

Leakage Prediction and Geometric Parameter Optimization of Stepped Labyrinth Seals Based on Data-Driven Surrogate Models

Xiaoling Li[1], Lixin Yang[1], Yuhang Wei[1]

[1] China, Beijing, Beijing Jiaotong University 1

Abstract

Leakage rate prediction and geometric parameter optimization are critical for the performance and reliability of stepped labyrinth seals in aero-engines and rotating machinery. These seals suppress leakage by creating successive throttling stages that dissipate kinetic energy; however, their leakage performance is governed by the highly nonlinear interaction of multiple geometric parameters, making accurate analytical prediction infeasible and direct CFD-driven optimization prohibitively expensive for high-dimensional design spaces. To overcome these challenges, this study proposes an integrated data-driven framework that combines computational fluid dynamics (CFD) simulation, surrogate modeling, physics-informed hybrid bias model, and genetic optimization for efficient leakage prediction and geometric design of stepped labyrinth seals. First, a CFD-based dataset of 4500 samples is constructed using a full-factorial design. Three surrogate models—a Multilayer Perceptron (MLP), eXtreme Gradient Boosting (XGBoost), and a Radial Basis Function (RBF) model—are then developed, and their hyperparameters are optimized using Bayesian optimization. To further improve prediction accuracy, a hybrid modeling strategy is introduced in which the surrogate model predicts the bias between CFD leakage and isentropic ideal leakage, thereby compensating for systematic bias in direct prediction. The MLP-based hybrid model achieves the best performance, with an R^2 of 0.9954 and Root Mean Squared Percentage Error (RMSPE) of 1.38% on the test set. To better understand the effects of design variables, SHapley Additive exPlanations (SHAP) and Sobol sensitivity analysis are subsequently employed to quantify parameter influence from complementary perspectives: SHAP analysis identifies tooth number N as the most influential structural parameter in surrogate prediction, while Sobol analysis reveals that forward expansion angle α and backward expansion angle β dominate the leakage variation across the design space. A Genetic Algorithm (GA) is finally applied based on the hybrid MLP surrogate to perform constrained geometric optimization, substantially reducing the computational cost relative to direct CFD-driven optimization. The proposed framework provides an accurate, interpretable, and computationally efficient tool for leakage prediction and geometry optimization of stepped labyrinth seals, offering practical support for the design of aero-engine sealing systems.

Keywords: Labyrinth seals; Surrogate model; CFD; Bayesian optimization; Genetic algorithm

1. Introduction

Sealing technology plays a vital role in reducing internal leakage and enhancing the efficiency, reliability, and safety of rotating machinery such as turbines and compressors [1–2]. Among non-contact sealing solutions, labyrinth seals are particularly effective at restricting leakage flow through a series of throttling cavities. Stepped labyrinth seals, a variant incorporating rotor-side teeth and stator-side steps, offer superior throttling capability and enhanced flow dissipation compared with conventional straight-through designs. However, their leakage performance is governed by the coupled influence of multiple geometric parameters—including tooth height, tooth pitch, step height, cavity angles, and tooth count—each of which interacts nonlinearly with the flow field. Identifying an optimal geometric configuration that minimizes leakage while satisfying structural and manufacturing constraints therefore remains a significant engineering challenge.

Traditionally, the performance evaluation and structural design of labyrinth seals have relied on experimental testing and CFD simulations [3–4]. While experimental methods provide high-fidelity measurements, they are both costly and time-intensive, especially when extensive multi-parameter sweeps are required. CFD offers detailed flow-field insights at a fraction of the experimental cost; nevertheless, a single high-resolution simulation can still require hours of computation, rendering direct CFD-driven optimization impractical for problems involving large design spaces or iterative search procedures.

To overcome these limitations, data-driven surrogate models have increasingly been adopted for performance prediction of labyrinth seals [5–10]. Machine learning approaches—including Multilayer Perceptron (MLP), eXtreme Gradient Boosting (XGBoost), and Radial Basis Function (RBF) models—have demonstrated strong capability in approximating complex nonlinear mappings while substantially reducing evaluation cost. These surrogates enable rapid screening of large design populations and thus provide an efficient engine for engineering optimization. Nevertheless, most existing surrogate-based studies adopt a purely data-driven paradigm, which inherently suffers from limited physical interpretability and degraded reliability in data-sparse or extrapolation regimes. To mitigate the black-box nature of standalone machine learning models, hybrid frameworks that embed physics-based prior knowledge into the data-driven bias process have been proposed and validated in several fluid machinery applications [11]. However, the application of such physics-augmented surrogate strategies to labyrinth seal leakage prediction remains largely unexplored, leaving considerable scope to enhance both the physical consistency and the generalization robustness of data-driven seal models.

To address the above issues, this study develops a hybrid methodology that combines CFD-generated data, machine-learning-based surrogate modeling, bias prediction, and optimization algorithms. Full-factorial design sampling is utilized to construct a geometric database of stepped labyrinth seals, and three surrogate models are trained and hyperparameter-optimized using Bayesian methods. Furthermore, a hybrid prediction framework is developed to improve accuracy through correction of the deviation between ideal theoretical leakage and CFD results. Key design parameters are analyzed using SHapley Additive exPlanations (SHAP) interpretability techniques, while Genetic Algorithm (GA) is adopted to search for the optimal seal configuration. Through CFD validation, the reliability of the proposed framework is demonstrated, addressing computational bottlenecks. This work provides an efficient and robust methodology for the performance prediction and design optimization of stepped labyrinth seals.

