0.95\lambda_g/2$ – $1.05\lambda_g/2$, respectively. The λ_g is the guided wavelength of rectangular waveguides.³⁷ The lower power limit of 700 W was selected because plasma does not emerge below 500 W due to shielding by the Faraday cage, which protects microwave radiation. Although the power supply capacity extends to 3000 W, the upper limit of the 1000 W range was chosen as it represents the optimal balance between energy efficiency and practical application needs. This power window achieves T_g within 600–1000 K, which is sufficient for various applications, including hydrogen production, CO₂ dissociation, water treatment, material processing and waste treatment while avoiding the excessive energy consumption and potential thermal equilibrium associated with higher power operation. The above power range, greater than 1000 W, is commonly used for increasing the production rate, a higher rate of molecular dissociation, and a higher volume for waste treatment

purposes. The gas flows below 5 lpm reduce electron-neutral collisions, which limits the ionization efficiency, while flows above 15 lpm cause axial and radial contraction, which causes dilution of reactive species and plasma parameters. The range for sliding short positions was selected through impedance matching optimization studies. These positions maintain minimum reflected power and operating outside this range (particularly beyond $(0.95-1.05)\lambda_g/2$) causes significant microwave reflection, drastically reducing power coupling efficiency. In our experimental setup, the tunable sliding short was adjusted to a value in millimeter scale, which indicates that the point close to 87 mm is actually the $\lambda_g/2$ position, which is the value of the guided wavelength for the WR340 rectangular waveguide. This parameter space provides a foundational framework that is both scientifically relevant for fundamental plasma studies and ensures optimal plasma parameters, and chemical output for targeted applications. The feature space is composed of 441 instances by varying the power at an interval of 50 W, Ar gas flow rate at an interval of 0.5 lpm and considering three different values of sliding short position ($0.95\lambda_g/2$, $\lambda_g/2$, $1.05\lambda_g/2$) as depicted in Fig. 1(a).

The plasma was generated in a MW-APPJ whose schematic is shown in Fig. 1(b). The setup consists of a power supply, magnetron, isolator, triple stub tuner, surfaguide launcher, and sliding short. The magnetron generates a 2.45 GHz frequency, taking power from the power supply and passing through a rectangular waveguide (WR340). By preventing microwave backflow from the applicator, the isolator safeguards the microwave source, while the triple stub tuner provides variable capacitive and inductive reactance for optimal impedance matching. Next to the triple stub tuner, a microwave applicator is connected, which is known as a

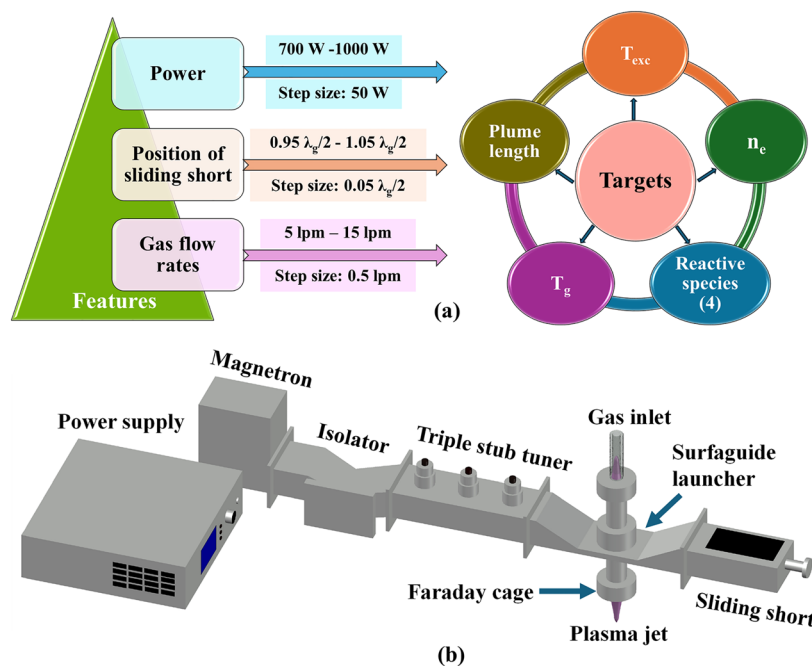


Fig. 1 (a) Dataset detail and (b) schematics of the MW-APPJ experimental set-up.

surfaguide launcher. The surfaguide launcher is a tapered rectangular (WR340) waveguide section designed to concentrate power in the applicator region, achieving significantly higher electric field intensities compared to conventional waveguide configurations. The launcher incorporates a centrally positioned, precision-aligned quartz tube (OD: 10 mm, ID: 6 mm, length: 400 mm) oriented for vertical gas flow (top to bottom configuration). High-purity argon (99.9999%) serves as the working gas for plasma generation. To ensure operational safety and prevent microwave leakage, the assembly features bilateral metallic extensions forming a Faraday cage enclosure. Adjacent to the surfaguide launcher, the adjustable sliding short circuit facilitates secondary impedance matching by reflecting unabsorbed microwave power. This enables the formation of standing waves and enhances the power coupling efficiencies for plasma initiation and maintenance.

The microwave power supply has a regulated 3 kW (maximum) power with 10 W resolution. The cooling system employs a dual-mode configuration: (1) a water-cooling circuit delivering flow rates of 6 lpm to the magnetron and power supply components, and (2) a secondary 2 lpm water cooling loop for the stub tuners and applicator, both maintained at 25 ± 2 °C. The applicator features additional thermal cooling through a 45 lpm compressed air swirl flow surrounding the quartz tube. Initially, plasma was generated at a very low power, *i.e.*, 10 W only. Microwave leakage is quantified using a standard calibrated radiation meter (confirmed < 5 mW cm⁻², at 5 cm from the leakage). Impedance matching is optimized *via* a triple stub tuner and sliding short adjustment *via* minimum reflected power. Power is gradually increased to the operational threshold (> 500 W) for measurement outside the applicator. The characterization is performed within the 700–1000 W range under stabilized conditions. The plasma characterization was performed using standard diagnostic techniques to quantify fundamental parameters. Optical emission spectroscopy (OES) served as the diagnostic tool, enabling non-invasive determination of the T_{exc} through Boltzmann plot analysis of atomic emission lines of Ar, n_e *via* the line intensity ratio methods,³⁸ and relative emission intensities of reactive species through calibrated emission line intensities. The T_g was acquired using a calibrated K-type (200–1200 °C) thermocouple. Plasma plume length was quantitatively assessed through direct imaging with a precision measuring scale, establishing the visible discharge boundary along the axial direction.

2.2 Machine learning models and hyperparameter tuning

The “standard scaler” module of scikit-learn was used to normalize the input dataset prior to ML modelling. This step is important for the diverse range and scale of input variables in the dataset. Such scaling ensures that each feature contributes equally during model training and prevents features with larger magnitudes from dominating the learning process. We have employed six different vanilla regression ML models, *viz.*, k nearest neighbour (KNN), extra tree (ET), random forest (RF), artificial neural network (ANN), gradient boosting (GB) and extreme gradient boosting (XGB), for the multi-target prediction task. The rationale behind using six different models is associated with the “no free lunch theorem”³⁹

which underscores the challenge of constructing a single generalized ML model for different domain prediction tasks. The selection of models encompasses a diverse set of learning algorithms such as instance-based (KNN), neural network-based (ANN), bagging-based ensembles (RF, ET) and boosting-based ensembles (GB, XGB). This step aids in assessing the suitability of each model for the proposed prediction task, *i.e.* plasma characteristic targets (output). Furthermore, to improve the generalization capability of the models and to avoid the models from memorizing instead of learning, a five-fold cross-validation strategy was implemented. The predictive performance and generalization capability of ML models are strongly influenced by the choice of hyperparameters. Therefore, to ensure a fair and unbiased comparison among different learning algorithms, systematic hyperparameter optimization was carried out for all six regression models considered in this study, namely KNN, ET, RF, ANN, GB, and XGB. This step was essential to avoid performance bias arising from the use of default (“vanilla”) model settings, particularly given the known sensitivity of certain models (*e.g.*, neural networks and boosting algorithms) to hyperparameter selection.

Hyperparameter tuning was performed using Bayesian optimization,⁴⁰ implemented *via* the BayesSearchCV framework. Compared to exhaustive grid search or random search strategies, Bayesian optimization offers improved sample efficiency by sequentially exploring the hyperparameter space based on a probabilistic surrogate model. This approach is especially well-suited for problems with limited datasets and moderately high-dimensional search spaces, as is the case in the present study. For each regression model, a compact yet physically and statistically meaningful hyperparameter search space was defined based on prior knowledge of model behavior and best practices reported in the literature. The hyperparameter space corresponding to each model can be seen from Table 1. All regression models were implemented in a unified multi-output learning framework, wherein a single model was trained to simultaneously predict all eight plasma characteristics. This strategy avoids training separate models for each target variable, thereby preventing an unnecessary increase in model complexity and hyperparameter dimensionality. By optimizing a single shared set of hyperparameters per model architecture, the proposed approach ensures consistency, reduces the risk of overfitting, and enables a fair comparison across different learning algorithms. Moreover, the multi-output formulation preserves interdependencies among plasma parameters and provides a computationally efficient alternative to independent single-target modeling.

KNN is a non-parametric, instance-based algorithm that generates predictions by averaging the target values of the “ k ” most similar training instances. A Minkowski distance metric with the k value of five was used for the model training. The locality-based decision-making approach offered by KNN is effective for modelling nonlinear relationships between features and targets. ANN is a universal function approximator composed of layers (input, hidden and output) of interconnected nodes (neurons) with non-linear activation functions. The ANN model learns by minimizing the loss function, *i.e.*

Table 1 Hyperparameter search space for different ML regression models

ML model	Hyperparameter space
ET	n_estimators: [100, 1000], max_depth: [3, 20], min_sample_split: [3, 20], min_sample_leaf: [3, 20], bootstrap: ['True', 'False'], criterion: ['squared_error', 'absolute_error']
GB	n_estimators: [50, 500], learning_rate: [0.01, 0.3], max_depth: [2, 10], min_sample_split: [2, 20], min_sample_leaf: [1, 10], subsample: [0.6, 1], loss: ['squared_error', 'absolute_error']
KNN	n_neighbors: [4, 20], weights: ['uniform', 'distance'], distance_function: ['euclidean', 'manhattan', 'minkowski'], p: [3, 15]
MLP	hidden_layer_sizes: [(20,), (30,), (40,), (50,), (20,10), (30,15), (40,20), (50,25)], activation: ['relu', 'tanh'], solver: ['adam', 'lbfgs'], alpha: [1×10^{-6} , 1×10^{-2}], learning_rate_init: [1×10^{-4} , 1×10^{-1}]
RF	n_estimators: [50, 500], max_depth: [2, 20], min_sample_split: [2, 20], min_sample_leaf: [2, 20], criterion: ['squared_error', 'absolute_error']
XGB	n_estimators: [50, 500], learning_rate: [0.02, 0.3], max_depth: [3, 10], min_child_weight: [1, 10], subsample: [0.6, 1], colsample_bytree: [0.6, 1]

mean squared error (MSE), *via* backpropagation and stochastic gradient descent. The MSE loss (L_{MSE}) for any kind of regression task is given in eqn (1).

$$L_{\text{MSE}} = \frac{1}{n} \sum_{i=1}^n [Y_i - F(X_i, \theta)]^2 \quad (1)$$

where n is the number of datapoints, Y_i is the actual value of the target and $F(X_i, \theta)$ is the predicted value of the target for the input feature X_i and weight θ . Both RF and ET are ensemble learning methods, which build multiple decision trees and aggregate their predictions to make a single prediction. However, both models differ in terms of the split criterion. Random forest constructs trees using bootstrapped samples of the data and selects optimal feature splits based on impurity reduction, thereby reducing variance. On the other hand, ET selects split thresholds completely at random without performing optimization of the loss function. RF ensures higher predictive accuracy by optimizing the splits to minimize MSE while ET sacrifices marginal precision for computational efficiency. GB and XGB are advanced ensemble learning techniques, which iteratively combine the weak learners to correct residual errors. GB operates by sequentially fitting decision trees to the negative gradients (pseudo-residuals) of MSE for regression.⁴¹ The pseudo residual for an instance i at the m th iteration ($r_{i,m}$) is given in eqn (2), which is the difference between the actual value and predicted value of instance i at the $(m - 1)$ th iteration. In the subsequent step, a new tree, *i.e.* $h_m(X)$ is trained on residuals, and the model prediction has been updated by using the following eqn (3).

$$r_{i,m} = Y_i - F_{m-1}(X_i) \quad (2)$$

$$F_m(X) = F_{m-1}(X) + \alpha \cdot h_m(X) \quad (3)$$

where $F_m(X)$ and $F_{m-1}(X)$ are the prediction of the m th iteration and $(m - 1)$ th iteration respectively, and α is the learning rate used to prevent overfitting. This gradient descent approach minimizes the loss function but lacks explicit regularization, relying instead on shrinkage and subsampling for robustness. XGB enhances the performance of GB by augmenting regularization and second order optimization. The equation of the regularized objective function (loss function) with L_1/L_2 penalties used in XGB is given by eqn (4).⁴²

$$\mathcal{L} = \sum_{i=1}^n L(Y_i, F(X_i)) + \sum_{k=1}^K (\beta T + 0.5\gamma \|w\|^2) \quad (4)$$

where β and γ are hyperparameters to control T (leaves) and w (leaf weight), respectively, and \mathcal{L} is the loss function. XGB incorporates a second-order Taylor expansion, incorporating both gradients and Hessians for precise updates. Additionally, XGB employs efficient split-finding algorithms using approximate quantile sketches and handles missing values natively through sparsity-aware partitioning.