2. Numerical method

2.1 Geometric Modeling and Parameter Definition

The stepped labyrinth seal considered in this study consists of rotor-side teeth and stator-side steps, which together form multiple sealing cavities. Six geometric design variables are considered: tooth height (H, *mm*), tooth pitch (B *mm*), step height (e, *mm*), forward expansion angle (α, *deg*), backward

expansion angle (*β*, *deg*), and the number of teeth (*N*). In addition, the seal clearance *c* is fixed at 0.2 mm and the tooth tip width *t* is set to 0.3 mm, both of which are treated as geometric constants. The step length *L* is defined as $L = x \cdot B$, where the proportionality factor *x* is set to 0.6, making *L* a derived parameter that scales proportionally with the tooth pitch. The variation ranges of these parameters are determined according to engineering constraints and practical manufacturability, as summarized in Table 1. The computational domain comprises the inlet settling section, sealing cavities, and outlet section.

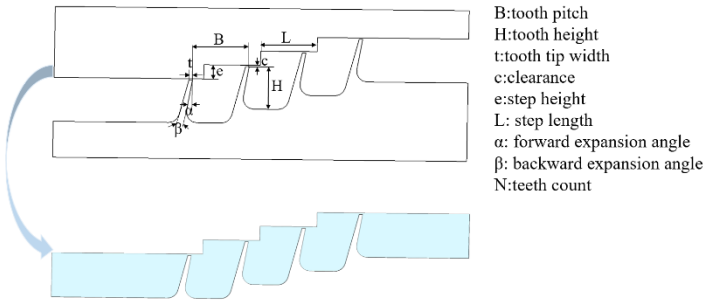


Figure 2 –Typical configuration of a stepped labyrinth seal

Table 1 –Geometric parameters of the stepped labyrinth seal and their variation ranges

parameter	symbol	range	unit
tooth height	*H*	4 ~ 8	*mm*
tooth pitch	*B*	6 ~ 10	*mm*
step height	*e*	0 ~ 2	*mm*
forward expansion angle	*α*	-30 ~ 30	*deg*
backward expansion angle	*β*	0 ~ 30	*deg*
teeth count	*N*	2 ~ 6	—

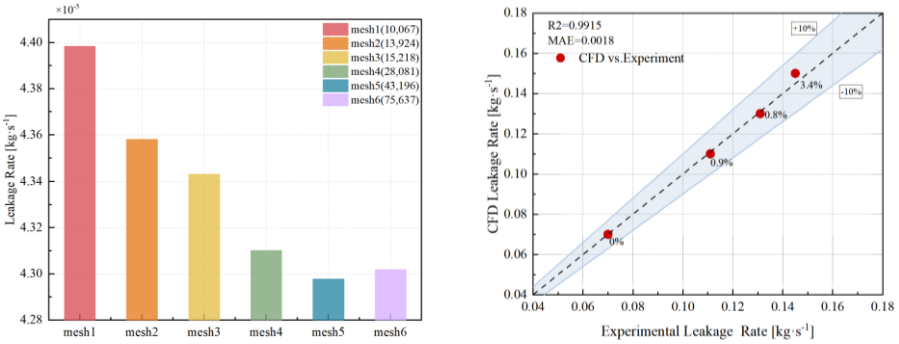
2.2 CFD model

ANSYS CFX is utilized to solve the steady Reynolds-Averaged Navier–Stokes (RANS) equations. The Realizable *k-ε* turbulence model with enhanced wall treatment is adopted to capture the complex flow features inside the seal cavities, including swirl, recirculation, and local separation. Structured quadrilateral meshes are generated, and a grid independence study is performed to ensure numerical reliability.

The boundary conditions are specified as follows: total pressure at the inlet (P_{in}=4 atm) and total temperature (T_{in}=300 K), static pressure at the outlet (P_{out}=1 atm), and no-slip adiabatic wall boundaries. The working fluid is ideal-gas air. Convergence is considered achieved when all residuals dropped below 10^{-5}.

2.3 Numerical model validation

To ensure grid independence, six mesh densities are tested. Mesh 5 is ultimately selected because its predicted leakage rate differs only marginally from that obtained using a finer mesh, while maintaining a reasonable computational cost. To further verify the accuracy of the CFD model, the numerical results are compared with the experimental data. As shown in Figure 2, the CFD predictions agree well with the experimental measurements under different operating conditions. The average deviation over the investigated range is approximately 1.28%, indicating that the CFD model has sufficient accuracy for generating the dataset required for surrogate-model development.



(a) grid independence (b) experimental verification

Figure 2 – Grid independence and experimental verification

3. Surrogate Model Construction

3.1 Dataset Generation and Partition

A full factorial design is employed to systematically explore the six-dimensional parameter space, in which all possible combinations of the specified factor levels are considered. Each geometric parameter is divided into several discrete levels within its respective variation range, yielding a total of 4500 sample points that comprehensively cover the design space. Each parameter combination is taken as input, and the steady leakage rate obtained from CFD simulations is used as output to construct the sample database.

The complete dataset is randomly shuffled and divided into training, validation, and test sets in proportions of 60%, 20%, and 20%, respectively. To eliminate the effect of dimensional inconsistency among the input variables, the input features are standardized as follows:

$$Z = \frac{(x - \mu)}{\sigma} \tag{1}$$

where x represents an individual data point from the raw dataset of each input feature. μ denotes the arithmetic mean of each input feature, while σ represents the standard deviation of each input feature.

3.2 Surrogate Model Structures

Three ML-based surrogate models are developed and compared in this work: MLP, XGBoost and RBF. All models are implemented in Python 3.12 using the scikit-learn framework and related libraries. Their structures and training strategies are tailored to the characteristics of each modeling approach.