2.3 Evaluation metrics

As our study focuses on predicting continuous target variables, *i.e.*, a regression task, multiple complementary evaluation metrics were employed to assess the predictive accuracy and generalization capability of the ML models. We have used mean squared error (MSE), mean absolute percentage error (MAPE) and coefficient of determination (R^2) to evaluate the performance of predictive ML models. The MSE metric quantifies the average squared deviation between predicted and actual values, thereby penalizing larger errors more strongly. The mathematical expression for MSE is given in eqn (5).

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2 \quad (5)$$

where n is the number of instances, Y_i is the actual value of target, \hat{Y}_i is the predicted value of target, and $(Y_i - \hat{Y}_i)$ is the residual corresponding to the instance, i .

MAPE was used to provide a scale-independent measure of prediction accuracy. MAPE expresses the prediction error as a percentage of the true value, enabling direct comparison of model performance across different plasma parameters with varying magnitudes and physical units. The expression for MAPE is given in eqn (6).

$$\text{MAPE (\%)} = \frac{1}{n} \sum_{i=1}^n \left| \frac{Y_i - \hat{Y}_i}{Y_i} \right| \times 100 \quad (6)$$

Furthermore, the R^2 was employed to evaluate the proportion of variance in the experimental data that is explained by the model. R^2 provides an intuitive measure of goodness-of-fit, with values closer to unity indicating stronger agreement

between predictions and observations. The mathematical formulation for R^2 is given in eqn (7).

$$R^2 = 1 - \frac{\sum_{i=1}^n (Y_i - \hat{Y}_i)^2}{\sum_{i=1}^n (Y_i - \bar{Y})^2} \quad (7)$$

Where \bar{Y} is the mean value of the observed target values.

2.4 Uncertainty quantification and model reliability assessment

While conventional regression metrics such as MSE, MAPE and R^2 provide a quantitative measure of predictive accuracy, they do not capture the reliability or confidence associated with individual predictions. For practical deployment of machine learning (ML) models in plasma diagnostics and process monitoring, it is essential to quantify the uncertainty associated with model predictions, particularly when predictions are used to guide experimental decisions or parameter selection. Therefore, in addition to point prediction accuracy, a systematic uncertainty quantification (UQ) and reliability assessment framework was implemented for all optimized regression models. In this study, epistemic uncertainty, arising from limited data availability and model uncertainty, was quantified using a bootstrap ensemble approach. Epistemic uncertainty is relevant for data-driven plasma modeling, where experimental datasets are often sparse and expensive to acquire. Bootstrap-based methods offer a practical, model-agnostic approach to approximating predictive uncertainty, eliminating the need for explicit probabilistic formulations of the learning algorithm.

2.4.1 Bootstrap-based epistemic uncertainty estimation. For each optimized ML model (KNN, ET, RF, ANN, GB, and XGB), uncertainty estimation was performed by constructing an ensemble of models through bootstrap resampling of the training dataset. Specifically, multiple bootstrap datasets ($B = 50$) were generated by sampling the original training set with replacement, and an identical model configuration (using the tuned hyperparameters) was trained on each resampled dataset. Each trained model then produced predictions on the fixed test set. The resulting ensemble of predictions for a given input instance forms an empirical predictive distribution. From this distribution, the mean prediction represents the final point estimate, while the standard deviation of predictions across the ensemble serves as a measure of epistemic uncertainty. This approach captures variability in predictions due to changes in the training data and provides a robust estimate of model confidence.

2.4.2 Prediction intervals and coverage analysis. To further assess the reliability of uncertainty estimates, prediction intervals (PIs) were constructed from the bootstrap prediction distributions. For a given confidence level α , the lower and upper bounds of the prediction interval were obtained from the corresponding percentiles of the bootstrap ensemble. In this work, a 90% prediction interval was primarily analyzed, defined by the 5th and 95th percentiles of the predictive distribution. The quality of the prediction intervals was evaluated using two complementary metrics, *i.e.* prediction interval coverage probability (PICP) and mean

prediction interval width (MPIW). The PICP is defined as the fraction of true test observations that fall within the predicted interval, whereas MPIW is defined as the average width of the prediction interval, reflecting the sharpness of the uncertainty estimate. An ideal uncertainty model simultaneously achieves a high PICP, close to the nominal confidence level, and a low MPIW, indicating informative yet reliable uncertainty bounds. These metrics were computed independently for each plasma parameter to account for differences in scale and variability among targets.

2.4.3 Calibration analysis and uncertainty-error consistency analysis. Beyond aggregate coverage metrics, the calibration behavior of uncertainty estimates was examined using calibration curves, which compare nominal coverage probabilities to empirically observed coverage. For multiple nominal confidence levels (*e.g.*, 50%, 70%, and 90%), empirical coverage was calculated as the fraction of test points enclosed by the corresponding prediction intervals, averaged across all target variables. A well-calibrated model is expected to produce a calibration curve that closely follows the diagonal (ideal calibration) line. Deviations from this line indicate systematic overconfidence (empirical coverage below nominal) or underconfidence (empirical coverage above nominal). Calibration analysis, therefore, provides insight into whether uncertainty estimates are statistically consistent and trustworthy. To assess whether the estimated uncertainty meaningfully reflects prediction reliability, an uncertainty-error relationship analysis was performed. For each test instance and target variable, the absolute prediction error $|Y - \hat{Y}|$ was plotted against the corresponding epistemic uncertainty (standard deviation of the bootstrap ensemble). A good model exhibits a positive correlation between uncertainty magnitude and prediction error, indicating that higher uncertainty is associated with less reliable predictions. Such behavior confirms that the uncertainty estimates are informative rather than arbitrary. This analysis was carried out for all models and targets, enabling qualitative comparison of uncertainty reliability across learning algorithms.

The combined evaluation of point prediction accuracy, uncertainty calibration, prediction interval quality, and uncertainty-error consistency enabled a holistic comparison of model performance. Rather than selecting the best-performing model solely based on error metrics, uncertainty-aware diagnostics were used to identify models that provide both accurate and reliable predictions. This integrated framework ensures that the selected ML model not only achieves high predictive accuracy but also delivers well-calibrated and interpretable uncertainty estimates, which are essential for robust plasma parameter prediction and subsequent experimental validation. The complete workflow for the plasma parameter prediction can be visualized in Fig. 2. The best-performing ML model was selected for further investigations to gain deeper insight into its predictive capability. Subsequently, the selected model was employed to predict plasma parameters for an independent experimental dataset comprising thirty points. Of these, fifteen data points lay within the training domain feature range of the training dataset (interpolated regime), while the remaining fifteen were located outside the training feature domain (extrapolated regime). The predictive performance of the model was

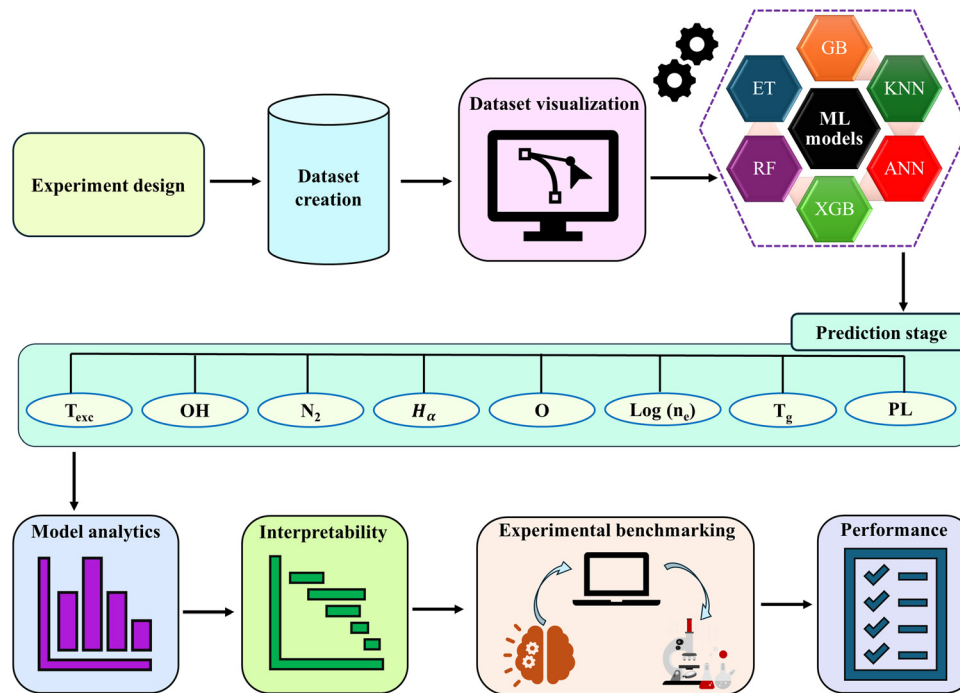


Fig. 2 Schematic of the machine learning framework used for the prediction of plasma parameters.

evaluated separately for the interpolated and extrapolated subsets to rigorously assess its stability, generalization capability, and robustness beyond the training domain.

3. Results and discussion

3.1 Data set visualization

The feature distribution in the used dataset can be visualised in Fig. 3. All the features exhibit a discrete type of distribution, which corresponds to the experimental design where the measurements were performed at fixed operating levels. The frequency along the y-axis in Fig. 3 denotes the number of data points associated with each set of input parameters. For example, there are 63 (3×21) data points available in Fig. 3(a) for three sliding short positions and 21 gas flow rates. This particular

representation of the process window can minimize the bias in the feature space, which is favourable for supervised ML models. The aim is to construct data-driven models that can accurately predict plasma parameters across and beyond the continuous input domain by capturing the governing trends embedded within discretely sampled experimental datasets.

Further Pearson pairwise correlation (PCC) analysis was performed for the target variables to understand the interdependencies among them. PCC analysis quantifies the strength and direction of linear relationships between variables. It is crucial for identifying parameter interdependencies, reducing feature redundancy, and improving the interpretability of experimental or data-driven models. Based on the correlation strength, the target pairs were categorised into highly correlated (Pearson's $r > 0.8$), moderately correlated ($0.5 \leq$ Pearson's $r \leq 0.8$) and weakly correlated (Pearson's $r < 0.5$). The



Fig. 3 Histograms showing the distribution of features: (a) power, (b) sliding short position, and (c) gas flow rates.

pairwise scatterplots with marginal histograms and kernel density estimation (KDE) contours for highly correlated, moderately correlated and weakly correlated features are illustrated in Fig. 4, 5 and 6, respectively. The KDE contours represent the estimated joint probability density between the pair of target variables. They provide a smooth estimation of concentrated data and reveal the data distribution across the feature space. Each successive contour encloses a higher concentration of datapoints, which allows the user to visualize the data cluster around a particular value combination. The presence of multiple nested contours reflects the gradual change in data density and highlights potential nonlinear or multimodal relationships

between target variables. Additionally, the distribution type of each target variable is given in Table 2. Considering the diverse distribution type of target variables, we have selected non-parametric and ensemble-based ML models such as RF, ET, GB and XGB, which are robust to distributional assumptions and capable of capturing complex non-linear relationships. Furthermore, KNN and ANN also do not rely on the data normality assumptions and are well-suited for capturing complex relationships. In contrast, linear models, which assume a Gaussian distribution, were not considered in this study owing to their limited capability in addressing the non-Gaussian nature observed in the dataset.⁴³