The MLP model is configured as a deep feedforward neural network composed of fully connected layers. The six standardized geometric parameters are fed into the input layer and transformed through multiple hidden layers to predict the scalar leakage rate. LeakyReLU is used as the activation function in the hidden layers to introduce nonlinearity while alleviating the dying-neuron issue associated with standard ReLU. The loss function employed is the Mean Squared Error (MSE), and the Adam optimizer is used to minimize this loss via adaptive moment estimation. To improve generalization and suppress overfitting, early stopping and learning-rate decay are introduced during training. The final network architecture is determined through Bayesian hyperparameter optimization.

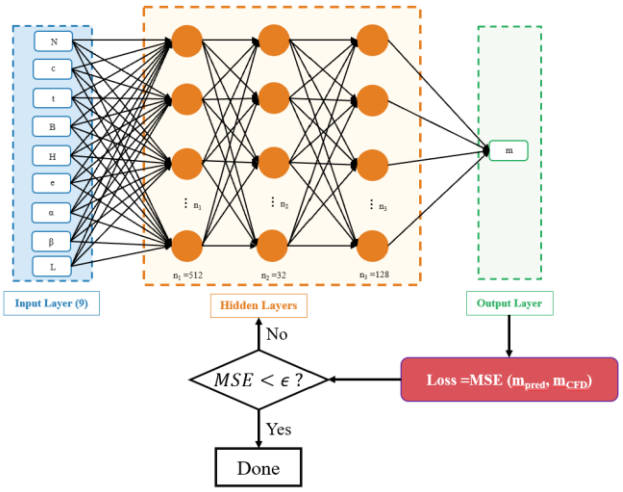


Figure 3 – Architecture of MLP model

The XGBoost model is an ensemble learning algorithm based on gradient-boosted decision trees. It constructs regression trees sequentially, with each new tree fitted to the residual errors of the current ensemble. By using the second-order Taylor expansion of the loss function, the model achieves efficient and accurate optimization. Regularization terms based on L1 and L2 penalties are embedded in the objective function to control model complexity and improve robustness. The final prediction is obtained by summing the weighted outputs of all trees.

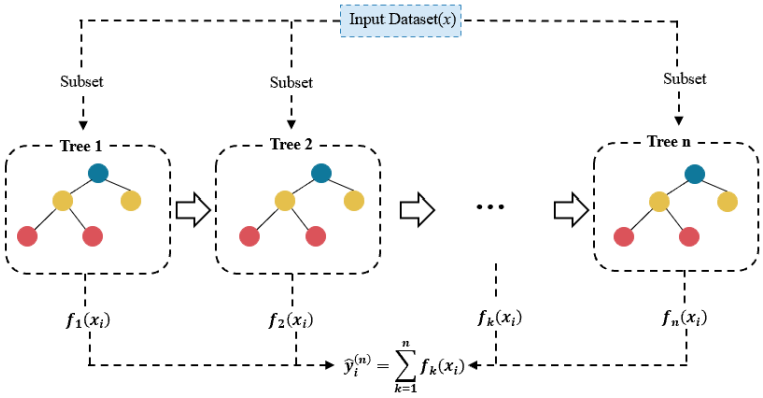


Figure 4 –Architecture of XGBoost model

The RBF model is formulated using a Gaussian-kernel-based Support Vector Regression (SVR) framework. The RBF network is a kernel-based surrogate model that approximates the target function through a weighted sum of radial basis functions centered at the training samples. In this formulation, the target function is approximated by a weighted combination of radial basis functions. The kernel parameter γ controls the locality of each basis function, while the insensitive loss width ε defines the tolerance band within which prediction errors are not penalized. The regularization parameter C governs the trade-off between model complexity and fitting accuracy. This structure offers good smoothness and robustness when modeling continuous leakage variations in the geometric design space.

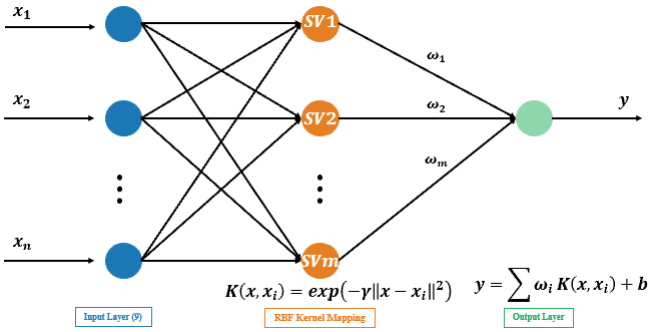


Figure 5 –Architecture of RBF model

Bayesian optimization is employed for hyperparameter tuning across all three surrogate models. This method constructs a probabilistic surrogate model of the objective function and iteratively refines it by incorporating observed evaluation results. In this study, the average Root Mean Squared Error (RMSE) from 10-fold cross-validation on the training set serves as the objective function for each hyperparameter configuration. A Gaussian process is utilized as the surrogate model to approximate the objective function landscape, while the Tree-structured Parzen Estimator (TPE) algorithm is adopted as the acquisition function to identify the most promising hyperparameter combination in each successive iteration. The maximum number of optimization iterations is set to 200. Detailed search ranges and the resulting optimal hyperparameter values for each model are summarized in Table 2.

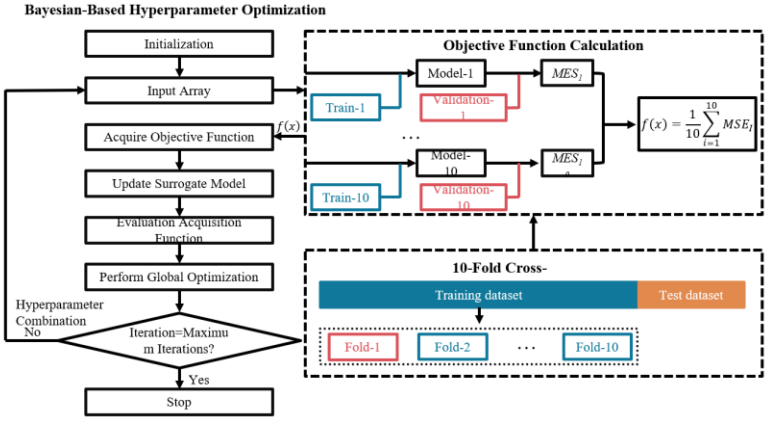
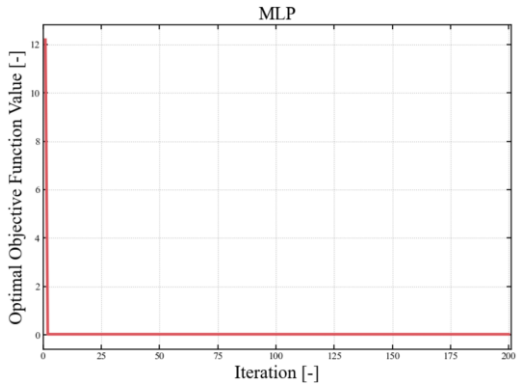
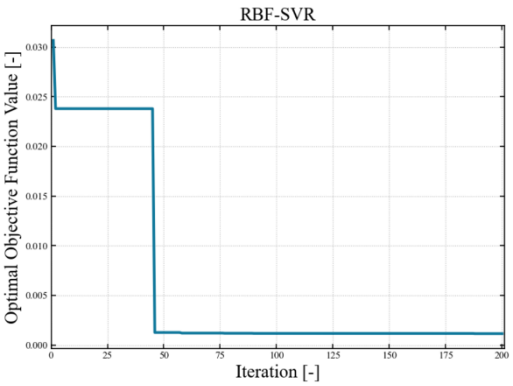


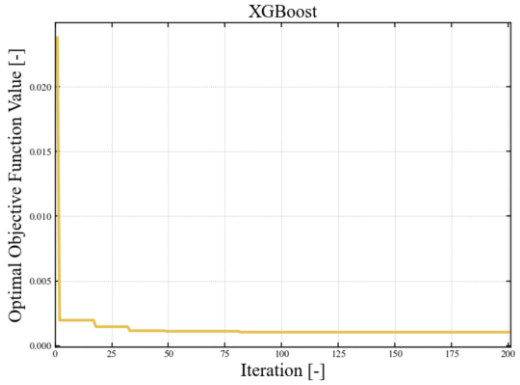
Figure 6 –Bayesian-based hyperparameter optimization flowchart



(a) Bayesian optimization for MLP



(b) Bayesian optimization for RBF



(c) Bayesian optimization for XGBoost

Figure 7 –Performance of Bayesian optimization over iterations

Table 2 Bayesian-based hyperparameter optimization: Search space and search results

Model	Search Range	Hyperparameter	Best Value
MLP	[0.0001, 0.01]	Learning rate	0.01
	[200, 2000]	Max epochs	2000
	[16, 128]	Batch size	128
	[[64, 64], [128, 64], [256, 128], [512, 32, 128], [256, 64, 128]]	Hidden dims	[512, 32, 128]
	[0, 0.3]	Dropout rate	0.01
	['ReLU', 'LeakyReLU', 'ELU']	Activation	LeakyReLU
XGBoost	[50, 600]	N_estimators	50
	[2, 12]	Max depth	12
	[0.01, 0.3]	Learning rate	0.0816
	[0.5, 1]	Colsample_bytree	1
	[0, 1]	Reg_alpha	0
	[0.5, 5]	Reg_lambda	5
	[0.5, 1]	Subsample	0.5
RBF	[0.1, 1000]	C	100
	[0.0001, 10]	Gamma	0.2326

[0.001, 0.5]	Epsilon	0.3861

To further improve predictive performance beyond that of the standalone surrogate models, a hybrid surrogate framework is proposed by introducing a physics-informed correction strategy based on isentropic leakage theory. Rather than directly predicting the absolute leakage rate, the MLP is trained to estimate a dimensionless bias $\zeta = m_{CFD}/ m_{ideal}$, which represents the deviation between CFD simulated leakage and the analytically derived isentropic ideal leakage. During prediction, the hybrid model first calculates the ideal leakage from inlet and outlet pressures and then multiplies it by the predicted correction factor to obtain the final leakage estimate. By embedding this physically meaningful reference into the surrogate framework, the model can more effectively capture deviations caused by viscous dissipation, geometric complexity, and flow separation. As a result, both interpolation accuracy and extrapolation robustness are enhanced compared with purely data-driven prediction.