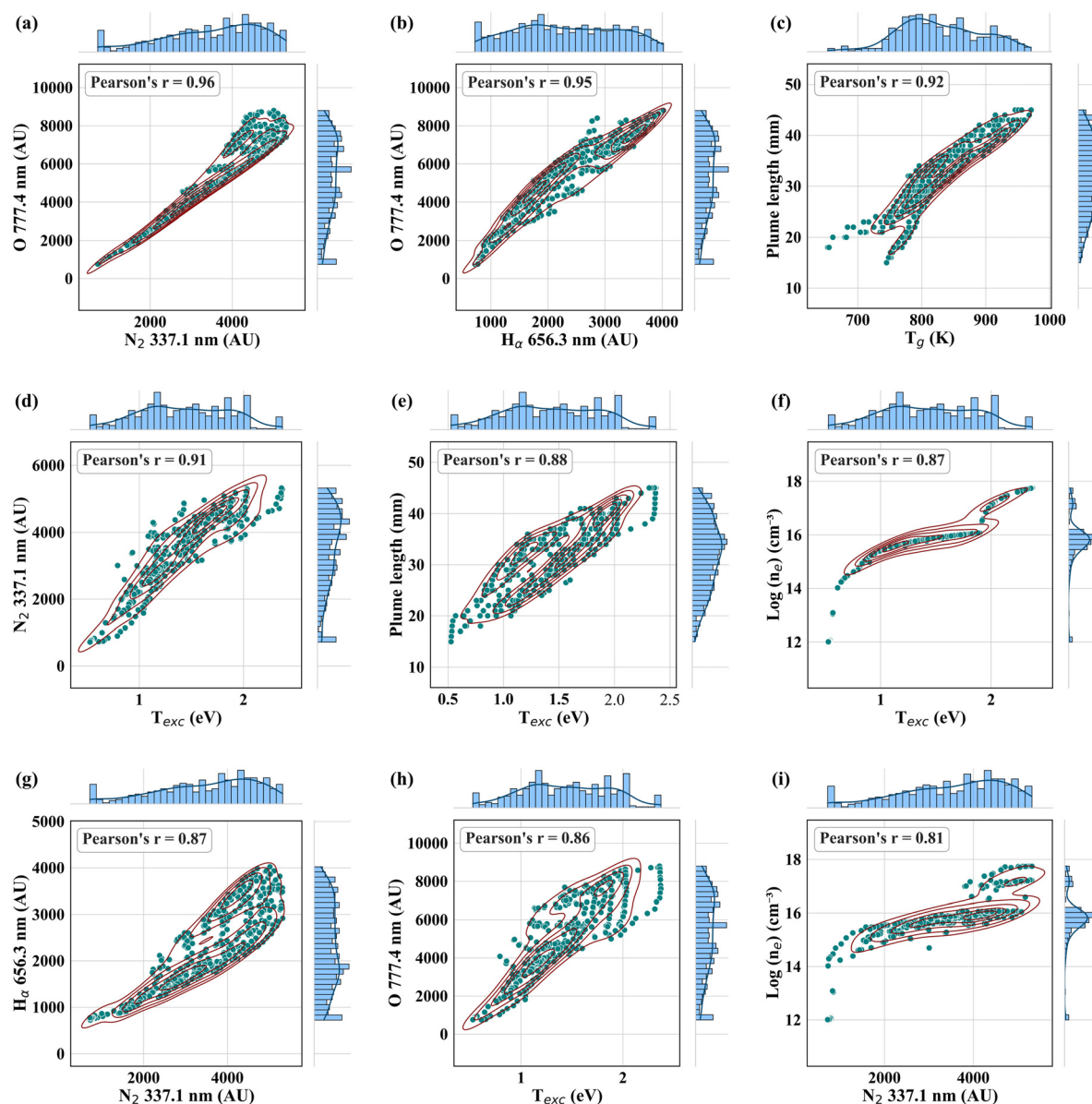


Fig. 4 Pairwise scatter plots with marginal histograms and kernel density estimation (KDE) contours are shown for strongly correlated output parameter pairs (Pearson's $r > 0.8$). Each subfigure displays the joint distribution along with the Pearson correlation coefficient. (a) O (777.4 nm) vs. N₂ (337.1 nm), (b) O (777.4 nm) vs. H_α (656.3 nm), (c) plume length vs. T_g, (d) N₂ (337.1 nm) vs. T_{exc}, (e) plume length vs. T_{exc}, (f) log(n_e) vs. T_{exc}, (g) H_α (656.3 nm) vs. N₂ (337.1 nm), (h) O (777.4 nm) vs. T_{exc}, (i) log(n_e) vs. N₂ (337.1 nm).

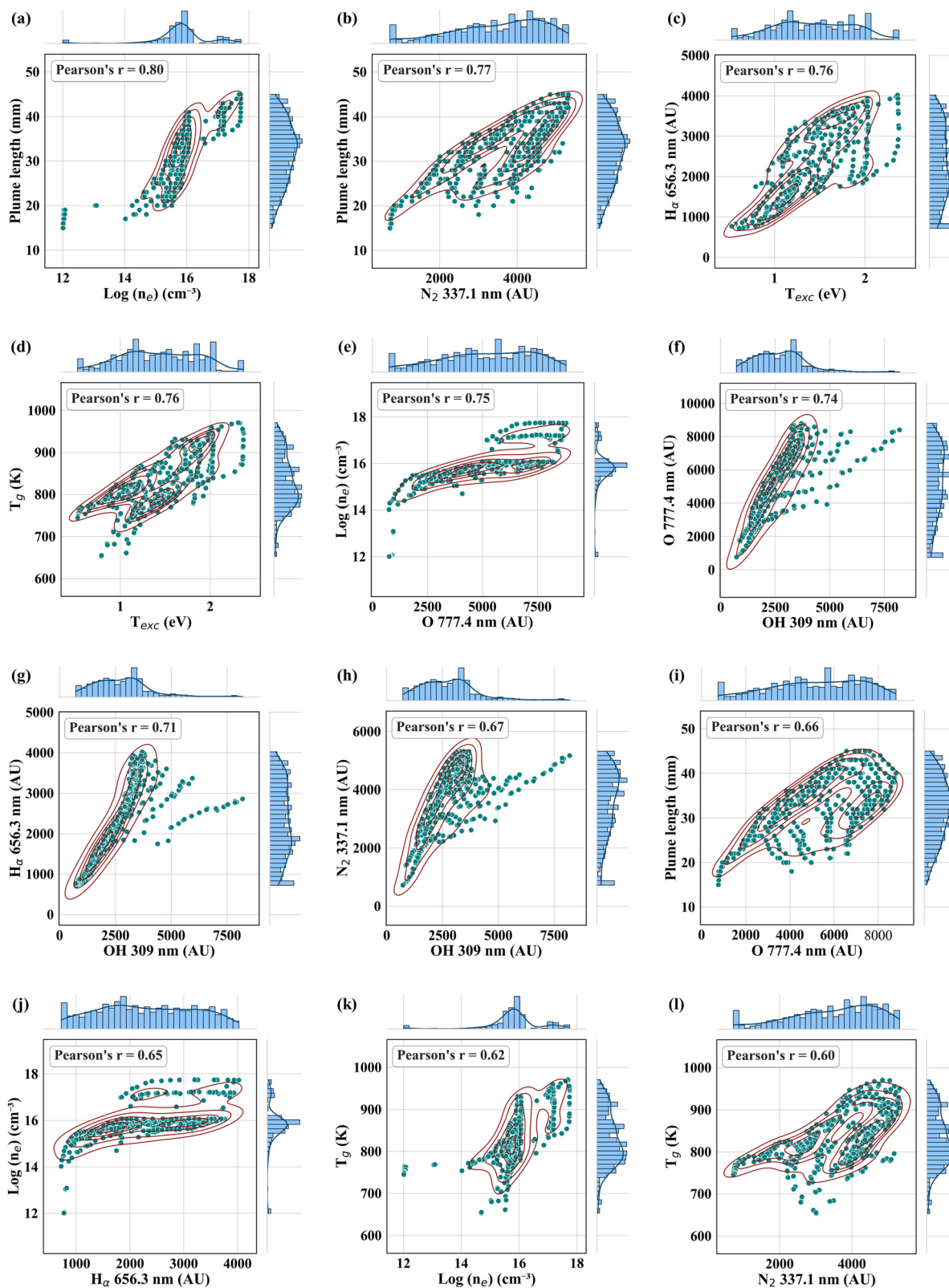


Fig. 5 Pairwise scatter plots with marginal histograms and kernel density estimation (KDE) contours are shown for moderately correlated output parameter pairs ($0.5 \leq \text{Pearson's } r \leq 0.8$). Each subfigure displays the joint distribution along with the Pearson correlation coefficient. (a) Plume length vs. $\log(n_e)$, (b) plume length vs. N_2 (337.1 nm), (c) H_α (656.3 nm) vs. T_{exc} , (d) T_g vs. T_{exc} , (e) $\log(n_e)$ vs. O (777.4 nm), (f) O (777.4 nm) vs. OH (309 nm), (g) H_α (656.3 nm) vs. O (777.4 nm), (h) N_2 (337.1 nm) vs. OH (309 nm), (i) plume length vs. O (777.4 nm), (j) $\log(n_e)$ vs. H_α (656.3 nm), (k) T_g vs. $\log(n_e)$, (l) T_g vs. N_2 (337.1 nm).

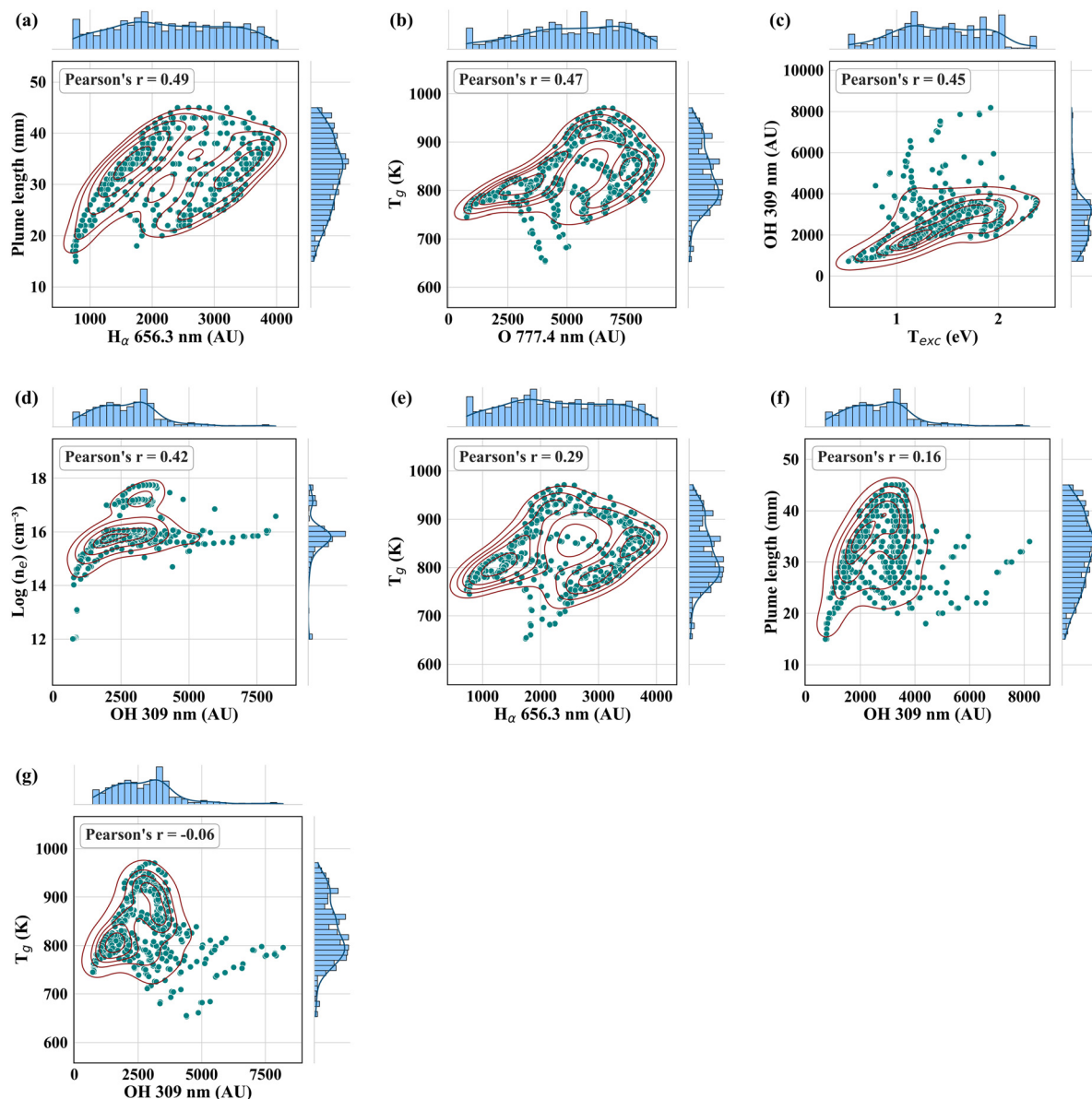


Fig. 6 Pairwise scatter plots with marginal histograms and kernel density estimation (KDE) contours are shown for weakly correlated output parameter pairs (Pearson's $r < 0.5$). Each subfigure displays the joint distribution along with the Pearson correlation coefficient. (a) Plume length vs. H_α (656.3 nm), (b) T_g vs. O (777.4 nm), (c) OH (309 nm) vs. T_{exc} , (d) $\log(n_e)$ vs. OH (309 nm), (e) T_g vs. H_α (656.3 nm), (f) plume length vs. OH (309 nm), (g) T_g vs. OH (309 nm).

Table 2 The distribution type of each target variable

Target variable	Distribution type
T_{exc}	Close to Gaussian
OH	Right skewed
N_2	Left skewed/multimodal
H_α	Uniform
O	Left skewed
$\log(n_e)$	Close to Gaussian
T_g	Left skewed
Plume length	Close to Gaussian

Some crucial targets are explained in Fig. 4–6. From Fig. 4, it can be inferred that the T_{exc} exhibited strong correlation with N_2 (Fig. 4(d)), plume length (Fig. 4(e)), $\log(n_e)$ (Fig. 4(f)) and O

(Fig. 4(h)). This can be fundamentally explained through kinetic and thermodynamic considerations. The rise in T_{exc} indicates higher energy input to plasma, which increases the collisional process, which elevates the N_2 , plume length and $\log(n_e)$. The rise in electron number density resulting from enhanced ionization rates leads to an increase in the plasma plume and gas temperature, as illustrated in Fig. 5 (plume length, (Fig. 5(a)), T_g (Fig. 5(k)). Furthermore, the OH parameter displayed weak correlation with T_{exc} , $\log(n_e)$, plume length and T_g (Fig. 6(c), (d), (f) and (g)), respectively. The weak dependence of OH intensities on plasma parameters stems from H_2O dissociation kinetics rather than electron-driven processes, and its short lifetime causes rapid quenching, making it less sensitive to T_{exc} , n_e , plume length, and T_g .⁴⁴

3.2 Optimized model performance and comparative assessment

To ensure a fair and unbiased comparison among the different learning algorithms, all six regression models (KNN, ET, RF, ANN, GB, and XGB) were subjected to hyperparameter optimization using Bayesian search, as described in Section 2.2. The optimized hyperparameter configurations obtained for each model are summarized in Table 3. The corresponding default ('vanilla') parameter settings provided by the used libraries are reported in Table ST-1 for a direct comparison. This side-by-side reporting highlights the extent to which model performance is influenced by hyperparameter selection and avoids performance bias arising from default configurations. The predictive performance of the optimized models was evaluated on an independent test set using MAPE and R^2 as the primary performance metrics. For this purpose, the complete dataset was randomly partitioned into training and test subsets in a 70:30 ratio. The test set was strictly excluded from model training, thereby providing an unbiased assessment of model generalization performance on previously unseen data. The comparative results for all optimized models are presented in Fig. 7, whereas the performance of the corresponding vanilla models is provided in Fig. S1.