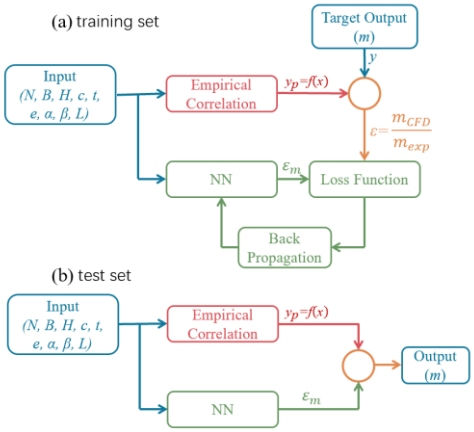


Figure 8 –Architecture of hybrid surrogate model

4. Results and Discussion

4.1 Surrogate Model Performance

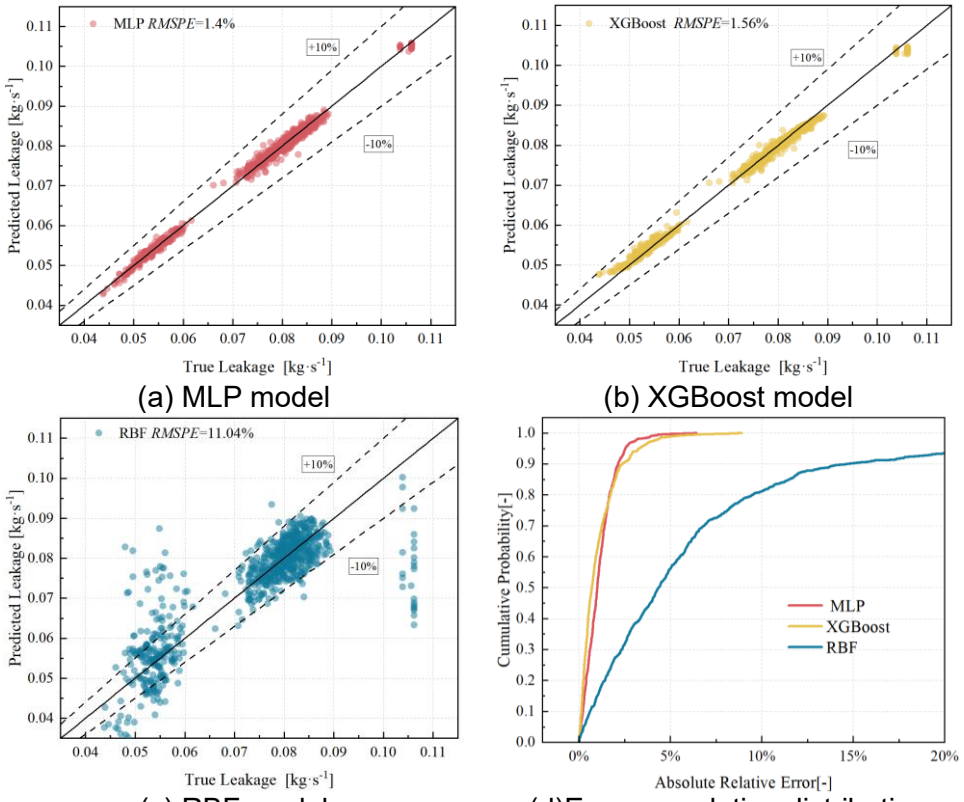
Figure 9 compares scatter plots comparing the predictions of the three surrogate models—MLP, XGBoost and RBF—against the ground-truth CFD-simulated leakage rates on the test set. The MLP and XGBoost models both perform well, each attaining an R^2 of 0.994, indicating that each model successfully captures the nonlinear mapping between seal geometry and leakage rate. By contrast, the RBF model exhibits substantially degraded performance on the test set, with $R^2 = 0.692$.

The MLP model achieves the lowest RMSE among the standalone models, reflecting its capacity to represent complex high-dimensional relationships through multi-layer feature abstraction. XGBoost also delivers competitive performance, especially in the mid-range of the leakage distribution, although slightly larger deviations are observed near the extremes where training samples are relatively sparse. The RBF model, while computationally efficient, shows somewhat wider scatter for samples located away from the training data centroids, indicating a limited generalization capability in the extrapolation regime. The mathematical formulations for these metrics are as follows:

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(\varphi_{i,cal} - \varphi_{i,exp}\right)^2} \tag{2}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}\left(\varphi_{i,exp} - \varphi_{i,cal}\right)^2}{\sum_{i=1}^{n}\left(\varphi_{i,cal} - \bar{\varphi}_{i,cal}\right)^2} \tag{3}$$

The parity plots and cumulative error distribution curves further confirm these observations. The MLP predictions are more tightly clustered around the ideal parity line, whereas XGBoost and RBF display slightly larger dispersion. These results indicate that although all models are effective for leakage prediction, the MLP provides the best balance between accuracy and stability.



(a) MLP model (b) XGBoost model

(c) RBF model (d)Error cumulative distribution

Figure 9 –Measured vs. predicted parity plot Error cumulative distribution function plot

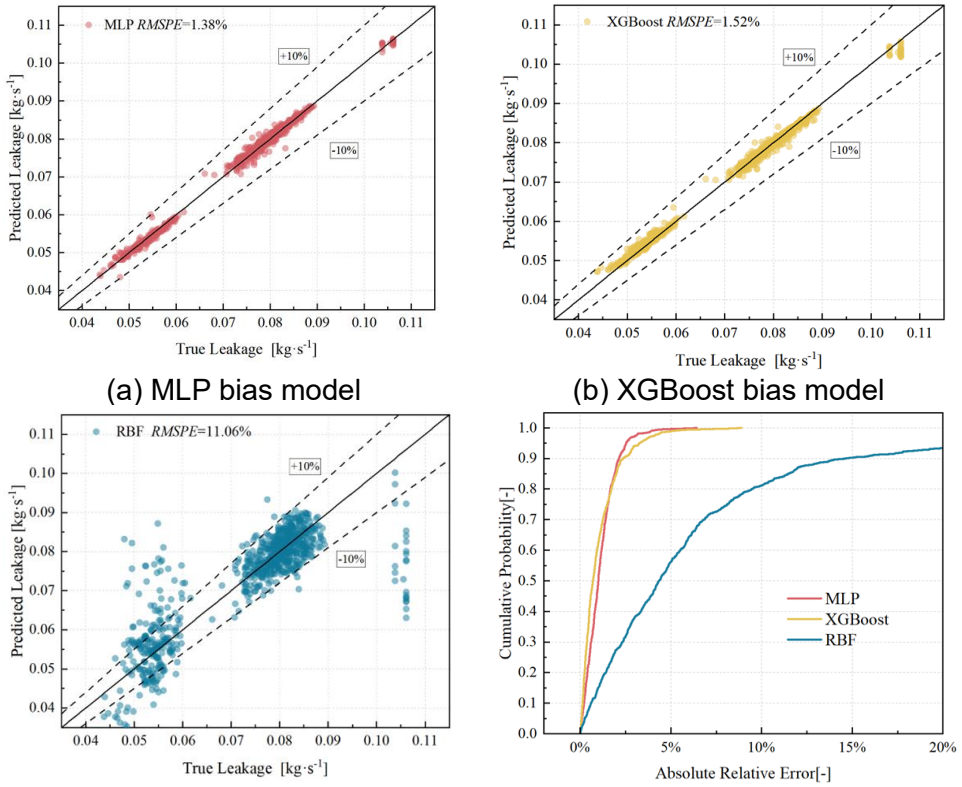
4.2 Performance Analysis of the Hybrid Model