A comparison between the optimized and vanilla models revealed a significant improvement in predictive accuracy across all regression algorithms after hyperparameter tuning except for tree-based bagging algorithms, *i.e.* ET and RF. The hyperparameter tuned GB, KNN, MLP and XGB models exhibit lower MAPE values than their untuned counterparts for almost all targets, reflecting better generalization to the validation data. In contrast, the tree-based bagging models (ET and RF) display a marginal deterioration in MAPE after optimization, suggesting that the imposed constraints in the tuning process limited their representational capacity. In the optimized ET and RF configurations, the maximum tree depth was capped at 20, whereas the vanilla implementations allowed unrestricted tree growth, enabling them to fit more complex nonlinear relationships specific to the present dataset. While deeper trees can capture complex nonlinear patterns, they also increase the risk of overfitting, particularly for limited datasets. Across the different algorithms, the magnitude of improvement in MAPE after hyperparameter tuning varies substantially, and this pattern is particularly informative about how each model responds to optimization. The most dramatic gains are observed for the MLP and KNN regressors, where hyperparameter tuning leads to a near order-of-magnitude reduction in error for certain

plasma parameters. For instance, in the MLP model, the MAPE for O decreases from 26% in the vanilla configuration to 2.56% after tuning, while N_2 drops from 20.28% to 2.69%, indicating that the optimized network parameters allow the model to capture the nonlinear dependence of these targets far more effectively. A similarly strong effect is seen for KNN, where parameters such as O and N_2 exhibit moderate reductions in MAPE, reflecting that an appropriate choice of neighbors and distance metric significantly improves the local representation of the feature space. On the other hand, the tree-based boosting, *i.e.* GB and XGB, showed minimal improvements, implying that their baseline configurations were already reasonably close to the optimal model configuration.

Although pointwise performance metrics such as MAPE and R^2 provide a quantitative measure of prediction accuracy, they do not convey information regarding the confidence or reliability of the model predictions. Hence, uncertainty quantification and model reliability assessment were performed in the subsequent steps to evaluate predictive confidence, calibration behaviour, and robustness under data variability.

3.3 Uncertainty quantification and model reliability assessment

3.3.1 Epistemic uncertainty estimation. Epistemic uncertainty reflects uncertainty arising from limited data availability and imperfect model knowledge and is particularly relevant for data-driven plasma diagnostics where experimental measurements are sparse, costly, and unevenly distributed across the input space. In this study, epistemic uncertainty for all optimized regression models (KNN, ET, RF, ANN, GB, and XGB) was quantified using a bootstrap ensemble approach, as described in Section 2.4. The resulting uncertainty estimates were analyzed through uncertainty–error relationship plots, shown in Fig. 8(a)–(f), where the absolute prediction error $|Y - \hat{Y}|$ is plotted against the corresponding epistemic uncertainty (ensemble standard deviation) for each test instance and plasma parameter.

Across all models, a non-uniform distribution of epistemic uncertainty is observed, indicating that model confidence varies remarkably across the feature space. This behavior is expected for experimental plasma datasets characterized by nonlinear dependencies and uneven sampling density. However, the extent to which uncertainty magnitude correlates with prediction error differs markedly between learning algorithms, providing insight into the reliability of the uncertainty estimates. For the tree-based ensemble models *i.e.* ET and RF (Fig. 8(a) and (e)), a broad spread of epistemic uncertainty

Table 3 Optimized hyperparameter values for each ML regression model

ML model	Optimized hyperparameters
ET	n_estimators: 1000, max_depth: 12, min_sample_split: 3, min_sample_leaf: 3, bootstrap: 'False', criterion: 'absolute error'
GB	n_estimators: 271, learning_rate: 0.122, max_depth: 10, min_sample_split: 4, min_sample_leaf: 2, subsample: 0.6, loss: 'squared_error'
KNN	n_neighbors: 4, weights: 'distance', distance function: 'manhattan', p: NA
MLP	hidden_layer_sizes: (50,), activation: 'relu', solver: 'lbfgs', alpha: 1×10^{-6} , learning_rate_init: 0.0024
RF	n_estimators: 278, max_depth: 20, min_sample_split: 2, min_sample_leaf: 2, criterion: 'squared_error'
XGB	n_estimators: 500, learning_rate: 0.067, max_depth: 10, min_child_weight: 4, subsample: 0.88, colsample_bytree: 1

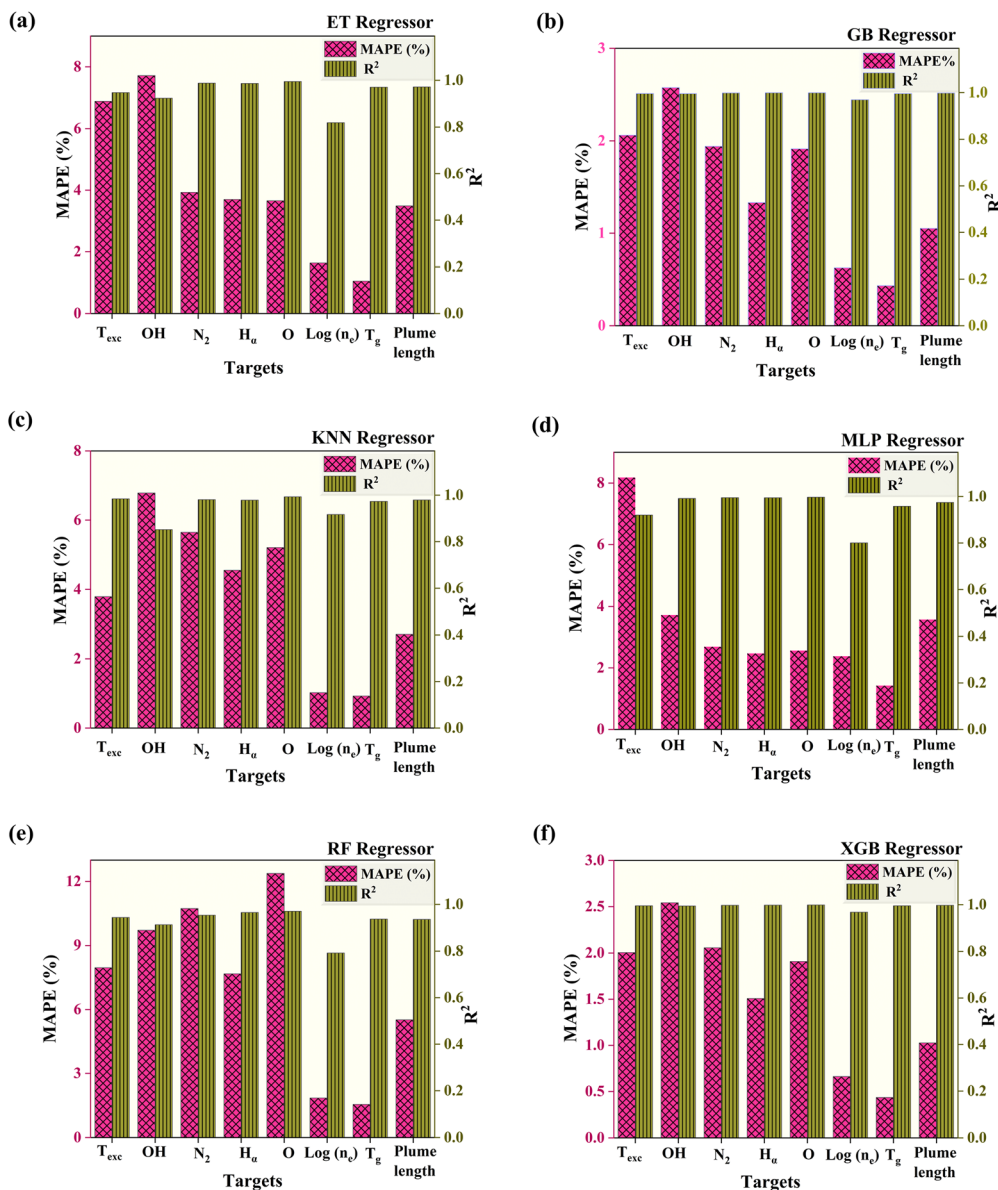


Fig. 7 The MAPE and R^2 value for the test set by using the hyperparameter tuned (a) ET, (b) GB, (c) KNN, (d) MLP, (e) RF and (f) XGB regressor ML model.

values is evident, extending up to approximately 600–700 units for ET and 400–500 units for RF. These models exhibit a moderate positive association between uncertainty and absolute error, particularly for emission-intensity related parameters (*e.g.*, OH 309 nm, O 777.4 nm), where higher uncertainty generally coincides with larger prediction errors. However, several low-error points are also associated with relatively high uncertainty, suggesting a degree of conservative uncertainty estimation inherent to bagging-based ensembles. This behavior is consistent with prior observations that randomization in tree construction can inflate variance estimates, especially when tree depth is constrained during hyperparameter optimization. The instance-based KNN model (Fig. 8(c)) displays the widest epistemic uncertainty range among all models, with uncertainty values exceeding 1000 units for certain test samples. Correspondingly, KNN also exhibits the

largest absolute errors, reaching up to ~ 3000 units for specific spectral emission parameters (OH 309 nm). While a general upward trend between uncertainty and error is visible, the dispersion is substantial, indicating that local data sparsity strongly influences both prediction accuracy and uncertainty. This highlights the sensitivity of KNN to the distribution of training samples in the feature space, where extrapolative predictions are inherently associated with high epistemic uncertainty.

In contrast, the boosting-based models, *i.e.* GB and XGB, depicted in Fig. 8(b) and (f), respectively, demonstrate comparatively compact uncertainty distributions, with epistemic uncertainty largely confined below ~ 300 –500 units. These models show a monotonic relationship between uncertainty magnitude and absolute error, where higher uncertainty consistently corresponds to increased prediction error across most

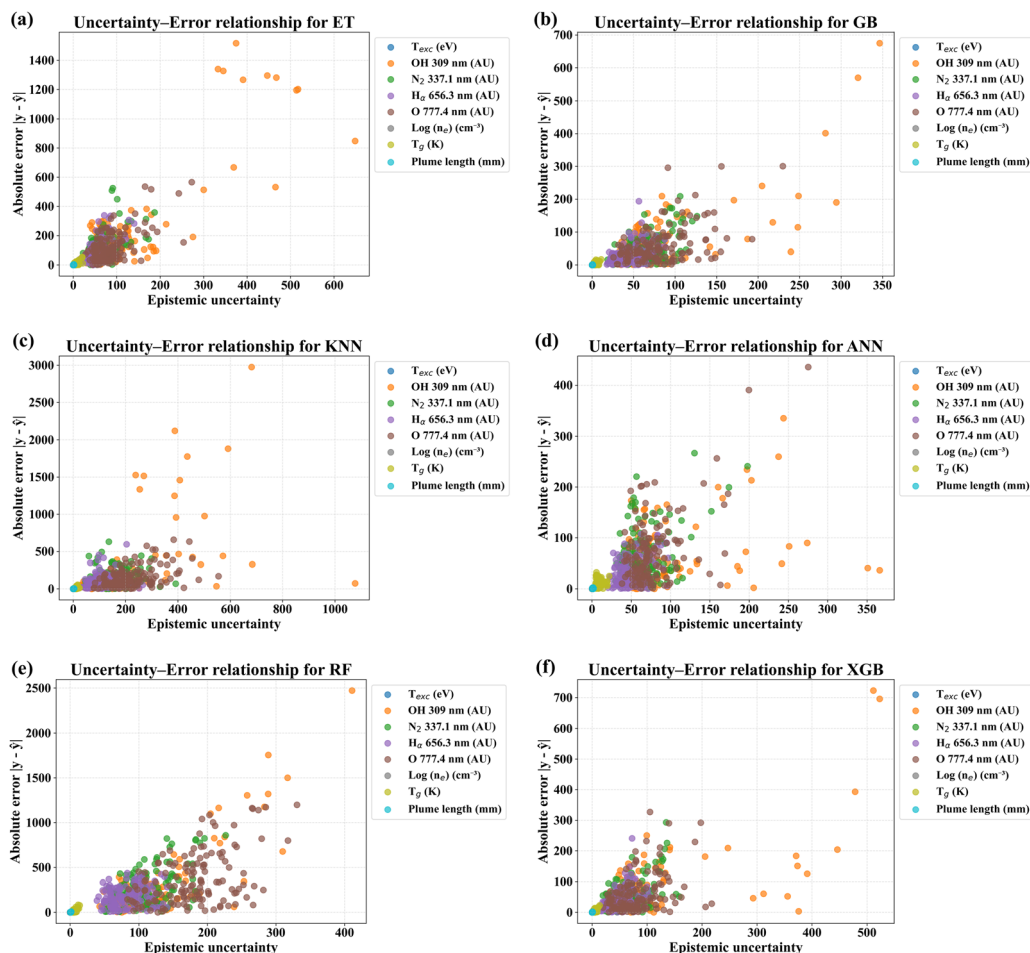


Fig. 8 Uncertainty–error relationship for the optimized regression models: (a) ET, (b) GB, (c) KNN, (d) MLP, (e) RF, and (f) XGB.

plasma parameters. This behavior suggests that boosting-based ensembles provide more informative and calibrated epistemic uncertainty estimates, likely due to their iterative residual-correction mechanism and regularization strategies, which stabilize predictions in regions supported by sufficient data. The MLP model (Fig. 8(d)) exhibits an intermediate performance, with epistemic uncertainty values primarily concentrated below 400 units and absolute errors mostly below ~ 500 units. A positive uncertainty–error relationship is evident, although some moderate-error points occur at relatively low uncertainty values, which indicates localized overconfidence. Nevertheless, compared to KNN and tree-bagging models, the ANN displays a more compact and structured uncertainty distribution.

3.3.2 Prediction interval reliability and sharpness. The reliability and informativeness of the uncertainty estimates were further assessed through prediction interval coverage probability (PICP) and mean prediction interval width (MPIW) at a nominal confidence level of 90%. While PICP quantifies whether the predicted uncertainty intervals are statistically consistent with the observed data, MPIW reflects the sharpness of these intervals. Table 4 summarizes the parameter wise PICP values obtained for all optimized regression models. A PICP value close to a nominal confidence level (~ 0.9) indicates well-calibrated

Table 4 Parameter-wise prediction interval coverage probability (PICP) at 90% confidence for the optimized regression models

Targets	ET	GB	KNN	MLP	RF	XGB
T_{exc}	0.47	0.8	0.72	0.95	0.28	0.72
OH	0.37	0.82	0.76	0.88	0.47	0.83
N_2	0.49	0.88	0.7	0.77	0.3	0.81
H_α	0.53	0.9	0.76	0.94	0.31	0.86
O	0.56	0.87	0.86	0.86	0.45	0.87
$\text{Log}(n_e)$	0.51	0.84	0.73	0.78	0.34	0.84
T_g	0.52	0.72	0.79	0.73	0.39	0.74
Plume length	0.53	0.86	0.86	0.8	0.28	0.84

uncertainty estimates, whereas lower values indicate under-coverage or overconfident predictions. Furthermore, a very high value indicates overly conservative uncertainty bounds. Among the evaluated models, the ANN, GB and XGB regressors exhibit the most consistent and near-nominal coverage across plasma parameters. For instance, the ANN model achieves PICP values of 0.95 for T_{exc} , 0.94 for H_α , and 0.88 for OH emission intensities, indicating that the constructed prediction intervals reliably enclose the true observations for the majority of targets. Similarly, GB demonstrates strong coverage performance, with PICP values exceeding 0.8 for most parameters and reaching 0.9 for

H_{α} . In contrast, the bagging-based ensemble models (ET and RF) exhibit significantly lower PICP values across most targets. For RF, coverage probabilities remain below 0.5 for all parameters, which indicates under-coverage and overconfident prediction intervals. ET shows moderately improved behavior relative to RF but still fails to reach nominal coverage, with PICP values clustering around 0.5–0.56. The instance-based KNN model displays intermediate performance, with PICP values ranging between 0.70 and 0.86 depending on the parameter, suggesting reasonable but inconsistent reliability across targets. These trends indicate that models relying heavily on local neighborhoods or bootstrap aggregation are more sensitive to data sparsity and distributional variability, leading to less reliable interval estimates.

The sharpness of the uncertainty bounds is illustrated in Fig. 9(a)–(f), which reports the MPIW values for each plasma parameter across all models. Noticeable differences in MPIW are observed both across models and across target variables, reflecting differences in output scale and intrinsic variability. For emission-intensity based targets (OH, N_2 , H_{α} , and O), KNN and RF exhibit the widest prediction intervals, with MPIW values exceeding 500 AU for O in KNN and RF, which implies

highly diffuse uncertainty estimates. In contrast, ANN, GB, and XGB produce narrower intervals (<270 AU) for the same parameters. For example, the MPIW for O is reduced from approximately 704 AU (KNN) and 568 AU (RF) to 241 AU (ANN), 262.5 AU (GB), and 258.1 AU (XGB), representing a reduction of more than 50% in uncertainty width. For thermophysical targets such as T_g , $\log(n_e)$, and plume length, all models yield comparatively narrow intervals. However, ANN and XGB again demonstrate the best balance between sharpness and reliability, maintaining MPIW values below 30 K for T_g and below 3 mm for plume length while simultaneously achieving high PICP values. This combination indicates that these models provide informative uncertainty bounds without sacrificing statistical coverage.

3.3.3 Calibration behavior. Fig. 10(a)–(f) presents the calibration curves for all optimized regression models, comparing nominal prediction interval coverage with the empirically observed coverage on the test set. An ideally calibrated model is expected to follow the diagonal line, indicating statistical consistency between predicted uncertainty levels and observed outcomes. Deviations below the diagonal indicate systematic

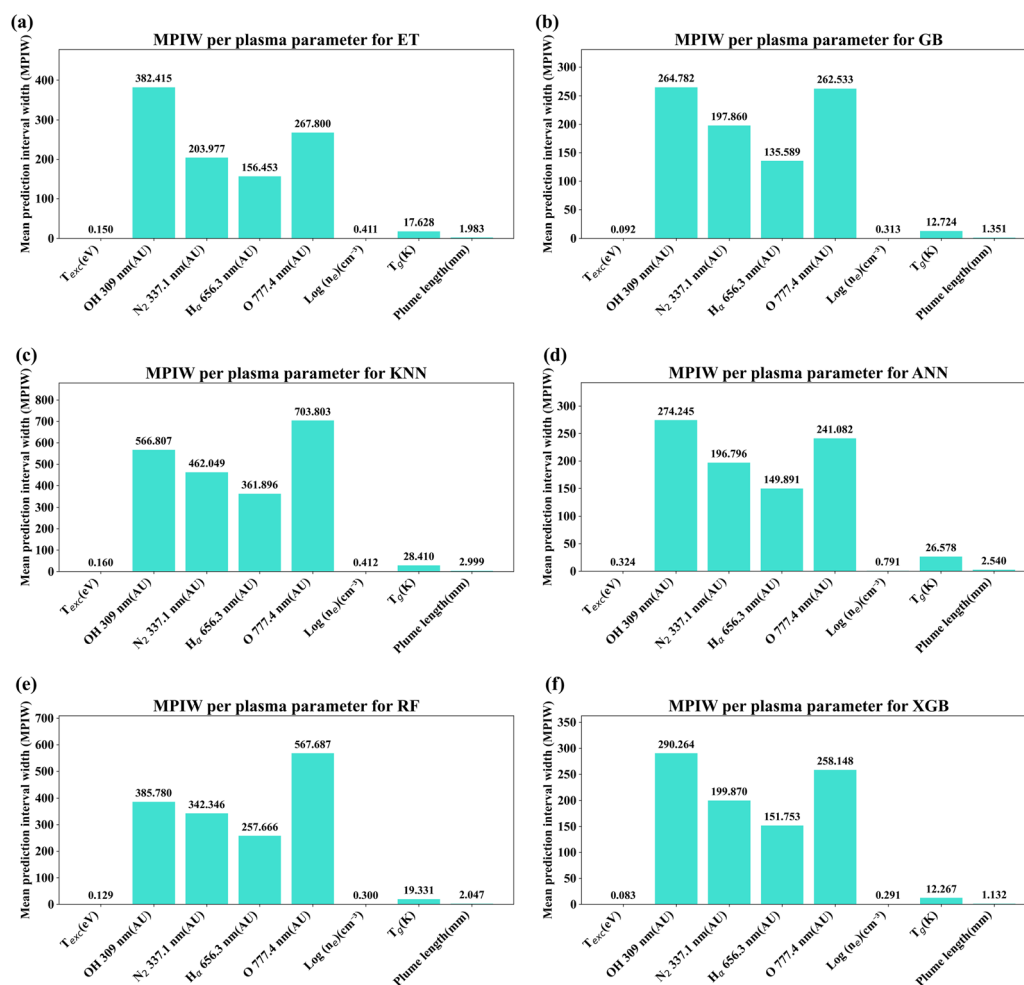


Fig. 9 Mean prediction interval width (MPIW) at 90% confidence for individual plasma parameters obtained using the optimized (a) ET, (b) GB, (c) KNN, (d) ANN, (e) RF, and (f) XGB regression models.

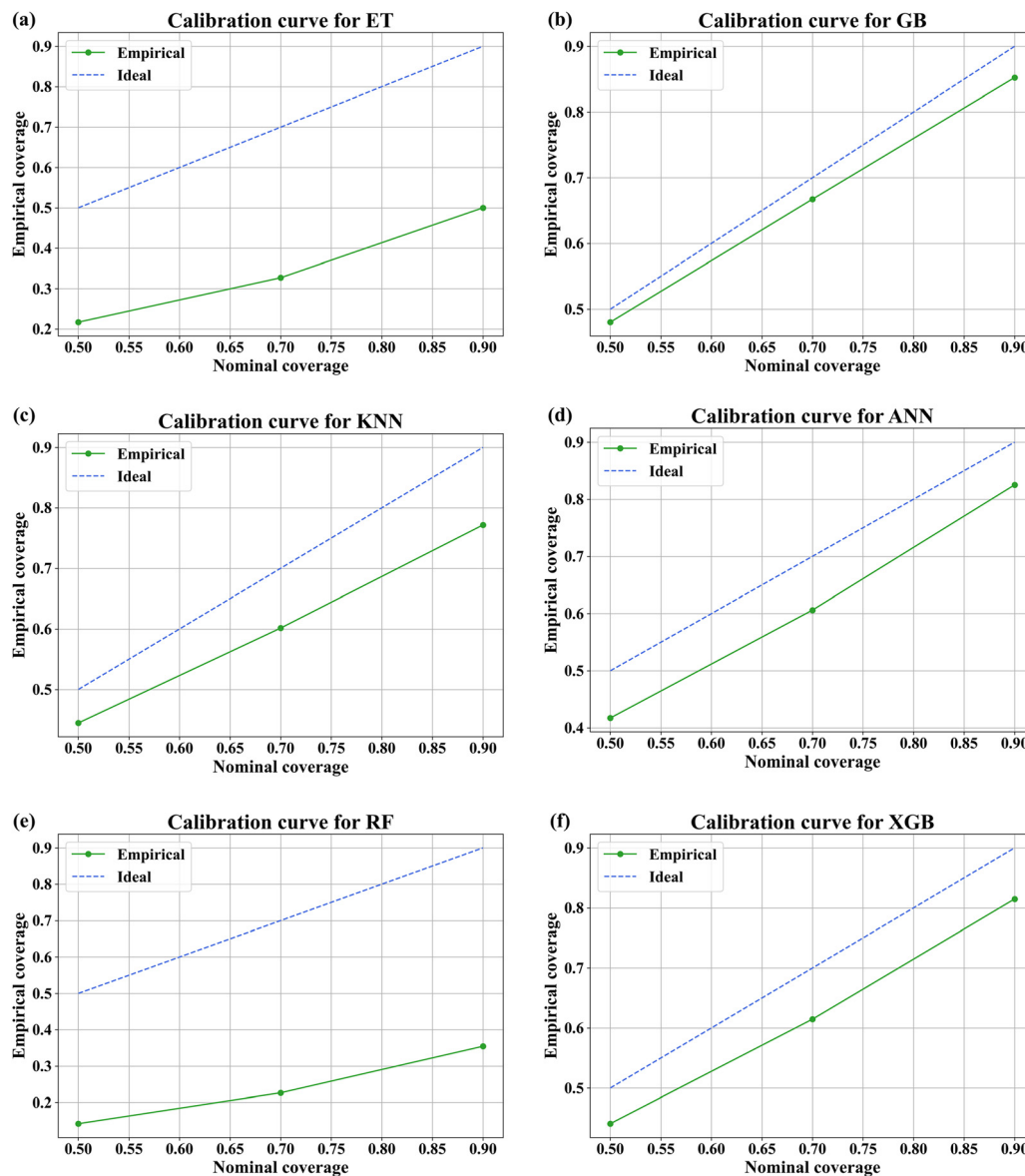


Fig. 10 Calibration curves comparing nominal coverage and empirical coverage for the optimized (a) ET, (b) GB, (c) KNN, (d) ANN, (e) RF, and (f) XGB regression models.

under-coverage (overconfident predictions), whereas deviations above the diagonal indicate overly conservative uncertainty estimates. The ET and RF models exhibit pronounced under-calibration across all nominal coverage levels (Fig. 10(a) and (e)). For these models, empirical coverage remains substantially below the nominal confidence even at the 90% level, with RF showing empirical coverage below 0.4 at all tested levels. This behavior is in line with the low PICP values reported in Table 4 and confirms that the uncertainty estimates produced by these bagging-based models are overconfident and unreliable for plasma parameter prediction. The KNN model depicted in Fig. 10(c) demonstrates moderate calibration performance, with empirical coverage increasing approximately linearly with nominal coverage. However, the calibration curve consistently lies below the ideal line, indicating persistent under-coverage.

This suggests that although KNN captures relative uncertainty trends, its prediction intervals remain insufficiently wide to achieve nominal statistical coverage.

In contrast, ANN, XGB and GB display the most favorable calibration characteristics among all evaluated models (Fig. 10(b), (d), and (f) respectively). For these models, the empirical coverage closely tracks the diagonal across all nominal confidence levels, particularly at 70% and 90%. This indicates that the predicted uncertainty intervals are statistically consistent and neither overly optimistic nor excessively conservative. Notably, GB exhibits tighter adherence to ideal calibration at each confidence level, while ANN and XGB maintain stable behavior across the full range, corroborating their strong PICP performance reported in Table 4.