To improve prediction accuracy further, the hybrid surrogate model based on isentropic leakage correction is evaluated against the standalone models. The corresponding parity plot and cumulative error distribution demonstrate that the hybrid model yields the closest agreement with CFD results. Compared with the direct-prediction models, the hybrid approach significantly reduces systematic deviation and improves predictive consistency across the full leakage range.

Among all examined models, the Hybrid MLP demonstrates the best overall predictive capability, achieving an R^2 of 0.995 and Root Mean Squared Percentage Error (RMSPE) of 1.38%, with 99.25% of predictions falling within ±5% of the CFD results. The Hybrid XGBoost model likewise exhibits highly competitive performance, yielding an R^2 of 0.995 and an RMSPE of 1.52%. Notably, it attains 100% coverage within ±10% error, indicating consistently bounded prediction deviations across the entire test set. In contrast, the Hybrid RBF model shows little to no improvement over its standalone counterpart, with an R^2 of only 0.6905 and an RMSPE of 11.06%. This limited performance suggests that the kernel-based interpolation mechanism of the RBF model lacks sufficient representational flexibility to fully exploit the physics-informed correction introduced in the hybrid framework.

The results confirm that combining data-driven learning with simplified physical knowledge is a practical

way to improve the predictive fidelity of surrogate models for complex flow systems. This hybrid structure therefore serves as the final predictive model for subsequent interpretability analysis and geometric optimization.



(a) MLP bias model (b) XGBoost bias model

(c) RBF bias model (d)Error cumulative distribution

Figure 10 –Measured vs. predicted parity plot Error cumulative distribution function plot

4.3 Parameter Importance Analysis

To quantitatively assess the influence of the geometric parameters on leakage prediction, SHAP is employed to interpret the trained hybrid surrogate model. Rooted in cooperative game theory, SHAP provides a unified and consistent measure of the contribution of each feature to the final prediction.

As illustrated in Figure 11, the number of teeth *N*, emerges as the most influential parameter governing leakage behavior, with a mean SHAP value of 0.5999, far exceeding that of the other geometric variables. The forward expansion angle *α*, and backward expansion angle *β*, constitute the second most significant group of factors. By comparison, the remaining parameters exhibit only limited individual influence on the predicted leakage rate. In particular, the seal clearance *c*, and tooth tip width *t*, make almost negligible contributions, which is consistent with their treatment as fixed geometric constants in the present design framework.

This feature ranking carries clear physical significance. The predominance of *N* reflects the cumulative throttling effect introduced by successive teeth: each additional tooth creates an extra stage of flow restriction and energy dissipation, thereby markedly suppressing the overall leakage. The strong sensitivity to *α* and *β* can be attributed to their direct regulation of the local flow incidence and expansion characteristics within each cavity, which in turn strongly influence vortex generation, flow separation, and kinetic energy dissipation. In contrast, the relatively modest importance of *H* and *B* indicates that, within the investigated parameter ranges, variations in tooth height and tooth pitch induce only minor changes in effective flow restriction when compared with the more dominant effects of tooth count and cavity-shaping angles. Beyond enhancing confidence in the predictive fidelity of the surrogate model, this interpretability analysis also provides physically meaningful and practically actionable guidance for subsequent seal design optimization.

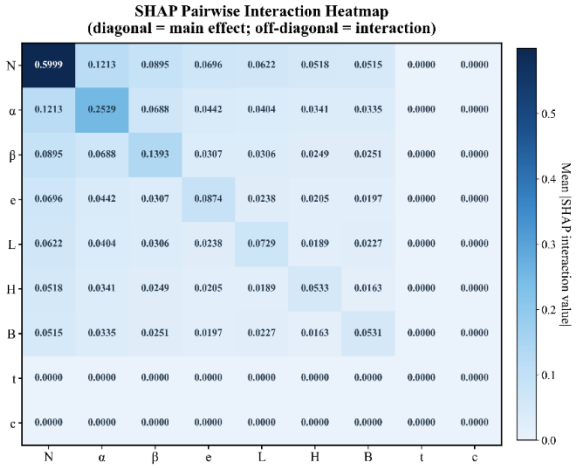


Figure 11 –SHAP Pairwise Interaction Heatmap

5. Geometric Parameter Optimization Based on Genetic Algorithm

5.1 Optimization Problem Formulation

Using the hybrid surrogate model as the fitness evaluation function, a single-objective optimization problem is formulated to minimize leakage rate: m_{min} (H,B,e,α,β,N) subject to all geometric parameters remaining within the ranges listed in Table 1.To verify the reliability of the optimal solution, the optimal geometric configuration obtained by GA is re-evaluated using CFD and compared with the surrogate model prediction to confirm its actual physical performance.

5.2 Optimization Results and Analysis

Figure 12 illustrates the convergence history of the GA, showing the evolution of the best-individual fitness and the mean population fitness over successive generations. During the first approximately 30 generations, the best fitness value drops sharply as the population explores the design space and eliminates clearly suboptimal configurations. Beyond generation 50, the best fitness curve gradually levels off, indicating that the algorithm has entered the exploitation phase and converged to the vicinity of the global optimum. The mean population fitness approaches the best-individual value at approximately generation 80, confirming that genetic diversity has been successfully maintained throughout the search and that premature convergence has been avoided. The use of the hybrid surrogate model as the fitness function is critical here, as it reduces the evaluation cost per generation from hours (CFD) to milliseconds (surrogate), making the evolutionary search computationally feasible.

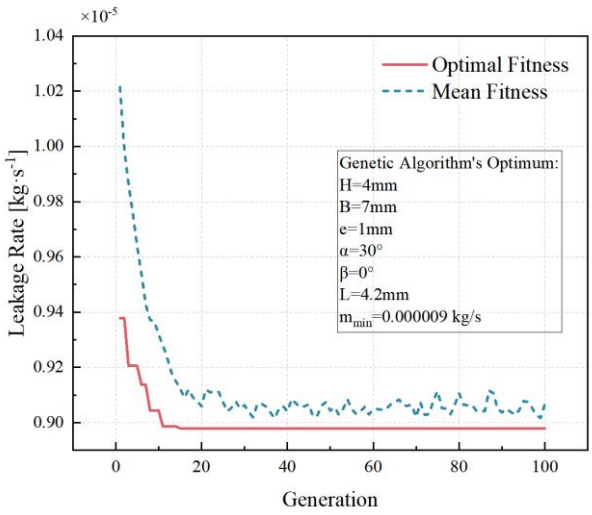


Figure 12 –Convergence curve of genetic algorithm

To further understand the influence of each design variable on leakage variation, Sobol sensitivity analysis is performed. The first-order and total-effect Sobol indices are presented in Figure 13. The forward expansion angle *α* exhibits the highest first-order sensitivity coefficient, indicating that it is the dominant factor affecting leakage. The backward expansion angle *β* also show relatively strong sensitivity and can therefore be considered important secondary variables. By contrast, step height *e*, tooth height *H* and tooth pitch *B* have weaker first-order effects.
It is noteworthy that the total-effect sensitivity coefficients of all parameters are slightly larger than their corresponding first-order values, indicating that interaction effects among variables do exist but are relatively limited in magnitude. This finding suggests that the geometric parameters largely influence leakage through independent mechanisms, and optimizing them individually would capture the majority of the achievable performance improvement.

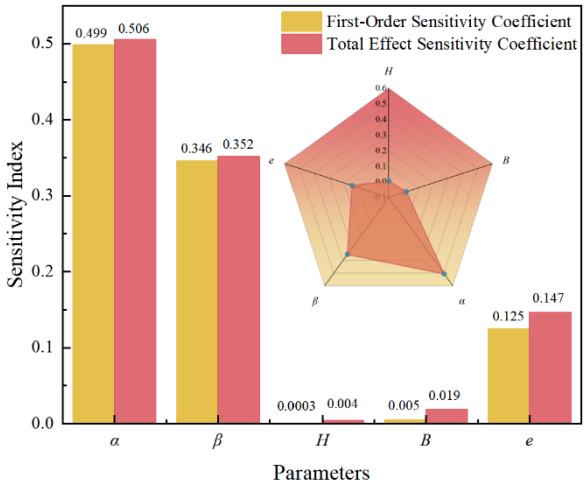


Figure 13 –First-order sensitivity coefficients and total effect sensitivity coefficients

SHAP reveals which parameters the surrogate model *relies on* most heavily for prediction, while Sobol analysis reveals which parameters *drive the most variation* in leakage rate across the design space. Together, they suggest that *N* is a structurally critical parameter whose categorical changes produce substantial shifts in flow physics, while α and *β* govern a large fraction of the continuous leakage variation across the design space. Both perspectives are valuable for design optimization: *N* should be prioritized as a primary design decision due to its dominant physical effect, while *α* and *β* offer the most effective continuous tuning knobs for fine-grained leakage control.

5.3 Second-Order Interaction Effect Analysis of Geometric Parameters

The second-order Sobol sensitivity coefficients and Spearman rank correlation coefficients are presented in Figure 14, providing complementary insights into pairwise interactions and monotonic relationships between geometric variables and leakage rate. The second-order sensitivity analysis reveals that the interactions between tooth height *H* and tooth pitch *B*, exhibit the largest second-order sensitivity coefficients. This indicates that while certain parameter pairs—particularly *B* and *H*—exhibit non-negligible coupling, the dominant parameters *α* and *β* show relatively limited interactions, consistent with the small differences observed between their first-order and total-effect sensitivity coefficients.
The Spearman rank correlation coefficients reveal that the forward expansion angle α exhibits the strongest negative correlation with leakage rate, indicating that increasing *α* consistently reduces leakage across the design space. In contrast, the backward expansion angle *β* exhibits the highest positive correlation with leakage, confirming that larger *β* is directly associated with increased leakage flow. Among the remaining parameters, tooth height H and step height e exhibit moderate positive

correlations with leakage rate ($\rho \approx 0.32$ and 0.28, respectively), while tooth pitch B shows only a weak correlation ($\rho \approx 0.13$), broadly consistent with their lower first-order sensitivity rankings.