Considering the combined insights from scoring metrics, calibration analysis, prediction interval coverage, and uncertainty-error

consistency, the GB model demonstrates the most reliable uncertainty characterization among the evaluated regressors. While ANN and XGB achieve competitive point prediction accuracy, the GB model exhibits the closest alignment between empirical and nominal coverage across confidence levels, indicating superior probabilistic calibration. Moreover, GB maintains high PICP values without resorting to excessively wide prediction intervals, reflecting an optimal balance between reliability and sharpness. The uncertainty–error relationship further confirms that regions of elevated prediction error are consistently associated with higher epistemic uncertainty for GB, reinforcing the physical interpretability of its uncertainty estimates. On the basis of these combined criteria, GB was selected as the most suitable model for subsequent generalization analysis and experimental validation. The parity plot for all target variables (Fig. 11(a)–(h)) depicts the correlation between predicted and actual values obtained from the GB model. The majority of data points align closely along the 45° diagonal, which indicates strong predictive accuracy across the target. Minor deviations observed in the case of $\log(n_e)$ suggest under- or over-prediction for a specific region of feature space (Fig. 11(g)). Such deviations are expected for parameters exhibiting higher intrinsic variability and reduced signal sensitivity and are consistent with the comparatively higher uncertainty bounds observed for this target in the uncertainty quantification analysis.

3.4 Model interpretability using SHAP

From the previous analysis, it was found that the GB model outperformed all other machine learning models considered in this study in terms of both predictive accuracy and uncertainty reliability. Therefore, to gain a deeper understanding of its predictive behaviour from a scientific standpoint, SHAP (Shapley additive exPlanations) analysis was conducted to interpret the model's outputs and assess the relative contribution of each input feature to the predictions.⁴⁵ Fig. 12 depicts the SHAP bar plot for all the output variables. These plots illustrate the average magnitude of each feature's contribution to the model predictions. Notably, among features, power exhibited the highest mean SHAP value across all outputs, except for OH 309 nm, where flow rate emerged as the most crucial feature. In contrast, sliding short position consistently showed the lowest mean SHAP value across all the targets, suggesting it had the least influence on the model's prediction. Although the bar plots provided a clear overview of feature importance, they did not capture the relationship between the individual features and model outputs. Therefore, SHAP summary plots for all target variables are generated and presented in Fig. 13(a)–(h). In these plots, the y-axis represents the input features, while the x-axis displays the SHAP values for each data point in the test set. Across all targets, higher power values (indicated by red markers) predominantly correspond to positive SHAP values, demonstrating a strong positive correlation between input power and plasma parameters. This indicates a strong positive correlation between power input and these plasma parameters, which is also studied by Zaplotnik *et al.*⁴⁴ The increase in input power significantly influences plasma parameters by enhancing the energy deposition per neutral particle within the

discharge. This elevates the electron-impact ionization and excitation rates, which leads to a rise in plasma parameters. In the case of OH (Fig. 13(b)), low flow rates (indicated by blue) are associated with high positive SHAP values, meaning that lower flow rates contribute more positively to the prediction of OH 309 nm emission intensity. In other words, there exists an inverse relationship between flow rate and the OH 309 nm spectral line. The observed reduction in the intensity of the OH 309 nm emission line with increasing gas flow rate can be attributed to the decrease in dissociation efficiency of precursor water molecules (H_2O) within the plasma discharge region. At higher flow rates, the residence time of H_2O molecules in the active plasma zone is significantly reduced, limiting the dissociation processes, which leads to diminished spectral emission intensity. Further, SHAP summary plots for N_2 , H_α , O, T_g (Fig. 13(c), (d), (e) and (g) respectively) reveal a wider SHAP value spread for both power and flow rate, which indicates nonlinear and complex interaction between these input features and the corresponding outputs. Both input parameters are inherently nonlinear and exhibit complex interactions due to competing physicochemical mechanisms arising from the interplay between energy coupling, transport phenomena, and reaction kinetics. The sliding short position exhibits a consistently narrow range of SHAP values with minimal deviation from zero across all targets, reaffirming its limited influence on model predictions. The observed minimal dependence of plasma parameters on sliding short position variations can be explained through microwave plasma coupling dynamics. The sliding short functions as a reflective boundary condition that determines the standing wave pattern and consequent power coupling efficiency to the plasma discharge, making it more stable. For given Ar gas, the system exhibits consistent impedance matching, and invariant power absorption, which leads to the discharge being stable.

As the SHAP analysis indicated a negligible contribution of the sliding short position to model predictions, an additional reduced-order modeling exercise was performed. The hyperparameter-tuned GB model was retrained using a reduced feature set comprising only power and flow rate. The predictive performance of the reduced model was then compared with that of the full feature model for all plasma parameters. The results of the reduced feature modelling analysis are summarized in Table 5. Upon removal of the sliding short position, a degradation in predictive performance is observed across all targets. The deterioration in performance metrics is severe in the case of T_{exc} , OH and $\log(n_e)$. The MAPE and MSE value showed a sharp increase for these targets whereas the R^2 significantly increased. These results are in line with the mean SHAP plot (Fig. 12) and SHAP summary plot (Fig. 13), where a higher contribution from the sliding short position can be clearly seen for T_{exc} , OH and $\log(n_e)$ targets. These results indicate that, although the sliding short position exhibits a low mean SHAP magnitude and appears less influential on a pointwise basis, it still contributes non-negligible contextual information that enhances the global predictive capability of the model. The degradation observed in the reduced-feature configuration suggests that the sliding short position indirectly stabilizes the learned relationships between power,

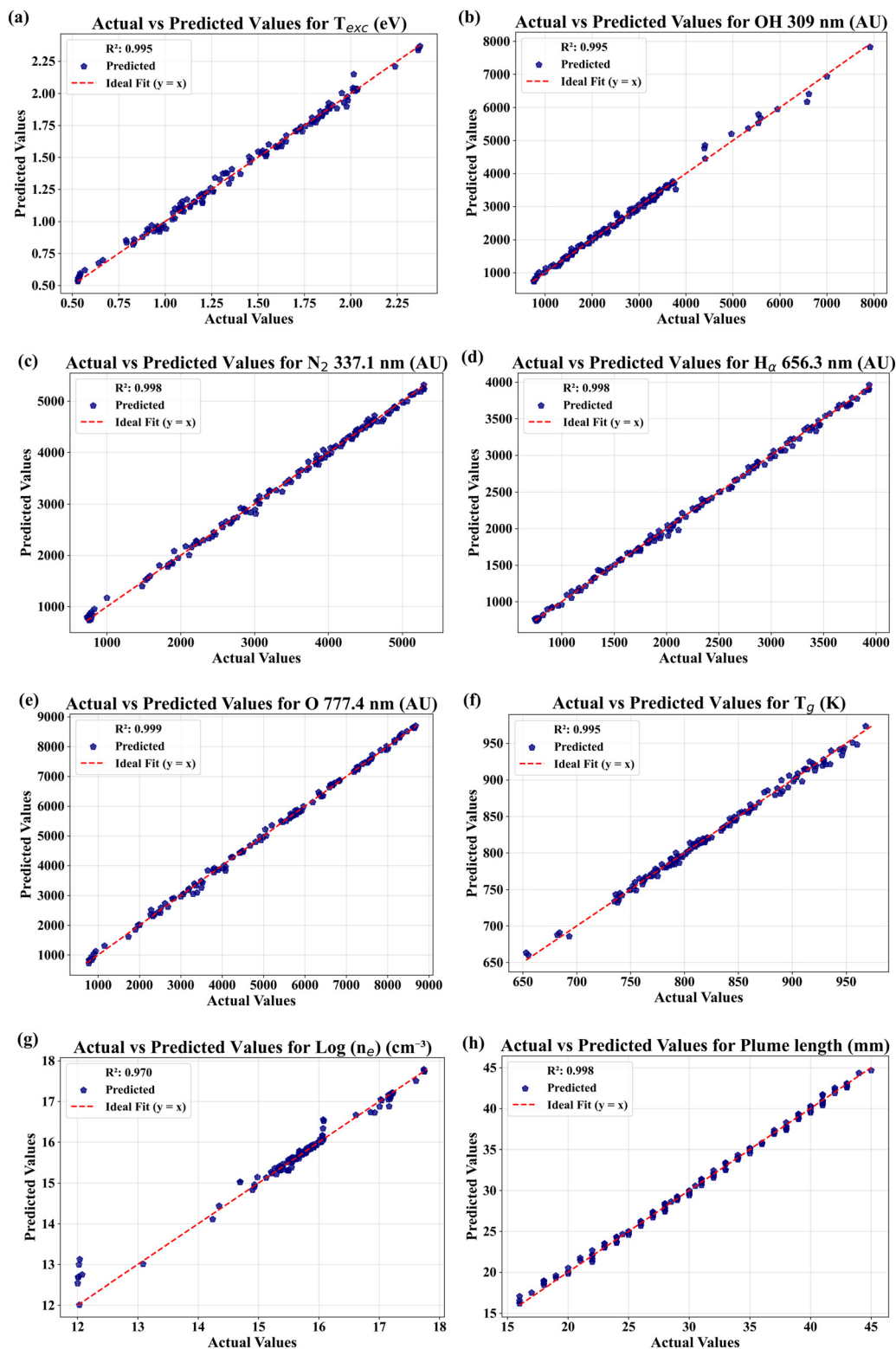


Fig. 11 Parity plots for (a) T_{exc} , (b) OH (309 nm), (c) N_2 (337.1 nm), (d) H_α (656.3 nm), (e) O (777.4 nm), (f) T_g , (g) $\log(n_e)$, and (h) plume length, predicted by the optimized GB model.

flow rate, and plasma response, likely by encoding impedance-matching and coupling conditions that are not fully captured by the remaining inputs alone. Therefore, despite its weak

direct attribution in SHAP analysis, retaining the sliding short position is necessary to preserve overall model accuracy and generalization.

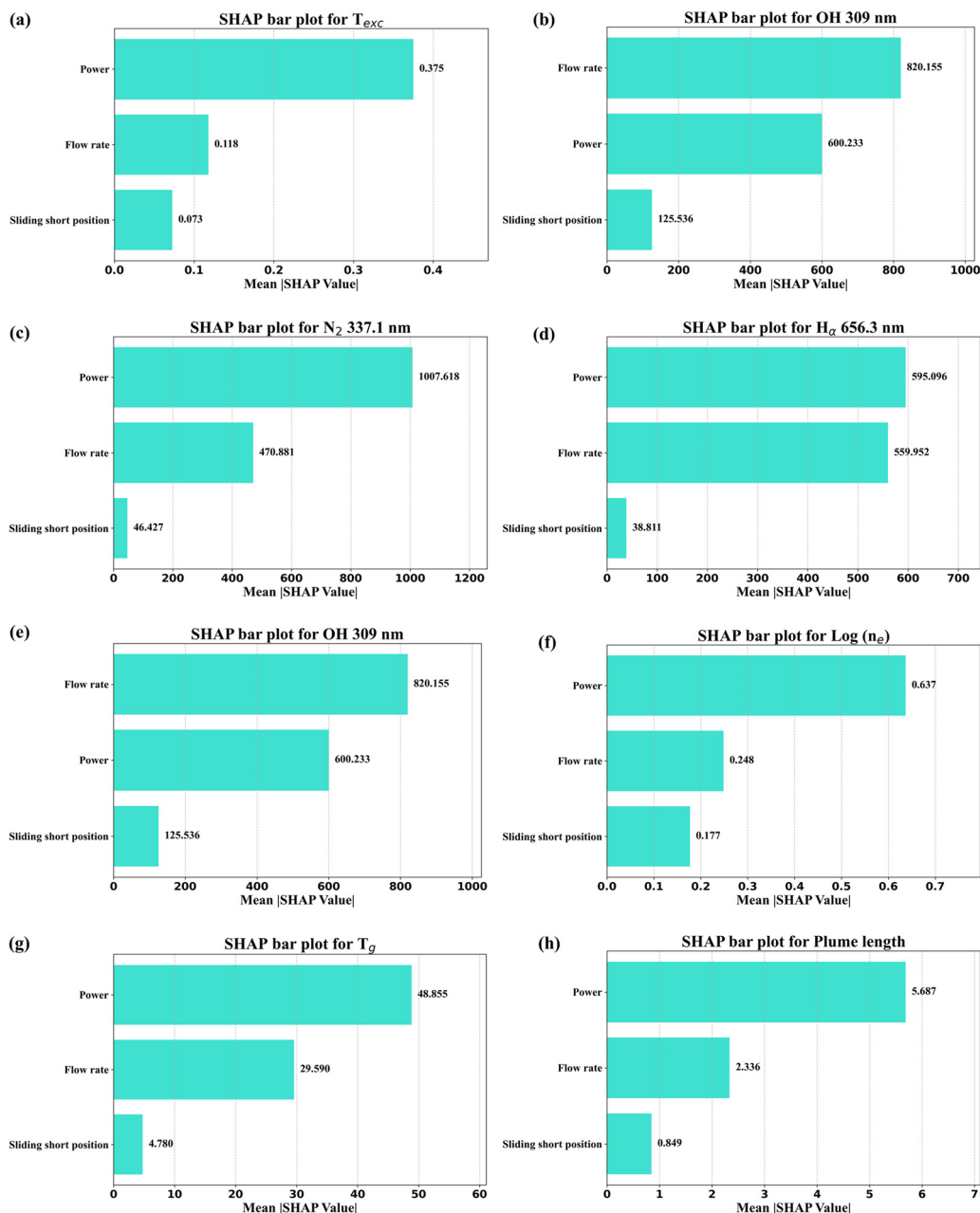


Fig. 12 SHAP bar plot showing mean absolute Shapley values for (a) T_{exc} , (b) OH 309 nm, (c) N_2 337.1 nm, (d) H_α 656.3 nm, (e) O 777.4 nm, (f) $\log(n_e)$, (g) T_g , and (h) plume length.

4. Experimental validation and generalization assessment of the ML models

To rigorously evaluate the real-world predictive capability and generalization behaviour of the optimized GB model, an independent experimental validation study was conducted. Experimental validation was performed using 30 independent plasma operating conditions, explicitly separated into 15 interpolated and 15 extrapolated points based on their location relative to the training-domain feature ranges. The interpolated points were selected within the operating window spanned by the

training data, whereas the extrapolated points intentionally lay outside this domain to assess model behaviour under out-of-distribution conditions. This separation enables a more nuanced assessment of model robustness beyond conventional random test-set evaluation. The experimental input conditions, along with the corresponding measured plasma parameters and model predictions, are provided in the gitfront repository. For each experimental instance, the GB model was used to generate point predictions as well as prediction intervals, allowing both accuracy and reliability to be assessed under unseen conditions. Performance was quantified using MAPE, MSE, PICP (90%) and MPIW under experimental deployment.

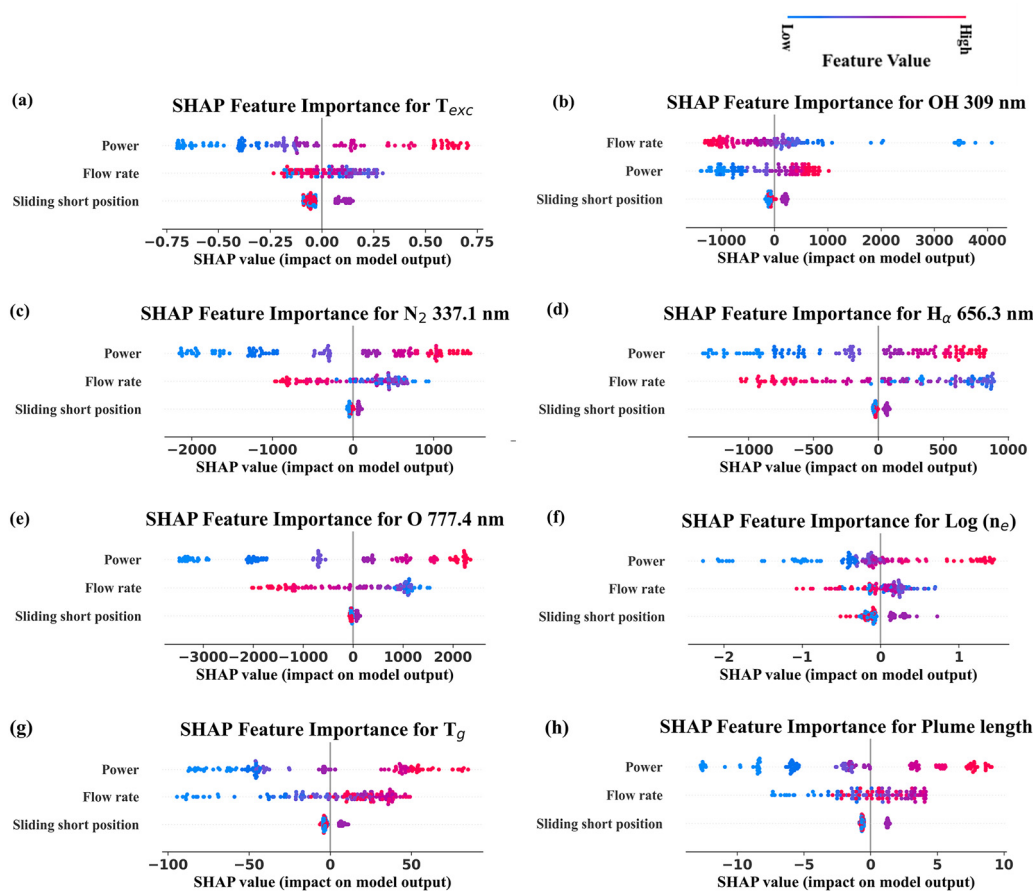


Fig. 13 SHAP summary plot for (a) T_{exc} , (b) OH 309 nm, (c) N_2 337.1 nm, (d) H_α 656.3 nm, (e) O 777.4 nm, (f) $\log(n_e)$, (g) T_g , and (h) plume length.

Table 5 Comparison of predictive performance of the optimized GB model using the full feature set and a reduced feature set excluding the sliding short position, evaluated using MAPE, R^2 , and MSE for each plasma parameter

Target	Full feature GB			Reduced feature GB		
	MAPE (%)	R^2	MSE	MAPE (%)	R^2	MSE
T_{exc}	2.06	0.995	0.001	9.03	0.908	0.019
OH	2.57	0.994	9352.15	8.99	0.963	64 083.97
N_2	1.94	0.997	3620.65	3.74	0.994	8700.36
H_α	1.33	0.998	1277.34	3.18	0.992	6166.44
O	1.91	0.998	6379.5	3.25	0.996	15 259.23
$\log(n_e)$	0.62	0.97	0.04	2.03	0.783	0.27
T_g	0.43	0.995	21.97	1.04	0.974	118.14
Plume length	1.05	0.997	0.13	4.5	0.96	2.19

Unlike the test dataset, which represents random samples drawn from the same distribution as the training data, the experimental dataset introduces additional sources of variability arising from measurement noise, environmental fluctuations, and potential deviations in plasma stability. Consequently, experimental validation provides a more stringent and practically relevant benchmark for model performance. To capture this behaviour, interpolated and extrapolated experimental points were evaluated separately, and their predictive accuracy and uncertainty characteristics were analysed independently.

The performance evaluation of the optimized GB regressor for the unseen dataset was done by determining the MAPE and MSE regression metrics. These metrics were compared with their respective values on the hold-out test set to evaluate model generalization. The comparison of MAPE and MSE is presented as a radar plot in Fig. 14(a) and 3D bar plot in Fig. 14(b), respectively. For the interpolation regime, the MAPE values for most plasma parameters remained below 6%, with the lowest error observed for $\log(n_e)$ and the highest for T_{exc} . For the extrapolated regime, the MAPE value showed relatively large deviation from the hold-out-test set MAPE value (seen from Fig. 7(b)), specifically for emission-intensity based targets (OH, N_2 , H_α and O).

To complement pointwise accuracy metrics, the reliability of the optimized GB model under experimental deployment was further evaluated using PICP (90%). While MAPE and MSE quantify average predictive accuracy, they do not capture whether the model's uncertainty estimates remain statistically consistent when transitioning from test data to real experimental conditions. Therefore, PICP analysis was employed to assess the robustness of the model's uncertainty quantification across different generalization regimes. Fig. 14(c) presents a comparative bar chart of the 90% PICP values obtained for interpolated experimental points and extrapolated experimental points for each plasma parameter. For the interpolation regime, the PICP

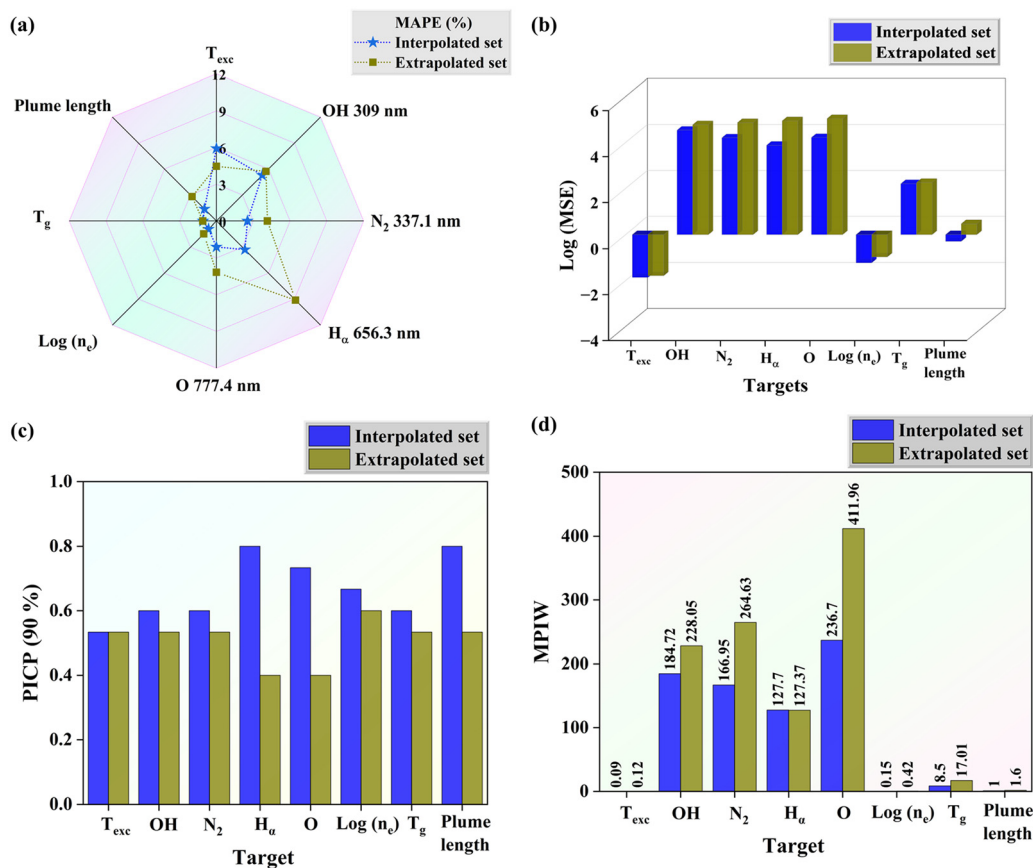


Fig. 14 Radar plots comparing (a) MAPE and (b) 3D bar plot of $\log(MSE)$, (c) PICP at 90% confidence and (d) MPIW for the experimental datasets for the experimental datasets.

values remained close to the nominal confidence level for most targets (between 0.53–0.8), indicating that the prediction intervals constructed from the bootstrap ensemble reliably enclosed the true experimental observations. The corresponding MPIW values for the interpolated points (Fig. 14(d)) remained relatively narrow, confirming that reliable coverage was achieved without excessively conservative uncertainty bounds. In contrast, the extrapolation regime exhibited a more pronounced deviation in PICP values, particularly for emission-intensity based targets (H_α , O) and plume length. For H_α , O and plume length, the PICP decreased to 0.4, 0.4, and 0.53, respectively, reflecting increased difficulty in capturing unseen nonlinear behaviour outside the training domain. Importantly, this reduction in coverage was accompanied by a widening of the prediction intervals, as evidenced by the increased MPIW values in Fig. 14(d), with the MPIW for O increasing from 236.7 (interpolation) to 411.96 (extrapolation). This behaviour indicates that the model appropriately responds to a distributional shift by expressing higher epistemic uncertainty rather than producing overconfident predictions. Altogether, the PICP comparison across the test, interpolation, and extrapolation regimes demonstrates that the optimized GB model maintains statistically meaningful uncertainty estimates under experimental deployment. While coverage naturally deteriorates for certain targets under extrapolation, the model responds by increasing uncertainty width, thereby

preserving reliability. This behaviour is critical for practical plasma diagnostics, where reliable uncertainty bounds are essential for informed decision-making, especially when operating near or beyond previously explored process conditions.

This work employs a data-driven, uncertainty-aware machine-learning framework for the simultaneous prediction of multiple plasma parameters from experimentally accessible operating conditions. The results demonstrate that the proposed approach achieves high predictive accuracy together with statistically meaningful uncertainty estimates across both interpolative and extrapolative regimes, thereby providing a reliable surrogate for plasma diagnostics under practical operating conditions. Building upon this foundation, the incorporation of physics-based descriptors or hybrid ML-physics surrogates represents a promising direction to further enhance extrapolation performance and physical interpretability. Moreover, the integration of inverse design^{46,47} and process-optimization strategies⁴⁸ will be pursued in future work, where the present predictive framework will serve as the surrogate model for identifying operating conditions that yield target species concentrations or desired plasma states. In addition, the active learning strategy can also be employed, which will enable iterative refinement of the plasma database by adaptively selecting the most informative experiments. Such strategies are expected to improve data efficiency and accelerate model improvement, with the current

supervised models providing a robust baseline and initial surrogate within an integrated active-learning loop.

5. Conclusions

This study presents a comprehensive, uncertainty-aware multi-output machine learning framework for predicting key plasma parameters in MW-APPJs. A curated experimental dataset comprising 441 instances was generated by varying microwave power, sliding short position, and argon flow rate, with simultaneous measurement of eight plasma characteristics. Six regression models (KNN, ET, RF, ANN, GB, and XGB) were systematically optimized using Bayesian hyperparameter tuning and evaluated under a unified performance and reliability framework. While the hyperparameter tuned XGB model achieved competitive pointwise accuracy, the optimized gradient boosting (GB) regressor emerged as the most balanced model when accuracy, calibration, and uncertainty reliability were jointly considered. On the held-out test set, the GB model achieved MAPE values below 3% for all the plasma parameters and R^2 values exceeding 0.97 across all outputs. Bootstrap-based uncertainty quantification revealed that GB provided near-nominal 90% prediction interval coverage (PICP varying within 0.8–0.9 for all targets) while maintaining comparatively narrow prediction intervals, outperforming instance-based and bagging-based models that exhibited under-coverage or excessively wide uncertainty bounds. Calibration curve analysis further confirmed that the GB model exhibited the closest alignment to the ideal diagonal among all evaluated models, indicating statistically consistent uncertainty estimates. Experimental validation was performed using 30 independent plasma operating conditions, explicitly separated into 15 interpolated and 15 extrapolated feature sets based on their location relative to the training-domain feature ranges. For interpolated conditions, experimental MAPE values remained below ~6% for all outputs with R^2 values exceeding 0.85, which confirmed the strong generalization within the training domain. Under extrapolation, increased errors were observed, specifically for emission-intensity based targets (H_{α} , O). However, this degradation was accompanied by an increase in epistemic uncertainty, preserving coverage and reliability. SHAP analysis of the optimized GB model identified microwave power as the dominant controlling feature for most plasma parameters, while gas flow rate governed OH emission intensity. The reduced feature set analysis implied that, although the sliding short position exhibited low feature importance in SHAP analysis, its inclusion improved predictive accuracy for specific plasma parameters such as T_{exc} , OH and $\log(n_e)$. Altogether, the uncertainty-aware hyperparameter tuned regression models demonstrated a pathway for accurate prediction and robust generalization of plasma parameters, which paves the way for optimized, energy efficient, and sustainable plasma processes in future industrial applications. This work also provides a strong foundation for future integration of physics-guided features, active learning strategies for plasma process control and optimization.