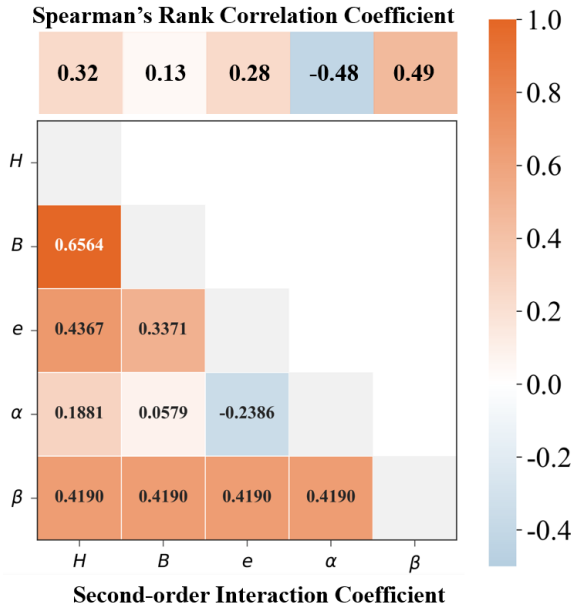


Figure 14 –Second-order sensitivity coefficients and correlation coefficients

6. Conclusions

A data-driven framework integrating CFD, surrogate modeling, physics-informed correction, interpretability analysis, and genetic optimization was developed in this study for leakage prediction and geometric parameter optimization of stepped labyrinth seals. The main conclusions are summarized as follows:

1) Based on a 4500-sample CFD dataset, three surrogate models—MLP, XGBoost, and RBF—were established and optimized using Bayesian hyperparameter tuning. Among the standalone models, the MLP achieved the best predictive accuracy, whereas the RBF model showed clearly inferior generalization capability.
2) A hybrid prediction strategy was introduced by modeling the bias between CFD leakage and isentropic ideal leakage, rather than directly predicting the leakage rate. This physics-informed formulation effectively reduced systematic prediction bias and improved predictive robustness. Among all candidate models, the hybrid MLP provided the best overall performance on the test set.
3) SHAP analysis indicated that the tooth number N is the most influential feature in the surrogate prediction process, while the forward expansion angle α and backward expansion angle β also exert substantial effects. Sobol sensitivity analysis further showed that α and β dominate the leakage variation within the investigated design space. These findings suggest that N is a key structural design variable, whereas α and β are the most effective continuous parameters for leakage regulation.
4) Correlation and interaction analyses showed that α is negatively correlated with leakage rate, whereas β is positively correlated. Overall, interaction effects among design variables are limited, although the coupling between tooth height H and tooth pitch B is relatively more evident than that of other parameter pairs.
5) By coupling the hybrid MLP surrogate with a GA, the geometric optimization problem was solved efficiently under the prescribed constraints. The proposed framework reduces the computational cost of design evaluation substantially compared with direct CFD-based optimization and provides a practical approach for the rapid design of stepped labyrinth seals.

Overall, the present study demonstrates that the combination of data-driven surrogate modeling and physics-informed correction offers an accurate and efficient route for leakage prediction and geometry

optimization of stepped labyrinth seals, with potential value for aero-engine sealing system design.

References

[1] Park G, Hur M S, Kim T S. Optimal design and performance prediction of stepped honeycomb labyrinth seal using CFD and ANN[J]. *Engineering Science and Technology*, an International Journal, 2025, 61: 101939.

[2] Liu H, Li G, Kang C, et al. Enhancing energy efficiency in aero-engines through multi-fidelity optimization of labyrinth seal design[J]. *Energy*, 2025, 335: 138142.

[3] Qin H, Lu D, Zhong D, et al. Experimental and numerical investigation for the geometrical parameters effect on the labyrinth-seal flow characteristics of fast reactor fuel assembly[J]. *Annals of Nuclear Energy*, 2020, 135: 106964.

[4] Čížek M, Pátek Z, Vampola T. Aircraft turbine engine labyrinth seal CFD sensitivity analysis[J]. *Applied Sciences*, 2020, 10(19): 6830.

[5] Liu H, Li G, Kang C, et al. Optimization of smooth straight-through labyrinth seal based on XGBoost and improved genetic algorithm[J]. *Journal of Engineering for Gas Turbines and Power*, 2025, 147(3): 031007.

[6] Cremanns K, Roos D, Hecker S, et al. Efficient multi-objective optimization of labyrinth seal leakage in steam turbines based on hybrid surrogate models[C]//Turbo Expo: *Power for Land, Sea, and Air. American Society of Mechanical Engineers*, 2016, 49712.

[7] Zhao Y, Wang C. Shape optimization of labyrinth seals to improve sealing performance[J]. *Aerospace*, 2021, 8(4): 92.

[8] Liu C, Amei B, Yi Z, et al. Surrogate-based optimization and experiment validation of a fan-shaped film cooling hole with a large lateral space[J]. *Applied Thermal Engineering*, 2022, 207: 118145.

[9] Li Z, Guan X, Zhang J, et al. Bayesian optimization of operational and geometric parameters of microchannels for targeted droplet generation[J]. *Chinese Journal of Chemical Engineering*, 2025.

[10] Cheng H, Wen Z, Zhao Y, et al. Effect and optimization of geometric parameters and arrangement on film cooling performance of fan-shaped holes based on generalized regression neural network[J]. *International Communications in Heat and Mass Transfer*, 2024, 158: 107868.

[11] Schweidtmann A M, Zhang D, Von Stosch M. A review and perspective on hybrid modeling methodologies[J]. *Digital Chemical Engineering*, 2024, 10: 100136.

Copyright Issues

The copyright statement is included in the template and must appear in your final pdf document in the position, style and font size shown below. If you do not include this in your paper, ICAS is not allowed and will not publish it.

Archiving

The ICAS 2026 proceedings will receive an ISSN number and all papers will get individual DOIs (unless you decline that option).

Contact Author Email Address

The contact author email address should appear explicitly in this section to facilitate future contacts. For example, LiXin Yang: lxyang1@bjtu.edu.cn.

Copyright Statement

The authors confirm that they, and/or their company or organization, hold copyright on all of the original material included in this paper. The authors also confirm that they have obtained permission, from the copyright holder of any third party material included in this paper, to publish it as part of their paper. The authors confirm that they give permission, or have obtained permission from the copyright holder of this paper, for the publication and distribution of this paper as part of the ICAS proceedings or as individual off-prints from the proceedings.