Author contributions

Suryasunil Rath: writing – original draft, software, methodology, investigation, formal analysis, and data curation. Priyabata Das: writing – original draft, software, methodology, formal analysis, and data curation. Pulak Mohan Pandey: supervision. Satyananda Kar: writing – review and editing, supervision, resources, and project administration.

Conflicts of interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data is available within this article and supplementary information (SI). Supplementary information is available. See DOI: <https://doi.org/10.1039/d5cp04364f>.

Code availability statement: the code used for data preprocessing, machine learning model development, and analysis can be accessed at <https://gitfront.io/r/daspriyabrata/tLZPyz1bCqNg/Plasma-ML-Prediction/>.

Acknowledgements

SR acknowledges the Ministry of Education, Govt. of India, for the stipend during his PhD tenure.

References

- 1 C. Tendero, C. Tixier, P. Tristant, J. Desmaison and P. Leprince, Atmospheric pressure plasmas: A review, *Spectrochim. Acta, Part B*, 2006, **61**(1), 2–30.
- 2 A. B. Mallick, G. V. Prakash, S. Kar and R. Narayanan, Development of a pulse modulated sub-radio frequency power supply for atmospheric pressure plasma devices, *Rev. Sci. Instrum.*, 2023, **94**(12), 123508.
- 3 T. P. Radhika and S. Kar, Effect of an additional floating electrode on radio frequency cross-field atmospheric pressure plasma jet, *Sci. Rep.*, 2023, **13**(1), 10665.
- 4 X. Lu, D. Liu, Y. Xian, L. Nie, Y. Cao and G. He, Cold atmospheric-pressure air plasma jet: Physics and opportunities, *Phys. Plasmas*, 2021, **28**(10), 100501.
- 5 X. Zhang and M. S. Cha, Ammonia cracking for hydrogen production using a microwave argon plasma jet, *J. Phys. D: Appl. Phys.*, 2023, **57**(6), 65203.
- 6 R. Antunes, A. Meindl, C. Kranig, A. Hecimovic and U. Fantz, Hydrogen production from ammonia using a microwave plasma torch at atmospheric pressure, *Int. J. Hydrogen Energy*, 2025, **170**, 151121.
- 7 A. Bogaerts, T. Kozák, K. van Laer and R. Snoeckx, Plasma-based conversion of CO₂: current status and future challenges, *Faraday Discuss.*, 2015, **183**, 217–232.

- 8 S. Das, B. Mishra, S. Mohapatra, B. P. Tripathi, S. Kar and S. Bhatt, Efficacy of argon cold atmospheric pressure plasma jet on hospital surface decontamination and its impact on the surface property, *Phys. Scr.*, 2024, **99**(2), 25601.
- 9 P. Mahreen, P. Kar and S. Sahu, Atmospheric Pressure Non-Thermal Plasma in Food Processing, *Food Process*, ImprintCRC Press, 1st edn, 2021, p. 29, DOI: [10.1201/9781003163213-10](https://doi.org/10.1201/9781003163213-10).
- 10 S. Das, S. Mohapatra, S. Kar, S. Bhatt and S. Pundir, Reactive species variation in cold atmospheric pressure plasma jet discharge under the influence of intrinsic parameters and its effect on *E. coli* inactivation, *Biointerphases*, 2023, **18**(6), 61003.
- 11 S. Das, S. Mohapatra and S. Kar, Elucidating the bacterial inactivation mechanism by argon cold atmospheric pressure plasma jet through spectroscopic and imaging techniques, *J. Appl. Microbiol.*, 2024, **135**(9), 1xae238.
- 12 T. Rana and S. Kar, Assessment of energy consumption and environmental safety measures in a plasma pyrolysis plant for eco-friendly waste treatment, *J. Energy Inst.*, 2024, **114**, 101617.
- 13 T. Rana and S. Kar, Study of preheating temperature and electrode consumption in a plasma gasification system for waste processing, *J. Energy Inst.*, 2025, **120**, 102122.
- 14 M. Leins, J. Kopecki, S. Gaiser, A. Schulz, M. Walker, U. Schumacher, U. Stroth and T. Hirth, Microwave Plasmas at Atmospheric Pressure, *Contrib. Plasma Phys.*, 2014, **54**(1), 14–26.
- 15 S. Zhang, Z. Chen, J. Yang, S. Chen, D. Feng, Y. Zhou, B. Wang and X. Lu, Study on discharge mode and transition mechanism of atmospheric pressure Ar/Zn pulsed microwave plasma jet, *AIP Adv.*, 2021, **11**(9), 95201.
- 16 C. M. Ferreira, Theory of a plasma column sustained by a surface wave, *J. Phys. D: Appl. Phys.*, 1981, **14**(10), 1811.
- 17 E. Tatarova, F. M. Dias, E. Felizardo, J. Henriques, C. M. Ferreira and B. Gordiets, Microwave plasma torches driven by surface waves, *Plasma Sources Sci. Technol.*, 2008, **17**(2), 24004.
- 18 M. Moisan and H. Nowakowska, Contribution of surface-wave (SW) sustained plasma columns to the modeling of RF and microwave discharges with new insight into some of their features. A survey of other types of SW discharges, *Plasma Sources Sci. Technol.*, 2018, **27**, 073001.
- 19 M. Moisan, I. P. Ganachev and H. Nowakowska, Concept of power absorbed and lost per electron in surface-wave plasma columns and its contribution to the advanced understanding and modeling of microwave discharges, *Phys. Rev. E*, 2022, **106**(4), 45202.
- 20 S. Kar, L. Alberts and H. Kousaka, Microwave power coupling in a surface wave excited plasma, *AIP Adv.*, 2015, **5**(1), 17104.
- 21 S. Kar, H. Kousaka and L. L. Raja, Spatio-temporal behavior of microwave sheath-voltage combination plasma source, *J. Appl. Phys.*, 2015, **117**(18), 183302.
- 22 S. Rath and S. Kar, Microwave atmospheric pressure plasma jet: A review, *Contrib. Plasma Phys.*, 2025, **65**(2), e202400036.
- 23 M. Moisan, Z. Zakrzewski and J. C. Rostaing, Waveguide-based single and multiple nozzle plasma torches: The TIAGO concept, *Plasma Sources Sci. Technol.*, 2001, **10**(3), 387–394.
- 24 A. Hamdan, C. Gagnon, M. Aykul and J. Profili, Characterization of a microwave plasma jet (TIAGO) in-contact with water: Application in degradation of methylene blue dye, *Plasma Processes Polym.*, 2020, **17**(3), 1900157.
- 25 H. S. Uhm, Y. C. Hong and D. H. Shin, A microwave plasma torch and its applications, *Plasma Sources Sci. Technol.*, 2006, **15**, S26.
- 26 T. van der Gaag, H. Onishi and H. Akatsuka, Arbitrary EEDF determination of atmospheric-pressure plasma by applying machine learning to OES measurement, *Phys. Plasmas*, 2021, **28**(3), 33511.
- 27 D. Gidon, X. Pei, A. D. Bonzanini, D. B. Graves and A. Mesbah, Machine Learning for Real-Time Diagnostics of Cold Atmospheric Plasma Sources, *IEEE Trans. Radiat. Plasma Med. Sci.*, 2019, **3**(5), 597–605.
- 28 J. Trieschmann, L. Vialetto and T. Gergs, Review: Machine learning for advancing low-temperature plasma modeling and simulation, *J. Micro/Nanopatterning, Mater., Metrol.*, 2023, **22**(4), 41504.
- 29 A. Mesbah and D. B. Graves, Machine learning for modeling, diagnostics, and control of non-equilibrium plasmas, *J. Phys. D: Appl. Phys.*, 2019, **52**(30), 30LT02.
- 30 A. N. Bakhtiyari, Y. Wu, D. Qi and H. Zheng, Modeling temporal and spatial evolutions of laser-induced plasma characteristics by using machine learning algorithms, *Optik*, 2023, **272**, 170297.
- 31 I. Suresh, P. S. N. S. R. Srikar and R. K. Gangwar, Integration of ML methods with CR model-based optical diagnostic for the estimation of electron temperature in Ga laser produced plasma, *Phys. Plasmas*, 2024, **31**(11), 113501.
- 32 J. Li, J. Xu, E. Rebrov, B. Wanten and A. Bogaerts, Machine learning-based prediction and optimization of plasma-based conversion of CO₂ and CH₄ in an atmospheric pressure glow discharge plasma, *Green Chem.*, 2025, **27**(15), 3916–3931.
- 33 J. Li, J. Xu, E. Rebrov and A. Bogaerts, Machine learning-based prediction and optimization of plasma-catalytic dry reforming of methane in a dielectric barrier discharge reactor, *Chem. Eng. J.*, 2025, **507**, 159897.
- 34 C. Bong, B. S. Kim, M. H. A. Ali, D. Kim and M. S. Bak, Machine learning-based prediction of operation conditions from plasma plume images of atmospheric-pressure plasma reactors, *J. Phys. D: Appl. Phys.*, 2023, **56**(25), 254002.
- 35 M. Witman, D. Gidon, D. B. Graves, B. Smit and A. Mesbah, Sim-to-real transfer reinforcement learning for control of thermal effects of an atmospheric pressure plasma jet, *Plasma Sources Sci. Technol.*, 2019, **28**(9), 95019.
- 36 L. Lin, S. Gershman, Y. Raitses and M. Keidar, Data-driven prediction of the output composition of an atmospheric pressure plasma jet, *J. Phys. D: Appl. Phys.*, 2023, **57**(1), 15203.
- 37 H. Jacobs, G. Novick, C. M. LoCascio and M. M. Chrepta, Measurement of Guide Wavelength in Rectangular Dielectric Waveguide, *IEEE Trans. Microwave Theory Tech.*, 1976, **24**(11), 815–820.
- 38 T. P. Rathika and S. Kar, Glow-to-arc discharge transitions in a radio frequency atmospheric pressure plasma jet, *Phys. Fluids*, 2024, **36**(8), 84119.

- 39 P. Das and P. M. Pandey, Machine learning based phase prediction and powder metallurgy assisted experimental validation of medium entropy compositionally complex alloys, *Modell. Simul. Mater. Sci. Eng.*, 2023, **31**(8), 085015.
- 40 S. P. Padhy, S. Saurabh, K. Choudhary, R. Kiran and N. Nguyen-Thanh, Automated regression workflow for interpretable deflection prediction in bio-inspired laminated composite plates, *Front. Struct. Civ. Eng.*, 2025, **19**(10), 1651–1668.
- 41 C. Zhang, Y. Zhang, X. Shi, G. Almpandis, G. Fan and X. Shen, On Incremental Learning for Gradient Boosting Decision Trees, *Neural Process. Lett.*, 2019, **50**(1), 957–987.
- 42 T. Chen and C. Guestrin, XGBoost: A Scalable Tree Boosting System, *Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discov. Data Min.*, Association for Computing Machinery, New York, NY, USA, 2016, pp. 785–794.
- 43 P. Akbari, M. Zamani and A. Mostafaei, Machine learning prediction of mechanical properties in metal additive manufacturing, *Addit. Manuf.*, 2024, **91**, 104320.
- 44 R. Zaplotnik, G. Primc and A. Vesel, Optical emission spectroscopy as a diagnostic tool for characterization of atmospheric plasma jets, *Appl. Sci.*, 2021, **11**(5), 1–22.
- 45 S. M. Lundberg and S.-I. Lee, A unified approach to interpreting model predictions, *Advances in Neural Information Processing Systems 30 (NIPS 2017)*, 2017, 1705.07874.
- 46 S. P. Padhy, K. P. Davidson, L. P. Tan, V. B. Varma, V. K. Sharma, X. Tan, Y. Wei, X. Xu, K. Hippalgaonkar, M. H. Jhon and R. V. Ramanujan, Integrated design framework for titanium aluminides through interpretable machine learning, *J. Alloys Compd.*, 2025, **1047**, 184937.
- 47 P. Das and P. M. Pandey, Multi-objective optimization and machine learning assisted design and synthesis of magnesium based novel non-equiatomc medium entropy alloy, *J. Alloys Compd.*, 2024, **985**, 174066.
- 48 S. P. Padhy, V. Chaudhary, Y. F. Lim, R. Zhu, M. Thway, K. Hippalgaonkar and R. V. Ramanujan, Experimentally validated inverse design of multi-property Fe-Co-Ni alloys, *iScience*, 2024, **27**(5), 109723.